

FINAL REPORT

to the Office of Life Science Programs
National Aeronautics and Space Administration

INVESTIGATION OF THERMODYNAMIC MECHANISMS FOR THE PRODUCTION
OF COMPLEX COMPOUNDS ESSENTIAL FOR THE ORIGIN OF LIFE

NSR 21-003-002

Oct. 1, 1965 to Sept. 30, 1966

GPO PRICE \$ _____

CFSTI PRICE(S) \$ _____

Hard copy (HC) 4.00

Microfiche (MF) 1.00

553 July 65

Margaret O. Dayhoff
Richard V. Eck
Ellis R. Lippincott

National Biomedical Research Foundation
8600 Sixteenth Street
Silver Spring, Maryland

FACILITY FORM 802	N66 28800	N66 28805
	(ACCESSION NUMBER)	(THRU)
	<u>127</u>	<u>6</u>
	(PAGES)	(COPIES)
<u>CR-74559</u>	<u>04</u>	
(NASA CR OR TMX OR AD NUMBER)	(CATEGORY)	

PROGRESS REPORT 1965-66

NASA Contract No. 21-003-002

Equilibrium Studies

It is evident that equilibrium processes have played a large role in the determination of compounds present in planetary atmospheres, meteorites, geological deposits and in the origin of life. The effects of equilibrium should be computed over broad ranges of conditions.

We have almost completed preparation of a monograph on our survey of the major compounds present in an ideal gas for all proportions of C, H, O, and N and trace quantities of S, P and Cl. This monograph describes the methods used and results for over 300 compounds of biochemical interest. These systems have been very interesting to astronomers, particularly with regard to planetary atmospheres, to space scientists because of the questions of remote measurements on planets and the chemical hallmarks of life on other planets, to meteoriticists, particularly Anders' group at Chicago, who have suggested equilibrium processes to account for carbonaceous compounds in meteorites, and to geobiochemists as Abelson and Blumer, interested in the stability of biological compounds over geological eras and in equilibrium processes in carbon metamorphosis on earth.

We have submitted for publication a paper entitled "Thermodynamic Equilibria in Planetary Atmospheres," E. R. Lippincott, R. V. Eck, M. O. Dayhoff, and C. Sagan.

Abstract. The thermodynamic equilibrium composition of all possible combinations of the elements C, H, O, and N at reasonable average pressures and temperatures have been calculated for the atmospheres of Earth, Venus, Mars and Jupiter. All regions consistent with the observed compound concentrations are located and discussed.

The constituents of the terrestrial atmosphere are in approximate thermodynamic equilibrium in spite of the continual addition of trace organic compounds from biological activity, lightning, radiation and vulcanism.

The atmosphere of Venus seems likely to be in thermodynamic equilibrium. It then follows that the measured upper limits on the abundances of CH_4 , NH_3 and CO are all inconsistent with the possibility of elemental carbon or hydrocarbons in contact with the atmosphere, either on the surface or as clouds. If the Cytherian atmosphere evolved from an original gas mixture much more reducing than the present, two processes must have occurred: the loss of hydrogen to space and the loss of O_2 , most probably by efficient reaction with a considerable layer of reduced surface materials. In addition, Venus must have begun its evolutionary history with a (C/O) abundance ratio ≤ 0.5 , a value consistent with several cosmic abundance estimates.

The evidence for Mars is not inconsistent with thermodynamic equilibrium. The computations do not indicate whether small quantities of free O_2 exist. The equilibrium abundance of all oxides of nitrogen is extremely low even if computed with a considerable excess of O_2 .

For Mars and Venus there is no molecular species with a large predicted equilibrium abundance and spectroscopically accessible absorption features which has not been already identified.

Under equilibrium conditions at low temperatures Jupiter could not contain significant amounts of any compound not observed. However, at high temperatures such as would be produced by lightning discharges, HCN , aromatics, and a variety of nitrogen compounds would be expected. These could contribute to the colors observed.

We have also submitted for publication the paper "Thermodynamic Equilibrium in the Origin of Organic Matter" by R. V. Eck, E. R. Lippincott, M. O. Dayhoff, and Y. T. Pratt. This included a preliminary report confirming the experimental evidence for equilibrium processes in a gas plasma. Calculations were made under the conditions of the heat catalysed reactions of Studier, Hayatsu and Anders, and of our plasma experiments.

Abstract. Theoretical and experimental support is presented for the hypothesis that organic compounds occurring in carbonaceous chondrites may have formed under equilibrium or near equilibrium conditions. The equilibrium distributions of organic compounds at temperatures between 300° and 1000° K and pressures of 10^{-6} to 50

atmospheres for the C-H-O system have been computed. At high temperatures and low pressures aromatic compounds may form even in the presence of excess hydrogen. Equilibrium concentrations of numerous compounds possible at 1000° K when N, S, and Cl are added to the system have also been determined. A limited equilibrium method is employed in which those few compounds which form with most difficulty are excluded from the computations. This approach is shown to be useful in the interpretation of certain experimental data. In preliminary experiments it has been found that gases, converted to the plasma state by high energy radio frequency discharge, yield product mixtures which are in qualitative agreement with those predicted.

The computer program to determine thermodynamic equilibrium in multiphase systems has been written and checked out. We have used it to determine the graphite phase boundaries for our survey monograph in preparation.

Biopolymer Studies

We have searched among the proteins of living organisms in the hope of finding relics of primitive structures preserved by the strong conservatism of natural selection. Ferredoxin is likely to have evolved from a prototype present very early in evolution, even before a genetic code evolved. Its structure may well preserve relics of the era of the evolution of the genetic code. Science will soon publish our paper on this subject, "Evolution of the Structure of Ferredoxin Based on Surviving Relics of Primitive Amino Acid Sequences," by R. V. Eck and M. O. Dayhoff.

Abstract. *The present-day structure of ferredoxin, with its simple, inorganic active site and its functions basic to photon-energy utilization, suggests the incorporation of its prototype into metabolism very early during biochemical evolution, even before complex proteins and the complete modern genetic code existed. The information in the amino acid sequence of ferredoxin enables us to reconstruct its evolutionary history. It has evolved by doubling a shorter protein, which may have contained only eight of the simplest amino acids. This shorter ancestor in turn developed from a repeating sequence of the*

amino acids alanine, aspartic acid or proline, serine, and glycine. The living relics of this primordial structure persist today because of a conservative principle in evolutionary biochemistry: Natural selection inhibits with extreme severity any change in a well-adapted system on which several other essential components depend. Many such intricate details of the earliest stages of life must still survive, awaiting detection and elucidation.

Talks

Since our last progress report, talks were given as follows:

NASA Bioscience Subcommittee of the Space Science Steering Committee meeting State College, Pennsylvania, April 1965, "Thermodynamic and Biochemical Studies of the Origin of Life" by M. O. Dayhoff, E. R. Lippincott, and R. V. Eck.

18th Annual Conference on Engineering in Medicine and Biology, Philadelphia, Pennsylvania, November 1965, "Predictions of the Compositions of the Atmospheres of the Planets" by R. V. Eck.

Meeting of the Biophysical Society, Boston, Massachusetts, February 1966, 2 papers, "Thermodynamic Equilibrium on the Prebiological Earth" by M. O. Dayhoff

and

"Protein Sequences Treated as Mathematical Puzzles: An Atlas of Protein Sequence and Structure," R. V. Eck.

Manuscripts Submitted for Publication

"Thermodynamic Equilibria in Planetary Atmospheres," E. R. Lippincott, R. V. Eck, M. O. Dayhoff, and C. Sagan, submitted 1965.

"Thermodynamic Equilibrium in the Origin of Organic Matter," R. V. Eck, E. R. Lippincott, M. O. Dayhoff, and Y. T. Pratt, submitted 1966.

"Evolution of the Structure of Ferredoxin Based on Surviving Relics of Primitive Amino Acid Sequences," R. V. Eck and M. O. Dayhoff, in press, Science, 1966.

PERSONNEL

Senior Scientists

Margaret O. Dayhoff
Richard V. Eck
Ellis R. Lippincott

Senior Programmer

Chan Mo Park

Literature Scientist

Minnie Sochard

Technical Assistants

Marie Chang
Lorrie Goldstein

BIOGRAPHICAL SKETCH

Margaret Oakley Dayhoff

Date of Birth: [REDACTED]. Place of Birth: [REDACTED]

Married: 2 Children.

U.S. Citizen; Female.

Chronology:

1960-present: Senior Research Scientist, Head, Department of Chemical Biology, National Biomedical Research Foundation.

1957-1959: Post-doctoral research fellowship, University of Maryland.

1948-1951: Research Assistant, Rockefeller Institute for Medical Research.

1947-1948: Watson I.B.M. Computing Laboratory Fellowship.

Education:

New York University, B.A. magna cum laude with honors; 1945, Mathematics.

Columbia University, M.A.; 1946, Chemistry.

Columbia University, Ph.D.; 1948, Quantum Chemistry.

Professional Societies:

Fellow of the American Association for the Advancement of Science

American Chemical Society

Association for Computing Machinery

Biophysical Society

New York Academy of Science

Society for Computer Science in Biology and Medicine, Secretary

Honors:

Phi Beta Kappa

Sigma Xi

Publications:

"Automatic Scanning of X-ray Diffraction Films into the Core Memory of a High Speed Computer", Proc. of 18th Ann. Conf. on Engineering in Medicine and Biology, Vol. 7, p. 165, 1965.

"Evolution of the Structure of Ferredoxin Based on Relics of Primitive Amino Acid Sequences" with R. V. Eck, Science, 1966 (in Press).

"Thermodynamic Equilibrium in the Origin of Organic Matter" with R. V. Eck, E. R. Lippincott, and Y. T. Pratt. (Submitted for publication)

"Thermodynamic Equilibria in Planetary Atmospheres" with E. R. Lippincott, R. V. Eck, and C. Sagan. (Submitted for publication)

Atlas of Protein Sequence and Structure 1965, with R. V. Eck, M. A. Chang, and M. Sochard, National Biomedical Research Foundation.

"Computer Aids to Protein Sequence Determination", J. of Theor. Biol., Vol. 8, pp. 97-112, 1965.

"Thermodynamic Equilibria in Prebiological Atmospheres" with E. R. Lippincott and R. V. Eck, Science, Vol. 146, No. 3650, pp. 1461-1464, Dec. 11, 1964.

"Computer Search for Active Site Configuration", J. of Amer. Chem. Soc., Vol. 86, pp. 2295-2297, 1964.

"Thermodynamic Equilibria in Prebiological Atmospheres, Program CHEM-EQUILIBRIUM" Proc. of 17th Ann. Conf. on Engineering in Medicine and Biology, p. 58, 1964.

"Simulation of Cellular Processes, Program CELSIM" with F. Heinmets, Proc. of 16th Ann. Conf. on Engineering in Medicine and Biology, Vol. 5, p. 54, 1963.

"A Contour-Map Program for X-ray Crystallography", Comm. of ACM, Vol. 6, No. 10, pp. 620-622, Oct. 1963.

"Delta Function Model of Chemical Binding" with E. R. Lippincott, 'Advance in Molecular Spectroscopy', Pergamon Press, 1962.

"Comprotein: A Computer Program to Aid Primary Protein Structure Determination" M. O. Dayhoff, R. S. Ledley, Proc. of Fall Joint Computer Conference, Vol. 9, pp. 262-274, 1962.

"The Use of Computers in Protein Biochemistry" with R. S. Ledley, Digest of the 15th Ann. Conf. on Engineering in Medicine and Biology, Vol. 4, p. 34, 1962.

"Delta Function Model of Chemical Binding" with E. R. Lippincott, Spec. Acta, Vol. 16, 807, 1960.

"The Apparent and Partial Molal Volumes of Potassium Iodide and of Iodine in Methanol at 25°C from Density Measurements" with D. A. MacInnes, J. Am. Chem. Soc., Vol. 75, p. 5219, 1953.

Publications:

"The Partial Specific Volumes, in Aqueous Solution of Three Proteins" with G. E. Perlman and D. A. MacInnes, J. Am. Chem. Soc., Vol. 74, p. 2515, 1952.

"The Partial Molal Volumes of MgI_2 and Iodine in Aqueous Solution" with D. A. MacInnes, J. Am. Chem. Soc., Vol. 74, p. 1917, 1952.

"A Study of Iodide-Iodine Solutions with the EMF Centrifuge" with D. A. Mac Innes J. Chem. Phys., Vol. 20, p. 1034, 1952.

"A Magnetic Float Method for Determining the Densities of Solutions" with D. A. MacInnes and B. R. Ray, Rev. Sci. Inst., Vol. 22, p. 642, 1951.

"Punched Card Calculation of Resonance Energies", Ph.D. Thesis, Columbia University, with G. E. Kimball, J. Chem. Phys., No. 17, p. 706, 1949.

BIOGRAPHICAL SKETCH

Richard V. Eck

Date of Birth: [REDACTED]. Place of Birth: [REDACTED]
U.S. Citizen; Male.

Chronology:

March 1965-present: Head, Department of Biology, National Biomedical Research Foundation.

Sept. 1963-present: Senior Research Scientist, National Biomedical Research Foundation.

1954-1963: Biologist, National Cancer Institute. Developed several animal experimental models of complications of cancer surgery, along with mathematical methods for interpreting them. Collaborated with surgeons in the use of these methods, and advised them on experimental design, etc. Worked with immunogenetics and tissue transplantation. Did theoretical-mathematical work on the nature of the genetic code.

1943-1946: Army of the United States, highest rank First Lieutenant.

Education:

Illinois Institute of Technology, 1938-1941, Chemical Engineering.

University of Maryland, 1941-1943, B.S. Agricultural Chemistry.

West Virginia University, 1943-1944, (A.S.T.P.) Mechanical Engineering.

University of Maryland Graduate School, 1946-1950. Approx. 100 hours of graduate courses in plant pathology, botany, biochemistry, statistics, experimental design. Taught laboratory courses in Elementary Botany, Plant Pathology, and Virology. Research on a semi-micro chemical method of assaying tobacco mosaic virus in small samples of leaf tissue.

Publications:

"Wound Seeding as a Cause of Failure in Surgical Therapy of Cancer" with A. W. Hilberg, M.D., R. R. Smith, M.D., and others, Proc. Third Nat. Canc. Conf., 568-571, 1957.

"Local Chemotherapy of Experimentally Tumor-Seeded Wounds" with A. G. Ship, M.D., and R. R. Smith, M.D., Cancer, 11, 687-695, 1958.

"Inhibition of Local Tumor Implantation with Triethylenethiophosphoramide (TSPA)" with W. M. Kramer, M.D., and R. R. Smith, M.D., Cancer, 11, 999-1002, 1958.

"Prevention of Experimental Lung Metastases with Triethylenethiophosphoramide (ThioTEPA)" with W. M. Kramer, M.D., and R. R. Smith, M.D., Surg., Gyn. Obst., 106, 427-434, 1958.

"An Experimental Study of the Effect of Irradiation on the Dissemination of Cancer" with P. D. Olch, M.D., and R. R. Smith, M.D., Canc. Res., 19, 464-467, 1959.

"An Experimental Study of the Effect of External Irradiation on a 'Primary' Tumor and Its Distant Metastases" with P. D. Olch, M.D., and R. R. Smith, M.D., Cancer, 12, 23-26, 1959.

"Non-Randomness in Amino-Acid 'Alleles'", Nature, Vol. 191, pp. 1284-1285, 1961.

"A Simplified Strategy for Sequence Analysis of Large Proteins", Nature, Vol. 193, pp. 241-243, 1962.

"The Protein Cryptogram: I. Non-Random Occurrence of Amino Acid 'Alleles'", J. Theoret. Biol., Vol. 2, pp. 139-151, 1962.

"Genetic Code: Emergence of a Symmetrical Pattern", Science, Vol. 140, pp. 477-481, 1963.

"Cryptogrammic Detection of a Pattern in Amino Acid 'Alleles'; Its Use in Tracing the Evolution of Proteins", Proc. of the 17th Ann. Conf. on Engineering in Medicine and Biology, Vol. 6, pp. 115, 1964.

"Thermodynamic Equilibria in Prebiological Atmospheres", with M. O. Dayhoff and E. R. Lippincott, Science, Vol. 146, No. 3650, pp. 1461-1464, Dec. 11, 1964.

"Predictions of the Compositions of the Atmospheres of the Planets", Proc. of the 18th Ann. Conf. on Engineering in Medicine and Biology, Vol. 7, p. 185, 1965.

"Evolution of the Structure of Ferredoxin Based on Relics of Primitive Amino Acid Sequences", with M. O. Dayhoff, Science, 1966 (in press).

"Thermodynamic Equilibrium in the Origin of Organic Matter", with E. R. Lippincott, M. O. Dayhoff, and Y. T. Pratt (submitted for publication).

"Thermodynamic Equilibria in Planetary Atmospheres", with E. R. Lippincott, M. O. Dayhoff, and C. Sagan (submitted for publication).

Atlas of Protein Sequence and Structure 1965, with M. O. Dayhoff, M. A. Chang, and M. R. Sochard, National Biomedical Research Foundation.

ELLIS R. LIPPINCOTT

Professor

I. Personal Data

Date of Birth - [REDACTED]
Place of Birth - [REDACTED]
Marital Status - Rita C. Clifton, August 14, 1948
Children - Melanie, Jennifer, Wendy, Robin and Holly

II. Education

B.A. in Chemistry, Earlham College, 1943
M.A. in Chemistry, Johns Hopkins University, 1944
Ph.D. in Chemistry, Johns Hopkins University, 1947
(With Minors in Physics and Mathematics)
Postdoctoral Research Associate, Massachusetts Institute of
Technology, June 1947-July 1948

III. Experience in Higher Education

A. Instructor in Chemistry, Loyola College, Baltimore, 1947
Instructor in Chemistry, University of Connecticut, 1948-51
Associate Professor in Chemistry, Kansas State College,
1951-55
Professor of Chemistry, University of Maryland, 1955 to present
Lecturer in Infrared Spectroscopy (two-week summer course)
Massachusetts Institute of Technology, 1953 to present

B. Courses Taught:

General Chemistry, Undergraduate Physical Chemistry, Graduate
Physical Chemistry, Molecular Structure, Quantum Chemistry,
Infrared and Raman Spectroscopy, Atomic and Molecular Spectra,
Statistical Mechanics

C. Research Interests:

Physical Chemistry, Quantum Chemistry, Biophysics, High Pres-
sure Optics, Molecular Spectra and Structure, Infrared and
Raman Spectroscopy, Hydrogen Bonding, Potential Energy Func-
tions and Applications of spectral Studies to Biochemistry

D. Research and Educational Activities:

Program Director for Public Health Service Physical Chemistry
Training Program
Research Director of 30 Ph.D. and M.S. candidates for which
degrees have been granted
Principal Investigator for several research grants over the
past sixteen years totalling more than one million dollars
in research and training funds with the following agencies:

(National Science Foundation, Atomic Energy Commission, Office of Ordnance Research, Office of Naval Research, Kansas State Bureau of General Research, University of Maryland Research Board, Research Corporation, Public Health Research, Public Health Training, Advanced Research Projects Agency (Department of Defense), North Atlantic Treaty Organization and National Aeronautical Space Administration, etc.) These funds were used for basic research and the financial support of students working for advanced degrees.

IV. Experience other than in Higher Education

Various part-time and summer industrial employments prior to the granting of the Ph.D. degree in 1947
Consultant to National Bureau of Standards and other governmental and industrial organizations

V. Publications, Patents and Other Creative Work

Author or co-author of approximately 120 publications which include technical papers, book reviews, patents, contributions to technical books and editor of book entitled "Proceedings of the Xth Colloquium Spectroscopicum Internationale"

VI. Professional Activities

A. Membership in professional organizations:

American Chemical Society
American Physical Society
Faraday Society
Coblentz Society
Philosophical Society of Washington
Washington Academy of Science
Society for Applied Spectroscopy
American Society for Testing and Materials

B. Approximately 60 technical or educational papers have been presented to national and international scientific groups. Invited lecture series and papers have been presented before:

National Academy of Science, 1956
International Symposium on "Hydrogen Bonding," Ljubljana, Yugoslavia, 1957
European Spectroscopic Conference, Freiburg, W. Germany, 1957 and Bologna, Italy, 1961
Gordon Research Conferences, 1956, 1957, 1962 and 1964
American Chemical Society, 1957, 1959, 1962 and 1964
Pittsburgh Spectroscopic Society, 1958 and 1963
Philosophical Society of Washington, 1958
National Science Foundation College Teachers Institute Lecture Series, University of North Carolina, 1959

In addition, invited lectures have been given to several University, industrial, governmental groups, etc. Popular lectures have been given to high school groups.

E. During 1965, invited papers were presented before several state colleges and universities, the American Chemical Society at its fall meeting in Atlantic City, New Jersey; The Naval Research Laboratory; the National Bureau of Standards, etc. Also, papers were contributed to the 8th European Congress on Molecular Spectroscopy in Copenhagen, Denmark; and the XII. International Spectroscopy Colloquium at the University of Exeter in London, England.

VII. Honors Received:

Visiting Scientist, Division of Chemical Education, American Chemical Society (1963-1966).

Chairman - Gordon Research Conference on Infrared Spectroscopy (1964).

Chairman - Symposium on Recent Developments in Research Methods, National Institutes of Health (1962).

Program Chairman - Xth Colloquium Spectroscopicum Internationale (1962).

Program Chairman - Spring Raman Institute and Fourth Annual Workshop (1966).

VIII. Honorary Societies:

Phi Beta Kappa, Sigma Xi, Alpha Chi Sigma, Gamma Alpha, and Pi mu Epsilon.

Advisory Editor for the scientific journal "Spectrochimica Acta".

IX. Awards Received:

1. The 1964 Hillebrand Award cited Professor Lippincott for "His investigations in infrared and Raman spectroscopy, particularly in the determination of molecular structures and the bond properties of molecules and his insight in determining correlations among molecular bond properties; for his contributions to pioneering work in spectra at very high pressure, and his contributions to spectroscopic techniques, especially the development of new Raman light sources."
2. The Certificate of Award presented by the Department of Commerce, National Bureau of Standards, in 1964 was "in recognition and appreciation of distinguished service in the performance of official duties."

Publications from January 1, 1962:

1. "Hydrogen Bonding in the Calcium Phosphates," J.M. Stutman, A.S. Posner and E.R. Lippincott, *Nature*, 193, 368 (1962).
2. "Lattice Frequencies and Rotational Barriers for Inorganic Carbonates and Nitrates," R.A. Schroeder, C.E. Weir and E.R. Lippincott, *J. Chem. Phys.*, 36, 2803 (1962).
3. "Lattice Frequencies and Rotational Barriers for Inorganic Carbonates and Nitrates from Low Temperature Infrared Spectroscopy," R.A. Schroeder, C.E. Weir and E.R. Lippincott, *J. Research Nat. Bur. Stds.*, 66A, 407 (1962).
4. "Comparative Study of Empirical Internuclear Potential Functions," D. Steele, E.R. Lippincott and J.T. Vanderslice, *Rev. Mod. Phys.*, 34, 239 (1962).
5. "The Vibrational Spectra and Geometric Configuration of cis-1,3,5-Hexatriene," E.R. Lippincott and T.E. Kenney, *J. Am. Chem. Soc.*, 84, 3641 (1962).
6. "Complex Formation of Monomeric Amides with Lithium Perchlorate," A.F. Diorio, E.R. Lippincott and L. Mandelkern, *Nature*, 194, 1296 (1962).
7. "Wave Functions for Anharmonic Oscillators by Perturbation Methods," A.M. Shorb, R. Schroeder and E.R. Lippincott, *J. Chem. Phys.*, 37, 1043 (1962).
8. "Delta Function Model of Chemical Binding," E.R. Lippincott and M.O. Dayhoff, "Advances in Molecular Spectroscopy" Pergamon Press (1962).
9. "Polymorphism in Fibrous Polypeptides: Transformation in Naturally occurring Keratin," A.F. Diorio, L. Mandelkern and E.R. Lippincott, *J. Phys. Chem.*, 66, 2096 (1962).
10. "Optical Studies at High Pressures Using Diamond Anvils," C.E. Weir, A. Van Valkenburg, and E.R. Lippincott, Modern Very High Pressure Techniques, Butterworth, Inc., Washington, D.C. (1962).
11. "High Intensity Electrodeless Sources for Raman Spectroscopy," F.X. Powell, O. Fletcher and E.R. Lippincott, *Rev. Sci. Instru.* 34, 36 (1963).
12. "The Limitations and Advantages of Infrared Spectroscopy in Patent Problems," Ellis R. Lippincott, *J. Pat. Office Soc.* 45, 380 (1963).
13. "Calculations of Higher-Order Spectroscopic Parameters Using Empirical Potential Functions," J.M. Stutman, E.R. Lippincott, and D. Steele, *J. Chem. Phys.*, 39, 564 (1963).
14. "Raman Spectra and Solvent-Extractions of Cuprous Halides," J.A. Creighton and E.R. Lippincott, *J. Chem. Soc.*, 983, 5134 (1963).

15. "The Vibrational Spectra and Molecular Configuration of Diphenyl Ether," J.E. Katon, W.R. Fearheller, Jr. and E.R. Lippincott, *J. Mol. Spectroscopy*, 13, 72 (1964).
16. "Optical Studies at High Pressures," L.S. Whatley, E.R. Lippincott and A. Van Valkenburg, *Science*, 144, 968 (1964).
17. "Pressure Distribution Measurements in Fixed-Anvil High-Pressure Cells," E.R. Lippincott and H.C. Duecker, *Science*, 144, 1119 (1964).
18. "Assembly and Performance of a Double Beam Microscope Spectrophotometer from Commercial Instruments" H.C. Duecker and E.R. Lippincott, *Rev. Scientific Instruments*, 35, 1108, (1964).
19. "The Molecular Spectra of the Thiotriithiazyl Cation," R.T. Bailey and E.R. Lippincott, *Spectrochim. Acta*, 20, 1327 (1964).
20. "Polarizabilities from Delta-Function Potentials," E.R. Lippincott and J.M. Stutman, *J. Phys. Chem.*, 68, 2926 (1964).
21. "Nuclear Magnetic Resonance Studies Concerning the Effect of Adjacent Hydrogen Atoms on the Hydrogen-Bond Properties," E.R. Lippincott and A.S. Rao, *J. Chem. Phys.*, 41, 3006 (1964).
22. "Shrinkages of the Internuclear Distances by Molecular Vibrations," G. Nagarajan and E.R. Lippincott, *J. Chem. Phys.* 42, 1809 (1965).
23. "Thermodynamic Equilibria in Prebiological Atmospheres", M.O. Dayhoff, E.R. Lippincott and R.V. Eck, *Science*, 146, 1461 (1964).
24. "Mean Amplitudes of Vibration, Thermodynamic Functions, Molecular Polarizability and Absolute Raman Intensities of $\sum g^+$ Modes in Carbon Subnitride", G. Nagarajan, E.R. Lippincott and J.M. Stutman, *Z. Naturforschung* 20a, 786 (1964).
25. "Pressure-Induced Trapping Phenomenon in Silver Iodide", H.C. Duecker and E.R. Lippincott, *Science*, 146, 1295 (1964).
26. "Mean Amplitudes of Vibration and Bastiansen-Morino Shrinkage Effects in some Linear Symmetrical Four Atomic Molecules," G. Nagarajan and E.R. Lippincott, *J. Chim. Physique*, 62 (1965).
27. "The Vibrational Spectrum of Dimethylferrocene," R.T. Bailey and E.R. Lippincott, *Spectrochim. Acta*. 21, 389 (1965).
28. "Raman Spectrum of Azulene" R.T. Bailey and E.R. Lippincott, *J. Chem. Phys.* 42, 1121 (1965).
29. "The Vibrational Spectra of Penta-Deutero-Chlorobenzene," T.R. Nanney, R.T. Bailey and E.R. Lippincott. *Spectrochim. Acta* 21, 1495 (1965).
30. "Mean Amplitudes of Vibration, Bastiansen-Morino Shrinkage Effect, Thermodynamic Functions and Molecular Polarizabilities of Sulfur Trioxide," E.R. Lippincott, G. Nagarajan and J.M. Stutman, *J. Phys. Chem.*, 69, 2017, (1965).
31. "Spectroscopy", E.R. Lippincott and L.S. Whatley, *Encyclopedia of Physics*, Edited by R. Besancon, Reinhold Publishers (1965).

32. "Infrared Absorption Spectrum and Ferroelectric Behavior of Sodium Trihydroselenite," R.K. Khanna, J.C. Decius and E.R. Lippincott. *J. Chem. Phys.* 43, 2974 (1965).
33. "The Vibrational Spectra and Structure of Cyclooctatetraene Iron Tricarbonyl," R.T. Bailey, E.R. Lippincott and D. Steele. *J. Am. Chem. Soc.* 87, 5346 (1965).
34. "Polarizabilities from Delta-Function Model of Chemical Binding - Part II: Molecules with Polar Bonds," J.M. Stutman, E.R. Lippincott and G. Nagarajan. *J. Phys. Chem.* 70, 78 (1966).
35. "Distribution of Pressure in an Opposed Anvil High Pressure Cell," E.R. Lippincott and H.C. Duecker. Accepted for publication in *A.S.M.E. J. of Basic. Eng.*
36. "The Raman and Low Frequency Infrared Spectra of C_6F_5Cl , C_6F_5Br and C_6F_5I ", I.J. Hyams, E.R. Lippincott and R.T. Bailey. Accepted for publication in *Spectrochim. Acta*.
37. "The Vibrational Spectra of Penta-Deutero-Bromobenzene and Penta-Deutero-Iodobenzene", T.R. Nanney, E.R. Lippincott and J.C. Hamer. Accepted for publication in *Spectrochim. Acta*.
38. "Absolute Raman Intensities of Symmetrical Stretching Modes in some Molecules and Ions from Delta-Function Model of Chemical Binding," G. Nagarajan and E.R. Lippincott. Accepted for publication in *Bull. Soc. Chim. Belg.*
39. "The Vibrational Spectrum and Geometrical Configuration of Decafluoro-biphenyl," D. Steele, T.R. Nanney and E.R. Lippincott. Accepted for publication in *Spectrochim. Acta*.
40. "The Vibrational Spectra of Thiophene Derivatives. Part I. The Monohalogenothiophenes", M. Horak, I.J. Hyams and E.R. Lippincott. Accepted for publication in *Spectrochim. Acta*.
41. "Low Temperature Infrared Spectra of Crystalline Inorganic Compounds Containing Tetrahedral Anions," R.A. Schroeder, E.R. Lippincott and C.E. Weir. Accepted for publication.
42. "Infrared Studies on Glycine and Its Addition Compounds," R.K. Khanna, M. Horak and E.R. Lippincott. Accepted for publication.
43. "Infrared Studies on Some Ferroelectric Compounds of Glycine," R.K. Khanna, M. Horak and E.R. Lippincott. Accepted for publication.
44. "Enthalpy, Free Energy, Entropy and Heat Capacity of Cyclohexane and Acetaldehyde, G. Nagarajan and E.R. Lippincott. Submitted for publication.
45. "Stimulated Raman Spectra from some Polyatomic Molecules", E.R. Lippincott, C.E. Myers, P.J. Hendra. Submitted for publication.
46. "Molecular Conformation and Vibrational Spectra of p-Halogeno Derivatives of Anisole", M. Horak, E.R. Lippincott and R.K. Khanna. Submitted for publication.

BIOGRAPHICAL SKETCH

Chan Mo Park

Date of Birth: [REDACTED] Place of Birth: [REDACTED]

Married: One Child
Korean Citizen; Male

Education:

Kyunggi High School, Seoul, Korea, 1951-1954
College of Engineering, Seoul National University, 1954-1958, B.S. Chemical Eng.
Graduate School, Seoul National University, 1958-1959
Graduate School, University of Maryland, 1960-present, M.S. Chemical Eng., 1964

Chronology:

1961-1962: Research Assistant, Dept. of Chem. Eng., University of Maryland

1962-1963: Graduate Teaching Assistant, Dept. of Chem. Eng., University of Maryland

1963-1964: Graduate Research Assistant, Computer Science Center, University of Maryland

1964-present: Research Associate from April 1964 to July 1964; Research Programmer since July 1964, Computer Science Center, University of Maryland

Present: Senior Research Scientist, National Biomedical Research Foundation

Publications:

M.S. Thesis, "Effective Thermal Conductivity in a Packed Bed", University of Maryland, January 1964.

"Interim Report on Scattering of Gammas through Ducts," with C. B. Agnihotri and J. Silverman, University of Maryland, June 1962.

"Single Reflection of Co⁶⁰ Gammas in Concrete Ducts", with J. Silverman, Transaction American Nuclear Society, Vol. 5, No. 2, November 1962 (Paper presented at the 1962 Winter Meeting of A.N.S.).

"Subroutines for Digital Computation of Chemical Engineering Problems - Part I", Technical Report No. 24, Computer Science Center, University of Maryland, October 1965.

BIOGRAPHICAL SKETCH

Minnie R. Sochard

Date of Birth: [REDACTED] Place of Birth: [REDACTED]

Married: 2 Children

U. S. Citizen: Female

Education:

Hunter College, New York City, New York, 1952, B.S., Zoology

Georgetown University, Washington, D. C., 1965, M.S., Microbiology

Chronology:

- 1964 - present: Junior Research Scientist, National Biomedical Research Fdn.
1962 - 1965: Graduate Assistant, Dept. Microbiology and Tropical Medicine, Georgetown University
1963 - 1964: Diagnostic Microbiology, NIH grant, Georgetown University.
1955 - 1958: Research Microbiologist, Gastroenteritis Section, Walter Reed Army Institute of Research
1954 - 1955: Research Microbiologist, NIH
1952 - 1954: Diagnostic Bacteriologist, Babies Hospital, College of Physicians and Surgeons, Columbia University, New York City, New York

Membership in Professional Societies:

American Society for Microbiology

Publications:

- Infectious Agents in Infant Diarrhea: I. A Hemagglutination-Inhibition Procedure for Detection of Bacterial Fractions in Infant Sera, V. M. Young, M. R. Sochard, H. C. Gillem, Proc. Soc. Exp. Biol. Med. (1960), 105: 635-638.
- Infectious Agents in Infant Diarrhea: II. Seriological Reactions with Eschenichia Coli O1 through O25, V. M. Young, M. R. Sochard, H. C. Gillem, and S. Ross, Proc. Soc. Exp. Biol. Med., (1960), 105:638-642.
- Studies of Infectious Agents in Infant Diarrhea: III. Bacterial, Viral and Parasitic Agents in Faces of Puerto Rican Children, V. M. Young, R. B. Lindberg, A. Ortiz, D. Jahiel, M. R. Sochard, J. J. Hemphill, Am. J. Trop. Med. in Hyg., (1962), Vol. 11, pp. 380-388.
- Studies of Infectious Agents in Infant Diarrhea: Seriological Behavior of Non-Pathogenic Eschenichia Coli, V. M. Young, M. R. Sochard, H. C. Gillem, and S. Ross, Bact. Proceedings, 59th Gen. Meeting, 1959, M-132.
- ATLAS of Protein Sequence and Structure 1965, Dayhoff, M. O., Eck, R. V., Chang, M. A., and Sochard, M. R.

BIOGRAPHICAL SKETCH

Marie A. Chang

Date of Birth: [REDACTED]. Place of Birth: [REDACTED]
Married
Korean Citizen. U. S. Permanent Resident. Female

Chronology:

April 1963 - Present: National Biomedical Research Foundation, Research Scientist

March 1962 - July 1962: Congressional High School, Math Teacher, Grades 10 - 12

July 1960 - March 1961: International Bank for Reconstruction & Development

Sept. 1959 - June 1960: Waxmann Institute of Microbiology

Education:

Georgian Court College, Lakewood, New Jersey, 1955-1959, A.B., Mathematics

Rutgers University, New Brunswick, New Jersey, 1959-1960, Applied Mathematics

Publications:

Atlas of Protein Sequence and Structure 1965, with M. O. Dayhoff, R. V. Eck, and M. Sochard, National Biomedical Research Foundation.

Correlation and Regression Analysis for IBRD Applications, with D. H. Niewiaroski, IBRD publication, 1961.

THERMODYNAMIC EQUILIBRIA IN PLANETARY ATMOSPHERES

Ellis R. Lippincott

Department of Chemistry
University of Maryland, College Park

Richard V. Eck

National Biomedical Research Foundation
Silver Spring, Maryland

Margaret O. Dayhoff

National Biomedical Research Foundation
Silver Spring, Maryland

Carl Sagan

Harvard University
and

Smithsonian Astrophysical Observatory
Cambridge, Massachusetts

THERMODYNAMIC EQUILIBRIA IN PLANETARY ATMOSPHERES*

Ellis R. Lippincott, Richard V. Eck, Margaret O. Dayhoff and Carl Sagan

Abstract

From present information on the composition, pressures, and temperatures of planetary atmospheres, we have calculated the expected thermodynamic equilibrium composition of the atmospheres of the Earth, Venus, Mars, and Jupiter. Departures from thermodynamic equilibrium must be attributed to special mechanisms, including, on the Earth, biological activity. The major constituents of the terrestrial atmosphere are found to be in approximate thermodynamic equilibrium; while many minor constituents have abundances exceeding their equilibrium values, there is a marked tendency for equilibrium to be restored. It appears that the atmosphere of Venus is in thermodynamic equilibrium, while that of Jupiter is not. The evidence for Mars is less conclusive, although not inconsistent with thermodynamic equilibrium. For none of these planets is there a molecular species with a large predicted equilibrium abundance and spectroscopically accessible absorption features which has not been already identified. The predicted equilibrium abundances of oxides of nitrogen are extremely low on all planets. The measured upper limits on the abundances of Cytherean CH_4 , NH_3 , and CO are all inconsistent with the possibility of elemental carbon or organic molecules in contact with the atmosphere of Venus either on the surface or as clouds. The possibility of small quantities of oxygen on Mars is not excluded. We suggest that the colored materials at the Jovian clouds are organic molecules, produced by non-equilibrium processes in a reducing environment.

If the Cytherean atmosphere evolved from an original gas mixture much more reducing than the present mixture, two processes are required to achieve the unique contemporary composition: the loss of hydrogen to space, and the efficient reaction of atmospheric oxygen with reduced surface materials. In addition, Venus must have begun its evolutionary history with a $[\text{C}/\text{O}]$ abundance ratio ≤ 0.5 , a value consistent with several cosmic abundance estimates.

* This work was supported in part by NASA Contract No. 21-003-002 with the National Biomedical Research Foundation, 8600 16th Street, Silver Spring, Maryland.

THERMODYNAMIC EQUILIBRIA IN PLANETARY ATMOSPHERES

Ellis R. Lippincott, Richard V. Eck, Margaret O. Dayhoff and Carl Sagan

1. Introduction

It is not known how closely the gross chemical composition of planetary atmospheres is approximated by thermodynamic equilibrium of the constituents. In the upper reaches of planetary atmospheres, where the optical depth at ultraviolet frequencies is small, relative molecular abundance may be dominated by photodissociation and recombination. In the lower portions of the atmospheres, which are shielded from photodissociating ultraviolet radiation, a closer approximation to thermodynamic equilibrium can be expected. Especially where surface temperatures are high or when catalysts such as water are present, there will be a strong tendency for the degradation of molecules to the lowest energy state. Even in the presence of special mechanisms which favor the production of specific compounds -- mechanisms such as ultraviolet radiation or lightning -- the thermodynamic equilibrium distribution of reaction products should be a useful first-order approximation. Since the equilibrium state is independent of particular reaction mechanisms and reaction rates, and since the computations of equilibrium states are quite straightforward, studies of thermodynamic equilibrium chemistry would appear prerequisite to considerations of non-equilibrium processes.

One of the earliest examples of a fruitful argument from thermodynamic equilibrium in planetary atmospheres was provided by Wildt (1937), who pointed out that the absence of detectable amounts of higher hydrocarbons, both saturated and unsaturated, in the atmospheres of the Jovian planets was an argument for the presence of a great excess of molecular hydrogen, a gas which had not been detected at that time. Further indirect evidence for the presence of low-mass constituents in the Jovian atmosphere was provided by observations of the occultation of σ Arietus, by Baum and Code (1953); but not until fairly

recently were the quadrupole lines of H_2 observed directly in the photographic infrared (Kiess, Corliss, and Kiess, 1960). Urey (1959) has argued from the large abundance of carbon dioxide on Venus that substantial quantities of CH_4 , NH_3 , H_2 , N_2O , and other oxides of nitrogen would not be expected on that planet. A very low upper limit on the allowed abundance of oxides of nitrogen in the Martian atmosphere was set by Sagan, Hanst and Young (1965) in a discussion of combined photochemical and thermodynamic equilibria. Wildt (1937) and Urey (1952), while performing thermodynamic equilibrium calculations in a planetary atmospheres context, stressed the possibility of significant departures from thermodynamic equilibrium.

Some more recent studies have suggested that thermodynamic equilibrium calculations may, under certain circumstances, be a useful approximation to reality. Thermodynamic equilibrium calculations by Dayhoff, Lippincott, and Eck (1964) for the relative abundances of a large number of compounds of biological interest have been performed under a wide variety of assumed conditions of temperature, pressure, and elemental composition. These machine computations have revealed a possible mechanism for the abiological formation of polycyclic aromatic hydrocarbons (asphaltic tars) and have shown the existence of an oxidation threshold where free oxygen appears and at which all but the simplest organic compounds disappear. Studier, Hayatsu, and Anders (1965) have recently called attention to a remarkable agreement between the distribution of organic compounds observed in the carbonaceous chondrites, and as we predicted from thermodynamic equilibrium calculations in the range of elemental abundances where asphaltic tars are expected. This success has encouraged us to extend thermodynamic equilibrium calculations to planetary atmospheres.

2. Methods of Calculation

Our computational procedures have been described previously (Dayhoff, Lippincott, and Eck, 1964) following a well-known method (White, Johnson, and Dantzig, 1958) which minimizes the free energy of the system, simultaneously satisfying the equilibrium constants of all possible reactions. The calculations were performed on an IBM 7094 computer.

At thermodynamic equilibrium, the distribution of molecular species is independent of the specific reaction pathways by which equilibrium is attained. The molecular balance depends upon the relative elemental abundances, pressure, temperature, and the standard free energies of formation of the compounds. All compounds present in significant concentrations must be included in order to arrive at the correct molecular distribution. Fortunately, there are only a small number of these major compounds; they are simple in structure and well-known to chemists. For a real system to approach equilibrium, there need be only one reversible reaction pathway which leads from the major constituents to the other compounds.

In our computations on planetary atmospheres, we explored the entire range of possible relative atomic compositions using a large number of molecular species in order to locate any specific ranges of atomic abundances which would be simultaneously compatible with all the existing estimates of molecular abundances or their upper limits. The results are presented in ternary diagrams which cover all possible proportions of the elements C, H, and O. Compounds of nitrogen can also be represented as a projection on this diagram. The astronomical observation of a specific relative abundance for a given molecular species will, in general, define a line on the ternary diagram. An upper or lower concentration limit will exclude a certain area. The application of several such constraints may operate in a mutually inconsistent manner, so that no point on the diagram simultaneously satisfies all observations. In such a case, we would conclude that thermodynamic equilibrium is not attained in this atmosphere. On the other hand, if the application of several such constraints can be performed in a consistent manner, yielding a region of the ternary diagram simultaneously compatible with all constraints, we might tentatively conclude that thermodynamic equilibrium is a useful first approximation to the atmosphere in question. Each point in the allowed region of the diagram would then correspond to a possible elemental composition in the planetary atmosphere. The calculations would place some upper and lower limits on the allowable relative abundances of other possible constituents of the atmosphere.

3. Observational Uncertainties

The initial conditions for these computations are based upon the results of astronomical spectroscopy. It is not a trivial matter to derive relative abundances from such observations. The conversion of equivalent widths to absolute abundances requires careful laboratory calibrations. Intercomparisons of abundances derived for the same molecule at different wavelengths must allow for the possibility that the effective reflecting atmospheric level is a function of wavelength. Existing observational discrepancies on, for example, the abundance of water vapor in the atmosphere of Venus (see, e.g., the discussion by Sagan and Kellogg, 1963) underscore the uncertainties of such reductions. Intercomparisons of the relative abundances of different molecular species is even more uncertain. Determinations of absolute abundances is complicated by the possibility of multiple scattering (see Chamberlain, 1962; 1965). Observations of very weak lines of a given molecular species may refer to substantial depths in the planetary atmosphere, where ultraviolet photodissociation is not a dominant process. Observations of strong lines may refer to high altitudes alone, where the relative distribution of molecular species may be far from that predicted by thermodynamic equilibrium. However, the species must be derivable from some compounds present extensively in the atmosphere.

Bearing these cautions in mind, we now proceed to a discussion of thermodynamic equilibrium in the atmospheres of the planets Earth, Venus, Mars, and Jupiter. In general, the conclusions we will draw will not be sensitively dependent on the exact values of the adopted mixing ratios of minor constituents; and variations in these mixing ratios even by several orders of magnitude will leave most of the results unchanged. Thus, despite the uncertainties in the observational material, it will nevertheless be possible to draw conclusions of some significance.

4. The Earth

As a test of our computation procedures, we first consider thermodynamic equilibrium in the terrestrial atmosphere. In addition to its major components, the atmosphere of the Earth contains a variety of compounds of diverse origins: methane and smaller amounts of other hydrocarbons from natural gas, petroleum and asphalt; sulfur dioxide, hydrogen sulfide, carbon monoxide and nitrogen oxides from vulcanism; terpenes and other volatile organics from vegetation; methane, N_2O , and hydrogen sulfide from the metabolic processes of microorganisms; ozone and nitrogen oxides from lightning and solar radiation, etc. Table 1 shows the amounts of some of these compounds which have been detected in the open air.

With the relative atomic composition and physical parameters implied by Table 1 we have calculated the chemical composition which the Earth's atmosphere should have in thermodynamic equilibrium. The results are displayed in Table 2. We see that the relative abundances of major constituents in the terrestrial atmosphere are completely compatible with thermodynamic equilibrium. The minor constituents, however, cannot exist at thermodynamic equilibrium in the presence of excess oxygen. At equilibrium, any carbon in the atmosphere would occur as CO_2 , and any sulfur as SO_3 . Since there is no reason to think that less oxidized materials are presently accumulating in the atmosphere, they must be in approximate dynamic equilibrium, being removed or destroyed as fast as they are being added. That is, even at the prevailing low temperatures there is a definite tendency for the Earth's atmosphere to approach thermodynamic equilibrium. Such reactions are no doubt catalyzed by solar radiation and by dust particles and droplets suspended in the air. This constitutes a specific example of our general contention that, even in the presence of special mechanisms which favor the production of certain specific compounds, the overall tendency will be for a planetary atmosphere to approach thermodynamic equilibrium. If the major components of the atmosphere were methane, hydrogen, water, and nitrogen (or ammonia) instead of the present composition, the details of the radiation-coupled reactions would be different, but the overall trend to thermodynamic equilibrium would still be present. It is not obvious at what

relative rates unstable prebiological compounds such as amino acids would be produced by nonequilibrium processes and degraded by the environment and by their interaction with each other.

The observation of complex, very energetic compounds such as terpenes in the contemporary atmosphere of the Earth would lead an extra-terrestrial observer to suspect the existence of life here. The lack of any obvious simple mechanism to form terpenes, either by thermodynamic equilibrium at some plausible combination of elemental composition, temperature, and pressure, or by photochemical reactions or other physical processes, would indicate the presence of biological activity on the Earth -- even in the absence of any other signs of life. Even the presence in the observed large amounts of much simpler, but highly reduced, compounds such as methane in our oxidizing atmosphere would be suggestive of life on Earth.

5. Venus

The adopted initial atmospheric composition of Venus is given in Table 3, constructed from the references there given. Some comments should be made on the values adopted for $[\text{CO}_2]^*$, $[\text{H}_2\text{O}]$, and $[\text{CO}]$. The 5% volume mixing ratio of carbon dioxide adopted in our analysis is a fairly conventional one, derived from Spinrad's (1962a) observations, and dependent upon the assumption that the near infrared CO_2 bands are formed by single scattering. Chamberlain (1965) has recently shown that the introduction of multiple scattering might reduce the CO_2 mixing ratio by as much as an order of magnitude. The precise reduction factor is at the present time unknown. However, similar correction factors for multiple scattering will be applicable to all other abundances and abundance upper limits of Table 3, and relative abundances of all minor constituents should be approximately those of Table 3, whether single or multiple scattering is assumed. The relative abundance of the major atmospheric constituents -- probably N_2 , with some admixture of noble gases -- would then vary from 95% to larger values. Such variations have no significant effects on our conclusions.

*In the following discussion square brackets denote volume mixing ratios.

Reported observations of water vapor in the spectrum of Venus have been made by Dollfus (1964) and by Bottema Plummer, and Strong (1964). The derived values of the water vapor mixing ratio depend on the scattering mechanism and the total atmospheric pressure at the effective level of band formation. The range of derived mixing ratio values shown in Table 3 is consistent (Sagan and Kellogg, 1963; Chamberlain, 1965) with a previous upper abundance limit on water vapor established by Spinrad (1962b). However, the water vapor which is observed by infrared spectroscopy is primarily in the region of the clouds. Since there is now fairly convincing evidence that the clouds are condensed water (Bottema, Plummer, Strong, and Zander, 1964; Sagan and Pollack, 1965), it is in general invalid to assume that the water vapor mixing ratio in the vicinity of the clouds applies to the lower Cytherean atmosphere. The amount of condensed water in the clouds required to explain the near infrared reflection spectrum of Bottema, Plummer, Strong, and Zander (1964), and also to explain the millimeter attenuation of the surface thermal microwave emission, is 1 gm cm^{-2} or less convective atmospheres which can maintain such clouds at the observed pressures and temperatures must have volume mixing ratios of water vapor in their lower reaches of a few times 10^{-4} (Sagan and Pollack, 1965). The water vapor mixing ratio may therefore range from a few times 10^{-6} in the vicinity of the clouds and above, to a few times 10^{-4} near the Cytherean surface.

The quantities of carbon monoxide in the spectrum of Venus observed by Moroz (1965), if related to a 1 atm pressure level near the cloudtops, give a volume mixing ratio $\sim 10^{-6}$. However, CO is a principal carbon dioxide photodissociation product, and amounts of carbon monoxide almost comparable to those reported by Moroz may arise from photodissociation (Moroz, 1965; Shimi, 1963). If the observed carbon monoxide abundance is significantly in excess of that expected from photochemical equilibrium, then a CO mixing ratio of 10^{-6} may be extended down to the lower Cytherean atmosphere. If, on the other hand, the observed carbon monoxide is due principally to photochemical processes, then the carbon monoxide abundance in the lower Cytherean atmosphere will be substantially less.

We adopt a Venus surface temperature of 700°K , and a surface pressure of 50 atm (Sagan, 1962; Pollack and Sagan, 1965). Moderately large departures from these values will not alter our conclusions significantly. In particular, the calculations performed at such pressures and temperatures should, because of the relative pressure and temperature insensitivity of the reaction thresholds and because of atmospheric convection, be applicable to a major fraction of the total mass of the Cytherean atmosphere.

A C-H-O ternary diagram is presented in Figure 1. The corners of the triangle represent pure carbon, pure hydrogen, and pure oxygen. The positions of pure carbon dioxide and pure water are also indicated. The line connecting the CO_2 and H_2O points is an oxidation threshold, which divides the diagram into an upper, reducing, portion in which sizeable numbers of different species of organic molecules may be expected at thermodynamic equilibrium, and a lower, oxidizing, portion, in which such molecules are unstable, and in which free oxygen exists. Also indicated in Figure 1 is the asphaltic tar threshold. Above this line, polycyclic aromatic hydrocarbons occur at thermodynamic equilibrium (in the absence of a mechanism to form graphite); below it, they are excluded.

The ammonia upper abundance limit of 4×10^{-7} restricts the atmosphere at thermodynamic equilibrium to the region below the line marked NH_3 in Figure 1. This line lies in the reducing region of the diagram, slightly above the oxidation threshold, and curves away from the carbon dioxide point, as illustrated. This one spectroscopic upper limit therefore excludes most of the reducing portion of the ternary diagram. The concentration of ammonia varies so rapidly in the region of the illustrated line that the width of this line represents several orders of magnitude in concentration -- an example of the insensitivity of these calculations to exact relative abundances.

We next consider the upper abundance limit on methane. This concentration yields a curve which closely follows the C-O edge until the CO_2 point is reached, whereupon it swings away and follows the oxidation threshold. The CH_4 curve is even more constraining than the NH_3 curve, and excludes almost the entire reducing portion of the diagram.

We now consider the constraints imposed by the water vapor mixing ratio. Figure 1 shows a line for $[\text{H}_2\text{O}] = 10^{-5}$, which, in the oxidizing region of the diagram, lies very close to the C-O edge. The curves for mixing ratios an order of magnitude larger would lie very close to the curve shown. Thus, the searches for methane, ammonia, and water have restricted the equilibrium chemistry of the Cytherean atmosphere to a very narrow region of the ternary diagram, hugging the C-O edge throughout the oxidizing region, and extending very slightly into the reducing region.

In Figure 2, where the Venus ternary diagram has been magnified 100 times, the line corresponding to a carbon monoxide volume mixing ratio of $[\text{CO}] = 10^{-6}$ is indicated. Since part of the observed carbon monoxide may be due to CO_2 photodissociation, the fraction of carbon monoxide relevant for thermodynamic equilibrium calculations will be somewhat less than 10^{-6} . The observations therefore further restrict the ternary diagram to the area below the line $[\text{CO}] = 10^{-6}$. We see that this line closely parallels the oxidation threshold. As with the NH_3 and CH_4 lines, the width of the $[\text{CO}] = 10^{-6}$ line actually corresponds to many orders of magnitude in the mixing ratio. The lower side of this line corresponds to the oxygen upper abundance limit, 8×10^{-5} . If observations can establish any detectable amount ($> 10^{-16}$) of CO in the lower atmosphere, the amount of O_2 must be extremely small. Conversely, if $[\text{O}_2] \geq 8 \times 10^{-5}$, then $[\text{CO}] \leq 4 \times 10^{-17}$. At any rate the upper limits on Cytherean $[\text{CO}]$ and $[\text{O}_2]$ restrict the thermodynamic equilibrium composition to the thickness of the thick slant line in Figures 1 and 2.

In Figure 2 is displayed a line above which a solid phase of graphite might form. The composition of the atmosphere is clearly incompatible with solid carbon in equilibrium in the atmosphere or on the surface. Similarly, any equilibrium accumulation of polycyclic aromatics, or any other hydrocarbons, is forbidden, a conclusion reached for a much smaller array of compounds by

Mueller (1964). The asphalt threshold seen in Figure 1 lies well into the region excluded by the unsuccessful searches for CH_4 and NH_3 . This conclusion is separately confirmed by the CO upper abundance limits, and is independent of the quantity of water present. This exclusion of hydrocarbons is relevant to some models of the Venus atmosphere which invoke hydrocarbon clouds (Kaplan, 1963). The absence of polycyclic aromatic hydrocarbons from the surface is consistent both with passive observations of the microwave phase effect and with active radar observations of Venus (Pollack and Sagan, 1965).

Table 4 shows the chemical composition of the atmosphere of Venus, at two extremes of CO and O_2 concentration, predicted from thermodynamic equilibrium. If the measured value of CO depends entirely on photodissociation, then there may be as much as $[\text{O}_2] = 8 \times 10^{-5}$. We emphasize again that any error in the relative abundance of nitrogen would cause negligible deviation in the relative abundance of carbon compounds, and indeed, would have very little effect on the calculated ammonia abundance. A smaller amount of nitrogen would only make the ammonia concentration a less severe restriction on the actual composition (cf. Figure 1). The constituents shown at the bottom of the table would be formed if 0.0001% sulfur were added to the atmosphere.

Formaldehyde was once suspected in the atmosphere of Venus (Wildt, 1940) but at thermodynamic equilibrium with the detected amounts of water and the observational upper limit on methane, the concentration of formaldehyde would be negligible. Non-equilibrium reactions in simulated Cytherean atmospheres are known, however, to produce formaldehyde (Sagan and Miller, 1960).

We now explore some other consequences of the foregoing abundances. Possible atmospheric compositions are limited to a segment of the $[\text{CO}] = 10^{-6}$ curve near $[\text{H}_2\text{O}] = 10^{-4}$, as shown in Figure 2. The composition of Venus is then specified by a segment of one thick line in the ternary diagram, near the CO_2 composition point, and slightly on the reducing side of the oxidation threshold. This conclusion is consistent with all abundance determinations and upper limits. The fact that none of the observations are mutually

inconsistent suggests that the atmosphere of Venus is, on the whole, close to that predicted by thermodynamic equilibrium.

In this case some remarks can be made on the original composition and subsequent evolution of the Cytherean atmosphere. From cosmic abundance considerations, it is clear that the atmosphere must originally have had a composition putting it in the extreme right-hand corner of the ternary diagram (cf. Figure 3). The chemical evolution of the atmosphere occurs initially through the escape of atomic hydrogen from the Cytherean exosphere. Escape of significant quantities of carbon or oxygen from Venus during geological time seems quite unlikely (see, e.g., Sagan, 1966). Any differential escape is correspondingly unlikely. The track along which the atmospheric composition of Venus evolves depends then on the initial carbon-to-oxygen abundance ratio. The cosmic [C/O] abundance ratio has been assigned a value of 0.20 by Suess and Urey (1956), while Cameron (1959) has suggested a value of 0.44 and more recently (1963), 0.67. For a ratio of 0.20 or 0.44, the evolution would proceed along the lines ending at α or β (Figure 3), in a region where free oxygen exists. For an initial [C/O] ratio of 0.67, the evolutionary track ends at γ , in the reducing portion of the diagram. The actual present atmosphere of Venus appears to have a [C/O] ratio that differs from that of CO_2 by a factor of 10^{-5} or less. If the initial atmosphere had a C/O ratio anywhere between 0.2 and 0.7, the probability that the present ratio would be as close to CO_2 as it appears to be, through no other process but the loss of hydrogen, would then be about 4×10^{-5} . This seems quite unlikely. However, if oxygen were also depleted from the atmosphere, then over a wide range of initial [C/O] ratios the atmosphere would finally settle at the CO_2 point. Since it seems unlikely that oxygen has escaped from Venus, the depletion of atmospheric oxygen must be attributed to chemical reactions, and the surface material of Venus, initially reducing, must now be partially oxidized. An independent argument for the presence of extensive oxygen sinks on the surface of Venus has been offered by Sagan (1966), in an attempt to explain the differential abundance of water on Venus and on Earth by differential rates of hydrogen escape.

Typical evolutionary tracks would then have resembled those illustrated by the arrows in Figure 3. Due to loss of hydrogen, the atmosphere evolves away from the hydrogen corner of the ternary diagram until it intersects the oxidation threshold. As soon as free oxygen is produced, it combines with surface material and the evolutionary track turns abruptly upward, following the oxidation threshold toward the CO_2 point.

If the initial $[\text{C}/\text{O}]$ ratio were 0.44, the initial evolutionary track due to the loss of hydrogen would twice intersect the graphite threshold (cf. Figure 3). There would then be an interlude in the evolutionary history of Venus in which graphite might be present. However, the activation energy for the formation of graphite is so great that it would very likely never precipitate directly. Such a track would follow the line marked H- β until it intersected the oxidation threshold, whereupon it would turn and proceed towards the CO_2 point.

If the initial $[\text{C}/\text{O}]$ ratio were 0.67, the evolutionary track would follow the line marked H- γ in Figure 3. This line intersects the graphite threshold once, and never crosses the oxidation threshold. The atmosphere would then always remain in the reducing portion of the diagram. In order to end at the CO_2 point, some mechanism must exist for the removal of carbon from the atmosphere. The asphalt threshold is not intersected, so no formation of polycyclic aromatics would be expected. If the removal of carbon from the atmosphere were possible -- for example, by the precipitation of graphite or other organic material -- the atmosphere would then evolve along the graphite line in Figure 2. The composition would end significantly above the actual composition point, in conflict with the spectroscopic observations.

Thus, it is probable that Venus began its evolutionary history with a $[\text{C}/\text{O}]$ abundance ratio ≤ 0.5 and that its evolutionary track in the ternary diagram was directed towards the CO_2 point by the simultaneous loss of hydrogen to space and oxygen to the surface.

6. Mars

Because of the low temperature and relative ultraviolet transparency of the Martian atmosphere, the applicability of thermodynamic equilibrium calculations should be more limited than in the case of Venus. The character of the equilibrium state is nevertheless of interest. A surface temperature of 240°K and a surface pressure of 50 mb was assumed for the Martian atmosphere (Kaplan, Munch, and Spinrad, 1964; Hanst and Swan, 1965). Our remarks on the insensitivity of the calculations to the precise values of pressures and temperatures apply again to this case and a surface pressure of 10 mb would not appreciably alter the results. The assumed atmospheric composition is shown in Table 5. The atmosphere is assumed to be approximately 90% molecular nitrogen and noble gases. The remainder of the atmosphere is expected to be compounds of carbon, oxygen, nitrogen, and hydrogen. The anticipated CO_2 mixing ratio is not known at the present time. Its value is tied closely to the value of the absolute surface pressure (Kaplan, Munch, and Spinrad, 1964; Hanst and Swan, 1965; Chamberlain and Hunten, 1965). The values of Table 5 correspond to the range of suggested values at the time of writing. As in the case of Venus, the absolute value of the CO_2 mixing ratio is not critical in computations of the relative values of the mixing ratios of minor constituents.

The ammonia upper abundance limit excludes the highly reducing portion of the ternary diagram, as shown in Figure 4. The upper limit on the methane mixing ratio imposes a more significant boundary condition, restricting the range of possible atmospheric compositions essentially to the portion below the oxidation threshold, with the exception of a small section of the reducing part of the diagram along the C-O axis. The carbon monoxide upper limit places no new constraint on the diagram; the upper limits on oxygen and ozone eliminate a region near the oxygen corner of the ternary diagram. The estimate for the abundance of water vapor in the Martian atmosphere places the equilibrium composition along a curve which crosses the oxidation threshold not far from the CO_2 composition point, quite analogous to the situation for Venus. This line quickly enters the region forbidden by the methane abundance upper limits.

The composition of the Martian atmosphere is therefore restricted to the solid portion of the water line of Figure 4, encompassing a small region of the oxidizing part of the diagram, and a very small segment, near CO_2 , in the reducing region.

It has been suggested that various oxides of nitrogen are produced by non-equilibrium processes in the atmosphere of Mars, and that these oxides of nitrogen play a major role in the total chemistry of the atmosphere and surface of Mars (Kiess, Karrer, and Kiess, 1960; 1963). In Table 6, we present some calculations of the equilibrium abundance of various compounds in the Martian atmosphere at representative points in Figure 4. Even in the favorable case where oxygen is present as a major constituent, we see that the resulting abundances of the oxides of nitrogen are extraordinarily small. These conclusions are consistent with those of Sagan, Hanst, and Young (1965), who used a much smaller array of equilibrium reactions than are implicitly included here and considered also photochemical production and breakdown of these oxides. The present results place the equilibrium upper limit on NO_2 several orders of magnitude below the previous calculated values (Sagan, Hanst, and Young, 1964); and below the most recent observational upper limits (Marshall, 1964) by the same factor.

Further conclusions on the nature of the Martian atmosphere in thermodynamic equilibrium could be drawn if reliable estimates were performed of the abundances of such molecules as CO , CH_4 , H_2S , SO_3 , or O_2 . The limited data available are consistent with the existence of thermodynamic equilibrium, but do not demonstrate thermodynamic equilibrium. As in the case of the atmosphere of Venus, the escape of hydrogen, plus the depletion of atmospheric oxygen (here, either by reaction with the crust or by escape from the planetary exosphere) could have led to an evolutionary track in the ternary diagram approximately along the oxidation threshold. The present information is consistent with a small proportion of molecular oxygen in the atmosphere, although the atmosphere could also be slightly reducing. The possibility of an equilibrium abundance of organic molecules on the Martian surface depends in part

on this question of which side of the oxidation threshold the total atmospheric composition lies. The long-term stability of organic molecules on the Martian surface also depends on the rate of ultraviolet photodissociation, if the atmosphere is transparent in the near ultraviolet (see Sagan and Kellogg, 1963).

7. Jupiter

The assumed composition of the Jovian atmosphere is shown in Table 7. Because there is no way for water to escape from Jupiter, it must be present in significant amounts below the visible clouds. At the clouds it is frozen out because of the low temperatures. We have arbitrarily assumed that the H_2O mixing ratio below the clouds is comparable to that of methane and ammonia. In the clouds it is fixed by the vapor pressure of ice. The thermodynamic equilibrium calculations were carried out for two sets of pressures and temperatures -- in the first case, for a pressure of 1 atm and a temperature of $200^\circ K$, corresponding approximately to the region of the clouds (Spinrad and Trafton, 1962; Owen, 1965); and in the second case, for a pressure of 1000 atm and a temperature of $350^\circ K$. This point corresponds to the level of the hypothesized water droplet clouds, predicted by Gallet (1963; see also Wildt, Smith, Salpeter, and Cameron, 1963). The concentrations of the major resulting constituents of the Jovian atmosphere at thermodynamic equilibrium are presented in Table 8. The addition of 10^{-6} sulfur, would give the other constituents shown. As first pointed out by Wildt (1957), the large excess of molecular hydrogen works to greatly reduce the concentration of even simple organic molecules at thermodynamic equilibrium. Essentially all carbon is present as methane, all oxygen as water, and all nitrogen as ammonia. The total atmospheric composition gives a point very near the H corner of the ternary diagram. The relative abundance of molecular nitrogen increases fairly rapidly with temperature, but even at $350^\circ K$, it is still a very minor atmospheric constituent.

Despite the extremely low computed mixing ratios of even the simplest organic molecules, there is nevertheless reason to believe that organic molecules may be present in the atmosphere of Jupiter. The bright and contrasting colors of the bands, belts, and spots -- particularly, the Great Red Spot -- surely betoken differences in molecular composition in the vicinity of the Jovian clouds. While some small concentration of minerals may be expected in this region, due to micrometeoritic infall, it seems unlikely that such materials are differentially distributed over the Jovian clouds; a more reasonable source of chromophores would appear to arise from the major atmospheric constituents, and therefore, be organic materials (see e.g., Urey, 1952; Sagan, 1963). Experiments attempting to simulate the Jovian atmosphere, in which a mixture of the major gases is supplied with energy from a corona discharge, have succeeded in producing such simple organic molecules as HCN, CH_3CN , C_2H_2 , C_2H_4 , and C_2H_6 (Sagan and Miller, 1960). Their interaction products may be brightly colored. Indeed, in experiments relevant to the origin of life on Earth, mixtures of materials which are in effect simulated Jovian environments have been subjected to a wide array of energy sources, and organic molecules have been consistently produced in high yield, provided only that the over-all conditions were reducing (see, e.g., Fox, 1965). Electric discharges and solar ultraviolet light -- both to be expected in the vicinity of the Jovian clouds -- will lead to the production of organic molecules in abundances far greater than those expected from thermodynamic equilibrium considerations.

REFERENCES

- Allen, C. W. 1963, "Astrophysical Quantities," second ed. (London: The Athlone Press, 1963).
- Baum, W. A., and Code, A. D. 1953, A.J., 58, 108.
- Bottema, M., Plummer, W., and Strong, J. 1964, Ap.J., 139, 1021.
- Bottema, M., Plummer, W., Strong, J., and Zander, R. 1964, Ap.J. 140, 1640.
- Cameron, A. G. W. 1959, Ap.J., 129, 676.
- _____. 1963, Lecture notes on Nuclear Astrophysics, Yale University, unpublished.
- Chamberlain, J. W. 1962, Appendix 9, in "The Atmospheres of Mars and Venus," ed. W. W. Kellogg and C. Sagan (Washington: National Academy of Sciences -- National Research Council, Publication 944).
- _____. 1965, Ap.J., 136, 582.
- Chamberlain, J. W., and Hunten, D. M. 1965, Revs. Geophys., 3, 299.
- Dayhoff, M. O., Lippincott, E. R., and Eck, R. V. 1964, Science, 146, 1461.
- Dollfus, A. 1963, Compt. Rend., 256, 3250.
- Fox, S., ed. 1965, "The Origins of Prebiological Systems," (New York: Academic Press).
- Gallet, R. M. 1963, private communication.
- Hanst, P. L., and Swan, P. R. 1965, Icarus, 4, 353.
- Hutchinson, G. E. 1954, "The Biochemistry of the Terrestrial Atmosphere," in The Earth as a Planet, ed. G. P. Kuiper (Chicago: University of Chicago Press) 371.
- Kaplan, L. D. 1963, J. Quant. Spectrosc. Radiat. Transfer, 3, 537.
- Kaplan, L. D., Münch, G., and Spinrad, H. 1964, Ap.J., 139, 1.
- Kiess, C. C., Corliss, C. H., and Kiess, H. K. 1960, Ap.J., 132, 221.
- Kiess, C. C., Karrer, S., and Kiess, H. K. 1960, P.A.S.P., 72, 256.
- _____. 1963, P.A.S.P., 75, 50.

- Kuiper, G. P. 1952, "Planetary Atmospheres and Their Origins," in The Atmospheres of the Earth and Planets, ed. G. P. Kuiper (Chicago: University of Chicago Press) 306.
- Marshall, J. V. 1964, Comm. Lunar and Planet. Lab., 2, 167.
- Moroz, V. I. 1965, Soviet Astronomy - A.J., 8, 566.
- Mueller, R. R. 1964, Icarus, 3, 285.
- Owen, T. C. 1965, Ap. J., 141, 444.
- Owen, T. C., and Kuiper, G. P. 1964, Comm. Lunar and Planet. Lab., 2, 113.
- Pollack, J. B., and Sagan, C. 1965, Icarus, 4, 62.
- Sagan, C. 1962, Icarus, 1, 151.
- _____. 1963, Proc. XI Internat. Astrophys. Colloq., Liège, 506.
- _____. 1966, "Origins of the Atmospheres of the Earth and Planets," in International Dictionary of Geophysics, S. K. Runcorn, ed. (London: Pergamon Press, in press).
- Sagan, C., Hanst, P. L., and Young, A. T. 1965, Planet. Space Sci., 13, 73.
- Sagan, C., and Kellogg, W. W. 1963, Ann. Rev. Astron. Ap., 1, ed. L. Goldberg, 235.
- Sagan, C., and Miller, S. L. 1960, A.J., 65, 499.
- Sagan, C., and Pollack, J. B. 1965, Icarus, to be published.
- Shimizu, M. 1963, Planet. Space Sci., 11, 269.
- Spinrad, H. 1962a, P.A.S.P., 74, 156.
- _____. 1962b, Icarus, 1, 266.
- Spinrad, H., and Richardson, E. H. 1965, Ap.J., 141, 282.
- Spinrad, H., and Trafton, L. M. 1962, Icarus, 2, 19.
- Studier, M. H., Hayatsu, R., and Anders, E., 1965, Science, 149, 1455.
- Suess, H. E., and Urey, H. C. 1956, Revs. Mod. Phys., 28, 53.
- Urey, H. C. 1952, "The Planets; Their Origin and Development," (New Haven: Yale University Press, 1952).
- _____. 1959, Handbuch der Physik, 52, 363.
- White, W. B., Johnson, S. M., and Dantzig, G. B., J. Chem. Phys. 28, 751.
- Wildt, R. 1937, Ap.J., 86, 321.
- _____. 1940, Ap.J., 92, 247.
- Wildt, R., Smith, H. J., Salpeter, E. E., and Cameron, A. G. W. 1963, Physics Today, 16, 19.

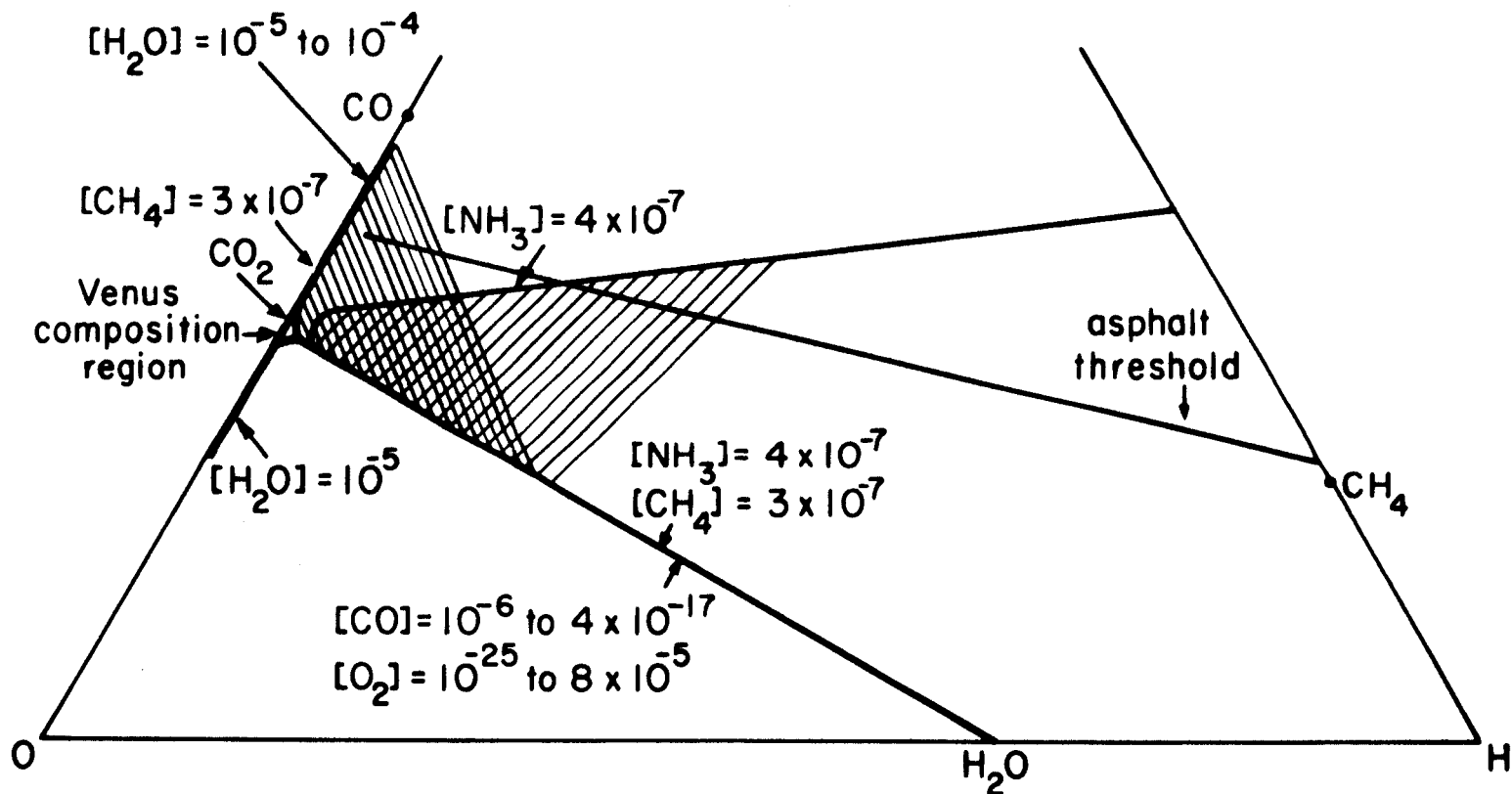


Fig. 1.--C-H-O ternary diagram for the Cytherean atmosphere. The shaded regions are excluded at thermodynamic equilibrium by the observational upper limits on $[CH_4]$ and $[NH_3]$. For clarity the shading has not been extended through the entire excluded region. The adopted values of $[H_2O]$ and the observed upper limits on $[CO]$ and $[O_2]$ are denoted by lines.

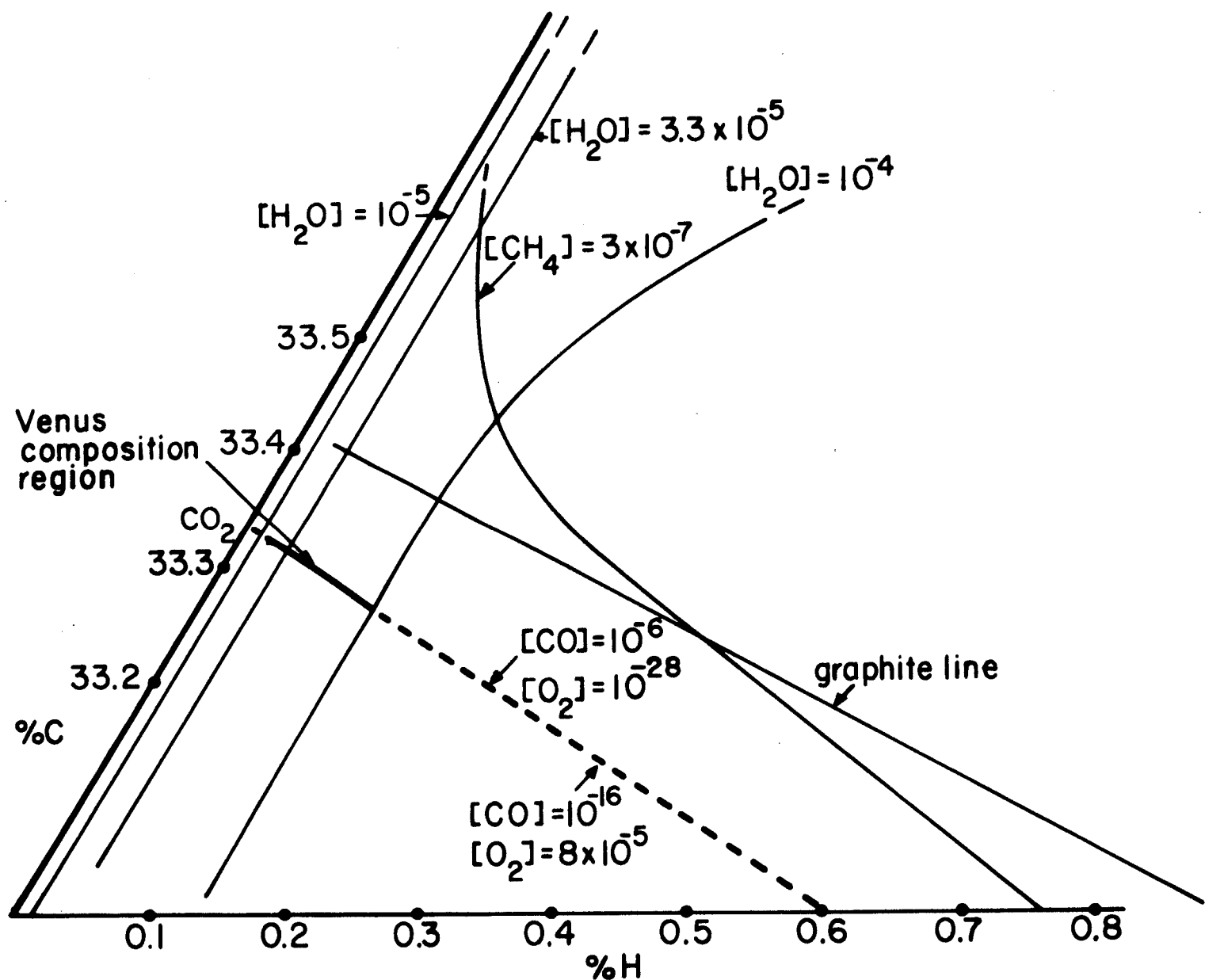


Fig. 2.--C-H-O ternary diagram of Figure 1, here enlarged one hundred-fold.

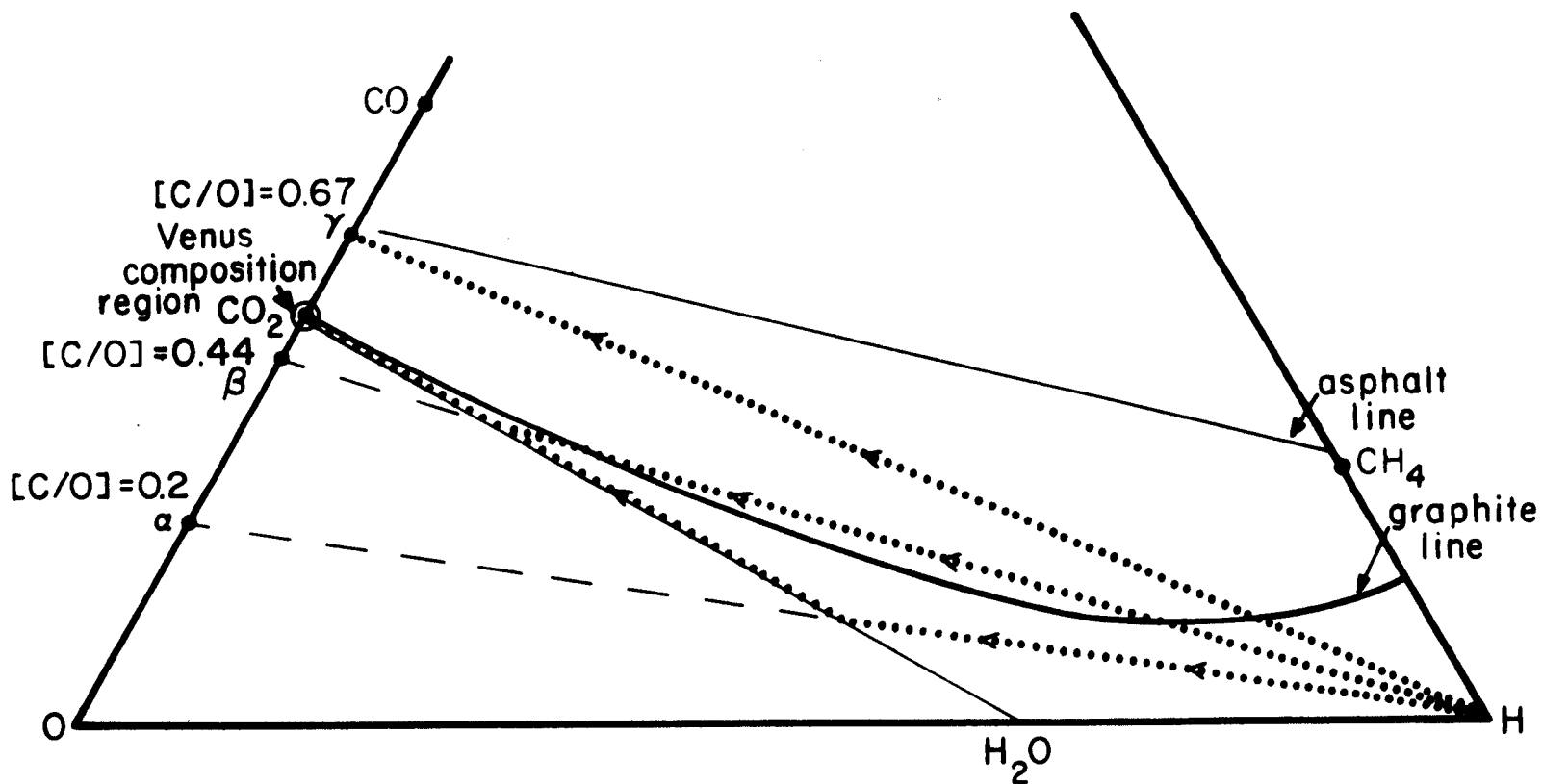


Fig. 3.--Evolution of the atmosphere of Venus. Initial conditions assumed include a large excess of H₂, and three [C/O] abundance ratios. Loss of hydrogen alone causes evolution along the tracks H-α, H-β, and H-γ. Additional loss of oxygen would tend to deflect the lower two evolutionary tracks along the oxidation threshold, the line CO₂ - H₂O. Tracks with initial [C/O] > 0.5 must lose carbon to arrive at the present composition; but the precipitation of graphite or polycyclic aromatics would not lead to the contemporary value of [C/O].

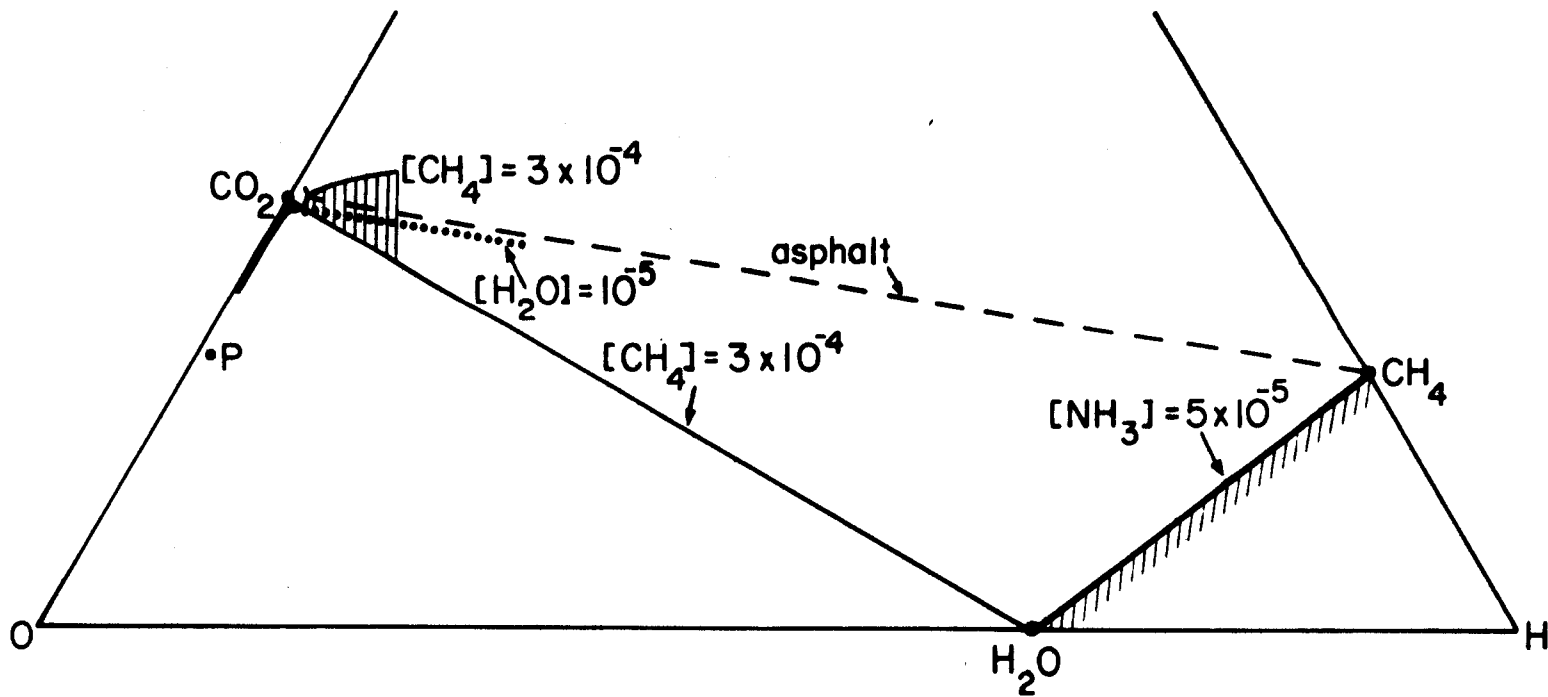


Fig. 4.--C-H-O ternary diagram for the Martian atmosphere. Areas not excluded by observation are localized about the solid portion of the water line, near the CO₂ point. One additional abundance determination can specify the thermodynamic equilibrium composition.

TABLE 1

ADOPTED CHEMICAL COMPOSITION AND PHYSICAL PARAMETERS OF THE
 TERRESTRIAL ATMOSPHERE (VOLUME MIXING RATIOS ARE DISPLAYED)

O_2	0.2095
O_3	10^{-8} to 10^{-7}
H_2	5×10^{-7}
H_2O	1×10^{-3} to 2.8×10^{-2}
N_2	0.7808
NO_2	5×10^{-10} to 2×10^{-8}
N_2O	5×10^{-7}
NH_3	1.7×10^{-8}
SO_2	1×10^{-6}
CO_2	3.30×10^{-4}
CH_4	1.5×10^{-6}
A	1×10^{-2}
He	5.2×10^{-6}
Ne	1.8×10^{-5}
Pressure	1.0 atm
Temperature	280°K

(References: Allen 1963; Hutchinson 1954)

TABLE 2

THE THERMODYNAMIC EQUILIBRIUM COMPOSITION OF THE TERRESTRIAL
ATMOSPHERE, COMPUTED FROM THE TOTAL ELEMENTAL COMPOSITION OF TABLE 1*

N ₂	0.78	H ₂ S	0
O ₂	0.21	CO	0
A	0.01	H ₂	0
H ₂ O	1.0 × 10 ⁻³	CS ₂	0
CO ₂	3.33 × 10 ⁻⁴	CH ₄	0
SO ₃	1 × 10 ⁻⁶	CH ₃ SH	0
HNO ₃	5.1 × 10 ⁻¹⁰	COS	0
NO ₂	6.2 × 10 ⁻¹¹	Benzene	0
NO	2.9 × 10 ⁻¹⁷	Formic acid	0
H ₂ SO ₄	4.1 × 10 ⁻¹⁸	HCN	0
SO ₂	8.6 × 10 ⁻²⁰	NH ₃	0
N ₂ O	3.8 × 10 ⁻²⁰	Formaldehyde	0
Ozone	6.6 × 10 ⁻³²	Methenol	0

* A mixing ratio of zero is an abbreviation for a value < 10⁻³⁵.

TABLE 3

ADOPTED CHEMICAL COMPOSITION OF THE CYTHEREAN ATMOSPHERE
(VOLUME MIXING RATIOS ARE DISPLAYED)

O_2	$< 8 \times 10^{-5}$
H_2O	1×10^{-6} to 1×10^{-4}
N_2	~ 0.95
NO_2	$< 1 \times 10^{-6}$
N_2O	$< 4 \times 10^{-7}$
NH_3	$< 4 \times 10^{-7}$
CO_2	$\sim 5 \times 10^{-2}$
CO	$\leq 1 \times 10^{-6}$
CH_4	$< 3 \times 10^{-7}$
C_2H_4	$< 2 \times 10^{-7}$
C_2H_6	$< 1 \times 10^{-7}$
$HCHO$	$< 3 \times 10^{-8}$
Pressure	~ 50 atm
Temperature	$700^\circ K$

(References: Kuiper 1952; Spinrad 1962a, 1962b; Sinrad and Richardson, 1965; Bottema, Plummer, and Stron 1964; Dollfus 1964; Sagan and Pollack 1965)

TABLE 4

THERMODYNAMIC EQUILIBRIUM COMPOSITION OF THE CYTHEREAN ATMOSPHERE,
 COMPUTED FROM TOTAL ELEMENTAL COMPOSITION OF TABLE 2*

<u>Compound</u>	<u>Oxidizing Limit</u>	<u>Reducing Limit</u>
O ₂	8×10^{-5}	1×10^{-25}
CO	4×10^{-17}	1×10^{-6}
CO ₂	5×10^{-2}	5×10^{-2}
H ₂ O	1×10^{-5}	1×10^{-5}
N ₂	0.95	0.95
H ₂	2×10^{-19}	5×10^{-9}
NH ₃	1×10^{-29}	4×10^{-14}
HCN	0	1×10^{-18}
NO	3×10^{-9}	1×10^{-19}
NO ₂	2×10^{-10}	3×10^{-31}
N ₂ O	2×10^{-12}	6×10^{-23}
HNO ₃	2×10^{-15}	0
O ₃	5×10^{-21}	0
Methane	0	7×10^{-21}
Ethane	0	0
Formic Acid	7×10^{-26}	2×10^{-15}
Formaldehyde	0	4×10^{-19}
Methanol	0	6×10^{-26}
Acetic Acid	0	5×10^{-23}
Methyl Amine	0	5×10^{-33}
Acetylene	0	8×10^{-34}
Benzene	0	0
Asphalt	0	0
SO ₂	2×10^{-7}	1×10^{-6}
SO ₃	8×10^{-7}	1×10^{-16}

TABLE 4 (Cont.)

<u>Compound</u>	<u>Oxidizing Limit</u>	<u>Reducing Limit</u>
COS	0	3×10^{-8}
H ₂ S	0	1×10^{-9}
CH ₃ SH	0	1×10^{-26}
H ₂ SO ₄	2×10^{-17}	3×10^{-27}

*Two extreme systems, $[CO] = 10^{-6}$, $[O_2] = 10^{-25}$ and $[CO] = 4 \times 10^{-17}$, $[O_2] = 8 \times 10^{-5}$ are shown. The hypothesized sulfur mixing ratio of 10^{-6} has a negligible effect except on compounds containing sulfur.

TABLE 5

ADOPTED CHEMICAL COMPOSITION OF THE MARTIAN ATMOSPHERE
(VOLUME MIXING RATIOS ARE DISPLAYED)

O_3	$< 1.3 \times 10^{-6}$
H_2O	$\sim 10^{-5}$
N_2	~ 0.9
NO_2	$< 3 \times 10^{-7}$
N_2O	$< 5 \times 10^{-3}$
NH_3	$< 5 \times 10^{-5}$
SO_2	$< 1 \times 10^{-7}$
H_2S	$< 2 \times 10^{-4}$
COS	$< 5 \times 10^{-5}$
CO_2	$\sim 1 \times 10^{-1}$
CH_4	$< 3 \times 10^{-4}$
C_2H_4	$< 5 \times 10^{-5}$
C_2H_6	$< 3 \times 10^{-5}$
$HCHO$	$< 1 \times 10^{-5}$
Pressure	~ 0.05 atm
Temperature	$240^\circ K$

(References: Kaplan, Munch, and Spinrad 1964; Owen and Kuiper 1964; Hanst and Swan 1965; Chamberlain and Hunten 1965; Marshall 1965)

TABLE 6

THERMODYNAMIC EQUILIBRIUM COMPOSITION OF THE MARTIAN ATMOSPHERE,
 COMPUTED FROM THE ELEMENTAL COMPOSITION OF TABLE 5*

	<u>Oxidizing Case</u>	<u>Slightly Reducing Case</u>
N_2	0.82	0.90
CO_2	0.09	0.10
CO	0	3×10^{-13}
O_2	0.09	0
H_2O	1×10^{-5}	1×10^{-5}
CH_4	0	1×10^{-6}
C_2H_6	0	5×10^{-18}
C_6H_6	0	0
H_2	0	1×10^{-10}
HCHO	0	3×10^{-28}
HCOOH	0	3×10^{-21}
CH_3COOH	0	9×10^{-19}
CH_3OH	0	4×10^{-27}
NH_3	0	7×10^{-12}
HNO_3	1×10^{-12}	0
NO_2	6×10^{-13}	0
NO	3×10^{-20}	0
N_2O_4	4×10^{-23}	0
N_2O	2×10^{-23}	0
NO_3	4×10^{-27}	0
N_2O_5	2×10^{-29}	0

TABLE 6 (Cont.)

	<u>Oxidizing Case</u>	<u>Slightly Reducing Case</u>
N_2O_3	6×10^{-32}	0
HCN	0	2×10^{-29}
SO_3	1×10^{-6}	0
SO_2	6×10^{-22}	1×10^{-27}
COS	0	6×10^{-10}
H_2S	0	1×10^{-6}
$(CH_3)_2SO$	0	3×10^{-12}
CS_2	0	8×10^{-19}
CH_3SH	0	4×10^{-19}
H_2SO_4	8×10^{-22}	

* Two extreme systems, one with $[O_2] = 0$, the other with $[O_2] = 0.09$ -- corresponding to the point P in Figure 4 -- are displayed. The more highly oxidizing conditions favor the presence of oxides of nitrogen, but even then their predicted equilibrium abundances are miniscule.

TABLE 7

ADOPTED CHEMICAL COMPOSITION OF THE JOVIAN ATMOSPHERE
(VOLUME MIXING RATIOS ARE DISPLAYED)

	High	Deep
H ₂	0.60	0.60
H ₂ O	$\sim 1 \times 10^{-6}$	$\sim 1 \times 10^{-4}$
NO ₂	$< 3 \times 10^{-6}$	$< 3 \times 10^{-6}$
NH ₃	2×10^{-4}	2×10^{-4}
CH ₄	5×10^{-3}	5×10^{-3}
C ₂ H ₆	$< 1.2 \times 10^{-4}$	$< 1.2 \times 10^{-4}$
CH ₃ NH ₂	$< 3 \times 10^{-4}$	$< 3 \times 10^{-4}$
C ₂ H ₂	$< 6 \times 10^{-5}$	$< 6 \times 10^{-5}$
He	0.36	0.36
Ne	0.03	0.03
Pressure	1 atm	1000 atm
Temperature	200°K	350°K

(References: Kuiper 1952; Spinrad and Trafton 1963; Owen 1965)

TABLE 8

THEMODYNAMIC EQUILIBRIUM COMPOSITION OF THE MAJOR CONSTITUENTS OF THE JOVIAN ATMOSPHERE, COMPUTED FROM THE ELEMENTAL COMPOSITION OF TABLE 7*

	<u>High</u>	<u>Deep</u>
H ₂	0.60	0.60
He	0.36	0.36
Ne	0.03	0.03
CH ₄	5 × 10 ⁻³	5 × 10 ⁻³
NH ₃	2 × 10 ⁻⁴	2 × 10 ⁻⁴
H ₂ O	1 × 10 ⁻⁴	1 × 10 ⁻⁴
H ₂ S	1 × 10 ⁻⁶	1 × 10 ⁻⁶
C ₂ H ₆	3 × 10 ⁻²³	1 × 10 ⁻¹⁵
N ₂	6 × 10 ⁻²²	1 × 10 ⁻¹⁶
CH ₃ SH	2 × 10 ⁻²⁸	7 × 10 ⁻²⁰
CH ₃ NH ₂	2 × 10 ⁻³¹	8 × 10 ⁻²¹
CH ₃ OH	0	3 × 10 ⁻²⁴
(CH ₃) ₂ SO	0	2 × 10 ⁻²⁹
CO ₂	0	1 × 10 ⁻³¹
CO	0	9 × 10 ⁻³²
HCHO	0	2 × 10 ⁻³³
HCOOH	0	0
HCN	0	0
COS	0	0

* For the sulfur compounds, addition of [S] = 10⁻⁶ was assumed.

THERMODYNAMIC EQUILIBRIUM IN THE ORIGIN OF ORGANIC MATTER*

R.V. Eck
National Biomedical Research Foundation
Silver Spring, Maryland

E. R. Lippincott
Department of Chemistry
University of Maryland, College Park

M.O. Dayhoff
National Biomedical Research Foundation
Silver Spring, Maryland

Y.T. Pratt
Department of Chemistry
University of Maryland, College Park

*This work was supported by NASA Contract NO. 21-003-002 to the National Biomedical Research Foundation and NASA Grant NO. NGR 21-002-059 to the University of Maryland.

THERMODYNAMIC EQUILIBRIUM IN THE ORIGIN OF ORGANIC MATTER

R.V. Eck, E.R. Lippincott, M.O. Dayhoff and Y.T. Pratt

ABSTRACT

Theoretical and experimental support is presented for the hypothesis that organic compounds occurring in carbonaceous chondrites may have formed under equilibrium or near equilibrium conditions. The equilibrium distributions of organic compounds at temperatures between 300 and 1000 K and pressures of 10^{-6} to 50 atmospheres for the C-H-O system have been computed. At high temperatures and low pressures aromatic compounds may form even in the presence of excess hydrogen. Equilibrium concentrations of numerous compounds possible at 1000 K when N, S, and Cl are added to the system have also been determined. A limited equilibrium method is employed in which those few compounds which form with most difficulty are excluded from the computations. This approach is shown to be useful in the interpretation of certain experimental data. In preliminary experiments it has been found that gases, converted to the plasma state by high energy radio frequency discharge, yield product mixtures which are in qualitative agreement with those predicted.

A major clue to the origin of organic compounds found in some meteorites⁽¹⁻⁶⁾ is the distribution of molecular species present. Studier, Hayatsu and Anders⁽¹⁾ have analyzed a number of carbonaceous chondrites with the aid of mass spectrometry and their results have led them to the interesting hypothesis that the formation of organic compounds may occur under equilibrium or near equilibrium conditions in solar nebulae. They point out that neither Miller-Urey reactions nor biological processes can account for their observed distribution of compounds, particularly the preponderance of aromatic hydrocarbons and the high methane-to-ethane ratios. Their data did agree rather well, however, with a distribution calculated by Dayhoff, Lippincott and Eck⁽⁷⁾ for conditions of thermodynamic equilibrium for a C-H-O-N ideal gas system at 500 K and 1 atm.

Urey and Lewis⁽⁸⁾, on the other hand, have expressed the view that the organic compounds detected by Studier et al. and other investigators could not form under the conditions present in solar nebulae, but could only have been produced by high-energy radiation or, in some instances, by living organisms.

In a subsequent paper, Studier, Hayatsu and Anders⁽⁹⁾ reported experimental support for the equilibrium theory by identification of products formed from the reaction of carbon monoxide and hydrogen using meteoritic material, inorganic compounds or heat alone as catalysts.

In our previous paper⁽⁷⁾ we presented the calculated distribution of elements among over 100 compounds at one atmosphere pressure and 500 K. At some C-H-O elemental proportions the formation of graphite was excluded because of the high activation energy of all reactions forming it. The present report describes calculations of thermodynamic equilibria in C-H-O systems at temperatures from 300 and 1000 K and

pressures from 10^{-6} to 50 atmospheres. Additional calculations including the elements N, S and Cl at 1000 K are reported. These computations support the theory of Studier et al(1). Preliminary experiments on equilibria produced in a plasma discharge system are also presented.

In the general ternary diagram (Fig. 1) the main characteristics of ideal gas systems containing all proportions of the elements C, H, and O are shown. There are two pronounced thresholds at which concentrations of many compounds change abruptly by many orders of magnitude. The first is the oxidizing threshold along the $\text{CO}_2\text{-H}_2\text{O}$ line, below which free O_2 appears and all organic compounds are oxidized; above this line traces of many organics appear in addition to CO_2 , CO , H_2O and CH_4 . The second is the "asphalt" threshold above which large concentrations of benzene and polynuclear aromatics appear. The position of this line varies with temperature and pressure, in contrast to the oxidizing threshold which remains essentially unchanged. There is a weak hydrogen threshold along the $\text{CH}_4\text{-H}_2\text{O}$ line which appears at low temperatures. The curved line represents a phase boundary above which solid graphite would precipitate in a complete equilibrium.

The positions of the graphite phase boundary at various temperatures and pressures are shown in Figure 2. In addition to the data illustrated, it may be noted that at 500 K and 10^{-6} atm and at 700 K at both 10^{-6} and 10^{-3} atm this line ends near the H corner.

Conventional equilibrium computations consider a very limited number of compounds, usually in single, reversible reactions. The balance of chemical species is computed for this extremely limited system as if no other reactions were occurring at the same time. Organic chemists know that this is frequently highly unrealistic; such

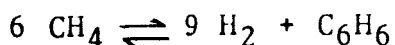
a simplification is justified by necessity because more complex systems could not be considered in detail. The method of thermodynamic equilibrium, which becomes feasible with the use of a high-speed digital computer, makes possible the simultaneous calculation of reactions among a large number of compounds. However, in cases where time becomes important, complete equilibrium may be an unrealistic ideal. With rapid quenching, or at temperatures at which many reactions require appreciable lengths of time to approach equilibrium, some reactions may occur only to a negligible extent. If the activation energy is high, and the temperature is sufficiently low, certain reaction products may be essentially "forbidden". If a particular compound is the end product of a long series of reactions, or if one of its essential intermediates occurs only in very small proportions, the rest of the system may have reached a metastable equilibrium after a certain time, while the particular compound is still present only in very small amounts.

As a way of simulating such conditions for purposes of computation, we use a method of limited equilibrium, in which all compounds of significant concentration are included except the few which are specifically excluded. In these calculations, the rate of production of each compound is assumed to be either rapid enough for equilibrium to be approximated, or negligible (zero). Results computed by this method are in reasonable agreement with certain experimental results⁽⁹⁾.

According to a formally strict definition of thermodynamic equilibrium, graphite must be included as a major constituent of the system^(8, 10, 11). However, at some temperatures, there are reasons to exclude graphite from the computation and to use the method of limited equilibrium. Graphite is the end product of the polymerization of

higher aromatic compounds, a relatively slow sequence of reactions. At temperatures below about 1000 K, it apparently cannot form directly from methane and other simple organics, because the activation energy of such reactions would be too high. At moderate temperatures, therefore, a metastable equilibrium of all other compounds could be approached before significant amounts of graphite could be produced. The elemental proportions represented by the region above the graphite line in the triangular diagram therefore give experimentally attainable gas-phase compositions. The exclusion of graphite permits these compositions to be studied, and the "asphalt region" to be investigated. If we had felt obliged to adhere to the strict definition of "equilibrium", these studies could not have been carried out. In the following computations, graphite has been excluded.

In Figure 3 are shown boundaries above which the benzene concentration is greater than 10^{-6} mole fraction, an amount readily measured experimentally. At high temperatures and low pressures, benzene (and asphalt) appear even in an atmosphere containing a large excess of free hydrogen -- a result consistent with the experimental findings of Studier et al⁽⁹⁾. Such conditions may have existed in solar nebulae⁽¹⁾. Two mutually reinforcing factors lead to this effect. First, methane becomes less stable with increasing temperature, and second, the equation



shifts toward the right as the pressure decreases. Benzene becomes measurable only in regions where solid graphite would be stable if it had been included in the equilibrium.

With graphite formation excluded, the concentrations of the other constituents change gradually in the neighborhood of the graphite

line; it is not a definite threshold. Above the graphite line there are regions where the computed benzene concentration would become very high if benzene were the only aromatic included in the equilibrium. When higher molecular weight aromatics are included also, they become more important than benzene -- the largest polynuclear aromatics becoming the most important, often so concentrated that they would precipitate. However, since the large molecules must build up from the small ones, the kinetics of a given situation dictates how far this polymerization will be able to go in the available time. We have approximated the condensation of polynuclear aromatics by including in our system benzene, naphthalene, and a mixture we call "asphalt", which is a composite of 100 isomers having six benzene rings with the molecular formula $C_{22}H_{12}$. The free energy of this composite is extrapolated from that of benzene and naphthalene by the method of group contributions⁽¹²⁾. The position of the asphalt threshold, where aromatics build up sharply, is insensitive to the exact formulas and free energies of the aromatics included in the equilibria. The threshold is sharper at lower temperatures and at higher pressures. Figure 4 shows lines above which the asphalt concentration is greater than 10^{-6} mole fraction for various conditions of temperature and pressure. Asphalt may form at all temperatures and pressures investigated. At high temperature and low pressure it is stable even where excess H_2 is present.

In the study of meteorites and geological deposits, the elements N, S, and Cl are also important. The addition of an amount of N equal to that of C produces very little change in the system. Molecular nitrogen remains principally as N_2 forming only traces of organics and oxides, a small amount of NH_3 in the presence of excess H_2 , and a little HCN in the asphalt region. Formation of HCN is particularly

avored by low pressure.

Table I shows the mole fraction abundances of various compounds of interest at a temperature of 1000 K and pressures of 1 and 10^{-6} atm. The molar abundances of the new elements relative to carbon were $[N] = [C]$, $[S] = .0001 \times [C]$, and $[Cl] = .0001 \times [C]$. Mole fractions of constituents at the points 1, 2, 3 and 4, whose elemental compositions are shown in Fig. 1 are tabulated in Table I. Point 1 is chosen below the graphite phase boundary and represents a strict, all-inclusive equilibrium. At the other three points, solid graphite formation is forbidden. It may be seen from the table that asphalt formation is encouraged by low pressure, so that it forms readily even in a system with a large excess of H_2 . Low pressure also favors the decomposition of CH_4 , CO_2 , and NH_3 .

In the asphalt region HCN and acetylene concentrations build up, as well as those of aromatics. Pyridine, the most stable of the nitrogen-containing aromatics is always less concentrated than benzene by several orders of magnitude.

The principal compound containing chlorine is HCl. Chloromethane, the most stable chlorinated hydrocarbon, is considerably less stable.

The most important sulphur compound is H_2S except at low pressure in the asphalt region where CS_2 is most important. Generally, organic sulphur compounds are much more stable than their oxygen analogs. Carbon disulphide, methanethiol, benzenethiol, and thiophene are among the most stable sulphur-containing organics.

The ethane-methane ratio is of some interest since this ratio was found to be 10^{-3} or less in some carbonaceous chondrites. It is smaller than that observed in fossil fuels or in photoactivated polymerizations or cracking processes⁽¹⁾. This ratio was computed to be less than 10^{-3} for all points illustrated in Table I.

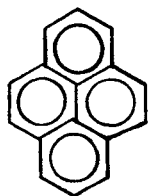
In the Fischer-Tropsch synthesis of hydrocarbons, the temperature can be kept low enough so that benzene does not form readily, presumably because the activation energies of all reactions leading to it are too high. In accordance with the method of limited equilibrium, we have therefore made computations in an otherwise all-inclusive system from which graphite, benzene, and all aromatic compounds are deleted. This leads to the formation of high yields of aliphatic compounds. A comparison of systems with and without aromatics is shown in Table II. When aromatics are omitted, cyclopentane and cyclohexane are important. The larger aliphatic hydrocarbons are much more concentrated in this case than when benzene is present -- the higher homolog mole fractions decreasing by a factor of ten for each additional carbon in the chain. Hydrocarbons with one double bond are slightly less stable than saturated ones. Those with a triple bond are considerably less stable. The ethane to-methane ratio is .014. In the equilibrium, there is little difference in concentration between normal isomers and branched chains, in contrast to the preference for normal isomers observed in the Fischer-Tropsch synthesis. Of the oxygenated compounds, acetone, acetic acid, ketene, and acetaldehyde are most stable. Acetone is about as concentrated as octane. Traces of long chain fatty acids are also observed.

When benzene and the other aromatics are included in this equilibrium, they are the most important organics. Higher homologs of all other families rapidly vanish. The ethane-to-methane ratio is 1.3×10^{-3} , and the concentrations of high homologs of the saturated hydrocarbons decrease by a factor of 300 for each increment in the number of carbon atoms. Here ethylene and acetylene are much less stable

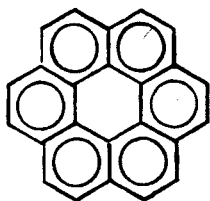
than ethane. If the temperature is kept sufficiently low, these results are in qualitative agreement with the experimental results reported by Studier et al⁽⁹⁾, in which aliphatics rather than aromatics form. We take this as experimental support for the practical value of our method of limited equilibrium, in which, under appropriate conditions, selected products are excluded from a system otherwise at equilibrium.

Preliminary laboratory experiments which we have undertaken to test the above theoretical conclusions have yielded the predicted results. By means of a radio-frequency oscillator and the auxiliary equipment described in part previously⁽¹³⁾, it is possible to generate a plasma consisting of radicals, ions and atoms from almost any compound having sufficient vapor pressure. When the energizing radio-frequency power is turned off, such a plasma must tend toward a state of equilibrium, although the rapid rate of cooling may prevent its complete attainment. By introducing a variety of elemental compositions into this reactor it should be possible to map out the entire range of the C-H-O ternary diagram shown in Figure 1, along with the effects of other elements, on the final products.

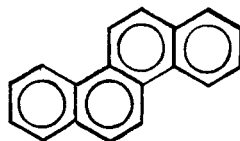
Initial results with a simple carbon-hydrogen system indicate the presence of the asphalt barrier. Excitation of gases in which the ratio of hydrogen-to-carbon is less than four-to-one leads to the formation of tars consisting primarily of polynuclear aromatic compounds. Among these, pyrene, coronene and chrysene



Pyrene



Coronene



Chrysene



Fluoranthene

have been identified so far; fluoranthene is also apparently present. Only minute traces of elemental carbon have been found. If the hydrogen-to-carbon ratio is somewhat greater than four-to-one, the system appears in the reducing region of the ternary diagram and only gaseous products are obtained. A detailed report of this work is in preparation.

These experiments demonstrate the plausibility of our assumption of the inaccessibility of graphite, even under the violent conditions of plasma excitation, as well as in the heat-catalyzed reactions studied by Studier, Hayatsu and Anders.

REFERENCES AND NOTES

1. M.H. Studier, R. Hayatsu, E. Anders, *Science*, 149, 1455 (1965).
2. E. Anders, *Ann. N.Y. Acad. Sci.*, 108, 514 (1963).
3. A.P. Vinogradov, G.P. Vdovkin, *Geokenmiya*, 843 (1964).
4. W.G. Meinschein, *Space Sci. Rev.*, 2, 653 (1963).
5. B. Nagy, D.J. Hennessy, *Ann. N.Y. Acad. Sci.*, 108, 553 (1963).
6. G. Mueller, *Geochim. Cosmochim. Acta.*, 26, 1085 (1962).
7. M.O. Dayhoff, E.R. Lippincott, R.V. Eck, *Science*, 146, 1461 (1964).
8. H.C. Urey*, J. Lewis, *Science*, (1966).
9. M.H. Studier, R. Hayatsu, E. Anders, *Science*, (1966).
10. H.E. Suess, *J. Geophys. Res.*, 67, 2029 (1962).
11. S.L. Miller, H.C. Urey, *Science*, 130, 245 (1959).
12. D.W. Van Krevelen, H.A.G. Charmin, *Chem. Eng. Sci.*, 1, 66 (1951);
J.L. Franklin, *Mod. Eng. Chem.*, 41, 1070 (1949).
13. F.X. Powell, O. Fletcher, E.R. Lippincott, *Rev. Sci. Inst.*, 34, 36
(1963); E.R. Lippincott, F.X. Powell, J.A. Creighton, D.G. Jones,
Developments in Applied Spectroscopy, Vol. 3, 1963.

* Dr. Urey has kindly pointed out a slight error in the standard free energies used in these computations, a total of .38 kilocalories per mole for H₂, O₂, and N₂. This leads to a change by a small factor in calculated equilibrium concentrations for a few compounds. However, none of the conclusions are affected.

LEGENDS

Fig. 1. Thermodynamic equilibria in atmospheres of varying elemental proportions. The ternary diagram provides a display of systems of all possible relative proportions of C, H, and O. The points corresponding to pure gases of the major compounds are indicated, and regions where different compounds are important are shown. The solid curve indicates the phase boundary along which graphite becomes stable at 1 atm. and 500 K. The activation energy for this reaction is so high that under many conditions it does not occur and gaseous equilibria above this line are observed experimentally. Above the line $\text{CH}_4\text{-CO}_2$, equilibrium favors the formation of large proportions of polycyclic aromatic compounds and a lesser increase in most of the other families of compounds. The graphite and asphalt lines are always present, but their positions vary with temperature and pressure. The numbered asterisks indicate the compositions of the equilibrium distributions shown in Table I.

Fig. 2. Graphite phase boundaries. In an all-inclusive equilibrium, the gas phase would never have a composition above this phase boundary. However, the activation energy barrier to graphite formation is very high so that under many conditions equilibrium may be approximated in metastable gases without any graphite formation. At 300° a liquid water phase forms over the lower central portion of the figure. Outside this region, the graphite phase boundary is indistinguishable from the boundary at 500° .

Fig. 3. Curves where benzene concentration is 10^{-6} mole fraction. Benzene is found at 1000° even in the presence of an excess of hydrogen. The curves for 50 atm. and 1 atm. at 1000° are indistinguishable except for a slight divergence at the upper end. At 300 K the benzene concentration is slightly lower than 10^{-6} atm above the $\text{CO}_2\text{-CH}_4$ line and much lower elsewhere.

Fig. 4. Curves above which asphalt concentration is greater than 10^{-6} mole fraction. Graphite has been omitted from these systems for kinetic reasons. At 300° the curves for pressures from 10^{-6} to 50 atm. are very close together. Asphalt forms at all temperatures and pressures investigated. At high temperature and low pressure it is stable even where excess H_2 is present.

Table I. Mole fraction compositions of compounds in equilibrium. The relative abundances of C, H and O for the four points are shown in Fig. 1. N is as abundant as carbon. Cl and S are .0001 as abundant as C. Temperatures of 1000°K and pressures of 1 atm. or 10^{-6} atm. are considered.

Elemental Composition Point (See Fig. 1)	1 Atm.				10^{-6} Atm.			
	1	2	3	4	1	2	3	4
H ₂	.54	.11	.86	.55	.55	.37	.91	.81
O ₂	$.59 \times 10^{-21}$	$.57 \times 10^{-26}$	$.15 \times 10^{-23}$	$.37 \times 10^{-25}$	$.52 \times 10^{-15}$	$.99 \times 10^{-35}$	$.25 \times 10^{-36}$	$.21 \times 10^{-36}$
H ₂ O	.15	$.92 \times 10^{-4}$	$.12 \times 10^{-1}$	$.12 \times 10^{-2}$.15	$.14 \times 10^{-10}$	$.53 \times 10^{-11}$	$.43 \times 10^{-11}$
CO ₂	$.55 \times 10^{-1}$	$.48 \times 10^{-3}$	$.56 \times 10^{-3}$	$.15 \times 10^{-3}$	$.53 \times 10^{-1}$	$.17 \times 10^{-10}$	$.32 \times 10^{-12}$	$.30 \times 10^{-12}$
CO	.14	.40	$.28 \times 10^{-1}$	$.48 \times 10^{-1}$.15	.34	$.39 \times 10^{-1}$	$.41 \times 10^{-1}$
CH ₄	$.55 \times 10^{-2}$.19	$.55 \times 10^{-1}$.25	$.63 \times 10^{-14}$	$.49 \times 10^{-4}$	$.21 \times 10^{-3}$	$.19 \times 10^{-3}$
C ₂ H ₆	$.11 \times 10^{-7}$	$.68 \times 10^{-4}$	$.70 \times 10^{-6}$	$.22 \times 10^{-4}$	$.14 \times 10^{-31}$	$.13 \times 10^{-11}$	$.96 \times 10^{-11}$	$.89 \times 10^{-11}$
N ₂	.10	.30	$.42 \times 10^{-1}$.15	.10	.25	$.37 \times 10^{-1}$.12
NH ₃	$.69 \times 10^{-4}$	$.10 \times 10^{-4}$	$.89 \times 10^{-4}$	$.86 \times 10^{-4}$	$.70 \times 10^{-10}$	$.62 \times 10^{-10}$	$.91 \times 10^{-10}$	$.14 \times 10^{-9}$
HCN	$.41 \times 10^{-6}$	$.28 \times 10^{-3}$	$.13 \times 10^{-5}$	$.21 \times 10^{-4}$	$.45 \times 10^{-12}$	$.10 \times 10^{-1}$	$.43 \times 10^{-2}$	$.82 \times 10^{-2}$
Benzene	$.20 \times 10^{-18}$	$.76 \times 10^{-3}$	$.29 \times 10^{-14}$	$.12 \times 10^{-8}$	$< 10^{-38}$	$.28 \times 10^{-5}$	$.53 \times 10^{-5}$	$.84 \times 10^{-5}$
Naphthalene	$.88 \times 10^{-30}$	$.41 \times 10^{-3}$	$.43 \times 10^{-23}$	$.18 \times 10^{-13}$	$< 10^{-38}$	$.11 \times 10^{-5}$	$.13 \times 10^{-5}$	$.31 \times 10^{-5}$
"Asphalt"	$< 10^{-38}$	$.33 \times 10^{-4}$	$< 10^{-38}$	$.46 \times 10^{-29}$	$< 10^{-38}$	$.68 \times 10^{-2}$	$.86 \times 10^{-3}$	$.80 \times 10^{-2}$
Formic Acid	$.15 \times 10^{-7}$	$.26 \times 10^{-10}$	$.25 \times 10^{-9}$	$.42 \times 10^{-10}$	$.15 \times 10^{-13}$	$.33 \times 10^{-23}$	$.15 \times 10^{-24}$	$.13 \times 10^{-24}$

Table I (contd)

Elemental Composition Point (See Fig. 1)	1 Atm.				10^{-6} Atm.			
	1	2	3	4	1	2	3	4
Acetic Acid	$.24 \times 10^{-10}$	$.70 \times 10^{-11}$	$.24 \times 10^{-11}$	$.28 \times 10^{-11}$	$.26 \times 10^{-28}$	$.66 \times 10^{-28}$	$.52 \times 10^{-29}$	$.44 \times 10^{-29}$
Formaldehyde	$.10 \times 10^{-6}$	$.54 \times 10^{-7}$	$.32 \times 10^{-7}$	$.34 \times 10^{-7}$	$.10 \times 10^{-12}$	$.17 \times 10^{-12}$	$.47 \times 10^{-13}$	$.43 \times 10^{-13}$
Methanol	$.24 \times 10^{-8}$	$.25 \times 10^{-9}$	$.12 \times 10^{-8}$	$.85 \times 10^{-9}$	$.26 \times 10^{-20}$	$.28 \times 10^{-20}$	$.19 \times 10^{-20}$	$.16 \times 10^{-20}$
Ethanol	$.34 \times 10^{-13}$	$.63 \times 10^{-12}$	$.11 \times 10^{-12}$	$.52 \times 10^{-13}$	$.41 \times 10^{-37}$	$.50 \times 10^{-27}$	$.60 \times 10^{-27}$	$.50 \times 10^{-27}$
Ethylene	$.72 \times 10^{-8}$	$.22 \times 10^{-3}$	$.28 \times 10^{-6}$	$.14 \times 10^{-4}$	$.90 \times 10^{-26}$	$.12 \times 10^{-5}$	$.36 \times 10^{-5}$	$.38 \times 10^{-5}$
Acetylene	$.27 \times 10^{-10}$	$.42 \times 10^{-5}$	$.65 \times 10^{-9}$	$.49 \times 10^{-7}$	$.33 \times 10^{-22}$	$.64 \times 10^{-2}$	$.80 \times 10^{-2}$	$.93 \times 10^{-2}$
Xylene	$.16 \times 10^{-27}$	$.19 \times 10^{-7}$	$.88 \times 10^{-22}$	$.18 \times 10^{-14}$	$< 10^{-38}$	$.37 \times 10^{-18}$	$.21 \times 10^{-17}$	$.35 \times 10^{-17}$
Acetone	$.97 \times 10^{-15}$	$.16 \times 10^{-10}$	$.12 \times 10^{-13}$	$.63 \times 10^{-12}$	$< 10^{-38}$	$.27 \times 10^{-24}$	$.23 \times 10^{-24}$	$.22 \times 10^{-24}$
Dimethyl Ether	$.81 \times 10^{-13}$	$.15 \times 10^{-11}$	$.26 \times 10^{-12}$	$.12 \times 10^{-11}$	$.97 \times 10^{-37}$	$.12 \times 10^{-26}$	$.14 \times 10^{-26}$	$.12 \times 10^{-26}$
Methylamine	$.23 \times 10^{-11}$	$.58 \times 10^{-10}$	$.18 \times 10^{-10}$	$.12 \times 10^{-9}$	$.26 \times 10^{-29}$	$.26 \times 10^{-19}$	$.67 \times 10^{-19}$	$.10 \times 10^{-18}$
Glycine	$.88 \times 10^{-20}$	$.20 \times 10^{-20}$	$.73 \times 10^{-21}$	$.13 \times 10^{-20}$	$< 10^{-38}$	$< 10^{-38}$	$< 10^{-38}$	$< 10^{-38}$
Pyridine	$.81 \times 10^{-22}$	$.14 \times 10^{-8}$	$.15 \times 10^{-18}$	$.14 \times 10^{-13}$	$< 10^{-38}$	$.12 \times 10^{-12}$	$.77 \times 10^{-13}$	$.20 \times 10^{-12}$
S ₂	$.74 \times 10^{-9}$	$.12 \times 10^{-6}$	$.50 \times 10^{-10}$	$.15 \times 10^{-8}$	$.32 \times 10^{-3}$	$.88 \times 10^{-7}$	$.18 \times 10^{-7}$	$.51 \times 10^{-7}$
CS ₂	$.12 \times 10^{-8}$	$.17 \times 10^{-3}$	$.31 \times 10^{-9}$	$.98 \times 10^{-7}$	$.55 \times 10^{-9}$	$.26 \times 10^{-2}$	$.39 \times 10^{-3}$	$.12 \times 10^{-2}$
H ₂ S	$.20 \times 10^{-2}$	$.50 \times 10^{-2}$	$.84 \times 10^{-3}$	$.29 \times 10^{-2}$	$.14 \times 10^{-2}$	$.15 \times 10^{-4}$	$.17 \times 10^{-4}$	$.25 \times 10^{-4}$
SO ₂	$.20 \times 10^{-10}$	$.25 \times 10^{-14}$	$.14 \times 10^{-13}$	$.18 \times 10^{-14}$	$.12 \times 10^{-4}$	$.37 \times 10^{-26}$	$.44 \times 10^{-28}$	$.59 \times 10^{-28}$
CO ₂	$.17 \times 10^{-4}$	$.59 \times 10^{-3}$	$.87 \times 10^{-6}$	$.80 \times 10^{-5}$	$.11 \times 10^{-4}$	$.44 \times 10^{-6}$	$.23 \times 10^{-7}$	$.40 \times 10^{-7}$

Table I (contd)

Elemental Composition Point (See Fig. 1)	1 Atm.				10^{-6} Atm.			
	1	2	3	4	1	2	3	4
Methane Thiol	$.20 \times 10^{-8}$	$.86 \times 10^{-6}$	$.51 \times 10^{-8}$	$.13 \times 10^{-6}$	$.15 \times 10^{-20}$	$.19 \times 10^{-12}$	$.38 \times 10^{-12}$	$.57 \times 10^{-12}$
Benzene Thiol	$.12 \times 10^{-23}$	$.56 \times 10^{-7}$	$.45 \times 10^{-20}$	$.10 \times 10^{-13}$	$< 10^{-38}$	$.18 \times 10^{-12}$	$.16 \times 10^{-12}$	$.41 \times 10^{-12}$
Thiophene	$.20 \times 10^{-17}$	$.63 \times 10^{-6}$	$.32 \times 10^{-15}$	$.98 \times 10^{-11}$	$< 10^{-38}$	$.13 \times 10^{-8}$	$.92 \times 10^{-9}$	$.21 \times 10^{-8}$
Cl_2	$.24 \times 10^{-15}$	$.11 \times 10^{-13}$	$.26 \times 10^{-16}$	$.50 \times 10^{-15}$	$.23 \times 10^{-15}$	$.23 \times 10^{-14}$	$.22 \times 10^{-16}$	$.24 \times 10^{-15}$
HCl	$.20 \times 10^{-2}$	$.59 \times 10^{-2}$	$.84 \times 10^{-3}$	$.29 \times 10^{-2}$	$.20 \times 10^{-2}$	$.51 \times 10^{-2}$	$.77 \times 10^{-3}$	$.24 \times 10^{-2}$
CH_3Cl	$.69 \times 10^{-9}$	$.36 \times 10^{-6}$	$.18 \times 10^{-8}$	$.44 \times 10^7$	$.77 \times 10^{-21}$	$.23 \times 10^{-10}$	$.61 \times 10^{-11}$	$.19 \times 10^{-10}$
CCl_4	$.35 \times 10^{-33}$	$.62 \times 10^{-27}$	$.16 \times 10^{-34}$	$.63 \times 10^{-31}$	$< 10^{-38}$	$.58 \times 10^{-33}$	$.38 \times 10^{-37}$	$.51 \times 10^{-35}$

Table II. Mole fraction composition of gas mixture C:H:O=30:40:30 at 1 Atm. pressure and 500° K. Graphite was omitted from the equilibrium in the first column. Graphite and all aromatic compounds were omitted from the equilibrium in the second column. Under these conditions, aliphatics are formed in high concentration.

	Aromatics Included	Aromatics Deleted		Aromatics Included	Aromatics Deleted
H ₂	.94x10 ⁻⁴	.96x10 ⁻⁶	Allene	.35x10 ⁻¹⁷	.19x10 ⁻⁹
O ₂	<10 ⁻³⁸	<10 ⁻³⁸	Ketene	.86x10 ⁻¹³	.50x10 ⁻⁹
H ₂ O	.11x10 ⁻³	.95x10 ⁻⁸	Formaldehyde	.20x10 ⁻¹¹	.14x10 ⁻¹¹
CO	.41x10 ⁻²	.29	Acetaldehyde	.75x10 ⁻¹¹	.44x10 ⁻⁹
CO ₂	.61	.38	Formic Acid	.13x10 ⁻¹⁰	.81x10 ⁻¹³
Methane	.38	.32	Acetic Acid	.20x10 ⁻⁸	.10x10 ⁻⁸
Ethane	.66x10 ⁻⁴	.46x10 ⁻²	Butyric Acid	.30x10 ⁻¹⁴	.10x10 ⁻¹⁰
Propane	.80x10 ⁻⁷	.46x10 ⁻³	Octanoic Acid	.14x10 ⁻²⁵	.23x10 ⁻¹⁴
Butane	.21x10 ⁻⁹	.99x10 ⁻⁴	Nonanoic Acid	.83x10 ⁻²⁹	.11x10 ⁻¹⁵
Pentane	.29x10 ⁻¹²	.11x10 ⁻⁴	Palmitic Acid	<10 ⁻³⁸	.54x10 ⁻²²
Hexane	.47x10 ⁻¹⁵	.15x10 ⁻⁵	Methanol	.21x10 ⁻¹²	.15x10 ⁻¹⁴
Heptane	.73x10 ⁻¹⁸	.19x10 ⁻⁶	Ethanol	.86x10 ⁻¹⁴	.52x10 ⁻¹⁴
Octadecane	<10 ⁻³⁸	.15x10 ⁻¹⁶	Octanol	.15x10 ⁻³¹	.29x10 ⁻²⁰
Cyclopropane	.76x10 ⁻¹³	.42x10 ⁻⁷	Acetone	.51x10 ⁻¹¹	.25x10 ⁻⁷
Cyclobutane	.23x10 ⁻¹¹	.11x10 ⁻³	Dimethyl Ether	.51x10 ⁻¹⁸	.31x10 ⁻¹⁸
Cyclopentane	.13x10 ⁻¹¹	.48x10 ⁻²	Lactic Acid	.87x10 ⁻²²	.33x10 ⁻²²
Cyclohexane	.15x10 ⁻¹³	.46x10 ⁻²	Oxalic Acid	.65x10 ⁻¹³	.26x10 ⁻¹⁵
Ethylene	.88x10 ⁻⁸	.59x10 ⁻⁴	Pyruvic Acid	.71x10 ⁻²⁸	.26x10 ⁻²⁶
Propene	.53x10 ⁻⁹	.29x10 ⁻³	Glycerol	.71x10 ⁻³⁸	<10 ⁻³⁸
Butene	.92x10 ⁻¹²	.42x10 ⁻⁴	Carbon Suboxide	.24x10 ⁻²⁷	.95x10 ⁻²⁰
Octene	.21x10 ⁻²³	.43x10 ⁻⁸	Benzene	.96x10 ⁻⁴	-
Acetylene	.55x10 ⁻¹⁶	.36x10 ⁻¹⁰	Naphthalene	.73x10 ⁻⁴	-
Propyne	.29x10 ⁻¹⁶	.16x10 ⁻⁸	Asphalt	.11x10 ⁻¹	-
Octyne	.15x10 ⁻³⁰	.31x10 ⁻¹³			

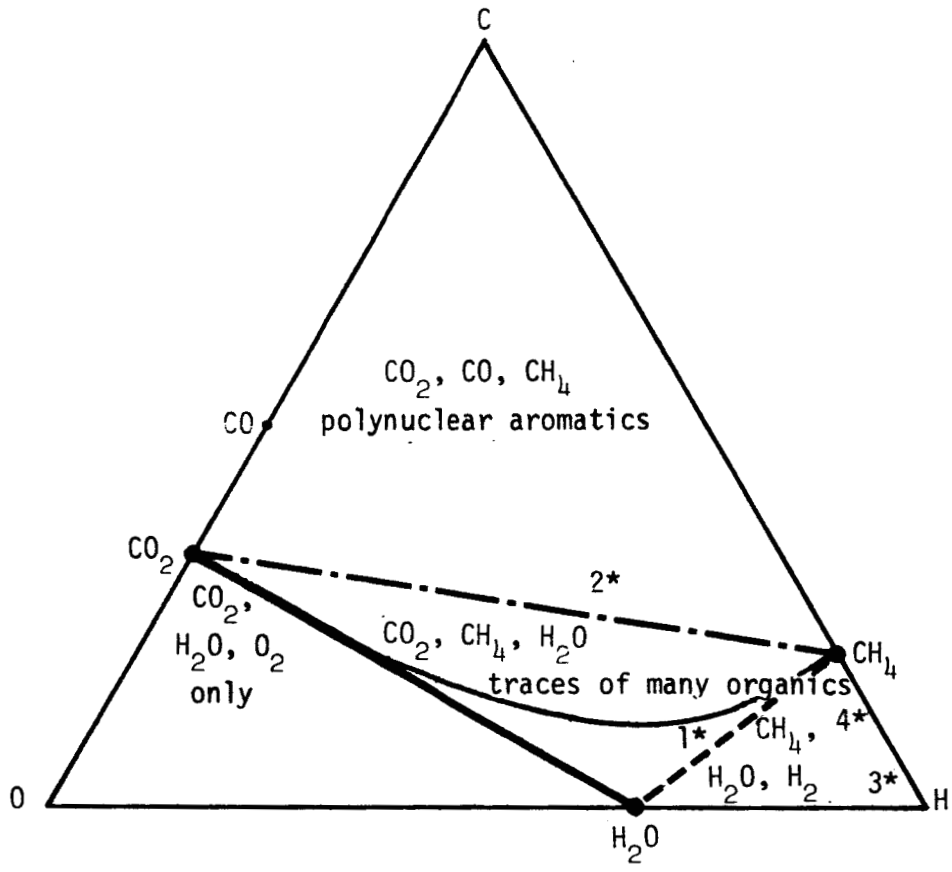
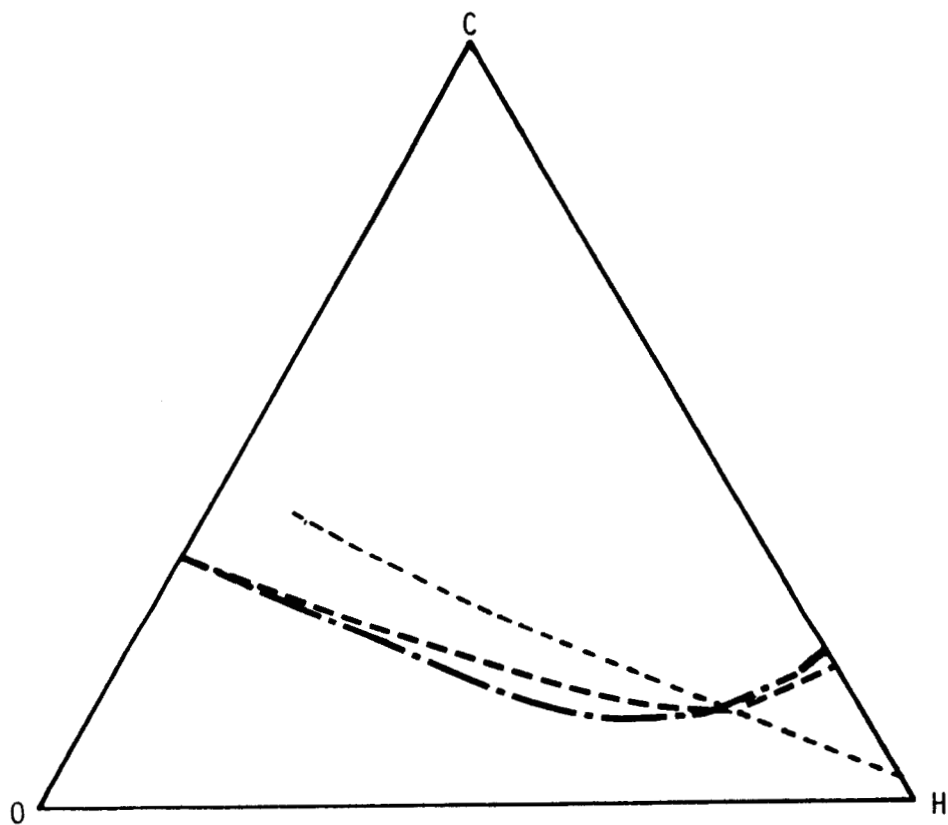
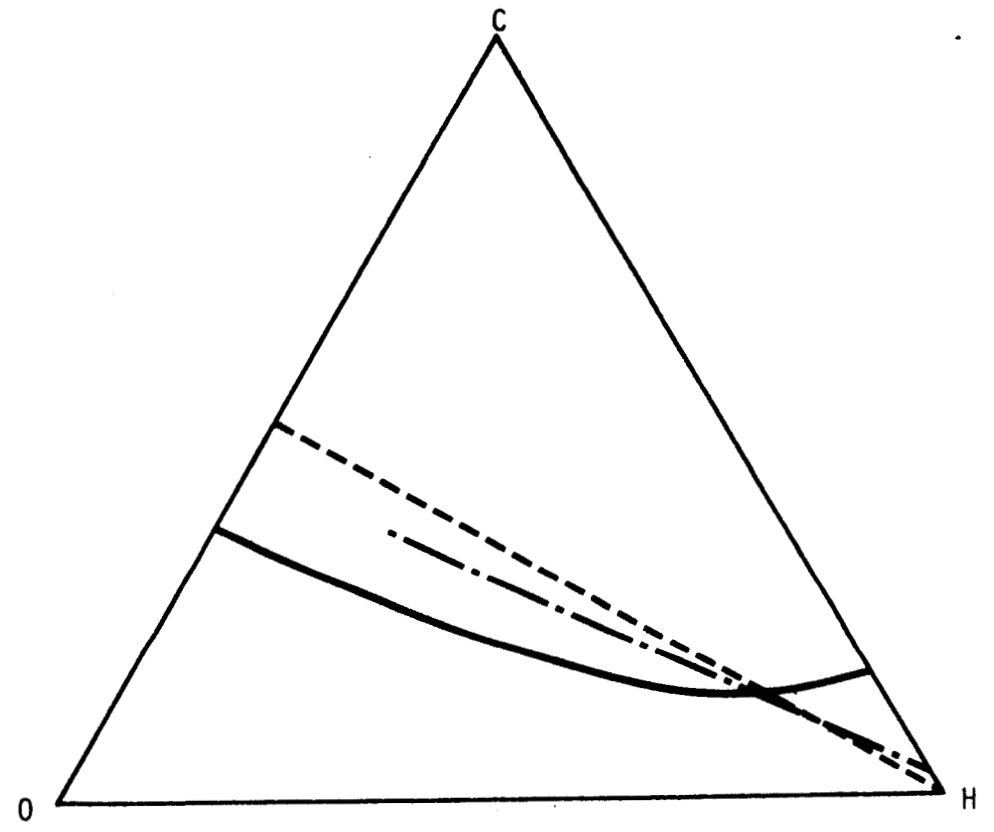
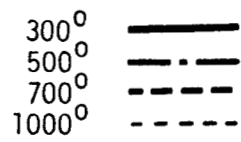


Fig. 1



(a)

1 Atm.



(b)

1000° K

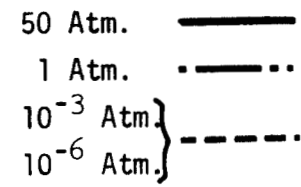
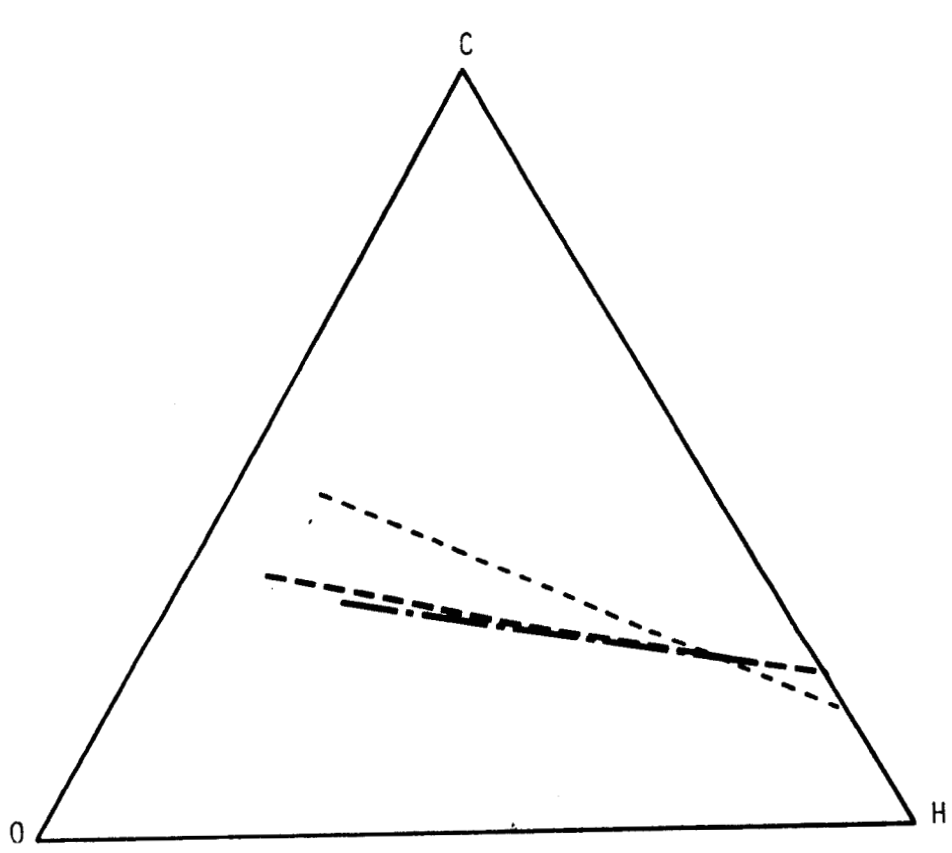


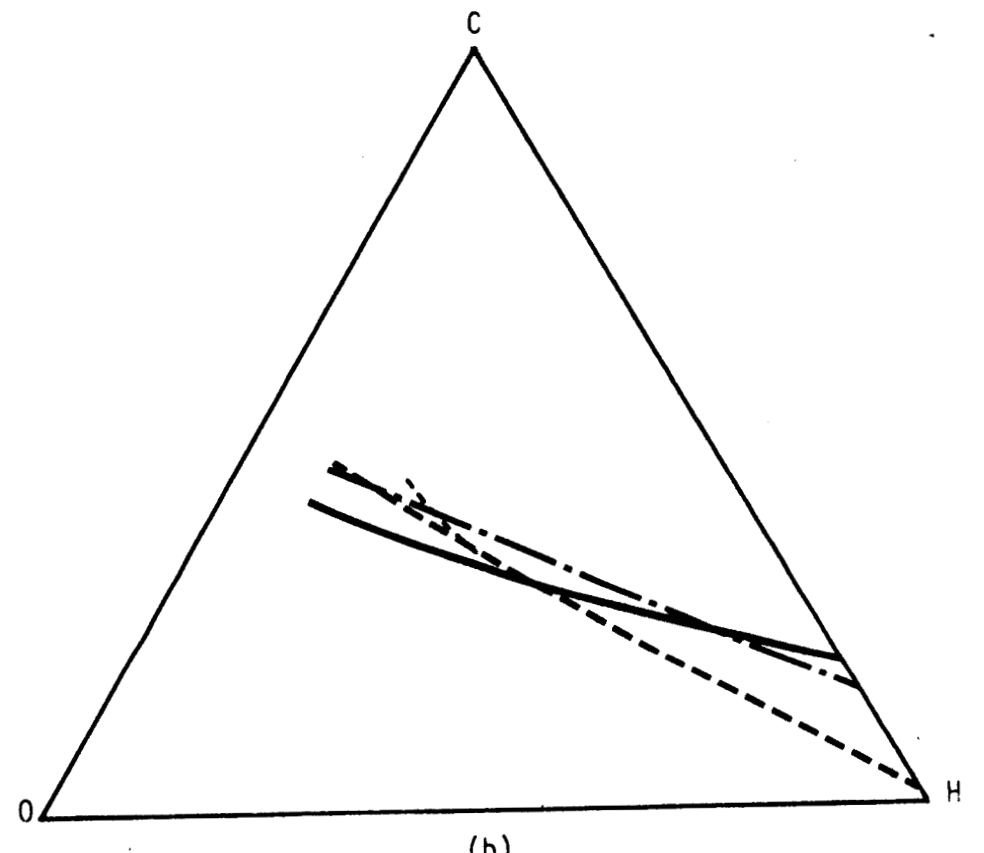
Figure 2.



(a)

1 Atm.

500° - - - - -
 700° - - - - -
 1000° - · - · -

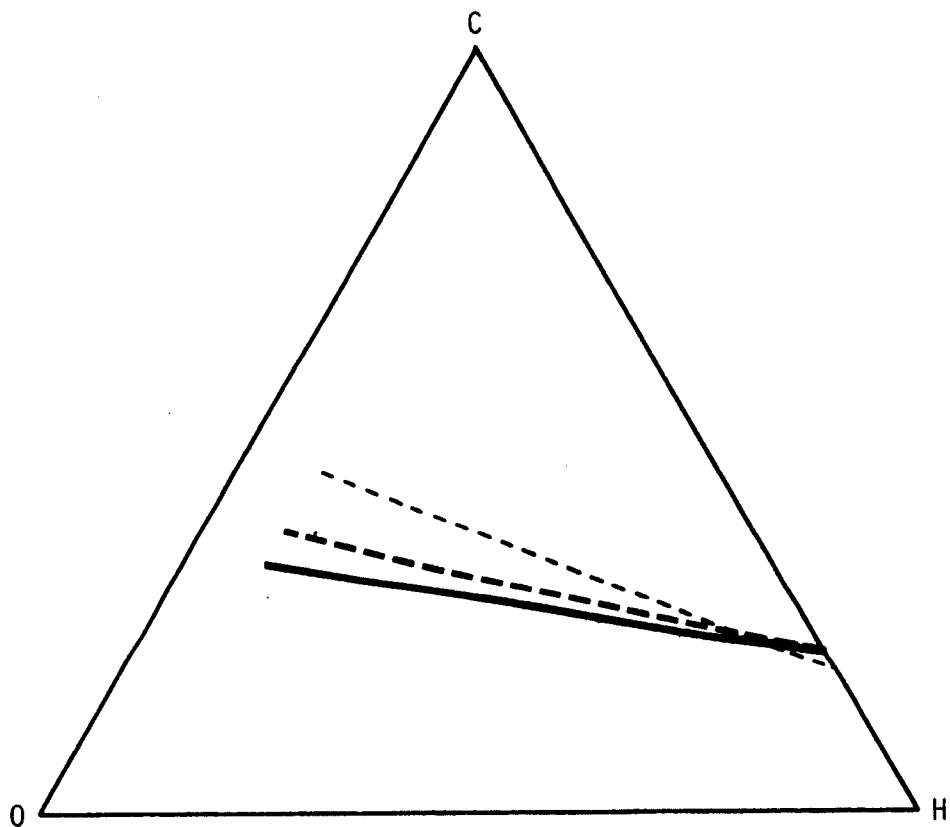


(b)

1000° K

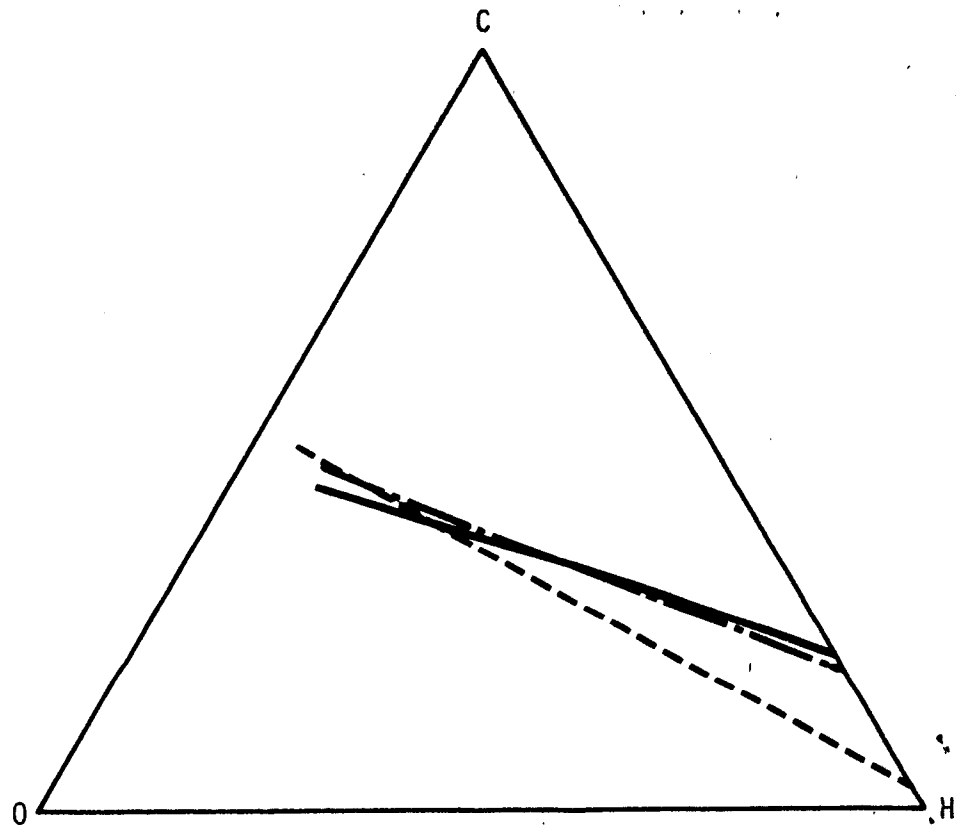
50 Atm. - - - - -
 1 Atm. - · - · -
 10⁻³ Atm. - - - - -
 10⁻⁶ Atm. - - - - -

Figure 3.



(a)

1 Atm.
 300° }
 500° } ———
 700° } - - -
 1000° } - · - ·



(b)

1000° K
 50 Atm. ———
 1 Atm. - - -
 10⁻³ Atm. }
 10⁻⁶ Atm. } - · - ·

Figure 4.

EVOLUTION OF THE STRUCTURE OF FERREDOXIN

BASED ON SURVIVING RELICS OF PRIMITIVE AMINO ACID SEQUENCES*

Richard V. Eck and Margaret O. Dayhoff
National Biomedical Research Foundation
8600 16th Street, Silver Spring, Maryland 20910

NBR Report No. 660101-20878

Abstract. The present-day structure of ferredoxin, with its simple, inorganic active site and its functions basic to photon-energy utilization, suggests the incorporation of its prototype into metabolism very early during biochemical evolution, even before complex proteins and the complete modern genetic code existed. The information in the amino acid sequence of ferredoxin enables us to reconstruct its evolutionary history. It has evolved by doubling a shorter protein, which may have contained only eight of the simplest amino acids. This shorter ancestor in turn developed from a repeating sequence of the amino acids alanine, aspartic acid or proline, serine, and glycine. The living relics of this primordial structure persist today because of a conservative principle in evolutionary biochemistry: Natural selection inhibits with extreme severity any change in a well-adapted system on which several other essential components depend. Many such intricate details of the earliest stages of life must still survive, awaiting detection and elucidation.

*This work was supported by NASA Contract No. 21-003-002 and NIH Grants No. GM-12168 and GM-08710 to the National Biomedical Research Foundation.

Many of the concepts of organic evolution which have long been known in biology are still little used in discussions of biochemistry. In nature, biochemistry is included in biology. An organism is a functioning system composed of the structures, organs, tissues, and organelles of classical biology. It also includes metabolites, macromolecules, enzyme aggregates, and biochemical feedback systems. Fine biochemical details have only recently become accessible. Potentially, a much greater amount of information is available in biochemistry than in classical biology.

According to evolutionary theory, each functioning system in an organism is subject to occasional changes or mutations, but the infrequency of these mutations necessitates that they will almost always occur, and be selected for, one at a time. Each change or addition must be an improvement, or at least not too severe a disadvantage, in order that natural selection permit its survival. This constraint has a very conservative effect. If its ecological niche stays the same, a well-adapted organism strongly resists change. Thus we find familiar looking fossil shells a third of a billion years old. If its niche changes, new functions evolve, but the most primitive structures tend to remain unchanged, since these older components have already come to be relied upon by several later additions. Any change in a very old component, even though it might be advantageous in some way, would coincidentally disturb so many other things that it would almost always be extremely disadvantageous to the organism. For example, we can compare the amino acid sequences of cytochrome-C from yeast (1) and from horse (2), position by position. In 64 of the 104 positions the amino acids in the two chains are identical. Between horse and human cytochrome-C (3) there are only 12 amino acid differences.

When we consider evolution retrospectively, the constraints are even more severe. One basic evolutionary principle is that every living organism or structure or function had ancestors which were very similar to itself, but simpler. (This is true even if it had more complex immediate ancestors.) In a particular case there are generally only a few plausible slightly simpler ancestors. As we trace the changes in a structure or function back through time, we must bear in mind that all of the structures and functions of the cell may be simpler. We are then dealing with components homologous to those seen today, not identical.

In the study of ferredoxin reported here, we have taken a number of the evolutionary principles from biology and applied them to biochemistry. In this way we have found an astonishingly primitive pattern still embedded in the molecular structure of a living organism.

The amino acid sequence of ferredoxin from Clostridium pasteurianum, a nonphotosynthetic anaerobic bacterium, has been reported (4). This protein seems to have evolved at an earlier time than many others which have been studied. We draw this inference from the following considerations (5).

1. Ferredoxin occurs in primitive anaerobic organisms, both photosynthetic and nonphotosynthetic. It must have been present in simpler organisms, the extinct common ancestors of these.

2. Ferredoxin contains iron and sulfur, bonded loosely to the protein. FeS is a chemically simple type of catalyst and would have been readily available to the most primitive organism.

3. The functions of ferredoxin are basic to cell chemistry:

The reduction of ferredoxin is the key photochemical event in photosynthesis by chloroplasts. All the energy is channeled through this

compound to other cellular energy-storage mechanisms.

Ferredoxin is the most highly reducing stable compound so far found in the cell, having a reducing potential near that of molecular hydrogen. This suggests that its function may have evolved at a very early time when the earth's atmosphere was still strongly reducing.

It reduces triphosphopyridine nucleotide (TPN), a ubiquitous reducing agent in the cell. Therefore, it may be even more primitive than TPN.

It catalyzes adenosine triphosphate (ATP) formation by radiation. This indicates its possible relatedness to primitive energy transfer processes.

It catalyzes the synthesis of pyruvate from carbon dioxide and acetylcoenzyme-A. This indicates its involvement with one of the simplest, most primitive synthetic processes in intermediary metabolism, the fixation of CO₂.

It participates in nitrogen fixation. (6)

4. Ferredoxin contains an unusually high proportion of the smaller, more thermodynamically stable amino acids, such as glycine, alanine, cysteine, serine, and aspartic acid. Their synthesis by autotrophic organisms requires only a small number of endothermic steps.

5. Ferredoxin is smaller than most other enzymes, being only 55 amino acid links long. It appears to have some sort of repeating structure, so that it may once have been still smaller.

Let us now consider the amino acid sequence of ferredoxin. Applying the constraints imposed by the principles of evolution, can we find traces of its ancestry?

Figure 1 shows the reported sequence of ferredoxin (4) and some manipulations with it. For the purposes of comparative studies like this, a single-letter notation (7) is much more suitable than the usual three-letter notation (row 1). If links 30 to 55 (row 3) are placed under links 1 to 29 (row 2), it is evident that the number of coincidences (row 4) far exceeds that which would be expected by chance ($P \ll 0.001$). It appears that this protein has evolved by doubling of the nucleic acid which determines it. This corresponds to the process of chromosome duplication which has long been known to geneticists. Presumably the ancestral sequence contained all those amino acid links which are common to both parts. Where the links are different, probably one of the two was present originally.

The ancestral sequence of 29 links must itself have had a simpler ancestor. An attempt was made to discover this by the same process, but no further regularities could be found by superimposing quarters of the chain.

Another kind of simplicity would be for the ancestor to have had fewer kinds of amino acids. During the evolution of the genetic code there must have been a time when the genetic mechanism could discriminate fewer amino acids. Let us conjecture that those which coincide in the two halves of the ferredoxin sequence (glycine, cysteine, alanine, proline, valine, aspartic acid, and glutamine) are survivors of this early stage. Perhaps when any of these occurs in either half, it is also likely to be a survivor of the earlier stage. Therefore, we note wherever any of these 7 amino acids occur in the paired ferredoxin sequences (row 5 of Fig. 1). In this arrangement, two regular patterns emerge. Cysteines occur in a cycle of 3 and alanines occur in a cycle of 4. A discontinuity in both patterns occurs at the midpoint, position 15. None of the other amino acids confirm a cycle of 3. However, the cycle of 4 is confirmed by glycine, proline,

aspartic acid, and serine. A repeating sequence of 4 amino acids (with a break at position 15) has been written in row 6. The combined halves agree with this cycle in 13 positions (row 7); they disagree in only 4 positions (row 8). Altogether 17 occurrences of these 4 simple amino acids in the living ferredoxin chain agree with this cycle, and 5 disagree with it. Figure 2 shows the reported sequence rewritten in groups of 4 for direct comparison with the repeating pattern.

The probabilities involved in a pattern of this kind are difficult to compute, because the pattern was discovered rather than predicted. Some sort of pattern can always be found in randomness, if it is examined in fine detail. In principle, one should modify the computed probability to reflect all the other patterns which would have been considered equally good if they had been discovered. This number is difficult to determine, and must be somewhat arbitrary. However, in this case it does not seem to be very large, and the number of coincidences still seems beyond ordinary chance. We consider it to be good evidence for the pattern of a cycle of 4.

Using a computer program, we have matched the sequence of ferredoxin against itself in all combinations, and against the various possible cycles. The result of this objective method is the same as we found by inspection. Cysteines occur only in a cycle of 3. Alanine, aspartic acid, serine, and glycine agree with a cycle of 4; and proline in 3 places occupies the position of aspartic acid. This substitution is reasonable stereochemically, since the two side-chains have a very similar conformation, when aspartic acid folds into a hydrogen-bonded intramolecular ring. Aspartic acid is metabolically simpler to synthesize than proline and therefore seems likely to have evolved earlier.

We will now use these patterns and the rules of evolution to reconstruct the history of the ferredoxin molecule. We first consider the prosthetic group, or inorganic part, in a prebiological environment. Then starting with an extremely primitive living system, we follow the development of the increasing intricacy of the protein.

At chemical equilibrium in an ideal gas mixture of a reducing nature at standard temperature and pressure, H_2S , CH_4 , CO_2 , H_2O , and N_2 predominate; NH_3 and organic compounds occur in small amounts (8). At equilibrium, FeS is stable in this gas. Possibly life may have organized about such a stable inorganic catalyst that could participate in capturing photons and in directing energy toward the reduction and fixation of CO_2 and toward the synthesis of pyrophosphate. In this photon-activated system, other, less-stable catalysts may then have been synthesized, permitting the development of more complex systems.

Regardless of how life originated, there was at one time a very primitive organism, far simpler than any known to be living today. It was capable of making and polymerizing some of the simplest amino acids and nucleotides. Perhaps the nucleic acids were made of only two nucleotides, simpler than the ones which occur in the present genetic mechanism. A variety of such polymers could be formed having useful structural or catalytic functions, for which natural selection preserved them.

One sequence of twelve nucleotides doubled and redoubled itself, making a longer, repetitive chain.

At about the same time, the primitive amino acid polymerizing mechanism of the cell began to utilize this nucleic acid chain as a template; it coded for the amino acids alanine, aspartic acid, serine, and glycine. If this occurred when the nucleotide chain was still only twelve links long, the resulting peptide would be A D S G, as in Figure 3, row 1.

After the nucleotide chain became longer, it coded for a simple repeating protein, A D S G A D S G . . . (row 2). This protein had some advantageous, perhaps structural, function in the cell, unrelated to the present energy-transfer function of ferredoxin.

An aberration in the nucleotide sequence produced a break in the cycle (row 2, underlined).

The synthesizing abilities of the organism became more versatile and more efficient. The genetic code became more complex, so that the genetic mechanism was able to incorporate other amino acids. Mutations occurred which modified and complicated the particular amino acid sequence which we are following (row 3).

Cysteine was among these new amino acids added to our sequence. The sulfide bond of the cysteine link became attached to iron sulfide. The protein thus cooperated in the photon-coupled catalytic function, which it still retains, and became protoferredoxin. On the principle that evolution proceeds one step at a time, we assume that the cell was already using iron sulfide as a catalyst, probably attached to cysteine alone, or to some peptide less suitable than protoferredoxin. This new attachment would merely have increased the efficiency of this function.

Eventually four cysteines were added by mutation, and two identical chains combined to make an intricate protein-iron-sulfide complex of greatly increased efficiency. It still retains essentially this structure.

The nucleic acid doubled in length by a process prototypical of a chromosome aberration, resulting in a protein of 58 links (row 4). In the three-dimensional structure, the effect of this change was to attach the two shorter chains end-to-end. They must already have been in a configuration which was only moderately disturbed by this new constraint. The attachment was an improvement but not a radical change. We predict that when the three-dimensional structure of ferredoxin is worked out, evidence will be found for the previous stage, with its two identical, cooperating, shorter chains.

The three end links may have been lost at this time or later, to give the present total length of 55 links.

Many functions of the cell improved in efficiency and complexity, evolving new capabilities. The genetic mechanism also evolved, and became capable of incorporating additional amino acids. Mutations occurred and were selected, each of which one/made a slight improvement in the overall function, until the present sequence was produced (row 5).

By this time there were many lines of descent in the phylogenetic tree, and different species must have produced different mutational variations on the earlier sequences. It will be extremely interesting to be able to compare amino acid sequences of ferredoxins from diverse species, to discover at what stage of this process they had a common ancestor. Perhaps most other species have mutated much more, so that only in such conservative primitive organisms as Clostridium or Chromatium will there be such clear "living-fossil" evidence of the earliest stages of protein evolution.

At any of the stages, additional duplication and separation of the genetic material may have occurred, followed by mutations. This would produce other proteins, which today may have various degrees of relatedness to ferredoxin. How far can we hope to trace this kind of relatedness before the evidence becomes so blurred by changes that the similarities are undetectable?

Just as the salt composition of our tissue fluids is supposed to represent a stabilized sample of ancient sea water, so the simplest, metabolically most ancient components of cellular metabolism preserve some aspects of their original environment. In modern organisms, primitive reactions, such as those involving glutathione or coenzyme-A operate under their primordial reducing conditions, isolated from the harsh outer environment by later adaptations. Such ancient

systems are extremely conservative, because so many diverse later reactions have become dependent on them that they are no longer "free" to mutate. A mutational change which might be beneficial in one way, in almost every case would be a strong disadvantage in many other ways. Natural selection would therefore reject it. This conservative principle enables us to comprehend why ferredoxin from a living organism could still retain detectable details of its ancient origin.

One of the most interesting implications of this study, we feel, is that in organisms still living there exist biochemical relics of the minutest details of the origin and evolution of the genetic mechanism. Determination of the sequences of other proteins and of nucleic acids from ^{various} primitive organisms should make possible a much more detailed reconstruction of these processes and events.

References and Notes

1. K. Narita, K. Titani, Y. Yaoi, H. Murakami, Biochim. Biophys. Acta., 77, 688 (1963).
2. E. Margoliash, E. Smith, G. Kreil, H. Tuppy, Nature, 192, 1121 (1961).
3. H. Matsubara, E. Smith, J. Biol. Chem., 237, 3575 (1962).
4. M. Tanaka, Y. Nakashima, Ann Benson, H. F. Mower, Y. T. Yasunobu, Biochem. Biophys. Res. Commun., 16, 422 (1964).
5. D. I. Arnon, Science, 149, 1460 (1965).
6. L. E. Mortenson, Proc. Natl. Acad. Sci. U. S., 52, 272 (1964).
7. A = alanine, C = cysteine, D = aspartic acid, E = glutamic acid, F = phenylalanine, G = glycine, H = histidine, I = isoleucine, K = lysine, L = leucine, M = methionine, N = asparagine, O = tyrosine, P = proline, Q = glutamine, R = arginine, S = serine, T = threonine, V = valine, W = tryptophan.
8. M. O. Dayhoff, E. R. Lippincott, R. V. Eck, Science, 146, 1461 (1964).

LEGENDS

Figure 1. Evidence of the primitive ancestry of ferredoxin. At the top the amino acid sequence of ferredoxin from Clostridium pasteurianum is shown in conventional notation (4). Row 1: The sequence is translated into a one-letter code (7), a notation more suitable for this type of study. Rows 2 and 3: The two halves of the chain compared by alignment. Row 4: Twelve amino acids which are identical in both halves. If the sequences were unrelated, one would expect only two or three such identities. The probability of finding this many coincidences by chance is negligible. Row 5: The same seven simple amino acids found in row 4, plus serine, whenever they occur in either chain. Row 6: A simple repeating sequence of four amino acids (with a discontinuity at position 15), from which row 5 appears to be derived. Row 7: Thirteen amino acids from row 5 which conform to row 6. Row 8: The 4 amino acids which do not conform to the cyclic pattern. The chance probability of this many coincidences would be very small if the pattern in row 6 had been independently given. Since it was derived from this study itself, the coincidence is less extreme, but still seems to be good evidence for the validity of the cyclic pattern.

Figure 2: Evidence for a cycle of four in ferredoxin. The two half chains, rows 2 and 3 in Fig. 1, are written in successive groups of four, with a break at positions 15 - 44. Alanine (A) occurs mainly in the first column, D and P in the second, S in the third, and G in the fourth. Exceptions are underlined. This good fit appears unlikely to be due to chance, but a numerical evaluation of the probability would involve several arbitrary assumptions.

Figure 3. Proposed origin and evolution of ferredoxin (see text for fuller details). Row 1: Originally, in an extremely primitive organism, a short

sequence of four of the simplest amino acids (alanine, aspartic acid, serine, and glycine) could be produced. Row 2: This sequence lengthened by doubling of the genetic material, and one discontinuity occurred (underlined). Row 3: The genetic code becoming more versatile, mutations (underlined) occurred, but only to relatively simple amino acids (the same four, plus cysteine, valine, proline, and glutamine). Iron sulfide was attached to the cysteines, which constituted the "active site" of the respiratory function of this primitive ferredoxin. This configuration still persists. Row 4: By "chromosome" duplication, the whole chain doubled. Row 5: The present more intricate genetic code having evolved, further mutations (underlined) to more complex amino acids, occurred. The last three links were deleted. The result was the present sequence of ferredoxin from C. pasteurianum (4).

Position	
1	A O K I
30	F V I <u>D</u>
5	A D S C
34	A D T C
9	V <u>S</u> C G
38	I D C G
13	A C
42	N C
15	A S E
44	<u>A</u> N V
18	C P V N
47	C P V G
22	A I S Q
51	A P V Q
26	<u>G</u> D S I
55	E
Cycle	A D S G P

Figure 2.

1. A D S G

2. A D S G A D S G A D S G A D D S G A D S G A D S G A D S G

3. A D S D A D S C V D C G A C A S V C P V G A P S Q G D S G

4. A D S D A D S C V D C G A C A S V C P V G A P S Q G D S G A D S D A D S C V D C G A C A S V C P V G A P S Q G D S G

5. A Q K I A D S C V S C G A C A S E C P V N A I S Q G D S I E V I D A D I C I D C G N C A N V C P V G A P V Q E

Fig. 3

THERMODYNAMIC AND BIOCHEMICAL STUDIES OF THE ORIGIN OF LIFE

by

Margaret O. Dayhoff
Ellis R. Lippincott
Richard V. Eck

National Biomedical Research Foundation
8600 16th Street, Silver Spring, Md.

The organic evolution of living cells from non-living compounds was a complex process involving a sequence of systems of successively higher organization. These "metabolons" are no longer found living independently, but their nature can be deduced from the structure and function of the organelles, macromolecules, and chemicals of the living cells today, and from a knowledge of the nature of the evolutionary process and the knowledge of which structures are simple. Many of these necessary facts and processes are poorly understood today.

Simple Structures

In order to investigate which chemicals are "simple," a broad survey of the most stable mixtures of compounds formed from carbon, hydrogen, oxygen, nitrogen, sulfur, phosphorus and chlorine, was conducted. The concentrations of biologically interesting compounds in an ideal gas under all elemental compositions and plausible conditions of temperature and pressure was noted. (Dayhoff, Lippincott, and Eck, Science, Dec. 11, 1964) These computations indicate that there were very few molecules of any complexity in an equilibrium primitive atmosphere. The probability of a whole reaction chain, such as those found in cells, coming together from such an extremely dilute mixture is exceedingly remote. Life must therefore have evolved starting as an organized functioning system of very simple compounds. The things available in major proportions in a reducing atmosphere are carbon dioxide, water, methane, hydrogen, carbon monoxide, hydrogen sulfide, nitrogen, HCl, P₄O₆, and possibly a "tar" phase of polyaromatic hydrocarbons. Ammonia is present in small concentration and organic compounds such as thiomethanol, acetic acid, formic acid, methanol and ethane are very dilute (about 10⁻¹⁰ mole percent). Such biologically interesting compounds as ribose and adenine would be so dilute that there would be only about a million molecules in the whole earth in a gaseous equilibrium.

These results lead us to search for some primordial aggregation of the simple compounds which could utilize radiant energy for metabolism.

Evolutionary Processes and Complex Structures

On the other hand, starting from the complexity of present living things, a number of inferences can be made about the structure of very early forms of life, by a mathematical analysis of the experimental results of the studies of the genetic code and of proteins. In spite of the billions of years that have passed, and the tremendous increase in complexity of organization of living chemistry that has occurred, there are still detectable (in the proportions and relationships of the components of nucleic acids and proteins) simple patterns which appear to be surviving traces of a very primitive mechanism from which the present extremely complex genetic mechanism must have evolved, step by step. The concept that the evolution of living things must have occurred one step at a time, and that every step must have been able to survive, becomes a very powerful approach to the unraveling of such patterns. If there were no such simple, general rule constraining the possible numbers of changes which could occur at each stage, it might seem impossible to find these still-living relics of the chemical organization of the very early forms of life.

Plasma Simulation of Equilibrium

According to our computations numerous simple organic compounds have been formed on the primitive earth merely on the basis of thermodynamic equilibrium. In order to verify these calculations experimentally, various gases and gaseous mixtures are being subjected to a high energy radio frequency plasma discharge. From this equilibrium in the gas phase, the product mixtures are analyzed.

Preliminary experimental results on systems containing only carbon and hydrogen have qualitatively corroborated some of these calculations. When a mixture of methane and hydrogen (H:C ratio above 4) was treated in the plasma reactor, no liquid or solid products were obtained. This is in accordance with our predictions. The decomposition of a mixture of methane and ethylene (H:C ratio 4 or below) resulted in the accumulation of tars consisting of numerous high molecular weight polynuclear aromatic hydrocarbons and a small amount of aliphatic material, as calculations had also predicted. Procedures have been developed which should provide efficient and precise analyses of these hydrocarbon mixtures on a routine basis when a suitable gas chromatograph becomes available.

This work was supported by NIH Grants Nos. GM-12168 and GM-08710 and by NASA Contract No. 21-003-002 to the National Biomedical Research Foundation.

Models of Organization and Growth; Information Dissemination ... Chairman: B. G. Farley

Predictions of the Compositions of the Atmospheres of the Planets**Richard V. Eck**

National Biomedical Research Foundation

Silver Spring, Md.

The availability of computers, and the new outlook on computational methods which have been developed for them, makes possible "a new way of doing chemistry." In general, it has been impractical to compute chemical equilibria by mass-action formulae involving more than a few different molecular species. This has inhibited the study of such systems as combustion, in which many reactions are occurring simultaneously. The reason is that one equation with its equilibrium constant must be included for each compound. These equations have unwieldy exponents and coefficients which makes their simultaneous solution difficult.

High-speed computers have stimulated the development of general iterative methods for the solution of mathematical problems, and this makes possible a great simplification of the formulation of problems of chemical thermodynamic equilibrium. It becomes necessary only to include the formulas of components, along with simple experimental or computed data on their thermodynamic free energies of formation. The calculation, done by the method of successive approximations, merely minimizes the total free energy of the system.¹ If any component happens to be omitted, it does not in general invalidate the result unless it is a "major" component. There are a limited number of these, which are generally simple compounds, well known to chemists. Given the atomic proportions of the system, and its temperature and pressure, a very large number of compounds can be computed simultaneously.

Using this procedure, we published a computation of the equilibrium amounts of a large number of biologically interesting compounds which would be formed by simple equilibrium processes in the atmosphere of the primordial Earth.² This suggested that we apply the same method to the atmospheres of the planets. We have now done this, and are able to reach a number of pertinent tentative conclusions: On Venus, it is impossible that there exist hydrocarbons or any solid carbon in contact with the atmosphere. Neither can there be any formaldehyde, or indeed any gases other than nitrogen and carbon dioxide, with only extremely small amounts of carbon monoxide and water, and even lesser amounts of a few other simple compounds. On Mars, there is somewhat less information available, so that the possibility of free oxygen existing is not excluded. Neither is it certain at what rate equilibrium processes can go in the low prevailing temperatures. In any case, however, any oxides of nitrogen which might be formed by energetic processes would tend to be destroyed by equilibrium-type reactions. On

Jupiter, the great excess of hydrogen forces all other elements to occur principally as simple hydrides, such as H₂O, CH₄, H₂S, and NH₃. The colors in the clouds must therefore be accounted for by the products of some special energetic reactions.

In general, if chemical equilibrium cannot account for the presence of a detected compound, some special reaction or other source must be sought. It is well known that the atmosphere of Earth is continually being augmented from a number of sources by such compounds as CO, H₂S, SO₂, CH₄, NH₃, etc. In our present oxidizing atmosphere, all such reducing compounds are degraded as rapidly as they are added, so that a dynamic equilibrium is maintained. Such compounds as might have been produced by physical chemical processes are generally very simple. If an extraterrestrial observer could analyse the atmosphere of Earth, and detect such complex compounds as terpenes (the blue haze in the Great Smoky Mountains, for example), he could very well infer the presence of life, even though he might conjecture that it might be based on some unknown chemistry. This conclusion need not consider the particular nature of the detected compounds, but merely their thermodynamic improbability, so that they could only be accounted for by some complex feedback-controlled process. On the other hand, the complex asphaltic compounds which have been found in the carbonaceous meteorites have been shown to be just the sort which could readily have been formed by thermodynamic equilibrium, according to these calculations.³

When it becomes possible to land instruments on Mars, equipment should be included capable of detecting organic compounds of molecular weight up to at least 150. Whatever such compounds may be, thermodynamic considerations could be applied to the question of their living origin, independent of the particular type of life which might exist.

¹W. B. White, S. M. Johnson, and G. B. Dantzig, *J. Chem. Phys.* 28, 751 (1958).

²M. O. Dayhoff, E. R. Lippincott, and R. V. Eck, *Science* 146, 1461 (1964)

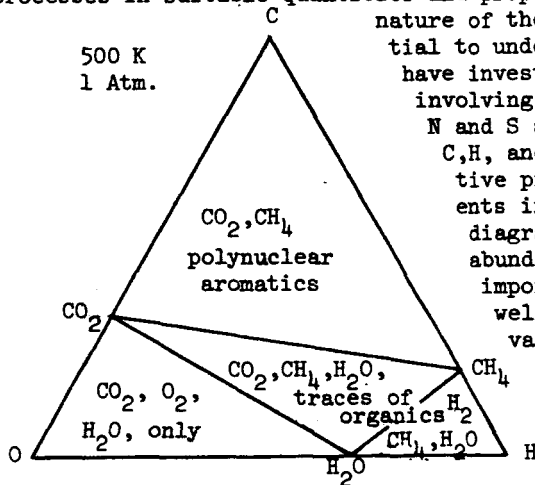
³M. H. Studier, R. Hayatsu, and E. Anders, *Science* (1965), in press.

*This work is supported by National Aeronautic and Space Administration Grant NSR 21-003-002.

TD4

THERMODYNAMIC EQUILIBRIUM ON THE PREBIOLOGICAL EARTH, M. O. Dayhoff and R. V. Eck*, National Biomedical Research Foundation, 8600 - 16th Street, Silver Spring, Md. 20910

Complex compounds were necessary for the origin, structure, and nutrition of the first living things and their predecessors undergoing chemical evolution. These compounds must have been produced by non-living processes in suitable quantities and proportions. A knowledge of the



Compounds present in ideal gas mixture at thermodynamic equilibrium.

nature of the earth's atmosphere is essential to understanding this evolution. We have investigated the gaseous equilibrium involving compounds of C,H,O,N,S, where N and S are in low concentration, and C,H, and O assume all possible relative proportions. The main constituents in each region of the ternary diagram are shown. The relative abundance of compounds of biological importance will be discussed, as well as possible consequences of various evolutionary paths of the atmosphere itself.

Support: NASA contract #NSR 21-003-002 to NBRF.

TA7

PROTEIN SEQUENCES TREATED AS MATHEMATICAL PUZZLES: AN ATLAS OF PROTEIN SEQUENCE AND STRUCTURE. Richard V. Eck* and Margaret O. Dayhoff. National Biomedical Research Foundation, 8600 - 16th St., Silver Spring, Md. 20910

Protein sequence determinations are now being published in increasing numbers because of improved methods and equipment. These sequences contain information about several important aspects of living chemistry. They appear to be random, and crude statistical tests have supported this interpretation. Logically, they cannot be random. Proteins are obviously related to one another in various ways. Furthermore, the hydrogen bonds and other intermolecular forces of amino acid side-groups determine the precise three-dimensional folding of a protein chain, which in turn determines its function. A randomly-assembled protein molecule, unlike a natural protein, would fold itself into any one of many possible alternative (generally useless) shapes. Protein sequences also contain an enormous amount of information about evolutionary relatedness, both of biological species and of biochemical processes. They may also help to elucidate the genetic mechanism. Our cryptogrammic studies require a large amount of sequence information. An ATLAS of PROTEIN SEQUENCE and STRUCTURE is being compiled in which we propose to include all the published sequences, as well as some related structural information. The data are being kept on punched cards, which will simplify updating and minimize typographical errors. Our computer programs for the "Cryptogram" use these data cards directly. This Atlas will be distributed to investigators who have published protein sequences, and after revision, it will be available to others as well. We welcome contributions of data, references, corrections, and suggestions of changes for further editions.

Support: NIH grants Nos. GM-12168 and GM-08710 to NBRF.

Program RELATE

M. O. Dayhoff and R. V. Eck

Program RELATE was written to permit an exhaustive comparison of the sequences of amino acids in two proteins, in a search for regions where they may be identical or related in structure, either through similarity of function or by evolution from a common ancestor.

A comparison is made of each possible fragment of a given length from one protein with each such fragment from the other. A relatedness score is accumulated for each fragment, successive values being determined by a matrix element between the two acids in corresponding fragment positions. Information concerning the five highest scoring matches of each fragment of the first protein are stored. All the high scoring matches from all fragments are then combined, ordered by score and the highest ones printed out, together with the positions of the two fragments in the proteins and the difference between these positions.

When an amino acid mutates, the replacement or "allele" is not random, but related to the original due to natural selection based on functional similarity and also on similarity of the codon which most probably has undergone only a single mutation of one of its three nucleotides. A matrix of relatedness is used to give a score estimating the probability of replacement of each amino acid by each other, and taking into account the probabilities of finding each such match by random chance.

Since gaps may occur in protein sequences as they evolve, it is risky to attempt to match the full length of two proteins suspected of homology. Therefore, in this program, fragments of arbitrary length are matched, in such a way that when a gap occurs it will introduce only a slight irregularity. The more gaps that are suspected, the shorter these test-matching fragments should be. On the other hand, shortening the fragments introduces additional random noise, so a compromise must be effected. Since the optimum length of test fragments will vary in different cases, this number is set separately for each run, depending on the anticipated relatedness of the two proteins. In order to soften the end effect of these essentially arbitrary fragments, a weighting function, tapering off at the ends of the fragment may be used in computing the scores.

A first approximation to an ideal relatedness-score matrix would be unitary, giving merely a count of amino acid identities in the two fragments. However, the various amino acids occur with different frequencies. Finding two matching cysteines by chance is much less likely than finding two matching alanines, and the strength of the inference to be drawn from such a coincidence is therefore much greater. The relatedness score should reflect this effect, as a second approximation. An even more delicate approximation should also contain scores for every possible pair of alleles, some negative, reflecting all the corresponding probabilities.

With experience, we expect to be able to adjust the matrix of relatedness scores so that unrelated sequences will score near zero. This can be simulated with quasi-random sequences, so that the expected standard errors can be found. Any relatedness scores which are sufficiently high can then be interpreted with confidence as indicating some kind and degree of relatedness.

The two proteins being compared are assumed to be of completely known sequence. Any experimental ambiguities may be ignored by substituting a fictitious amino acid which has all zero matrix elements.

PROGRAM RELATE

Illustrative Computer Output

Relate was designed to automatically line up homologous polymers and insert gaps.

The sequence of human hemoglobin alpha is compared with that for human hemoglobin beta. All sequences 7 acids long from the first sequence are compared with all sequences 7 long from the second, a total of 135 X 140 or 18,900 comparisons.

Each comparison accumulated a score from the matrix elements shown on the second page. This matrix assigns a high score to identities, depending somewhat on the frequency with which these acids occur in proteins. Negative values are sometimes given between certain mutations, reflecting the fact that they occur less often than by random chance.

The third page shows the five highest scoring fragments in the second polymer (out of a possible 140) corresponding to each fragment in the first polymer. The difference in position of the first acids of the fragments in the two polymers is given. From this the register and positions of gaps can be deduced to be as follows:

V	L	S	P	A	D	K	T	N	V	K	A	A	W	G	K	V	G	A	H	A	
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	
V	H	L	T	P	E	E	K	S	A	V	T	A	L	W	G	K	V	D	V	.	.
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20		
G	E	O	G	A	E	A	L	E	R	M	F	L	S	F	P	T	T	L	T	O	F
22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43
D	E	V	G	G	E	A	L	G	R	L	L	V	V	O	P	W	T	E	R	F	F
21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42
P	H	F	.	D	L	S	H	G	S	A	Q	V	etc	No further		
44	45	46		47	48	49	50						51	52	53	54	55		gaps are		
E	S	F	G	D	L	S	T	P	D	A	V	M	G	D	P	K	V	etc	indicated		
43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60				

We are experimenting to decide upon the best algorithm to make this last step automatic.

Finally the 150 best scores are shown in numerical order.

A comparison was also made of hemoglobin alpha with myoglobin. Here only one change in register, with the introduction of gaps is indicated.

SEQUENCES OF THE TWO PROTEINS

HEMOGLOBIN ALPHA - HUMAN

V	L	S	P	A	C	K	T	N	V	K	A	A	W	G	K	V	G	A	H	A	G	E	D	G	A	E	A	L	E	
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	
R	F	F	L	S	F	F	T	T	L	T	D	F	P	H	F	D	L	S	H	G	S	A	Q	V	K	G	H	G	K	
31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	
K	V	A	C	A	L	T	N	A	V	A	H	V	D	D	M	P	N	A	L	S	A	L	S	D	L	H	A	H	K	
61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	
L	R	V	C	P	V	N	F	K	L	L	S	H	C	L	L	V	T	L	A	A	H	L	P	A	E	F	T	P	A	
91	92	93	94	95	96	97	98	99	100	101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120	
V	F	A	S	L	C	K	F	L	A	S	V	S	T	V	L	T	S	K	C	R										
121	122	123	124	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140	141										

HEMOGLOBIN BETA - HUMAN

V	F	L	T	P	E	E	K	S	A	V	T	A	L	W	G	K	V	D	V	C	E	V	G	G	E	A	L	G	R	
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	
L	L	V	V	C	P	K	T	E	R	F	F	E	S	F	G	D	L	S	T	P	D	A	V	M	G	C	P	K	V	
31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	
K	A	F	C	K	K	V	L	G	A	F	S	D	G	L	A	H	L	D	D	L	K	G	T	F	A	T	L	S	E	
61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	
L	F	C	C	K	L	F	V	C	P	E	D	F	R	L	L	G	D	V	L	V	C	V	L	A	H	H	F	G	K	
91	92	93	94	95	96	97	98	99	100	101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120	
E	F	T	P	P	V	E	A	A	C	E	K	V	V	A	G	V	A	D	A	L	A	H	K	D	H					
121	122	123	124	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143	144	145	146					

FRAGMENTS 7 ACIDS LONG ARE COMPARED

5 SOLUTIONS HAVING THE HIGHEST SCORES ARE STORED FOR EACH FRAGMENT OF THE FIRST PROTEIN

THE 150 HIGHEST SCORES ARE KEPT

DICTIONARY OF ACIDS A C D E F G H I K L M N O P Q R S T W V *

MATRIX OF SCORES

	A	C	D	E	F	G	H	I	K	L	M	N	O	P	Q	R	S	T	W	V
A	15.0	1.0	5.4	2.4	-6.5	7.4	-3.5	-2.3	-1.9	-5.8	-1.4	1.3	-5.6	2.6	1.2	-0.9	5.6	2.1	-5.2	-1.5
C	1.0	30.0	-1.2	-4.0	2.8	0.	0.2	5.0	-2.3	2.4	1.8	-1.7	1.4	-1.3	1.1	1.0	1.1	2.8	1.1	3.8
D	5.4	-1.2	20.0	11.8	-9.9	10.7	0.5	-9.0	2.2-14.5	-4.7	8.2	-4.5	6.5	7.1	-3.4	2.6	-0.3	-1.4	-11.5	
E	2.4	-4.0	11.8	20.0	-8.3	8.0	0.7	-7.6	0.7-11.6	-1.7	7.5	-3.6	5.7	5.7	-5.8	1.8	-2.6	-0.2	-9.6	
F	-6.5	2.8	-9.9	-8.3	27.0-12.7	0.	11.6	-0.1	14.1	3.7	-9.2	11.4	-1.3	-3.2	4.2	-9.3	-6.6	4.6	7.8	
G	7.4	0.	10.7	8.0-12.7	15.0	-2.7	-7.3	-3.4-10.4	-2.3	5.6	-6.8	5.3	6.1	-8.0	3.7	2.1	-3.4	-4.3		
H	-3.5	0.2	0.5	0.7	0.	-2.7	27.0	-2.0	3.6	0.7	-1.3	1.9	1.5	-0.9	-0.7	3.3	-0.5	-3.0	1.6	-5.1
I	-2.3	5.0	-9.0	-7.6	11.6	-7.3	-2.0	20.0	-1.7	10.0	3.4	-8.4	5.6	-0.3	-4.0	3.4	-5.6	-0.3	1.2	13.2
K	-1.9	-2.3	2.2	0.7	-0.1	-3.4	3.6	-1.7	10.0	-2.5	-2.1	1.5	0.9	-1.6	-1.6	7.9	1.3	0.4	2.4	-4.7
L	-5.8	2.4-14.5	-11.6	14.1-10.4	0.7	10.0	-2.5	17.0	7.2-11.5	8.3	-3.1	-5.6	0.6	-6.7	-0.9	3.1	10.7			
M	-1.4	1.8	-4.7	-1.7	3.7	-2.3	-1.3	3.4	-2.1	7.2	30.0	-3.8	6.9	-1.9	0.2	-0.2	-4.7	0.3	2.6	3.4
N	1.3	-1.7	8.2	7.5	-9.2	5.6	1.9	-8.4	1.5-11.5	-3.8	20.0	-3.8	1.5	6.0	-1.5	6.4	0.2	0.7	-7.4	
O	-5.6	1.4	-4.5	-3.6	11.4	-6.8	1.5	5.6	0.9	8.3	6.9	-3.8	27.0	0.9	0.4	2.5	-5.4	-4.5	6.8	3.2
P	2.6	-1.3	6.5	5.7	-1.3	5.3	-0.9	-0.3	-1.6	-3.1	-1.9	1.5	0.9	20.0	3.6	-5.2	0.1	-5.6	-2.2	-2.4
Q	1.2	1.1	7.1	5.7	-3.2	6.1	-0.7	-4.0	-1.6	-5.6	0.2	6.0	0.4	3.6	20.0	-3.9	-1.7	-1.9	1.7	-4.3
R	-0.9	1.0	-3.4	-5.8	4.2	-8.0	3.3	3.4	7.9	0.6	-0.2	-1.5	2.5	-5.2	-3.9	27.0	3.9	2.4	2.9	1.2
S	5.6	1.1	2.6	1.8	-9.3	3.7	-0.5	-5.6	1.3	-6.7	-4.7	6.4	-5.4	0.1	-1.7	3.9	17.0	4.2	-0.3	-2.2
T	2.1	2.8	-0.3	-2.6	-6.6	2.1	-3.0	-0.3	0.4	-0.9	0.3	0.2	-4.5	-5.6	-1.9	2.4	4.2	17.0	-1.2	3.2
W	-5.2	1.1	-1.4	-0.2	4.6	-3.4	1.6	1.2	2.4	3.1	2.6	0.7	6.8	-2.2	1.7	2.9	-0.3	-1.2	30.0	0.1
V	-1.5	2.8-11.5	-9.6	7.8	-4.3	-5.1	13.2	-4.7	10.7	3.4	-7.4	3.2	-2.4	-4.3	1.2	-2.2	3.2	0.1	17.0	

NUMBERS OF FIRST ACID CF FRAGMENT IN FIRST AND SECOND POLYMERS AND SCORES

POSITIONS	SCORES	PCISIONS	SCORES	POSITIONS	SCORES	POSITIONS	SCORES	POSITIONS	SCORES
1 2 -1	61.30	1 97 -96	51.50	1 134-133	45.60	1 34 -33	45.00	1 133-132	44.70
2 3 -1	70.60	2 98 -96	59.00	2 48 -46	52.30	2 75 -73	37.20	2 54 -52	36.40
3 4 -1	54.90	3 44 -41	45.00	3 99 -96	36.80	3 50 -47	36.20	3 136-133	34.80
4 5 -1	67.70	4 92 -88	48.10	4 100 -96	44.90	4 127-123	42.90	4 14 -10	40.80
5 53 -88	51.20	5 6 -1	48.10	5 15 -10	46.10	5 76 -71	41.60	5 138-133	38.40
6 7 -1	59.70	6 16 -10	54.70	6 94 -88	52.80	6 56 -50	50.80	6 19 -13	43.50
7 123-116	44.00	7 17 -10	42.50	7 8 -1	42.10	7 134-127	38.50	7 57 -50	36.60
8 9 -1	62.10	8 124-116	50.40	8 135-127	46.30	8 65 -57	43.40	8 79 -71	42.70
9 10 -1	72.90	9 125-116	64.00	9 136-127	51.60	9 80 -71	50.40	9 66 -57	46.70
10 11 -1	81.60	10 126-116	72.50	10 137-127	49.60	10 67 -57	47.40	10 133-123	44.40
11 12 -1	81.60	11 127-116	72.50	11 25 -14	51.70	11 82 -71	42.60	11 17 -6	42.40
12 13 -1	91.90	12 128-116	67.50	12 73 -61	48.30	12 18 -6	47.40	12 43 -31	45.20
13 14 -1	75.40	13 129-116	67.50	13 19 -6	56.30	13 56 -43	49.90	13 74 -61	48.30
14 15 -1	81.70	14 57 -43	69.50	14 20 -6	51.60	14 130-116	49.80	14 110 -96	45.10
15 58 -43	78.30	15 21 -6	66.50	15 16 -1	55.10	15 136-121	50.20	15 43 -28	43.80
16 59 -43	69.60	16 67 -51	52.30	16 19 -3	50.60	16 133-117	47.70	16 113 -97	45.40
17 20 -3	68.40	17 68 -51	65.00	17 60 -43	60.30	17 134-117	55.80	17 23 -6	52.70
18 139-121	66.40	18 69 -51	62.60	18 75 -57	56.60	18 90 -72	49.90	18 135-117	47.10
19 76 -57	63.60	19 70 -51	55.00	19 140-121	53.00	19 136-117	47.10	19 125-106	46.70
20 77 -57	56.00	20 19 1	55.30	20 4 16	47.20	20 92 -72	43.50	20 97 -77	42.50
21 20 1	74.80	21 115 -94	57.50	21 119 -98	45.80	21 123-102	42.70	21 138-117	42.00
22 21 1	91.30	22 46 -24	52.00	22 16 6	43.30	22 69 -47	41.10	22 89 -67	40.20
23 22 1	97.60	23 97 -74	56.70	23 90 -67	53.50	23 135-112	51.10	23 127-104	45.00
24 23 1	85.60	24 136-112	51.10	24 98 -74	50.20	24 125-101	46.70	24 18 6	46.30
25 24 1	109.40	25 137-112	61.20	25 126-101	53.70	25 99 -74	47.60	25 34 -9	36.70
26 25 1	101.60	26 138-112	63.40	26 127-101	61.40	26 72 -46	55.70	26 42 -16	45.00
27 26 1	108.30	27 36 -9	77.30	27 128-101	65.80	27 139-112	59.80	27 79 -52	54.20
28 27 1	59.00	28 37 -9	60.00	28 129-101	56.60	28 62 -34	54.30	28 117 -89	50.90
29 38 -9	82.20	29 28 1	81.80	29 118 -89	54.50	29 106 -77	50.90	29 130-101	45.30
30 39 -9	110.10	30 29 1	76.20	30 65 -35	48.30	30 131-101	44.80	30 108 -78	42.30
31 40 -9	95.40	31 30 1	88.20	31 66 -35	47.70	31 104 -73	45.20	31 45 -14	35.40
32 41 -9	68.10	32 31 1	60.00	32 81 -49	47.50	32 44 -12	45.70	32 67 -35	35.50
33 22 1	64.80	33 42 -9	63.50	33 45 -12	44.80	33 103 -70	43.20	33 82 -49	35.40
34 48 -14	59.90	34 116 -82	44.40	34 33 1	44.10	34 8 26	41.70	34 103 -69	41.50
35 117 -82	60.70	35 44 -9	58.40	35 70 -35	53.60	35 80 -45	50.40	35 40 -5	47.40
36 118 -82	62.10	36 71 -35	49.50	36 35 1	49.40	36 85 -49	44.70	36 28 8	44.30
37 36 1	65.00	37 100 -63	56.00	37 85 -48	49.50	37 82 -45	43.30	37 65 -28	42.50
38 20 8	56.30	38 37 1	50.70	38 86 -48	49.90	38 83 -45	45.00	38 66 -28	44.20
39 110 -71	59.50	39 31 8	55.50	39 38 1	51.40	39 140-101	50.50	39 87 -48	48.00
40 39 1	61.40	40 112 -72	60.50	40 111 -71	60.40	40 75 -35	55.80	40 42 -2	52.00
41 40 1	63.70	41 73 -32	70.70	41 113 -72	68.80	41 13 28	51.70	41 93 -52	49.90
42 41 1	66.80	42 114 -72	63.10	42 14 28	60.30	42 74 -32	56.50	42 94 -52	44.00
43 115 -72	56.60	43 75 -32	56.60	43 15 28	54.60	43 95 -52	50.30	43 42 1	48.70
44 51 -47	64.50	44 116 -72	63.10	44 96 -52	50.90	44 16 28	50.70	44 71 -27	48.60
45 117 -72	66.10	45 92 -47	63.30	45 77 -32	58.50	45 45 0	43.60	45 102 -57	42.40
46 46 0	46.20	46 73 -27	42.60	46 18 28	41.20	46 41 5	40.10	46 103 -57	39.70
47 47 0	73.90	47 70 -23	60.30	47 74 -27	57.90	47 80 -33	43.40	47 136 -89	43.20
48 71 -23	54.20	48 48 0	49.60	48 114 -66	43.40	48 75 -27	41.60	48 61 -13	40.20
49 62 -13	53.40	49 72 -23	50.80	49 131 -82	37.00	49 49 0	36.00	49 5 44	33.90
50 63 -13	44.40	50 55 -5	44.30	50 15 35	42.70	50 73 -23	36.00	50 6 44	34.20
51 56 -5	53.00	51 16 35	49.10	51 19 32	48.20	51 74 -23	46.20	51 7 44	40.90
52 57 -5	65.00	52 137 -85	50.30	52 110 -58	41.80	52 134 -82	41.20	52 78 -26	38.50
53 58 -5	77.40	53 138 -85	49.10	53 21 32	47.80	53 111 -58	45.80	53 18 35	42.70
54 59 -5	84.80	54 112 -58	47.20	54 53 1	46.90	54 19 35	40.80	54 22 32	39.90
55 60 -5	96.40	55 98 -43	43.70	55 113 -58	42.70	55 54 1	41.00	55 3 52	36.40
56 61 -5	96.40	56 90 -34	40.20	56 26 30	39.90	56 99 -43	37.40	56 128 -72	37.30
57 62 -5	80.60	57 129 -72	54.20	57 136 -79	43.30	57 27 30	37.70	57 76 -19	37.20
58 63 -5	83.90	58 130 -72	57.50	58 118 -60	42.10	58 77 -19	40.50	58 133 -75	37.30

-1

1

-5

-5

59 64 -5	71.90	59 134 -75	57.40	59 131 -72	54.50	59 47 12	50.30	59 119 -60	44.70
60 125 -75	78.70	60 65 -5	71.00	60 104 -44	59.00	60 69 -9	52.50	60 48 12	50.30
61 136 -75	82.70	61 66 -5	65.20	61 70 -9	58.00	61 117 -56	55.50	61 105 -44	54.30
62 127 -75	86.00	62 67 -5	63.40	62 71 -9	61.80	62 50 12	57.40	62 106 -44	55.10
63 138 -75	69.10	63 51 12	59.60	63 68 -5	53.80	63 119 -56	48.20	63 72 -9	48.20
64 69 -5	70.30	64 135 -71	59.50	64 79 -15	58.30	64 120 -56	57.80	64 139 -75	57.30
65 70 -5	74.60	65 136 -71	69.10	65 121 -56	59.00	65 13 52	56.90	65 80 -15	53.30
66 127 -71	88.70	66 71 -5	86.60	66 110 -44	69.70	66 133 -67	55.10	66 122 -56	52.10
67 72 -5	83.20	67 138 -71	73.30	67 25 42	56.70	67 8 59	48.10	67 111 -44	47.60
68 73 -5	99.00	68 139 -71	66.70	68 93 -25	57.20	68 132 -64	51.70	68 18 50	48.80
69 74 -5	110.80	69 133 -64	70.20	69 19 50	66.90	69 94 -25	65.40	69 140 -71	59.00
70 75 -5	110.60	70 134 -64	70.30	70 20 50	59.80	70 95 -25	58.30	70 42 28	56.40
71 76 -5	98.30	71 96 -25	69.50	71 18 53	56.50	71 91 -20	53.50	71 70 1	53.30
72 77 -5	88.90	72 97 -25	66.10	72 19 53	63.60	72 92 -20	51.90	72 2 70	49.40
73 20 53	66.50	73 78 -5	64.00	73 134 -61	62.30	73 45 28	48.60	73 68 5	46.20
74 79 -5	67.40	74 135 -61	62.30	74 69 5	52.50	74 5 69	47.10	74 124 -50	46.70
75 126 -61	62.50	75 66 9	54.20	75 80 -5	53.00	75 49 26	52.10	75 43 32	50.80
76 4 72	60.70	76 67 9	57.40	76 50 26	56.90	76 137 -61	48.30	76 7 69	48.00
77 5 72	71.10	77 8 69	66.70	77 25 52	64.60	77 100 -23	62.10	77 85 -8	56.50
78 43 35	70.30	78 9 69	68.00	78 66 12	66.80	78 83 -5	63.50	78 86 -8	57.30
79 67 12	85.70	79 10 69	72.30	79 84 -5	69.70	79 74 5	66.10	79 44 35	62.50
80 85 -5	84.60	80 68 12	76.40	80 42 38	76.30	80 75 5	75.70	80 8 72	66.80
81 86 -5	97.50	81 9 72	70.90	81 76 5	62.30	81 43 38	61.70	81 69 12	60.10
82 87 -5	92.90	82 70 12	71.40	82 77 5	64.10	82 44 38	62.00	82 10 72	61.30
83 88 -5	91.30	83 71 12	83.40	83 137 -54	73.20	83 110 -27	67.00	83 78 5	64.60
84 89 -5	84.30	84 138 -54	72.50	84 72 12	66.80	84 58 26	55.20	84 111 -27	53.60
85 90 -5	84.30	85 139 -54	75.20	85 112 -27	69.90	85 61 24	55.10	85 59 26	52.60
86 91 -5	75.80	86 112 -27	63.10	86 140 -54	58.50	86 60 26	58.30	86 98 -12	57.80
87 92 -5	75.80	87 61 26	64.60	87 99 -12	57.80	87 63 24	55.00	87 114 -27	47.70
88 93 -5	68.80	88 115 -27	58.80	88 13 75	52.60	88 102 -14	46.60	88 62 26	46.50
89 94 -5	87.80	89 16 72	57.30	89 30 59	51.50	89 101 -12	45.50	89 52 37	44.80
90 95 -5	77.70	90 17 72	77.00	90 90 0	53.90	90 104 -14	49.50	90 120 -30	49.30
91 96 -5	75.90	91 18 72	72.60	91 91 0	55.10	91 32 59	51.00	91 76 15	46.90
92 97 -5	85.90	92 122 -30	51.90	92 19 73	49.20	92 17 75	46.30	92 92 0	45.90
93 98 -5	90.50	93 20 73	53.30	93 123 -30	45.80	93 93 0	44.80	93 15 78	44.40
94 99 -5	90.50	94 139 -45	52.90	94 124 -30	50.90	94 26 68	49.40	94 79 15	40.50
95 100 -5	87.50	95 27 68	48.30	95 36 59	48.10	95 108 -13	45.20	95 62 33	40.70
96 101 -5	71.20	96 83 12	54.10	96 109 -13	44.30	96 28 68	43.50	96 63 33	41.80
97 102 -5	81.30	97 84 12	59.10	97 57 40	57.70	97 110 -13	54.30	97 44 53	46.50
98 103 -5	76.90	98 111 -12	66.00	98 85 13	61.30	98 58 40	49.50	98 42 56	41.70
99 112 -13	72.30	99 104 -5	66.90	99 139 -40	52.00	99 108 -9	48.70	99 59 40	48.30
100 105 -5	69.70	100 113 -13	64.20	100 140 -40	50.50	100 27 73	48.30	100 108 -8	42.10
101 28 73	71.10	101 106 -5	56.50	101 109 -8	55.10	101 105 -4	51.10	101 114 -13	48.80
102 29 73	49.60	102 28 74	45.80	102 62 40	43.70	102 107 -5	42.70	102 110 -8	41.40
103 29 74	60.80	103 92 11	58.40	103 108 -5	56.00	103 107 -4	54.30	103 82 21	49.30
104 109 -5	70.50	104 30 74	66.10	104 80 24	55.10	104 83 21	51.30	104 108 -4	51.20
105 109 -4	67.40	105 110 -5	63.20	105 31 74	59.90	105 81 24	58.40	105 84 21	54.70
106 110 -4	83.70	106 111 -5	73.20	106 85 21	56.30	106 48 58	44.00	106 82 24	42.10
107 112 -5	76.60	107 111 -4	67.40	107 135 -28	45.20	107 86 21	42.90	107 104 3	42.00
108 113 -5	78.10	108 136 -28	49.30	108 112 -4	49.10	108 73 35	49.00	108 84 24	47.90
109 114 -5	73.00	109 74 35	54.70	109 113 -4	53.70	109 122 -13	49.30	109 68 41	46.60
110 115 -5	76.00	110 95 15	59.10	110 75 35	53.50	110 20 90	48.00	110 114 -4	43.70
111 116 -5	88.00	111 76 35	59.20	111 79 32	53.10	111 39 72	51.70	111 1 110	51.20
112 117 -5	108.50	112 66 46	53.20	112 2 110	53.10	112 40 72	50.40	112 97 15	50.30
113 118 -5	101.50	113 67 46	56.10	113 41 72	53.60	113 4 109	49.20	113 133 -20	47.70
114 119 -5	90.00	114 100 14	55.70	114 68 46	52.80	114 134 -20	52.00	114 5 109	48.60
115 120 -5	101.70	115 69 46	66.60	115 135 -20	65.10	115 79 36	46.90	115 5 110	46.80
116 121 -5	104.30	116 70 46	55.70	116 2 114	49.20	116 140 -24	47.00	116 136 -20	46.60
117 122 -5	99.30	117 3 114	54.10	117 71 46	48.80	117 92 25	43.60	117 41 76	42.20
118 123 -5	77.90	118 57 61	52.90	118 38 80	50.00	118 93 25	46.20	118 4 114	45.60

-5

119	124	-5	69.20	119	69	50	53.70	119	79	40	51.90	119	135	-16	51.50	119	39	80	47.10
120	13	107	66.70	120	125	-5	61.00	120	136	-16	54.30	120	70	50	53.80	120	73	47	53.30
121	126	-5	68.40	121	41	80	55.20	121	1	120	51.80	121	137	-16	50.50	121	74	47	50.10
122	75	47	68.50	122	97	25	66.00	122	42	80	61.50	122	127	-5	59.20	122	25	97	56.00
123	26	97	75.70	123	128	-5	69.20	123	76	47	65.30	123	36	87	62.20	123	43	80	54.80
124	27	97	70.80	124	129	-5	69.20	124	37	87	63.00	124	77	47	57.70	124	44	80	53.50
125	38	87	80.30	125	130	-5	67.30	125	28	97	63.00	125	78	47	62.40	125	68	57	55.20
126	39	87	89.00	126	131	-5	76.00	126	65	61	57.80	126	103	23	56.10	126	107	19	54.30
127	40	87	80.90	127	66	61	72.60	127	132	-5	69.80	127	104	23	59.30	127	83	44	51.00
128	41	87	72.70	128	67	61	62.30	128	44	84	59.50	128	133	-5	59.50	128	105	23	54.60
129	85	44	69.10	129	45	84	66.40	129	42	87	56.40	129	75	54	55.20	129	28	101	52.50
130	135	-5	56.50	130	86	44	55.70	130	128	2	53.70	130	26	104	53.50	130	69	61	50.20
131	27	104	53.30	131	44	87	49.20	131	109	22	45.20	131	70	61	44.90	131	10	121	44.80
132	106	26	52.40	132	110	22	46.90	132	28	104	45.50	132	38	94	43.70	132	88	44	41.90
133	65	68	43.30	133	11	122	42.80	133	89	44	41.20	133	39	94	40.40	133	111	22	35.80
134	139	-5	53.80	134	112	22	53.40	134	12	122	48.20	134	108	26	41.60	134	9	125	40.30
135	140	-5	57.40	135	34	101	43.20	135	36	99	43.00	135	113	22	42.60	135	109	26	41.00

NUMBERS OF FIRST ACID OF FRAGMENT IN FIRST AND SECOND POLYMERS AND SCORES

POSITIONS	SCORES	POSITIONS	SCORES	POSITIONS	SCORES
69 74	110.80	15 58	78.30	16 59	69.60
70 75	110.60	108 113	78.10	71 96	69.50
30 39	110.10	118 123	77.90	14 57	69.50
25 24	109.40	90 95	77.70	119 124	69.20
112 117	108.50	53 58	77.40	123 128	69.20
27 26	108.30	27 36	77.30	124 129	69.20
116 121	104.30	90 17	77.00	65 126	69.10
115 120	101.70	98 103	76.90	63 128	69.10
26 25	101.60	107 112	76.60	129 85	69.10
113 118	101.50	80 68	76.40	41 113	68.80
117 122	99.30	80 42	76.30	88 93	68.80
68 73	99.00	30 29	76.20	122 75	68.50
28 27	99.00	126 131	76.00	121 126	68.40
71 76	98.30	110 115	76.00	17 20	68.40
23 22	97.60	91 96	75.90	32 41	68.10
81 86	97.50	86 91	75.80	78 9	68.00
56 61	96.40	87 92	75.80	4 5	67.70
55 60	96.40	80 75	75.70	13 129	67.50
31 40	95.40	123 26	75.70	12 128	67.50
82 87	92.90	13 14	75.40	107 111	67.40
12 13	91.90	85 139	75.20	74 79	67.40
22 21	91.30	21 20	74.80	105 109	67.40
83 88	91.30	65 70	74.60	125 130	67.30
94 99	90.50	47 47	73.90	83 110	67.00
93 98	90.50	67 138	73.30	69 19	66.90
114 119	90.00	106 111	73.20	99 104	66.90
126 39	89.00	83 127	73.20	80 8	66.80
72 77	88.90	109 114	73.00	78 66	66.80
66 127	88.70	9 10	72.90	84 72	66.80
31 30	88.20	128 41	72.70	42 41	66.80
62 127	88.00	91 18	72.60	77 8	66.70
111 116	88.00	127 66	72.60	120 13	66.70
89 94	87.80	84 138	72.50	68 139	66.70
95 100	87.50	11 127	72.50	115 69	66.60
66 71	86.60	10 126	72.50		
92 97	85.90	79 10	72.30		
24 23	85.60	99 112	72.30		
79 67	85.30	59 64	71.90		
54 59	84.80	82 70	71.40		
80 85	84.60	96 101	71.20		
85 90	84.30	101 28	71.10		
84 89	84.30	77 5	71.10		
58 63	83.90	60 65	71.00		
106 110	83.70	81 9	70.90		
41 40	82.70	124 27	70.80		
83 71	83.40	41 73	70.70		
67 72	83.20	2 3	70.60		
61 126	82.70	104 109	70.50		
29 38	82.20	78 43	70.30		
29 28	81.80	64 69	70.30		
14 15	81.70	70 134	70.30		
10 11	81.60	69 133	70.20		
11 12	81.60	85 112	69.90		
97 102	81.30	33 32	69.80		
127 40	80.90	127 132	69.80		
57 62	80.60	79 84	69.70		
125 38	80.30	100 105	69.70		
60 125	78.70	66 110	69.70		

SEQUENCES OF THE TWO PROTEINS

HEMOGLOBIN ALPHA - HUMAN

V	L	S	P	A	E	K	T	N	V	K	A	A	W	G	K	V	G	A	H	A	G	E	D	G	A	E	A	L	E	
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	
R	F	F	L	S	F	P	T	T	L	T	D	F	P	H	F	D	L	S	H	G	S	A	Q	V	K	G	F	G	K	
31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	
K	V	A	E	A	L	T	N	A	V	A	H	V	D	D	M	P	N	A	L	S	A	L	S	D	L	H	A	H	K	
61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	
L	R	V	E	P	V	N	F	K	L	L	S	H	C	L	L	V	T	L	A	A	F	L	P	A	E	F	T	P	A	
91	92	93	94	95	96	97	98	99	100	101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120	
V	F	A	S	L	E	K	F	L	A	S	V	S	T	V	L	T	S	K	D	R										
121	122	123	124	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140	141										

MYOGLOBIN - SPERM WHALE

V	L	S	E	G	E	W	Q	L	V	L	H	V	W	A	K	V	E	A	D	V	A	G	H	G	Q	D	I	L	I
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
R	L	F	K	S	F	P	E	T	L	E	K	F	D	R	F	K	H	L	K	T	E	A	E	M	K	A	S	E	D
31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60
L	K	K	F	G	V	T	V	L	T	A	L	G	A	I	L	K	K	K	G	P	F	E	A	E	L	K	P	L	A
61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90
C	S	F	A	T	K	F	K	I	P	I	K	D	L	E	F	I	S	E	A	I	I	H	V	L	H	S	R	H	P
91	92	93	94	95	96	97	98	99	100	101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120
G	N	F	G	A	E	A	Q	G	A	M	N	K	A	L	E	L	F	R	K	D	I	A	A	K	D	K	E	L	G
121	122	123	124	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143	144	145	146	147	148	149	150
D	G																												

NUMBERS OF FIRST ACID OF FRAGMENT IN FIRST AND SECOND POLYMERS AND SCORES

POSITIONS	SCORES	POSITIONS	SCORES	POSITIONS	SCORES	POSITIONS	SCORES	POSITIONS	SCORES
1 1 0	78.30	1 106-105	44.80	1 68 -67	43.80	1 49 -48	42.80	1 75 -74	42.30
2 2 0	59.40	2 89 -87	47.40	2 86 -84	40.50	2 118-116	37.70	2 61 -59	36.20
3 17 -14	41.50	3 92 -89	39.30	3 119-116	38.40	3 20 -17	37.60	3 35 -32	37.50
4 83 -79	42.60	4 124-120	40.60	4 143-139	39.30	4 18 -14	38.60	4 93 -89	35.50
5 84 -79	35.00	5 90 -85	34.90	5 94 -89	34.80	5 144-139	33.30	5 125-120	33.20
6 65 -59	41.70	6 119-113	40.40	6 51 -45	38.40	6 82 -76	34.10	6 138-132	32.90
7 120-113	45.30	7 52 -45	44.30	7 14 -7	44.10	7 139-132	40.90	7 69 -62	35.80
8 53 -45	43.40	8 15 -7	41.80	8 121-113	41.70	8 70 -62	41.40	8 140-132	35.80
9 54 -45	52.00	9 67 -58	49.90	9 16 -7	47.10	9 122-113	45.70	9 12 -3	42.50
10 10 0	56.90	10 68 -58	47.80	10 3 7	42.90	10 17 -7	42.20	10 55 -45	42.00
11 4 7	62.10	11 11 0	56.90	11 69 -58	44.00	11 93 -82	43.60	11 128-117	37.50
12 12 0	67.40	12 5 7	51.00	12 94 -82	45.30	12 38 -26	45.10	12 126-114	44.20
13 13 0	85.90	13 119-106	48.60	13 68 -55	42.50	13 39 -26	40.80	13 6 7	40.10
14 14 0	87.90	14 18 -4	68.20	14 120-106	52.60	14 76 -62	43.50	14 87 -73	42.40
15 19 -4	75.80	15 121-106	69.80	15 15 0	56.40	15 88 -73	55.00	15 59 -44	50.00
16 20 -4	74.50	16 122-106	60.90	16 16 0	56.40	16 78 -62	54.60	16 89 -73	51.80
17 21 -4	84.10	17 123-106	67.40	17 79 -62	64.60	17 90 -73	55.00	17 17 0	54.40
18 80 -62	77.60	18 22 -4	72.70	18 117 -99	66.30	18 91 -73	58.00	18 124-106	54.00
19 118 -99	77.60	19 81 -62	59.20	19 35 -16	56.90	19 23 -4	54.90	19 144-125	52.90
20 119 -99	93.50	20 82 -62	65.30	20 124-104	56.80	20 36 -16	49.40	20 142-122	48.10
21 120 -99	78.30	21 79 -58	58.30	21 125-104	57.80	21 53 -32	55.20	21 52 -31	51.10
22 121 -99	90.70	22 80 -58	54.40	22 104 -82	54.30	22 53 -31	51.10	22 54 -32	45.60
23 105 -82	74.70	23 122 -99	70.10	23 55 -32	54.60	23 23 0	52.20	23 132-109	38.30
24 123 -99	70.60	24 56 -32	57.00	24 35 -11	53.60	24 106 -82	47.10	24 130-106	46.10
25 39 -14	67.30	25 36 -11	66.90	25 57 -32	64.00	25 25 0	62.10	25 124 -99	58.30
26 125 -99	73.30	26 37 -11	73.30	26 40 -14	68.90	26 132-106	58.30	26 58 -32	55.30
27 27 0	80.10	27 41 -14	74.60	27 133-106	61.20	27 38 -11	60.80	27 129-102	58.30
28 28 0	65.80	28 134-106	58.00	28 42 -14	55.30	28 71 -43	54.10	28 98 -70	43.50
29 29 0	85.10	29 43 -14	50.50	29 99 -70	47.20	29 102 -73	46.80	29 135-106	45.60
30 30 0	68.10	30 100 -70	64.20	30 27 3	54.90	30 117 -87	49.20	30 26 4	47.80
31 31 0	95.70	31 101 -70	58.20	31 118 -87	52.70	31 104 -73	47.20	31 67 -36	44.00
32 32 0	66.10	32 102 -70	59.00	32 45 -13	53.30	32 105 -73	46.30	32 28 4	44.60
33 33 0	75.90	33 103 -70	58.50	33 46 -13	50.90	33 106 -73	47.70	33 29 4	45.40
34 34 0	65.90	34 66 -32	55.70	34 70 -36	42.90	34 97 -63	42.00	34 104 -70	41.30
35 35 0	65.80	35 67 -32	47.10	35 71 -36	44.20	35 118 -83	42.40	35 98 -63	40.40
36 36 0	49.70	36 145-109	44.70	36 72 -36	39.50	36 68 -32	37.30	36 99 -63	35.50
37 37 0	76.70	37 66 -29	60.30	37 132 -95	53.70	37 27 10	53.00	37 100 -63	50.90
38 67 -29	68.00	38 38 0	63.20	38 99 -61	51.30	38 133 -95	47.00	38 94 -56	46.10
39 42 -3	69.10	39 39 0	69.10	39 100 -61	51.60	39 75 -36	50.20	39 134 -95	50.20
40 43 -3	82.80	40 40 0	79.10	40 101 -61	68.80	40 69 -29	55.90	40 76 -36	50.50
41 44 -3	70.90	41 41 0	64.30	41 38 3	63.30	41 102 -61	61.40	41 67 -26	50.00
42 45 -3	70.30	42 42 0	67.60	42 39 3	66.50	42 100 -58	52.40	42 103 -61	49.40
43 46 -3	69.60	43 40 3	61.70	43 43 0	60.00	43 137 -94	51.60	43 111 -68	49.30
44 87 -43	55.40	44 41 3	51.20	44 30 14	48.70	44 110 -66	43.70	44 112 -68	41.00
45 88 -43	64.40	45 31 14	54.30	45 48 -3	48.70	45 111 -66	44.80	45 103 -58	43.40
46 89 -43	69.50	46 32 14	52.80	46 112 -66	50.70	46 115 -69	47.80	46 104 -58	47.50
47 90 -43	53.50	47 88 -41	46.60	47 78 -31	44.50	47 60 -13	41.40	47 33 14	40.80
48 2 46	66.30	48 79 -31	48.00	48 91 -43	47.40	48 34 14	45.10	48 138 -90	42.90
49 80 -31	61.20	49 3 46	60.00	49 23 26	56.10	49 63 -14	49.60	49 92 -43	48.30
50 81 -31	67.50	50 56 -6	59.20	50 12 38	53.50	50 24 26	50.70	50 64 -14	48.70
51 57 -6	52.20	51 17 34	49.70	51 82 -31	45.80	51 51 0	45.40	51 88 -37	42.40
52 10 34	81.00	52 58 -6	71.80	52 83 -31	49.20	52 52 0	42.80	52 143 -91	40.90
53 19 34	94.20	53 59 -6	69.80	53 84 -31	54.80	53 144 -91	50.30	53 53 0	49.00
54 20 34	77.80	54 60 -6	61.70	54 32 22	48.90	54 77 -23	43.70	54 8 46	40.70
55 21 34	72.70	55 61 -6	55.00	55 33 22	54.90	55 90 -35	45.40	55 78 -23	43.40
56 22 34	88.80	56 62 -6	61.80	56 34 22	57.80	56 91 -35	41.80	56 79 -23	38.50
57 23 34	65.00	57 35 22	51.20	57 63 -6	45.50	57 92 -35	41.50	57 81 -24	35.90
58 116 -58	55.40	58 119 -61	51.00	58 36 22	49.70	58 82 -24	49.10	58 64 -6	48.60

59 121 -62	62.10	59 18 41	47.50	59 124 -65	40.70	59 54 5	37.20	59 15 44	36.60
60 105 -45	56.20	60 140 -80	52.20	60 15 45	50.00	60 87 -27	47.50	60 55 5	46.20
61 106 -45	55.20	61 16 45	54.00	61 56 5	48.70	61 67 -6	45.40	61 141 -80	42.60
62 107 -45	57.20	62 17 45	49.60	62 142 -80	47.90	62 68 -6	46.30	62 124 -62	43.50
63 128 -65	50.80	63 125 -62	46.40	63 108 -45	42.50	63 127 -64	42.10	63 19 44	39.10
64 129 -65	60.30	64 105 -41	55.20	64 109 -45	47.60	64 70 -6	43.50	64 136 -72	42.60
65 120 -65	53.00	65 137 -72	45.80	65 67 -2	45.20	65 16 49	44.90	65 88 -23	63.20
66 107 -41	74.60	66 138 -72	48.10	66 17 49	44.10	66 91 -25	43.80	66 68 -2	39.60
67 108 -41	81.60	67 7 60	54.20	67 92 -25	44.70	67 43 24	44.30	67 69 -2	42.10
68 109 -41	62.90	68 38 30	55.10	68 8 60	54.00	68 44 24	53.10	68 12 56	47.90
69 110 -41	55.90	69 21 48	54.20	69 9 60	53.40	69 13 56	51.40	69 20 49	47.40
70 22 48	59.10	70 10 60	57.10	70 14 56	48.20	70 40 30	45.80	70 34 36	44.20
71 23 48	57.50	71 35 36	54.60	71 19 52	52.60	71 118 -47	51.60	71 144 -73	48.30
72 119 -47	53.80	72 12 60	51.00	72 36 36	50.50	72 48 24	48.00	72 116 -44	45.30
73 128 -55	59.80	73 21 52	48.50	73 124 -51	43.90	73 52 21	42.90	73 33 40	40.90
74 129 -55	81.10	74 125 -51	55.40	74 34 40	50.10	74 22 52	41.50	74 53 21	40.90
75 120 -55	72.20	75 126 -51	56.40	75 35 40	49.70	75 118 -43	40.00	75 54 21	38.10
76 131 -55	61.00	76 119 -43	58.40	76 68 8	51.20	76 36 40	45.20	76 17 59	39.50
77 37 40	60.60	77 69 8	57.80	77 66 11	49.50	77 83 -6	48.30	77 125 -48	46.10
78 67 11	55.60	78 70 8	54.20	78 126 -48	49.90	78 84 -6	48.20	78 133 -55	47.50
79 68 11	60.80	79 71 8	56.20	79 85 -6	54.00	79 134 -55	48.20	79 65 14	47.40
80 69 11	72.20	80 66 14	57.00	80 55 25	55.60	80 43 37	48.70	80 104 -24	45.30
81 70 11	56.00	81 87 -6	53.90	81 56 25	52.00	81 84 -3	47.30	81 110 -29	45.30
82 88 -6	67.60	82 68 14	54.40	82 71 11	49.90	82 57 25	48.80	82 114 -32	47.80
83 89 -6	62.00	83 58 25	60.80	83 107 -24	55.70	83 18 65	55.60	83 69 14	53.90
84 59 25	64.10	84 19 65	63.80	84 90 -6	55.00	84 140 -56	47.30	84 113 -29	44.50
85 60 25	73.00	85 80 5	56.00	85 20 65	52.60	85 91 -6	50.10	85 117 -32	45.70
86 112 -26	60.10	86 61 25	55.40	86 81 5	53.20	86 92 -6	50.90	86 115 -29	45.40
87 53 -6	70.80	87 24 63	59.70	87 62 25	55.40	87 95 -8	50.10	87 82 5	50.10
88 96 -8	55.30	88 94 -6	50.30	88 135 -47	48.80	88 35 53	44.30	88 134 -46	41.00
89 57 -8	58.10	89 136 -47	54.30	89 47 42	49.70	89 82 7	49.00	89 116 -27	42.00
90 127 -47	52.10	90 60 30	45.20	90 136 -46	45.00	90 105 -15	43.60	90 98 -8	41.80
91 138 -47	55.90	91 137 -46	45.60	91 61 30	43.20	91 116 -25	42.00	91 102 -11	40.80
92 117 -25	68.30	92 100 -8	56.30	92 145 -53	46.70	92 39 53	43.50	92 45 47	41.70
93 118 -25	61.00	93 101 -8	59.80	93 86 7	43.00	93 40 53	41.00	93 146 -53	37.20
94 119 -25	54.00	94 26 68	51.80	94 129 -35	51.70	94 5 89	48.90	94 97 -3	43.50
95 98 -3	60.00	95 27 68	58.80	95 83 12	49.70	95 100 -5	46.10	95 132 -37	44.00
96 59 -3	63.40	96 28 68	53.60	96 84 12	49.60	96 102 -6	47.40	96 133 -37	46.40
97 42 55	73.70	97 6 91	58.50	97 134 -37	54.70	97 103 -6	52.80	97 45 52	51.50
98 43 55	74.60	98 104 -6	57.60	98 31 67	56.00	98 7 91	54.80	98 101 -3	53.70
99 60 39	55.70	99 105 -6	53.50	99 8 91	53.30	99 136 -37	48.10	99 109 -10	47.10
100 106 -6	62.80	100 61 35	52.60	100 9 91	49.10	100 110 -10	47.10	100 45 55	42.00
101 62 39	52.60	101 111 -10	50.70	101 95 6	45.20	101 107 -6	43.60	101 24 77	42.70
102 63 39	54.20	102 96 6	46.50	102 25 77	44.40	102 112 -10	43.10	102 6 96	39.90
103 26 77	57.70	103 97 6	53.50	103 64 39	52.00	103 113 -10	49.40	103 7 96	48.80
104 67 37	60.20	104 27 77	51.90	104 9 95	46.30	104 28 76	44.50	104 8 96	42.00
105 68 37	72.40	105 9 96	55.90	105 28 77	51.20	105 138 -33	49.70	105 29 76	45.10
106 69 37	59.70	106 30 76	55.10	106 66 40	51.90	106 75 31	51.10	106 113 -7	49.20
107 70 37	59.70	107 66 41	54.50	107 19 88	47.90	107 140 -33	46.90	107 7 100	45.60
108 71 37	54.90	108 20 88	53.00	108 141 -33	50.00	108 32 76	46.50	108 114 -6	46.30
109 9 100	60.20	109 21 88	58.70	109 13 96	55.50	109 142 -33	53.70	109 78 31	51.40
110 79 31	73.90	110 120 -10	58.10	110 14 96	56.60	110 46 64	56.10	110 10 100	43.90
111 80 31	89.90	111 117 -6	71.50	111 15 96	69.60	111 47 64	56.10	111 11 100	53.20
112 81 31	82.90	112 118 -6	68.00	112 16 96	56.70	112 12 100	56.40	112 48 64	55.40
113 82 31	75.90	113 119 -6	67.30	113 17 96	58.40	113 51 62	46.50	113 142 -29	45.80
114 120 -6	72.00	114 83 31	65.40	114 52 62	53.00	114 103 11	49.30	114 18 96	44.20
115 84 31	62.20	115 121 -6	50.50	115 22 93	47.30	115 15 100	49.90	115 105 10	44.80
116 6 110	55.80	116 110 6	50.00	116 18 98	48.50	116 105 11	46.70	116 85 31	46.58
117 138 -21	57.90	117 111 6	52.20	117 9 108	44.30	117 107 10	44.20	117 114 3	42.40
118 89 29	46.90	118 112 6	46.50	118 52 66	42.40	118 20 98	40.70	118 115 3	38.30

-6

119	15	104	46.70	119	129	-10	46.30	119	37	82	46.10	119	113	6	45.50	119	90	29	45.30
120	67	53	55.20	120	12	108	53.90	120	114	6	52.90	120	130	-10	52.80	120	71	49	51.60
121	13	108	55.50	121	92	29	52.00	121	68	53	51.20	121	115	6	51.00	121	39	82	50.60
122	40	82	74.40	122	37	85	72.50	122	93	29	65.80	122	132	-10	58.90	122	83	39	49.20
123	41	82	71.20	123	70	53	62.10	123	38	85	58.90	123	133	-10	57.60	123	129	-6	48.90
124	42	82	64.30	124	71	53	58.10	124	134	-10	57.60	124	39	85	54.60	124	130	-6	44.90
125	29	96	59.20	125	43	82	56.30	125	135	-10	54.60	125	72	53	53.80	125	86	39	44.10
126	126	-10	50.80	126	26	100	50.20	126	15	111	46.20	126	121	5	44.90	126	27	99	42.60
127	102	25	62.70	127	12	115	49.60	127	67	60	47.30	127	16	111	46.40	127	31	96	46.20
128	103	25	50.10	128	138	-10	49.20	128	68	60	49.00	128	13	115	48.10	128	33	95	46.00
129	69	60	54.40	129	40	89	42.30	129	43	86	40.20	129	66	63	38.30	129	106	23	37.70
130	70	60	54.40	130	109	21	47.60	130	66	64	45.50	130	63	67	42.10	130	5	125	40.70
131	64	67	61.00	131	71	60	52.70	131	67	64	49.10	131	110	21	41.20	131	65	66	40.00
132	65	67	67.10	132	111	21	52.60	132	68	64	50.50	132	29	103	49.70	132	72	60	48.40
133	66	67	68.90	133	73	60	47.70	133	112	21	47.30	133	30	103	42.60	133	69	64	31.80
134	67	67	64.30	134	113	21	54.40	134	31	103	49.10	134	70	64	46.80	134	140	-6	43.60
135	114	21	52.20	135	141	-6	51.10	135	28	107	50.60	135	68	67	46.40	135	39	96	41.30

ILLUSTRATIVE OUTPUT FROM PHYLOGENETIC TREE PROGRAM

Sequences of cytochromes C from widely diverse species are selected to illustrate the operation of this program. The sequences are aligned, with gaps if necessary. The amino acids identical in all sequences are noted. The alternative amino acids or alleles are given at each position in order of their frequency. A matrix of the number of amino acid differences between each pair of species is given. A tentative phylogenetic tree is drawn. Ultimately, the computer will produce the topology of this tree. In addition quantitative aspects such as common ancestor sequences and the number and kind of mutations between species and ancestors will be calculated.

HOMOLOGOUS CYTOCHROMES

	1									2									3									4									5																			
	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9							
HUMAN	-	-	-	-	G	D	V	E	K	G	K	K	I	F	I	M	K	C	S	Q	C	H	T	V	E	K	G	G	K	H	K	T	G	P	N	L	H	G	L	F	G	R	K	T	G	Q	A	P	G	O	S	O	T	A	I	
HORSE	-	-	-	-	G	D	V	E	K	G	K	K	I	F	V	Q	K	C	A	Q	C	H	T	V	E	K	G	G	K	H	K	T	G	P	N	L	H	G	L	F	G	R	K	T	G	Q	A	P	G	O	S	O	T	A	I	
CHICKEN	-	-	-	-	G	D	I	E	K	G	K	K	I	F	V	Q	K	C	S	Q	C	H	T	V	E	K	G	G	K	H	K	T	G	P	N	L	H	G	L	F	G	R	K	T	G	Q	A	P	G	O	S	O	T	A	I	
RATTLESNAKE	-	-	-	-	G	D	V	E	K	G	K	K	I	F	I	T	K	C	S	Q	C	H	T	V	E	K	G	G	K	H	K	T	G	P	N	L	H	G	L	F	G	R	K	T	G	Q	A	P	G	O	S	O	T	A	I	
TUNAFISH	-	-	-	-	G	D	V	A	K	G	K	K	I	F	V	Q	K	C	A	Q	(C.H)	T	V	E	N	G	G	K	H	K	(V.G.P.N)	L	H	G	L	F	G	R	K	T	(G.Q)	A	P	G	O	S	O	T	A	I						
SAMIA MOTH	-	G	V	P	A	G	N	A	E	N	G	K	K	I	F	V	Q	R	C	A	Q	C	H	T	V	E	A	G	G	K	H	K	V	G	P	N	L	H	G	L	F	G	R	K	T	G	Q	A	P	G	O	S	O	S	N	A
NEUROSPORA	-	G	F	S	A	G	D	S	K	K	G	A	N	L	F	K	T	R	C	A	E	C	H	G	E	G	G	N	L	T	Q	K	I	G	P	A	L	H	G	L	F	G	R	K	T	G	S	V	D	G	O	A	O	T	D	A
YEAST	T	E	F	K	A	G	S	A	K	K	G	A	T	L	F	K	T	R	C	E	L	C	H	T	V	E	K	G	G	P	H	K	V	G	P	N	L	H	G	I	F	G	R	H	S	G	Q	A	Q	G	O	S	O	T	D	A
COMMON					G					G			F				C		C	H								K	G	P	L	G		GR		G					G			O		A										
ALLELES	-	-	-	-	D	V	E	K		K	K	I	V	Q	K	A	Q				T	V	E	K	G	G	K	H	T	N	H	L	F		K	T	Q	A	P		Q	S	T	D												
	T	G	F	P	A	N	A	K	N	A	N	L	I	T	R	S	E				G	E	G	N	N	L	T	Q	V	A	M	F	O		H	S	S	V	E		F	T	S	A												
	E	V	S		S	I	A		S		T	T	K	M		E	L				A						P	I				I																								

NUMBER OF AMINO ACID DIFFERENCES BETWEEN SEQUENCES.

		H	H	C	R	T	M	N	B
		U	O	H	S	F	T	E	Y
HUMAN	HU	0	12	11	14	21	31	48	45
HORSE	HO	12	0	9	22	19	29	46	46
CHICKEN	CH	11	9	0	19	18	28	47	45
RATTLESNAKE	RS	14	22	19	0	26	31	47	47
TUNAFISH	TF	21	19	18	26	0	32	48	47
SAMIA MOTH	MT	31	29	28	31	32	0	47	47
NEUROSPORA	NE	48	46	47	47	48	47	0	41
YEAST	BY	45	46	45	47	47	47	41	0

(1)

PROGRAM ALLELE

R. V. Eck and M. O. Dayhoff

Proteins which are closely related, such as hemoglobins from different species, may be lined up and the substitutions of one amino acid for another observed. These pairs are known as alleles. From the observation of a great many such alleles, a matrix of relative deviations from random pairing is formed by program ALLELE. This matrix can now be used in determining a score of relatedness of two amino acid sequences of more distant kinship, such as myoglobin and hemoglobin. Here identities are not sufficient to choose positioning and to indicate where gaps should come.

Based on the frequency of occurrence of each acid in all alleles, an expected frequency of each pair can be computed. This would presume that the replacements were entirely random.

The observed values deviate from the expected values since the functional replacements depend on structures of the two acids and on the codons which produce them. A matrix of these deviations is derived. High matrix values reflect the relatedness of the two acids. Negative values represent a poor ability of the acids to substitute for each other.

Since the amount of allele data is scanty and subject to high digital noise, it is desirable to smooth it. This may be done by including corrections to the terms based on the correlation of the behavior of each pair of amino acids with each of the others. Those of similar structure to the pair should give a positive correlation with both. Those of dissimilar structure should give a negative correlation with both. There will be less noise in these values since they are based on many additive terms. However, there are more assumptions involved in their validity. Therefore, the two matrices are added, giving the second one a weight so that its effect is comparable with the digital noise.

Further a grouping of the acids for similarity of function can be made on the basis of the correlations.

From the observed matrix B, an expected matrix E is obtained such that

$$E(I,J) = \frac{\sum_I E(I,J) \sum_J E(I,J)}{\sum_{I,J} E(I,J)}$$

and the sum of all non-diagonal terms is the same in E and B.

Deviations from expected are obtained by subtracting:

$$D = B - E$$

Relative deviations are next calculated. If there were a large amount of data, this would just involve dividing each term of D by the corresponding term of E or B, whichever is smaller. (This alternative makes a symmetrical arrangement from positive and negative deviations.) However, since this value is extremely sensitive to digital noise for small expected values, a 1 is added.

$$C(I,J) = D(I,J)/B(I,J)+1 \text{ or } D(I,J)/E(I,J)+1$$

Now, the alternative derivation of this matrix yielding S is performed by considering the correlations of the alleles formed with each other acid by both acids of a pair.

First we form $Q = \sqrt{C}$

$$\text{Then } S(I,J) = \sum_K Q(I,K) Q(K,J) / 18$$

S is highly correlated with C. An improved matrix of probable deviations C' is obtained by adding these two matrices with a scale factor α making S comparable with the digital noise in the observations.

$$C' = C + \alpha S$$

From this smoothed matrix C', inverse operations may be applied to generate improved "observed" allele data.

=ALLELE=

Illustrative Computer Output

A table of allele frequencies was obtained by tabulating the change in the homologous families cytochrome, hemoglobin, trypsin, etc. This input data is shown below.

Based on random pairing of these acids, the expected allele frequencies shown on the first page are calculated. Deviations of the observed values from those expected are shown next. A smoothed matrix of deviations is found finally. This has included values derived from the correlation of the behavior of each pair of amino acids with each of the others.

From the final matrix of deviations, a smoothed table of allele frequencies can be derived which is more correct than the original observed data. Patterns of correlation of behavior of the acids may be derived.

OBSERVED ALLELE FREQUENCIES

	D	E	G	N	P	Q	A	S	T	C	F	I	L	M	O	W	V	R	K	H
D	10	6	9	1	1	5	6	2	1				1		1	1	5		3	1
E	10		3		4	11	8	6	1				1				4		4	1
G	6	3		1	3		8	7	4		2		1		1		3		3	2
N	9		1				3	4	1				1		1		1	2	1	2
P	1	4	3			1	5	2			1		1		1	1	2			1
Q	1	11			1		2	1	1					1			2	2	1	
A	5	8	8	3	5	2		11	8			1	3		1		9		2	2
S	6	6	7	4	2	1	11		7		1		1	1			3		3	1
T	2	1	4	1		1	8	7		1		3	1				5		1	2
C	1								1				1				1			
F			2		1			1				1	3		2		1			1
I							1		3		1		4				11	1		1
L	1	1	1	1	1		3	1	1	1	3	4		1	1		8	2	2	1
M						1		1					1				2			
O	1		1	1	1		1				2		1							
W	1				1															1
V	5	4	3	1	2	2	9	3	5	1	1	11	8	2						
R				2		2						1	2						4	2
K	3	4	3	1		1	2	3	1				2					4		1
H	1	1	2	2	1		2	1	2		1	1	1			1		2	1	

EXPECTED ALLELE FREQUENCIES

D	5.43	5.43	4.42	2.53	2.22	2.22	7.21	5.54	3.67	0.37	1.14	2.12	3.25	0.47	0.75	0.28	5.89	1.23	2.42	1.82
E	5.43	5.43	4.42	2.53	2.22	2.22	7.21	5.54	3.67	0.37	1.14	2.12	3.25	0.47	0.75	0.28	5.89	1.23	2.42	1.82
G	4.42	4.42	3.60	2.06	1.81	1.81	5.87	4.52	2.99	0.30	0.93	1.73	2.65	0.38	0.61	0.23	4.80	1.01	1.97	1.48
N	2.53	2.53	2.06	1.17	1.03	1.03	3.35	2.58	1.71	0.17	0.53	0.99	1.51	0.22	0.35	0.13	2.74	0.57	1.13	0.85
P	2.22	2.22	1.81	1.03	0.91	0.91	2.95	2.27	1.50	0.15	0.47	0.87	1.33	0.19	0.31	0.11	2.41	0.51	0.99	0.75
Q	2.22	2.22	1.81	1.03	0.91	0.91	2.95	2.27	1.50	0.15	0.47	0.87	1.33	0.19	0.31	0.11	2.41	0.51	0.99	0.75
A	7.21	7.21	5.87	3.35	2.95	2.95	9.57	7.36	4.87	0.50	1.51	2.82	4.31	0.62	1.00	0.37	7.82	1.64	3.22	2.42
S	5.54	5.54	4.52	2.58	2.27	2.27	7.36	5.66	3.75	0.38	1.16	2.17	3.32	0.48	0.77	0.29	6.02	1.26	2.47	1.86
T	3.67	3.67	2.99	1.71	1.50	1.50	4.87	3.75	2.48	0.25	0.77	1.43	2.19	0.32	0.51	0.19	3.98	0.83	1.64	1.23
C	0.37	0.37	0.30	0.17	0.15	0.15	0.50	0.36	0.25	0.03	0.08	0.15	0.22	0.03	0.05	0.02	0.41	0.09	0.17	0.13
F	1.14	1.14	0.93	0.53	0.47	0.47	1.51	1.16	0.77	0.08	0.24	0.44	0.68	0.10	0.16	0.06	1.23	0.26	0.51	0.38
I	2.12	2.12	1.73	0.99	0.87	0.87	2.82	2.17	1.43	0.15	0.44	0.83	1.27	0.18	0.29	0.11	2.30	0.48	0.95	0.71
L	3.25	3.25	2.65	1.51	1.33	1.33	4.31	3.32	2.17	0.22	0.68	1.27	1.94	0.28	0.45	0.17	3.52	0.74	1.45	1.09
M	0.47	0.47	0.38	0.22	0.19	0.19	0.62	0.46	0.32	0.03	0.10	0.18	0.28	0.04	0.06	0.02	0.51	0.11	0.21	0.16
O	0.75	0.75	0.61	0.35	0.31	0.31	1.00	0.77	0.51	0.05	0.16	0.29	0.45	0.06	0.10	0.04	0.82	0.17	0.34	0.25
W	0.28	0.28	0.23	0.13	0.11	0.11	0.37	0.29	0.19	0.02	0.06	0.11	0.17	0.02	0.04	0.01	0.30	0.06	0.13	0.09
V	5.89	5.89	4.80	2.74	2.41	2.41	7.82	6.02	3.98	0.41	1.23	2.30	3.52	0.51	0.82	0.30	6.39	1.34	2.63	1.98
R	1.23	1.23	1.01	0.57	0.51	0.51	1.64	1.26	0.83	0.09	0.26	0.48	0.74	0.11	0.17	0.06	1.34	0.28	0.55	0.41
K	2.42	2.42	1.97	1.13	0.99	0.99	3.22	2.47	1.64	0.17	0.51	0.95	1.45	0.21	0.34	0.13	2.63	0.55	1.08	0.81
H	1.82	1.82	1.48	0.85	0.75	0.75	2.42	1.86	1.23	0.13	0.38	0.71	1.09	0.16	0.25	0.09	1.98	0.41	0.81	0.61

D E G N P Q A S T C F I L M O W V R K H

58.43 58.43 47.60 27.17 23.91 23.91 77.57 59.66 39.48 4.03 12.24 22.83 34.94 5.04 8.10 3.01 63.39 13.28 26.08 19.61

GRAND TOTAL = 628.73

TOTALS OF ABSOLUTE VALUES OF CELLS

58.43 58.43 47.60 27.17 23.91 23.91 77.57 59.66 39.48 4.03 12.24 22.83 34.94 5.04 8.10 3.01 63.39 13.28 26.08 19.61

DEVIATIONS FROM EXPECTED

O	0.00	4.57	1.58	6.47	-1.22	-1.22	-2.21	0.46	-1.67	0.63	-1.14	-2.12	-2.25	-0.47	0.25	0.72	-0.89	-1.23	0.58	-0.82
E	4.57	0.00	-1.42	-2.53	1.78	8.78	0.79	0.46	-2.67	-0.37	-1.14	-2.12	-2.25	-0.47	-0.75	-0.28	-1.89	-1.23	1.58	-0.82
G	1.58	-1.42	0.00	-1.06	1.19	-1.81	2.13	2.48	1.01	-0.30	1.07	-1.73	-1.65	-0.38	0.39	-0.23	-1.80	-1.01	1.03	0.52
N	6.47	-2.53	-1.06	0.00	-1.03	-1.03	-0.35	1.42	-0.71	-0.17	-0.53	-0.99	-0.51	-0.22	0.65	-0.13	-1.74	1.43	-0.13	1.15
P	-1.22	1.78	1.19	-1.03	0.00	0.09	2.05	-0.27	-1.50	-0.15	0.53	-0.87	-0.33	-0.19	0.69	0.89	-0.41	-0.51	-0.99	0.25
Q	-1.22	8.78	-1.81	-1.03	0.09	0.00	-0.95	-1.27	-0.50	-0.15	-0.47	-0.87	-1.33	0.81	-0.31	-0.11	-0.41	1.49	0.01	-0.75
A	-2.21	0.79	2.13	-0.35	2.05	-0.95	-0.00	3.64	3.13	-0.50	-1.51	-1.82	-1.31	-0.62	0.00	-0.37	1.18	-1.64	-1.22	-0.42
S	0.46	0.46	2.48	1.42	-0.27	-1.27	3.64	0.00	3.25	-0.38	-0.16	-2.17	-2.32	0.52	-0.77	-0.29	-3.02	-1.26	0.53	-0.86
T	-1.67	-2.67	1.01	-0.71	-1.50	-0.50	3.13	3.25	0.00	0.75	-0.77	1.57	-1.19	-0.32	-0.51	-0.19	1.02	-0.83	-0.64	0.77
C	0.63	-0.37	-0.30	-0.17	-0.15	-0.15	-0.50	-0.38	0.75	0.00	-0.08	-0.15	0.78	-0.03	-0.05	-0.02	0.59	-0.09	-0.17	-0.13
F	-1.14	-1.14	1.07	-0.53	0.53	-0.47	-1.51	-0.16	-0.77	-0.08	0.00	0.56	2.32	-0.10	1.84	-0.06	-0.23	-0.26	-0.51	0.62
I	-2.12	-2.12	-1.73	-0.99	-0.87	-0.87	-1.82	-2.17	1.57	-0.15	0.56	0.00	2.73	-0.18	-0.29	-0.11	8.70	0.52	-0.95	0.29
L	-2.25	-2.25	-1.65	-0.51	-0.33	-1.33	-1.31	-2.32	-1.19	0.78	2.32	2.73	0.00	0.72	0.55	-0.17	4.48	1.26	0.55	-0.09
M	-0.47	-0.47	-0.38	-0.22	-0.19	0.81	-0.62	0.52	-0.32	-0.03	-0.10	-0.18	0.72	0.00	-0.06	-3.02	1.49	-0.11	-0.21	-0.16
O	0.25	-0.75	0.39	0.65	0.69	-0.31	0.00	-0.77	-0.51	-0.05	1.84	-0.29	0.55	-0.06	0.00	-0.04	-0.82	-0.17	-0.34	-0.25
W	0.72	-0.28	-0.23	-0.13	0.89	-0.11	-0.37	-0.29	-0.19	-0.02	-0.06	-0.11	-0.17	-0.02	-0.04	0.00	-0.30	-0.06	-0.13	0.91
V	-0.89	-1.89	-1.80	-1.74	-0.41	-0.41	1.18	-3.02	1.02	0.59	-0.23	8.70	4.48	1.49	-0.82	-0.30	-0.00	-1.34	-2.63	-1.98
R	-1.23	-1.23	-1.01	1.43	-0.51	1.49	-1.64	-1.26	-0.83	-0.09	-0.26	0.52	1.26	-0.11	-0.17	-0.06	-1.34	0.00	3.45	1.59
K	0.58	1.58	1.03	-0.13	-0.99	0.01	-1.22	0.53	-0.64	-0.17	-0.51	-0.95	0.55	-0.21	-0.34	-0.13	-2.63	3.45	0.00	0.19
H	-0.82	-0.82	0.52	1.15	0.25	-0.75	-0.42	-0.86	0.77	-0.13	0.62	0.29	-0.09	-0.16	-0.25	0.91	-1.98	1.59	0.19	0.00

D E G N P Q A S T C F I L M O W V R K H

-0.00 -0.00

GRAND TOTAL = -0.00

TOTALS OF ABSOLUTE VALUES OF CELLS

30.49 35.90 22.77 22.25 14.95 22.36 25.83 25.51 22.99 5.49 13.89 28.71 26.77 7.08 8.73 5.02 34.92 19.47 15.79 12.55