Edited by Ian Houson

Process Understanding

For Scale-Up and Manufacture of Active Ingredients

Edited by
Ian Houson

Process Understanding

Related Titles

Helmus, F. P.

Process Plant Design
Project Management from Inquiry to Acceptance

2008
ISBN: 978-3-527-31313-6

Bröckel, U., Meier, W., Wagner, G. (eds.)

Product Design and Engineering
Best Practices

2007
ISBN: 978-3-527-31529-1

Rao, S.

The Chemistry of Process Development in Fine Chemical and Pharmaceutical Industry

2007
ISBN: 978-0-470-31995-6

Zlokarnik, M.

Scale-up in Chemical Engineering

2006
ISBN: 978-3-527-31421-8

Vogel, G. H.

Process Development
From the Initial Idea to the Chemical Production Plant

2005
ISBN: 978-3-527-31089-0

Wiley-VCH (ed.)

Ullmann's Processes and Process Engineering

2004
ISBN: 978-3-527-31096-8

Koolen, J. L. A.

Design of Simple and Robust Process Plants

2001
ISBN: 978-3-527-29784-9

Edited by Ian Houson

Process Understanding

For Scale-Up and Manufacture of Active Ingredients

The Editor

Dr. Ian Houson
Giltech Limited
12 North Harbour Estate
Ayr KA8 8BN
United Kingdom

Contents

Preface

When I was asked to be the editor for a book on Process Understanding, I was delighted as it provided me with an opportunity to cover something that I have found challenging throughout my career as an industrial process development chemist. During my doctoral studies, I had specialized in one discipline and was encouraged to work very much on my own. However, when I started working in industry, I was suddenly being asked to work with a whole range of people and disciplines, often with no detailed knowledge of what they did. Then, as I gained experience, I learned that the other disciplines with whom I worked often have information that can be really helpful to me in the work I did (occasionally, I even had useful information for them!).

Even after 15+ years of working in active ingredients development and manufacture, I am still learning about what is important to other disciplines and how aspects of their work can really help me in the work I do. This book is a continuation of that learning and is intended to be relevant to both people who start new and experienced process technologists.

This book is not designed to be a detailed technical treatise on each of the subject areas, but to provide a valuable introduction to a range of subject areas that are vital to the successful development and manufacture of active chemical ingredients. The reader will be introduced to the areas that must be understood throughout the active ingredients lifecycle right from the route selection through to established manufacture. This book should help the reader understand what is important to other/all disciplines involved in the lifecycle, leading to improved interdisciplinary working, smoother technical transfer between disciplines, and more efficient process development and manufacture.

Process understanding is the underpinning knowledge that allows the manufacture of chemical entities to be carried out economically, sustainably, robustly, and to the required quality. This area has risen in importance in the last few years, particularly, with the recent impetus from the ''Quality by Design'' initiative from the US Food and Drug Administration. This move to a more science- and risk-based approach is already well entrenched in a number of fine chemicals companies and it is heartening to see fundamental scientific understanding being placed back at the core of process development and manufacture.

Many process development/scale-up books focus on specific products and tell you the story of one chemical entity. There is relatively little written about the

general principles and underlying philosophy of what information was required to underpin the decisions made. This book will seek to provide a broad view of what process understanding means to different disciplines and gives readers the opportunity to think about what is important to other people/disciplines and stages throughout the product life cycle. This book will seek to show how process understanding is, not only necessary, but can also deliver a real competitive advantage within the pharmaceuticals and fine chemicals industry.

Although the authors were chosen primarily for their technical expertise, they have also been selected to provide a balanced view owing to their geographical spread and with a mix of academic and industry, pharmaceuticals, and fine chemicals backgrounds. It is hoped that the reader will benefit from such a breadth of experience. I have tried to include both established areas for process development such as safety and scale-up of equipment as well as examining some of the more emerging topics such as Quality by Design, semi-quantitative modeling, and outsourcing (contract manufacture).

And finally, I leave with you this thought . . . .

*We know there are **known knowns;** there are things we know we know. We also know there **known unknowns;** that is to say we know there are some things we do not know. But there are also **unknown unknowns** – the ones we don't know we don't know . . . .*

12 February, 2002   *Donald Rumsfeld, The Pentagon*

The latter are the ones we should worry about and are why I agreed to be the editor of this book!

Acknowledgments

I hope that you enjoy reading this book as much as I have enjoyed putting it together!

October 2010   *Ian Houson*

List of Contributors

David J. Ager
DSM Pharmaceutical Chemicals
PMB 150 9650 Strickland
Road Suite 103
Raleigh, NC 27615
USA

John Atherton
University of Huddersfield
Department of Applied Sciences
Queensgate
Huddersfield HDI 3DH
UK

Colm Campbell
BioMarin Europe Ltd.
29–31 Earlscourt Terrace
Dublin 2
Ireland

Leroy Cronin
University of Glasgow
School of Chemistry
Joseph Black Building
University Avenue
Glasgow G12 8QQ
UK

Mark J. Dickson
Morgan Sindall
Professional Services Ltd.
20 Timothys Bridge Road
Stratford Enterprise Park
Stratford-Upon-Avon
Warwickshire CV37 9NJ
UK

Wilfried Hoffmann
Pfizer Global Research &
Development
Sandwich Laboratories
B530, IPC 533
Ramsgate Road
Sandwich CT13 9NJ
UK

Ian Houson
Giltech Limited
12 North Harbour Estate
Ayr KA8 8BN
UK

Dylan Jones
Genzyme Haverhill Operations
Technical Department
12 Rookwood Way
Haverhill
Suffolk CB9 8PB
UK

Brian Keaveny
Plant Director
Clarochem Ireland Limited
Damastown
Mulhuddart
Dublin 15
Ireland

Philip J. Kitson
University of Glasgow
School of Chemistry
Joseph Black Building
University Avenue
Glasgow G12 8QQ
UK

Vince McCurdy
Pfizer Inc.
558 Eastern Point Rd
Groton, CT 06340-5196
USA

Stephen Rowe
Chilworth Technology Ltd.
Beta House
Southampton Science Park
Southampton
Hampshire SO16 7NS
UK

Paul Sharratt
Institute of Chemical and
Engineering Sciences (ICES)
1 Pesek Road
Singapore, 627833
Singapore

Mark J. H. Simmons
University of Birmingham
School of Chemical Engineering
Edgbaston
Birmingham B15 2TT
UK

E. Hugh Stitt
Johnson Matthey
Technology Centre
PO Box 1
Billingham TS23 1LB
UK

Mark Talford
10 Fern Grove
Whitehaven
Cumbria CA28 6RB
UK

Chick C. Wilson
University of Glasgow
School of Chemistry
Joseph Black Building
University Avenue
Glasgow G12 8QQ
UK

Steve Woolley
Shasun Pharma
Solutions Limited
Dudley Lane
Dudley
Cramlington
Northumberland NE23 7QG
UK

1
Quality by Design
Vince McCurdy

1.1
History

The pharmaceutical industry has been a highly regulated industry in the past for many good reasons [1]. While pharmaceuticals have greatly improved the mortality and morbidity rates, there is still some element of risk to the patients. These risks are greatly mitigated with the delivery of medicine at the appropriate purity, potency, delivery rate, and so on. While pharmaceutical regulations have clearly protected the population from much of the needless harm such as that incurred early in the twentieth century, there has been a concern more recently that overregulation may be associated with stifling innovation that can improve pharmaceutical quality even further [2] – innovation that has the potential to greatly improve the quality, cost, and time to market new and improved medicines. The twenty-first century began with the pharmaceutical industry using manufacturing technologies that have been employed since the 1940s and did not make significant changes in manufacturing process unless significant compliance or costs saving advantages could justify the high costs and long cycle time needed to gain approval. This often resulted in inefficient, overly expensive processes that were ultimately not in the best long-term interests of patients. As a result, the FDA (Food and Drug Administration) and other agencies around the world have embraced a new paradigm for regulation [3]. The "desired state" was to shift manufacturing from being empirical to being more science, engineering, and risk based. Another regulatory guidance that had major impact was the Process Analytical Technology (PAT) Guidance [9]. The continuous, real-time monitoring of manufacturing processes is a key enabler to achieve greater process control. Finally, the current Good Manufacturing Practices (cGMPs) for the Twenty-First Century Guidance acknowledged the undesired impact of good manufacturing practices (GMPs) on understanding manufacturing science and sought to set the framework for additional guidances that encouraged risk- and science-based understanding in exchange for more freedom to introduce innovations and improvements that will result in enhanced quality, cost, or timing.

Table 1.1 Comparison of the current state to the future desired QbD state.

Aspect	Current state	Desired QbD state
Pharmaceutical development	Empirical; typically univariate experiments	Systematic; multivariate experiments
Manufacturing process	Locked down; validation on three batches; focus on reproducibility	Adjustable within design space; continuous verification within design space; focus on control strategy
Process control	In-process testing for go/no-go; offline analysis	PAT utilized for feedback and feed forward in real time
Product specification	Primary means of quality control; based on batch data	Part of overall quality control strategy; based on product performance
Control strategy	Mainly by intermediate and end product testing	Risk-based; controls shifted upstream; real-time release
Lifecycle management	Reactive to problems and OOS; postapproval changes needed	Continual improvement enabled within design space

Juran is often credited with introducing the concepts behind Quality by Design (QbD) [4]. Pharmaceutical QbD is a systematic approach to development that begins with pre-defined objectives and emphasizes product and process understanding based on sound science and quality risk management (ICH Q8R2). The holistic and systematic approach of QbD was relatively new to the pharmaceutical industry at the beginning of the twenty-first century. However, elements of QbD were certainly being applied across the industry long before then. QbD was put into practice in a big way with the advent of the FDA CMC pilot program in 2005. Nine companies participated in the program and eventually submitted regulatory filings based on a QbD framework [1, 2, 5–7]. Much was learned from these initial filings that help steer the industry and regulators toward a common vision for QbD. A comparison of the "current state" to the future "desired state" was succinctly summarized by Nasr in Table 1.1 [8].

A process is well understood when

- all the critical sources of variability are identified and explained;
- variability is managed by the process, and;
- product quality attributes can be accurately and reliably *predicted* over the *design space* established for materials used, process parameters, manufacturing, environmental, and other conditions [9].

Process understanding is the major goal of a QbD program. A complete list of characteristics of a successful QbD program is summarized in Table 1.2.

Table 1.2 The characteristics of a successful QbD program.

Involves product design and process development
Risk-based, science based
Primary focus is patient safety and product efficacy
Business benefits are also drivers
Results in improved process understanding
Results in improved process capability/robustness
Systematic development
Holistic – applies to all aspects of development
Multivariate – interactions are modeled
Provides PAR, design space, or suitable equivalent
Requires a significant reduction in regulatory oversight postapproval

1.2
Defining Product Design Requirements and Critical Quality Attributes

In order to design quality into a product, the requirements for the product design and performance must be well understood in the early design phase. In pharmaceuticals, these product requirements can be found in a Quality Target Product Profile (QTPP). The QTPP is derived from the desired labeling information for a new product. Pharmaceutical companies will use the desired labeling information to construct a target product profile that describes anticipated indications, contraindications, dosage form, dose, frequency, pharmacokinetics, and so on. The target product profile is then used to design the clinical trials, safety and ADME studies, as well as to design the drug product, that is, the QTPP.

In addition to defining the requirements to design the product, the QTPP will help identify critical quality attributes such as potency, purity, bioavailability or pharmacokinetic profile, shelf-life, and sensory properties as shown in Figure 1.1. In some cases, these attributes are directly measurable, for example, potency. In other cases, surrogate measurements are developed indirectly to measure the quality or performance, for example, *in vitro* dissolution for a controlled release product.

There are numerous ways to represent a QTPP. Another example of a QTPP for a lyophilized sterile vial is shown in Table 1.3.

A crucial element of QbD is to ensure that the measurement systems being used are truly assessing the quality of the product or performance. Very often it is the case that attributes that have little to do with quality are measured, for example, dissolution test for an immediate release Biopharmaceutical Classification System (BCS) class I drug (high aqueous solubility and high permeability). Drugs of this type are rapidly and completely absorbed; therefore, a dissolution test provides little value from a quality control perspective. Quality attributes can sometimes be modeled on the basis of first principles or other multivariate analysis. Predictive models are extremely important components of QbD [10]. In the case of bioperformance, predictive statistical, mechanistic, and analytical tools

QTPP to Critical Quality Attributes								
Quality Target Product Profile	Critical Quality Attributes							
	Assay/Potency	Impurities	Content uniformity	Stability	Dissolution	Average weight	API solid form on bead	Water content
Paediatric sprinkle dosage form								
2 mg, 4 mg & 8 mg dose	✔		✔	✔				
Oral, once-daily dosing					✔	✔		
Shelf life at least not less than 2 years at 25C/60% RH				✔			✔	
Blister and bottle packaging				✔			✔	
Same in vivo performance as adult product					✔	✔		
No food effect					✔			
Degradants/impurities below safety threshold or qualified		✔		✔			✔	✔
Meets Pharmacopoeial requirements for oral solid dosage forms								

Figure 1.1 Product requirements from QTPP help to identify potential critical quality attributes.

are being applied, which can guide Active Pharmaceutical Ingredient (API) particle size selection, dissolution method design, and setting specifications [11].

While a QTPP is basic to QbD, additional product or process design requirements may need to be considered while designing the manufacturing process for a new API or drug product. In API route design, major decisions need to be made regarding which chemistry will yield a synthetic route that delivers high purity at an acceptable cost [12]. Likewise, a drug product formulation and process technology decision needs to be made that also delivers a drug product that conforms to the quality requirements at an acceptable cost. An understanding of the product (formulation) design is critical to product performance. A clear rationale for why excipient types, grades, and amounts are selected is part of the product understanding. An understanding of which material attributes contribute most to the excipient functionality is important to performance. Supplier specifications may be a poor indicator of excipient functionality in a dosage form and hence may not be critical material attributes. In some cases, it may be necessary to introduce additional testing on incoming materials that are more relevant to how the excipient impacts the dosage form performance [13]. Likewise, the solid form of the API needs to be engineered for quality. The selection of the proper

Table 1.3 Quality target product profile for a lyophilized sterile vial.

Quality target product profile for a lyo vial for sterile injectable	Requirement
Indication	Chronic disease (treatment of nervous breakdown)
Dosage form	Lyophilisate for solution for injection
Dosage strength	Nominal dose 20 mg/vial
Administration route	Subcutaneous (0.8 ml)
Reconstitution time	Not more than 2 min
Solution for reconstitution	1 ml 0.9% saline (provided by the pharmacy)
Packaging material drug product	2R glass vial, rubber stopper, meets pharmacopoeial requirement for parenteral dosage form
Shelf life	Two yr 2–8 °C
Drug product quality requirement	Meets pharmacopoeial requirement for parenteral dosage form as well as product specific requirements
Stability during administration	Reconstituted solution is stable for 24 h at temperature $\leq$30 °C

salt, solid form (amorphous, polymorph), particle size and morphology, and degree of aggregation will impact critical quality attributes such as solubility, dissolution rate, chemical and physical stability as well as manufacturability (bonding index, stickiness, flow, filterability). Advances in crystal engineering enable better control and understanding of how to achieve targeted API particle properties (Chapter 7).

Finally, the role of the packaging systems for the raw material, in-process materials, and final drug product needs to be understood. All packaging systems should be demonstrated to protect the materials and not introduce contamination, for example, leachables or extractables, during transport and handling. The QTPP will set expectations for the final drug product packaging. True product understanding should translate into design spaces for the API properties, formulation, manufacturing process, and the packaging systems.

One of the biggest challenges is to integrate the design and process development at the key interfaces in the supply chain. Interfaces that present significant challenges to process understanding and hence process control are highlighted in Figure 1.2.

While QbD does target designing quality into processes, it can also be equally effective in identifying methodologies directed at reducing the high costs of development and manufacture of pharmaceuticals. Inclusion of attributes that measure costs directly or indirectly is essential to optimize the quality, time, cost, and risk relationships. Figure 1.3 shows the ''cost of quality rework'' relative to the stages of the R&D and manufacturing lifecycle [14]. The greatest opportunity to manage process costs and the product quality of a pharmaceutical is in the early process and product design phase when decisions are made about technologies and materials to be used. Although these are major decisions for pharmaceutical

Figure 1.2 Key material-process interfaces in a pharmaceutical product.

Figure 1.3 Cost of product quality or rework.

companies, they are often made implicitly rather than explicitly. Interestingly, few companies actively manage this phase of design and assume that decisions made in a vacuum were appropriate (Chapter 12).

1.3
The Role of Quality Risk Management in QbD

ICH Q9 discusses the role of risk management in pharmaceutical development as follows:

> To select the optimal product design (e.g., parenteral concentrates vs. pre-mix) and process design (e.g., manufacturing technique, terminal sterilization vs. aseptic process).

To enhance knowledge of product performance over a wide range of material attributes (e.g., particle size distribution, moisture content, flow properties), processing options, and process parameters.

To assess the critical attributes of raw materials, solvents, Active Pharmaceutical Ingredient (API)-starting materials, API's, excipients, or packaging materials.

One role for management in QbD is to ensure that teams utilize risk assessment tools that are capable of providing risk- and science-based reviews at critical milestones in the R&D lifecycle. One such critical milestone is prior to finalization of process technology, synthetic route, or a qualitative formulation. Decisions made at these milestones will generally impact the quality and costs attributes to a much greater extent than decisions made during process development and later in the product lifecycle. As with any rigorous risk assessment, it is important to include appropriate subject matter experts to obtain prior knowledge and apply feedback learnings to these major decisions.

Process understanding is achieved when the relationship between critical quality attributes (CQAs, y) and all the sources of variation (x) in the manufacturing process are understood:

$$y = f(x)$$

The principle sources of quality variations (examples) or inputs to a process include

- material attributes (peroxides, water content, impurities);
- process parameters (temperature, force, speed);
- equipment design (baffles, agitator type, surface type);
- measurement system (sample prep, extraction time);
- environment (relative humidity, temperature, oxygen content);
- person (operator, analyst).

It is important to note that the total process variation as measured by the variance or standard deviation (σ) of the average batch data is a function of all sources:

$$\sigma_{\text{Total}} = f(\sigma_{\text{Material}} + \sigma_{\text{Process}} + \sigma_{\text{Equipment}} + \sigma_{\text{Measurement}} + \sigma_{\text{Environment}} + \sigma_{\text{Person}})$$

The goal of process understanding is to be able to predict how the sources of variation (x) will impact the CQA performance (y) and be able to control these parameters to control quality. One of the initial challenges to design and develop a new API or drug product is to identify all the possible sources of variation for a particular new manufacturing process. The list of possible sources of variation will be very large, too large to study experimentally. The challenge presented to a scientific team is to sort out which inputs are at highest risk for impacting the process. Fortunately, QbD (e.g., ICH Q9) provides tools to systematically risk assess all the possible inputs to a process to identify those relatively few that have the greatest potential to impact the process. Table 1.4 provides an ISO 3100 list of

Table 1.4 Success factors in risk management.

Risk management should
    Create value
    Be an integral part of organizational processes
    Be part of decision making
    Explicitly address uncertainty
    Be systematic and structured
    Be based on the best available information
    Be tailored
    Take into account human factors
    Be transparent and inclusive
    Be dynamic, iterative, and responsive to change
    Be capable of continual improvement and enhancement

---

success factors for successful risk management [15]. Any organization embarking on QbD and or a QRM program could use this list as an internal quality check for their QRM program.

Ishikawa (fishbone) diagram is a very effective tool to capture a brainstormed list of potential process inputs impacting variation. Mapping the manufacturing process using a process flow diagram (PFD) is helpful to define the scope of the risk assessment and to identify possible process inputs. API mapping may include unit operation, chemistry pathways, and an impurities cascade. An example of mapping API and drug product processes is shown in Figure 1.4.

FMEA (failure modes and effects analysis) or use of a prioritization matrix (cause and effect matrix, Figure 1.5) is helpful in identifying the process inputs that impact on quality attributes. In some cases, a deeper dive into the driving forces at critical control points in the manufacturing process can yield a more fundamental understanding of sources of variation.

Once the CQAs and process performance attributes (PPAs) are associated with inputs to the process, $Y_i = f(x_1, x_2, \ldots, x_n)$ through a risk assessment process, experiments can be efficiently designed to develop predictive models and confirm causal relationships.

Before embarking on extensive experimentation, a critical next step is to make sure that critical measurements are made using ''fit for purpose'' methodology. A comprehensive risk assessment should identify those measurements that are suspect. A simple frequency plot of the data with specification limits will provide an indication of when variation is a potential problem (Figure 1.6).

The time spent improving a nonrobust analytical method can provide significant return on investment when experimental results yield true process understanding and control [16]. In this author's experience, sampling and sample preparation are typically high-risk areas for product quality measurements, for example, chromatography. Gage R&R studies are useful QbD tools to assess the relative contribution of the measurement system to the total variation of a manufactured product [17]. If the measurement contributes more than 10% of the total variability, additional method development is often warranted. However, some methods must contribute

Figure 1.4 Process map of API and drug product manufacturing processes.

a much lower variance to the total. Measurement of trace levels of genotoxic impurities is often a particularly challenging method development exercise since safety limits are approaching the limits of quantitation [18]. The opportunity to improve analytical methods or implement a totally new method may be more rapidly achievable in the future if the concept of an "analytical target profile" is adopted. The ATP defines the analytical criteria necessary to achieve equivalent or better analytical performance [19]. Analytical method understanding is crucial to

Importance of quality attribute		
Quality attributes		Score
Influence matrix		

(Process parameters)

Figure 1.5 Cause and effect matrix related process parameters (inputs) of quality attributes (outputs) of a process.

Consider measurement system analysis

X
X X   X
X X   X X X X X X X X
LSL   USL

Figure 1.6 Frequency plot of data from tablet potency measurements with specification limits. Note that this distribution of potency data is off-centered and relatively wide compared to the specification range, leading to questions about recovery and reproducibility of the method.

QbD. For example, how the materials are processed can impact the capability of the method to accurately quantitate an analyte. Compaction pressure is known to impact the near-infrared (NIR) spectra and may need to be included as a parameter in an NIR calibration program [20].

Ideally, these relationships are modeled such that interactions among the input parameters are known. Simple or complex models can then be used to create a design space that defines an acceptable operating region for the process.

Combining formal risk ranking and a statistical design of experiments (DoEs) is a powerful duo of tools in QbD, which is used extensively in the industry today (Figure 1.7). One of the reasons for this combination to be so popular is that most companies have access to the expertise required to utilize this combination; it is

Risk assessment (ranking) ➡ Brainstorm risks, focus on higher risks

Create experimental plan ➡ Craft experiments needed to understand high risks

Screening DoE ➡ Reduce risk uncertainty, confirm high/low risks

Response surface DoE ➡ Achieve process understanding

Design space ➡ Integrate knowledge, establish boundaries for process

Control strategy ➡ Identify critical control points and apply appropriate monitoring and control systems

Figure 1.7 Combination of risk assessment and statistical design of experiments (DoE).

also highly effective and efficient. A typical sequence of study is discussed in the example below.

A risk assessment ranked the process parameters likely to impact charge heterogeneity of a monoclonal antibody (MAb) as measured by the ion-exchange chromatography (IEC). The CQA of interest was charge heterogeneity. Multiple screening and response surface DoEs were performed that included testing of charge heterogeneity to confirm which process parameters impacted charge heterogeneity. The DoE analysis eventually enabled identification of process ranges that would control charge heterogeneity to an acceptable value [21].

Additional knowledge can be extracted by applying multivariate analysis [LVM, principal component analysis (PCA)] and data mining to integrated batch, process, stability, and bioperformance datasets. These tools have the benefit of extracting knowledge from a single product database or a portfolio of products with similar processes and technologies.

Another application of risk management tools is deciding which attributes and parameters are ''critical'' from a regulatory perspective. There has been much discussion and debate within the industry on how criticality should be defined and practiced. The ramifications of the critical designation are quite significant in the pharmaceutical industry as it defines the composition of the design space and the focus for the control strategy. The CQAs and critical process parameters (CPPs) are the foundation from which regulatory commitments are made. Changes to the design space or the control strategy would typically require a prior approval from regulators. Process validation protocols typically stipulate what are the CQAs and CPPs and monitor and control their performance.

The ISPE PQLI subcommittee on criticality has attempted to establish guidance on deciding critical parameters and attributes. Criticality is viewed on a continuum from low to high criticality. The realization that a parameter or attribute criticality can vary over a wide range was viewed as a breakthrough. However, the reality is that regulators expect pharmaceutical companies to draw a clear distinction between noncritical and critical to assist with the application of regulations.

FMEA and FMECA (Failure Mode, Effects, and Criticality Analysis) are useful as decision-making tools and also as risk mitigation tools. An example of how FMECA can be employed as a criticality decision-making tool is shown in Table 1.5.

1.4
Design Space and Control Strategy

ICH Q8(R2) defines design space as:

> . . . the multidimensional combination and interaction of input variables (e.g., material attributes) and process parameters that have been demonstrated to provide assurance of quality. Working within the design space is not considered as a change. Movement out of the design space is considered to be a change and would normally initiate a regulatory postapproval change process. Design space is proposed by the applicant and is subject to regulatory assessment and approval (ICH Q8(R2).

In some cases, boundaries will be identified that are known to be an *edge of failure*. In these situations, it may be important to set boundaries at acceptable tolerance intervals around the edges of failure to better mitigate the risks near such edges (Figure 1.8). Application of a tolerance interval is generally not necessary when the edges of failure are not in play at design space boundaries.

To make matters more complicated, an understanding of how the CQAs interrelate is important. If multiple CQAs are impacted by one or more of the same process parameters, the acceptable operating region can be greatly limited. A variety of multifactorial and multivariate modeling approaches should be considered. Modeling based on first principles, for example, reaction rate kinetic model, is the preferred approach; however, empirical methods can also be very effective. In order to establish acceptable boundaries, that is, design space for multiple interrelated CQAs, the response surfaces of these CQAs should be overlaid upon one another using the same parameter axes. CQA trade-offs may be required. As an example, the high cationic concentration of pDNA favored the biological activity of a vaccine but was deleterious to the physical stability of the liquid product. Trading some stability for biological activity was necessary to finalize the design space and optimize the formulation [22]. Modeling approaches and examples will be discussed in more detail in other chapters.

Once a sufficient level of process understanding is achieved, a control strategy should be developed that assures that the process will remain in control within the normal variation in material attributes and process operating ranges. The process understanding will identify where the appropriate control points are in

Table 1.5 FMECA used to assess the criticality of a process parameter.

Parameter name	Potential failure mode	Potential failure effect	SEV (1–10)	Potential causes (optional)	OCC (1–10)	Current controls	DET (1–10)	RPN (S*O*D)	Criticality designation	Justification
Reaction temperature	High temperature	Increased levels of impurity "x"	10	• Recipe error • Temperature sensor issue • Pump issue	3	• Automated recipe control • Sensor calibration • Back-up pump available	3	90	CPP	Impurity "x," which is a CQA, formed during API step where purge is minimal. No rework procedure available presently to reduce elevated levels. Deemed a CPP as a result, in spite of good controls described
	Low temperature	Slow reaction rates	5	• Recipe error • Temperature sensor issue • Pump issue	3	• Automated recipe control • Sensor calibration • Back-up pump available	1	15	Non-critical	Reaction completion; IPC in place, which will mitigate against incomplete reactions. Plant controls in place including demo run prior to first batch deemed sufficient to minimize risk

Figure 1.8 Design space with an edge of failure (EoF) and use of tolerance interval to mitigate risk.

the manufacturing process. Typically, these control points would be located where the variation is highest or where a CPP dominates control of the resultant product quality. For example, critical raw material attributes may be critical inputs to a process step. One mechanism to control the process is to control the quality of that material such that it always delivers a consistent product. Impurity fate mapping (IFM) is such an example in which the raw material and process impurity sources are identified and their fate mapped throughout the process. The process capability to remove these impurities at CCPs is an essential element of the control strategy [23].

Another control strategy could be to adjust the process parameters to accommodate the variation in the raw material attributes. This latter strategy would be dependent on having measurements systems in place that could measure critical material attributes, which then adjust other critical process parameters accordingly to maintain process control. For example, the amount of water and granulation mixing endpoint may vary batch to batch based on the granule size and count [24]. Control strategy is a cornerstone of a modern quality system. It can be a combination of parametric and attribute-based controls. Generally, real-time monitoring and control of the process is preferred over relying on end product testing. For example, the logical place to test for a major process impurity would be at the last step at which the impurity is purged from the process. Spiking studies could be performed to demonstrate the robustness of the process to purge high levels of the impurity. Over time, it may also be possible to demonstrate high process capability (Cpk > 2) and reduce or eliminate the test and rely more on parametric control. The control strategy should allow adjustments in testing plans based on commercial batch experience, that is, process capability and process understanding.

1.5
Quality Systems

While QbD is most effective when it is employed at a product/process design level, it should also be accomplished in the manufacturing and quality assurance environments. The authors of ICH Q10 foresaw the need to provide guidance on a modern quality system that would be critical to support QbD and continuous improvement

of pharmaceutical products over their lifecycle. Continuous improvement of a product and process should be employed throughout the lifecycle of a product. Process capability (CpK) is an extremely valuable metric to indicate which CQAs or other PPAs are least robust. CI efforts generally focus on the low CpK attributes.

A modern quality system may necessitate retooling the quality assurance workforce to be capable of interpreting more complex technical reports that rely more on predictive models, multivariate analysis, simulations, and advance process controls. Some of the PAT and design space models may require periodic updating. Interpreting the risks associated with process changes may be more complicated, as the risks change depending on how close the process is to an edge of failure.

As regulators entrust industry to make significant improvements in product and process quality, quality systems become more important to manage the changes that occur in pharmaceutical manufacturing. The FDA utilizes a postapproval management plan (PMP) to clearly articulate under what conditions the FDA will need to be informed or approve of such changes. Hudson has proposed a more detailed structure on how to format a PMP [25].

Finally, as Janet Woodcock, MD, Deputy Commissioner for Operations/Chief Medical Officer at FDA, stated at the 2008 PDA meeting, ''QbD is an evolution and not a revolution'' – an evolution that is in response to the increasing cost pressures on both the regulatory agencies and industry to control the escalation of drug prices [26]. QbD will continue to evolve for years to come as new tools and technologies advance to improve the way we mitigate risks and increase our understanding and control of the manufacturing processes. In addition to increasing quality, the pharmaceutical industry will reduce development and manufacturing cycle times as well as costs in the process.

References

1. Immel, B.G. (2005) A brief history of GMPs. *Regul. Compliance Newslett.*, Winter.
2. U.S. Department of Health and Human Services Food and Drug Administration (2004) *Pharmaceutical cGMPS for the 21st Century – A Risk-Based Approach: Second Progress Report and Implementation Plan.*
3. U.S. Department of Health and Human Services Food and Drug Administration (2007) *Pharmaceutical Quality for the 21st Century: A Risk-Based Approach. http://www.fda.gov/oc/cgmp/report0507.html.*
4. Juran, J.M. (1992) *Juran on Quality by Design – The New Steps for Planning Quality into Goods and Services*, The Free Press.
5. Venkateshwaran, T.G. (2009) Global regulatory submissions for QbD: Wyeth's experience in the CMC pilot. *Pharm. Technol.*, **33** (10), 96–102.
6. Drakulich, A. (2009) Critical challenges to implementing QbD: A Q&A with FDA. *Pharm. Technol.*, **33** (10), 90–94.
7. McCurdy, V., am Ende, M.T., Bush, F.R., Mustakis, J., Rose, P., and Berry, M. (2010) Quality by design using an integrated active pharmaceutical ingredient – drug product approach to development. *Pharm. Eng.*, **30** (4), 12–32.
8. Nasr, M.M. (2007) Quality by design (QbD) – a modern system approach to pharmaceutical development and manufacturing – FDA perspective. FDA Quality Initiatives Workshop, February 28.
9. U.S. Department of Health and Human Services Food and Drug Administration

(2004) *Guidance for Industry: PAT – A Framework for Innovative Pharmaceutical Development, Manufacturing and Quality Assurance*, U.S. Department of Health and Human Services Food and Drug Administration, Pharmaceutical CGMPs.

10. Tong, C., Lozano, R., Mao, Y., Mirza, T., Ldbenberg, R., Nickerson, B., Gray, V., and Wang, Q. (2009) The value of in Vitro dissolution in drug development: a position paper from the AAPS in Vitro release and dissolution focus group. *Pharm. Technol.*, **33** (4), 52–64.

11. Selen, A., Cruañes, M.T., Müllertz, A., Dickinson, P.A., Cook, J.A., Polli, J.E., Kesisoglou, F., Crison, J., Johnson, K.C., Muirhead, G.T., Schofield, T., and Tsong, Y. (2010) Meeting report: applied biopharmaceutics and quality by design for dissolution/release specification setting: product quality for patient benefit. *AAPS J.*, **12** (3), 465–472.

12. Aamir, E., Nagy, Z.K., and Rielly, C.D. (2010) Optimal seed recipe design for crystal size distribution control for batch cooling crystallization processes. *Chem. Eng. Sci.*, **65** (11), 3602–3614.

13. Kirdar, A.O., Chen, C., and Weidner, J., Rathore, A.S. (2010) Application of near-infrared (NIR) spectroscopy for screening of raw materials used in the cell culture medium for the production of a recombinant therapeutic protein. *Biotechnol. Prog.*, **26** (2), 527–531. ISSN: 1520-6033.

14. Philip, C. (1979) *Quality is Free*, McGraw-Hill, New York.

15. ISO 31000:200. (2009) *Risk Management*, International Standards Organization.

16. Zhou, L., Socha, J.M., Vogt, F.G., Chen, S., and Kord, A.S. (2010) A systematic method development strategy for water determinations in drug substance using Karl Fischer titrations. *Am. Pharm. Rev.*, **13** (1), 74–84.

17. Burdick, R.K., Borror, C.M., and Montgomery, D.C. (2005) *Design and Analysis of Gauge R&R Studies*, Society for Industrial and Applied Mathematics, Philadelphia, PA.

18. Sun, M., Liu, D.Q., and Kord, A.S. *A Systematic Method Development Strategy for Determination of Pharmaceutical Genotoxic Impurities Source*, Organic Process Research & Development ACS ASAP CODEN: OPRDFK, ISSN: 1083-6160.

19. Pohl, M., Schweitzer, M., Hansen, G., Hanna, B.M., Borman, P., Smith, K., Larew, J., and Nethercote, P. (2010) Implications and opportunities of applying the principles of QbD to analytical measurements. *Pharm. Technol. Eur.*, **22** (2), 29–36.

20. Alcalà, M., Ropero, J., and Vázquez, R., Romañach, R.J. (2009) Deconvolution of chemical and physical information from intact tablets NIR spectra: two- and three-way multivariate calibration strategies for drug quantitation. *J. Pharm. Sci.*, **98** (8), 2747–2758.

21. Horvath, B., Mun, M., and Laird, M.W. (2010) Characterization of a monoclonal antibody cell culture production process using a quality by design approach. *Mol. Biotechnol.*, **45** (3), 203–206.

22. Toke, E.R., Lorincz, O., Somogyi, E., and Lisziewicz, J. (2010) Rational development of a stable liquid formulation for nanomedicine products. *Int. J. Pharm.*, **392** (1–2), 261–267.

23. Li, Y., Liu, D.Q., Yang, S., Sudini, R., McGuire, M.A., Bhanushali, D.S., and Kord, A.S. (2010) Analytical control of process impurities in Pazopanib hydrochloride by impurity fate mapping. *J. Pharm. Biomed. Anal.*, **52** (4), 493–507.

24. Huang, J., Kaul, G., Utz, J., Hernandez, P., Wong, V., Bradley, D., Nagi, A., and O'Grady, D. (2010) A PAT approach to improve process understanding of high shear wet granulation through in-line particle measurement using FBRM C35. *J. Pharm. Sci.*, **99** (7), 3205–3212.

25. Hudson, P.S. and Baker, D.D. (2009) The postapproval management plan: a tool to apply science- and risk-based approaches. *Pharm. Technol.*, **33** (1), 82–86.

26. Rathore, A.S. and Devine, R. (2008) PDA workshop on quality by design for biopharmaceuticals: concepts and implementation, May 21–22, 2007, Bethesda, Maryland. *PDA J. Pharm. Sci. Technol.*, **62** (5), 380–390.

2
Route and Process Selection

David J. Ager

2.1
Introduction

Process research and development for pharmaceutical products are often under considerable time and cost constraints, especially if the material has to be made for clinical and biological testing. With the pressure to use the final manufacturing route for phase 3 materials, the synthesis of a drug candidate may not be optimal. Yet, the drug innovator has to compete in later years with generic manufacturers who have had the relative luxuries of time to develop better routes and technological advances in the intervening years. This chapter outlines the factors that must be weighed when a route to a drug candidate is selected. Safety, environmental, and patent ''freedom to operate (FTO)'' factors have to be taken into account along with more traditional aspects to synthesis, such as expediency in the number of steps, high convergence, avoidance of protection, and redox sequences. For some transformations, there may be equipment limitations. Route selection must take all these factors into account while ensuring a high chance of success and, if more than one approach is feasible, then working toward a common intermediate so that options can be kept open without jeopardizing regulatory filings or time. The exercise is one of risk management and determining which factors are important to the success of the drug launch. Once the route has been selected, it then needs to be turned into a viable process and this may deviate from the initial thoughts as data is obtained. The route and process selection involves a wide variety of disciplines, such as chemistry, chemical engineering, environmental, safety, purchasing, regulatory, and legal, which must all work together to achieve success within a short timeframe. These constraints mean that differences have to be drawn between data that must be obtained, as with regulatory or safety procedures, to ensure success and data that gives a larger comfort zone but its knowledge is not critical.

The impact of route selection on the success of a drug cannot be understated. Most of the costs for drug development occur in late phase 2 and phase 3. The cost of the chemical is relatively less compared to the total developmental costs.

Table 2.1   Comparison viewpoints of parameters during scale-up [1].

	Discovery	**Development**	**Manufacturing**
Amounts	Grams	Kilograms	MTs
Cost	Trivial	Critical	Consistent/minimal
Purity	Fair	Excellent	Excellent
Purification	Any	Limited	None/procedures available
Conditions	Any	Limited	Standard
Raw materials	Catalog	Bulk	Established suppliers
Waste	Trivial	Critical	Known/measured
Reproducibility	Moderate	Critical	Exact
Scalability	Trivial	Critical	Established

This is illustrated by the different viewpoints of discovery, development, and manufacturing (Table 2.1) [1, 2].

Over the years, the pressure of time-to-market has changed the dynamics of the process, and companies have adopted different methods. The change from a pharmaceutical company preparing a drug candidate knowing that it was going to go in a plant owned by that company, and involving all transformations from readily available starting materials to the final active pharmaceutical ingredient (API) or even the final formulated drug form is becoming rare.

Thus, in the period from around 1960 to 1980, process development was kept off the critical path, as companies did not want to invest in drugs that would fail early. This led to intense activity to design a manufacturing process around phase 3, and, with time now as the enemy, development was often conservative and relied upon an adaptation of the original process rather than looking at alternative approaches. Cost reductions were achieved by running reactions on a larger scale and going down the learning curve. If the commercial market grew, the addition of more capacity, which does not require regulatory approval or interruption of supply if problems are encountered, was the next phase. This led to an undesirable state of generally inefficient and expensive routes being used to manufacture. Only when generic competition was imminent were new chemistries looked at [3].

This also led to major custom chemical manufacturers, who operated on a large scale, essentially ignoring starting materials and intermediates until the drug candidate was entering phase 3. This also allowed ''technology'' companies to play a significant role, especially as stereochemistry was becoming more important to control, because they might have a unique methodology to provide a rapid solution to a problem that had, to a large degree, been ignored.

How times have changed! Although large pharmaceutical houses may still have production capacity, the emphasis is often on the last few steps. The emergence of small biotech companies, and even some larger pharmaceutical ones, without their own manufacturing plants, places the emphasis on working with outside

manufacturing companies (see Chapter 11). The approaches to route and process selection can, therefore, vary widely from company to company and even within a single company from one drug candidate to another. The choice of a specific route is the result of a compromise between opposing risk factors.

The variables that need to be considered when choosing a route and process are outlined in this chapter. The interactions between the factors that can influence the success of a route are highlighted so that a minimal number may be considered when making a decision. Although this chapter focuses on route selection, this is not the end of the story, as a process will have to be developed to take to manufacturing. This is analogous to looking at a map and finding a route between points, knowing that you have to move a large amount of material between these places. A road may already exist and the best solution could be to hire a truck. If the terrain is mountainous with no road, pack mules may be the answer. If a large amount of material needs to be moved, building a road or railway may be a longer term but overall cheaper solution. Finally, a helicopter may be the way to proceed if speed is of essence, the price of the merchandise is high, and volumes are low. The route is still the same; the process of accomplishing the goal, traversing the route with material, can be different from what was originally envisioned. Other chapters in this book will discuss the move from process research to development, while this chapter concentrates on the initial factors to consider when developing a feasible route.

It must be remembered that most compounds do not complete the drug development process. Quantification of the number of compounds entering each development phase varies by source, but, for phase 1 drugs, many (from 4 to 99) will not complete the phase 1 hurdle to become a successful candidate. The number of compounds that succeed and have to be prepared a second time varies from company to company. Smaller companies will have financial concerns if their drug does not reach the appropriate goals and there are no backup compounds. By contrast, large pharmaceutical companies usually have more stringent criteria for a compound to enter the development pipeline and the success rate may be higher. However, the uncertainty should be a risk factor that determines if work has to be ''front loaded'' into process development [4].

The data in Figure 2.1 has been taken from various sources and includes compounds in development phases for second indications. However, it can be seen that the attrition rate for preclinical compounds to enter phase 1 is 60–80%. Comparing annual numbers also does not take into account the time lag for compounds to move down the development pipeline. This implies that one in three to five compounds will not be considered for a second synthetic campaign. The trends suggest that there is now a better chance of success than there was 10 years ago. The time lag makes interpretation of the movement from phase 1 to 2 difficult. The latest trend is that there is a high probability of moving to phase 2. However, only about one-third of the compounds in phase 2 will go to phase 3. Although there is attrition in phase 3, the manufacturing route for launch has been set, even for companies where a cheaper route will be found and used postlaunch; so, the number is not relevant in these discussions.

Figure 2.1 Number of compounds in development phases over the past 10 years.

Although this chapter concentrates on route selection, this cannot be decoupled from process selection. A one-step process to make the target molecule in a high-temperature vapor-phase reaction might look fantastic on paper, but the risk and capital equipment questions to implement it will be significant if the reaction is to be implemented on a large scale, to say the least! However, for smaller volumes, it might be possible to use microreactors or flow systems. Although the capital investment may be reduced compared to a batch process, the company now has to be willing to invest (and to them unproven) in technology.

The factors influencing process development and design are discussed in more detail in Chapter 12.

2.2
Route Evaluation

The need to develop a low cost, simple process, and get it first-time right, can be considered to be the panacea of process development. There are a number of methods to perform route selection [5, 6]. New drug candidates are invariably new chemical entities and, thus, even when analogy exists, the chemistry has not been proved. Compared to other chemical industries, the control of impurities and regulatory considerations mean that the initial route may be multistep to ensure that the appropriate quality can be achieved. Refinements and improvements can be introduced as the understanding of the chemistry is gained. The evolution of the process from route selection to the final manufacturing process must take these changes into account. The process is often shown in graphical form (Figure 2.2). Once commercial material has been produced, looking for better methods and improvements should not stop, even if only from a defensive position, once the compound goes generic.

Figure 2.2 Cycles within the development process.

Larger pharmaceutical companies can streamline the approach to a degree. Rather than the more traditional approach to process development – find the optimal route upfront and then be on the critical path for compound supply – it is now accepted that cycles should be undertaken. Process research should ''reach back'' into discovery and look at key reactions and intermediates. Process development should then look at the molecule itself and come up with early route(s). Chemical development should then put the process into the plant, tweaking as necessary to fit the site and equipment. This is a three-cycle process. The first cycle looks at a technically feasible method to produce enough material for early phase testing and makes it available as soon as the candidate has been identified. With provision of material as the primary goal, this first cycle can use an inefficient method. The second cycle begins when the data from the early tests show enough promise and is aimed at producing material for the next phases of testing and pilot-scale manufacturing. In many ways, this cycle is a reality check that the approach is feasible and can be scaled up for commercial production, although more experimentation and data collection will probably be required. The final cycle is to find a process that is commercially viable and can be used for phase 3 materials and transfer to manufacturing [3].

Route selection is one of the most important decisions to be taken when a new drug candidate moves into development. Many factors will determine whether the route selected is ultimately good or not. In many cases, the decision will not be called into question, as the drug candidate will fall out of the development pipeline owing to some biological or other property of the compound, such as stability. Little information is available; yet the route-selection decision will have long-term consequences. A large number of factors have to be taken into account and these are outlined in this chapter, while many of the other chapters in this book cover these topics in more depth. As a consequence, the reader should refer to these

chapters. Only the key points that will result in either a fast go/no-go decision or will have an impact on the weighting of a particular route are discussed herein.

As there are many factors to consider in the route-selection process, the larger the number of people who can provide expertise and opinions, the better the decision will be. The group should contain, in addition to chemists, a chemical engineer, manufacturing, quality, and safety, health, and environment (SHE) representatives, with access to legal, business, and supply chain resources also being important. At present, timing is considered to be the most important as making material for clinical testing is invariably on the critical path. Company philosophy determines the role of a project manager. Some companies assign this role to a single person to carry the project from cradle to grave. Others use a different person during the various development and commercialization stages.

Other factors are just as important and the major issues associated with each are discussed below. Green chemistry is usually associated with environmental concerns. However, many of the factors are interrelated and "green chemistry" also has throughput and safety components. Outside of timing, the acronym SELECT has been proposed for the important parameters that need to be considered during the route-selection process. SELECT stands for safety, environment, legal, economics, control, and throughput [7]. The criteria are summarized in Table 2.2.

Table 2.2 The SELECT criteria [7].

Criteria	Subcriteria	Examples of potential issues
Safety	Process safety	Explosions or exotherms
		Threat to workers or plant
	Exposure to substances harmful to health	Carcinogens or sensitizers
		Threat to workers
Environmental	Volume of wasted natural resources	Quantity and variety of solvents
	Substances harmful to the environment	Aquatic toxins and ozone-depleting chemicals
Legal	Infringement of intellectual property rights	Key intermediate patented by competitor
	Regulations that control use of reagents and intermediates	NONS (notification of new substances, EU legislation)
Economics	Meeting cost of goods target for future market	Long synthesis using expensive materials
	Investment costs to support development quantities	High cost of process cannot be changed in near term
Control	Control of quality parameters	Meeting specification and GMP requirements
	Control of chemistry and physical parameters	Nonselective reactions, unstable intermediates
Throughput	Time scale of manufacture in available plant	Long route with dilute stages
	Availability of raw materials	Rare natural products

GMP, good manufacturing practice.

As a pharmaceutical is being made, quality cannot be jeopardized. However, safety, environment, and health (of plant workers) also cannot be compromised.

The final outcome of the route-selection exercise should be a cost-effective manufacturing process that provides the desired material in high quality, at low cost while fulfilling the criteria of being a green or sustainable process. As the route is modified through the various scale-up stages, information has to be gathered and decisions made. These are summarized in Table 2.3 [8]. Some of the early changes are discussed in Section 2.5.

Table 2.3 Evolution of a process [8].

Target molecule	Factors considered	Outcome
Route selection	Overall strategy Perform number of ''killer experiments'' to determine best options	Most promising routes Cost estimation Health and environment classification Project plan

Initial optimization phase (initial move from route to process)	Initial definition of step parameters:
	    Solvents
	    Reagents
	    Catalysts
	    Conditions
		Process flow chart
		Analytical methods
		Basic process safety
		Suppliers

Process parameters defined	Studies put limits on control parameters
	Step parameters defined and possibly combined including
	    Work-ups
	    Purification steps
	    Telescoping
		Critical quality attributes
		Critical process parameters
		In-process controls
		Impurity tracking
		Solvent recovery
		Complete process safety data
		Materials for CT I/II

Process refining and scale-up	Refine process and control parameters	Physical parameters Specifications API/Registered starting material(s) Patent submission Materials for CT II/III
Commercial production	Learning curve and other improvements	Investments for large scale Validated manufacturing process DMF submission and approval Stability data Commercial material
API manufacture	Robust process with limited improvement potential	Commercial material at constant quality

2.3
Factors to Consider

As already noted, there are conflicting factors that need to be considered during the route-selection process. There is no easy answer nor is there a ''magic formula'' to follow. Different companies and even different people could come to different conclusions, as the underlying driving forces may not be the same.

2.3.1
Timing

Perhaps timing is the most difficult parameter to control and reach a satisfactory conclusion. The outcome will depend on the complexity of the molecule, current knowledge, scale, resources available, company philosophy, and stage in the development pipeline as they all have major influences on route selection and the likelihood of success at the next decision point.

For a compound that has just entered into the development phase, a medicinal chemistry route is available, even if only as a description from another source. Obviously, the clinical indication and potency will determine how much material is needed for the next study. For very potent compounds, this may be in the gram range and the current medicinal chemistry approach may just need tweaking. It may not be cost effective to embark on a change to the route even if a new one is more efficient. In other cases, bioavailability may be known to be an issue and large amounts may be required. The medicinal chemistry route may not be capable of delivering these quantities because of a number of reasons, such as low throughput, a hazardous intermediate, or reagent, or a starting material not being available. In many cases, the situation is between two extremes; with some modifications, the current route can be adapted to meet the immediate requirements but future needs will require a new route.

It is common for the delivery of the first batch of a drug entering the development phase, as a new chemical entity (NCE), to be on the critical path. The material from this first batch is needed for toxicological studies, for phase 1, and in many cases for formulation studies. Medicinal chemistry may have provided large laboratory amounts of the compound, but it is usually less than 1 kg. Much of the chemical and physical properties of the compound and intermediates are not known, as the compounds are novel. This implies moving in a rapid manner at this early stage taking huge risks owing to lack of knowledge of key process parameters [4].

Different companies wrestle with these decisions and come up with different approaches. As many compounds fail in phase 1, the use of the current route is the most expeditious as little time and resources need to be spent doing process research. There is often little time or resources to develop a completely new route. Smaller companies tend to adopt this approach but there is a trap waiting down the line (*vide infra*). However, a little forethought can pay dividends in the future.

The move from a medicinal chemistry route to kilogram batches is the time to look at the final steps of the sequence. Is a simple transformation involved, such as the hydrolysis of an ester? Are polymorphs a problem and does the crystallization have to be closely controlled? Is salt formation required for the API? If the last step can be locked in at this stage, yet allow a wide range of approaches to be used to the penultimate compound, then there may be some leeway to look at variations in the current route, even as the first larger-scale material is being made.

With the first delivery out of the way, or if the original route just cannot be used for the first delivery, route selection can begin in earnest. Many groups will also do this even when a workable route is available, as some small modifications can pay dividends, but this should not detract from the goal of developing a manufacturing process.

The approach will depend on the company, its size, and management philosophy. Some will do minimal amounts of process research to get the original synthesis to a state where it can be used for commercial manufacture. The thinking here is along the lines, ''It doesn't pay to delay product launch to get a better process. We're better off launching with an acceptable process and then investing in process development refinements to drive down costs.'' In other words, ''design it fast now, design it right later'' [9]. Real costs will be incurred by the use of a less than optimal route. There is no guarantee that less process development will be needed when compare to implementing a better alternative. In addition, vendors may need to be found for a new starting material, while ones for the first route, if this philosophy is known, will not look at process improvements themselves for bespoke products. Smaller companies, as they do not have the cash or resources to invest in significant process development, often take this approach. In addition, smaller companies tend to partner or sell their products to larger companies who then undertake the required development. This, however, is a double-edged sword, as the purchasing company often reduces the reward if a significant amount of work needs to be undertaken.

2.3.2
Costs

In addition to timing, cost is the other major variable that needs to be taken into account. Here there may be a delicate interplay between these two factors. Some companies are extremely conservative; while aiming to get the process right the first time, an inordinate amount of time is spent in addressing all perceived and real problems. This is also a costly exercise and, almost invariably, some unforeseen factor or problem still arises. However, getting it right the first time can save money; there are no failed batches, for example. This last method also reduces the lead time for getting the candidate to market. However, the approach requires intense effort early on to ensure that reactions are robust. The key is to have an experienced team that can meet the challenges of problems that arise during the scale-up process and adapt accordingly. This experience can manifest itself even at the gram scale. As an example, a highly exothermic reaction may not be applied

at 100 g even when successful at <10 g owing to heat removal problems caused by the lower contact surface areas between the reaction flask and cooling liquid. The author saw this when performing alkyllithium reactions when larger-scale reactions resulted in the hexane boiling even though the flask was immersed in a dry ice/acetone bath.

Good risk assessments at this early stage also help with the development of quality by design (QbD) criteria.

Another aspect associated with costs is resource availability. For a small group, performing the first large-scale syntheses may take almost all of the available resources and there may be none left to look at second-generation approaches. Outsourcing these first campaigns may be a cost-effective long-term solution, but this can be a difficult ''sell'' to the management. Outsourcing the route selection and process research for a new approach should only be done with a company that has had a successful track record of making materials at the projected commercial scale. This mitigates the risk of working with a company that makes intermediate scales and then having to repeat the whole exercise if the second-generation route is found to have scale-up issues.

2.3.2.1 Removal of a Chromatography Step

One of the common steps process chemists have to ''remove'' from a sequence is chromatography. Many companies do not have large-scale equipment to perform this separation technique. In addition, large amounts of solvents and stationary phase can be involved. In some cases, this technique may have been used as a matter of course to purify an intermediate or target molecule. In other instances, it may have been necessary to remove an impurity that was difficult to remove by other means. Understanding why the impurity is formed may give the lead to solve the problem and allow use of what initially looked like an unworkable route at scale to become a contender for a manufacturing process.

Such an example is provided in the synthesis of TAK-779 (**1**), a nonpeptide CCR5 antagonist [10]. In the original route to the α,β-unsaturated acid **3**, the β-keto ester prepared from **2** and dimethyl carbonate was reduced with $NaBH_4$ followed by dehydration and hydrolysis. The β-keto ester reduction was performed in CH_2Cl_2 and involved a portion-wise addition of the reductant to minimize formation of the diol, which was still formed as a by-product and required chromatography for separation. A change of solvent to THF/H_2O (10 : 1) at $-10\ ^\circ C$ still gave some of the diol, but dehydration followed by hydrolysis with NaOH allowed the desired **3** to be separated by simple extraction followed by crystallization (Scheme 2.1).

2.3.3
Safety, Health, and Environment (SHE)

None of these factors can be compromised. Reactions that involve hazardous reagents, such as azide, or a nitration are simple to flag. Some companies have the capabilities to run these reactions, while for other companies use of a reducing agent

Scheme 2.1

such as diborane or lithium aluminum hydride may be a nonstarter. Experience within the company is usually the knowledge pool.

Other reactions may not be as obvious and this can be where an experienced team reaps rewards. A laboratory run may show a temperature spike (and sometimes not in all runs). More than one person running a reaction at laboratory scale can have the advantage of negating an individual's experimental "quirks," as well as providing more than one pair of eyes for observation. Any questionable reaction should be run in a system where heat flow can be monitored, such as an RC-1, so that engineers can determine whether it is safe to perform in the available equipment. If the reaction has to be performed in a specific manner, slow addition of a reagent, for example, safety must be built into the plant with engineering safeguards to ensure that a runaway reaction does not ensue. This might involve additional costs or an unacceptable lead time to install the equipment.

The health of the operators cannot be compromised. Not only do they have to be aware of the hazardous materials they are handling but they also cannot be exposed to them. In addition to solvents and reagents, this can include intermediates and the API.

The SELECT parameters overlap with safety, environmental, and control to a certain degree [7]. The main type of issues associated with process and worker safety are as follows:

- Thermal runaway
- Gas evolution
- Potentially explosive or shock-sensitive materials
- Highly corrosive materials
- Acute or chronic toxicity
- Gentoxicity
- Pyrophoric and highly flammable materials.

For a specific chemical, the COSHH (control of substances hazardous to health) procedures can be used to minimize risks during handling. This three-tier approach is as follows:

- Where possible, substitute the chemical for a less hazardous one.
- If this is not possible, reduce the quantity of that chemical (e.g., catalytic rather than stoichiometric).
- If neither is possible, then use engineering controls and personal protective equipment.

In addition to the handling problems and costs, there will probably also be an environmental price to pay through waste disposal and control of emissions unless the material can be easily recycled.

As many compounds in a route are new, databases are available to determine some degree of predictive safety or hazard assessment for that chemical [7].

Microreactors are beginning to be accepted in pharmaceutical processes and can even be used under current Good Manufacturing Practice (cGMP) [11, 12]. The technology is particularly beneficial for hazardous reactions, including exothermic and fast reactions, where additions may be extremely difficult to perform in batch mode, and ones where a hazardous intermediate or reagent is used. If the method of numbering up is used to make larger amounts, the advantage of finding conditions in the laboratory can be quickly transferred to the plant with confidence [13]. The need to move from the laboratory to pilot plant to manufacturing is alleviated together with the somewhat unpredictable nature of moving between scales, although this is getting better as the understanding of the process is increasing. The number of industrial examples is now increasing [13–15].

Waste disposal is often forgotten by the laboratory chemist, but heavy metals, halogens, and aromatic by-products, to name but a few, can also present costly problems. In other cases, a solvent, such as dichloromethane, may need to be used to obtain a high yield. This substance will need to be captured and recycled; can the plant where the process is to be run accommodate this? If it cannot, then an alternative must be found.

2.3.3.1 Safer Processes

For a more in-depth discussion on the critical stages of safety assessment, see Chapter 3.

As with other aspects of route selection, a formalized methodology has been proposed with ''expert rules'' to alleviate some of the time pressures associated with gathering experimental safety data. Although the approach covers reactions outside of those used in pharmaceutical manufacture, such as vapor-phase reactions, many of the parameters are relevant [16, 17]. The method also sets up a reaction ''network'' where all possible reactions and by-product formations are considered. Each of these reactions is considered with the criteria in Table 2.3. In addition, there are factors to consider for the overall process and the development of flowsheets.

Table 2.4 incorporates all the factors and issues relevant to batch processing and for pharmaceutical manufacture. Those relating to large-scale commodity and petrochemical production have been omitted.

Table 2.4 Heuristics for inherent safety analysis during route selection.

Item	Condition	Safety issue	Alternative
Raw material Intermediate Catalyst Solvent	Hazardous material	Use of hazardous material in reaction Increased inventory of hazardous material	Look for a process that uses safer material Replace with a safer alternative Move to a safer alternative by changing the form of the material, structure of the material, or masking the material Dilution Couple reactions where the hazardous intermediate is made with one that consumes it to minimize buildup Change reaction conditions and catalyst as necessary
	Flammable material	Possibility of fire or explosion	Look for alternative
	Corrosive material	Equipment corrosion	Look for alternative
	Toxic, mutagenic, carcinogenic, or teratogenic material	Possibility of worker exposure	Look for alternative
	Reactive, polymerizable, pyrophoric, peroxide forming, water reactive, or thermally unstable	Possibility of unintended reaction	Look for alternative
Solvent Reaction	Hazardous solvent	Use of hazardous solvent	Use safer solvent
	Liquid phase catalyst	Increases exposure potential	Use solid or polymer supported catalyst
	High-temperature reaction	Use of extreme operating conditions	Look for alternative with safer operating conditions
	Reaction temperature > auto-ignition temperature of material	Material under ignition conditions	Change operating conditions to safer regime
	Heat of reaction	Release of large amounts of energy and potential for runaway	Look for lower energy alternative Use smaller reaction inventory

Table 2.4 (*continued*)

Item	Condition	Safety issue	Alternative
	Exothermic reaction with liquid phase	Loss of temperature control could lead to uncontrolled boiling, overpressure, and rupture	Increase robustness of reactor Use solvent to remove heat of reaction Use smaller inventory
	Exothermic reaction and vapor-phase reaction	Loss of temperature control and could lead to runaway reaction	Add nonhazardous inert material or diluent to remove heat of reaction Add excess nonhazardous reactants to feed stream to remove heat of reaction
	High-pressure reaction	Use of extreme operating conditions	Look for alternative with safer operating conditions
	Low conversion	Reduced conversion of raw materials resulting in the need for recycle or recovery unit and an increase in in-process inventory of materials	Use excess nonhazardous reactants to increase conversion Change reactor to improve heat and mass transfer Improve contact between reactants to ensure proper distribution and mixing of reactants and by avoidance of stagnant zones
	Process yield is low for catalytic reaction	Low throughput with large volumes	Look for alternative catalyst to get better yield
	Process yield is low for noncatalytic reaction	Low throughput with large volumes Stoichiometric reagents often required	Look for catalytic process to get better yield
	By-product formation from a main (desired) reaction	By-product formation results in a need for more separation units, use of more raw materials and waste of energy in the process	Change reaction conditions to minimize by-product formation Use nonhazardous or less hazardous materials that are easy to separate in excess Look for alternative processes with higher yield
	Reactants for side reactions are raw materials or intermediates	–	Limit the use of excess reactants

Table 2.4 (*continued*)

Item	Condition	Safety issue	Alternative
	Large amounts of materials in process	Large inventory of chemicals	Reduce residence time in reactor Change configuration or size of equipment without affecting throughput
Work-up and separations	Boiling point of material < operating temperature	Possibility of flash in case of leak	Reduce inventory of flammable material in separator
	Presence of thermally/polymerizable unstable materials	Possibility of thermal decomposition and polymerization	Keep operating temperature away from the decomposition temperature Conduct the separation process under low pressure Add inhibition materials to avoid unintended reactions
	Use of mass-separating agent	Use of mass-separating agent leading to need for an additional unit to recover such agent and increase in process inventory	Use a process that does not use a mass-separating agent Replace mass-separating agent in separation unit with less hazardous material Replace mass-separating agent in separation unit with less hazardous material and with a material that is easy to recover and has higher selectivity Replace the mass-separating agent in the separation unit with an in-process material
Heat exchangers and utilities	Use of a heat-transfer medium other than steam or cooling water	Use of a heat transfer liquid or refrigerant	Change the heat-transfer medium
	Presence of hazardous chemicals	Inventory of hazardous chemicals	Use compact heat exchangers

(*continued overleaf*)

Table 2.4 (*continued*)

Item	Condition	Safety issue	Alternative
	Temperature > 150 °C or pressure > 25 bar	Use of extreme operating conditions	Decrease temperature in equipment Decrease pressure in equipment
	Use of a hazardous solvent	Handling of hazardous solvent	Use in-process materials (raw materials, product, by-products, intermediates) as solvents in the process Use water as the solvent Use a less hazardous solvent with desirable properties
Storage	Hazardous intermediates/raw material supply from outside boundary limits/tanker/ pipeline/shipment	Inventory of hazardous chemicals	Reduce the inventory to the minimum required level by adopting a just-in-time approach *In situ* manufacture of material
	Hazardous intermediate/raw material supply from inside boundary limits	Inventory of hazardous chemicals	Reduce the inventory of the material to the minimum practicable level
	Hazardous intermediate from inside boundary limits	Inventory of hazardous chemicals	Modify the reactor, which allows the in-process consumption of the intermediate as soon as it is produced Reduce the inventory by use of a close coupled reactor arrangement instead of an individual reactor arrangement Modify the process configuration such that downstream units draw directly from the plant Change the operation philosophy such that downstream would be shut down or run at the lowest possible throughput when the upstream unit is shut down and *vice versa*

Adapted from Refs. 16, 17.

One aspect of route selection that is relatively simple to assess is when a reactive intermediate or reagent is involved. If the proposed route contains a reactive reagent, such as diazomethane, concentrated nitric acid, or phosgene, or if an intermediate contains a reactive group, such as azide or diazo, then alternatives should be looked at. Although there are ways of minimizing the safety problems, for example, by the use of microreactors, this will still require significant time and resources to develop a robust and safe method.

2.3.3.2 Green Chemistry

With a number of large-volume drugs becoming generic, environmental impact and cost issues are of current importance; this is also applicable for new compounds entering the route-selection process. One measure to address this issue is to strive toward a ''green'' or sustainable process [4, 18]. Chemists tend to focus on green chemistry, which relates to the chemical aspects of a synthesis or manufacturing process. The industry and other disciplines tend to adopt a broader approach and look for sustainable processes. The two overlap in many ways and green engineering also has its principles (*vide infra*).

The fundamentals of green chemistry are, to a degree, covered by the SELECT process, but with emphasis on the environment, social impacts, and cost [19]. The green chemistry concepts listed below can be interpreted and refined in various ways, such as the atom economy approach, where the vast majority of atoms in the starting materials should end up in the product.

1) It is better to prevent waste than to treat or clean up waste after it has been formed.
2) Synthetic methods should be designed to maximize the incorporation of all materials used in the process into the final product.
3) Wherever practicable, synthetic methodologies should be designed to use and generate substances that possess little or no toxicity to human health and the environment.
4) Chemical products should be designed to preserve efficacy of function while reducing toxicity.
5) The use of auxiliary substances (e.g., solvents separation agents, etc.) should be made unnecessary wherever possible and innocuous when used.
6) Energy requirements should be recognized for their environmental and economic impacts and should be minimized. Synthetic methods should be conducted at ambient temperature and pressure.
7) A raw material of feedstock should be renewable rather than depleting, wherever technically and economically practicable.
8) Unnecessary derivatization (blocking group, protection/deprotection, temporary modification of physical/chemical processes) should be avoided whenever possible.
9) Catalytic reagents (as selective as possible) are superior to stoichiometric reagents.

10) Chemical products should be designed so that at the end of their function they do not persist in the environment and break down into innocuous degradation products.
11) Analytical methodologies need to be further developed to allow for real-time, in-process monitoring and control prior to the formation of hazardous substances.
12) Substances and the form of the substance used in a chemical process should be chosen so as to minimize the potential for chemical accidents, including releases, explosions, and fires.

Twelve more principles have also been proposed that augment the original 12 [20].

1) Identify and quantify by-products.
2) Report conversions, selectivities, and productivities.
3) Establish full mass balance for process.
4) Measure catalyst and solvent losses in air and aqueous effluent.
5) Investigate basic thermochemistry.
6) Anticipate heat and mass transfer limitations.
7) Consult a chemical or process engineer.
8) Consider the effect of the overall process on the choice of chemistry.
9) Help develop and apply sustainability measures.
10) Quantify and minimize the use of utilities.
11) Recognize where safety and waste minimization are incompatible.
12) Monitor, report, and minimize laboratory waste emitted.

The 12 principles of green chemistry have also been summarized by the mnemonic PRODUCTIVELY [21]:

- P – Prevent wastes
- R – Renewable materials
- O – Omit derivatization steps
- D – Degradable chemical products
- U – Use safe synthetic methods
- C – Catalytic reagents
- T – Temperature and pressure ambient
- I – In-process monitoring
- V – Very few auxiliary substances
- E – E-factor, minimize feed in product
- L – Low toxicity of chemical products
- Y – Yes, it is safe.

The principles of green chemistry can be related to the potential liability or benefits of a particular route. In addition to environmental impact, economic factors can become apparent as summarized in Table 2.5 [22].

Table 2.5 Relationship between green and economic factors [22].

	Environmental	Economic
Atom economy	Minimal by-product formation Reduced environmental burden	Maximized use of starting materials, reagents, catalysts Reduced costs
Solvent reduction	Less solvent waste Reduced environmental burden	Reduced volumes Higher throughput Less energy requirements Reduced costs
Reagent optimization	Use of catalytic reactions Recycle possible Reduced environmental burden	Higher efficiency Higher selectivities Reduced costs
Convergency	Improved process efficiency Lower number of overall steps Reduced environmental burden	Higher efficiency Higher overall yield Fewer operations Reduced costs
Energy reduction	Reduced environmental burden due to improvements in power generation, transportation, and so on	Milder conditions Shorter process times Increased efficiency Reduced costs
In situ analysis	Reduced risk of exposure or environmental releases	Increases throughput and efficiency Less materials wasted Fewer reworks Reduced costs
Safety	Use of nonhazardous materials and processes reduces exposure, release, explosion, and fire risks	Worker safety improvements Fewer engineering control measures required Reduced downtime Reduced costs

The 12 principles of green engineering look at the sustainability issues from a different perspective [23]. These engineering principles complement and enhance those of chemistry in many ways.

1) Designers need to strive to ensure that all materials and energy inputs and outputs are as inherently nonhazardous as possible.
2) It is better to prevent waste than to treat or clean up waste after it is formed.
3) Separation and purification operations should be designed to minimize energy consumption and materials use.
4) Products, processes, and systems should be designed to maximize mass, energy, space, and time efficiency.
5) Products, processes, and systems should be "output pulled" rather than "input pushed" through the use of energy and materials.

6) Embedded entropy and complexity must be viewed as an investment when making design choices on recycle, reuse, or beneficial disposition.
7) Targeted durability, not immortality, should be a design goal.
8) Design for unnecessary capacity or capability (e.g., ''one design fits all'') solutions should be considered a design flaw.
9) Material diversity in multicomponent products should be minimized to promote disassembly and value retention.
10) Design of products, processes, and systems must include integration and interconnectivity with available energy and material flows.
11) Products, processes, and systems should be designed for performance in a commercial ''afterlife.''
12) Material and energy inputs should be renewable rather than depleting.

The ''greenness'' of a process has many aspects including the use of raw materials, solvent usage (and recycle), and reagent amounts whether catalytic, stoichiometric, or in excess. Among the metrics that have been put forward are Trost's atom economy [24, 25], Sheldon's environmental impact factor (E) [26], and reaction mass efficiency [27]. This last factor has been related to the other metrics, and can be used to assess the ''greenness'' of alternative route options. A related approach to the latter method is EATOS, the environmental assessment tool for organic syntheses [28].

As an example, catalytic reactions should be performed rather than using stoichiometric reagents [29]. A 1,4-reduction of an enone with hydrogen is greener than the use of a hydride donor reagent. In addition, hydrogen is also cheaper. For oxidations, hydrogen peroxide delivers more oxygen as a percentage of the oxidant than any other [26]. Biocatalysis should always be considered, where possible, as enzymes often force reactions to be run in aqueous media, and, of course, catalytic reactions are usually cheaper and greener than stoichiometric counterparts [26]. Despite some long established enzymatic reactions, the area of ''white'' (industrial) biotechnology is still seen as an emerging field and many process chemists and engineers tend to avoid it. Again, this illustrates the importance of having a group with diverse backgrounds participating in the route-selection process and having the willingness to embrace new technologies that are commercially viable.

A list of reactions that could benefit from having ''greener'' alternatives has been published. In route selection, additional thought should be given before one of these reactions is incorporated into a sequence, as most result in an expensive process or the need to have specialized equipment. The list of reactions where better alternatives are preferred is as follows [30]:

- Mitsunobu reactions
- Amide reductions with stoichiometric hydride reagents
- Bromination reactions
- Sulfonation reactions
- Amide formation reactions with poor atom economy reagents
- Nitration reactions
- Demethylation reactions

- Friedel–Crafts reactions on unactivated substrates
- Ester hydrolyses
- Hydroxy group activation for nucleophilic substitution
- Epoxidation
- Wittig chemistry with Ph_3P
- Radical chemistry with tin reagents.

Some of these ''problem'' reactions are being addressed by the use of flow reaction technologies, as this can reduce the environmental impact significantly, such as by recycling unused nitrating reagents, accessing high temperatures or pressures, and increasing reaction selectivities through improved heat and mass transfer.

2.3.4
Legal

In some ways, intellectual property (IP) is linked to FTO. The latter considers what competitors and others working in the field have protected and the ability to still make the target molecule without infringing any of those patents. However, the process and compound should still be protected. In most cases, a composition of matter or use of the material in a specific application provides the first line of defense. The process can provide a second line, which often extends past the compound becoming generic. Although a new route may provide a process patent, it tells competitors the process and helps them find ways to get round their patent. FTO is more important. The pros and cons of patenting a process are still being discussed [4]. The philosophy of developing a good process after a compound has been launched extends this patent coverage period. However, a good company does not stop looking at new methods and processes to commercial drugs as new methodologies are developed. It may be cost effective to change a process long after launch. With the rapid development of new synthetic methods and technologies, the chances are high that a new approach will have been developed before the drug becomes generic.

The chemistry precedence is a big driver in route selection. However, the downside to this is that the reaction, reagents, intermediate, or even transformation may be patented. In other cases, such as an asymmetric hydrogenation, it may not be too difficult to find an alternative catalyst that is off-patent or is owned by the company. In some cases, licensing may not be too expensive, but this must be established before committing to the final process. There are waivers for using a patented process under certain circumstances if it is for research purposes. In a few cases, this may be acceptable if the patent will expire before the drug candidate is likely to reach the market, but this can still be a risky venture.

The other legal issues that need to be addressed when considering a route are the use of regulated substances. These could be controlled or banned substances. In addition, hazardous materials can have legal constraints, such as not being able to transport them by air. This can add significant amounts of time if transoceanic travel

is involved. Customs clearance can also be a problem of time. In addition, some chemicals, such as thionyl chloride, cyanogen chloride, and phosgene can have severe restrictions outside of those imposed by their reactivity. In the European Union, it may be a problem if a process were to use unacceptable amounts of COMAH (control of major accidents and hazards) listed chemicals or the materials had third-party restrictions, such as notification of new substance regulations (NONS) data. Registration, evaluation and authorization of chemicals (REACH) is being implemented in the EU to cover some legal aspects as well as safety, and use of, or movement of, large amounts of materials may trigger the need for a registration.

2.3.5
Other Considerations

There are other factors that still need to be considered at this stage where an opinion can save time and cost if the proposed route has a step outside of standard chemistry and where an experts may have insight into ways to perform a specific step. As an example, if a catalytic step only gave a low conversion or selectivity in the "killer experiment," what are the chances of finding a catalytic system that would fulfill the goals for the step? What cost and time would it take to perform a screen? Could this screen be done in-house or does it have to be outsourced? If it is the latter, are the paperwork and other details, such as confidentiality agreements, in place so that it can be done quickly? This is also an area where FTO and IP issues need to be recognized, and agreements, at least in principle, made before embarking on the experimental pathway. In addition to chemical catalysts, there are other scenarios to consider.

The implementation of a single biocatalyst may require some process development. For bioprocesses where more than one transformation is enzyme catalyzed, a number of approaches may be viable and these may need to be prioritized. Although the development of screening, evolution, and genetic modifications has led to many more biocatalysts being available or accessible [31, 32], engineering aspects of the approach are as important for the successful development of a scalable process [33]. Thus, while a biocatalyst may give a high yield, the kinetics of the reaction, throughput, ease of implementation, and downstream processing also need to be considered. A semi-quantitative method to screen potential biocatalytic processes has been proposed to address the many variables and provide guidance to the preferred routes [33].

Some classes of compounds have a limited number of methodologies that can be used to prepare clinical and commercial supplies. Oligonucleotides, peptides, and carbohydrates can fall under this heading. Although the methodologies are expanding, pushed by the need to make larger amounts and the processes greener, the chemistries are still dominated by the methods used to couple units [34–36].

As already discussed, route selection and moving a process forward in the early stages of development are in a delicate balance with getting materials in time and costs. In many cases, the "best" manufacturing route will not result because of

these early constraints. The pharmaceutical company's process group, therefore, should always be looking at better routes and process improvements even if they are not implemented. The learning alone could lead to a new cost-effective route when the drug becomes generic.

One aspect of route selection that is often ignored is the time taken to perform an individual step (chemical transformation). An exothermic reaction may demand a slow addition time to control the heat output of the reaction. For example, a reaction that takes 30 min in the laboratory because of slow addition to control and exotherm may require 6 h in the plant because of poorer heat transfer at the larger scale. In addition, the workup of the reaction may be problematic. Here, experienced chemists can provide great insight: Is a solid difficult to filter? Is separation of layers slow? Sometimes a simple change in solvent can alleviate the bottleneck and it is quicker to do this in the laboratory to find the answer rather than run into the problem magnified many fold in the pilot plant.

To perform a single chemical transformation requires a number of "tasks" or "operations" to be carried out: The reactor must be charged with solvent, reactants, and reagents; heating or cooling may be needed. The reaction workup and isolation of the product involve additional tasks. For a single transformation, 60–80% of the tasks may not be directly related to the reaction itself [37]. In route selection, hindsight can be a wonderful thing. When embarking on a route scouting and scale-up program, it is imperative to be safe and have an action plan. Before performing experiments, the group should think about what useful data should be collected and how it will be used in future experimental design and campaigns. One aspect of early campaigns is the isolation of intermediates that may be redundant when steps are telescoped and the process is better understood. This approach not only allows for intermediate fixes in the process, by remedial purification of an intermediate, but also allows for analytical data about impurities to be obtained. This information can then be used to streamline the process in future campaigns and to decide whether the conservative approach is necessary. This aspect of route and process selection is not always appreciated as illustrated in an article where the author asks the question "So why does the US pharmaceutical industry persist in using complex manufacturing processes to make APIs? Consider . . . API process chemistry: typically, each intermediate must be isolated, a cumbersome and costly process" [38]. The answer lies in the approach and stage of development. Little is known in the early stages and the information required for second-generation and generic manufacture is not available. In addition, patents are often filed soon after the route has been shown to be viable rather than give away the manufacturing method. Perhaps pharmaceutical companies will learn from the current situation with many large-volume drugs going generic, and will protect their processes, and potential processes, better in the future.

With a medicinal chemistry route that needs significant work to make material, little is known about the process chemistry, impurities, and control factors and parameters. Analysis is needed to understand what is going on before the experiments are performed. A process can be simplified later on and, perhaps, analyses removed, as they do not contribute to quality control.

A $\xrightarrow{\text{95%}}$ B $\xrightarrow{\text{95%}}$ C $\xrightarrow{\text{95%}}$ D $\xrightarrow{\text{95%}}$ E $\xrightarrow{\text{95%}}$ F $\xrightarrow{\text{95%}}$ Product
      95%   90%   86%   81%   77%   74%
  (based on A)   (based on A)   (based on A)   (based on A)   (based on A)   (based on A)

A $\xrightarrow{\text{95%}}$ B $\xrightarrow{\text{95%}}$ C $\xrightarrow{\text{95%}}$ E $\xrightarrow{\text{95%}}$ Product
      95%   90%   86%   81%
  (based on A)   (based on A)   (based on A)   (based on A)

A $\xrightarrow{\text{95%}}$ B $\xrightarrow{\text{95%}}$ C
      95%   90%
  (based on A)   (based on A)
                                    $\xrightarrow{\text{95%}}$ Product
                                                86%
                                         (based on A or D)
D $\xrightarrow{\text{95%}}$ E $\xrightarrow{\text{95%}}$ F
      95%   90%
  (based on D)   (based on D)

Figure 2.3 Comparison of a shorter linear sequence and a convergent approach.

Some points are proposed that do offer guidance when looking at each process step [38]:

- Can each step be performed in time? If not, can time and costs be reduced?
- How will reaction kinetics affect the total reaction time?
- Are the best solvents for the process being used? This includes using solvents that offer maximal density differentials to aid phase separations and workup in general.
- Will the intermediate require isolation?
- Can the same solvent be used throughout the process or through a number of steps?
- Are commercial conditions being replicated in the development process?
- Is the process such that it delivers quality product rather than quality being achieved by product testing?

2.3.5.1 Throughput

If the medicinal chemistry route is taken as the initial starting point, then any subsequent process routes should be cheaper and more efficient. A reduction in the number of chemical steps can be achieved by short cuts, say by avoiding a change in protecting group, or better by the use of a convergent synthesis. These scenarios are summarized in Figure 2.3 where all chemical yields are an optimistic 95% for each step. The convergent sequence has the same number of steps as the longer linear sequence. The overall goal is to rapidly build up molecular complexity with a minimal number of steps [39].

Although many issues that affect throughput cannot be determined until experimentation has been performed, such as the need for a slow addition to avoid by-product formation, or even when transferred to manufacturing, some concerns

Scheme 2.2

can be evident at the route selection stage [7]. The key variables that influence throughput are as follows:

- Chemical yield
- Availability and capacity of available vessels
- Cycle time for a step (reaction time, work-up time including crystallization time, distillations, drying, and cleaning)
- Concentrations and volumes
- Number of unit operations
- Use of specialized equipment
- Use of protection or salt formation where the formula weights are increased significantly by material that does not end up in the product
- Low availability of a starting material or reagent.

An example of a reaction that has low throughput but is currently popular in academia and drug discovery is a ring-closing metathesis. Although formation of larger rings is a very useful reaction, as with the HCV protease inhibitor BILN 2061 (Scheme 2.2), a large amount of work was required to alleviate dimer formation at higher concentrations [40] – a problem arising from the reversible nature of the metathesis reaction itself. Even with optimization work, the concentration is still 0.2 M.

AstraZeneca has established a relationship between the number of steps and the amount of API that can be manufactured in a specific time. A simple model has been generated that calculates the number of batches required in a particular campaign [7].

$$N_B = \sum_1^n \left(\frac{R_n M_{wt,n-1}}{M_{wt,n} Y_n} \right) \frac{V_n}{V_{P,n}}$$

$$T_m = \frac{N_B}{P} + \text{misc}$$

where $P \equiv \dfrac{A_P}{T_c}$

The variables in these equations are

Y = chemical yield (%)

M_{wt} = molecular weight (g mol^{-1})

V = bottleneck operational volume (L)

V_P = plant volume (L)
A_P = plant availability
T_c = process cycle time (weeks)
P = productivity (batches per week)
n = number of isolated steps
N_B = number of batches for required amount of material
R_n = amount of material required ($n = n$ for API, $n = 1$, for first stage isolation, etc.)
T_m = length of manufacturing campaign (weeks)
misc = time for other activities such as interstage cleaning.

To improve throughput, a number of factors can be considered and this is useful at the route selection stage, as it may prove prudent to adopt a different synthetic strategy. For example, an intermediate may not be very soluble and avoiding this particular compound may be the best solution. Factors to consider are as follows:

- Chemical yield can often be improved through a better understanding of reaction kinetics and mechanism. The screening of parameters such as solvent, reagents, and catalysts can lead to improvements.
- If the capacity, number, and types of vessels are the limiting factors, consider using an alternative plant.
- Reduce the number of the most time-consuming unit operations through ''telescoping.'' These operations include solvent replacement, extractions, crystallization, filtrations, and drying.
- Poor solubility can lead to dilute reactions. A solvent change or derivatization of material may help but avoiding the situation could be the best solution.
- Specialized techniques, such as chromatography, can be slow. Look for alternatives or adapt to be continuous.
- High molecular weight protecting groups, salt forms, and reagents can decrease throughput. Try and avoid these.
- Raw material suppliers with long lead times can have a large impact. An efficient supply chain must be implemented.

2.3.5.2 Solvents
For some processes, solvent usage can be a big issue. During the route-selection process, the potential to use the same solvent for two or more steps should be considered positive, as this could also result in telescoping. In most cases, solvent recycle will reduce solvent usage and costs. Thus, steps that require mixed solvent systems in either the reaction or workup, and where subsequent separation of these solvents for recycle is not easy, should be investigated for alternatives. However, looking for these problems at the planning stage may not be easy.

As noted under safety and green processes, nonhazardous and safer solvents should be sought and their usage should be as low as possible. Water is often forgotten as a useful solvent or cosolvent for organic reactions (see Chapter 12).

2.3.5.3   Raw Materials

Raw materials play a key role in the assessment of initial routes and it is often difficult to make comparisons with the limited amount of knowledge available. Some raw materials are commodity chemicals and the amounts used, even for a large volume and successful pharmaceutical, will have little impact on the current market. If available, large-scale prices should be used, although attention should be paid to the quality and stability of supply. On occasion, it may be necessary to contact a company that makes a similar product to see if it can apply its technology to the required raw material. Reputable companies usually give a price indication for large volumes based on current knowledge and scale-up experience even though the price for the initial small volumes may be several orders of magnitude higher than this price. It is unrealistic to expect a few kilograms of a compound that has not been made at scale to be only slightly higher than the projected cost for tons!

In the long term, with a successful drug, the raw material costs will reduce as more competitors enter the market and new technologies, which may not have been available when the original route was selected, are used.

2.3.5.4   Intermediates

Although processes often benefit from the telescoping of steps – performing two or more steps in a sequence without isolation of an intermediate – purification of the resultant material may be made more difficult as impurities may not be purged. This is of particular concern when the sequence product is not a solid, although alternative purification techniques might still be available to accomplish the purification, such as acid/base extractions, distillations, and so on.

2.4
Route Selection

The variables to be considered have been discussed, to a degree, above. Whether it is for a process that will replace a workable first generation method or to come up with the first process, the time factor will still play an important role. The approach is iterative and the potential for alternatives – chemical process selection rather than route selection – will allow for contingencies and help succeed in doing it right the first time at scale.

With any group of chemists, there will always be differences of opinion about the best way to prepare a compound. There should be no shortage of ideas. If possible, the medicinal chemists who have worked on the project should be included in the ideation process. This group can usually give insight into the robustness of the final step and whether this has the potential to be fixed.

With a number of feasible suggestions in hand, and thoughts about the pros and cons of the different approaches under discussion, how are the permutations prioritized? The project, company needs, and resources available will dictate the method used, be it formal or not. A general flow scheme outlines some of the selection processes (Figure 2.4).

Figure 2.4 Flow scheme for route selection.

The key to success is to look at a reasonable number of alternatives while not spending an inordinate amount of time finding the best route. The approaches can be grouped in a number of ways: For example, is there a key reaction in a number of approaches, such as an asymmetric reduction? Do a number of methods have formation of the same bond as a pivotal step? This is a simple way to reduce the number of ideas to workable groups.

Looking within these groups of reactions, a key transformation or transformations can usually be identified. If there is no close literature precedence, then these should be the subjects of ''killer experiments.'' In other words, the transformation needs to be tried with the real system or a close analogue to ensure that there will be success. If the chemistry is problematic, then the approach is dropped. It may be possible to return to this chemistry in a second-generation process. Routes that converge on a common intermediate should also be grouped especially if the downstream steps have a high chance of being performed successfully. The focus for the preliminary screen should be that the chemistry has a high probability of working. Of course, the medicinal chemistry and any subsequent processes that have been used have already demonstrated this attribute.

Once the general chemical strategies have been gathered, the next screening criteria are cost and time. In reality, not all ideas can be run through in the

laboratory. Some companies have sophisticated spreadsheets to perform this part of the process but some simple methods can also be employed. The following are the questions to answer:

- Does the chemistry work?
- Is there precedence for the reaction?
- If not, will a killer experiment be simple to run?
- Is there a showstopper?

If the answer to any of these questions is no, then that approach has to be given a low priority. Here, a ''showstopper'' is one that involves SHE or legal issues. Cost still has to be a key factor, and the key questions here are whether an expensive starting material or reagent has to be used. Some idea of the relative costs of the route ideas can be obtained by comparing the number of steps for each approach; longer sequences will cost more. The final factor to consider at this stage is whether specialized equipment is needed to perform a transformation or to handle a hazardous reagent or intermediate. If the process will not fit into already accessible plant equipment, there will be costs associated with obtaining that equipment or tolling that step out to a third party.

After this step, the groups of reactions should now have some degree of priority associated with them. It is worthwhile spending a little time at this stage to consider some of the factors that can make or break a good route: Can the final product and key intermediates be reworked? If all the intermediates in a proposed route are thick oils, then this could cause purification problems and it may be necessary to make a crystalline derivative so that purification can be achieved even if this will add extra steps. Compared with a route that provides solid intermediates, an approach that only involves liquid or oily intermediates will drop in priority. As a general rule, however, protection and oxidation–reduction sequences should be avoided, as these add two steps to a sequence. Although not crucial at this stage, it is still a good idea to look at the potential of telescoping steps. Just reversing the sequence of two steps in a sequence may provide this. One way to do this is to consider the type of solvent needed for each reaction. Put the reactions that need nonpolar solvents together, especially if the reactions that require polar solvents are related to chemistry on a different part of the molecule.

By now, the various ideas and groups will have found some sort of priority listing. The top candidates should have a high degree of chemical success and potentially be the lowest cost approaches. If insurmountable SELECT criteria have been identified with a specific route, or a group of approaches, then these should be dropped from consideration at this stage.

Now is the time to perform the killer experiments on the top priority ideas.

The final ranking process can now be undertaken. Again, a simple process will often suffice, especially if ideas have been grouped. A representative member of a group can be used to represent the whole group, and the best (shortest with the best chance of success) is often chosen.

Raw material costs will play a significant role in the overall cost of the process and, if a low-yielding step is anticipated, this can be factored in at this stage. This is different to a low chance of chemical success for a reaction far removed from any precedence.

The baseline is the route that has already worked in making the material. This has 100% probability of working; this is the medicinal chemistry route. From the yields of the individual steps, the cost of each intermediate can be determined. For simplicity, solvents and cheap reagents should be ignored. A weighting factor can be used if an expensive reagent of procedure is used, so that the costs for this step are multiplied by say 1.5. This approach also takes into account the length of the sequence. The various routes can then be compared and the cost savings compared to the route that works. The chance of chemical success then gives the indication about which candidates should be followed. It is not uncommon for a breakthrough idea with only a 10% chance of success providing only a 10% reduction in costs under optimistic conditions. Note that the chemical yield in the calculation for this breakthrough step might be used as 95% in the cost part of the calculation, while realistically it has only a 10% chance of working. The interplay between the chances of chemical success and potential cost savings is the decisive factor.

The available resources will determine how many routes can then be taken further. The assessment can also take into account groups of similar ideas; usually these will have a higher chance of chemical success. In a few cases, a proposed route may fall into two groups and, perhaps, has two breakthrough reactions. Alternative scenarios will only have one of these breakthrough reactions. As experimental results confirm the feasibility of one of these reactions, the ''double breakthrough'' will move up in the rankings if it has significant cost-reduction consequences.

Sufficient experiments should be run to determine the best route to follow. This is also the stage where IP, safety, environmental, and health issues need to be addressed. Unless these are simple to overcome, the route should be given a low priority.

The method outlined above allows routes to be compared. It does not reflect the overall cost of the final product. No allowance is made for going down the learning curve and for reductions in costs of raw materials, while costs such as plant time and utilities and waste treatment are also ignored.

A formalized method to rank possible routes has been used by AstraZeneca and is based on a Kepner–Tregoe decision analysis [41, 42]. These are outlined in Table 2.6 [43].

Whether a more formal process or a looser system is employed will depend on a specific company's circumstances. If the route-selection process involves many experts from different backgrounds and disciplines from multiple sites, then a more formalized approach is required. If the exercise involves a small group at the same site, then a less formal approach will be quicker and just as productive.

Table 2.6 Route-selection criteria [43].

Criteria	Explanation
Accommodation	Are there any steps that would be difficult to accommodate?
Back-ups	Is the route applicable to any backup compounds?
Chemical feasibility	How likely are the proposed reactions to work?
Chiral integrity	How well will any chirality survive transformations in the route?
Chirality	What is the enantiomeric excess of any introduced chiral centers?
Convergence	How convergent is the route?
Cost of goods	What is the cost of goods for the route?
Effluent	What is nature of the effluent cost of disposal?
Environment	Do any of the steps on the route pose a significant environmental hazard?
Flexibility	Will the route allow delivery of different compounds if the choice has not been narrowed down to one?
Health	Do any of the steps on the route pose a significant health hazard?
Intellectual property	Are there any intellectual property issues or opportunities?
Meets existing API specification	Will the route afford material that meets the existing API specification?
Number of steps	How many chemical steps does the route contain?
Number of steps to key step	How much time/effort is required to investigate the key step on a route?
Potential genotoxic impurities	Are there any issues with potential genotoxic impurities on the route?
Potential yield (overall/individual step)	What is the potential yield of individual stages? What is the potential overall yield? (Data can be updated as experimental work is completed.)
Purification points	How many, and where, are the purification points on the route?
Raw material availability	Can the required raw materials be sourced in bulk?
Robustness	Are the chemical transformations robust?
Safety	Are all the transformations on the route safe to operate?
Solubility of intermediates	Are there any issues with the solubility of any intermediates?
Throughput	What is the throughput of the route?

For a less formal process, factors such as cost, timing, and SELECT still need to be considered. Chemical feasibility must still be at the top of the list. The key questions to address are the following:

- Will the chemistry supply the appropriate amount of material within quality criteria and at acceptable cost?
- Are SHE and legal requirements met?

2.4.1
Sildenafil

An excellent example of a change in route between medicinal chemistry and commercial is provided by sildenafil (Viagra®) (**4**) [44]. The medicinal chemistry route is linear (Scheme 2.3) but was good enough to provide early development quantities. However, the use of the sulfonyl chloride **9** in the final bond-forming reaction is problematic as it is a potentially toxic material; multiple crystallizations were required to ensure purity. In addition, this sulfonyl chloride is hydrolytically unstable and any losses due to this unwanted reaction are expensive at such a late stage in the synthesis. The formation of **9** at such a late stage and with a relatively high molecular weight requires larger quench volumes.

The overall yield from **5** to the product **4** was 7.5%. The key finding that allowed for a route change was the cyclization to convert the amide **7** to the pyrimidinone **8**. The use of an aqueous system resulted in the concurrent formation of the acid. However, use of anhydrous conditions, such as KOBu-*t* in *t*-BuOH gave a quantitative conversion with no impurities being detected. This observation allowed for a reordering of the steps so that the potentially toxic material is handled earlier in the synthesis, and the sequence becomes more convergent. The commercial route is outlined in Scheme 2.4. The use of solvents was also greatly reduced.

Scheme 2.3

Scheme 2.4

Scheme 2.5

The amine **10** is prepared by a palladium-catalyzed reduction of the nitro compound **6** rather than the tin reduction. The sulfonamide **11** was prepared as outlined in Scheme 2.5. The overall yield of **4** from **5** was increased to 75.8%.

2.5
Process Selection

So far, we have been concerned with route selection – how to get from available starting materials to the final product. The details now need to be put onto the plan. For example, if an aromatic coupling reaction is needed, which reaction should be used, Heck, Suzuki, or Stille? This is where grouping of reactions can help as the conversion of one intermediate to another may be similar. Some experimentation may be necessary. Even the order of steps can play a role; is it better to do an enone reduction before hydrolyzing an ester or *vice versa*? The problematic steps need to be addressed at this time. A good process group will not ignore potential problems, as these can grow to an insurmountable size later when it is expensive both in terms of cost and time to fix.

It is not uncommon for the unexpected to happen; Murphy's Law will be obeyed. The earlier these problems are addressed, the more options are open to achieve a solution. The final chemical process needs to be understood (see Chapter 4).

The selection process is an iterative one. The initial preparation provides information for a more informed route-scouting exercise. Experimentation leads to the preferred route. Implementation both in the laboratory and to make material will show up shortcomings and problems that then need to be addressed, perhaps by changing the method of performing one step, such as changing a reduction to a catalytic hydrogenation rather than using sodium borohydride. This is one of the major reasons why the final step in the overall sequence should be fixed and understood as soon as possible, as this allows some leeway to perform these earlier step changes.

It is during this process stage that the groups of reactions can be ungrouped and thought of as individual alternatives and put through the selection criteria once more. In most cases, some experimental data will need to be obtained to determine which is the lowest cost option available. For the example above concerning the Heck, Suzuki, and Stille reactions for a coupling reaction, preparation of the substrates for these reactions will be different and so will have cost and raw material differentiators. The use of tin in the Stille method is a SHE flag. For another example, consider putting the three components, A, B, and C together, with A and B being joined by an amide bond and B and C by an ester bond to give the final product A−B−C. If there are other functional groups present in the component parts, the paper exercise may have already led to a single approach through reagent and functional group compatibility issues. If not, then the experiments need to be run.

The process selection exercise refines some of the variables already considered. Solvent usage is examined in two consecutive steps to see if there is a potential for telescoping, for example, at this stage, reagents are examined and the best ones chosen. The aim is to define the process so that thermochemical, impurity, solubility, and other important data can be collected. Compared with the route scouting exercise, which can, for the most part, be considered a paper one, early process selection depends on experimental results. In some cases, a potential option may have to be dropped because there is not enough time or resources available to evaluate it.

When a process is running, costs can be reduced by going down the learning curve; operators are more familiar with the operations; scheduling limits downtimes and maximizes use of equipment and raw material, and hence reagent prices can decrease. If process improvement is not undertaken, then management may not appreciate the potential cost benefits of the exercise and, if volumes increase and a bottleneck is seen in the process, can concentrate on solving the wrong problem [3]. Once a process has been selected, options still need to be kept open and modifications made. In many companies, there is a reluctance to go back in the process development sequence. However, as the process is run for a period of time, problems that have not been noticed during the transfer and early runs will become apparent. Small fixes may solve these problems, but again a wider vision should

also be used. Why fix the current problem if an alternative and better method circumvents it with better yield or throughput?

2.5.1
Pregabalin

The development of the anticonvulsant CI-1008 (**12**) illustrates some of the points made above [45]. Route selection was primarily based on the low-cost manufacturing process based on ''ideal process'' cost projections. Four routes were evaluated in the laboratory and, of these, two were scaled up in the pilot plant to result in the selection of a resolution method.

The discovery method was based on the use of chiral auxiliary chemistry and had issues with low temperature reactions, chromatography, side reactions, and a low overall yield. However, this approach was modified to provide the initial few kilograms of material. The methodology was modified again to use the *tert*-butyl ester rather than benzyl; the reduction method was improved to avoid the smell associated with the use of $BH_3 \cdot SMe_2$, and the final ester hydrolysis and reduction was changed to help ease of isolation (Scheme 2.6).

Although there was potential for recycle of the chiral auxiliary, the improved route of Scheme 2.6 still did not meet the cost goals. Other routes were investigated. One started from the cheap chiral starting material, L-leucine. However, the sequence was long, nine steps, and was not developed past the proof of principle phase. A similar approach with the same number of steps from isobutyraldehyde involved a Stobbe condensation and a resolution.

Enzymatic resolution provided a much shorter sequence (Scheme 2.7). However, the enzyme is animal derived and, thus, not allowed for use in pharmaceutical applications owing to the potential of harmful by-products such as prions being present. In addition, pig liver esterase (PLE) has ethnic implications. The last step is from analogy with a route using a classical resolution approach of the racemic amide acid (**rac-13**) with α-phenylethylamine.

Scheme 2.6

Scheme 2.7

Scheme 2.8

The knowledge gained from the previous studies prompted the use of malonate chemistry and a screen to find a resolution agent (Scheme 2.8). The two chemicals resolution approaches were almost comparable in cost, but the method of Scheme 2.8 was chosen, as it did not use chlorinated solvents.

The routes were compared with the assumptions of 100% yields, no labor or overhead costs, no waste disposal costs, and bulk prices for raw materials. The oxazolidinone route (Scheme 2.6) was 12.2 times more expensive than the chemical resolution routes (*cf.* Scheme 2.8). Use of L-leucine was 6.1 times more expensive, the Stobbe condensation 2.2 times, and the malonate route of Scheme 2.7 1.5 times, again compared to Scheme 2.8.

Scheme 2.9

Scheme 2.10

This is an excellent example of adaptation to meet needs, as the compound **12**, Pregabalin, is now on the market as Lyrica®. The process has not stood still through the commercialization process. Two approaches were considered for the manufacturing process: The first used an asymmetric hydrogenation as the key step (Scheme 2.9). To avoid complications from the use of proprietary ligands, Trichickenfootphos (TCFP) was developed [46]. Although catalyst usage is low (substrate : catalyst = 27 000 : 1), an enzymatic method is now employed (Scheme 2.10, *cf.* Scheme 2.7) [47]. The keys to the success of this approach are the use of a low-cost enzyme and the ability to easily separate the product and racemize the undesired isomer; there is also a simple, cheap entry to the cyano diester starting material (*cf.* Scheme 2.8) [47, 48].

2.5.2
NK1 Receptor Antagonist

The modification of a route from the one initially used to prepare a few grams (Scheme 2.11) is provided by GW-597599 (**14**), an NK1 receptor antagonist. As with many of these medicinal chemistry or small-scale approaches, there are a number of issues. In the first step, use of low temperature is required to minimize overreaction of the Grignard reagent to give a tertiary alcohol. For the imine reduction to give **15**,

Scheme 2.11

high pressures and long reaction times are necessary. Safety concerns are present with the use of borane and triphosgene. Finally, the resolution is performed at the end of the sequence after the diastereoisomers have been separated [49].

The low yield for the isomer separation can be avoided by preparing the two components **15** and **16** by asymmetric synthesis. The convergent approach (Scheme 2.12) is also enhanced if the amide reduction of **17** is performed prior to the coupling of the components [49].

Optimization of the reaction steps, including telescoping, as well as a dynamic kinetic resolution method provided an efficient method to the amide **17** as the

Scheme 2.12

Scheme 2.13

mandelate salt (Scheme 2.13) [49]. The borane reduction was modified to generate the reductant *in situ* [50].

2.5.3
Data

During the route scouting and the process selection experiments, a considerable amount of knowledge about individual steps and the overall process is gathered. These data can be useful (see Chapter 4). There are a large number of techniques available to collect data about various reaction parameters [51].

Obviously, impurities coming from a step and how they go through subsequent steps can have significant impact on the final steps, especially if they are difficult to remove.

Outside of safety information, there is a tendency to gather too much data in the early stages of process research and development. What question needs to be answered? In many cases, the primary need is to understand the cause of a low yield or why an impurity is forming. Design experiments to answer the specific question. With the move to in-process analysis, such an approach should be high on the solution list. It is easy to get bogged down collecting information about a specific reaction while forgetting the overall picture or knowing why the data is being collected. In some cases, a later step may be able to remove an impurity if it is below 1%, so time need not be taken to get it below 0.1% in the reaction where it is formed. The overall sequence has to be optimized, not a specific step.

A program should be designed so that experiments answer a specific question associated with the immediate decision-making process. In other words, they should be on or close to the critical path. If the data are useful later on for process optimization, then the experiment is not in the ''critical'' category and can be run as the process development progresses. If additional data can, however, be collected during these killer experiments, then its value should not be overlooked as long as it is obtained for specific reasons.

2.6
Summary

Overall, safety has to be the overriding factor for route and process selection. This not only includes the workers who are to be intimately involved in the execution of the method in the pilot plant and manufacturing plants but also the end-users and the patients who will see safety through the quality of the product. Thus, although speed, cost, and quality interplay, quality has to be the most important factor for drug process development. The end product ends up in humans and cannot be compromised. As understanding of the process, the impurities formed, and the properties of the molecule develops, specifications can be set and tightened as necessary. The quality aspect can impact the process. For example, elegant syntheses are short with a minimal number of steps, and telescoping can be useful in reducing the overall number of unit operations. However, if a problematic impurity cannot be removed during the sequence, or by-product formation is outside of tolerable limits, a purification step may be needed to ensure quality control of the API. Once again, a compromise may have to be used. However, in the route-selection process, paying attention to low-yielding reactions, potentially toxic by-product formation, or even just looking at a long sequence where oils are the products can pay dividends in time and money by addressing these problems early.

The key factors driving the route selection decision are timing, cost, SHE, legal, and, above all, a chemistry fit. Risk management gives a large weighting to

chemistries that work and provide precedence. Time and costs have to be traded, and this can change as a candidate moves down the drug development pipeline. SHE factors cannot be compromised, but compliance may involve significant costs. Once a number of potential routes fulfill the necessary criteria, a few selective experiments can then show the best option to follow. Process selection will refine the approach and provide more specific parameters.

Route and process selection are iterative exercises and should be continued throughout the life cycle of a product. If the candidate makes it to being a blockbuster drug, you can be certain that others are doing the same exercise anticipating when the compound would become generic.

References

1. Brummel, R.G. (2002) in *From Bench to Pilot Plant* (eds M. Nafissi, J.A. Ragan, and K.M. DeVries), American Chemical Society, Washington, DC, p. 61.
2. Petousis, N.H. (2002) in *From Bench to Pilot Plant* (eds M. Nafissi, J.A. Ragan, and K.M. DeVries), American Chemical Society, Washington, DC, p. 1.
3. Pisano, G.P. (1997) *The Development Factory: Unlocking the Potential of Process Innovation*, Chapter 4, Harvard Business School, Boston, MA.
4. Federsel, H.-J. (2009) *Acc. Chem. Res.*, **42**, 671.
5. Pisano, G.P. (1997) *The Development Factory: Unlocking the Potential of Process Innovation*, Harvard Business School, Boston, MA.
6. Anderson, N.G. (2000) *Practical Process Research and Development*, Academic Press, San Diego, CA.
7. Butters, M., Catterick, D., Craig, A., Curzons, A., Dale, D., Gillmore, A., Green, S.P., Marziano, I., Sherlock, J.-P., and White, W. (2006) *Chem. Rev.*, **106**, 3002.
8. Adler, C., Brunner, J., Fichtner, C., Küng, P., Levis, M.K., Ruchtl, H.-R., Sjöberg, A., and Weber, B. (2006) *Chimia*, **60**, 523.
9. Pisano, G.P. (1997) *The Development Factory: Unlocking the Potential of Process Innovation*, Chapter 7, Harvard Business School, Boston, MA.
10. Ikemoto, T., Ito, T., Hashimoto, H., Kawarasaki, T., Nishiguchi, A., Mitsudera, H., Wakimasu, M., and Tomimatsu, K. (2000) *Org. Proc. Res. Dev.*, **4**, 520.
11. Thayer, A.M. (2009) *Chem. Eng. News*, **87** (11), 17.
12. Braune, S., Pöchlauer, P., Reintjens, R., Steinhofer, S., Winter, M., Lobet, O., Duidat, R., Woehl, P., and Guermeur, C. (2008) *Chim. Oggi*, **26** (5), 1.
13. Wiles, C. and Watts, P. (2008) *Eur. J. Org. Chem.*, 1655.
14. Zhang, X., Stefanick, S., and Villani, F.J. (2004) *Org. Proc. Res. Dev.*, **8**, 455.
15. Markowz, G., Schirrmeister, S., Albrecht, J., Becker, F., Schütte, R., Caspary, K.J., and Klemm, E. (2005) *Chem. Eng. Technol.*, **28**, 459.
16. Palaniappan, C., Srinivasan, R., and Tan, R. (2002) *Ind. Eng. Chem. Res.*, **41**, 6698.
17. Palaniappan, C., Srinivasan, R., and Tan, R.B. (2002) *Ind. Eng. Chem. Res.*, **41**, 6711.
18. Anastas, P.T. and Eghbali, N. (2010) *Chem. Soc. Rev.*, **39**, 301.
19. Anastas, P.T. and Kirchhoff, M.M. (2002) *Acc. Chem. Res.*, **35**, 686.
20. Winterton, N. (2001) *Green Chem.*, **3**, G73.
21. Tang, S.L.Y., Smith, R.L., and Poliakoff, M. (2005) *Green Chem.*, **7**, 761.
22. Tucker, J.L. (2006) *Org. Proc. Res. Dev.*, **10**, 315.
23. Anastas, P.T. and Zimmerman, J.B. (2003) *Environ. Sci. Technol.*, **5**, 94A.
24. Trost, B.M. (2002) *Acc. Chem. Res.*, **35**, 695.
25. Trost, B.M. (1991) *Science*, **254**, 1471.

26. Sheldon, R.A. (2008) *Chem. Commun.*, **29**, 3352.

27. Andrews, J. (2005) *Org. Proc. Res. Dev.*, **9**, 149.

28. Eissen, M. and Metzger, J.O. (2002) *Chem.Eur.J.*, **8**, 3581.

29. Trost, B.M. (1995) *Angew. Chem., Int. Ed. Engl.*, **34**, 259.

30. Constable, D.J.C., Dunn, P.J., Hayler, J.D., Humphrey, G.R., Leazer, J.I. Jr., Lindermann, R.J., Lorenz, K., Manley, J., Pearlman, B.A., Wells, A., Zaks, A., and Zhang, T.Y. (2007) *Green Chem.*, **9**, 411.

31. Panke, S. and Wubbolts, M. (2005) *Curr. Opin. Chem. Biol.*, **9**, 188.

32. Schultze, B. and Wubbolts, M.G. (1999) *Curr. Opin. Biotechnol.*, **10**, 609.

33. Shaeri, J., Wohlgemuth, R., and Woodley, J.M. (2006) *Org. Proc. Res. Dev.*, **10**, 605.

34. Bray, B. (2003) *Nat. Rev. Drug Discov.*, **2**, 587.

35. Wang, C.-C., Lee, J.-C., Luo, S.-Y., Kulkarni, S.S., Huang, Y.-W., Lee, C.-C., Chang, K.-L., and Hung, S.-C. (2007) *Nature*, **446**, 896.

36. Verma, S. and Eckstein, F. (1998) *Ann. Rev. Biochem.*, **67**, 99.

37. Atherton, J.H., Double, J.M., and Gourlay, B. (2005) Survey of PI equipment and requirements in the fine chemicals and pharmaceuticals sector. 7th Annual World Congress of Chemical Engineering, 11–14 July, Glasgow.

38. Malhotra, G. (2005) Pharmaceutical Manufacturing, July/August, 50, *http://www.pharmamanufacturing.com/articles/2005/294.html* (accessed 5 Febuary 2011).

39. Newhouse, T., Baran, P.S., and Hoffmann, R.W. (2009) *Chem. Soc. Rev.*, **38**, 3010.

40. Shu, C., Zeng, X., Hao, M.-H., Wei, X., Yee, N.K., Busacca, C.A., Han, Z., Farina, V., and Senanayake, C.H. (2008) *Org. Lett.*, **10**, 1303.

41. Moseley, J.D., Brown, D., Firkin, C.R., Jenkin, S.L., Patel, B., and Snape, E.W. (2008) *Org. Proc. Res. Dev.*, **12**, 1044.

42. Parker, J.S., Bower, J.F., Murray, P.M., Patel, B., and Talavera, P. (2008) *Org. Proc. Res. Dev.*, **12**, 1060.

43. Parker, J.S. and Moseley, J.D. (2008) *Org. Proc. Res. Dev.*, **12**, 1041.

44. Dale, D.J., Dunn, P.J., Golihtly, C., Hughes, M.L., Levett, P.C., Pearce, A.K., Searle, P.M., Ward, G., and Wood, A.S. (2000) *Org. Proc. Res. Dev.*, **4**, 17.

45. Hoekstra, M.S., Sobieray, D.M., Schwindt, M.A., Mulhern, T.A., Grote, T.M., Huckabee, B.K., Hendrickson, V.S., Franklin, L.C., Granger, E.J., and Karrick, G.L. (1997) *Org. Proc. Res. Dev.*, **1**, 26.

46. Hoge, G., Wu, H.-P., Kissel, W.S., Pflum, D.A., Greene, D.J., and Bao, J. (2004) *J. Am. Chem. Soc.*, **126**, 5966.

47. Martinez, C.A., Hu, S., Dumond, Y., Tao, J., Kelleher, P., and Tully, L. (2008) *Org. Proc. Res. Dev.*, **12**, 392.

48. Hu, S., Martinez, C.A., Tao, J., Tully, N.E., Kelleher, P.G.T., and Dumond, Y.R. (2006) PCT Patent Appl. WO 2006000904.

49. Guercio, G., Bacchi, S., Goodyear, M., Carangio, F., Tinazzi, F., and Curti, S. (2008) *Org. Proc. Res. Dev.*, **12**, 1188.

50. Guercio, G., Manzo, A.M., Goodyear, M., Bacchi, S., Curti, S., and Provera, S. (2009) *Org. Proc. Res. Dev.*, **13**, 489.

51. Rubin, A.E., Tummala, S., Both, D.A., Wang, C., and Delaney, E.J. (2006) *Chem. Rev.*, **106**, 2794.

3
Critical Stages of Safety Assessment in Process Design and Scale-Up

Stephen Rowe

3.1
Reaction Safety Concepts

Explosion hazards associated with processes are a major consideration when seeking to scale-up. The criticality of assessing these risks during early development cannot be understated. This chapter deals with the risks posed by runaway exothermic or gas-generating reactions, decomposition of unstable substances and gas, and vapor and dust explosion hazards, and how they are assessed. Occupational safety and general health and safety hazards are not considered in this chapter.

A number of decisions have to be made during route selection and early development, which have a huge bearing on the ultimate process risk. A number of simple concepts are outlined to assist readers in selecting safer routes and developing safer processes. The use of prediction techniques is introduced, followed by laboratory tests used to quantify reaction hazards. Importantly, guidance is provided on how to interpret the data and make decisions that will make the process safer. For scale-up to larger scales, the process of hazard identification is briefly introduced with prevention and protection strategies identified. The chapter concludes with a summary of flammability issues including data requirements associated with various methods for operating safely.

The critical element of safety evaluation is that it must be an integral part of the development process and not regarded as an ''add-on.'' The examination of safety concepts should begin right at the beginning of the development process during route selection – this is where hazards with proposed routes can be identified and avoided most easily and cost-effectively. At every subsequent step of development, there are further decisions that will have a bearing on the residual risk of the process on scale-up and production. Developing a process and then retrospectively conducting a hazard assessment may be extremely expensive and highly undesirable. It would be madness to fully develop a process and only then check the quality of the product after development is complete – it is the same with safety aspects.

We start by looking at reaction hazards and the hazards posed by unstable substances.

3.1.1
What Is the Hazard?

During my early years in safety, one of my learned colleagues had a mantra that still rings true – ''A little bit of heat never hurt anybody. Pressure's the problem.'' Much of our work in assessing the hazard of a reaction involves quantifying heat effects but it is pressure effects that will ultimately cause loss of containment – albeit, often caused by heat generation.

During the course of a reaction, pressure can arise from two sources:

- **Gas generation** (for example, a reaction that generates a permanent gas such as carbon dioxide or nitrogen)
- **Vapor pressure generation** (for example, where the temperature of a reaction mixture exceeds its atmospheric boiling point in a closed reactor).

Differentiating between these two effects becomes critically important in the sizing of emergency relief systems. Vapor pressure reactions can be ''tempered'' by the loss of solvent during the relief process where the vented vapor takes with it the latent heat of vaporization. A balance of heat generation versus heat loss prevents further escalation of temperature, and hence pressure. For gas-generating reactions, there is no latent heat loss, only sensible heat loss, and hence no control of reaction temperature during the venting process.

Chemical reaction hazard assessment should ensure that all potential causes of overpressurization are known, the consequence understood (that is, to say, quantified), and either prevented from occurring or protected against. Gas generation or vapor pressure effects can result from the normal process, a deviation from the normal process, side reactions, or thermal decomposition of unstable species (starting materials, products, or intermediates).

Gas generation, from the normal process or foreseeable deviation, presents an immediate and obvious potential for pressurization in an inadequately vented or sealed reactor. If such behavior is present, it must be protected against either through adequately designed containment or the presence of an *adequately sized* relief system. Quantification of the amount and rate of gas generation will normally be required in such circumstances, typically employing adiabatic calorimetric techniques.

3.1.2
The Critical Effects of Scale-Up on Thermal Behavior

When a reaction is scaled up from the laboratory scale to the industrial scale, there are two important changes that must be considered in the assessment of thermal data. Heat generated within a reactor is distributed in three ways (see Figure 3.1):

- Heat retained by the reaction mass, increasing the batch temperature.
- Heat consumed in heating the reactor (to achieve equilibrium with the reaction mass). The combination of heat distribution between the batch and the reactor is often quantified in a term known as the *phi factor*.
- Heat lost to the environment by radiation from the outer walls of the reactor.

Heat retained in the reaction mass | Heat distributed to the reactor | Heat losses to the environment

Figure 3.1 Heat distribution from an exothermic chemical reaction.

Typically, as the scale of reaction increases, the heat consumed in heating the reactor and the heat losses to the environment will decrease in proportion to that consumed in heating the reactor contents because of the relative lowering of the surface area to volume ratio. Thus, at larger scales, a higher proportion of the heat stays within the reaction mass, thereby causing the attainment of higher temperatures. Further discussion and quantification of heat distribution can be found in Barton and Rogers [1].

An exothermic reaction will be "out of control" or in a "thermal runaway" situation when the rate of heat generation exceeds the rate of heat removal (either through forced cooling or atmospheric heat losses). Since most exothermic reactions will follow Arrhenius kinetics, increasing the temperature will cause exponential acceleration of the reaction rate and hence of the heat-generation rate. In contrast to this, cooling or atmospheric heat losses will usually only increase linearly as a function of the global heat transfer coefficient (typically constant) and the temperature difference between the reaction mass and surroundings. Figure 3.2 demonstrates the balance between heat generation and heat losses for an exothermic batch reaction.

Figure 3.2 Heat removal and heat loss for a batch exothermic reaction.

Effective simulation of runaway reactions can only be achieved on the basis of data that mimics the industrial-scale failure case. That is to say, when attempting to simulate a large-scale loss of cooling scenario in laboratory scale equipment, the heat losses and thermal efficiency of the equipment used should mimic the large scale. Typically, this requires the use of adiabatic (zero heat loss) calorimetry using low thermal inertia (low phi factor) reactors. Apparatus that is not representative of plant-scale heat loss and thermal inertia will require mathematical correction or application of a safety margin.

3.1.3
Safety Features of a Reaction

The thermal safety of a reaction can be crudely, but very effectively, described by knowledge of four key temperatures, and their relation to one another (described in more detail in Stoessel [2]). These are as follows:

- Normal process temperature (T_P)
- Maximum temperature of the synthesis reaction (MTSR)
- Maximum temperature for technical reasons (MTT)
- The onset temperature of decomposition or secondary reaction (T_{dec}).

The MTSR is the maximum temperature that could be achieved if the exothermic reaction occurs with no heat loss (for example, in the event of a complete loss of cooling). This is the sum of the normal process temperature and the adiabatic temperature rise of the reaction (ΔT_{ad}). The adiabatic temperature rise is given by

$$\Delta H_r \cdot N = m \times Cp \times \Delta T_{ad} \tag{3.1}$$

where
ΔH_r = Overall heat of reaction (kJ mol^{-1} of limiting reactant)
N = Number of moles of the limiting reactant (mol)
m = Mass of the entire reaction mixture (kg)
Cp = Heat capacity of the reaction mixture (kJ kg^{-1} K^{-1})
ΔT_{ad} = Adiabatic temperature rise (K)

The ΔT_{ad} assumes that there are no secondary reactions or side reactions at elevated temperatures and that there is zero heat loss or removal from the reaction mass.

The MTT is the temperature at which the vapor pressure of the mixture poses a potential risk to the integrity of the vessel. For an open reactor (or low pressure vessel), the MTT would be taken as the boiling point of the mixture. For a closed pressure vessel, the MTT would typically be the temperature equivalent to either the relief device set pressure or, less conservatively, the design pressure of the reactor. It is conservative, and easier, to define the MTT as the boiling point of the mixture at atmospheric pressure.

The T_{dec} is the onset temperature of secondary or decomposition reactions. This data can be determined through thermal stability screening tests or more

sensitive adiabatic calorimetry. If thermal screening tests are used, then it would be necessary to correct the measured onset temperature of a reaction using an appropriate and conservative safety margin. This is necessary to account for the nature of the test (ramped, heat–wait–search, or isothermal) and the sensitivity of the test method. It is common, for example, to find safety margins of up to $100\,^{\circ}C$ applied to measured onset temperatures from differential scanning calorimetry (DSC) tests. If more sensitive adiabatic techniques are used to determine the onset temperature of decomposition, then smaller margins can be applied. In such cases, it is possible to calculate the temperature from which it takes 24 h for the reaction to reach its maximum rate (referred to as T_{D24}), assuming that cooling is lost in a large-scale reactor. This value can be used directly as T_{dec}.

Using these four conceptually simple temperatures, the relative thermal risk of a process can be classified into one of five "Criticality Classes" (see Figure 3.3).

The Criticality Classes are based on thermal effects only and do not consider gas generation directly. As previously noted, gas generation at any stage of the process must be quantified and accounted for in the safe operation of the process.

In terms of thermal effects, Criticality Class 1 is an inherently safe position whereby the exothermic reaction generates insufficient temperature rise to reach conditions necessary for boiling or decomposition. In Class 1, even overheating of the mixture by an external heat source will invoke boiling in an adequately vented vessel before decomposition, thereby providing a thermal barrier to initiating decomposition. In this case, only gas generation from the normal reaction could pose a risk to reactor integrity. Class 2 is a slightly increased risk but only by overheating, and, in this case, decomposition will occur prior to attaining boiling.

Criticality Classes 1 and 2 are inherently safe (in the absence of gas generation) and require no technical safety systems under normal conditions. However, a thorough hazard assessment of such processes should still be performed as a single, foreseeable process deviation may be capable of causing a shift in the

Figure 3.3 Risk ranking of reactions using Criticality Classes.

Criticality Class. For example, omission of the solvent charge in a process could cause a severe increase in the potential adiabatic temperature owing to a decrease in the heat capacity of the vessel contents, decrease in the onset of thermal decomposition, and elevation of the boiling simultaneously. This could conceivably result in a normal Class 1 reaction shifting, in one step, to Class 5.

Reactions of Class 3 or 4 require provision of procedural or, more reliably, technical safety measures. In these classes, the normal reaction combined with cooling failure presents a pressurization risk. Protection or prevention systems are required to either prevent these scenarios from manifesting or provide measures to protect the reactor against the consequences. Reactions of Class 5 should be avoided and modifications to the process considered if encountered.

The Criticality Class concept can be taken much further with mathematical criteria applied to severity and probability (not all decompositions may be catastrophic if the energy is low and there is no gas or volatile formation). The overall risk, the product of severity and probability, will dictate the level of detail required in the classification process. Nevertheless, the simple concept of ranking the four key temperatures can be utilized at all levels of safety evaluation – including during route selection and initial research and development.

Having noted the nature of the hazard, the normally adverse effects of scaling up on heat removal, heat loss, and the critical reaction safety characteristics of a process, it is necessary to have a rigorous assessment procedure in place. This is necessary to ensure that all reactions are designed and operated with a robust ''Basis of Safety.'' Basis of Safety is the collection of appropriate safety measures (both organizational and technical) that has the ultimate objective of preventing overpressurization in the first place. This is frequently impossible and so the Basis of Safety may include protection methods, such as emergency relief venting, to mitigate the consequences of loss of control of a process.

The remainder of this chapter deals with specific safety decisions that should be made at each stage of R&D, process development, and operation. Fundamentally, the emphasis should be placed on developing processes that are relatively robust and fault tolerant (Criticality Classes 1 and 2). This is an inherently safer approach than leaving safety decisions until the process is fixed and ready for scale-up.

3.1.4
Stages of Safety Assessment

The assessment of reaction safety should be an integral part of the process lifecycle – commencing when process development commences. Once the process is through development, the hazards are intrinsic – built-in. Good development practices will result in a process that is more inherently safe by design. The stages of safety assessment linked with the development lifecycle are schematically represented in Figure 3.4.

The decisions that can be made, and the investigations conducted, at each stage of development are discussed in the following sections. The safety of the plant process will depend on each phase being performed competently and thoroughly.

| Chemical route selection | • Heat of reaction prediction
• Adiabatic temperature rise prediction
• MTSR estimation | • Unstable functional group analysis |

| Process development/ optimisation | Normal process
• Reaction heat measurement
• Gas generation quantification
• Adiabatic temperature rise calculation | Thermal stability
• Explosivity hazard assessment
• Preliminary thermal stability screening
• Definition of safe process temperatures |

| Pilot scale production | Basis of safety : pilot scale
• Identification of hazardous deviations
• Adiabatic calorimetry on deviations
• Definition of the basis of safety |

| Large scale production | Basis of safety : production
• Identification of hazardous deviations
• Adiabatic calorimetry on deviations
• Definition of the basis of safety |

Figure 3.4 Stages of safety assessment.

Deficiency in any area is likely to compromise the overall effectiveness of the study. Many case histories of industrial incidents bear witness to this fact.

3.2
Pre-Laboratory Safety Studies

In the initial route identification and development phases, critical decisions are made that dictate the risk of the resulting plant process. Safety is a critical component of the decision-making processes in route, process method, and condition selection. Most hazards can be identified prior to laboratory testing so this is the stage where safety-related decisions define the level of intrinsic process risk. Poor decision making – or lack of safety consideration at this stage – will mean that intrinsic hazards remain a burden throughout the lifecycle of the process. A high-risk process may require extensive safety features, which have a significant adverse effect on process economics. The early identification of hazards requires a relatively small amount of effort without time-consuming and expensive laboratory work. It is therefore an economically sensible approach to pursue.

3.2.1
Predicting Reaction Safety Characteristics

The following safety characteristics of a reaction can be predicted on the basis of a balanced chemical reaction equation for the desired reaction(s) and known side reactions:

- Heat of reaction (ΔH_r) – which permits determination of ΔT_{ad} and MTSR
- Potential for permanent gas generation – obvious from identifying gaseous products in the balanced equations

• Identification of any energetic (potentially explosive) molecules.

Reaction thermodynamics can normally be readily predicted. Fundamentally, simple bond energy calculations (available in open literature) can be performed to estimate the heat of reaction (Hess's method). This can be refined by using the heats of formation (ΔH_f) of the products and reactants.

Hess's method	$\Delta H_r = \sum(\text{energy of bonds broken}) - \sum(\text{energy of bonds made})$
ΔH_f method	$\Delta H_r = \sum(\Delta H_f \text{ products}) - \sum(\Delta H_f \text{ reactants})$

Given that heats of formation are largely unavailable for novel molecules, prediction methods may be used for estimation such as the CHETAH [3] computer program developed by the ASTM International. The program utilizes Benson's method of group contributions and facilitates heat-of-reaction calculation from its predicted heats of formation for products and reactants. Despite some limitations (regarding predictions for salts, the absence of some functional groups, and the use of only gas-phase data) CHETAH provides a useful preliminary tool in predicting the heat of reaction based solely on chemical structures.

The relative accuracy of the various estimation methods, compared with reaction calorimetry measurement, for the esterification reaction between methanol and acetic anhydride can be gauged from Table 3.1. In this case, the CHETAH data is based on gas-phase thermochemistry data, whereas the experimental and heat of formation predictions are based on liquid-phase data.

Typical heats of reaction can range from $-70\,\text{kJ}\,\text{mol}^{-1}$ for a typical mineral acid/base reaction to $-500\,\text{kJ}\,\text{mol}^{-1}$ for an aromatic nitro-group reduction.

Once the heat of reaction has been estimated, it can be readily converted into a theoretical adiabatic temperature rise (ΔT_{ad}) – this is the temperature rise that will occur if the reaction is performed without heat loss, assuming that there are

Table 3.1 Heat of reaction data for methanol/acetic anhydride esterification.

Method	ΔH_r (kJ mol^{-1})	References
Heat of formation data	−75.8	ΔH_f data from NIST [4] for liquid-phase reactants and products
CHETAH prediction	−83.0	[3]
Reaction calorimetry	−67.0	Measured using Mettler Toledo RC1*e* reaction calorimeter

no secondary or side reactions initiated at the elevated temperature. From this, the theoretical MTSR is readily computed.

In addition to the hazards that may be posed by the desired chemical reaction and potential side reactions, the possibility exists that one or more of the process materials may contain inherently unstable functional groups. In extreme cases, explosive properties can be associated with such groups, which can have major implications for handling, processing, storage, and transport. If such unstable groups are present, they will impart an instability hazard to the process and will require maximum permissible handling temperatures (T_{dec}) to be defined and avoided.

Early identification of such substances is important for several reasons:

- If identified early enough, consideration can be given to changing the route or materials to exclude highly energetic functional groups.
- If they cannot be excluded, it is essential that small-scale hazard studies, possibly including formal explosivity testing, be undertaken at an early stage to indicate the magnitude of the hazard. In any case, precautions can be specified for synthesis, which is designed to mitigate any such issues.

Potentially energetic functional groups can usually be readily identified. A selection of the most commonly encountered energetic functional groups is provided in Table 3.2 together with the typical range of decomposition energies associated with the groups (using data from Ref. [5]).

Table 3.2 Commonly encountered energetic functional groups.

Name/structure	Range of decomposition energies (kJ mol^{-1})
Alkenes ($R_2C{=}CR_2$)	50–90
Alkynes/acetylenes (R–C≡C–R)	120–170
Epoxides	70–100
Organic/inorganic peroxides/hydroperoxides (R–O–O–R/R–O–O–H)	230–360
Organic sulfoxides ($R_2S{=}O$)	40–70
Organic sulfonyl chlorides (R–SO_2Cl)	50–70
Hydrazines (R–NH–NH–R)	70–90
Diazo/diazonium (R–N=N–R/R–N≡N$^+$)	100–180
Azides (R–N_3)	200–240
Oxime ($R_2C{=}NOH$)	110–140
N-oxides ($R_2N{:}O$)	100–130
Nitroso (R_2C–N=O)	150–290
Isocyanate (R–N=C=O)	50–75
Nitro (R_3C–NO_2)	310–360
N-nitro (R_2N–NO_2)	400–430
Acyl nitrates (–O–NO_2)	400–480

R, in most cases, represents an organic fragment.

The impact of the energetic group in a molecule depends on the size of the molecule. For high molecular weight organic compounds, the presence of a single energetic functional group is unlikely to present a significant hazard. It is therefore of greater benefit if the decomposition energy of a substance is quoted in joules per gram rather than kilojoules per mole (values of $> 500 \, J \, g^{-1}$, for example, indicate potentially explosive behavior, while values of $> 300 \, J \, g^{-1}$ may indicate potentially dangerous self-heating). While it is possible to identify energetic functional groups, it is rarely possible to predict the temperature under which such activity may commence (T_{dec}) and hence experimental techniques are usually required to derive this value. For materials that are more common, reference literature such as Brethericks Handbook of Reactive Chemical Hazards [6] may provide valuable information on thermal stability and reactivity.

At the chemical route selection stage, emphasis is placed on preliminary identification of hazardous reactions or materials. In selecting the most suitable route to manufacture, each route will be assessed against a matrix of criteria (economic and safety) as discussed in Chapter 1. For safety, the following situations must be identified:

- Potentially highly energetic materials must be identified and, if possible, avoided. Where avoidance is not possible, substances should be highlighted for early testing and classification and consideration should be given to processing methods avoiding isolation (for example, processing as a solvent solution rather than an isolated product).
- Any reactions that could cause overpressurization of a vessel including
    - desired reactions, or side reactions that generate permanent gas;
    - desired reactions where the predicted MTSR is above the MTT, thereby posing a vapor pressure hazard;
    - desired reactions where the predicted MTSR is above T_{dec} (if known), thereby presenting a risk of secondary decomposition.

The presence of any of the above criteria in a process does not necessarily suggest that the process is not viable. It does, however, indicate that a more detailed study and possible process changes will be necessary to significantly reduce the intrinsic risk.

3.2.2
Selecting Inherently Safer Processing Conditions

For potentially hazardous reactions identified through prediction of ΔH_r and ΔT_{ad}, it is critical, at the earliest stage possible, to consider elements of inherent safety. These are decisions regarding process design, which can eradicate or mitigate hazardous scenarios. There are a large number of choices that dictate the hazard of the resulting process. Fundamentally, we should seek to develop reactions that, under normal conditions, fall into Criticality Classes 1 or 2.

- If the MTSR is greater than MTT or T_{dec}, consider
    - always using semibatch instead of batch processing methods,

- dividing additions into portions to reduce the MTSR of each portion,
- introduction of catalysts to reduce T_P and hence the MTSR,
- using more solvent (or less reactants) or a solvent with a higher heat capacity to reduce the MTSR, and
- using a higher boiling solvent to increase MTT (while being mindful to avoid transgressing T_{dec}!).

• If the MTT is above T_{dec}, or even if energetic functional groups are present, consider
  - using a lower boiling solvent to ''protect'' hazardous decomposition reactions from being initiated by overheating.

The consequence of most of these decisions can be readily assessed using a small amount of predicted data. Ignorance is the only excuse for not doing this! Following this simple guidance, and applying it at the route selection and early research stage, is likely to result in the specification of an inherently much safer process – and one that is more likely to be successful.

3.3
The Synergies of Safety and Optimization – Together

Once the process reaches the development and optimization stage, physical safety testing can commence. The aim should be to collect adequate information during this phase such that a complete safety dossier exists at the end of the phase. It is crucial that safety testing is not left until the end of development. In this case, process changes may be required on the basis of hitherto unknown safety problems with the original process. This leads to delays in scale-up and always results in additional costs. Similarly, too much data should not be collected too early in development when subsequent changes to the process may render the data irrelevant.

The phases of testing during development and optimization entail the following:

• Testing of potentially explosive compounds
• Thermal stability assessment (to determine T_{dec})
• Definition (or confirmation) of reaction thermodynamics, kinetics, and gas-generation potential.

Toward the end of development, and prior to pilot scale, an assessment of foreseeable process deviations should be undertaken, the consequences assessed, and a basis of safety specified for the scaled-up process.

3.3.1
Testing of Potentially Explosive Compounds

If energetic functional groups are identified in the initial screening procedure, small-scale quantification of the risk is required. The starting point should be a small-scale thermal stability assessment using DSC. This instrument measures

energy changes from a material over time during a temperature ramped or isothermal test. Importantly, the test only requires a small sample quantity (5–10 mg) and thus can be performed at a very early stage of development.

DSC tests for safety purposes should be performed in high pressure, closed test cells to prevent endothermic evaporative losses, which may mask underlying exothermic reactions. It is generally accepted that, if a material exhibits a heat of decomposition of less than 500 J g^{-1}, then it will not possess explosive properties (albeit, it may still present a serious decomposition risk). This threshold is used in UN procedures for classification of dangerous goods for transport [7]. If the decomposition energy is above 500 J g^{-1}, the effect of mechanical (impact and friction) stimulation and thermal sensitivity should be fully evaluated employing the tests and criteria as set out in Ref. [7].

Processing and handling of materials that are explosive and mechanically or thermally sensitive to ignition may prove prohibitive for many companies. If such behavior is exhibited, consideration should be given to methods not requiring isolation of the energetic compound, redesign of the process to avoid such moieties, or even contracting the process stage to a company who specializes in hazardous chemistry and the processing of energetic compounds.

3.3.2
Thermal Stability Assessment

Once explosivity issues have been explored, it is necessary to determine the thermal limits of the process at all stages. The thermal limits will typically be dictated by the temperature at which decomposition or side reactions commence (T_{dec}). For starting materials and products, this data can be collected at an early stage of development. For intermediate mixtures, testing should be conducted when the process is nearing the end of development (to avoid wasting data through subsequent process changes).

Preliminary (quick, small-scale, and low-cost) tests are often employed to screen for thermal instability. Differential thermal analysis (DTA) such as the Carius Tube test or RADEX, or smaller scale DSC are commonly employed. These tests are designed to provide an indication of the thermal behavior but they are neither adiabatic nor have a low phi factor.

The temperature at which a reaction is first observed in a test (the often misquoted ''onset temperature'') will vary for different techniques and is not a constant. The apparent ''onset'' temperature cannot therefore be used directly and requires provision of a conservative safety margin. The magnitude of the safety margin should depend on the sensitivity of the method and the experimental profile. For this, the temperature ramp rate, sample size, and air availability need to be considered (especially for powders). The magnitude of safety margins applicable to a range of thermal stability tests is discussed by Rowe [8]. Application of the safety margin results in the definition of a crude value of T_{dec}.

At this stage, it is not economical or practical to conduct detailed thermal stability studies using adiabatic calorimetry on all process streams and materials, although

they provide the most sensitive process safety analytical technique. The aim of using a screening test is to identify materials where the decomposition "onset" is far in excess of MTSR and MTT – such materials do not normally require further study using techniques that are more accurate. Should the T_{dec} value (apparent "onset" temperature with applied safety margin) be near or below the MTSR or MTT, then more rigorous examination is warranted using adiabatic or higher sensitivity methods to determine the T_{D24} value. Such methods will also provide consequence data in terms of peak conditions achievable by the reaction (pressure and temperature), together with kinetic data for potential design of protection systems.

As mentioned above, adiabatic calorimeters are most commonly used for detailed thermal stability assessment and T_{D24} evaluation. A number of commercially available units exist for such studies, the most common being the accelerating rate calorimeter (ARC [9]), VSP II [10], Phi Tec II [11], and the adiabatic Dewar calorimeter II (ADC) II [12].

In the special case of powders, molecular fragmentation (decomposition) or self-reaction may not be the overriding thermal hazard. Owing to the large surface area/weight ratio, powders may be more susceptible to oxidation. For powder-drying operations in air (such as fluid bed dryers or tray dryers), a range of thermal stability techniques are available that can be used to determine oxidation onset temperatures [13]. The use of contained test methods (DSC or DTA) for powder-drying applications can lead to specification of erroneous and nonconservative safe temperature limits and must therefore be avoided.

3.3.3
Reaction Thermodynamic, Kinetic, and Gas-Generation Quantification

Unless predictive methods, coupled with process observation, suggest no or a low exothermic potential, reaction calorimetry should always be performed [14]. The technique seeks to verify the heat of reaction and provides critical kinetic information on the reaction. When combined with ancillary equipment, the test can also be used to quantify rates and quantities of gas formation from the normal reaction.

There is a range of reaction calorimeters commercially available including the Mettler Toledo RC1e, HEL Simular, ChemiSens CPA 202, and others. Although the systems differ in their measurement method, they all seek to generate the same information. Most systems work under isothermal (constant reaction mass temperature) or isoperibolic (constant jacket temperature) conditions and are suitable for mobile, low-to-medium-viscosity reaction systems. Tests are usually performed by semibatch addition of the final reactant or catalyst and should seek to directly replicate the proposed plant process. For batch chemical reactions, particularly those where the plant process allows the batch temperature to rise, it is more normal to conduct adiabatic calorimetry.

Reaction calorimetry can be conducted in the early stages of development, for example, when seeking to compare synthetic routes, but is more normally

conducted when the process is nearing the end of development. At whatever stage the investigation is conducted, the critical kinetic data obtained as part of the method can provide significant process improvement opportunities. For this reason, reaction calorimetry can be viewed as a very useful tool for process optimization as well as for safety assessment.

The data generated from reaction calorimetry can be used in the generation of complex mathematical models for process simulation assuming that all reactions occurring can be detailed in balanced chemical reaction equations. However, it is more normal for the following empirical data to be derived and evaluated:

- Heat of reaction (ΔH_r)
- Heat capacity of the reaction mass after reaction (Cp)
- Adiabatic temperature rise (ΔT_{ad})
- Reactant accumulation (%)
  - The percentage of the heat released from the system as a function of the quantity of reagent added, compared to the overall heat evolution from the process is termed *accumulated heat* or *reactant accumulation*.
- Changes in physical properties
  - Information on the physical properties of the reaction mixture can be crucial in ensuring safe and robust scale-up. If a mixture becomes more viscous, the measured heat transfer coefficient (U) will decrease. While the absolute value of U in the reaction calorimeter is almost irrelevant for scale-up, changes may require modification to the plant process to enable effective heat removal. In the case of decreasing U value, the addition rate may need to be decreased in the latter stages of addition to match the process heat release to the diminishing heat removal capability of the reactor.
- Rate and quantity of permanent gas generation
  - Normally evaluated using ancillary equipment such as automated gas burette or thermal mass flow meter or by inference from pressure measurement in contained reaction systems.

The adiabatic temperature rise, based on reaction calorimetry data, provides an accurate value for MTSR in the thermal safety evaluation, refining any initial predictions. It should be noted, however, that this is only true if there are no secondary or side reactions that may be initiated at an elevated reaction temperature. This permits a more robust assessment of the true Criticality Class and facilitates process changes to modify the risk. However, the reactant accumulation data provides a key opportunity to enhance the safety and productivity of the process – together.

For an ideal semibatch process, the added reagent should react as soon as it enters the reaction mass. If the material reacts instantaneously, when a process deviation occurs during the feeding, then stopping the addition will cause immediate cessation of the reaction and the deviation will be nonhazardous. However, for a variety of reasons, some of the added material may not react instantaneously and may accumulate. In this case, if a process deviation occurs, there will be continued reaction even after the addition is halted. In extreme cases, this accumulated

reactant may introduce a significant hazard if the MTSR of the accumulated reaction is above MTT or T_{dec}. The extra time in the reactor required to stir the reaction mass after the end of the addition to achieve 100% conversion makes the process less economical as well as potentially less safe.

Development work for semibatch processes should focus on establishing conditions with minimal accumulation. Accumulation can be caused by

- the temperature being too low (kinetics too slow to match the feed rate);
- addition time being too short;
- incorrect initiation, presence of inhibiting species, or absence of catalyzing species;
- inadequate agitation causing mass-transfer-induced accumulation.

Accordingly, accumulation can be reduced by using higher process temperatures, longer feed durations, better agitation, or inclusion of catalysts.

The area of accumulation is one where safety and optimization work synergistically – a safer process is generally a more productive process. The methanol/acetic anhydride esterification reaction exemplifies reactant accumulation and process modifications, which can make the process safer and faster. Using stoichiometric quantities of the two reagents, the process is conducted by adding acetic anhydride to methanol over 20 min. Figure 3.5 illustrates the reaction profile under two different conditions:

- At 55 °C without catalyst and
- At 20 °C with 1% w/w sulfuric acid catalyst.

Figure 3.5 Comparison of reaction calorimetry data for the methanol/acetic anhydride esterification.

Table 3.3 Comparison of methanol/acetic anhydride reaction data.

Reaction temperature (°C)	50	20
Catalyst	No	Yes
ΔH_r (kJ mol^{-1})	−64.6	−64.3
Accumulation (%)	85	33
ΔT_{ad} (full reaction; K)	169.0	173.4
ΔT_{ad} (accumulation only; K)	143.7	57.2
MTSR (full reaction; °C)	219.0	193.4
MTSR (accumulation only; °C)	197.7	77.2
MTT (boiling point; °C)	80	80
Time to reach 95% thermal conversion after addition ends (min)	115	26

Although the heat of reaction is consistent between the two methods as would be expected, the extent of accumulation is markedly different. The time required to reach 95% conversion after completion of the addition, together with other derived data, is demonstrated in Table 3.3.

The process at 50 °C without catalyst is highly inefficient (taking 115 min to reach 95% conversion from the end of the addition period) and unsafe (Criticality Class 3). Despite being intended as a semibatch reaction, there is so much accumulation that it is, in practice, a batch reaction with 85% of the heat being evolved after the addition is completed. By increasing the reaction rate through use of higher temperature or, better still, using a catalyst as in the data shown, the heat flow becomes more feed rate controlled with less accumulation. Any process deviation after the end of dosing is non-safety critical.

Such an intimate link between safety and efficiency suggests that reaction calorimetry should start as early as possible in development and be used as a development tool. Conducting such calorimetry after development and just prior to scale-up is not good practice and may result in an inherently unsafe or inefficient process, or increased costs in repeating development work.

3.3.4
Developing Fault-Tolerant Processes – by Design

Once the thermal stability of process materials and the kinetics and thermodynamics of the process have been evaluated, the Criticality Class of the desired reaction will be established. At this point, it is essential to consider methods and modifications for making the process more inherently safe and robust. While doing this, common failure situations should be assessed to determine if changes can make the process more fault tolerant. Section 3.2.2 proposes methods for reducing the Criticality Class of a process based on the four critical process parameters. This assessment, originally considered during process design, should be repeated during development to ensure that the residual process risk is low.

Armed with the data generated during design and development, it will be possible to evaluate the effect of common process deviations. The potential for a specific process deviation to occur varies considerably between different scales and different plants. However, some deviations are likely to remain common and foreseeable for most plants. These include, but are not limited to, issues relating to the following:

- Reactant, solvent, and catalyst additions – for example, too much/little, too fast/slow, omission, or wrong sequence
- Equipment failure – agitation failure/fault, cooling failure, and overheating
- Temperature – too low/high.

The aim is to develop a process that is sufficiently well designed to contend with such foreseeable deviations. The key principles of inherent safety should be considered during process development. It is unlikely that the residual process risk can be reduced to zero (inherently safe) but it is possible to reduce the process risk considerably (making it inherently safer). The guiding principles of inherent safety are as follows:

- **Substitution** – changing a hazardous material for a less hazardous one
  - for example, using toluene/tetrahydrofuran as a solvent for a Grignard process instead of the more flammable diethyl ether
- **Intensification** – reducing the quantity of hazardous materials processed
  - for example, using microreactor systems for highly hazardous chemistry instead of traditional stirred tank reactors
- **Attenuation** – changing to less hazardous conditions
  - for example, including a catalyst to reduce the required process temperature and/or pressure for an exothermic reaction or to change a batch reaction into a semibatch one
- **Control** – providing instruments of a suitable integrity level, or procedures, to eliminate potential deviations
  - for example, using a low-temperature interlock to stop an addition if the process temperature is too low to reduce the risk of accumulation.

Armed with these tools, it is possible to develop robust, fault-tolerant, and efficient chemical processes. For scale-up, however, confirmation and quantification of residual risks is imperative so that a reliable basis of safety can be implemented.

3.4
Establishing a Reliable Basis of Safety for Scale-Up

The safety work performed during design and development is wasted if this is not translated into a reliable safety system for industrial operations. The usual first stage of scale-up is to pilot scale (typically between 50 and 1000 l) followed to production scale (>1000 l). Pilot-scale facilities are generally characterized by highly trained operators (usually qualified scientists), a high level of parameter variability, predominantly manual operation, and minimal presence of hardwired

safety systems. This combination of conditions implies that deviation scenarios (the occurrence of a deviation from the planned processing instructions) would not be uncommon – although close supervision by highly trained operators may reduce the frequency of such scenarios.

Many incidents at pilot scale highlight the need to treat the pilot scale as "small-scale production" rather than "large-scale laboratory." Making minor modifications to the process at pilot scale, without thorough prior safety evaluation, must be strictly prohibited.

The critical stages of prepilot plant assessment are as follows:

- Examining the existing thermochemical data for "obvious" hazards inherent in the process
- Conducting a thorough hazard-identification exercise to identify foreseeable, and realistic scenarios that may present an overpressurization hazard
- Determining the consequences of foreseeable deviations and defining the worst-case overpressurization scenario.
- Specifying and implementing safety measures to protect the vessel(s) from all foreseeable conditions that may present a risk of overpressurization.

For mildly exothermic processes operated at high dilution in the absence of any energetic functional groups, there is clearly a case for a more superficial assessment, but this should never be interpreted as "no assessment." A single deviation for a reaction of Criticality Class 1, for example, omission of the solvent, could result in a shift to Criticality Class 5.

3.4.1
Hazardous Scenario Identification

In order to derive a list of potentially hazardous scenarios for the pilot plant, it is necessary to combine the thermochemical data relating to the process with an intimate knowledge of the pilot plant configuration and control strategy. That is, gaining an understanding of what can realistically go wrong with the operation of the vessel resulting in a potentially hazardous event. Methods for hazard identification [15] include the following:

- Hazard and operability (HAZOP) studies
- Checklist assessments
- Informal "what if?" assessments
- Failure modes and effects analysis (FMEA)
- Fault tree analysis.

The assessment procedure selected will be dictated by the magnitude of scale-up and/or the intrinsic risk of the process. Less formal analysis such as "checklist" or "what if?" analysis may be applied to scale-up to pilot, whereas it is more common for the more formal "HAZOP" technique to be applied for production scale-up. Whichever method is used, the outcome should be a list of potential scenarios that are feasible, credible, and may give rise to a hazardous consequence.

3.4.2
Determining the Consequences of Hazardous Scenarios

Once a shortlist of hazardous scenarios is available, it is necessary to conclusively ascertain whether the consequences of the scenarios are hazardous or benign. This can be achieved through computational simulation, estimation based on existing process safety data, or experimental simulation. Computational simulation is feasible, but requires substantial information on physicochemical and kinetic properties. A fundamental understanding of the mechanism of the reaction – and all conceivable side/secondary reactions – along with formal kinetic parameters for each reaction would be required. For a small-volume product, the complexity of collecting the necessary data would prove prohibitive. In some cases, for example, the scale-up to a high-throughput continuous reactor, this rigorous approach may be warranted.

Estimation of scenario consequences may be possible using existing data. Heat of reaction and heat capacity data can be manipulated to evaluate the consequences of certain deviations. A good example here would be a change in quantities of solvents or reactants. As a screening exercise, this may be sufficient to rank deviations in terms of their likely severity. Combined with adequate thermal stability data, the potential of scenarios to initiate undesirable secondary reactions can also be assessed and any change in Criticality Class quantified. Any such calculations are likely to yield thermodynamic information regarding the overall magnitude of thermal change and the probability of initiating other events. This approach is likely to have merit for qualitative assessment but is unlikely to provide enough kinetic data for safety system design. Thus, it is an option for highlighting a scenario that ''is likely to have significant consequences'' but is unlikely to adequately quantify the kinetics of the resulting event. Typically, this approach would be reserved for ranking deviation potential.

In some cases, this approach may not be appropriate. For example, where loss of agitation has been identified as safety critical, a decision is required regarding the potential of a reaction system to stratify. Simple thermodynamic evaluation will not answer this question. In this instance, the failure should be examined under appropriate experimental, that is, adiabatic conditions. This will provide a full understanding, both thermodynamic and kinetic, of the consequences of stratification.

3.4.3
Experimental Simulation – Adiabatic Calorimetry

The importance of the impact of heat loss and thermal inertia on plant behavior has already been highlighted. To simulate a runaway reaction under plant-scale conditions, adiabatic and low thermal inertia test methods are required using adiabatic calorimeters as discussed in Section 3.3.2. In addition to having effectively zero heat loss and low thermal inertia (phi factor), these calorimeters are designed to resist high pressures, simulate plant-scale agitation systems as closely as possible,

Figure 3.6 Adiabatic Dewar calorimetry data for the
methanol/acetic anhydride reaction from 25 °C.

and permit heating and material additions. The test procedure employed must
closely mimic the process deviation under investigation including, where possible,
the use of plant-grade materials.

The data obtained from testing should provide a direct measurement of the
consequences of a failure case (kinetics as well as thermodynamics), which can be
used directly in the design of safety systems. For reactions and their associated
credible deviations falling within Criticality Classes 3, 4, or 5, procedural or
engineered safety systems are required. These will either prevent the scenario
from occurring or protect against the consequences. Figure 3.6 illustrates adiabatic
Dewar calorimetry data for the methanol/acetic anhydride reaction discussed in
Section 3.3.3. The test simulates the effect of loss of process vessel cooling on the
uncatalyzed reaction following completion of the addition.

The MTSR measured for the reaction (178 °C) is considerably above the MTT
(conservatively considered the system boiling point of 65 °C). The starting materials
and products of the reaction are thermally stable to well in excess of 200 °C (that
is, to say, $T_{dec} > 200$ °C). This would place this scenario in Criticality Class 3 and
would suggest the need for engineering safety measures to protect the reactor from
this deviation. The time to maximum rate (TMR) for the reaction is 100 min (this is
the time taken from initiation to reaching maximum rate of temperature rise). The
pressure generated in this case is purely due to vapor pressure of the products – no
permanent gas is formed by the reaction.

3.4.4
Specify and Implement Safety Measures

Once the consequences of all the worst-case candidates have been quantified, the final task is to specify which safety measures are required to protect the reactor from the consequences or to validate if existing protection measures and protocols are acceptable. There are numerous options available including the following:

- Process control
- Design for containment
- Reaction dumping/passive quenching
- Reaction inhibition/active quenching
- Emergency pressure relief (venting).

In the process industries, pressure relief systems via bursting discs or relief valves are the normal ultimate basis of safety. However, with increasing environmental pressures and legislation, it is no longer sufficient to size an orifice large enough to prevent the vessel exceeding its design pressure. The design must consider treatment of the discharged stream. For this and other protection systems, accurate kinetic information on the runaway reaction is required, and validation of the design is essential to confirm that it is sufficiently reliable.

In some cases, process control can be employed as the ultimate basis of safety – that is, reliance on instrumentation and control systems to prevent a scenario from materializing. Any such systems should be developed to the principles of best practice laid down by engineering standards. For example, if the functional safety relies on safety instrumented systems, then the level of protection afforded by the instrumented system should conform to the methods set out in IEC 61508/IEC 61511. For some scenarios, the outcome of the deviation may be sufficiently severe that it cannot be permitted to happen. In this case, control systems would be the only basis of safety available and the criticality of having a reliable system would be evident.

It is common for a combination of layers of protection to be employed rather than rely solely on one basis of safety. Layers of protection analysis (LOPA [16]) has recently found prominence in extending the hazard identification and risk-assessment process to demonstrate that a systematic assessment of multiple independent safety features achieves an acceptable level of safety. If a safeguard is effective in preventing a scenario from reaching its consequence and it is independent of the initiating event including other layers of protection, then it is considered an independent protection layer (IPL). The combination of IPLs, general design features, and procedural and other such layers are assessed to yield an overall credit. The frequency of the initiating event, the assessed risk reduction or probability of failure on demand (PFD), and the severity of the undesired consequence are used to judge the acceptability of an identified risk against tolerable safety, environment, and commercial criteria.

At the point of scale-up, the adequacy of the protection or prevention measures will be directly proportional to the adequacy of the underpinning stages of

specification. If any phase of the procedure is deficient, this will have a detrimental impact on the adequacy of the final design [17]. A written safety dossier must exist, which demonstrates that the assessment procedure has been followed completely. The basis of safety for the pilot-scale operation to protect against all the credible failures should be clear and unambiguous – as should the important procedural/engineering control measures be in place as part of this basis of safety.

3.4.5
Large-Scale Production

The procedure for safety evaluation of large-scale production should be similar to that for pilot scale but would generally be more rigorous. The most important differences between pilot and production scale-up include the following:

- The consequences of a deviation will be more dramatic owing to the larger inventory. This implies the need for a more rigorous and exhaustive hazard-identification exercise.
- The variability of the production plant is likely to be less than that for the pilot plant.
- The need for instrumented safety systems to comply with best practice will require assessment of safety systems with respect to international best practice methods such as IEC 61508/61511.

A critical element of any safety system is that its suitability must be reconfirmed following any process change. A review of the impact of any change to the process or plant should be accompanied by a review of the potential consequences of that change and the adequacy of the corresponding safety systems in light of the modification.

3.5
Flammability Hazards

Along with reactivity hazards, flammability and fire properties of the process materials also present a potential process risk. In a laboratory environment, control of ignition sources is the generally accepted basis of safety. However, at pilot and production scale, the potential risk posed by flammability increases substantially, including the significant risk posed by flammable dusts clouds. The most common flammability parameters associated with gases, vapors, and dusts are highlighted in Table 3.4.

Flammability data for gases and vapors is, in many cases, available from reliable literature sources; however, for dusts and powders their considerable variability is such that this is invariably not the case. This means that measurement of properties specific to a material will normally be required.

Almost all organic or metal powders, when finely divided and dispersed, are capable of igniting and propagating an explosion. Whether this will pose a risk in

Table 3.4   Important parameters for characterizing flammability hazards.

Parameter group	Dusts/powders	Gases/vapors
Ignition sensitivity	Minimum ignition energy (MIE)	Minimum ignition energy (MIE)
	Minimum (cloud) ignition temperature (MIT)	Autoignition temperature (AIT)
	Layer (5 mm) ignition temperature (LIT)	
Ignition severity	Maximum explosion pressure (P_{max})	Maximum explosion pressure (P_{max})
	Explosion severity constant (K_{st})	Explosion severity constant (K_g)
Flammable range	Minimum explosible concentration (MEC)	Upper explosive limits (UELs) and lower explosive limits (LELs)
	Limiting oxygen concentration for combustion (LOC)	Minimum oxygen concentration for combustion (MOC)
		Flash point

the production facility will depend on many factors. The ignition sensitivity and explosion severity of a particular substance can be highly variable and is influenced greatly by parameters such as the moisture content, its particle size, and even particle geometry. As a consequence of this, when a decision is made as to whether a flammable powder poses a significant risk when processed, it must be based on flammability data relating to the powder concerned and not on generic data. Only the tests needed to specify and confirm the acceptability of the basis of safety are required, not necessarily all parameters indicated in Table 3.4.

Powders can be much less sensitive to ignition than gases or vapors. In the latter case, the high sensitivity to ignition normally eliminates avoidance of ignition sources as a robust or reliable basis of safety. For powders, avoidance of ignition sources can be reliably employed – especially for ignition-insensitive powders.

The consequence of an undesirable event will dictate the level of expense and time allocated to addressing it. The consequence may be trivial (e.g., small, localized flash fire) or catastrophic (e.g., reactor explosion resulting in fatalities, environmental contamination, and commercial loss). For gas, vapor, or dust explosion hazards, the consequences of an event may be evaluated using prediction software (such as PHAST [18]). Such software is well developed, readily available, and provides a rapid overview of the impact of an event.

The data requirements associated with bases of safety for flammability hazards are highlighted in Figure 3.7 along with test parameters required for assessment. Different stages of a process (for example, powder charging to a vessel, powder blending, micronizing, and drying) may require different bases of safety, so a range

Figure 3.7 Data requirements for the Basis of Safety for flammability hazards.

of parameters may be required to validate each of the bases of safety for a given substance.

The same approach is applied to flammability hazards as applied to exothermic and gas-generating reaction hazards. That is,

- material/process characterization,
- hazard identification and determination of hazardous scenarios, and
- provision of a basis of safety.

While some early development decisions will affect the final process flammability risk, the majority of material flammability data will require collection when sufficient quantities of material are available – typically late development or early pilot production. If this data is only available after the commencement of pilot scale operation, how do we set a robust basis of safety for avoidance of flammability hazards at pilot scale?

3.5.1
Assessing Pilot-Scale Flammability Hazards

In essence, the operations performed at pilot scale are generic. Each process involves a collection of a relatively limited number of discrete unit operations including (but not limited to) vessel inerting, vessel charging (liquids, solids, and

gases), reactions, sampling, distillation/reflux, vessel discharging (liquids, slurries), centrifugation/filtration, drying, and vessel/equipment cleaning and maintenance.

The following procedure is recommended for establishing a pilot-scale basis of safety that readily facilitates the implementation of new processes:

- Compile standard operating procedures (SOPs) for all foreseeable unit operations on the pilot plant.
- Conduct a detailed hazard and risk assessment for each unit operation with regard to flammability and ignition source identification.
  - Base the risk assessment on demanding (worst case) material properties (for example, for a liquid, assume the material is flammable and has a subambient temperature flash point). Document the material properties that have been considered in the assessment.
  - Undertake a hazardous area classification for vapors, gases, and powder for both normal and expected abnormal scenarios.
  - Conduct a full and detailed audit of intrinsic ignition sources associated with the plant and those not particular to the equipment but which may occur in the pilot plant (including electrical/mechanical equipment and electrostatic ignition sources).
- Determine an acceptable Basis of Safety for each unit operation.
- Implement any recommendations or actions required to support the selected Basis of Safety.

Once this overall pilot plant assessment is in place, the introduction of each new process becomes relatively straightforward:

- Confirm that standard unit operations are proposed.
- Confirm that the (flammability and reactivity) properties of materials used in the process are within the limits of those used in the generic assessment.
  - Collect necessary material data for all chemicals used. For dust explosion testing, sufficient quantities for full testing ($>500\,$g) may only be available after a pilot-scale batch has been produced. Limit tests, at the limits assumed in the generic assessment, may be possible with smaller samples available at laboratory scale, hence facilitating confirmation of the basis of safety for each generic unit operation.
- For each new unit operation, conduct a hazard and risk assessment, identify any variant that may introduce new ignition sources and, if so, take remedial action to eliminate them and specify a Basis of Safety.
- Confirm whether any new ignition sources are introduced by the process (for example, pyrophoric materials or new packaging materials with different electrostatic properties).
- Confirm that the Basis of Safety specified in the generic unit operation assessment remains valid and robust.
- Implement special measures where the generic unit operation Basis of Safety is invalidated or requires supplementary measures.

- Ensure that the final operating instructions contain the necessary safety measures (procedural or engineering) to meet the requirements of the generic basis of safety and any special (process-specific) measures identified.

Following this procedure, the explosive atmosphere assessment can be condensed considerably. While there is likely to be some effort required in setting up the SOPs, generic risk assessments, and ignition source audits, the process for introduction of new processes should be streamlined considerably and may be condensed to within a matter of days.

3.6
Summary

Protecting against overpressure hazards arising from gas, vapor, or dust explosion, and thermal stability and reaction hazards is a prerequisite for the process industries. The critical phases in the process are as follows:

- Process/material characterization
- Hazard and risk identification
- Consequence analysis
- Safety system specification, design, and implementation.

At the end of this process, a robust basis of safety should be specified and implemented, which protects against all foreseeable overpressure hazards. A rigorous exercise will dictate the extent to which overpressure hazards are identified. Consequence analysis will identify the magnitude of the manifested hazard and will dictate the effort and measures imposed to mitigate the risk. The criticality of having the appropriate experimental data on the process and/or material cannot be understated. Deficiency in safety data can lead to underdesign of the safety system – rendering it potentially unsafe – or can lead to overdesign of the safety system – adding unnecessary expense and potentially complexity.

For reaction hazard and thermal stability assessment, emphasis should be placed on developing inherently safer processes at the route selection and R&D stages. Prediction techniques to provide an early indication of reaction and thermal stability hazard are available and should be employed. The use of such techniques, with the conceptual understanding of classification of reactions using the Criticality Class concept should facilitate development of more inherently robust chemical processes.

One of the key aspects in developing safer chemical processes involves education. Undergraduate chemistry courses do not typically contain significant content in assessing and understanding the hazards of chemical reactions. This makes it fundamentally important for the industry to educate and train graduate recruits at the earliest opportunity possible. Chemical engineering courses do tend to focus more on hazard awareness. However, it is common for chemical engineers only to become involved in scale-up once the process is fixed – and after the

important safety-related decisions in development have already been, sometimes unknowingly, made.

References

1. Barton, J. and Rogers, R.L. (1997) *Chemical Reaction Hazards – A Guide to Safety*, 2nd edn, IChemE, Rugby. ISBN: 0 85295 341 0.
2. Stoessel, F. (2008) *Thermal Safety of Chemical Processes*, Wiley-VCH Verlag GmbH. ISBN: 978-3-527-31712-7.
3. ASTM (2005) Stock # DS51E. *ASTM Computer Program for Chemical Thermodynamic and Energy Release Evaluation - CHETAH Version 8.0*, ASTM, Philadelphia, PA. ISBN: 0 8031 3366 9.
4. Heat of formation data, for liquid phase molecules, extracted from the NIST Chemistry Webbook, *http://webbook.nist.gov/chemistry*.
5. Grewer, T. (1994) *Thermal Hazards of Chemical Reactions*, Industrial Safety Series, Vol. **4**, Elsevier Science BV. ISBN: 0-444-89722-4.
6. Urben, P.G. (ed.) (2008) *Bretherick's Handbook of Reactive Chemical Hazards*, 7th edn, Academic Press. ISBN: 0-123-72563-1.
7. (2003) *Recommendations on the Transport of Dangerous Goods – Manual of Tests and Criteria*, 4th Revised edn, United Nations, New York and Geneva. ISBN: 92-1-139087-7.
8. Rowe, S.M. (2002) Thermal stability: a review of methods and interpretation of data. *Org. Proc. Res. Dev.*, **6**, 877–883.
9. Townsend, D.I. and Tou, J.C. (1980) Hazard evaluation by an accelerating rate calorimeter. *Thermochim. Acta*, **37**, 1–30.
10. Askonas, C.F., Burelbach, J.P., and Leung, J.C. (2000) The versatile VSP2: a tool for adiabatic thermal analysis and vent sizing applications. Presented at North American Thermal Analysis Society, 28th Annual Conference, Orlando, October 2000.
11. Singh, J. (1993) Reliable scale-up of thermal hazards using the Phi Tec II calorimeter. *Thermochim. Acta*, **226** (1–2), 211–220.
12. Rowe, S.M. and Middle, K.V. (1999) Advances in adiabatic dewar calorimetry. International Conference on Runaway Reactions, St Petersburg (Russia), June 1999.
13. Abbott, J.A. (1990) *Prevention of Fires and Explosions in Dryers – A User Guide*, 2nd edn, IChemE, Rugby. ISBN: 085295 257 0.
14. Rowe, S.M. and Starkie, A.J. (2000) Calorimetric characterisation of the controlled process – a fundamental ingredient for safe processing. International Conference On Assessment and Control of Chemical Processing Hazards, IBC Global Conferences Ltd, London, October 2000.
15. (1993) *Risk Analysis in the Process Industries*, EFCE Publication Series, vol. 45, IChemE, Rugby. ISBN: 0 85295 183 3.
16. Dowell, A.M. and Hendershot, D.C. (2002) Simplified risk analysis – Layer of Protection Analysis (LOPA). AIChE 2002 National Meeting, Indianapolis, 3–8 November, Paper 281a.
17. Rowe, S.M. (2008) *Rock-solid Safety*, vol. 802, The Chemical Engineer (IChemE), April, pp. 24–28.
18. Process Hazard Analysis Software Tool (PHAST), Available from DNV, *http://www.dnv.com/software/all/phast/productInfo.asp* (accessed 2011).

4
Understanding the Reaction

John Atherton, Ian Houson, and Mark Talford

4.1
Introduction

What constitutes the understanding of a chemical process? Although the exact definition will change as one moves through the process lifecycle, for the purposes of this chapter, a definition is as follows:

Process understanding is

a conceptual model of the process of sufficient complexity to understand the factors that control the process outcome and to be able to predict successfully what happens when process changes are made.

There are four main reasons why we need to "understand" a process: to make sure that it is

1) safe to operate;
2) optimized within the technocommercial constraints externally imposed, for example, capital availability, time, and resources;
3) scaleable from laboratory to the required manufacturing scale; and
4) robust with respect to the expected variation in input materials and the expected variation in operating parameters.

Best practice in process development is moving away from a phenomenological approach, based on cause and effect, to a methodological approach based on achieving an appropriate level of scientific understanding of the physics and chemistry that determines process performance. (ICH Q8 and Q9 guidelines *http://www.ich.org/cache/compo/363-272-1.html* and Chapter 1.) With the exception of slow homogeneous chemical reactions, the performance of chemical processes depends on how the reactive materials are contacted. Even at the laboratory scale, process performance can be critically dependent on the chosen equipment or on the physical state of a solid reactant. Definition of "critical

process parameters'' or ''design space'' cannot be meaningful without specification of relevant hydrodynamic, mass transfer, or heat transfer characteristics of the equipment used for the process. Some understanding of process kinetics, including that of reactions leading to ''specification critical'' impurities, is needed in order to ensure selection of appropriate equipment and process conditions such as pH (for aqueous systems) and reactant contacting method (batch, fed batch, continuous).

The first part of this chapter discusses the chemical aspects that need to be understood while the second half introduces the physical aspects that require consideration.

At a fundamental level, this chapter focuses on two questions

- What do we need to control at the molecular level?
  - Or ''How do we get the molecules to react in the way that we want them to: to maximize yield and purity?''
- How can we deliver the required conditions at the meso and macro levels (equipment scale)?

In order to do this, we first look at the underlying chemical and physical rate processes and use our understanding of these to determine how we can best deliver the conditions from the resources at our disposal.

This chapter is divided into four sections:

1) Concepts of chemical complexity and ordering of data requirements
2) Discussion of the impact of chemical rate processes
3) Discussion of the impact of physical rate processes
4) Concepts of scale and structure to aid in equipment selection.

4.2
Process Complexity

In order to better comprehend the information requirements to provide adequate process understanding, it is helpful to consider the factors that lead to process complexity. Figure 4.1 illustrates one way of approaching this.

We deal briefly with the three axes in this diagram.

4.2.1
Number of Phases

The majority of processes in the pharmaceutical, agrochemical, and fine chemical industries are multiphase. A recent survey has shown that two-thirds of pharma processes have at least two phases and one-third have three or more phases present at the reaction stage [1]. System complexity increases rapidly as the number of phases increases. Interphase mass transfer is always an issue, and can be rate and selectivity controlling when the solubility of a reactant in the reacting phase is low. Gas/liquid and solid/liquid systems are the most problematic.

Figure 4.1 Some elements of process complexity.

4.2.2
Physical and Dynamic Complexity

Some of the fundamental physical parameters that interact in chemistry are shown on the right-hand side of this axis. They are independent of the equipment used, but interact with the equipment to produce equipment-dependent variables (expressed in engineering terms) shown on the left of the axis.

4.2.2.1 Length Scales
The length scale can affect processing characteristics in a number of ways, and therefore great care has to be taken in extrapolating process performance data from one item of equipment to another. For example, the *volume : surface area* ratio influences the heat transfer performance, as shown in Table 4.1.

Table 4.1 Change in volume : surface area ratio with scale.

Volume	10 l	100 l	1 m³	10 m³
	Change in volume/surface area ratio			
1 l	2.15	4.64	10.00	21.5
10 l	–	2.15	4.64	10.0
100 l	–	–	2.15	4.64
1 m³	–	–	–	2.15

The *solution depth* will influence the phase separation times in a two-phase system. This is because the linear movement of the interfaces is approximately constant, whereas the distance the interfaces must move increases with scale.

4.2.2.2 Time
Residence times, including unplanned hold times, are key in determining process performance. The overall reaction time in a batch or semibatch process usually increases on increasing the scale. Material-handling issues and heat-removal requirements contribute significantly to this.

4.2.2.3 Solubility
Low solubility of a reaction component can limit the overall reaction rate, even under conditions where the mass transfer coefficient is high.

4.2.2.4 Density
The density difference between two phases is important for determining the settling or phase separation rates in two-phase systems. The likelihood of emulsion formation is greatly enhanced if the densities of two liquid phases are similar. Densities of liquid phases can change during reactions, as reagents are used up and products formed.

4.2.2.5 Rheology
Materials with difficult rheological properties, for example, a yield stress (as in a Bingham plastic), can cause serious problems with heat transfer, mixing, and flow in pipes, particularly after a shutdown. Rheological properties can change throughout a process.

4.2.2.6 Heat Transfer
This is probably the most difficult aspect to deal with at the early stage of process design and equipment selection, as the required heat transfer rate will depend on reaction rate, which itself may be concentration dependent, and dependent on the feed rate (or on the flow rate in a continuous process). The heat of reaction should be measured or calculated at an early stage of development (Chapters 3 and 5).

4.2.2.7 Mass Transfer/Interfacial Area
In multiphase systems, there will be a minimum requirement for mass transfer to achieve an acceptable reaction rate. High $k_L a$ values may be needed when one reagent has low solubility (discussed in more detail later in the chapter). This is common in catalytic hydrogenation processes, where ''hydrogen starvation'' can lead to undesirable side reactions. This may also have an effect on selectivity if there are competing reactions where (say) the desired reaction requires good mass transfer but the side reaction is a single-phase reaction.

4.2.2.8 Mixing Time

A small but significant proportion (possibly 10–15%) of reactions are sensitive to how the reactants are mixed. Depending on the kinetics of competing reactions, the required mixing time may vary from milliseconds to minutes. Particular attention needs to be paid to this if the intrinsic chemical rates are similar to, or faster than, the mixing times.

4.2.3
Chemical Complexity

The components of chemical complexity are generally well understood, and identification of by-products and understanding of reaction pathways and mechanisms of catalysis are meat and drink to development chemists. Less well understood are situations where very fast reactions occur on the timescale of mixing, or where the product formation rate is driven by physical interactions, for example, a solubility equilibrium.

4.3
Topics for Data Acquisition

Therefore, our recommended list of topics to *consider* for data acquisition, not all of which may be relevant for a particular process, is as follows:

1) Literature survey for physicochemical and physical property data: pK_a, solubilities, kinetics
2) Reaction monitoring to get time/composition profiles in order to determine the rates of formation of product and by-products, and the rate of consumption of the reactant
3) Identification and characterization of pre- and postreaction equilibria and their likely influence on overall reaction rates
4) Identification of factors influencing main and side reactions
5) If necessary, investigation of the kinetics of specific aspects of the process to deconvolute complex features
6) Heats of reaction, from literaure, calculation or experiment
7) Consideration/investigation of the possibility of mixing effects degrading selectivity
8) For multiphase processes, acquisition of solubility or partition data, identification of the reacting phase, and identification of the relationship between performance and mass transfer conditions. Measurement of the reaction rate in the reacting phase
9) Generation of a picture, incorporating relevant chemical and physical characteristics of the process, in order to display and better understand the various interacting factors contributing to process performance.

A summary of the data acquisition strategy is shown in Figure 4.2, and this provides the plan for this chapter.

Figure 4.2 Outline of data requirements for process understanding.

4.4
Reaction Profiles

It is now widely recognized that the acquisition of composition/time profiles during a batch or fed-batch reaction is essential to gain adequate process understanding. (Based on the author's consultancy experience with >30 fine chemical and pharmaceutical companies over the past five years.) This involves acquiring the batch composition for the starting material, product, significant by-product and intermediates during the reaction. Chromatographic methods are usually required to give the level of information necessary for understanding the chemical complexity involved. Once this understanding is gained, much useful information can be acquired from thermochemical and *in situ* spectroscopic methods [2]. For fed-batch reactions, useful information can be gleaned by tracking the disappearance rate of a small portion of the starting material added quickly to the reaction mass. A rapid rise in temperature, even of a few degrees, shows that an exothermic process is taking place, but does not by itself discriminate between one that is caused simply by heat of dilution and one that is due to an exothermic reaction. Any rapid heat rise suggests that the temperature is nonisotropic, and further investigation of the possibility of mixing effects on selectivity should be made.

It is common practice to track the disappearance of the starting material, and there is a natural tendency to assume that a good time to stop the reaction is when

Figure 4.3 Example reaction profile for consecutive first-order reactions.

the concentration of the key raw material has fallen to a predetermined low level. This is not necessarily the case. At the simplest level, the yield of a product of a first-order reaction is at a maximum when all the reactant has been consumed only if the product itself is stable. Figure 4.3 shows the case where the decomposition rate of the product **B** is one-twentieth of its formation rate. In this case, a maximum yield of 85.4% occurs when there is still 4% of unreacted starting material.

Therefore, meaningful reaction profiling must also include measurement of the product formation rate, as well as tracking the concentrations of significant intermediates and impurities.

4.5
Reaction Pictures

Specification of data requirements to provide an adequate understanding is an iterative and beneficial process that begins with asking questions based on the complexity indicators. Answering the questions is best done by developing a picture of the reaction system that includes the chemistry, the phase behavior, and the relevant physical interactions.

An example (Figure 4.4) that shows several common features in the catalytic reduction of a nitrile, the desired product being a primary amine.

Once this picture has been developed, a simple inspection reveals the numerous individual processes that contribute to the overall process performance.

This analysis facilitates the identification of key process parameters, and the identification of "branch points" – the points at which a species can be transformed in the direction of either the product or the by-product. This is very helpful in the diagnosis of rate or selectivity problems. Potential causes of unexpectedly slow reactions are also identifiable (Table 4.2): in this case, low hydrogen availability due to inadequate mass transfer, low substrate availability due to poor solubility or a slow dissolution rate, or low catalyst availability if it is not adequately suspended in the vessel (e.g., in a pile at the bottom).

Figure 4.4 Reaction picture for the catalytic reduction of a nitrile.

Table 4.2 Table of potential causes of slow reaction or poor selectivity.

Parameter	Relevant items	Influences	Comment
Heat transfer	From liquid to reactor	Cycle time	–
	Cooling fluid		–
Solubility	H_2	Availability of H_2 for reaction	H_2 solubility in millimolar range
Solubility	Nitrile	Availability of nitrile for reaction	–
k_La	H_2	H_2 transfer rate into solution	–
k_La	Nitrile	Nitrile transfer rate into solution (dissolution)	Nitrile particle size, polymorph
k_La	Catalyst	Limits chemical reaction rates (suspension of catalyst within reactor)	Catalyst particle size
Rate constants	For R1–R7	Rates and relative rates influence selectivity	–
Local concentration	Of all seven reacting species	Local concentrations influence selectivity	–

4.6
Ionic Equilibria and Reaction Selectivity

The performance of many reactions is dependent on the equilibrium processes that influence the availability of the reactive form of the starting material. A few examples using nitration, acylation, and Strecker reactions will be used to illustrate the principles involved. An example involving organic solvents is the chemistry of organometallic lithium salts in apolar organic solvents, which is dominated by aggregation effects.

4.6.1
Nitration

Most nitration processes proceed via the nitronium ion, which is formed and reacts as shown below.

$$HNO_3 + H^+ \rightleftharpoons NO_2^+ + H_2O$$
$$Ar\text{-}H + NO_2^+ \rightarrow Ar\text{-}NO_2 + H_2O$$

Pure nitric acid is only about 2% ionized as shown above at $-10\,°C$; if the water content is increased to 5%, then the nitronium ion becomes spectroscopically undetectable [3]. Nitration with neat nitric acid is therefore unsatisfactory, as the water generated by the reaction reduces the availability of the nitronium ion. Use of sulfuric acid/water as a solvent/desiccant is usually more satisfactory. Figure 4.5 shows how the nitration rate of *p*-dichlorobenzene tracks the availability of the nitronium ion [4]. Nitric acid in sulfuric acid/water is half ionized when the mixture contains 88% sulfuric acid; ionization is essentially complete at an acid strength of 94% [5]. Note the very sharp fall in the rate as the acid concentration decreases below 88%.

Figure 4.5 Effect of sulfuric acid concentration on the ionization of nitric acid to provide nitronium ion, and on the rate constant ($k2$) for nitration of *p*-dichlorobenzene.

4.6.2
Acylation

A qualitative understanding of the principles involved is very helpful in assessing the process options. Amine acylation is a common procedure used to exemplify the principles involved [6]. This process can often be carried out in aqueous or two-phase organic/aqueous systems. Competitive solvolysis of the acylation agent causes yield loss.

Over the pH range commonly used for acylation, the predominant side reaction of the acylating agent is hydrolysis by water or hydroxide ion. A plot of $\log_{10}$ (solvolysis rate) versus pH shows a discontinuity at the pH level where the hydroxide rate begins to dominate (pH$_i$ is the pH at which the rate of the water reaction and the hydroxide reaction are the same). Figure 4.6 shows such a plot for benzene sulfonyl chloride. Availability of the amine increases with pH up to the point where no amine remains unprotonated. Figure 4.7 shows a plot of $\log_{10}$ (fraction of amine as free base/total amine) for 0.01 M allylamine (pK_a 9.49). (When the pH equals the pK_a, half the amine is protonated and half unprotonated.)

The effect of these two competing factors is that the selectivity of the process shows a maximum with respect to pH (Figure 4.8).

The pH providing the maximum selectivity is halfway between pH$_i$ for the electrophile and pK_a for the amine; thus awareness of this analysis enables choice of the optimum pH for the process simply by finding these two values, often available in the literature, for example, [6].

A similar analysis has been applied recently to optimize the reaction between 4-methylsulfonylaniline and cyanamide [7].

Many reactions involving additions to carbonyl groups are reversible, and the equilibrium position of such reactions will therefore be pH and concentration sensitive. Figure 4.9 shows the effect of these parameters on the calculated fraction of 4-chlorobenzaldehyde oxime formed at equilibrium, using a pK_a of 6.52 for

Figure 4.6 Solvolysis rate of benzenesulfonyl chloride versus pH.

Figure 4.7 Log10 (fraction as free base) versus pH for allylamine.

hydroxylamine and an equilibrium constant K_e of 24 for the formation of the oxime from neutral hydroxylamine [8].

Figure 4.8   Graph showing how selectivity for acylation versus solvolysis varies with pH.

Figure 4.9   Fraction of oxime formed at equilibrium versus pH and final hydroxylamine concentration.

These data show that, for a homogeneous reaction, the equilibrium position is highly sensitive to pH, and requires a high concentration of hydroxylamine to drive the reaction to completion. In practice, the low solubility of the oxime will permit the achievement of high conversion at a lower pH and hydroxylamine concentration than might be expected on the basis of the homogeneous data.

4.6.3
Association Equilibria – Lithium Diethylamide (LDA)

Lithium enolate chemistry is widely used in synthesis, and is dominated by association/aggregation effects. The work of Collum provides many useful insights into this complex subject, for example, Ref. [9].

4.7
Kinetics

Whether a reaction is operated in batch or fed-batch mode will itself provide some information as to the process kinetics and thermochemistry [10]. A "slow" reaction requiring significant driving concentrations of both reactants is unlikely to have fast kinetics (although a multiphase reaction with poorly soluble reactants is an exception to this generalization). "Fast" in this context means a reaction taking a few minutes or less to complete. While a detailed understanding of the homogeneous reaction kinetics may not be necessary for the optimization of a batch process, answers to some general questions will be helpful:

• Is the reaction reversible?
• What is the reaction order?
• Is the reaction slow enough for one or more reactants or intermediates to accumulate?
• For a fed-batch reaction, at what rate is a small amount of added reactant consumed?
• Is the product stable on the timescale of the process?

4.7.1
Order of Reaction

The reaction order has a great influence on completion times and on the effect of concentration on process performance. Figure 4.10 shows the profiles for first- and second-order reactions and batch reactions.

Figure 4.10 Reaction profiles for starting material loss, rate constant unity.

Figure 4.11 Simple first-order reversible reactions with A <-> B

The time to complete 99% of a first-order reaction is 4.6 half-lives, whereas for a second-order reaction with equal concentrations of reactants it is 99 half-lives (Figure 4.10). Knowledge of reaction order is therefore essential for predictive purposes.

More complex profiles are frequently seen, and useful diagnoses can be made from the overall shape of the profiles. Some examples are shown below:

Figure 4.11 shows the behaviour of a simple first order reaction with $k_{forward} = 2 \times k_{reverse}$. The observed rate of approach to equilibrium, measured by the approach rate of either component to the equilibrium value, is the same, and has a rate constant equal to the sum of the individual rate constants.

When there are a different number of species on either/each side of the equilibrium, then the equilibrium position will be concentration sensitive. Common examples are carbonyl condensation reactions such as Mannich and Strecker reactions, and cyanohydrin formation. For the oxime formation shown earlier, the water concentration is effectively constant and so the equilibrium constant K at a high pH is given by $K = \frac{[Ox]}{[NH_2OH][ClBA]}$. If the process is operated at equal starting concentrations of hydroxylamine and aldehyde, then the equilibrium yield in solution is as shown in Figure 4.12.

In this situation, a good option is to operate under conditions where the product has low solubility, so that the reaction is driven closer to completion by precipitation of the product.

A reversible reaction with a parallel first-order reaction of starting material is C ← A ⟷ B, where $k_{AB} = 1.0$, $k_{BA} = 0.05$, and $k_{AC} = 0.5$ (Figure 4.13).

In this case, a pseudoequilibrium is reached rapidly and thereafter C is formed in a reaction in which the rate determining step is B → A.

A + B $\xrightarrow{k_1=1}$ I + A $\xrightarrow{k_2=1}$ P is a consecutive reaction that gives a desired product P, where initial [A] = 1, [B] = 2 (Figure 4.14).

The early lag in the formation of the product is due to build up of intermediate *I*.

Figure 4.12 Equilibrium yield of oxime in solution vs starting reactant concentrations.

Figure 4.13 Reversible reaction with parallel first-order reaction of starting material.

Figure 4.14 Graph showing how material fractions change over time for consecutive reactions.

Reaction products may be unstable, or may undergo further reaction with an added reagent. Evaluation of extended hold times, such as those that may occur due to unplanned delays, is therefore essential.

4.8
Catalyzed Processes

The behavior of catalyzed processes can be complex, and some knowledge of the catalysis mechanism is essential for a confident understanding of the process. This area is complex, but has its own extensive literature. Blackmond [11] has described an efficient methodology for acquiring an understanding of the effect of process variables on performance.

Workers from Eli Lilly have provided an instructive example of the application of mechanistic understanding from the literature, combined with carefully targeted experimentation, to greatly reduce problem-causing impurity in a Suzuki cross-coupling [12].

The desired cross-coupled product was contaminated with an unacceptable amount of the dimer derived from the boronic acid (Imp 6). The initial process used Pd(OAc)$_2$ as the catalyst. It was shown that the undesired dimer resulted from a stoichiometric reaction between Pd(II) and boronic acid, and the formation of homo-dimer was exacerbated by the presence of oxygen. The use of Pd black as a catalyst, in conjunction with removal of oxygen by nitrogen sparge, reduced the impurity to an acceptable low level.

Reaction pictures, as exemplified earlier, are invaluable in developing and sharing an understanding of complex processes.

4.9
The Rate-Determining Step

Considerations of the Interplay between Chemical and Physical Rate processes and their Impact on the Process Outcome.

Much of the first section of this chapter focuses on understanding the chemical rate processes that occur. In most chemical reactions, however, there is another set of processes that takes time to occur; these are physical rate processes and include the following:

- Rates of solid dissolution
- Heat transfer
- Mass transfer
- Rate of crystal growth
- Mass transport (transport of species within a single phase leading to local concentration effects, e.g., in a mixing plume where reagents are added).

Many process development chemists have not been taught about these physical processes as part of their studies and this is often dismissed as ''just engineering.'' The authors strongly disagree with this point of view and hope to demonstrate in this section that they can have a profound effect on the process outcome.

It is important to first know that:

- Chemical reaction rate constants *do not* change with scale or equipment.
- Physical rate processes *do* change with scale and equipment.

This means that, although a chemical process may deliver the desired selectivity, purity, and yield at a certain scale and item of equipment, if the scale or equipment is changed without sufficiently understanding the physical rate processes, then the process outcome can be severely impacted.

This can be exemplified by the reaction scheme described in Figure 4.4 in Section 4.5 of this chapter.

The reaction is a heterogeneous Pd/C-mediated reduction of a nitrile to the corresponding amine using molecular hydrogen. The main impurity generated in this process is the dimeric amine formed when the product reacts with the intermediate amine.

Let us first identify the chemical rate processes:

Notes

1) The loss of ammonia from the diamine is fast compared to the formation of the diamine.
2) For the sake of simplicity, we assume that the reactions are of first order in each of the reagents.
3) For simplicity, we have also not discriminated between reagents that are adsorbed onto the catalyst surface and/or are free in solution.

Figure 4.15 Nitrile reduction showing chemical rate processes.

From the reaction scheme above (Figure 4.15), it can be seen that the competing reactions are between the reduction of the imine to desired amine product and the reaction between the imine and the amine product (the ''branch point'' discussed earlier). Considering the competing rate equations:

Rate desired amine formation α [imine].[Pd].[H$_2$]

Rate undesired diamine formation α [imine]·[product]

Thus, at the *molecular level*, a *high* concentration of Pd, a *high* concentration of H$_2$ on the catalyst surface, and a *low* concentration of the product are desired. All of these factors are controlled by physical rate processes, which must be understood to achieve the desired process outcome of high yield and low diamine formation.

If we add the physical rate processes to the chemical rate processes, we can see that the number of rate processes has increased. This means that the complexity of the system has also increased. We can also see that, to control the local concentrations of the reactants, we now need to consider how to control the occurrence of physical rate processes (Figure 4.16).

Figure 4.16 Nitrile reduction scheme with chemical and physical rate processes.

From the scheme above, it quickly becomes apparent that the concentration of hydrogen in solution is controlled by the mass transfer rate of the hydrogen from the gaseous phase into solution and the available concentration of palladium. The process outcome is now entirely dependant on the physical rate process.

Assuming that all other aspects are kept constant:

1) If the mass transfer rate of hydrogen into solution is *fast*, then the rate of imine reduction will be significantly faster than the diamine formation, leading to *high* yields and *low* impurity formation.

2) If, however, the mass transfer rate of hydrogen into solution is *slow*, then formation of the diamine will dominate, leading to *low* yields and *high* impurity formation.

While this example may seem trivial and it is well known that hydrogenation rates are strongly dependant upon the mass transfer rates from the gaseous phase into the liquid phase, many process technologists understand this only at an empirical level without going into the detailed reasoning.

Indeed, in this example it is also equally important (though frequently overlooked) to ensure that the catalyst is well suspended in solution; from an engineering standpoint, it is difficult to design equipment that provides both good G/L mass transfer *and* good solid suspension.

It is vital to understand that

chemical reaction rate constants ***are not*** scale and equipment dependent

but

physical rate processes ***are*** scale and equipment dependent.

This means that as the equipment or scale is changed, the intrinsic chemical rates will not change, whereas the physical rates can change. This means that the rate-determining step can also change; needless to say, this can be bad news if you are not expecting it!

4.10
Mixing in Chemical Reactors

Consider a simple single-phase bimolecular reaction of the form

$$A + B \rightarrow P$$

The primary need for this reaction to proceed is to bring the molecules of A and B into contact with each other. This contact is provided by the mixing of the reactants, and hence the mixing capability of the reaction equipment can have a significant influence on the performance of the reaction. In this section, we provide an overview of the effects that mixing can have on a reaction. Further information on mixing can be found in the texts by Paul [13] and Harnby *et al.* [14].

Standard approaches to reactor design assume that the contents of a reactor are well mixed, and hence provide homogeneous conditions for the reaction to take place. However, in real systems where separate feed streams containing the reactants are brought into contact, there will inevitably be a period of time when the conditions in the reactor are not homogeneous. For the simple bimolecular single-phase reaction system described above, if there are no possible side reactions, then this does not matter – if the intrinsic chemical reaction rate[1] is faster than the rate at which the contents of the reactor can be mixed, then the observed reaction rate is limited to the rate of mixing. However, if one of the reagents is able to react with product, say

$$A + P \rightarrow D$$

and this reaction also has fast intrinsic kinetics, then the rate of mixing determines the selectivity. Consider what happens in the region of the reactor immediately surrounding the initial contact between A and B. In this region, A and B react to form P. If the rate of mixing in this area is sufficiently fast (of the same order of magnitude as the intrinsic reaction rate), then P is likely to be dispersed into the bulk mixture and is not likely to undergo any further reaction. However, if the rate of mixing in the reaction zone were significantly less than the intrinsic reaction rate, then before P is dispersed into the bulk mixture it could come into contact with A and undergo the undesired reaction. The greater the difference between the intrinsic chemical rates, the greater is the effect on selectivity – for very fast reactions in equipment with very slow mixing, all the desired products could undergo undesired reactions. While this example has considered a case where a consecutive reaction is undesired, similar considerations apply to cases where there are competing parallel reactions; see Bourne [15].

4.11
Mixing Theory

Chemical engineers are generally more familiar with the science of mixing than chemists. At the molecular scale, mixing occurs by diffusion, and the mixing time at the molecular scale, t_m, is related to the diffusivity of the diffusing material in the bulk material, D, and the length of the diffusion pathway, x:

$$t_m \propto \frac{D}{x^2}$$

Table 4.3 presents the diffusion times calculated for a liquid–liquid and gas–gas system for a range of diffusion path lengths. It is seen that in order to comprise rapid mixing it is necessary to have very small diffusion pathways.

---

1) The intrinsic chemical reaction rate is the natural rate at which the reaction would take place if all residence time, heat transfer, mixing and mass transfer requirements could be met.

Table 4.3 Diffusion times for liquid–liquid and gas–gas systems.

Diffusion path length (mm)	Diffusion time nitric acid–water (s)	Diffusion time carbon dioxide–air (s)
0.01	0.03	7×10^{-6}
0.1	3	7×10^{-4}
1.0	340	0.07
10	34 000	7

Table 4.4 Flow regimes in process equipment.

Flow regime	Characteristics	Characteristic of equipment/material	*Re* in pipe flow	*Re* in stirred tanks
Laminar	Flow in streamlines; no natural random movement; poor mixing in large-scale equipment	Microreactors and micromixers Low-velocity flows High-viscosity fluids	<2000	<10
Transitional	In-between laminar and turbulent regimes; potentially not reproducible		2000–4000	10–10 000
Turbulent	Random fluctuating velocities imposed on the main flow direction, structures at a range of lengths meters to micrometers, high energy, well mixed.	Larger-scale traditional process equipment High-velocity flows Low viscosity fluids	>4000	$>10\,000$

The manner in which these path lengths are achieved is dependent on both the type of equipment employed for the reaction and the flow regime within it. Table 4.4 presents the key characteristics of the flow regimes encountered in a range of process equipment types. *Re* is the Reynolds Number (see also Chapter 6), a dimensionless group used to identify the flow regime in a given system.

In pipes and channels, *Re* is given by

$$Re = \frac{\rho u d}{\mu}$$

where

ρ is the density of the fluid,
u is the velocity of the fluid in the pipe,

d is the diameter of the pipe, and
μ is the dynamic viscosity of the fluid.

The value of *Re* in a stirred tank is given by

$$Re = \frac{\rho N D^2}{\mu}$$

where

N is the rotational speed of the agitator, and
D is the diameter of the agitator.

Any consistent system of units can be used.

The assessment of *Re* using the above equations assumes that the fluid rheology is Newtonian – that is, there is a linear relationship between the shear stress imparted on the fluid and the shear rate induced by the agitator, the gradient of which is the viscosity. This can be assumed for most simple liquid systems, though there are a number of non-Newtonian rheologies that could be exhibited by more complex systems, for example:

- **Bingham plastics** – the fluid exhibits a yield stress (a minimum shear stress that has to be imparted by the mixer) before flow will occur.
- **Shear thickening fluids** – the viscosity increases with increasing shear rate.
- **Shear thinning fluids** – the viscosity decreases with increasing shear rate.

These fall outside of the scope of this text, but would need special consideration during process development to ensure that adequate mixing is delivered by the equipment if such systems were found to exist.

The streamline nature of laminar flow means that for successful mixing we need to engineer the contact between the fluid streamlines to ensure that diffusion can occur over appropriate path lengths to deliver an acceptable mixing time. This is the principle on which micromixers and microreactors work.

In stirred tanks and larger-scale continuous tubular reactors, turbulent flow is required in order to deliver the necessary diffusion path lengths. The random nature of turbulent flow is due to eddies that deliver locally short diffusion path lengths, thereby achieving good local mixing in the eddy. The energy imparted by the mixer will determine the size distribution of these eddies within the reaction mass, and hence the effectiveness of the mixing. In continuous flow systems, this mixing energy is determined by the pressure drop (a straightforward chemical engineering calculation). In stirred tanks, the mixing power is determined by the agitator type and internal arrangement of the tank itself. This is discussed further in Chapter 6.

4.11.1
Mixing Regimes

In order to understand the impact of mixing on the outcome of the reaction, four mixing regimes have to be considered. These are outlined in Table 4.5.

Table 4.5 Mixing regimes.

Mixing regime	Order of magnitude characteristic time	Reaction performance sensitive to . . .	Chemical selectivity in stirred tank reactors determined by . . .	Chemical selectivity in turbulent flow continuous reactors determined by . . .
Micromixing	Milliseconds	Mixing effects close to the point of reagent addition	Local turbulence	Local effects of bulk turbulence
Mesomixing	Seconds	Mixing effects close to point at which mixing energy is introduced	Reaction kinetics, circulation, feed rate, local turbulence at point of introduction of fed materials	Reaction kinetics, feed rate, local turbulence at point of introduction of fed materials
Macromixing	Tens of seconds to minutes	Mixing effects in the bulk fluid	Reaction kinetics, bulk circulation	Reaction kinetics, bulk mixing time, residence time. (Not likely to be an issue in continuous equipment because of the fast mixing compared with stirred tanks.)
Independent of mixing	10 min or greater	Reaction kinetics only		

The case where the reaction is independent of mixing has been covered by the simple bimolecular reaction discussed in the introduction to this section, and will not be discussed further. However, the other mixing regimes will now be considered in more detail.

4.11.2
Micromixing

Reaction selectivity is likely to be sensitive to micromixing when the product and by-product reactions have intrinsic reaction times of the order of milliseconds, and the concentration gradients are at length scales equivalent to the turbulent eddies at the point of mixing in the stirred tank, or to the size of the channels in a micromixer or microreactor. From Table 4.3, we can see that the diffusion path length is of the order of magnitude of tens of micrometers. The Baldyga and Bourne mixing model (see Baldyga and Bourne [16]) offers a method for determining whether a reaction might be micromixing sensitive if carried out in a stirred tank. Their approach is

based on the Kolmogorov length scale for turbulent eddies described further in Chapter 6.

4.11.3
Macromixing

Reactions are likely to be sensitive to macromixing, or bulk mixing, when the intrinsic reaction rate of the undesired reaction is of the order of tens of seconds to minutes. For these reactions, it is essential to achieve a well-mixed system, by which we mean a system where the variance in concentration between samples taken from that system is small (<5%) regardless of the location of the sample or the sample size. In a stirred tank, where the mixing occurs in turbulent flow, the time taken to achieve this variance is the bulk mixing time, $t_{95\%}$ – the fluids can be said to be 95% mixed at $t_{95\%}$. While the bulk mixing time is an arbitrary parameter, it is defined such that vessels can be compared on a consistent basis and can therefore be used as a scale-up parameter. This is discussed further in Chapter 6. $t_{95\%}$ can be measured experimentally by adding an inert tracer to a liquid of the same physical properties and sampling the change in concentration with time.

Bulk mixing is also an important consideration for reactions that are particularly exothermic and/or are sensitive to temperature. The bulk mixing time will affect the rate at which heat can be dissipated from the reaction zone to the cooling jacket and/or coils. With inadequate bulk mixing, heat transfer occurs primarily by conduction, a process analogous to diffusion. Over long conduction pathways, the rate of heat transfer will be slow. Turbulent eddies introduce convective heat transfer, and increasing turbulence significantly improves the heat transfer rate. Therefore, adequate bulk mixing should be provided to ensure that there are no local temperature ''hot spots'' at any point in the reactor where the local reaction rate could be higher. The author recalls one process where bulk mixing was not given adequate consideration – the temperature controller (located away from the agitator) indicated the desired bulk set-point temperature but it was later found that there was a poorly mixed zone in the vessel at a temperature high enough to boil away the process solvent and thermally degrade the product, at a significant cost to the company!

4.11.4
Mesomixing

Mesomixing-sensitive (mixing at the droplet scale) reactions are characterized by the rate of the undesired reaction being of a timescale similar to the rate at which the reacting materials are drawn through the initial mixing zone, that is, the agitator in a stirred tank. A typical example would be a fed-batch reaction that is sensitive to a high feed rate of the fed component, where the undesired reaction takes place before the plume of the fed material present at the inlet can be dissipated by the turbulence within the vessel. Whether a reaction system might be mesomixing

sensitive can be known by determining the micro- and bulk-mixing times for the reactor and checking whether the intrinsic reaction times fall between these values.

4.11.5
Determining Mixing Sensitivity in the Laboratory

A simple analysis of mixing sensitivity can be carried out during process development to avoid many problems that could occur during any subsequent scale-up activity. The answers to the simple questions below can give an indication of whether there are any mixing sensitivities to consider:

• Does the reaction perform differently if the reactants are added to each other in reverse order?
• Is any by-product formation a result of consecutive reactions occurring early in the process?
• Does the reaction performance change if the agitator is run much slower than normal?

Bourne [15] developed a straightforward laboratory procedure for determining whether reactions could have any mixing sensitivities. This is shown in Figure 4.17.

4.11.6
Comments on Scalability of Mixing

Chapter 6 discusses scale-up rules for mixing in chemical reactors. However, when carrying out laboratory experiments during process development, it can be useful to apply such scale-up rules in reverse. In effect, this means using well-designed lab equipments that can deliver mixing performance characteristic of the equipment available in the pilot plant and/or in the production plant. There are some important points to consider when considering equipment for lab experiments:

Figure 4.17 Experimental flowchart for characterizing mixing sensitivity [15].

- Baffles can be used to enhance mixing performance in stirred tanks at all scales in most of the cases. They are useful even in lab glass equipment to improve the understanding of the likely mixing behavior.
- High agitator speeds require excessive power at larger scales. The performance of the magnetic flea cannot be reproduced even in a pilot-scale vessel!
- Using lab vessels with geometry similar to that of likely plant equipment will result in fewer issues during scale-up.

4.12
Multiphase Processes

As noted earlier in this chapter, two-thirds of industrial processes involve at least two material phases during the reaction. However, in most of these processes, the reaction itself occurs in only one of the phases. If the intrinsic rate of reaction is fast, then the rate of mass transfer of material into the reacting phase will become the rate-determining step.

For many processes, the key issue during development is to maximize the selectivity of the desired reaction, and the presence of multiple phases gives rise to issues similar to those of single-phase mixing sensitivity discussed above. However, the physical complexity is potentially much greater, and an appreciation of all the fundamental physical phenomena that could occur is vital for complete understanding of such processes. Of course, for real industrial processes, it is often not possible to get this full understanding, even in qualitative terms, because the extensive research that would be needed is difficult and expensive. However, some understanding of the fundamental rate processes is useful in selecting the important factors for study in process development and in interpreting the results.

Factors which may be important for study include the following:

- In which phase does each reaction (desired and undesired) occur?
- How are the species present in the reaction mixture distributed between phases, that is, what are the phase equilibria?
- For multiple fluid phases, which is continuous and which is dispersed? Do the droplets, bubbles, or particles of the dispersed phase have varying experiences in the reactor which gives rise to varying selectivity?
- How do the local concentrations of the various species vary as a result of the mass transfer and reaction?

To interpret these issues, the process technologist clearly needs an understanding of the physics of reaction kinetics, phase equilibria, mass transfer, and so on. A useful activity to help identify potential issues in a multiphase system is to draw a reaction picture such as that illustrated in Figure 4.4, showing the distribution of phases, species, and reactions and the physical interactions (and therefore rate processes) between them. While chemists will appreciate the chemical rate

processes, it is important to involve a chemical engineer in the development of such a reaction picture in order to help identify the various physical rate processes that could be important for study.

4.13
Mass Transfer Theory

Mass transfer is the chemical engineering science of the transport of material between phases. Several researchers have proposed models to describe mass transfer (see Treybal [17]), for example, the Whitman two-film theory illustrated for two liquid phases in Figure 4.18.

In Figure 4.18, C_x refers to the concentration of the reagent in the bulk phase x, with $C_{x,\,i}$ being the saturated concentration at the interface between the two phases. The distance between the dotted lines and the solid line represents the diffusion, or mass transfer, film in each phase – as with mixing, interphase mass transfer is a diffusion-controlled process. The rate of diffusion through each film is proportional to the mass transfer coefficient k, represented by the concentration gradient through the mass transfer film. Typically, the diffusion film thickness would be between 50 and 100 μm, with a diffusion time of the order of seconds. Note that similarly the Whitman two-film theory applies to liquid–solid and liquid–gas systems.

The rate at which mass transfer occurs through either film, r, is proportional to the product of the overall mass transfer coefficient of the system (meter per second) and the specific area of the interface between the two phases (meter square per meter cube).

$$r \propto k_L a$$

Note that k_L is the overall liquid phase mass transfer coefficient, but this could similarly be defined for a gas phase as k_G.

Figure 4.18 Whitman two-film theory of mass transfer.

It is seen that, by increasing or decreasing a, the process technologist can increase or decrease the mass transfer rate (e.g., by increasing or decreasing liquid droplet, solid particle, or gas bubble size).

4.13.1
Effect of Mass Transfer on Chemical Reaction Rates

Chemical reaction rates can be affected by the rate of mass transfer in multiphase systems. The effect of mass transfer will depend on which of the following cases best describes the nature of the reaction.

4.13.1.1 Chemical Rate-Limited Reaction
In chemical rate-limited reactions, the reaction proceeds at the intrinsic reaction rate with no effect of mass transfer. This is typical not only of single-phase systems but also of multiphase systems where the rate of mass transfer is faster than the intrinsic reaction rate. Here the observed reaction rate can be considered to be proportional to the chemical rate constant.

4.13.1.2 Mass Transfer Rate-Limited Reaction
In these reactions, the intrinsic reaction rate is faster than the rate at which mass transfer occurs; hence, the observed reaction rate is limited by the transfer of material into the reacting phase. Here the observed rate of reaction is proportional to $k_L a$.

4.13.1.3 Solubility-Limited Reaction
This category is a special case of the chemical rate-limited reaction where the solubility of one reactant in the reacting phase is low and hence its concentration limits the reaction rate. A change in reaction solvent might help increase the reaction rate but any changes in reaction mechanism as a result of changing solvent (and hence the overall effect on selectivity) need to be understood.

Other physical effects on the chemical reaction rate that should also be considered include the following:

- Heat transfer rate-limited reaction – this is the case where an exothermic reaction is deliberately slowed down so that the rate of heat evolution can be accommodated by the heat transfer equipment without causing a significant change in the reaction temperature (typical of the fed-batch reaction).
- Mixing limited reaction – the intrinsic chemical rate is faster than the rate at which mixing can occur. This is not common as these reactions are usually highly exothermic, and therefore tend to fall into the heat transfer limited category.
- Reactions that generate gas – the reaction rate has to be limited to the rate at which gas can disengage from the bulk liquid phase safely.

There are multiphase reaction systems where the intrinsic reaction rate is fast enough such that the reaction goes to completion inside the mass transfer film. In these cases, an alternative treatment is needed as the mass transfer and reaction rate processes can no longer be considered as taking place in series. The observed reaction rate is proportional to both the square root of the chemical rate constant and the specific interfacial area between the phases.

4.13.2
Phase Equilibria

The mass transfer rate process is analogous to a reversible chemical reaction such that it will reach an equilibrium state. There are a number of simple equilibrium calculations that can help us understand system behavior.

4.13.2.1 Partitioning
In multiphase systems, neutral species will partition themselves between aqueous and organic phases. The partition coefficient, K_D, which is typically available from the literature, is the ratio of concentration of material, c, in each phase:

$$K_D = \frac{c_{\text{org}}}{c_{\text{aq}}}$$

The octanol/water partition coefficient, P (often published as $\log_{10} P$ values) is often used. This is a good analogy for biological systems and enables the process technologist to understand how materials might partition in such systems:

$$P = \frac{c_{\text{octanol}}}{c_{\text{aq}}}$$

4.13.2.2 "Salting Out"
Salting out is a technique that can be used to alter the solubility of a material in a particular phase where this may lead to, for example, more favorable separation conditions. The principle is that increasing the salt concentration in a solution will attract water molecules, thus decreasing the number of "free" water molecules available to interact with any non-electrolytic species in the system. Stetschenow's correlation relates the solubility of the nonelectrolytic species with the concentration of dissolved salt:

$$\log_{10}\left(\frac{S_w}{S_e}\right) = k_s c_s$$

where S_w and S_e are the solubilities of the nonelectrolyte in water and in saline solution with a salt concentration c_s. k_s is a constant specific to the system being studied.

4.13.2.3 Common Ion Effect
The common ion effect is similar to salting out in that the solubility of a species in a solution is altered by adding another species with a common ion. The increase

in the concentration of the common ion affects the solubility of the species initially present. The common ion effect is used in water softening, where highly soluble sodium carbonate is added to the water to precipitate out sparingly soluble calcium carbonate. The key factor is the solubility product, *S*, which remains constant with changing concentrations of the common ion. For a salt with chemical formula M_mX_n:

$$S = \left[M^+\right]^m \left[X^-\right]^n$$

Thus, as the concentration of X^- increases another X^- salt is added to keep *S* constant; the concentration of M^+, and thereby the solubility of M_mX_n, must be reduced.

4.14
Mass Transfer and Mixing Requirements in Multiphase Systems

Consideration of mixing requirements for multiphase systems still requires an understanding of the mixing sensitivity of the reaction itself. However, there are additional complexities to consider as well, which are described for various systems below.

4.14.1
Liquid–Liquid Systems

Systems involving two immiscible liquids rely on the dispersion of droplets of one liquid phase (the dispersed phased) into another (the continuous phase) for mass transfer to take place. For a given liquid–liquid system, it is possible to create two different physical systems, as shown in Figure 4.19, which exhibit different physical behaviors.

The dispersion formed is a function of the density and viscosity of the two liquid phases, the mixing energy imparted to the system, and also the order of addition. It is important to understand which phase, when dispersed, offers the best reaction

Figure 4.19 Two possible liquid–liquid dispersions.

Figure 4.20 Sample phase continuity diagram.

outcome as the chemical performance of the system can be significantly affected if the wrong dispersion is generated.

To understand the effect of changing the volume fractions of the two phases, it will be useful to refer to a phase continuity diagram of the system. An example is shown in Figure 4.20.

In Figure 4.20, there are three distinct regions – one where the light phase is always the continuous phase irrespective of mixing intensity, one where the heavy phase is always the continuous phase, and the ambivalent region where either phase could be continuous.

Consider what happens when a light phase material is added to a continuous heavy phase, illustrated by the arrow between points (i) and (ii) in Figure 4.20. As the volume fraction of light phase material is increased, the heavy phase will remain continuous from point (i) through the ambivalent region to point (ii) as long as the mixing intensity remains constant. At point (ii), the volume fractions meet the heavy to light continuous phase inversion boundary. Phase inversion will occur when the volume fraction of the light phase increases beyond point (ii), and the light phase will become the continuous phase. The same is true for the opposite case moving along the line between points (iii) and (iv) in Figure 4.20 – as long as the mixing intensity remains constant, the light phase will remain continuous until the heavy phase volume fraction increases beyond point (iv) at the light to heavy continuous phase inversion boundary.

For many processes, the volume fractions of light and heavy phases are likely to be within the ambivalent region. For these processes, the material first charged to the reactor will be the continuous phase. Because the process lies within the ambivalent region it is important to understand what will happen if the agitation is stopped and restarted. It is possible that if mixing is lost and phase separation occurs, then when mixing is reinstated the previously dispersed phase could become the continuous phase, and vice versa, with a detrimental effect on the reaction performance.

The scale-up of vessels for liquid dispersions can be carried out using the scale-up rules for stirred tanks as discussed in Chapter 6. The important parameters are the just dispersed agitator speed, N_{JD}, and the specific power requirement for the agitator to deliver either light-in-heavy or heavy-in-light dispersions. Empirical approaches to calculating these were developed by Lines [18].

4.14.2
Liquid–Solid Systems

Atherton *et al.* [1] showed that more than 60% of processes involve a solid material, and more than 40% involve a solid feed that is insoluble in the process solvent. As the solid material is generally denser than the liquid phase, a key mixing requirement to deliver adequate reaction and mass transfer performance in liquid–solid systems is that the agitator is capable of keeping the solids both suspended and well dispersed. A useful parameter to be determined for liquid–solid systems is the just suspended agitator speed, N_{JS}, this being the minimum required agitator speed to suspend the solids:

$$N_{JS} = \frac{S v^{0.1} d_p^{0.2} \left(\frac{g\Delta\rho}{\rho_l}\right)^{0.45} X^{0.13}}{D^{0.85}}$$

where

- S is a mixing-geometry-dependent constant,
- v is the kinematic viscosity of the liquid phase, m^2/s (= dynamic viscosity/density),
- d_p is the particle diameter, m,
- g is acceleration due to gravity, m/s^2,
- $\Delta\rho$ is the density difference between the phases, kg/m^3,
- ρ_l is the liquid phase density, kg/m^3,
- X is the mass percent of solid in the mixture, and
- D is the agitator diameter, m.

A useful rule of thumb is to employ an agitator speed of 10–20% more than N_{JS} in order to obtain a good distribution of solids.

Smaller, lighter solids that have a tendency to float at the surface of the bulk liquid phase can sometimes be observed. If this a common feature of the system, then employing a second agitator near the surface of the liquid should be considered in order to draw the lighter solids down into the bulk phase. As with all mixing problems, chemical engineers will be able to advise on possible solutions.

4.14.3
Gas–Liquid Systems

Good gas–liquid dispersions in stirred tanks result from the following:
- Selection of an appropriate gas dispersion agitator and baffles.
- Careful balancing of agitator speed and gas flow rate.

Assuming that an appropriate agitator and baffle combination has been chosen, then simplistically, increasing the gas flow rate at constant agitator speed results in a poorer dispersion of gas bubbles through the liquid, whereas increasing the agitation speed at constant gas flow rate results in a better dispersion of gas bubbles. Essentially, a good gas dispersion results from well-distributed fine gas bubbles throughout the bulk liquid phase.

Two useful parameters to be determined, which can help ensure good gas dispersion generation, are the flow number *Fl* and the Froude number *Fr*:

$$Fl = \frac{Q}{ND^3}, \ Fr = \frac{N^2 D}{g}$$

where Q is the volumetric gas flow rate, in m^3/s.

The agitator speed at which the gas begins to disperse can be determined from the following empirical relationship:

$$Fl \leqslant 30 Fr \left(\frac{T}{D}\right)^{-3.5}$$

where T is the agitated vessel diameter in m.

The agitator speed at which the gas is fully dispersed can be determined from another empirical relationship:

$$Fl \leqslant 0.2\sqrt{Fr}\sqrt{\left(\frac{T}{D}\right)}$$

Note that in plant equipment it could be difficult to obtain an agitator speed to deliver the fully dispersed gas condition without requiring excessive power, although it should be possible to exceed the just dispersed speed.

The best means of introducing gas for dispersion within a stirred tank are submerged inlets close to the agitator (spargers, dip pipes, or even hollow agitator drive shafts dispersing gas at or through the agitator itself). However, entrainment of gas from the head space is still common in many process vessels used for gas–liquid mass transfer operations. In these cases, it is useful to employ an agitator at the liquid surface to improve the rate of gas entrainment (which is otherwise limited by solubility and diffusion at the surface of the liquid).

4.14.4
Solid–Liquid–Gas Systems

Solid–liquid–gas systems encompass all the complexities inherent in the binary solid–liquid and gas–liquid systems discussed above. If we consider the reaction picture in Figure 4.4 above, then the mass transfer rate processes we need to consider are solid–liquid (solid nitrile subtrate dissolution and transport to the catalyst surface) and gas–liquid (hydrogen dissolution and transport to the catalyst surface). The agitation needs to be capable of

• suspending and dispersing the catalyst (to ensure even distribution of the reaction throughout the bulk solvent volume),

- suspending and dispersing the solid nitrile substrate (so that its dissolution is also well distributed throughout the bulk solvent volume),
- maximizing hydrogen entrainment at the liquid surface, and
- addressing any mixing sensitivity inherent in the pseudo-homogeneous phase chemistry.

This system is a good example where two agitators might be provided, one in the bulk liquid that delivers an acceptable balance between solid dispersion and mixing sensitivity, and one at the liquid surface to provide adequate gas entrainment.

4.15
Concepts of Structure and Scale for Equipment Selection

This section seeks to introduce two relatively new concepts that are useful when considering equipment selection. This chapter has already introduced many of the areas where process understanding is required to successfully design and operate a process. The next section covers how these various strands can be used to decide which equipment characteristics are required to best deliver the optimal process. It does not discuss specific equipment types in detail but provides an overview of what factors should be taken into account when considering equipment choices. This approach requires a multidisciplinary team to discuss these factors early during the development program so that key decisions can be made before the barriers to change become too large.

4.15.1
What Do We Mean by "Structure"?

Most value-adding operations in the chemical process industries occur at a molecular level; for example, an intramolecular reaction, or the movement of molecules relative to each other to achieve the separation of dissimilar molecules. Such molecular interactions are influenced solely by the conditions local to the molecule(s) involved and not by any equipment dimension. However, local conditions are determined by the mass and energy movement within the equipment through convection and/or diffusion.[2] The level of "structure" in a device is an indication of the level of control over the mass and energy movement.

> *Increasing the "**structure**" means increasing the*
> ***predictability*** *and* ***intensity*** *of the flow of reaction materials*
> *(can be solid, liquid, or gas).*

---

2) It should be noted that thermal conduction
   is analogous to material diffusion.

4.15.2
What is "Predictability"?

Increasing the predictability of flow of a reaction medium is important because it means that there is less variability in the processing conditions that the reaction mass will experience. In other words, all the molecules in the reaction mass will experience the same physical conditions as the reaction progresses (either spatially, temporally, or flowing down a continuous reactor). The benefit of this is that there will be less inherent variability in process outcome and the yield and purity are likely to be more consistent (and usually better).

Also, if the scale of structure can be the same at the development and production scales, then, by inference, the scaled-up process will provide the same behavior as in the laboratory.

4.15.3
What is "Intensity"?

In this context, intensity refers to the rate/speed at which the following physical rate processes occur:

1) Heat transfer
2) Mass transfer/mass transport.

In batch equipment, increased structure would typically mean adding baffles, using a more intensive type of agitator (e.g., moving from an anchor to a retreat curve to a four-blade pitched turbine agitator) or adding heating/cooling coils. This can lead to incremental improvements, but step-change performance can only be achieved by moving to smaller scale and more structured devices.

Increased structure can often lead to improved mixing, but this is not always the case.[3] For further information on this subject, ask a chemical engineering colleague!

Optimal process performance will usually be achieved when all molecules experience the same history and uniformity of local conditions; this is most readily achieved at small scale or high structure although it is not always required.

Increasing the structure will increase the rates of associated physical processes vs the related chemical rate constants (Figure 4.21). It will not necessarily improve performance. For example, there will be no benefit in increasing the level of structure beyond a batch stirred tank when the intrinsic chemical rates are slower (e.g., hours) than the bulk and local mixing rates (minutes to tens of minutes).

---

3) At macro-scale it can be as easy to make negative changes to equipment as positive changes (radial vs axial mixing, for example).

Figure 4.21 Effects of increasing structure.

Figure 4.22 Graph showing how appropriateness of structure can change with scale [19].

Increasing the structure usually also means increasing cost. Thus it is vital to balance the increased benefit from increasing the structure over the increased cost of the equipment, as can be seen in Figure 4.22.

4.15.4
Scales of Structure

The IMPULSE project (integrated multiscale processing units with locally structured elements, Sixth Framework Program of the European Commission, Project no. NMP 011816) has developed a number of methodologies that are useful in the development and implementation of a multiscale process, that is to say processes utilizing equipment with an appropriate level of structuring to deliver the required processing conditions. The IMPULSE project provides a very

Figure 4.23 Scales of structure and typical characteristics [20].

useful *aide memoir* to identify which types of equipment may be useful depending upon the scale of structure required (Figure 4.23).

The traditional production-scale chemical process equipment used in the pharmaceutical and fine chemicals industries falls mainly into the macrostructure category. The systems of equations derived for modeling and designing the equipment tend to consider the important physical phenomena at the macroscale. As the scale of process equipment gets smaller, different physical phenomena become important in the modeling and designing of equipment; these have not been covered in this chapter, but must be considered, particularly when moving to the microscale.

4.15.5
How Susceptible to Variability is the Process; When Would Different Equipment Help?

Some processes will be adversely affected by variations in local conditions (e.g., temperature or concentration) and it is important to find out if your process is one of these. Collecting the data identified earlier in this chapter will help in doing that. Considering each of the following areas in turn:

1) **Concentration**: What effect will any local variations in reagent concentration have on your process? This should be considered either at the meso or at the droplet scale and particularly around mixing/addition plumes.
2) **Heat**: Should be considered at two levels:
   a. **Macro or bulk scale**: Is the operating strategy and equipment able to provide sufficient overall heat transfer and take the peak heat load (energy flux)?

b. **Meso or droplet scale**: Examination of localized heating – particularly around droplets (or addition plumes) of reagents that are being charged. Does local heating from reaction at the edge of the droplet cause undesired reactions? This phenomenon is usually important only for medium to highly exothermic reactions.

3) **Time**: Is the process adversely affected by shorter or longer residence times? An example of this would be when a reagent is charged over time (fed- or semibatch) such that the product will be present in the reaction mixture for longer time than if all the reagents were charged at the start. This is likely to be detrimental if the product decomposes over time or can react with the reagents.

The authors recommend using reaction pictures, as described earlier in this chapter, to help visualize what spatial variations in local conditions may be occurring in a particular equipment item. Different items of equipment will inherently have different levels and types of variability.

If local conditions exist within your proposed equipment that may adversely affect the process, then an alternative processing strategy is likely required. This can be achieved by a change in operating strategy or a change in equipment (increase in structure).

As a rule of thumb:

If your chemical reaction rates are similar to, or faster than, mixing times offered by your proposed equipment then an intensified process (increased heat or mass transfer) may significantly improve your process outcome (yield, purity, etc).

If your reaction intermediates or products are known to be heat sensitive, then medium to highly exothermic reactions are likely to benefit from a high level of structure.

4.16
Conclusion

This chapter has introduced the importance of understanding both the chemical and physical rate processes that occur during chemical reactions. Historically, chemical reaction development has been the domain of the chemist and it is only after the "chemistry has been fixed" that chemical engineers become involved in the scale-up and manufacture of the reaction. This chapter has shown that this is NOT the most efficient way of carrying out process development and that it is vital to have chemical engineering input even at an early stage of project development.

This chapter has introduced a large number of phenomena that may be occurring during a reaction, which need to be understood. Not all of these phenomena will affect the rate, yield, or selectivity of every process. However, it is important that the process technologist consciously considers whether the phenomenon is likely to occur and then to test whether it does, rather than to carry on blindly and only find out later when things go wrong.

The authors have endeavored to demonstrate that only through understanding both the fundamental chemical and physical rate processes can a process be successfully designed and scaled-up without trusting to luck – and we all know what happens if we hope to be lucky all the time!

References

1. Atherton, J.H., Double, J.M., and Gourlay, B. (2005) Survey of PI equipment requirements in the fine chemicals and pharmaceuticals sector. paper presented at World Congress of Chemical Engineering, Glasgow.

2. (a) Landau, R.N., Blackmond, D.G., and Tung, H.-H. (1994) *Ind. Eng. Chem. Res.*, **33**, 814–820; (b) Blackmond, D.G. (2005) *Angew. Chem. Int. Ed.*, **44**, 4302–4320.

3. Hoggett, J.G., Moodie, R.B., Penton, J.R., and Schofield, K. (1971) *Nitration and Aromatic Reactivity*, Cambridge University Press.

4. Coombes, R.G., Grout, D.H.G., Hoggett, J.G., Moodie, R.B., and Schofield, K. (1970) *J. Chem. Soc. (B)*, 347.

5. Sampoli, M., De Santis, A., Marziano, N.C., Pinna, F., and Zingales, A. (1985) *J. Phys. Chem.*, **89**, 2864–2869.

6. (a) King, J.F., Rathore, R., Lam, J.Y.L., and Klassen, D.F. (1992) *J. Am. Chem. Soc.*, **114**, 3028–3033; (b) King, J.F., Guo, Z.R., and Klassen, D.F. (1994) *J. Org. Chem.*, **59**, 1095–1101; (c) King, J.F., Gill, M.S., and Ciubotaru, P. (2005) *Can. J. Chem.*, **83**, 1525–1535.

7. Chan, L.C., Cox, B.G., Currie, A., and Yates, S.D. (2007) *OPRD*, **11**, 984–984.

8. Sander, E.G. and Jencks, W.P. (1968) *J. Am. Chem. Soc.*, **90**, 6154–6162.

9. Sun, X. and Collum, D.B. (2000) *J. Am. Chem. Soc.*, **122**, 2459.

10. Cox, B.G. (1994) *Modern Liquid Phase Kinetics*, OUP.

11. Blackmond, D.G. (2005) *Angew. Chem. Int. Ed.*, **44**, 4302–4320.

12. Miller, W.D., Fray, A.H., Quatroche, J.T., and Sturgill, C.D. (2007) *Org. Proc. Res. Dev.*, **11**, 359–364.

13. Paul, E.L., Atiemo-Obeng, V.A., and Kresta, S.M. (eds) (2004) in *Handbook of Industrial Mixing: Science and Practice*, John Wiley & Sons, Ltd, Hoboken, NJ.

14. Harnby, N., Edwards, M.F., and Nienow, A.W. (eds) (1992) *Mixing in the Process Industries*, 2nd edn, Butterworth-Heinemann, Oxford.

15. Bourne, J.R. (2003) *Org. Proc. Res. Dev.*, **7** (4), 471–508.

16. Bałdyga, J. and Bourne, J.R. (1999) *Turbulent Mixing and Chemical Reactions*, John Wiley & Sons, Ltd, Chichester.

17. Treybal, R.E. (1980) *Mass Transfer Operations*, 3rd edn, McGraw-Hill, New York.

18. Lines, P.C. (1990) *IChemE Symp. Ser.*, **121**, 167.

19. Houson, I. and Gourlay, B. (2008) Introduction to Multi-scale Processing for Process Technologists, IMPULSE Project Training Materials. Available at *http://moodle-impulse.inpl-nancy.fr/ file.php/43/Index.htm* (accessed 2010).

20. Double, J.M. and Talford, M.C. (2008), Scale-related Phenomena in Multi-scale Processes, IMPULSE Project Training Materials. Available at *http://moodle-impulse.inpl-nancy.fr/ file.php/43/Index.htm* (accessed 2010).

5
Use of Models to Enhance Process Understanding

Wilfried Hoffmann

5.1
Introduction

One essential part of process development is the design of a process which is able to provide material in quantities of market volume with a defined quality and with additional constraints of cost, safety, and ecology. In an ideal situation, this process is developed in the laboratory and is then transferred to a manufacturing environment and can be used there without major adjustments or process changes.

In reality, this ideal situation is hard to achieve and the main reason is the influence of scale and equipment on quality attributes, the different weighting of cost, safety, and ecology factors and how these factors change with time, scale, business circumstances, and the complex interplay of all these components in a real process.

Cooling capacity, for example, is reduced on a larger scale, so that feed and temperature profiles and therefore the kinetics will change; mass transfer rates change with equipment, again leading to different reaction mixture profiles that may have consequences for the work-up and isolation steps.

There is obviously a desire to understand the reasons for this different behavior and to explore the options so that, by applying certain principles and procedures, it may become possible to predict the outcome of scale-up and to integrate this knowledge into the design of the manufacturing process.

These principles and procedures are the kinetic and thermodynamic characterization of the process with respect to scale and equipment. The knowledge that is required to apply these principles and procedures, is called, in the context of this chapter, *process understanding.*

In the following sequences, these principles and procedures are introduced mainly by using the chemical reaction step for illustration, as this is the area where most of these principles were developed first and where most of the knowledge has accumulated. This does not mean that these principles cannot be applied in other areas of a process, but the kinetics of a chemical system in a solution is more readily accessible than the kinetics of crystallization or the complex thermodynamics of multicomponent distillation.

On the basis of these insights, it is not surprising that one of the most powerful tools in these applications is process modeling. This modeling can be carried out at the qualitative, semiquantitative, and quantitative levels. Qualitative models are briefly discussed in the Chapter 4 by Atherton, Houson, and Talford while this chapter focuses on the semi-quantitative level of process modeling. The relationship between process understanding and process modeling is the theme of this chapter and a helpful structural tool to link these together is the concept of process characterization elements.

5.2
The Process Characterization Elements of a Chemical Reaction

If we want to develop a system to allow for the application of our principles to answer the question of the scale-up behavior, a closer look at a chemical process is required.

In the simplest case of a homogeneous liquid-phase reaction system, the major characterization elements are as follows:

1) **The chemical reaction networks** – this is the description of the chemical reactions based on mass balance, that is, the networks should show a closed system (balanced equations) and it should show the connections between the different species.
2) **The energy balance** – the description of the heat balance of the chemical reactions and the external balance by an external cooling or heating system.
3) **The mass flows entering or leaving the system** – in a simple case, the feed of a reagent over time (fed-batch or semibatch reaction).

    With the presence of more than one phase, the interactions between these phases come into play, and the existing three elements can be enlarged by the appropriate elements. More details about complexity can be found in Chapter 4.
4) **Solid/liquid** – this is the element that is responsible for the dynamic behavior of the dissolution of a solid into a liquid or the crystallization of a solid from a liquid.
5) **Liquid/liquid** – this element is used to describe the dynamics and thermodynamics seen, for example, in phase transfer reactions or in extractions used in the work-up.
6) **Gas/liquid** – this is the summary element for systems with the generation of a gas, the dissolution of a gas (for example, hydrogenation systems), and all systems where vapor–liquid equilibrium is important (reflux, distillation, etc.).

    In the case of fed batch reactions with fast kinetics, mixing effects can be very important [1] and this element may be included as well.
7) **Homogeneous mixing** – mixing effects in the zone of addition may have an impact on the selectivity. This can also include local heating effects where the reactions are exothermic.

A reaction at the laboratory scale can generally be analyzed for these elements. The question about the behavior of these elements on scale-up then is regarding how these elements translate from lab scale to large scale.

With respect to these elements, process understanding is the required knowledge of how these elements translate. This exercise needs to be done element per element, and the complexity of this task is caused by the interplay between these elements.

A simple example with the interaction of points 1 and 3 above illustrates this fact. Starting with element 1, we have to look at the chemical networks. Here, as long as chemical rates are only a function of concentrations, we should expect a simple 1 : 1 translation, that is, the same conversions can be expected at the same time for small scale and large scale. This is the traditional basis of scale-up. A reaction in the lab that gives 85% yield in 4-h reaction time is expected to give the same performance on scale. Now let us consider element 3. This probably will not scale 1 : 1, as unit operations generally take longer on a larger scale. It takes longer to charge the reagents, to bring them to the desired temperature and feed times, and final isolation times may be longer, and so on. As a consequence, there is more time for the reactions in element 1 to proceed. If the desired reaction is the only one with a significant rate, this may not be of concern, but if, for example, the starting material or the product exhibits some instability; this will change the concentration profiles so that element 1 may no longer translate 1 : 1. However, without the specific process understanding, in this case the understanding of the specific kinetics, this question cannot be answered.

Figure 5.1 shows schematically the connection of process understanding (the translation) as the link between the different elements of a process as they appear in the lab and their translation of how they would appear on scale.

The importance of process understanding cannot be underestimated (see Section 10.2 and Chapter 12), and any tool that helps analyze the elements, translates them to a large scale, and constructs new large-scale processes should be greatly appreciated. The standard approach to select the important elements is a series of

Figure 5.1 Process characterization elements and process understanding.

lab experiments. Here, the design and the final data interpretation are of major importance, and this is the stage where modeling tools enter the scene.

5.3
The Impact of Modeling

In the context of this approach, a model is any calculation or algorithm that can be used to reproduce experimental data and to predict the outcome of experiments under different conditions by application of first principles. These first principles are the laws of mass and energy conservation and the various rate descriptors required to reproduce the dynamics of system changes. In order to be useful as a support of process understanding, these models should be able to give insights into the translation of the various elements to a different scale or equipment.

Many of the models being discussed in this chapter use differential equation systems to describe the rates of the different parallel reactions and physical processes. There are many software products on the market with more or less user-friendly interfaces, which generate the appropriate differential equations in the background, and present the solutions in a graphical form for ease of comparison. Many of these have sophisticated functionalities such as kinetic fitting algorithms, which are extremely helpful in matching the model parameters to a set of experimental data, or they have access to large databases, which are used to calculate physical properties of a reaction system as a function of temperature and composition. But whatever the degree of sophistication (and cost) of these software packages, the way these can be used in process development is as follows.

The model has to be constructed to describe the elements as a result of the lab reaction analysis, the parameters of the system have to be adjusted from the analysis of the correct experimental data, physical properties and thermodynamic data from literature are added as required, and the system behavior on different scales and equipment is then calculated.

In this chapter, the important aspects of modeling are introduced by means of a model starting from a simple chemical rate description developing into a complex process, and the relationship of this development and the increase in process understanding is explained. The model presented is close to reality and is constructed from elements of different processes of Pfizer. This method avoids problems with disclosure of unpublished information and allows to focus on important features of process understanding. There is a generic benefit over using real experimental data as real data suffer from ''pollution,'' caused by experimental errors, impure reagents, and several nonrelevant impurities. Of course, these real experimental data need to be considered, and methods have been developed to treat experimental errors in process modeling; it is much easier to start with ''cleaned-up'' data generated by an underlying model to introduce the concepts of a model-based development approach.

5.4
Understanding the Chemistry

5.4.1
A Simple Start

The simple reaction to be considered is the bimolecular reaction of molecules A and B to form a product $A + B \Rightarrow P$, for example, a Diels Alder reaction. To predict the product concentration as a function of time at a constant temperature, all we need is the rate law, the rate constant at the temperature of interest, and the starting concentrations. The differential equation describing this reaction kinetics is a model and solving this equation as a function of starting conditions is a process simulation.

This simple model can also contribute to process understanding. Figure 5.2 shows the product formation profile of our hypothetical reaction at 60 °C with a starting concentration of $1\,\text{mol}\,l^{-1}$ for both reactants (rate constant = $2.7 \times 10^{-3}\,l\,\text{mol}^{-1}\,s^{-1}$).

This model shows that the reaction quickly gets to about 90% conversion (55 min), but it takes a much longer time to achieve 95% (2 h) or 98% conversion (5 h). So there are immediately two questions to be discussed. Is it worth investing 3 h more of reactor time to get 3% better conversion, or is there a way to get a higher conversion in the same time?

It is no secret that excess of one of the reactants increases the rate and therefore the conversion. Once the model is available, the conversion of any combination of starting concentrations as a function of time can be easily calculated. This would

Figure 5.2 Isothermal batch reaction.

allow the user to predict the cheapest overall process based on this model and knowledge of the raw material and plant-operating costs.

An important feature of a model is its capability of problem visualization. In the previous example, a rate constant value was required to predict the absolute time data of a desired conversion. However, the overall problem could have been visualized with any rate constant. This would have shown the effect of starting concentrations on the conversion as a property of second-order reactions, although only on a relative time basis. The advantage of using this ''visualization mode'' is the lack of the requirement to determine the model parameters, which is the most time consuming and most difficult part of modeling. The benefit of this for process understanding becomes more obvious with increasing complexity of the models. Estimation of model parameters in this case can give valuable insights into the sensitivity of parameters to the overall response, and from these trends, directions of process design can be explored while minimizing relatively expensive experimentation.

This is demonstrated in the next paragraph, when another important feature of a model is introduced, namely, the capability to grow with the growing process understanding.

Let us assume that experimental data of our bimolecular reaction system show that the product reacts with starting material A to an impurity I1.

To update our model to these facts, one additional mass balance and a rate law are added. Thus, the updated model may now look as follows:

$$A + B \xrightarrow{\ k_1\ } P \qquad r_1 = k_1[A][B]$$
$$A + P \xrightarrow{\ k_2\ } I1 \qquad r_2 = k_2[A][P]$$

It becomes obvious that the relative rates of these two reactions and again the starting concentrations of A and B eventually determine the product yield. From the process understanding point of view, this model reveals that high concentrations of A will favor I1. So probably a good way to carry out this reaction is by using an excess of B to accelerate the desired reaction. This may bring us to the idea of not running this reaction in batch mode, but of slowly adding a solution of A to an excess of B to ensure that at any time the concentration of A is low and therefore the rate of I1 formation is slow.

Even with the estimated rate constants k_1 and k_2, the above can be tested with the model.

5.4.2
Getting Real Rate Parameters

However, the feature of kinetic fitting allows a rapid evaluation of the rate constants from experimental data. On the basis of an underlying kinetic scheme (the rate laws), the rate constants are varied automatically till the sum of least squares is a minimum.

Here we run into a serious problem. How do we know that the underlying rate law scheme is appropriate? Of course, we may speculate that if the fitting is bad this might be caused by a wrong model, but here the influence of the experimental errors can be large as well. It is evident that large experimental errors will not give a good fit even with a "perfect" model. Experimental data therefore should be as accurate as possible and the first thing to do is to check if the experimental data are in agreement with the mass balance expressed by the model.

The mass balance is a list of components originating from each of the starting materials. In our example, we got two starting materials, namely, A and B. The reaction scheme shown above, which is a presentation of the mass balance, tells that at any time each molecule of A will either be unreacted A or be present in P or I1. As every molecule of I1 is made up of two moles of A, we can write the mass balance for component A as $A + P + 2 \times I1 = A_0$. Similarly, we get the mass balance for component B as $B + P + I1 = B_0$. When the experimental data are available, the mass balance for each component should be constant within a small percentage. If one of the mass balances is dropping, this is an indication that the reaction scheme has ignored an additional reaction of this component.

Here, we can use mechanistic models as much as possible. Mechanistic models describe the kinetic interaction on a molecular level, that is, one line per reaction, and here the rate laws are always strictly first order in each of the reacting components, that is, we get an overall first-order reaction for the type $A \rightarrow B$, and we get an overall second-order reaction for the type $A + B \rightarrow C$.

Complex overall rate laws and fractional overall reaction orders in empirical models are the result of multiple lines in a mechanistic model. The use of modeling software, which can handle these multiple line models, can thus render the use of overall rate descriptions obsolete, thus eliminating a big hurdle for chemists not too familiar with kinetics.

Thus, mechanistic models can be developed as a set of single-line elementary irreversible or reversible reactions and as a first guess these models can be set up by chemists from their mechanistic knowledge.

The disadvantage of mechanistic models is the much larger number of parameters (every line has a rate constant). However, not all rate constants have the same sensitivity on the overall rate behavior. In general, only the slow reaction rates control the overall rate, so that the rate constants of faster steps can be estimated or even guessed without a change of the overall kinetic behavior. Acid/base equilibrium reactions are generally considered to be very fast reactions, and if the rate of the reaction of one of the species involved in this equilibrium (for example, the nucleophilic reaction of a phenolate from a phenol/phenolate equilibrium) is much slower, the numerical value of the rate constants for equilibration can be fixed to a high value (may be >1000 times the rate constant of the slower reaction).

Looking at this the other way round, a kinetic fitting exercise will not be able to determine the value of the fast rate constants, as these concentration profiles will not be sensitive to them. The challenge here is to design a set of experiments that allows the extraction of the rate-controlling steps and the

numerical assignment of the associated rate constants. These experiments should deliver the information necessary and may look different to experiments designed to increase the yield.

Going back to our reaction, a single experiment was performed isothermally at $60\,^\circ$C and samples were taken at the given times and analyzed. The mass balances for A and B were checked as described above and were accepted to be in agreement with the postulated model. (Keeping a batch reaction of an exothermal reaction isothermal, in particular at the start, is difficult and in reality it is favorable to record the temperature together with reaction data at different temperatures, which can then be used to get activation parameters with the same fit.)

Figure 5.3 shows the result of a kinetic fit with the experimental data points and the underlying simulation with the obtained rate constants $k_1 = 2.7 \times 10^{-3}\,\mathrm{l\,mol^{-1}\,s^{-1}}$ and $k_2 = 5 \times 10^{-4}\,\mathrm{l\,mol^{-1}\,s^{-1}}$.

Although in this single experiment the data seem to be in excellent agreement with the postulated model, it is good practice to confirm this agreement with at least one second experiment with different starting concentrations.

In this batch mode, the undesired impurity forms to about 16% and it becomes clear that unreacted B is left over. So 1 mol of A only generates 0.68 mol of the product.

Figure 5.3 Fitting of rate data to experimental data points.

Feed of A at 60 °C

Figure 5.4 The effect of changing to feed batch mode.

As suggested earlier, with these rate constants we can now simulate different scenarios of feed-batch reactions. Figure 5.4 shows the result of the dosing of 1 mol of A in a 200 ml solution within 30 min to a solution of 1.5 mol of B in 1 l solution.

This time, with the same rate constants, 1 mol of A generates 0.83 mol of the product and the amount of I1 has fallen to 0.086 mol. We have to pay for this increase in conversion and yield with an increase in the leftover of B from 0.16 to 0.59 mol.

It is now easy to play with different excesses and feed times to get a feeling of the sensitivity of these parameters. It also allows the development team to take a whole process approach depending on the exact nature and costs of the various materials to identify the optimal overall process.

Until now, all our considerations related to an isothermal process at a single temperature. The next step is the implementation of the temperature dependence of the two rate constants into the model. We have to now introduce the activation energy E_a into the system. High activation energy means a strong increase in the rate constant with temperature. Reactions with low temperature-sensitive rate constants have low activation energies.

5.4.3
Introduction of Temperature Dependence

A model can capture information of any complexity in a very convenient form. This is an important feature of models, as the interactions of the elements in the translation will increase the complexity, and a system that can cope with this complexity is needed.

Our model will expand as follows:

$$A + B \xrightarrow{k_1} P \qquad r_1 = k_1[A][B]\ k_1(T_{\text{ref}}), Ea_1$$

$$A + P \xrightarrow{k_2} I1 \qquad r_2 = k_2[A][P]\ k_2(T_{\text{ref}}), Ea_2$$

We can still continue in the visualization mode when we have a feel of the activation energies. In case $Ea_2 > Ea_1$, the model indicates that a larger degree of I1 formation can be expected at a higher temperature, and to get high yields of product the reaction temperature should be low. As all the reaction rates drop with temperature, the reaction times will increase and there are also natural limits to the lowest feasible temperature (solubility, process safety aspects, etc.). Further on, the longer the reaction times, the higher will be the cost for large-scale manufacture. How much excess of B can be tolerated will depend on the relative cost of B and A and how easy it is to separate the product from the reaction mixture.

Thus, from the model-enhanced process understanding, there are optimum conditions that need to be found. This cannot be done with the visualization model; here we need the parameters and we already know that this information can be readily obtained by kinetic fitting.

Again, there are some options delivered by enhanced process understanding. A method that is currently becoming more and more standard is to run data-rich experiments, that is, to try to get as much information from an experiment, and then by the method of kinetic fitting adjust the rate parameters in a model to reproduce all the experimental data.

The information we can get from a reaction can be discontinuous at discrete time points, such as off-line analytical information obtained using sampling and chromatographic analysis or using continuous on-line methods such as IR, or calorimetric information. In the case of reactions involving gases, the uptake of the gas as a function of time, for example, with hydrogenation reactions, or the generation of gas as a function of time, for example, CO_2 generation in decarboxylation reactions, are monitored. In special cases, other analytical data such as pH, other electrode potentials, or on-line measurements of conductivity can be performed.

For the reaction in our example, we could measure concentrations of all species A, B, P, and I1 at different temperatures and get the desired set of rate constants and activation energies. As the temperature is measured, we no longer need to have an eye on good isothermal data, as modeling can handle any temperature profiles. It is possible to get Arrhenius data even from a single temperature ramp experiment [2].

Evaluation of a set of concentration data at different temperatures (one experiment for each temperature) revealed the following information:

$$A + B \xrightarrow{k_1} P \qquad r_1 = k_1[A][B]\ k_1 = 2.7 \times 10^{-3}\ \text{l mol}^{-1}s^{-1}\ @\ 60\,^\circ\text{C},$$
$$Ea_1 = 60\ \text{kJ mol}^{-1}$$

$$A + P \xrightarrow{k_2} I1 \qquad r_2 = k_2[A][P]\ k_2 = 5.0 \times 10^{-4}\ \text{l mol}^{-1}s^{-1}\ @\ 60\,^\circ\text{C},$$
$$Ea_2 = 90\ \text{kJ mol}^{-1}$$

Another option is to measure parts of the reaction system independently and use the results in the more complex model. For example, the reaction of the product with starting material A can be measured separately from the desired reaction. Rather than tracking the small amounts of I1 observed in the reaction between A and B, an alternative would be to start with P and A and follow the formation of I1 with no interference with the reaction of B and A. This could deliver a better data quality as much higher concentrations of I1 with less analytical error might be obtained.

Regardless of the method used, the result should be the same rate constant and activation energy, which could be directly used in the more complex model.

This modular approach used in models offers the use of literature data in combination with measured ones and, indeed, most of the physical properties information required in more complex models can be found in the literature.

We have so far seen how modeling can enhance process understanding in the case of a chemical reaction system and that the best conditions depending on complex criteria that may include cost can then be obtained by modeling.

An objection to modeling is that all this information might have been obtained by performing small-scale screening experiments with state-of-the-art parallel reaction systems with automatic sampling and on-line or off-line analytics. Indeed, in many systems, it appears to be a quicker and easier method of carrying out a real experiment rather than developing a model first with all its assumptions and uncertainties and then simulating this experiment. However, the limitations of experiments and the advantages of modeling become significant when physical rates are involved and the scale and the equipment have an impact on the results. In all these cases where experiments become too expensive or have a high risk involved, for example, in ''what if'' safety-testing scenarios, the prediction of the results by modeling is the method of choice.

An obvious case is the prediction of temperature profiles and the cooling capacity required for scale-up of exothermic reactions. Again, we demonstrate this with our model reaction, which now has to be expanded by results from calorimetric measurements.

5.4.4
Including Reaction Heats

From heat rate profiles (obtained by a calorimeter), it is easy to get the overall reaction heat and, with knowledge of the rates of the individual steps, it is possible to partition the overall heat to the individual reactions.

$$A + B \xrightarrow{k_1} P \qquad r_1 = k_1[A][B] \; k_1(T_{ref}), Ea_1 \quad \Delta H_1$$

$$A + P \xrightarrow{k_2} I1 \qquad r_2 = k_2[A][P] \; k_2(T_{ref}), Ea_2 \quad \Delta H_2$$

The desired reaction is exothermal by $\Delta H_{r1} = -150\,\text{kJ}\,\text{mol}^{-1}$ and the side reaction to form I1 by $\Delta H_2 = -80\,\text{kJ}\,\text{mol}^{-1}$.

Figure 5.5 Isothermal dosing of A at 60 °C including heat generation.

Now the heat flow can be simulated together with concentration data as a function of conditions, and Figure 5.4 can be updated to Figure 5.5.

As mentioned above, all this information could, in principle, be obtained without modeling, just by experimental work. Indeed, experiments are the basis for the generation of the models and for extracting the kinetic and thermodynamic parameters.

5.4.5
Putting Elements Together: Large-Scale Simulations

Once we have a reasonable satisfactory model, then we can perform simulations where experiments are no longer possible, particularly on a large scale. Thus we can say that the domain of modeling is the prediction of large-scale behavior.

Here we see a good match with the objective of translation of the characteristic elements defined previously. Process modeling is the tool to perform this translation.

The question here is, ''what will change when we scale our exothermic reaction up by a factor of 1000?'' Rather than adding 1 mol of A to a solution of 1.5 mol of B in a 1 l of solution, we will add 1 kmol of A to 1.5 kmol of B in a 1000 l solution.

The first nontrivial translation of the recipe is the temperature. We have to be aware that, in contrast to the impression we have with small-scale reactions, temperature is not automatically controlled, but is the result of an interplay of the heat generation and the cooling capacity.

The heat-generation rate is a function of the kinetics and the reaction heats and has already been implemented in our model.

In a conventional jacketed batch vessel, the cooling capacity is a function of the equipment and the jacket temperature provided. In its simplest form, the rate of heat removed through the reactor wall can be expressed by $Q_{flow} = AU(T_r - T_j)$, where A is the heat exchange area, that is, the wetted part of the reactor and a simple function of the reactor geometry and the filling level. U is called the *heat transfer coefficient*, a complex function of physical properties of reaction mixture, reactor wall, and heat transfer fluid and fluid dynamics. $T_r - T_j$ is the difference between the reaction temperature T_r and the jacket temperature T_j.

With the knowledge of the reactor geometry, physical properties (specific heat, density, viscosity, thermal conductivity) of the materials, and fluid dynamic data available from well-established correlations of agitator geometry and speed, AU can be calculated and the results of such calculations are in fair agreement with experimentally accessed heat transfer data (see Chapter 6).

With the reactor geometry and materials data of one of our 1600 l reactors in our pilot plant, the available heat transfer fluid data, and a combination of measured (specific heat, density) and estimated (viscosity, thermal conductivity) physical properties data of our reaction mixture, the overall heat transfer coefficient with a 90 rpm agitator speed was calculated to be $U = 267\,W\,m^{-2}\,K^{-1}$.

Figure 5.6 shows a prediction of the temperature profile in the 1600 l reactor with a feed rate of 200 l in 1 h. The specific heat of the reaction mixture and of the feed was $1.9\,J\,g^{-1}\,K^{-1}$ and the temperature of the feed was assumed to be 20 °C.

In the first stage of this model, a jacket temperature profile was imposed with a constant jacket temperature of 40 °C throughout the dosing and 60 °C before and thereafter.

Figure 5.6 Scale-up prediction with imposed jacket temperature profile.

From the analysis of the temperature profile, a high level of process understanding can be extracted. The first drop in reaction temperature is caused by the overcompensation of the reaction heat by cooling with cold feed. Then heat flow increases owing to the increase in the reaction rate, with a subsequent increase in the reaction temperature, and at the end of the feed the higher cooling capacity caused by the increased ΔT eventually reduces the temperature. After the sudden stop of the cold feed, the net heat generation leads equally suddenly to another increase in the reaction temperature, which then approaches the jacket temperature with the reaction going toward completion.

Thus, instead of a constant temperature, we see a temperature profile with a maximum of about 70 °C and this has consequences for the performance. Compared with the isothermal small-scale run (Figure 5.4), the yield of the product dropped from 83.0 to 79.7% and the number of moles of the generated side product I1 increased to 101.5 mol rather than the expected 86 mol.

This simple coolant temperature model with an imposed jacket temperature can now be further improved by adding a real temperature control algorithm to the model.

One option is to control the reactor temperature to a set value by varying the jacket temperature. Figure 5.7 shows the result after the implementation of a simple P-controller.

With a lowest jacket temperature of about 30 °C, the reaction temperature can now be limited to 62 °C. It is obvious that the present model allows the simulation of very different reaction conditions on a large scale, testing the effect of temperature, concentrations, and feed time.

Figure 5.7 Including a simple temperature controller.

5.4.6
Thermal Process Safety Simulations

Before moving forward, however, important thermal safety aspects should be implemented in the model. As detailed in Chapter 3 on Process Safety, the thermal stability of the reacted mixture has strong impacts on the design of the process.

The consequences of a loss of cooling capacity at the worst possible situation need to be considered and potential instabilities of the reaction mixture at higher temperatures have to be included.

Let us assume that from the analysis of an adiabatic experiment it becomes evident that our reaction mixture is not stable but exhibits a heavily exothermic decomposition reaction ($\Delta H_r = -420 \, \text{kJ} \, \text{mol}^{-1}$) with a low rate at the desired reaction temperature ($k_3 = 5 \times 10^{-7} 1 \, \text{s}^{-1}$ @ 60 °C), but with a high energy of activation of $Ea_3 = 140 \, \text{kJ} \, \text{mol}^{-1}$.

This reaction rate is so slow that the decomposition of the product is hardly to be noticed in the isothermal lab batches and a simulation reveals that, from the 0.83 mol of product, only 0.0013 mol would have been decomposed.

However, a kinetic simulation shows that this decomposition reaction will have a severe impact on the overall thermal safety of the process in the case of a loss of cooling capacity just at the end of the dosing.

Figure 5.8 shows the simulated temperature profile for the reaction of Figure 5.7, when at the end of dosing there is a complete loss of cooling capacity and the overall heat transfer coefficient is assumed to go to $U = 0 \, \text{W} \, \text{m}^{-2} \, \text{K}^{-1}$.

How can Figure 5.8 be interpreted? It clearly shows that after cutting off the cooling capacity caused by any reason, the reaction mixture will exhibit a thermal runaway with a time to explosion of a little more than 4 h! This time may be

Figure 5.8 Thermal runaway scenario for 60 °C reaction with 1-h feed time.

Figure 5.9 Thermal runaway scenario.

considered to be too short in relation to the severity of the incident and we need to look for ways to extend this time to get more time for countermeasures.

How can we change our reaction conditions?

The fundamental relationships between the reaction conditions and the time to maximum rate (TMR), as this critical time till runaway is described in literature, are best presented in a thermal runaway scenario shown in Figure 5.9.

In the classical approach by Gigax [3], the first (green) part of the thermal excursion is accessible by reaction calorimetry, which delivers the maximum temperature of the synthesis reaction (MTSR) and the red part is accessible among other methods by adiabatic calorimetry, which gives the kinetics of the decomposition reaction.

As kinetic data of the synthesis reaction were not readily available at that time, it was assumed that the desired reaction is usually fast so that the temperature excursion in case of a cooling capacity loss will reach the temperature MTSR in a time $t_1 - t_0$, which can be neglected compared to TMR.

This assumption is not well met in the case of relatively slow desired reactions, so that the time to reach MTSR can no longer be neglected, as it increases the actual TMR by $t_1 - t_0$. Another, more severe consequence of the traditional approach is the kinetic separation of the kinetics to reach MTSR and the TMR as a function of this MTSR. With high exothermal reactions and large energy accumulations or relatively fast decomposition reactions, significant decomposition already occurs on the way to MTSR and this kinetic overlap results in much smaller TMR values. In other words, there is no longer a clear separation between the desired reaction and the decomposition reaction, and the definition of MTSR as a step in the runaway scenario loses its meaning.

Kinetic modeling is a way to overcome these difficulties and the simulation of runaway scenarios will consider these interactions, as the limitations of the MTSR

Figure 5.10 TMR (adiabatic, dashed lines) and reaction time (in normal mode, dotted lines) as a function of reaction temperature.

concept no longer appear. This way, kinetic modeling will greatly enhance the accuracy of thermal hazard related predictions.

The important result of an analysis of the dependence of TMR as a function of the reaction conditions shows that there is a reaction temperature, where this TMR has the largest value and, by simulation, this temperature and the associated TMR (including $t_1 - t_o$) can be obtained.

Figure 5.10 shows this TMR for our reaction with a 1-h feed time as a function of the reaction temperature. In addition, the time for getting the reaction to 99% of consumption of A after the end of dosing is shown.

As expected, there is a temperature where the associated TMR is maximum, and this is in the range of 42 °C with a TMR of about 9 h. This is a significant improvement compared to the 4 h at 60 °C and may reduce the thermal risk to an acceptable level. With a lower reaction temperature, 2–3 more hours of agitation are required at the end of dosing to get the conversion of A to >99%.

We can now explore the reaction conditions to see the effects of all parameters and to actually design the large-scale process.

Process modeling enables this design in one go with the translation of all elements simultaneously and avoids the traditional concept of lab design first and then scale-up.

The beneficial use of process modeling with the prediction of large-scale behavior has been demonstrated so far with a single-phase chemical reaction including process-safety-related statements.

In the next part, the inclusion of multiphase systems with physical rates of mass transfer is shown.

5.5
Physical Rates (the Elements of Mass Transfer)

The modeling of the chemistry in detail to get the desired understanding is straightforward. An important extension to these single-phase homogeneous systems would be the incorporation of multiple phases with a description of the mass transfer between the phases. In a simplified approach, mass transfer between two phases can be described by a two parameter model – one parameter describing the chemical equilibrium of a species in the two phases and the other parameter describing the rate, that is, how quickly this equilibrium is achieved.

Figure 5.11 shows a general presentation of the dynamic process to achieve an equilibrium with two different rates.

This rate process could be the dissolution of a solid up to its solubility limit, the dissolution of a gas in a liquid, or the distribution of a species in one of the phases in a liquid/liquid equilibrium.

It is shown how the rate of this process influences the shape of this curve, but not the thermodynamics. In contrast to chemical rate constants, which are mostly a function of temperature, these physical rate constants (generally called *mass transfer coefficients or mass transfer rate constants $k_L a$*) are strongly dependent on physical energy interactions and equipment and therefore scale.

Some of these mass transfer coefficients can be easily measured on scale, for example, $k_L a$ values for hydrogen uptake can be obtained from recording the pressure drop after switching on the agitator or following the H_2 uptake of a

Figure 5.11 Rate process to achieve a thermodynamic equilibrium without a chemical reaction.

reaction under mass transfer controlled conditions. Others are more difficult to obtain, as many different parameters have an impact, for example, on particle size distribution in the case of a solid/liquid mass transfer, whereas agitation rate beyond a certain limit is usually of no importance.

In most cases, there are engineering correlations available that at least give an estimate of the values of these physical rate constants under different conditions (Chapter 6).

The interesting question in the case of two-phase systems is the interaction of these physical rates with a chemical system as described earlier. Here modeling gives valuable insights and this should be demonstrated by two typical examples, a hydrogenation reaction (Section 5.5.1) and a base-catalyzed reaction with a solid base (Section 5.5.2)

5.5.1
Gas/Liquid Mass Transfer

A prototype reaction with a gas/liquid physical rate interaction with a chemical process is a hydrogenation reaction. Because of the low solubility of hydrogen in organic solvents (typically on the order of a few millimoles per liter at 1 atm pressure), H_2 has to be continuously supplied to the reaction mixture to convert all the starting material to product. This is, in most cases, done by drawing in H_2 from the gas phase, which can be described by $d[H_2]/dt = k_L a([H_2]_{sat} - [H_2])$. As the H_2 concentration is likely to appear in the rate law of the chemical reaction as well, for example, rate $= k [A] \cdot [H_2] [Cat]$, the overall shape of the H_2 uptake curve will depend on the relative rates of these different processes. In the case of a fast chemical reaction, H_2 enters the liquid phase and a low steady-state concentration of H_2 is established, so that the H_2 uptake curve is expected to be a straight line with a slope $d[H_2]/dt = k_L a \times [H_2]_{sat}$.

In the case of a slow chemical reaction or a fast H_2 supply, the H_2 concentration will be close to the saturation limit at any time and the H_2 uptake rate will therefore follow the intrinsic chemical rate, which, if the substrate concentration is rate controlling, will give a curved shape due to the steady decrease of substrate A.

Figure 5.12 shows this behavior with the modeling of a simple hydrogenation, the conversion of cyclohexene to cyclohexane in ethanol with a Pd/C catalyst.

Modeling is a powerful tool to visualize the interactions of mass transfer rates and chemical rates, and, as gas solubilities are proportional to pressure (Henry's law), the increase in the mass transfer rate and in the chemical reaction rates due to the increase of the dissolved H_2 concentration the effect of pressure can easily be modeled as well.

Whereas in this simple example with cyclohexene, the effect of different mass transfer rates may only change the overall reaction time (which can, however, be significant on scale), more interesting are systems where the selectivity of a reaction may become a function of mass transfer and pressure.

A typical representative of these types of reactions is the reduction of nitriles to primary amines. These reactions, in general, suffer from a side reaction, where

Figure 5.12 Measured H_2 uptake (dashed lines) versus modeled uptake with reduced catalyst to simulate chemical rate becoming slower than mass transfer rate (dotted lines), and reduced mass transfer by factor 10 (black line) with chemical rate as fast as that in blue line.

the primary amine product adds to the imine intermediate, which, after loss of ammonia and further hydrogenation, results in the formation of a secondary amine side product.

Literature that investigates the overall kinetics of this process [4], including the influence of catalyst absorption processes for the substrate and the effect of added ammonia, is available.

In order to demonstrate the effect of mass transfer and H_2 pressure, a simpler model, which fixes the concentration of substrate and catalyst and does not include the effects of ammonia, can be used.

The minimum set of equations to get the characteristics captured are as follows:

$$R{-}CN + H_2 \xrightarrow{\text{Cat},k_1} R{-}CH{=}NH$$

$$R{-}CH{=}NH + H_2 \xrightarrow{\text{Cat},k_2} R{-}CH_2NH_2$$

$$R{-}CH{=}NH + R{-}CH_2NH_2 \xrightarrow{k_3} R{-}CH{=}NH{-}CH_2R + NH_3$$

$$R{-}CH{=}NH{-}CH_2R + H_2 \xrightarrow{\text{Cat},k_4} R{-}CH_2NHCH_2R$$

There are three hydrogenation reactions and one non-hydrogenation reaction (the addition of the product to the imine), which should be independent of H_2 concentration, H_2 pressure, and catalyst. Depending on the relative rates of this

Figure 5.13 Kinetic fitting of experimental data to simplified model for nitrile reduction.

reaction compared to the hydrogenation reactions, more or less secondary amines should be formed.

Figure 5.13 shows a nitrile reduction under kinetic control (the reaction was performed in a loop reactor with 24.6 kg substrate and 4.86 kg catalyst at 50 psi pressure and 35 °C). With these conditions, 20% of secondary amine was formed.

The rate constants can be obtained from a kinetic fitting of the model to the experimental data and within the assumptions made this model should be good enough to show the influence of a change in mass transfer and pressure on the yield of the desired product.

The model may not be very useful to optimize the concentration of substrate versus catalyst, as this would require more detailed experimentation to evaluate the chemical rate laws, and the model would fail completely to give any idea of the impact of ammonia.

However, to answer the question, ''how would this reaction with the same concentration of substrate and catalyst behave under different hydrogenation conditions?'' for example, by transferring to another site with a conventional hydrogenator with a lower mass transfer or to see the effect of pressure, this simple model may be appropriate.

Thus, a set of reactions was carried out with a variation of $k_L a$ and H_2 pressure. The reaction time was limited to a maximum of 10 h, so that for a slower reaction the yield after 10 h could be evaluated.

Figure 5.14 shows the results of such a design of experiment (DoE)-like approach.

Figure 5.14 clearly shows, within the accuracy of the model, that there is a combination of $k_L a$ and H_2 pressure, which results in consistent yields of >90%. These relationships, which can be achieved by process modeling, are becoming

Figure 5.14 Effect of pressure and $k_L a$ on yield of desired product.

increasingly important under the topic of Quality by Design and Design Space approaches and are more detailed in the Chapter 1 by Vince McCurdy.

In the previous model, a simple rate law for the hydrogenation step was assumed: $d[P]/dt = k\,[A]\,[H_2]\,[Cat]$, which, in the case of a solid catalyst, ignored all potential absorption and desorption processes. Thus, this model described the rate of the hydrogenation in the case of a kinetic control as being first order in the substrate, catalyst, and in hydrogen, or, as the solubility of H_2 is proportionate to pressure, as a linear function of the H_2 pressure.

Investigations of the kinetics of hydrogenation reactions, however, very often show a different picture.

As pointed out earlier, the advantage of modeling is that complexity can easily be added. In the following sequence, the consequences of substrate absorption on the catalyst will be added to the hydrogenation model.

On the basis of the work by Langmuir, the interaction of the substrate may be described by an absorption isotherm for substrate A on the catalyst surface with an equilibrium constant K_A. This absorbed substrate can then react with H_2 to form a product and regenerate the catalyst.

This rate formulation was introduced by Hinshelwood, so that the whole approach is named *Langmuir–Hinshelwood kinetics*.

$$A + Cat \overset{k_1}{\rightleftharpoons} CatA \qquad K_A$$
$$CatA + H_2 \xrightarrow{\ k_2\ } P + Cat$$

In general, the assumption made is that the absorption equilibrium rate constants are faster than the hydrogenation reaction, so that at any time the concentration of the absorbed species has achieved its equilibrium state, that is, a simple mass balance holds:

$$[Cat]_0 = [Cat] + [CatA] = [Cat] + K_A [Cat] [A] = [Cat] (1 + K_A [A])$$

This approach is often used as a first choice for fitting experimental rate data to a model.

If we assume that not only the substrate but also the product and hydrogen are competing for the free sites on the catalyst surface, Langmuir's absorption equilibriums can be extended:

$$A + Cat \rightleftharpoons CatA \qquad K_A$$
$$H_2 + Cat \rightleftharpoons CatH_2 \qquad K_{H2}$$
$$P + Cat \rightleftharpoons CatP \qquad K_P$$

with the extension of the catalyst mass balance:

$$[Cat]_0 = [Cat] (1 + K_A [A] + K_{H2} [H_2] + K_P [P])$$

Interestingly, there are also different choices for the Hinshelwood rate assumptions, and in many cases approaches with a dual-site mechanism are more successful. Here, the rate equation is based on the reaction of two absorbed species: $CatA + CatH_2 = CatP + Cat$.

These approaches can explain situations where the overall rate is not dependent on H_2 pressure or slows down with the product formation, as the product is competing successfully with the occupancy of free sites on the catalyst.

The following experimental data (C. Stoneley and W. Hoffmann, Pfizer Inc., Sandwich, UK, unpublished results) for a debenzylation reaction with a large substrate were fitted to a Langmuir–Hinshelwood model with a dual-site mechanism to explain the observed H_2 pressure and catalyst dependence (Figure 5.15).

The absorption equilibrium data suggest that a much larger part of the catalyst surface is covered with H_2 than that covered by the substrate. This explains the high sensitivity of the rate due to the catalyst loading and the very moderate sensitivity of the applied H_2 pressure. With the same catalyst loading (20%), it appears that the 100 psi pressure gives a slightly lower rate than the 50 psi run, which can be explained by a lower concentration of the absorbed species A in the case of a higher concentration of absorbed H_2.

5.5.2
Solid/Liquid Mass Transfer

Very often, one of the reagents is not completely soluble under the reaction conditions and thus forms a solid phase. Similar to the gas/liquid equilibrium, the resulting solid/liquid equilibrium can be described by a mass transfer rate and

Figure 5.15 Langmuir–Hinshelwood kinetics for a debenzylation reaction.

a thermodynamic equilibrium constant, which, in this case, is the solubility – in general, a strong function of temperature.

A typical reaction is the deprotonation of a soluble phenol with an inorganic base, for example, K_2CO_3, which has a limited solubility in the organic solvent.

The resulting phenolate can then undergo the desired reaction (for example, a Williamson ether synthesis with a halide).

Depending on the reaction rate, the mass transfer k_La, the solubility, and the concentration profile of the phenolate in solution can be understood by the modeling approach. Again, similar to the dissolution of H_2, the ratio of the physical rate of dissolution and the chemical reaction rate is important.

The following example shows a reaction of 1.2 mol of a phenol and 1 mol of a halide in 1 l reaction volume with added 1.1 mol of K_2CO_3 as a finely powdered material. The solubility of the base is only 0.05 mol l^{-1}. The product is profiled and then the reaction is repeated with a different K_2CO_3 quality, which was to be used in the pilot plant (Figure 5.16).

These two lab experiments show that the dissolution rate, which is influenced by the particle size and morphology, has an impact on the overall rate. Although pilot plant material was used in the second experiment, these experiments do not show the expected kinetics on scale-up, as the solid/liquid mass transfer depends not only on the particle size but also on fluid dynamics and energy dissipation. To visualize the importance of k_La for scale-up predictions, the following simulation shows the calculated concentration profiles of the dissolved carbonate as a function of k_La (Figure 5.17).

Figure 5.16 Product profile with two K_2CO_3 qualities: solid line: pilot plant material, dotted line: finely powdered material.

Figure 5.17 Concentration of dissolved carbonate as a function of solid/liquid mass transfer.

It is obvious that potential side reactions that depend on the carbonate reaction will run with different rates (and give different yields) when the reaction is transferred to another scale [5]. When side reactions play a role, the use of Cs_2CO_3 instead of K_2CO_3 very often leads to improved yields in the lab principally due to the much better solubility of Cs_2CO_3 (about $10\times$ in organic solvents). However, because of the high density of Cs_2CO_3, k_La values may be lower than those for K_2CO_3 on scale-up, and the minimum agitation speed to get the solid lifted from the bottom is definitely higher. Thus, together with the higher cost, the use of Cs_2CO_3 on scale needs to be carefully analyzed, and in many cases may not be that superior as expected from lab experiments. Here again, modeling can be the method of choice to evaluate the alternatives. In general, once the solids are suspended, agitation rate does not have a big impact on k_La. The more important factor is the particle size.

For a realistic prediction of solid/liquid mass transfer coefficients, empirical correlations have been developed and can be accessed by software packages such as VisiMix or DynoChem.

5.6
Summary and Outlook

The previous chapters have given a brief introduction on the application of dynamic modeling in various aspects of process development. Starting with the capture of simple kinetic rate laws in a model, this starting point model was expanded by additional features such as temperature dependence, heat effects, and site reactions. It was shown as to how this model could be used to visualize and predict the interactions with scale- and equipment-dependent parameters such as heat and mass transfer rates. Kinetic data could be extracted from carefully performed lab experiments. By making preferential use of literature available, data on physical properties and equipment-dependent physical rate information, either measured or estimated by engineering correlations, performance predictions on a large scale could then be modeled by a combination of these chemical and physical rate data.

Although the examples given in this chapter focused on the dynamic aspects of a process, it should be pointed out that systems with a more thermodynamic domain, for example, the modeling of gas/liquid equilibriums in the case of distillations, can be modeled as well. Many thermodynamic parameters to describe binary and ternary nonideal gas/liquid systems are available in literature. These equilibrium processes can be combined with energy balances to account for temperature profiles during distillations, and to simulate, for example, the design of a solvent swap from a multiple distil off of solvent A and replace by solvent B sequence toward a continuous feed of solvent B to distilling solvent A.

Another outlook on the application of modeling is the performance prediction of a kinetic system in various continuous reactor setups, which has recently gained considerable interest in pharmaceutical industries to fight costs. On the basis of a kinetic model developed for the batch/semibatch standard process development

plug-flow or continuous stirred tank reactor (CSTR), reactor operation conditions can be developed without going back to the lab. This was successfully demonstrated by transferring a semibatch process to a 10-kg scale pilot plant plug flow reactor without any lab reaction flow experiment (F. Susanne, W. Hoffmann, and T. Moran, Pfizer Inc., Sandwich, UK, unpublished results).

References

1. John, R. (2003) Bourne: mixing and selectivity of chemical reactions. *Org. Proc. Res. Dev.*, **7**, 471.
2. Hoffmann, W., Kang, Y., Mitchell, J.C., and Snowden, M. (2005) Kinetic data by nonisothermal reaction calorimetry: a model-assisted calorimetric evaluation. *Org. Proc. Res. Dev.*, **11**, 25.
3. Gygax, R. (1988) Chemical reaction engineering for safety. *Chem. Eng. Sci.*, **43**, 1759.
4. Hoffer, B.W., Schoenmakers, P.H.J., Mooijman, P.R.M., Hamminga, G.M., Berger, R.J., van Langeveld, A.D., and Moulijn, J.A. (2004) Mass transfer and kinetics of the three-phase hydrogenation of a dintrile over a Raney-type nickel catalyst. *Chem. Eng. Sci.*, **59**, 259.
5. Moseley, J.D., Bansal, P., Bowden, S.A., Edward, A., Couch, M., Hubacek, I., and Weingärtner, G. (2006) Trouble with potassium carbonate and centrifuges: Mass Transfer and scale-Up effects in the manufacture of ZD9331 POM quinacetate. *Org. Proc. Res. Dev.*, **10**, 153.

6
Scale-Up of Chemical Reactions

E. Hugh Stitt and Mark J. H. Simmons

6.1
Introduction

The key issues in scaling up chemical reactions have long been recognized. They are conveniently introduced and summarized in a seminal paper by Paul [1] who notes that many reactions require no special design or operational considerations once the reacting system has been established and its requirements determined. For these reactions, a laboratory-scale sensitivity analysis and pilot plant evaluation may be sufficient to demonstrate the feasibility of successful direct scale-up to production. Not all reaction systems are, however, that sympathetic to the scale-up practitioner.

How can successful scale-up be defined? Paul [1] identifies this as ''plant operation that achieves the same conversion, selectivity, and product distribution as defined in the laboratory.'' In the simple cases described above, reactor design may be accomplished through straightforward scale-up based on the volume and use of standard batch reactor configurations.

The scale-up of many reactions is, however, not so straightforward and variations in performance (rate, selectivity) are observed as a function of operating scale. This is a result of a number of factors that may include the following:

- Reduction in surface area/volume ratio that results in limitations on the heat transfer rate, in turn influencing heat-up, cool-down, or temperature maintenance, and gas or vapor dissolution and/or evolution
- Sensitivity to mixing (i.e., circulation time, shear, mass transfer between phases, etc.)
- Time of addition of a reactant and/or removal of a product in semibatch.

This chapter explores how these variables can influence the scale-up of a reaction, and how process understanding can be utilized to minimize the risk or losses resulting from scale-up, with particular focus on batch and semibatch operations in agitated reactors. It is not the intention of this chapter to provide complete approaches to a scale-up solution. For this the reader is referred to more comprehensive texts [2−4]. Rather, the intent of this chapter is to provide a

basic introduction to mixing alongside a pragmatic approach to achieving process understanding by assessing the likelihood of scale-up problems, and specifically to identify key scale-up parameters.

6.2
Case Study – Batch Hydrogenation

It can be stated flippantly that all of mass transfer, heat transfer, and mixing time get worse as the scale of the reactor increases. But how can each of these influence the progress and selectivity of a given reaction? As a means of demonstrating the effects that scale-up parameters can have on a batch reaction, a case study is presented and discussed. The case study is loosely based on a parallel selectivity hydrogenation example. The data presented are derived from computer simulations of the reaction and reactor.

The two reactions are

$$A + H_2 \Longrightarrow B \quad \text{and} \quad A + H_2 \Longrightarrow C$$

and their respective kinetics are represented by

$$R_1 = k_1 C_A C_{H_2}{}^2 \quad \text{and} \quad R_2 = k_2 C_A C_{H_2}$$

where C_A is the substrate concentration and $C_{H2} C_{H2}$ is the concentration of hydrogen in the liquid phase. The reaction rate constants, k_1 and k_2, are each represented by an Arrhenius type equation and have different activation energies ($E_{A2} = 2.5 E_{A1}$). Both reactions are significantly exothermic, but have similar reaction enthalpies. The reactions are carried out in batch, jacketed autoclaves with pressure-controlled hydrogen addition (variously referred to as "*dead end operation*" or "*semibatch operation*").

The global results for a simulated pilot (2 l) scale run are given in Figure 6.1. The reactor design variables assigned for the simulation were within normal engineering expectations for the scale of operation, offering reasonable heat transfer and good mixing (gas dispersion and gas–liquid mass transfer). The profiles show reaction completion after approximately 50 min and an end of run selectivity of 97%. The heat transfer provision is not sufficient to avoid a significant temperature rise during the period of fast reaction early on the batch, although the temperature rise is constrained to less than 4 °C at this scale.

The manufacturing scale results for a 2000 l reactor are given in Figure 6.2 and show a very different profile. Completion of reaction is at approximately 70 min. The temperature plot is very different, not showing a peak but rather a near monotonic increase in temperature through the batch. That is, the heat transfer is never sufficient to adequately remove the reaction heat except in the very final stages. The selectivity is notably poorer, and end of run selectivity is only 90%. This magnitude of difference between pilot and manufacturing scale is not uncommon. Increases in the required reaction time much larger than that shown here are frequent. This, however, is notably a function of the reaction and reactor characteristics.

Figure 6.1 Pilot-scale batch hydrogenation data.

Figure 6.2 Manufacturing-scale batch hydrogenation data.

Figure 6.3 Pilot-scale hydrogenation with reduced heat transfer.

In order to better demonstrate the scale-up effects, the mass and heat transfer effects will be decoupled using simulation at the pilot scale but with, in the first case, the heat transfer set to a large scale, while the mass transfer is maintained at its pilot scale value, and then the converse. The results are shown in Figures 6.3 and 6.4. Both profiles show significant deviation from the ''pilot scale'' base case in Figure 6.1. The loss of heat transfer capability results in a higher temperature, and thus ironically faster reaction, but exacerbating the problem of loss of selectivity. This loss of selectivity is a direct result of the higher temperature and the relative activation energies of the competitive reactions. The reduction in the rate of mass transfer conversely leads to a significant extension of the reaction time as well as a major loss of selectivity. Both of these are a direct result of the reduced liquid phase (thence catalyst surface) hydrogen availability.

In this case, therefore, all of the observed scale-up effects can be attributed to the impact of changing temperature and hydrogen availability on the two competitive reactions. Establishing this clear link would, however, not be possible without a fundamental understanding of not only the scale-up of the stirred reactor but also the reaction chemistry. Process understanding is therefore essential for successful scale-up.

The next sections address the key areas of required process understanding and specifically explore the scale-up of stirred reactors, the influence of the chemistry, and how to establish the critical parameters.

Figure 6.4 Pilot-scale batch hydrogenation with reduced mass transfer.

6.3
Scale-Up of Stirred Tank Reactors (STRs)

Despite the mechanical simplicity of stirred tank reactors (STRs), the motion of the fluid, and hence the degree of mixing and homogeneity within the vessel, is highly complex, which creates nonlinearities in the behavior of the fluid motion with changing scale of operation. This means that molecules within the vessel will experience different conditions depending on their location, which can lead to variable (usually undesired) yields and selectivities. This complexity is enhanced further if additional (dispersed) phases are introduced into the liquid flow. This presents considerable challenges for the development team in terms of

- scale-up of new processes from the laboratory scale (10^{-4} to 10^{-2} m^3) to the process scale (10^{-1} to 10^3 m^3) and
- adaptation of process conditions for alternative reactions.

This section develops and presents simple design rules that enable the practicing process engineer to perform scale-up of STRs on a sound basis, while highlighting various difficulties and problems to be avoided. The fundamentals of turbulent mixing in STRs are described and key dependent variables, which can be used to develop strategies for scale-up, are presented, which are pertinent to the applications below.

- Single liquid-phase mixing
  - blending two miscible liquids;

- blending and chemical reaction;
- blending with heat transfer.
- Solid–liquid mixing
  - resuspension of solids for transport or removal (e.g., storage tank);
  - suspension of a solid reactant, which may or may not dissolve;
  - enhancement of a liquid-phase reaction using a particulate heterogeneous catalyst.
- Gas–liquid (solid)mixing
  - gas–liquid reactions (possibly with suspended particulate catalyst, e.g., hydrogenations and oxidations).

6.3.1
Fundamentals of Flow Regimes, Turbulence, and Turbulent Mixing

The nature of fluid flow was investigated in pipe flow by Osborne Reynolds in 1883 [5]. He introduced a stream of dye into the axis of a pipe flow, and observed the motion of the dye as it passed down the pipe. He observed two distinct regimes, first where the stream of dye remained intact down the length of the pipe and, second, where the dye was dispersed by the motion of random three-dimensional whorling motions (eddies) in the flow. He defined these regimes as follows:

- Laminar flow: the motion of the fluid molecules always follows the fluid streamlines.
- Turbulent flow: dispersion occurs by the action of irregular eddies.

The onset of turbulent flow can be determined via the Reynolds number, Re, the ratio of inertial to viscous forces, defined as

$$Re = \frac{UL}{\nu} = \frac{UL\rho}{\mu} \tag{6.1}$$

where U is characteristic velocity of the flow (m s^{-1}), L is the characteristic length scale of the flow, and $\nu = \mu/\rho$ is the kinematic viscosity (m^2 s^{-1}) (μ is the dynamic viscosity (Pa s) and ρ is the density (kg m^{-3})). The transition from laminar to turbulent flow occurs as inertial effects cause instabilities in the flow to grow. Viscous effects attenuate these instabilities but are insufficient to prevent their growth once the Reynolds number exceeds a critical value, Re_{crit}. Generally, due to the large scale of the vessels ($L \sim 1$–10 m) and close to aqueous viscosities of the solvents used within the fine chemicals industries ($\mu \sim 10^{-3}$ to 10^{-2} Pa s), the turbulent regime is by far the most commonly observed regime, found for $Re > 20\,000$ in a stirred vessel (see later Section 6.3.2).

Checking the value of Reynolds number within a reactor is a vital first step since the flow regime greatly influences the mixing and thus the reactor performance. Owing to the random motion in turbulent flows, mixing is much enhanced compared to laminar flows.

6.3.1.1 Mixing Mechanisms in Laminar Flows

Within a fluid, convective (sometimes called *advective*) mixing, where molecules are dispersed by the action of the moving fluid, is caused by the motion of molecules of fluid relative to each other. This movement is above and beyond the Brownian motion, which drives diffusion and is caused by velocity gradients within the flow. Determination of these velocity gradients and how they influence mixing performance is dependent on the flow regime and whether the flow is steady or unsteady with time.

To illustrate this, consider a simple example of steady (time-invariant) laminar flow between two infinitely long plates shown in Figure 6.5, where the bottom plate is fixed and the top plate is moving at a constant velocity, U. The fluid moves only in the x direction, with its maximum velocity at the top plate, U, and a velocity of zero at the bottom plate. This is due to the *no-slip condition* caused by friction of the fluid molecules at the wall surface; thus the fluid closest to the wall travels at the velocity of the wall.

This sets up a constant linear velocity gradient in the gap of $\mathrm{d}u/\mathrm{d}y = U/h$, *orthogonal* to the flow direction, x. This is an example of *shear flow*. Shear flow also occurs for laminar flow in a pipe, although the velocity gradient in this case is not constant because of the parabolic velocity profile developed.

Returning to the two plates, consider now what happens if a rectangular fluid element is placed between them. The top of the element will be subject to a greater velocity than the bottom of the element; thus the top moves further than the bottom of the element as it travels down the pipe. As the mass of the element must be conserved, the element is stretched; it becomes thinner and its surface area increases as shown in Figure 6.5. This stretching thus drives mixing in the x direction and the magnitude of this effect is simple to calculate because the velocity gradient is constant. Using Pythagoras theorem, the amount of stretching from an initial element length of l_{o} to a length l after a time t can be written as $\frac{l(t)}{l_{\mathrm{o}}} = \left(1 + \frac{U}{h}t\right)$. This implies that the amount of stretching is linear with time, that is,

$$\frac{l(t)}{l_{\mathrm{o}}} \propto t \tag{6.2}$$

An alternative mechanism to shear flow occurs if a flow passing through a contraction within a pipe is considered. As the mass flow must be conserved, the fluid accelerates as the cross-sectional area of the pipe decreases. There now

Figure 6.5 Stretching of fluid elements in a laminar shear flow.

exists an axial gradient of velocity, which is *parallel* to the flow direction. This is an example of *elongational* flow, which is also a linear mechanism of the form given by Equation (6.2)

Since for steady laminar flows the velocities are constant in time, any velocity gradients present are also independent of time and the analysis is relatively straightforward, as in the above example. However, matters are complicated if the flows are time dependent, as is the case within a stirred vessel, because of the periodic passage of the impeller blades. This can lead to *chaotic instabilities*, which greatly enhance mixing according to an exponential behavior,

$$\frac{l\,(t)}{l_{\mathrm{o}}} \propto \mathrm{e}^{\lambda_{\mathrm{L}} t} \tag{6.3}$$

where λ_{L} is a characteristic of the chaotic flow component called the *Lyapunov exponent*. Prediction of mixing in chaotic flows is nontrivial and requires solution of the mass and momentum equations governing the flow using advanced mathematical methods, which are beyond the scope of this chapter. Chaotic mixing effects are exploited in the blending of highly viscous materials in laminar flow using static mixers such as the Kenics or SMX designs [6].

6.3.1.2 Mixing Mechanisms in Turbulent Flows

Most reactions of interest to the fine chemicals and pharmaceutical industries occur in a turbulent environment within STRs; the distribution of the rate of mixing (or homogeneity) within the vessel is critical to achievement of the required rate of reaction or selectivity. As for laminar flows above, to understand the mixing behavior within an STR, it is necessary to know the distribution of velocities within the flow so that the velocity gradients can be calculated. This process is considerably more complicated for turbulent flows as they are subject to local time-dependent fluctuations in the flow velocity, which generate eddies in the flow field. An eddy can be thought of as a whirling motion in the flow, which has a characteristic length scale, λ, associated with it and prediction of the motion of these structures is greatly complicated by their random nature.

To understand how turbulence causes mixing, it is necessary to determine a phenomenological picture of how these eddies are generated, how they propagate, and what overall bulk parameters in the flow can be used to predict their properties. Within an STR, the generation of eddies can be described by considering the scales of fluid motion within the vessel. As the impeller rotates, the rotational energy of the blades is transferred to kinetic energy within the fluid. The energy is supplied at a rate equal to the power P from the impeller motor (minus any mechanical losses). For a vessel with fluid of density ρ and volume V the average power imparted to the fluid per unit mass, $\overline{\varepsilon}_{T}$, can be written as

$$\overline{\varepsilon}_{T} = \frac{P}{\rho\,V} \tag{6.4}$$

This important quantity is termed the *average specific energy dissipation rate* for the vessel.

Once the flow reaches a steady state after switching on the motor, the energy inputted at a rate ε_T causes the formation of large-scale eddy structures on the scale of the order of the impeller blade width. These break down into smaller scale structures and transfer all of their energy to these without loss. Since when a large eddy breaks down it passes on its energy to several smaller eddies, the energy content of these eddies decreases with size. The energy is finally dissipated as heat at the smallest scales due to the viscosity of the fluid (called *viscous dissipation*). This concept is termed an *energy cascade* and was first proposed by Lewis Fry Richardson (1922) and is illustrated in Figure 6.6.

The smallest scales in the flow are identified as the Kolmogorov scales. Kolmogorov reasoned that viscous dissipation would occur once inertial and viscous forces were of the same order, that is, at a Reynolds number equal to unity. He defined a length scale, λ_K, based on this assumption. Hence,

$$Re_K = \frac{\lambda_K u_K}{\upsilon_K} = 1 \tag{6.5}$$

He also reasoned that, at these small scales, the motion is independent of the macroscale geometry and only depends upon the energy input, ε_T, and the fluid properties represented by the kinematic viscosity of the fluid ($\nu = \mu/\rho$). On the basis of scaling arguments, the length scale, λ_K, and also a velocity scale, u_K, and a timescale, τ_K, can be defined as

$$\lambda_K = \left(\frac{\upsilon^3}{\varepsilon_T} \right)^{1/4} \qquad u_K = (\varepsilon_T \upsilon)^{1/4} \qquad \tau_K = \left(\frac{\upsilon}{\varepsilon_T} \right)^{1/2} \tag{6.6}$$

The Kolmogorov length scale is a representation of the size of the smallest eddies that can exist in the flow. Thus, turbulent mixing is only able to provide mixing down to this length scale. Below this scale, mixing occurs by molecular diffusion only.

Figure 6.6 Cascade of turbulent energy in an STR.

To give an idea of the value of λ_K, consider the agitation of water ($\nu = 10^{-6}$ m^2 s^{-1}) at a typical power input of $\varepsilon_T = 1$ W kg^{-1}. Using Equation (6.6), λ_K can be estimated as 32×10^{-6} m, or 32 μm. Clearly, this is several orders of magnitude larger than the molecular scale, which is of order 10^{-10} to 10^{-9} m.

Although a crude estimate for the overall Kolmogorov scale within an STR can be made by using the average specific energy dissipation rate (Equation 6.4), the reality is that the energy imparted by the impeller is not evenly dissipated over the whole vessel. Local (spatially evaluated) values of ε_T can change drastically from the average value $\overline{\varepsilon}_T$. Generally $\varepsilon_T > \overline{\varepsilon}_T$ near the impeller ($\varepsilon_T/\overline{\varepsilon}_T \approx 50$) and $\varepsilon_T \ll \overline{\varepsilon}_T$ away from impeller ($\varepsilon_T/\overline{\varepsilon}_T \approx 0.1$). Local values of λ_K can therefore vary by up to an order of magnitude from the average value in the most extreme cases. In addition, as the scale of the tank increases, spatial heterogeneity in energy dissipation also increases. In particular, for fed-batch or continuous systems, better performance is generally observed if addition of reagents is made into regions of the vessel experiencing high local values of ε_T [7, 8]. Thus, accurate estimates for mixing length scales and motions causing mixing require the turbulent flow field within the vessel to be determined, from which local values of turbulent energy and thus local specific energy dissipation rate can be evaluated.

6.3.1.3 Estimating Energy Dissipation and Mixing Length Scales from Turbulent Flow Fields

The typical behavior of the fluid velocity, U, taken at an arbitrary point in a turbulent flow of constant flow rate is shown in Figure 6.7. The velocity signal illustrates that the fluid velocity is subject to random multiple frequency disturbances due to the presence of the eddies.

Provided that the flow is at a constant flow rate and that the total sample time, $t = T$, is much larger than the smallest characteristic frequency of the fluctuations, τ, that is, $T \gg \tau$, a time-average velocity at this point, $\overline{u}$ (see Figure 6.7), can be calculated. The time-average mean is simply calculated as

$$\overline{u} = \frac{\sum_{i=1}^{N} U_i}{N} \tag{6.7}$$

Figure 6.7 Typical velocity signal measured at a point in a turbulent flow with constant flowrate.

where N is the number of samples. Considering now an arbitrary instantaneous point value of U, this can be split into the time-averaged part above (which is independent of the fluctuations) and a fluctuating part (Reynolds decomposition).

$$U = \overline{u} + u'$$ (6.8)

The fluctuations can be averaged over the whole sample using the root mean square (rms), $\tilde{u}$. The overbar in Equation (6.9) below indicates time averaged over all samples.

$$\tilde{u}^2 = \overline{u'^2} = \overline{(u - \overline{u})^2}$$ (6.9)

In order to be able to predict the rate of dissipation of energy, we can begin by calculating the kinetic energy directly from the velocity field. We can define the kinetic energy per unit mass (KE, J kg^{-1}), which can be defined in Cartesian (x, y, z) coordinates as

$$KE = \frac{1}{2}(u^2 + v^2 + w^2)$$ (6.10)

where u, v, and w are the components of velocity in the x, y, and z directions, respectively. Performing a similar analysis using Equation (6.8) above, the kinetic energy can be separated into two parts, one due to the mean flow and the other due to the fluctuations. Concentrating on the latter, we can define the turbulent kinetic energy or TKE, given the symbol k. This is representative of the turbulent energy within the flow that eventually dissipates as heat once it cascades to the Komogorov length scale (Figure 6.6a).

$$k = \frac{1}{2}\left(\tilde{u}^2 + \tilde{v}^2 + \tilde{w}^2\right)$$ (6.11)

It only remains now to estimate values of ε_T from k to allow λ_K to be determined. As the energy is injected at the largest scales (Figure 6.6a) and dissipated at the smallest (of order λ_K), obtaining ε_T accurately requires resolution of the velocity field to below λ_K. This is beyond the capabilities of most instrumentation, and several different methodologies of varying complexities have been proposed to overcome this limitation [9, 10]. The simplest and widely adopted approach (although not perhaps the most accurate in absolute terms [9]) is by Wu and Patterson [11], who proposed an equation to evaluate the local energy dissipation from the turbulent kinetic energy,

$$\varepsilon_T = A\frac{k^{3/2}}{\Lambda}$$ (6.12)

where A is a constant of proportionality (taken generally as being equal to 0.85 but other values have been used) and Λ is a macro *integral length scale* (ILS), which is representative of the largest structures in the flow. Various constant values of ILS have been adopted throughout the flow field, from $\Lambda = D/6$ to $D/12.5$ [9]. Also, the assumption that the value of Λ is spatially constant is clearly not the case in reality because in the bulk region of the flow the typical dimension of the turbulence is greater than that in the proximity of the impeller. Various methods of estimation of Λ from experimental data are given in Gabriele *et al*. [10]. In the following

section on stirred vessel design and scale-up, the fundamentals described above are developed into engineering design rules that can be used by the practicing chemical or process engineer.

6.3.2
Stirred Vessel Design and Scale-Up

A generic schematic of a stirred vessel, with geometric and important parameters defined, is shown in Figure 6.8. Generally, four longitudinal baffles of width T/10, mounted at 90° round the circumference of the vessel, are installed to prevent gross rotation of the fluid in the azimuthal direction (and hence vortex formation) along with promoting three-dimensional circulation and mixing. Typical ratios for the dimensions defined in Figure 6.8 are

- impeller diameter/tank diameter: $D/T \sim 0.3$–0.5;
- fill height/tank diameter: $H/T \sim 1$ for single impeller systems;
- baffle width/tank diameter: ~ 0.1; and
- impeller clearance/tank diameter $C/T \sim 0.2$–0.25.

This type of reactor is used extensively for fine chemicals and petrochemicals production. For the pharmaceutical industry, where cleaning efficiency and contamination between batches may be an issue, the vessels may be glass lined and without baffles (allegedly to improve cleaning). Common alternative configurations for such vessels include use of single baffles, either fixed to the wall or removable ("beaver-tail"), and conical vessel bases for supposed ease of discharge [12].

The flow can be laminar or turbulent, calculated on the basis of Reynolds number (Equation 6.13), which is a modification of Equation 6.1 where the characteristic velocity is taken as ND and length scale as the impeller diameter D. Values of $Re < 10$ indicate laminar flow. Values of Re between 10 and 20 000 indicate

- Power, P (W)
- Rotation rate, N (rev s^{-1})
- Volume, V (m^3)
- Fluid dynamic viscosity, μ (Pa. s)
- Fluid density, ρ (kg m^{-3})
- Impeller diameter, D, (m)
- Tank diameter, T, (m)
- Impeller width, W, (m)
- Fill height, H (m)
- Impeller clearance from bottom, C (m)

Figure 6.8 Characteristic dimensions of a cylindrical stirred tank.

transitional flow with fully developed turbulence assumed for $Re > 20\,000$.

$$Re = \frac{\rho N D^2}{\mu} \tag{6.13}$$

6.3.2.1 Impeller Flow Patterns

Depending on the type of impeller used, different impeller flow patterns may be observed as shown in Figure 6.9.

- Radial flow. Four flow loops are formed – two below and two above the impeller plane. This is generally recommended for gas dispersion as well as single-phase operation. This pattern is formed using radial flow impellers, which comprise vertical straight or curved blades. The blades are attached to a disk that prevents pumping of fluid through the impeller, for example, Rushton disk turbine (RDT).
- Axial flow. Two flow loops are formed with the flow pumping through the impeller plane near the shaft. Axial flow impellers use blades angled to the vertical. No disk is present, for example, pitched blade turbine (PBT). Other axial devices include marine propellers and hydrofoils. Retreat curve impellers (RCIs) (used in glass-lined vessels) produce a flow pattern that is dominated by flow in the tangential direction unless baffles are employed [12]; the pattern in the axial plane generally conforms to a two-loop structure.

    The flow can be

- up-pumping – increasingly popular for gas–liquid duties and may also be applicable for solid–liquid systems and
- down-pumping – used for suspension of solids – discharge near shaft is assumed to help dispersion of solids from the tank bottom.

6.3.2.2 Power Input and Specific Energy Dissipation Rate

The power P drawn by an impeller is a critical quantity in terms of engineering requirements (selection of drive motor size, installation, and running costs) and in understanding the overall flow behavior and mixing within the vessel as outlined in

Figure 6.9 Flow regimes in STRs (a) radial; (b) axial up-pumping; and (c) axial down-pumping.

Section 6.3.1.2. Using dimensional reasoning, it is possible to determine the overall power requirement in terms of the above-defined parameters using dimensional analysis.

$$P = f\left(\rho, \mu, N, g, D/T, H/T, C/T, \text{other geometric ratios}\right)$$

Assuming baffles are fitted so that the effect of gravitational acceleration, g (which causes vortexing), can be ignored and applying *geometric similitude*, where geometric ratios are kept constant as the scale of the vessel is changed (i.e., constant D/T, H/T, etc.), we obtain for geometrically similar systems

$$Po = \frac{P}{\rho N^3 D^5} = f\left(Re\right) \tag{6.14}$$

The relationship between Po and Re, expressed by Equation (6.14), is shown for some typical impellers in Figure 6.10 [13]. The general findings are that

- $Po \propto Re^{-1}$ for laminar flow
- $Po = f\left(Re\right)$ for transitional flow
- $Po = $ constant for turbulent flow.

Constant values of Po for turbulent flows can vary by an order of magnitude depending on impeller configuration; Po is strongly influenced by the number of blades and the blade width, W. The effect of blade number can be expressed as $Po \propto (n/D)^{\beta}$, where $\beta = 0.8$ for 3–6 blades and $\beta = 0.7$ for 6–12 blades. This relationship holds for straight blades and for curved blade impellers with four to eight blades. The effect of blade width can be expressed as $Po \propto (W/D)^{1.45}$ for six-bladed Rushton turbines and $Po \propto (W/D)^{0.65}$ for a four-bladed 45° pitched blade.

Figure 6.10 Power curves for some typical impellers, from Hemrajani and Tatterson [13].

Table 6.1 Turbulent power numbers for some common impeller types.

Impeller	Po (turbulent)
Rushton disk turbine (RDT) six-blade version shown	5[a]
Concave-blade turbine (Scaba SRGT or Chemineer CD6)	4.4[a]
Pitched blade turbine (PBT) $6 \times 45°$ blades W = D/5 (shown left) $4 \times 45°$ blades W = D/5	1.64[a] 1.27[a]
Lightnin A310 three blade hydrofoil Chemineer HE3 (similar to A310, not shown) © P Csisar [14]	0.3[a] 0.2−0.3[a]
Lightnin A315 four-blade hydrofoil (down-pumping) Lightnin A345 up-pumping version of A315 (not shown)	0.75[a] 0.75[a]
Marine propeller: (down- and up-pumping) (1.0 pitch, W/T = 0.1) (1.5 pitch, W/T = 0.1) (2.0 pitch, W/T = 0.1) © P Csisar	0.34[a] 0.62[a] 1.00[a]

(continued overleaf)

Table 6.1 (*Continued*)

Impeller							Po (turbulent)
	Three-blade retreat curve impeller (RCI) glass-coated steel	C/T (tank base)	0.16 flat	0.16 conical	0.31 flat	0.31 conical	
		Four wall baffles	0.99[b]	0.87[b]	0.93[b]	0.89[b]	
		One wall baffle	0.56[b]	0.51[b]	0.56[b]	0.43[b]	
		Beaver tail	0.46[b]	0.41[b]	0.42[b]	0.33[b]	

[a] Hemrajani and Tatterson [13].
[b] Rielly *et al.* [12].

Po is in general only weakly affected by D/T (0.33–0.5) and C/T within the ranges used in STR. Commonly used values for the most common designs of impeller are shown in Table 6.1. Included in Table 6.1 are data for RCIs often employed in glass-lined vessels for pharmaceutical production.

In general, stirred vessels should not be operated at $Re < 1000$ for most duties using these impellers, as the lack of turbulence reduces the level of mixing performance to an unacceptable level. For mixing of highly viscous materials where laminar flow is unavoidable, low-wall-clearance-type designs such as the anchor impeller should be considered.

The power drawn, P, can thus be calculated for the turbulent regime if the value of turbulent *Po* for the impeller used is known and thus $\overline{\varepsilon}_T$ can be calculated from Equation 6.4

6.3.2.3 Mixing Times

Mixing within STRs is nonideal. The mixing time, θ_m, is a measure of the time required to achieve homogeneity within the vessel becuase of convective turbulent mixing. Evaluation of mixing time and the way mixing times change with scale is therefore critical to the understanding of STR performance. For a single impeller vessel filled to a height $H = T$, the following correlations have been presented for mixing time.

$$\theta_m = 5.3 \frac{Po^{-1/3}}{N} \left(\frac{D}{T}\right)^{-2} \qquad \text{Nienow [15]} \tag{6.15}$$

$$\theta_m = 5.9 T^{2/3} \left(\overline{\varepsilon}_T\right)^{-1/3} \left(\frac{D}{T}\right)^{-1/3} \qquad \text{Ruszkowski [16]} \tag{6.16}$$

Consequences: for geometrically similar vessels (i.e., constant D/T)

- $\theta_m N$ constant for all impellers with the same power number (Equation 6.15), regardless of the impeller shape. Hence, keeping mixing time constant on scale-up requires constant N.

- At the same total power input per unit mass $\overline{\varepsilon}_T$ (Equation 6.16),
  - for a given tank size, *T*, all impellers give the same mixing time;
  - $\theta_m \propto T^{2/3}$; hence mixing time increases upon scale-up at constant power per unit mass.

6.4
Stirred Tank Scale-Up

6.4.1
Choice of Criterion for Scale-Up in Turbulent Flow

For a geometrically similar system, the choice of scale-up criterion can have a drastic effect on the mixing performance. The key decision rests on the choice of the scaling parameter to be kept constant. In the exercise below, the implications of two different choices are considered for scale-up of a single impeller vessel filled to a height $H = T$ at the following conditions:

- Constant mixing time: from Equation (6.15), this requires the rotational speed, *N*, to be kept constant. Application of this scaling criterion aims to ensure that the macroscale mixing behavior, as represented by the mixing time, is maintained.
- Constant'' turbulent mixing behavior: Although the distribution of ε_T is known to change with scale, the crude assumption made is that scale-up at constant $\overline{\varepsilon}_T$ should lead to similar overall values of turbulent mixing length scale. Use of this scaling criterion therefore concentrates on maintaining similar mixing performances at the smallest scales in the flow, which is important for complex chemical reactions where selectivity toward the desired product is critical.

6.4.1.1 Constant Mixing Time (Constant *N*)
Consider a small lab scale (1) and a larger process scale (2). Therefore, if *N* is constant it holds that

$$N_2 = N_1$$

For turbulent flow, *Po* is constant. Hence, from the definition of *Po*

$$\frac{P_1}{\rho N_1^3 D_1^5} = \frac{P_2}{\rho N_2^3 D_2^5} \tag{6.17}$$

Therefore,

$$\frac{P_2}{P_1} = \left(\frac{D_2}{D_1}\right)^5$$

Since vessel volume $V \propto D^3$, then

$$\frac{P_2}{P_1} = \left(\frac{V_2}{V_1}\right)^{5/3}$$

As *D* increases as *N* is kept constant, the Reynolds number for the flow (represented by Equation 6.13) increases with the square of the diameter, that is, $Re \propto D^2$, on scale-up.

Table 6.2 Comparison of scale-up strategies.

Constant parameter	Effect on N $N_2 = N_1 \left(\frac{D_1}{D_2}\right)^{\alpha}$	Effect on θ_m $\theta_{m2} = \theta_{m1} \left(\frac{D_2}{D_1}\right)^{\beta}$	Effect on P $\frac{P_2}{P_1} = \left(\frac{V_2}{V_1}\right)^{\gamma}$	Effect on $\bar{\varepsilon}_T$ $\frac{\varepsilon_2}{\varepsilon_1} = \left(\frac{V_2}{V_1}\right)^{\delta}$
	α	β	γ	δ
N	0		5/3	2/3
$\bar{\varepsilon}_T$	2/3		1	0
$U = ND$	1		2/3	−1/3
P	5/3		0	−1

6.4.1.2 "Constant" Turbulent Mixing Behavior (Constant $\bar{\varepsilon}_T$)

$$\varepsilon_T = \frac{P}{\rho V} = \frac{Po\rho N^3 D^5}{\rho \frac{\pi}{4} T^3}$$

As $D \propto T$:

$$\frac{P}{\rho V} \propto \frac{N^3 D^5}{T^3} \propto \frac{N^3 D^5}{D^3} \propto N^3 D^2 = \text{constant}$$

Therefore,

$$N_2 = N_1 \left(\frac{D_1}{D_2}\right)^{2/3}$$

and

$$\frac{P_2}{P_1} = \left(\frac{D_2}{D_1}\right)^3 = \left(\frac{V_2}{V_1}\right)$$

As $N \propto D^{-2/3}$, from Equation (6.13), $Re \propto D^{4/3}$ on scale-up. The scaling rules for different scale-up criteria are summarized in Table 6.2, showing the implications of each criterion on rotational speed, tip speed, power, average power input per unit mass, and mixing time as the volume of the vessel increases. The choice of different scaling rules for different mixing duties is summarized in Section 6.4.4.

6.4.2
Heat Transfer

Heating or cooling duties may be supplied to STRs via a number of methods, including heating jackets on the vessel walls or internal coils. The latter are generally avoided if cleaning and contamination is an issue; this section therefore focuses on jacketed vessels. A diagram of a typical jacketed vessel is shown in Figure 6.11.

Figure 6.11 Jacketed vessel schematic for heat transfer.

Applying Newton's law of cooling, the heat transferred to a jacketed vessel may be written for a batch vessel as

$$Q = \frac{M}{\theta} c_P \left(T_{P2} - T_{P1}\right) = UA\Delta T_{LM} \tag{6.18}$$

where

$$\Delta T_{LM} = \frac{(T_S - T_{P2}) - (T_S - T_{P1})}{\ln\left(\frac{(T_S - T_{P2})}{(T_S - T_{P1})}\right)} \tag{6.19}$$

where Q is the heat input (W), θ is the heating time (s), M and c_P are the mass (kg) and specific heat capacity (kJ kg^{-1} K^{-1}) of the fluid in the vessel, respectively, T_{P1} and T_{P2} are the temperatures of the vessel fluid at the start and end of the heating period, and T_S is the temperature of the service side heating medium (e.g., steam, T_S assumed constant). A is the inner heating surface area of the jacket (m^2), which can be determined from the vessel geometry.

With reference to Figure 6.11, the overall heat transfer coefficient, U, depends on

- the process side thermal boundary layer, represented by a process side heat transfer coefficient, h_P;
- any fouling layer formed on the inside of the vessel, represented by a fouling resistance, f_P;
- the thickness, x_w, and the thermal conductivity, k_w, of the vessel wall;
- the service side fouling layer, resistance, f_S; and
- the process side thermal boundary layer, with heat transfer coefficient, h_S.

The overall heat transfer coefficient can be obtained from

$$\frac{1}{U} = \frac{1}{h_P} + f_P + \frac{x_w}{k_w} + f_S + \frac{1}{h_S} \tag{6.20}$$

Generally, the value of U is governed by the largest term on the right-hand side of Equation 6.20 For steam heating, assuming the vessel is reasonably clean, this is usually h_P (h_S may control vessel cooling). h_P can be predicted using relationships of the following form [17]:

$$Nu = A Re^a Pr^b Vi^c G^d \tag{6.21}$$

where

Nu (the Nusselt number) $= h_P T/k$;

Pr (the Prandtl number) $= c_p \mu/k$;

Vi (the viscosity ratio) $= \mu/\mu_w$ (where μ_w is the fluid viscosity at the wall temperature).

G is a geometric function or ratio, which may be taken as $\left(\frac{T}{H}\right)^{0.15} \left(A\frac{W}{D}\right)^{0.2}$, where $A = 5$ for a six-blade RDT, 5.88 for a four-blade PBT, and equal to D/W for an RCI.

a, b, and c usually take values of $a = 0.67$, $b = 0.33$, and $c = 0.14$, but can vary according to the geometry and duty to be performed. A detailed summary of heat transfer calculations in stirred vessels may be found in Ref. [18].

Hence, on scale-up, the relationship in terms of ε_T and T is

$$h_P \propto \varepsilon_T^{2/9} T^{-1/9} \left(\frac{D}{T}\right)^{2/9} \tag{6.22}$$

6.4.3
Multiphase Systems: Solid–Liquid Systems

Many unit operations involve solid–liquid mixing including dispersion of solids, dissolution and leaching, crystallization and precipitation, and solid catalyzed reactions. For the STR to operate effectively, the solid phase must be dispersed through the fluid volume; dispersion is caused by transfer of mechanical energy from the agitator to kinetic energy in the liquid and particles.

Several factors affect the dispersion of the solid. These include the mechanical design of the tank (T; impeller type, D/T, C/T, etc.), the physical properties of the liquid (ρ, μ) and solid (ρ_S, particle size, d_P, or particle size distribution), as well as the operating parameters (e.g., solids concentration, solids volume fraction, impeller speed, power input, and liquid fill height).

The suspension of a dense solid ($\rho_S > \rho$) is shown with increasing impeller speed in Figure 6.12. As the impeller speed is increased, the particles leave the bottom of the vessel (Figure 6.12a) and become suspended in the liquid. Eventually, a condition is reached where no particle spends more than 1–2 s at the bottom of the vessel, and the entire surface area of the particles is exposed for mass transfer or reaction. This is termed the *Zwietering condition* (Figure 6.12b) and is characterized by the speed required to just suspend all the particles, N_{JS}. It should be noted that under this condition a vertical concentration gradient of particles is

Figure 6.12 States of solid suspension (a) on bottom; (b) just suspended; and (c) homogeneous dispersion. (From Atiemo-Obeng *et al.* [19]).

still observed; the height reached by the particles within the vessel is termed the *cloud height*. Increasing N beyond N_{JS} improves the homogeneity of the suspension (Figure 6.12c) but does not greatly improve the mass transfer characteristics.

6.4.3.1 Particle Suspension and Flow Patterns

N_{JS} can be calculated using the Zwietering correlation [20]. This holds for STRs with diameters from 0.15 to 0.6 m (3–170 l), particle sizes from 125 to 850 μm, liquid viscosities from 0.0003 to 0.0093 Pa s, and solid mass percentages between 0.5 and 20%. The best operating conditions are generally taken as 10% above N_{JS}

$$N_{JS} = S \left(\frac{g\Delta\rho}{\rho_L} \right)^{0.45} d_P^{0.2} X^{0.13} \left(\frac{\mu}{\rho_L} \right)^{0.1} D^{-0.85} \tag{6.23}$$

where S = the suspension parameter. This is related to *Po* and *Fl*, thus it is impeller and geometry specific, but independent of the scale for turbulent flow. Selected values from Mak 1992 [21] are tabulated in Table 6.3a. X = mass of solids/mass of liquid × 100 (i.e., %), $\Delta\rho = \rho_S - \rho_L$. Thus, for geometrically similar systems,

$$N_{JS} \propto D^{-0.85} \tag{6.24}$$

While this correlation has been proved to be highly robust at larger scales, wall effects can cause significant deviations when extrapolating the correlation to smaller laboratory scales (<3 l), which is important for scale-down. A modified Zwietering criterion is proposed for scale-down from the 10 to 1 l scale (assuming a geometrically similar system and identical properties of the solid and liquid phases). The constant S in Equation 6.23 is replaced by a constant Q, which is a function of scale and particle size as well as geometry [22]. This approach could be criticized as the influence of parameters such as d_P is included twice (via Q and $d_P^{0.2}$), which is a consequence of modification of the Zwietering relationship. Values of Q for pitched blade and RCI impellers typically used for solids suspension are given in Table 6.3b.

Table 6.3 Parameters for Zwietering correlation.

(a) $N_{JS} = S \left(\frac{g \Delta \rho}{\rho_L} \right)^{0.45} d_P^{0.2} X^{0.13} \left(\frac{\mu}{\rho_L} \right)^{0.1} D^{-0.85}$

Impeller type	S
Lightnin A-310, D = T/2, C = T/4	7.1
45° PBT, D = T/3, W = D/3, C = T/4	4.8
45° PBT, D = T/3, W = D/3, C = T/6	4.6
45° PBT, D = T/3, W = D/3, C = T/8	4.2

(b) $N_{JS} = Q \left(\frac{g \Delta \rho}{\rho_L} \right)^{0.45} d_P^{0.2} X^{0.13} \left(\frac{\mu}{\rho_L} \right)^{0.1} D^{-0.85}$

Particle size fraction (μm)	PBT 1 l	PBT 10 l	RCI 1 l	RCI 10 l
<80	1.85	3.09	1.86	2.86
80–125	1.99	–	1.96	–
125–200	2.68	3.25	2.46	3.7
200–315	3.48	3.76	3.05	4.23
>315	3.81	4.17	3.52	4.27

Choice of Geometry is Critical for Efficient Suspension   Experiments have shown that for axial devices, optimal geometric ratios are $D/T \approx 0.4$ $C/T \approx 0.2$ (e.g., for Lightnin A310, PBT). For the particle to remain in suspension, it is necessary for the hydrodynamic forces to overcome the particles' propensity to settle; the equilibrium-settling velocity of a particle in free fall is the terminal velocity, $V_t (\text{ms}^{-1})$. While V_t is normally evaluated assuming an infinite bounded fluid, the settling is hindered by both interactions with other particles and particle–fluid coupling. For monodisperse suspensions, Maude [23] gives a preliminary estimate of this effect as

$$V_{ts} = V_t (1 - \varphi)^n \tag{6.25}$$

where V_{ts} = hindered settling velocity (m s^{-1}). The values of n are dependent upon the particle Reynolds number, defined as $Re_P = \frac{\rho_L V_t d_P}{\mu}$

$$\begin{aligned} Re_P &< 0.3 & n &= 4.65 \\ 0.3 &< Re_P < 1000 & n &= 4.375 Re_P^{-0.0875} \\ Re_P &> 1000 & n &= 2.33 \end{aligned}$$

Unhindered settling velocities for particles are given by Ref. [24]

$$50 < d_P < 1500 \,\mu\text{m}: \quad V_t = \frac{0.152 g^{0.71} d_P^{1.14} \Delta \rho^{0.71}}{\rho_L^{0.29} \mu^{0.43}} \tag{6.26}$$

$$d_P > 1500\ \mu m: \quad V_t = \left(\frac{4}{3}\mathrm{g}d_P\,\Delta\rho/\rho_L\right)^{1/2} \tag{6.27}$$

6.4.3.2 Solid–Liquid Mass Transfer

The mass transfer is characterized using a solid–liquid mass transfer coefficient, k_{SL} (m s^{-1}), where

$$R = k_{SL}a_P V(C_S - C) \tag{6.28}$$

where R is the rate of mass transfer (kg s^{-1}), C_S is the concentration at the solid surface (kg m^{-3}), C is the concentration in the bulk liquid (kg m^{-3}), V is the vessel volume (m^3), and a_P is the interfacial area of solid per unit volume (m^2 m^{-3}) defined as

$$a_P = \frac{6\varphi}{d_P} \tag{6.29}$$

where φ is the solid's volume fraction in the vessel.

For fast reactions, solid–liquid mass transfer limits the reaction rate. For a first-order reaction, this can be assumed if $K/k_{SL} > 100$, where K is the true kinetic rate constant. Solid–liquid mass transfer can be predicted using a modified *Froessling equation* for $N = N_{JS}$ [24].

$$Sh = 2 + 0.72\,\mathrm{Re}_P^{1/2}\,Sc^{1/3} \tag{6.30}$$

where $\mathrm{Re}_P = \frac{\rho_L V_t d_P}{\mu}$; $Sh = \frac{k_t d_P}{D_M}$; $Sc = \frac{\mu}{\rho_L D_M}$ and k_t is the solid–liquid mass transfer coefficient corresponding to a particle terminal velocity V_t (calculated from Equations 6.26 and 6.27). The mass transfer coefficient at the just-suspended condition, k_{JS}, is then obtained by calculation of the enhancement factor, E, observed because of the turbulent nature of the flow [24].

$$\frac{k_{JS}}{k_t} = E = \left(\frac{d_P}{40 \times 10^{-6}}\right)^{0.08} \tag{6.31}$$

The above N_{JS} experiments show that $k \propto N^{1/2}$.

6.4.4
Multiphase Systems: Gas–Liquid Systems

Gas–liquid reactions are often performed for large-scale hydrogenations or oxidations, which generally exhibit significant exotherms and selectivity issues. At a smaller scale, gasification may be used to create product microstructure (bubbles), for example, for foams, mousses, and so on. Gas–liquid mixing is usually performed using a standard geometry of $C/T = \frac{1}{4} - \frac{1}{2}$, $D/T = \frac{1}{4} - \frac{1}{2}$ with a dip pipe or sparger ring being used to introduce gas beneath the impeller. The gas flow, Q_G, is generally defined in vessel volumes per minute (vvm), and conversion to SI units of m^3 s^{-1} is necessary for use in dimensional equations.

Impeller choice encompasses both radial and axial designs: for the former, concave (hollow) blade designs will give more stable operation by preventing the

formation of gas cavities behind the blades. The cavities can dramatically reduce the transfer of energy from the impeller to the liquid phase in conventional flat-blade designs such as the Rushton turbine. For the latter, up-pumping designs are preferable to down-pumping, which have been shown under certain circumstances to exhibit flow instabilities that affect efficiency [25]. Examples are given below:

Radial – concave blade designs (e.g., Chemineer BT6, Scaba SRGT)

Axial – up-pumping wide blade hydrofoil (Lightnin A315/A340, Prochem Maxflo).

The role of the agitator is to break the gas into small bubbles for high interfacial area and disperse the bubbles through the liquid, while keeping the bubbles in the liquid for sufficient time for a reaction to occur. Of course, the agitator must still perform necessary liquid-mixing duties.

6.4.4.1 Power Consumption

The power consumption within a gas–liquid mixing system also needs to consider the contribution of the rising gas stream, that is,

$$\frac{P_T}{V_L} = \frac{P_g}{V_L} + \rho_L v_s g \tag{6.32}$$

where P_T is the total gassed power input, P_g is the mechanical power from the impeller when gas is present, v_s is the gas superficial velocity ($v_s = 4Q_g/\pi T^2$), V_L is the liquid volume, and g is the acceleration due to gravity. However, generally the second term on the right-hand side of Equation (6.32) is small compared to the first term.

P_g can be observed to drop on addition of gas because of the formation of gas cavities around the impeller blades; hence the turbulent power number of the impeller, Po, used for single-phase calculations must be applied with caution. This effect is minimized by the use of modern impeller designs described above but should, nevertheless, be factored into calculations using manufacturers' data.

6.4.4.2 Gas Hold-Up and Flow Patterns

The gas holdup, φ, is defined as the volume fraction of gas in the system. Gross measurement may be obtained from the height change within the vessel on introduction of gas, although due to instabilities at the top surface of the vessel this is prone to error.

$$\varphi = \frac{H_G - H_U}{H_G} \tag{6.33}$$

where H_G is the height of liquid + gas in the vessel and H_U is the height of liquid in the vessel alone. Ranges of φ are usually 10–20% – extremes are 5–50%. φ is normally correlated to the average energy input per unit mass $\overline{\varepsilon}_T$, v_S, and the absolute temperature, θ (K).

$$\varphi = A \left(\overline{\varepsilon}_T\right)^B v_S{}^C \theta^D \tag{6.34}$$

For air–water systems with multiple impellers, $A = 70 \times 10^6$, $B = 0.20$, $C = 0.55$, and $D = -3.2$. For single impeller systems, this value of φ will be reduced by 35%.

Figure 6.13 Flow patterns in gas–liquid systems. (From Nienow *et al.* [26].)

Similar to solid–liquid mixing, an optimum flow pattern for mass transfer operation can be defined. Figure 6.13 illustrates the observed flow patterns in a gas–liquid mixing vessel with increasing N at constant Q_G (or decreasing Q_G at constant N). The following features are identified:

– (a) *Flooding*: Impeller flooded. Gas passes through agitator. Liquid flows over outer blades of agitator.
– (b–c) *Loading point*: $N = N_F$. Gas is captured by impeller blades.
– (d) Condition is reached where gas is dispersed throughout the vessel – all the vessel volume is being used. Denoted as *complete dispersion* $N = N_{CD}$.
– (e) *Gross recirculation* of gas back into impeller. $N = N_R$.

The operating point for a solid–liquid mixing process is generally 10–20% above N_{CD}. Correlations for N_{CD} (and also N_F, N_R) are generally expressed in terms of the gas flow number as a function of D/T and the Froude number, for example, for a single six-blade Rushton turbine when $T < 1.8$ m [26]

$$\frac{Q_g}{N_{CD}D^3} = 0.2 \left(\frac{D}{T}\right)^{0.5} \left(\frac{N_{CD}^2 D}{g}\right)^{0.5} \tag{6.35}$$

6.4.4.3 Mass Transfer

As for solid–liquid mass transfer, we can define an overall transfer rate based on a gas–liquid mass transfer coefficient, k_{LA}; thus

$$R = k_L a V \left(C^* - C\right) \tag{6.36}$$

where C^* is the concentration in the gas phase (kg m^{-3}), C is the concentration in the bulk liquid (kg m^{-3}), and a is the interfacial area of gas per unit volume (m^2 m^{-3}) defined as

$$a = \frac{6\varphi}{d_{32}} \tag{6.37}$$

where d_{32} is the volume to surface area mean diameter, known as the *Sauter mean diameter* of the gas bubbles. For gas–liquid systems, because a is determined

by the multiphase hydrodynamics within the vessel (unlike solid–liquid systems where a_P is determined *a priori* by the properties of the solid phase added), it is difficult to decouple k_L from a. Correlations for $k_L a$ are generally of a form similar to Equation 6.34

$$k_L a = \alpha \, (\bar{\varepsilon}_T)^{\beta} \, v_S{}^{\gamma} \tag{6.38}$$

It should be noted that owing to extreme nonlinearities in the behavior of gas–liquid systems, one should never extrapolate beyond the ranges of these correlations. For air–water systems at $20\,^{\circ}$C, $\alpha = 1.2$, $\beta = 0.7$, and $\gamma = 0.6$ [27].

6.4.5
Summary

A summary of the above arguments on scaling can be shown in Figure 6.14 [28], which plots the ratio $(\bar{\varepsilon}_T)_{\text{PLANT}}/(\bar{\varepsilon}_T)_{\text{LAB}}$ versus $V_{\text{PLANT}}/V_{\text{LAB}}$. The issue of scaling at constant mixing time is self-evident: a 100-fold increase in $\bar{\varepsilon}_T$ for a 1000 times increase in volume requires a power increase of 100 000! This precludes use of this criterion unless the change in scale is very small. Scaling up at constant solid suspension, shear rate, or heat transfer causes $\bar{\varepsilon}_T$ to decrease on scale-up, with consequent changes in the small-scale turbulent behavior. For most general duties, scale-up at constant $\bar{\varepsilon}_T$ provides the best compromise and, as a first approach, it is clearly important to ensure that the flow regime is the same (i.e., both in the turbulent regime) at both scales by calculation of the Reynolds number (Equation 6.13). It should be noted that, for specialist reactions, particularly bioreactions, it may be necessary to consider alternative scale-up methods where geometric similarity is abandoned in order to ensure that local species concentrations or shear do not cause death of the organism.

Figure 6.14 Scale-up diagram (redrawn from Penney [28]).

6.5
Chemistry Effects in Scale-Up

Different chemistry demands different approaches to scaling of the reaction environment. The subject of liquid-phase reactions is very varied. It can be readily associated with long, slow, homogeneous-phase organic reactions that are limited by the slow kinetics and the low concentrations that are frequently used for selectivity reasons. This type of reaction, of course, in principle gives no significant issues on scale-up, because the rate-determining process is the reaction itself and changing the scale of operation does not change this fact. If, therefore, an irretrievably slow reaction is demonstrated, then this is probably sufficient to allow a simple scale-up based on the required reaction time and thus batch volume to achieve a given production rate.

It is when the overall process rate is not solely controlled by the reaction kinetics that the multivariable problems of scale-up must be considered in more detail. It may also be noted that where the rate-determining step is chemical at the small scale but becomes a physical rate process on scale-up, then this can change the process outcome (yield or selectivity). The following paragraphs give an overview of dependencies and sensitivities that are commonly encountered, but without reference to specific chemistries.

6.5.1
"Fed-Batch" Liquid-Phase Reactions

Scale-up problems in single-phase liquid reactions generally occur when there is a relatively fast reaction. This is frequently characterized by a fed-batch operation; that is, one of the reagents is fed into the reaction vessel during the reaction.

The slow liquid reagent feed is used essentially to ensure good mixing of the fed reagent, that is, mixing of the reagent at a rate equivalent to or faster than the reaction. High local concentration due to an excessive feed rate or slow dissipation of that feed, can result in poor reaction selectivity. This is especially true for oxidation and hydrogenations using stoichiometric reagents. In the case of reactive crystallization and precipitation, poor quality product may arise, especially where polymorphism is prevalent or where solution occlusion may have a significant impact on purity. Equally, the fed-batch design may be used to limit the rate of reaction in order to mitigate a very high exotherm and facilitate better temperature control. In these cases, it is effective to control the rate of reagent addition using temperature.

The feed rate may be a critical design and operating variable. If it is not correctly tuned to the rate of feed dispersion or mixing time of the vessel, then changes in reaction performance will most likely be observed.

One very emphatic piece of advice is on the use of dip-pipes versus surface feed. If mixing of the reagent is critical to reactor performance, then subsurface feed into the impeller region should be recommended as it provides faster dispersion of the feed "plume." The magnitude of these effects has been demonstrated for precipitation

by Peleka *et al.* [29] and the large differences in local mixing energy reported by Edwards *et al.* [30]. There is a common operational preference to avoid the use of dip-pipes: fouling and blockage, interbatch cleaning, and breakage. To paraphrase Ed Paul, however, ''Don't think why you need to use a dip pipe – consider rather why you don't need one'' [1]; in other words, the subsurface feed should be the default design.

There is an enormous volume of research work that has been devoted to the subject of mixing effects on competitive reactions. Experimentally, this is for the most part based on the so-called Bourne reactions (there are others!), the earliest of which were developed in the 1970s. These deliberately competitive reactions in general feature reactions of different rates such that selectivity can be ascribed to different scale-mixing effects. There are many research papers and reviews on this subject area. For a comprehensive treatise, Baldyga and Bourne's book on the subject is seminal [31]. For a version focused on the fine chemicals industry, readers are referred to a more recent review [32].

6.5.2
Liquid–Solid Reactions

In liquid–solid reactions, the key design issue tends to achieve adequate suspension of the solid. As an aside, conical-based vessels will give a much higher tendency to solids accumulation in the reactor base (and blockage of any drain valves) than dished- or flat-bottomed vessels. Inadequate suspension can result in slow reaction and extreme reagent concentrations in the particle-rich zone of the reactor. In some cases, agglomeration of the solids can exacerbate these problems. In other cases, the solid may tend to float rather than sink and a different mixing strategy is required. For all of these cases, the ''slurryability'' of the particulate solid is an important characteristic and this may dominate the design requirements of scale-up if a dry solid or wet solid (e.g., filter cake) is to be added at the start of or during the reaction. This should be the first consideration in scale-up studies: characterizing the solid–liquid mixing.

There is a huge variety in liquid–solid reactions (i) where the solid is a dissolving reagent, (ii) where the solid is a catalyst (e.g., solid acid catalyzed esterification), or (iii) where the solid is a reaction product (e.g., crystallization). All present somewhat different problems.

For solid feed, as noted above, its slurryability is critical. This must be tested and, if problematic, then a systematic and fundamental approach should be taken to scale-up. Equally, if breakdown of solids agglomerates is required, then specialist mixing design is required. These may involve, for example, high shear (or high power per unit volume) mixing for an agglomerated feed, or a specific design for ''draw down'' of floating solids if that situation prevails. Specific guidance and underpinning principles [19] need to be considered if the solid–liquid mixing aspect is not to limit and control the reaction on scale-up. It should be remembered that while it is easy to ''move the beaker'' to achieve dispersion of a difficult solid

in the laboratory, this is not an effective strategy at manufacturing scale! Such practical mixing issues must be addressed at an early stage.

Does this really affect the resulting chemistry? In many cases, the answer can be yes.

For a dissolving solid-based reaction, the rate of dissolution will presumably control the reaction rate, and therefore there will be sensitivity to particle size (and its distribution) and the solids loading. This may not be overly sensitive to scale-up in a well-designed vessel. The basics of solids suspension and its scale-up are presented in Section 6.4.3.

For a solid catalyzed liquid-phase reaction, once adequate solid suspension is achieved then mass transfer rates are also likely sufficient, with the majority of heterogeneously catalyzed liquid-phase reactions being significantly influenced, if not controlled, by the intraparticle diffusion. This effect should be checked by varying particle sizes of course.

Solids forming reactions are very different, and mixing is generally a critical factor in determining the primary particle size and shape, agglomeration, and crystal morphology of the particles. A detailed discussion of this particular process is beyond the scope of this text.

6.5.3
Gas–Liquid (–Solid) Reactions

The predominant case here is for a gaseous reactant: hydrogenation, oxidation, carbonylation, and hydroformylation. Brief consideration is also given to the less common case of gas-evolving reactions such as dehydrogenation.

6.5.3.1 Gaseous Reactant
The challenge for the reaction engineer is to achieve adequate dispersion of, generally, a sparged gas. With scale-up this becomes increasingly difficult and the associated gas–liquid mass transfer coefficient declines. This reduces the availability of the gaseous reactant in the liquid phase, which is then adsorbed at any catalyst surface. The availability of the reactant will naturally reduce the observed reaction rate, leading to longer batch times. Equally importantly, the change in relative dissolved concentrations can frequently affect selectivity – typically adversely.

Evidently, the response of the chemistry to deterioration in the rate of mass transfer is a vital consideration for scale-up. It should also be remembered that for exothermic reactions heat removal will also decline with scale and thus temperature sensitivity is also very important.

6.5.3.2 Gaseous Product
The fundamental difference here is that the object is to allow the gaseous product to escape from the liquid phase, while providing sufficient agitation to maintain the catalyst suspended. The ability to perform this task will change with scale. Dehydrogenations can be equilibrium limited and thus the disengagement of the

gas can be a limiting factor. This may not only retard the rate [33, 34] but also lead to alternative reactions and loss of selectivity.

6.5.3.3 Catalysis

In general, catalytic effects will be considered *inter alia* as part of the reaction and chemistry characterization. Two key variables can, however, be overlooked:

- Catalyst loading: has a significant effect on overall rate – linear, of course under intrinsic kinetic conditions. Catalyst loading is a useful design variable to limit the reaction rate, say, to restrict an exotherm generation rate and thus enable improved temperature control on scale-up. Reducing catalyst concentration can, however, have unexpected effects such as limiting total conversion. This may arise where the product or an impurity is adsorbed onto the catalyst surface, effectively blocking access to active sites. Occasionally, catalyst loading can also affect the selectivity of a reaction.
- Particle size – can have significant effects on the reaction rate and selectivity. It is not unreasonable to presume until proven otherwise that a heterogeneously catalyzed liquid-phase reaction will be influenced or limited by intraparticle diffusion. Increasing particle size will, in these circumstances, reduce the ''pellet effectiveness'' and ostensible activity per unit mass of catalyst. Changing particle size may also change the intraparticle and surface concentrations and thus selectivity. Such effects should be thoroughly assessed.

6.5.4
Impurities

It may seem trite, but the reagent quality can be an important issue. It is common to carry out laboratory-scale development work with ''research'' grade chemicals. It is important to consider, however, what effect impurities in a ''commercial'' grade feedstock may have, irrespective of whether that feed is manufactured internally or purchased.

6.6
Achieving Process Understanding for Reactor Scale-Up

From the above, it is clear that there are many factors to consider in the scale-up of a chemical reaction, and that these factors vary from one reaction to another. In order to take a rational and ultimately successful approach, it is essential to be systematic in the search for relevant and critical information and understanding. The fact is that of the myriad of variables only a relatively small number will control and determine scale-up behavior (or misbehavior if you will). The focus of activity should therefore be on identifying the key scale-up parameters (chemistry and engineering) and gaining an understanding of them. To paraphrase, which variable or parameter will hurt me most if I get it wrong?

In additional to this, it is essential to understand the performance characteristics and limitations of the scaled-up reactor. In many instances, this equipment will be extant; in fewer instances, it will require design and construction.

6.6.1
Chemistry Scale-Up Sensitivity

The best way to explore the sensitivity of a given chemistry to scale-up is to experimentally investigate the reaction response to the typical scale-up variables. This need not necessarily involve a full kinetic study, but does need to be adequate to fully interrogate the key parameters. Experimental techniques and how to approach intrinsic kinetic measurements are presented elsewhere [35, 36] and will not be further discussed here. Suffice to say it is always preferable to have intrinsic kinetic data, and to have some knowledge of where transport effects become significant. This may seem an esoteric requirement, but an objective will be to decouple kinetic effects from transport effects. Paul [1] notes that a development strategy must focus on defining the kinetic relationships of the reaction system so that the strictly chemical issues can be addressed and separated from the scale-up issues. He notes further that this type of analysis can lead to a further broad characterization of complex reaction systems as follows:

- Reactions that require resolution of kinetic problems in order to be run successfully in the laboratory
- Reactions that can be run successfully in the laboratory but which require special plant design considerations and equipment
- Reactions that pose both special laboratory and scale-up problems.

The key point is that an experimental strategy that allows the chemical and scaling effects to be decoupled is an absolute requirement. This can only be achieved with confidence if the chemistry-oriented experiments are carried out under conditions where the reaction is not influenced by transport and mixing effects.

In terms of achieving adequate understanding of the chemistry, the following variables should be considered:

- **Reaction network** – what are the significant by-products? The strategy for an $A \Rightarrow B \Rightarrow C$ (series) selectivity will be different to that for an $A \Rightarrow B$ and $A \Rightarrow C$ (parallel) system.
- **Reagent concentration (if a solvent is used)** – a straightforward power law kinetic model may be readily derived for simple reaction networks with relatively few experiments. The key, however, is to establish quickly whether the reaction kinetic effects are ''simple.''
- In the case of a fed reagent, the feed rate is the analog of feed concentration and experiments are carried out specifically to vary this and establish sensitivity.
- In the case of a gas–liquid reaction, the pressure of the gas phase is of course the analog of concentration.
- **Product concentration** – does the product inhibit the reaction? If yes, then is this an equilibrium limitation or, say, adsorption on the catalyst? Does product

reaction or degradation occur, and, if yes, then is this sensitive to temperature and other reactants?

- **Spectator effects** – what effect do components present in the reaction mixture, but not ostensibly reactants, have on the reaction rate and selectivity?
- **Solvent selection** – this is a reaction feature that is commonly set early and in a very empirical and experiential manner using a set of solvent screening tests. The role of the solvent in the reaction chemistry can be, however, far more complex than simple solvation effects [37] including influencing the mass transfer [38, 39] or playing a mechanistic role in the reaction [40].
- **Calorimetry** – gaining an understanding of the rate of heat evolution or adsorption of a reaction under the controlled isothermal conditions of a well-designed calorimeter is essential information. It allows not only a basic understanding of the heating/cooling needs of the reaction (vital for scale-up) but also may provide essential mechanistic information, courtesy of differing heats of reaction.
- **Temperature** – a well-known "rule of thumb" is that the reaction rate doubles for every $10\,^{\circ}C$ increase in temperature. This in fact applies to a reaction with an activation energy of circa $75\,kJ\,mol^{-1}$ at $100\,^{\circ}C$. Rates of reaction will of course increase to a lesser or greater extent than this according to whether their activation energy is below or above the given value. Selectivity may therefore be especially sensitive to temperature. This may be due to competitive reactions of differing activation energy, catalyst effects, or due to consecutive reaction of the product.
- **Safety and runaway** – this tends to be a hazard resulting from highly exothermic reactions, characterized by a progressive increase in the rate of heat generation, temperature, and pressure (the latter generally caused by vaporization or decomposition of the reacting mass) and starts when the rate of heat generation exceeds the rate of heat removal [41].

The above may seem like a very long shopping list; it can, however, be classified into some basic interrogative queries that imply the critical information required for successful scale-up:

- What is the reaction network? What are the key side reactions and by-products?
- Are the reaction kinetics simple – namely, power law kinetics, with little or no complication from adsorption and inhibition effects? This includes reagents, products, and "spectator" species.
- Is the reaction selectivity sensitive to mixing and/or mass transfer effects?
- Does the reaction evolve significant heat and is the reaction sensitive to available heat transfer and temperature excursions?
- Is there a potential for reaction runaway?

6.6.2
On the Acquisition of Relevant Chemical Information

A detailed experimental program exploring the kinetic and reactor design variables systematically and in a statistically relevant manner would always be desirable.

In practice, however, this is not always feasible or necessary. The experimental program may be conveniently tailored to answering the critical questions posed above - namely, is the reaction performance (in terms of rate and selectivity) sensitive to the presence of contaminants, reduced heat transfer, and lower rates of mixing and mass transfer? If yes, then to what extent and which are the prime influences?

In terms of characterizing the reaction kinetics, there are a number of key factors in the design of the experiments to consider:

1) It is always recommended, where feasible, to carry out these experiments under intrinsic kinetic conditions. In gaining understanding for process scale-up, it is absolutely essential to decouple the reaction kinetics from reactor behavioral characteristics.

    In some cases, for very fast reactions where the dispersion of the fed reagent is inherently limiting, such as precipitations or sulfonations, this will not be possible. Knowing this is, however, a first step in characterizing the reaction to be scaled!

2) Samples should be taken as frequently as possible through the reaction. End of run data do not yield rate or kinetic information. By contrast, batch runs with good temporal resolution of the data do yield sufficient kinetic information, and also allow the fitting of kinetic models [42].

3) For heterogeneously catalyzed reactions, consider the effects of catalyst deactivation and reuse. Is an apparent non-first-order dependency on catalyst concentration a result of a mass transfer limitation or poisoning of the catalyst by a contaminant? Catalyst deactivation can be falsely ascribed to excessive but reversible adsorption effects; see, for example, Mounzer *et al.* [43].

The specifics of suitable equipment and experimental methods are dealt with elsewhere [35, 36] and introductions to methods for evaluating transport and mixing influence on the reaction kinetics are also widely available, and it is not the role of this text to explore these further.

''Design of experiments'' techniques can be beneficially used. This gives information on the cross correlation of variables as well as allowing their importance ranking. Success in such approaches is, however, heavily dependent on good selection of the variable values (which requires a significant amount of prework) and in the happenstance of the system having clear demarcation of the variable influence. Oversimplification, such as Taguchi, can, however, lead to obscuration of results, leading, in many cases, to a conclusion that all variables are of similar importance.

6.6.3
On the Acquisition of Relevant Reactor Design Information

Once the role of the key chemical variables is understood, the effects of the key reactor operating or design variables can be explored. This is best and simply done by a systematic variation of those variables over a relevant range at the laboratory scale. The critical output here is not a complete model of the process design and

operating variables but rather an understanding of which process and scale-up variables are the most important.

The key scale-up variables are heat transfer, mixing, and/or mass transfer. These can be adequately evaluated by systematically varying their defining parameters at the laboratory scale before any attempts to scale-up are made.

- Rate of heat transfer, which will decrease with increasing scale, $Q = UA \, \Delta T$, where U is the overall heat transfer coefficient, A is the available heat transfer area, and ΔT is the temperature difference. It is very difficult to satisfactorily vary "U." A commercial laboratory reactor design is available that allows variation of "A" [44], and this is of course a major effect in scale-up. The temperature-driving force may of course be varied by changing the heating/cooling medium temperature. It may be far simpler, however, to consider the effect of temperature on the reaction rate, heat evolution, and selectivity. If the reaction is insensitive to temperature, then reduction of heat transfer capability on scale-up is unlikely to be an issue.
- Rate of mass transfer – for G/L, G/L/S, and Liquid/Liquid (L/L) reactions, the rate of mass transfer of a reagent or product across an interface will be an important influence on the reaction rate and may influence selectivity. The rate of mass transfer $M = k_l a \, \Delta C$. The mass transfer coefficient ($k_l a$) can be varied by changing the impeller speed. Alternatively, and in a far more linear manner, for a gas/liquid reaction, the mass transfer driving force (ΔC) can be modified simply by changing pressure.
- Rate of mixing – for homogeneous liquid-phase and fed-batch reactions this will affect local concentration gradients through the vessel. The dispersion of a feed plume takes a finite time and as such there are concentration gradients around the plume as it emerges from its feed pipe (irrespective of whether it is submerged or above surface). There are different concentration ratios of the reactant in different parts of the vessel, and therefore different selectivities can prevail. The dispersion of a feed plume is affected by the location of the feed pipe, the volumetric feed rate, the feed velocity, and the mixing strategy (vessel design plus impeller type and speed). At the laboratory scale, for process investigation purposes, the simplest variables are volumetric feed rate and impeller speed. If the reaction is sensitive to feed rate or impeller speed at the laboratory scale, then it will be sensitive to the same and mixing efficiency when scaled up.

6.7
Reactor Selection

6.7.1
So Which Reactor Can I Use?

The most common scenario of batch reaction scale-up is for the use of existing reactors; and that is the scenario addressed below. The key question here is therefore which available reactor should best be used, or maybe more pertinently,

which of the available reactors should absolutely not be used. It is common (mal-) practice to consider the following:

• Reactor volume versus batch size required
• Design pressure and temperature
• Materials of construction and suitability for the given reagents/solvents
• Environment, health and safety (EHS) considerations such as area classification and containment
• Batch scheduling of the site (viz. availability).

While all of these elements are essential, and none, with the possible exception of the first and last, can be deprioritized, this approach does not guarantee successful scale-up. The correct reactor must additionally match the process needs for mixing and heat and mass transfer in order to successfully scale-up a given chemistry. To do this effectively, of course, means that the characteristics of each reactor be known. That is, a reactor's capability for mixing, heat and gas–liquid mass transfer must be known, at least to a reasonable approximation, in a quantitative manner. Factors that are relevant in this context include the following:

• Generic duty suitability
  – single phase – high or low shear mixing;
  – solid–liquid versus gas–liquid mixing.
• Heat transfer capability – kilowatts per cubic meter of operating volume
• Gas–liquid mass transfer capability ($k_L a$).

This then brings up the question as to how reactors may be characterized and successfully matched to a required reaction duty. For a new reactor, these data should be available from the design. For an extant plant, however, they are probably best measured. Many experimental techniques are available for measuring mixing and mixing-relevant parameters [45]. Relatively few of these, however, translate well into characterizing reactors on a manufacturing plant.

6.7.2
Generic Duty

Solids suspension is best achieved using downward pumping impellers, while gas–liquid dispersion requires up-pumping impellers. These are not interchangeable! Pure radial flow impellers may be used for both duties – but at reduced effectiveness. Low shear impellers will not give high rates of micromixing and are unlikely therefore to be well suited to mixing limited reactions. A clear definition of the mixing suitability of reactors should therefore be compiled and retained.

6.7.3
Characterizing Mixing Rate

Mixing rates are best established by tracer addition. The simplest approach is to add a conducting tracer such as a brine solution to the vessel filled with water (via the feed pipe or dip pipe) and monitor the variation in conductivity, giving

the temporal variability of conductivity and its approach to a new mean value. These data are normally interpreted as a t_{95}; the time to achieve a 95% approach to full homogeneity (see, for example, p 172 in Brown *et al.* [45]). At greater degrees of homogeneity, signal-to-noise ratio and scatter tend to dominate the data. It is generally better to use multiple rather than single probes.

Electrical resistance tomography (ERT) has been demonstrated in the laboratory at plant scale as a useful technique to obtain information on rates of mixing [46], and has been used specifically in mixing-based studies of precipitation [7, 47]. The recent commercial implementation of linear ERT probes [48] has made this a useful plant tool, and these probes can be obtained with good chemical resistance and to good manufacturing practice (GMP) specification. Furthermore, it has been shown that this approach can be used to detect mixing problems, such as caverns and stratification, when processing non-Newtonian fluids [49].

6.7.4
Characterizing Solids Suspension

While observations on ''cloud height'' would be useful, the reality is that manufacturing vessels tend to be opaque. Solids suspension is therefore best evaluated from theory. The Zwietering correlation (Equation 6.23), while old, is reliable and will yield a very good approximation of a vessel's capability at suspending solids. Values for the Zwietering parameter (S) are widely available in many mixing texts [24] and also from mixing equipment suppliers' brochures and web sites.

If the solid has significantly different conductivity than the conducting or dielectric liquid continuous phase, then the electrical resistance or capacitance tomography techniques referred to above may be beneficially used here as well [45].

6.7.5
Characterizing Heat Transfer

Heat transfer basics and correlations have been introduced earlier in this chapter. They provide a useful understanding of the key scale-up and operating parameters that determine the overall heat transfer capability. Many fine chemical reactors are glass lined, in which case the process side heat transfer coefficient is unlikely to be the controlling resistance, but rather the conduction through the glass wall. There will also be uncertainty regarding fouling of the jacket for steel-walled vessels. Overall, therefore, and from a perfectly pragmatic standpoint, it is best to measure the heat transfer – or deduce it from plant batch records.

The easiest approach is to use batch heating or cooling data and deriving the overall heat transfer parameter ''UA'' (see Equation 6.18) using classical batch cooling or heating analysis (see any undergraduate heat transfer text, for example [50]). The value of UA will change as a function of impeller speed, fluid properties (viscosity, conductivity, and presence of gas or solids), and fill level (which changes not only ''A'' but also the fluid flow patterns) but is a good quantitative general

guide to the heat transfer capability achievable that can be compared with laboratory and pilot-scale equipment.

Data from reaction conditions are harder to interpret unless the rate of any reaction heat absorption/evolution can also be assessed; but this will generally require on-line or periodic chemical analysis.

6.7.6
Characterizing Mass Transfer

Mass transfer measurements are harder to make on a plant without deliberate trials.

Reaction rate data may be useable if hydrogen (or other reactive gas) uptake rate data are available for a fast (mass transfer limited) reaction at constant pressure. If this is the case, then the fastest achievable hydrogen uptake rate can be used to estimate the volumetric mass transfer coefficient ($k_L a$ in Equation 6.36) from $N = k_L a \, p$, where N is the hydrogen uptake rate (mol s^{-1}), $k_L a$ is the volumetric mass transfer coefficient (mol s^{-1} bar^{-1}), and P is the hydrogen pressure (Pa). This is a significant simplification as it ignores hydrogen solubility (assume liquid-phase concentration is zero), an expedient in the absence of hard-to-measure liquid-phase dissolved gas concentrations. Probes are, however, now commercially available to measure the dissolved concentration of various common gaseous reactants, such as hydrogen [51], and thus this approach may be taken more rigorously.

Off-line methods for mass transfer measurement are several. Dynamic methods, such as oxygen absorption, do not have a good reputation for reliability and accuracy and are generally not recommended.

A good alternative for plant vessels is ''pressure step'' methods [52–54]. In these cases either, for a surface-aeration-based mass transfer the vessel is pressured up, the agitation started, and the decline in pressure noted, or, for a sparged vessel the set pressure is given a step increase at fixed agitation.

6.8
Exploiting Process Understanding in Scale-Up

It must be emphasized that for fully reliable scale-up a rigorous approach to kinetics, mixing, and transport is required, as advocated in classic reaction engineering [55] and mixing texts [2, 3]. The current authors would also support such a rigorous approach. The reality, however, is that in many instances the fine and intermediate scale chemical industry ''does not have time'' to do it properly. It must be emphasized that this may in fact be a false economy in that the incidental costs of bad scale-up can far exceed those of doing it properly and the elapsed time of fire-fighting can easily approach that for a more thorough scale-up study. That said, this single text is unlikely to change industry perception, and therefore what is presented here is a rational approach to scale-up that reduces risk and maximizes probability of success by understanding the critical process parameters in reaction

scale-up. It is not, however, foolproof and does not guarantee success. Furthermore, it does not eradicate the need for careful, scale-up specific experimentation at the laboratory and pilot scale, but rather is totally dependent on carrying this out at the appropriate time.

The preceding sections of this chapter have presented an example of how scale-up can go wrong, an introduction to the relevant essentials of mixing, a strategic approach to understanding the critical reaction parameters for scale-up and characterizing extant reactors. Scale-up itself, at its most simple, matched the last two of these. It is not possible to present a universal solution, as this varies from one chemistry or reaction to another.

In summary, however trite as they may sound, the following are some examples:

- If the reaction is mixing time limited, then the reaction must be slowed in a controlled manner to accommodate the slower mixing at the plant scale.
- If the reaction is heat transfer limited and temperature sensitive, then the reaction must be slowed to accommodate the slower heat transfer of the larger reactor, thus maintaining the operating temperature within the preferred range.
- If the reaction is mass transfer limited and sensitive to this parameter, then the reaction must be slowed to accommodate the lower mass transfer rates at the larger scale.

The sections below give a little more background thinking on these that will assist in describing the general principles.

6.8.1
Mixing Rate-Limited Reactions

The classic example here is a precipitation or a homogeneous liquid-phase reaction where lab and plant operation are in the ''fed-batch'' mode, where one reactant is added to the reactor containing the remaining reactants. The progress of the reaction is in essence dependent on the dispersion of the feed plume. The approach to scale-up is to balance the rate of addition to the rate of mixing. In short, if the mixing time of the plant scale is 10 times that for the laboratory or pilot scale, then to a first approximation the rate of addition should be slowed by a factor of 10. A more normal conservative approach will (arbitrarily or experientially) put a (say) 20% margin on this, thus extending the addition and reaction time by a factor of 12. This is why it is essential to characterize the mixing parameters (t_{95} in this instance) at the various scales. Without this understanding, accurate scale-up is not possible. In this type of reactor, mixing time will also be a key parameter in reactor selection.

The above strategy may not realize adequate scale-up. One additional important aspect to consider is the effect of shear and the scale of mixing (Section 6.3.1.2). Reaction occurs at the molecular scale, and therefore mixing at the Kolmogorov length and below is important. In low shear mixing, the Kolmorgorov length is larger, and thus the diffusive timescale associated with sub-Kolmogorov mixing is longer, so concentration gradients are able to persist to a greater extent. This can

in turn lead to variation in the selectivity of mixing-limited reactions. The effects of shear, not just global mixing times, need to be assessed at the laboratory and/or pilot scale by the use of different impellers. If this is a key parameter in scaling the reaction, then scale-up should be considered from a rigorous mixing design basis.

6.8.2
Solid−Liquid Mixing-Limited Reactions

The key scenario here is maintaining adequate or full suspension of the solids. The critical consideration is Zwietering: see Equation 6.23 Section 6.4.3.1 presents additional information on scaling for constant solids suspension and this should be adhered to. The key in this is selecting a vessel equipped with an impeller possessing a high Zwietering constant ''S'' value.

If, however, adequate suspension does not appear feasible in the scaled-up reactor, alternative approaches must be considered. To identify these, refer again to the Zwietering equation, which indicates that there is a weak dependency of N_{JS} on particle size and solids loading.

- Reducing particle mean diameter will have a beneficial effect on solids suspension (only $d_p^{0.2}$ though). The increased specific solid surface area may, however, affect the reaction and this should be checked at the laboratory scale.
- Reducing solids loading also has a slightly beneficial effect ($X^{0.13}$) and will provide some mitigation unless the liquid-phase reagent is present in significant excess; however, it will be necessary to consider the effect of changing solids/liquid reagent ratio. Use of a reduced reagent concentration may be required (increased solvent use) to balance the chemistry effect.

6.8.3
Heat-Transfer-Limited Reactions

As noted previously, deterioration in the overall rate of available heat removal with scale is almost inevitable. Many reactions in the fine chemicals industry exhibit strong exotherms, and, to a lesser extent, endotherms. The problem on scale-up is to manage the exotherm in terms of maintaining the reaction mixture within a temperature range identified at the laboratory scale. There are essentially two ways to achieve this:

- To slow the reaction rate to match the reduced rate of heat removal. This can be achieved using one of several levers such as reacting gas pressure, catalyst loading, and reagent concentrations.
- To mitigate the adiabatic temperature rise by using a reduced concentration in a solvent, the solvent substantially acting as a heat sink. Introduction of solvent, or increasing the amount of solvent of course has a direct impact on downstream product isolation process needs and costs.

Either of these strategies will potentially affect the chemistry. For example, based on simple reaction kinetic considerations, changing the reactant concentration will alter selectivity unless the main and side reactions have the same reaction order and likewise reducing gas pressure will influence the liquid phase and catalyst surface concentration of the gaseous reactant (introducing potentially a mass transfer limitation). These effects will need to have been explored at the laboratory scale in order to determine the most promising (or least damaging) strategy.

If heat transfer limitations cannot be satisfactorily managed, then it may be necessary to consider alternative reactors. Flow reactors and continuous processing [56, 57], for example, are eminently capable of handling extreme heat transfer requirements.

6.8.4
Mass-Transfer-Limited Reactions

Scale-up will normally be accompanied by deterioration in G/L mass transfer rates. If the reaction selectivity is affected by this, then scale-up must be sympathetic. There are three strategies that may achieve this, or at least offer some mitigation:

- Operate at higher gas pressure. Mass transfer rates are ostensibly linear with pressure and this may offer some mitigation but is unlikely to fully compensate.
- Reduce the reaction rate by operating at lower catalyst loading or reduced reagent concentration (solvent use).
- Operate at lower temperature, thus not only reducing the reaction rate but also increasing (in most cases) gas solubility. Note also, though, that deceased temperature will reduce diffusion rates and thus mass transfer rates and will also affect gas bubble sizes. The overall relationship between these parameters is not simple, but generally the reaction rate effect is the most important.

All of the above potentially affect the chemistry, thence reaction selectivity, in addition to overall rate. This sensitivity will need to have been explored as part of the laboratory experimental program once mass transfer limitations have been identified as a critical scale-up parameter.

6.9
Conclusions

The early part of this chapter presented a simulation-based case study of how scale-up can go wrong in reaction engineering. It highlighted the global effects of changes in the heat and mass transport coefficients as the reactor scale changes. This was followed by an overview of mixing theory, and especially as it applies to fine chemicals reactor-type vessels (batch stirred tanks) and typical mixing scenarios in terms of single- and multiphase mixing problems. The fundamentals of mixing scale-up for these systems were also addressed. Subsequently, the effects of changing scale on the observed and actual chemistry were discussed,

highlighting the interaction of mixing and transport with reaction, for a range of reaction scenarios. These sections set the groundwork for achieving adequate understanding of the reaction process to allow a rational approach to scale-up.

A critical step is to decouple reaction and mixing/transport rate effects experimentally at the laboratory scale. Only by doing this can sufficient understanding be achieved to reliably predict or accommodate the effects of changes in mixing and transport with scale. Once, and only once, the chemistry is understood, in terms of the reaction network, some approximate kinetics, and its thermochemistry, the additional mixing and transport effects can be effectively explored. An outline approach to the exploration of the scale-up variables and parameters was discussed. The key output of this is to identify the critical scaling parameters for the given chemistry – the parameters that have the highest priority and control over the reaction. This is the most important step, but for the reaction to be performed reliably requires the predetermination of the chemistry. This aspect cannot be overstated. Once the key scale-up parameters are identified, scale-up itself can start to be addressed.

Two aspects were considered here. First, for the most common scenario of scaling into existing plant, criteria for vessel selection was discussed, and this highlighted that the commonly used criteria are not adequate. Second, the mixing and transport characterization of the plant vessels was discussed. It is essential for scale-up that the mixing and mass transfer performance of the plant vessels be known, preferably quantitatively but at worst in relative terms.

Finally, the practice of scale-up was considered, and it was shown how the blend of chemistry and transport effects obtained at the small scale can be used to determine the operation of a plant-scale vessel. The act of scale-up itself is vitally dependent on adequate understanding of the process fundamentals and the plant equipment.

A relatively simple and largely empirical approach has been described in the chapter. This is done with acknowledgement that more in-depth studies are generally not favored or likely to be adopted in the fine and intermediate scale chemicals manufacture where multiproduct plant dominates. Where possible, a more thorough and fundamentally based approach is recommended. The ''short-cut'' approach described herein may not always be adequate, and indeed a more thorough approach is then required. Bad scale-up is not bad luck; it is bad engineering practice. Avoiding the major pitfalls of scale-up is dependent on achieving adequate process and equipment understanding to recognize how rigorous an approach is required according to the complexity of the given reaction process.

References

1. Paul, E.L. (1988) Design of chemical reaction systems for specialty organic chemicals. *Chem. Eng. Sci.*, **43**, 1773–1782.

2. Patterson, G.K., Paul, E.L., Kresta, S.M., and Etchells, A.W. (2004) Mixing and chemical reaction, in *Handbook of Industrial Mixing: Science and*

Practice, Chapter 13 (eds E.L. Paul, V.A. Atiemo-Obeng, and S.M. Kresta), John Wiley & Sons, Inc., 755–861.

3. Paul, E.L., Midler, M., and Sun, Y. (2004) Mixing in the fine chemicals and pharmaceutical industries, in *Handbook of Industrial Mixing: Science and Practice*, Chapter 17 (eds E.L. Paul, V.A. Atiemo-Obeng, and S.M. Kresta), John Wiley & Sons, Inc., 1027–1064.

4. Patterson, G. (2005) Modeling and scale-up of mixing- and temperature-sensitive chemical reactions. *Ind. Eng. Chem. Res.*, **44**, 5325–5341.

5. Reynolds, O. (1883) An experimental investigation of the circumstances which determine whether the motion of water in parallel channels shall be direct or sinuous and of the law of resistance in parallel channels, Philos. *Trans. Royal Soc.*, **174**, 935.

6. Szalai, E.S. (2004) Laminar mixing: a dynamical systems approach, in *Handbook of Industrial Mixing: Science and Practice* (eds E.L. Paul, V.A. Atiemo-Obeng, and S.M. Kresta), John Wiley & Sons, Inc., 89–117.

7. Edwards, I., Axon, A.A., Barigou, M., and Stitt, E.H. (2009) The combined use of PEPT and ERT in the study of aluminium hydroxide precipitation. *I.E.C.Res.*, **48**, 1019–1028.

8. Assirelli, M., Bujalski, W., Eaglesham, A., and Nienow, A.W. (2002) Study of micromixing in a stirred tank using a Rushton turbine – comparison of feed positions and other mixing devices. *Trans. IChemE Part A: Chem. Eng. Res. Des.*, **80** (A8), 855–863.

9. Khan, F.R. (2005) Investigation of turbulent flows and instabilities in a stirred vessel using particle image velocimetry, PhD thesis, Loughborough University.

10. Gabriele, A., Nienow, A.W., and Simmons, M.J.H. (2009) Use of angle resolved PIV to estimate energy dissipation rates for up- and down-pumping pitched blade agitators in a stirred tank. *Chem. Eng. Sci.*, **64** (1), 126–143.

11. Wu, H. and Patterson, G.K. (1989) Laser-Doppler measurements of turbulent flow parameters in a stirred mixer. *Chem. Eng. Sci.*, **44** (10), 2207–2221.

12. Rielly, C.D., Habib, M., and Sherlock, J.P. (2007) Flow and mixing characteristics of a retreat curve impeller in a conical-based vessel. *Trans. IChemE Part A: Chem. Eng. Res. Des.*, **85** (A7), 953–962.

13. Hemrajani, R.R. and Tatterson, G.B. (2004) Mechanically stirred vessels, in *Handbook of Industrial Mixing: Science and Practice* (eds E.L. Paul, V.A. Atiemo-Obeng, and S.M. Kresta), John Wiley & Sons, Inc., p. 361.

14. Images from Post Mixing Website, (c) Pete Csiszar *http://www.postmixing.com/mixing%20forum/impellers/impellers.htm.* (accessed 4 February 2011).

15. Nienow, A.W. (1997) On impeller circulation and mixing effectiveness in the turbulent flow regime. *Chem. Eng. Sci.*, **52** (15), 25557–22565.

16. Ruszkowski, S. (1994) A rational method for measuring blending performance and comparison of different impeller types. In Proceedings of 8th European Mixing Conference, Institution of Chemical Engineers, Rugby, UK, pp. 283–291.

17. Chilton, T.H., Drew, F.B., and Jebens, R.H. (1944) Heat transfer coefficients in agitated vessels. *Ind. Eng. Chem.*, **36**, 510–516.

18. Penney, W.R. and Atiemo-Obeng, V.A. (2004) Heat transfer, in *Handbook of Industrial Mixing: Science and Practice* (eds E.L. Paul, V.A. Atiemo-Obeng, and S.M. Kresta), John Wiley & Sons, Inc., 869–885.

19. Atiemo-Obeng, V.A., Penney, W.R., and Armenante, P. (2004) in *Handbook of Industrial Mixing: Science and Practice* (eds E.L. Paul, V.A. Atiemo-Obeng, and S.M. Kresta), John Wiley & Sons, Inc., pp. 543–584.

20. Zwietering, T.N. (1958) Suspending of solid particles in liquid by agitators. *Chem. Eng. Sci.*, **8** (3-4), 244–253.

21. Mak, A.T.C. (1992) Solid liquid mixing in a mechanically agitated vessel, PhD thesis, University College London.

22. Custers, J.P.A., Hersmis, M.C., Meuldijk, J., Vekemans, A.J.M., and Hushof, L.A. (2002) 3,4,5-Tridodeccylooxybenzoic acid: combining reaction engineering and

chemistry in the development of an attractive tool to assist scale-up of solid-liquid reactions. *Org. Proc. Res. Dev.*, **6** (5), 645–651.

23. Maude, A.D. (1958) Cited in *Chemical Engineers' Handbook*, (eds R.H. Perry and D. Green), McGraw-Hill, New York, 1984.

24. Nienow, A.W. (1997) The mixer as a reactor, liquid-solid systems, in *Mixing in the Process Industries*, 2nd edn (eds N. Harnby, M.P. Edwards, and A.W. Nienow), Butterworth-Heinemann, 401–405.

25. Chapman, C.M., Nienow, A.W., Middleton, J.C., and Cooke, M. (1983) Particle-gas-liquid mixing in stirred vessels. *Chem. Eng. Res. Des. (Trans. IChemE Part A*, **61** (71-82), 167–182.

26. Nienow, A.W., Wisdom, D.J., and Middleton, J.C. (1978) Proceedings of 2nd European Conference on Mixing, April 1977, BHRA Fluid Engineering, Cranfield, pp. F1-1–F1-16 and X54.

27. Middleton, J.C. (1997) Gas-liquid dispersion and mixing, in *Mixing in the Process Industries*, 2nd edn (eds N. Harnby, M.P. Edwards, and A.W. Nienow), Butterworth-Heinemann, 346–348.

28. Penney, R.W. (1971) Scale-up for mixing operations. *Chem. Eng.*, 44.

29. Peleka, F., Mavros, P., and Zamboulis, D. (2004) Adsorbent morphology: influence of transient mixing conditions and feed location on semi-batch goethite precipitation. ISMIP 5, Seville.

30. Edwards, I., Axon, S.A., Barigou, M., and Stitt, E.H. (2009) The combined use of PEPT and ERT in the study of aluminium hydroxide precipitation. *Ind. Eng. Chem. Res.*, **48**, 1019–1028.

31. Bałdyga, J. and Bourne, J.R. (1999) *Turbulent Mixing and Chemical Reactions*, John Wiley & Sons, Inc.

32. Bourne, J.R. (2003) Mixing and selectivity in chemical reactions. *Org. Proc. Res. Dev.*, **7**, 471–508.

33. Crawford, P., Burch, R., Hardacre, C., Hindle, K.T., Hu, P., Kalirai, B., and Rooney, D.W. (2007) Understanding the dehydrogenation mechanism of tetrahydrocarbazole over palladium using a combined experimental and density functional theory approach. *J. Phys. Chem. C*, **111** (17), 6434–6439.

34. Hindle, K.T., Burch, R., Crawford, P., Hardacre, C., Hu, P., Kalirai, B., and Rooney, D.W. (2007) Dramatic liquid-phase dehydrogenation rate enhancements using gas phase hydrogen acceptors. *J. Cat.*, **251** (2), 338–344.

35. Atherton, J., Houson, I., and Talford, M. (2011) Understanding the reaction, Wiley-VCH Verlag GmbH, Weinheim, p. 87.

36. Berger, R.J., Stitt, E.H., Marin, G.B., Kapteijn, F., and Moulijn, J.A. (2001) Chemical reaction kinetics in practice. *Cat. Tech*, **5** (1), 30–60.

37. Nova, E., Rooney, D.W., Sinha, N., Neurock, M., and Stitt, E.H. Hydrogenation of propionic acid on a promoted ruthenium/carbon catalyst, In preparation.

38. Hu, B., Nienow, A.W., Stitt, E.H., and Pacek, A.W. (2007) Bubble sizes in agitated water-hydrophilic organic solvents for heterogeneous catalytic reactions. *Ind. Eng. Chem. Res.*, **46** (13), 4451–4458.

39. Hu, B., Fishwick, R.P., Nienow, A.W., Pacek, A.W., Winterbottom, J.M., Wood, J., and Stitt, E.H. (2007) Simultaneous measurement of in-situ bubble size and reaction rates with a heterogeneous catalytic hydrogenation reaction. *Chem. Eng. Sci.*, **62**, 5392.

40. Hindle, K., Rooney, D.W., Akpa, B.S., Weber, D., McGregor, J., Gladden, L.F., Sinha, N., Neurock, M., and Stitt, E.H. Influence of Propan-2-ol/Water mixed solvent on the hydrogenation of Butan-2-one over a Ru/SiO_2 catalyst: an experimental and theoretical study, *J. Phys. Chem.*, to be submitted.

41. Barton, J.A. and Nolan, P.F. (1991) in *Safety of Chemical Batch Reactors and Storage Tanks* (eds A. Beruzzi and J.M. Zaldivar), Kluwer Academic Publishers, pp. 1–17.

42. Crezee, E., Hoffer, B.W., Berger, R.J., Makkee, M., Kapteijn, F., and Moulijn, J.A. (2003) Three-phase hydrogenation of D-glucose over a carbon supported ruthenium catalyst-mass tranfer and

kinetics. *App. Catal. A-Gen.*, **251** (1), 1–17.

43. Mounzer, H.N., Wood, J., and Stitt, E.H. (2010) Heterogeneous oxidation of 2-octanol on 5wt%Pt – 1wt%Bi / Carbon catalyst. *Chem. Eng. Sci.*, **65**, 179–185.

44. Ashe-Morris Constant flux reactor, *http://www.theengineer.co.uk/channels/process-engineering/constant-flux-reactor/268494.article* (accessed 1 March 2011).

45. Brown, D.A.R., Jones, P.N., and Middleton, J.C. (2004) Experimental methods, in *Handbook of Industrial Mixing: Science and Practice*, Chapter 4 (eds E.L. Paul, V.A. Atiemo-Obeng, and S.M. Kresta), John Wiley & Sons, Inc..

46. Mann, R., Dickin, F.J., Wang, M., Dyakowski, T., Williams, R.A., Edwards, R.B., Forrest, A.E., and Holden, P.J. (1997) Application of electrical resistance tomography to interrogate mixing processes at plant scale. *Chem. Eng. Sci.*, **52**, 2087–2097.

47. Kagoshima, M. and Mann, R. (2005) Interactions of precipitation and fluid mixing with model validation by electrical tomography. *Trans. IChemE Part A: Chem. Eng. Res. Des.*, **83**, 806–810.

48. Bolton, G. (2006) A review of linear electrical; electrical tomography probes for monitoring the behaviour of multiphase mixing processes. Paper A17, 12th European Conference on Mixing, Bologna, June.

49. Simmons, M.J.H., Edwards, I., Hall, J.F., Fan, X., Parker, D.J., and Stitt, E.H. (2009) Techniques for visualisation of cavern boundaries in opaque industrial mixing systems. *AIChE J.*, **55**, 2765–2772.

50. Coulson, J.M. and Richardson, J.F. (1999) *Coulson and Richardson*, Fluid Flow, Heat Transfer and Mass Transfer, Vol. **1**, 6th edn, Butterworth Heinemann, Oxford.

51. Meyberg, M. and Roessler, F. (2005) In Situ measurement of steady-state hydrogen concentrations during a hydrogenation reaction in a gas-inducing stirred slurry reactor. *Ind. Eng. Chem. Res.*, **44**, 9705–9711.

52. Teramoto, M., Tai, S., Nishii, K., and Teranishi, H. (1974) Effects of pressure on liquid phase mass transfer coefficients. *Chem. Eng. J.*, **8**, 223–226.

53. Dietrich, E., Mathieu, C., Delmas, H., and Jenck, J. (1992) Raney-Nickel catalyzed hydrogenations: gas-liquid mass transfer in gas-induced stirred slurry reactors. *Chem. Eng. Sci.*, **13/14**, 3597–3604.

54. Crezee, E., Hoffer, B.W., Berger, R.J., Makkee, M., Kapteijn, F., and Moulijn, J.A. (2003) Three-phase hydrogenation of D-glucose over a carbon supported ruthenium catalyst-mass transfer and kinetics. *Appl. Catal. A: Gen.*, **251**, 1–17.

55. Levenspiel, O. (1998) *Chemical Reaction Engineering*, John Wiley & Sons, Inc.

56. Cabassud, M. and Gourdon, C. (2010) in *Novel Concepts in Catalysis and Chemical Reactors* (eds A. Cybulski, J.A. Moulijn, and A Stankiewicz), Wiley-VCH Verlag GmbH, Weinheim, pp. 261–287.

57. Stitt, E.H. and Rooney, D.W. (2010) in *Novel Concepts in Catalysis and Chemical Reactors* (eds A. Cybulski, J.A. Moulijn, and A. Stankiewicz), Wiley-VCH Verlag GmbH, Weinheim, pp. 309–330.

7
Process Understanding – Crystallization

Leroy Cronin, Philip J. Kitson, and Chick C. Wilson

7.1
Introduction

The emergence of molecular-based compounds being developed for applications as advanced materials and pharmaceuticals necessitates an ever more advanced approach to the control and characterization of solid-state properties; one key route to accomplish this goal is by understanding the crystallization of these compounds. Crystalline and solid-form manufacturing is a universal application – more than 80% of pharmaceutical products and 60% of fine and specialty chemical products are made in crystalline form, for example, pharmaceuticals, agrochemicals, fats and oils, foods and confectionary, pigments and inks, consumer products, materials, and metals. This has led to a dramatically increased interest and prominence in crystallization processes, in the discovery of new forms, control of polymorphism, design of morphology for particular modes of delivery, and for formulation and scale-up, in optimizing the crystallization process in batch or flow modes. Further, there has been a general realization that the crystallization process is core to the production of solid-state materials, at a fundamental level, and is set to make an increasing contribution to coatings, especially epitaxially grown layers. Among the issues that require to be addressed in understanding the crystallization process include physical form discovery and screening, particle technology analytical methods for characterizing the synthesized products, backed up by modeling and theory of both microscopic and macroscopic processes.

In more detail, and to set the scope for the discussions introduced below, each of these raise a range of technical and scientific challenges that must be addressed to inform a full understanding of particular crystallization processes geared at the production of selected solid material forms.

- Physical form discovery and screening: Crystallization screening for physical form discovery; crystal structure determination; crystal structure comparison, analysis of three-dimensional structures; polymorphism; effects of solvation; transformations and phase transitions; and physicochemical characterization.

- Particle technology: Nucleation, crystal growth studies, agglomeration/attrition, morphology control, particle characterization, bulk characterization, crystallizer design, additive synthesis, and processing and scale-up.
- Analytical methods: X-ray diffraction (XRD), single crystal and powder; spectroscopy, Fourier transform infrared (FTIR), Raman, near infrared (NIR), ultra violet (UV), solid-state nuclear magnetic resonance (NMR); thermal analysis/calorimetry; focused beam reflectance microscopy (FBRM); process analytical technologies; chemometrics; optical and scanning electron microscopy (SEM); and surface analysis (atomic force microscopy (AFM)).
- Modeling and theory: Molecular structure, quantum chemistry; intermolecular interactions; crystal structure prediction; morphology prediction; solvation; computational fluid dynamics; molecular dynamics simulations; and chemometrics/design of experiments.

As an example of the importance of control of molecular association in crystallization processes, the following physical and chemical properties can differ enormously among crystal forms and cocrystals of the same material, dramatically affecting their processing, formulation, and delivery modes [1]:

- Physical and chemical properties: Density and refractive index, thermal and electrical conductivity, hygroscopicity, free energy and chemical potential, melting point, heat capacity, vapor pressure, solubility, thermal stability, and chemical and photochemical reactivity.
- Kinetic properties: Rate of dissolution, kinetics of solid-state reactions, and stability.
- Surface properties: Surface free energy, crystal habit, surface area, and particle size distribution.
- Mechanical properties: Hardness, compression, and thermal expansion.

The interplay of some of these effects, and some of the techniques that can be used in their study, are given in Table 7.1; many of these are discussed further in this chapter, which reviews some of the underpinning science and techniques relevant to the crystallization process.

7.1.1
Crystal Definition and Structure – Crystal Defects and Basics of Crystal Growth

Macroscopically, a large single crystal is among the great beauties of nature, with aesthetic properties that pervade many cultures both historically and currently. Microscopically, the crystal retains much of this simplistic beauty. The regularity is retained, although there are microscopic defects even in the most ''perfect'' of gem-like crystals, as is the simplicity, manifest in the repeated pattern of the units forming the crystal. The regularity of a crystal gives a clear indication of the underlying construction, which is based on a periodic array of lattice points populated by a structural unit that is repeated at the lattice points. The lattice in any crystal is defined by the unit cell, principally by its external dimensions (defining the crystal system) and also by aspects of its internal constitution (defining the lattice type).

Table 7.1 Interplay of crystallization parameters, physical effects, and the techniques used in their study.

Primary physical parameters controllable during crystallization	Secondary effect	Methods of analysis (see Section 7.5)
Particle size	Filterability, dissolution rate/profile, flow of bulk powder	FBRM, DLS
Particle size distribution	Flow of bulk powder, filterability, consistency of processing	FBRM, DLS
Crystal habit	Filterability, mechanical strength, ease of downstream physical processing, flow of bulk powder	XRD, optical microscopy, electron microscopy
Polymorphic form and phase changes	Physical properties, melting point, solubility, availability of active ingredient	XRD, DSC
Solvate formation, desolvation of solid forms	Stability, bioavailability, processing characteristics	DSC/TGA, XRD
Surface effects – roughness, charge characteristics	Flow of bulk powder, ease of formation of agglomerates	–
Cocrystallization	Bioavailability, dissolution rate/profile	DSC/TGA, XRD, chemical analysis

This internal constitution introduces the concept of internal symmetry within the unit cell. There are seven crystal systems, triclinic, monoclinic, orthorhombic, tetragonal, cubic, trigonal, and hexagonal, which are defined on the basis of the external geometry of the unit cell [2], while the introduction of lattice centrings (lattice points not at the corners of the unit cell) leads to the presence of 14 so-called Bravais lattices. Introduction of internal symmetry within the unit cell yields a total of 230 space groups in 3D, with many thousands of space groups possible for >3D systems.

This internal regularity leads to the macroscopic shape of crystals, defined by the crystal faces, whose formation is governed by the minimization of surface energies. In general, these faces tend to be defined by low values of Miller indices (defining the lattice planes forming the faces). Microscopically, this growth mechanism results from the lowest attachment energy of molecules to the growing surface, defined as the interaction energy per molecule between a depositing slice and the crystal face. In this model, crystal faces with the lowest attachment energies tend to dominate the resulting morphology. One simple model directly relates the growth rate of a face to the attachment energy, $R \propto E_{att}$ [3]. The macroscopic shape of a crystal is known as its *morphology*, and is dictated by the rate of crystal growth on each of the binding faces. The morphology is a critical factor in the physical properties of crystalline materials used for applications, affecting important factors such as compaction, flow characteristics, physical properties, anisotropies, and so on.

However, not all crystal faces end up being regular in real crystals, instead containing terraces and steps and, importantly, other forms of defects. The

Figure 7.1 Block crystal growth, terraces, and kinks.

attachment discussed above tends to occur at sites where the attachment energy can be minimized. At these kink sites, there are only half the number of neighbors present at the attachment site than would be present in the crystal bulk (Figure 7.1). In addition to minimizing the attachment energy, the kinks are retained following the attachment, this being available for further growth deposition processes. For crystals growing from solution, the density of kinks on the surface tends to be rather low, comprising around 0.1% of all molecular sites on the surface. It is clear from this that the low number of kinks present is likely to be a major factor in controlling the rate of growth of the crystal from solution.

The nature of the growing face significantly affects the mechanism of crystal growth. Broadly speaking, crystal faces can be classified into two categories: atomically rough faces where the growth is diffusion controlled and continuous, with the growth rate depending linearly upon the supersaturation; and flat faces, in which the growth is dependent on the rate of formation of critical nuclei to overcome the energy barriers in the construction of the layer. The need for nuclei formation in the latter case can often be mitigated by the presence of defects, which effectively act as nucleation centers. Crystal morphology can also be tuned by introduction of additives, which can be incorporated in the growth process, and have two distinct potential effects. Once present, some tailor-made additives inhibit the growth of unfavored faces, while others can act as effective nucleation centers and encourage enhanced face growth [4] (Figure 7.2).

7.1.2
Thermodynamics of Crystallization

The equilibrium form of a crystal can be considered to be governed by the Gibbs–Wulff Law, $\sum_n \gamma_n dA_n = 0$, where with reference to the nth face of the crystal, A is the area and γ the surface tension. This takes into account only thermodynamics; in practice, however, kinetic effects can often dominate, particularly as the crystal becomes larger. The interplay of thermodynamic and kinetic effects is one of the real challenges in understanding crystal growth and crystallization

Figure 7.2 Tailor-made additives – inhibition or enhancement of growth on selected faces depending on the nature of the additive.

processes, with factors such as impurity level, mixing regime, vessel design, and cooling profile often having a major impact on the nature of the crystals produced.

In general thermodynamic terms, the crystallization process is a multiphase multicomponent system. For such a general system of C components and P phases, the amount of each component in a single phase, at a given value of temperature (T) and pressure (p), is given by its molar quantity n_i. The Gibbs energy of the phase ϕ, G^ϕ, can then be expressed as a function of temperature, pressure, and these molar quantities of the various components:

$$G^\phi \left(T, p, n_1, n_2 \ldots, n_C \right)$$

with differential form

$$dG^\phi = \left(\frac{\partial G^\phi}{\partial T} \right)_{p,n_i} dT + \left(\frac{\partial G^\phi}{\partial p} \right)_{T,n_i} dp + \left(\frac{\partial G^\phi}{\partial n_1} \right)_{T,p,n_{i\neq1}} dn_1 + \ldots$$

$$= -S^\phi dT + V^\phi dp + \mu_1^\phi dn_1 + \ldots$$

Figure 7.3 Simple phase diagram for a one-component system, showing phase boundaries. The chemical potentials of the relevant phases are equal at these phase boundaries.

where there is a term for each of the *C* components. The terms μ_i are called the *chemical potentials*, representing the contribution of each component to the Gibbs function, and thermodynamic theory states that the system will attempt to minimize the value of the Gibbs free energy – these chemical potentials thus become the driving force for thermodynamic processes such as, in this case, crystallization. At equilibrium, the chemical potentials of all phases present will be equal.

A one component system is fairly easy to interpret conceptually through use of a simple schematic phase diagram, expressed in terms of thermodynamic variables or chemical potential (Figure 7.3).

The crystal (solid) phase is stable at high *p* and low *T*, and at the phase transition point between liquid and solid, the chemical potentials of these two phases will be equal. μ decreases at different rates for solid, liquid, and gas $\left(\left(\frac{\partial \mu}{\partial T} \right)_p = -S \right)$, and the phase transition (in this case crystallization) occurs when $\mu_S = \mu_L$ (and then when $\mu_L = \mu_G$), that is, the chemical potentials are equal (points at which $\Delta G = 0$). From this general framework, the thermodynamics of any specific crystallization process can be developed.

7.1.3
Kinetics of Crystal Growth, Nucleation

We have seen from the above simple consideration of phase diagrams that crystallization is a condensation phase transition process involving the creation of a solid crystalline phase from a parent phase. At the microscopic level, the formation of a new solid phase in this way requires an input of work to create an interfacial region (formation of nuclei). There is thus an energy barrier to the formation of the solid phase, which is related to the surface area of the newly forming phase. Since the nuclei can be very small in the initial stages of crystallization, this surface energy can be correspondingly large. Introduction

of such an energy barrier, or activation energy for the formation of nucleation sites, naturally leads to a chemical kinetic interpretation of the process. In an analogous way to the simple collision theory of gas kinetics, statistical fluctuations in a liquid or solution phase may lead to a locally sufficient energy to overcome the activation energy barrier. Kinetic effects are thus largely bound up with the nucleation process.

Nucleation processes are of two main types: homogeneous nucleation occurs within the single component phase in the system, while heterogeneous nucleation will be induced on a distinct substrate within the system, for example, a surface such as the wall of the crystallization vessel, a mechanical fluctuation, or at the micro and nano positions of particle impurities that are always present in the solution. Following nucleation, the process of crystal growth involves the expansion of these nucleating centers that have achieved a critical size to the macroscopic crystal, once again with energy input required to overcome energy carriers to association of molecules on the growing surface. Thus, the rate of nucleation can be expressed in terms of a two-step process of the formation of a concentration of critical nuclei from the equilibrium system and the subsequent rate at which further molecules accrue upon these nuclei, leading to the growth of the bulk phase.

A simple visual way of depicting the kinetics of crystal growth is presented in Figure 7.4, with particular reference to the production of the final solid phase in a polymorphic system.

In the crystallization process to one of two possible polymorphic crystal structures (I and II), the free energy of the starting liquid phase is given by G_0, and the crystallization can result in the formation of one of two crystalline products (I or II) in which I is more stable ($G_{II} > G_I$). The kinetics of crystallization in this case is complicated by the often complex structure of the activated site. The crucial

Figure 7.4 Simple schematic representing the kinetic effect in favoring a less stable (metastable) polymorph (II) over a more stable form by the existence of lower activation energy for the production of the less stable form (a). Such situations can lead to coexistence of different polymorphs (b), the composition of the final mixture again being heavily influenced by the kinetics.

factors that result in polymorphism are crystal nucleation and growth, which are under kinetic control as indicated above. Figure 7.4 shows that the final product may result from the less stable but faster growing nuclei, forming a metastable phase, with the transformation to the thermodynamically stable phase forbidden by a considerably higher energy barrier. There is also a substantial energy barrier for possible transformation from the metastable phase to the thermodynamically stable phase.

7.1.3.1 Metastable States

We have seen above that the factors controlling crystallization can often involve a subtle balance between thermodynamic and kinetic factors, and can be dramatically affected by the physical conditions under which the crystallization process occurs. In addition, the occurrence of metastable states can have a significant effect on the outcomes of a crystallization experiment. Metastable thermodynamic states are frequently encountered in a wide range of systems including pharmaceuticals, and can have dramatic effects on the crystallization process. They can, for example, affect the creation of supersaturation conditions, alter the solid-state form produced, and introduce issues with the control of solid-phase conversions during isolation, manufacturing, storage, and dissolution. Since the process of crystallization offers a way of reducing the free energy of metastable thermodynamic states, the occurrence and stability of these states are determined by the crystallization mechanism and kinetics; hence, it is an important factor in the control of many crystallization processes [5].

7.1.4
Nucleation and Crystal Growth

Nucleation has been referred to above several times, as the kinetically controlled first step in the crystallization process. Nucleation remains one of the most challenging aspects of crystal growth to understand, characterize, and control, with many solutions to this based on empirical investigations tuning experimental variables to achieve the desired results reproducibly. The basic process in primary nucleation is the formation of nuclei of sufficient size to be able to sustain growth from solution, thus becoming critical nuclei. We have seen above that the nucleation can be homogeneous or heterogeneous. The nature of homogeneous nucleation is such that the critical nuclei will represent, at the microscopic level, the equilibrium form of the crystal, while this is not true for heterogeneously generated nuclei.

7.1.4.1 Supersaturation and Metastable Zone Width

There are three well-defined regions associated with solution crystallization. The first is a stable, unsaturated zone where crystallization is impossible. The second is the metastable zone between the solubility and supersolubility curves (Figure 7.5), where crystallization is improbable, although growth would occur if seeds were to be planted in a metastable solution to act as heterogeneous nucleation sites. The width of the metastable zone indicates the stability of the solution; the larger the

Figure 7.5 The metastable zone between the solubility and supersolubility curves, showing the metastable zone with (MSZW).

zone width, the more stable the solution. This is the preferred region in which to carry out controlled crystallization processes; that is, by adjustment of system parameters, as this helps to prevent any unwanted nucleation from occurring. Crystallization within the metastable zone is not common due to the stability of the solution; beyond this metastable zone, the system is said to be labile, and this is where spontaneous nucleation can occur.

Tuning the cooling rate of a supersaturated solution until the first indications of crystallization can be seen can control the onset of nucleation. This is referred to as the *nucleation point*. A solute is maintained in solution until a sufficiently high level of supersaturation has been developed; this in turn encourages spontaneous nucleation to occur. It is thus important to characterize the metastable zone width (MSZW) under a specific set of operating conditions [6], which relate closely to the conditions of the final-scale crystallization. The polythermal technique involves cooling a saturated solution at a fixed rate until nucleation occurs. This is repeated several times at a variety of cooling rates until a reliable MSZW can be determined. The MSZW can be considered to be characteristic of each unique crystallization system. The *induction period of nucleation* is defined as the time that elapses between the instant when the supersaturated state is generated and the point of time at which solid phase particles become detectable. This includes the time required for the generation of a critical nucleus in supersaturation and the growth to a detectable range, which can be as low as $1\,\mu m$ in, for example, the FBRM method (see below).

Understanding the MSZW is of fundamental importance to be able to control crystal growth and is thus widely studied. In-depth solubility studies and supersolubility studies of a single compound are needed and temperature control is crucial. Reliable dissolution profiles can be determined and these are fundamental, in particular in the pharmaceutical industry, due to the increase in discovery of new

polymorphic forms and the corresponding changes in their key physical properties such as solubility. For this reason it is vitally important to be able to carry out these experiments in a clean and controlled environment, as even the smallest contaminant such as a speck of dust can initiate nucleation.

In solution crystallization, the strongest reduction of supersaturation takes place during the rate-determining step, which is the slowest step in the crystallization reaction mechanism. The potential processes involved comprise transport to the growth site, overcoming surface energy barriers, and removal of heat of crystallization from the system. For solution crystallization, this tends to be dominated by the first two of these processes, while for the case of melt crystallization, the heat transport tends to be the rate-determining step [7].

Detailed expressions for the thermodynamics of nucleation have been developed. For example, Kachiev *et al.* consider the thermodynamics of the nucleation process and present expressions for the supersaturation, the nucleation work, and the size of the nucleus in homogeneous or heterogeneous nucleation [8]. This leads to the design of experimental methods for the determination of the size of the nuclei. The mechanism and kinetics of nucleation are also considered in this work, with expressions given for the supersaturation dependence of the monomer attachment frequency and the rate of homogeneous or heterogeneous nucleation, together with other kinetic aspects of the process. Methods are also available for estimating the width of the metastable zone from solubility data.

7.2
Crystallization Processes

7.2.1
Crystallization from the Melt

The principles of nucleation and growth control in melt crystallization were initially formulated by Verneuil in 1902, in the work on the growth of synthetic rubies, and are adapted in most later growth methods from melt. This original work established flame-fusion processes, allowing very high temperatures of more than 2000 °C to be reached through designing apparatus that only required a small amount of fine-grained material to be melted using an oxyhydrogen burner. Once the raw material is melted, a seed crystal is used to control the crystal growth [9]. Melt crystallization is still widely used for the production of large single crystals, particularly of metallic or inorganic crystals. Widely used adaptations of the methods introduced by Verneuil include the Bridgeman and Czochralski methods discussed in the following two sections.

7.2.1.1 **The Bridgman Method**
The Bridgman method, and its closely related variant the Bridgman–Stockbarger method, is an important melt-crystallization technique. As in the original experiments of Verneiul, the polycrystalline sample material is heated above its melting

point, but in this case it is held inside a container and a cooling regime is instigated from one end of this container, where a seed crystal is present. By these means, crystal growth takes place along the container in the form of good quality single crystals that tend to have low concentrations of defects.

7.2.1.2 The Czochralski Method

The Czochralski process is an alternative method for producing single crystals from the melt, which involves pulling a crystal from the melt [10]. Frequently used to produce a range of often large single crystals (at the centimeter level or greater), the Czochralski method is often used to produce single crystals of semiconductors, in cases where a low defect density is not a prerequisite. In addition to large single crystals (that can be up to 2 m long), the technique offers the opportunity for defect engineering, and thereby control of the temperature-dependent properties of crystal defects [11].

7.2.1.3 Crystallization of Organic Materials from the Melt

Although used less frequently, the nucleation and growth of single component and binary organic materials are also susceptible to study using melt-crystallization techniques [12], including growth of crystalline polymer materials [13]. Although these can be used to produce large single crystals as for metallic and inorganic materials, melt conditions have also recently been increasingly shown to be a potential source of new crystal forms, and are now beginning to be used in polymorph screening and discovery. Techniques such as Kofler hot-stage microscopy can allow the growth of new phases to be followed, thereby allowing nucleation and crystal growth kinetics to be followed in these materials.

7.2.2
Crystallization from Solution

Crystallization from solution is the most common technique for producing a wide range of materials, in particular, organic materials such as pharmaceuticals that have driven many investigations into improving crystallization processes in production and manufacturing. The technique is reviewed in several recent publications [14], including consideration of the relevance to large-scale plant crystallization [15].

7.2.2.1 Single Solvent Crystallization

7.2.2.1.1 Temperature-Controlled Crystallization　There are many ways to attempt to grow single crystals, from the traditional slow evaporation to more modern apparatus including multiple parallel automatic and robotic crystallizers, and medium-throughput semimanual devices. The characteristics of these systems include the ability to screen crystallization from multiple solvent options, automatic dispensing, and in some cases solid-form characterization, and programmable temperature control. One typical semiautomatic medium throughput instrument

Figure 7.6 A typical parallel laboratory crystallizer, allowing flexible, programmable temperature control for small-scale crystallizations.

is illustrated (Figure 7.6); this device allows 12 separate rows of independently temperature-controlled conditions for crystallization to be trialed simultaneously: with 4 individual vessels in each row this allows for up to 48 different experiments in 12 separate temperature conditions to be carried out simultaneously. These are small-scale vessels, typically containing milligram quantities of material, and can give results in a far shorter period than larger-volume slow-evaporation techniques. The instrument allows ramping and cooling profiles to be set up in each row and can lead to positive identification of the conditions that may lead to the best crystal growth.

7.2.2.1.2 Evaporation-Controlled Crystallization This is amongst the simplest of crystallization methods, and forms the basis of many high-throughput and high-volume processes. On the laboratory scale, the experiment starts with a small amount of the sample being placed in a sample vial (Figure 7.7). Sufficient solvent is added to dissolve the contents of the vial. The vessel is covered, or partially covered, allowing the sample to evaporate slowly and, hopefully, allowing crystals to grow. Evaporation-controlled crystallization will often be carried out under closely

Figure 7.7   Schematic illustrating the principle of simple, slow-evaporation crystallization.

regulated temperature conditions, though for many growth processes accurate temperature control is not required.

7.2.2.2   **Multiple Solvent Crystallization**

Multiple solvent crystallization techniques can take two main forms: one in which the sample is dissolved in two mutually miscible liquids in both of which it has significant solubility, which are then allowed to evaporate as described above; and the other in which a solvent diffusion process is adopted, involving solvents in which the sample has significantly different solubilities. Solvent diffusion is usually employed if conventional evaporation is unsuccessful. The technique (Figure 7.8) involves dissolving the sample in a solvent in which it is readily soluble (good solvent). Another solvent is then added to this sample vessel, the solvent to be added usually being a solvent in which the sample will not be readily soluble (poor solvent). The two solvents used must be comiscible. The poor solvent is added slowly to ensure that there are two layers present. These two layers will then mix slowly and the crystals will thus be encouraged to grow at the interface. If necessary, cooling of the tube can be done to slow the diffusion rate and to reduce solubility.

Figure 7.8   Schematic illustrating the principle of solvent diffusion crystallization. The blue solvent is the good solvent, the yellow solvent is the poor solvent, and the green area is the interface.

Figure 7.9 Schematic illustrating the principle of vapor diffusion crystallization. Blue represents the good solvent and yellow the poor solvent.

7.2.2.3 **Vapor Diffusion**

Vapor diffusion is a technique closely related to that of solvent diffusion, and is also known as *isothermal distillation*. The setup is as shown in Figure 7.9, where the good solvent in which the material is dissolved is placed in the crystallization vessel, which is placed in a larger vessel containing the desired poor solvent. The whole system is then covered, and the poor solvent diffuses through the vapor phase into the solution of the target compound in the good solvent, thus reducing the solubility and leading to the precipitation of crystalline material. The main reason for using this technique is that it offers slow rate of diffusion, has high controllability, and is very adaptable.

7.2.3
Crystallization from Vapor

Chemical vapor deposition (CVD) is a further widely used technique for the growth of inorganic or metallic materials, and is often used in the production of, for example, diamond materials [16]. Specific CVD reactors for growth of particular materials have been designed and used. Variants of CVD including pulsed vapor deposition (PVD) have also been developed, and have particular utility in the fabrication of complex-oxide heterostructures, superlattices, and well-controlled interfaces, with deposition methods that approach true layer-by-layer growth [17]. As can be imagined, the understanding of surface chemical processes is vital to the understanding of such crystal growth processes, and approaches such as time-resolved surface XRD can be employed for detailed structural mechanistic studies.

7.3
Batch Crystallization Techniques

In industrial process environments, crystallization would generally be undertaken in a batch crystallizer, allowing for high-volume production. In such environments, the process is usually based on precipitation from a solvent with controllable parameters that allow for phase separation of the deposited solid phases, and, in favorable circumstances, control of important factors such as particle size and morphology. These are vital aspects in producing crystalline materials with appropriate physical properties for the desired application. The control of particle sizes in batch crystallizers is a function of a wide range of factors that include cooling, evaporation, and dilution regimes employed. Modeling of batch crystallizer operation can be undertaken and allows this to be tuned to select appropriate particle sizes [18]. The batch crystallization operation has also been developed into more sophisticated environments for specific applications, for example, in the use of liquid-phase-coupled batch crystallizers enhancing the driving forces for preferential crystallization of enantiomeric forms allowing for enantiomeric separation [19].

The principles of chemical engineering are thus added to the fundamental understanding of crystallization processes in designing and optimizing such batch crystallizers. The rapid expansion of the chemical and other related industries have required increased study of the mechanism and design of the crystallization process, and design of a production plant increasingly takes account of this. The design of industrial crystallizers has reflected the evolution of new crystallization and manufacturing processes. Early designs were based around simple stirred tanks (which have been essentially in existence for hundreds of years) where cooling, evaporation, or pH control were used to induce crystallization, while variants of this design have been implemented for specific applications and have improved product characteristics. Modern approaches to continuous crystallization processes have continued this development, offering various advantages, including substantial improvements in product quality control.

7.3.1
Tank Crystallizers

The simplest cooling crystallizers are tanks provided with a mixer for internal circulation, where temperature decrease is obtained by heat exchange with an intermediate fluid circulating in a jacket. These simple machines are used in batch processes, as in the processing of pharmaceuticals, and are prone to scaling problems. While providing large-volume production capability for crystalline materials, batch processes normally provide a relatively variable quality of product.

There are a series of variants in this approach, which are commonly implemented for bulk crystallization processing. Among these are the following:

• Scraped surface crystallizers. The characteristics of a scraped-surface crystallizer are of a trough within which slow-speed blades rotate, agitating the forming

crystal, and ensuring that any deposits on the walls are returned to the crystallizer body, usually resulting in a wide distribution of crystal sizes. The process of crystallization in a scraped surface system can be complex and considerable investigation is still being carried out to obtain an understanding at the fundamental level of the various factors influencing crystal growth, including fluid flow and heat transfer [20].

- Double-pipe scraped surface crystallizers. These systems are different from the scraped-surface device in having an annular space within which cooling water circulates. This type of device is used in many manufacturing processes including the crystallization of ice cream and plasticizing margarine.
- Circulating-liquid evaporator crystallizer. This device uses carefully designed flow of liquid and vapor phases to induce supersaturation, in a controlled manner, by evaporation processes. The circulating-liquid evaporator crystallizer was one of the earliest batch of crystallizers and is often called the *Oslo crystallizer*, as it was first implemented there.
- Circulating-magma vacuum crystallizer. In this device, a hot suspension or magma of feedstock material is passed through a heater to produce a heated surface liquor that then mixes with the body of the material causing supersaturation, depositing crystals that are then harvested.

7.3.2
Continuous (Flow) Crystallizers

The evolution of crystallization in manufacturing technology for the chemical and other related industries has been rooted in the use of the stirred-tank reactor as the normal approach for manufacture. Although hugely successful, this type of application offers particular challenges around the scale-up of processes, is also expensive, results in wastage of materials, has a large plant footprint, and offers poor control over quality.

In particular, for crystallization, approaches using stirred-batch vessels or through a series of tanks, requires large vessels and offers poor control over crystal form. The industry has therefore begun to move toward leaner and more sustainable forms of manufacturing, with some success (Figure 7.10). However, the real benefit lies in moving toward a fully continuous process, and there are key scientific and technical developments that are required to enable this shift to happen. Indeed, in the area of nanoparticle crystallization, continuous manufacture using crystallization under flow is a new and possible key area [21].

Continuous Crystallization is one approach that overcomes many of these limitations, offering benefits in terms of sustainability, including substantial footprint and capital cost reduction, lower running costs, speed of scale-up of platform technologies, and controllable quality of the crystals formed. There are a range of continuous crystallizer technologies now being adopted in the industry, including devices such as the continuous oscillating baffled crystallizer (COBC; Figure 7.11).

Figure 7.10 The move from batch to continuous manufacturing principles, through leaner methods, offers advantages in terms of cost, efficiency, and sustainability. Major benefits are anticipated in moving from lean to continuous processes, once the technologies are sufficiently developed.

Figure 7.11 Schematic illustrating the construction of industrial-scale COBC devices.

7.3.2.1 Continuous Oscillatory Baffled Crystallizer

The COBC consists of a continuous cylindrical tube or column with baffles periodically spaced along the inside walls. The solution is oscillated axially by means of a diaphragm, bellows, or pistons at one or both ends of the tube. The sharp edges of the baffles are positioned transverse to the oscillating flow of the liquid. On the upstroke the liquid passing through the baffles forms vortex rings and on the downstroke, these rings are swept into the bulk and unraveled (Figure 7.12).

The generation of these eddies in the cells allows for very uniform mixing throughout the column. The mixing can be controlled so that plug flow conditions can be achieved even at low flow rates, that is, radial mixing is uniform and axial dispersion is at a minimum. This also allows for very precise cooling profiles along the length of the reactor. Being able to exercise such precise control over crystal growth conditions therefore allows for the development of processes capable of producing very consistent products in terms of crystal morphology and quality. The quality and consistency of crystals produced in a COBC is enhanced over batch crystallizers, and the particle size can also be altered through altering the oscillation velocity or changing the baffle set-up, that is, the spacing between the baffles or the size of the baffles.

Net flow

Forward
oscillation

Reverse
oscillation

Figure 7.12 Mixing mechanism in an oscillating baffled cell,
showing the characteristic eddies generated in the reaction
mixture.

Having already highlighted the importance of the relationship between crystal morphology and properties, the benefits of this technology are obvious and it can be used in both cooling crystallization and antisolvent crystallization. Studies of this technology using paracetamol show that the product recrystallized from a COBC is of much better quality than that from conventional methods [22]. Studies of ʟ-glutamic acid have also shown that precise selective polymorph formation can be achieved by altering the solution concentration, and the crystal size can also be varied by altering the cooling rate [23]. These studies emphasize the value of the technology in developing processes for precise morphology control of active pharmaceutical ingredients (API s) and other materials, and in offering significant advantages in terms of process, operation, and cost [24].

7.4
Process Control of Crystallization

Despite the important role that crystallization has in process control and in determining solid-phase outcomes, until recently, understanding and controlling crystallization phenomena has often been neglected in the pharmaceutical industry (and others), only becoming a significant consideration when problems in manufacturing or processing were encountered [5]. Now, however, the realization that understanding and control of this critical phenomenon is vital, has led to a dramatic increase in awareness and the development of advanced techniques for crystallization process monitoring and control.

There has been significant development of *in situ* techniques for monitoring crystallization processes that do not require sampling and give real-time data analysis, such as attenuated total reflectance – ultra violet (ATR-UV), FBRMs and attenuated total reflectance – Fourier transform infrared (ATR-FTIR), and techniques such as dynamic light scattering (DLS) to allow particle sizing during the crystallization process. These techniques are discussed in Section 7.6 below; here we discuss the parameters that are often essential to control during crystallization processing.

7.4.1
Crystal Growth Rate and Morphology Control

Modifying crystal growth processes using additives is a well-established approach, with additives of two forms – the predominant inhibitors and the lesser known promoters of crystal growth. The principles of additives are very well established, with many naturally occurring materials being constructed in this way; for example, the area of biomineralization where inorganic structures are controlled by organic materials. The additive design allows the control of crystal growth rates and the control of morphology by selective inhibition (or enhancement) of growth of selected faces. These techniques can, of course, also often result in interesting hybrid materials that are of less interest in a process environment [25].

The solvent also has a strong influence on the habit of crystalline materials, with an important role played by the surface chemistry; models describing the effect of solvent–surface interactions in enhancing or inhibiting crystal growth have been developed [26–30]. This surface chemistry also has, of course, implications on the effects of ''tailor-made'' additives as discussed above [31]. The interplay between solvent–surface interactions, additives, and their implications for the effect of solvent on crystal growth and morphology have been discussed [32].

7.4.2
Particle and Crystal Size

Crystal growth is based on the principles of precipitation, and control and monitoring of the shape and size of the growing particles is vital in controlling the properties of the resultant product. Consideration of particle morphology and size distributions is important both in laboratory and industrial environments, and in considering the implications for process control, understanding these factors in industrial reactor environments is important [33].

During crystallization processes, it is possible to manipulate the growth of individual microscopic particles to exercise control over the overall dimension to which these are likely to grow under set crystallization conditions, and also the morphology of these, for example, through the introduction of tailor-made additives. As mentioned above, these alter molecular recognition and can reduce the growth rate of a specific face, offering control over of both shape and size of the final crystalline product. Particle sizing is frequently carried out using DLS (see below), while recent developments in on-line methods for characterizing particle shape and size during crystallization process allow the design and optimization of crystal shape control within crystallizers [34]. In a similar vein, the principles of crystal engineering, including multicomponent crystallization, can also be effective in offering routes to the control of crystal habit and particle morphology and ultimately an improved solubility and dissolution rate [35].

In addition to the usual thermodynamic variables discussed above, and together with effects such as rector design (affecting flow conditions, agitation, etc.), ultrasound has also been found to be an effective way of controlling particle and crystal size in ''sono-crystallization'' processes [36].

7.4.3
Crystal Purity

In many cases, the phase compositional purity of a single crystal – and its mechanical perfection – can be important for applications. However, in many other situations, impurities at the microscopic level can be vital to allow the material to function. Introduction of these so-called defects into crystalline materials is an important aspect of materials design that is most often carried out under carefully controlled crystallization conditions. However, the incorporation of impurities, whether by intent or not, can also be discussed in the context of process crystallizers, with, for example, the effect of crystal surface roughness on adsorption of impurities during the crystallization of sucrose in a fluidized bed crystallizer and in a batch crystallizer [37]. In situations such as these, the impurity adsorption is growth rate dependent and is strongly influenced by the crystal surface properties, with high surface roughness correlating with lower impurity adsorption.

7.4.4
Composition Control (Cocrystallization)

Cocrystallization is a method for controlling crystallization that has gained increasing interest recently. The technique offers the potential for control of the physical properties of crystalline products that is of great interest. For example, in the pharmaceutical industry, favorable physical properties of a pharmaceutical cocrystalline material, such as solubility, can be designed to be more favorable for drug products than the pure material. In addition, cocrystallization – multiple component crystallization – can drastically affect both the thermodynamics and kinetics of crystallization processes, and offers a further dimension to the variables to be optimized during the process of obtaining a desired crystallization product, on which systematic investigations have been carried out in a range of studies [38].

These considerations are important, as the increasing prevalence of poorly soluble drugs in development can cause problems with bioavailability, particularly for orally administered drugs. Although a range of methods has been implemented for enhancing the bioavailability of drugs with low aqueous solubility, these do not guarantee success [35]. Techniques such as cocrystallization – an example of an application of the growing field of ''Crystal Engineering'' – offer an alternative route to addressing these issues. A further potential advantage of cocrystallization methods is that derivatization of APIs (or indeed materials in general) through noncovalent interactions is considered a greener approach than making and breaking covalent bonds as it minimizes the formation of by-products. The potential number of true cocrystals for a given API also far exceeds that of salts due to the

relatively low number of acceptable counterions when compared with the large number of cocrystal counter molecules that would be acceptable. Pharmaceutical cocrystals therefore represent a broad patent space and have broad scope for ''noncovalent derivatization'' without compromising the structural integrity of the API and thus, possibly, the bioactivity.

7.4.5
Polymorphism Control

Many organic molecules are emerging as having many crystalline forms, including polymorphs and solvates, as more techniques are being used to generate and characterize the organic solid state. The fundamental scientific and industrial interest in controlling crystallization and polymorph formation has inspired a very substantial research effort on this area, which has at its heart experimental methods for comprehensive polymorph screening, involving crystallization and characterization methods outlined above. Most new solid-form materials are now subjected to full polymorph, salt and solvate screening, for reasons both of ensuring sample purity, selecting the most effective formulation, and allowing for patent protection of discovered materials. Computational methods are also under development as an intended complement to these experimental investigations [39].

   The identification of polymorphs is critical, but once discovered, and with a given polymorph being desired for a particular application of the molecular material in the solid state, the ability to control polymorph formation becomes vital. This control can take the form of careful choice of solvent, temperature, crystallization regime, and the judicious choice of additives in appropriate cases. Seeding is also regularly employed, particularly for high-volume polymorph selection in batch crystallization processes [40].

7.5
Analytical Techniques for Product Characterization

As outlined above, increasing in-line ''*in situ*'' analytical techniques are being implemented in the monitoring of crystallization processes, allowing real-time intervention and control over the process. These include methods based on ATR and DLS. The analytical techniques discussed here can often be used both on-line and off-line, and methods for implementing more on-line solutions are being developed continuously. In any case, following recrystallization to produce a solid, often polycrystalline, sample, it is vital for product control and for the understanding of the process involved to undertake a series of characterization experiments using a wide range of analytical techniques, including thermal methods such as differential scanning calorimetry (DSC) and thermogravimetric analysis (TGA), various spectroscopies including infrared spectroscopy (IR), X-ray powder diffraction (XRPD),

SEM, and a range of light scattering and optical methods including DLS and low angle laser light scattering (LALLS). Some of these are also outlined here.

7.5.1
Focused Beam Reflectance Measurements and Attenuated Total Reflectance Ultra Violet

The theory of ATR is based on light passing from a material of high refractive index, for example, a crystal, to a material of lower refractive index, for example, a solution. Light travels to the reflection surface and can be partially absorbed by the solute before being reflected back to the probe. The reflected light is therefore attenuated, causing a measurable reduction in the output signal dependent on the absorbance of the solution [41]. Each probe has two fiber optic cables, one for transmission of the light from the light source to the measuring head of the immersion probe, and the other for conducting the signal, which is the light that has passed through the sample and back to the spectrophotometer. ATR-UV probes are suitable for the direct measurement of strongly absorbing solutions in which the UV absorption of the solvent does not mark the solute absorption.

The main advantage of using an ATR-UV probe for measuring solubility and crystallization in solution is that no sampling is required and that this is a real-time process. Other advantages are that the probe is relatively insensitive to the presence of particles in solution as the probe is based on surface measurements and it is suitable for an easy setup. However, in deciding to use the ATR-UV probe, it must also be considered whether the solute has a significant UV absorption compared to the solvent in which the process is carried out so that it can be measured in the presence of the solvent. The UV absorption of the solute is directly proportional to the concentration of the solution according to the Beer–Lambert law and therefore provides a convenient method for accurate *in situ* real-time measurement of solute concentration when compared to other techniques. This process comes into its own during cooling crystallization processes, enabling a significant solubility profile to be compiled from solubility to crystallization.

ATR-UV can also be used in connection with FBRM. FBRM uses a highly focused laser beam projected through the sapphire window of the FBRM probe to rapidly scan over a small region. The beam is rotated at a fixed high velocity, allowing rapid scans across particles flowing across the path of the beam. This high-speed scanning movement of the beam is significant as this means that the motion of the particles is insignificant. A particle entering the beam path produces back-scattered light, which is picked up by a stereoscopic system. The crystal continues to backscatter light until the beam reaches the other edge of the crystal. The time period of backscattering is recorded and multiplied by the scan speed of the beam to give the distance between one edge of the crystal and the other; this is known as the *chord length*. These chord lengths that are measured are profiled in a chord length distribution (CLD) plot. FBRM has been used to measure solubility and MSZW for potash alum using a ramping method.

7.5.2
Dynamic Light Scattering

DLS can be used for measuring the size and size distribution of molecules and particles in the submicron region, down to $\sim$1 nm, with obvious applications in the understanding of the particle formation and association that underpins crystallization processes. The technique is based on the interaction of an incident beam of laser light with particles that are intrinsically in motion through the Brownian motion indices by their collisions with solvent molecules that are moving due to their thermal energy. In such a situation, the fluctuation in the light scattered from the particles changes depending on the amount of motion of these particles, and hence on their size, since smaller particles will be affected more by the Brownian motion of the surrounding solvent. The particle size is extracted from such experiments using the Einstein–Stokes relationship for the diffusion constant:

$$D = \frac{k_B T}{6\pi \eta r}$$

where η is the viscosity, and r the radius of the spherical particle, which is valid in regions of low Reynolds number, where the viscous forces dominate over the inertial forces.

The particle dimension measured in DLS experiments is the hydrodynamic diameter as it refers to particle diffusion in a fluid. As can be seen from the above equation, the dimension obtained is that of a sphere with the same translational diffusion coefficient as the particle being measured. This interpretation is thus an approximation, as the diffusion coefficient will depend not only on the diameter of the scattering particle but also on other effects such as surface structure, charge, and so on. The particle sizes extracted from DLS experiment, however, are extremely useful in many contexts and are widely applied.

7.5.3
Ultrasound Methods

High-frequency ultrasound has a wide range of applications in science, medicine, and industry, including in crystallization and materials processing. Waves of high-intensity ultrasound generating cavitation effects in liquids locally – the formation of small vacuum bubbles or voids in the liquid – leads to extreme nonuniform effects in temperature and pressure, and also mechanical effects due to acoustic cavitation. These effects can be used in processing, where their influence on particle surfaces and collisions and local mixing and milling effects can be used in the control of crystallization [42].

7.5.4
X-Ray Methods

XRD is the definitive method for determining the structural characteristics of the materials formed during crystallization processes. XRD techniques are based on

the fact that when a beam of X rays is passed through a crystalline material, the beam is scattered by the electrons of the atoms in definite directions with varying intensities. This process is called *diffraction* and the scattered rays can be recorded with a detector giving a pattern reflective of the symmetry of the unit cell of the crystalline material or materials present, their atomic composition, and the positions of the atoms and molecules in the unit cell. There are two major techniques used in XRD studies in crystallization science, powder X-ray diffraction (PXRD) and single-crystal XRD.

When a crystalline material is produced from a crystallization process, it is frequently in the form of polycrystalline material. These so-called powder materials consist of a large collection of very small crystallites, which lead to diffraction patterns that are ideally suited to rapid identification. PXRD is primarily used for identification of new cocrystals and/or phases, although full structure determination is possible in some cases. In this method, the polycrystalline sample is loaded into a thin capillary or on a flat plate. The sample holder is placed in the X-ray beam on a powder diffractometer and the diffraction pattern collected as a function of scattering angle 2θ. In the case of a capillary mounted sample, the holder is rotated to minimize preferred orientation effects. The polycrystalline nature of the sample material means that, instead of a pattern of discrete spots, as is produced with a single crystal, a pattern of concentric rings is produced. As the powder contains randomly orientated polycrystallites, each individual crystal generates each Bragg reflection at the same Bragg angle but is diffracted in a different direction. The diffracted beams are measured by a detector and the intensity is recorded as a function of angle for each reflection, that is, a cross section through the rings. Figure 7.13 shows how the powder diffraction pattern (Figure 7.14) is built up.

Figure 7.13 The principles of powder diffraction and the construction of Debye–Scherrer rings from a polycrystalline sample containing many tiny crystallites in random orientations.

Figure 7.14 Typical X-ray powder diffraction pattern of intensity versus 2θ.

This technique is ideally suited to the rapid identification of crystalline phases present in the product sample, and modern high-throughput PXRD techniques, including instrumentation to allow automated multiple sample data collections, such as the Bruker General Area Detector Diffraction System (GADDS) system, that allow tens to hundreds of samples to be analyzed in a day. The technique is nowadays utilized together with sophisticated software for phase analysis including powder profile analysis software (Rietveld), extensive database of powder patterns International Centre for Diffraction Data Powder Diffraction File (ICDD-PDF), and multiple sample recognition software (PolySNAP). It allows almost on-line analysis of outputs including identification of new phases, polymorph identification, detailed quantitative analysis of crystalline and possible amorphous phases present, and subsequent analysis of any new phases synthesized in the crystallization process.

In the more specialist cases, particularly in the discovery aspects of crystallization process development, full characterization of new phases depends on the determination of crystal structures. For this, although high-resolution PXRD is sometimes able to allow for full structure determination, single-crystal XRD is the technique of choice. Provided that a single crystal of sufficient size can be obtained (for modern CCD-based diffractometers, this is typically 0.1 mm sized crystals), a full 3D diffraction pattern can be measured and, from measurements of the intensities of the reflections, the molecular and crystal structure can be routinely obtained, allowing full definition of molecular geometry and analysis of the interactions controlling the formation of the crystal structure.

7.5.5
DSC/TGA

DSC is a thermoanalytical technique, which is routinely used in the identification and characterization of crystalline materials, including multiple-component crystals and polymorphs of materials. The method is based on the principle that a change

Figure 7.15 Schematic view of the setup for carrying out differential scanning calorimetry (DSC).

in the physical state of a material is accompanied by the liberation or absorption of heat. In practice, it is used to measure the heat energy necessary to establish a near-zero temperature difference between a substance and an inert reference material, as the two specimens are subjected to identical temperature regimes in an environment heated or cooled at a controlled rate. One of the common analysis methods used in DSC is heat flux DSC, in which the sample and reference are enclosed in a single furnace and connected by a low-resistance heat-flow path (Figure 7.15) with a purge gas flow for efficient heat transfer and removal of volatiles.

As the temperature is changed in a linear manner, any heat changes in the sample result in a difference in the energy needed to maintain the sample and reference at the same temperature. As they are in good thermal contact, any excess heat energy will flow into the metallic disc and this heat flow is measured as it is directly proportional to the small temperature difference between the sample and the reference. This technique is used to observe phase transitions, polymorphs, and new phases in the materials resulting from crystallization processes as they undergo heating or cooling, as endothermic processes such as melting will result in a negative heat flow, whereas exothermic processes such as crystallization will result in a positive heat flow (Figure 7.16).

TGA can be used to monitor weight changes as the sample is heated up toward the melting point. The TGA instrument typically consists of a high-precision balance with a typically platinum pan. Once loaded with the sample, the balance is placed in an oven with thermocouples allowing for accurate temperature reading. On heating, the weight of the sample is recorded as a function of the temperature, with weight loss correlated with change in sample composition, for example, a dehydration step in which solvent water was vaporized would result in large weight loss. In many cases, combined DSC/TGA is used for sample analysis, the complementary information available from the two techniques offering further insights into sample behavior.

Figure 7.16 Example of a DSC trace exhibiting endothermic and exothermic processes.

7.6
Conclusions

It is clear that the importance of crystallization in manufacturing and processing is set to increase in the next few years, traditional batch crystallization methods being joined by a much enhanced emphasis on continuous and flow processing, facilitated by technological developments especially in automated control systems with multiple feedback routes from sensor arrays. Such moves to more dynamic crystallization environments make it still more important that both the fundamentals of the crystallization process and the translation of these into scaled-up environments are fully understood. A battery of process technologies and analytical techniques are available to the researcher, process engineer, or manufacturer in this area, and with the help of these techniques and the associated understanding of the fundamentals, a rational and rewarding approach can be taken to developments in this area.

Acknowledgments

References

1. Bernstein, J. and Braga, D. (2007) in *Making Crystals by Design*, vol. 1.295 (eds F. Grepioni and D. Braga), WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim.

2. Wilson, C.C. (2000) *Single Crystal Neutron Diffraction from Molecular Structures*, Chapter 1, World Scientific. ISBN: 981-02-3776-6.

3. Hartman, P. and Bennema, P.J. (1980) The attachment energy as a habit controlling factor. 1. theoretical considerations. *J. Cryst. Growth*, **49** (1), 145–156.

4. Black, S.N., Davey, R.J., and Halcrow, M. (1986) The kinetics of crystal-growth in the presence of tailor-made additives. *J. Cryst. Growth*, **79** (1–3), 765–774. [Proceedings of the Eighth International Conference on Crystal Growth].

5. Rodríguez-Hornedo, N. and Murphy, D. (1999) Significance of controlling crystallization mechanisms and kinetics in pharmaceutical systems. *J. Pharm. Sci.*, **88**, 651–660.

6. Barrett, P. and Glennon, B. (2002) Characterizing the metastable zone width and solubility curve using lasentec FBRM and PVM. *Chem. Eng. Res. Design*, **80**, 799–805.

7. Lacmann, R., Herden, A., and Mayer, C. (1999) Kinetics of nucleation and crystal growth. *Chem. Eng. Technol.*, **22**, 279–289.

8. Kashchiev, D. and van Rosmalen, G.M. (2003) Review: nucleation in solutions revisited. *Cryst. Res. Technol.*, **38**, 555–574.

9. (a) Scheel, H.J. (2000) Historical aspects of crystal growth technology. *J. Cryst. Growth*, **211**, 1–12; (b) Ueltzen, M. (1993) The verneuil flame fusion process – substances. *J. Cryst. Growth*, **132**, 315–328.

10. Müller, G. (2007) The Czochralski method – where we are 90 years after Jan Czochralski's invention. *Cryst. Res. Technol.*, **42**, 1150–1161.

11. Evers, H., Klufers, P., Staudigl, R., and Stallhofer, P. (2003) Czochralski's creative mistake: a milestone on the way to the gigabit era. *Angew. Chem.-Int. Ed.*, **42**, 5684–5698.

12. Schreiner, A. and Konig, A. (2002) Influence of impurities on nucleation and growth rates of organic melts. *Chem. Eng. Technol.*, **25**, 181–187.

13. Hoffman, J.D. and Miller, R.L. (1997) Kinetics of crystallization from the melt and chain folding in polyethylene fractions revisited: theory and experiment. *Polymer*, **38**, 3151–3212.

14. Sangwal, K. (1998) Growth kinetics and surface morphology of crystals grown from solutions: recent observations and their interpretations. *Prog. Cryst. Growth Charact. Mater.*, **36**, 163–248.

15. Ulrich, J. (1994) Recent developments in crystallization from solution. *Chem. Ing. Tech.*, **66**, 1341–1345.

16. Butler, J.E., Mankelevich, Y.A., Cheesman, A., Ma, J., and Ashfold, M.N.R. (2009) Understanding the chemical vapor deposition of diamond: recent progress. *J. Phys.-Condens. Matter*, **21**, 20.

17. Christen, H.M. and Eres, G. (2008) Recent advances in pulsed-laser deposition of complex oxides. *J. Phys.-Condens. Matter*, **20** (26).

18. (a) Tavare, N.S., Garside, J., and Chivate, M.R. (1980) Analysis of batch crystallizers. *Ind. Eng. Chem. Proc. Design Dev.*, **19**, 653–665; (b) Tavare, N.S. (1987) Batch crystallizers: a review. *Chem. Eng. Commun.*, **61**, 259–318.

19. Elsner, M.P., Ziomek, G., and Seidel-Morgenstern, A. (2009) Efficient separation of enantiomers by preferential crystallization in two coupled vessels. *AIChE J.*, **55**, 640–649.

20. Lian, G.P., Moore, S., and Heeney, L. (2006) Population balance and computational fluid dynamics modelling of ice crystallisation in a scraped surface freezer. *Chem. Eng. Sci.*, **61**, 7819–7826.

21. Miras, H.N., Cooper, G.J.T., Long, D.L., Bogge, H., Muller, A., Streb, C., and Cronin, L. (2010) Unveiling the transient template in the self-assembly of a molecular oxide nanowheel. *Science*, **327**, 72–74.

22. Chew, C.M., Ristic, R.I., Dennehy, R.D., and De De Yoreo, J.J. (2004) Crystallisation of paracetamol under oscillatory flow mixing conditions. *Cryst. Growth Des.*, **4** (5), 1045–1052.

23. Ni, X. and Liao, A. (2008) Effects of cooling rate and solution concentration on solution crystallisation of L-glutamic acid in OBC. *Cryst. Growth Des.*, **8** (8), 2875–2881.

24. Lawton, S., Steele, G., Shering, P., Zhao, L.H., Laird, I., and Ni, X.W. (2009) Continuous crystallization of pharmaceuticals using a continuous oscillatory baffled crystallizer. *Org. Proc. Res. Dev.*, **13**, 1357–1363.

25. Jones, F. and Ogden, M.I. (2010) Controlling crystal growth with modifiers. *Crystengcomm*, **12**, 1016–1023.

26. Leamy, H.J., Glimer, G.H., Jackson, K.A., and Bennema, P., (1973) Lattice-gas Interface structure – Monte-Carlo simulation. *Phy. Rev. Lett.*, **30** (13), 601–603.

27. Bourne, J.R. and Davey, R.J. (1976) Role of solvent-solute interactions in determining crystal-growth mechanisms from solution. 1. Surface entropy factor. *J. Cryst. Growth*, **36** (2), 278–286.

28. Bourne, J.R. and Davey, R.J. (1976) Role of solvent-solute interctions in determining crystal-growth mechanisms from solution. 2. Growth kinetics of hexamethylene tetramine. *J. Cryst. Growth*, **36** (2), 287–296.

29. Elwenspoek, M., Bennema, P., and Vandereerden, J.P. (1987) Orientational order in naphthalene crystal solution interfaces. *J. Cryst. Growth*, **83** (3), 297–305.

30. Bennema, P. (1992) Theory of growth and morphology applied to organic-crystals – possible applications to protein crystals. *J. Cryst. Growth*, **122** (1–4), p. 110–119 (Abstract | PDF (999 K) | View Record in Scopus | Cited By in Scopus (10)).

31. Weissbuch, I., Addadi, L., Lahav, M., and Leiserowitz, L. (1991) Molecular recognition at crystal interfaces. *Science*, **253** (5020), 637–645.

32. Lahav, M. and Leiserowitz, L. (2001) The effect of solvent on crystal growth and morphology. *Chem. Eng. Sci.*, **56**, 2245–2253.

33. Dirksen, J.A. and Ring, T.A. (1991) Fundamentals of crystallization – kinetic effects on particle-size distributions and morphology. *Chem. Eng. Sci.*, **46**, 2389–2427.

34. Wang, X.Z., Roberts, K.J., and Ma, C. (2008) Crystal growth measurement using 2D and 3D imaging and the perspectives for shape control. *Chem. Eng. Sci.*, **63**, 1173–1184.

35. Blagden, N., de Matas, M., Gavan, P.T., and York, P. (2007) Crystal engineering of active pharmaceutical ingredients to improve solubility and dissolution rates. *Adv. Drug Deliv. Rev.*, **59**, 617–630.

36. Abbas, A., Srour, M., Tang, P., Chiouc, H., Chan, H.K., and Romagnoli, J.A. (2007) Sonocrystallisation of sodium chloride particles for inhalation. *Chem. Eng. Sci.*, **62**, 2445–2453.

37. Ferreira, A., Faria, N., and Rocha, F. (2009) The effect of crystal surface roughness on impurity adsorption. *Cryst. Res. Technol.*, **44**, 521–533.

38. (a) Gagnière, E., Mangin, D., Puel, F., Rivoire, A., Monnier, O., Garcia, E., and Klein, J.R. (2009) Formation of co-crystals: kinetic and thermodynamic aspects. *J. Cryst. Growth*, **311**, 2689–2695; (b) Blagden, N., Berry, D.J., Parkin, A., Javed, H., Ibrahim, A., Gavan, P.T., De Matos, L.L., and Seaton, C.C. (2008) Current directions in co-crystal growth. *New J. Chem.*, **32**, 1659–1672.

39. Price, S.L. (2008) From crystal structure prediction to polymorph prediction: interpreting the crystal energy landscape. *Phys. Chem. Chem. Phys.*, **10**, 1996–2009.

40. Beckmann, W. (2000) Seeding the desired polymorph: background, possibilities, limitations, and case studies. *Org. Proc. Res. Dev.*, **4**, 372–383.

41. Mirabella, F.M. Jr. (1993) Practical spectroscopy series; Internal reflection spectroscopy: Theory and applications, Marcel Dekker, Inc., 17–52.

42. Mason, T.J. (2007) Developments in ultrasound – non-medical. *Prog. Biophys. Mol. Biol.*, **93**, 166–175.

8

Key Technologies and Opportunities for Innovation at the Drug Substance–Drug Product Interface

Colm Campbell and Brian Keaveny

8.1
Introduction

An extremely important, yet poorly understood, area in drug development and manufacturing is the continuum that constitutes the drug substance–drug product interface [1]. Attempts to examine the potential for value-adding opportunities would lead one to ask the fundamental question: why is it necessary to dry a drug substance, prior to dispatching to the drug product manufacturing sites, if the material is going to be dissolved in exactly the same solvent? The answer is, assuming the drug substance is stable in the solvent during the hold period, ''no reason whatsoever!'' This is one example of how the disconnect between the two types of pharmaceutical manufacturing is based on historical and not necessarily on scientific or economically rational considerations. It would seem logical that organizations that successfully bridge this ''drug substance–drug product interface'' are likely to yield benefits from increased efficiency.

This chapter sets out to describe this space, along with the supporting technologies, to include both manufacturing methods and analytical techniques. Opportunities for innovation and adding value, by exploiting these technologies, will also be explored. But first, we will define this area of pharmaceutical processing and why it is of importance for optimizing formulation operations and, more importantly, the performance of the drug product itself.

8.1.1
The Drug Substance–Drug Product Interface

Figure 8.1 sets out some of the key operations performed during typical pharmaceutical process chains. Primary manufacturing, which occurs in the drug substance facility, involves construction of the molecular framework by synthetic organic chemistry and workup, all the way to crystallization, isolation, and drying. The dry product is then dispatched to the drug product customer, who may or may not be part of the same organization. Even within the same organization, drug substance and drug product sites tend to be situated in different locations. Any blending that

Chemical reactions
Extractions
Distillations
Separations
Crystallisation
Isolation
(Spray)drying
Blending
Granulation
Tabletting
Filling
Packaging

Interface

Primary
manufacturing

Boundary

Secondary
manufacturing

Figure 8.1 The drug substance–drug product interface, in the context of key pharmaceutical manufacturing operations.

may occur on the primary site would tend to be of the single component variety, for the purpose of ensuring powder homogeneity. Multicomponent blending would traditionally be seen as part of secondary operations and would, therefore, occur on the drug product site. In addition, many other operations may occur, ultimately resulting in the approved, final dosage form. Particle comminution or size reduction methods, although of huge importance, are not included in the diagram, as this can occur on either side of the boundary. It is, however, discussed in some detail in Section 8.4. Spray drying is also a common activity in secondary processing sites.

The *interfacial area* can be defined as the final operations in drug substance manufacturing and the initial formulation operations. Where this interface starts and finishes is a matter of debate. It may or may not include crystallization and granulation, depending on one's own perspective. The key message is that, although the operations are carried out on different manufacturing sites or by different organizations, a good understanding of the science and processing issues, by both primary and secmaondary manufacturers, is critical to success. If the formulator understands what particulate characteristics the active pharmaceutical ingredient (API) should possess for optimal process and product performance, an effective API manufacturer can modify crystallizations, or introduce powder processing steps, to provide an API that meets expectation. It should be possible to produce different crystal sizes to order by altering supersaturation build-up rates in crystallization and milling; or compaction can be used to provide fine particles or granules, respectively. However, more subtle requirements, such as generating particles with particular pore distribution properties, or particular types of surface groups, generally require more complex particle engineering solutions.

This can be facilitated by constructive dialogue between the primary supplier and the secondary customer, thus ensuring that the drug substance meets the requirements of the end user. All too often, the primary producer believes that the attainment of a quality specification, generally based on chemical purity, represents a job well done. The additional requirement of ensuring an easily formulated product, an issue that gives rise to huge resource wastage in secondary plants is, sometimes, not part of the success criteria. This problem is exemplified by the fact that ''particle science'' activities, the field that involves provision of the correct particulate properties for a particular drug, tend to occur during early development

only [2]. This means that in-depth particle properties are typically finalized for a particular drug substance, many years before scale-up in manufacturing sites, when some of the real processing issues first start to show up. So, it is not difficult to envisage scenarios where specifications that are not actually relevant to the performance of the final product can be set. As our understanding is often limited during early development, it is important to challenge the specifications that are set, why they are set, and whether they are actually relevant to the product's efficacy as the drug moves toward commercialization!

It is apparent that the current process chemistry/quality-oriented specifications that are typically observed for APIs should be augmented by a second set of specifications that are set with the specific formulation process in mind. These ''functional'' specifications should be developed to identify particular physical characteristics that the API should possess to meet the requirements of the formulation process, for example, flow performance, dissolution performance, and so on. Indeed, the fact that most of the other raw materials used by formulators – fillers, binders, colors, flavors, disintegrants, lubricants, glidants, and so on – typically have functional specifications is another indication of the artificial dichotomy that exists between the primary and secondary manufacturing worlds.

Confining the discussion to tabletting alone as an example, the bulk powder requirements for typical secondary manufacturing would include optimal feed rates from the hopper and into the dies, good compressibility, and tablet strength [3]. The ability of the primary manufacturer to produce material that functions well in these categories can greatly improve the efficiency in the secondary plant.

Clearly, crystallization strategies, or particle engineering methods such as milling, roll compaction, and spray drying can be used to provide API with the optimal characteristics for secondary processing. For example, needle morphologies tend not to compress well into tablets, so crystallizations that favor plate-like growth are preferred by the secondary customer. That said, filtration, washing, and drying of plates on the primary plants can present problems. Only by getting the two disciplines together to discuss the issue and decide which option is preferable can the optimal process be identified. It is worth noting that additives can be used to deliberately block growth on particular faces, thereby preventing proliferation of a needle habit [4]. Some of these technologies will be discussed in more detail in Sections 8.3 and 8.4.

8.1.2
Physical Characteristics and Bioavailability

Perhaps more importantly, effective presentation of suitable API for formulation also includes providing material with characteristics that maximize bioavailability. Once again, the primary manufacturer has an opportunity to influence the drug product performance by providing material that will possess properties that maximize drug performance. Practically, the form accessed by crystallization and the associated particle engineering activities, as well as the particular particle

Figure 8.2 The Biopharmaceutical Classification System (BCS).

morphology in which the chosen form is presented, will profoundly affect solubility and permeability of the drug, the two drivers for bioavailability.

The Biopharmaceutical Classification System (BCS), illustrated in Figure 8.2, shows the four different classes of drug substances, as categorized by intestinal absorption. This is a useful guide for assessing the key drivers of bioavailability, for a particular drug. Class I compounds show both good permeability and solubility and are, therefore, well absorbed, while Class IV compounds, which show poor solubility and permeability, are poorly absorbed. Classes II and III represent intermediate conditions, with class II compounds showing a correlation between bioavailability and solvation rate and class III drugs requiring a formulation that enhances permeability, for optimal efficacy [5].

Largely arising from these considerations, different API morphologies tend to be associated with particular delivery modes. For example, inhalation grade actives are typically presented as fine, micronized particles, available for absorption via pulmonary surfaces. Of course, oral delivery products can cope with a wide range of morphologies and in this case, the morphology is often tailored to the peculiarities of the process, rather than particular bioabsorption aspects. Intravenous or intramuscular delivery requires no special API structural attributes, as the material is dissolved prior to delivery.

When choosing a particular form during preformulation, to progress to commercialization, a form with a weaker crystal lattice will exhibit a lower melting point and therefore, a higher aqueous solubility [6]. Bioavailability can be driven by solubility, permeability, or a combination of the two. In the former case, progression of a metastable polymorph will result in bioavailability benefits. During drug development, it is always important to balance this with the processing imperatives – more stable forms are easier to reproduce and keep intact, throughout processing and storage.

Dissolution rate alone is often a poor indicator of drug absorption performance, with overall solubility and permeability also being important. More than 90%

of all drugs are absorbed in the GI tract, further enforcing the importance of permeability. A drug is considered to undergo ''rapid dissolution'' if >85% of ''full dissolution'' is achieved after 30 min.

Aqueous solubility is, very often, quite poor in commercial APIs, as at a molecular level, these drugs tend to be hydrophobic. That said, there are approximately 10^{13} orders of magnitude in aqueous solubility, when surveying the myriad of commercial drugs available. At this stage, it is worth pointing out that *in vivo* dissolution performance in mammalian systems is greater than that in aqueous solutions. For example, lipophilic drugs dissolve well in bile, which contains large amounts of lecithin. Most reasonable model studies should employ ''simulated gastrointestinal fluid'' (SGF).

Once the particular form is chosen and filed with the regulators, there are still ample opportunities to optimize aqueous solubility. Once again, drug substance development and manufacturing personnel, who are conscious of the needs of the end user, can modulate API properties by carefully designed processes. This may involve smart crystallization strategies, well-designed comminution methods, and optimized secondary processes, or preferably, a combination of all three.

The general principle that powders, made up of more finely divided particles, will tend to exhibit greater dissolution rates (but not solubility) by virtue of possessing greater surface area, often proves not to be the case. Quite often, ''primary particle size'' is irrelevant to dissolution rate and where more subtle forces are at play. Of course, particle size distribution is often a key specification for an API, although the relationship between particle size and dissolution rate can be overestimated.

In more general terms, surface wettability is a key factor in dissolution rate. This can be improved by using surfactants in the formulation or by applying crystallization and comminution strategies that expose ionizable groups on the surface. In many cases, it is not the amount of surface available that alters dissolution rate, but the chemical nature of the surface itself. The Noyes–Whitney equation, shown below, captures the essentials of dissolution rate *in vivo*:

$$\frac{\mathrm{d}m}{\mathrm{d}t} = \frac{A * D * (C_s - C_t)}{l}$$

m mass of dissolving solute,
T time,
A wettable surface area,
D diffusion constant of solute in solvent ($\mathrm{m^2\ g^{-1}}$),
C_s saturation solubility concentration,
C_t concentration at time (in vivo, function of permeability),
l thickness of unstirred diffusion layer.

8.2
Opportunities for Innovation

Two value-adding areas of innovation in the drug substance–drug product interface will be considered: tailored APIs and part-formulated APIs.

8.2.1
Tailored APIs

Despite the efforts of intensive, combinatorial chemistry initiatives to push more candidates into pharmaceutical pipelines, around 75% of all new chemical entities approved by the Food and Drug Administration (FDA) in 2008 involved reformulation [7]. This is not surprising, when we consider the relative lack of success in the NCE drive, coupled with the pragmatism of the industry globally.

One issue that needs to be overcome, however, by the formulators with proprietary technology (the so-called ''super generics'') is the poor suitability of many off the shelf generic APIs for the particular secondary technologies in question. The generic API producer, by and large, has not got the economic incentive or, in many cases, technical know-how to produce tailored API solutions. Consequently, a potentially lucrative market exists for organizations that are capable of connecting with formulators, to provide material with physical characteristics that match the technology in question.

8.2.2
Part-Formulated APIs

A very different market space exists for organizations capable of part-formulating API. This value-adding activity involves carrying out some of the formulation activity on the API site, thereby capturing some of the drug product value. No better example of successfully occupying this interfacial area exists than where, in some cases, the primary manufacturing equipment is used to carry out classical secondary operations.

An example, performed at the authors' organization, involves the use of a Hosokawa Vrieco-Nauta dryer, to part-formulate or more specifically, granulate a particular API [8]. The value of such a process clearly derives from the capturing of much of the formulation value on the API site, where two processes are effectively merged into one. Ordinarily, the API is dried on the primary site and granulated at the secondary facility, in a Glatt Fluidized Bed Dryer. Trials on pilot scale Vrieco Nauta dryers showed that the granulation could be performed reproducibly, leveraging the excellent mixing afforded by the rotating Auger screw/arm system. The granule size could be tuned by adjusting the agitation rates, and the output from many of the trials possessed comparable granule size to material from the established process, along with comparable dissolution rates.

8.3
Crystallization

The contribution of crystallization to pharmaceutical processing is an enormously important and detailed field, which is beyond the scope of this discussion and forms the subject of Chapter 7. For a recent industrial perspective on API crystal

product and process design within the pharmaceutical industry – current state of play and continued barriers to the improved production of crystalline products [9].

However, we will take this opportunity to discuss three techniques: spherical, ultrasonic, and continuous crystallizations, as they are capable of furnishing API with characteristics that provide specific formulation advantages and, very definitely, are of importance at the drug substance–drug product interface.

Before this, one example of how understanding basic crystallization theory can contribute to large processing efficiencies will be described [10]. While this example refers to an intermediate, the rationale can equally be applied to APIs, thereby illustrating how physical habits can be dramatically altered, either to improve processing performance or to acquire particular characteristics for optimal formulation performance.

In this case, many tens of tonnes of an intermediate were isolated for many years, without serious difficulty, as part of a very well-established chemical step. In the process, the sodium salt of the intermediate is cooled to $<30\,^{\circ}$C, whereupon the salt crystallizes. Concentrated hydrochloric acid is then added, to convert the salt to the free sulfonanilide and the product is isolated by centrifugation. However, crystallization of much of the sodium salt, prior to adding the acid, results in the addition of an unwanted additive to the precrystallization solution, which in turn, leads to the favoring of nucleation over crystal growth. By carrying the process out at $>50\,^{\circ}$C, thereby keeping the salt dissolved throughout the process, a ''cleaner'' crystallization is achieved, where a product with a macrocrystalline, granular habit results. A comparison of particle size distributions and physical habit is presented in Figure 8.3, illustrating the improved morphology.

Clearly, output from the modified process possesses many advantages over the existing process, in a pharmaceutical plant. These include

- faster isolation;
- less dusty;
- de-liquors more on the centrifuge, so less time is needed in the dryer operation;
- the product cake compresses more, leading to a higher bulk density. This means that the batch can be isolated in less loads and it also requires less intermediate storage capacity (fewer drums or IBCs).

8.3.1
Spherical Crystallization

This technique started to receive literature attention in the early 1980s but, to date, has not become a mainstream industrial methodology. The main benefit is that the time-consuming granulation step can be avoided, if the granules (or spherical agglomerates of crystals) are produced directly in the crystallization.

Initial reports described how agglomeration of salicylic acid crystals into spherical arrangements could be compressed directly into tablets, thereby bypassing the granulation step [11]. A more targeted approach appeared in a later report, where the drug substance, ibuprofen, was crystallized as spherical agglomerates, in

Figure 8.3 Particle size distributions and optical micrographs of product from the "old" and "modified" processes.

the presence of EudragitR S100, a polymer which appears to induce spherical agglomeration, by inducing habit and growth rate changes on the ibuprofen microcrystals [12]. This reference also pointed out that flow rates are much enhanced for spheres over other particle types, a point that was elaborated upon in a later article, in the context of other physical properties [13].

Using mefenamic acid as an example, among other things, the authors showed enormous increases in particle size, surface area, and aqueous solubility, in the spherical crystals, relative to the standard powder. Also, bulk density decreased, reflecting the tighter packing of the spherical agglomerates. Common techniques used now, to achieve this spherical effect, include emulsion solvent diffusion, ammonia diffusion, and neutralization. The former, often called "*emulsion crystallization*" has received some literature coverage, where a dispersed phase, stabilized by a suitable surfactant, can provide spherical particles by crystallization, if a supersaturated solution is maintained within the droplet [14].

8.3.2
Ultrasonic Crystallization

Certain drug delivery modes, in particular, inhalation, require finely divided particles, to ensure effective and safe deposition in the lungs. This can be achieved using comminution techniques, such as jet milling (micronization) or the various nanosuspension technologies (see Section 8.4). These methodologies

are, however, prone to introducing crystal defects, via plastic deformation and ultimately, amorphous content, when it may not be desired. However, nanocrystals can be prepared directly from the crystallizing solution, using solution atomization and crystallization by sonication (SAX) technology, providing particles with well-designed surface characteristics and geometries. The technology was invented by Robert Price, University of Bath, and is now licensed by Prosonix. The basis involves the use of ultrasound energy, where cavitation mechanisms promote nucleation. Also known as ''*sonochemistry*,'' this results in a high degree of process control, where nanometer-sized seeds can be tuned to provide particles with specific habits, amorphous states, polymorphs, and particle size distributions. Particle habits of many types have been described, including spheres, which possess the advantages described above. Prosonix have developed reactor systems that facilitate scale-up [15].

8.3.3
Continuous Crystallization

Continuous crystallization processes have been developed, although they are not commonly practiced in the pharmaceutical industry. Continuous or steady state operation is not always the ideal mode for crystallizations and they can suffer with difficulties such as encrustation on heat exchanger surfaces or undesired self-seeding. However, they can be useful when conventional batch crystallization and postcrystallization milling cannot address the desired particle size requirement; there are many instances where dry milling as a means of achieving correct particle size is either dangerous, as a result of the dust explosion potential of the powder, or extremely inefficient because of the need for multiple passes. Furthermore, continuous operations can offer significant cost savings compared with batch operations and reduce the likelihood of lot to lot variability. A typical approach involves the use of a high-shear rotor-stator chamber to crystallize compounds at a high degree of supersaturation under good mixing. Semicontinuous crystallization processes have also been utilized in an attempt to overcome the distinct limitations of batch and continuous operations. This area is addressed in more detail in Chapter 7.

8.4
Selected Manufacturing Technologies at the Drug Substance–Drug Product Interface

This section covers the principles and applications of selected technologies in this area, namely, micronization, nanonization, blending, and roll compaction. These technologies, along with specialist crystallization methods, can be used to provide particles and bulk properties that suit particular applications, the overall field often being referred to as ''*particle engineering*.'' Filtration and (spray) drying are not covered, partly because the key know-how in these areas typically fits into the category of ''institutional knowledge'' and partly because these techniques tend to

be well understood by manufacturers. Wet granulation, a vast area which arguably goes beyond the interface into secondary manufacturing proper, is also beyond the scope of this review.

Because of the nature of powder materials, processes and facilities must be carefully designed to avoid accident scenarios that can lead to combustible dust explosions. Such dust explosion scenarios include explosions within process equipment, explosion propagation into interconnected equipment, and secondary dust explosions in buildings. Many of the techniques discussed below can create conditions that can increase the risk of dust explosions, when there is a particulate cloud with a concentration between its minimum explosible concentration (MEC) and its upper explosible concentration (UEC), in contact with an appropriate ignition source. A thorough assessment of these risks is beyond the scope of this chapter [16].

8.4.1
Micronization

At present, about 40% of the drugs in development pipelines exhibit poor aqueous solubility, largely due to their increasing complexity and hydrophobicity. This, in turn, leads to poor bioavailability. This can be improved by modifying the drug "chemically," which increases the saturation solubility. Some obvious ways of doing this include forming water-soluble complexes (e.g., with with β-cyclodextrins), soluble salts, or preparing some type of drug conjugate, with water-soluble functions [17]. A second approach involves increasing the dissolution velocity by increasing the surface area of the drug powder [18]. Traditionally, micronization provides a way of providing "micron" sized particles. There is some disagreement in the literature regarding the typical size distribution of a drug after micronization, but 0.1–25 μm spans most of the suggested ranges.

This technique can employ either high pressure (jet) milling, or very fast precipitation, typically using supercritical fluids (SCFs) [19]. This is usually motivated either by the requirements of the particle size for the dosage form in question (e.g., inhalation) or to increase dissolution rates in poorly soluble APIs, by increasing the surface area and wettability. Of course, it is possible to produce very fine particles by solvent/antisolvent (SAS) precipitations in conventional systems, but standard filtration technology is not designed to filter such fine material, leading to very slow and costly filtration times. A positive feature of SCF precipitations is that the particle size and morphology of the output can be controlled more favorably than in conventional precipitation processes [20].

8.4.1.1 Jet Milling
Different types of jet milling arrangements are available, but all of them operate on the principal of size reduction by high pressure particle collisions. This type of technology can be a very effective way of consistently producing fine, sub-10 μm particles. Micronized material can often flow very poorly, owing to the high level of aeration imparted by the process, leading to cohesive bulk properties. In addition,

the high frictional impacts can induce crystal defects and amorphous content, giving rise to less stable phases. However, with the correct engineering controls and process understanding, this type of processing can be used to effectively provide large amounts of API to a narrow, reproducible particle size range.

8.4.1.2 SCF Precipitation

An SCF is any material maintained above a critical pressure and temperature, so that it has the properties of both a liquid and a gas. This technology can be used to create micron-sized particles that can be administered with pulmonary drug delivery technology. Pulmonary drug delivery is becoming increasingly popular because it provides a noninvasive route with rapid drug uptake. It can be used for the systemic delivery of substances, as well as the delivery of respirables. In addition to affording sufficiently small particles for use in dry powder inhalers (DPIs), SCF also provides uniform crystalline particles with smooth surfaces. This is not achieved by milling micronization and it has been shown that smooth-surfaced particles improve dispersion during inhalation. As an added advantage, smooth particles do not interact as readily with excipients as rough-surfaced particles during storage, resulting in increased stability [21].

Typically, the SCF is CO_2, which can also be used as a dispersing antisolvent. Either the rapid expansion of the supercritical solution (RESS) or the SAS process can be applied. A third but less widespread technique involves particles from gas saturated solutions (PGSSs). CO_2 is generally used because it is nontoxic, cheap, readily available, and it has a critical that is point easy to reach (31 °C, 74 bar) [22].

In the RESS technique, the drug is dissolved in an appropriate organic solvent and the solution is co-introduced, along with the SCF, through the annuli of a two-fluid coaxial nozzle, where thorough mixing of the two streams occurs. Variations, such as the use of a second solvent, can be used in a more sophisticated set-up, using a three-fluid coaxial nozzle. The solvent is extracted with the SCF and the resulting dry particles are collected in a separate vessel.

The SAS method involves dissolution of the API in a suitable organic solution, followed by rapid injection of the SCF (antisolvent), giving very fast supersaturation buildup and therefore, very small particles.

To date, however, this technology has not enjoyed widespread application in industrial processes, and scale-up issues are discussed [23]. That said, it was reported that, following almost three decades of academic and bench scale research, industrial applications are imminent in the food area [24].

8.4.1.3 Contrast between Jet Milled and *In situ* Micronized Material

While jet milling is a well-established, proven technology, albeit with some disadvantages, SCF micronization is still in its infancy. It does, however, offer potential advantages, owing to the high level of size and morphology control possible. Jet milling results in thermodynamically activated surfaces, which can cause a high degree of agglomeration behavior [24, 25].

Some literature examples, involving direct comparison of the two techniques, drawing from particular processes, are summarized below:

- A comparison was made between untreated, jet milled, and *in situ* micronized Fluticasone-17-propionate. The *in situ* micronization involved an SAS technique, in the presence of a cellulose ether (HPMC) stabilizing colloid, to prevent agglomeration and crystal growth. The product was isolated by spray drying. X-ray patterns of the jet milled material showed differences, compared to the untreated and *in situ* micronized material, suggesting that the process resulted in some phase changes. Also, the fine particle fraction ($<5\,\mu m$) was fourfold greater for the *in situ* micronized material [24].
- In a similar study to that described above, disodium chromoglycate was micronized by jet milling and SAS SCF (stabilized with the same HPMC colloid) methodologies. Once again, comparison strongly favored the *in situ* micronized product, where the fine particle fraction ($<5\,\mu m$) increased from 7% in the jet milled output to 75% in the *in situ* micronized material. Aerodynamic properties were also significantly better in the latter case [26].
- RESS SCF processing of budenoside resulted in material with significantly different aerodynamic properties than the jet milled equivalent [25].

8.4.2
Nanonization

Accessing nanoparticles, by either grinding in a suitable mill (bottom up) or creating by an *in situ* technique (top down), represents the next logical step in size decrease, from micronization. Nanoparticles, which range in size from 200 to 600 nm offer advantages over microparticles, in principle allowing more widespread applicability in inhalation delivery technologies and faster dissolution rates [27].

The main production technologies currently in use to produce drug nanocrystals employ nanosuspensions (registered as Dissocubes®), where the nanoparticles are suspended in water. The technology has also been extended to the formulation of drug nanocrystals in tablets and capsules.

Nanocrystals are produced in a pearl mill, using milling balls that consist of glass and zirconium oxide. However, because of toxicity concerns, resulting from erosion of the materials of construction and contamination of the drug substance, the preferred material of construction for pharmaceutical applications is a hard polymer, such as cross-linked polystyrene. Depending on the hardness of the powder and the required fineness of the particle size, the milling times range from hours to days [28]. A literature example showed that fluticasone and budenoside could both be formulated as nanosuspensions, by wet milling with glass beads. Moreover, intratracheal administration in animal models showed deep lung deposition and fast lung absorption [29].

Dissocubes® are produced in a homogenizer. The material should be jet milled and suspended in an aqueous surfactant, prior to processing, although some homogenizers can use material that has not been pretreated in this way. Typical operating pressures are 100–1500 bar. The technique can be extended to aseptic production or for processing cytotoxics [28].

Other methods of nanoparticle preparation include spinning disc processing (SDP), which employs flash nanofabrication technology and is capable of producing excellent particle control (size, shape, surface characteristics). It has been reported that this technique can produce particles in the range 5–200 nm [30].

Nanocrystals and Dissocubes® find application in the traditional dosage forms such as topicals, tablets, or other solid forms that go directly to the gastrointestinal tract, pulmonary delivery, and parenteral administration.

8.4.2.1 Contrasting Performance of Micro- and Nanoparticles

Dissocubes® offer a major advantage in pulmonary delivery over microparticles. They can, of course, be delivered in the same way, using mechanical or ultrasonic nebulizers, but owing to their small size, it has been claimed that each aerosol droplet will contain at least one drug particle, leading to a more even distribution of the drug in the lungs.

One example, involving a comparison of nano- and microsalbutamol delivery using dry powder inhalation, showed that the nanoparticles achieved more than twice as much lung deposition over the microparticles [31], while another study describes the production and characterization of mucoadhesive nanosuspnesions of the naphthoquinone antibiotic, bupravaquone [32]. This was prompted by the poor oral bioavailability of the drug and it points to a useful synergy between nanosuspensions and hydrogels in circumventing oral administration.

The literature also contains impressive data in the parenteral area. Immuno-compromised patients are at risk of developing toxoplasma encephalitis (TE) and atovaquone shows potent *in vitro* activity against the associated pathogen, *Toxoplasma gondii* [33]. However, it is poorly absorbed when orally administered, resulting in poor therapeutic efficacy. Murine studies showed excellent activity against *T. gondii*, when the drug was administered intravenously as a nanosuspension. However, the article does not explicitly predict the likely result of intravenous injection of a suspension of microparticles, but it is stated elsewhere that microparticles are too large for this route of administration [28].

Clinical trials, involving DPI delivery of nanoatropine sulfate, were performed, with the aim of assessing advantages over the conventional intramuscular delivery [34]. The study appeared to show bioavailability advantages, by virtue of a quicker delivery to the blood, via the lungs, as well as sustained action, due to absorption from the gut, of the remaining portion.

The widely known oncology drug, paclitaxel, is administered as nanocrystals, with preferred particle sizes of 100 nm [35]. The importance of this technology to BMS, the brand leader, is emphasized by their licensing agreement with Elan's drug delivery business unit, Nanosystems [36]. This unit has already commercialized several nanonized drugs and they claim to have many more in the pipeline. BMS plans to use the collaboration to formulate drug candidates and add market value to existing products, using this technology. Baxter, through NanoEdge technologies, has also invested heavily in the area, using the technology to prepare lipid emulsions, as well as nanosuspensions [37].

Table 8.1 Different blending types and associated equipment.

Mechanism	Description	Equipment
Diffusion	Particles redistributed by random motion	V-blender Double cone Bin
Convection	Low shear transfer of particles by moving parts	Ribbon Planetary
High shear	Particle movement via slip planes	Shaft pressurized diffusion
Low shear	Pneumatic transfer	Fluidized bed (dryer)

8.4.3
Blending

This ubiquitous methodology involves dry mixing of two or more components together, where an API is typically mixed with one or more excipients, using one of the mechanisms outlined in Table 8.1. The main aim is to achieve a homogeneous dispersion of the different materials, prior to tabletting or some other secondary manufacturing operation. Sometimes, agitation of a single component to achieve physical homogeneity is also described as blending.

Different types of blending mechanisms define the type of equipment to be used. The common types are summarized in Table 8.1.

By far the most common type is the diffusion method and therefore, it will be discussed in some detail. In primary manufacturing plants, this often follows a drying step to ensure powder homogeneity, while in formulation operations; this technique is used to homogeneously disperse the active among one or more excipients.

Generally, particles of similar size will tend to blend well with each other, so an accurate measurement of particle size, by microscopy, laser diffraction, or directly, via sieve plates analyses, are important (see Section 8.5.1.5). However, other factors can complicate performance, such as surface roughness, shape, particle density, or more subtle issues, such as stickiness, surface charge, or surface porosity [38].

Good blending is most easily achieved with good flowing constituent powders, while poor flowing cohesive powders tend to blend more poorly. However, the same can be said for segregation, the process by which the same mechanisms that give rise to homogeneous blending also give rise to "de-blending" [39]. This can happen, even upon storage in static powder containers, if the constituents are sufficiently free flowing. Simple desegregation (percolation) mechanisms can be quite effective in facilitating powder segregation, especially in noncohesive powders, where the fine ones settle at the bottom. Particle separation, resulting from differing surface charges within the powder mixture, can also be problematic when it is considered that tumbling can help build up the static charges. This can result in different

levels of adherence to, for example, the product liner, ultimately resulting in a loss of homogeneity.

Within the blender, mixing is achieved by the constant dividing and intermeshing of particles. In addition to particle–particle interactions, a well-designed process should take into account the dead volume in the blender, number, speed, and total time of revolutions, as well as sampling. This latter parameter is associated with many issues, as many off-line samples, removed by sample thieves such as the Globe Pharma thief, groove thief, end cup sampler, and core sampler, might be required to successfully demonstrate homogeneity [40]. It has been asserted that well-mixed powders can be accurately characterized with 30 samples, while poorly mixed systems may require hundreds [41].

Process analytical technology (PAT) techniques, principally in-line NIR spectroscopy, and to a lesser extent, thermal effusivity, have been shown to be very useful for confirming the blend end point. This is discussed further in Section 8.5.3.

8.4.4
Roller Compaction

This technique, surprisingly, is not used as widespread in pharmaceutical processing as perhaps it should be, finding more common application in the food sector. That said, its potential is now being realized by several organizations. It involves compressing product into sheets under pressure and then milling, to afford granules of different size ranges, as required. This is of great advantage in providing free-flowing agglomerates with high bulk densities, occupying the minimum of storage volumes. This type of operation is often achieved in the pharmaceutical industry by wet granulation, a more complex process that usually involves the introduction of excipients.

It can also provide a way to furnish powder with specific requirements, to interface with super-generics. For example, tablet tensile strength can be increased if the force used in compacting the powder is increased. Also, as it uses fairly low energies, it is less likely to introduce phase changes and amorphous content generation than other powder manipulation techniques, such as jet milling or nanonization.

The technique is the focus of some studies in the academic literature. For example, NIR spectroscopy has been used to correlate particle size distribution and compact strength for roller compacted powders, milled under different roll speeds and feed rates [42]. Thermal effusivity (see Section 8.4.3) has also been studied, to assess powder homogeneity during roll compaction [43].

8.5
Analytical Techniques

This field is populated by a variety of analytical techniques, many of which find widespread use in the pharmaceutical industry, while many others are only

in their infancy, although well applied in other industries. While the standard spectroscopic techniques are well established for studying molecular properties in the drug substance environment, many of the techniques relevant to this area would be considered highly specialized. This can often mean that during preformulation or other phases of drug substance material analysis, only certain types of analyses are carried out by particular companies.

The types of analyses relevant to this area can be broadly divided into surface/particulate and bulk, although there is some overlap. Nevertheless, this section will briefly describe some of the key analytical tools, while emphasizing relevance to pharmaceutical applications. Methods that are particularly applicable to blends will also be reviewed along with emerging technologies. Technologies concerned with molecular and supramolecular understanding, though relevant to drug development and understanding, are beyond the scope of this review, which surveys techniques that can be used to match drug substance characteristics with formulation performance.

8.5.1
Surface/Particulate

8.5.1.1 Atomic Force Microscopy (AFM)
Also known as scanning force microscopy (SFM), this surface technique, which studies the surface monolayer only, measures force and surface stress, in the context of surface interparticle interactions [44]–[46]. It is a microscope-based technique, which is very sensitive and works best when the surfaces are fairly smooth. Unlike scanning electron microscopy (SEM), this apparatus provides a 3D map of the surface particles. The technique finds application in studying crystal growth, polymorphism, and as it involves the surface, not surprisingly, coatings.

8.5.1.2 Dynamic Vapor Sorption (DVS)
This widely used technique can provide information on amorphous content (to below 0.2%), polymorphism, surface energies, hydrate and solvate formation, as well as moisture-induced crystallization kinetics [47], [48]. In essence, it involves the study of water uptake on a solid material, such as a drug substance or product, to understand propensity for water-mediated changes, such as surface recrystallization, which can result in the formation of a different solid phase, on storage or handling. Due diligence studies frequently apply this technique to assess the range of relative humidity conditions that a drug can be subjected to.

8.5.1.3 Focussed Ion Beam (FIB)
This technique, which functions similarly to the SEM, can provide microstructural analysis of chemical composition upon surfaces. It can be used for microscopic examination of microspheres. While SEM employs a beam of electrons, this technology beams ions most commonly arising from a gallium source. It finds most widespread application in semi-conductor applications, where it can be used to micromachine surfaces, modifying materials at the micro- and less commonly,

nanoscale. A novel pharmaceutical application involves analysis of surface porosity of spray-dried particles from an aerosol formulation [49].

8.5.1.4 Inverse Gas Chromatography (iGC)

This method, which involves packing the GC column with the analyte and injecting acidic, basic, and amphoteric probes to ascertain the nature of the surface, by assessing their retention behavior, is becoming quite widely used in the pharmaceutical industry. By understanding the nature of the surface, which has both dispersive (nonpolar) and nondispersive (polar) components, insight into surface wettability, dissolution rate, and coating properties can be elucidated. When applied to different batches of the same product, this can be an excellent method for troubleshooting batch to batch variation in, for example, dissolution rates [50].

One study, designed to understand the differences in surface energies before and after milling D, L-propranolol hydrochloride, showed that the surfaces became more energetic and electron donating after milling [51]. This was rationalized with the aid of molecular modeling, where it was hypothesized that the increase in surface energy, postmilling, was largely due to the increase in exposure of the 101 face, which is populated with the electron-rich naphthalene moiety.

By using different crystallization systems to prepare different morphological variants of ibuprofen, different values of the nondispersive and dispersive components resulted. This was correlated with the relative preponderance of the 001 face when polar interactions were favored; and the relative preponderance of the apolar 110 face when dispersive interactions dominated. It was hypothesized that the different habits would exhibit different dissolution rates owing to the relative exposure of different surface groups [52].

8.5.1.5 Particle Size

Arguably the most popular methods of material analysis, used in pharmaceutical development and routine manufacture, are those used to measure particle size [53]. This is often used to assess the degree of agglomeration, reproducibility of crystallization or milling processes, as well as the suitability of a material for secondary operations. Unimodal distributions, where one discrete population of particle sizes persists, is generally considered a desirable outcome, with the absence of filter clogging "fines" and difficult to dissolve agglomerates.

While many different types of analysis exist, they can generally be classified as direct or indirect. Examples of the former include qualitative microscopic techniques, such as optical or scanning electron microscopy and sieving, while indirect methods use laser diffraction and suitable algorithms (for example, Fraunhoffer for Malvern and Mie for Sympatec) to approximate the distributions.

The main techniques are described in the following subsections.

8.5.1.5.1 Microscopy The key methods are the traditional, optical $(1-150\,\mu m)$ and scanning electron $(0.001\,\mu m)$ techniques. Both are semiquantitative at best, but are useful in that they provide a feeling for shape, defects, and other visual observations, which the quantitative methods cannot do. Sophisticated SEM systems can give

fine particle detail easily showing, for example, particle indentation. Interfacing with other techniques, such as Energy Dispersive X-ray analysis (EDAX) or FT-IR, can result in compositional analysis of single particles. This can be useful for determining the nature of small amounts of impurities in drug substance batches.

8.5.1.5.2 Sieving   Sieve analysis is performed using a nest or stack of sieves, where each lower sieve has a smaller aperture size than that of the sieve above it. Sieves can be referred either by *aperture size* or by *mesh size*. The aperture size, generally in microns, refers to the actual size of the hole, while the mesh size is the number of wires per linear inch. This methodology is a useful compliment to the laser diffraction counterparts, as it gives direct measurements, not relying on algorithms.

8.5.1.5.3 Sedimentation Distribution   The particle size distribution of fine powder can be determined by examining a sedimenting suspension of the powder. This method depends on the fact that the terminal velocity of a particle in a fluid increases with size. Stokes's law defines the rate of sedimentation. This technique is not commonly used in the pharmaceutical industry.

8.5.1.5.4 Laser Light Scattering   The most widely used technique involves laser light scattering (0.02–2000 μm), while nano-sized particles (1–5 μm) can be analyzed by photocorrelation spectroscopy. Both methods contrast with sieving and sedimentation distribution, in that they are volume distribution based, as distinct from weight distribution based. These techniques are capable of producing fast analyses of wet, dry, or dispersed/suspended powders, although microscopic techniques should be used to compliment, as the algorithms assume the particles to be perfect spheres. Attachments, such as the Scirocco 2000, in the case of Malvern, can be used for analyses at pressures up to 4 bar, causing deagglomeration, allowing examination of the primary particles only. The Sympatec Helos-Rodos system is capable of the same manipulation.

8.5.1.5.5 In-Line Technology   Lasentec focused beam reflectance measurement (FBRM) systems and, to a lesser extent, particle visualization measurements (PVMs) systems are widely used to analyze crystallization processes *in situ*. By measuring the chord length, FBRM is capable of acquiring data every few seconds, resulting in close to continuous distribution measurements. The software also allows chord lengths to be grouped, allowing trending of different particle populations, throughout a crystallization. This technology can easily be applied to metastable zone width determination, the detection of secondary nucleation events, as well as periods of rapid growth and nucleation and is now considered an essential tool for crystallization development. PVM acts as an *in situ* microscope, allowing detection of, for example, transitory oil phases. These instruments, as well as the second generation Lasentrac systems, can be used on all practical scales, ranging from laboratory glassware to large-scale plant vessels.

8.5.1.6 Particle Shape

This is every bit as important as particle size, when trying to understand solid state characteristics and their relationship to function, in pharmaceutical powders. Knowledge of shapes can help predict blend performance and flow properties.

While microscopic examination is useful for a qualitative picture of particle morphology, Malvern have recently introduced a microscope-/laser-based instrument, called the *MorphologiG3*, which is capable of providing shape information on individual particles, as well as user-defined distribution data, for chosen parameters. Some of the parameters measured include the following:

- **Aspect ratio**: width divided by length;
- **Circularity**: a measure of the closeness to a perfect circle;
- **Covexity**: a measure of the surface roughness.

8.5.1.7 Pycnometry

This nondestructive technique is not widely used in the pharmaceutical industry [54]. Its main application is the determination of the true density of a solid particle, although bulk density, which takes into account entrained air, is a more ubiquitously measured parameter. It can also be used to measure open pores and is unsuitable for materials that agglomerate or possess closed pores.

8.5.1.8 Surface Area (BET)

Since bioavailability can, to some extent, depend on the surface area, this parameter is often studied [55]. The BET technique, named after its originators, Brunauer, Emmett, and Teller, which involves measuring surface gas absorption following a degassing procedure, uses the BET equation to provide values measured in square meters per gram. Usually, measurements in the range $0.01–2000\,\mathrm{m}^2\,\mathrm{g}^{-1}$ are recorded for pharmaceutical powders.

8.5.1.9 X-ray Tomography (XRT)

This methodology allows mapping of crystal imperfections on particle surfaces [56]. In the pharmaceutical context, although not yet widely used, it is potentially applicable to multicomponent granulates, where pockets of crystalline populations can be mapped. This has the potential to compliment inverse gas chromatography (iGC), allowing the possibility of understanding the distribution of an active or excipient, within the granulate matrix.

8.5.2
Bulk

The main, but not exclusive, property one is concerned with when assessing bulk pharmaceutical powders, is flow. Compressibility, a parameter that affects how well a material will compress into a tablet, is another key bulk property and this is related to powder flow.

Flow is a function of the particle size, shape, charge, surface roughness, and other particulate characteristics, which profoundly influence the degree of entrained air, interparticle friction, and degree of cohesion in the bulk and adhesion to the surface of the holding vessel. It is clear, therefore, that crystallization, comminution, and blending processes can greatly influence flow properties. Once again, the connection between the activities of the drug substance manufacturing plant and the performance in the formulation process is noted, emphasizing the often understated reality of the drug substance–drug product continuum.

Poorly flowing powders tend to be cohesive and, most of the time, are highly aerated, while good flowing powders are not cohesive and tend to be less aerated. As mentioned above, the bulk properties often originate from the nature of the particles present. The following sections summarize the techniques that are used to measure bulk properties in the pharmaceutical industry.

8.5.2.1 Angle of Repose, Carr's Index, and Hausner Ratio

These classical, non-instrument-based methods can be used to obtain empirical flow and compressibility information on pharmaceutical powders [57]. These are easily and quickly measured with simple laboratory apparatus.

When a powder is poured onto a horizontal surface, the angle made between the resulting conical pile and the surface is called the *angle of repose*. Free-flowing powders tend to have lower angles or, put another way, tend to form a lower pile. Conversely, cohesive powders will form a more upright pile.

Carr's index and Hausner ratio are two closely related, empirically derived methods used for assessing compressibility. The bulk density of a powder (V_B) is easily measured by weighing a known volume. Tap density, (V_T), can then be obtained by tapping the material, where some compression is likely to occur. The extent of this compression is obtained in the Carr's Index, C_I, which is given by the expression

$$C_I(\%) = \frac{V_B - V_T}{V_B} \times 100$$

General conclusions about flow, which is closely connected to compressibility, can be gauged from the calculated C_I value and these are summarized in Table 8.2.

Table 8.2 Powder flow performance, as predicted by C_I values.

C_I value (%)	Powder flowability
5–15	Excellent
12–16	Good
18–21	Fair to passable
23–25	Poor
33–38	Very poor
>40	Extremely poor

The Carr's Index is related to the Hausner ratio, *H*, by the following expression:

$$C_I = 100 \times (1 - 1/H)$$

8.5.2.2 Dynamic Mechanical Analysis (DMA)

This technique is widely applied in the polymer industry and is, at least, beginning to undergo evaluation, for pharmaceutical powder analysis [58–60]. The basis arises from the application of a controlled stress, as a sinusoidal deformation, to a bulk powder. This can provide information on the nature of any elastic behavior, which may be inherent to the powder. It is also capable of measuring the glass transition (T_g), a parameter which can also be obtained from differential scanning calorimetry (DSC) analysis. T_g, which normally applies to polymers or pharmaceutical powders that can access an amorphous phase, is the temperature below which the phase transition to the glassy, amorphous phase occurs.

8.5.2.3 Dry Powder Rheology and Dynamic Avalanching

Instrumentation is available to provide more quantitative information on powder flow. For example, Malvern and Quantrachrome offer commercial solutions for powder rheological analysis.

The Freeman FT4 is well-publicized as an instrument capable of measuring fundamental flow properties. Six very accessible measurements are easily obtained by a basic treatment using this technology and these can provide quite a lot of fundamental information. These key measurements are as follows:

Basic Flowability Energy (BFE): The energy, measured in millijoules, needed to displace a conditioned and stabilized powder at a given flow pattern and flow rate.

Stability Index (SI): The factor by which the BFE changes during repeat testing (BFE of last test/BFE of first test). Seven tests are normally used during this initial conditioning. A value close to 1 suggests a stable, homogeneous powder, largely free from lumps and other perturbations.

Flow Rate Index (FRI): The factor by which the flow energy requirement is changed, when the flow rate is reduced by a factor of 10. This gives information on powder flow over a range of conditions. A value close to 1 represents ideal behavior.

Mass Change Ratio (MCR): Change in sample mass following consolidation (at a constant volume). Consolidation can be by direct pressure (mimicking charging operations) or tapping (mimicking transport/storage).

Aeration Ratio (AR): The factor by which the BFE is changed by aeration. Highly cohesive, aerated powders will entrain less air than nonaerated counterparts during this test, resulting in a lower change to the BFE value and therefore, a lower AR.

Compaction Index (CI): The factor by which consolidation (direct pressure or tapping) affects the BFE. This can provide information on compressibility and likely tablet performance.

Figure 8.4 BFE profiles for five API batches.

In addition to providing fundamental information, the technique can also be used to study batch to batch variation [61]. Figure 8.4 graphs BFE profiles for five batches, three of which were crystallized under subtly different conditions to the other two. The "3+2" split is very clear, as indicated by the profiles, emphasizing the utility of the technique for detecting subtle bulk changes, arising from slightly different crystallizations. These measurements were correlated with bulk density and particle size distribution.

A variation on this theme is "dynamic avalanching" where the behavior of the powder in a rotating drum can be used to understand powder flow properties [62]. Although not yet widely used in pharmaceutical analysis, the technique is receiving growing attention from the scientific community. A case study is described, where the antibiotic, Cefaclor, is studied by this technique to assess fines content [63]. In this example, knowledge of fines content is important as large amounts negatively impact the tablet production process by diminishing flow properties via the forces of agglomeration and cohesion, ultimately leading to nonuniform drug dispersion in the formulation blend.

8.5.3
Blends

Drug substance powders are typically blended with other powders in the first stage of formulation. In relation to bulk properties, the consistency of the input materials is critical to the quality of the blend produced and the ultimate efficacy of the final drug product, irrespective of the formulation used. In relation to blended powders, the overriding quality concern is normally blend uniformity, and specific techniques can be utilized to determine these key characteristics.

In addition to the techniques already discussed, there is particular interest in a number of in-line techniques that offer innovative advantages, as they facilitate the manufacturer to move away from the traditional approach of blending the constituent materials for a fixed period of time and speed. This has suffered from the weakness that it could not adapt to variability in the physical characteristics of the blend constituents. In-line monitoring of blending operations generates real-time measurements, which facilitates the achievement of processing consistency. It is also consistent with the US FDA draft guidance on PAT.

8.5.3.1 Near-IR
The most common PAT technique used for determining blend end point is in-line NIR spectroscopy. By building simple models, organizations such as ABB and Perkin Elmer can deliver validated solutions that are acceptable to the needs of the regulators. This technology, in conjunction with experimental design, has been used to study the effect of processing conditions such as humidity, blender speed, and component concentration on lactose–salicylic acid blends, in a V-blender [64, 65].

8.5.3.2 Thermal Effusivity
Thermal effusivity is an inherent property of materials, which is generated from three key parameters, namely, thermal conductivity, heat capacity, and density [66]. Each powder has its own thermal effusivity value as determined by the formula

$$\text{Thermal Effusivity} = \sqrt{k}\, r\, c_{\text{p}}$$

where

k = thermal conductivity ($\text{Wm}^{-1}\,\text{K}^{-1}$),
r = density (kg m^{-3}),
and = c_{p} heat capacity ($\text{J kg}^{-1}\,\text{K}^{-1}$).

In effect, thermal effusivity differentiates between solids or powders, on the basis of the level of heat transfer that is occurring in the mixture. As a result, it has the potential for in-line analysis of blending operations, where sensors can be fitted directly onto blending equipment. The sensors monitor the blend mixture and determine when the powder has come to a ''steady state,'' which indicates that blend uniformity has been achieved. Although a relatively new technique, it has already shown promising results in actual manufacturing operations.

8.5.3.3 **Laser Light Scattering**

The techniques discussed in Section 8.5.1.5.4 can also be utilized in real time, although this is typically done off-line in blending operations.

8.6
Conclusions

This chapter endeavors to explain the link between drug substance properties and drug product performance, set against the backdrop of opportunities for innovation, by improving efficiencies in the drug substance–drug product interface. Academic and industrial examples are presented to illustrate the manufacturing and analytical techniques.

We believe that over the coming decades, techniques providing material characterization will become commonplace in drug substance manufacturing and research environments. This will be driven by the ever-increasing importance of supergenerics and will, in turn, be accompanied by a greater interaction between the synthetic chemists and the formulators as opportunities to increase efficiency are consolidated.

Acknowledgments

References

1. Campbell, C. (2005) Speciality Chemicals Magazine, **25** (5), p. 20.
2. Docherty, R. (2008) Driving innovation, control and performance improvement on the critical path- the pivotal role of particle and powder technologies in dosage form manufacture. (Pfizer), 13th Arden House European Conference.
3. Campbell, C. (2008/2009) Manufacturing Chemist, p. 31.
4. Veintemillas-Verdaguer, S. (1996) *Prog. Cryst. Growth Characterizat. Mater.*, **32**, 75.
5. Amidon, G., Lennernas, H., Shah, V., and Crison, J. (1995) *Pharm. Res.*, **12**, 413.
6. Dearden, J. (2006) Pharma IQ Conference, Improving Solubility & Permeability in Drug Candidates, University of Liverpool.
7. Murti, Y. (2008) *Pharmainfo. Net.*, **6** (3).
8. Un-published proprietary know-how, Clarochem (Ireland) Limited.
9. Doherty, M.F., Variankaval, N., and Cote, A.S. (2008) *AIChE J.*, **54** (7), 1682–1688.

10. Campbell, C. (2008) Irish Chemical News, p. 24.

11. Kawashima, Y., Okumura, M., and Takenaka, H. (1982) *Science*, **216**, 1127.

12. Kachrimanis, K., Ktistis, G., and Malamataris, S. (1998) *Int. J. Pharm.*, **173** (1–2), 61.

13. Bhadra, S., Kumar, M., Jain, S., Agrawal, S., and Agrawal, G.R. (2004) *Pharm. Technol.*, **28** (2), 66–76.

14. Sjöström, B., Kronberg, B., and Carlfors, J. (1993) *J. Pharm. Sci.*, **82** (6), 579.

15. (a) Ruecroft, G., Hipkiss, D., Ly, T., Maxted, N., and Cains, P. (2005) *Org. Process Res. Dev.*, **9**, 923; (b) Ruecroft, G. (2006) *Manuf. Chem.*, **77** (6), 41–43; (c) Ruecroft, G. (2008) *Speciality Chem. Mag.*, **28** (5), 54–57.

16. Zalosh, R., Grossel, S., Kahn, R., and Sliva, D. (2005) *Chem. Eng. Prog.*, **101**, 29.

17. Müller, R.H. and Keck, C.M. (2004) *J. Nanosci. Nanotechnol.*, **4** (5), 471–483.

18. York, P. (1999) *Phasm. Sci. Technol. Today*, **2** (11), 430–440.

19. Rasenack, N. and Muller, B. (2004) *Pharm. Dev. Technol.*, **9** (1), 1.

20. Martin, A. and Cocero, M. (2008) *Adv. Drug Deliv. Rev.*, **60**, 339.

21. Rehman, M. and Kippax, P. (2003) *Pharm. Technol. Eur.*, **15** (8), 34–39.

22. Marciacq, F. and Freiss, B. (2008) *Speciality Chem. Mag.*, **28** (5), 58–59.

23. Perrut, M., Jung, J., and Leboeuf, F. (2005) *Int. J. Pharm.*, **288**, 3.

24. Steckel, H., Rasenack, N., Villax, P., and Muller, B. (2003) *Int. J. Pharm.*, **258**, 65.

25. Steckel, H., Pichert, L., and Muller, B. (2004) *Eur. J. Pharm. Biopharm.*, **57**, 507.

26. Steckel, H., Rasenack, N., and Muller, B. (2003) *Eur. J. Pharm. Biopharm.*, **55**, 173.

27. Kharb, V., Bhatia, M., Dureja, H., and Kaushik, D. (2006) *Pharm. Technol.*, **30** (2), 5–10.

28. Hecq, J. (2005) *Int. J. Pharm.*, **299** (1–2), 167–177.

29. Yang, J., Young, A., Chiang, P.O.-C., Thurston, A., and Pretzer, D. (2008) *J. Pharm. Sci.*, **97** (11), 4869–4878.

30. Swaminathan Iyer, K., Raston, C., and Saunders, M. (2007) *Lab Chip*, **7**, 1800.

31. Bhavna, M., Ahmad, F.J., Mittal, G., Jain, G., Malhotra, G., Khar, R., and Bhatnagar, A. (2009) *Eur. J. Pharm. Biopharm.*, **71**, 282.

32. Jacobs, C., Kayser, O., and Muller, R.H. (2001) *Int. J. Pharm.*, **214**, 1–2.

33. Scholer, N., Krause, K., Kayser, O., Muller, R., Borner, K., Hahn, H., and Liesenfeld, O. (2001) *Antimicrob. Agents Chemother.*, **45** (6), 1771.

34. Raisuddin, A., Jain, G., Zeenat, I., Talegaonkar, S., Pandit, P., Sule, S., Malhotra, G., Khar, R., Bhatnagar, A., and Ahmad, F. (2009) *Nanomedicine*, **5**, 55.

35. Liversidge, G. and Cundy, K. (1995) *Int. J. Pharm.*, **125**, 91.

36. Press release, Elan Corporation, September 30 (2003).

37. Rabinow, B.E., Gupta, P., Wong, J., Papadopoulos, P., and Chaubal, M. (2004) *Proc. Am. Assoc. Cancer Res.*, **45**, (abstract #632).

38. Vollmer, R. (2007) (JRS Pharma), Excipient Fest Ireland.

39. Harnby, N. (2006) *Chem. Eng.*, (783), 32–34.

40. Muzzio, F., Goodridge, C., Alexander, A., Arratia, P., Yang, H., Sudah, O., and Mergen, G. (2003) *Int. J. Pharm.*, **250**, 51.

41. Muzzio, F., Robinson, P., Wightma, C., and Brone, D. (1997) *Int. J. Pharm.*, **155**, 153.

42. Gupta, A., Peck, G., Miller, R., and Morris, K. (2004) *J. Pharm. Sci.*, **93** (4), 1047.

43. Ghorab, M., Chatlapalli, R., Hasan, S., and Nagi, A. (2007) *AAPS PharmSciTech*, **8** (1), Article 23. 155–161.

44. Butt, H., Berger, R., Bonaccurso, E., Chen, Y., and Wan, J. (2007) *Adv. Colloid Interface Sci.*, **133**, 91–104.

45. Ralston, J., Larson, I., Rutland, W., Feiler, A., and Keijn, M. (2005) *Pure Appl. Chem.*, **77** (12), 2149.

46. Edwardson, J. and Henderson, R. (2004) *Drug Discov. Today*, **9** (2), 64.

47. Giron, D., Mutz, M., and Garnier, S. (2004) *J. Therm. Anal. Calorim.*, **77**, 709.

48. Shah, B., Kakumanu, K., and Bansal, K. (2006) *J. Pharm. Sci.*, **95** (8), 1641.

49. Heng, D., Tang, P., Cainey, J., Chan, H.-K., Cutler, D., Salama, R., and Yun, J. (2007) *Pharm. Res.*, **24** (9), 1608.

50. Ticehurst, M., Rowe, R., and York, P. (1994) *Int. J. Pharm.*, **111**, 241.

51. York, P., Ticehurst, M., Osborn, J., Roberts, R., and Rowe, R. (1998) *Int. J. Pharm.*, **174**, 179.

52. Grimsey, I., Feeley, J., and York, P. (2002) *J. Pharm. Sci.*, **91** (2), 571.

53. Shekunov, B., Chattopadhyay, P., Tong, H., and Chow, H. (2006) *Pharm. Res.*, **24** (2), 203.

54. Krus, M., Hansen, K., and Künzel, H. (1997) *Mater. Struct.*, **30**, 394.

55. Brunauer, S., Emmett, P., and Teller, E. (1938) *J. Am. Chem. Soc.*, **60**, 309.

56. Barrera-Medrano, D. (2007) *Eur. Congr. Chem. Eng.*, **6**.

57. Gibson, M. (2001) *Pharmaceutical Preformulation and Formulation: A Practical Guide from Candidate Drug Selection to Commercial Dosage Form*, CRC Press, Boca Raton, FL. ISBN: 1-57491-120-1.

58. Menard, K. (2008) *Dynamic Mechanical Analysis: A Practical Introduction*, 2nd edn, CRC Press.

59. Turi, E. (1997) *Thermal Characterisation of Polymeric Materials*, 2nd edn, Academic Press.

60. Gabbott, P. (2008) *Principles and Applications of Thermal Analysis*, Blackwell Publishing, Oxford.

61. Campbell, C. (2006) *Speciality Chem. Mag.*, **26** (10), 30–32.

62. Lavoie, F., Cartillier, L., and Thibert, R. (2002) *Pharm. Res.*, **19** (6), 887.

63. Bonfichi, R., Cloralio, G., and Rainoldi, A. (2009) *Pharm. Technol.*, **21** (7), 26.

64. El-Hagrasy, A. and Drennen, J. III (2006) *J. Pharm. Sci.*, **95** (2), 422.

65. Arwa, S., El-Hagrasy, A., Delgado-Lopez, M., and Drennen, J. III (2006) *J. Pharm. Sci.*, **95** (2), 237.

66. Léonard, G., Bertrand, F., Chaouki, J., and Gosselin, P.M. (2008) *Powder Technol.*, **181** (2), 149.

9
Process Understanding Requirements in Established Manufacture

Dylan Jones

9.1
Introduction

This chapter discusses the requirements for *process understanding* as it pertains to continuous manufacturing of active pharmaceutical ingredients (APIs). It looks at these requirements from the perspective of *process analytical technology* (PAT) [1]. It goes on to discuss the practicalities of implementing these concepts into commercial production.

Superficially, it is all very simple. For continuous processes there is a requirement to understand the process, understand the measurement, and understand the control element. The more lengthy discussion is what exactly ''understand'' means in the context of maintaining regulatory compliance for the pharmaceutical industry.

There are many disciplines that fall somewhat within the scope of this chapter. These include analytical chemistry, engineering, information technology, automation, quality, and regulatory compliance. They are closely interwoven, and in many respects it makes more sense to study them together rather than in isolation.

Before launching into the discussion, it is worth making two points. First, it should be appreciated that there is variation in all the manufacturing processes: variation in the raw materials (inputs), variation in the process (process parameters), and therefore variation in the finished product (outputs). The objective is not so much to eliminate variation, as to manage it adequately and in the best interests of the patient.

Second, throwing the whole PAT toolbox as described herein is not presented as the best solution for all processes. Certain elements can be useful for certain applications under the right business circumstances. If implemented well, they bring tangible benefits to both the manufacturer and the patients.

In established manufacture, PAT falls conveniently into the sphere of c*ontinuous improvement*, and is a powerful vehicle for innovation. In the future, as an element of *quality by design* (*QbD*) [2], PAT may be central to a new generation of regulatory submissions.

At the time of writing, PAT concepts are still developing and evolving. Much of the regulatory guidance is still being drafted. Some of the ideas herein are prospective, and the author regrets that if this chapter seems very much ''of its time,'' it is unavoidably so.

9.2
The Status Quo

The big question for a patient on medication is: how do you know whether or not what you are taking is what it says on the box? The answer is you do not, and have to take a lot on trust.

In 1974, Ted Byers of the Food and Drug Administration (FDA) presented a paper entitled ''Design for Quality'' [3]. He suggested that validation of processes used in the manufacture of parenterals would be a good idea. Since the adoption of these ideas, the pharmaceutical industry has applied validation to all the manufacturing processes and has attempted to ensure final product quality through minimizing the variation in the raw materials and clearly defining (or fixing) the operating parameters of the process.

If you can demonstrate enough control of the process to successfully manufacture three ''good'' batches in a row, you fix all the settings within the range of tolerances you have explored and call it validated. Ultimately, the process needs only be understood to the extent that it can be made to satisfy this three-batch criterion, and in practice this is probably what happens.

While this regulatory approach has been a success, it does have three key weaknesses:

- First, it limits opportunity to capture *process understanding* to a relatively narrow period of time.
- Second, once validation has been performed and marketing approval has been given by the regulators, the process is fixed and cannot later be easily changed.
- Third it has contributed to a culture of conservatism that effectively stifled innovation in pharmaceutical manufacturing for decades. As a result, pharmaceutical processes tend to be inefficient, with low levels of hardware utilization, poorly controlled (two to three sigma), and wasteful (on average, 5–10% of any product is reworked or discarded) [4]. Compare this with productivity gains enjoyed by industries that have adopted innovation in manufacturing, such as semiconductor manufacturing which routinely achieve Six Sigma processes (i.e., a 0.00034% defect rate).

When process development is confined to a relatively short period, it is not possible to account for long-term trends and unforeseeable step changes. It does not allow an accurate assessment of *process capability* or, worse, creates an optimistic impression of it.

But products cannot stay in development indefinitely. Long lead-times in development are expensive, patent restrictions provide a strong incentive to get the

product to market as soon as possible, and if it is a treatment for a life-threatening condition then any delays will inevitably have a human cost. Against these constraints, the ability to demonstrate the control of a process three times in a row is as good a measure of success as any.

The practice of ''fixing the process'' must be appreciated in the context of the way final release testing is used to assess the product quality. If a product meets its specifications, you can sell it; if it fails, you have to throw it away. The issue is that it is retrospective and the quality judgment can only be made once all the value-adding activities have been performed and all the time or money has been spent.

There are a number of consequences to this strategy later in the product life-cycle. When unforeseeable change happens post regulatory approval (e.g., in some property of a raw material) and the product fails to meet its specifications, the defined process may need to be formally changed to compensate. The regulatory burden associated with such changes can be considerable. They are known in the industry as post-approval variations and it is a very expensive practice.

There is a consensus across the industry that there is a lot of unnecessary wastage arising from these practices. The FDA is of the opinion that too much of its time and resource is allocated to processing post approval variations and supplements [5]. Clearly, not enough quality is being designed into processes during development. By implication, final release testing is the primary safety mechanism protecting the patient – and it has its limitations [6]. What is to be done?

If you can accept that variance in the raw materials (inputs) is inevitable, one way to resolve this is to learn how to vary process parameters in response to these variable inputs, with the expectation of producing consistent product (fixed outputs).

In order to achieve this requires up-to-date information – sufficient analyzers in place to detect when the change is happening (process knowledge). Once change is detected, a plan (control algorithm) is required to alter the process to achieve the desired outputs (prediction). That plan represents your *process understanding*. This is the essence of PAT, and it is eminently suited to continuous processing.

9.3
Risk and Reward

Putting analyzers into a manufacturing process requires some understanding of the commercial risks. The challenge with continuous multivariate analysis and monitoring can be summarized thus:

What happens when you start to look at some aspect of your process and start seeing things you did not anticipate, or cannot explain, or cannot do anything about? How will you handle that information, and how will you estimate its significance?

If you ask the open question of whether there is any variability in the process, the mathematical answer will always be ''yes''. What will you do with that answer?

Even if you choose your questions carefully, there are pitfalls. If you ask the closed question of how much change is there in a given variable, the answer you receive will give you confidence: this much change; a quantifiable measure that can elicit an appropriate response. However, when you take multiple measurements you will (almost certainly) get different answers and your confidence will evaporate rapidly. Fortunately, it is when you take enough measurements and you start to understand the distribution of values that a realistic picture of the process emerges.

You are then afforded the opportunity to assess whether that variability is real and if it is, whether it is significant. Appreciate the need to have mechanisms in place to help guide you through these thought processes. You need a workable response to all contingencies to prevent you from painting yourself into any corners. Once you have the rationale in place, then process variability becomes manageable.

The management of risk and uncertainty is a recurrent theme in PAT. We are continually drawn back to the question of how it pertains to the potency or efficacy of the drug and ultimately to the safety of the patient.

That PAT finds variability does not initially strengthen the case for taking such an approach, but as the *process understanding* matures the ability to measure and hence control change becomes a powerful tool in the mitigation of risk. It reduces the risk of the process going out of control, the risk of that loss of control not being detected, and then the risk of relying only on finished product testing to detect these failures before the product goes for sale.

9.4
Terms and Definitions

A detailed consideration of the subject matter is contained within the ASTM International (formerly American Society for Testing and Materials) standards written specifically for PAT and QbD. These include ASTM E2363, E2474, E2500, and E2537. A selection of some of the most relevant concepts contained therein is provided below:

9.4.1
PAT

PAT is a system for designing, analyzing, and controlling manufacturing through timely measurements of critical quality and performance attributes of raw materials and in-process materials and processes with the goal of ensuring final product quality (ASTM E2363) [7].

9.4.2
Process Understanding, Critical Quality Attributes, and Critical Process Parameters

These terms hold particular meanings and significance to the FDA. Basically, process understanding is the ability to identify critical quality attributes (CQAs)

in your raw and in-process materials and critical process parameters (CPPs) in the manufacturing process and somehow relate them to properties in the finished product that have an impact on drug potency, efficacy, and patient safety. Normally this means the product release specifications.

9.4.3
Quality by Design

QbD is an alternative approach to the three-batch validation strategy for submitting a new drug application to the FDA. In this approach, the applicant endeavors to build in quality from the development phase and to continue throughout the product life-cycle. It requires the ability to demonstrate a sufficient level of *process understanding* to the regulators to ensure delivery of product that provides the required effect and is safe for the patients.

9.4.4
Design Space

Design space and operating space are concepts critical to PAT in continuous manufacture. Design space is the range of variables over which relationships can be independently verified; and operating space is a further subset within which ranges of variables your process makes good product.

9.4.5
Design Space as Applied to Spectral Analyzers

The concept of design space can be used to define the range over which a multivariate calibration (MVC) is suitable for use (e.g., the concentration range for a given analyte). Plant-based instrumentation may also be exposed to greater changes in environmental and sample conditions than are typically encountered in laboratories. The spectra from in-line measurements will contain not only the relevant information but also irrelevant variations. MVCs need to be able to compensate for such variations as temperature or humidity at the time of measurement. Finally, there will be a certain amount of random noise and possibly even drift associated with the instrument hardware.

9.4.6
Fitness for Purpose

This is a formal assessment of suitability to perform the activity for which it is intended. This may be applied to variously product, process, systems, and components.

9.4.7
Spectral Analyzers

Real time or near-real time analysis can be achieved by taking in-line measurements of materials. Techniques such as Ultraviolet, Raman, Mid and Near Infrared spectroscopy are ideally suited to such applications, particularly when used with MVCs.

9.4.8
Multivariate Calibrations

MVC is defined as an analyzer calibration that relates the spectrum at multiple wavelengths or frequencies to the physical, chemical, or quality parameters. The multivariate model is a mathematical formula that calculates these parameters from the measured spectrum (ASTM E6122) [8].

9.4.9
Process Capability

Process capability is a statistical estimate of the outcome of a characteristic from a process that has been demonstrated to be in a state of statistical control (ASTM E2281) [9]. In simple terms, for a given process it is a probability of a unit of product being out of specification.

9.4.10
Process Knowledge

Process models may be empirical or mechanistic in nature, and occasionally even derived from first principles. In addition to an understanding of relevant mathematical processing techniques, there are considerations of the design space of models and the process knowledge available at the time. That knowledge may be hard, such as data, or soft, such as the experience of an operator.

9.4.11
Continuous Quality Verification

This is a PAT concept. If implemented, it would become a company's program for quality management throughout the product life-cycle. It is not only concerned with ongoing evaluation of the performance of the process, acceptance criteria, and product release but also provides a mechanism to facilitate feedback of continuous learning into continuous improvement (ASTM E2537) [10].

9.5
Process Understanding Requirements

9.5.1
Start with the End in Mind

''A process is generally considered well understood when (i) all critical sources of variability are identified and explained; (ii) variability is managed by the process; and (iii) product quality attributes can be accurately and reliably predicted over the design space established for materials used, process parameters, manufacturing, environmental, and other conditions. The ability to predict reflects a high degree of process understanding. Although retrospective process capability data are indicative of a state of control, these alone may be insufficient to gauge or communicate process understanding.'' *Guidance for Industry PAT* [1].

This brings us to a concept in PAT known as the *desired state*. ''In the desired state of a process, all sources of variation are defined and controlled and the end-product variation is minimal. That implies that critical control attributes are controlled to target for all individual units of a product. As a result, processes are capable of consistently supplying, unit to unit and batch to batch, the desired quality.'' *ASTM Standard E 2474* [11]. I would qualify this statement to critical sources of variation. The ability to calibrate for levels of risk in the context of patient safety is important. If all uncertainty is treated equally, it is possible to quickly lose sight of the factors that actually matter.

Batch manufacturing makes use of a concept called ''the *golden batch*.'' This refers to a batch whose manufacture was particularly successful (by a combination of luck and design), and whose process signature becomes the standard against which all other batches are contrasted. This usually has limited value in continuous processing except in unit operations where the process trajectory is largely defined by events in its early stages (the start-up phase). Chaos theory says that models of this nature are characterized by a particular sensitivity to starting conditions that may be difficult to reproduce. This phenomenon may be particularly true of biological systems, such as during the start-up phase of bioreactors.

In order to develop *process understanding* of a unit operation you need to have data. That implies a need for sufficient analyzers to capture change when it occurs. Given the complexity and variability of in-process conditions it is common to interpret analyzer data using MVCs. In addition, for automation there is the not-so-small matter of the control system.

So, to reiterate, from the PAT perspective, *process understanding* is being aware of the important variation and uncertainty that exists in the process (process knowledge), understanding how it relates to the final release specifications (prediction), and having taken a measurement, being able to take an appropriate action (control), if required.

In the context of commercial manufacturing, this definition is still incomplete. There is also a requirement to understand the performance characteristics of the

analyzers and the behavior of the control system. These will both contribute to variation and uncertainty.

Moreover, to develop a robust process there is a requirement to understand how equipment breaks down, how chemical processes fail, how faults may be detected, and then knowing what to do about it when it happens (remediation).

Bear in mind that if a process begins to drift or goes out of control, then it will be due to any of three factors: (i) the variability is real, (ii) the process hardware is changing (e.g., breaking down), or (iii) the analyzer is failing (misinformation). All these eventualities must be accounted for.

9.5.2
General Considerations

PAT is a multidisciplinary effort that requires input from a number of organizational functions (Research and Development, Information Technology, Regulatory, Engineering) and will have an impact on the behavior and work practices of a number of others (Production and Quality). It is important to get conditions right that will allow the technical aspects of any project to flow smoothly. How well these disparate requirements are collated and addressed in a coherent manner is the true measure of the success of a project.

9.5.2.1 Regulatory
Any development group's first customer is usually regulatory compliance. The consequence of global product filings means that the documentation will be reviewed by many agencies, and the views of each of these agencies regarding MVCs may vary. Submission of variations to the regulatory authorities comes with substantial administrative costs.

9.5.2.2 Information Technology
The abundance of unprecedented amounts of raw data generated by spectral analyzers in particular will place a burden on the IT infrastructure. The collection, transmission, archival, retrieval, and reprocessing of spectral data in a manner that satisfies good manufacturing practice (GMP) may not be trivial to achieve. The simplest solution is to treat the spectral analyzer as though it were any traditional process measurement. The PC associated with the spectral analyzer is a stand-alone system that provides a 4–20 mA output, much like a pH meter or temperature probe. However, for fully integrated systems, the International Society of Automation's ISA 95 [12] provides a workable implementation strategy.

9.5.2.3 R&D/Engineering
During the development of a continuous manufacturing unit operation destined for automation (closed loop control), at least four lines of development will occur in parallel. The chemical engineer and chemist will design the process scale-up and will manage the effects of these changes on the chemistry; the automation engineer will begin to consider the control philosophy; and the PAT scientist will

attempt to develop the in-line assays. A fair amount of iteration and recursion is to be expected.

9.5.2.4 Quality Assurance

Given the availability of up-to-date in-process information, integration of the MVC with QA systems to give assurance that the requirements of the product license have been fulfilled is an important consideration. Consider how calibration updating and maintenance fit into the existing change control procedures, how this is to be documented.

9.5.2.5 Production

This requires an understanding of how analytical methods may be deployed in practice. In addition to routine use, this requires consideration of how the calibrations react to exceptional process scenarios, and indeed how the process and operators then react to that information. The ongoing obligations that arise from the implementation of spectral analyzers need to be clearly understood.

9.5.3
Characteristics of Continuous Processes

In essence, a given unit operation in a continuous process is concerned with moving the material from point A to B, while it undergoes a (bio) chemical transformation, within a set of controlled conditions. Continuous processes usually require a series of discrete unit operations, with the output of one step becoming the input for the next step. Elements of the operation may even be batch processes, particularly at the start of the process, when the raw materials are introduced. There may also be buffering capacity between unit operations, inserted to ensure continuity by allowing certain activities in the process to be suspended periodically without bringing the whole process to a halt.

9.5.3.1 Phases of Operation

For each unit operation there tend to be three normal phases of operation: start-up, steady-state operation, and shutdown. Each of these phases of operation may need to be considered separately.

Start-up can be the most dangerous period, as the system is out of necessity going through changes in state. The rate and extent of the change must be closely supervised. These will also have some bearing on the exact time at which the next link in the chain of unit operations begins its own start-up activities. Generally, it is undesirable to have out of trend or out of specification material flowing through a continuous process.

The design space and operating space requirements for each of these phases are unlikely to be the same. Some analyzers may only have application in certain phases of operation.

Fortunately, for most of the time the normal operational phase for continuous processes is the steady state.

9.5.3.2 Mass and Energy Balance

Two very basic and crucial elements for understanding unit operations and their operational modes are equations of state. These can be used to account for mass and energy changes that occur within a given system. Mass balance and yield are the most common models for assessing the performance of a phase of operation. How much did I put in, how much did I get out, and how much did I throw away? And then, occasionally, what happened to the rest of it? Understanding reaction enthalpies is important for engineers during process development when designing out the risks of ''thermal runaway.''

To fully understand the fate of each batch of raw material introduced to a continuous process, mathematical models can be derived to estimate mean residence times and the time distribution over which elements of a batch are expected to remain within a system. There are good quality-based reasons for wanting to know this. In the event that a batch of raw material is found to be unsuitable post processing, knowing what product to quarantine, and being able to justify it, are important pieces of information.

9.5.3.3 Fluid Dynamics

The flow of material through continuous processes is a key engineering consideration during plant design. For dilute solutions, it may not be complicated. Complex rheology can occur where a solution is very concentrated, contains polymeric molecules, where phase changes occur *in situ*, or where multiple phases are present, for example, slurries or fluidized powder. The rheological profile of the material may have an enormous impact on processing behavior, the reaction kinetics, and therefore process design. Mixing may require static mixers, baffles, or oscillatory flow depending upon the required intimacy of the mixing, and whether laminar or turbulent flow is desirable.

While this subject falls outside the scope of this chapter, such techniques as computational fluid dynamics and tomography can be employed to create mathematical models to improve understanding for process optimization. This type of process understanding (refer to Chapter 10) tends to manifest itself in the process hardware and has less to do with demonstrating ongoing regulatory or quality compliance.

Fluid dynamics is relevant where a complex rheological profile is highly characteristic of a process, and bulk properties provide useful (inferential) measurements for control.

9.5.4
Measurable Variation in a Process

Finding measurable variation in a process is easy to do, but not so easy to do well. A variable should satisfy three conditions: it needs to matter; it needs to be something that you can do something about, and it should be predictive of a final release specification.

Measurements of process variables (temperatures, pressures, flow rates) are usually freely available, and when contemplating control strategies these are often a good place to start.

The analysis of waste streams and ancillary recycling operations may afford a good opportunity to study process variability, while neatly sidestepping some of the procedural issues around maintaining the GMP-validated status of a commercial manufacturing plant. For example, the immersion of probes into pharmaceutical concoctions may introduce unacceptable risk.

Indirect measurements such as conductivity and pH, if effective, are inexpensive. The problem with indirect measurements is that they find variation, but may not tell you what it is or how relevant. They need to be interpreted in the context of *process understanding* but may not, of themselves, supply it.

If none of the above suggestions prove helpful, then you may be forced down a route of identifying what *critical quality attributes* are impacted, finding some means of measuring them directly with an in-line assay, establishing how they relate to the final release specifications, and finally learning how to control them with the *critical process parameters.* This is the true realm of PAT.

9.5.5
Uncertainty in the Analytical Measurements

9.5.5.1 Analyzer Design
Understanding uncertainty in process analyzers is a subject best grappled from the outset: having established ''Fitness for Purpose'' criteria is helpful during analyzer selection.

The question will arise as to what performance is required of a unit operation to ensure it stays in control and meets the desired specifications. What are the CQAs and is an in-line assay really needed to measure them? If so, exactly how good does the measurement have to be? What criteria matter and what should be their specifications? For processes in early development, this can be a challenging question. ASTM E2500 [13] refers to this as ''design review'' and suggests it is revisited as understanding evolves throughout the product life-cycle.

There are also commercial considerations: cost of ownership and ongoing maintenance are the obvious ones. The trade-offs are usually between cost, speed, and sensitivity. The size of the supplier organization, its commitment, its people, and its ability to provide adequate support, are all useful when assessing risk. Whether they are still likely to be in business in five years is a concern when dealing with novel technology platforms.

From the regulatory perspective, the question is how easily they will navigate the various safety, quality, and regulatory guidance notes, such as ATEX [14], GAMP [15], ICH, and 21 CFR part 11 [16].

Scientifically, four requirements that should be understood quite early on include precision, range, sampling frequency, and sampling size.

9.5.5.2 **Precision**

One thing to bear in mind is that the analytical method is actually only one element of the control system, so the first design review is simply an exercise to get the analytical precision into the right order of magnitude. The philosophy then is to start by setting specifications reasonably loosely. These will allow acceptance criteria for screening technology platforms without introducing the greater risk of accidentally dismissing potential methods too early during development, particularly when experience in deploying the technology is limited.

The requirements for analytical precision can be estimated from the process capability and the process capability index for a given unit operation.

The quality of this early estimation will be dependent upon how much process variability can be estimated while the unit operation is still in the design-phase. Where certain off-the-shelf manufacturing hardware has defined operating specifications, these can be used.

In the first pass, the required measurement capability provides three sigma control of the process, or an estimated process capability of approximately one. These limits can be gradually tightened to 1.3 or 1.6, if required, as the appreciation of the unit operation improves. The relative standard deviation limits for precision of a measurement can be determined from the process capability requirements.

Because of the way the industry operates, moves to loosen specifications will generally be regarded with suspicion and will therefore be resisted, so it is better if they are not set too ambitiously early on.

The use of precision as it relates to measurement capability is a useful mechanism for translating the mathematics of engineers and the ASTM into the specification requirements for validating analytical methods in accordance with ICHQ2 (R1) [17], and the chemometrics in CPMP/QWP/3309 [18].

9.5.5.3 **Range**

The range of a model is constrained by the inherent limitations of the instrument, the reference method, the chemistry of the process (failure boundaries), or the availability of samples. It is inadvisable to cut corners here, as a requirement to change the design space late in the program can meet strong regulatory resistance. As a rule of thumb when asking ''How big should be the design space of my MVC?'' the answer will go along the lines of ''As big as possible.''

Design of experiments (DOEs) is a good tool for maximizing the information content of a model with the minimum number of samples while maintaining the lowest possible correlation between variables. In a perfect world, the creation of calibrations occurs via a series of DOE experiments that explore not only the operating space but also identify the relevant failure boundaries (of both process and analytics).

9.5.5.4 **Sampling Frequency**

This is dependent upon the anticipated rate of change of the process. There is a parameter called the *Nyquist rate* which determines the minimum frequency of

measurements required to correctly interpret and mathematically extrapolate the periodic changes.

How to estimate the periodicity of a process without the benefit of the analyzer in the first place is an interesting challenge. A consideration of flow rates, dwell times, and change volumes may be enlightening. Ultimately it all depends on how dynamic is the system. The measurement should be in keeping with the required responsiveness of the control system.

9.5.5.5 Sample Size

The subject of process sampling can be important. Basically, any in-line assay should ensure representative sampling of the in-process material. Deciding what exactly is representative will be driven by a consideration of control requirements or any patient safety implications, and will need to be performed to the satisfaction of any regulatory dictates.

Consider an analyzer placed in-line in a continuous process stream. The analyzer may measure 100% of the material that flows past, but does so in discrete packets of time. The duration of a measurement multiplied by the rate at which material flows past the sampling point will define the sample size. The measurement is in fact an average of the entire sample. If the sample size is too large, it may ''average out'' real variability, and if too small it may pick up on inhomogeneity at a scale that is quite irrelevant. For example, Raman spectrometers with focused beams can have spot sizes between 5 and 50 µm.

The criterion of patient safety is very situation-dependent. During API manufacture, this can be a difficult variable to relate back to patient safety and more often the control aspects remain the principal consideration. At the other extreme, for example, during secondary manufacture, where PAT is used to estimate blend uniformity prior to tabletting, sample size is an important parameter.

9.5.6
Understanding the Control System

The control system brings its own set of challenges. Some changes are slow, while others are extremely rapid. Understanding is required around how the control system behaves under normal operating conditions and just as importantly how it will react under exceptional conditions. During plant commissioning the dynamics of the process may not be completely understood, so control parameters tend to be set to best estimates. The performance of the control system in commercial manufacturing is monitored and the control parameters can be optimized based on this information. A few points to consider are discussed below.

9.5.6.1 Lag

There may be delays in response to change known as *lag periods*. Basically you decide you need to move the process from one set of conditions to another, so change some process parameters. There will be a period of time between making

the change and seeing a measurable response. Proportional integral differential (PID) control alone may not cope with these situations very well.

9.5.6.2 Oscillations

(Poorly tuned) PID controllers have a tendency to oscillate around the set-point (also called *hunting*). This can occur particularly during start-up. In the worst case, oscillation can occur between sequential unit operations where the periodicity in oscillations in one operation influences the behavior of the controller on the next unit operation. Like two pendulums oscillating at different frequencies on the same string, behavior can become erratic and unpredictable.

9.5.6.3 Tuning

Controllers, particularly PID controllers, will most likely require occasional maintenance. Poorly tuned controllers can introduce oscillations into a process, and reduce performance. This can be observed using Fourier transform to resolve different frequency oscillations and new periodicity is instantly recognizable in the power spectrum.

9.5.7
Failure Modes and Effects Analysis (FMEA)

Eventually something will go wrong with a process. Failure modes and effects analysis FMEA is an attempt to anticipate and quantify the risk of an event occurring based on three criteria: probability of occurrence, ease of detection, and the resulting severity. PAT may be deployed to mitigate such risk elements.

Accurate risk analysis implies good process understanding, so it is worth doing well. For a detailed discussion, refer to ICH Q9 [19]. FMEA provides a good format for summarizing not only how things break down but also what happens when they do, and most importantly, how serious a problem that is going to be. Once this is appreciated, all that remains is to figure out what to do about it. FMEA can be applied to the process, the analytics, or the control system.

Safety elements are managed by *process hazard analysis*. FMEA can be used here, as can a system known as hazard and operability studies (HAZOP) [20]. This is a qualitative ''what if'' assessment of how things might go wrong in a manner that endangers human safety.

Quality risk assessments are used to focus attention on the elements of the FMEA that pertain to the product meeting its final release specifications and the conditions of the marketing authorization to ensure product efficacy and patient safety. These risks should be calibrated in the context of their potential impact (if any) on patient safety.

For example, consider what happens when FMEA is applied to a process analyzer destined for use in a PAT application. Once the CQA (and a suitable place to measure it) in a manufacturing process has been identified, its importance needs to be assessed and the effects of instrument failure understood. Here, FMEA facilitates the examination of the potential downsides to employing process

analyzers. Once the analyzers are in place, the operation will become dependent upon them functioning correctly. Reliability and contingency planning will become an important issue.

If the function of the measurement is sufficiently critical, redundancy in instrumentation can be employed. This could mean having a second analyzer placed in series beside the primary analyzer: if one instrument fails, the second activates.

Having two instruments with the same failure modes in series may not be ideal, though. In these circumstances, consider purchasing instruments from different vendors, or even better, develop two assays based on completely different (or orthogonal) technologies that derive the same attribute from independent aspects of the sample. For example, concentration can be determined by both spectroscopy and density.

In manufacturing operations, reliance on a single supplier has always been a risky strategy. Finally, having orthogonal methods in place allows independent verification of unusual process behavior.

9.6
Method Development and Installation

9.6.1
Starting on the Plant

There are two proven methods of implementing MVCs into manufacturing processes. The first is that described in ASTMD 3764 [21]. Here the probes are installed in the process as the first step and data are collected over a period of time. The design space covers the random variability in the process encountered by these probes over time, as well as any deliberate variability introduced as part of the calibration development exercise. Development is an incremental process and validation is only accomplished once the acceptance criteria for range and performance have been met.

This approach is pragmatic, efficient, and the usual problems of scale-up and transfer are neatly avoided. Unfortunately, an established process is a prerequisite. As regards process understanding, it is unlikely to allow the premeditated exploration of the boundaries of process or analyzer failure.

9.6.2
Starting in the Lab

Development starts in the laboratory and graduates over time into production, often via a pilot plant. This approach requires a much greater investment of time and resources, but has a distinct advantage – there is greater control over the samples and sample variability, allowing better definition of the design space. At the end of the exercise, the calibration curves could be deployed in a variety of scenarios, from in-line analyzers to rapid QC laboratory assays.

PAT-based spectroscopic assays are powerful tools in the laboratory, simply because of the speed and the reduction in reliance on the performance of the analytical chemist. One reciprocal measure of the success of a lab assay method is the amount of erroneous data it generates. ''Out of trend'' and ''out of specification'' investigations are time-consuming, and very expensive. Experience indicates that PAT-based methods tend to reduce the frequency of erroneous measurements.

9.6.3
Scaled-Down Models

When creating calibrations for a PAT application, scaled-down models are very useful. It is important to ensure that the scaled-down model will reflect process (not least in the physical effects on the chemistry) sufficiently to minimize the effort required later for effective calibration transfer and to avoid regulatory challenges.

Laboratory scale models are the most flexible workhorses, but care must be taken to ensure that they adequately model the larger scale equipment in terms of matching physical phenomena, for example, heat and mass transfer. Pilot plants are useful if you have access to them, but they tend to lack flexibility. They remain fairly large scale, are expensive to run, are run intermittently, have a heavy burden on instrument safety and validation, frequently require zoning (see ATEX regulations) and are, by design, focused on scale-up of the chemistry. This does not always suit MVC development.

Scaled-down models can vary massively in their degree of sophistication or indeed the extent to which they are actual scale models or mimics of the process. In the end it is all about weighing up the importance of time, cost, and risk. Below are examples of both extremes.

9.6.3.1   Simple
The simplest scaled-down model of a stirred tank reactor is a round-bottomed flask and a stirrer. It may be deployed in the time it takes to cross the laboratory (rapid prototyping!). A spectral analyzer probe inserted into the flask was used to demonstrate that a certain reaction occurred via the formation of a stable, if short-lived, intermediate. This information may not have been gleaned by at-line sampling and testing. The experiment provided sufficient insight to determine that to convert the batch reaction to a continuous process, the formation and consumption of the intermediate should be handled in separate tube reactors, with each stage optimized independently (Figure 9.1).

9.6.3.2   Complex
There is value in going to the other extreme. If PAT principles are such good tools for productivity and efficiency, then they should provide payback if used in development. Also, if you know how to deploy in R&D, then you know something about what will be required to deploy in production.

To test this notion, a custom-built, scaled-down test rig analogous to a unit operation in the process was assembled. Process analyzers, services (RS232,

Figure 9.1 Identification of a reaction intermediate through real-time monitoring.

Ethernet communications), utilities (heat, water, vacuum, air), and a closed loop control system were built into the design on a small scale.

Initially, it permitted the collection of data *in situ* to build calibration curves. Once predictive calibrations were established, the system could be fully automated, allowing exploration of the design space to continue in real time, and with significantly little user intervention.

The payback for the initial investment in the test rig has been long term.

- It has allowed the ruggedness of analyzers for long-term deployment to be understood even before the production plant was built. It is important to have some practical insights into what problems to expect and be able to weigh risks appropriately.
- It has been used to manufacture on a small scale, batches of material used to create calibration curves for other downstream unit operations that were not part of the original plan. They proved to be great enablers for activities that might otherwise have been too costly, complex, and time-consuming to contemplate.
- The capability to manufacture the ''desired variability'' at the push of a button gives very reliable control of the design-space. Together with DOE this reduces correlation between variables, and generally improves the quality of the MVC.

So to summarize, scaled-down lab models involving automation are useful for understanding relationships, exploring the design space (particularly when creating MVCs), testing control philosophies, and signal conditioning algorithms.

9.7
Statistical Process Control

To demonstrate some level of *process understanding*, information needs to be presented in a format and articulated in language that can be understood by others. Statistical process control (SPC) and multivariate statistical process control (MSPC) provide industry standard mechanisms for data collection, interpretation, and the extraction of knowledge.

9.7.1
Time Series Charts

It is worth reiterating that there are a number of tried and trusted techniques for summarizing and trending data. Sometimes a time series chart is adequate. The variable of interest is plotted against time; the target value, alert limits, and alarm limits may be superimposed to give it context.

Most unit operations will be monitored by charting a series of discrete process variables. This is done even for "fixed processes" that have to tolerate some level of variability, with each CPP independently contained within acceptable limits.

Traditionally, this is a perfectly good means of maintaining control. It begins to fail when significant aspects of a process are influenced by a combination of variables.

9.7.2
2D Plots

Consider a kinetic reaction in a simple tube reactor. It involves mixing two chemicals and allowing them to react together. The process parameters are trended on a time series chart. For some reason, the flow rate tends toward the higher limit while the jacket temperature is at the low end of its limits. High flow reduces dwell time, while low temperature reduces the reaction rate. Refer to Figure 9.2. Taken individually, the variables remain within their alert limits (a,b) and are normally distributed (d). The combined effect, however, is more extreme, possibly reducing the yield to below desirable levels. This particular example may appear trivial, but

Figure 9.2 Comparison of time series and 2D plots.
(Acknowledgments: Dr Ahmed Al-Alawi.)

the message is that you are not likely to spot the event and its causes immediately unless you happen to be looking at both variables simultaneously.

9.7.3
Parallel Coordinate Geometry

What happens when there are more than two variables? Let us revisit the kinetics problem. What if temperature and time are not the only variables that matter? The concentrations and individual feed-rates of the chemicals may vary. Many chemical reactions are pH-dependent in a manner that is quite nonlinear. Let us say that there is also a catalyst to consider and that catalyst quality is variable. Finally, let us say that the catalyst is provided by two suppliers and the activity differs for reasons no one has ever been able to explain. How are you going to de-convolute all of that information and (equally as important) how are you going to convince people that your interpretation is correct?

For small multivariate problems polar coordinates provide an excellent graphical representation (Figure 9.3). Essentially, it is simply a flavor of parallel coordinate geometry.

Figure 9.3 Comparison of polar coordinate and PCA plots.
(Acknowledgements: Giuseppe Elia)

When multiple variables are concerned, parallel coordinate geometry is a useful technique. Its advantages lie in that as a graphical representation of the process it is simple to interpret, and it is computationally light. Irregularities in the sampling frequency and gaps in the data do not pose significant problems. It can also be used to identify ''white holes,'' those areas within the design space where the system simply does not operate.

9.7.4
Multivariate Analysis

If none of the previous approaches are satisfactory, consider multivariate statistics. Principal component analysis (PCA) (in its many flavors) is a powerful technique for reducing the number of variables in a problem and for finding the underlying nature of the relationships between them. However, a certain level of experience and competence is required to use them reliably.

Using a combination of process and material measurements, a multivariate model of any given unit operation can be derived. This model, while not necessarily providing a control function, allows comparisons of the process over time. Does the unit operation still behave in the same way as it did last week, or last year?

Critically, it can be used to diagnose plant data as it is collected in real time. It answers the question: are any of these numbers unexpected? Should the process drift or a sensor fail, this will be reflected in unusual data patterns. There is a concept in ASTM E2537 called *continuous quality verification* and this approach is an efficient, automated means to address the subject. In this scenario, data are scrutinized at every single time they are sampled. This happens in real time and allows data to be effectively ''validated'' before the control system makes a decision on how to respond.

Allied to this, multivariate tools can provide steady state data reconciliation, which can be used for early fault detection and diagnosis. Data tend to be correlated, so one variable can often be predicted from the others. If a given variable is out of trend, there may be grounds to suspect its validity. This too can be achieved in real time.

9.7.5
The Analysis of Noise

So much effort is spent extracting the signal from the background noise that it is easy to forget that residual noise is very characteristic of any PAT system. Its amplitude and frequency are reflection of the way the instrumentation, the process, and the data gathering operations are assembled. Changes in the behavior of process noise are very powerful tools for fault detection.

For example, consider a near infrared spectrometer in transmission mode, shining a beam of light across a solution moving through a tube. Under correct

operating conditions, the analyzer can be used to predict the concentration of one (or even all) of the components in the solution.

The instrument produces its own inherent noise; bubbles in the medium introduce noise, and the process may experience undulations and step-changes over time. Under normal operating conditions, each of these effects will cause the signal to meander in a time series chart.

If the noise decreases below a certain threshold, this tells you that the spectrometer is no longer responding to the process. This is a very robust and inexpensive way of detecting analyzer failure.

Should the noise increase in amplitude, it could reflect real changes in concentration. However, if this behavior is contrary to expectation, there might be cause to be skeptical. What could be wrong?

It could be that too many bubbles become entrained in the medium, creating a transient and trivial event. The scattering could arise from changes (failure) in a physical property such as refractive index due a failure in an upstream mixing system?

Alternatively, the spectral analyzer source intensity may be falling, resulting in a loss of measurement sensitivity and an increase in noise. Each of these failure modes will have a different effect on the noise, so it is possible to diagnose the exact cause from the signal alone. The ability to predict impending failures based on subtle changes in the residual error matrices of multivariate models has great potential [22].

9.8
Automation

9.8.1
Business Drivers for Automation

9.8.1.1   Cost Benefits
Continuous processes tend to be somewhat periodic in nature – they oscillate. Low frequency variations that may be influenced by something as slow changing as the seasons, undulate through the process causing efficiency to rise and fall. The ability to reduce this variation may provide an opportunity to increase yields. Frequently, places of maximum yield are close to boundaries of process failure (in the way putting more material through a pipe increases throughput but also increases the risk of blockage). Where periodicity can be reduced, the operating space can be moved closer to the boundaries of process failure without introducing additional risk.

9.8.1.2   More Consistent Quality
An automated process will tend to respond faster and with better consistency to high-frequency step-changes to keep the process in control.

Most variability falls into one of the following six categories: people, methods, raw materials, equipment, metrology, and the environment. Automation reduces the number of potential sources of variability (i.e., people and methods), usually resulting in a product of more consistent quality.

9.8.1.3 Improved Compliance

A pharmaceutical manufacturer is required to keep detailed records of process information, including exceptions and deviations. Automated data-logging reduces data entry errors. Records of exceptions and responses are more consistent.

9.8.2
The Control Philosophy

The control philosophy is the manner in which control is exercised. It is defined by the inputs and outputs, specifically variation in the inputs and describes how the process must be shifted in response to that variation to ensure that the target range for the outputs can be maintained.

Control systems are used in two scenarios [23]: (i) to steer the process to a desired end point (servo control) or (ii) to maintain the process at a desired steady state (regulatory control).

A common example of servo control deployed during batch manufacturing is endpoint titration. During continuous manufacturing, servo control may be required during the start-up phase to achieve a certain set-point, with regulatory control being required to maintain steady state conditions in the presence of small disturbances.

There are a number of types of controllers, from PID, multiloop, and cascade, to inferential and model-based predictive control.

The most common systems use PID control. Some controllers use transfer functions (which are computationally light) to describe the dynamic characteristics of continuous systems and embedded in the Laplace transform is no small measure of *process understanding*.

At the more sophisticated end of the control philosophy there is inferential control. This approach is set to become increasingly significant, if parametric release is ever to become widely accepted in the industry. In this context, the CQA is not measured directly, but is rather estimated from a number of indirect measurements.

A mixture of direct sample and process measurements, all collated into one coherent picture, provides the basis of model-based predictive control.

In its most extreme manifestation, *process understanding* based on PAT and QbD concepts could eventually result in the use of closed loop, model-based predictive control systems to enable process control, quality assurance, and final product release to all occur in real time.

9.8.3
Signal Conditioning

Signal conditioning is a term used to cover the multiplicity of techniques designed to manipulate an analogue signal to improve its quality. In time series data, signal conditioning can be applied to extract a specific signal from a multitude of chatter. This is analogous to a listener in a room full of voices being able to tune in and hear to just one.

The reason this is applied to control signals is to ensure that the system only responds to the desired change, and not to noise, or erroneous spikes (gross errors) or real but inconsequential process perturbations. We wish the controller to follow the trajectory of the process and to not chase its inherent noise.

The easiest example of conditioning is the simple moving average. Above this, there are linear filters: low-pass, high-pass, or band-pass and finally a diversity of nonlinear filtering techniques of varying sophistication. Hampel filters [24, 25] are an effective means of removing spurious signals. Refer to Figure 9.4a,b. This example compares two orthogonal spectral analyzers (primary and backup) placed in series in a continuous reactor. The raw data for the backup analyzer could not be used for control on account of the noise and the data spikes. In order to reduce the noise and remove the spikes while introducing the least possible time lag, a Hampel filter was used to condition the signal.

Both analyzers now respond in a similar manner. The lag times, residual noise differences, and signal bias are sufficiently small that the system can be shown to be capable, and that no further optimization is essential. In short, the control system can switch between these analyzers and the process will remain in control.

Other useful algorithms for de-noising data are Kalman filters, Fourier transform, and wavelet compression, although it may be less desirable to deploy them in real time.

9.8.4
Univariate Control

Open loop control is the simplest mechanism of operation. Data are presented to an operator, usually in the form of a trend chart. The operator makes a decision as to whether the process is wandering and performs an adjustment to the process. The size of the response and the time taken to do what needs doing is then up to the individual.

In closed loop control, the operator is taken out, and the decision-making progress is performed with an algorithm.

Commonly, the control aspect is simply a matter of adjusting one variable (the manipulated variable) in response to changes in another to elicit the desired change in the responding variable.

For single adjustments made on single variables, how the unit operation is controlled depends on where in the process the measurement is placed. If it is placed at the start of the unit operation to measure a particular input variable, all

Figure 9.4 (a) Comparison of raw data from two analyzers monitoring the same process flow. (b) Comparison of two analyzers, with the noisier signal conditioned by filters. (Acknowledgments: Giuseppe Elia and Dr. James Reynolds.)

than can be done is to change another variable proportionately in response to it (ratio control). An example of where this system works well is keeping the ratios of two reagents the same.

When the measurement is placed at the end of the unit operation, that is, measuring an output, feedback control is possible. PID feedback controllers have been around since the start of the industrial revolution. The manipulated variable is consistently tweaked to achieve a target value. Temperature control is a good example of where PID control works well.

Feed-forward control happens when a quality attribute from one unit operation is applied to a downstream unit operation which modifies its process parameters in an effort to bring the process back to a desired state. This should only ever be

employed within a feedback loop, that is, when there is an analyzer at the outlet of the said downstream unit operation that is the final arbiter of control.

9.8.5
Model Predictive Control

I venture the opinion that very few problems in pharmaceuticals are sufficiently complex that model predictive control (MPC) would be the only practical means of resolving it. That said there are a few drivers: for an existing process, MPC may of itself provide a tangible cost benefit, a solution to an FMEA issue, and real-time QA for continuous quality verification.

GPC (geometric process control or generalized parallel coordinates), with its relative simplicity, is a well-proven technique for handling multiple variables [26]. Another excellent example of its reliability is where it has been used in collision avoidance algorithms for air traffic control systems.

Whenever MPC is required, the dynamics of the process need to be well understood in advance of implementation. One recommended way to accomplish this is to run the process in open loop, and to force change in every relevant dimension, and record how the process responds. Once the design space and the response surface are established, the operating space and any failure boundaries can be identified, the process can be modeled and the control system can be developed. The actual model can take many forms, using calculus, statistics, GPC, or neural networks. Once you have this model you have a truly PAT application.

9.9
Conclusion

Where is this all going? There is a school of thought that reasons that once PAT measurements have provided sufficient process understanding, the expensive, detailed measurements may be removed in preference for simpler, more reliable analyzers that enable such concepts as parametric and real-time release.

There is another school of thought that suggests the person to figure out how to implement QbD (including where and how to frontload the spend) within the risk, cost, and time constraints of patent life and process development, while also putting together a convincing argument about how a profit is to be recouped within the life-cycle of the product without the need to drive a complete overhaul of the way the industry goes about its business, is yet to be found.

Somewhere in the middle ground, meanwhile, there are opportunities to invest in increasing *process understanding* and turning that understanding into commercial, regulatory, and quality success.

Measurable improvements can be found in-process efficiency by reducing cycle times, and increasing plant capacity. Yields can be improved, and wastage reduced. Advantages can also be found in reducing compliance failures. By reducing the time among manufacture, testing and product release, inventory, and hence storage

requirements can be reduced. Any safety risks associated with in-process sampling of particularly toxic materials can be effectively eliminated.

Critically though, it reduces both the risk of product failure and the risk of that failure going undetected. It is consistent with our efforts to provide the assurance that patients are getting the quality of medication they require.

References

1. FDA (2004) Guidance for Industry PAT – A Framework for Innovative Pharmaceutical Development, Manufacturing, and Quality Assurance (September 2004).
2. FDA (2009) ICH Q8 (R2) Pharmaceutical Development.
3. Berry, I.R., Nash, R.A., and Wachter, A.H. (2003) *Pharmaceutical Process Validation*, 3rd edn, Illustrated, Published by Informa Health Care.
4. Dean, D. and Bruttin, F. Productivity and the Economics of Regulatory Compliance in Pharmaceutical Production, PwC Consulting Team, Basel, Switzerland. *http://www.fda.gov/ohrms/dockets/ac/02/briefing/3841B1_07_Price WaterhouseCoopers.PDF*.
5. Woodcock, J. (2002) FDA Regulation of Drug Quality: New Challenges, presentation by Janet Woodcock MD, Director, CDER, FDA.
6. Office of Pharmaceutical Science CDER, FDA (2004) PAT Guidance: Going from the Current Negative Vocabulary to an Manufacturing Science Enabling Vocabulary. A presentation by: Hussain, A.S. and Pharm, B., Ph.D. Office of Pharmaceutical Science CDER, FDA, London.
7. ASTM Standard, E2363. *Standard (2006) Terminology Relating to Process Analytical Technology in the Pharmaceutical Industry*.
8. ASTM Standard D6122 (2010) *Standard Practice for Validation of the Performance of Multivariate Process Infrared Spectrophotometers*.
9. ASTM Standard E2281 (2008) *Standard Practice for Process and Measurement Capability Indices*.
10. ASTM Standard E2537 (2008) *Standard Guide for Application of Continuous Quality Verification to Pharmaceutical and Biopharmaceutical Manufacturing*, American Society for Testing and Materials.
11. ASTM Standard E2474 (2006) *Standard Practice for Pharmaceutical Process Design Utilizing Process Analytical Technology*, American Society for Testing and Materials.
12. International Society for Automation (200?) ISA-95. *The International Standard for the Integration of Enterprise and Control Systems.*
13. ASTM Standard E2500 (2007) *Standard Guide for Specification, Design, and Verification of Pharmaceutical and Biopharmaceutical Manufacturing Systems and Equipment*, American Society for Testing and Materials.
14. European Commission (2009) ATEX Guidelines – Guidelines on the Application of Directive 94/9/EC of the European Parliament and the Council of 23 March 1994 on the Approximation of the Laws on the Member States Concerning Equipment Intended for Use in Potentially Explosive Atmospheres. *http://ec.europa.eu/enterprise/sectors/mechanical/files/atex/guide/atexguidelines_june2009_en.pdf*
15. International Society for Pharmaceutical Engineering (2008) Good Automated Manufacturing Practice (GAMP) 5: A Risk Based Approach to Compliant GxP Computerised Systems.
16. FDA (1997) 21 CFR Part 11: Electronic Records; Electronic Signatures; Final Rule.
17. FDA (2005) ICH Q2 (R1) Validation of Analytical Procedures: Text and Methodology.
18. EMEA (2003) Note for Guidance on the Use of Near Infrared Spectroscopy by the Pharmaceutical Industry and the

Data Requirements for New Submissions and Variations CPMP/QWP/3309.

19. FDA (2006) ICHQ9 *Guidance for Industry Q9 Quality Risk Management*, CDER, FDA.

20. International Electro technical Commission (2001) International Standard IEC 61882 Hazard and Operability Studies Application Guide.

21. ASTM Standard D3764. (2009) *Standard Practice for Validation of the Performance of Process Stream Analyzer Systems*.

22. Alawi, A., Choi, S.W., Martin, E.B., and Morris, A.J. (2005) Batch Modeling and Performance Monitoring Using Multiscale Data Analysis. Advances in Chemical Process Control (ADCHEM), 2006, Brazil.

23. Morris, J. (2007) Introduction to Process Control (for Chemists), Presentation, University of Newcastle.

24. Allen, D.A. (2009) A frequency domain Hampel filter for blind rejection of sinusoidal interference from electromyograms. *J. Neurosci. Methods*, **177**, 303–331.

25. Pearson, R. (2002) *Scrub Data with Scale-invariant Nonlinear Digital Filters*, Swiss Federal Institute of Technology.

26. Brooks, R., Wilson, J., and Zhao, N. (2008) Visualising Design Space, Control Space and Operating Envelopes make QbD and PAT Easy, Innovations in Pharmaceutical Technology, 88–92.

10
Plant Design

Mark J. Dickson

10.1
Introduction

The success of the process industries is based on supplying final products that meet the needs of consumers. A key step in delivering this objective is reliably converting economic processes from research and development to manufacturing. In the past, this has been achieved through stagewise scale-up of processes, with each process being proved at one scale before progressing to the next. Critics suggest that knowledge gained and reported at each development stage is often observational and not based on a thorough understanding of underlying scientific principles. The result is that process plant is designed on the basis of observational data and not underlying science, and variations in starting materials and operating conditions lead to ineffective and inefficient processes.

In this section, we focus on how processes and products are introduced into the manufacturing environment, and the implications this has for the design of process plant. We start by reviewing the options available to business managers when making manufacturing decisions. We then take a detailed look at how processes are transferred from laboratory to full-scale manufacturing (a change of focus from *process* to *plant*), followed by the restrictions placed on this transfer by current knowledge and regulations. We conclude by looking at the role of *process understanding* in enabling agility to be designed into a plant to meet current and future business needs.

The ideas in this chapter are based on the author's recent experiences in the design, construction, and operation of novel process plant facilities, based on improved process understanding. The move to science-based manufacturing assets in the process industries is a developing field. We expect a greater level of understanding as a result of the current interest and investment in research for this area.

In many process sectors, for example, high-volume petrochemicals, there is a financial incentive to improve the process through enhanced understanding. A small increase in productivity can relate to millions of dollars of additional revenue. This provides an incentive, and funds, to gain this understanding and

there are numerous software tools available that collate and trend this data into real information. By contrast, complex and low-volume products do not generate the required financial benefit to justify the expenditure to improve our understanding. However, increases in productivity in lower volume processes can still make a significant impact to the bottom line profit of companies in these sectors.

It is important to define what is meant by *process understanding*. From a scientific and engineering perspective, we often believe that more knowledge will always lead to better processes, and this would suggest that more *process understanding* is always preferred. From a business perspective, we need an appropriate level of understanding to make the best decision at the lowest cost. This suggests that we should undertake the fewest experiments possible and be specific about the knowledge we require. Therefore, I use the term *process understanding* here to mean *"gaining only the required pieces of information to make the correct decisions and identify the optimum process (best value) from a business/consumer perspective."*

10.1.1
CAPEX Project Phases

The introduction of new assets through capital expenditure (CAPEX) requires input from a range of stakeholders, including business managers, multiple engineering disciplines, operational staff, and multiple scientific fields. The implementation of process plant projects follows the phases highlighted in Figure 10.1. As the project progresses, there is a change of emphasis in the skill set required to successfully deliver each phase. In the early stages, the work is mainly science and chemical engineering based, as this requires us to resolve the process options to find the optimum process. As the project continues, other engineering disciplines

ID		Period							
		1	2	3	4	5	6	7	
1	Discovery								
2	Decision point								
3	Research								
4	Development								
5	Design								
6	Decision point								
7	Engineering								
8	Procurement								
9	Construction								
10	Commissioning								
11	Operation								
12	Operational improvements								

Figure 10.1   The key stages of process plant projects.

(mechanical, electrical, civil, etc.) take on further responsibility to turn the process concept into a real plant. The input of a wide range of engineering skills in later phases of a project is essential to delivering a flexible, functional, and controllable plant. As the project progresses, the size of the team required to deliver the plant increases significantly, increasing the incremental time cost. The quality of the final plant and overall cost, is therefore dependent on the quality of information and knowledge available earlier in the project.

Figure 10.1 shows the overall process from discovery to operation. All companies have intermediate decision points where the project is evaluated against business criteria. At each decision point, more information is known about the process and costs of manufacturing; and these can either be advantageous or prohibitive to the project progressing. The key decision point (item 6 in Figure 10.1) is taken with a detailed understanding of the market conditions and is where a company decides whether to produce in-house or outsource.

When a project moves from concept to construction, the costs incurred by the business increases exponentially. Figure 10.2, and many variations of it, indicate the high costs of change at later stages of a project. The move from design to engineering is accompanied by a change in emphasis from development to delivery. Continued changes to the basis of design are accompanied by large rework costs, due to a large resource working to deliver the project, having to undertake rework. The cumulative effect of multiple changes can cause significant cost overruns and schedule delays in delivering the functional plant. At this phase, any noncritical process improvements are best delivered as improvement projects to be implemented later.

Figure 10.2 The potential for reducing costs reduces as the project progresses, while the cost of making changes also increases.

Science

The process of identifying and developing technology and products.
Influenced by the experience of scientists and available tools.

Business

The systems which take products to the customer, and identify what customers want and need.
Fixed resources available which must be balanced.
Businesses are a group of people with capabilities and limitations.

Engineering

Combining technologies to make products.
Can be limited by existing assets or knowledge of technologies.

Optimum process

Figure 10.3 Science, engineering, and business disciplines must work together to find the optimum process for a specific company.

Figure 10.2 highlights that decisions made early in the development of a process have a significant potential to lock in costs, and this is where we find the real value of *process understanding*. The more we understand the process in earlier phases, the greater chance we have of making informed decisions leading us to the best value process. This is why we extend the definition of *process understanding*, beyond technical insight, to include financial and business information. We need to use technical knowledge alongside financial information to get a clearer picture of the risk and potential costs associated with a process.

Enhanced *process understanding* early in the project is of most use when combined with multidisciplinary decision making. Previous chapters have highlighted the need for engineering and science to work together, and here we add business managers to that mix (Figure 10.3). It is only through analysis by a multidisciplinary team (a team of specialists from different functions rather than multidisciplinary individuals!) can the optimum solutions be identified.

10.1.2
Starting Plant Design

The outcome of the process research and development phase includes

• process description – usually biased toward reaction stages;

- individual process stage laboratory reports;
- physical property data for major components;
- transformation kinetics (mainly reaction stages but occasionally major separation processes);
- trial data from laboratory equipment;
- early equipment selection decisions;
- development team with lots of tacit knowledge;
- market research and analysis reports;
- keenness within the organization to move into production.

Plant design is an iterative process that starts very early in the development of a new process and continues until the plant is fully operational. Early in development we have partial knowledge in all areas of the process with lots of options. As we narrow the options to identify the optimum process, it is important to consider plant design constraints and opportunities.

10.1.3
Equipment Selection

The key objective in equipment selection is to understand the physical, chemical, and/or biological transformations that are occurring and find the most appropriate equipment to deliver the required conditions to meet the business and economic requirements at the manufacturing scale. The optimum equipment in the laboratory may not be the same as the optimum equipment for manufacture, and therefore the emphasis should always be on the capability of manufacturing equipment.

In a sequential scale-up model, the selection of equipment is typically made early, and subsequent scale-up stages are forced to use the same equipment. In this model, the level of knowledge available about the process is insufficient at the early stage to make alternative or informed choices about equipment selection; therefore, the selections are often inefficient and suboptimal.

If the development team has identified the right process information, then the equipment selection decisions can be flexible, even as the project moves into the design phase. This gives the design team the opportunity to identify the best value equipment, considering the wider business environment. The information required from a development team is dependent on the complexity of the process but would include the information identified in Table 10.1.

If the information in Table 10.1 is available, then we can use comparison tables (e.g., Quality Function Deployment (QFD) or Kepner Tregoe) to match the process requirements to equipment capable of delivering the required conditions. Any gaps in this information will result in further experimentation to determine that parameter, although order of magnitude data is often sufficient for decision making for less critical parameters.

Enhanced process understanding is only a technical nice-to-have unless we are able to make use of that knowledge. For equipment selection, we also require an

Table 10.1 Information required about the process to allow equipment selection decisions to be flexible until design stage.

Information required	Why is this useful
The main transformations and significant side reactions/transformations	Defines the process and significant sources of impurities
	Provides opportunity for multidisciplined thinking to identify opportunities to maximize yield
System phases (e.g., 1 water 1 solvent). Also, in which phase do transformations take place?	Defines contacting pattern and identifies mass transfer inhibited systems
Exotherm or endotherm against time	Determines the required level of heat transfer and also where/when that heat transfer is required
Inherent kinetics (effect of mixing and heat transfer)	Helps select between short/long residence time technologies
	Mixing – select high-intensity mixing to improve limited systems
	Heat transfer – improve control of heat transfer to remove process limitations, for example, effects of drying conditions on product morphology
The optimum temperature range ($\pm$)	Defines required level of control or if a natural heat sinks are needed to stabilize the process
Stability of materials at all conditions	Defines contacting patterns and any safety/operability issues
Equilibrium details (which component must be removed to drive the favored route?)	Need to identify ways to continuously remove the inhibitor as it is generated, or get the correct contacting pattern
Are significant or difficult to remove impurities observed in the process?	Impurities from one stage can have a significant impact further down the line. Continuous removal of these impurities or adjustments to the process to prevent their formation can prevent issues in downstream processes
Does the transformation involve any very toxic reagents and in what quantities?	Chosen technology may limit amount of toxic material in process or link further reactions to avoid toxic intermediates
Does the transformation include solids? If so, what percentage, in what form, and have they been identified?	Determines the type of technology and equipment
Is a catalyst/promoter present and at what concentration? What is the rate of catalyst degradation/loss of activity?	Identifies if fixed/mobile catalyst options are possible and how they may be removed from a process stream
Opportunity to operate at lower efficiency and recycle for higher throughput	For example, to reduce excess or for chiral products

Table 10.1 *(Continued)*

Information required	Why is this useful
Any additional reasons for a high operating cost	For example, higher hazard materials requiring special equipment/handling
Any difficulties in achieving isolation purity	–
Environmental factors	Will current/future release limits require complicated/expensive downstream abatement?
Use of reagents/solvents	New/future legislation implications, for example, REACH

REACH, Registration, Evaluation, and Authorization of Chemicals.

understanding of available technologies and more detailed knowledge of specific equipment items. Only then can we adequately match the most appropriate equipment to the process needs. These decisions influence the overall plant design process, in that inappropriate equipment selection can either result in the plant being unable to meet the process specification or, just as important, cause the plant to use more energy/cost/time to meet the process specification.

10.1.4
Assets (Existing or New)

Businesses operate with limited assets and resources. If a single product or project has the potential to finance its own manufacturing plant, which will be fully utilized, then a dedicated and optimized plant may be constructed for this product. If we use the *runners/repeaters/strangers* analysis from operations management, then a product that justifies its own plant is usually considered a *runner*. However, many products will not achieve the level of sales required to justify dedicated investment, and market uncertainty may also delay investment until a product is established and the market value is known. Most products are likely to start as *strangers*, moving to *repeaters*, and hopefully to *runners* eventually.

It is therefore accepted that most companies have multiproduct assets, which allows them to run new processes in flexible assets with either minimal or no additional capital investment. It is also recognized that running a process in existing assets may result in suboptimal technical performance. However, it is important to note that suboptimal from a technical perspective is different from suboptimal from a business perspective.

Example: Company A has a reaction step that takes 4 h to achieve 80% completion and 8 h to reach 95%. To achieve the same throughput the reaction volume for 8 h is double that required for 4 h. If company A is limited in reactor volume, then a business driver may be to maximize the conversion **output per hour** *from the reactor volume available. Therefore, providing the opportunity cost (remaining 15% conversion) is not*

prohibitively expensive. Company A may be more profitable by limiting this reaction to 4 h.

Using existing assets is not always the best option, particularly if the resultant suboptimal process causes additional costs, for example, quality or control issues. Using existing assets also determines the minimum cost for a specific business to produce a particular product, while a company that has a different asset base may be able to produce the same product at a significantly lower internal cost. This can lead to make-or-buy decisions for a particular product if an outside company can produce more cost-effectively than using in-house assets.

This is where *process understanding* can really make a difference. If an organization has the required level of knowledge about a process to understand the effect that changing some equipment will have on the overall cost of manufacture, then the business has a range of options available to find competitive advantage. This can be through the integration of new assets with the existing available equipment or, if the benefits can be quantified, by moving to a completely new asset.

Using *process understanding* to move to a new manufacturing asset can significantly reduce the cost of manufacturing; however, consideration also needs to be given to the need to modify the new asset to accommodate process improvements.

The decision on whether to use existing or new assets is often driven by wider business issues, for example, availability of fully depreciated assets, supply chain integration (e.g., Just In Time, produce for storage), available resources (e.g., skills available to handle novel technology), and knowledge development/retention. This also needs to be balanced with the level of acceptable risk across an organization and this is often a point of much debate. As manufacturing technologies develop, there are often a range of novel technologies that have the technical promise of lowering overall costs and improving quality; however, these technologies may not have been proven in this application. Successful organizations are those that manage their exposure to unproven technologies by identifying the ones that have the highest potential impact (across multiple products in the company portfolio) and focusing attention on these areas. Specialist technologies that have a narrower range of future applicability represent higher risk than technologies with extensive future potential.

10.2
Developing Process Concept to Plant Concept

10.2.1
Process Information

The design and engineering phase of a project should be based on a commonly understood description of the process. This typically takes the form of a block flow or process flow diagram, preliminary plant mass balance, and detailed stage

Table 10.2 Typical documents available at the start of design.

Document	Limitations
Block/process flow diagram	Typically equipment based rather than process based, this means equipment selection decisions are often fixed by process development teams rather than plant designers. Choosing equipment early in development can often lock inefficiencies and high capital/operating costs into the process
Mass balance	Virtually always based on optimized individual stages rather than an optimized system. This means that each stage is pushed to maximize yield, often incurring high costs. Also the process stream is then conditioned between stages, again incurring additional cost
	Mass balances focus on main components and do not provide sufficient information on side reactions
Laboratory stage descriptions	Laboratory reports contain the data required for proving a process concept and this is not the same as the data needed for plant design
	Laboratory reports tend to focus on the main chemical transformations and contain less information about side reactions and physical processes
	The most useful laboratory reports include annotations highlighting key observations from experiments. While these may be viewed as issues during development, they are critical knowledge to a plant designer
Business case	Provides some process data but often makes assumptions about by-products, yields, and so on. Typically, a business case is focused on one process route that can again force the design down a suboptimal route

descriptions from laboratory experiments. Each of these key documents has commonly encountered limitations as detailed in Table 10.2. One of the key issues is that these documents record what is known about the process in terms of discrete data points. However, the design is about bringing all the information together into a model of the process in which design in one area will have a direct effect on other areas of the process. Therefore, increased *process understanding* should allow us to generate a process model; this model should record all the known data and where assumptions have been made. This highlights what is fixed information and what data has been assumed to fit the system together.

A model of the process based on a more thorough understanding of the underlying science allows the plant design team to make better judgments of the likely performance of different sections of the process (Chapter 5). This knowledge helps the plant design team to understand what options are available to improve

the process and what critical parameters must not be changed. In this respect, *process understanding* can be thought of as informing the design team of the current state of knowledge about the process, thereby assisting in the knowledge transfer and communication of the process, as it moves through the development chain.

10.2.2
Physical Properties

The physical property information typically available at the start of design and engineering is often limited to pure component data. However, actual process fluids do behave very differently to pure components and, given the complex nature of many materials, this physical property data is often best measured rather than predicted. Software tools in design are very accurate at predicting physical properties of commonly used materials and pure components but not of actual process fluids.

The quality of physical property information available for design is often poor, as this information is not required for setting up laboratory equipment. However, it could significantly influence the project economics, and basic physical property data is essential to ensuring a suitable hydraulic design; see Table 10.3.

For transformations more in-depth physical property data is required and specific information for example, heat transfer capacity or rate of reaction, is essential for plant design. The specific information needs of a process stage will often depend on the technology that is used to undertake it rather than on the transformation, and therefore it is difficult to generate a complete list of physical property information needed in all cases.

Table 10.3   Key physical property information used during design.

Physical property	Essential information	
	All materials	**During transformation**
Density	✓	✓
Viscosity	✓	✓
Presence of solids	✓	✓
Presence of dissolved gas or off-gas	✓	✓
Operating and max pressures	✓	✓
Operating and max temperatures	✓	✓
Heat transfer capacity	✓	✓
Boiling point	✓	✓
Freezing point	✓	✓
Degradation temperatures	✓	✓
Toxicity	✓	✓
Mixing intensity (energy input)	–	✓
Selectivity	–	✓
Presence of catalysts	–	✓

10.2.3
Impact of Observations on Design

During development, the process will undergo a series of optimization experiments that may include changing solvents, running steps in different orders, and so on. Throughout these experiments, the development team must record as much information as possible. There are many examples of scale-up and plant design problems that could have been resolved if the minor deviations observed in trials had been recorded (Table 10.4).

The observations from ''failed'' experiments are just as useful to plant designers as ''successful'' experiments. This is particularly true if alternative reaction paths lead to generation of solids, off-gas, or large temperature spikes. The design team can then ensure safe design by either avoiding the conditions that cause these reactions or designing the equipment to cope with these deviations. Plant designers often request photos and videos of experiments so they can interpret the observations from a designer perspective.

10.2.4
Equipment Selection Decisions in Process Development

In the previous section, we highlighted the need to define the process needs and match these with appropriate equipment. A key part of plant design is therefore to make these equipment selection decisions. In the majority of cases, there will be multiple technologies available to meet each process duty; in some cases, a technology or equipment items will need to be modified to be able to meet the process needs.

When multiple options are available, a comparison analysis should help identify the preferred technology. Where the analysis indicates multiple preferred options, and if our knowledge is incomplete to make the final selection, then we should specify clear experiments to find the required data to make the best choice. These are often referred to as ''*killer experiments*'' as the results are used to kill off certain options.

When a technology needs to be modified to meet the process needs, there is a requirement for the plant designer to engage quickly with expert suppliers of that equipment. It is important to understand the restrictions that have shaped the current design of equipment in that technology area and whether these will restrict the modifications.

10.2.5
Combining and Splitting Tasks

One of the most frequently encountered opportunities in plant design is to reduce the number of equipment items by combining sequential tasks. Conventional thinking appears to favor each subsequent unit operation taking place in its own dedicated equipment item. The process then becomes a series of equipment items

Table 10.4 Typical examples of statements in laboratory reports and the implications for design.

Observation	Potential cause	Implication for design
Minor pressure increase in headspace	Off-gas from the reaction	Exacerbated at larger scale to cause significant operational and safety concerns. Venting and relief system design needs to account for this off-gas
Occasional experiments have short temperature spikes during the reaction	Side reactions occurring locally with a higher exotherm	The reactor design needs to be capable of controlling the rapid exotherm caused by the side reaction else the reaction mixture may have thermal runaway
Short temperature spikes controlled by cooling system	Small volume can be controlled by spare capacity in cooling system	Cooling capacity of manufacturing scale may be limited relative to reactor volume, hence longer time to control the temperature with potential off-spec product
Solids layer formed during reaction, which later dissolves	Insoluble intermediate	Agitating slurries requires more power than agitating liquid phases. Therefore, equipment designed for liquid case may not be able to mix and dissolve the solids in the larger plant
Minor quantity of unknown solids remain in flask	Side reaction generates solid by-product	Plant design may need separate cleaning system/flushing system to remove solids. Alternatively, system must be designed for higher velocity to prevent settling out
Increased viscosity observed during the reaction	Distillation causes increased viscosity or polymerization of intermediate	Plant design may need to incorporate higher power agitation/improved heat transfer for viscous fluids

with transfer systems (pumps and pipes or mechanical transfer equipment) to move the process from one unit operation to the next. Increased understanding of the process allows the plant design team to utilize available parameters to make sequenced operations happen simultaneously; for example, a reaction can take place in the pump and transfer line between two other operations, or relatively new technologies can be used, such as reactive distillation, dividing wall distillation, and so on.

Reviews of whether process steps can be combined or split should be undertaken throughout the development of the process. If this information is recorded throughout the process, then it would allow the plant design team to understand what attempts have been made to simplify the process in previous work and identify the potential restrictions. This type of review very rarely takes place in process development; however, it is a good starting point for development teams to understand the interactions that changes in one area can make to upstream and downstream process steps. This is also another perspective from which to view the process and it encourages improvement through process challenge.

10.2.6
Batch versus Continuous Processing

Conventional chemical engineering teaching indicates that processes are often run in batch operation until a sufficient volumetric throughput is required, at which point continuous processing is more cost effective. However, small-scale continuous processing steps have been used to solve specific issues in batch operations for years. Significant advances over the last decade have resulted in many proven technologies being available to run more processes in a small-scale continuous mode. The decision between batch and continuous operation must be based on *process understanding,* and Table 10.5 highlights some of the differences from a plant design perspective.

The decision of whether a process should be run in a batch or continuous mode has both business and technical implications. The technical implications are related to process conditions (Chapters 2, 4, 6, 12) and equipment considerations. It is worth highlighting that choosing between batch and continuous modes is not a simple binary decision. Instead, there are multiple variations of batch technology (e.g., fed-batch) and even more variations on continuous operation (plug-flow or continuous stirred tank reactor (CSTR), steady-state, or non steady-state). The distinction between batch and continuous operation is often blurred, for example, running a single process stage through a plug-flow pipe or in a CSTR could be achieved using upstream feed system and downstream collection vessels. Although the reactor is considered to be in continuous mode, the system as a whole is very much a batch operation. This unit may or may not achieve steady-state operation, and therefore variable processing conditions can introduce the same variable quality issues as batch operations.

For the pharmaceutical industry, a series of sequential batch operations provides the opportunity for intermediate quality control. However, the advances in Process Analytical Technology (PAT) will enable online quality control of continuous processes.

Deciding which is the most appropriate of these categories requires a detailed level of understanding of the process; however, the mode of operation is less important from a technical perspective than the choice of technology. There is no reason why an optimum facility should not include a variety of batch and continuous technologies for the same process. The optimum facility may also

Table 10.5 A comparison of batch and continuous technology from a plant design perspective.

Project/process drivers	Batch	Continuous
Capital cost	High	Low but requires more information and engineering time
Operating cost	High	Lower energy but more control required
Toxic inventory	Set by batch size	Lower in reactor (potential for higher overall quantity in storage but this is held at safer conditions)
Heat transfer ability	Limited by surface to volume	Higher surface to volume
Reaction rate	Fixed by batch conditions, options to use fed-batch	Design equipment to optimize and option for product removal
Mixing ability	Limited by batch agitation	High to very high
Energy efficiency	Large peak demands	Smaller but continuous loads
QA requirements	Potential batch variability	Fine tune process to meet QA
Range of process options	Options limited to those in batch vessel	Multiple possible answers
Multiproduct capability	Highly versatile	Need to design for multiproduct use else single product
Lead time – development to production	Well known	Reduced for smaller equipment/buildings but increased data requirements
Batch traceability	Easily defined based on vessel contents	Time-related batches must be linked to process data collection while the material was in processing

incorporate equipment items of multiple scales including micro scale structured up to conventional macro scale or nonstructured equipment.

From a business perspective, the decision to move from batch to continuous operation has some interesting implications. The low-volume process industries are almost exclusively operated in batch mode under current operation. The advantage of this operation is that processes have natural break points that are embedded into current working practices. For example, there is a conscience decision by the operations team to move a batch from one vessel to the next and this only occurs once the receiving system has been prepared. When running a continuous operation the full process plant must be available, as stopping the process is likely to cause quality perturbations.

There are also wider implications of running continuous processes that include the following:

- **Staff training** – operators used to batch processing will need to understand how continuous operations are different.
- **Staff welfare** – a continuous plant may need constant supervision that can often change the working patterns of staff.
- **Emergency support** – if plants are working at full capacity on a 24/7 basis, then support services need to ensure availability.
- **Quality analysis** – batch processes go through validation (three repeatable batches) and then rely on the outputs of quality samples tested in the laboratory. For continuous processes, online sampling is required to ensure steady-state operation, and online PAT becomes important.
- **Control system update** – controlling a continuous plant automatically is well proven in large volume industries, but the basis of how the control system works is different to batch operation.
- **Integrating the supply chain** – continuous steady-state plants need to have raw materials continually fed and products removed; this may require more frequent supply and collection of materials.
- **Availability of site utilities** – many sites designed for batch operation have infrastructure designed for peak loads followed by lower operation. The option of running utility units continuously needs to be assessed carefully. A further consideration is the impact of running both batch and continuous processes from the same utility systems, as often large batch loads on a system can upset the steady-state conditions of a continuous process using the same utilities.

10.2.7
Sustainability

The move toward a sustainable process industry involves many different aspects from sourcing of sustainable raw materials through to maximizing the efficiency of processes such that we minimize waste, and, importantly, it includes the need to choose the most appropriate synthesis routes to minimize the energy input into processing. From a plant design perspective, the key issues around sustainability are efficiency of processing, treating/reusing/minimizing by-products and waste materials, and ensuring efficiency in plant cleaning.

Efficient processes can be achieved if they are run at optimum conditions and this is facilitated by the level of knowledge gained about the process. In particular, plant designers are often required to include design margins to ensure that all process parameters can be controlled. Design margins are required because our process understanding is not accurate enough to be confident in the specified operating parameters, and increased understanding does mitigate the need to introduce design margin inefficiency.

The treatment and reuse of by-products and waste materials requires some innovative thinking. There are limited examples of organizations choosing chemical

routes specifically because they produce a more useful by-product; however, the industry trend is to push toward this.

Sustainability however, is not just about the main process or its waste products, but maybe, more significantly, includes off-quality product and cleaning solvents. *Process understanding* again has a role to play in plant cleaning, which might not be immediately obvious. The selection of the process route and therefore process solvents, plus the choice of technology used to manufacture a product, has a strong link to the potential cleaning agents that will be used and the quantities of these materials. Fine and specialty chemicals and pharmaceuticals are notorious for using large quantities of solvents per kilogram of product, with much of this excess solvent being used to dilute processes for control or used as a cleaning solvent. Enhanced *process understanding* will help reduce this usage. There are also practical considerations that can be built into plant design to ensure the use of cleaning solvents is minimized (e.g., clean in place (CIP)), thereby reducing the environmental impact, the cost, and also the time taken for cleaning.

10.2.8
The Opportunity for Innovation

As manufacturing technologies develop, organizations are often criticized for not following the latest trends. Some organizations become the key innovators and help push technologies forward, addressing the development challenges, and being the first to overcome these. A second group is the followers, who wait for new technologies to be proved before they adopt them. This group is often cynical about the benefits of a new technology, which may be inconsistent with their corporate strategy, until it becomes accepted in industry.

The first innovators on new technology often bear high development costs, as no matter how promising a technology is in the laboratory it has to be commercialized to make money for the innovator. There is also risk in the development of technology, and one of the most challenging decisions is to stop programs that are not delivering the expected benefits such that resources can be utilized to develop other technologies.

Despite the high cost of innovating, business theory warns against the risk of not innovating. The lead innovators in an industry are likely to achieve competitive advantage and it is important for all organizations to be aware of industry trends (in the local and global markets).

10.3
Regulations

The designers of advanced manufacturing plant face some difficult regulatory and industry acceptance challenges. The current regulatory framework in each country ensures that all companies achieve a minimum standard and is based on the

prevailing dominant technologies; however, much of the discussion in this book would push the chemical manufacturer toward novel technology. While many emerging technologies show significant promise in terms of process capability they are yet to be proved in a manufacturing environment; hence they are difficult to regulate.

Plant designers often look toward industry guidelines and standards when specifying equipment; however, these guidelines typically only cover well-established technology. The widespread use of novel technology will therefore need a specific level of industry testing to ensure that equipment is reliable and provides repeatable results. Without test data, it is sometimes more difficult to justify the unknown risk of using novel technology over more traditional and well-established equipment. While this is not a fully justified or even optimum way of approaching a problem, it highlights one of the difficulties faced by companies wanting to introduce novel technology.

10.3.1
Legal Requirements

The key regulations for the process industries cover the safety and containment requirements for hazardous materials. Many of the hazardous materials used are toxic and/or flammable at operating conditions and equipment must be suitably designed to contain these hazards.

The European directive 94/9/EC (ATEX) provides guidelines from the specification of electrical components for use in hazardous areas. Manufacturers of traditional scale equipment have incorporated the requirements of ATEX; however, achieving the required physical separation distances and enclosure types for intensified and close-coupled equipment is often difficult.

The Pressure Equipment Directive (PED) can also prohibit intensified processing, as it specifies minimum material thicknesses from a safety perspective, which can conflict with the design intent to minimize material thickness for improved heat transfer.

In each of these examples, designers of novel equipment need to provide risk assessments to identify why their product can be operated safely owing to reduced inventory or alternative form of containment. This approach can only be used with a thorough understanding of the process to quantify the potential risks.

10.3.2
Industry Standards

Industry standards have evolved over time to reflect operational experience and basic science. These standards are used by regulators, insurers, and industry bodies to evaluate the safety of a process plant. For the design of novel equipment, it is often necessary to deviate from industry standards and therefore gaining acceptance by the wider stakeholder groups can only be achieved by providing answers to the likely concerns. These answers can be provided through a more

detailed understanding of how the process will behave in the novel technology. This understanding is needed to compensate for the relative lack of operational experience.

As novel technologies become more widely accepted, the current state of industry knowledge will develop such that novel technology today becomes standard industry practice in the future.

10.3.3
Developing the Knowledge Base

Although underlying science can often mitigate risk and perception of risk, innovators must also recognize the need for organizational and industry knowledge to develop. This change process will inevitably take time and has led to some stakeholders feeling frustrated at the slow take up of innovative plant designs utilizing novel and intensified technologies.

The process of introducing new technology raises many important issues for developing the industry skill base. First, current scientists and engineers need to recognize the need for enhanced *process understanding* as the basis of all projects. Second, we need to analyze current processes for opportunities to improve processes. Third, the training of future generations of scientists and engineers must refer to the needs for enhanced understanding. Finally, we must remember that the industry is not just about technical expertise but also requires process operations and maintenance staff to gain a more detailed understanding of how the plant is working, such that they have the expertise to ensure correct operation.

10.3.4
Quality Control

There is an essential need to ensure that our process plant produces repeatable quality products. Traditionally, many processes rely on experimental and operational observation as the basis of control. A move to *process understanding* allows more advanced control methods, including predictive control that can keep a manufacturing facility within quality critical parameters. A more thorough explanation of process control issues including PAT initiatives is included in previous chapters.

There is a strong link between PAT and plant design (Chapter 9) as there is a need to integrate the measurement sensors into the equipment. For traditional technology, this has resulted in a new generation of advanced analyzers including inline/online high performance liquid chromatography (HPLC), gas chromatography (GC), and so on. For intensified equipment, the sensor needs to be designed such that it measures a representative sample of the process. The next generation of small-scale sensors to achieve this is currently in development; however, the development of these sensors has to comply with many of the legislative requirements previously noted.

10.4
Infrastructure Design

The term multiproduct plant (MPP) has been used extensively in the process industries to describe flexible facilities that are used to produce multiple products. The MPP concept is useful for organizations with a large range of products to produce. However, it must be highlighted here that an MPP is not an all-product plant; this means there are limits on what can be achieved in an MPP. Plant designers are often confronted with the question of how much flexibility should be built into a plant.

First, we must understand what is meant by flexibility; this can mean the plant can be easily reconfigured, can operate across a wide pressure and temperature envelope, can accommodate future equipment, has a control system that is easy to recode, has multiple components at different scales of operation, can accommodate all materials and solvents, and so on. Designing a plant to accommodate all of these criteria is not possible, and approaching this level of flexibility is extremely costly and often unnecessary.

A better approach is to design for known flexibility, that is, the processes that you know will be run in the facility. Then ensure the design is adaptable and requires low investment to be optimized for new processes. Designing for future expansion rather than installed flexibility results in the best value initial capital investment and ensures the design team is thinking about ways of adapting the facility in future. The concept of future expansion should extend beyond process equipment and incorporate the wider facility and buildings.

10.4.1
Plug'N'Play

One way of achieving future adaptability is to consider the so-called plug'n'play concepts (Figure 10.4). The key objective here is to divide the plant into manageable sections. These sections may be fully dressed equipment items (i.e., complete with all associated piping and instrumentation), unit operations (i.e., a collection of equipment always used together), or a section of a whole process (e.g., crystallization and solids separation). Once the appropriate level of flexibility has been identified, each of these units can be built in a modular form, minimising the number of electrical and instrumentation cables/utility of connection points. The facility can then be designed as a series of hook-up locations that can accommodate one of these modules.

The same issues surrounding built-in flexibility apply to the plug'n'play concept as other MPP ideas. If the manageable sections are too small, then a significant number of hook-up points are required, adding cost through redundancy. This concept is also restricted in that many of the components that make up a process plant are by nature variable in size and therefore providing uniform hook-up locations may not be the most effective use of space.

Figure 10.4 Options for modular plug'n'play plant, including a single unit operation per module through to individual equipment sub-modules configured in a larger module. The system needs flexible utilities and services.

The most significant advantage of the plug'n'play concept in plant design is that it allows the plant to be reconfigured easily and therefore optimized for each new process. Keeping the components small and mobile also ensures that individual sections can be replaced as required for maintenance or for process changes.

It is important to note here that plug'n'play facilities can be highly automated and do not need to be manual operations. Control system technology within the process industries has developed, as it has in other information technology (IT) industries, and therefore a plug'n'play facility can still retain the essential control and safety functionality that we design into process plant facilities. A flexible automated facility can be sequence linked with islands of automation or it can utilize the IT concept of installing drivers in a central system to replicate the automation installed on modular units.

10.5
Portfolio Analysis and Asset Planning

Plant design is often considered in terms of the needs of a current or immediately known future process. In other cases, multiproduct facilities look to address a broad range of potential future processes, often based on historic knowledge of what the

organization has produced in the past. This has often meant that companies use extremely flexible batch process vessels, although these have limited capability and often result in processes being run under suboptimal conditions.

As companies learn more about advanced processing technologies based on scientific understanding, they are suddenly faced with a larger range of potential technologies, using either more structured equipment or different modes of operation, for example, continuous. A key issue is to identify which of the numerous technologies available can best meet the *process* needs, whereas the real question should be which of the technologies can best meet the *business* needs. The second question is about choosing manufacturing technologies that will be applicable across a range of processes within a company portfolio. This links technical decisions to market-driven strategic change and means organizations have to understand the various factors that influence change.

10.5.1
Finding the Optimum Process for My Company

A very interesting observation about *process understanding* is the recognition that *"the optimum process for one company may be different to the optimum process for another company for exactly the same product."* This is because the optimum process is not just a technical concept but linked to business constraints.

We have established that the optimum process is sought by recognizing and managing the constraints imposed on the process. These constraints are often technical trade-offs within the process; however, many of the constraints are company specific, for example:

- In-house expertise available in a certain technology
- Accessible expertise in the local industry
- Level of supply chain integration
- Working practices of the company
- Level of risk acceptable to the company
- Available infrastructure and utilities on the site.

Given the influence of these constraints, it should be no surprise that the optimum process for producing a given product may be different for each company. The key issue is to find the optimum solution considering the constraints on the company and determine whether this is competitive against the market.

10.5.2
Agile and/or Lean Manufacturing

In a competitive chemical industry, many companies have turned to management tools to make step change process improvements, rather than focusing on the technical aspects of *process understanding*. The business theory used in other manufacturing industries can help make significant step changes in the process sectors.

Lean manufacturing ideas, originally developed from the Japanese motor industry, are now widely used in companies to make staff aware of the cost of quality. Six sigma initiatives are combined with reducing process waste, ''muda'' in lean terminology, to ensure more efficient processes. While the move to lean manufacturing is welcomed for the process sectors, it does raise an interesting issue. Lean is all about reducing waste in a given process, and the process sectors such as other industries, apply the tools successfully. However, step change improvements in process terms are often more about doing something different, rather than just doing it with less waste (better). This means there is a limit to how lean a given process can be; *however, changing the synthesis route may be fundamentally more effective than reducing waste in one particular route.*

The ability to change the synthesis route and still maintain production requires agile assets. Agile assets are defined here as having a large operating envelope and contain the components required to generate a large range of potential configurations. An agile plant is also capable of rapid changeover so that production can be easily adjusted to meet market demands.

Supply chain management ideas are also successfully applied in the process sectors; this is evident in the bulk chemicals industries where new process plants are being built close to required raw materials. The transport of hazardous materials between sites, and even within a site, introduces additional costs into the final product and in many cases this can be avoided by making the right choices of synthesis route and raw materials.

We should remember at this point that plant design is about getting the manufacturing cost right to deliver the required quality. However, the products from the chemical industry are also sensitive to raw materials costs, distribution costs, and selling costs. Raw material and distribution costs can be affected by security of supply while *process understanding* can help mitigate the minor differences between raw material supplier's products – differences that affect the process quality and performance.

This brings us to the conclusion of whether successful process sector companies need to have lean processes to do things right, or an agile process plant that allows them to be flexible and therefore do the right things. An agile and responsive advanced manufacturing facility in the process industry would allow the manufacturer to be competitive in a global marketplace.

10.5.3
Asset Planning Options

All organizations have to manage the capital assets they hold to maximize the revenue they can generate. The following are the key questions about managing assets:

1) Do we have the right equipment?
2) Are we using it in the right way?

This is a complex decision process within any organization, as the ideal way to run new processes is often using technology outside the company's asset base. The ideal asset base also changes over time, as technology develops and a company's portfolio of products and processes changes. Therefore, we need to take a holistic and future view of developing new technologies and acquiring new assets.

The emergence of "real options" concepts within financial services can be extended here to the process sectors. In this concept, we would take equipment selection decisions that maximize the number and range of options available to us in the future.

Example: A company has two options for a new technology – option A, which analysis indicates will be the optimal technology for the new process but with limited other applications, and option B, which analysis indicates would be adequate for the new process but could also solve problems in current and future processes. The "real options" would recognize that the company is likely to benefit more in the longer term by choosing option B.

As organizations develop their asset base, they need to develop their own logic for making key decisions such as the one highlighted above. *Process understanding* helps inform the decision maker which technologies are likely to be most applicable to the types of processes the company will run in future. It is worth remembering at this point that many companies can access technology externally rather than developing their own capability through the make-or-buy decision.

10.5.4
The Contract Manufacturer

The contract manufacturer's perspective on *process understanding* has already been discussed (Chapter 11). From a plant design viewpoint, the contract manufacturer has tough decisions on the level of flexibility and adaptability that must be included in the plant and the future portfolio of products is far more difficult to predict. This means that investment in novel plant design concepts and technologies often has to be justified against a specific contract product. This leads to contract manufacturers being more likely to introduce cost-saving technologies in standard operations, for example, better heat transfer equipment or more mobile equipment, rather than in novel reactor designs.

Process understanding for contract manufacturers is particularly helpful in being able to make improvements for running a process in existing assets. This involves gaining a thorough understanding of the operating capability of installed equipment and limiting process changes to within this operating envelope.

11
Contract Manufacture

Steve Woolley

11.1
Introduction

This chapter aims to capture many of the issues, challenges, and risks that can and do exist in the often complex relationship between a customer and a contract manufacturer. It sets out to look at why companies contract and what is meant by contractor and client or customer, followed by the need for scope definition. The importance of process understanding, both before and during the technology transfer, is discussed with the inclusion of some real case studies. The chapter also looks to highlight the crucial role that both intra- and inter-company personal relationships play in a successful project.

11.2
Why Contract?

Companies often look to contractors or ''tollers'' to make general chemicals, intermediates, active pharmaceutical ingredients (APIs), or formulated compounds for a whole variety of reasons:

- to release capacity for new compounds in their own facility – this may be laboratory resources, pilot, or manufacturing plant;
- to reduce current onsite risk by contracting out hazardous stages of chemistry;
- to achieve an overall reduction in the cost of the chemical;
- to enable closure of a facility;
- to counter competition from generics;
- to look for an alternative route (with an associated price reduction or robustness);
- to access new technology – both synthetic methodology and specialist equipment or processing;
- to obtain a price for use in assessing the viability of a route, there is no current manufacturing business associated with the request.

Whatever the reason for contracting, the technical and business processes to achieve the desired endpoint can be very varied depending on the nature of both contractor and client. The level of process understanding at both the outset and at the end can also be very varied.

11.3
The Contractor

A *contractor* can be defined as *"a person, business, or corporation that provides goods or services to another entity under terms specified in a contract or within a verbal agreement."*

Typically, the goods or services required from the contractor can range from grams to tonnes of a given compound using an existing route or to develop a route for providing the material.

There exist a large and varied number of contractors spread across the world and they differentiate themselves both in the type of service offered and their size or capabilities. These capabilities may well include an expertise in areas such as fluorination, nitration, cryogenics, high potency, quality, etc.

Over the last few years many companies have looked toward the Far East, predominately India and China, to provide a low cost base for their requirements. This option has proved to be especially attractive when the process is labour intensive or generates significant quantities of waste, although many clients are still very nervous about the protection of their intellectual property. When the cost of the final product is heavily dominated by the cost of the raw materials or requires highly specialized equipment, then the Far East option is often less attractive for many companies based in the West.

The services offered can be broken down into several categories, although these are by no means exhaustive:

- Lab only facilities, providing small quantities of material often against a tight timeline. This may involve following a client's recipe or using in-house knowledge and expertise. The amount of understanding at this point may be very limited.
- Large lab and pilot facilities for kilogram samples.
- Manufacturing facilities, backed up by limited chemistry support.
- Full range of facilities from lab to commercial manufacturing plant. The so-called "one stop shop" contractor. This contractor has everything required to take a process at any stage of its life and develop, optimise or improve it further or just run a clients registered process.

If the product is for pharmaceutical use, then there is also the distinction between current Good Manufacturing Practice (cGMP) and non-cGMP manufacture. A number of the contractors in the cGMP area will also have been inspected and have accreditation by the Food and Drug Administration (FDA), Medical and Healthcare products Regulatory Agency (MHRA), and other regulatory bodies. The

understanding and interpretation of quality guidelines across the world can also be seen as an additional area of expertise and competitive edge by many contractors. This can be of great benefit to many clients who lack the breadth of experience and understanding that has been accumulated over many years and projects by the contractor.

11.4
The Client

In the same way as there is great variation in the type of contractor, there is also great variation in the range of clients looking to outsource business:

- ''Virtual'' companies that have no manufacturing capability and often very limited in-house experience of running processes at scale. The company has no production facilities and their expertise lies in the discovery of new products. The usual way forward lies in employing a variety of consultants together with an appropriate manufacturing contractor to provide the practical input and physical product. In many cases, the products discovered by these companies will be sold to a larger company as time goes on.
- Companies that have extensive lab and pilot facilities may choose to use a contractor in order to; access a specialist technology, to fast track the provision of a compound or to achieve a cost reduction for larger quantities of material.
- Large companies who have the full range of assets available in-house but they are not available as required due to other priorities.

11.4.1
Scope Required by a Client

Depending on the type of client, the scope required can cover a large array of options. Defining, agreeing, and understanding the scope is an important thing to do both at the proposal and delivery stages.

Examples of scope are as follows:

- Lab work to provide small gram quantities of material via a current route or as part of a proof of concept for a route. This route may often be proposed by the contractor based on their wide ranging experience. At this stage, the focus is usually more on providing material as opposed to any in-depth understanding of the synthesis route.
- Lab work to take a current or new route idea and develop it to provide material for subsequent trials. In the case of pharmaceuticals, this material may be well used for toxicity, clinical, or formulation trials.
- To run a small-scale process exactly as provided to give kilograms of a material; this process may or may not be the subject of process development in the future.
- Initial lab work at the contractor followed by manufacture to produce larger quantities of product. In this case, the client may want to own the process or is

just interested in a price per kilogram regardless of the route of manufacture. This option may involve the use of a contractor's proprietary technology or expertise in areas such as halogenation, chiral chemistry, nitration, cryogenics, potents, etc.

• To run a process at full manufacturing scale to make the product with no process improvement; this is often the case with a registered route for a pharmaceutical intermediate or a recipe-driven process where the chemistry has never really been fully understood. If the commercial incentive exists, then there may be medium or longer term work to optimize the process.

The client may require the contractor to provide pricing not only for the job in hand but also to indicate areas of process development (if required), to identify key cost-drivers, and to present options to a customer involving a jointly beneficial development program that could bring savings in time and cost if the project grows and larger quantities are required in the future.

It may be important to the client that the contractor not only has the technical capability, both chemistry and analytical, but also the support of an in-house hazard evaluation laboratory. This allows the evaluation work to be carried out on the site by the personnel responsible for the safety of the operators and chemists that will carry out the reaction. This provides confidence to the customer that the contractor not only has the chemistry capability, but also the understanding of the associated hazards involved.

If the product is for pharmaceutical use, then the process may well require formal registration and ultimately validation at the contractors' premises. An understanding of the procedures involved and the regulatory requirements is vital to this activity and can be a competitive advantage to the contractor.

It should be remembered that there is a need for the contractor to have the capability not only to manufacture the product, but also to have available the appropriate facilities and expertise to carry out all the associated analytical method development to support lab and manufacturing operations and validation required. This may include support to stability studies, polymorph work, and salt-screen evaluations to name a few.

In some cases, the client is not always sure of what exactly they want in terms of scope. This understanding generally develops and is clarified in conjunction with the contractors as both the proposal and delivery processes proceed.

The final proposal scope needs to contain a balance between quality, timeline, and price to deliver the contract. This is often portrayed in project management terms as the "iron triangle or triple constraint" (Figure 11.1)

The triangle highlights that the scope must be able to deliver the required quality within the agreed timeline and cost. However, it is not unusual for the scope of a project to change and then both the client and contractor need to understand the impact that potential changes will have on all sides of the triangle.

Figure 11.1   Iron triangle or triple constraint.

11.5
Technology Transfer

Technology transfer from client to customer, either at the proposal stage or the subsequent project stage, usually involves the signing of a confidentiality disclosure agreement (CDA). It is important for both contractor and client to protect their in-house knowledge and expertise. Indeed, the ideal situation is for both parties to sign a two-way CDA at the outset. This protects both the client's current expertise in the synthesis route and also the contractor's if they have unique technology or novel synthetic ideas, which they can use in the project.

In some companies this process is relatively easy, for others it can get lost in the realms of the legal department and significantly slow the whole process. The situation can then be further complicated, when the legal requirements of different countries are added. This can be a source of frustration for both technical teams, who in many cases want to get on with the job in hand.

As well as a CDA, the majority of client pharmaceutical companies will insist on a satisfactory quality audit as a precursor to placing business with a contractor. Depending on the type of business that the contractor is involved in these audits can amount to over 25 per year.

The ''transfer of technology or understanding'' comes in a variety of forms, quality, and quantity, and often at a variety of times during the relationship, the approach is very client dependent.

11.5.1
Prior to Winning the Contract

The technical information available to the contractor at this point is often limited to the chemical route and a basic process description. This may also be available at a scale totally inappropriate to the amount of material required by the client's proposal request.

The client may also decide, for whatever reason at this stage, to limit the information available, or indeed that may be all they have. In some cases, even the point of contact with the client is limited. This could be due to the business

currently being with another contractor, a site closure assessment or just the culture of the organization and how it deals with external parties.

In some cases a client provides more detailed data than the contractor can deal with and assess within the time available to respond with a proposal. It is part of the skill base of a good contractor to take the data provided and provide a proposal which clearly identifies the key risks and provides assumptions for discussion by the client. The better the process understanding that both parties have at this point, the more accurate will be the identified risks and scope that lead to the written proposal.

Access by the contractor to the client technical team to clarify points in the technical package is always beneficial at this time. It can help to determine early on the level of current understanding and thereby the risk associated with the chemistry; it can also help in clarifying any points that appear unclear or whether processing options have been tried in the past. For example, if the product is currently centrifuged, then is this as a consequence of a technical assessment or is it as a consequence of what equipment was available on the plant at that time.

Personal relationships and an understanding of working practices from past projects or involvements can often considerably help during this stage of the proposal process.

11.5.2
After Winning the Contract

Depending on the history of the project to date with the client, the level of process understanding available from the client can be very varied.

- Client has no manufacturing capability and very limited in-house experience of running processes at scale. They have developed the chemistry at a lab scale to give sufficient material for initial trials. The process as it stands at this point will almost certainly not be commercially viable and may not even respond very well to running at a pilot scale. The proposal here is a balance between doing enough for the current stage of development for the successful delivery of the product and the cost of developing the process to such a state. In pharmaceuticals, where typically the chance of long-term success is quite small, there has to be a balance between the money spent or work done and the likelihood of the drug surviving the next decision gate.
- Client has a pilot scale or manufacturing scale process which they own or have recently acquired but it has never been run at the client's facility. In this case, a significant level of process knowledge and understanding can lie with the current manufacturer. This information may or may not be readily available and will depend on the circumstances under which the manufacture is being transferred.
- Client has previously run this at scale at their own facility and are looking to transfer out for one of a variety of reasons (mentioned earlier under ''Why contract''). In this instance, current processing information should be relatively

easy to obtain; however, some of the development history of the process may have been lost or is no longer obvious.

– Information has been forgotten or lost as staff have moved on to new projects or have left the company.
– Documentation has got diluted over the years, especially if an internal site transfer has already occurred at the client or the process has not been run for a long time.
– There was a time pressure, so depth of understanding and optimization were never there.

11.6
What Makes a Good Technical Package?

The technical package, both during the scoping stage and in the subsequent project stage, together with personal interactions is vital to the success of the project for both the client and contractor.

In a perfect world, there are great many pieces of information that make up a full technical package for a process that is to be run at a commercial scale. In reality, due to the need to get a product made and be available for sale, this data is generally accumulated through the life of a project and so is only ever fully available once commercial manufacture has been in operation for a period of time.

• Process chemistry
  – Main chemical route
  – Knowledge of chemical and physical rate processes (e.g., Britest methodologies)
  – Impurity generation
  – Development and campaign reports
  – Critical process parameter or quality by design (qbd) work.
• Hazard evaluation (Chapter 3)
  – Thermal events via RC1 or other such equipment
  – Stability, differential scanning calorimetry (DSC), and accelerating rate calorimeter (ARC)
  – Dust explosion potential of raw materials and products.
• Process flow scheme
  – Block diagram
  – Mass balance
  – Equipment details of reactors, isolation equipment, and so on.
• Process record sheets
• Peoples' memories or anecdotal evidence and stories
• Analytical methods.
  – Method development
  – Validated methods
  – Availability of reference standards.
• Specification of what is required and can be reliably produced.

On many projects it can be very beneficial to facilitate a face-to-face discussion between the two teams early on; this can avoid potential duplication of work done in the past (which may have been forgotten until now) and clarify points related to the current work. This discussion can also help to jog forgotten memories and also cement a trust between the client and contractor teams.

11.7
Client Process Understanding

The level of chemistry and process understanding that exists in the client can be very variable. This is not surprising as in some instances the specialized role of the clients' R&D is to discover new molecules and not necessarily take them through to manufacture. Typically, the road from conception and discovery to commercial manufacture involves several people or teams of people all of whom possess different as well as complementary core competencies. Indeed it is of great benefit to the products life and speed to commercialization, if it is passed from team to team in a controlled manner.

In the case of the contractor delivering just the product, then the client may have no fundamental interest in the process route or technology and indeed the contractor may want to keep it a secret as proprietary technology. In this case all the process understanding, which in some cases may not be extensive, will rest with the contractor.

If the process route belongs to the client, then the contractor might hope that a high level of understanding exists in the client. This, however, can prove to be very variable due to several reasons as given below:

- The level of understanding available via the client can greatly depend on the history of the product.
- The process or molecule may not have been developed by the client from the start.
- Product inherited as part of a merger/takeover or bought from elsewhere or reenergized.
- Although belonging to one client, the product has been worked on by an array of contractors over the years. This has resulted in a loss of knowhow/anecdotal data due to an inadequate transfer of process understanding between various contractors and client over the years.
- The process development team has since been disbanded and the process has been run on the plant for many years only with local support. This can sometimes result in the process ''drifting'' away from the original as it gets tweaked and much of the original reasoning being lost.

The development time for a pharmaceutical product can be very long and involve several groups of people in developing different parts of the synthesis route. The overall co-ordination of this will vary from company to company and due to both the time involved and the inevitable change in resources that will occur, the

documenting of the development history and the reasons behind decisions have to be carefully captured. It has to be remembered that no new product makes money until it is on the market, so there is always a balance and a pressure between development time and launching the product.

By using its expertise, it is important for the contractor to be able to identify the different types of information it is getting from the client. On many occasions it will be a mixture of scientific fact, opinion, and myth. It can be very difficult to differentiate between them and face-to-face meetings will always help with this differentiation. The difficulty is often further complicated by the need to work across continents and cultures with documentation in a language that is not common to both client and contractor.

11.8
Case Studies

The following case studies are real examples related to projects that the author has been involved in at some point in the past:

1) A few years ago there was a technical transfer review (using the Britest toolkit to facilitate the discussion) with a client and the process had been developed to pilot scale by the client over a number of years. As the review was held at the client's premises, there were significant numbers and diversity of people at the meeting. As the review proceeded and the function of each reagent and solvent came under discussion, the role of the added iso-propyl alcohol (IPA) came up. The solid product being produced was to be run at a 500 kg batch size by the contractor and involved the use of a very significant amount and variety of solvents. During the course of the work approximately 1 Mt of IPA was added per batch to the reactor. Within the technical package provided and during initial project discussions, there was no reason given for this addition. The client's current development or pilot team at the meeting had no direct knowledge as to why this had been added, until a semiretired chemist who was present at the meeting said "we panicked!"

When questioned further, it transpired that during the scale-up lab trials to make early clinical material an undesired and unexpected impurity had been produced. There had been a lot of pressure to solve this quickly which resulted in the addition of some IPA to the reactor to fix the problem. From that day on the IPA became part of the process and the quantity had certainly not been optimized.

During this valuable face-to-face meeting, the level of process understanding was discussed in detail. This resulted in a strengthening of the trust between the two teams and ultimately led to a realization that the current route was suboptimal and an alternative route proposed by the contractor was piloted leading to a superior product in terms of purity, chirality, and cost. The chemical usage dropped from 60 to 27 tonnes per tonne of product. This

was an excellent example of success by collaboration between the client and contractors teams.

2) On another occasion the commercial registered process that was supplied by the client to a contractor required a large number of large vessels during the final stage to wash and separate the product layer several times. To achieve the desired plant velocity, there would have been the need to have six large stirred vessels available just to carry out the washes and separations. Although this was a commercial product, it appeared that no significant steps had ever been considered by the client or other contractors to streamline or improve this step in any way by the use of countercurrent extraction technology. Indeed, sometime later, an old R&D report from the client showed that the original lab process required the solution to be washed several times in a beaker and the wash liquor to be decanted off. In this instance, the particular contractor involved had some expertise in countercurrent liquid extraction and applied it to good effect resulting in the saving of several large vessels while maintaining plant velocity.

The process was successfully run at a multi-tonne scale using a small countercurrent column and finally broke the direct scale-up approach between the original lab process and the optimal plant process.

11.9
Winning and Delivering the Project

For a contractor, there are basically two parts to the overall business process – the first is winning the business and the second is delivering the completed project to the client. The time between starting the discussions and delivering the business can be either quite short or may extend to many months. It is not unusual for there to be several changes in scope during that time as each side adds more detail and the scope clarifies, or the business circumstances change.

11.9.1
Winning the Business

At a high level, this simply involves taking the client's requirements and their current level of understanding together with that from in-house sources and providing a price for the service. Sounds simple.

It is probably worth mentioning that the contractor is providing a level of service and this does vary from contractor to contractor. Depending on the relationship between the client and contractor, the winning of the business can be based on a whole range of both simple and complex bits of both commercial and technical information:

- Is this a one off project or is it likely to grow and repeat?
- Will there be room to develop the process further as time goes on?

- Is this part of a bundle of business with this client?
- Is the price to be used for costing purposes only and there is no immediate "real business" attached to the price request?
- What is the current development or market status of the product?

Having received the technical package, the evaluation is typically carried out by a group of experienced people at the contractors. The areas of consideration during the evaluation, some or all of which may be understood and available, include the following:

- The state of understanding of the chemistry, is there a risk attached to the yield, side reactions, etc.
- The contractors experience or lack of it of this type of chemistry and processing.
- What scale of processing is best suited to both the client and the contractor in terms of batch size and timing? How well will the process scale from its current scale to the new scale?
- Reaction hazards associated with both the chemistry and the raw materials utilized in the overall process.
- Toxic hazards associated with the process, what is known and more importantly not known. There is obviously a serious regulatory requirement in Europe and the USA to protect the employee and the environment. Where data is not available, then an assessment needs to be made by the contractor and shared with the client as appropriate.
- Waste streams and ease/cost of disposal or on-site treatment/recovery, this is becoming more expensive and specialized in Europe and the USA.
- With the changing legislation on solvents (for example, the Solvents Emission Directive in Europe, red list, etc.) it may no longer be possible to economically produce a compound using the existing or previously used solvents. Moving a process to a different location is often more difficult in terms of emission compliance than leaving it where it is. Solvents that were common place (carbon tetrachloride, chloroform, benzene, etc.) are now being increasingly regulated and can cause problems in terms of emissions at a new location. This means that the transfer of a process frequently is not just about the chemistry and the processing equipment: it can contain a significant element of abatement costs, for example, scrubbing of atmospheric vent streams.
- Registration, Evaluation and Authorization of Chemicals (REACH) responsibilities in Europe for registering both substances and mixtures that you may import or produce.
- The batch time based on the information provided and past experience of the contractor across an array of jobs. The contractor will have built up considerable in-house expertise in terms of timings and risks associated with many unit operations such as distillation, azeotroping, filtration, and drying.
- The availability and scheduling of the processing equipment versus other operations known and forecasted by the contractor.
- How well will the process fit the available assets and what modifications may need to be made to accommodate the process.

- Materials of construction required.
- Filtration and drying, what equipment has been used to date and why, what is available now. This particular data is often anecdotal and heartfelt.
- Are the analytical methods available, do they work, what equipment is required to run them. A typical API may require in the order of 15 separate tests to release the batch against an agreed specification. At the evaluation stage this is an area, that is, often unclear or left until later.

Having taken account of all of the above, together with their experience, the contractor will calculate the cost in terms of raw materials purchased, plant time utilized, manpower, and other miscellaneous items. This will provide a view of the costs he will incur should he win the project and the level of risk involved at this stage based on current understanding.

To this cost is added a profit margin and, this, together with an appropriate level of supporting information is communicated back to the client for their consideration. This may be in the form of a written proposal, completing a client-provided template or verbally. It is important for the contractor to be clear in their scope and assumptions and for the client to ensure that they compare competitive quotes on a like for like basis.

It is not unusual at the kilo proposal stage to ask the contractor for an estimate of long-term pricing. It is very important to state the assumptions on which this price is based, so as the risks associated with a given estimate are fully understood and can be compared between competing contractors on a like for like basis. In this instance, the client needs to understand that each contractor will almost certainly have a different level of process understanding and risk assessment based on the data provided and their own internal expertise. This can then be further compounded by assumptions on raw material prices at larger quantities.

11.9.2
Delivering the Scope

Once the client has awarded the project, both the contractor and the client organizations move into delivery mode.

As mentioned previously, it is very important at the start of the project to agree to the scope. This is obviously easy if the client is expecting a quantity of routine material (to an agreed specification) made using the contractor's in-house technology to be delivered by a certain date.

A large proportion of contract work, especially in the pharmaceutical and speciality chemicals industry, is significantly more complicated than this!

The project will often involve the setting up of project teams from both organizations with agreed lines of communication. A very useful technique to ensure common understanding is to arrange a joint project kick off meeting involving the client and the contractor teams. This allows the two project teams to get to know one another and provides a forum for sharing of information assumptions and

risks. It is often surprising what information comes from this meeting and both parties should not underestimate the benefit of good personal relationships and trust, which can arise from this meeting. Setting up good and effective communications between the project teams may also involve handling the challenges of language, culture, and time-zone especially between East and West where limited time windows exist during the normal working day.

It is at this point that reality often arrives and the very fact that the project is now live results in a real need for more clarity on quantity, timing, specification, and what the client and contractor actually do and do not understand. This meeting is often the first opportunity for both parties to request or be offered supplementary information. Some of this supplementary information can lead to greater clarification and a generally perceived lowering of risk or indeed vice versa. This is where good inter company relationships are important between the teams.

The opportunity for the contractor to see the process operating at the client's facility should always be looked upon as a bonus by the contractor and should be taken up whenever possible.

As the project starts, there is the initial technical challenge related to the inevitable changing of scale and equipment and delivering to a timeline.

In every case there is obviously a requirement to get the processing right first time, however, the amount of time required to fully understand a process for scale-up and the time available for delivery do not always go hand in hand.

One of the most undesirable things that can happen to both the contractor and client is the need to rework a batch of product that has failed specification, in some cases by a very small amount due to a lack of process understanding. This often takes both parties into unknown territory in terms of developing a rework procedure often against a very tight timeline with the inevitable loss in yield.

The sourcing of raw materials for a project can often be a major constraint on the start time for the project. During the proposal process, the contractor will have often made assumptions as to the availability of the principle raw materials. Depending on the scale at which the process has previously been run and the understanding of the customer, the raw material may or may not be readily available. In some cases, both the client and the contractor may also be initially unaware of the difference that small variations in the purity or impurity profile can make to the final product. This is especially important to identify early on when the chemistry has only been run previously on a small scale using laboratory reagents. It is normal for the contractor to obtain an early representative sample of the critical raw materials and use to test these in the lab. In some cases, this is an integral part of the release testing for that particular raw material.

In addition to transferring any process understanding from the client to the contractor regarding the chemistry there is also any information regarding analytical methods and the provision of reference standards.

These methods relate to in-process tests, intermediates, and final product tests to confirm that the material has been made to the required specification. In the case of an established product, then the contractor would expect to receive well-documented validated methods suitable for transferring together

with appropriate reference standards and markers. For less well-established development projects, the analytical methods and often the results will need to be the source of significant discussion as the project progresses.

11.9.2.1 Case Study

Following the initial project meetings via teleconference the client invited the contractors team to witness a pilot trial at their facility as a way of seeing the equipment and meeting the team. The time spent in unhurried discussions between the teams and also the team building that occurred was very beneficial. The meeting also resulted in the gathering of what appeared to be very minor pieces of information that would prove crucial in the way the process was transferred. One of these involved the potential for the process to ''produce'' HF on acidification of the reactor solution at high temperature with hydrochloric acid. This unexpected observation came to a head, while the visit was going on and enabled a pooling of data and a joint face-to-face brainstorming session. It transpired that the initial understanding of what was happening by both client and contractor was limited and it took several months of laboratory work to fully understand the mechanism and find a way to limit the HF production such that the process could be run safely at scale without damaging either a glass-lined or a hastelloy reactor. The teamwork coming from the initial brainstorming and the recognition that the joint teams owned the problem finally led to it being solved.

11.10
Project Timing

The contractor, as a service provider, is well versed in the running of projects and will usually have many established processes and new processes running in parallel at any moment in time. The management of this portfolio in terms of resource availability is part of the contractors expertise.

When the project proposal was originally discussed with the client, the contractor would have made various assumptions based on the technical package and other information. From this information and maybe some further assumptions, a timeline would have been proposed. Ideally, this timeline should really be treated as an elapsed time based on the customer's decision point to proceed. It may, in many cases, also have the added complexity of being tied to plant availability.

As part of the project kickoff meeting any revised time plan can be shared and agreed. It is then the task of both project managers to keep the project to time.

The contractor needs to keep the client up to date with progress against this time plan and also be flagging any points at which input from the client is critical to the achievement of the overall timeline.

In general a time plan should always contain some degree of contingency; this is a prudent measure as on most projects a problem will occur, which was

unanticipated. Basically the plan should allow for something to happen, although at the start of the project you almost certainly do not know what it is or where it will happen! The manufacturing plan will normally consist of two distinct types of task, those required to be done prior to manufacture starting and then the batch plan for the manufacture.

Many books have been written on the principles of project management and it is not considered to be within the scope of this book to discuss these in detail.

At the conclusion of a jointly managed project, it is best practice for both parties to meet up face to face to review the project. This post project meeting should be seen as an opportunity for feedback related to the success or failure of a project. The meeting should capture what worked well and what did not from both sides. It may be that the process was not as well understood as the client's thought or that the contractor failed to implement it successfully. The object of the meeting is to enable both parties to learn and improve in relation to service and also help in the overall relationship between the companies. In general, a face-to-face meeting is to be preferred whether the news is good or bad.

11.11
Challenges of Multiproduct Plant Scheduling against an Uncertain Background

The scheduling of processes in a large multiproduct facility which may contain up to 30 or 40 process streams is a complex operation. The near time part of the schedule is involved with delivering confirmed business, whereas the remainder of the schedule will be a mixture of confirmed and as yet unconfirmed business.

The managing of this ever changing real and potential schedule is something that lies within the contractor's area of expertise. Every company has its own way of accomplishing this and also making it visible within its own business.

However, it will involve considerations of things as follows:

- planning, including dealing with possible conflicts and managing deliverables so that a reserved plant slot is not delayed;
- forecasting of future activity as accurately as possible;
- tracking plant availability against unconfirmed demands;
- changes in the clients' quantity requirement and delivery time. Skill of contractor to manage and see or suggest ways forward;
- cleaning from product to product; on some very short pharmaceutical campaigns it can take longer to clean the plant than to actually make the product;
- managing unexpected delays caused by raw material delays, other project interactions and the chemistry not performing as expected by contractor and client.

While the client often does not see this complexity, as they often only have visibility of their project, it is up to the contractor to manage this complexity as part of their service to its large number of customers.

11.12
Conclusion

The business and technical work that goes into the winning and delivering a product to the customer is a complex combination of shared process understanding and personal relationships. For many products, the level of process understanding may not be as high as both contractor and client would like, or alternatively they may believe it is high until something unexpected happens. The importance of team work, flexibility, and good interpersonal relationships is vital to the success of the overall project, particularly, when sharing and managing the risks associated with the manufacture of chemicals at any scale from grams to tonnes.

12
Whole Process Design

Paul Sharratt

This chapter examines the challenges of capturing and representing process understanding for ''whole process'' design in the specialty and fine chemicals sectors. First, various relevant tools and concepts are reviewed. Then, the stages of whole process design (WPD) are explored, making explicit the key decisions.

12.1
Process Understanding for Whole Process Design

A chemical process typically consists of a combination of processing steps. Usually, we need to do more than simply mix two reactants together – there will also be feed preparation, addition of ancillary materials such as solvents, heating and cooling, product separation and purification, waste treatment, and many other operations. However, in many parts of the chemical process industry, the focus of process development is only on parts of the process, without fully recognizing the contribution of other parts to the overall outcome. For example, in organic chemical synthesis, there tends to be an emphasis on maximizing the yield of the reaction, implicitly assuming that the recovery from the post-reaction mixture is not as important and that recovery will be maximized by simply maximizing chemical yield in the reaction step. One problem with this is that maximum reaction yield does not necessarily correspond to the most profitable performance. The reality is that the yield losses can be as much or larger in the separations. The separations may also require a larger part of the total capital expenditure; a recent survey [1] indicated that each reaction required several processing steps to recover the product, with about 40% of processes needing five or more steps – each of which could require as much capital equipment as the reaction itself. Further, achievable performance in some parts of the process can be strongly influenced by performance in other parts. For example, the ease, cleanness, and cost of a separation may be influenced by the conversion of the reaction, particularly if overreaction generates troublesome products such as tars, close analogs of the product, or surface-active materials.

An example of this from a real process (disguised to protect confidential information) involved a nitro reduction to the corresponding amine using hydrogen

Figure 12.1 Anonymized reaction scheme for nitro reduction reaction.

Figure 12.2 Graph of concentrations with reaction progress. Note that the dechlorinated species is shown at 10 times its actual concentration for ease of viewing.

over palladium on carbon as shown in Figure 12.1. This was carried out using Pd/C and hydrogen in ethanol at less than 10 °C. The reaction yield was high, but a by-product formed by chlorine removal was very difficult to remove and resulted in 20% loss of the desired material in the separation process. The overall process cycle time was about 35 h.

The conversion profile for the reaction step is shown in Figure 12.2. Note that the des-chloro impurity was formed only at the end of the reaction after approximately 85% conversion of the raw material.

The team charged with improving process performance identified that stopping the reaction at 85% conversion allowed a much improved separation step, with an overall cycle time of only 12 h and a recovery of about 97% of the desired product. Although the des-chloro impurity was still present, it was below the specification limit so it did not need to be removed. The unreacted starting material was easily

Table 12.1 Whole process performance for low- and
high-conversion versions of the hydrogenation process.

Process	Reaction yield	Work-up yield	Whole process yield	Cycle time(h)
Original	97	80	78	35
New	85	97	82	12

removed by reactive extraction – using dilute acid to protonate the amine and
extract it into an aqueous phase, leaving the starting material in the organic phase.

The two processes are compared in Table 12.1. The ''New'' process produced
significant benefits, even though the reaction yield was significantly less than that
of the original process!

Other deviations from the intended processing outcome can arise from trans-
formations of materials in storages, pipes, and equipment that are unintended
and may well not have been identified during laboratory experimentation. These
issues are prone to reveal themselves at full scale, when delays between processing
steps may be longer or where the process materials are exposed to materials of
construction not present in the laboratory.

Whole process design is a philosophy that requires the process designer to take
a holistic view. It requires the capture and exploitation of a wide range of process
understanding that supports a series of design tasks. Key activities in generating
whole process understanding are as follows:

• identifying the overall performance characteristics of the process that are re-
  quired – these will likely include cost, safety, robustness, and environmental
  performance, but may have other features, depending on the product, business,
  and situation;
• determining how the ''variables'' in the process – the choice of operations and
  conditions, order of operations, types of separation techniques, for example,
  impact on those characteristics;
• generation of one or more feasible overall processes;
• developing sufficient understanding of individual processing steps to be able to
  relate their performance to their inputs and outputs;
• identifying interactions between the processing steps;
• finding the best set of operating conditions.

To adopt this whole process philosophy, it is important to have tools and
techniques to capture and exploit the underpinning process understanding. Some
sectors of the process industry have well-established tools that can readily allow
process developers to take a whole process view. In the oil and petrochemicals
sectors, the use of process simulation is ubiquitous to model of the performance
of the plant. However, the penetration of similar tools into the conventionally
batch sectors, as well as formulated products, is low. In these sectors, the ''whole

process'' representation will typically be limited to a chemistry recipe together with a spreadsheet containing the mass and energy balances. There may be some reports on ''the chemistry'' at quite an abstract level and some reports on conditions/situations to be avoided because of operating and safety problems that have been identified. *However, much of the core understanding may remain in the heads of the designers.*

It is also usual to explore whole process performance by carrying out laboratory-scale experiments on the whole process, and possibly pilot studies. This is inevitably quite late in design and requires most of the whole process design decisions to have been taken before it can be carried out. It is unsuitable as a way of examining the very large number of feasible process options, as it is costly and time consuming.

Good process understanding allows us to predict the impact of changes in feedstock, operating conditions, throughput, and many other foreseeable events. It also allows us to manipulate the process to achieve modified outputs – increased throughput, higher purity, for example – with confidence that the changes we make will result in the desired outcomes. However, process understanding is generally never complete. On occasions, we can make very precise and accurate predictions, for example, the variation of boiling point of a liquid mixture with composition. On other occasions, we may only be able to make crude estimates of the likely outcomes or even only be able to suggest the likely direction of change. Of course, complete understanding is also not desirable in most processes – not only would it require very large time and cost to obtain but also much of the understanding would not support improvement in or better control of the process.

Incomplete and/or inadequate understanding can be characterized as being combinations of four distinct issues. These are shown in Table 12.2.

The last three are the most dangerous, leading to situations where unexpected behaviors are more likely to occur, rather than variations within the expected

Table 12.2 Sources of failure of process understanding.

Parametric uncertainty	We know what is happening and the mathematical equations that should govern it, but we do not have exact values of the relevant parameters to put into the equations and we therefore cannot predict accurately
Insufficient understanding	We have not identified all the governing phenomena in a process
Erroneous understanding	Where we have wrongly identified governing phenomena (e.g., believing a process is kinetically controlled when it is controlled by the rate of mixing)
Complexity	Where the interactions of the multiple phenomena present allow the system as a whole to behave in ways that are unpredictable or hard to predict with confidence, even though the core phenomena are known

behavior. Unfortunately, it is common to assume that all problems are of the first type – until an intractable problem is encountered.

At the core of the whole process design is the ability to track the propagation of influences through the process – whether this is the impact of impurities on subsequent operations or other properties such as temperature, flow, viscosity, particle size distribution and so on. The understanding required is often difficult to obtain, consisting of many different individual pieces of information, each of which may be itself difficult to investigate experimentally.

12.1.1
Process Complexity and Its Impact on Data Needs for Understanding

The amount of experimental information needed to characterize and therefore to understand a whole process rises with the number of processing operations.

The number of potential factors to consider grows very rapidly with the scope of the process and the level of complication of individual steps. Suppose we have a single simple, well-mixed isothermal liquid phase true batch reaction with two reactants (A and B react to make C) and a solvent. This first step is followed by a simple batch extraction to remove product C as a solution in a second solvent (Figure 12.3). Success is quantified as a yield of product C and its purity level in the extract. The influence of the first step on the second is through the composition, the temperature, and the time of transfer to the second step. There are at least five variables that may control the reaction's influence on the extraction: the concentrations of A and B in solvent, the reaction time, temperature, and time of transfer into the extractor. To obtain ''complete'' understanding, we would need to consider the potential impact on the second step of variation of each of these five variables. The overall outcome (yield and purity) would be influenced by these as well as the extraction parameters. If we assume that only one batch wash is to be used, then we have about five more relevant variables: the wash quantity (ratio to the reaction product), temperature, mixing intensity, mixing time, and settling time.

Even the variables of this simple process would require a large amount of experimentation to explore purely by laboratory work. However, we can use prior

Figure 12.3 Relevant variables for a more simple process.

knowledge of the chemistry and chemical engineering to ''model'' the behavior of the system – either conceptually or mathematically. In effect, this reduces the number of independent variables and it becomes practicable to gain a good understanding of the system. For example, using a stoichiometric mass balance of the system and reaction kinetics would allow us understand the range of possible outcomes from the reaction. That would restrict the set of experiments that would then be needed to be carried out on the extraction step. Likewise, by measuring partition coefficients for the various components, we could design the extraction to achieve the desired performance without having to try all feasible conditions for every conceivable feed input.

We can characterize the complexity in a simplistic way by estimating the number of experiments required to explore the space using simple two-level ''high–low'' factorial experimentation. We might carry out the reaction with low and high concentrations of A, low and high concentrations of B, low and high temperatures, and so on. (Obviously, this is a gross oversimplification but helps to make the point.) For 10 variables with experiments carried out at two levels, this would be $2^{10} \approx 1000$. If we had stoichiometric and kinetic information about the reaction, then we could reduce the number of variables by 2 – leaving the A/B ratio, degree of conversion, and total solute concentration fed forward. This reduced the number of notional experiments required to $2^8 \approx 250$. If we have a model of the partition behavior of solutes between the solvents, then the system is almost completely defined (i.e., representable by a mathematical model). We might then only require a couple of confirmatory experiments. In other words, given some a priori understanding about the behavior of reactions and extractions, we can quickly gain a good understanding of the system. Of course, if we do not understand the basic kinetics and liquid–liquid phase equilibrium, our task is harder. The use of models to capture reaction understanding is dealt with in more detail in Chapter 5.

Consider a more realistic case (Figure 12.4) where the reaction can produce three impurities by reactions that are not fully understood, and the reaction is exothermic. Suppose that the formation of impurities is influenced by the mixing intensity in the reactor, as well as the extraction, there is a crystallization stage to recover the product after the extraction. The crystallization produces a distribution of particle sizes. The product specification will now include not only yield and purity but also a requirement for the content of fines and oversized particles.

There are now seven variables for the first stage. Further, the temperature is not a ''single value variable'' but a profile that changes with time. We could represent temperature profile simplistically as an initial, intermediate, and final temperature – that is, say that it is equivalent to three high/low variables. The second stage still has five controlling variables. The crystallization also has a profile variable (temperature as a function of time) as well as the mixing intensity. In this case, we have a much more difficult situation. The reaction product composition influences not only the outcome of the extraction but also the crystallization. Because trace impurities can influence crystallization significantly, we might need to look at the

Figure 12.4 Relevant variables for a more complex process.

impact of each of the upstream operations on the outcome. The crystallization outcome might be measured both as a yield of solid and the proportion in fines (too small), medium (acceptable), and large (too big) fractions – again three variables.

The initial step has in effect nine variables. Even having the main kinetics only helps a little, as we do not know the impact of the conditions on the impurities. We would need to consider also the five extraction variables and the three crystallization variables, one of which is a temperature profile. The total is 19 variables or about 500 000 experiments! Further, it is much harder to use modeling to reduce the complexity. Modeling could certainly still handle this problem, provided the input data were available, but we now need further information on the kinetics of the impurity reactions, more partition coefficients, and, most difficult, a quantitative understanding of the impact of impurities on the crystallization.

If we added a few more processing steps, it is clear that the result would be a level of complexity that would require a large experimental program combined with extensive modeling. This ''combinatorial'' explosion of complexity is the main reason that complete understanding is essentially impossible for all but the most trivial of processes. It is also why many processes pass through regular separation and isolation stages. At each isolation (typically in pharmaceuticals manufacture, this would be crystallization, filtration, washing, and drying), we are trying to produce a material that carries forward no unwanted features that impact on the following stages. This approach attempts to decouple the prior stages from the subsequent ones. Of course, it is hard to achieve this as even trace impurities may still have significant downstream effects; crystallization, for example, may be greatly influenced by the presence of trace quantities of materials similar in structure to the desired product.

12.1.2
Process Design Philosophies

In the last few years, there has been increasing interest in improving the performance of processes, with developments in green chemistry [2], process intensification [3], microreactors [4], and continuous processing (5, 6). In principle, all of these philosophies have something to offer to chemical processing. However, they have in general had little impact on manufacturing practices and processes in the traditional batch sectors. One of the key reasons for this is their frequent failure to address whole process outcomes, focusing rather on reactions and reactors. While it may be possible to use a much smaller reactor than a batch stirred tank, this change might bring little benefit if, for example, the whole process cost is dominated by a crystallization that has to be carried out batchwise because of its low rate. Whole process design allows targeting of novel approaches and concepts in areas where they bring business benefit. It attempts to enable ''market pull'' as a driver for process development and innovation rather than ''technology push.''

12.2
Process Outcomes

It is not true that the only criterion for a successful process is manufacturing cost. In most cases, there will be a wide range of desired outcomes – and in many cases, these may be more important than squeezing the cost to an absolute minimum. Throughout process design, it is essential to know (and keep in view) the key success criteria for a project. These may be any of or a combination of:

• capital cost of the project
• operating (manufacturing) cost
• scale of production
• seasonal or other variability in production and the need to produce at variable rate
• time to market for the product
• qualities of the product (purity, physical form, particle size, etc.)
• scalability of production
• establishment of production equipment that can be used for other products (multipurpose plant)
• demonstration of a new chemistry or technology of strategic importance
• robustness of manufacture (high reproducibility of manufacture)
• ability to cope with variable feedstocks
• ability to adjust the recipe and process
• safety, health, and environmental performance
• other criteria specific to the product, business, or situation.

This set forms a basis against which the process understanding needs can be judged. Different criteria will push design in a particular direction. Given that

complete process understanding is essentially impossible, the purpose of design is to provide sufficient understanding that there is a low risk of failing to meet the criteria.

Douglas [7] looked at the design of continuous petrochemical processes, where capital and operating cost are the key. For those cases, a simple measure (revenue minus the sum operating cost and annualized capital cost) was sufficient to discriminate between options and to indicate the viability of the project. He advocated the reconsideration of the value of the investment at each stage of design, and if at any stage the economic criteria were not met, this showed the design had become nonviable. This approach works well in areas where cost is the main driver, but when there are additional criteria, it becomes increasingly difficult to maintain a single criterion for process outcome. Of course, many of the criteria can be converted to a financial cost; for example, time to market can be represented in terms of opportunity cost. However, it is often difficult to do this and alternative approaches are often used.

Multicriteria decision making is often used [8], where the process options are rated against the outcome criteria using expert judgment. Alternatively, the project may set limiting values and characteristics for the nonfinancial outcomes and simply find the cheapest way of achieving those. At an overall project level, many companies use a Stage-Gate® approach, with the project (and associated process/plant design) having to meet a set of criteria at each "gate."

However, the list of criteria is assembled and used to support decisions; it is important that it is visible to the technical team who are taking process development decisions. Without a clear view of the design targets, or with an overly simplistic view as to what is to be achieved, there is a strong likelihood of taking poor design decisions.

12.3
Organization of the Design Activity

Process design always passes though a number of stages. Often, these stages are formally aligned with a company's internal decision-making processes. For example, each stage might be terminated with a go/no go decision, as part of a Stage-Gate® process for the progression of projects. A set of typical chemical process design stages is shown in Table 12.3. Note that different organizations and different business sectors will use different stages and give them different names. Also, stages may be combined. For example, it would be usual for the process concept stage to be combined with process development, particularly if the use of batch processing is assumed from the outset. Issues associated with these various stages are discussed in more detail in other chapters.

At each stage, several "degrees of freedom" are available and the activity of that design stage is to fix them – ideally definitively. It is not desirable to carry forward multiple process options between stages, as this substantially increases the cost and time required for design. Typically, the cost of process and plant design is in the

Table 12.3 Typical stages in whole process design, with associated degrees of freedom.

Stage	Typical degrees of freedom	Understanding typically needed to support decisions
Route selection	Which main chemical reactions will be selected for the synthetic route Selection of raw materials	Business needs/criteria Wide chemistry knowledge Knowledge of ''manufacturability'' of chemistries
Process concept	Batch/continuous Make/buy Intensive/traditional Candidate manufacturing operations	Cost data for raw materials Business needs/criteria Basic kinetic data Understanding of chemical complexity/difficulty of proposed process Capabilities of different available processing equipment Existing facility capabilities
Process development	Compositions Addition profiles Conditions Equipment type Separation techniques Solvent selection	As for process concept, plus Additional physicochemical data (process specific) Comparative performance of process under different conditions Individual equipment performance capabilities under required conditions General chemical engineering and physical chemistry
Flowsheet design	Final major equipment selection Equipment sizing Schedules Layout concept Control concept	Equipment detailed design – may require further physicochemical data
Detailed design	Vessel and pipe finishes Pipe sizes Minor equipment selection Detailed layout Control and electrical details	Company standards Corrosion/compatibility data Design of existing infrastructure

range 10–30% of the project cost, and the effort available for process development is limited. It is thus not feasible for two or three alternative designs to be taken through and a comparison made on the basis of complete information. Similarly, once a decision is taken in design, the cost of reversing it rapidly rises with time. This is illustrated in Figure 12.5.

It is important to note that the degrees of freedom depend on the project. For example, the chemical route may already be fixed, as is often the case for toll

Figure 12.5 Costs, effort levels, and decision impact in process design.

manufacturing. Another typical constraint is the desire to use existing process vessels.

12.4
Risk and Uncertainty in WPD

Inevitably, it is not possible to examine all possible process alternatives or to obtain absolutely accurate physicochemical data. This in turn leads to incomplete or erroneous understanding, as already discussed. It is thus important to have a rational approach to deal with lack of understanding. This ''risk analysis'' should play an important part in guiding the direction of process development – with work being targeted to resolve serious uncertainties and ultimately to give confidence in the final design.

One simple and widely used approach for prioritizing uncertainty issues is the seriousness/likelihood analysis. Once issues have been identified, they are rated (often by expert judgment) as to their potential impact and the likelihood that they will impact on design. They can be plotted on a ''risk matrix'' as shown in Figure 12.6. Clearly, serious problems that are likely to occur must be resolved – particularly if they are potential ''stoppers'' that would make the whole process nonviable. For serious issues that are judged to be unlikely, it is appropriate to seek further understanding. This may well be aimed at understanding whether the problem really exists – that is, either moving it to the serious/high-probability category or removing it. For the low-impact problems then, effort may be more directed at confirming that they really are of low impact, since if they are then their likelihood is of less importance.

Ideally, risk analysis would be carried out at various stages of design, as new uncertainties appear as understanding grows. Early on, the main uncertainties are likely to be around whether the chemistry can be made to operate at a viable yield and whether recovery of product at the required purity and good yield is possible. Later in development, smaller issues will appear – for example, suitable materials

	Low probability or frequency	High probability or frequency
Serious impact	Gather further understanding and resolve if necessary	Resolve as a priority
Minor impact	Ignore	Ignore or resolve if effort allows

Figure 12.6 Typical risk matrix and associated actions.

of construction for the plant, deposition of solids in lines leading to blockage, and so on. The smaller issues are less likely to be terminal to the project (though they can be) but are likely more numerous. They are generally picked up (or should be) through tools such as Hazard and Operability (HAZOP) and Failure Mode and Effect Analysis, which are widely used to explore the detailed design for safety and operability problems. However, depending on the timing of those studies, the identification of problems may be rather late and lead to significant project delays.

It is worth remembering the different types of uncertainty when deciding what should be done about a particular risk. If the risk is associated with parametric uncertainty, then better measurement (in both the laboratory and the plant) and the use of modeling will provide a good solution. However, in the other cases (where understanding is incomplete or simply wrong), we have more of a problem. The usual approach in most cases is to repeat the process in the laboratory several times, looking for reproducibility. This is sometimes done using different values of key processing parameters to identify an ''operating envelope'' within which the process should give the desired outcome. If, however, we have incomplete or erroneous understanding, we do not have the correct list of controlling parameters, and problems may not be identified. Further, the large number of relevant variables may mean that important areas of the envelope are simply not mapped.

A key activity in carrying out trials of the process is to look carefully for outcomes or behaviors that cannot be accounted for by the current level of understanding. These form evidence that either the measurements/observations are flawed (which is clearly serious) or the understanding is incomplete. Of course, the larger the number of experiments, the better the chance of identifying ''unexplained'' behaviors. To look for unexplained behaviors, it is also important to have an idea of the accuracy of the experimental results, so that random errors are not identified as behaviors. In many companies, the two key activities – assessing experimental and analytical accuracy and looking for unexplained behaviors – are paid little attention. Unfortunately, these are at the heart of being able to make claims of genuine understanding.

To deal with uncertainty and lack of understanding, it is common to introduce a number of features into processes:

- isolation of intermediates to a high level of purity to prevent propagation of problems through the process (decoupling consecutive stages);
- use of excesses of reagents, long processing times, and extremes of conditions to ensure that the process is definitely forced to the desired state;
- in-process testing.

All of these have a significant cost and productivity impact – in effect, the cost of dealing with the risk arising from incomplete understanding.

12.5
Whole Process Representations

Traditional process representations have grown from the individual disciplines that contributed to design. Process recipes are the chemist's representation and are, in essence, the instructions to carry out the chemistry. Flowsheets are the set of connected unit operations used by the chemical engineer. Line diagrams also come from engineering – this time mostly mechanical engineering, to show how to put the plant together. These traditional representations, while being useful for the tasks that spawned them, have significant limitations and are at best partial. Recipes do not fully represent processing conditions – for example, it would be very unusual for a recipe to say how intensely materials should be mixed; it is just that the agitation should be ''on.'' Flowsheets assume the palate of processing equipment available, so tend to restrict thinking about innovative options. Line diagrams are purely mechanical and do not consider chemistry at all.

For the crucial early parts of whole process design, we need representations that have a number of features:

- They should represent explicitly the degrees of freedom that exist at that stage of design.
- They should represent the connectivity of the process so as to be able to deal with propagation of materials and conditions (and thus problems) through a process.
- They should be accessible to all of the disciplines involved at the time.
- They should be flexible in use, both to a variety of different process types and to use with different levels of understanding – for example, qualitative, semiquantitative, model-based, or empirical representations of process behavior.
- They should be easily linked to more detailed levels of analysis later in design.

A number of representations that meet some or all of these criteria exist, and some are described below.

The terms input–process–output (IPO) diagrams or hierarchy–input–process–output (HIPO) diagrams come from a method developed by IBM in the 1970s for software design and documentation [9]. Essentially, the method breaks down a process into a series of consecutive processes, each of which may have material,

Inputs
Solvent
Reactant A
Reactant B
Heat
energy

Process
Add solvent, A, B
Mix
Heat to 50°C
Wait 1 h
Discharge

Outputs
Solvent
Reactant A
Reactant B
Products C
Unknown
by-products

Figure 12.7 Simple IPO diagram.

control, and information inputs and outputs. The process is responsible for the transformation of inputs to outputs. The attraction of this representation is that it allows for arbitrary amounts of information and knowledge. Also, there is a clear focus on the inputs and outputs that connect process stages. A simplistic example is shown in Figure 12.7.

A variety of other related representations have been proposed and are used. One more advanced version is the Process Definition Diagram that allows representation of additional important process features such as the nature and number of phases (gas/liquid/solid) present in a processing step and the contacting pattern of the materials involved [10].

Quality function deployment (QFD) is a technique that originated in the engineering industries for developing complex products ranging from jet fighters to oil tankers. It recognizes that a product can have multiple desired characteristics and that these are controlled by the many variables that determine the processing (process conditions, processing times, etc.). QFD involves the creation of matrices that link the desired characteristics to the conditions and process variables. While perhaps being more easily deployed in the design of production facilities, the tool can be adapted to map the interrelationships in complex processing tasks [11].

Computer-based representations of chemical processes, including batch processes, are readily accessible in commercial software such as Aspen Batch Process Developer, SuperPro Designer®. These can be useful, particularly in the later stages of design, but have a significant training requirement for a user to become competent. For this reason, spreadsheet-based representations of process are still widely used. A general problem with specialist software is that it requires a uniform level of information across a process, or if that is not available, for assumptions to be made to allow the calculation of mass and energy balances, device performance, and so on. Unfortunately, the process of making assumptions is often time consuming and forces the user to embody unrealistic assumptions in order to obtain results. Also, where the programs have ''default settings,'' these may become implicit assumptions without the knowledge of the user. One critical aspect is to have information accessible to all disciplines, so that decisions are more explicit and people have the information they need to understand the implications of decisions they may take.

General-purpose modeling codes such as MATLAB® and gPROMS are often the tool of choice for specialist modelers and can be useful where the benefit in process understanding (and hence performance) warrants the very high levels of effort to

generate detailed, faithful models of the process. Such tools are more likely to be used to investigate in detail one part of the process where there are difficulties than to be used for whole process design in the low tonnage sectors (Chapter 5).

Statistical models such as response surfaces that link the process outcomes as a function of inputs are also widely used, although by themselves they do not represent understanding – rather simply correlated data. A wide range of products is available for the design of multivariable experiments, as well as the presentation and analysis of the large data sets that typically come from process monitoring. They do bring benefits to whole process design. Clear thinking often emerges from the discipline of experimental design, and multivariable statistical methods can identify patterns that motivate the search for relevant understanding. However, statistics are only valid if the data used are complete and accurate. Ensuring completeness itself requires the use of understanding. Something about needing to spend time thinking where to start and what parameters to look at is important. Where to start your design space is covered in Chapter 1 but is worth mentioning here.

12.6
Decision Making in WPD

Many decisions are made in whole process design – both about the process itself and in the direction of the design activity. The latter is often ignored, with companies working to a ''template'' approach with the same activities programmed for equivalent stages of process design. However, there is value in being able to adjust the design process to meet business needs and to trade off cost, time, and risk.

12.6.1
Decisions about the Design Activity

Table 12.4 outlines how the activities required and tools used might vary according to the extent of process understanding sought. The table reflects a key point – as we seek more understanding, we see an increase in the need for experimental data, prior knowledge, and tools. For low-risk, well-understood chemistries, a company might well want to carry out a single experiment and accept the risk that the large-scale process does not perform similarly. Normally, multiple experiments are carried out to confirm process robustness. Only when the process outcome is critically dependent on the performance of a particular unit would it be sensible to invest in a detailed analysis and model of that unit. Likewise, the very high time and cost of producing a detailed whole process model would only be justified if both the performance of individual operations *and* the coupling between them were critical to the outcome.

An early decision (and one that needs to be reviewed during design) is the set of tools that are relevant to be used. This requires an ongoing review of the current level of understanding against the level needed to bring the risk to the desired level.

Table 12.4 Inputs and outputs in whole process design.

Level	Data inputs	Knowledge inputs	Typical outputs
Extreme	Multiple preparative experiments plus properties and kinetic data for all operations	Prior knowledge of chemistry and plant performance Operation-specific knowledge for all operations (chemical engineering) Whole process simulation software Knowledge capture tools[a]	Full computer model of whole process Model of key processing steps to allow optimization/control Qualitative representation of the process understanding[a]
High level	Multiple preparative experiments plus properties and kinetic data for specific operations	Prior knowledge of chemistry and plant performance Operation-specific knowledge (chemical engineering) Unit simulation tools Knowledge capture tools[a]	Spreadsheet mass balance Recipe Response-surface representation of process outcome Model of key processing steps to allow optimization/control Qualitative representation of the process understanding[a]
Typical level	Multiple preparative experiments Hazard assessment experimentation	Prior knowledge of chemistry and plant performance Knowledge capture tools[a]	Spreadsheet mass balance Recipe Response-surface representation of process outcome Qualitative representation of the process understanding[a]
Base level	Single preparative experiment Hazard assessment experimentation	Prior knowledge of chemistry and plant performance	Spreadsheet mass balance Recipe Summary of chemistry (text report)

[a] Recently, qualitative/semiquantitative knowledge capture tools have been developed to capture and represent process understanding – the BRITEST toolkit – and are being used increasingly [12].

This decision cycle is illustrated schematically in Figure 12.8. Such reviews might be aligned with the stages of process design described above but might equally be carried out between stages.

12.6.2
Decisions in Process Development

Many of the well-known decisions in process development and design have a whole process character; in other words, the decision impacts on the process in multiple stages or ways. Examples of such decisions include the following:

- chemical route selection;
- solvent selection;

Figure 12.8 Risk management decision cycle.

- use of recycles (or not);
- the batch/continuous decision;
- trade-offs between driving reactions to completion and the cost/efficiency of product purification and isolation;
- intellectual property;
- requirements to meet regulatory standards (quality, environmental, and safety).

Table 12.5 Decision support approaches for whole process design.

Tool	Useful for	Weaknesses
Optimization methods (rigorous optimization to maximize/minimize a numerical objective such as profit), for example, gPROMS	Problems where the process model is mathematical and where a precise answer is needed	Data hungry and required high-level modeling effort
Monte Carlo methods, for example, Px-Sim	Complex problems where there is significant uncertainty	Tends to give rather uncertain answer
Structured decision-making tools such as the Kepner Tregoe tool	Problems with multiple criteria and a mixture of quantitative and qualitative information	For very complex problems can become immensely time consuming

In order to take good decisions, it is important to have suitable whole process representations, sufficient understanding of key operations, and defined performance metrics. However, decisions may still require specialized techniques to make them effectively. Some relevant decision support tools are listed in Table 12.5.

12.7
Summary

Whole process design has emerged recently as a clear and useful concept within the design of low tonnage chemicals processes. Increasingly, with more pressure on process cost and performance, it is vital to consider carefully the process as a whole. A new suite of tools and approaches is emerging to tackle this extremely difficult and complex area, rather than relying on the traditional methods that often resulted in the deployment of highly inefficient manufacturing processes as a result of a series of "local" decisions taken almost in isolation. Progress is being made through careful reassessment of process design methodologies to take better decisions earlier, to make maximum use of information and prior knowledge, and better representations that can reveal the complex interdependencies within processes. Key to improving performance is to consider process understanding carefully as having a value and importance of its own, rather than it being the incidental product of a predefined series of development tasks. By designing carefully the activities to collect and exploit process understanding, better processes can readily be delivered.

References

1. Atherton, J.H., Double, J.M., and Gourlay, B. (2006) *Survey of PI Equipment Requirements in the Fine Chemicals and Pharmaceuticals Sector*, BRITEST Limited, Runcorn.

2. Anastas, P.T. and Warner, J.C. (1998) *Green Chemistry: Theory and Practice*, Oxford University Press, New York.

3. Reay, D., Ramshaw, C., and Harvey, A. (2008) *Process Intensification: Engineering for Efficiency, Sustainability and Flexibility*, Butterworth Heinemann.

4. Ehrfeld, W., Hessel, V., and Löwe, H. (2000) *Microreactors: New Technology for Modern Chemistry*, Wiley-VCH Verlag GmbH, Weinheim.

5. Anderson, N.G. (2001) Practical use of continuous processing in developing and scaling up laboratory processes. *Org. Process Res. Dev.*, **5**, 613–621.

6. Berkovitch, I. (1986) *From Batch to Continuous Pharmaceutical Engineering*, Manufacturing Chemist, pp. 43–45.

7. Douglas, J. (1988) *Conceptual Design of Chemical Processes*, McGraw-Hill Higher Education.

8. Moseley, J.S. and Parker, J.D. (2008) Kepner-Tregoe decision analysis as a tool to aid route selection. Part 1. *Org. Process Res. Dev.*, **12** (6), 1041–1043.

9. IBM Corporation (1974) *HIPO – A Design Aid and Documentation Technique*, Publication Number GC20-1851, IBM Corporation, White Plains, NY.

10. Wall, K., Sharratt, P.N., Sadr-Kazemi, N., and Borland, J.N. (2001) Plant-independent process representation. *Org. Proc. Res. Dev.*, **5** (4), 434–437.

11. Sheikhzeinoddin, T. and Sharratt, P.N. (2008) A product-centred approach to a multi-stage task in pharmaceuticals: isolation. 10th World Filtration Congress, Leipzig.

12. Sharratt, P.N., Wall, K., and Borland, J.N. (2003) Generating innovative process designs using limited data. *J. Chem. Technol. Biotechnol.*, **78**, 156–160.

Index