Space Telescope
Science Institute
Symposium Series

16

Astrophysics of Life

Edited by
Mario Livio, Neill Reid, and William Sparks

CAMBRIDGE

This page intentionally left blank

SPACE TELESCOPE SCIENCE INSTITUTE

SYMPOSIUM SERIES: 16

Series Editor S. Michael Fall, Space Telescope Science Institute

ASTROPHYSICS OF LIFE

This volume is based on a meeting held at the Space Telescope Science Institute, which aimed to lay the astrophysical goundwork for locating habitable places in the Universe. Written by leading scientists in the field, it covers a range of topics relevant to the search for life in the Universe, including: extrasolar planet searches and properties; the history of the solar system; star and planet formation; the habitability of planets; and strategies for searches. This is an indispensable collection of articles for researchers and graduate students.

SPACE
TELESCOPE
SCIENCE
INSTITUTE

Other titles in the Space Telescope Science Institute Series.

1 Stellar Populations
Edited by C. A. Norman, A. Renzini and M. Tosi 1987 0 521 33380 6

2 Quasar Absorption Lines
Edited by C. Blades, C. A. Norman and D. Turnshek 1988 0 521 34561 8

3 The Formation and Evolution of Planetary Systems
Edited by H. A. Weaver and L. Danly 1989 0 521 36633 X

4 Clusters of Galaxies
Edited by W. R. Oegerle, M. J. Fitchet and L. Danly 1990 0 521 38462 1

5 Massive Stars in Starbursts
Edited by C. Leitherer, N. R. Walborn, T. M. Heckman and C. A. Norman
1991 0 521 40465 7

6 Astrophysical Jets
Edited by D. Burgarella, M. Livio and C. P. O'Dea 1993 0 521 44221 4

7 Extragalactic Background Radiation
Edited by D. Calzetti, M. Livio and P. Madau 1995 0 521 49558 X

8 The Analysis of Emission Lines
Edited by R. E. Williams and M. Livio 1995 0 521 48081 7

9 The Collision of Comet Shoemaker-Levy 9 and Jupiter
Edited by K. S. Noll, H. A. Weaver and P. D. Feldman
1996 0 521 56192 2

10 The Extragalactic Distance Scale
Edited by M. Livio, M. Donahue and N. Panagia 1997 0 521 59164 2

11 The Hubble Deep Field
Edited by M. Livio, S. M. Fall and P. Madau 1998 0 521 63097 5

12 Unsolved Problems in Stellar Evolution
Edited by M. Livio 2000 0 521 78091 8

13 Supernovae and Gamma-Ray Bursts
Edited by M. Livio, N. Panagia and K. Sahu 2001 0 521 79141 3

14 A decade of *Hubble Space Telescope* science
Edited by M. Livio, K. Noll and M. Stiavelli 2002 0 521 82459 1

15 The Dark Universe: Matter, Energy, and Gravity
Edited by M. Livio 2003 0 521 82227 0

Astrophysics of life

Proceedings of the
Space Telescope Science Institute Symposium,
held in Baltimore, Maryland
May 6–9, 2002

Edited by
MARIO LIVIO
Space Telescope Science Institute, Baltimore, MD 21218, USA

I. NEILL REID
Space Telescope Science Institute, Baltimore, MD 21218, USA

WILLIAM B. SPARKS
Space Telescope Science Institute, Baltimore, MD 21218, USA

Published for the Space Telescope Science Institute

SPACE
TELESCOPE
SCIENCE
INSTITUTE

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE UNIVERSITY PRESS
Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore, São Paulo

Cambridge University Press
The Edinburgh Building, Cambridge CB2 2RU, UK

Published in the United States of America by Cambridge University Press, New York

www.cambridge.org
Information on this title: www.cambridge.org/9780521824903

First published in print format 2005

ISBN-13 978-0-511-08002-9 eBook (EBL)
ISBN-10 0-511-08002-6 eBook (EBL)

ISBN-13 978-0-521-82490-3 hardback
ISBN-10 0-521-82490-7 hardback

Contents

Participants

Aloisi, Alessandra	The Johns Hopkins University
Avera, Randy	Randolph Publishing/NASA
Babiuc, Maria Cristina	Technical University “Gh. Asachi” Iasi
Bergeron, Pierre	Université de Montréal
Bianchi, Luciana	The Johns Hopkins University
Blair, William	The Johns Hopkins University
Blitz, Balir	Radio Astronomy Laboratory, University of California
Bresolin, Fabio	Institute for Astronomy
Brown, Thomas	Space Telescope Science Institute
Bullock, James	Harvard-Smithsonian Center for Astrophysics
Cacciari, Carla	INAF-Osservatorio Astronomico di Bologna
Chandar, Rupali	Space Telescope Science Institute
Crowther, Paul	University College London
Clampin, Mark	Space Telescope Science Institute
Cody, George	Carnegie Institute of Washington
Cooper, John	NASA/Goddard Space Flight Center
Correia, José Carlos	University of Lisbon
Coude du Foresto, Vincent	Observatoire de Paris
D’Amario, James	Harford Community College
de Grijs, Richard	University of Cambridge
de la Fuente Acosta, Eduardo	Instituto de Astronomia, UNAM
Dellorusso, Neil	Catholic University of America/Goddard Space Flight Center
DeMarchi, Guido	Space Telescope Science Institute
Desch, Steven	Carnegie Institution of Washington
des Marais, Dave	NASA/Ames Research Center
Disanti, John	Catholic University of America/Goddard Space Flight Center
Duerbeck, Hilmar W.	Brussels Free University (VUB)
Dwek, Eli	NASA/Goddard Space Flight Center
Ehrenfreund, Pascale	Leiden Observatory
Felten, James	NASA/Goddard Space Flight Center
Foing, Bernard	SCI-SR, ESA/ESTEC
Ford, Holland	The Johns Hopkins University
Ford, K. E. Saavik	The Johns Hopkins University
Fraaije, Johannes	Leiden Institute of Chemistry
Frisch, Priscilla	University of Chicago
Fruchter, Andy	Space Telescope Science Institute
Gaudi, Scott	Institute for Advanced Study
Gibb, Erika	NASA/Goddard Space Flight Center
Gilliland, Ronald	Space Telescope Science Institute
Gimenez, Alvaro	ESA/ESTEC
Godfrey, J. Terry	Information Systems Laboratories, Inc.
Godon, Patrick	Space Telescope Science Institute
Gonzalez, Guillermo	Iowa State University
Graff, David	Ohio State University
Greyber, Howard	
Grogan, Keith	NASA/Goddard Space Flight Center

Haghighipour, Nader	Carnegie Institute of Washington
Hansen, Brad	University of California, Los Angeles
Hartnett, Kevin	NASA/Goddard Space Flight Center
Hasan, Hashima	NASA Headquarters
Hauser, Mike	Space Telescope Science Institute
Heap, Sara	NASA/Goddard Space Flight Center
Hillman, John	NASA Headquarters
Illingworth, Garth	University of California, Santa Cruz
Jeletic, Jim	NASA/Goddard Space Flight Center
Justh, Hilary	Pennsylvania State University
Kasting, James	Pennsylvania State University
Kharecha, Pushker	Pennsylvania State University
Kobulnicky, Chip	University of Wisconsin
Koerner, David	University of Pennsylvania
Krelove, Kara	Pennsylvania State University
Kress, Monika	University of Washington
Lecar, Mike	Center for Astrophysics
Leckrone, David	NASA/Goddard Space Flight Center
Lissauer, Jack	NASA/Ames Research Center
Livio, Mario	Space Telescope Science Institute
Lubow, Steve	Space Telescope Science Institute
Lunine, Jonathan	University of Arizona
Macchetto, Duccio	Space Telescope Science Institute
Magee-Sauer, Karen	Rowan University
Marcy, Geoffrey	University of California, Berkeley
Margon, Bruce	Space Telescope Science Institute
Mathews, Grant	University of Notre Dame
Matrajt, Graciela	University of Washington
Mazzuca, Lisa	NASA/Goddard Space Flight Center
McGrath, Melissa	Space Telescope Science Institute
McGuire, Patrick	Center for Astrobiology, Madrid
McKay, Christopher	NASA/Ames Research Center
McLean, Brian	Space Telescope Science Institute
McLeod, Kim	Wellesley College
Meadows, Victoria	Jet Propulsion Laboratory
Meech, Karen	Institute of Astronomy
Melnick, Gary	Harvard-Smithsonian Center for Astrophysics
Mercader-Perez, Juan	CSIC/INTA
Moore, Marla	NASA/Goddard Space Flight Center
Morrison, David	NASA Astrobiology Institute
Mumma, Michael	NASA/Goddard Space Flight Center
Nealson, Kenneth	University of Southern California
Neufeld, David	The Johns Hopkins University
Niedner, Mal	NASA/Goddard Space Flight Center
Noll, Keith	Space Telescope Science Institute
Nota, Antonella	Space Telescope Science Institute
Noyes, Robert	Harvard Smithsonian Center for Astrophysics
Pearlstein, Robert	IUPUI
Peeters, Zan	Leiden University
Pilcher, Carl	NASA Headquarters

Reid, Neill	Space Telescope Science Institute
Reitsema, Harold	Ball Aerospace Technologies Corporation
Rhie, Sun Hong	University of Notre Dame
Rubenstein, Eric	Smith College
Sahu, Kailash	Space Telescope Science Institute
Sandford, Scott	NASA/Ames Research Center
Scalo, John	University of Texas
Schreier, Ethan	Space Telescope Science Institute
Seager, Sara	Institute for Advanced Study
Seitter, Waltraut C.	Muenster University
Sozzetti, Alessandro	Smithsonian Astrophysical Observatory
Sparks, Bill	Space Telescope Science Institute
Suchkov, Anatoly	Space Telescope Science Institute
Swade, Daryl	Space Telescope Science Institute
Tanner, Angelle	University of California, Los Angeles
Tarter, Jill	SETI Institute
Tej, Anandmayee	Observatoire Astronomique de Strasbourg
Traub, Wesley	Harvard-Smithsonian Center for Astrophysics
Trauger, John	Jet Propulsion Laboratory
Turnbull, Margaret	University of Arizona
Turner, Edwin	Princeton University Observatory
Valenti, Jeff	Space Telescope Science Institute
Vieira, Gladys	NASA/Goddard Space Flight Center
Vishniac, Ethan	The Johns Hopkins University
Weinberger, Alycia	Carnegie Institution of Washington
Wheeler, J. Craig	University of Texas at Austin
Wilkinson, David	Princeton University
Wootten, Alwyn	National Radio Astronomy Observatory
Zahnle, Kevin	NASA/Ames Research Center
Zubko, Viktor	NASA/Goddard Space Flight Center

Preface

The Space Telescope Science Institute Symposium on "Astrophysics of Life" took place during 6–9 May 2002. Unlike other astrobiology symposia, the emphasis here was on astronomical observations and astrophysical research. With the discovery of more than a hundred extrasolar planets on one hand, and recent progress in the understanding of the evolution of the universe on the other, the "astro" part of astrobiology has advanced to the forefront of astronomical investigation.

These proceedings represent only a part of the invited talks that were presented at the symposium. We thank the contributing authors for preparing their manuscripts.

We thank Sharon Toolan of ST ScI for her help in preparing this volume for publication.

Mario Livio
I. Neill Reid
William B. Sparks
Space Telescope Science Institute
Baltimore, Maryland

A voyage from dark clouds to the early Earth

By P. EHRENFREUND,[1,2] S. B. CHARNLEY,[3] and O. BOTTA[2]

[1]Leiden Observatory, P.O. Box 9513, 2300 RA Leiden, The Netherlands

[2]Soft Matter/Astrobiology Group, Leiden Institute of Chemistry, P.O. Box 9502, 2300 RA Leiden, The Netherlands

[3]Space Science Division, NASA AMES Research Center, MS 245-3, Moffett Field, CA 94305, USA

Stellar nucleosynthesis of heavy elements, followed by their subsequent release into the interstellar medium, enables the formation of stable carbon compounds in both gas and solid phases. Spectroscopic astronomical observations provide evidence that the same chemical pathways are widespread both in the Milky Way and in external galaxies. The physical and chemical conditions—including density, temperature, ultraviolet radiation and energetic particle flux—determine reaction pathways and the complexity of organic molecules in different space environments. Most of the organic carbon in space is in the form of poorly-defined macromolecular networks. Furthermore, it is also unknown how interstellar material evolves during the collapse of molecular clouds to form stars and planets. Meteorites provide important constraints for the formation of our Solar System and the origin of life. Organic carbon, though only a trace element in these extraterrestrial rock fragments, can be investigated in great detail with sensitive laboratory methods. Such studies have revealed that many molecules which are essential in terrestrial biochemistry are present in meteorites. To understand if those compounds necessarily had any implications for the origin of life on Earth is the objective of several current and future space missions. However, to address questions such as how simple organic molecules assembled into complex structures like membranes and cells, requires interdisciplinary collaborations involving various scientific disciplines.

1. Introduction

Life in the Universe is the consequence of the increasing complexity of chemical pathways which led to stable carbon compounds assembling into cells and higher organisms. The starting point of this fascinating evolution is the synthesis of elements that play key roles in life as we know it: hydrogen, carbon, oxygen, nitrogen, sulfur and phosphorus. Whereas hydrogen is primordial, having been formed during the Big Bang together with helium and traces of other species, the more heavy elements such as carbon, oxygen and nitrogen have been formed by nucleosynthesis in stellar interiors. During the course of stellar evolution those heavy elements are redistributed into the interstellar medium (ISM) and are incorporated into solid and gaseous chemical networks. The elemental abundances in space, scaled to the most abundant element, hydrogen, are listed in Table 1.

Understanding interstellar organic chemistry may yield important insights into the chemical conditions prevalent at the birth of the Solar System (see Ehrenfreund & Charnley 2000, Ehrenfreund et al. 2002a for reviews). Observations of the interstellar medium, primarily at microwave frequencies, have led to the identification of 123 molecules (Wootten 2002). Most of these molecules are organic and many of them have also been detected in the outgassing comae of comets, supporting the view that cometary ices contain some pristine interstellar matter (Bockelée-Morvan et al. 2000). The high isotopic fractionation of deuterium found in extracts of carbonaceous meteorites indicates that they represent a highly processed sample of interstellar material (Cronin & Chang 1993). Material from cometary and asteroidal impacts is believed to have been important for providing the

Element	Abundance
H	1
He	7.5×10^{-3}
O	8.3×10^{-4}
C	4.0×10^{-4}
N	1.0×10^{-4}
Ne	0.8×10^{-4}
Si	4.3×10^{-5}
Mg	4.2×10^{-5}
S	1.7×10^{-5}
Fe	4.3×10^{-5}
Na	2.1×10^{-6}
P	3.0×10^{-7}

Table 1. Elemental abundances in space

molecular inventory available at the beginning of prebiotic evolution on the early Earth (e.g. Chyba et al. 1990). Hence, studies of the organic chemistry in the interstellar medium and the Solar System form the interface between astrochemistry and prebiotic evolution. They are an active area of research in Astrobiology (Ehrenfreund & Charnley 2000) and address issues such as the chemistry of molecules which could act as precursors or building blocks of biologically-important molecules, e.g. amino acids for proteins, purines and pyrimidines for RNA and DNA bases (Ehrenfreund et al. 2002a).

2. Organic molecules in dense interstellar clouds

In cold molecular clouds, simple molecules form through a chemistry network dominated by gas phase ion-molecule and neutral-neutral reactions (e.g. Herbst 2000). As external starlight is not penetrating into the cloud interior, cosmic ray ionization of molecular hydrogen is the process which energetically drives this chemistry. Rather large organic molecules can form in these environments, and it appears that the production of long carbon chain species is favored (e.g. Langer et al. 2000). Infrared observations show that ice mantles cover interstellar dust grains and indicate that a rich catalytic chemistry is possible on them (Ehrenfreund & Charnley 2000). The efficient adsorption due to the low temperatures in these molecular clouds, diffusion of atoms and small molecules, subsequent reactions on the grain surface and desorption are the prevailing processes responsible for the formation of new molecules and the enrichment of the gas phase inventory, see Figure 1.

The infrared satellite *Infrared Space Observatory* (*ISO*) observed a large number of high-mass star-forming regions and allowed to establish an inventory of the most abundant ice species in dense clouds (Ehrenfreund & Schutte 2000, Gibb et al. 2000). A few low-mass star-forming regions, representative of Sun-type stars have also been observed (e.g. Boogert et al. 2000). Apart from water ice, CO, CO_2 and CH_3OH are the most abundant ice species which were observed with the *ISO*. Ice species which show a tentative signature and/or low abundances are CH_4, NH_3, H_2CO, OCS and OCN^-, see Table 2. On cold grain surfaces, reactions occur in two steps: after accretion atoms and molecules diffuse (by tunneling or hopping) and react following essentially Langmuir-Hinchelwood kinetics. For molecular hydrogen formation on bare silicates or carbon grains, which is relevant for the diffuse ISM, experiments indicate that H diffusion appears to be slower

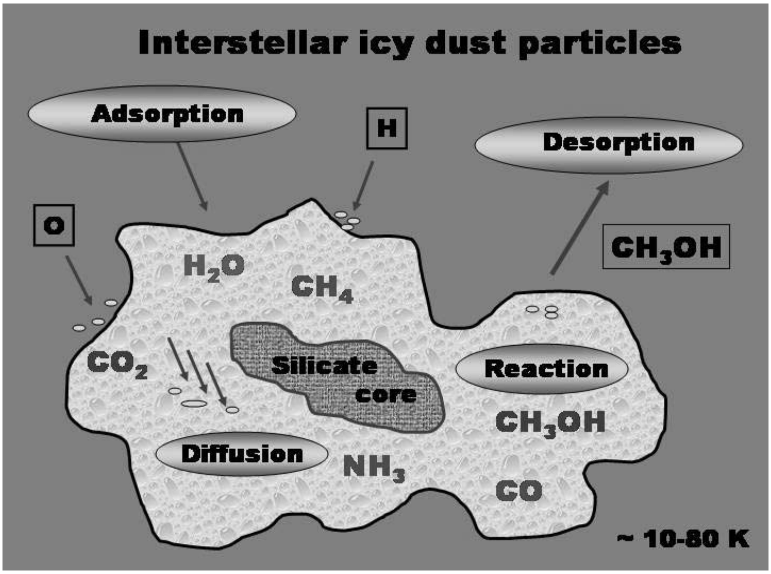

FIGURE 1. The catalytic surface of an interstellar icy grain. The efficient accretion of all atoms and molecules in dense clouds (H_2 excluded) leads to an active surface chemistry forming molecules which can not be produced via gas-phase reactions.

	W33A	Elias 29	Elias 16
Ice species	high-mass protostar	low-mass protostar	field star
H_2O	100	100	100
CO	9	5.6	25
CO_2	14	22	15
CH_4	2	<1.6	–
CH_3OH	22	<4	<3.4
H_2CO	1.7–7	–	–
OCS	0.3	<0.08	–
NH_3	3–15	<9.2	<6
C_2H_6	<0.4	–	–
HCOOH	0.4–2	–	–
O_2	<20	–	–
OCN^-	3	<0.24	<0.4

TABLE 2. Abundances of interstellar ices

(Pirronello et al. 1997, 1999) than expected by quantum diffusion. This may raise difficulties for producing H_2 in diffuse media. However, in dense molecular clouds at temperatures around 10 K, after formation of at least one molecular ice monolayer (Manico et al. 2001), quantum diffusion by H atoms is very rapid, and an H atom will almost always meet a coreactant before another particle arrives from the gas, i.e. to scan the entire surface. Reduction of O, N and C atoms by H addition reactions with zero activation energy form water, ammonia, and methane (Ehrenfreund & Charnley 2000). Direct accretion of CO forms molecular ices which can in principle be reduced and deuterated by quantum tunneling of H and D atoms through the activation energy barriers (Tielens & Hagen 1982; Charnley et al. 1997). This hydrogenation of CO is understood to be the source of the large abundances of solid methanol seen in many lines of sight, as well as for the extremely large deuterium fractionation observed in both formaldehyde and methanol (Loinard et al. 2001; Parise et al. 2002). Experimentally, hydrogenation of CO to form formyl radical, formaldehyde and methanol can occur at low temperatures (van Ijzendoorn 1985; Hiraoka et al. 1994, 1998), but the high efficiencies required to fit the observations have yet to be demonstrated unequivocably (cf. Watanabe & Kouchi 2002). Carbon dioxide can also be formed from CO in an oxygen addition reaction that appears to possess a small activation barrier (Grim & d'Hendecourt 1986; Roser et al. 2001; Charnley 2001a).

A surface chemistry based on these kinetics, with the additional constraint of radical stability (cf. Allen & Robinson 1977), leads to surface schemes like that of Figure 2. This shows how methanol as well as other simple alcohols and aldehydes could be formed by a sequence of reactions starting from CO. The surface reaction network of Figure 2, and extensions of it (e.g. Charnley 2001b), have several promising characteristics. From the perspective of the organic composition in the Murchison meteorite, the scheme depicted in Figure 2 is consistent with the decline in concentration observed within homologous series, such as amino acids, with increasing number of carbon atoms, as some of these compounds (e.g. aldehydes, ketones) are believed to have been precursors for the formation of these species during aqueous alteration (Cronin & Chang 1993).

Hot molecular cores are dense, warm regions heated by young protostars (Brown, Charnley & Millar 1988). Observations show that they are particularly rich in molecules believed to have been synthesized on grains, as well as in larger, more complex, organic molecules. These include methanol, ethanol, dimethyl ether, methyl formate, ketene, formaldehyde, acetaldehyde, formic acid and several nitriles (Langer et al. 2000). These regions are a natural laboratory in which the most complex molecules observable in the interstellar medium can be studied and therefore provide an important point of contact for comparative studies of Solar System organics, such as those found in meteorites and comets (Cronin & Chang 1993; Crovisier 1998). Several of the larger organic species probably formed in ice mantles. However, many more probably formed after the evaporation of simple molecular mantles, through ion-molecule reactions in the warm gas (Charnley, Tielens & Millar 1992; Charnley et al. 1995). In particular, alkyl cation transfer reactions involving surface-formed alcohols and other products of grains surface chemistry could form an extensive suite of very large interstellar molecules (Figure 3). Recent support for this picture comes from observations which indicate that methyl formate ($HCOOCH_3$) and acetic acid (CH_3COOH) have a common formation mechanism (Remijan et al. 2002), from correlations between formic acid (HCOOH) and CH_3COOH (Liu et al. 2001), as well as from the detection of many molecules predicted to form in grain-surface schemes similar to that of Figure 2 (see Charnley 2001b): vinyl alcohol (CH_2CHOH; Turner & Apponi 2001), glycolaldehyde ($HOCH_2CHO$; Hollis et al. 2000) and ethylene glycol ($HOCH_2CH_2OH$; Hollis et al. 2002).

3. Interstellar amino acids?

The most abundant amino acids found in meteorites are typically Glycine (Gly, NH_2CH_2COOH), β-Alanine (β-Ala, $NH_2CH_2CH_2COOH$), α-Alanine (α-Ala, $NH_2(CH_3)CHCOOH$), α-aminoisobutyric acid (AIB, $NH_2C(CH_3)_2COOH$), and 2-amino-*n*-butyric acid (α-ABA, $NH_2(C_2H_5)CHCOOH$); their presumed interstellar precursors for Strecker-cyanohydrin synthesis are formaldehyde, acetaldehyde, acetone and propionaldehyde. Figures 2 and 3 show that interstellar chemistry can clearly provide all the required amino acid precursors to start parent body chemistry. The question naturally

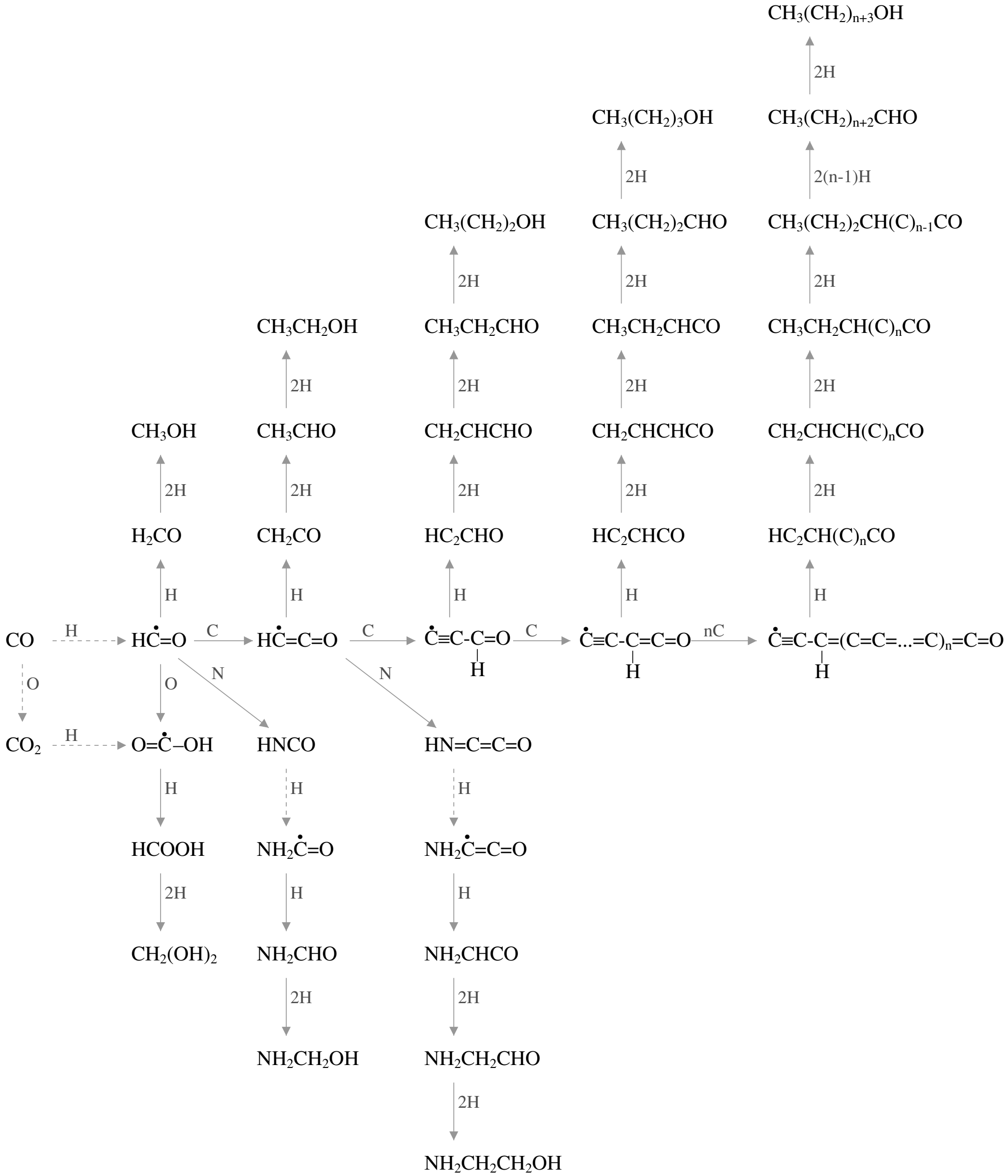

FIGURE 2. Interstellar grain surface chemistry. Hydrogen atom addition to unsaturated molecules creates reactive radicals and a rich organic chemistry seeded by carbon monoxide ensues. Broken arrows indicate reactions with activation energy barriers; where 2H is shown, a barrier penetration reaction followed by an exothermic addition is implicitly indicated (adapted from Charnley 1997).

arises as to whether amino acids could actually form in the interstellar medium and then become directly incorporated in meteorites.

3.1. *Grain-surface synthesis of amino acids?*

Interstellar ice mantles are also subject to energetic processing in dense clouds by a weak ambient flux of UV photons (Prasad & Tarafdar 1983). This processing potentially opens up more pathways for surface organic chemistry. Recent experimental and theo-

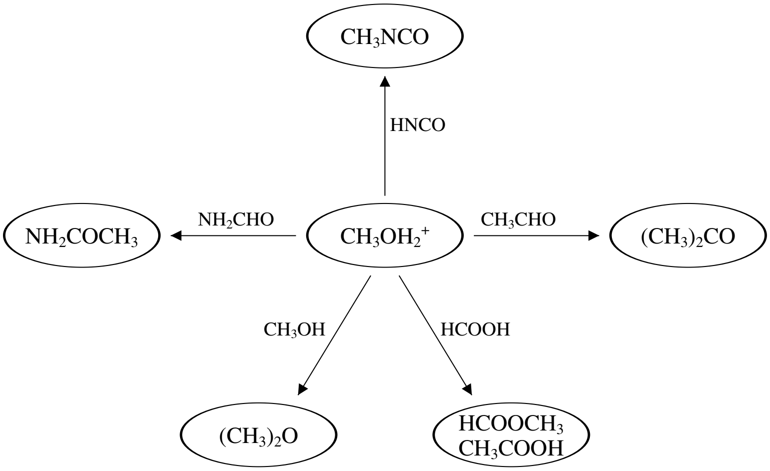

FIGURE 3. Methyl cation transfer in hot cores (Charnley 2001b). For clarity, formation of the molecular ion and the subsequent electron recombination step have both been omitted.

retical work suggests that UV photolysis of interstellar ice analogs produces amino acids (Bernstein et al. 2002; Munoz-Caro et al. 2002; see also Sorrell 2001 and Woon 2002). However, the applicability of these results to the interstellar medium is compromised by the need for a large UV flux in molecular clouds, and the fact that amino acids are photochemically very unstable (Ehrenfreund et al. 2001a).

We find that it is not possible to form amino acids in highly-restrictive surface reaction schemes like that of Figure 2. Although many of these pathways have unknown, and probably large, energy barriers, isomerization could be relevant for the production of some organic precursors. For example, in Figure 2 N atom addition to the ketyl radical (HC=C=O) has been depicted to lead to HN=C=C=O, and further C additions to longer cumulenone radicals will generally generate HNC_nO compounds. Saturation of these compounds leads, in analogy to the alcohols (see Figure 2), to aminoalcohols such as aminomethanol (NH_2CH_2OH), which was identified as a component of Comet Halley's tail (Kissel & Krueger 1987), and aminoethanol ($NH_2CH_2CH_2OH$). An isomer of HNCCO is formyl cyanide (HCOC≡N). It has been demonstrated experimentally that cold H atoms cannot add efficiently to the nitrile bond (Hiraoka et al. 1998) and, in this case, the end product will be cyanomethanol. Hence, this route could lead to interstellar cyanohydrins and could perhaps allow part of the Strecker-cyanohydrin route to hydroxy acids to be bypassed. However, it does not seem possible to efficiently produce the corresponding Strecker intermediates for amino acid synthesis in this way.

3.2. *Gas phase synthesis of amino acids?*

As there are no routes to amino acids in surface schemes based on Figure 2, we consider gas phase synthesis. Evaporation of alcohols, aminoalcohols and formic acid in hot molecular cores may produce amino acids through exothermic alkyl and aminoalkyl cation transfer reactions analogous to those of Figure 3 (Charnley 1997; 2001b). Gas phase synthesis of amino acids provides an important counterpoint to the recent work on solid-state production of interstellar amino acids (Bernstein et al. 2002; Munoz-Caro et al. 2002). Hot cores were identified as one of the few regions where the UV flux was sufficiently low (about 300 mag of extinction) that amino acids could survive in the gas (Ehrenfreund et al. 2001a).

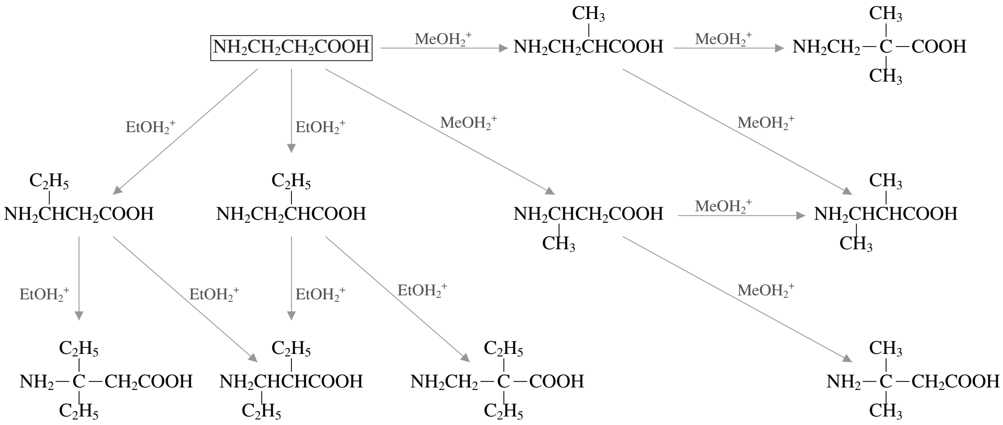

FIGURE 4. Possible amino acids formed from β-Alanine (Charnley 2001b).

A first possible step is the formation of protonated Gly and β-Ala via aminoalkyl cation transfer from aminomethanol and aminoethanol

$$\mathrm{NH_2CH_2OH_2^+} \;+\; \mathrm{HCOOH} \;\longrightarrow\; \mathrm{NH_2CH_2COOH_2^+} \;+\; \mathrm{H_2O} \qquad (3.1)$$

$$\mathrm{NH_2(CH_2)_2OH_2^+} \;+\; \mathrm{HCOOH} \;\longrightarrow\; \mathrm{NH_2(CH_2)_2COOH_2^+} \;+\; \mathrm{H_2O} \qquad (3.2)$$

followed by electron recombination producing the neutral acid. Further alkylation can theoretically produce a large suite of amino acids (see Figure 4) and could in principle explain the high degree of branching seen in the Murchison amino acids. Generally, the proposed reactions involve the hydrogen atoms of amino acids being replaced by alkyl groups, through elimination of a water molecule, as occurs between the alcohols themselves (Charnley et al. 1995).

In theory one can calculate the distribution of amino acids expected in such a scheme. Precursor abundances can be determined based on observations or from model calculations. Clearly, based on Figure 4, there are interrelationships between specific amino acids. However, one can 'solve' the networks (e.g. Figure 4) by inspection. There is no experimental information currently available, therefore one must simply assume similar rates for all amino acid formation and destruction processes (e.g. Mautner & Karpas 1986). Furthermore, one generally expects that surface-formed molecules will be less abundant the more heavy atoms they contain. Observations do show that MeOH/EtOH > 1, and we also expect that $\mathrm{NH_2CH_2OH}$/ $\mathrm{NH_2CH_2CH_2OH} > 1$. In this case, the scheme of Figure 4 predicts the same distribution of amino acids as Strecker-cyanohydrin synthesis, i.e. Gly $>$ α-Ala $>$ AIB $>$ α-ABA. Hence, although this chemistry may produce interstellar amino acids, as it stands it cannot account for either the similarity of the Gly and AIB abundances in Murchison, or the prevalence of β-Ala in Orgueil.

Although most of the reactions depicted in Figure 4 are exothermic, the existence of activation energy barriers is unknown. In fact, the lack of good experimental or theoretical 'ground truths' for any of the proposed reactions, means that any hot core chemistry modeling of amino acid chemistry would be largely unconstrained. For example, variation of reaction rates could reproduce a different amino acid distribution than that inferred above. It has been assumed that electron dissociative recombination of these large organics leads only to the loss of a hydrogen atom (e.g. Herbst 1978); different branching ratios would also have a significant effect. Specifically, we need to know the molecular structure of the product ions, whether or not energy barriers exist, the product branching ratios, and also the rate coefficients. These quantities could be measured experimentally or estimated theoretically to allow us to develop reliable the amino acid reaction networks.

A prime motivation for constructing this scheme was that proton transfer to HCOOH from protonated alcohols and aminoalcohols is endothermic, whereas alkyl or aminoalkyl cation transfer is exothermic (Lias et al. 1984; Hunter & Lias 2001), hence the latter process should be favored in hot cores. This fact also makes these ion-molecule reactions amenable to study in the laboratory. Preliminary studies using a flow reactor/tandem mass spectrometer have confirmed the general viability of the proposed aminoalkyl transfer reactions (using aminoethanol, D. K. Bohme, private communication). These results indicate that, contrary to Figure 4, significant molecular rearrangements can occur, and that some product channels possess large energy barriers. This has important consequences as it means that the amino acid distribution can be much different from that naively expected by inspection (see above).

Acetic acid is present in hot cores at abundances that make compounds formed from it relevant for meteoritic composition. The observation of $HCOOH/CH_3COOH\sim 1$ (Remijan et al. 2002) could have important consequences for the proposed theory (see Figure 3). It suggests that the kinetics can be such that a ‘parent’ (HCOOH) can form a ‘daughter’ (CH_3COOH) which persists longer due to differences in the destruction rates of the molecules. It has already been noted that there is a significant discrepancy between high abundances of HCOOH in interstellar ices (Schutte et al. 1996) and gaseous HCOOH in hot cores (Ehrenfreund et al. 2001b). This process, as well as molecular rearrangements in ion-molecule reactions involving CH_3COOH, could also play a role in amino acid synthesis. Hopefully, these issues will be quantified experimentally in the near future to allow an accurate calculation of the possible amino acid distribution in hot cores.

4. Complex organic molecules in diffuse interstellar clouds

Diffuse clouds have moderate extinctions (<1 mag), densities of roughly 100–300 cm^{-3} and are dominated by photochemistry. Lines of sight through diffuse clouds allow measurements of the extinction curves (Jenniskens & Greenberg 1993). The extinction curve measured towards many targets seems rather constant in the long wavelength part ($\lambda > 2500$ Å). However, substantial changes characterize the short wavelength behavior, including the 2200 Å bump. The differences in the extinction curve reflect changes in the composition and size distribution of local dust particles (Draine 1990). Hydrogenated amorphous carbon (HAC) is considered as a potential carrier for the 2200 Å bump observed in the interstellar extinction curve (Mennella et al. 1998), while a variety of complex aromatic networks are likely to be present on carbonaceous grains (see Papoular et al. 1996 and Henning & Salama 1998 for reviews).

Organic molecules present in the diffuse medium can be formed by gas phase reactions, either directly or indirectly through formation in circumstellar envelopes followed by subsequent mixing into the diffuse medium, or by photoreactions of carbonaceous particles and sputtering by grain-grain collisions. Since the initial discovery of simple diatomic molecules in interstellar space, many more gas phase molecules have been detected in the diffuse medium, such as HCO^+, CO, OH, C_2, HCN, CN, CS, H_2C, C_3H_2 (Lucas & Liszt 1997, 2000) and C_3 (Roueff et al. 2002).

Among the large organic molecules observed or suspected in diffuse clouds are polycyclic aromatic hydrocarbons (PAHs), fullerenes, carbon-chains, diamonds, amorphous carbon (hydrogenated and bare), and complex kerogen-type aromatic networks. The formation and distribution of large molecules in the gas and solid state is far from understood. In the envelopes of carbon-rich late-type stars, carbon is mostly locked in CO and

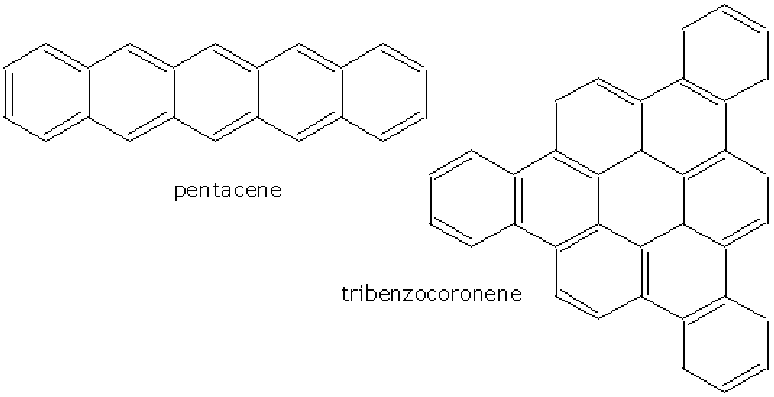

FIGURE 5. Examples of polycyclic aromatic hydrocarbons (PAHs), probably the most abundant organic gas phase molecules in space.

acetylene (C_2H_2). Soot formation involves polymerization of C_2H_2 in these envelopes, with PAHs as important intermediates (Frenklach & Feigelson 1989). PAHs are observed ubiquitously in our galaxy and beyond (Tielens et al. 1999) through their signature in the near and mid-infrared bands. They seem to be stable and abundant components in diffuse clouds and other space environments (see Figure 5).

4.1. *The diffuse interstellar bands*

The diffuse interstellar bands (DIBs) are a large number of absorption lines between 4000–10000 Å that are superimposed on the interstellar extinction curve (Herbig 1995). Since the discovery of the first two DIBs in the 1920s, their identification remains an important problem in astronomy (Herbig 1995, Ehrenfreund et al. 2001c). In the last 75 years DIBs have been observed toward more than a hundred stars. The number of known DIBs present in current observational data is ∼300 and continuously increases due to the higher sensitivity of detectors (Jenniskens et al. 1994, O'Tuairisg et al. 2000). At present, no definitive identification of any of the carriers of the DIBs exists. The detection of substructures in some of the narrow, strong DIBs strongly suggests a gas phase origin and a stable nature of the carrier molecules. Consequently, good candidates are large carbon-bearing molecules which reside ubiquitously in the interstellar gas (Ehrenfreund & Charnley 2000). Polycyclic Aromatic Hydrocarbons (PAHs), fullerenes and carbon chains are among the most promising carrier candidates (see Ehrenfreund & Charnley 2000 for a review). The same unidentified absorption bands are also observed in extragalactic targets (Snow 2002). We have recently observed with the VLT/UVES at unprecedented spectral resolution the absorption spectrum toward reddened stars in the Magellanic Clouds. This range covers the strong transitions associated with neutral and charged large carbon molecules of varying sizes and structures. We reported the first detection of diffuse interstellar bands (DIBs) at 5780 and 5797 Å in the Small Magellanic Cloud (SMC), (see Figure 6) and measured the variation of DIBs in the SMC, the Large Magellanic Cloud (LMC) and our galaxy (Ehrenfreund et al. 2002b). The variation of DIBs in the Magellanic Clouds compared to Galactic targets may be governed by a combination of the different chemical processes prevailing in low metallicity regions and the local environmental conditions. However, the formation pathways of the DIB carriers seems to be ubiquitous throughout the universe.

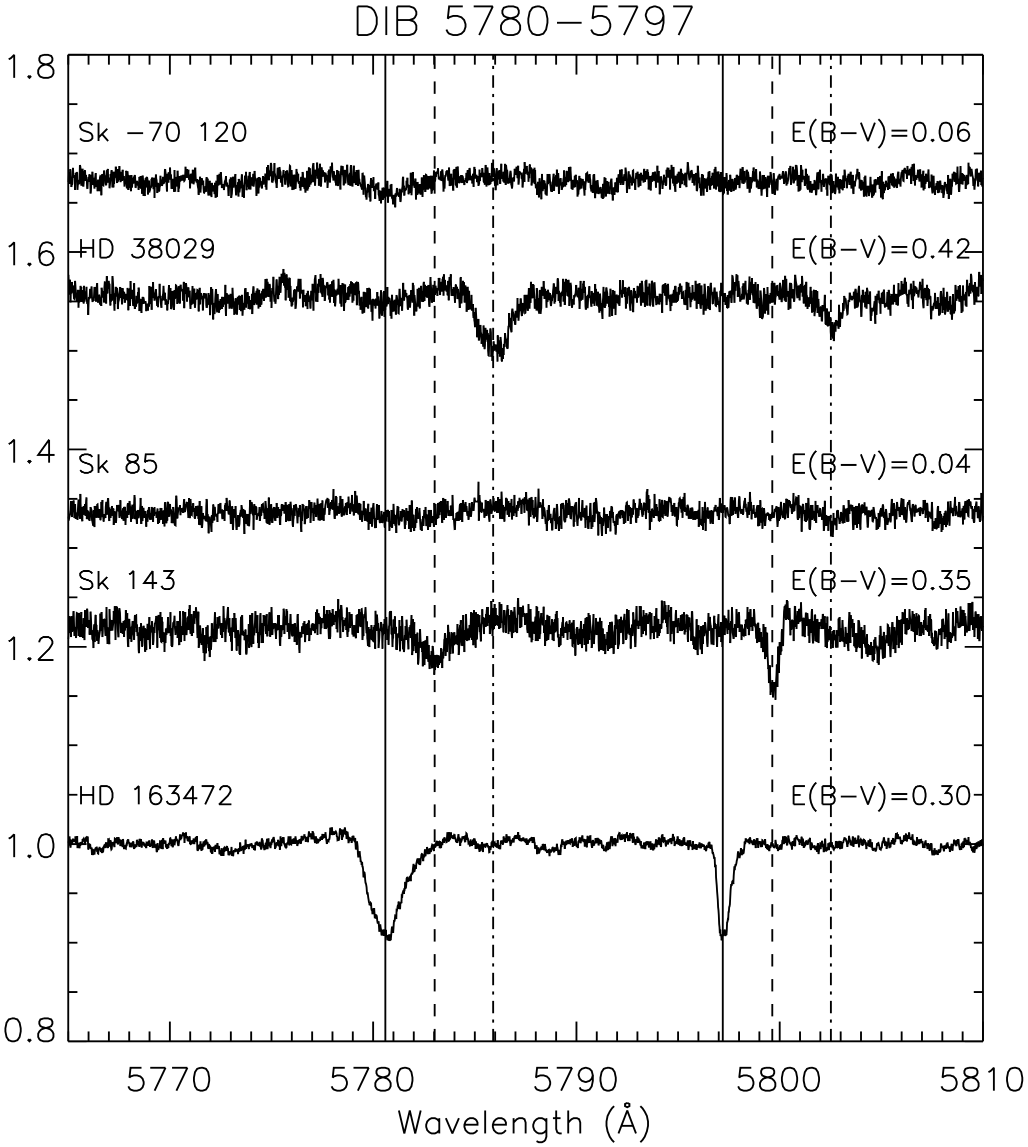

FIGURE 6. This figure shows spectra of the two best known DIBs, the $\lambda\lambda$5780 and 5797 DIB. All spectra are normalized to the continuum, and shifted for display. The spectrum at the bottom shows a Galactic translucent cloud source (HD 163472). The two middle spectra show an SMC target Sk 143 (AzV456) and an unreddened SMC standard Sk 85 (AzV242); the two top spectra show an LMC target (HD 38029) and an unreddened LMC standard Sk −70 120. The full vertical lines indicate the Galactic rest wavelengths for those DIBs, the dashed lines indicate the wavelengths expected for these DIBs at SMC velocities, as determined from the NaD lines and the dash-dotted lines are the same for the LMC velocities. Both DIBs are clearly detected in both the SMC and LMC reddened targets (Ehrenfreund et al. 2002b).

5. The evolution of organic molecules during Solar System formation

The interstellar gas and dust (discussed in Sections 2 and 3) provide the raw material for the formation of stars and planetary systems as outlined in the following scenario. Interstellar clouds undergo a gravitational collapse to form protostars. During the protostellar phase the central protostar is strongly embedded in its parent cloud. It is accreting surrounding material while ejecting a powerful bipolar outflow which regulates the angular momentum resulting from the original contraction of the rotating parent cloud. After about one million years, low-mass protostars (solar-type stars) enter the T-Tauri phase and are surrounded by a disk of dust and gas. Finally these objects are reaching the Main Sequence of the temperature-luminosity diagram and start nuclear burning. In the meantime, the surrounding protoplanetary disk has dissipated and planets may have formed in orbit around the star.

The infalling interstellar dust particles and molecules are certainly affected by different processes dominating at various radial distances from the forming star. It is assumed that the outer solar nebula was an environment of low temperature and pressure. Dutrey et al. (2000) performed a deep survey of the protoplanetary disk surrounding DM Tau, a T-Tauri star of 0.5 solar mass. Because of sensitivity limitations, the observed molecules only trace the outer disk chemistry (radius >50 AU). Simple organic molecules like H_2CO or HCN have been detected, suggesting that even more complex molecules may exist. Some regions in the disk may be very cold (T < 20 K) and dense ($n_N = 10^6$–10^9 cm^{-3}) and will trigger a strong condensation of molecules onto dust particles. This could explain the depletions of certain species compared to interstellar abundances (Aikawa et al. 2002). Thi et al. (2002) reported abundant solid CO abundances in the outer circumstellar disk around an edge-on class I object. In the inner disk heating and thermochemical reactions are important. In the regions <10 AU the chemistry is dominated by the thermal desorption of species from dusty grains. Models suggest that not all the gaseous matter is frozen onto grain surfaces at 10 AU (Markwick et al. 2002). This indicates that certain abundant organic species may be detected in the inner disk when higher sensitivity and angular resolution will become available with new radio astronomical facilities.

Isotopic analyses (e.g. of C, N, ^{26}Al, Sr, Zr, Mo, etc.) of presolar grains allow the reconstruction of the nucleosynthetic processes which occurred in the environment in which such particles were formed. Those analyses show that certain grains are formed in stellar atmospheres and thus represent samples of ancient stardust (Zinner & Amari 1999). However, most of the material has been altered by chemical and physical processes in the solar nebula before becoming incorporated into small bodies.

In comets and outer Solar System asteroids, organic molecules formed in the pre-solar interstellar nebula may have survived solar system formation in relatively pristine form. Therefore, these small bodies carry important evidence on the formation of our Solar System.

6. Comets

Cometary nuclei formed in the outer Solar System environments and are porous aggregates of ice and refractory material. In 1986 several spacecraft performed a close fly-by of Comet Halley in order to perform *in situ* measurements that revealed the structure and composition of this comet. Those *in situ* measurements performed with the mass spectrometers PUMA-1, PUMA-2, and PIA, flown on the *VEGA-1*, *VEGA-2*, and *GIOTTO*, respectively, showed that about 70% of the dust grains comprised a mixed phase of organic (CHON) material and refractory silicates. Current and future space missions such as *STARDUST* and *ROSETTA* will help us to reveal the composition of cometary refractories and ices in the nucleus.

Since then observations of volatiles in the cometary coma over large parts of the electromagnetic spectrum are a crucial tool to obtain indirect information on cometary nuclei. The composition of comets has been recently reviewed (e.g. Irvine et al. 2000). The substantial number of bright comets in recent years allowed high resolution measurements and provided stronger evidence for chemical differentiation amongst the comet population.

Cometary ices are dominated (more than 50%) by water ice, but by now more than 25 other small molecules have been identified, among them CO_2, H_2CO, NH_3, HCN, CH_3CN, C_2H_2, HNCO and H_2S. Their abundances relative to water range from 20–30% for CO (although this may include a significant non-nuclear contribution from distributed sources) down to 0.01%, which is at present the lowest abundance detected. More complex species observed are CH_3OH, CH_3CHO, $HCOOCH_3$, and NH_2CHO. An upper limit for glycine, the simplest amino acid, of $< 0.5\%$ has been obtained.

Revealing the composition of comets provides clues to the formation of our Solar System. By comparing the distribution and abundances of species observed on interstellar icy grains with cometary observations the amount of processing of those objects have undergone during Solar System formation can be estimated.

7. Organic molecules in meteorites

It was established over a century ago that some meteorites contain carbonaceous material. Organic compounds that have been identified in carbonaceous C1 and C2 chondrites include amines and amides; alcohols, aldehydes, and ketones; aliphatic and aromatic hydrocarbons; sulfonic and phosphonic acids; amino, hydroxycarboxylic, and carboxylic acids; purines and pyrimidines; and kerogen-type material (Cronin & Chang 1993; Botta & Bada 2002). For a given number of carbon atoms there is complete structural diversity within most classes of organic compounds, indicating random formation mechanisms involving free single C-bearing radicals. For the amino acids, all stable isomeric forms are present, and branched-chain isomers are more abundant than linear ones. The organic inventories of Murchison and other primitive meteorites, display large and variable enrichments in deuterium, ^{13}C and ^{15}N (e.g. Cronin & Chang 1993), which is indicative of their retention of an interstellar heritage. Carbonaceous meteoritic material thus represents a mixture of highly processed material of interstellar matter as well as pristine pieces. An important goal for theoretical astrochemistry is to elucidate which organics are of true interstellar origin, and to identify possible interstellar precursors and reaction pathways for those molecules which are the result of aqueous alteration.

7.1. *Amino acids in carbonaceous chondrites*

The total amino acid abundances measured in several CM chondrites are highly variable ranging from ~15,300 parts-per-billion (ppb) for Murchison to ~3,200 ppb for Essebi (Botta et al. 2002). Individual amino acids are detected in concentrations up to 2600 ppb (AIB in Murchison, Ehrenfreund et al. 2001d), and the total amino acids abundance can reach more than 10 parts-per-million (ppm), which is ~0.1% of the total soluble organic carbon in a carbonaceous chondrite (Botta & Bada 2002). Generally all of the CM meteorites show a complex distribution of amino acids. There are three main indications for the presence of indigenous extraterrestrial amino acids: (1) a high abundance of α-aminoisobutyric acid (AIB) and isovaline, which are both very rare amino acids on the Earth and therefore are not likely to be terrestrial contaminants; (2) the presence of racemic (D/L ~1) amino acids, such as alanine, α-ABA, β-ABA, β-AIB and isovaline; and (3) enrichments in deuterium, ^{13}C and ^{15}N in amino acid fractions, which support the view that they were derived from interstellar precursors through subsequent chemical processing on the parent body (Pizzarello et al. 1991). Relative amino acid abundances

Compound class	Concentration (ppm)
Amino acids/CM meteorites	17–60
Amino acids/CI meteorites	∼5[a]
Aliphatic hydrocarbons	>35
Aromatic hydrocarbons	3319[b]
Fullerenes	>100[c]
Carboxylic acids	>300
Hydrocarboxylic acids	15
Dicarboxylic acids and Hydroxydicarboxylic acids	14
Purines and Pyrimidines	1.3
Basic N-heterocycles	7
Amines	8
Amides linear	>70
Amides cyclic	>2[d]
Alcohols	11
Aldehydes and Ketones	27
Sulphonic acids	68
Phosphonic acids	2

[a] average of the abundances in the CI carbonaceous chondrites Orgueil and Ivuna (Ehrenfreund et al. 2001d)

[b] for the Yamato-791198 carbonaceous chondrite (Naraoka et al. 1988)

[c] 0.1 ppm estimated for C60 in Allende (Becker et al. 1994)

[d] Cooper & Cronin 1995

Table 3. Abundances of soluble organic compounds in the Murchison meteorite (taken from Botta & Bada 2002) and PAHs and fullerenes (taken from Becker et al. 2000). Amino acid concentrations for several CI chondrites are also listed.

are a recently discovered tool to help to distinguish between different types of carbonaceous chondrites and their parent bodies (Figure 7; Botta et al. 2002).

Although CI and CM meteorites are considered to be the most pristine samples from the early Solar System, both meteorite classes show mineralogical evidence for alteration by liquid water percolating through the parent body in the form of hydrated minerals like clays, carbonates, sulfates and phyllosilicates (McSween, 1979, Endress & Bischoff 1996). In the currently favoured interstellar/parent body formation pathway, precursor material synthesized in the interstellar medium by either gas-phase reactions or processed in ice mantles would be trapped in the meteorite parent body and undergone subsequent reactions in the aqueous solution to form secondary products. Energy sources for these processes include radioactive decay of short-lived isotopes like ^{26}Al and ^{60}Fe, but also low-energy impacts and perhaps other, unknown sources of energy. While early quantitative model calculations point to temperature estimates for these aqueous solutions of less than 20 C for CM and 100–150 C for CI carbonaceous chondrites (Clayton & Mayeda 1984), subsequent refinements of the models have reduced the maximum temperature reached for the CIs to ∼50 C (Leshin et al. 1997).

The reaction of HCN, NH_3 and carbonyl compounds to form α-amino acids, known as the Strecker-cyanohydrin synthesis, leads to an equilibrium between cyanohydrins and aminonitriles in aqueous solution. These intermediates will then undergo irreversible hydrolysis yielding α-hydroxy- and α-amino acids, respectively (Peltzer et al. 1984). The Strecker synthesis is considered to be the most probable pathway for the synthesis of α-amino acids in CM meteorites.

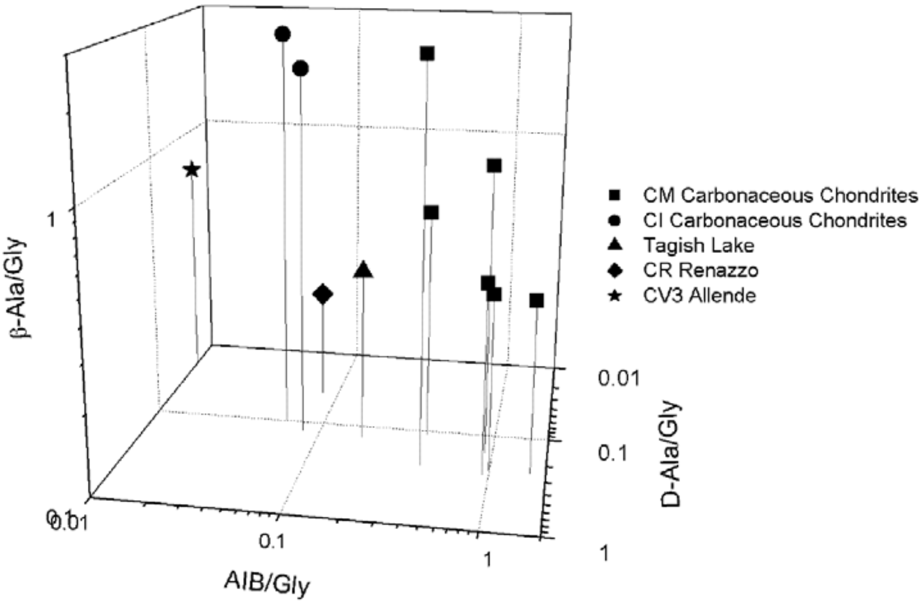

FIGURE 7. 3-dimensional logarithmic diagram of the amino acid abundance ratios AIB/Gly, D-Ala/Gly and β-Ala/Gly in the CM (Murchison, Murray, Mighei, Nogoya, Essebi, LEW90500), CI (Orgueil, Ivuna), CR (Renazzo), and CV3 (Allende) type carbonaceous chondrites as well as the Tagish Lake meteorite. Note that the AIB/Gly ratios for Renazzo, Allende and Tagish Lake are upper limits due to the non-detection of AIB in these samples. Also, the D-Ala/Gly ratios in Renazzo and Allende are upper limits.

It is possible to construct expected product ratios from recent interstellar observations (Remijan et al. 2000; Snyder et al. 2002; Liu et al. 2001). For $H_2CO/(CH_3)_2CO \sim 5 - 20$ and $CH_3CHO/(CH_3)_2CO \sim 0.6 - 2$, Strecker-cyanohydrin synthesis from these molecules should exhibit the amino acid trend Gly > α-Ala > AIB > α-ABA. However, the distribution in Murchison is Gly $\sim$ AIB > α-Ala $\gg$ α-ABA. Proposed explanations of this discrepancy involve selective destruction of precursors (CH_3CHO) or products (α-Ala), enhanced $(CH_3)_2CO$ formation or differences in relative formation efficiencies (Botta et al. 2002). The viability of any of these scenarios is completely unknown at present. In contrast to the CMs the abundances of complex amino acids are by one order of magnitude lower in the CIs, such as Orgueil and Ivuna. The amino acid distribution in these meteorites is completely different: β-Ala > Gly > α-Ala $\gg$ more complex amino acids (Ehrenfreund et al. 2001d). This difference between the CI and CM amino acid distribution has been interpreted as there being different parent bodies for CM and CI chondrites, with the latter possibly originating from extinct comets (Ehrenfreund et al. 2001d). On the CI parent bodies, the amino acid distribution indicates that the Strecker synthetic pathway was not active.

Understanding the origin of the extensive range of amino acids found in Murchison, accounting for their distribution, explaining the compositional relations with other CM chondrites, such as Murray, and the differences with CI chondrites, such as Orgueil and Ivuna, are central problems in Astrobiology (Botta & Bada 2002). Although isotopic differences observed between α-hydroxy and α-amino acids in Murchison may simply be due to differences in kinetics of the individual Strecker-cyanohydrin reactions, relating the relative abundances of the most abundant amino acids to those of their putative interstellar precursors is still a non-resolved problem (Botta et al. 2002).

7.2. *Amino acids as extraterrestrial biomarkers*

Amino acids are the building blocks of proteins and enzymes in life on Earth and are therefore essential for biology as we know it. Because amino acids can be synthesized under a variety of prebiotic conditions (Miller, 1953; Ferris et al. 1978) it is generally assumed that these compounds could have been present on the early Earth, independent of their synthetic origin. Being synthesized under simulated prebiotic conditions in the famous Miller-Urey experiment about 50 years ago, these compounds are also produced from the hydrolysis of tholins, which are residues from chemical reactions in gaseous mixtures that simulate the atmospheric conditions on Saturn's moon Titan and Neptune's moon Triton (McDonald et al. 1994). Shortly after its fall in Australia in 1969, several non-biological amino acids, including AIB and isovaline, which are extremely rare on Earth, were detected in the first organic analysis of the Murchison meteorite (Kvenvolden et al. 1970). Since then, more than 70 amino acids have been identified in carbonaceous chondrites, the most primitive type of meteorites. Only eight of these compounds are identical to those used by terrestrial organisms, while the rest is very rare in the biosphere. The molecular architecture of most amino acids provides a powerful means of discriminating between a biological and non-biological origin of these compounds in meteoritic extracts. In all organisms on Earth, only the L-enantiomers of chiral amino acids are incorporated into proteins and enzymes. In contrast, the abiological synthesis of chiral amino acids always yields a 1:1 mixture of the D- and L-enantiomers (a racemic mixture). In Murchison extracts, alanine was found to be racemic, indicating the presence of indigenous extraterrestrial amino acids in this meteorite. Small enantiomeric excesses have recently been found for α-methylated amino acids in Murchison, indicating that there might be an extraterrestrial mechanism that could lead to a prebiotic presence of "chirality" (Pizzarello & Cronin 2000). However, considering the current understanding of the formation of amino acids found in meteorites (e.g. Strecker synthesis on the meteorite parent body), it is not clear what that mechanism could be. In summary, the detection of amino acids in meteorites provides unequivocal evidence for the extraterrestrial synthesis of biologically relevant compounds.

7.3. *Nucleobases in meteorites*

Like amino acids, purines and pyrimidines play a major role in terrestrial biochemistry. They are central components of DNA and RNA, molecules that are used in the storage, transcription and translation of genetic information in all organisms on the Earth. Unlike amino acids, nucleobases do not exhibit molecular chirality, which makes it difficult to distinguish between abiotic and biotic origins of these compounds. The abundances of uracil (Stoks & Schwartz 1979) as well as adenine, guanine, xanthine and hypoxanthine (Stoks & Schwartz, 1981) were measured in the Murchison, Murray and Orgueil meteorites. None of these nucleobases was detected in the CV3 meteorite Allende. More complex N-heterocyclic compounds in the formic acid extract of the Murchison meteorite were also identified, including 2,4,6-trimethylpyridine, quinoline, isoquinoline, 2-methylquinoline and 4-methylquinoline (Stoks & Schwartz, 1982), as well as derivatives of quinolines and isoquinolines have also been detected in this meteorite (Krishnamurthy et al. 1992). No isotopic measurements have been made for the N-heterocyclic compounds found in meteorites that would provide evidence for their extraterrestrial origin. However, based on the very low contamination levels for amino acids in these meteorites, a low terrestrial contamination for the nucleobases can be inferred. The determination of the carbon and nitrogen isotopic composition of these compounds in meteorites would help to place constraints on the conditions at their place of origin.

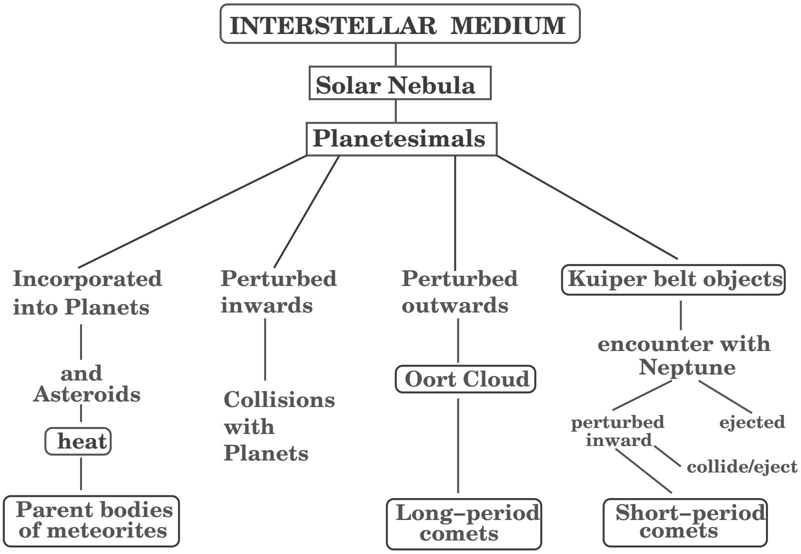

FIGURE 8. The interstellar medium provides the raw material for solar and extrasolar systems. During the formation of a protostar the infalling gas and dust is moderately to heavily processed before incorporation into planetesimals. Planetesimals are either incorporated into forming planets and small bodies, or remain remnant objects which are impacting the newly formed planets or are ejected beyond the Solar System. The analysis of Solar System objects provides therefore important constraints for reconstructing the origin of our Solar System (adapted from Cruikshank 1997).

7.4. *Discussion*

Figure 8 shows the connection between interstellar and Solar System material. Life on Earth is one of the outcomes of the origin and evolution of the Solar System. The sequence of the formation of heavy elements, subsequent gas and solid state chemistry in interstellar and circumstellar environments and the formation of stars and planets may have been complemented by the extraterrestrial delivery of organic material to the early planets. Early Earth was a hostile place and barely able to form efficiently organic molecules in the atmosphere or on its surface. The constant delivery of cosmic carbon during the heavy bombardment phase in the first 700 million years of Earth's history may have enhanced the speed of molecular assembly. Though the fossil and isotopic data of ancient rocks are currently debated, there is no doubt that life originated about 1 billion years after the formation of the Earth. The ingredients for life include the building blocks of our genetic material DNA/RNA: sugars, purines, pyrimidines and phosphates; amino acids (building up proteins) and membrane components (e.g. fatty acids, phosphates, glycerol). Whereas many of these components are found in carbonaceous meteorites, hardly any of them has been detected anywhere else in space.

It appears possible that amino acids could form in hot molecular cores. Of course, this population may have nothing to do with the amino acid composition of primitive Solar System material. In order to relate these interstellar processes with meteoritic amino acids requires consideration of how these molecules could become incorporated in the parent body (asteroid or comet). Ideally one would like to observe amino acids in the interstellar medium. Despite much effort, Glycine has evaded detection (e.g. Snyder 1997; Ceccarelli et al. 2000).

One of the most important open questions remaining is the fraction of hidden cosmic carbon. We can estimate the carbon which is included in CO, in ices species (CO, CO_2, CH_3OH), in PAHs and carbon chains but a large fraction of the cosmic carbon (more than 50%) is likely in the form of aromatic networks. The same is evident for meteoritic matter, where more than 80% of the carbon is in the form of a kerogen-type network. It will require a synergy of astronomical observations and nanotechnology to shed light on the solid carbon structures in space and what role this macromolecular carbon played for chemical pathways. Future astronomical studies using new satellites and telescopes may provide important answers concerning the inventory of organics in space and the role of extraterrestrial delivery. To establish an inventory of organics formed on the early Earth in order to determine how easy it is to start life is an important step. Two approaches are currently used to reveal the beginning of our molecular existence: the "bottom up" model which investigates the molecular evolution from simple molecules to simple life, which includes the pathway to confinement and the "top down" model where a modern living cell is reduced to a minimal cell.

Together with the search for extrasolar planets and the ongoing exploration of our Solar System (e.g. using biosensors to search for life on Mars) the interdisciplinary approach using nanotechnology, combinatorial chemistry and high throughput screening may provide new insights into the origin of life in the following decades.

This work has been supported by NOVA, VIDI/NWO and ESA. Theoretical organic astrochemistry at NASA Ames is supported by NASA's Exobiology Program through funds allocated by NASA Ames under Interchange No. NCC2-1162.

REFERENCES

Aikawa, Y., van Zadelhoff, G. J., Herbst, E., & van Dishoeck, E. F. 2002 *A&A* **386**, 622.

Allen, M. & Robinson, G. W. 1977 *ApJ* **212**, 396.

Becker, L., Bada, J. L., Winans, R. E., & Bunch, T. E. 1994 *Nature* **372**, 507.

Becker, L., Poreda, R. J., Bunch, T. E. 2000 *Proc. Natl. Acad. Sci.* **97**, 2979.

Bernstein, M., et al. 2002 *Nature* **416**, 401.

Bockelée-Morvan, D., Lis, D. C., Wink, J. E., Despois, D., Crovisier, J., et al. 2000 *A&A* **353**, 1101.

Boogert, A., Hogerheijde, M. R., Ceccarelli, C., Tielens, A. G. G. M., van Dishoeck, E. F., Blake, G. A., Latter, W. B., & Motte, F. 2002 *ApJ* **570**, 708.

Botta, O. & Bada, J. 2002, *Surveys in Geophysics*, **23**, 411.

Botta, O., et al. 2002 *Origins of Life and Evolution of the Biosphere* **32**, 143.

Brown, P. D., Charnley, S. B., & Millar, T. J. 1988 *MNRAS* **231**, 409.

Ceccarelli, C., Loinard, L., Castets, A., Faure, A., Lefloch, B. 2000 *A&A* **326**, 1122.

Charnley, S. B. 1997, in *Astronomical & Biochemical Origins and the Search for Life in the Universe* (eds. C. B. Cosmovici, S. Bowyer & D. Wertheimer). p. 89. Editrice Compositori.

Charnley, S. B. 2001a *ApJ* **562**, L99.

Charnley, S. B. 2001b, in *The Bridge Between the Big Bang and Biology* (ed. F. Giovannelli). Special Volume, p. 139. Consiglio Nazionale delle Ricerche President Bureau.

Charnley, S. B., Kress, M. E., Tielens, A. G. G. M., & Millar, T. J. 1995 *ApJ* **448**, 232.

Charnley, S. B., Tielens, A. G. G. M., & Millar, T. J. 1992 *ApJ* **399**, L71.

Charnley, S. B., Tielens, A. G. G. M., & Rodgers, S. D. 1997 *ApJ* **482**, L203.

Chyba, C., Thomas, P. J., Brookshaw, L., & Sagan, C. 1990 *Science* **249**, 366.

Clayton, R. N. & Mayeda, T. K. 1984 *Earth Planet. Sci. Lett.* **67**, 151.

Cooper, G. W. & Cronin, J. R. 1995 *Geochim. Cosmochim. Acta* **59**, 1003.

Cronin, J. R. & Chang, S. 1993, in *The Chemistry of Life's Origins* (eds. J. M. Greenberg, et al.). p. 209. Kluwer Academic Publishers.

Crovisier, J. 1998 *Faraday Discuss.* **108**, 437.

Cruikshank, D. 1997, in *From Stardust to Planetesimals* (eds. Y. Pendleton, A. G. G. M. Tielens). p. 315. Astron. Soc. Pac.

Draine, B. 1990, in *The Evolution of the Interstellar Medium.* p. 193. Astron. Soc. Pac.

Dutrey, A., Guilloteau, S., & Guelin, M. 2000, in *Astrochemistry: From Molecular Clouds to Planetary Systems* (eds. Y. C. Minh and E. F. van Dishoeck). IAU Symp. 197, p. 415. Astron. Soc. Pac.

Ehrenfreund, P., Bernstein, M., Dworkin, J. P., Sandford, S. A., & Allamandola, L. 2001a *ApJ* **550**, L95.

Ehrenfreund, P., Cami, J., Jimenez-Vincente, J., Foing, B. H., Kaper, L., van der Meer, A., Cox, N., d'Hendecourt, L., Maier, J. P., Salama, F., Sarre, P., Snow, T. P., & Sonnentrucker, P. 2002b *ApJ* **576**, L117.

Ehrenfreund, P. & Charnley, S. B. 2000 *ARA&A* **38**, 427.

Ehrenfreund, P., d'Hendecourt, L., Charnley, S. B., & Ruiterkamp, R. 2001b *Journal Geophysical Research* **106**, E12, 33291.

Ehrenfreund, P., Glavin, D., Botta, O., Cooper, G. W., & Bada, J. L. 2001d *Proc. Natl. Acad. Sci. USA* **98**, 2138.

Ehrenfreund, P., Irvine, W., Becker, L., Blank, J., Brucato, J., Colangeli, L., Derenne, S., Despois, D., Dutrey, A., Fraaije, H., Lazcano, A., Owen, T., & Robert, F. 2002a *Reports on Progress in Physics* **65**, 10, 1427.

Ehrenfreund, P. & Schutte, W. A. 2000, in *Astrochemistry: From Molecular Clouds to Planetary Systems* (eds. Y. C. Minh and E. F. van Dishoeck). IAU Symp. 197, p. 135. Astron. Soc. Pac.

Ehrenfreund, P., Sonnentrucker, P., O'Tuairisg, S., Cami, J., & Foing, B. H. 2001c, in *The Bridge between the Big Bang and Biology* (ed. F. Giovannelli). Special Volume, p. 150. Consiglio Nazionale delle Ricerche Italy, President Bureau.

Endress, M. & Bischoff, A. 1996 *Geochim. Cosmochim. Acta* **60**, 489.

Ferris, J. P., Joshi, P. C., Edelson, E. H., & Lawless, J. G. 1978 *J. Mol. Evol.* **11**, 293.

Frenklach, M. & Feigelson, E. D. 1989 *A&A* **341**, 372.

Gibb, E., Whittet, D. C. B., Schutte, W. A., Chiar, J., Ehrenfreund, P., et al. 2000 *ApJ* **536**, 347.

Grim, R. & d'Hendecourt, L. B. 1986 *A&A* **167**, 162.

Henning, T. & Salama, F. 1998 *Science* **282**, 2204.

Herbig, G. H. 1995 *ARA&A* **33**, 19.

Herbst, E. 1978 *ApJ* **222**, 508.

Herbst, E. 2000, in *Astrochemistry: From Molecular Clouds to Planetary Systems* (eds. Y. C. Minh and E. F. van Dishoeck). IAU Symp. 197, p. 147. Astron. Soc. Pac.

Hiraoka, K., Miyagoshi, A., Takayama, T., Yamamoto, K., & Kihara, Y. 1998 *ApJ* **498**, 710.

Hiraoka, K., Ohashi, N., Kihara, Y., Yamamoto, K., Sato, T., & Yamashita, A. 1994 *Chem. Phys. Lett.* **229**, 408.

Hollis, J. M., Lovas, F. J., & Jewell, P. R. 2000 *ApJ* **540**, L107.

Hollis, J. M., Lovas, F. J., & Jewell, P. R. 2002, *ApJ* **571**, L59.

Hunter, E. P. & Lias, S. 2001, in *NIST Chemistry WebBook, NIST Standard Reference Database Number 69* (eds. P. J. Linstrom and W. G. Mallard). National Institute of Standards and Technology; http://webbook.nist.gov.

Irvine, W. M., Schloerb, F. P., Crovisier, J., Fegley, B. Jr., & Mumma, M. J. 2000 in *Protostars and Planets IV* (eds. V. Mannings, A. Boss, and S. Russell). p. 1159. University of Arizona Press.

Jenniskens, P. & Désert, F. X. 1994 *A&AS* **106**, 39.

Jenniskens, P. & Greenberg, J. M. 1993 *A&A* **274**, 439.

Kissel, J. & Krueger, F. R. 1987 *Nature* **326**, 755.

Krishnamurthy, R. V., Epstein, S., Cronin, J. R., Pizzarello, S., & Yuen, G. U. 1992 *Geochim. Cosmochim. Acta* **56**, 4045.

Kvenvolden, K., Lawless, J., Pering, K., Peterson, E., Flores, J., Ponnamperuma, C., Kaplan, I. R., & Moore, C. 1970 *Nature* **228**, 923.

Langer, W. D., van Dishoeck, E. F., Bergin, E. A., Blake, G. A., Tielens, A. G. G. M., Velusamy, T., & Whittet, D. C. B. 2000, in *Protostars and Planets IV* (eds. V. Mannings, A. Boss, and S. Russell). p. 29. University of Arizona Press.

Leshin, L. A., Rubin, A. E., & McKeegan, K. D. 1997 *Geochim. Cosmochim. Acta* **61**, 835.

Lias, S., et al. 1984 *J. Chem. Phys. Ref. Data* **17**, 1.

Liu, S.-Y., Mehringer, D. M., Snyder, L. E. 2001 *ApJ* **552**, 654.

Loinard, L., Castets, A., Ceccarelli, C., Caux, E., Tielens, A. G. G. M. 2001 *ApJ* **552**, L163.

Lucas, R. & Liszt, H. S. 1997, in *Molecules in Astrophysics: Probes and Processes* (ed. E. F. van Dishoeck). p. 421. Kluwer Academic Publishers.

Lucas, R. & Liszt, H. S. 2000 *A&A* **355**, L327.

Manico, G., et al. 2001 *ApJ* **548**, L253.

Markwick, A. J., Ilgner, M., Millar, T. J., & Henning, T. 2002 *A&A* **385**, 632.

Mautner, M. & Karpas, Z. 1986 *J. Phys. Chem.* **90**, 2206.

McDonald, G. D., Thompson, W. R., Heinrich, M., Khare, B. N., & Sagan, C. 1994 *Icarus* **108**, 137.

McGlone, S. J. & Godfrey, P. D. 1994 *J. Am. Chem. Soc.* **117**, 1043.

McSween, H. Y. 1979 *Rev. Geophys. Space Phys.* **17**, 1059.

Mennella, V., Colangeli, L., Bussoletti, E., Palumbo, P., & Rotundi, A. 1998 *ApJ* **507**, L177.

Miller, S. L. 1953 *Science* **117**, 528.

Munoz-Caro, G. M., et al. 2002 *Nature* **416**, 403.

Naraoka, H., Shimoyama, A., Komiya, M., Yamamoto, H., & Harada, K. 1988 *Chem. Lett.* **831**.

O'Tuairisg, S., Cami, J., Foing, B. H., Sonnentrucker, P., & Ehrenfreund, P. 2000, *A&AS* **142**, 225.

Papoular, R., Conard, J., Guillois, O., Nenner, I., Reynaud, C., & Rouzaud, J. N. 1996 *A&A* **315**, 222.

Parise, B., Ceccarelli, C., Tielens, A. G. G. M., Herbst, E., Lefloch, B., Caux, E., Castets, A., Mukhopadhyay, I., Pagani, L., & Loinard, L. 2002 *A&A* **393**, L49.

Peltzer, E. T., Bada, J. L., Schlesinger, G., & Miller, S. L. 1984 *Adv. Space Sci.* **4**, 69.

Pirronello, V., Biham, O., Liu, C., Shen, L., & Vidali, G. 1997 *ApJ* **483**, L131.

Pirronello, V., Liu, C., Roser, J., & Vidali, G. 1999 *A&A* **344**, 681.

Pizzarello, S. & Cronin, J. R. 2000 *Geochim. Cosmochim. Acta* **64**, 329.

Pizzarello, S., Krishnamurthy, R. V., Epstein, S., & Cronin, J. R. 1991 *Geochim. Cosmochim. Acta* **55**, 905.

Prasad, S. S. & Tarafdar, S. P. 1983 *ApJ* **267**, 603.

Remijan, A., et al. 2002 *ApJ* **576**, 264.

Roser, J. E., Vidali, G., Manico, G., & Pirronello, V. 2001 *ApJ* **555**, L61.

Roueff, E., Felenbok, P., Black, J., & Gry, C. 2002 *A&A* **384**, 629.

Schutte, W. A., et al. 1996 *A&A* **315**, 333.

Snow, T. P. 2002, in *Gaseous Matter in Galaxies and Intergalactic Space* (eds. R. Ferlet, M. Lemoine, J.-M. Désert & B. Raban). p. 63. Frontier Group.

Snyder, L. 1997 *Origins of Life and Evolution of the Biosphere* **27**, 115.

Snyder, L. E., Lovas, F. J., Mehringer, D. M., Miao, N. Y., Kuan, Y.-J., Hollis, J. M., & Jewell, P. R. 2002 *ApJ*, **578**, 245.

Sorrell, W. 2001 *ApJ* **555**, L129.

Stoks, P. G. & Schwartz, A. W. 1979 *Nature* **282**, 709.

Stoks, P. G. & Schwartz, A. W. 1981 *Geochim. Cosmochim. Acta* **45**, 563.

Stoks, P. G. & Schwartz, A. W. 1982 *Geochim. Cosmochim. Acta* **46**, 309.

Thi, W. F., Pontoppidan, K., van Dishoeck, E. F., Dartois, E., & d'Hendecourt, L. 2002 *A&A*, **394**, L27.

Tielens, A. G. G. M. & Hagen, W. 1982 *A&A* **114**, 245.

Tielens, A. G. G. M., Hony, S., van Kerckhoven, C., & Peeters, E. 1999, in *The Universe as seen by ISO*. **427**. p. 579. ESA Publ. Div.

Turner, B. E. & Apponi, A. 2001 *ApJ* **561**, 207.

van Ijzendoorn, L. 1985 *PhD Thesis*. Leiden.

Watanabe, N. & Kouchi, A. 2002 *ApJ* **571**, L173.

Woon, D. E. 2002 *ApJ* **571**, L177.

Wootten, H. A. 2002, http://www.cv.nrao.edu/∼awootten/allmols.html.

Zinner, A. & Amari S. 1999, in *Asymptotic Giant Branch Stars*, (eds. T. Le Bertre, A. Lebre, and C. Waelkens). IAU Symposium 191, p. 59. Astron. Soc. Pac.

Galactic environment of the Sun and stars: Interstellar and interplanetary material

By PRISCILLA C. FRISCH,[1] HANS R. MÜLLER,[2] GARY P. ZANK,[3] AND C. LOPATE[1]

[1]University of Chicago, Department of Physics & Astronomy, 5640 S. Ellis Avenue, Chicago, IL 60637, USA

[2]Bartol Research Institute, University of Delaware, Newark, DE 19716, USA

[3]IGPP, University of California, Riverside, CA 92521, USA

Interstellar material surrounding an extrasolar planetary system interacts with the stellar wind to form the stellar astrosphere, and regulates the properties of the interplanetary medium and cosmic ray fluxes throughout the system. Advanced life and civilization developed on Earth during the time interval when the Sun was immersed in the vacuum of the Local Bubble and the heliosphere was large, and probably devoid of most anomalous and galactic cosmic rays. The Sun entered an outflow of diffuse cloud material from the Sco-Cen Association within the past several thousand years. By analogy with the Sun and solar system, the Galactic environment of an extrasolar planetary system must be a key component in understanding the distribution of systems with stable interplanetary environments, and inner planets which are shielded by stellar winds from interstellar matter (ISM), such as might be expected for stable planetary climates.

1. Introduction

Our solar system is the best template for understanding the properties of extrasolar planetary systems. The interaction between the Sun and the constituents of its galactic environment regulates the properties of the interplanetary medium, including the influx of interstellar matter (ISM) and galactic cosmic rays (GCR) onto planetary atmospheres. In the case of the Earth, the evolution of advanced life occurred during the several million year time period when the Sun was immersed in the vacuum of the Local Bubble (Frisch & York 1986, Frisch 1993). Here we use our understanding of our heliosphere to investigate the astrospheres around extrasolar planetary systems.†

The heliosphere, or solar wind bubble, is dominated by interstellar matter (a visualization of the heliosphere is shown in Fig. 1). Interstellar gas constitutes ∼98% of the diffuse material in the heliosphere, and the solar wind and interstellar gas densities are equal near the orbit of Jupiter, beyond which the ISM density dominates. The solar wind and photoionization prevents nearly all ISM from reaching the Earth. Interstellar ions and the smallest interstellar dust grains ($<0.1\ \mu$m) are deflected around the heliosphere. Neutral ISM, however, enters the heliosphere where it dominates the interplanetary environment throughout most of the heliosphere, except for the innermost regions where the solar wind dominates. Inner and outer planets experience radically different exposures to raw ISM over the lifetime of a planetary system. The exposure levels of the Earth to galactic cosmic rays and raw and processed ISM depends sensitively on heliospheric properties.

Longstanding theories suggest that interstellar material has the potential to modify the terrestrial climate. These theories have recently become less speculative because of the improved understanding of cosmic ray modulation in a time-varying heliosphere and

† This paper is based on the talk presented at the Space Telescope Science Institute May, 2002 Symposium on the *Astrophysics of Life*. See “Interstellar and Interplanetary Material,” linked to http://ntweb.stsci.edu/sd/astrophysicsoflife/index.html.

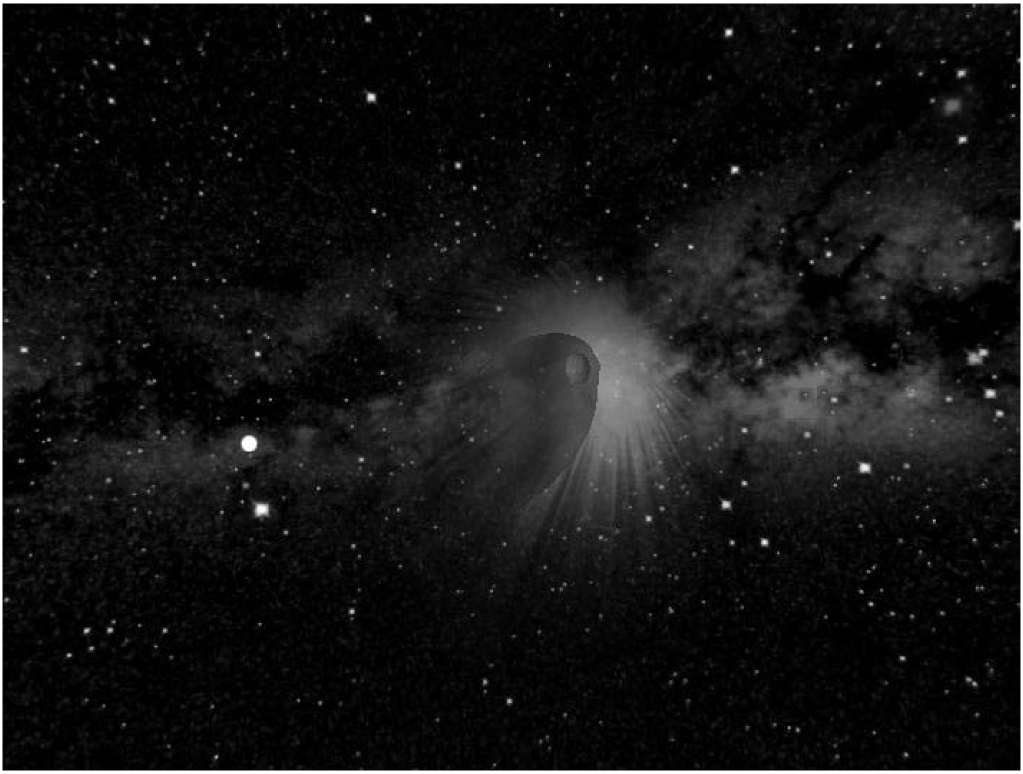

FIGURE 1. *The Galactic Environment of the Sun—Upstream viewpoint:* Visualization of heliosphere moving through our galactic neighborhood, based on an MHD simulation of the heliosphere morphology which includes the relative orientation and ram pressures of the interstellar and solar wind magnetic fields due to the ecliptic tilt with respect to the galactic plane (Linde et al. 1998). There is a north-south asymmetry in the heliosphere from the ecliptic tilt with respect to the interstellar magnetic field. The Mach ∼1 bow shock around the heliosphere is apparent, as is the termination shock of the solar wind (the smaller rounded surface inside of the heliopause where the solar wind transitions to subsonic). This figure is excerpted from a movie showing a 3D visualization of the heliosphere and the Milky Way Galaxy, which can be viewed at http://cs.indiana.edu/∼soljourn.

the relation between cosmic rays fluxes and atmospheric electricity and tropospheric cloud cover (Section 6). Extrasolar planetary systems are surrounded by astrospheres formed by the interaction between stellar winds and interstellar material. In turn, these astrospheres modulate the entrance and transport of galactic cosmic rays, anomalous cosmic rays, neutral interstellar (IS) atoms, and IS dust into and within the planetary system. Planet habitability has been evaluated in terms of atmospheric chemistry and energy budget (see other papers in this volume). However by analogy with the solar system, an historically stable astrosphere may also be a predictor for stable planetary climates and thus the conditions which promote the development of advanced life. It is this relation between the galactic environment of a star, the stellar astrosphere, and the properties and prehistory of the interplanetary medium of planetary systems that are of the greatest interest.

2. Heliosphere and interstellar matter

The heliosphere is the region of space filled by the solar wind, which is the expanding solar corona. The solar wind corresponds to a solar mass loss rate of ~10^{-14} $M_{\rm Sun}$ year^{-1}. The solar wind density decreases with R^{-2} as the solar wind expands, and the solar wind and interstellar medium pressures are equal at a plasma contact discontinuity known as the "heliopause" (e.g. Axford 1972, Holzer 1989). The basic properties of the heliosphere are shown in Fig. 2 (Zank et al. 1996). At the solar wind termination shock the solar wind becomes subsonic and the cool supersonic solar wind plasma is shock-heated to a hot (T $\sim 2 \times 10^6$ K) subsonic plasma. Interstellar neutrals cross the plasma regions with interaction mean free paths ~100 AU. If the relative Sun-cloud velocity (26 km s^{-1}) exceeds the fast magnetosonic speed of the surrounding interstellar cloud, a bow shock will form around the heliosphere.

Solar wind properties vary with the 22-year magnetic activity cycle of the Sun, with the solar magnetic polarity changing every 11 years during the period of the maximum in solar activity. During solar minimum, high speed low density solar wind forms in coronal holes at the solar poles ($n(p^+) \sim 2.5$ cm^{-3}, velocity $V \sim 770$ km s^{-1}, McComas et al. 2001). During solar maximum conditions, high speed stream material expands to the equatorial regions and the 1 AU ecliptic solar wind properties are: density $n(p^+) \sim$ 4–8 cm^{-3}, velocity $V \sim$ 350–750 km s^{-1}, and magnetic field $B \sim 2$ nT (or 20 μG). The activity cycle of the Sun is known to produce small modifications in the heliosphere over the 11-year solar cycle, with the termination shock moving outwards ~10 AU in the upwind direction, and outwards by ~40–50 AU in the downstream direction during solar minimum.

The Sun is presently in a low density, warm, partially ionized interstellar cloud with $n_H \sim 0.24$ cm^{-3}, $n(e^-) \sim 0.1$ cm^{-3}, and T $\sim 6,500$ K (Slavin & Frisch 2002). The upstream direction of the surrounding cloud, known as the Local Interstellar Cloud (LIC), is towards $l^{II} = 3.3^o$, $b^{II} = +15.9^o$ (in the rest frame of the Sun) and the relative Sun-LIC velocity is 26.4 ± 0.5 km s^{-1} (Witte, private communication). The LIC upstream direction in the local standard of rest (LSR, after removing the solar apex motion) is $l = 346^o$, $b = -1^o$ with a LIC velocity through the LSR of -15 km s^{-1}. The LIC is a member of a cluster of cloudlets flowing at -17 ± 5 km s^{-1} from the LSR upstream direction of $l^{II} = 2^o$, $b^{II} = -5^o$ (Frisch et al. 2002). The LSR upstream direction is sensitive to the assumed solar apex motion.†

The present-day Galactic environment of the Sun yields a highly asymmetrical heliosphere that is much larger than the planetary system. A range of multifluid, Boltzmann-kinetic, and MHD models of the heliosphere has been developed (see Zank, 1999, for a review). In the upstream direction, the solar wind termination shock (where the solar wind becomes subsonic) is at about 75–90 AU. The heliopause is located near 140 AU and represents the contact discontinuity between the solar wind and interstellar plasma component. The Sun is moving supersonically with respect to the LIC (sound speed is ~10 km s^{-1}), however a weak interstellar magnetic field (~3 μG, fast mode velocity ~23 km s^{-1}) may yield a barely supersonic heliosphere (M $\sim$ 1) with a bow shock. Several heliosphere models place a weak bow shock at ~250 AU in the upstream direction (see Zank 1999). For comparison, the planet Pluto is at 39 AU, and the *Voyager 1* and *Voyager 2* spacecraft are at 84 AU and 65 AU, respectively. In the downstream direction, the termination shock is elongated by a factor of ~2 compared to the upstream direction.

† These quoted values use a solar apex motion derived from Hipparcos data (Dehnen Binney 1998). The basic solar apex motion yields the LIC LSR upstream direction $l^{II} \sim 326^o$, $b^{II} \sim +4^o$ (Frisch 1995).

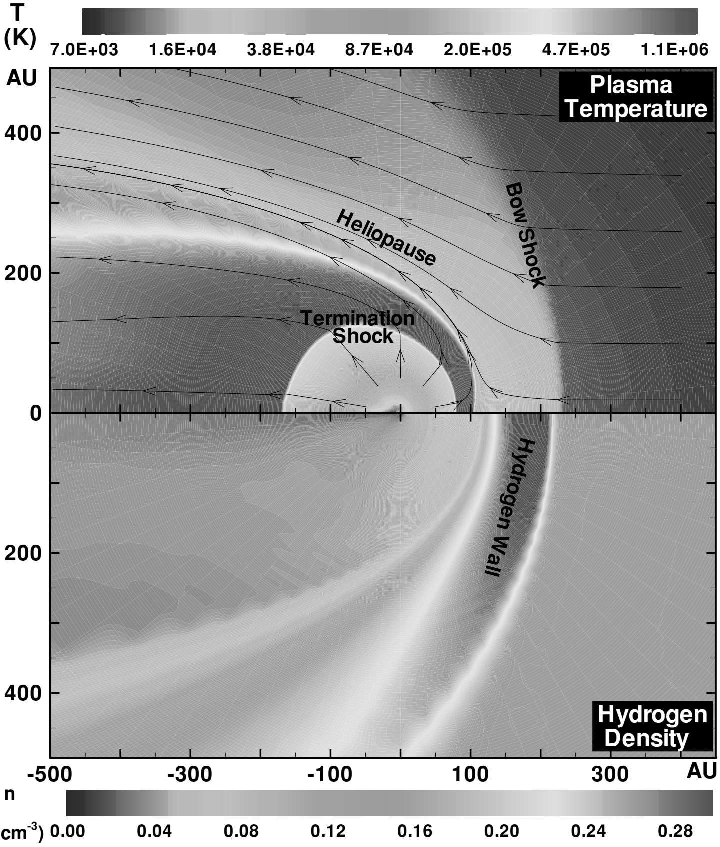

FIGURE 2. This figure displays the neutral hydrogen density (bottom panel) and plasma temperature (top panel) of the heliosphere immersed in the LIC, which has properties T $\sim 6,500$ K, $n(H^o) \sim 0.24$ cm^{-3}, $n(H^+) \sim 0.1$ cm^{-3}, and an unknown but probably weak magnetic field. The hydrogen wall is formed by charge exchange coupling between weakly decelerated and deflected interstellar protons, and interstellar H^o.

The north ecliptic pole points towards the galactic coordinates $l = 96^o$, $b = +30^o$, so the ecliptic plane is inclined by ~60^o with respect to the plane of the galaxy. A pronounced asymmetry between the northern and southern ecliptic is predicted for the heliosphere because of this tilt and the LIC upstream direction (e.g. Linde et al. 1998), combined with the likelihood that the localized interstellar magnetic field is in the galactic plane (Frisch 1990).

Interstellar plasma piles up against the compressed solar wind in the outer heliosphere, and charge-coupling between interstellar H^o and interstellar H^+ produces a low column density ($N(H^o) \sim 3 \times 10^{14}$ cm^{-2}), decelerated ($\delta V \sim 8$ km s^{-1}), heated (~29,000 K) H^o component that is visible as a redshifted shoulder in the Lyα absorption profile towards α Cen (Linsky Wood 1996, Gayley et al. 1997, the "hydrogen wall"). Similar pileups of interstellar H^o have been detected against the astrospheres around several nearby cool stars (Section 5).

The charged component of the ISM is deflected by the tightly wound solar wind magnetic field in the heliosheath region. The smallest interstellar dust grains(<0.1 μm) are also deflected around the heliopause (Frisch et al. 1999). Neutral ISM, however, enters the heliosphere where it dominates the interplanetary environment throughout most of the heliosphere, with the exception of the innermost regions where the solar wind dominates.

The *Voyager 1* and *Voyager 2* spacecraft are sending back data from the frontiers of the outer heliosphere, and future spacecraft may penetrate interstellar space (e.g. the *Interstellar Probe* mission, Liewer & Mewaldt 2000) and provide the first *in situ* measurements of the galactic environment of the Sun. These spacecraft, and others (e.g. *Ulysses*, *Galileo*, *Cassini*) have provided a wealth of data which clearly demonstrate that the ISM dominates the interplanetary environment throughout most of the solar system and heliosphere.

3. Historical variations of the heliosphere

The Galactic environment of the Sun and stars vary with the motions of the stars and interstellar clouds through space. The Sun itself has been immersed in the vacuum of the Local Bubble ($n(H^o) < 0.0005$ cm^{-3}, $n(H^+) \sim 0.005$ cm^{-3}, and T $\sim 10^6$ K) during the millions of years over which homo sapiens developed and civilization emerged (Frisch & York 1986, Frisch 1993). The Sun has recently (2,000–10^5 years ago) entered an outflow of diffuse ISM from the Sco-Cen Association (Frisch 1994, Frisch et al. 2002), and is now surrounded by a warm low density partially ionized cloud. The Sun may encounter other possibly denser cloudlets in the flow, with one possibility being the "Aql-Oph" cloudlet that is within 5 pc of the Sun near the solar apex direction. A study of nearby ISM shows 96 interstellar absorption components are seen towards 60 nearby stars sampling ISM within 30 pc (Frisch et al. 2002). Since the nearest stars show ~1 interstellar absorption component per 1.4–1.6 pc, relative Sun-cloud velocities of 0–32 km s^{-1} suggest variations in the galactic environment of the Sun on timescales <50,000 years.

The galactic environment of an astrosphere has a striking effect on the resulting astrosphere. This is illustrated in Fig. 3 for the heliosphere, which shows the heliosphere properties several million years ago when the heliosphere was embedded in the Local Bubble (left), and at some time in the future when it might be embedded in a cloud with density $n(H^o) = 15$ cm^{-3} (but otherwise like the LIC). During the time the Sun was embedded in the fully ionized Local Bubble Plasma, described by T $= 10^6$ K, $n(\mathrm{p}^+) = 0.005$ cm^{-3}, there were no interstellar neutrals in the heliosphere, and hence very few pickup ions or anomalous cosmic rays (very small quantities of each may have been present from a poorly understood inner source that may be related to either interplanetary dust or outgassing from planetary atmospheres). An increase to $n = 10$ cm^{-3} for the cloud around the Sun would contract the heliopause to radius of ~14 AU, increase the density of neutrals at 1 AU to 2 cm^{-3}, and create a Rayleigh-Taylor unstable heliopause from variable mass loading of solar wind by pickup ions (Zank & Frisch 1999). Models with higher densities (e.g. $n = 15$ cm^{-3}, T $= 3,000$ K) show that planets beyond ~15 AU (Uranus, Neptune, Pluto) will be outside of the heliosphere for moderate density diffuse clouds, and thus exposed to raw ISM. The Sun is predicted to encounter about a dozen giant molecular clouds, with much higher densities (>10^3 cm^{-3}) over its lifetime (Talbot & Newman 1977), but encounters with diffuse clouds ($n \sim 10$ cm^{-3}) will occur more frequently.

4. Interstellar and interplanetary matter

Components of the interstellar medium which enter the heliosphere from deep space include neutral gas atoms, larger interstellar dust grains, and galactic cosmic rays. The products created by the interactions of the ISM and solar wind create an ISM-dominated

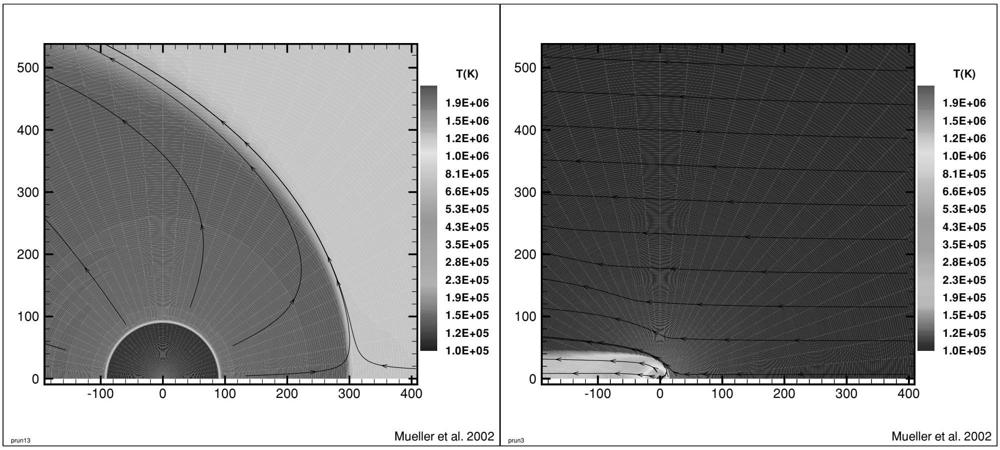

FIGURE 3. Heliosphere predicted for a Sun immersed in the hot Local Bubble (left) and immersed in a $n_H = 15$ cm^{-3} diffuse cold cloud (right). Figure from Müller et al. (2002).

heliosphere. Fig. 4 shows an overview of the heliosphere, with the products of the interaction between the ISM and solar wind identified.

4.1. *High Energy Galactic Cosmic Rays in the Heliosphere*

Galactic cosmic rays with energies less than ~100 GeV/nucleon are modulated by the increasingly nonuniform structure of magnetic fields embedded in the outward flowing solar wind during solar maximum (Fig. 6). The result of this modulation is a well known anti-correlation between the solar activity cycle and the cosmic ray flux at the Earth's surface. The anti-correlation is illustrated in Fig. 5, which shows neutron monitor counts, from secondary particles produced by cosmic ray interactions at the top of the atmosphere, versus the sunspot number. This anticorrelation reflects variations in the heliospheric modulation of the galactic cosmic ray flux as a function of the solar wind magnetic activity. Anomalous cosmic rays (see below), formed by accelerated pickup ions, experience modulation in the heliosphere similar to GCRs. Most cosmic ray modulation occurs in the outer part of the heliosphere, so that evidence of CR interactions on meteorites or planetary surfaces should contain fossil evidence on the heliosphere radius. The heliosphere varies with the solar cycle, as does cosmic ray modulation. Disorder in the solar wind magnetic field at sunspot maximum corresponds to an increase in cosmic ray modulation, although the heliosphere is smaller than at solar minimum. GCRs are capable of changing the flow pattern of the solar wind and the surrounding local ISM provided the particles' coupling to the plasma is sufficiently strong. The interstellar cosmic-ray spectra and the diffusion coefficients and cosmic-ray pressure gradients within the heliosphere are now becoming better understood (e.g. Ip & Axford 1985).

4.2. *Raw ISM in the heliosphere: H^o, He^o*

Neutral interstellar H and He atoms enter and penetrate the solar system, and are ionized by charge exchange with the solar wind or photoionization. A weak interplanetary glow from the fluorescence of solar Lyα radiation off of interstellar H^o, and solar 584 Å radiation off of interstellar He^o, led to the discovery of interstellar matter in the solar system in 1971 (Thomas & Krassa 1971, Bertaux & Blamont 1971, Weller & Meier 1974). H^o is ionized at ~4 AU by charge exchange with the solar wind and photoionization, while He^o penetrates to ~0.4 AU before becoming photoionized. The flux of He^o atoms has been measured directly by Ulysses, yielding values $n(He^o) = 0.014 \pm 0.002$ cm^{-3}, temperature

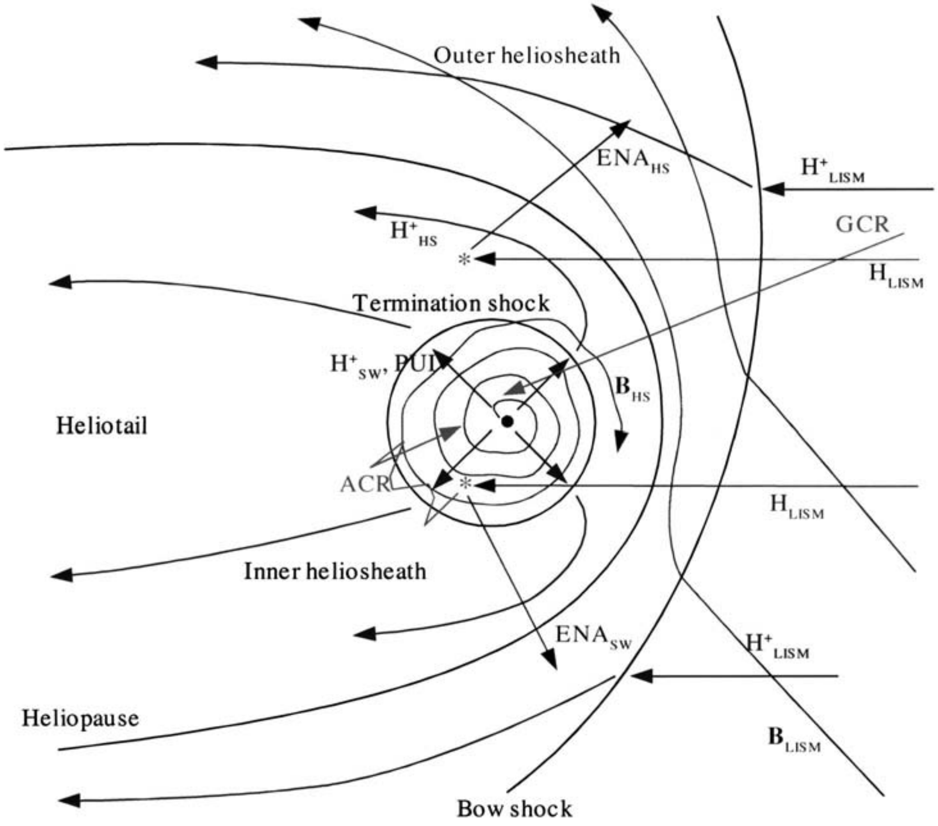

FIGURE 4. Overview of the heliosphere, with termination shock, heliopause, bow shock, and outer and inner heliosheath (HS). Some sample plasma (H^+), pickup ion (PU Ion), and solar wind plasma (v_{HS}) trajectories are shown, as well as trajectories of neutral hydrogen (H) coming from the interstellar medium (H_{ISM}) and experiencing charge exchange (*), and galactic cosmic rays (GCR). The solar and interstellar magnetic fields (B) are sketched (based on a plot by J. R. Jokipii).

6,500 K, and velocity of 26.4 km s^{-1} and and upstream direction $l^{II} = 3.3^o$, $b^{II} = +15.9^o$ (Witte et al. 1996 and private communication). The first spectral observations of interstellar H^o in the solar system observed a projected velocity -24.1 ± 2.6 km s^{-1} during solar minimum towards the direction $l^{II} = 16.8^o$, $b^{II} = +12.3^o$ (Adams & Frisch 1977). Correcting this velocity towards the He^o upstream direction gives a cloud velocity 24.8 ± 2.6 km s^{-1}, in agreement with the He^o velocity (since during solar minimum radiation pressure and gravity are approximately equal). The LSR upstream direction of the LIC is $l^{II} \sim 346^o$, $b^{II} \sim -1^o$.

Interstellar H^o and He^o behave differently in the heliosphere. About 20%–40% of the H^o is lost in the outer heliosheath through charge-exchange with interstellar H^+, and once in the solar system the H^o trajectory is governed by the relative strengths of the solar Lyα radiation pressure force and gravity. Interstellar He^o passes through the heliosheath unaltered, and the trajectory in the heliosphere is governed by gravity so that interstellar He is gravitationally focused downstream of the Sun. The Earth passes through the He focusing cone about December 1 of each year. The He^o cone density is enhanced at 1 AU

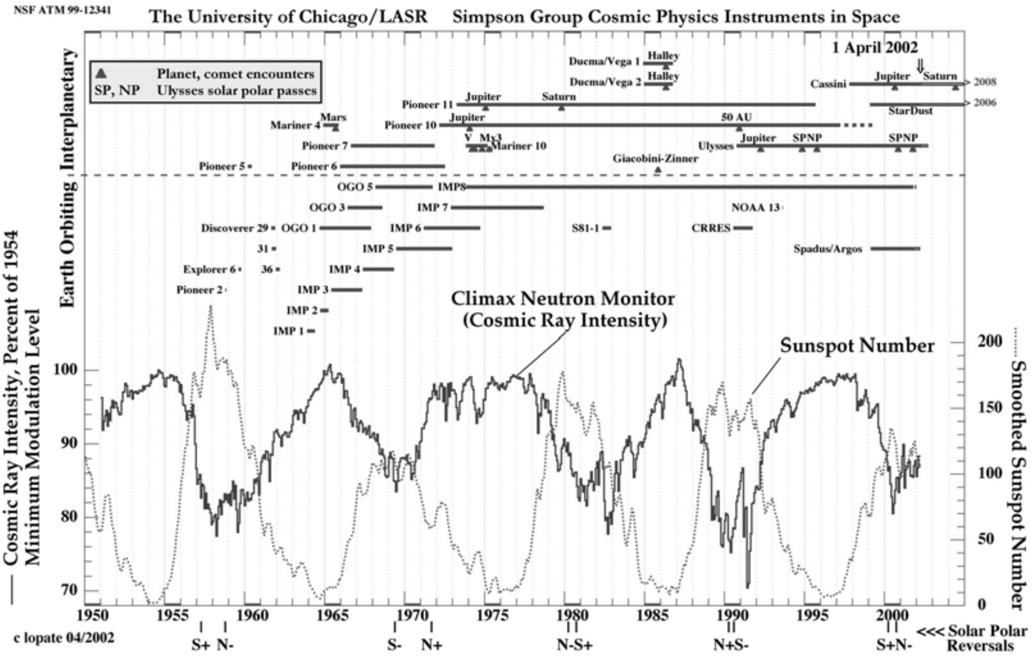

FIGURE 5. Solar cycle modulation of >3 GeV galactic cosmic rays: Sunspot number versus modulated galactic cosmic ray intensity. This figure is also available at http://ulysses.uchicago.edu/NeutronMonitor/neutron_mon.html along with related data.

by a factor of ∼250 over the value at infinity, but the peak density of the focusing cone is inside 1 AU (Michels et al. 2002).

4.3. *Raw ISM in the heliosphere: Dust*

Interstellar dust grains (ISDG) with radii >0.2 μm enter the heliosphere and have been detected by instruments on board *Ulysses*, *Galileo*, and *Cassini* (e.g. Baguhl et al. 1996, Frisch et al. 1999, Landgraf 2000). The mass flux distribution of these grains is shown in Fig. 7. Smaller grains (< 0.1 μm) are deflected in the heliosheath region and do not enter the heliosphere.

Large ISDGs (radii >0.35 μm) are focused downstream of the Sun, in a prominent gravitational focusing cone which is more extensive than the He focusing cone, extending over 10 AU in the downstream direction (Landgraf 2000). Large ISDGs constitute ∼30% of the interplanetary grain flux with masses $> 10^{13}$ gr (or radius>0.2 μm) at 1 AU (Gruen & Landgraf 2000).

ISDGs in the size range comparable to classical dust particles (0.1–0.2 μm, charge ∼1 eV) show a distribution in the heliosphere which varies with time because of Lorentz coupling to a solar wind magnetic field which changes in polarity every 11-year solar cycle. These positively charged grains alternately are focused and defocused towards the ecliptic plane. The 1996 solar minimum corresponded to a defocusing phase (Landgraf 2000).

The gas-to-dust mass ratio (R_{gd}) in the LIC is $R_{gd} = 125^{+18}_{-14}$, based on comparisons between interstellar dust in the solar system and the properties for the gas in the LIC, or $R_{gd} = 158$ based on missing mass arguments (Frisch & Slavin, 2002).

Radar measurements of micrometeorites show sources from outside the solar system. Interstellar micrometeorites with masses $\sim 10^{-7}$ g are detected by radar observations

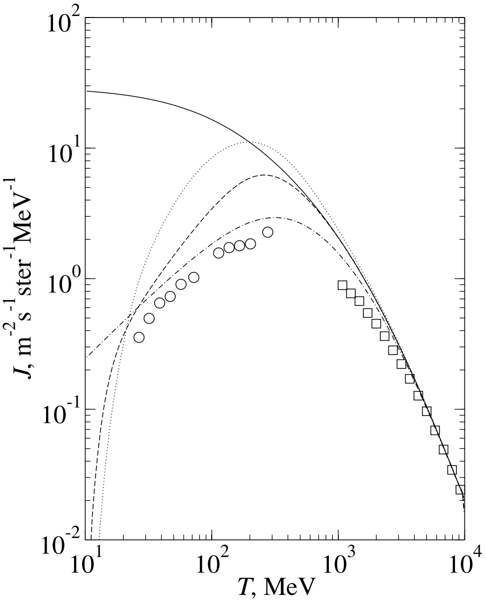

FIGURE 6. An example of the modulated cosmic ray spectrum at different locations in the heliosphere. The dashed line shows the modulated proton spectrum in the heliosheath ($\theta = 0^0$) at 110 AU, the dash-dotted line is for the supersonic solar wind at 10 AU, and the dotted line is for the heliotail ($\theta = 180^0$) at 650 AU. The un modulated interstellar spectrum is shown as a solid line. Experimental data from BESS (squares) and IMP8 (circles) are shown for comparison. Figure from Florinski et al. (2002). (At 10^2 MeV, from top to bottom the lines are: solid, dotted, dashed, dot-dashed.)

of the atmospheric trajectories and velocities (Baggaley 2000, Landgraf et al. 2000). A discrete source is seen at the location of β Pic (determined after solar motion is removed). These observations from the southern hemisphere also show an enhanced flux from the southern ecliptic. In the northern hemisphere, Doppler radar measurements of micrometeorites provide evidence for a radiant direction towards the Local Bubble (Meisel et al. 2002).

4.4. *Solar wind-ISM interactions products: Pickup ions and anomalous cosmic rays*

Interstellar atoms with first ionization potentials $\gtrsim$13.6 eV enter and penetrate the solar system, and are ionized by charge exchange with the solar wind. The resulting ions are coupled to the solar wind by the Lorentz force, where they are observed as a population of pickup ions (PUI, Gloeckler & Geiss 2002). PUIs of H, He, N, O, and Ne provide a direct sample of ionization levels in the LIC (Slavin & Frisch 2002). PUIs are accelerated to cosmic ray energies in the region of the termination shock of the solar wind, forming an anomalous population of cosmic rays (Garcia-Munoz et al. 1973, McDonald et al. 1974, Fisk et al. 1974). Anomalous cosmic rays, which are "anomalous" because of composition and energy, typically have lower energies than galactic cosmic rays. The anomalous cosmic ray H, He, N, O, Ne, and Ar populations have an interstellar origin, and thus provide an additional tracer of the neutral species in the LIC (Cummings & Stone 2002). Anomalous cosmic rays with energies >1 MeV/nucleon and an interstellar origin are also found trapped in the radiation belts of the Earths magnetosphere (e.g. Adams & Tylka 1993, Mazur et al. 2000).

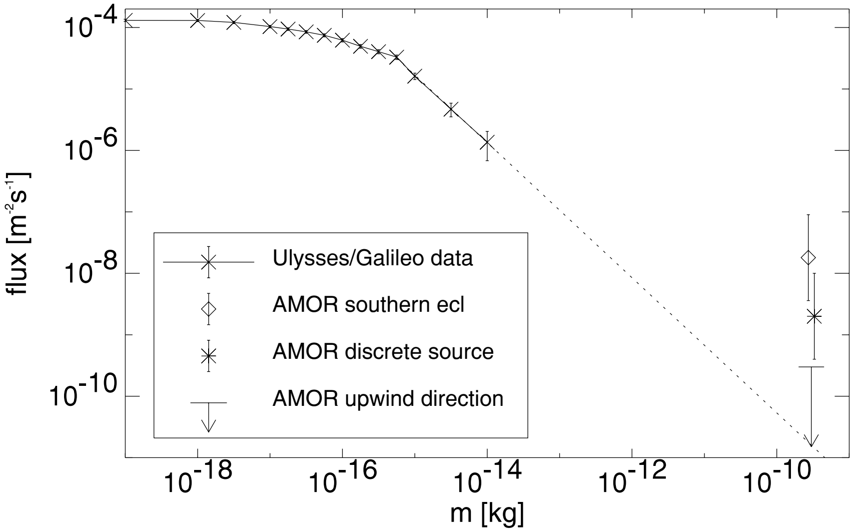

FIGURE 7. Mass flux of interstellar dust grains observed within the solar system by the *Ulysses*, *Galileo*, and *Cassini* spacecraft (Baguhl et al. 1996, Landgraf et al. 2000). The AMOR radar data points are of extrasolar micro-meteorites, and the point source corresponds to a direction towards β Pic (Baggaley 2000).

5. Astrospheres and extrasolar planetary system

An astrosphere is the stellar wind bubble around a cool star. Cool stars with stellar winds will have astrospheres regulated by the physical properties of the interstellar cloud surrounding each star (Frisch 1993), and stellar mass loss properties can be inferred from H^o Lyα absorption formed in the hydrogen wall region in the compressed heliosheath gas (Wood et al. 2002). The nearest star α Cen AB (1.3 pc) has a mass loss rate ∼2 times greater than the solar value (Wood et al. 2001). The pileup of interstellar H^o in the nose region of astrospheres surrounding nearby cool stars (e.g. α Cen, ϵ Eri, 61 CygA, 36 OphAB, 40 Eri A; Gayley et al. 1997, Wood et al. 2002), indicates that other cool stars have astrospheres which can be modeled using methodology developed for the heliosphere.

The astrosphere configuration for extrasolar planetary systems will vary with the individual properties of each system. The Sun moves through the local standard of rest with a velocity of V∼13 km s^{-1}, but many cool stars have larger velocities. Typical diffuse interstellar clouds move through space with velocities 0–20 km s^{-1} (or more), and the dynamical ram pressure (∼V^2) may vary by factors of ∼10^3, and cause variations in the astrosphere radius of factors of >30. The result is that inner and outer planets of extrasolar planetary systems will be exposed to different amounts of raw interstellar matter over the lifetime of the planetary system. Frisch (1993) estimated astrosphere radii and historical galactic environments of ∼70 G-stars within 35 pc of the Sun from the basic Axford-Holzer equation using the correct stellar dynamics, a solar-like stellar wind, and a realistic guess for the cloud properties. However, this primitive approach can now be improved upon with sophisticated multifluid astrosphere models (e.g. Zank 1999), improved data from the *Hipparcos* catalog, and improved understanding of the nearby ISM.

Astrosphere models, based on self-consistent algorithms for the coupling of interstellar and secondary neutrals and ions through charge exchange, predict observable signatures of the interaction of stellar winds and the ISM. The interaction products contain several

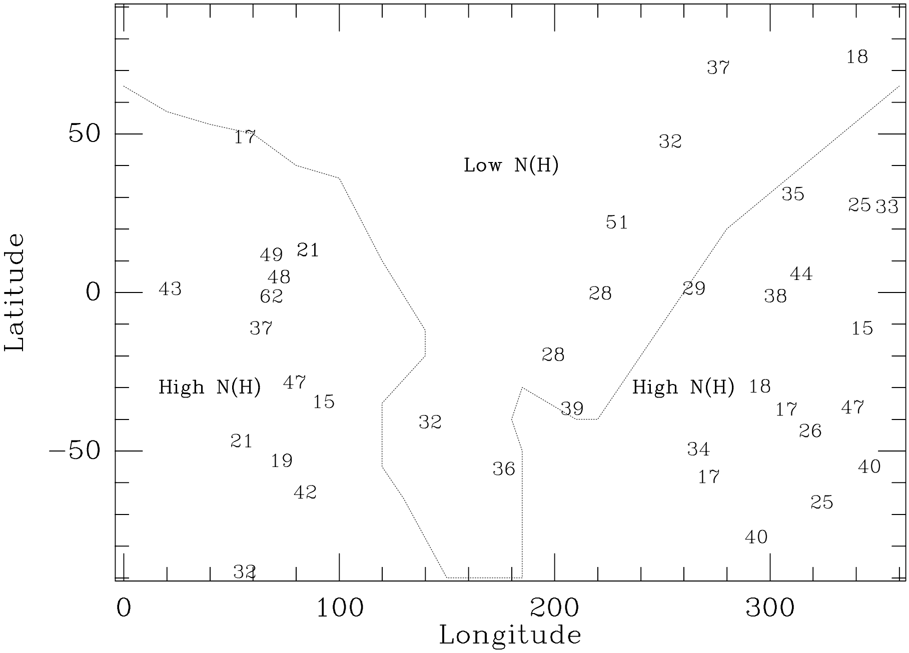

FIGURE 8. locations of ∼40 extrasolar planetary systems in galactic longitude and latitude. The plotted numbers are the star distance. The regions marked “High N(H)” show the upstream direction of the cluster of local interstellar clouds, towards which stars within ∼30 pc are likely to be embedded in a diffuse interstellar cloud. The direction towards “Low N(H)” shows the direction towards the interior of the Local Bubble or the north pole of Gould’s Belt, towards which stars beyond ∼5 pc are more likely to be embedded in the hot gas of the Local Bubble or high-latitude very low density ISM. The N(H) regions are based on Genova et al. (1990).

distinct populations which trace both ISM kinematics and the underlying donor plasma population. Comparisons between predictions of global astrospheric models and Lyα absorption lines towards nearby cool stars demonstrate that external cool stars have astrospheres with detectable hydrogen walls.

The modulation of GCRs and ACRs in the heliosphere indicates that the cosmic ray fluxes in an astrosphere will depend on the characteristics of the stellar wind interaction with the surrounding interstellar cloud. Stellar activity cycles give information on the mass loss from external cool stars. Activity cycles are observed towards many G-stars, although true solar analogues are not obvious (e.g. Baliunas & Soon 1995, Henry et al. 2000).

The galactic positions and distances of ∼40 nearby planetary systems are shown in Fig. 8. The same figure illustrates the asymmetric distribution of interstellar matter within ∼35 pc of the Sun, with most of the material located in the upstream direction towards the galactic center (labeled “High N(H)”) and very little ISM in the downstream direction (towards the interior of the Local Bubble, “Low N(H)”) or near the North Pole (“Low N(H)”). Stars beyond ∼5 pc towards low-N(H) directions are likely to be embedded in the Local Bubble, while stars within ∼40 pc in the high-N(H) directions are likely to be in diffuse clouds (which may have densities of up to several particles cm^{-3}). By analogy with the Sun, the galactic environments of extrasolar planetary systems will change with time.

6. Connections between astrospheres and planetary climates

Building on the knowledge that the Sun is receding from the constellation of Orion, an area of active star formation and giant molecular clouds, Shapley (1921) speculated that the ice ages on Earth resulted from a solar encounter with the molecular clouds in Orion. Since this earliest speculation, there have been a number of attempts to link cosmic phenomena and the terrestrial climate. The investigated phenomena include (but are not limited to) studies of encounters with molecular clouds that may be in spiral arms (Thaddeus 1986, Scoville & Sanders 1986, Innanen et al. 1978, Begelman & Rees 1976, McCrea 1975, Talbot & Newman 1977), changes in atmosphere chemistry due either to energetic particles from supernova or the accretion of ISM (Brakenridge 1981, McKay & Thomas 1978, Butler et al. 1978, Fahr 1968), nearby supernova (Sonett et al. 1987, Sonett 1997), or variations in the global electrical circuit or tropospheric cloud cover from cosmic ray flux variations in the atmosphere (Roble 1991, Rycroft et al. 2000, Tinsley 2000, Marsh & Svensmark 2000).

Marsh and Svensmark (2000) presented plausible evidence that a correlation is present between cosmic ray fluxes and low altitude (<3.2 km) cloud cover, which they attribute to cloud condensation around ionized aerosol particles. They also argue that low optically thick clouds cool the climate. The correlation was observed for low altitude clouds over the 1980–1995 interval, and the correlation is dominated by a cosmic ray flux minimum corresponding to the ∼1991 solar maximum, using Huancayo neutron counts (cutoff rigidity 13 GeV) as the cosmic ray monitor. This correlation, apparently related to water nucleation on ionized aerosols, provides a possible mechanism for an astrosphere-climate connection which can be quantitatively evaluated.

The evolution of advanced life has occurred while the Sun was immersed in the vacuum of the Local Bubble, and the anomalous cosmic ray population inside the heliosphere would have nearly vanished and the enlarged heliosphere would have yielded an effective cosmic ray modulation (Mueller et al. 2002). Such a galactic environment may have promoted stability in the terrestrial climate.

7. Conclusions

The evolution of advanced life has occurred during a time when the Sun was immersed in the vacuum of the Local Bubble, so that the enlarged heliosphere would have yielded effective modulation of galactic cosmic rays. In contrast, an encounter with a modest density diffuse cloud (n(HI) ∼10 cm^{-3}) is possible within 10^4–10^5 years, and would destabilize the heliosphere and modify cosmic ray fluxes impinging on the Earth. The modulation of both galactic and anomalous cosmic rays by solar wind magnetic fields, and the emerging link between cosmic ray fluxes and climate forcing, suggests that a stable heliosphere, and by analogy stable astrospheres, are significant factors in maintaining climatic stability as is necessary for sustainable civilization.

The Galactic environment of a star determines interplanetary medium properties, including the distribution of cosmic rays in the astrosphere. How does this affect the "Astrophysics of Life," which is the topic of this conference? Over the past century many suggestions have been made regarding Galactic effects on Earth's climate. Recent work has demonstrated that the global electrical circuit is moderated by the cosmic ray flux (Roble 1991), and that, for instance, cloud cover in the lower troposphere (<3.2 km) correlates with cosmic ray flux (Marsh & Svensmark 2002). The fact which is clear, however, is that at the present time the solar wind shields the Earth from most ISM products. Relatively low fluxes of energetic particles, including galactic cosmic rays (>1 GeV/nucleon) and anomalous cosmic rays (<0.5 GeV/nucleon), are able to penetrate to the Earth however.

Simulations which describe the interaction between interstellar clouds and stellar winds will provide valuable information on the properties of the astrospheres of extrasolar planetary systems, as well as a basis for evaluating the interplanetary environment. Understanding the historical properties of astrospheres around extrasolar planetary systems will provide a basis for evaluating the climatic stability on possible Earth-like extrasolar planets. The differences in exposure to raw ISM for inner and outer planets over the planet lifetimes may be significant.

PCF would like to thank NASA for research support through grants NAG5-8163, NAG5-1105, and NAG5-6405 to the University of Chicago. GPZ and HRM acknowledge the partial support of an NSF-DOE grant ATM-0296114 and NASA grant NAG5-11621. CL thanks and acknowledges NSF grant ATM 99-12341 for providing support for the cosmic ray research.

REFERENCES

Adams, J. H. & Tylka, A. J. 1993, in *Back to the Galaxy* (eds. S. S. Hold & F. Verter). AIP Conference Proceedings, Vol. 278, p. 186. American Institute of Physics.

Adams, T. F. & Frisch, P. C. 1977 *ApJ* **212**, 300.

Axford, W. I. 1972, in *Solar Wind* (eds. C. P. Sonnet, P. J. Coleman Jr., & J. M. Wilcox), p. 609. NASA Spec. Publ., SP-308.

Baggaley, W. J. 2000 *J. Geophys. Res.* **105**, 10353.

Baguhl, M., Grun, E., & Landgraf, M. 1996 *Space Science Reviews* **78**, 165.

Baliunas, S. & Soon, W. 1995 *ApJ* **450**, 896.

Begelman, M. C. & Rees, M. J. 1976 *Nature* **261**, 298.

Bertaux, J. L. & Blamont, J. E. 1971 *A&A* **11**, 200.

Brakenridge, G. R. 1981 *Icarus* **46**, 81.

Butler, D. M., Newman, M. J., & Talbot, R. J. 1978 *Science* **201**, 522.

Cummings, A. C., Stone, E. C., & Steenberg, C. D. 2002 *ApJ* **578**, 194.

Dehnen, W. & Binney, J. J. 1998 *MNRAS* **298**, 387.

Fahr, H. J. 1968 *Ap&SS* **2**, 474.

Fisk, L. A., Kozlovsky, B., & Ramaty, R. 1974 *ApJ* **190**, L35.

Florinski, V., Zank, G. P., & Pogorelov, N. V. 2002, *American Geophysical Union, Fall Meeting 2002, abstract #SH71A-01.*

Frisch, P. C. 1990, in *Physics of the Outer Heliosphere.* p. 19. Pergamon Press.

Frisch, P. C. 1993 *ApJ* **407**, 198.

Frisch, P. C. 1994 *Science* **265**, 1423.

Frisch, P. C. 1995 *Space Sci. Rev.* **72**, 499.

Frisch, P. C., Dorschner, J. M., Geiss, J., Greenberg, J. M., Grün, E., Landgraf, M., Hoppe, P., Jones, A. P., Krätschmer, W., Linde, T. J., Morfill, G. E., Reach, W., Slavin, J. D., Svestka, J., Witt, A. N., & Zank, G. P. 1999 *ApJ* **525**, 492.

Frisch, P. C., Grodnicki, L., & Welty, D. E. 2002 *ApJ* **574**, 834.

Frisch, P. C. & Slavin, J. D. 2002 *ApJ*, **565**, 364.

Frisch, P. & York, D. G. 1986, in *The Galaxy and the Solar System.* p. 83. University of Arizona Press.

Garcia-Munoz, M., Mason, G. M., & Simpson, J. A. 1973 *ApJ* **182**, L81.

Gayley, K. G., Zank, G. P., Pauls, H. L., Frisch, P. C., & Welty, D. E. 1997 *ApJ* **487**, 259.

Genova, R., Beckman, J. E., Molaro, P., & Vladilo, G. 1990 *ApJ* **355**, 150.

Gloeckler, G. & Geiss, G. 2002, in *American Geophysical Union, Spring Meeting 2002, abstract #SH31B-01*, p. B1.

Gruen, E. & Landgraf, M. 2000 *J. Geophys. Res.* **105**, 10291.

Henry, G. W., Baliunas, S. L., Donahue, R. A., Fekel, F. C., & Soon, W. 2000 *ApJ* **531**, 415.

Holzer, T. E. 1989 *ARA&A* **27**, 199.

Innanen, K. A., Patrick, A. T., & Duley, W. W. 1978 *Ap&SS* **57**, 511.

Ip, W.-H. & Axford, W. I. 1985 *A&A* **149**, 7.

Landgraf, M. 2000 *J. Geophys. Res.* **105**, 10303.

Landgraf, M., Baggaley, W. J., Grün, E., Krüger, H., & Linkert, G. 2000 *J. Geophys. Res.* **105**, 10343.

Liewer, P. C., Mewaldt, R. A., Ayon, J. A., & Wallace, R. A. 2000, in *The Proceedings of the Space Technology and Applications International Forum* (ed. M. S. El-Genk). AIP Conference Proceedings, Vol. 504, p. 911. American Institute of Physics.

Linde, T. J., Gombosi, T. I., Roe, P. L., Powell, K. G., & DeZeeuw, D. L. 1998 *J. Geophys. Res.* **103** (A2), 1889.

Linsky, J. L. & Wood, B. E. 1996 *ApJ* **463**, 254.

Marsh, N. D. & Svensmark, H. 2000 *Physical Review Letters* **85**, 5004.

Mazur, J. E., Mason, G. M., Blake, J. B., Klecker, B., Leske, R. A., Looper, M. D., & Mewaldt, R. A. 2000 *J. Geophys. Res.* **105**, 21015.

McComas, D. J., Barraclough, B. L., Funsten, H. O., Gosling, J. T., Santiago-Muñoz, E., Skoug, R. M., Goldstein, B. E., Neugebauer, M., Riley, P., & Balogh, A. 2000 *J. Geophys. Res.* **105**, 10419.

McCrea, W. H. 1975 *Nature* **255**, 607.

McDonald, F. B., Teegarden, B. J., Trainor, J. H., & Webber, W. R. 1974 *ApJ* **187**, L105.

McKay, C. P. & Thomas, G. E. 1978 *Geophys. Res. Lett.* **5**, 215.

Meisel, D. D., Janches, D., & Mathews, J. D. 2002 *ApJ* **567**, 323.

Michels, J. G., Raymond, J. C., Bertaux, J. L., Quémerais, E., Lallement, R., Ko, Y.-K., Spadaro, D., Gardner, L. D., Giordano, S., O'Neal, R., Fineschi, S., Kohl, J. L., Benna, C., Ciaravella, A., Romoli, M., & Judge, D. 2002 *ApJ* **568**, 385.

Mueller, H. R., Zank, G. P., & Frisch, P. C. 2001, in *The Outer Heliosphere: The Next Frontiers* (eds. K. Scherer, H. Fichtner, H. J. Fahr, & E. Marsch). COSPAR Colloquia Series 11, p. 329. Pergamon Press.

Roble, R. G. 1991 *Journal of Atmospheric and Terrestrial Physics* **53**, 831.

Rycroft, M. J., Israelsson, S., & Price, C. 2000 *Journal of Atmospheric and Terrestrial Physics* **62**, 1563.

Scoville, N. Z. & Sanders, D. B. 1986, in *The Galaxy and the Solar System* (eds. R. Smoluchowski, J. M. Bahcall, & M. S. Matthews). p. 69. University of Arizona Press.

Shapley, H. 1921 *J. Geology* **29**.

Slavin, J. D. & Frisch, P. C. 2002 *ApJ* **565**, 364.

Sonett, C. P., McHargue, L., & Damon, P. E. 1997, in *25th Intl. Cosmic Ray Conf., Durban*, preprint.

Sonett, C. P., Morfill, G. E., & Jokipii, J. R. 1987 *Nature* **330**, 458.

Talbot, R. J. & Newman, M. J. 1977 *ApJS* **34**, 295.

Thaddeus, P. 1986 *The Galaxy and the Solar System* (eds. R. Smoluchowski, J. M. Bahcall, & M. S. Matthews). p. 61. University of Arizona Press.

Thomas, G. E. & Krassa, R. F. 1971 OGO *A&A* **11**, 218.

Tinsley, B. A. 2000 *Space Science Reviews* **94**, 231.

Weller, C. S. & Meier, R. R. 1974 *ApJ* **193**, 471.

Witte, M., Banaszkiewicz, M., & Rosenbauer, H. 1996 *Space Sci. Rev.* **78**, 289.

Wood, B., Linsky, J., Müller, H., & Zank, G. 2001 *ApJ* **547**, L49.

Wood, B. E., Müller, H., & Zank, G. P. 2000 *ApJ* **542**, 493.

Wood, B., Müller, H., Zank, G. & Linsky, J. 2002 *ApJ* **574**, 412.

Zank, G. P. 1999 *Space Science Reviews* **89**, 413.

Zank, G. P. & Frisch, P. C. 1999 *ApJ* **518**, 965.

Zank, G. P., Pauls, H. L., Williams, L. L., & Hall, D. T. 1996 *J. Geophys. Res.* **101** (A10), 21639.

Transits

By RONALD L. GILLILAND

Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, MD 21218, USA

Transits of the planets Mercury and especially Venus have been exciting events in the development of astronomy over the past few hundred years. Just two years ago the first transiting extra-solar planet, HD 209458b, was discovered, and subsequent studies during transit have contributed fundamental new knowledge. From the photometric light curve during transit one obtains a basic confirmation that the radial velocity detected object is indeed a planet by allowing precise determination of its mass and radius relative to these stellar quantities. From study of spectroscopic changes during transit it has been possible to probe for individual components of the transiting planets atmosphere. Planet transits are likely to become a primary tool for detection of new planets, especially other Earth-like planets with the *Kepler Discovery Mission.* Looking ahead, the additional aperture of the *James Webb Space Telescope* promises to allow the first possibility of studying the atmosphere of extra-solar Earth-analogue planets, perhaps even providing the first evidence of direct relevance to the search for signs of life on other planets.

1. Transits in history

Transits happen when an obscuring body passes in between us, the observers, and a background luminous source. Historically, both of the planets interior to Earth in the solar system have been observed while transiting the Sun. Mercury transits the Sun from our perspective frequently, Venus transits the Sun from the vantage point of the moving Earth only twice in every 130 years given current orbits.

Johannes Kepler, the great theoretical astronomer of the 17th century best known for enumerating laws of planetary motion, was the first to predict that transits of the inner planets should occur. His predictions made in the 1620s for a Mercury transit of 7 November 1631 were confirmed with observations, while a coincidentally close in time transit of Venus in December 1631 was not successfully observed. The next Venus transit of 4 December 1639 was not in fact predicted by Kepler, but was predicted by Jeremiah Horrocks, and observed by himself and local English colleague William Crabtree. Observations of Venus transits were off to a modest start having been witnessed by only two astronomers. The next chance for observing a transit of Venus would be 122 years later, on 6 June 1761, by which time its observation would become a major event worthy of arduous sailing expeditions to the South Seas. A primary motivating factor for observing the Venus transit from different locales on the Earth was to detect the parallax effect provided by the Earth's radius and use this to establish the fundamental scale of the solar system. This was successfully done, and the transit of Venus might arguably be taken as the first primary rung of an astronomical distance scale, the extended definition of which has remained a central astronomical endeavor to this day, although now it's the metric of the Universe at large, rather than the solar system at study. For additional historical perspective see the book, *June 8, 2004: Venus in Transit* by Maor (2000), which was used as the primary reference here.

More relevant to a meeting on the astronomy of astrobiology, continuing this brief historical perspective, is that the 1761 transit of Venus resulted in the first detection of an atmosphere associated with another planet. During the transit of 1761, Mikhail Lomonosov noted that just before the time of planet ingress with the solar disk and just after egress, there was a teardrop shaped pattern of light encompassing the planet resulting from refraction of light through an atmosphere.

For the next fundamental contribution from transits the wait was even longer than the 122 years between Venus transits; 360 years were to pass after 1639 before an additional member of the family of transiting planets was detected.

2. HD 209458b

Only a few years ago, 1995 to be precise, the first unambiguous discovery of an extrasolar planet based on the radial velocity technique was made by Mayor and Queloz (1995). The great surprise associated with this detection was that the planet inferred by a slight periodic wobble as evidenced by the Doppler effect in radial velocities, was very, very close to its host star, much closer even than Mercury to the Sun. Since the mass of Mayor and Queloz's first planet, 51 Peg b, was inferred to be about that of Jupiter, and several additional detections about other stars followed in short order, the members of this class became known as 'hot Jupiters,' planets with periods of 3–5 days at orbital separations only a few hundredths of an AU from the parent star. It was recognized very early that with a planet so close to the star it orbits, simple geometry yields a significant chance, given random orbital inclinations on the sky, that transits would occur. While for the Earth-Sun scale of separation for extrasolar planets there is only a 0.5% chance that random inclinations would yield a transit for any system, for the hot Jupiters this probability ballooned to fully 10%. The predicted magnitude of a transit in time-series photometry is simply the ratio of the area of the planet to the area of the star it obscures during transit, with Jupiter having a radius of about 10% that of the Sun, one would expect a hot Jupiter transit to show transits (or eclipses) about 1% deep. Detection of extrasolar planets opened the floodgates of theoretical studies of how they might form and evolve, and in particular led to an early prediction by Guillot et al. (1996) that hot Jupiters would have somewhat bloated radii, and thus would be even easier to detect.

With each newly detected hot Jupiter from the radial velocity surveys, observers would obtain photometric observations near the time of predicted inferior conjunction to search for the characteristic dip of 1–2% of light for 2–3 hours expected for transits of hot Jupiters. With each newly detected hot Jupiter the aggregate probability that one in the family of such planets should be transiting its host star grew by 10%. At just the right time for a 50/50 chance of expecting a transit to exist, the newly detected planet via the radial velocity technique, HD 209458 b, was shown (Charbonneau et al. 2000; Henry et al. 2000) to have transits about 1.6% deep at the expected phase for transits given the radial velocity ephemeris.

Observation of a transit with accurate photometry rather immediately yields a fundamental measure associated with the planet that is of great interest, namely the ratio of the planet area to the star area. And since in general, thanks to great successes in stellar structure and evolution theory and classical stellar astronomy over the past several decades, we can estimate stellar radii to order 10% with input of colors, brightness and distance, we could in this case expect comparably good knowledge for the first time of an extrasolar planet radius. Since the radial velocity technique provides a measure of the planet mass uncertain by the sine of the orbital inclination, the mere existence of a transit immediately implies $\sin(i) = 1$ is closely satisfied, thus pinning down the mass. With mass and radius we have density, and thus the chance to provide serious grist for the theoretical mill. In the case of HD 209458 b, even the rough photometry available from small ground based telescopes in less than ideal conditions was sufficient to prove that the planet's volume was bloated as had been predicted, due to its proximity to the star. This bloating results not from the high stellar energy input to the planet increasing its atmospheric temperature and swelling the atmosphere, but rather follows from

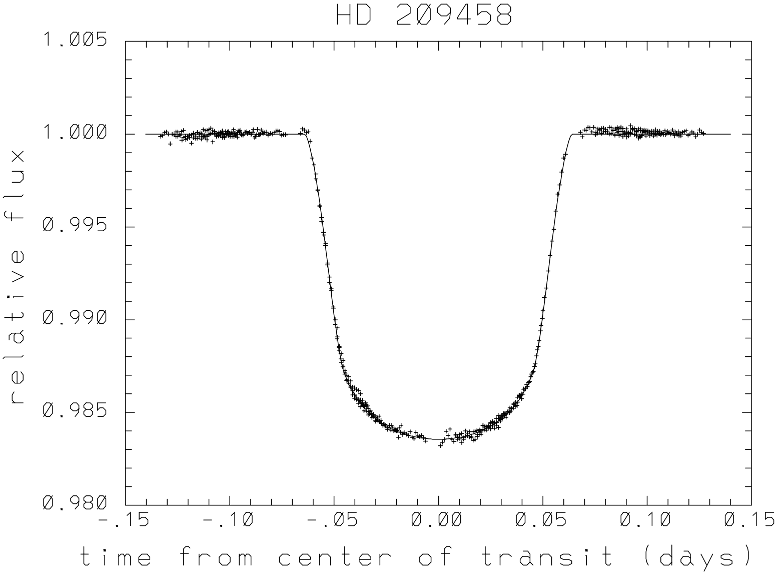

FIGURE 1. Photometric time-series for HD 209458 as a composite of *HST* observations made with STIS during April and May 2000.

retardation due to the high energy input, of the rate at which the planet contracts after formation.

Transits of Venus were exciting enough in the 18th century for international expeditions to be executed for their observation. At the turn of the millennium, recognition of an extrasolar planet with conveniently frequent transits occurring every 3.5 days inspired astronomers to bring the best resources of the day to bear in obtaining yet more precise observations. At this time the availability of *HST* was recognized as providing an opportunity with one set of observations to refine knowledge of the planet size, search for the presence of rings and moons, and possibly to obtain direct evidence of the existence of an atmosphere on the planet. Figure 1 shows the light curve of HD 209458 b in transit as made with a composite of three separate transits as observed in April and May 2000 with *HST* (Brown et al. 2001). These observations were successful in refining the error on the planet radius by a factor of three, thus providing a better challenge for theoretical modeling efforts. In addition, through a detailed study of the light curve shape near ingress and egress, it was inferred that no moons existed with an upper limit of about 1.2 $R_\oplus$, and that no rings existed with an extent greater than 1.8 planetary radii. Through precise timings of the transits, arguments were also provided, setting an upper limit on possible moons of 3 $M_\oplus$. The superior observing conditions provided by an orbiting observatory had already allowed a search for companions to the planet down to a scale comparable to the Earth, with very modest investments of observing time.

Observed with a spectrometer during transit, it is possible to search for evidence of an atmosphere about the planet. A general characteristic of an atmosphere (assuming that a substantial height not dominated by a high cloud deck exists) is that some wavelengths of light will pass more freely through the atmosphere than others. In particular, if viewed from a great distance during transit, the planet would appear larger, i.e. block relatively more light during transit in the core of a strong absorption line associated with its atmosphere, relative to wavelengths where light can pass freely through the atmosphere. Early modeling (Brown 2001) had predicted that for hot Jupiters the planet's radius would appear significantly larger in the core of the Na D resonance doublet—in this case significant implies order 1%. Coupled with the planet area of 2% relative to the star, one might expect spectra acquired during transit to be some 2×10^{-4} times fainter than stellar spectra taken out of transit (after normalizing out the overall 2% change independent of wavelength).

By co-adding a large number of individually high signal-to-noise spectra acquired both in and out of transit, Charbonneau et al. 2002 found evidence for Na in the atmosphere of HD 209458 b using the Space Telescope Imaging Spectrograph on *HST*. In this case, *HST* was used for a purpose it was not designed for—to search for a very small signal associated with a relatively very bright nearby star that happens to have a transiting hot Jupiter. Further STIS observations of HD 209458 b will be obtained to search for the planet during the secondary eclipse (if a yet smaller signal of a few $\times 10^{-5}$ can be found, this would establish the amount of light from the star reflected backwards and hence constrain the albedo and energy input level). Observations will also be conducted to search for other atmospheric constituents during transit; if observations with the *HST* infra-red instrument, NICMOS, prove to be sensitive enough it may be possible to demonstrate the existence of water in the atmosphere. Within two years of discovery we have taken the first modest, but surprisingly successful, steps in characterizing an extrasolar planet's atmosphere.

3. Transits as extrasolar planet discovery technique

All of the 80-odd extrasolar planets detected to date around ordinary stars have been found using the radial velocity technique. In this approach, one observes a single star at a time using a large telescope and a state-of-the-art spectrograph. The observations must then be followed by a very exacting analysis in which variations of the star's speed are measured at the level of only a few meters per second. With dedication of significant astronomical resources, the 2–3 thousand best suited stars are being monitored, and many more extrasolar planet detections are yet to follow from this exciting recent development.

With transits of 'hot Jupiters,' one can in principle observe about 10,000 bright, relatively nearby stars simultaneously for signs of transits using only a modest telescope and CCD camera. At a 10% chance per existing system of transiting, expected ~2% signal amplitudes, and repeated dips on a fairly convenient cadence of every 3–5 days, ground-based searches for additional 'hot Jupiters' should become a productive means of discovery.

With *HST*, this same approach can be applied to more distant, fainter stars. In July 1999 I used WFPC2 on *HST* to monitor 34,000 dwarf stars in the globular cluster 47 Tucanae for a nearly continuous period of 8.3 days. This experiment was sufficient to find close-in giant planets of the hot Jupiter variety if they existed in this cluster. If as common in 47 Tuc as locally, we calculated after the fact based on our realized data properties that we should have seen 17 planets (Gilliland et al. 2000). We saw none. There are unfortunately a number of reasons why planets might not exist in 47 Tuc: (1) It has a metallicity a few times lower than the Sun, perhaps formation of planets requires a certain threshold of metals; this would be consistent with observations showing a correlation of metallicity with detected planets via the radial velocity technique. (2) Perhaps the crowded stellar environs of a globular cluster prevents planets from forming, or if formed, allows for their destruction. The null experiment has raised questions which can only be answered with more extensive observations.

The truly exciting prospect offered by transits as a detection technique is their proposed utility in finding analogues of the Earth. In the last round of Discovery Mission competition, NASA awarded the tenth mission to: "Kepler: A Search for Terrestrial Planets," which is on track for a 2007 launch on a bold quest to find other Earths in the extended solar neighborhood. *Kepler* will be built by Ball Aerospace and Technologies Corporation with William Borucki of NASA Ames Research Center as the Principal Investigator.

Kepler is a 0.95 meter Schmidt telescope with a focal plane of 42 CCDs containing a total of 10^8 pixels covering a field of 100 square degrees. A photometric precision of better than 100 parts per million is expected for 12th magnitude stars (middle of 9–15 magnitude band covered) that will be observed once every 15 minutes. The near galactic equator field to be observed will contain 100,000 dwarf stars that will be viewed continuously for four years, building up remarkably precise and extensive sets of photometric time series. The signal from an Earth-analogue transit is 84 parts per million deep, and obviously would repeat once per year. The chance of transits per existing Earth-analogue system is 0.5%.

Although, as described, the technical challenge for *Kepler* is daunting, we expect discoveries to follow, if solar systems like our own are the rule, rather than the exception:

- $\sim$50 planets if the same radius as Earth in 1 AU orbits.
- $\sim$200 planets if radii are 33% larger than Earth, 1 AU.
- 1000s of terrestrial planets if orbits much smaller than 1 AU are common.
- Several hundred 'hot Jupiters,' most detected via a sinusoidal reflected light signal, several tens of which will be aligned tightly enough in inclination to yield transits.

The *Kepler* mission will provide us with knowledge of how common planets similar to our own are in the local galactic neighborhood. This would, in turn, be one of the more important steps in defining the frequency of sites we believe to be hospitable to life.

4. Transits and the search for life

We have seen that a transit observation of Venus in the 18th century provided the first detection of a planet atmosphere beyond Earth. A little over 240 years later another transit observation, this time of a planet some 150 light years away from us, revealed the first detection of an extrasolar planet atmosphere. In coming years, we expect that we will take advantage of transits to learn more about the atmosphere of HD 209458 b and other gas giant transiting planets surely to be discovered, and that transits will allow us to quantify the frequency of Earth-like planets. These developments will almost surely happen; utilization of transits for discovery will, we hope, become mundane, although we're not yet in that position. And transits will surely allow further atmospheric probes of 'hot Jupiters' to be made.

Being provocative, let us consider what further advances might be in store from the study of transits, as we have sometimes done with *HST*, utilizing an instrument for a task it was not designed for. The holy grail of the astronomical component of astrobiology is the detection of extrasolar life. Transits may well have a role to play in this. If Earth analogues are common, then the nearest transiting terrestrial planet is likely with a star of 6–7th magnitude, i.e. very bright by the standards most astronomers in this building are used to thinking about. Assuming that we can somehow find this nearest and brightest star with a transiting terrestrial planet, what might we do then? Relative to the HD 209458 b planet observed with *HST* that had a radius about 10% of its host star, an Earth analogue would only have a radius some 1% that of the star for an overall photometric diminution of one part in ten thousand during the 10-hour transit. An Earth-like atmosphere would provide only a thin veneer above the solid/liquid body of the planet itself, and it's only this thin veneer which would provide the opportunity of probing its atmosphere during transit—the expected strength of atmospheric features would be a substantial part of one-part-per-million deep comparing spectra taken in transit to those out of transit. Can we imagine detecting relative signals at the level of one part per million? Well sure, it's not all than outrageous to think such might just be possible.

The *James Webb Space Telescope* is an IR-optimized telescope planned for a 2011 launch. If implemented with the capability of recording nearly all available photons from this nearest star with a transiting Earth-like planet, then with 30 hours of observation through a 10-hour transit it seems possible to contemplate distinguishing at about the 3 σ level between a Venus-like and an Earth-like atmosphere by cross-correlating the in-transit versus out-of-transit spectrum with model predictions for the rather different Earth (dominated by Oxygen and water) and Venus (dominated by Carbon Dioxide) atmospheres in transmission. Is this a long shot? Well yes, certainly. We would need things to break favorably with respect to existence and detection of a favorable target, and we would need *JWST* and its NIRSPEC instrument to be able to maintain high efficiency on what is, for it, an outrageously bright star. But none of these things are harder to fathom than would have been the case in the 1970s if someone had proposed that *HST*, in its early stages of planning, would one day detect the atmosphere of an extrasolar planet. And we still have time to influence the capabilities realized for NIRSPEC so that it can deal with the high photon flux, allowing collection of very high signal-to-noise spectra that can be used to probe for atmospheric constituents around a faint blue planet transiting a solar cousin. I would not be surprised for the unique circumstances afforded by transits to provide us with the first opportunities for observations of direct relevance to the search for (signs of) life on a distant planet.

I would like to thank in particular Bill Borucki for his long advocacy of the transit technique in detecting Earth-like planets, that has now resulted in the approved *Kepler* mission, and Tim Brown for the planetary atmosphere work and many discussions related to the speculation on the possible role of *JWST* in probing planet atmospheres via transits. Observations with the NASA/ESA *Hubble Space Telescope* were obtained at St Sci, which is operated by AURA Inc., under NASA contract NAS 5-26555.

REFERENCES

Brown, T. M., Charbonneau, D., Gilliland, R. L., Noyes, R. W., & Burrows, A. 2001 *ApJ* **552**, 699.

Brown, T. M. 2001 *ApJ* **553**, 1006.

Charbonneau, D., Brown, T. M., Latham, D. W., & Mayor, M. 2000 *ApJ* **529**, L45.

Charbonneau, D., Brown, T. M., Noyes, R. W., & Gilliland, R. L. 2002 *ApJ* **568**, 377.

Gilliland, R. L., et al. 2000 *ApJ* **545**, L47.

Guillot, T., Burrows, A., Hubbard, W. B., Lunine, J. I., & Sauman, D. 1996 *ApJ* **459**, L35.

Henry, G. W., Marcy, G. W., Butler, R. P., & Vogt, S. S. 2000 *ApJ* **529**, L41.

Maor, E. 2000. *June 8, 2004: Venus in Transit.* Princeton University Press.

Mayor, M. & Queloz, D. 1995 *Nature* **378**, 355.

Planet migration

By EDWARD W. THOMMES[1] AND JACK J. LISSAUER[2]

[1]Astronomy Department, University of California, Berkeley, CA 94720, USA

[2]Space Sciences Division, NASA Ames Research Center, Moffett Field, CA 94035, USA

A planetary system may undergo significant radial rearrangement during the early part of its lifetime. Planet migration can come about through interaction with the surrounding planetesimal disk and the gas disk—while the latter is still present—as well as through planet-planet interactions. We review the major proposed migration mechanisms in the context of the planet formation process, in our Solar System as well as in others.

1. Introduction

The word planet is derived from the Greek word "planetes," meaning wandering star. Geocentric views of the Universe held sway until the Middle Ages, when Copernicus and Kepler developed a better phenomenological explanation of planetary wanderings, which with small modifications has withstood the test of time. Kepler's first law of planetary motion states that planets travel along elliptical paths with one focus at the Sun. Thus, although planets wander about the sky, in this model their orbits remain fixed and they do not migrate. In his physical model of the Solar System, Newton theorized that planets gradually altered one another's orbits, and he felt compelled to hypothesize occasional divine intervention to keep planetary trajectories well-behaved over long periods of time. In the early 1800s, Poisson pointed out that planetary-type perturbations cannot produce secular changes in orbital elements to second order in the mass ratio of the planets to the Sun, but Poincare's work towards the end of the 19th century suggests that the Solar System may be chaotic. The stability of mature planetary systems is a fascinating topic, but we shall be concerned herein with the potentially much more rapid migration of planets during and immediately following the epoch of their formation. Modern research into this topic began when Goldreich and Tremaine (1980) showed that density wave torques could have led to significant orbital evolution of Jupiter within the protoplanetary disk on a timescale of a few thousand years, and research accelerated when giant planets were found much closer to their stars (Mayor & Queloz 1995) than predicted by models of their formation (Lin et al. 1996, Bodenheimer et al. 2000).

In Section 2, we discuss models for the migration of giant planets within our own Solar System which may have occurred as the results of interactions of the planets with one another and with small solid bodies. Section 3 summarizes models of the potentially substantial planetary migration that results from (primarily gravitational) interactions between a planet and a gaseous protoplanetary disk. The predictions of this model are compared to observations of Saturn's rings and moons in Section 4. In Section 5, the models are applied to extrasolar planetary systems.

2. Migration of Jupiter, Saturn, Uranus, Neptune and Pluto

In studying the accretion of the giant planets, Fernandez and Ip (1984) first noted that proto-Uranus and -Neptune could undergo significant orbital migration due to angular momentum exchange with a (sufficiently massive) planetesimal disk. The mechanism operates as follows: Gravitational stirring by Uranus and Neptune imparts high eccentricities on the surrounding planetesimals. Those which acquire sufficiently small perihelia can be "handed off" to the next-innermost planet, with a resultant gain in angular momentum for the first planet. In this way, planetesimals get passed inward from Neptune to Uranus to Saturn and finally to Jupiter, which is massive enough to readily eject them from the Solar System. (The other giant planets are also massive enough to eject planetesimals from the Solar System; however, the characteristic timescales for direct ejection are longer than for passing the planetesimals inwards to the control of Jupiter.) It is generally accepted that planetesimal scattering by the giant planets, principally Jupiter, is what formed the Oort cloud, the quasi-spherical distribution of comets orbiting at $\sim 10^3$–10^5 AU from the Sun (Duncan et al. 1987). Since Jupiter does the lion's share of the work, it undergoes a net loss of angular momentum and its orbit shrinks; Saturn, Uranus and Neptune mainly contribute by passing planetesimals down, so they gain angular momentum. For a planet on a circular orbit, change in angular momentum L is related to change in semimajor axis a according to

$$\Delta L = \frac{1}{2} M \sqrt{\frac{G\mathrm{M}_\odot}{a}} \Delta a \quad , \tag{2.1}$$

where M is the planet's mass, so Jupiter, being the innermost and most massive, migrates the shortest distance, while Neptune migrates farthest. Hahn and Malhotra (1999) investigated this migration mechanism in detail, using it to try and reproduce the eccentricities of Pluto and the other Kuiper belt objects sharing the Neptune exterior 3:2 resonance (Plutinos). They found that with a planetesimal disk of about 50 Earth masses ($\mathrm{M}_\oplus$), Neptune migrates the right distance to pump the eccentricities of objects carried along in its 3:2 resonance to the observed values, $e \sim 0.3$. Larger/smaller disk masses produced too much/not enough migration. Migration takes place over a timescale of tens of millions of years, after which time it stalls due to depletion of the planetesimals.

A preceeding phase of migration for Neptune, Uranus and Saturn, shorter but more violent, was proposed in the model of Thommes et al. (1999, 2002). In trying to account for the formation of the "ice giants," Uranus and Neptune (14.6 and 17.2 $\mathrm{M}_\oplus$ respectively, $\sim$90% of which are condensibles) at their present orbital radii (19 and 30 AU respectively), one runs into serious timescale problems (Lissauer et al. 1995, Levison & Stewart 2001, Thommes et al. 2002b). At such large heliocentric distances, the accretion timescale of such massive bodies far exceeds the lifetime of the nebular gas (about 10^7 years or less, e.g. Strom et al. 1990); once planetesimal random velocities are no longer damped by aerodynamic gas drag, it may not be possible to produce an ice giant-sized body on *any* timescale. Indeed, at 20 AU, the mass at which the surface escape velocity from a (Uranus/Neptune-density) body equals the escape velocity from the Sun, is only a bit over an Earth mass. In the model of Thommes et al., the Jupiter-Saturn region serves as the birthplace of *all* the giant planets, thus alleviating the timescale problem. Assuming that gas giant planets form by core accretion (e.g. Pollack et al. 1996), one of the protoplanets—likely proto-Jupiter—eventually reaches runaway gas accretion and acquires a massive gas envelope, abruptly (over $\sim 10^5$ years) expanding its gravitational reach and stabilizing its neighbors' orbits. As a result the other giant protoplanets are scattered, predominantly outwards, ending up with aphelia in the still accretionally unevolved outer planetesimal disk. Dynamical friction with the planetesimals then reduces the eccentricities of these scattered giant protoplanets, decoupling them from Jupiter and from each other on a timescale of a few million years. Numerical simulations show that this sequence of events commonly results in an outer planetary system similar to our own, with the scattered protoplanets eventually settling down to nearly circular orbits of radii comparable to those of Saturn, Uranus and Neptune. An example is shown in Fig. 1. A frequent side effect in the (stochastic) simulations is strong gravitational stirring of the

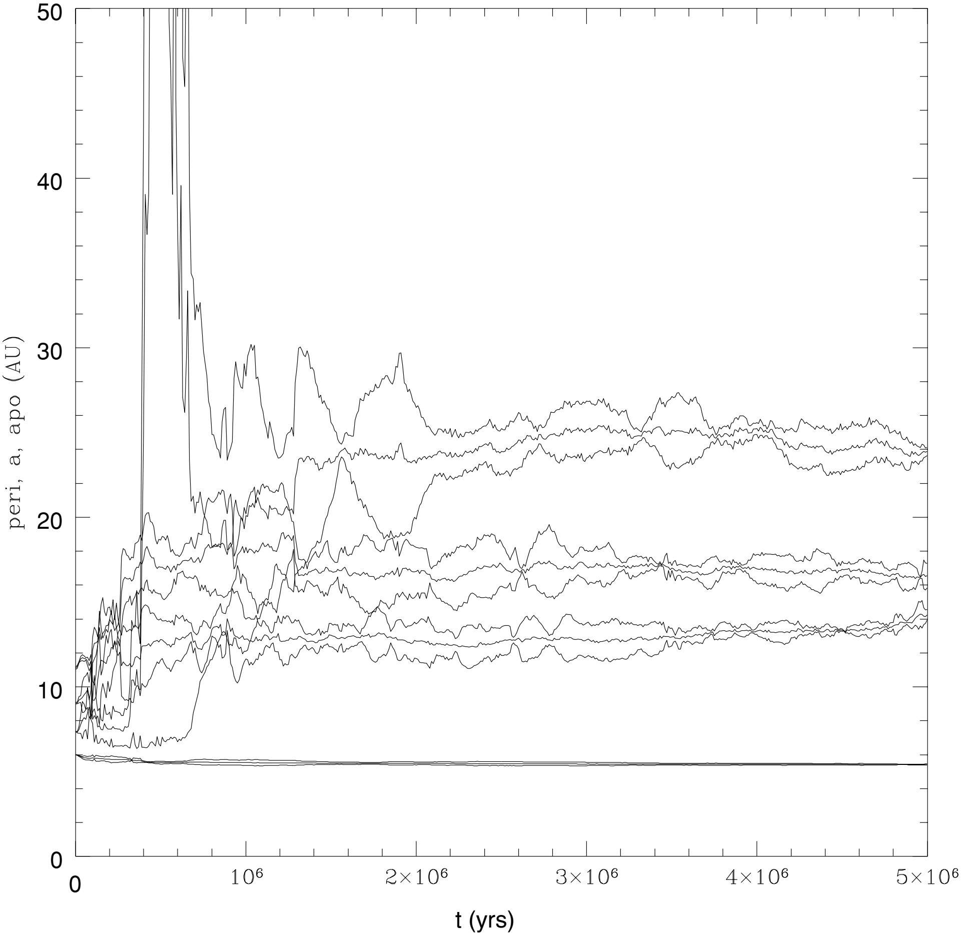

FIGURE 1. A numerical simulation showing how Uranus and Neptune could have originated among Jupiter and Saturn. Semimajor axis, perihelion and aphelion distance as a function of time are plotted for each body. Initially, four 10 $M_\oplus$ bodies are placed on circular orbits in the Jupiter-Saturn region. The innermost one's mass is increased over a 10^5 year timescale to that of Jupiter, to simulate the accretion of a massive gas envelope. As a result, the remaining bodies are scattered outward; their initially high eccentricities are damped by interaction with the planetesimal disk which exists beyond ~12 AU. After 5 Myrs of evolution, the result is a giant planet system which looks qualitatively similar to our own, with the scattered and recircularized bodies standing in for Saturn, Uranus and Neptune. See Thommes et al. (2002) for details.

Kuiper belt region; this may help to account for the high random velocities of objects in the "classical" Kuiper belt (e.g. Petit et al. 1999, Kuchner et al. 2002).

In situ accretion of Uranus and Neptune could still have occurred if the growing ice giants had somehow continued to be supplied with planetesimals of relatively low velocity dispersion. There are several ways in which this could have happened. One possibility is that differential migration between the growing protoplanets and the planetesimals caused the former to sweep, predator-like, through heretofore dynamically cold parts of the planetesimal disk (Ward & Hahn 1995, Tanaka & Ida 1999, Bryden et al. 2000). Accretion would stall when the protoplanet mass becomes sufficiently large that it no longer ploughs through the planetesimals, but instead captures the planetesimals in its mean-motion resonances and pushes them ahead of it (Kary et al. 1993). Insofar as the differential migration is to come about from interaction with the gas disk (aerodynamic gas drag on the planetesimals, gravitational interaction of the protoplanets with the disk (see below), or some combination thereof), this mechanism will only operate for $\sim 10^7$ years. After that, migration will become much slower, since migration-inducing interactions would only involve the much less massive solids component of the protoplanetary disk. Another possibility is that collisions among the planetesimals acted to damp their velocities sufficiently for significant further accretion to take place after dissipation of the gas. Thommes et al. (2002b) make a simple estimate of post-gas growth timescales in the limit of "maximally effective" collisional damping (meaning simply that the random velocity damping timescale is taken to be equal to the inter-planetesimal collisional timescale), and show that planetesimals of order ten kilometers in size or smaller—collision rate is inversely proportional to planetesimal radius—could conceivably permit the accretion of $\sim 10\ M_\oplus$ bodies in the 20–30 AU region in of order 10^8–10^9 years.

3. Disk-Planet interactions

A disk with nonzero viscosity will transport angular momentum outward, resulting in the spreading of the disk; less and less of the disk material ends up with more and more of the angular momentum and as a result, there is a net inward transport of mass (Lynden-Bell & Pringle 1974). In modeling disk viscosity ν, the prescription of Shakura and Sunyaev (1973) is commonly used:

$$\nu = \alpha h^2 \Omega \quad , \tag{3.1}$$

where α is a dimensionless parameter, h is the disk half-thickness, and Ω is the disk angular velocity. Estimates for α in the protosolar nebula range from 10^{-4} to 10^{-2} (e.g. Cabot et al. 1987, Dubrulle 1993). The source of viscosity is unknown; it may be turbulent motion resulting from convective instability in the disk (Lin & Papaloizou 1980), damping of density waves launched by embedded planetary bodies (Larson 1989, Goodman & Rafikov 2001), or magnetohydrodynamic turbulence (Balbus & Hawley 1991). The latter mechanism requires that the disk be sufficiently ionized to be strongly coupled to the magnetic field. At a radius of less than about 0.1 AU from the star, collisional ionization ought to be able to accomplish this; further out, ionization by cosmic rays near the disk surface will dominate. Thus it is possible that beyond $\sim$0.1 AU, a protoplanetary disk only transports angular momentum within relatively thin layers on the outer surfaces of the disk (Gammie 1996).

Individual gas molecules move along paths which are nearly Keplerian ellipses. However, the picture becomes more complex once protoplanets form and perturb the disk. This problem was first investigated in the context of satellite interactions with planetary rings by Goldreich and Tremaine (1980), and extended by Ward (1986, 1997) and Artymowicz (1993) to planetary interactions with a protoplanetary disk. A review of the theory of planet-disk interactions is given by Lin et al. (2000), as well as by Goldreich and Sari (2002), who investigate the resulting evolution of planet eccentricities. We provide an abbreviated summary below.

It is convenient to begin by expressing the gravitational potential of the planet as a Fourier series:

$$\psi = \Sigma_l \Sigma_{m=0}^{\infty} \psi_{l,m}(r) \cos[m(\phi - \Omega_{l,m} t)] \quad , \tag{3.2}$$

where ϕ is the azimuthal angle, $\psi_{l,m}(r)$ is an amplitude which depends on radius, and $\Omega_{l,m} = \Omega_p + (l-m)\kappa_p/m$ is the pattern speed, with Ω_p being the angular frequency of a planet on a circular orbit, and κ_p being the planet's epicyclic (radial oscillation) frequency. Both l and m are integers; for each value of the azimuthal mode number m, there are—to lowest order in eccentricity—three components that contribute to the pattern speed: $\Omega_{l=m} = \Omega_p$, and $\Omega_{m\pm1,m} = \Omega_p \pm \kappa_p/m$. Each component, in turn, has three associated resonances: a corotation resonance, at which the pattern speed matches the disk angular velocity, $\Omega_{l,m} = \Omega_d$, and an inner and outer Lindblad resonance, at which the difference between the disk angular velocity and the pattern speed is a harmonic of the epicyclic frequency, $\Omega_d - \Omega_{l,m} = \pm\kappa_d/m$. For Keplerian orbits, $\kappa_p = \Omega_p$ and $\kappa_d = \Omega_d$. At each of the resonance sites, angular momentum and energy are exchanged between the disk and the planet, and significant epicyclic motion is excited in the gas. We can view the situation as a resonating disk particle always being at the same phase in its radial oscillation when it experiences a particular phase of a Fourier component of the satellite's forcing. This enables continued coherent kicks from the satellite to build up the particle's radial motion, and significant forced oscillations may thus result. In this way, spiral density waves—horizontal density oscillations which result from the bunching of streamlines of gas molecules on eccentric orbits—are launched. If the perturbing planet is inclined relative to the disk, resonances involving the vertical frequencies will also arise, and these will launch spiral bending waves, which are vertical corrugations of the disk plane resulting from the induced coherent inclinations of gas molecule orbits. We do not consider vertical resonances further, since they do not directly affect planet migration; the topic is investigated in detail by, e.g. Lubow and Ogilvie (2001).

4. Saturn's rings—Observational tests of disk-satellite interactions

Several of Saturn's numerous moons orbit near or within the planet's spectacular main ring system. These moons excite spiral density waves at resonant locations within the rings of Saturn. Gaps (or ring edges) are produced at strong resonances and close to moons where resonances overlap. Pan, a small moon within the A ring, has cleared a gap around its orbit, and Pandora and Prometheus, somewhat larger moons orbiting just exterior Saturn's main ring system, confine particles to the narrow, irregular F ring. Spiral density waves and sharp ring boundaries were detected in Saturn's rings by four instruments on the Voyager spacecraft. These data provide for valuable tests of models of interactions between secondaries and astrophysical disks.

4.1. *Wave characteristics and ring properties*

Several dozen density waves in Saturn's rings have thus far been identified with exciting resonances and analyzed to determine the local surface mass density of the rings (e.g. Esposito et al. 1984, Rosen et al. 1991). The observed and predicted resonance locations agree within observational uncertainties, which are generally less than one part in 10^4. The surface density, σ, at most wave locations in the optically thick A and B rings is of order 50 g/cm^2. Measured values in the optically thin C ring are $\sigma \sim 1$ g/cm^2; an intermediate value of $\sigma \sim 10$ g/cm^2 has been estimated for Cassini's Division.

The outer edges of the B and A rings are maintained by the Mimas 2:1 and Janus 7:6 resonances, which are the strongest resonances within the ring system (Smith et al. 1981, Holberg et al. 1982, Porco et al. 1984a, Lissauer & Cuzzi 1982, Borderies et al. 1982). Nearly empty gaps with embedded optically thick ringlets have been observed at strong resonances located in optically thin regions of the rings (Holberg et al. 1982). These features are probably caused by a resonance-related process; however, no explanation for the embedded ringlets currently exists. Nearly empty gaps with embedded ringlets have also been observed at nonresonant locations (Porco et al. 1984b).

Although resonance features in Saturn's rings are well understood in many respects, there remain several major outstanding issues. Some of these problems are related to angular momentum transport, a key factor in planetary migration.

The damping behavior of most observed spiral waves differs significantly from that predicted by the simple fluid approximation in a constant viscosity disk (Goldreich & Tremaine 1978). This has led to severe problems with attempts to estimate the viscosity of the rings using observations of wave damping. Some of the factors which can influence the damping behavior of waves are variations in the background surface density and velocity dispersion, interference with other waves and wave nonlinearities (Lissauer et al. 1984). The damping of nonlinear density waves has been studied in detail by Shu et al. (1985), who find that damping rates can be very sensitive to particle collision properties and to optical depth. Thus, the anomalous damping behavior of many spiral waves presents both a challenge and an opportunity for researchers attempting to deduce ring properties other than surface mass density from the study of nonlinear waves (Longaretti & Borderies 1986). Spiral waves carry (positive or negative) angular momentum, and deposit it in regions of the disk in which they damp, so the poor correspondence between theory and observation is of concern to planet migration modelers.

The cause of the observed enhancement of material in regions where strong waves propagate is poorly understood. Density waves excited at inner Lindblad resonances carry negative angular momentum, i.e. the angular momentum of the ring particles is temporarily reduced by the passage of these waves. When such waves damp, particles drift inwards. Resonant removal of angular momentum causes sharp outer ring edges and gaps to be produced at the strongest resonances within the ring system (Borderies et al. 1982). Waves are observed to be excited at the strongest resonances which do not produce gaps. However, waves do not appear to deplete material from the regions in which they propagate; on the contrary, a surface density enhancement is often observed in such regions (Holberg et al. 1982, Longaretti & Borderies 1986, Rosen et al. 1991).

The problem of "dredging" of ring material due to wave damping has never been solved in a self-consistent manner, in which the evolution of a region due to wave propagation, wave damping and general ring viscosity is following until a quasi-steady state is attained. (The most detailed study of this problem thus far attempted is presented by Borderies et al. 1986.) Damping of the wave in the outer portion of the region in which it propagates could bring material towards resonance, but what stops (or at least slows) this material from further inward drift? The answer to this question may be relevant to two of the major questions facing ring theorists today: (i) What maintains inner edges of the major rings of Saturn? and (ii) Do the short timescales for ring evolution due to density wave torques imply Saturn's rings are much younger than the planet itself?

4.2. *Migration of moons*

Before Voyager arrived at Saturn, Goldreich and Tremaine (1978) predicted that torques due to density waves excited by the moon Mimas at its 2:1 resonance had removed sufficient angular momentum from ring material to have cleared out Cassini's Division, a 4000 km wide region of depressed surface mass density located between the broad high density A and B rings. Although the hypothesis of density waves clearing Cassini's Division remains unverified, Voyager found a multitude of density waves excited by small newly-discovered satellites orbiting near the rings. Although the torque at these individual resonances is less than that at Mimas' 2:1 resonance, the sum of their torques is much greater. Moreover, the waves are observed and amplitudes agree with theory to within a factor of order unity.

Goldreich and Tremaine (1982) pointed out that the back torque the rings exert on the inner moons causes them to recede on a timescale short compared to the age of the Solar System; estimates suggest that all of the small moons orbiting inside the orbit of Mimas should have been at the outer edge of the A ring within the past 10^8 years, with Prometheus' journey outward occurring on a timescale of only a few millions years; more recent, lower estimates for the masses of these moons increase these timescales by a factor of a few, but do not solve the problem. Resonance locking to outer more massive moons could slow the outward recession of the small inner moons; however, angular momentum removed from the ring particles should force the entire A ring into the B ring in $\sim 10^9$ yr. Lissauer et al. (1984) and Borderies et al. (1984) quote somewhat shorter times based on larger masses of the moons.

If the calculations of torques are correct, and if no currently unknown force counterbalances them, then small inner moons and/or the rings must be new, i.e. much younger than the age of the Solar System. Both of these possibilities appear to be *a priori* highly unlikely. Rings could be remnants of Saturn's protosatellite disk which never accreted into moon-sized bodies due to the strong tidal forces of Saturn inside Roche's limit, in which case they would be $\sim 4.5 \times 10^9$ yr old. Alternatively, Saturn's ring particles could be part of the debris from a moon that was collisionally disrupted, in which case they would most likely date from the first $\sim 10^9$ years of the Solar System, when much more debris large enough to cause such a disruption was available than is today (Lissauer et al. 1988), or from a tidally disrupted giant comet, although this possibility is also *a priori* unlikely (Dones 1991). Recent accretion of the inner moons within the ring system may be possible (Borderies et al. 1984), but why did such accretion only occur during the past $\sim 10^8$ yr? For these reasons, the issue of short timescales due to density wave torques is a major outstanding problem in the field of planetary rings.

5. Migration of extrasolar planets

5.1. *Type I migration*

Goldreich and Tremaine (1980) pointed out that interactions with the gas disk could move planets large distances during the lifetime of the gas. Relative to the location of the planet, the inward/outward-propagating density waves carry a net negative/positive flux of angular momentum. If this angular momentum is deposited in the disk (as opposed to reflecting at the disk boundaries and returning to the planet), there will be a positive/negative torque on the planet from the interior/exterior parts of the disk. Variations in disk properties ought, in general, to produce a mismatch between the net interior and exterior torques, and a resultant migration rate for the planet of order

$$v_{\rm I} = k_1 \frac{M}{M_*} r\Omega \frac{\Sigma_d r^2}{M_*} \left(\frac{r\Omega}{c}\right)^3 \quad , \tag{5.1}$$

where k_1 is a measure of the torque asymmetry, M is the mass of the planet, M_* is the mass of the primary, r is the distance from the primary, Σ_d is the disk surface density, and c is the gas sound speed. From dimensional arguments, k_1 should scale with the disk aspect ratio h/r (Goldreich and Tremaine 1980). This type of orbital drift is commonly referred to as Type I migration. Ward (1997) showed that the outer torque ought to dominate in general, resulting in orbital decay for the planet.

5.2. *Gap opening and Type II migration*

The migration rate continues to increase linearly with planet mass as per Eq. (5.1), until the torque saturates and the planet pushes the inner and outer parts of the disk apart, opening a gap for itself. Having thus established a (perhaps imperfect) barrier to the passage of disk material across its orbit, it is then (more or less) locked to the viscous evolution of the disk, i.e. carried inward along with the net inward flow of the disk material—assuming, of course, that this is all happening in a part of the disk with nonzero viscosity. This type of orbital evolution is called Type II migration. The characteristic velocity for this mode of migration is

$$v_{\rm II} = k_2 \alpha r \Omega \left(\frac{c}{r\Omega}\right)^2 \quad , \tag{5.2}$$

where k_2 is a constant of order unity.

It is unclear when the formation of a gap occurs; this depends on both the viscosity of the disk, and on the way in which the density waves are dissipated. Lin and Papaloizou (1993) required the density waves to be strongly nonlinear as soon as they are launched, so that they immediately shock and deposit their angular momentum in the disk. They showed that the minimum planetary mass to accomplish this, in the limit of zero disk viscosity, is such that the Hill (or Roche) radius of the planet, defined as

$$r_{\rm H} = \left(\frac{M}{3M_*}\right)^{1/3} r \quad , \tag{5.3}$$

is equal to the half-thickness of the disk, h. This criterion—often called the thermal condition—yields a gap-opening mass of

$$M^{\rm crit}_{\rm thermal} \sim M_* \left(\frac{h}{r}\right)^3 \sim 100 \left(\frac{r}{1\ {\rm AU}}\right)^{3/4} {\rm M}_\oplus \quad , \tag{5.4}$$

where the numerical estimate is obtained using a Solar-mass primary and the standard Hayashi (1981) nebula model, which has a half-thickness of

$$h = 0.0472 (a/1\ {\rm AU})^{-5/4}\ {\rm AU} \quad , \tag{5.5}$$

and a minimum surface density—obtained from smoothly spreading out the mass contained in the planets and enhancing the gas content to solar abundance, of

$$\Sigma_{\rm min} = 1.7 \times 10^3 \left(\frac{a}{1\ {\rm AU}}\right)^{-3/2} {\rm g/cm}^2 \quad . \tag{5.6}$$

In a disk with nonzero viscosity, an additional condition for maintaining a gap is that the rate of angular momentum transfer across the gap exceed the intrinsic viscous angular momentum transport rate of the disk (Lin & Papaloizou 1993, Bryden et al. 1999). This leads to a critical mass of

$$M^{\rm crit}_{\rm viscous} \sim 40 M_* \alpha \left(\frac{h}{r}\right)^2 \sim 300 \left(\frac{\alpha}{10^{-2}}\right) \left(\frac{r}{1\ {\rm AU}}\right)^{1/2} {\rm M}_\oplus \quad . \tag{5.7}$$

In a viscous disk, the minimum gap opening mass is then ${\rm Min}[M^{\rm crit}_{\rm thermal}, M^{\rm crit}_{\rm viscous}]$.

On the other hand, Ward and Hourigan (1989) assumed that damping is linear and independent of the perturber's mass. In this way, they obtained a much smaller critical mass of

$$M^{\rm crit}_{\rm inertial} \sim \frac{h^3 \Sigma}{r} \sim 0.007 \left(\frac{r}{1\ {\rm AU}}\right)^{5/4} {\rm M}_\oplus \quad , \tag{5.8}$$

not for gap-opening, but for the planet to induce a density contrast in the disk—essentially to pile up disk mass ahead of itself—that is strong enough to stall its migration. They referred to this as the inertial mass.

The result of Rafikov (2002) is intermediate between $M^{\rm crit}_{\rm thermal}$ and $M^{\rm crit}_{\rm inertial}$. In this model, density waves travel some distance in the disk before dissipating via weak nonlinearity. The critical mass obtained in this way is around 2 $M_\oplus$ at 1 AU in an inviscid minimum-mass disk.

5.3. *Migration versus formation?*

Type I migration timescales can becomes as short as $\sim 10^4$ years for a planet of mass $\sim$10 $M_\oplus$ in a sufficiently viscous disk (Ward 1997), two to three orders of magnitude less than the lifetime of the gas disk. In the core accretion model, the critical mass needed to initiate runaway gas accretion is thought to be of order 10 $M_\oplus$; the formation timescale for a body of this size is of order a million years or more (e.g. Lissauer 1987, Lissauer & Stewart 1993, Weidenschilling 1998, Kokubo & Ida 2000, Thommes et al. 2002b). So, unless gap formation/stalling occurs at masses far below 10 $M_\oplus$, a growing core would spiral into its parent star long before it got large enough to accrete a massive atmosphere. It has been proposed that rapid migration could actually speed up the accretion process (Ward 1986), in the same way described above for the formation of Uranus and Neptune. However, since formation timescales are shorter at smaller radii, an inward-migrating protoplanet is likely to encounter not a pristine planetesimal disk, but instead the stirred-up remains of a disk which has itself already produced large protoplanets. Furthermore, simulations have shown that even in the idealized case, accretion efficiency is low and the disk has to be enhanced by at least a factor of five relative to minimum mass in order to allow a protoplanet starting at 10 AU to grow to $\sim$10 $M_\oplus$ before it falls into the star (Tanaka & Ida 1999).

Clearly, rapid Type I migration constitutes a major potential problem for the core accretion model of giant planet formation. However, there may be other ways out, apart from the possibility that this phase ends quickly. For one thing, the analysis thus far has been restricted to a single body interacting with the disk; it is unclear whether the coupling to the disk remains as strong when multiple protoplanets are launching density waves in close proximity to each other. Also, Tanaka et al. (2002) demonstrated that reflection of the density waves at the outer edge of the disk could permit a non-migrating steady state to be attained without the formation of a gap.

5.4. *Migration and the properties of extrasolar planets*

Even after locking themselves in a gap, giant planets are likely to undergo significant migration (e.g. Ward 1997). The ensemble of extrasolar planets detected by radial velocity searches (Mayor & Queloz 1995, Marcy et al. 2000) provides considerable observational support for migration playing a large role in the formation of planetary systems. The most direct clue is the large number planets on close-in orbits. Nearly half of the known planets orbit closer than 1 AU to their parent star. Although there is certainly a strong selection effect favoring detection of short-period planets, it is clear that a non-negligible fraction of giant planets somehow end up on such orbits. *In situ* formation is one possibility, but there are a number of problems in trying to construct such models; in particular, a very high surface density of solids and/or substantial planetesimal migration would be required to form a giant planet core-sized body (Bodenheimer et al. 2000). Formation at larger stellocentric distances followed by inward migration seems to provide a more natural explanation for planets like those orbiting 51 Peg and ρ CrB.

Migration may also have facilitated interactions between planets in some of the detected multiple-planet systems. The two planetary companions of Gliese 876 are in a 2:1 mean-motion resonance with each other (Marcy et al. 2001). Lee and Peale (2002) showed that capture into this resonance would occur if the two planets were originally farther apart in orbital period, and were induced to migrate toward each other, assuming their eccentricities remained low throughout ($e \sim 10^{-2}$ or less). Bryden et al. (2000) and Kley (2000) performed hydrodynamic simulations of two gap-opening planets embedded in a gas disk, and demonstrated that, if the planets' orbits are sufficiently close together, they will tend to clear the annulus of gas between them on timescales as short as thousands of years. The result is that both planets end up sharing a gap, and are pushed toward each other by the inner and outer parts of the disk. Aside from the Gliese system, the two planets discovered in HD 82943 appear to also be in a 2:1 mean motion resonance; numerical simulations suggest that, given the inferred orbital parameters, only resonant configurations are stable (Jianghui et al. 2002). Furthermore, the periods of the inner two companions of the putative three-planet system 55 Cnc are very close to a 3:1 commensurability (Marcy et al. 2002). Capture into this resonance, and other higher-order resonances, can occur if initial eccentricities of the convergently migrating planets are nonzero.

Divergent migration in multiple-planet systems is another possible mechanism for reproducing some of the observed properties of extrasolar planets. Chiang et al. (2002) show that if planets on initially circular orbits move apart in orbital period, the noncapturing resonance passages they encounter can induce significant eccentricities in the orbits of both bodies. Such a mechanism may help to account for the generally large eccentricities of those extrasolar planets with orbital radii greater than a few hundredths of an AU, out of range for tidal circularization by their parent star (e.g. Lin et al. 2000). Divergent migration may come about if there is a sufficiently steep gradient in disk viscosity, so that the inner planet migrates faster than the outer one. In particular, with reference to the model of Gammie (1996), if one planet is interior to the collisional ionization radius, where viscosity will be high, and the other is exterior to it, where viscosity is low, divergent migration may result. However, it is required that the gas annulus between the two planets persist, notwithstanding the results cited above.

5.5. *Type III migration*

In a low-mass disk—one in which the giant planet mass is comparable to or greater than the disk mass—the planet's inertia will slow the evolution of the disk. Thus the planet's migration speed will be inversely proportional to its mass:

$$v_{\rm III} \sim \frac{M_{\rm disk}}{M_{\rm planet} + M_{\rm disk}} v_{\rm II} \quad . \tag{5.9}$$

At the same time, accretion of disk material onto the star will trail off as the part of the disk interior to the planet drains onto the star. This mode can be referred to as Type III migration. Observations hint that something like this may be happening in TW Hydra, a ten million year old T Tauri star (Calvet et al. 2002). TW Hydra's accretion rate is very low compared to younger T Tauri stars, and at the same time, the spectral energy distribution of its infrared excess is best fit by a disk with a sharp inner edge at 4 AU, perhaps signifying truncation by a giant planet. Interestingly, however, TW Hydra's disk mass is estimated to be quite large, $\sim 0.6 M_*$. A 10 million year old disk still containing this much mass implies a low disk viscosity, whether the disk is simply accreting slowly, or whether it is indeed being kept at bay by a planet. Calvet et al. estimate $\alpha < 10^{-3}$.

6. Conclusions

The longstanding view of planet formation as an orderly process, involving little radial migration of material, is being made to look increasingly inaccurate by both observational and theoretical findings. In our own Solar System, the high eccentricities of objects in exterior mean-motion resonances with Neptune imply an outward migration of several AU by the outer ice giant; modeling suggests that Uranus, and to a lesser degree Jupiter and Saturn, likewise underwent migration as they cleared the surrounding planetesimal disk. An earlier and more violent period of migration could have occured if Uranus and Neptune originally formed among proto-Jupiter and -Saturn; such a model would alleviate the longstanding formation timescale problem of Uranus and Neptune, and could simultaneously help to account for the gravitationally stirred-up state of the Kuiper belt.

In the first ten million years or so of a planetary system's life, the nebular gas is still present and provides a much larger sink/source of angular momentum than the planetesimal disk. Growing protoplanets exchange angular momentum by launching density waves at resonance sites in the disk. From theoretical consideration, this ought to bring about a net loss of angular momentum and rapid orbital decay—Type I migration—for bodies of order a few Earth masses. For a sufficiently massive body, the torques between it and the disk will be strong enough to open a gap, thus locking the body into the subsequent viscous evolution of the disk in what is called the Type II mode of migration. The core accretion model of giant planet formation seems to require that the Type I to II transition occur at small masses, otherwise growing cores will spiral into the star before they can acquire a massive envelope. Alternatively, it is possible that planet formation simply is an enormously wasteful process, which dumps a steady stream of growing protoplanets onto the primary, and the end result is whatever happens to be left over when the gas fades away. This is sometimes called the "last of the Mohicans" scenario. Significant post-formation migration is quite likely responsible for the large number of planets detected on close-in orbits ("giant Vulcans," also referred to as "hot Jupiters"). In multiple-planet systems, convergent and divergent migration of planets can be invoked to explain, respectively, resonant capture and eccentricity excitation. As the nebular gas dissipates, it is likely that the tables are eventually turned; the planets, heretofore at the mercy of the gas, assert themselves and serve as anchors to slow down the viscous evolution of the last remains of the disk, so that migration ends in a Type III phase. Clearly, the present observational and theoretical "state of the art" still requires us to use a liberal amount of conjecture in attempting to sketch a coherent picture of planet migration. However, it seems equally clear that migration is intimately linked with the formation of planetary system, and a complete picture of the latter will require a full understanding of the former.

REFERENCES

Artymowicz, P. 1993 *ApJ* **419**, 155.

Bodenheimer, P., Hubickyj, O., & Lissauer, J. J. 2000 *Icarus* **143**, 2.

Borderies, N., Goldreich, P., & Tremaine S. 1982 *Nature* **299**, 209.

Borderies, N., Goldreich, P., & Tremaine, S. 1984, in *Planetary Rings* (eds. R. Greenberg, A. Brahic, & M. S. Matthews). IAU Colloq. 75, p. 713. University of Arizona Press.

Borderies, N., Goldreich, P., & Tremaine, S. 1986 *Icarus* **68**, 522.

Bryden, G., Chen, X., Lin, D. N. C., Nelson, R. P., & Papaloizou, J. C. B. 1999 *ApJ* **514**, 344.

Bryden, G., Lin, D. N. C., & Ida, S. 2000 *ApJ* **544**, 481.

Bryden, G., Różyczka, M., Lin, D. N. C., & Bodenheimer, P. 2000 *ApJ* **540**, 1091.

Cabot, W., Canuto, V. M., Hubickyj, O., & Pollack, J. B. 1987 *Icarus* **69**, 387.

Calvet, N., D'Alessio, P., Hartmann, L., Wilner, D., Walsh, A., & Sitko, M. 2002 *ApJ* **568**, 1008.

Chiang, E. I., Fischer, D., & Thommes, E. W. 2002 *ApJ* **564**, L105.

Dones, L. 1991 *Icarus* **92**, 194.

Dubrulle, B. 1993 *Icarus* **106**, 59.

Duncan, M., Quinn, T., & Tremaine S. 1987 *AJ* **94**, 1330.

Esposito, L. W., Cuzzi, J. N., Holberg, J. B., Marouf, E. A., Tyler, G. L., & Porco, C. C. 1984, in *Saturn* (eds. T. Gehrels & M. S. Matthews). p. 463. University of Arizona Press.

Fernandez, J. A. & Ip, W.-H. 1984 *Icarus* **58**, 109.

Gammie, C. F. 1996 *ApJ* **457**, 355.

Goldreich, P. & Sari, R. 2003 *ApJ* **585**, 1024.

Goldreich, P. & Tremaine, S. D. 1978 *Icarus* **34**, 227.

Goldreich, P. & Tremaine, S. 1980 *ApJ* **241**, 425.

Goldreich, P. & Tremaine, S. 1982 *ARA&A* **20**, 249.

Hahn, J. M. & Malhotra, R. 1999 *AJ* **117**, 3041.

Hayashi, C. 1981 *Prog. Theor. Phys. Suppl.* **70**, 35.

Holberg, J. B., Forrester, W. T., & Lissauer, J. J. 1982 *Nature* **297**, 115.

Jianghui, J., et al. 2002 *ApJ*, submitted; astro-ph/020825.

Kary, D. M., Lissauer, J. J., & Greenzweig, Y. 1993 *Icarus* **106**, 288.

Kley, W. 2000 *MNRAS* **313**, L47.

Kuchner, M. J., Brown, M. E., & Holman, M. 2002 *AJ* **124**, 1221.

Larson, R. B. 1989, in *The Formation and Evolution of Planetary Systems* (eds. H. A. Weaver & L. Danly). p. 31. Cambridge University Press.

Lee, M. H. & Peale, S. J. 2002 *ApJ* **567**, 596.

Levison, H. F. & Stewart, G. R. 2001 *Icarus* **153**, 224.

Lin, D. N. C., Bodenheimer, P., & Richardson, D. C. 1996 *Nature* **380**, 606.

Lin, D. N. C. & Papaloizou, J. C. B. 1980 *MNRAS* **191**, 37.

Lin, D. N. C. & Papaloizou, J. C. B. 1993, in *Protostars and Planets III* (eds. E. H. Levy & J. I. Lunine). p. 749. University of Arizona Press.

Lin, D. N. C., Papaloizou, J. C. B., Terquem, C., Bryden, G., & Ida, S. 2000, in *Protostars and Planets IV* (eds. V. Mannings, A. P. Boss, & S. S. Russell). p. 1111. University of Arizona Press.

Lissauer, J. J. 1987 *Icarus* **69**, 249.

Lissauer, J. J. & Cuzzi, J. N. 1982 *AJ* **87**, 1051.

Lissauer, J. J., Pollack, J. B., Wetherill, G. W., & Stevenson, D. J. 1995, in *Neptune and Triton*, (ed. D. P. Cruikshank). p. 37. University of Arizona Press.

Lissauer, J. J., Shu, F. H., & Cuzzi, J. N. 1984, in *Planetary Rings* (eds. R. Greenberg, A. Brahic, & M. S. Matthews). IAU Colloq. 75, p. 385. University of Arizona Press.

Lissauer, J. J., Squyres, S. W., & Hartmann, W. K. 1988 *J. Geophys. Res.* **93**, 13776.

Lissauer, J. J. & Stewart, G. R. 1993, in *Protostars and Planets III* (eds. E. H. Levy & J. I. Lunine). p. 1061. University of Arizona Press.

Longaretti, P.-Y. & Borderies, N. 1986 *Icarus* **67**, 211.

Lubow, S. H. & Ogilvie, G. I. 2001 *ApJ* **560**, 997.

Lynden-Bell, D. & Pringle, J. E. 1974 *MNRAS* **168**, 603.

Marcy, G. W., Butler, R. P., Fischer, D. A., Laughlin, G., Vogt, S. S., Henry, G. W., & Pourbaix, D. 2002 *ApJ* **581**, 1375.

Marcy, G. W., Butler, R. P., Fischer, D., Vogt, S. S., Lissauer, J. J., & Rivera, E. J. 2001 *ApJ* **556**, 296.

Marcy, G. W., Cochran, W. D., & Mayor, M. 2000, in *Protostars and Planets IV* (eds. V. Mannings, A. P. Boss, & S. S. Russell). p. 1285. University of Arizona Press.

Mayor, M. & Queloz, D. 1995 *Nature* **378**, 355.

Petit, J., Morbidelli, A., & Valsecchi, G. B. 1999 *Icarus* **141**, 367.

Pollack, J. B., Hubickyj, O., Bodenheimer, P., Lissauer, J. J., Podolak, M., & Greenzweig, Y. 1996 *Icarus* **124**, 62.

Porco, C., Danielson, G. E., Goldreich, P., Holberg, J. B., & Lane, A. L. 1984a *Icarus* **60**, 17.

Porco, C., Nicholson, P. D., Borderies, N., Danielson, G. E., Goldreich, P., Holberg, J. P., & Lane, A. L. 1984b. *Icarus* **60**, 1.

Rafikov, R. R. 2002 *ApJ* **572**, 566.

Rosen, P. A., Tyler, G. L., Marouf, E. A., & Lissauer, J. J. 1991 *Icarus* **93**, 25.

Shakura, N. I. & Sunyaev, R. A. 1973 *A&A* **24**, 337.

Shu, F. H., Dones, L., Lissauer, J. J., Yuan, C., & Cuzzi, J. N. 1985 *ApJ* **299**, 542.

Smith, B. A. & 26 colleagues 1981 *Science* **212**, 163.

Strom, S. E., Edwards, S., & Skrutskie, M. F. 1990. in *Cool Stars, Stellar Systems, and the Sun* (ed. G. Wallerstein). ASP Conf. Ser. 9, p. 275. Astronomical Society of the Pacific.

Tanaka, H. & Ida, S. 1999 *Icarus* **139**, 350.

Tanaka, H., Takeuchi, T., & Ward, W. R. 2002 *ApJ* **565**, 1257.

Thommes, E. W., Duncan, M. J., & Levison, H. F. 1999 *Nature* **402**, 635.

Thommes, E. W., Duncan, M. J., & Levison, H. F. 2002 *AJ* **123**, 2862.

Thommes, E. W., Duncan, M. J., & Levison, H. F. 2003 *Icarus* **161**, 431.

Ward, W. R. 1986 *Icarus* **67**, 164.

Ward, W. R. 1997 *Icarus* **126**, 261.

Ward, W. R. & Hahn, J. M. 1995 *ApJ* **440**, L25.

Ward, W. R. & Hourigan, K. 1989 *ApJ* **347**, 490.

Weidenschilling, S. J. 1998 *BAAS* **30**, 1050.

Organic synthesis in space

By SCOTT A. SANDFORD

NASA Ames Research Center, Mail Stop 245-6, Moffett Field, CA 94035, USA

It is becoming increasingly clear, based on a combination of observational, theoretical, and laboratory studies, that the interstellar medium (ISM) is not chemically "inert." Instead, it contains a variety of distinct environments in which chemical synthesis and alteration are constantly occurring under the aegis of a number of different processes. The result of these different processes is an interstellar medium rich in chemical diversity. The discussion found here will concentrate on those materials and molecular species built from the elements C, H, O, and N, with particular emphasis on those compounds that may be of prebiotic interest. Furthermore, there is excellent evidence that the products of interstellar chemistry are not restricted solely to the ISM, but that some fraction of these materials survive the transition from interstellar dense clouds to planetary surfaces when new stars and planets form in these clouds. This raises the interesting possibility that molecules created in the interstellar medium may play a role in the origin and evolution of life on planetary surfaces.

1. Introduction

A variety of organic and volatile compounds are now known or suspected to exist in a number of different space environments including stellar outflows, the diffuse interstellar medium, dense molecular clouds, and protostellar nebulae. On the basis of isotopic studies of meteoritic materials, it is now also understood that some fraction of these interstellar materials can survive the transition from the star formation environment of dense clouds, through the formation of a protostellar disk, into planetesimal parent bodies (comets and asteroids), with ultimate delivery to a planetary surface. This raises the possibility that interstellar molecules of prebiotic interest may play a role in the formation of life on planetary surfaces.

In the sections that follow, I will review some of our current knowledge of the chemical processes and molecular inventories of organics found in different interstellar environments. These environments span the entire evolutionary sequence from stellar death, through the diffuse ISM, on into dense molecular clouds, and through the process of new star formation. Each of these environments is dominated by different physical and chemical processes and each contains distinctly different populations of molecular materials. Emphasis will be placed on materials that are of potential prebiotic importance and their chemical precursors. This discussion will be followed by a review of the evidence, largely gathered from the laboratory isotopic study of extraterrestrial materials (meteorites and cosmic dust) and simulated analogs, that interstellar materials, including organics, can and do survive the transition from the interstellar space into forming stellar systems. Certainly many interstellar materials survive incorporation into our own Solar System.

Considered as a whole, this information suggests that interstellar organics may have played important roles in the origin and subsequent evolution of life on the Earth. Given these prebiotically important species were present in the dense cloud from which our Solar System was made, and given that star formation in dense clouds appears to be a universal process, this suggests that some of the principle starting components of life may be universally available in new stellar systems. This suggests that life may be reasonably common, at least on planets that contain conducive environments.

2. A partial inventory of organics and related materials in the Galaxy

Molecular material in the galaxy is constantly being created, modified, and destroyed as it is recycled through the processes of ejection into circumstellar space during stellar death, is dispersed into the general diffuse ISM, accumulates into dense clouds, and is subsequently either redispersed into the diffuse ISM or incorporated into new stellar systems. Each of these environments is dominated by very different physical conditions, and it is not surprising that these different environments contain different populations of molecular species.

2.1. *Organic species in stellar outflows*

The life cycle of organic materials in space begins with the ejection of material from dying stars into the ISM. In the case of organic molecules, much of this material comes from C-rich stars evolving from their asymptotic giant branch (AGB) stage into protoplanetary and planetary nebulae (PPN/PNe). Mass loss from the AGB star sends a dust and gas envelope into the ISM at the same time an increasing flux of ultraviolet photons from the dying star begin to ionize the surrounding material, creating a protoplanetary, and then planetary, nebula (Kwok 1993).

Extreme AGB carbon stars, PPN, and PNe exhibit the spectroscopic signatures of organic molecules such as acetylene (H-C≡C-H), polycyclic aromatic hydrocarbons (PAHs), aliphatic (singly C-C bonded) hydrocarbons, and C-rich dust (Cernicharo et al. 2001). Thus, the spectra of these objects are unique tracers of the birth and early evolution of these important interstellar species.

Of particular interest to astrobiology are PAHs, whose abundance makes them the largest reservoir of organic carbon compounds in the ISM. The spectra of these materials are dominated by discrete bands at 3.3, 6.2, 7.7, 8.6, and 11.2 μm, but both laboratory and telescopic studies demonstrate that they also possess an array of more subtle features (see, for example, Allamandola et al. 1989; Hudgins et al. 1994; Hudgins & Allamandola 1995; Roelfsema et al. 1996; Hudgins & Sandford 1998a,b,c). Despite their acknowledged *presence*, the *chemistry* of interstellar PAHs—how they form, which species dominate the population, and how the sizes and structures of the dominant species evolve over time—remains poorly understood. However, the availability of a large and growing laboratory database of the IR spectroscopic properties of PAHs, and their comparison with increasingly available telescopic IR spectra suggest that the material ejected from stars in the planetary nebula phase does evolve during the brief (1000–3000 years) of the AGB-to-PPN-to-PNe transition (Kwok et al. 1999; Figure 1). In particular, it appears that the number of different molecular species present in the nebula decreases with time, while at the same time the fraction of material that is ionized increases (Allamandola et al. 1999). This is presumably because the increasing UV output as the central star evolves into a white dwarf results in the photodestruction of molecular species in such a manner that only the more stable species are able to survive.

2.2. *Organics in the diffuse ISM*

Material ejected from stars is ultimately dispersed into the general diffuse interstellar medium. Here it is mixed with material already present in the diffuse ISM, including materials from the disruption of dense molecular clouds. Conditions in the diffuse ISM are harsh, and only the hardiest species in these ejecta survive the radiation fields, shock waves, etc. of this environment. Thus, it is not surprising that the diffuse ISM is not observed to contain a wide variety of different molecular species here. Nonetheless, organic species are found in the diffuse ISM.

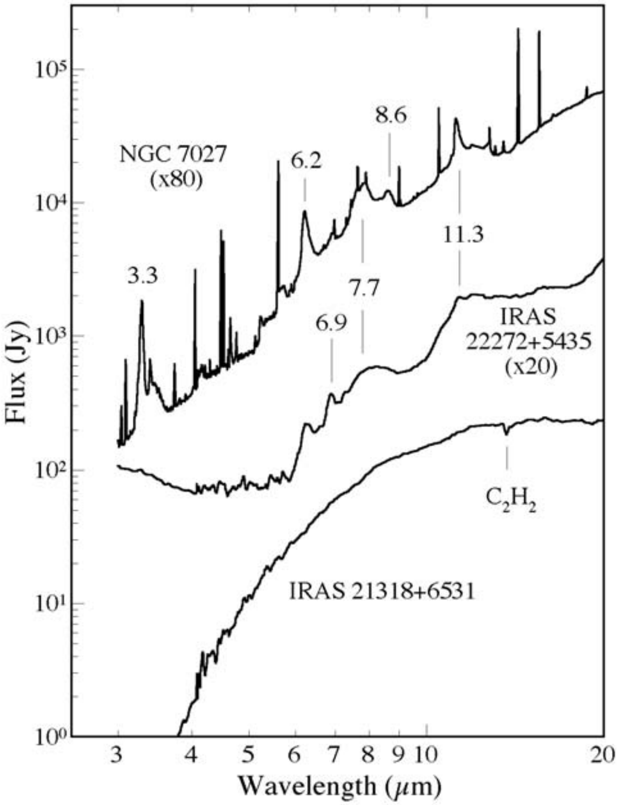

FIGURE 1. The molecular population and hence, infrared spectra, of planetary nebulae change as these objects evolve. Figure adapted from Hony (2002).

The principal evidence for carbonaceous dust in the diffuse ISM is an absorption band near 3.4 μm (Figure 2) seen along only a few lines of sight toward extinguished background stars (see, for example, Sandford et al. 1991; Pendleton et al. 1994; Pendleton & Allamandola 2002). Interestingly, the current limited set of measurements suggests this material is not uniformly distributed throughout the Galaxy, but shows higher relative concentrations in the inner most parts of the galaxy (Sandford et al. 1995). The band's profile indicates the presence of -CH_3 (methyl) and -CH_2- (methylene) groups in aliphatic hydrocarbons (molecules having only single C-C bonds) attached to perturbing chemical groups (Sandford et al. 1991). The presence of weak 3.3 μm absorption bands characteristic of aromatic CH stretching vibrations suggests that the principle perturbing chemical groups may be aromatic hydrocarbons. Thus, the currently available (but limited) spectral information is consistent with this material being similar to meteoritic kerogen, which consists of aromatic chemical units that are interlinked by aliphatic bridges (Pendleton & Allamandola 2002; Figure 3).

In addition, emission from high latitude dust—the infrared cirrus—demonstrates that the diffuse ISM also contains individual gas phase PAHs (Onaka et al. 1996). The relationship, if any, between the gas phase, emitting PAHs and the absorbing aliphatic and aromatic components is currently unknown.

As this material contains the end products of circumstellar outflows and is an important feedstock for the formation of new dense molecular clouds, these organics are an integral

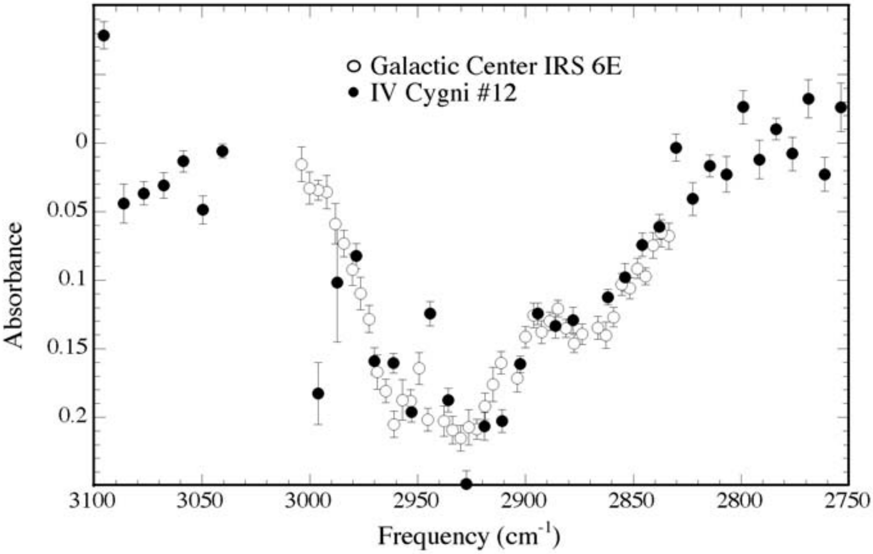

FIGURE 2. The C-H stretch band of organics in the diffuse ISM towards the galactic center and VI Cyg#12. The positions and strengths of the features indicate the presence of aliphatic -CH_2- and -CH_3 groups in a ~2.5/1 proportion. Adapted from Sandford et al. (1991).

part of the lifecycle of complex interstellar organics. Although based on data from a limited number of objects, it is now thought that ~30% of the cosmic C in the diffuse ISM is in the form of gas phase PAHs intermixed with particles containing aliphatic and aromatic hydrocarbons. However, the specific identity of these materials, their sources, inter-relationships, and their galactic abundance and distribution, are not currently well understood.

2.3. *Molecules in interstellar dense molecular clouds*

Independent of the formation site and evolution of cosmic organics, they must ultimately pass through the dense molecular cloud phase if they are to be incorporated into a planetary system. Important, distinct dense cloud environments include the quiescent dense cloud medium, the higher temperature and radiation infall zones surrounding forming protostars, the planet-forming disks around Young Stellar Objects (YSOs), and the zones in which winds and H II regions produced by young hot stars interact with nearby dense cloud material. Astronomical data and astrochemical laboratory simulations suggest that these environments alter the initial molecular species present and produce a wide variety of new molecular species, many of which are of astrobiological significance.

Eventually, the material in dense clouds experiences one of two fates. The majority of the material is ultimately dispersed back into the diffuse ISM as star formation disrupts the original dense cloud. This material will undergo additional rapid processing as it is exposed to more intense ionizing radiation and warming during dispersal of the cloud. Small portions of the material in dense clouds will suffer a different fate, however, and be incorporated into newly forming stellar systems. This material may also experience elevated radiation fields and will undergo some amount of warming above the temperature of the general cloud. These materials are of particular potential astrobiological interest since they represent the population of species that may ultimately end up on planetary surfaces.

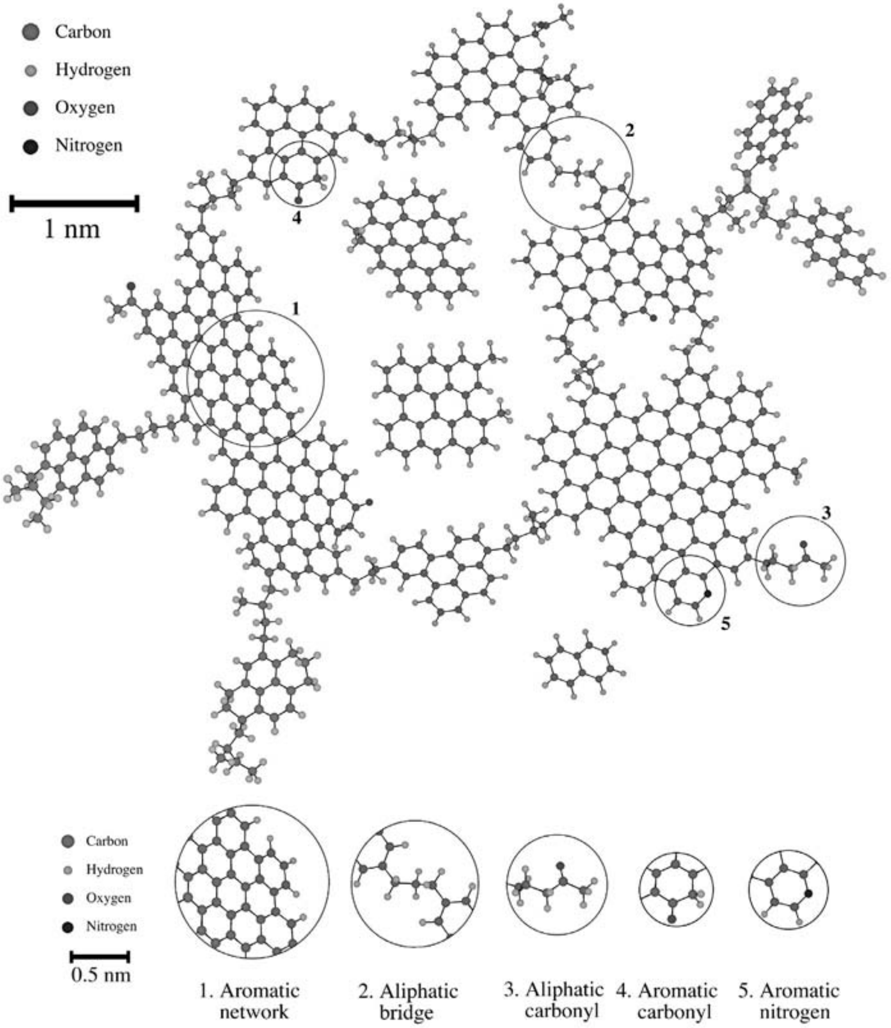

FIGURE 3. The C-H stretching infrared absorption spectrum of diffuse interstellar dust is consistent with a material like meteoritic kerogen, i.e. a complex network of aromatic groups interlinked by a variety of highly branched, largely aliphatic bridges. Figure adapted from Pendleton & Allamandola (2002).

The chemical diversity of dense interstellar clouds is far larger than most other astronomical environments. This is because the greater densities found in these objects allow for far richer and more rapid production of new species. In addition, the optical depth of these clouds screens out much of the diffuse medium radiation field, thereby protecting newly formed species from immediate destruction. Indeed, the vast majority of gas phase molecular species identified in space by sub-millimeter and radio spectral techniques are found in dense clouds (see, for example, Turner 1991, 2001; van Dishoeck & Blake 1998). Many of these gas phase species are thought to been formed by gas phase ion-molecule reactions (Herbst 1987; Langer & Graedel 1989), although contributions from grain mantle accretion and evaporation also play an important role (Brown & Charnley 1990, 1997; Charnley et al. 1992).

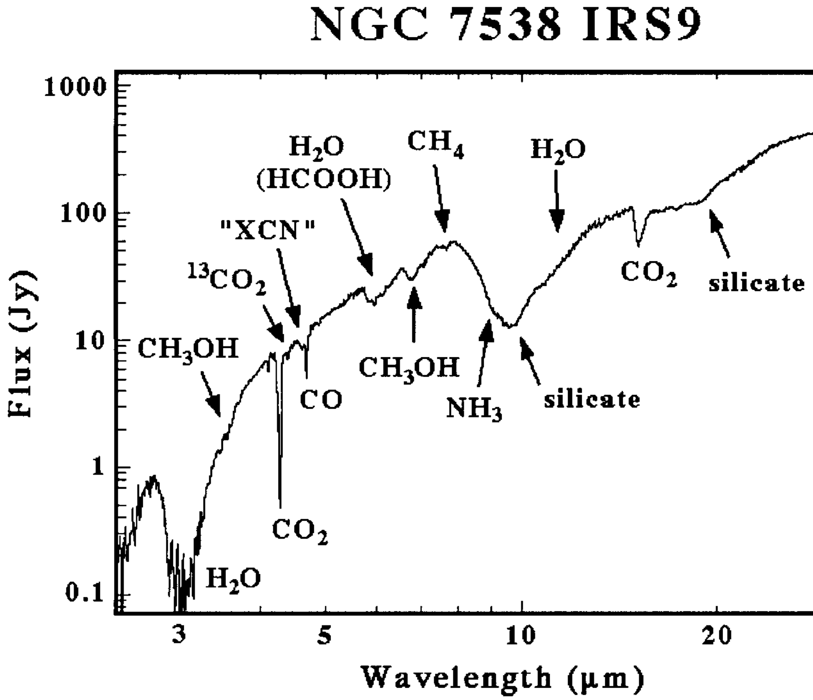

FIGURE 4. The infrared spectrum of the protostar NGC 7338 IRS9. The absorption features are cause by materials, largely volatile ices frozen onto dust grains, along the line of sight to the star. Figure adapted from Whittet et al. (1996).

Because the optical depth of dense clouds is sufficient to screen out ambient radiation, the interiors of these clouds cool to very low temperatures ($T < 50$ K). At these temperatures, most gas phase molecules are expected to condense onto grains and most of the products of gas phase chemistry in the ISM should spend the majority of their lifetimes in icy grain mantles (Sandford & Allamandola 1993). This is, in fact, observed to be the case. Infrared spectra of protostars embedded within dense clouds show a number of absorption bands that are produced by icy materials that lie in the cloud along the line of sight to the embedded star. The most abundant species in these ices are typically H_2O, CO, CO_2, and CH_3OH, but a host of other species, such as HCO, H_2CO, HCOOH, CH_4, OCS, and possibly ketones and/or aldehydes are seen at lower abundances (Whittet et al. 1985; Tielens & Allamandola, 1987; Sandford et al. 1988; d'Hendecourt & Jourdain de Muizon 1989; Lacy et al. 1991, 1998; Allamandola et al. 1992; Schutte et al. 1994; Palumbo et al. 1995; Gibb et al. 2000; Figure 4). Larger molecules, if present, should also be efficiently condensed into these ices. In this regard, it is perhaps not too surprising that absorption features have also been detected in dense clouds that are consistent with the presence of significant amounts of polycyclic aromatic hydrocarbons (Sellgren et al. 1995; Brooke et al. 1996; Bregman et al. 2000).

Icy grain mantles in dense clouds serve as more than simple storage sites for molecules created in the gas phase, however. Their surfaces are capable of mediating gas-grain reactions that can further enrich the molecular diversity of the cloud. The compositions of new species produced in this manner depends largely on the local H/H_2 ratio (Tielens & Hagen 1982; Tielens & Allamandola 1987). In environments where the H/H_2 ratio is large, surface reactions with H atoms are important and atoms like C, N, and O will be converted to simple hydrides like CH_4, NH_3, and H_2O. If the H/H_2 ratio is substantially less than one, reactive species such as O and N are free to react with one another forming molecules such as O_2 and N_2. Thus, two qualitatively different types of ice mantle may be produced by grain surface reactions, one dominated by polar, H-bonded molecules and the other dominated by non-polar or only slightly polar, highly unsaturated molecules. Observational evidence of the band profiles of CO-containing interstellar ices seem to support this dichotomy (Sandford et al. 1988; Tielens et al. 1991).

By their very nature, the formation processes of gas phase ion-molecule reactions and gas-grain reactions in dense clouds do not lead to the formation of particularly large molecules. The formation of more complex species is likely dominated by chemistry driven by the irradiation of the ices in dense clouds. Since many of the molecular products of ice irradiation are of direct prebiotic interest, an entire section of this paper has been devoted to this subject.

2.4. *Molecules formed by the irradiation of interstellar ices*

Energetic *in situ* processing of interstellar ices in dense clouds is driven by cosmic rays, cosmic ray induced UV, the significantly enhanced UV field in star forming regions, and high energy particle bombardment and UV radiation from the T-Tauri phase in stellar birth. This radiation results in the breaking and rearrangement of chemical bonds within the ice, causing the destruction of some species and the creation of others (Hagen et al. 1979). Many of the new species are highly reactive radicals and ions, and additional chemistry occurs if the ices are warmed and these species become mobile. This overall process is an important source of molecular diversity since it can create molecular species, particularly complex ones, that cannot be made via gas phase and gas-grain reactions at the low temperatures and pressures characteristic of dense clouds.

The main evidence that such processing occurs in dense clouds comes from the detection of a broad, often weak, feature centered at about 4.62 μm in the spectra of materials along the lines of sight to some protostars (Lacy et al. 1984; Tegler et al. 1993; Weintraub et al. 1994). This feature is thought to be due to C≡N stretching vibrations in a molecule produced when ices are exposed to ion bombardment or ultraviolet radiation (Moore et al. 1983; Lacy et al. 1984; Grim & Greenberg 1987; Tegler et al. 1993; Bernstein et al. 1995, 2000; Palumbo et al. 2000).

Laboratory studies have shown that energetic processing of realistic interstellar ice analogs produces large numbers of new organic compounds in these ices, including species far more complex than the starting materials (Hagen et al. 1979; Moore et al. 1983; Agarwal et al. 1985, d'Hendecourt et al. 1986; Allamandola et al. 1988; Bernstein et al. 1995). Species produced in this manner include, but are not limed to, ethanol (CH_3CH_2OH), amides (such as formamide, $HC(=O)NH_2$, and acetamide, $CH_3C(=O)NH_2$), nitriles and isonitriles (R-C≡N and R-N≡C), ketones R-C(=O)-R', hexamethylenetetramine (HMT, $C_6H_{12}N_4$), and compounds related to polyoxymethylene POM, $(\text{-}CH_2O\text{-})_n$ (Schutte et al. 1993; Bernstein et al. 1995; Cottin et al. 2001).

Three families of species made during these irradiation experiments are of particular interest to astrobiology. First, it has recently been found that the UV photolysis of simple ices containing H_2O, CH_3OH, and NH_3 or HCN results in the production of amino acids (Bernstein et al. 2002). The UV photolysis of more complex ices also produces amino acids (Muñoz Caro et al. 2002). The most abundant amino acids formed in this manner are glycine, alanine, and serine—some of the most abundant amino acids found in primitive meteorites and within living organisms on Earth.

Irradiation of these same ices also produces significant amounts of amphiphiles (Dworkin et al. 2001). Amphiphiles are long chain molecules that contain both hydrophilic and

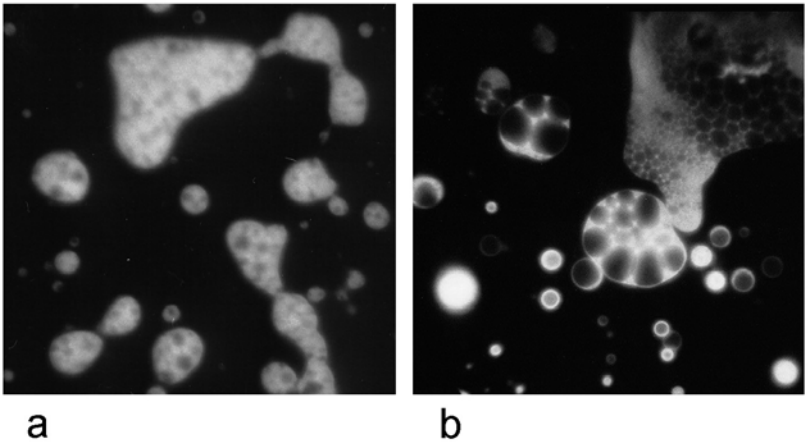

FIGURE 5. Vesicles produced in liquid water by amphiphilic molecules extracted from (a) the organic residue produced when an interstellar ice analog is irradiated with UV photons and then warmed, and (b) the Murchison meteorite. Images adapted from Dworkin et al. (2001) and Deamer (1985).

hydrophobic end groups. Such molecules, when exposed to water, can spontaneously self-assemble to form vesicles (Figure 5a). Amphiphilic compounds that exhibit similar behaviors are found in primitive meteorites (Figure 5b) (Deamer 1985). Delivery of such molecules to the early Earth may have played a key role in the production of the first membranes and vesicles, key protocellular structures that many believe represent a critical step towards the formation of life (Koch 1985; Morowitz 1992; Bernstein et al. 1999a; Segre et al. 2001).

Finally, if PAHs are present in these ices, irradiation results in the alteration of their edge structures through oxidation and reduction reactions (Bernstein et al. 1999b, 2001). Oxidation reactions result in the production of new aromatic alcohols, ketones, and ethers, while reduction reactions lead to aromatics with partially aliphatic rings (Hn-PAHs) (Bernstein et al. 1996). The aromatic ketones (quinones) are of particular astrobiological interest since members of this family of molecules play key roles in the biochemistry of all living systems on Earth (see, for example, Suttie 1979; Thurl et al. 1985). As with the amino acids and amphiphiles mentioned above, oxidized and reduced aromatic species are found in primitive meteorites (Basile et al. 1984; Krishnamurthy et al. 1992).

2.5. *Interstellar nanodiamonds*

One carbon-dominated phase that is worthy of additional note, even though it is not an "organic" in the generally recognized sense of the word, is nanodiamonds.

In the late 1980s it was demonstrated that a significant fraction of the carbon in many meteorites resided in nanodiamonds (e.g. Lewis et al. 1987; Blake et al. 1988; Fraundorf et al. 1989). These nanodiamonds are extremely small (typically 10–15 Å in diameter) and a major fraction of their atoms lie near their surfaces (Bernatowicz et al. 1990). These nanodiamonds were extracted from their parent meteorites by a succession of chemical processes that etched away the other components of the meteorites. The remaining diamond-rich residues carry certain isotopically anomalous noble gas fractions (e.g. Ming & Anders 1988) suggesting the meteoritic nanodiamonds, or at least a fraction of them, predate the Solar System and must have at one time resided in the interstellar medium.

There is also telescopic spectral evidence for the presence of nanodiamonds in extrasolar space. In 1992, Allamandola et al. serendipitously detected a new, weak absorption feature near 3.47 μm in the spectra of 4 embedded protostars embedded in dense interstellar clouds. The band falls at a position characteristic of the C-H stretch of tertiary carbon, i.e. a carbon atom bonded to one hydrogen and three other carbons. The tertiary carbon feature, in conjunction with an absence of strong features due to primary (-CH_3) and secondary (-CH_2-) carbons, led Allamandola et al. (1992) to suggest that the carrier has a diamond-like structure. Subsequent (unpublished) work by this group and others (Sellgren et al. 1995; Brooke et al. 1996) has shown that the feature exists in the spectra of virtually all protostars in dense clouds studied to date. More recently, very compelling matches have been made between the spectra of nanodiamonds and the telescopic spectra of a few young stars thought to be surrounded by dust disks (Guillois et al. 1999).

The ultimate source of the interstellar nanodiamonds remains uncertain although several formation processes, including chemical vapor deposition, photolysis of hydrocarbons, grain collisions produced by interstellar shocks, and irradiation have been suggested (see, for example, Anders & Zinner 1993; Dai et al. 2002). Transmission electron microscope studies of the meteoritic nanodiamonds suggest that formation by chemical vapor deposition may be the most likely (Daulton et al. 1996). In any event, the ubiquity of the band in the spectra of dense clouds and the high abundances of nanodiamonds found in meteorites and inferred in dense clouds suggests that their formation involves a relatively common environment rather than an exotic one.

3. Evidence of the survival of circumstellar/interstellar materials in planetary systems

A key question in Astrobiology is “To what extent did extraplanetary/extrasolar organics play a role in the formation of life on Earth?” To address this issue we must establish the extent to which interstellar organics survive incorporation into protostellar nebulae and ultimate delivery to planetary surfaces. The best way to pursue this issue is to search for links between interstellar organics and those found in primitive extraterrestrial objects like meteorites and interplanetary dust particles (IDPs), materials that are thought to sample comets and asteroids. In the case of organic materials, one of the best tracers of such connections is *deuterium (D).*

The current cosmic D/H ratio is $\sim 1.5 \times 10^{-5}$ (Vidal-Madjar et al. 1998). However, both observations and theory indicate that in the ISM this ratio can reach, and even exceed, a value of 0.1 for certain interstellar molecular species (Tielens 1983, 1992, 1997; Turner 2001). These enrichments result from isotopic fractionation that occurs during at least four different interstellar chemical processes: gas phase ion-molecule reactions, gas-grain surface reactions, unimolecular photodissociation reactions, and the irradiation of mixed molecular ices (Sandford et al. 2000, 2001). Since these are some of the same processes that make complex organic molecular species in the interstellar medium, many interstellar organics should be strongly D-enriched, and the presence of D-rich species in meteoritic samples is probably indicative of the presence of material with an interstellar heritage.

Deuterium enrichments are, in fact, seen in meteorites. In some cases the enrichments are seen in bulk meteoritic materials (Zinner 1988), but D enrichments have also been observed in meteoritic subfractions and even within specific classes of molecular species, such as amino and carboxylic acids (Epstein et al. 1987; Pizzarello et al. 1991; Krishnamurthy et al. 1992). The aromatic fraction of meteorites is known to be a major carrier of excess deuterium (Robert & Epstein 1982; Yang & Epstein 1983; Kerridge et al. 1987; Krishnamurthy et al. 1992).

IDPs are also known to contain D-enriched components, and these components are often distributed heterogeneously within the particles on very small scales (McKeegan et al. 1985; Messenger 2000). Ion microprobe studies of the D distribution and correlation with other elements suggest tentative links between the D and a carbonaceous carrier (McKeegan et al. 1985; Messenger et al. 1996; Aleon et al. 2000; Messenger 2000). Aleon et al. (2000) reported that the D enrichment increases in IDPs with increased C/H, with most enrichments occurring when C/H is well above one, suggesting the carrier has an aromatic nature. IDPs are known to contain abundant aromatics (PAHs) (Allamandola et al. 1987; Clemett et al. 1993), but the relationship between them and D-enrichments is not currently clear.

The detection of highly elevated D/H ratios in organics from meteorites and IDPs, and the elevated terrestrial D/H ratio of $\sim 1 \times 10^{-4}$, suggest that both our Solar System as a whole, and the Earth in particular, received a significant portion of D-enriched materials during their formation. The arrival of interstellar organics as part of meteorites and IDPs continues today, but during the earliest times after the formation of our Solar System, i.e. during the period of the Late Bombardment when the last remaining vestiges of the solar nebula were being swept up or ejected, these materials would have been arriving at the surface of the Earth at rates many orders of magnitude higher than we see today (Chyba et al. 1990).

4. Summary

A combination of observational, theoretical, and laboratory studies clearly demonstrate that a number of different environments in the interstellar medium are sites of active astrochemistry. The result of these different processes is an interstellar medium rich in chemical diversity, much of which is manifested in the form of organic species which may be of prebiotic interest. Furthermore, there is ample isotopic evidence from meteorites and IDPs, particularly in the form of species enriched in deuterium, that some fraction of these materials can survive the transition from dense clouds to planetary surfaces when new stars and planets form in dense interstellar molecular clouds. This raises the interesting possibility that molecules created in the interstellar medium may play a role in the origin and evolution of life on planetary surfaces.

The author would like to acknowledge support of NASA grants 344-37-44-01 (Origins of Solar Systems), 344-38-12-04 (Exobiology), and 344-50-92-02 (Astrobiology).

REFERENCES

Agarwal, V. K., Schutte, W., Greenberg, J. M., Ferris, J. P., Briggs, R., Connor, S., Van De Bult, C. P. E. M., & Baas, F. 1985 *Origins of Life* **16**, 21.

Aleon, J., Engrand, C., Robert, F., & Chaussidon, M. 2000 *Meteoritics and Planetary Science* **35**, A19.

Allamandola, L. J., Hudgins, D. M., & Sandford, S. A. 1999 *ApJ* **511**, L115.

Allamandola, L. J., Sandford, S. A., Tielens, A. G. G. M., & Herbst, T. M. 1992 *ApJ* **399**, 134.

Allamandola, L. J., Sandford, S. A., & Valero, G. J. 1988, in *Dust in the Universe* (eds. M. E. Bailey & D. A. Williams). p. 548. Cambridge University Press.

Allamandola, L. J., Sandford, S. A., & Wopenka, B. 1987 *Science* **237**, 56.

Allamandola, L. J., Tielens, A. G. G. M., & Barker, J. R. 1989 *ApJS* **71**, 733.

Anders, E. & Zinner, E. 1993 *Meteoritics* **28**, 490.

Basile, B. P., Middleditch, B. S., & Oro, J. 1984 *Org. Geochem.* **5**, 211.

Bernatowicz, T. J., Gibbons, P. C., & Lewis, R. S. 1990 *ApJ* **359**, 246.

Bernstein, M. P., Dworkin, J., Sandford, S. A., & Allamandola, L. J. 2001 *Meteoritics and Planetary Science* **36**, 351.

Bernstein, M. P., Dworkin, J. P., Sandford, S. A., Cooper, G. W., & Allamandola, L. J. 2002 *Nature* **416**, 401.

Bernstein, M. P., Sandford, S. A., & Allamandola, L. J. 1996 *ApJ* **472**, L127.

Bernstein, M. P., Sandford, S. A., & Allamandola, L. J. 1999a. *Scientific American* **281**, #1, 42.

Bernstein, M. P., Sandford, S. A., & Allamandola, L. J. 2000 *ApJ* **542**, 894.

Bernstein, M. P., Sandford, S. A., Allamandola, L. J., Chang, S., & Scharberg, M. A. 1995 *ApJ* **454**, 327.

Bernstein, M. P., Sandford, S. A., Allamandola, L. J., Gillette, J. S., Clemett, S. J., & Zare, R. N. 1999b *Science* **283**, 1135.

Blake, D. F., Freund, F., Krishnan, K. F. M., Echer, C. J., Shipp, R., Bunch, T. E., Tielens, A. G. G. M., Lipari, R. J., Hetherington, C. J. D., & Chang, S. 1988 *Nature* **332**, 611.

Bregman, J. D., Hayward, T. L., & Sloan, G. C. 2000 *ApJ* **544**, L75.

Brooke, T. Y., Sellgren, K., & Smith, R. G. 1996 *ApJ* **459**, 209.

Brown, P. D. & Charnley, S. B. 1990 *MNRAS* **244**, 432.

Cernicharo, J., Heras, A. M., Tielens, A. G. G. M., Pardo, J. R., Herpin, F., Guelin, M., & Waters, L. B. F. M. 2001 *ApJ* **546**, L123.

Charnley, S. 1997, in *Astronomical and Biochemical Origins and the Search for Life in the Universe* (eds. C. B. Cosmovici, S. Bowyer, & D. Werthimer). p. 89. Proc. 5th International Conf. on Bioastronomy, IAU Coll. #161.

Charnley, S. B., Tielens, A. G. G. M., & Millar, T. J. 1992 *ApJ* **399**, L71.

Chyba, C. F., Thomas, P. J., Brookshaw, L., & Sagan, C. 1990. *Science* **249**, 366.

Clemett, S., Maechling, C., Zare, R., Swan, P., & Walker, R. 1993 *Science* **262**, 721.

Cottin, H., Szopa, C., & Moore, M. H. 2001 *ApJ* **561**, L139.

Dai, Z. R., Bradley, J. P., Joswiak, D. J., Brownlee, D. E., Hill, H. G. M., & Genge, M. J. 2002 *Nature* **418**, 157.

Daulton, T., Eisenhour, D., Bernatowicz, T., Lewis, R., & Buseck, P. 1996 *Geochim. Cosmochim. Acta* **60**, 4853.

Deamer, D. W. 1985 *Nature* **317**, 792.

d'Hendecourt, L. B., Allamandola, L. J., Grim, R. J. A., & Greenberg, J. M. 1986 *A&A* **158**, 119.

d'Hendecourt, L. B. & Jordain de Muizon, M. 1989 *A&A* **223**, L5.

Dworkin, J. P., Deamer, D. W., Sandford, S. A., & Allamandola, L. J. 2001 *Proc. Nat. Acad. Sci. USA* **98**, 815.

Epstein, S., Krishnamurthy, R. V., Cronin, J. R., Pizzarello, S., & Yuen, G. U. 1987 *Nature* **326**, 477.

Fraundorf, P. Fraundorf, G., Bernatowicz, T., Lewis, R., & Ming, T. 1989 *Ultamicroscopy* **27**, 401.

Gibb, E. L., Whittet, D. C. B., Schutte, W. A., Boogert, A. C. A., Chiar, J. E., Ehrenfreund, P., Gerakines, P. A., Keane, J. V., Tielens, A. G. G. M., van Dishoeck, E. F., & Kerkhof, O. 2000 *ApJ* **536**, 347.

Grim, R. J. A. & Greenberg, J. M. 1987 *ApJ* **321**, L91.

Guillois, O., Ledoux, G., & Reynaud, C. 1999 *ApJ* **521**, L133.

Hagen, W., Allamandola, L. J., & Greenberg, J. M. 1979 *Astrophys. Spa. Sci.* **65**, 215.

Herbst, E. 1987. In *Interstellar Processes* (eds. D. J. Hollenbach & H. A. Thronson, Jr.). p. 611. D. Reidel.

Hony, S. 2002. Ph.D. thesis, University of Amsterdam.

Hudgins, D. M. & Allamandola, L. J. 1995 *J. Phys. Chem.* **99**, 3033.

Hudgins, D. M. & Sandford, S. A. 1998a *J. Phys. Chem.* **102**, 329.

Hudgins, D. M. & Sandford, S. A. 1998b *J. Phys. Chem.* **102**, 344.

Hudgins, D. M. & Sandford, S. A. 1998c *J. Phys. Chem.* **102**, 353.

Hudgins, D. M., Sandford, S. A., & Allamandola, L. J. 1994 *J. Phys. Chem.* **98**, 4243.

Kerridge, J. F., Chang, S., & Shipp, R. 1987 *Geochim. Cosmochim. Acta* **51**, 2527.

Koch, A. L. 1985 *J. Mol. Evol.* **21**, 270.

Krishnamurthy, R., Epstein, S., Cronin, J., Pizzarello, S., & Yuen, G. 1992 *Geochim. Cosmochim. Acta* **56**, 4045.

Kwok, S. 1993 *Ann. Rev. Astr. Ap.* **31**, 63.

Kwok, S., Volk, K., & Hrivnak, B. J. 1999 *A&A* **350**, L35.

Lacy, J. H., Baas, F., Allamandola, L. J., Persson, S. E., McGregor, P. J., Lonsdale, C. J., Geballe, T. R., & van der Bult, C. E. P. 1984 *ApJ* **276**, 533.

Lacy, J., Carr, J., Evans, N., Baas, F., Achtermann, J., & Arens, J. 1991 *ApJ* **376**, 556.

Lacy, J. H., Faraji, H., Sandford, S. A., & Allamandola, L. J. 1998 *ApJ* **501**, L105.

Langer, W. D. & Graedel, T. E. 1989 *ApJS* **69**, 241.

Lewis, R. S., Tang, M., Wacker, J. F., Anders, E., & Steel, E. 1987 *Nature* **326**, 160.

McKeegan, K. D., Walker, R. M., & Zinner, E. 1985 *Geochim. Cosmochim. Acta* **49**, 1971.

Messenger, S. 2000 *Nature* **404**, 968.

Messenger, S., Walker, R. M., Clemett, S. J., & Zare, R. N. 1996 *Lunar Planet. Sci. Conf.* **XXVII**, 867.

Ming, T. & Anders, E. 1988 *Geochim. Cosmochim Acta* **52**, 1235.

Moore, M. H., Donn, B., Khanna, R., & A'Hearn, M. F. 1983 *Icarus* **54**, 388.

Morowitz, H. J. 1992 *Beginnings of Cellular Life.* Yale University Press.

Muñoz Caro, G. M., Meierhenrich, U. J., Schutte, W. A., Barbier, B., Arcones Segovia, A., Rosenbauer, H., Thiemann, W. H.-P., Brack, A., & Greenberg, J. M. 2002 *Nature* **416**, 403.

Onaka, T., Yamamura, I., Tanabe, T., Roellig, T. L., & Yuen, L. 1996 *Publ. Astron. Soc. Japan* **48**, L59.

Palumbo, M. E., Pendleton, Y. J., & Strazzulla, G. 2000 *ApJ* **542**, 890.

Palumbo, M. E., Tielens, A. G. G. M., & Tokunaga, A. T. 1995 *ApJ* **449**, 674.

Pendleton, Y. J. & Allamandola, L. J. 2002 *ApJS* **138**, 75.

Pendleton, Y. J., Sandford, S. A., Allamandola, L. J., Tielens, A. G. G. M., & Sellgren, K. 1994 *ApJ* **437**, 683.

Pizzarello, S., Krishnamurthy, R. V., Epstein, S., & Cronin, J. R. 1991 *Geochim. Cosmochim. Acta* **55**, 905.

Robert, F. & Epstein, S. 1982 *Geochim. Cosmochim. Acta* **46**, 81.

Roelfsema, P. R., Cox, P., Tielens, A. G. G. M., Allamandola, L. J., Baluteau, J.-P., Barlow, M., Beintema, D., Boxhoorn, D., Cassinelli, J. P., Caux, E., Churchwell, E., Clegg, P. E., de Graauw, T., Heras, A. M., Huygen, R., van der Hucht, K., Hudgins, D. M., Kessler, M., Lim, T., & Sandford, S. A. 1996 *A&A* **315**, L289.

Sandford, S. A. & Allamandola, L. J. 1993 *ApJ* **417**, 815.

Sandford, S. A., Allamandola, L. J., Tielens, A. G. G. M., Sellgren, K., Tapia, M., & Pendleton, Y. 1991 *ApJ* **371**, 607.

Sandford, S. A., Allamandola, L. J., Tielens, A. G. G. M., & Valero, G. J. 1988 *ApJ* **329**, 498.

Sandford, S. A., Bernstein, M. P., Allamandola, L. J., Gillette, J. S., & Zare, R. N. 2000 *ApJ* **538**, 691.

Sandford, S. A., Bernstein, M. P., & Dworkin, J. P. 2001 *Meteoritics and Planetary Science* **36**, 1117.

Sandford, S. A., Pendleton, Y. J., & Allamandola, L. J. 1995 *ApJ* **440**, 697.

Schutte, W. A., Allamandola, L. J., & Sandford, S. A. 1993 *Icarus* **104**, 118.

Schutte, W. A., Gerakines, P. A., van Dishoeck, E. F., Greenberg, J. M., & Geballe, T. R. 1994, in *Physical Chemistry of Molecules and Grains in Space* (ed. I. Nenner). Conf. Proceedings No. 312, p. 73. American Institute of Physics.

Segre, D., Deamer, D. W., & Lancet, D. 2001 *Orig. Life Evol. Biosphere* **3**, 119.

Sellgren, K., Brooke, T. Y., Smith, R. G., & Geballe, T. R. 1995 *ApJ* **449**, L69.

Suttie, J. W., ed. 1979 *Vitamin K Metabolism and Vitamin K Dependent Proteins*, Proc. 8th Steenbock Symp., Univ. Park Press.

Tegler, S. C., Weintraub, D. A., Allamandola, L. J., Sandford, S. A., Rettig, T. W., & Campins, H. 1993 *ApJ* **411**, 260.

Thurl, S., Buhrow, I., & Schafer, W. 1985 *Biol. Chem. Hoppe Seyler* **366**, 1079.

Tielens, A. G. G. M. 1983 *A&A* **119**, 177.

Tielens, A. G. G. M. 1992, in *Astrochemistry of Cosmic Phenomena* (ed. P. D. Singh). p. 91. Kluwer.

Tielens, A. G. G. M. 1997, in *In Astrophysical Implications of the Laboratory Study of Presolar Materials* (eds. T. J. Bernatowicz & E. K. Zinner). p. 523. Amer. Inst. of Physics.

Tielens, A. G. G. M. & Allamandola, L. J. 1987, in *Physical Processes in Interstellar Clouds* (eds. G. E. Morfill & M. Scholer). p. 333. D. Reidel.

Tielens, A. G. G. M. & Hagen, W. 1982 *A&A* **114**, 245.

Tielens, A. G. G. M., Tokunaga, A. T., Geballe, T. R., & Baas, F. 1991 *ApJ* **381**, 181.

Turner, B. E. 1991 *ApJS* **76**, 617.

Turner, B. E. 2001 *ApJS* **136**, 579.

van Dishoeck, E. F. & Blake, G. A. 1998 *Annu. Rev. Astron. Astrophys.* **36**, 317.

van Dishoeck, E. F., Blake, G. A., Draine, B. T., & Lunine, J. I. 1993, in *Protostars and Planets III* (eds. E. H. Levy & J. I. Lunine). p. 163. University of Arizona Press.

Vidal-Madjar, A., Lemoine, M., Ferlet, R., Hibrard, G., Koester, D., Audouze, J., Cassi, M., Vangioni-Flam, E., Webb, J. 1998 *A&A* **338**, 694.

Weintraub, D. A., Tegler, S. C., Kastner, J. H., & Rettig, T. 1994 *ApJ* **423**, 674.

Whittet, D. C. B., Longmore, A. J., & McFadzean, A. D. 1985 *MNRAS* **216**, p. 45.

Yang, J. & Epstein, S. 1983 *Geochim. Cosmochim. Acta* **47**, 2199.

Zinner, E. 1988, in *Meteorites and the Early Solar System* (eds. J. F. Kerridge & M. S. Matthews). p. 956. University of Arizona Press.

The vegetation red edge spectroscopic feature as a surface biomarker

By S. SEAGER[1,2] AND E. B. FORD[3]

[1]Department of Terrestrial Magnetism, Carnegie Institution of Washington, Washington, DC 20015, USA

[2]School of Natural Sciences, Institute for Advanced Study, Princeton, NJ 08540, USA

[3]Princeton University Observatory, Princeton, NJ 08544, USA

The search for Earth-like extrasolar planets is in part motivated by the potential detection of spectroscopic biomarkers. Spectroscopic biomarkers are spectral features that are either consistent with life, indicative of habitability, or provide clues to a planet's habitability. Most attention so far has been given to atmospheric biomarkers, gases such as O_2, O_3, H_2O, CO, and CH_4. Here we discuss surface biomarkers. Surface biomarkers that have large, distinct, abrupt changes in their spectra may be detectable in an extrasolar planet's spectrum at wavelengths that penetrate to the planetary surface. Earth has such a surface biomarker: the vegetation "red edge" spectroscopic feature. Recent interest in Earth's surface biomarker has motivated Earthshine observations of the spatially unresolved Earth and two recent studies may have detected the vegetation red edge feature in Earth's hemispherically integrated spectrum. A photometric time series in different colors should help in detecting unusual surface features in extrasolar Earth-like planet spectra.

1. Introduction

One hundred extrasolar giant planets are currently known to orbit nearby sun-like stars. These planets have been detected by the radial velocity method and so, with the exception of the one transiting planet, only the minimum mass and orbital parameters are known. Many plans are underway to learn more about extrasolar planets' physical properties from ground-based and space-based observations and via proposed or planned space missions. Direct detection of scattered or thermally emitted light from the planet itself is the only way to learn about a variety of the planet's physical characteristics. Direct detection of Earth-size planets, however, is extremely difficult because of the proximity of a parent star that is 10^6 to 10^{10} times brighter than the planet.

Terrestrial Planet Finder (*TPF*), with a launch date in the 2015 timeframe, is being planned by NASA to find and characterize terrestrial-like planets in the habitable zones of nearby stars. The ESA mission *Darwin* has similar goals. The motivation for both of these space missions is the detection and spectroscopic characterization of extrasolar terrestrial planet atmospheres. Of special interest are atmospheric biomarkers—such as O_2, O_3, H_2O, CO and CH_4—which are either indicative of life as we know it, essential to life, or can provide clues to a planet's habitability (Des Marais et al. 2002). In addition, physical characteristics such as temperature and planetary radius could be constrained from low-resolution spectra.

We have shown (Ford, Seager, & Turner 2001) that planet characteristics could also be derived from photometric measurements of the planet's variability. A time series of photometric data of a spatially unresolved Earth-like planet could reveal a wealth of information such as weather, the planet's rotation rate, presence of large oceans or surface ice, and existence of seasons. The amplitude variation of the time series depends on cloud-cover fraction; more cloud cover makes a more photometrically uniform Earth and so reduces variability. The signal-to-noise necessary for photometric study would be obtained by a mission capable of measuring the sought-after atmospheric biomarker spectral features. Furthermore the photometric variability could be monitored concurrently with a spectroscopic investigation, as was done for the transiting extrasolar giant planet HD209458b (Charbonneau et al. 2002).

To detect and study surface properties only wavelengths that penetrate to the planetary surface are useful. Visible wavelengths are more suited than mid-IR wavelengths for such measurements for several reasons. First, the albedo contrast of surface components is much greater than the temperature variation across the planet's surface. Second, at visible wavelengths the planet's flux is from scattered starlight and hence at some configurations the planet is only partially illuminated. This allows a more concentrated signal from surface features, such as continents, as they rotate in and out of view. Furthermore, the non-uniform illumination and non-isotropic scattering of different surface components mean much of the scattered light can come from a small part of the planet's surface. At mid-IR wavelengths the planet has, to first order, uniform flux across the planet hemisphere. In addition, the narrow transparent spectral window at 8–12 μm will close for warmer planets than Earth and for planets with more water vapor than Earth. However, further study at the mid-IR "window" needs to be investigated.

An extremely exciting possibility, aided by a photometric time series, is the detection of surface biomarkers in the spectrum of an extrasolar planet. This would be possible at wavelengths that penetrate to the planet's surface, and for surface features that have large, distinct, abrupt changes in their spectra. Although most surface features (e.g. ice, sand) show very little or very smooth continuous opacity changes with wavelength, Earth has one surface feature with a large and abrupt change: vegetation (Figure 1). In this paper we discuss Earth's vegetation red-edge spectroscopic feature as a surface biomarker.

2. The vegetation red edge spectral feature

All chlorophyll-producing vegetation has a very strong rise in reflectivity at around 0.7 μm by a factor of five or more. This red edge spectral signature is much larger than the familiar chlorophyll reflectivity bump at 0.5 μm, which gives vegetation its green color. In fact, if our eyes could see a little further to the red, the world would be a very different place: plants would be very red, and very bright. The glare from plants would be unbearably high, like that of snow. The red edge is caused both by strong chlorophyll absorption to the blue of 0.7 μm, and a high reflectance due to plant cell structure to the red of 0.7 μm. Figure 1 shows a deciduous plant leaf reflection spectrum. The high absorptance at UV wavelengths (not shown) and at visible wavelengths is by chlorophyll and is used by the leaf for photosynthesis. Photosynthesis is the process by which vegetation and some other organisms use energy from the sun to convert H_2O and CO_2 into sugars and O_2. The primary molecules that absorb the light energy and convert it into a form that can drive this reaction are chlorophyll A (0.450 μm) and B (0.680 μm).

As seen in Figure 1, between 0.7 μm and 1 μm the leaf is strongly reflective. Not shown in Figure 1 is that the leaf also has a very high transmittance at these same wavelengths, such that reflectivity plus transparency is near 100%. Interestingly, the bulk of the energy of solar radiation as it reaches sea level is at approximately 0.6 to 1.1 μm. If plants absorbed with the same efficiency at these wavelengths as at visible wavelengths they would become too warm and their chlorophyll would degrade. A specific plant must balance the competing requirements of absorption of sunlight at wavelengths appropriate for photosynthesis reactions with efficient reflectance at other wavelengths

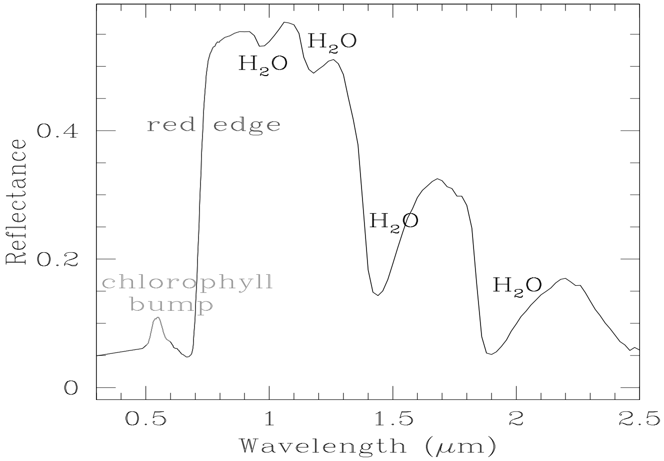

FIGURE 1. Reflection spectrum of a deciduous leaf. The small bump near 0.5 μm is a result of chlorophyll absorption (at 0.45 μm and 0.68 μm) and gives plants their green color. The much larger sharp rise (between 0.7 and 0.8 μm) is known as the red edge and is due to the leaf cell structure.

to avoid overheating (Gates et al. 1965). Therefore the exact wavelength and strength of the red edge depends on the plant species and environment. Although negligible from the TPF view point, it is interesting to note that the specific wavelength and strength of the red edge feature is used for remote sensing of specific locations on Earth to identify plant species and also to monitor a field of vegetation's (such as crops) health and growth during the growing season.

In the near-infrared, as shown in Figure 1, plants have water absorption bands. The band strength depends on plant water content, weather conditions, plant type, and geographical region. These absorption features can be quite strong, but are not very useful for identifying life, since they would only be indicative of water and would not be distinguishable from atmospheric water vapor.

Plant leaves are very reflective away from chlorophyll absorption and water absorption wavelengths due to the internal leaf structure (Gates et al. 1965). Light partially scatters off of the leaf surface but also scatters efficiently inside the leaf. Light reflects off of and refracts through cell walls from the surrounding air gaps between cells. Inside cells themselves the high change in the index of refraction from 1.33 for water to 1.00 for air causes an efficient internal reflection at the interface between cell walls and the surrounding air gaps. Also, inside cells light can Mie or Rayleigh scatter off of cell organelles which have sizes on the order of the wavelength of light. The overall reflectance and transmittance is a complex function of the cell size and shape and the size and shape of the air gaps between the cells (see, e.g. Govaerts et al. 1996). Because there is little absorption away from the chlorophyll and water absorption wavelength regions, light will eventually scatter out of the leaf at the top (reflection) or bottom (transmission).

3. Plants as an Earth surface biomarker

The red-edge spectroscopic feature is very strong for an individual plant leaf, at a factor of five or more. Averaged over a (spatially unresolved) hemisphere of Earth, however, the vegetation red-edge spectral feature is reduced from this high reflectivity down to a few percent. This is because of several effects including the forest canopy architecture, soil

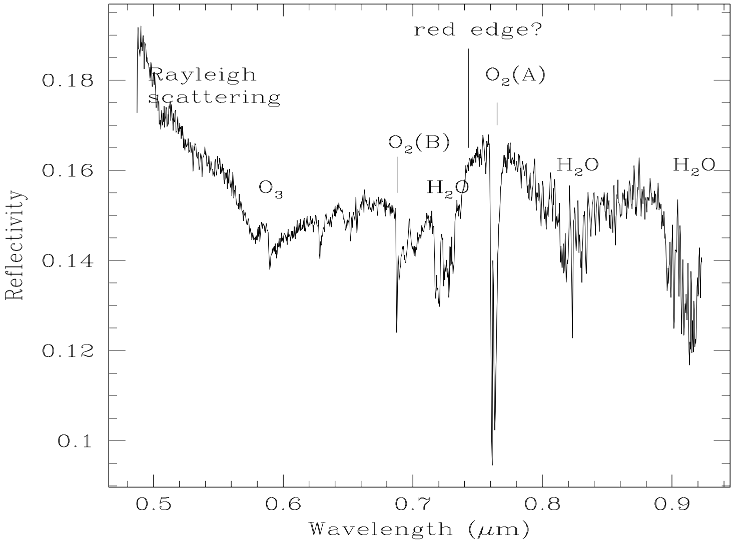

FIGURE 2. A visible wavelength spectrum of the spatially unresolved Earth, as seen with Earthshine (adapted from Woolf et al. 2002). The viewpoint is largely centered equatorially on the Pacific ocean. The major atmospheric features are identified. The reflectivity scale is arbitrary. Data courtesy of N. Woolf and W. Traub. For details see Woolf et al. 2002.

characteristics, the non-continuous coverage of vegetation across Earth's surface, and the presence of clouds which prevent viewing the surface. In addition the reflectance of vegetation is anisotropic and so the illumination conditions and viewing angle are important. Nevertheless at a signal of a few percent Earth's vegetation red edge may be a viable surface biomarker to a distant, telescope-bearing civilization. The chlorophyll bump at 0.5 μm, however, is negligible in a hemispherically averaged spectrum. The spectral signature of oceanic vegetation or plankton is also unlikely to be detectable, due to strong absorption by particles in the water and also by the strong absorptive nature of liquid water beyond red wavelengths.

Using vegetation's red edge as a surface biomarker is not a new idea. Early last century the high near-infrared reflection signature was used to test the hypothesis that the changing dark patches on Mars were due to seasonal changes of vegetation (Slipher 1924; Millman 1939; Tickhov 1947; Kuiper 1949). Not surprisingly, only negative results were obtained.

More recently Sagan et al. (1993) used the *Galileo* spacecraft for a "control experiment" to search for life on Earth using only conclusions derived from data and first principle assumptions. En route to Jupiter, the *Galileo* spacecraft used two gravitational assists at Earth (and one at Venus). During the December 1990 fly-by of Earth, the *Galileo* spacecraft took low-resolution spectra of different areas of Earth. In addition to finding "abundant gaseous oxygen and atmospheric methane in extreme thermodynamic disequilibrium", Sagan et al. (1993) found "a widely distributed surface pigment with a sharp absorption edge in the red part of the visible spectrum" that "is inconsistent with all likely rock and soil types." Observing ~100 km^2 areas of Earth's surface the vegetation red edge feature showed up as a reflectance increase of a factor of 2.5 between a band centered at 0.67 μm and one at 0.76 μm. In contrast there was no red edge signature from non-vegetated areas.

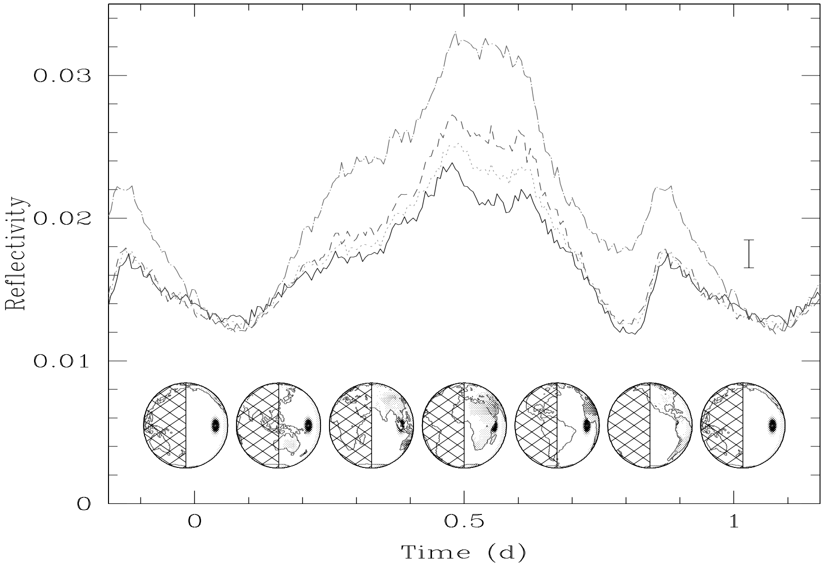

FIGURE 3. A light curve for a cloud-free Earth model for one rotation. The x-axis is time and the y-axis is the reflectivity normalized to a Lambert disk at a phase angle of 0°. The viewing geometry is shown by the Earth symbols, and a phase angle of 90° is used. Note that a different phase angle will affect the reflectivity due to a larger or smaller fraction of the disk being illuminated; because of the normalization the total reflectivity is $\ll$ in this case of phase angle of 90°. From top to bottom the curves correspond to wavelengths of 0.75, 0.65, 0.55, and 0.45 μm, and their differences reflect the wavelength-dependent albedo of different surface components. The noise in the light curve is due to Monte Carlo statistics in our calculations. The images below the light curve show the viewing geometry (cross-hatched region is not illuminated) and relative contributions from different parts of the disk (shading ranges from <3% to >40%, from white to black) superimposed on a map of the Earth. At time = 0.5 day, the Sahara desert is in view and causes a large peak in the light curve due to the reflectivity of sand which is especially high in the near-infrared (top curve).

A new area of extrasolar planet research is now emerging: using Earthshine to study the spatially unresolved Earth. Earthshine is light from the sun that has been scattered off of Earth onto the moon and then back to Earth. It appears as a faint glow on the otherwise dark part of the moon during the crescent phase, but can be studied with a CCD camera and specialized coronagraph even as the moon waxes (Goode et al. 2001). Satellite data of Earth is not as useful as Earthshine because it is highly spatially resolved and limited to narrow spectral regions. Also, since most satellite data is collected by looking straight down at specific regions of Earth hemispherical flux integration with lines-of-sight through different atmospheric path lengths is not available. Recent spectral observations of Earthshine have tentatively detected the red-edge signature at the few percent level. Woolf et al. (2002) observed the setting crescent moon from Arizona which corresponds to Earth as viewed over the Pacific Ocean. Nevertheless their spectrum (Figure 2) shows a tantalizing rise just redward of 0.7 μm that is tentatively the spectroscopic red-edge feature. Figure 2 also shows other interesting features of Earth's visible-wavelength spectrum, notably O_2 and H_2O absorption bands (note that spectral lines of both O_2 and H_2O cut into the red-edge signature.) Arnold et al. (2002) have made observations of Earthshine on several different dates. With observations from France the Earthshine is from America and the Pacific Ocean (the evening moon) and

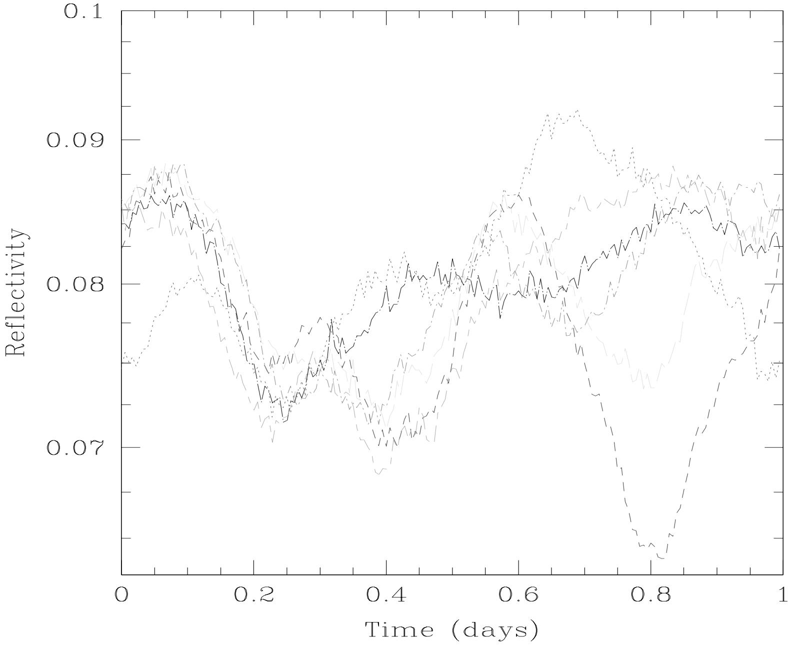

FIGURE 4. Rotational light curves for model Earth with clouds. This figure shows six different daily light curves at 0.55 μm for our Earth model with clouds, as viewed from a phase angle of 90°. These theoretical light curves use cloud cover data from satellite measurements taken on six consecutive days.

Europe and Asia (the morning moon). After subtracting Earth's spectrum to remove the contaminating atmospheric absorption bands they find a vegetation red edge signal of 4 to 10%.

4. Temporal variability to detect surface biomarkers

A small but sharp spectral feature from a component of the planet's surface should be more easily identified by temporal variation. As the continents rotate in and out of view, the planet's reflectivity will change, causing a change in the measured spectrum. Recent Earthshine measurements have shown that detection of Earth's vegetation-red edge is tricky due to smearing out by other atmospheric and surface features. Trying to identify such small features at unknown wavelengths in an extrasolar planet spectrum may be very difficult. We propose that such spectral features could be much more easily identified by the increased temporal variability at a carefully chosen color. In particular, any changes associated with a rotational period would be highly relevant. Since the wavelength of any surface biomarkers would not be known *a priori*, flexible data acquisition is essential. For example low-resolution spectra could be later integrated into narrow-band photometry of many different bands. Here we discuss simulations of Earth's temporal variability, including preliminary calculations of Earth's vegetation red-edge variability.

We model the photometric flux from a rotating Earth by a Monte Carlo code using a spherical map of Earth which specifies the scattering surface type at each point on the sphere and a set of wavelength-dependent bidirectional reflectance distribution functions which specify the probability of light incident from one direction to scatter into another direction for each type of scattering surface (see Ford et al. 2001 for details). We

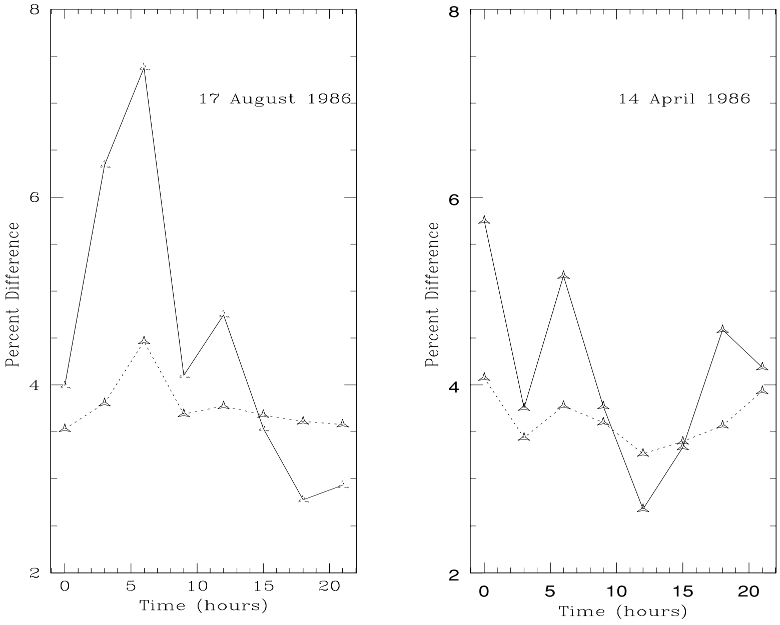

FIGURE 5. Variability of Earth's color. The solid line shows a color [(I(0.75–0.8)–I(0.7–0.65))/I(0.75–0.8)] chosen to emphasize variability of vegetation's red edge. For comparison, the dotted line shows a color [I(0.85–0.8)–I(0.75–0.8)/I(0.75–0.8)], which is less sensitive to vegetation. These colors include theoretical spectra from Des Marais et al. 2002 modulated by our Earth rotational surface and cloud model. The cloud cover for the model in the left panel is from the ICSSP database from 17 August 1986 and in the right panel from 14 April 1986. Earth is more variable in the color sensitive to the red edge vegetation feature.

use a map of Earth from a one-square-degree satellite surface map that classifies each pixel as permanent ice, dirty/temporary ice, ocean, forest, brush, or desert. We consider cloudy models using the scattering properties of Earth clouds and we also include an approximation of atmospheric Rayleigh scattering. We focus our attention to quadrature (a phase angle of 90°) for which the planet-star separation is largest and the observational constraints thus least severe.

The existence of different surface features on a planet may be discernable at visible wavelengths as different surface features rotate in and out of view. Considering a cloud-free Earth, the diurnal flux variation caused by different surface features rotating in and out of view could be as high as 200% (Figure 3). This high flux variation is not only due to the high contrast in different surface components' albedos, but also to the fact that a relatively small part of the visible hemisphere dominates the total flux from a spatially unresolved planet. Clouds interfere with surface visibility and in the presence of clouds the diurnal light curve shown in Figure 3 becomes that shown in Figure 4. It is very interesting to note that an extrasolar Earth-like planet certainly could have a lower cloud cover fraction than Earth's 50% cloud cover. The cloud pattern and cover fraction are influenced by a variety of factors including the planet's rotation rate, continental arrangement, obliquity, and presence of large bodies of water.

A time series of data in different colors (Figure 5) may help make it possible to detect a small but unusual spectral feature, even with variable atmospheric features. Most of Earth's surface features, such as sand or ice, have a continuous increase or minimal change with wavelength, in contrast to the abrupt vegetation red edge spectral feature. We have generated spectrophotometric variability of Earth by using theoretical spectra (for cloud and non-cloud atmosphere from Traub (private communication) as included in (Des Marais et al. 2002)) modulated by our Earth rotational surface and cloud model. We use cloud data from the ISCCP database (Rossow & Schiffer 1991) such that the rotating Earth also has changing cloud patterns. We have chosen to integrate the spectrum into colors, the first [(I(0.75–0.8)–I(0.7–0.65))/I(0.75–0.8)] chosen to emphasize variability of vegetation's red edge and the second, for comparison, a color [I(0.85–0.8)–I(0.75–0.8)/I(0.75–0.8)], which is less sensitive to vegetation. Figure 5 shows that Earth is more variable in a color across the red edge than for colors with similar wavelength differences in other parts of Earth's spectrum. For extrasolar planet measurements spectra or spectrophotometric data would be most useful in the form of a spectrum so that the photometric bands can be chosen after data acquisition.

5. Extrasolar Plants?

It is difficult to speculate on extrasolar plants and we will not do so here. Some might argue that the vegetation red edge differences among coniferous, deciduous, and desert plants are meaningful. However, all Earth vegetation has almost certainly evolved from the same ancestor and it is not a fair evolutionary experiment. Nevertheless a few interesting facts are suggestive and useful to those who wish to speculate on the possible existence extrasolar plants or light harvesting organisms:

- Plants absorb very efficiently throughout the UV and the visible wavelength regions of the spectrum where the energy is required for photosynthesis (involving molecular electronic transitions);
- At sea level (after atmospheric extinction) the solar energy distribution peaks at 1 μm and approximately 50% of the energy is redward of 0.7 μm;
- Plants reflect and transmit almost 100% of light in the wavelength region where the direct sunlight incident on plants has the bulk of its energy;
- Considering these last three points, Earth's primary “light harvesting organism,” vegetation, has evolved to balance the competing requirements of absorption of sunlight at wavelengths appropriate for photosynthesis reactions with efficient reflectance at other wavelengths to avoid overheating (Gates et al. 1965); and
- Other pigments involved in vegetation's light harvesting process also absorb in the 0.44 μm wavelength regime (but only chlorophyll B absorbs near 0.68 μm). However, some other organisms have pigments that absorb at other wavelengths.

6. Summary and Conclusions

The vegetation red edge spectroscopic feature is a factor of five or more change in reflection at ~0.7 μm. This red edge feature is well-used in satellite remote sensing studies of Earth's vegetation. Earthshine observations have been used to detect the vegetation red edge signature in the spatially unresolved spectrum of Earth where it appears at the few percent level.

When discovered, observations of extrasolar Earth-like planets at wavelengths that penetrate to the planet's surface will be very useful, especially for planets with much lower cloud cover than Earth's 50%. A time series of spectra or broad-band photometry could reveal surface features of a spatially unresolved planet, including surface biomarkers. Earth's hemispherically integrated vegetation red-edge signature is weak (a few to ten percent), but Earth-like planets with different rotation rates, obliquities, land-ocean fraction, and continental arrangement may well have lower cloud-cover.

While it is near impossible to speculate on spectral features of light harvesting organisms on extrasolar planets, flexible data acquisition will maximize scientific return. The detection of an unusual spectral signature that is inconsistent with any known atomic, molecular, or mineralogical signature would be fantastic. Combinations of unusual spectral features together with strong disequilibrium chemistry would be even more intriguing and would certainly motivate additional studies to better understand the prospects for such a planet to harbor life.

We thank the conference organizers for a very interesting meeting. We thank Ed Turner and other members of the Princeton *Terrestrial Planet Finder* team for many useful discussions and Wes Traub and Ken Jucks for use of their theoretical spectra. S. S. is supported by the W. M. Keck Foundation and the Carnegie Institution of Washington.

REFERENCES

Arnold, L., Gillet, S., Lardiere, O., Riaud, P., & Schneider, J. 2002 *A&A* **392**, 231.

Charbonneau, D., Brown, T. M., Noyes, R. W., & Gilliland, R. L. 2002 *ApJ* **568**, 377.

Des Marais, D. J., Harwit, M. O., Jucks, K. W., Kasting, J. F., Lin, D. N. C., Lunine, J. I., Schneider, J., Seager, S., Traub, W. A., & Woolf, N. J. 2002 *Astrobiology* **2**, 153.

Ford, E. B., Seager, S., & Turner E. L. 2001 *Nature* **412**, 885.

Gates, D. M., Keegan, H. J., Schleter, J. C., & Weidner, V. R. 1965 *Applied Optics* **4**, 11.

Goode, P. R., Qiu, J., Yurchyshyn, V., Hickey, J., Chu, M.-C., Kolbe, E., Brown, C. T., & Koonin, S. E. 2001 *GeoRL* **28**, 1671.

Govaerts, Y. M., Jacquemoud, S., Verstraete, M. M., & Ustin, S. L. 1996 *Applied Optics* **35**, 6585–6598.

Kuiper, G. P. 1949 *The Atmospheres of the Earth and Planets.* p. 339. University of Chicago Press.

Millman, P. M. 1939 *The Sky* **3**, no. 10, 11.

Rossow, W. B. & Schiffer, R. A. 1991 *Bull. Amer. Meteor. Soc.* **72**, 2.

Slipher, J. M. 1924 *PASP* **36**, 261.

Tikhov, G. A. 1947 *Bull. Astr. and Geodet. Soc. U.S.S.R.* **8**, 8.

Woolf, N. J., Smith, P. S., Traub, W. A., & Jucks, K. W. 2002 *ApJ* **574**, 430.

Search for extra-solar planets through gravitational microlensing

By KAILASH C. SAHU

Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, MD 21218, USA

Gravitational microlensing offers a powerful technique to search for extra-solar planets around lensing stars via short-timescale amplifications produced by the planet on the microlensing lightcurve. This method is technologically simple, can be carried out with a network of relatively small ground-based telescopes, and is sensitive down to earth-mass planets.

More than 100 microlensing events towards the Galactic bulge have been monitored by the PLANET collaboration to look for such planetary signals. No clear planetary signal has been detected, which implies that less than 33% of the lensing stars have Jupiter-mass planets with orbital radii of 1.5–4 AU. Since other techniques are currently not sensitive to the outer portion of these orbital radii, these are the best current limits on extra-solar planets at these orbital separations.

Isolated planetary-mass objects can also reveal themselves as short timescale microlensing events in a monitoring program. Lack of such short-timescale events in the MACHO and EROS database towards the LMC suggests that the contribution of planetary-mass objects is less than 10% of the halo dark matter.

1. Gravitational microlensing as a tool

Uranus is roughly a 6th magnitude object, and is almost a naked-eye object. Yet it was discovered only in 1791, long after the telescope was invented, and it took a great astronomer like Sir William Herschel to do so (at least by some accounts). Uranus was the last planet to be discovered by its direct light. An overwhelming majority of all the subsequent planet discoveries have been made by the gravitational effect of the planet.

In 1845, the French astronomer Leverrier and the British astronomer John Adams predicted the position of Neptune from the orbital perturbations of Uranus. The prediction was then observationally followed up by Johan Galle, who discovered Neptune in a single night of observations. The story of the discovery of Pluto is similar, although a bit more complicated. In 1930, Percival Lowell predicted the position of Pluto from the orbital perturbations of Neptune. This was observationally followed up by Clyde Tombaugh, which eventually led to the discovery of Pluto. The discovery of the first definitive extra-solar planet around the pulsar PSR1257+12 was again through the gravitational effect of the planet (Wolszczan & Frail 1992). The recent discoveries of planets around the nearby stars have made use of the gravitational effect through the radial velocity perturbation caused by the planet on the parent star (Mayor & Queloz 1995; Marcy & Butler, 1995; Marcy et al. 2001; also see http://exoplanets.org/planet_table.html). A tremendous amount of effort has been spent in looking for planets around other stars through other esoteric means, such as spatial interferometry or adaptive optics. While some of these efforts will no doubt bear fruit in the future as we overcome the technical challenges they pose, they have borne very little fruit so far. The reason is not difficult to understand: the gravitational effect, in almost all cases, makes use of the bright nearby object, whereas the other methods seek to overcome the effect of the bright nearby object through technology. In the case of spatial interferometry or adaptive optics, one must always fight to keep the light of the bright star down in order to detect the faint planetary signal in the presence of this highly dominant bright source. In other words, the bright star always acts as a hindrance to the search, and is always something that one must win over in order to be able to detect the much fainter planet nearby. The situation is reversed in case of the gravitational effect of the planet, in which case one simply uses the effect of the planet on the brighter object to look for the planet.

It is only recently that ground-based transit experiments such as OGLE have announced the detection of over 100 planet-like transits in bulge and disk stars (Udalski et al. 2002; 2003). Many of them have been shown to be due to stellar-mass companions, and two have been confirmed as having planetary mass from their radial velocity signatures (Konacki et al. 2003a,b; Sasselov, 2003; Sirko & Paczyński 2003; Dreizier et al. 2003). These are the first extra-solar planetary mass objects to be discovered by a technique other than gravitational effect of the planet.

The technique of microlensing uses a different aspect of the gravitational effect, and the star in this case too helps in the search for the planet. This may potentially be a very powerful tool to look for extra-solar planets, and, as discussed in more detail later, this is the only method sensitive to the search for Earth-like planets around normal stars using ground based observations. It must be noted, however, that microlensing does have its selection effects, and this method is more sensitive to detection of planets around *low* mass stars since, statistically, a large fraction of the lenses are expected to be low mass stars.

2. Microlensing due to stars

The idea of microlensing by stars is not new. In 1936, Einstein wrote a small paper in *Science* where, he did 'a little calculation' at the request of his friend Robert Mandl and showed that if a star happens to pass very close to another star in the line of sight, then the background star will be lensed (Einstein 1936). However, he also dismissed the idea as only a theoretical exercise and remarked that there was 'no hope of observing such a phenomenon directly.' He was right at that time; the probability of observing is less than one in a million even towards the Galactic bulge where the density of stars is about the highest. With the technology of 1936, there was no way one could observe such a phenomenon directly.

Paczyński, in two papers written in 1986 and 1991, noted that if one could monitor a few million stars, one could observe microlensing events, perhaps as a signature of the dark matter towards the LMC, or by known stars towards the Galactic bulge (Paczyński 1986; Paczyński 1991; also see Griest 1991). The project was taken up immediately by three groups—MACHO, EROS and OGLE—who first reported their detections of microlensing events in 1993 (Alcock et al. 1993; Aubourg et al. 1993; Udalski et al. 1993). By now, more than 1000 events have been discovered by the collaborations, mostly towards the Galactic bulge. In case of the Galactic bulge events, the lenses are known to be mostly low-mass stars in the line-of-sight. It is then a logical step to look for planets around these lensing stars through microlensing.

3. Theoretical aspects of microlensing

It is useful to review some of the basics of the lensing by a single star and a binary star. For the details of the theoretical aspects of the lensing by a star, the reader may refer to the excellent review article by Paczyński (1996) and the very exhaustive monograph devoted to the subject of Gravitational Lensing by Schneider, Ehlers and Falco (1992), and Petters, Levine and Wambsganss (2001).

With the lensing geometry as described in Figure 1, the Einstein ring radius R_E can be written as

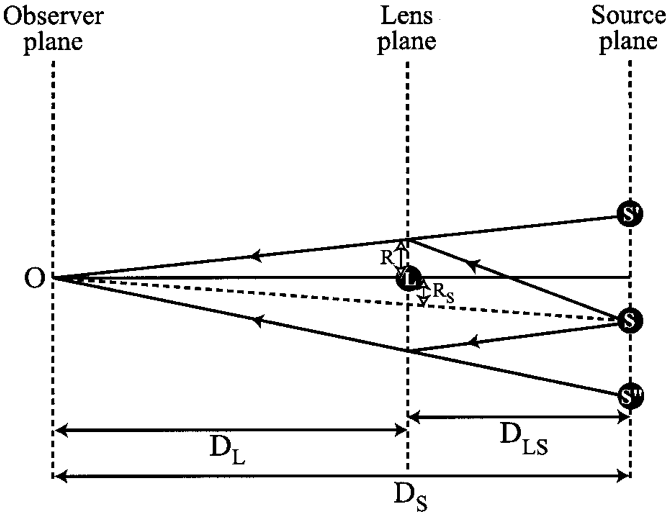

FIGURE 1. Schematic geometry of the gravitational microlensing. The observer, the lens and the source are located at positions O, L and S, respectively. D$_L$ is the distance to the lens, D$_S$ is the distance to the source, and D$_{LS}$ is the distance from the lens to the source. The lens (L) produces two images of the source at positions S$'$ and S$''$. At the lens plane, the lens, source and the two images lie on a straight line. S$'$ corresponds to the brighter image which is formed outside the Einstein ring and is at least as bright as the source itself. S$''$ corresponds to the fainter image which is formed inside the Einstein ring.

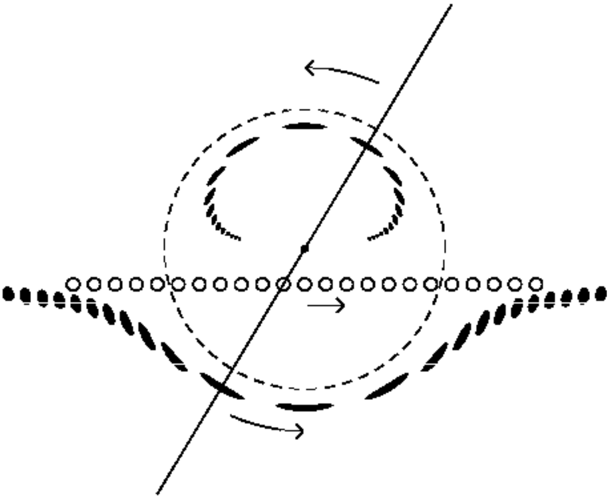

FIGURE 2. This figure shows how the apparent positions and the sizes of the images change at various stages of the microlensing. In this geometry the position of the lens, indicated by a solid dot, is fixed, and the open circles show the actual positions of the source. The filled circles show the images of the source as the source passes close to the lens in the plane of the sky. The dashed circle is the Einstein ring of the lens. At any instant, the source, the lens and the two images lie on a straight line. (Taken from Paczyński, 1996)

$$R_E^2 = \frac{4GM_LD}{c^2}, \qquad D \equiv \frac{D_{LS}D_L}{D_S} \tag{3.1}$$

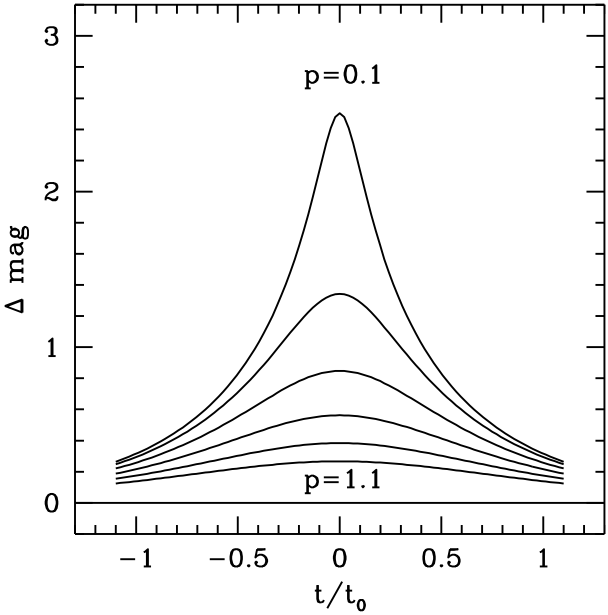

FIGURE 3. The microlensing light curves as a function of impact parameter.

where M$_L$ is the mass of the lens, D$_L$ is the distance to the lens, D$_{LS}$ is the distance from the lens to the source, and D$_S$ is the distance from the observer to the source. The positions of the two images as the source moves through the Einstein ring are shown in Fig. 2.

The amplification due to the microlensing depends only on the impact parameter, which can be written as

$$A = \frac{u^2 + 2}{u(u^2 + 4)^{1/2}} \tag{3.2}$$

where u is the impact parameter in units of R_E.

The expected microlensing lightcurves for different impact parameters are shown in Fig. 3.

This equation can be easily inverted to derive the impact parameter from a given amplification

$$u = 2^{1/2} \left[A \left(A^2 - 1 \right)^{-1/2} - 1 \right]^{1/2} \tag{3.3}$$

which can be used to derive the minimum impact parameter u_m from an observed light curve.

The timescale of microlensing is the time taken by the source to cross the Einstein ring radius, which is given by

$$t_0 = \frac{R_E}{V_e} \tag{3.4}$$

where V_e is the tangential velocity of the lens with respect to the source. The impact parameter at any time during the microlensing event can be expressed as

$$u = \left[u_m^2 + \left(\frac{t - t_m}{t_0} \right)^2 \right]^{1/2} \tag{3.5}$$

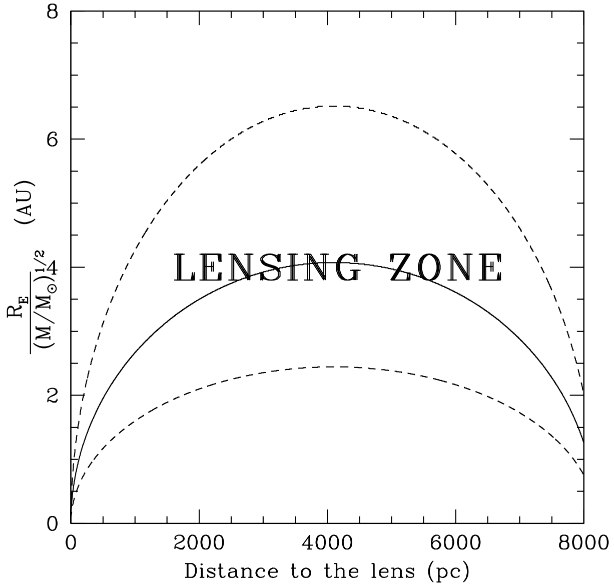

FIGURE 4. The solid line shows the size of the Einstein ring (R_E) as a function of the distance to the lens, assuming the source to be at the Galactic bulge. The two images of the source generally lie within 0.6 R_E and 1.6 R_E, which are shown by two dotted lines. The follow-up monitoring program is most sensitive to planets within this region which is sometimes referred to as the "lensing zone."

where t_m is the time corresponding to the minimum impact parameter (or the maximum amplification).

From Eqs. 3.1 and 3.4, the mass of the lens can be expressed as

$$M = \frac{[tV_e c]^2}{4GD} \tag{3.6}$$

4. Planets as lenses

The light curve due to a binary lens, unlike the single lens, can be complex and can be very different from the mere superposition of two point lens light curves. In case of a double lens, the lens equation, which is a second order equation for a single lens, becomes two 5^{th} order equations (or one 5^{th} order equation in the complex plane; Witt & Mao, 1994). The most important new feature is the formation of caustics, where the amplification is infinite for a point source, but finite for a finite size source. When the source crosses a caustic, an extra pair of images forms or disappears. Useful treatments of microlensing due to a double lens include Schneider and Weiss (1986), Asada (2003), and Dominik (1999). There are several papers which specifically deal with the theoretical predictions of planetary signals on the microlensing light curve (e.g. Bolatto & Falco 1994; Bennett & Rhie 1996; Wambsganss 1996; Peale 1997; Dominik 1999). The signature of the planet can be seen, in most cases, as sharp extra peaks in the microlensing light curve. Computer codes for analysis of such data have been developed by Mao and Di Stifano (1995) and Dominik (1996).

It was first shown by Mao and Paczyński (1991) that about 10% of the lensing events should show the binary nature of the lens, and this effect is strong even if the companion is a planet. The problem of microlensing by a star with a planetary system towards the Galactic bulge was further investigated by Gould and Loeb (1992). They noted that,

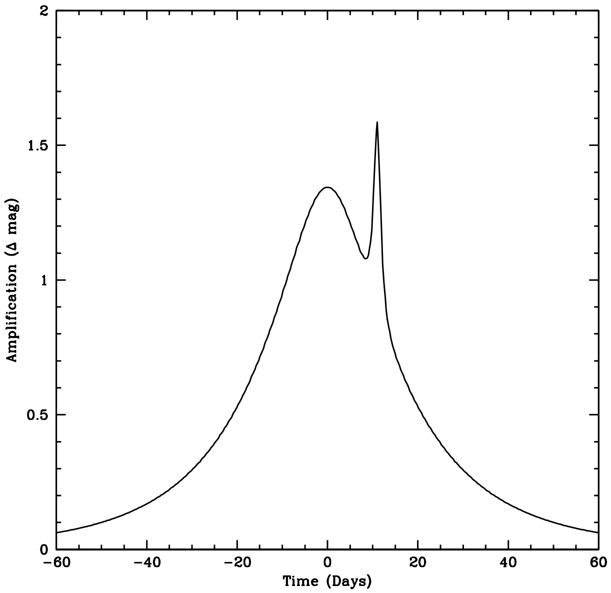

Figure 5. The figure shows the possible effect of a planet on the microlensing light curve. The mass of the planet is about one thousandth the mass of the primary, and is situated close to the Einstein ring. (adapted from Gould & Loeb 1992).

for a solar-like system half way between us and the Galactic bulge, Jupiter's orbital radius coincides with the Einstein ring radius of a solar-mass star. Such a case is termed 'resonant lensing' which increases the probability of detecting the planetary signal. In ∼20% of the cases, there would be a signature with magnification larger than 5%. The importance of the resonant lensing can be qualitatively understood as follows. In Fig. 2, the impact parameter changes through a large range as the source passes close to the lens. The positions of two images formed by the lensing effect change continuously, but they remain close to the Einstein ring for a large range of impact parameters. So, the effect of the planet can be large if the planet happens to be close to the Einstein ring, which causes a further amplification. This also qualitatively explains why the probability of observing the effect of the planet increases if it is close to the Einstein ring. The follow-up monitoring program is most sensitive to planets if the planet lies within about 0.6 R_E and 1.6 R_E, which is sometimes referred to as the "lensing zone" (Fig. 4).

In a large number of cases, however, the resulting light curve due to a planet-plus-star system is close to the superposition of two point lens light curves (Fig. 5). This is particularly true when the star-planet distance is much larger than R_E. In such a case, the timescale of the extra peak due to the planet, t_p, and the timescale of the primary peak due to the star, t_s, are related through the relation $t_p/t_s = \sqrt{(m_p/M_s)}$, where m_p is the mass of the planet and M_s is the mass of the star. Figure 6 shows the sizes of the Einstein ring radii due to planetary and stellar mass lenses as a function of distance to the lens. The typical sizes of the main-sequence and giant sources are also shown. In such a case, it is clear that the size of the source is almost always smaller than the Einstein ring radius of a Jupiter-mass planet; as a result the amplification due to the planet can be large. The amplification can also be large for an Earth-mass planet if the source is a main-sequence star. However, if the source is a giant-type star, then there is only a fixed range of D_d where the amplification due to an Earth-size planet can be large.

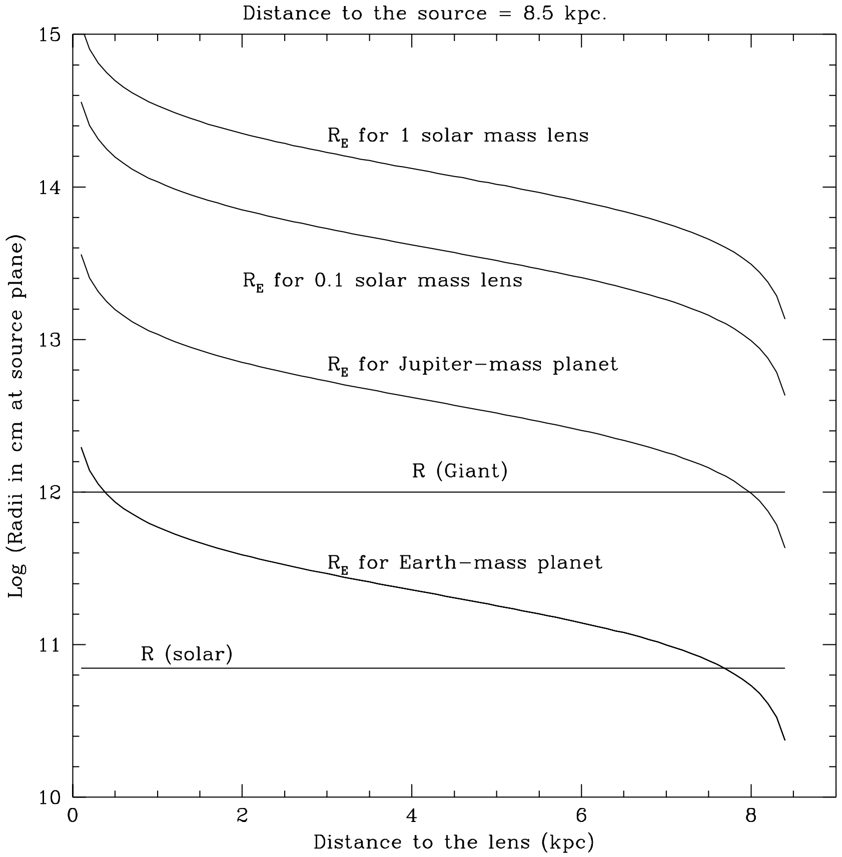

FIGURE 6. The figure shows the sizes of the Einstein ring radii R_E at the source plane, for a lensing event towards the Galactic bulge. R_E for an Earth-mass planet a solar mass star are shown. Also shown are the actual radii of a solar type star and a typical giant star as projected onto the source plane, which are denoted by R(solar) and R(giant) respectively. For a Jupiter-mass planet, R_E is almost always larger than the radius of a giant star, so the effect of Jupiter can always be significant. But for an Earth-mass planet, there is only a small parameter space where the Einstein ring radius is larger than the size of a giant star. If the source is a main-sequence star like our Sun, the Einstein ring radius due to an earth-mass planet is almost always larger than the source size, and hence the amplification can be large.

The minimum duration of the extra feature due to the planet, to a first approximation, is the time taken by the source to cross the caustic, which can be about 1.5 to 5 hrs. The maximum duration of the spike is roughly the time taken by the planet to cross its own Einstein ring. Using a reasonable set of parameters (the lower mass of the planet is taken as that of the Earth, the higher mass is assumed to be that of Jupiter) this can be a few hours to about three days. Any follow-up program must be accordingly adjusted so that the extra feature due to the planet is well sampled.

5. Requirements for a follow-up monitoring program

The first requirement for a follow-up monitoring program is access to the 'alert' events. With the alert capability of the survey programs firmly in place, it is now possible to carry out dedicated follow-up programs. At present, the alert events from OGLE, EROS and MOA collaborations at a given time are sufficient to carry out a ground-based follow-

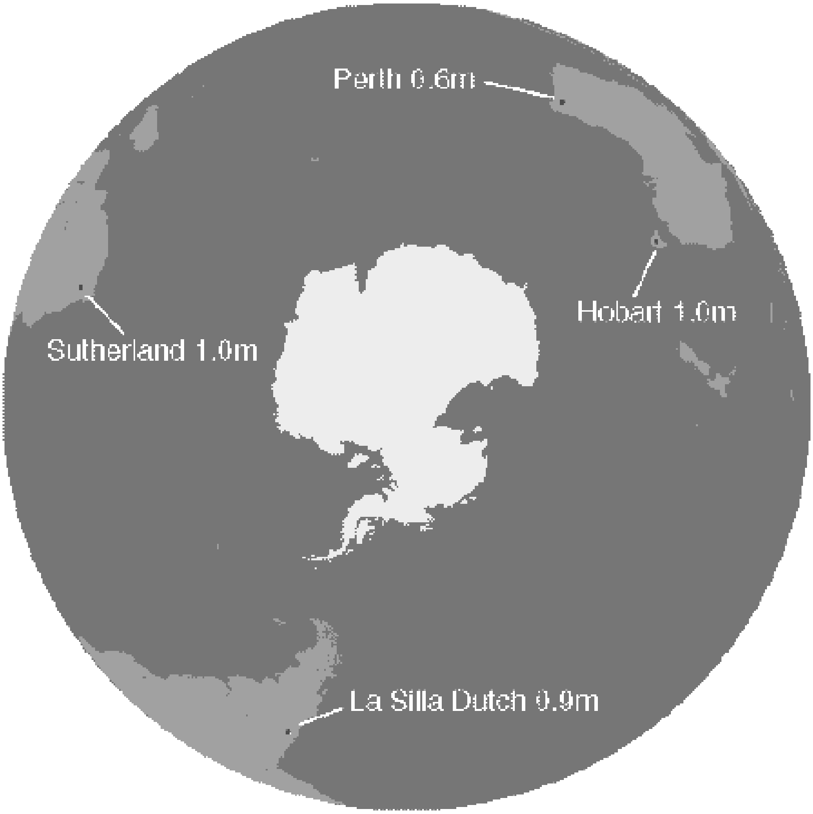

FIGURE 7. The network of telescopes used by PLANET, superposed on a south-pole centered view of the Earth. The telescopes situated at these longitudes enable PLANET to achieve a near-continuous coverage in the follow-up monitoring program.

up program towards the Galactic bulge using a network of relatively small ground-based telescopes. In the future, it may be possible to extend such follow-up networks to larger telescopes, and also perhaps towards the LMC where the many of lenses have been found to be stars within the LMC (Sahu 1994; Wu 1994; Sahu 2003).

The second requirement is the ability to monitor hourly. It should be noted that, assuming that the longer timescale events are mostly due to slower proper motion of the lensing star, the timescale of the planetary signal approximately scales with the timescale of the main event. So the monitoring, in general, can be less frequent for longer timescale events. But typically, as noted before, the timescales of the planetary event can be a few hours to a few days. The follow-up monitoring program must have the capability to do hourly monitoring so that the extra feature due to the planet is well sampled. For discrimination against any other short term variations, some color information is also useful, since the microlensing is expected to be achromatic, whereas most other types of variations are expected to have some chromaticity. Thus, it is preferable to have a few observations in two colors.

The third requirement is to have 24-hour coverage in the monitoring program. This calls for telescopes at appropriately spaced longitudes around the globe.

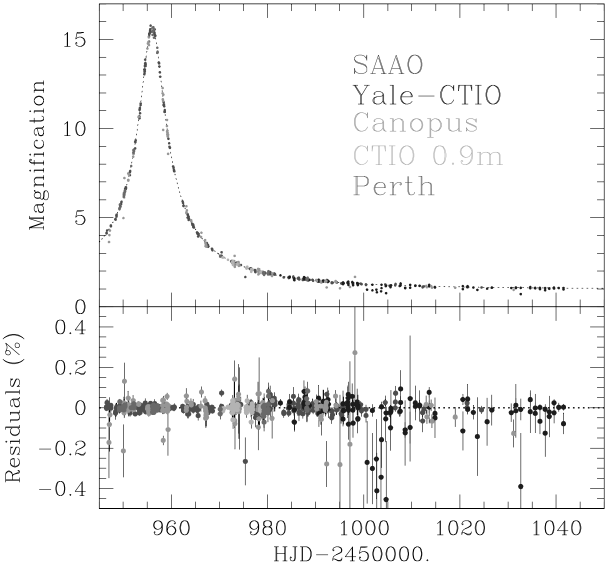

Figure 8. Top: PLANET data for OGLE-1998-BUL-14. The entire data set consists of 461 I-band and 139 V-band data points. The median sampling interval during this time span is about one hr, or 10^{-3} t$_E$, with no gaps greater than four days. The data are from the Yale-CTIO 1 m, South African Astronomical Observatory 1 m, the Perth 0.6 m, the Canopus 1 m, and the CTIO 0.9 m. Bottom: Residuals from the best-fit point-source-point-lens model. (Taken from Albrow et al. 2000)

6. Results from PLANET Collaboration

PLANET (Probing Lensing Anomalies NETwork), which is a world-wide collaboration of astronomers with access to a set of 1-m class telescopes situated in Chile, South Africa and Australia (Fig. 7), was established in 1995 with its main aim as looking for extra-solar planets through frequent monitoring of ongoing microlensing events. PLANET has intensely monitored more than 100 microlensing events, during which several binary events have been discovered. The data are reduced through a semi-automated pipeline in real time using the DoPhot photometric reduction package (e.g. Saha et al. 1997). If an anomaly is observed, the sampling frequency is appropriately increased in all sites so that the data can be used to better characterize the nature and cause of the anomaly. The exposure times are adjusted so that about 2% photometric accuracy is achieved in the monitoring program. Figure 8 shows an example lightcurve of OGLE-BULGE-1998-14 as observed by PLANET, which demonstrates its capability to carry out dense sampling, 24-hour coverage, and better than ∼2% photometric accuracy, which would be adequate to detect any planetary signals.

The intense monitoring of the ongoing microlensing events has led to many interesting scientific results. These include the determination of the lens location for a microlensing event towards the SMC (Albrow et al. 1999); limb darkening measurements of several K-giants (Albrow et al. 1999a, 2001), mass measurements of a lens (An et al. 2002), etc. However, no clear signature of a planet has been detected in any of the 100-odd microlensing events (Albrow et al. 2001a; Gaudi et al. 2002). This implies that less than 33% of the lensing stars have Jupiter-mass planets with orbital radii of 1.5–4 AU (Fig. 9). Since other techniques are currently not sensitive to the outer portion of these orbital radii, these are the best current limits on extra-solar planets at these orbital separations.

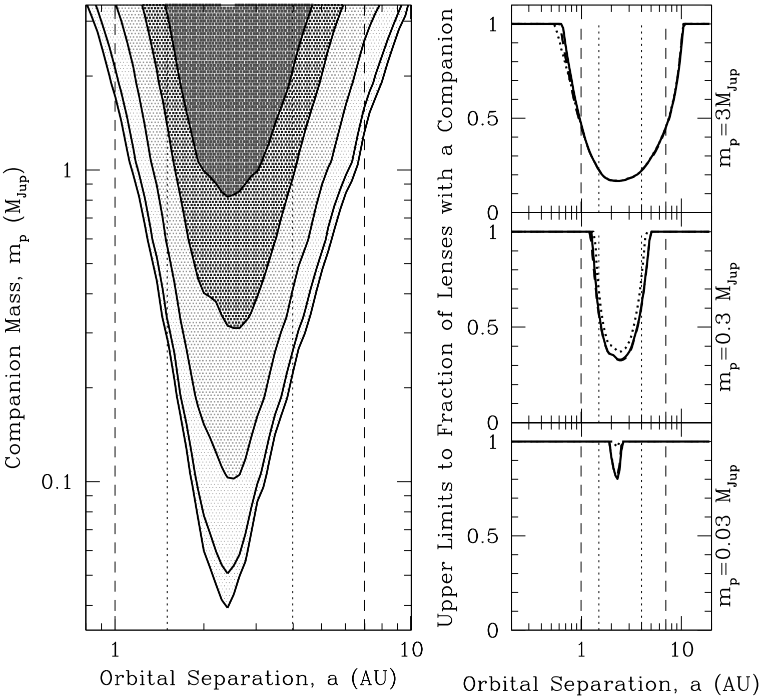

Figure 9. Left panel: Planet exclusion contours as a function of planet mass and orbital separation. The solid black lines show exclusion contours of 75%, 66%, 50%, 33% and 25% (outer to inner). Right panel: Horizontal cross-sections through the left panel. The curves show upper limits to the fraction of stars with planetary mass companions as a function of orbital separation.

The data obtained by the microlensing monitoring programs MOA and OGLE can also be used to look for planetary signals. Indeed, some of the observed data are consistent with planetary signatures (Jaroszyński & Paczyński 2002, Bond et al. 2003; Tsapras et al. 2003), although better sampling would be necessary for a robust detection, and to characterize any possible planetary companion.

7. Isolated planetary-mass objects

Isolated planetary-mass objects can reveal themselves as short timescale microlensing events in a monitoring program. The data from the microlensing networks can thus be used to get an estimate of abundance of such objects in different directions. Lack of such short timescale events in the MACHO and EROS database towards the Magellanic Clouds suggests that the contribution of planetary-mass objects is less than 10% of the halo dark matter (Fig. 10).

8. Future prospects

Although monitoring of microlensing events has so far produced mainly null results on the detection of extra-solar planets, microlensing promises to be a very efficient technique in the future. Among all the currently available techniques, microlensing is the only technique capable of detecting extra-solar planets at *any* orbital separation, although the probability of detection is a strong function of the orbital separation. Furthermore, microlensing is sensitive down to Mercury-mass planets, although, again, the probability

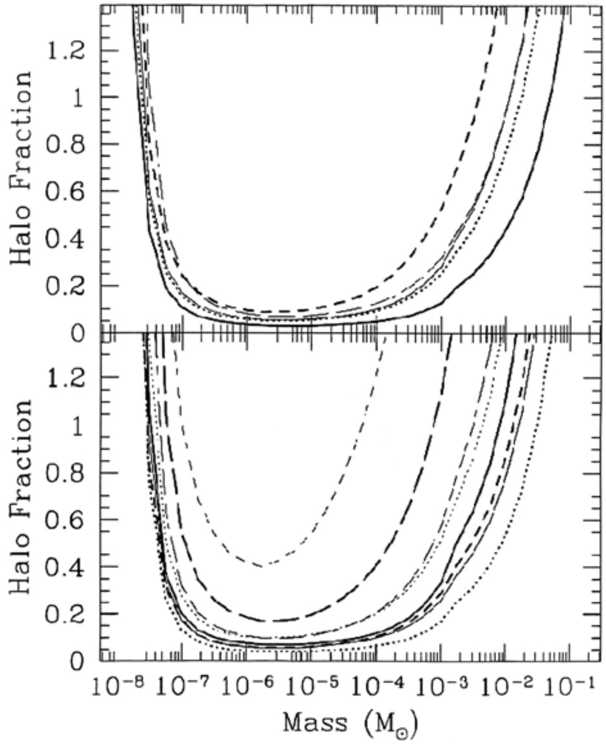

FIGURE 10. Halo fraction upper limit (95% c.l.) vs. lens mass for the five EROS models (top) and the eight MACHO models (bottom). (Taken from Alcock et al. 1998).

of detection decreases with the mass of the planet. And microlensing is the only currently available technique capable of detecting isolated planetary-mass objects. Larger telescopes with large fields of view, dedicated to microlensing, will greatly contribute to the success of such a project.

If a network of 4-meter class telescopes with large fields of view can be dedicated to microlensing observations, it would be possible to detect and monitor many microlensing events simultaneously. The same data used for discovering microlensing events can also serve the purpose of the follow-up program, and can be used to detect planetary signatures in the microlensing lightcurves. With such a network, it would be possible to detect and monitor about 200 microlensing events towards the Galactic bulge per year. Extending the current planet frequency statistics to large radii should lead to the detection of a few planetary-mass objects every year. With a few years' data, such a program would give a statistical estimate of the planetary frequencies as a function of their mass and orbital radius down to Mercury-mass planets. The project can be carried out even more efficiently with a single telescope in space where the superior image quality, continuous coverage in the monitoring program unaffected by the weather patterns of the ground, and the stable PSF unaffected by the Earth's atmosphere should enormously help in detecting even small-amplitude planetary signatures.

Finally, microlensing being a purely gravitational effect, the technique is, in principle, sensitive to detecting planetary-mass objects even at cosmological distances. Monitoring of lensed quasars, for example, should reveal the presence of planetary-mass objects in the lensing galaxy through short timescale microlensing events.

I would like to thank all the members of the PLANET collaboration, for their help and camaraderie during the years of persevering efforts on the follow-up monitoring of microlensing events.

REFERENCES

Albrow, M., et al. (PLANET Collaboration) 1999 *ApJ* **512**, 672.

Albrow, M., et al. (PLANET Collaboration) 1999a *ApJ* **522**, 1011.

Albrow, M., et al. (PLANET Collaboration) 2000 *ApJ* **535**, 176.

Albrow, M., et al. (PLANET Collaboration) 2001 *ApJ* **549**, 759.

Albrow, M., et al. (PLANET Collaboration) 2001a *ApJ* **556**, L113.

An, J. H., et al. 2002 *ApJ* **572**, 521.

Alcock, C., et al. 1993 *Nature* **365**, 621.

Alcock, C., et al. 1998 *ApJ* **499**, L9.

Asada, H. 2003 *A&A* **390**, L11.

Aubourg, E., et al. 1993 *Nature* **365**, 623.

Bennett D. P. & Rhie, S. 1996 *ApJ* **472**, 660.

Bolatto, A. D. & Falco, E. E. 1994 *ApJ* **436**, 112.

Bond, I., et al. 2002 *MNRAS* **333**, 71.

Dominik, M. 1996 *Ph.D. Thesis*, Univ. Dortmund.

Dominik, M. 1999 *A&A* **349**, 108.

Dreizler, S., Hauschildt, P. H., Kley, W., et al. 2003 *A&A* **402**, 791.

Einstein, A. 1936 *Science* **84**, 506.

Gaudi, B. S., et al. 2002 *ApJ* **566**, 463.

Gould, A. & Loeb, A. 1992 *ApJ* **396**, 104.

Griest, K. 1991 *ApJ* **366**, 412.

Jaroszyński, M. & Paczyński, B. 2002 *Acta Astron.* **52**, 361.

Konacki, M., Torres, G., Sasselov, D. D., & Jha, S. 2003a *ApJ* **597**, 1076.

Konacki, M., Torres, G., Jha, S., Sasselov, D. D. 2003b *Nature* **421**, 507.

Mao, S. & Paczyński, B. 1991 *ApJ* **374**, L37.

Mao, S. & Di Stifano, R. 1995 *ApJ* **440**, 22.

Marcy, G. W. & Butler, R. P. 1996 *ApJ* **464**, L153.

Marcy, G. W., et al. 2001 *ApJ* **556**, 296.

Mayor, M. & Queloz, D. A. 1995 *Nature* **378**, 355.

Paczyński, B. 1986 *ApJ* **304**, 1.

Paczyński, B. 1991 *ApJ* **371**, L63.

Paczyński, B. 1996 *ARA&A* **34**, 415

Peale, S. J. 1997 *Icarus* **127**, 269.

Petters, A. O., Levine, H., & Wambsganss, J. 2001. *Singularity Theory and Gravitational Lensing.* Birkhäuser.

Saha, A., et al. 1996 *ApJ* **466**, 55.

Sahu, K. C. 1994 *Nature* **370**, 275.

Sahu, K. C. 2003, in *The Dark Universe: Matter, Energy, and Gravity.* Cambridge University Press, in press (astro-ph/0302325).

Schneider, P., Ehlers, J., & Falco, E. E. 1992. *Gravitational Lensing.* Springer-Verlag.

Sasselov, D. D. 2003 *ApJ 596*, 1327.

Schneider, P. & Weiss, A. 1986 *A&A* **164**, 237.

Sirko, E. & Paczyński, B. 2003 *ApJ* **592**, 1217.

Tsapras, Y., Horne, K., Kane, S., & Carson, R. 2003 *MNRAS* **343**, 1131.

Udalski, A., et al. 1993 *Acta Astron.* **43**, 289.

Udalski, A. 2002 *Acta Astron.* **52**, 1.

Udalski, A. 2003 *Acta Astron.* **53**, 133.

Wambsganss, J. 1997 *MNRAS* **284**, 475.

Werner, K. & Wolff, B. 2003 *A&A* **402**, 791.

Witt, H. & Mao, S. 1994 *ApJ* **430**, 505.
Wolszczan, A. & Frail, D. A. 1992 *Nature* **355**, 145.
Wu, X.-P. 1994 *ApJ* **435**, 66.

The Galactic Habitable Zone

By GUILLERMO GONZALEZ

Department of Physics and Astronomy, Iowa State University, Ames, IA 50011, USA

Galactic scale phenomena relevant to life on a terrestrial planet are reviewed. The habitability of the Earth for complex life is surprisingly dependent on a diverse collection of processes ranging from Galactic chemical evolution to Galactic nuclear activity to comet impacts. The combined effect of these is to restrict the time and space that complex life can exist on a terrestrial planet. That region in the Milky Way is termed the Galactic Habitable Zone.

1. Introduction

The introduction of the Circumstellar Habitable Zone (CHZ) concept in the late 1950s (Huang 1959) and later refinements (Hart 1979; Kasting et al. 1993; Franck et al. 2000) have permitted the study of life in the universe to be systematized to some degree. However, discussion of habitability on the scale of the Milky Way Galaxy has received less attention. Trimble (1997a) considered habitability in the context of Galactic chemical evolution. Clarke (1981) discussed the possible effects on habitability of a Seyfert-like outburst in the Galactic center. In addition, many papers have been written about the possible threats to life by nearby supernovae (e.g. Ellis & Schramm 1995). While these studies have been helpful studies, they do not attempt to systematize the concept of habitability on the Galactic scale.

Before beginning any discussion about habitability, it is important to be up front about assumptions regarding life. As in CHZ studies, we assume Earth-like life in exploring Galactic-scale habitability constraints. This assumption is partly for convenience (we understand its chemistry fairly well) and partly because alternatives to carbon and water are extremely unlikely (see Chapter 8 in Barrow & Tipler 1986). We also take a terrestrial planet in the CHZ of its host star as the best type of habitat for complex life.

This review is not meant to be an exhaustive treatment of the Galactic Habitable Zone (GHZ), but it is intended as a systematized treatment of the subject. Hopefully, it will then serve as a framework to guide future research in this field. In the following, we present a general outline of the two broad categories of processes that go into defining the GHZ: 1) planetary building blocks, and 2) threats to existing complex life.

2. Building blocks

2.1. *Building terrestrial planets*

By building blocks we mean all the chemical elements and processes that go into the creation of a habitable planetary system.Gonzalez et al. 2001a discuss the mix of elements required to build a terrestrial planet like the Earth and organisms. While the bulk composition of the Earth is dominated by iron, oxygen, and silicon, carbon is only a trace element. The abundance of carbon is larger in the crust, where it is essential for organisms and climate regulation via the carbon-silicate cycle. The crust must contain a rich mix of elements given that simple life requires 16 elements and the most complex life requires about 26 (Davies & Koch 1991; Trimble 1997b). The Earth's crust is believed to have resulted from a late veneer from asteroid and comet impacts.

Because the terrestrial planets are composed almost exclusively of metals,† it is impossible to form one out of matter formed in the moments immediately following the Big Bang event. A few generations of stellar nucleosynthesis and mixing with the interstellar medium are required before a sufficient metal abundance is built up to make terrestrial planet comparable in size to the Earth. Thus, globular clusters and other metal-poor environments are unlikely to have Earth-size terrestrial planets. Gonzalez, Brownlee & Ward (2001a) assume that terrestrial planet mass scales with the local surface density of solids in a protoplanetary disk to the 1.5 power.

What might be called second-order effects concern the mix of elements that go into building a terrestrial planet. For example, the ratio of Fe to Si+Mg would determine the ratio of the mass of the iron core to the silicate mantle; this would, in turn, determine the type of convection in the planet. Mantle convection, and thus plate tectonics, is particularly sensitive to the concentration of the long-lived radioisotopes in a terrestrial planet (e.g. ^{40}K, ^{232}Th, ^{235}U, and ^{238}U). The average half-life of these isotopes is a few billion years, so this should also be the timescale for the weakening of convection in the Earth. One could argue that a smaller initial endowment of these radioisotopes could be compensated for with a larger planet. This is true, but planet size affects the habitability in several ways. Perhaps most importantly, a larger terrestrial planet would be more likely to be ocean-covered and therefore less likely to have continents. Such an environment is unlikely to support complex life.

2.2. *Giant planets and habitability*

While there is at present no observational data on terrestrial planets around other stars, there is a large and rapidly growing database of extrasolar giant planets. By July 2002, the number of giant planet candidates had topped the century mark. The sample displays a number of interesting trends. The one that most concerns us here is the high incidence of giant planets among metal-rich stars relative to field stars without any known giant planets (Gonzalez et al. 2001b; Santos et al. 2001). While the cause(s) of this trend is still under debate, a leading contender is that a higher initial metallicity makes it more likely that giant planets will form around a star (see review by Gonzalez 2002).

The final state of giant planets in a given system affects the overall habitability of that system in several ways. The giant planets in our Solar System "shield" the inner planets from comets (Wetherill 1994). Less massive giant planets provide less protection to terrestrial planets. In addition, giant planets that migrate or end up in high eccentricity orbits make it less likely that a system will contain habitable planets. Lineweaver (2001) has quantified the metallicity effects on habitable planet formation over the history of the universe, assuming the observed metallicity trend among giant planets is due to the phenomenon we noted above. Taken together, these various effects result in a finite range in metallicity over which habitable planetary systems can form. Thus, metal-rich stars may be predominantly accompanied by giant planets in terrestrial planet-disturbing orbits, while metal-poor stars may be accompanied by small terrestrial planets and small giant planets.

2.3. *Galactic chemical evolution*

Galactic chemical evolution is controlled by the infall of fresh hydrogen and helium onto the Galaxy, nucleosynthesis inside stars, and return of processed matter to the interstellar medium (ISM). Most of the synthesized elements are returned to the ISM via supernova explosions, but mass loss from intermediate mass stars contributes significantly to a

† Here, we are employing the astronomer's definition of metals—elements heavier than He.

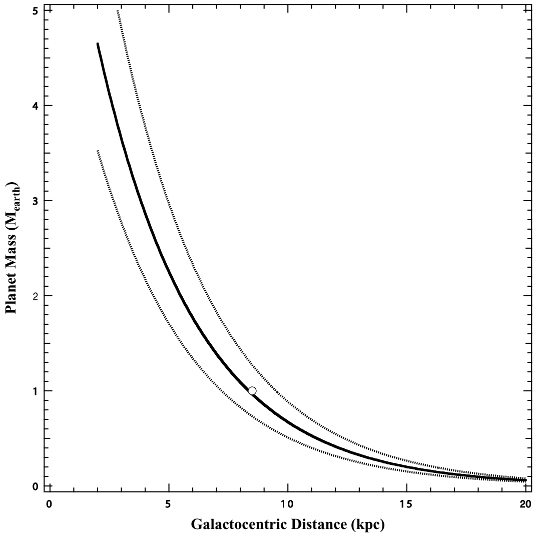

FIGURE 1. Mass of terrestrial planets forming at present as a function of distance from the Galactic center. The dotted curves show the one-sigma dispersion in masses expected from the primordial spread in metallicity at a given Galactocentric distance.

few light elements (e.g. C and N). The two primary types of supernovae are core-collapse massive stars (Type II) and nuclear explosions of degenerate stars (Type Ia). The Type Ia supernovae are considered to produce mostly Fe-peak elements, while Type II supernovae produce relatively more O (and Mg and Si) and r-process elements (Timmes et al. 1995; Samland 1998).

While Galactic chemical evolution models are best constrained from spectroscopic observations of stars in the solar neighborhood, observations of more distant regions show that it varies with location. The halo contains only metal-poor stars; star formation there has not continued into the present. Star formation started early in the bulge and continues today. Among the bulge stars, the relative abundances of O, Mg, Si, and Ti to Fe are observed to be greater for stars of solar metallicity compared to the solar neighborhood. This is interpreted as resulting from different star formation histories for the two regions. The particular composition of the bulge stars is consistent with a greater input from Type II relative to Type Ia supernovae as compared to the solar neighborhood.

Even the thin disk (where the Sun resides) displays variations. Most significantly, there is a radial metallicity gradient in the thin disk amounting to -0.07 dex kpc^{-1} (see Maciel 2002). The Sun's metallicity is very close to that of the local interstellar medium, but it is significantly greater than that of nearby stars comparable in age to the Sun. There is also an age-metallicity trend among thin disk stars amounting to about 0.035 dex Gyr^{-1} (Gonzalez 1999b). Figure 1 shows the expected variation in terrestrial planet mass with distance from the Galactic center. The calculation assumes that the mass of a terrestrial planet scales with metallicity to the 1.5 power, and a star with the metallicity of the Sun is accompanied by an Earth-mass terrestrial planet.

Observations of abundance patterns in nearby stars indicate that Type II supernova are declining in number relative to Type Ia supernovae (see, for example, Figure 39 of Timmes et al. 1995). This means that the ratios of Mg+Si to Fe and r-process elements to Fe are declining in the ISM. Thus, the relative abundances of the long-lived radioisotopes important for heating the interior of a terrestrial planet decline with time. An Earth-size planet forming today will generate less internal heat in 4.5 Gyr than the Earth does

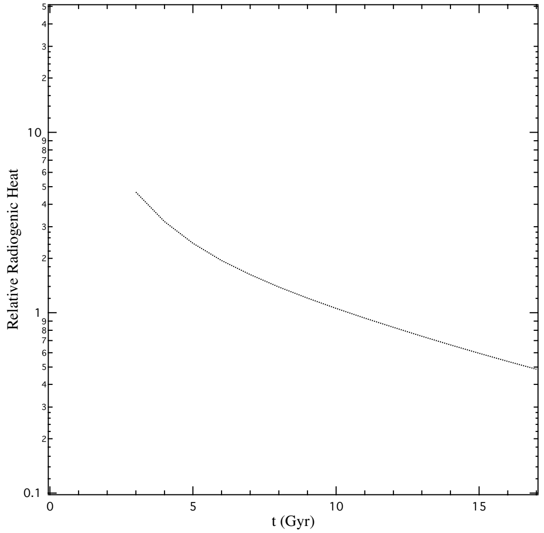

FIGURE 2. Radiogenic heat in a terrestrial planet 4.5 Gyrs after its formation relative to the present radiogenic heating in the present Earth as a function of formation time. The vertical offset has been adjusted so that the curve passes through unity at t = 10.5 Gyr (adapted from Gonzalez et al. 2001a).

today (Gonzalez et al. 2001a). Figure 2 shows the evolution of the relative radiogenic heating from radioactive decay in a terrestrial planet the mass of the Earth as a function of formation time. The calculation assumes the Earth's initial allotment of the four geologically-important isotopes listed above was typical. However, the best observations available at this time indicate that the Solar System might have a larger allotment (Gonzalez et al. 2001a).

While abundance ratios (e.g. Si/Fe) tend to display little scatter at a given age and Galactocentric distance, the "absolute" abundances (e.g. Fe/H) display considerable scatter. There are two sources for the scatter, primordial and "radial mixing." Stars forming at the same Galactocentric distance and time are not all born with the same metallicity. Observations indicate that the primordial scatter in the metallicity in the solar neighborhood is somewhat less than 0.1 dex (Gonzalez 1999b); Gaidos & Gonzalez 2002). Once born, stars wander in the disk via gravitational perturbations by spiral arms and GMCs, changing their mean Galactocentric distances. This stellar diffusion, combined with the radial metallicity gradient, results in increasing scatter with age (Wielen et al. 1996).

3. Threats to complex life

3.1. *Transient radiation events*

Threats to existing complex life on a terrestrial planet come in many forms, but Galactic threats can be put in one of two broad categories: 1) transient radiation events, and 2) comet impacts. The radiation events come in several forms: supernovae, Active Galactic Nucleus (AGN) outbursts, and gamma ray bursts (GRBs). These events are not uniformly distributed in the Milky Way. Type II supernovae occur mostly in the spiral arms and other regions with ongoing star formation (such as the Galactic center). Type Ia supernovae, being older than Type II supernovae, are more randomly distributed in the Milky Way. Both types of supernovae are concentrated towards the Galactic cen-

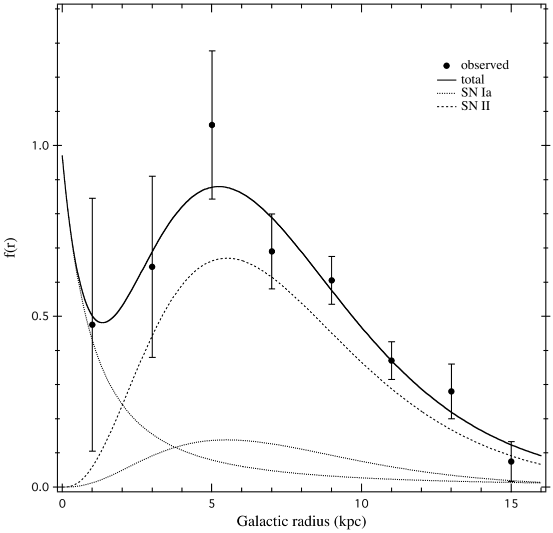

FIGURE 3. Observed radial distribution of surface density of supernova remnants from Case & Bhattachrya (1998). The dashed and dotted lines are the Type II and Type Ia supernovae, respectively. The Type II supernovae follow the mass distribution of the disk (with a hole at the center), and the Type Ia supernovae follow the bulge and disk mass distributions.

ter. Figure 3 shows the observed radial distribution of supernova remnants in the Milky Way.

Very few supernovae occur in the halo. Assuming GRBs are due to the core collapse of very massive stars, then their distribution in the Milky Way should be similar to that of Type II supernovae. However, the radiation GRBs is probably highly beamed, requiring careful attention to geometry of the targets. While AGN outbursts are restricted to the nucleus of the Milky Way, the cosmic ray particles they emit travel along the Galaxy's magnetic field lines and fill most of its volume (Clarke 1981).

Each of these sources of radiation has a distinct time evolution. AGN outbursts are expected to be long-lived, lasting perhaps a million years on average. Supernovae are bright only for a few months, but their remnants maintain a high local radiation level for several thousand years. GRBs are the shortest-lived events, lasting only a few minutes.

There has been much debate in the literature on the specific ways that radiation from supernovae and GRBs interacts with a planetary environment. One type of threat results from damage to the ozone layer from formation of nitrous oxides in the stratosphere (see Crutzen & Bruhl and references cited therein). More direct threats result from the secondary cosmic rays that reach the surface of a planet (Dar et al. 1998) and neutrinos from core collapse events (Collar 1996; Collar 1997). Even for very short duration radiation events, damage to the ozone layer and pollution of the environment by radioactive nuclei produced by shower particles on one side of a planet will quickly have global effects.

Benitez et al. (2002) have presented evidence for a minor recent extinction event possibly caused by supernovae in the Scorpius-Centaurus OB association. If confirmed with additional paleontological data, this could become a useful empirical calibrator of the biological effects of a supernova of known distance.

3.2. *Comet impacts*

Assuming planetary systems like ours are accompanied by Oort clouds, then their terrestrial planets should also suffer from comet impacts. Because they are weakly bound

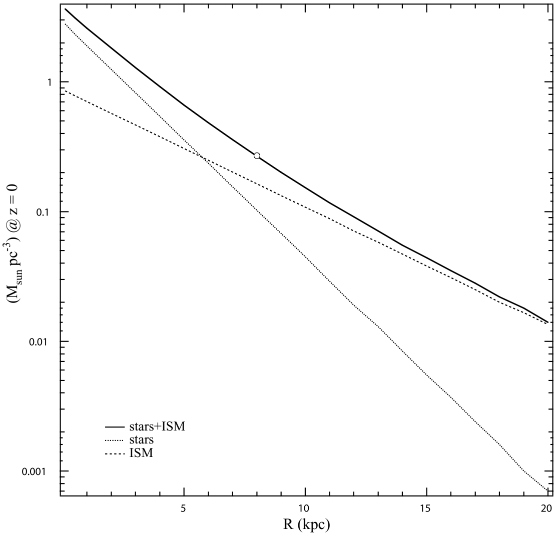

FIGURE 4. Mass density (M$_\odot$ pc^{-3}) of stars (dotted) and the ISM (dashed) in mid-plane of Milky Way. The Sun's location is shown as an open circle. Notice the steep increase in the star density towards the Galactic center.

to their host star, comets in an Oort cloud can be perturbed by passing stars (Matese & Lissauer 2002), nearby Giant Molecular Clouds (GMCs; Hut & Tremaine 1985), and Galactic tides (Matese et al. 2001). All three types of perturbations are more important closer to the Galactic center. Of course, a greater rate of perturbation of the outer Oort cloud comets will deplete this reservoir more quickly, but, at the same time, the inner Oort cloud will also be perturbed more often and replenish the outer cloud. For much denser stellar environments, the Kuiper Belt comets will also be perturbed. Figure 4 shows the mid-plane variation of mass density with distance from the Galactic center.

The ratio of the number of comet to asteroid impacts on the Earth is still an unsettled issue. If comet impacts contribute less than 10%, then changes to the comet flux through the inner Solar System will have to be relatively large to have a significant influence on an inhabited planet.

To date, the only widely-accepted evidence for a major extinction caused by an impact remains the Cretacious/Tertiary event 65 Myrs ago. However, the source object of this event remains uncertain. Some have claimed that the cratering record contains a periodicity near 30 Myrs, which is interpreted as resulting from the variation of the Galactic tides from the Solar System's motion perpendicular to the plane (Clube & Napier 1986). The faunal extinction periodicity is claimed to be about 26 Myrs. However, recent analyses the cratering record indicate a period close to 37.5 Myrs, which is not statistically significant (Yabushita 2002) but clearly different from the extinction periodicity.

4. Putting it all together

4.1. *The GHZ*

Combining the two main types of Galactic-scale constraints on complex life given above, we arrive at a more complete picture of the GHZ. Its inner boundary is set primarily by threats to existing complex life, while its outer boundary is set by the minimum requirements to build a large terrestrial planet. Of course, the Milky Way's disk is not homogeneous, the motions of the stars within it are not perfect circles, and not all the stars formed at a given Galactocentric distance and time have the same composition. Several of the relevant factors are still too poorly constrained to establish its precise boundaries, but we can paint a crude picture of the GHZ. Basically, it forms an annulus of optimized habitability in the thin disk, and it has “fuzzy” boundaries.

4.2. *Other factors*

The spiral arms form the most obvious deviation from a simple homogeneous disk. Vallée (2002) has reconstructed the spiral structure of the Milky Way based on the most recent observational studies of its large-scale structure. He shows that the Sun is about half-way between two major arms, Perseus and Sagittarius. However, local arms are known to have spurs and others fragments that intrude into the interarm region. The largest nearby one is the Orion spur. Additional details of the environment immediately surrounding the Sun and its possible consequences for the terrestrial environment are given by Zank & Frisch (1999). Still, the probability of encounters with nearby Type II supernovae and GMCs is far greater inside the major arms. However, the star density is only a few percent greater in the arms, so comet showers resulting from nearby star passages will not be significantly greater there.

Because the scale-height of GMCs and massive stars is small, stars with large vertical oscillation amplitudes will spend relatively less time near the mid-plane, where they could encounter them. On the down side, stars with large vertical oscillation amplitudes will pass through the mid-plane at high velocities. This could prove damaging to a planet’s atmosphere if the plane-crossing occurs at a region of relatively high dust content.

Motion within the plane of the disk is also important. Of particular relevance is the location of the Sun relative to the corotation circle (i.e., the radius wherein the orbital period of stars equals that of the spiral arm pattern). Balázs (1988) and Marochnik (1984) have argued that if the Sun is near the corotation circle, then the transit time between spiral arm crossings will be longer than stars farther from the corotation circle. Of course, the transit time through spiral arms will also be longer. Mishurov & Zenina (1999), using observations of Classical Cepheids, determined that the Sun is very near the corotation circle. Finally, the eccentricity of a star in the disk must be small if it is to avoid frequent spiral arm crossings, even if is at the corotation circle. If proximity to the corotation circle is indeed important for habitability, then the GHZ could be much narrower than implied by the main factors given above.

4.3. *Guidance from the Weak Anthropic Principle*

The fact that some of the Sun’s properties are atypical relative to some suitably chosen nearby star comparison sample may be taken, at least in part, as evidence that they are critical for habitability (see Gustafsson 1998; Gonzalez 1999a; Gonzalez 1999b). Thus, application of the Weak Anthropic Principle (WAP) may at least give some direction to research on habitability. There are some questions that may lend themselves to this kind of application of the WAP: “Why is the Sun among the most massive $\sim$8% stars?”; “Why is the Sun so close to the Galactic mid-plane?”; “Why is the Sun’s Galactic orbit more nearly circular than most nearby stars of similar age?”; Why is the Sun’s metallicity significantly higher than the mean of nearby stars of similar age?”; “Why is the Sun so close to the corotation circle?”.

5. Conclusions

The fact that other regions of the Galaxy is different from the solar neighborhood should provide motivation for us to consider how habitability for complex life may vary with place and time. Upon considering Galactic chemical evolution and the distribution of transient radiation events, we arrive at the concept of the Galactic Habitable Zone. The GHZ has the shape of an annulus in the disk and is broken by the major spiral arms. The GHZ probably first appeared at least 3 or 4 Gyrs after the Milky Way's birth, at first narrow and then gradually expanding and probably migrating outward. Eventually, the GHZ will fade as the supply of the geologically-important radioisotopes dwindles.

There is much need for additional research on phenomena relevant to the GHZ concept. In particular, the metallicity dependence of terrestrial planet formation needs to receive more attention from theorists. The *Kepler Planet Transit Telescope* promises to produce observations of terrestrial planets around nearby stars. A spectroscopic survey of stars with terrestrial planets should allow us to determine the precise functional dependence of terrestrial planet mass on metallicity. There is also considerable uncertainty concerning the dynamical relationships among orbits of individual stars, their locations relative to the corotation circle and spiral arms, and possible evolution of the spiral arm pattern. More research is also required on Galactic chemical evolution. Does the radial metallicity gradient change significantly over time?

The GHZ concept can also be extended to other galaxies. Is a galaxy's Hubble type relevant to its habitability? An elliptical galaxy lacks a "cold" component like the Milky Way's thin disk, which means that most of its stars will periodically visit its dense inner regions. Likewise, the disordered orbits of stars in many irregular galaxies makes it less likely that they contain "safe zones." Given the known correlation between metallicity and luminosity among galaxies of all types (Pilyugin & Ferrini 2000), it is likely that most galaxies less luminous than the Milky Way will lack sufficient metals to build Earth-size planets. Research is also needed on the relationship between habitability and cluster membership. Is membership in a small group to be preferred over membership in a large cluster, like Virgo? Finally, the GHZ concept can be extended to the broader universe, given that other large galaxies have likely experienced chemical evolution histories similar to that of the Milky Way. The primary determinant on the evolution of the overall habitability of the universe is the star formation rate.

REFERENCES

Balázs, B. S. 1988, in *Bioastronomy–The Next Steps* (ed. G. Max). p. 61. Kluwer.

Barrow, J. D. & Tipler, F. J. 1986 *The Anthropic Cosmological Principle* pp. 510–575. Oxford.

Benitez, N., Maiz-Apellaniz, J. & Canelles, M. 2002 *Phys. Rev. Lett.* **88**, 081101.

Case, G. L. & Bhattahrya, D. 1998 *ApJ* **504**, 761.

Clarke, J. N. 1981 *Icarus* **46**, 94.

Clube, S. V. M. & Napier, W. M. 1986 Galactic dark matter and terrestrial periodicities. *Quart. J. R. Astron. Soc.* **37**, 617–642.

Collar, J. I. 1996 Biological effects of stellar collapse neutrinos. *Phys. Rev. Lett.* **76**, 999.

Collar, J. I. 1997 *Phys. Rev. Lett.* **78**, 1395.

Crutzen, P. J. & Bruhl, C. 1996 *Pub. Nat. Acad. Sci.* **93**, 1582.

Dar, A., Laor, A. & Shaviv, N. J. 1998 *Phys. Rev. Lett.* **80**, 5813.

Davies, R. E. & Koch, R. H. 1991 *Phi. Trans. R. Soc. Lon. B* **334**, 391.

Ellis, J. & Schramm, D. N. 1995 *Proc. Nat. Acad. Sci.* **92**, 235.

Franck, S., Block, A., von Bloh, W., Bounama, C., Steffen, M., Schonberner, D., & Schellnhuber, H. J. 2000 *J. Geophys. Res.* **105**, 1651.

Gaidos, E. J. & Gonzalez, G. 2002 *New Astron.* **7**, 211.

Gonzalez, G. 1999a *Astron. Geophys.* **40**, 5.25.

Gonzalez, G. 1999b *MNRAS* **308**, 447.

Gonzalez, G. 2003 *Rev. Mod. Phys.*, **75**, 100.

Gonzalez, G., Brownlee, D., & Ward, P. 2001a *Icarus* **152**, 185.

Gonzalez, G., Laws, C., Tyagi, S., & Reddy, B. E. 2001b *AJ* **121**, 432.

Gustafsson, B. 1998 *Space Sci. Rev.* **85**, 419.

Hart, M. H. 1979 *Icarus* **37**, 351.

Huang, S.-S. 1959 *Am. Sci.* **47**, 397.

Hut, P. & Tremaine, S. 1985 *AJ* **90**, 1548.

Kasting, J. F., Whitmire, D. P., & Reynolds, R. T. 1993 *Am. Sci.* **47**, 397.

Lineweaver, C. H. 2001 *Icarus* **151**, 307.

Maciel, W. J. 2002 *Rev. Mex. Astron. Astrophys.* **12**, 207.

Marochnik, L. S. 1984 *Astrophys.* **19**, 278.

Matese, J., Iannanen, K. A., & Valtonen, M. J. 2001, in *Collisional Processes in the Solar System* (eds. M. Y. Marov & H. Rickman). Astrophysics and Space Science Library, vol. 261, p. 91. Kluwer.

Matese, J. J. & Lissauer, J. J. 2002 *Icarus* **157**, 228.

Mishurov, Y. N. & Zenina, I. A. 1999 *A&A* **341**, 81.

Pilyugin, L. S. & Ferrini, F. 2000 *A&A* **358**, 72.

Samland, M. 1998 *Astrophys. J.* **496**, 155–171.

Santos, N. C., Israelian, G., & Mayor, M. 2001 *A&A* **373**, 1019.

Timmes, F. X., Woosley, S. E., & Weaver, T. A. 1995 *ApJS* **98**, 617.

Trimble, V. 1997a, in *Extraterrestrials–Where are they?* (eds. B. Zuckerman & M. H. Hart), p. 184. Cambridge University Press.

Trimble, V. 1997b *Orig. Life Evol. Bios.* **27**, 3.

Vallée, J. P. 2002 *ApJ* **566**, 261.

Wetherill, G. W. 1994 *Astrophys. Space Sci.* **212**, 23.

Wielen, R., Fuchs, B., & Dettbarn, C. 1996 *A&A* **314**, 438.

Yabushita, S. 2002 *MNRAS* **334**, 369.

Zank, G. P. & Frisch, P. C. 1999 *ApJ* **518**, 965.

Cosmology and life

By MARIO LIVIO

Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, MD 21218, USA

I examine some recent findings in cosmology and their potential implications for the emergence of life in the Universe. In particular, I discuss the requirements for carbon-based life, anthropic considerations with respect to the nature of dark energy, the possibility of time-varying constants of nature, and the question of the rarity of intelligent life.

1. Introduction

The progress in cosmology in the past few decades leads also to new insights into the global question of the emergence of intelligent life in the Universe. Here I am not referring to discoveries that are related to very localized regions, such as the detection of extrasolar planetary systems, but rather to properties of the Universe at large.

In order to set the stage properly for the topics to follow, I would like to start with four observations with which essentially all astronomers agree. These four observations *define* the cosmological context of our Universe.

(i) Ever since the observations of Vesto Slipher in 1912–1922 (Slipher 1917) and Hubble (1929), we know that the spectra of distant galaxies are redshifted.

(ii) Observations with the *Cosmic Background Explorer* (*COBE*) have shown that, to a precision of better than 10^{-4}, the cosmic microwave background (CMB) is *thermal*, at a temperature of 2.73 K (Mather et al. 1994).

(iii) Light elements, such as deuterium and helium, have been synthesized in a high-temperature phase in the past (e.g. Gamow 1946; Alpher, Bethe, & Gamow 1948; Hoyle & Tayler 1964; Peebles 1966; Wagoner, Fowler, & Hoyle 1967).

(iv) Deep observations, such as the Hubble Deep Field, have shown that galaxies in the distant Universe look younger. Namely, their sizes are smaller (e.g. Roche et al. 1996), and there is a higher fraction of irregular morphologies (e.g. Abraham et al. 1996). This is what one would expect from a higher rate of interactions, and from "building blocks" of today's galaxies.

When the above four observational facts are combined and considered together, there is no escape from the conclusion that our Universe is *expanding and cooling.* This conclusion is entirely *consistent* with the "hot big bang" model. Sometimes the stronger statement, that these observations "prove" that there was a hot big bang, is made. However, the scientific method does not truly produce "proofs."

During the past decade, deep observations with a variety of ground-based and space-based observatories have advanced our understanding of the history of the Universe far beyond the mere statement that a big bang had occurred (see, e.g. the determination of cosmological parameters by the *Wilkinson Microwave Anisotropy Probe (WMAP)*; Spergel et al. 2003). In particular, remarkable progress has been achieved in the understanding of the cosmic star formation history.

Using different observational tracers (e.g. the UV luminosity density) of star formation in high-redshift galaxies, tentative plots for the star formation rate (SFR) as a function of redshift have been produced (e.g. Lilly et al. 1996; Madau et al. 1996; Steidel et al. 1999). There is little doubt that the SFR rises from the present to about $z \approx 1$. What happens in the redshift range $z \approx 1$–6 is still somewhat controversial. While some studies suggest that the SFR reaches a peak at $z \approx 1$–2 and then declines slightly toward higher redshifts (e.g. Steidel et al. 1999; or maybe even more than slightly, Stanway, Bunker, & McMahon 2003), or stays fairly flat up to $z \approx 6$ (e.g. Calzetti & Heckman 1999; Pei, Fall, & Hauser 1999; Giavalisco et al. 2004), others claim that the SFR continues to rise to $z \approx 8$ (Lanzetta et al. 2002). The latter claim is based on the suggestion that previous studies had failed to account for surface brightness dimming effects. For my present purposes, however, it is sufficient that the history of the *global* SFR is on the verge of being determined (if it has not been determined already). A knowledge of the SFR as a function of redshift allows for the first time for meaningful constraints to be placed on the global emergence of carbon-based life.

2. Remarks about carbon-based life

The main contributors of carbon to the interstellar medium are intermediate-mass (1–8 $M_\odot$) stars (e.g. Wood 1981; Yungelson, Tutukov, & Livio 1993; Timmes, Woosley, & Weaver 1995), through the asymptotic giant branch and planetary nebulae phases. A knowledge of the cosmic SFR history, together with a knowledge of the initial mass function (presently still uncertain for high redshift), therefore allows for an approximate calculation of the rate of carbon production as a function of redshift (Livio 1999). For a peaked SFR, of the type obtained by Madau et al. (1996), for example, the peak in the carbon production rate is somewhat delayed (by $\lesssim 1$ billion years) with respect to the SFR peak. The decline in the carbon production rate is also shallower for $z \lesssim 1$ (than the decline in the SFR), owing to the buildup of a stellar reservoir in the earlier epochs.

Assuming a "principle of mediocrity," one would expect the emergence of most carbon-based life in the Universe to be perhaps not too far from the peak in the carbon production rate—around $z \approx 1$ (for a peak in the SFR at $z \approx 1$–2). Since the time scale required to develop intelligent civilizations may be within a factor of 2 of the lifetime of F5 to mid-K stars (the ones possessing continuously habitable zones; Kasting, Whitmore, & Reynolds 1993; and see § 1.5 below), it can be expected that intelligent civilizations have emerged when the Universe was $\gtrsim 10$ Gyr old. A younger emergence age may be obtained if the SFR does not decline at redshifts $1.2 \lesssim z \lesssim 8$ (e.g. Lanzetta et al. 2002).

Carbon features in most anthropic arguments. In particular, it is often argued that the existence of an excited state of the carbon nucleus (the 0_2^+ state) is a manifestation of fine-tuning of the constants of nature, which allowed for the appearance of carbon-based life.

Carbon is formed through the triple-α process in two steps. In the first, two α particles form the unstable (lifetime $\sim 10^{-16}$ s) ^{8}Be. In the second, a third α particle is captured, via $^8\mathrm{Be}(\alpha,\gamma)^{12}\mathrm{C}$. Hoyle argued that in order for the 3α reaction to proceed at a rate sufficient to produce the observed cosmic carbon, a resonant level must exist in ^{12}C, a few hundred keV above the ^{8}Be + ^{4}He threshold. Such a level was indeed found experimentally (Dunbar et al. 1953; Hoyle, Dunbar, & Wenzel 1953; Cook, Fowler, & Lauritsen 1957).

The question of how fine-tuned this level needs to be for the existence of carbon-based life has been the subject of considerable research. The most recent work on this topic was done by Oberhummer and collaborators (e.g. Oberhummer, Csótó, & Schlattl 2000; Csótó, Oberhummer, & Schlattl 2001; Schlattl et al. 2003). These authors used a model that treats the ^{12}C nucleus as a system of 12 interacting nucleons, with the approximate resonant reaction rate

$$r_{3\alpha} = 3^{3/2} N_\alpha^3 \left(\frac{2\pi\hbar^2}{M_\alpha k_B T}\right)^3 \frac{\Gamma_\gamma}{\hbar} \exp\left(-\frac{\varepsilon}{k_B T}\right) \quad . \tag{2.1}$$

Here M_α and N_α are the mass and number density of α particles, respectively, ε is the resonance energy (in the center-of-mass frame), Γ_γ is the relative width, and all other symbols have their usual meaning. Oberhummer et al. introduced small variations in the strengths of the nucleon-nucleon interaction and in the fine structure constant (affecting ε and Γ_γ), and calculated stellar models using the modified rates. In their initial work, Oberhummer et al. (2000) concluded that a change of more than 0.5% in the strength of the strong interaction or more than 4% in the strength of the electromagnetic interaction would result in essentially no production of carbon or oxygen [considering the $^{12}\mathrm{C}(\alpha,\gamma)^{16}\mathrm{O}$ and $^{16}\mathrm{O}(\alpha,\gamma)^{20}\mathrm{Ne}$ reactions] in any star. More specifically, a decrease in the strong-interaction strength by 0.5%, coupled with an increase in the fine structure constant by 4%, resulted in a decrease in the carbon production by a factor of a few tens in 20 $\mathrm{M}_\odot$ stars, and by a factor of $\sim$100 in 1.3 $\mathrm{M}_\odot$ stars. Taken at face value, this seemed to support anthropic claims for extreme fine-tuning necessary for the emergence of carbon-based life.

Earlier calculations by Livio et al. (1989) indicated less impressive fine-tuning. Livio et al. showed that shifting (artificially) the energy of the carbon resonant state by up to 0.06 MeV does not result in a significant reduction in the production of carbon. Since this 0.06 MeV should be compared to the *difference* between the resonance energy in ^{12}C and the 3α threshold (calculated with the basic nucleon-nucleon interaction), it was not obvious that a particularly fantastic fine-tuning was required. Most recently, however, Schlattl et al. (2003) reinvestigated the dependence of carbon and oxygen production in stars on the 3α rate. These authors found that following the entire stellar evolution was crucial. They concluded that in massive stars the C and O production strongly depends on the initial mass. In intermediate- and low-mass stars, Schlattl et al. found that the high carbon production during He shell flashes leads to a *lower* sensitivity of the C and O production on the 3α rate than inferred by Oberhummer et al. (2000). Schlattl et al. (2003) concluded by saying that "fine-tuning with respect to the obtained carbon and oxygen abundance is more complicated and far less spectacular" than that found by Oberhummer et al. (2000).

3. Dark energy and life

In 1998, two teams of astronomers, working independently, presented evidence that the expansion of the Universe is accelerating (Riess et al. 1998; Perlmutter et al. 1999). The evidence was based primarily on the unexpected faintness (by $\sim$0.25 mag) of distant ($z \approx 0.5$) Type Ia supernovae, compared to their expected brightness in a universe decelerating under its own gravity. The results favored values of $\Omega_m \approx 0.3$ and $\Omega_\Lambda \approx 0.7$ for the matter and "dark energy" density parameters, respectively. Subsequent observations of the supernova SN 1997ff, at the redshift of $z \simeq 1.7$, strengthened the conclusion of an accelerating Universe (Riess et al. 2001). This supernova appeared *brighter* relative to SNe in a coasting universe, as expected from the fact that at $z \approx 1.7$ a universe with $\Omega_m \approx 0.3$ and $\Omega_\Lambda \approx 0.7$ would still be in its decelerating phase. The observations of SN 1997ff do not support any alternative interpretation (such as dust extinction or evolutionary effects) in which supernovae are expected to dim monotonically with redshift. Measurements of the power spectrum of the cosmic microwave background (e.g. Abroe et al. 2002; de Bernardis et al. 2002; and Netterfield et al. 2002; and, most recently the *WMAP* results, Bennett et al. 2003) provide strong evidence for flatness ($\Omega_m + \Omega_\Lambda = 1$). When combined with estimates of Ω_m based on mass-to-light ratios, X-ray temperatures of intracluster gas, and dynamics of clusters (all of which give $\Omega_m \lesssim 0.3$; e.g. Strauss &

Willick 1995; Carlberg et al. 1996; Bahcall et al. 2000), again a value of $\Omega_\Lambda \approx 0.7$ is obtained (Spergel et al. 2003).

Arguably the two greatest puzzles physics is facing today are:

(1) Why is the dark energy (vacuum energy) density, ρ_v, so small, but not zero? (Or, why does the vacuum energy gravitate so little?)

(2) Why *now*? Namely, why do we find at present that $\Omega_\Lambda \approx \Omega_m$?

The first question reflects the fact that taking graviton energies up to the Planck scale, M_P, would produce a dark energy density

$$\rho_v \approx M_P^4 \approx (10^{18}\ \mathrm{GeV})^4 \quad , \tag{3.1}$$

which misses the observed one, $\rho_v \approx (10^{-3}\ \mathrm{eV})^4$, by more than 120 orders of magnitude. Even if the energy density in fluctuations in the gravitational field is taken only up to the supersymmetry-breaking scale, M_{SUSY}, we still miss the mark by a factor of 60 orders of magnitude, since $\rho_v \approx M_{\mathrm{SUSY}}^4 \approx (1\ \mathrm{TeV})^4$. Interestingly, though, a scale $M_v \approx (M_{\mathrm{SUSY}}/M_P)M_{\mathrm{SUSY}}$ produces the right order of magnitude. However, while a few attempts in this direction have been made (e.g. Arkani-Hamed et al. 2000), no satisfactory model has been developed.

The second question is related to the anti-Copernican fact that Ω_Λ may be associated with a cosmological constant, while Ω_m declines continuously (and in any case, ρ_v may be expected to have a different time behavior from ρ_m), and yet the first time that we are able to measure both reliably, we find that they are of the same order.

The attempts to solve these problems fall into three general categories:

(1) The behavior of "quintessence" fields

(2) Alternative theories of gravity

(3) Anthropic considerations

The attempts of the first type have concentrated in particular on "tracker" solutions (e.g. Zlatev, Wang, & Steinhardt 1998; Albrecht & Skordis 2002), in which the smallness of Ω_Λ is a direct consequence of the Universe's old age. Generally, a uniform scalar field, ϕ, is taken to evolve according to

$$\ddot{\phi} + 3H\dot{\phi} + V'(\phi) = 0 \quad , \tag{3.2}$$

where $V'(\phi) = \frac{dV}{d\phi}$ and H is the Hubble parameter. The energy density of the scalar field is given by

$$\rho_\phi = \frac{1}{2}\dot{\phi}^2 + V(\phi) \quad , \tag{3.3}$$

and that of matter and radiation, ρ_m, by

$$\dot{\rho}_m = -3H(\rho_m + P_m) \quad , \tag{3.4}$$

where P_m is the pressure. For a potential of the form

$$V(\phi) = \phi^{-\alpha} M^{4+\alpha} \quad , \tag{3.5}$$

where $\alpha > 0$ and M is an adjustable constant ($M \ll M_P$), and a field that is initially much smaller than the Planck mass, one obtains a solution in which a transition occurs from an early ρ_m-dominance to a late ρ_ϕ-dominance (with no need to fine-tune the initial conditions). Nevertheless, for the condition $\rho_\phi \approx \rho_m$ to actually be satisfied at the present time requires (Weinberg 2001) that the parameter M would satisfy

$$M^{4+\alpha} \simeq (8\pi G)^{-1-\alpha/2} H_0^2 \quad , \tag{3.6}$$

which is not easily explicable.

In order to overcome this problem, some quintessence models choose potentials in which the Universe has periodically been accelerating in the past (e.g. Dodelson, Kaplinghat, & Stewart 2000), so that the dark energy's dominance today appears naturally.

A very different approach regards the accelerating expansion not as being propelled by dark energy, but rather as being the result of a modified gravity. For example, models have been developed (Deffayet, Dvali, & Gabadadze 2002), in which ordinary particles are localized on a three-dimensional surface (3-brane) embedded in infinite-volume extra dimensions to which gravity can spread. The model is constructed in such a way that observers on the brane discover Newtonian gravity (four dimensional) at distances that are shorter than a crossover scale, r_c, which can be of astronomical size. In one version, the Friedmann equation is replaced by

$$H^2 + \frac{k}{a^2} = \left(\sqrt{\frac{\rho}{3M_P^2} + \frac{1}{4r_c^2}} + \epsilon \frac{1}{2r_c^2} \right)^2 \quad , \tag{3.7}$$

where ρ is the total energy density, a is the scale factor and $\epsilon = \pm 1$.

In this case, the dynamics of gravity are governed by whether ρ/M_P^2 is larger or smaller than $1/r_c^2$. Choosing $r_c \approx H_0^{-1}$ preserves the usual cosmological results. At large cosmic distances, however, gravity spreads into extra dimensions (the force law becomes five dimensional), and becomes weaker—directly affecting the cosmic expansion. Basically, at late times, the model has a self-accelerating cosmological branch with $H = 1/r_c$ (to leading-order Equation 3.7 can be parameterized as $H^2 - H/r_c \simeq \rho/3M_P^2$). Interestingly, it has recently been suggested that the viability of these models can be tested by lunar ranging experiments (Dvali, Gruzinov, & Zaldarriaga 2003). I should also note that the *WMAP* results indicated an intriguing lack of correlated signal on angular scales greater than 60 degrees (Spergel et al. 2003), reinforcing the low quadrupole seen already in *COBE* results. One possible, although at this stage speculative, interpretation of these results is that they signal the breakdown of conventional gravity on large scales.

A third class of proposed solutions to the dark energy problems relies on anthropic selection effects, and therefore on the *existence* of intelligent life in our Universe. The basic premise of this approach is that some of the constants of nature are actually random variables, whose range of values and *a priori* probabilities are nevertheless determined by the laws of physics. The observed big bang, in this picture, is simply one member of an ensemble. It is further assumed that a "principle of mediocrity" applies; namely, we can expect to observe the most probable values (Vilenkin 1995). Using this approach, Garriga, Livio, & Vilenkin (2000; following the original idea of Weinberg 1987) were able to show that when the cosmological constant Λ is the only variable parameter, the order of magnitude coincidence $t_0 \approx t_\Lambda \approx t_G$ (where t_0 is the present time; t_Λ is the time Ω_Λ starts to dominate; t_G is the time when giant galaxies were assembled) finds a natural explanation (see also Bludman 2000).

Qualitatively, the argument works as follows.

In a geometrically flat universe with a cosmological constant, gravitational clustering can no longer occur after redshift $(1 + z_\Lambda) \approx (\rho_\Lambda/\rho_{m0})^{1/3}$ (where ρ_{m0} is the present matter density). Therefore, requiring that ρ_Λ does not dominate before redshift $z_{\max}$, at which the earliest galaxies formed, requires (e.g. Weinberg 1987)

$$\rho_\Lambda \lesssim (1 + z_{\max})^3 \rho_{m0}. \tag{3.8}$$

One can expect the *a priori* (independent of observers) probability distribution $P(\rho_\Lambda)$ to vary on some characteristic scale, $\Delta\rho_\Lambda \approx \eta^4$, determined by the underlying physics. Irrespective of whether η is determined by the Planck scale ($\sim 10^{18}$ GeV), the grand unification scale ($\sim 10^{16}$ GeV) or the electroweak scale ($\sim 10^2$ GeV), $\Delta\rho_\Lambda$ exceeds the anthropically allowed range of ρ_Λ (Eq. 3.8) by so many orders of magnitude that it looks reasonable to assume that

$$P(\rho_\Lambda) = const \quad , \tag{3.9}$$

over the range of interest. Garriga & Vilenkin (2001) and Weinberg (2001) have shown that this assumption is satisfied by a broad class of models, even though not automatically. With a flat distribution, a value of ρ_Λ picked randomly (and which may characterize a "pocket" universe) from an interval $|\rho_\Lambda| \lesssim \rho_\Lambda^{\rm max}$, will, with a high probability, be of the order of $\rho_\Lambda^{\rm max}$. The principle of mediocrity, however, means that we should observe a value of ρ_Λ that maximizes the number of galaxies. This suggests that we should observe the largest value of ρ_Λ that is still consistent with a substantial fraction of matter having collapsed into galaxies—in other words, $t_\Lambda \approx t_G$, as observed. In §2 I argued that the appearance of carbon-based life may be associated roughly with the peak in the star formation rate, $t_{\rm SFR}$. The "present time," t_0, is not much different from that (in that it takes only a fraction of a stellar lifetime to develop intelligent life), hence $t_0 \approx t_{\rm SFR}$. Finally, hierarchical structure formation models suggest that vigorous star formation is closely associated with the formation of galactic-size objects (e.g. Baugh et al. 1998; Fukugita, Hogan, & Peebles 1998). Therefore, $t_G \approx t_{\rm SFR}$, and we obtain $t_0 \approx t_G \approx t_\Lambda$.

Garriga et al. (2000) further expanded their discussion to treat not just Λ, but also the density contrast at recombination, $\sigma_{\rm rec}$, as a random variable (see also Tegmark & Rees 1998). The galaxy formation in this case is spread over a much wider time interval, and proper account has to be taken for the fact that the cooling of protogalactic clouds collapsing at very late times is too slow for efficient fragmentation and star formation (fragmentation occurs if the cooling time scale is shorter than the collapse time scale, $\tau_{\rm cool} < \tau_{\rm grav}$). Assuming an *a priori* probability distribution of the form

$$P(\sigma_{\rm rec}) \sim \sigma_{\rm rec}^{-\alpha} \quad , \tag{3.10}$$

Garriga et al. found that "mediocre" observers will detect $\sigma_{\rm rec} \approx 10^{-4}$, $t_0 \approx t_G \approx t_\Lambda \approx t_{\rm cb}$, as observed, *if* $\alpha > 3$ (here the "cooling boundary" $t_{\rm cb}$ is the time after which fragmentation is suppressed).

Other anthropic explanations for the value of the cosmological constant and the "why now?" problem have been suggested in the context of maximally extended ($N = 8$) supergravity (Kallosh & Linde 2003; Linde 2003). In particular, the former authors found that the Universe can have a sufficiently long lifetime only if the scaler field satisfies initially $|\phi| \lesssim M_P$, and if the value of the potential $V(0)$, which plays the role of the cosmological constant, does not exceed the critical density $\rho_0 \approx 10^{-120} M_P^4$.

Personally, I feel that anthropic explanations to the dark energy problems should be regarded as the *last resort*, only after all attempts to find explanations based on first principles have been exhausted and failed. Nevertheless, the anthropic explanation may prove to be the correct one, if our understanding of what is truly *fundamental* is lacking. A historical example can help to clarify this last statement. Johannes Kepler (1571–1630) was obsessed by the following two questions:

(1) Why were there precisely six planets? (only Mercury, Venus, Earth, Mars, Jupiter and Saturn were known at his time).

(2) What was it that determined that the planetary orbits would be spaced as they are?

The first thing to realize is that these "why" and "what" questions were a novelty in the astronomical vocabulary. Astronomers before Kepler were usually satisfied with simply recording the observed positions of the planets; Kepler was seeking a theoretical explanation. Kepler finally came up with preposterously fantastic (and absolutely wrong) answers to his two questions in *Mysterium Cosmographicum*, published in 1597. He suggested that the reason for there being six planets is that there are precisely five Platonic solids. Taken as boundaries (with an outer spherical boundary corresponding to the fixed stars), the solids create six spacings. By choosing a particular order for the solids to be embedded in each other, with the Earth separating the solids that can stand upright (cube, tetrahedron, and dodecahedron) from those that "float" (octahedron and icosahedron), Kepler claimed to have explained the sizes of the orbits too (the spacings agreed with observations to within 10%).

Today we recognize what was the *main* problem with Kepler's model—Kepler did not understand that neither the number of planets nor their spacings are *fundamental* quantities that need to have an explanation from first principles. Rather, both are the result of historical accidents in the solar protoplanetary disk. Still, it is perfectly legitimate to give an anthropic "explanation" for the Earth's orbital radius. If that orbit were not in the continuously habitable zone around the sun (Kasting & Reynolds 1993), we would not be here to ask the question.

It is difficult to admit it, but our current model for the composition of the Universe: ~73% dark energy, ~23% cold dark matter, ~4% baryonic matter, and maybe ~0.5% neutrinos, appears no less preposterous than Kepler's model. While some version of string (or $M-$) theories may eventually provide a first-principles explanation for all of these values, it is also possible, in my opinion, that these individual values are in fact not fundamental, but accidental. Maybe the only fundamental property is the fact that *all the energy densities add up to produce a geometrically flat universe*, as predicted by inflation (Guth 1981; Hawking 1982; Steinhardt & Turner 1984) and confirmed by *WMAP* (Spergel et al. 2003). Clearly, for any anthropic explanation of the value of Ω_Λ to be meaningful at all, even in principle, one requires the existence of a large ensemble of universes, with different values of Ω_Λ. That this requirement may actually be fulfilled is precisely one of the consequences of the concept of "eternal inflation" (Steinhardt 1983; Vilenkin 1983; Linde 1986; Goncharov, Linde, & Mukhanov 1987; Linde 2003). In most inflationary models the time scale associated with the expansion is much shorter than the decay time scale of the false vacuum phase, $\tau_{\rm exp} \ll \tau_{\rm dec}$. Consequently, the emergence of a fractal structure of "pocket universes" surrounded by false vacuum material is almost inevitable (Garcia-Bellido & Linde 1995; Guth 2001; for a different view, see, e.g. Bucher, Goldhaber, & Turok 1995; Turok 2001).

This ensemble of pocket universes may serve as the basis on which anthropic argumentation can be constructed (even though the definition of probabilities on this infinite set is nontrivial; see, e.g. Linde, Linde, & Mezhlumian 1995; Vilenkin 1998).

4. Varying constants of nature?

Another recent finding, which, *if confirmed*, may have implications for the emergence of life in the Universe, is that of cosmological evolution of the fine structure constant $\alpha \equiv e^2/\hbar c$ (Webb et al. 1999, 2001, and references therein). Needless to say, life as we know it places significant anthropic constraints on the range of values allowed for α. For example, the requirement that the lifetime of the proton would be longer than the main sequence lifetime of stars results in an upper bound $\alpha \lesssim 1/80$ (Ellis & Nanopoulos 1981; Barrow, Sandvik, & Magueijo 2002a). The claimed detection of time variability was based on shifts in the rest wavelengths of redshifted UV resonance transitions observed in quasar absorption systems. Basically, the dependence of observed wave number at redshift z, w_z, on α can be expressed as

$$w_z = w_0 + a_1 w_1 + a_2 w_2 \quad , \tag{4.1}$$

where a_1 and a_2 represent relativistic corrections for particular atomic masses and electron configurations, and

$$w_1 = \left(\frac{\alpha_z}{\alpha_0}\right)^2 - 1 \tag{4.2}$$

$$w_2 = \left(\frac{\alpha_z}{\alpha_0}\right)^4 - 1 \quad . \tag{4.3}$$

Here α_0 and α_z represent the present day and redshift z values of α, respectively. By analyzing a multitude of absorption lines from many multiplets in different ions, such as Fe II and Mg II transitions in 28 absorption systems (in the redshift range $0.5 \lesssim z \lesssim 1.8$), and Ni II, Cr II, Zn II, and Si IV transitions in some 40 absorption systems (in the redshift range $1.8 \lesssim z \lesssim 3.5$), Webb et al. (2001) concluded that α was *smaller* in the past. Their data suggest a 4σ deviation

$$\frac{\Delta\alpha}{\alpha} = -0.72 \pm 0.18 \times 10^{-5} \tag{4.4}$$

over the redshift range $0.5 \lesssim z \lesssim 3.5$ (where $\Delta\alpha/\alpha = \frac{\alpha_z - \alpha_o}{\alpha_o}$). It should be noted, however, that the data are consistent with *no* variation for $z \lesssim 1$, in agreement with many previous studies (e.g. Bahcall, Sargent, & Schmidt 1967; Wolfe, Brown, & Roberts 1976; Cowie & Songaila 1995).

Murphy et al. (2001) conducted a comprehensive search for systematic effects that could potentially be responsible for the result (e.g. laboratory wavelength errors, isotopic abundance effects, heliocentric corrections during the quasar integration, line blending, and atmospheric dispersion). While they concluded that isotopic abundance evolution and atmospheric dispersion could have an effect, this was in the direction of actually amplifying the variation in α [to $\Delta\alpha/\alpha = (-1.19 \pm 0.17) \times 10^{-5}$]. The most recent results of Webb et al. are not inconsistent with limits on α from the Oklo natural uranium fission reactor (which was active 1.8×10^9 years ago, corresponding to $z \approx 0.1$) and with constraints from experimental tests of the equivalence principle. The former suggests $\Delta\alpha/\alpha \simeq (-0.4 \pm 1.4) \times 10^{-8}$ (Fuji et al. 2000), and the latter *allows* for a variation of the magnitude observed in the context of a general dynamical theory relating variations of α to the electromagnetic fraction of the mass density in the Universe (Bekenstein 1982; Livio & Stiavelli 1998).

Before going any further, I would like to note that what is desperately needed right now is an independent confirmation (or refutation) of the results of Webb et al. by other groups, both through additional (and preferably different) observations and via independent analysis of the data. In this respect it is important to realize that the reliability of the SNe Ia results (concerning the accelerating Universe) was enormously enhanced by the fact that two separate teams (the Supernova Cosmology Project and the High-z Supernova Team) reached the same conclusion independently, using different samples and different data analysis techniques. A first small step in the direction of testing the variable α result came from measurements of the CMB. A likelihood analysis of BOOMERanG and MAXIMA data, allowing for the possibility of a time-varying α (which, in turn, affects the recombination time) found that in general the data may prefer a smaller α in the past (although the conclusion is not free of degeneracies; Avelino et al. 2000; Battye, Crittenden, & Weller 2001). A second, much more important step, came through an extensive analysis using the nebular emission lines of [O III] $\lambda\lambda 4959$, 5007 Å (Bahcall, Steinhardt, & Schlegel 2003). Bahcall et al. found $\Delta\alpha/\alpha = (-2 \pm 1.2) \times 10^{-4}$ (corresponding to $|\alpha^{-1} d\alpha/dt| < 10^{-13}$ yr^{-1}, which they consider to be a null result, given the precision of their method) for quasars in the redshift range $0.16 < z < 0.8$. While this result is not formally inconsistent with the variation claimed by Webb et al., the careful analysis of Bahcall et al. has cast some serious doubts on the ability of the "many-multiplet" method employed by Webb and his collaborators to actually reach the accuracy required to measure fractional variations in α at the 10^{-5} level. For example, Bahcall et al. have shown that to achieve that precision, one needs to assume that the velocity profiles of different ions in different clouds are essentially the same to within 1 km s^{-1}. Clearly, much more work on this topic is needed. I should also note right away that, in order not to be in conflict with the yield of ^{4}He, $|\Delta\alpha/\alpha|$ cannot exceed $\sim 2\times10^{-2}$ at the time of nucleosynthesis (e.g. Bergström, Igury, & Rubinstein 1999).

On the theoretical side, simple cosmological models with a varying fine structure constant have now been developed (e.g. Sandvik, Barrow, & Magueijo 2003; Barrow, Sandvik, & Magueijo 2002b). They share some properties with Kaluza-Klein-type models in which α varies at the same rate as the extra dimensions of space (e.g. Damour & Polyakov 1994), and with varying-speed-of-light theories (e.g. Albrecht & Magueijo 1999; Barrow & Magueijo 2000).

The general equations describing a geometrically flat, homogeneous, isotropic, variable-α universe are (Beckenstein 1982; Livio & Stiavelli 1998; Sandvik et al. 2002) the Friedmann equation (with $G = c \equiv 1$)

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi}{3}\left[\rho_m\left(1 + |\zeta_m| e^{-2\psi}\right) + \rho_r e^{-2\psi} + \rho_\psi + \rho_\Lambda\right] \quad , \tag{4.5}$$

the evolution of the scalar field varying α ($\alpha = \exp(2\psi) e_0^2/\hbar c$)

$$\ddot{\psi} + 3H\dot{\psi} = -\frac{2}{w} e^{-2\psi} \zeta_m \rho_m \quad , \tag{4.6}$$

and the conservation equations for matter and radiation

$$\dot{\rho}_m + 3H\rho_m = 0 \tag{4.7}$$

$$\dot{\rho}_r + 4H\rho_r = 2\dot{\psi}\rho_r \quad . \tag{4.8}$$

Here, ρ_m, ρ_r, ρ_ψ, ρ_Λ are the densities of matter, radiation, scalar field ($\frac{w}{2}\dot{\psi}^2$), and vacuum, respectively, $a(t)$ is the scale factor ($H \equiv \dot{a}/a$), $w = \hbar c/l^2$ is the coupling constant of the dynamic Langrangian (l is a length scale of the theory), and ξ_m is a dimensionless parameter that represents the fraction of mass in Coulomb energy of an average nucleon compared to the free proton mass.

Equations 4.5–4.8 were solved numerically by Sandvik et al. (2002) and Barrow et al. (2002), assuming a negative value of the parameter ξ_m/w, and the results are interesting both from a purely cosmological point of view and from the perspective of the emergence of life. First, the results are consistent with both the claims of a varying α of Webb et al. (which, as I noted, badly need further confirmation) and with the more secure, by now, observations of an accelerating Universe (Riess et al. 1998; Perlmutter et al. 1999; Spergel et al. 2003), while complying with the geological and nucleosynthetic constraints. Second, Barrow et al. find that α remains almost constant in the radiation-dominated era, experiences a small logarithmic time increase during the matter-dominated era, but approaches a constant value again in the Λ-dominated era. This behavior has interesting anthropic consequences. The existence of a nonzero vacuum energy contribution is now *required* in this picture to dynamically stabilize the fine structure constant. In a universe with zero Λ, α would continue to grow in the matter-dominate era to values that would make the emergence of life impossible (Barrow et al. 2001).

Clearly, the viability of all of the speculative ideas above relies at this point on the confirmation or refutation of time-varying constants of nature.

5. Is intelligent life extremely rare?

With the discovery of ~120 massive extrasolar planets (Mayor & Queloz 1995; Marcy & Butler 1996, 2000; Schneider 2003), the question of the potential existence of extraterrestrial, Galactic, intelligent life has certainly become more intriguing than ever. This topic has attracted much attention and generated many speculative (by necessity) probability estimates. Nevertheless, in a quite remarkable paper, Carter (1983) concluded on the basis of the near-equality between the lifetime of the sun, $t_\odot$, and the time scale of biological evolution on Earth, t_ℓ, that extraterrestrial intelligent civilizations are exceedingly rare in the Galaxy. Most significantly, Carter's conclusion is supposed to hold even if the conditions optimal for the emergence of life are relatively common.

Let me reproduce here, very briefly, Carter's argument. The basic, and very crucial assumption on which the argument is based is that the lifetime of a star, t_*, and the time scale of biological evolution on a planet around that star, t_ℓ (taken here, for definiteness, to be the time scale for the appearance of complex land life), are *a priori entirely independent*. In other words, the assumption is that land life appears at some *random* time with respect to the main sequence lifetime of the star. Under this assumption, one expects that generally one of the two relations $t_\ell \gg t_*$ or $t_\ell \ll t_*$ applies (the set where $t_\ell \approx t_*$ is of negligible measure for two independent quantities). Let us examine each one of these possibilities. If *generally* $t_\ell \ll t_*$, it is very difficult to understand why in the first system found to contain complex land life, the Earth-Sun system, the two time scales are nearly equal, $t_\ell \approx t_*$. If, on the other hand, *generally* $t_\ell \gg t_*$, then clearly the first system we find must exhibit $t_\ell \approx t_*$ (since for $t_\ell \gg t_*$ complex land life would not have developed). Therefore, one has to conclude that *typically* $t_\ell \gg t_*$, and that consequently, complex land life will generally not develop—the Earth is an extremely rare exception.

Carter's argument is quite powerful and not easily refutable. Its basic assumption (the independence of t_ℓ and t_*) appears on the face of it to be solid, since t_* is determined primarily by nuclear burning reactions, while t_ℓ is determined by biochemical reactions and the evolution of species. Nevertheless, the fact that the star is the main energy source for biological evolution (light energy exceeds the other sources by 2–3 orders of magnitude; e.g. Deamer 1997), already implies that the two quantities are not completely independent.

Let me first take a purely mathematical approach and examine what would it take for the condition $t_\ell \approx t_*$ to be satisfied in the Earth-Sun system *without* implying that extraterrestrial intelligent life is extremely rare. Imagine that t_ℓ and t_* are not independent, but rather that

$$t_\ell/t_* = f(t_*) \quad , \tag{5.1}$$

where $f(t_*)$ is some *monotonically increasing* function in the narrow range $t_*^{\min} \lesssim t_* \lesssim t_*^{\max}$ that allows the emergence of complex land life through the existence of continuously habitable zones (corresponding to stellar spectral types F5 to mid-K; Kasting et al. 1993). Note that, for a Salpeter (1955) initial mass function, the distribution of stellar lifetimes behaves as

$$\psi(t_*) \approx t_* \quad . \tag{5.2}$$

Consequently, if relation 5.1 were to hold, it would in fact be the *most probable* that in the first place where we encounter an intelligent civilization we would find that $t_\ell/t_* \approx 1$, as in the Earth-Sun system. In other words, if we could identify some processes that are likely to produce a monotonically increasing $t_* - t_\ell/t_*$ relation, then the near equality of t_ℓ and t_* in the Earth-Sun system would find a natural explanation, with no implications whatsoever for the frequency of intelligent civilizations. A few years ago, I proposed a simple toy-model for how such a relation might arise (Livio 1999). The toy-model was based on the assumption that the appearance of land life has to await the build-up of a sufficient layer of protective ozone (Berkner & Marshall 1965; Hart 1978), and on the fact that oxygen in a planet's atmosphere is released in the first phase from the dissociation of water (Hart 1978; Levine, Hayes, & Walker 1979). Given that the duration of this phase is inversely proportioned to the intensity of radiation in the 1000–2000 Å range, a relation between t_ℓ and t_* can be established. In fact, a simple calculation gave

$$t_\ell/t_* \simeq 0.4(t_*/t_\odot)^{1.7} \quad , \tag{5.3}$$

precisely the type of monotonic relation needed.

I should be the first to point out that the toy-model above is nothing more than that—a toy model. It does point out, however, that at the very least, establishing a link between the biochemical and astrophysical time scales may not be impossible. Clearly, the emergence of complex life on Earth required many factors operating together. These include processes that appear entirely accidental, such as the stabilization of the Earth's tilt against chaotic evolution by the Moon (e.g. Laskar, Joutel, & Boudin 1993). Nevertheless, we should not be so arrogant as to conclude everything from the one example we know. The discovery of many "hot Jupiters" (giant planets with orbital radii $\lesssim 0.05$ AU) has already demonstrated that the solar system may not be typical. We should keep an open mind to the possibility that biological complexity may find other paths to emerge, making various "accidents," coincidences, and fine-tuning unnecessary. In any case, the final scientific assessment on life in the Universe will probably come from biologists and observers—not from speculating theorists like myself.

This work has been supported by Grant 938-COS191 from the Templeton Foundation. I am grateful to Andrei Linde and Heinz Oberhummer for helpful comments.

REFERENCES

Abraham, R. G., van den Bergh, S., Glazebrook, K., Ellis, R. S., & Santiago, B. X. 1996 *ApJS* **101**, 1.

Albroe, M. E., et al. 2002 *MNRAS* **334**, 11.

Albrecht, A. & Magueijo, J. 1999 *Phys. Rev. D* **59**, 043516.

Albrecht, A. & Skordis, C. 2002 *Phys. Rev. Lett.* **84**, 2076.

Alpher, R. A., Bethe, H., & Gamow, G. 1948 *Phys. Rev.* **73**, 803.

Arkani-Hamed, N., Hall, L. J., Colda, C., & Murayama, H. 2000 *Phys. Rev. Lett.* **85**, 4434.

Avelino, P. P., Martins, C. J. A. P., Rocha, G., & Viana, P. 2000 *Phys. Rev. D* **62**, 123508.

Bahcall, J. N., Sargent, W. L. W, & Schmidt, M. 1967 *ApJ* **149**, L11.

Bahcall, J. N., Steinhardt, Cl. L. & Schlegel, D. 2003; *astro-ph/0301507.*

Bahcall, N. A., Cen, R., Davé, R., Ostriker, J. P., & Yu, O. 2000 *ApJ* **541**, 1.

Barrow, J. D. & Magueijo, J. 2000 *ApJ* **532**, L87.

Barrow, J. D., Sandvik, H. B., & Magueijo, J. 2002a; *Phys. Rev. D* **65**, 123501.

Barrow, J. D., Sandvik, H. B., & Magueijo, J. 2002b *Phys. Rev. D* **65**, 063504.

Battye, R. A., Crittenden, R., & Weller, J. 2001 *Phys. Rev. D* **63**, 043505.

Baugh, C. M., Cole, S., Frenk, C. S., & Lacey, C. G. 1998 *ApJ* **498**, 504.

Bekenstein, J. D. 1982 *Phys. Rev. D* **25**, 1527.

Bennett, C. L., et al. 2003 *ApJ*, in press; *astro-ph/0302207.*

Bergström, L., Iguri, S., & Rubinstein, H. 1999 *Phys. Rev. D* **60**, 045005.

Berkner, L. V. & Marshall, K. C. 1965 *J. Atmos. Sci.* **22**, 225.

Bludman, S. 2000 *Nucl. Phys. A* **663–664**, 865.

Bucher, M., Goldhaber, A., & Turok, N. 1995 *Phys. Rev. D* **52**, 3314.

Calzetti, D. & Heckman, T. M. 1999 *ApJ* **519**, 27.

Carlberg, R. G., Yee, H. K. C., Ellingson, E., Abramham, R., Grabel, P., Morris, S., Pritchet, C. J. 1996 *ApJ* **462**, 32.

Carter, B. 1983 *Philos. Trans. R. Soc. London A* **310**, 347.

Cook, C. W., Fowler, W. A., & Lauritsen, T. 1957 *Phys. Rev.* **107**, 508.

Cowie, L. L. & Songaila, A. 1995 *ApJ* **453**, 596.

Csótó, A., Oberhummer, H., & Schlattl, H. 2001 *Nucl. Phys. A* **688**, 560.

Damour, T. & Polyakov, A. M. 1994 *Nucl. Phys. B* **423**, 532.

Deamer, D. W. 1997 *Microb. and Molec. Bio. Rev.* **61**, 239.

de Bernardis, P., et al. 2002 *ApJ* **564**, 559.

Deffayet, C., Dvali, G., & Gabadadze, G. 2002; *Phys. Rev. D* **65**, 044023.

Dodelson, S., Kaplinghat, M., & Stewart, E. 2000 *Phys. Rev. Lett.* **85**, 5276.

Dunbar, D. N. F., Pixley, R. E., Wenzel, W. A., & Whaling, W. 1953 *Phys. Rev.* **92**, 649.

Dvali, G., Gruzinov, A., & Zaldarriaga, M. 2003; *hep-ph/0212069.*

Ellis, J. D. & Nanopoulos, D. V. 1981 *Nature* **292**, 436.

Fuji, V., et al. 2000 *Nucl. Phys. B* **573**, 377.

Fukugita, M., Hogan, C. J., & Peebles, P. J. E. 1998 *ApJ* **503**, 518.

Gamow, G. 1946 *Phys. Rev.* **70**, 527.

Garcia-Bellido, J. & Linde, A. O. 1995 Phys. Rev. D **51**, 429.

Garriga, J., Livio, M., & Vilenkin, A. 2000 *Phys. Rev. D.* **61**, 023503.

Garriga, J. & Vilenkin, A. 2001 *Phys. Rev. D.* **64**, 023517.

Giavalisco, M., et al. 2004 *ApJ*, in press.

Goncharov, A. S., Linde, A. D., & Mukhanov, V. F. 1987 *Int. J. Mod. Phys. A* **2**, 561.

Guth, A. H. 1981 *Phys. Rev. D* **23**, 347.

Guth, A. H. 2001, in *Astrophysical Ages and Time Scales* (eds. T. von Hippel, C. Simpson, & N. Mansit). p. 3. ASP Conf. Ser. Vol. 245.

Hart, M. H. 1978 *Icarus* **33**, 23.

Hawking, S. W. 1982 *Phys. Lett. B* **115**, 295.

Hoyle, F., Dunbar, D. N. F., & Wenzel, W. A. 1953 *Phys. Rev.* **92**, 1095.

Hoyle, F. & Tayler, R. J. 1964 *Nature* **203**, 1108.

Hubble, E. 1929 *Proc. Nat. Acad. Sci.* **15**, 168.

Kallosh, R. & Linde, A. 2002 *Phys. Rev. D.* **67**, 023510.

Kasting, J. F. & Reynolds, R. T. 1993 *Icarus* **101**, 108.

Kasting, J. F., Whitmore, D. P., & Reynolds, R. T. 1993 *Icarus* **101**, 108.

Lanzetta, K. M., Yahata, N., Pascarelle, S., Chen, H.-W., & Fernández-Soto, A. 2002 *ApJ* **570**, 492.

Laskar, J., Joutel, F., & Boudin, F. 1993 *Nature* **361**, 615.

Levine, J. S., Hayes, P. B., & Walker, J. C. G. 1979 *Icarus* **39**, 295.

Lilly, S. J., Le Fèvre, O., Hammer, F., & Crampton, D. 1996 *ApJ* **460**, L1.

Linde, A. 2003, in *Science and Ultimate Reality: From Quantum to Cosmos* (eds. J. D. Barrow, P. C. W. Davies, & C. L. Harper). Cambridge University Press; in press.

Linde, A., Linde, D., & Mezhlumian, A. 1995 *Phys. Lett. B* **345**, 203.

Linde, A. D. 1986 *Mod. Phys. Lett. A* **1**, 81.

Livio, M. 1999 *ApJ* **511**, 429.

Livio, M., Hollowell, D., Weiss, A., & Truran, J. W. 1989 *Nature* **340**, 281.

Livio, M. & Stiavelli, M. 1998 *ApJ* **507**, L13.

Madau, P., Ferguson, H. C., Dickinson, M., Giavalisco, M., Steidel, C. C., & Fruchter, A. 1996 *MNRAS* **283**, 1388.

Marcy, G. W. & Butler, R. P. 1996 *ApJ* **464**, L147.

March, G. W. & Butler, R. P. 2000 *PASP* **112**, 137.

Mather, J. C., Cheng, E. S., Cottingham, D. A., et al. 1994 *ApJ* **420**, 439.

Mayor, M. & Queloz, D. 1995 *Nature* **378**, 355.

Murphy, M. T., Webb, J. K., Flambaum, V. V., Churchill, C. W., & Prochaska, J. X. 2001 *MNRAS* **327**, 1223.

Netterfield, C. B., et al. 2002 *ApJ* **571**, 604.

Oberhummer, H., Csótó, A., & Schlattl, H. 2000 *Science* **289**, 88.

Peebles, P. J. E. 1966 *ApJ* **146**, 542.

Pei, Y., Fall, S. M., & Hauser, M. G. 1999 *ApJ* **522**, 604.

Perlmutter, S., et al. 1999 *ApJ* **517**, 565.

Riess, A. G., et al. 1998 *AJ* **116**, 1009.

Riess, A. G., et al. 2001 *ApJ* **560**, 49.

Roche, N., Ratnatunga, K. U., Griffiths, R. E., Im, M., & Neuschaefer, L. W. 1996 *MNRAS* **282**, 1247.

Salpeter, E. E. 1955 *ApJ* **121**, 161.

Sandvik, H. B., Barrow, J. D., & Magueijo, J. 2002 *Phys. Rev. Lett.* **88**, 031302.

Schlattl, H., Heger, A., Oberhummer, H., Rauscher, T., & Csótó, A. 2003 *MNRAS*, in press.

Schneider, J. 2003 *Extra-Solar Planets Catalog*, http://www.obspm.fr/encycl/catalog.html/.

Slipher, V. M. 1917 *Proc. Amer. Phil. Soc.* **56**, 403.

Spergl, D. N., et al. 2003 *ApJS*, **148**, 175.

Stanway, E. R., Bunker, A. J., & McMahon, R. G. 2003 *MNRAS*, in press; *astro-ph/0302212.*

Steidel, C. C., Adelberger, K. L., Giavalisco, M., Dickinson, M., & Pettini, M. 1999 *ApJ* **519**, 1.

Steinhardt, P. J. 1983, in *The Very Early Universe* (eds. G. W. Gibbons, S. Hawking, & S. T. C. Siklos). p. 251. Cambridge University Press.

Steinhardt, P. J. & Turner, M. S. 1984 *Phys. Rev. D* **29**, 2162.

Strauss, M. A. & Willick, J. A. 1995 *Phys. Rep.* **261**, 271.

Tegmark, M. & Rees, M. J. 1998 *ApJ* **499**, 526.

Timmes, F. X., Woosley, S. E., & Weaver, T. A. 1995 *ApJS* **98**, 617.

Turok, N. 2001, in *Birth and Evolution of the Universe* (eds. K. Sato & M. Kawasaki). p. 1. Universal Academy Press.

Vilenkin, A. 1983 *Phys. Rev. D* **27**, 2848.

Vilenkin, A. 1995 *Phys. Rev. Lett.* **74**, 846.

Vilenkin, A. 1998 *Phys. Rev. Lett.* **81**, 5501.

Wagoner, R. V., Fowler, W. A., & Hoyle, F. 1967 *ApJ* **148**, 3.

Webb, J. K., Flambaum, V. V., Churchill, C. W., Drinkwater, M. J., & Barrow, J. D. 1999 *Phys. Rev. Lett.* **82**, 884.

Webb, J. K., Murphy, M. T., Flambaum, V. V., Dzuba, V. A., Barrow, J. D., Churchill, C. W. Prochanska, J. X., & Wolfe, A. M. 2001 *Phys. Rev. Lett.* **87**, 091301.

Weinberg, S. 1987 *Phys. Rev. Lett.* **59**, 2607.

Weinberg, S. 2001, in *Sources and Detection of Dark Matter and Energy in the Universe* (ed. D. B. Cline). p. 18. Springer.

Wolfe, A. M., Brown, R. L., & Roberts, M. S. 1976 *Phys. Rev. Lett.* **37**, 179.

Wood, P. R. 1981, in *Physical Processes in Red Giants* (eds. I. Iben, Jr. & A. Renzini). p. 205. Reidel.

Yungelson, L., Tutukov, A. V., & Livio, M. 1993 *ApJ* **418**, 794.

Zlatev, I., Wang, L., & Steinhardt, P. J. 1998 *Phys. Rev. Lett.* **82**, 896.