Reports

Dihydrofolate Reductase: X-ray Structure of the Binary Complex with Methotrexate

Abstract. *A central eight-stranded β-pleated sheet is the main feature of the polypeptide backbone folding in dihydrofolate reductase. The innermost four strands and two bridging helices are geometrically similar to but are connected in a different way from those in the dinucleotide binding domains found in nicotinamide-adenine dinucleotide–linked dehydrogenases. Methotrexate is bound in a 15-angstrom-deep cavity with the pteridine ring buried in a primarily hydrophobic pocket, although a strong interaction occurs between the side chain of aspartic acid 27 and $N_{(1)}$, $N_{(8)}$, and the 2-amino group of methotrexate.*

Dihydrofolate reductase (DHFR) catalyzes the NADPH-dependent reduction of dihydrofolate to tetrahydrofolate (*1*, *2*). The latter is utilized as a one-carbon carrier in several pathways of purine and pyrimidine biosynthesis, including that of thymidylate. Thus, a blockade of DHFR causes depletion of tetrahydrofolate, leading in turn to a deficiency of thymidylate. Since DNA synthesis is dependent on a continuing supply of thymidylate, such a deficiency can lead to cessation of growth in a rapidly proliferating cell line.

A variety of DHFR inhibitors are clinically useful as antibacterials, antiprotozoals, immunosuppressants, and antineoplastic agents (*3*). One important practical reason for interest in the three-dimensional structure of DHFR is that this group of enzymes exhibits considerable species-to-species variability in sensitivity toward different inhibitors (*4*, *5*). Thus knowledge of the three-dimensional structure of DHFR may provide an opportunity for rational design of specific chemotherapeutic agents.

Two widely used anticancer drugs, aminopterin and amethopterin [methotrexate (MTX)], very closely resemble folate and are extremely potent, essentially stoichiometric inhibitors of all DHFR's. In fact these compounds bind to the enzyme about 10^3 times more strongly than dihydrofolate itself. This is another reason for examining the three-dimensional structure of the DHFR-MTX complex.

A third reason for our interest in these enzymes is that they utilize NADPH instead of NADH. Although a great deal has been learned in recent years about the structures of several NADH-linked oxidoreductases (*6*), to our knowledge there have been no previous structural studies of NADPH-linked enzymes. It is generally accepted that the function of NADH-linked enzymes is energy production whereas NADPH-linked enzymes are biosynthetic, but it is not clear what role the 2′-phosphate group in NADPH may play in this distinction. We hope that a structural comparison of the two groups of pyridine nucleotide–linked oxidoreductases will shed light on this and the related question of whether they evolved independently or from a common ancestor.

Complete amino acid sequences have been determined for DHFR from the following: an MTX-resistant strain of *Escherichia coli* (*7*), a trimethoprim-resistant strain of *E. coli* (*8*), an MTX-resistant strain of *Streptococcus faecium* (*9*), and an MTX-resistant strain of *Lactobacillus casei* (*10*). The first 25 residues of a second, high-activity DHFR from *S. faecium* (*10*) and the first 31 residues of DHFR from bovine liver (*11*) have been sequenced as well.

Dihydrofolate reductase for this study was obtained as previously described (*12*) from the MTX-resistant strain of *E. coli*. Large crystals, about 0.8 mm on a side, of the binary MTX-enzyme complex were grown from 20 percent ethanol in water, buffered at *p*H 6.8 with 0.05*M* histidine-HCl. The complex crystallizes in space group $P6_1$, with unit cell parameters $a = b = 93.2$ Å and $c = 73.6$ Å, and two molecules in the asymmetric unit.

Two isomorphous heavy-atom derivatives were prepared by soaking selected crystals in $10^{-3}M$ solutions of sodium diuranate and of sodium ethylmercurithiosalicylate. X-ray diffraction intensities were obtained by using a newly developed multiwire area detector diffractometer (*13*). This technique permitted very rapid collection of data and made possible extensive replication to a resolution of 2.5 Å with only one crystal each of the parent and the heavy-atom derivatives. The parent crystal yielded 12,000 symmetry-independent reflections to this resolution. In a typical run 80,000 intensities were measured on a single crystal during a period of 5 days. Difference Patterson functions revealed a single site of substitution in the uranium derivative and five sites of varying relative occupancy in the mercury derivative.

Electron density maps were prepared both on a scale of 0.25 cm/Å for overall tracing of the chain and on the conventional scale of 2 cm/Å for model building. The two molecules within an asymmetric unit were built independently by different people. No obvious differences between the two molecules were observed.

As work progressed, some apparent discrepancies were observed between the experimental electron density and the primary sequence published by Bennett (*7*), particularly in the COOH-terminal half of the sequence. However, in these regions of disagreement our interpretation of the electron density is more consistent with a sequence recently published by Stone *et al.* (*8*) for the DHFR from a trimethoprim-resistant *E. coli* strain. Additional studies by Bennett (*14*) largely support our interpretation, although at present two possible sequence differences between DHFR's from the MTX- and the trimethoprim-resistant strains appear to remain unresolved. Amino acid identification and sequence numbering in this report follow those in Stone *et al.* (*8*). We believe that the DHFR's from these two sources may have the same 159-residue sequence and that the mutations imparting drug resistance probably cause only an increase in

Scoreboard for Reports. We have an uncomfortably large backlog of accepted Reports that await publication. For the past several months we have accepted about 17 Reports per week, a little more than 25 percent of those submitted. In order to reduce the backlog and shorten the publication delay, we will accept only 12 papers per week for the next few months.

the rates of DHFR synthesis and not a change in the amino acid sequence of the enzyme.

The overall folding of the polypeptide backbone (Fig. 1) is dominated by an eight-stranded β sheet beginning at the amino terminus and ending with a single antiparallel strand at the carboxy terminus. Thirty percent of the backbone chain is involved in this piece of secondary structure. The sheet shows the usual left-handed twist from one strand to the next (*15*), amounting to 130° between the two extremes. The molecule contains three helical regions, which together incorporate about 18 percent of the backbone chain.

The MTX molecule is bound in a cavity which is 15 Å deep and cuts across one whole face of the enzyme. Residues involved in this binding are not limited solely to the first half of the sequence, as has been suggested by Blakley and coworkers (*9, 16, 17*).

At least 13 amino acid residues are involved in binding MTX to DHFR. The drug is held in an open conformation with its pteridine ring nearly perpendicular to the aromatic ring of its *p*-aminobenzoyl glutamate group. The MTX molecule is draped over α helix αB with the pyrimidine end deeply buried in a primarily hydrophobic pocket formed by side chains from helix αB and from the amino terminal β strand βA. Side chains of Ile-5, Ala-7, Leu-28, Phe-31, and Ile-94 are in van der Waals contact with the pteridine ring. In addition, the plane of the peptide bond between Ala-6 and Ala-7 is nearly parallel to the plane of the pteridine ring, and about 3.5 Å distant, suggesting the existence of a pi-pi interaction with $N_{(1)}$, $C_{(2)}$ and its attached amino group, and $N_{(3)}$. Threonine 113 is so situated that its side chain hydroxyl can accept a hydrogen bond from the 2-amino substituent. The 4-amino group donates a hydrogen bond to the carbonyl oxygen of Ile-5. The face of the pyrazine ring not in contact with αB is completely exposed to solvent, permitting close approach in the ternary complex to a bound NADPH molecule, as required for reduction of dihydrofolate at $N_{(5)}$-$C_{(6)}$ (*18, 19*).

The side chain carboxyl of Asp-27 appears to play a very important role in the binding of MTX and presumably, by analogy, dihydrofolate. The $N_{(1)}$ of MTX is approximately equally distant (2.7 Å) from both Oδ1 and Oδ2 of Asp-27, which in turn are about 3.3 Å from the 2-amino group and $N_{(8)}$, respectively. As further discussed below, we believe that $N_{(1)}$ is protonated in the MTX-enzyme complex and that this protonation must account for much of the enzyme's enhanced affinity for the drug as compared with the substrate. Moreover, sequence alignments suggested by Freisheim (*10*) show that Asp-27 is conserved.

The aromatic ring of the *p*-aminobenzoyl portion of MTX resides in a second hydrophobic pocket formed by the side chains of Leu-28, Ile-50, Leu-54, and Ile-94. The β carbon of Ser-49 is in van der Waals contact with the methyl group at $N_{(10)}$ in MTX. The glutamate portion of the inhibitor is bound at the enzyme surface, with the side chain of Arg-57 hydrogen-bonded to the α-carboxyl group. The γ-carboxyl group of the glutamate in one of the two DHFR molecules in the asymmetric unit is hydrogen-bonded to an exterior solvent molecule. In the other complex within the asymmetric unit, the corresponding γ-carboxyl group may be hydrogen-bonded to the side chain of Lys-32. Sequence comparisons with DHFR from *L. casei* and *S. faecium* suggest that lysine and arginine residues can occur interchangeably at these two positions.

Methotrexate binds very tightly to all known DHFR's. The inhibitor dissociation constant K_I, for the *E. coli* enzyme is $2.4 \times 10^{-9} M$, about three orders of magnitude smaller than K_I for folate or than the substrate dissociation constant, K_S, for dihydrofolate (*12, 20*). There has been considerable discussion concerning the reason for this great affinity of DHFR for MTX. The only differences between folate and MTX are that the hydroxyl group at $C_{(4)}$ in folate is replaced by an amino group in MTX, and $N_{(10)}$ is methylated. It is important to bear in mind, however, that the most stable tautomer for a 4-hydroxypteridine has a double-bonded oxygen at $C_{(4)}$ and a hydrogen atom attached instead to $N_{(3)}$ (*21*). Thus, unlike a 2,4-diaminopteridine such as MTX, a 2-amino-4-hydroxy derivative such as folate cannot donate a hydrogen bond to the carbonyl oxygen of Ile-5. Nevertheless there is a serious difficulty in supposing that this offers an adequate explanation for the greater binding affinity of MTX. Namely, we find that the backbone amide of Gly-95 is only 3.2 Å from the 4-amino group in MTX and is positioned in such a way that it could donate a hydrogen bond to the 4-keto group of a similarly bound folate molecule. We therefore tentatively conclude that hy-

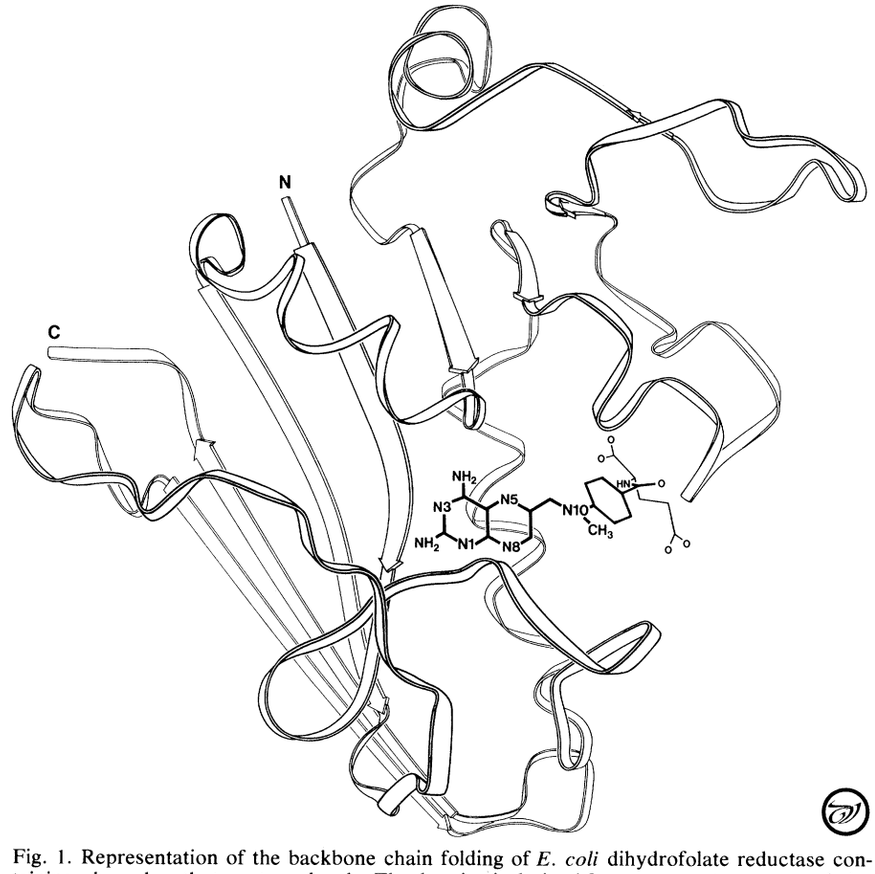

Fig. 1. Representation of the backbone chain folding of *E. coli* dihydrofolate reductase containing a bound methotrexate molecule. The drawing is derived from a computer-generated plot of all atoms in the drug and all α-carbon atoms of the enzyme. Strands of the central pleated sheet are shown as heavy arrows.

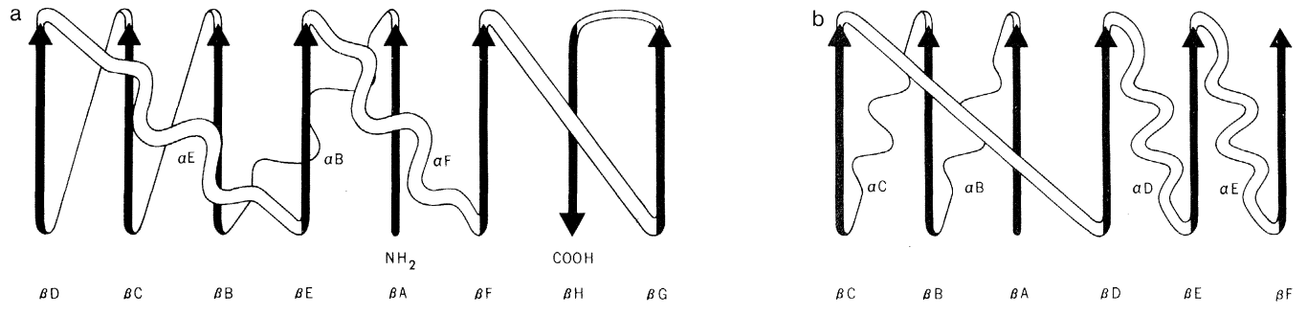

Fig. 2. (a) Topological diagram showing how individual strands are connected to form the central β-pleated sheet in DHFR. No attempt is made to describe the lengths of the individual strands or the conformation of the connections, except that connections involving α helices are represented by wavy lines. Connections drawn with thin lines are below and those drawn with thick lines are above the sheet. (b) Topological diagram for the NAD^+ binding region in dehydrogenases; adapted from Richardson (*26*).

drogen bonding of the 4-amino substituent by itself would not enhance the binding of MTX relative to folate or dihydrofolate.

On the other hand, substitution of a hydroxyl group by an amino group at $C_{(4)}$ does increase the basicity of the pteridine ring by about three *pK* units, and it is thought that protonation of the ring occurs most readily at $N_{(1)}$ (*21, 22*), which, as mentioned above, interacts strongly with the side chain of Asp-27. Moreover, it is noteworthy that Asp-27 is located on an inside surface of the MTX binding cleft, and thus inhibitor binding occludes water from the area surrounding the aspartate side chain. Consequently, a negatively charged carboxylate group so situated will stabilize the protonated form of the pteridine ring and, thus, enhance binding of MTX relative to folate or dihydrofolate. This interpretation is consistent with ultraviolet spectroscopic evidence showing enhanced protonation of MTX on binding to the enzyme (*23, 24*). Clearly, however, Asp-27 is not present for the purpose of binding MTX. Both its position in relation to the bound inhibitor and its conservation in four known DHFR sequences strongly suggest that Asp-27 must play some role in the catalytic mechanism.

The hydrophobic interaction between the methyl group on $N_{(10)}$ of MTX and Cβ of Ser-49 is apparently of marginal importance since aminopterin, which lacks this methyl group, is an equally potent inhibitor of DHFR's.

Figure 2a is a topological diagram showing the connectivity between strands which together make up the large central β sheet in DHFR. Note that all crossover connections are right-handed in the sense defined by Richardson (*25*). The longest stretch of helix, αB, includes residues 26 to 35; helix αE runs from 77 to 85; and helix αF from 97 to 105. The first six strands of the eight-stranded β-pleated sheet are all parallel and are buried within the molecule. Individual strands include the following residue numbers: 2 to 8 (A), 39 to 43 (B), 57 to 62 (C), 73 to 75 (D), 92 to 95 (E), 108 to 115 (F), 136 to 143 (G), and 149 to 156 (H).

Richardson (*26*) and Levitt and Chothia (*27*) have published connectivity diagrams similar to Fig. 2a for all β sheets known to occur in protein structures. The β-sheet topology found for DHFR is unlike any of these. However, in phosphoglycerate mutase (*28*) the first four parallel strands in a six-stranded mixed β sheet are topologically similar to the first five parallel strands in DHFR if the fifth and shortest strand in DHFR is considered to be an excursion. This is most probably a coincidence due to the requirements of efficient protein folding (*26, 27*).

The NAD^+-linked dehydrogenases represent a class of dinucleotide binding proteins which has been extensively studied by x-ray crystallographic techniques. High-resolution structures are now available for lactate dehydrogenase (LDH) (*29*), liver alcohol dehydrogenase (ADH) (*30*), soluble malate dehydrogenase (*31*), and lobster glyceraldehyde phosphate dehydrogenase (*32*). In each case the dinucleotide binding domain consists of a six-stranded parallel sheet and four connecting α helices, two on each side of the sheet (Fig. 2b).

Differences in relative positions of the various structural elements among the four dehydrogenases led Rossmann *et al.* (*6*) to conclude that the four central strands and two connecting helices have been more strongly conserved during the course of evolution than have been the extremities of the structure. Furthermore, Rao and Rossmann (*33*) showed that the dinucleotide binding domain consists of a pair of similar mononucleotide binding domains related by a twofold rotation about an axis parallel to and lying between βA and βD. Similarly folded tertiary structures are found in a number of other proteins, which led Rossmann *et al.* (*6*) to propose that the $\beta\alpha\beta$ structure may be a mononucleotide binding unit of widespread occurrence and common evolutionary origin.

In DHFR there are also two helices, αB and αF, with their axes running almost parallel to the nearby β strands, and there is also an approximate twofold axis parallel to and lying between the two innermost β strands, βA and βE (Fig. 2a). It is most important to note, however, that the helices do not connect β strands that are adjacent but rather those that are separated by one intervening strand.

Although topological connectivities are thus seen to be totally different when comparing the NAD^+-linked dehydrogenase with DHFR, we have found a remarkable similarity in the placement of the two connecting α helices relative to the parallel sheet. This similarity was discovered initially by simultaneously displaying the backbone chains of both LDH and DHFR on an Evans and Sutherland picture system and attempting to superimpose their respective central four-stranded parallel sheets. When this was done the two connecting α helices in DHFR were found to coincide with the corresponding helices in LDH almost as well as those in ADH coincide with those in LDH. However, the twist between the innermost strands βA and βE is larger in DHFR than in LDH because of steric interactions with the pteridine ring bound at the carboxy end of the four-stranded sheet by amino acids from βA, βE, and αB. This effect does not occur in the dehydrogenases because they have separate substrate binding domains.

Adenosine 5′-monophosphate binds in the same way to all four known dehydrogenase structures (*6*), and in each case the last residue in βA is glycine. It has been pointed out by Rossmann *et al.* (*6*) that glycine must be conserved at this position to avoid unfavorable steric in-

teraction with the ribose ring of the nucleotide. When the $(\beta\alpha\beta)_2$ fold in DHFR is rotated onto the $(\beta\alpha\beta)_2$ fold in LDH, Gly-96 in DHFR lies within 3 Å of the conserved glycine residue. A second conserved residue among all NAD^+-dependent dehydrogenases is an aspartate that forms a hydrogen bond with the 2′-hydroxyl of the ribose. The superposition described above places Arg-44 of DHFR within 3 Å of the conserved aspartate residue. This is precisely what would be expected for interaction between the enzyme and the 2′-phosphate in NADPH. Moreover, sequence alignments suggested by Freisheim (*10*) show that the glycine and arginine residues under consideration are conserved in the other two DHFR's for which sequence information is complete.

A comparison of the geometries of LDH and DHFR shows that the spatial arrangement, but not the connectivity, of the $(\beta\alpha\beta)_2$ structure is nearly the same for the two enzymes, and that this $(\beta\alpha\beta)_2$ structure is also the nucleotide binding site in DHFR. Thus, DHFR cannot have evolved from the common ancestral $\beta\alpha\beta$ mononucleotide-binding protein postulated by Rossmann *et al.* (*6*). We suggest that a better explanation for the existence of a pair of $\beta\alpha\beta$ structures in DHFR which are geometrically similar to those in LDH but have different connectivities is that such a structure represents a preferred and perhaps unique way of binding mononucleotides which has been arrived at in these two cases by convergent evolution.

D. A. MATTHEWS, R. A. ALDEN
J. T. BOLIN, S. T. FREER
Department of Chemistry, University of California, San Diego, La Jolla 92093
R. HAMLIN
Department of Physics, University of California, San Diego
N. XUONG
Departments of Chemistry, Biology, and Physics, University of California, San Diego
J. KRAUT
Department of Chemistry, University of California, San Diego
M. POE, M. WILLIAMS
K. HOOGSTEEN
Merck Institute for Therapeutic Research, Rahway, New Jersey 07065

References and Notes

1. Abbreviations used in this report: NADPH, reduced nicotinamide-adenine dinucleotide phosphate; NADH, reduced nicotinamide-adenine dinucleotide; Ile, isoleucine; Ala, alanine; Leu, leucine; Phe, phenylalanine; Asp, aspartic acid; Arg, arginine; Lys, lysine; Gly, glycine; and Ser, serine.
2. M. Osborn and F. M. Huennekens, *J. Biol. Chem.* **233**, 969 (1958).
3. J. Bertino and D. Johns, in *Cancer Chemotherapy*, I. Brodsky, Ed. (Grune & Stratton, New York, 1972), vol. 2, pp. 9–22.
4. B. R. Baker, *Design of Active-Site Directed Irreversible Enzyme Inhibitors* (Wiley, New York, 1967), pp. 192–266.
5. J. J. Burchall, *J. Infect. Dis.* **128**, S437 (1973).
6. M. G. Rossmann, A. Liljas, C. I. Brändén, L. J. Banaszak, in *The Enzymes*, P. D. Boyer, Ed. (Academic Press, New York, ed. 3, 1975), pp. 61–102.
7. C. D. Bennett, *Nature (London)* **248**, 67 (1974).
8. D. Stone, A. W. Phillips, J. J. Burchall, *Eur. J. Biochem.* **72**, 613 (1977).
9. J. M. Gleisner, D. L. Peterson, R. L. Blakley, *Proc. Natl. Acad. Sci. U.S.A.* **71**, 3001 (1974).
10. J. Freisheim, private communication.
11. R. Blakley, private communication.
12. M. Poe, N. J. Greenfield, J. M. Hirshfield, M. N. Williams, K. Hoogsteen, *Biochemistry* **11**, 1023 (1972).
13. C. Cork, R. Hamlin, W. Vernon, N. H. Xuong, V. Perez-Mendez, *Acta Crystallogr. Sect. A* **31**, 702 (1975).
14. C. D. Bennett, private communication.
15. C. Chothia, *J. Mol. Biol.* **75**, 295 (1973).
16. D. L. Peterson, J. M. Gleisner, R. L. Blakley, *J. Biol. Chem.* **250**, 4945 (1975).
17. J. M. Gleisner and R. L. Blakley, *ibid.*, p. 1580.
18. E. J. Pastore, M. Friedkin, O. Jardetzky, *J. Am. Chem. Soc.* **85**, 3058 (1963).
19. E. Pastore, *Abstr. Pap. 154th Meet. ACS* (1967), abstr. 107.
20. M. N. Williams, M. Poe, N. J. Greenfield, J. M. Hirshfield, K. Hoogsteen, *J. Biol. Chem.* **248**, 6375 (1973).
21. R. L. Blakley, *The Biochemistry of Folic Acid and Related Pteridines* (North-Holland, Amsterdam, 1969), pp. 58–105.
22. K. L. Kapoor, *Int. J. Quant. Chem.* **7**, 27 (1973).
23. J. S. Erickson and C. K. Mathews, *J. Biol. Chem.* **247**, 5661 (1972).
24. M. Poe, N. J. Greenfield, J. M. Hirshfield, K. Hoogsteen, *Cancer Biochem. Biophys.* **1**, 7 (1974).
25. J. S. Richardson, *Proc. Natl. Acad. Sci. U.S.A.* **73**, 2619 (1976).
26. _____, private communication.
27. M. Levitt and C. Chothia, *Nature (London)* **261**, 552 (1976).
28. J. W. Campbell, H. C. Watson, G. I. Hodgson, *ibid.* **250**, 301 (1974).
29. M. J. Adams *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **70**, 1968 (1973).
30. H. Eklund, B. Nordström, E. Zeppezauer, G. Söderlund, I. Ohlsson, T. Boiwe, C. I. Brändén, *FEBS Lett.* **44**, 200 (1974).
31. L. E. Webb, E. J. Hill, L. J. Banaszak, *Biochemistry* **12**, 5101 (1973).
32. M. Buehner, G. C. Ford, D. Moras, K. W. Olsen, M. G. Rossmann, *J. Mol. Biol.* **90**, 25 (1974).
33. S. T. Rao and M. G. Rossmann, *ibid.* **76**, 241 (1973).
34. Supported by grants from the National Institutes of Health (GM 20102, CA 17374, GM 10928, and RR 00757), the National Science Foundation (BMS 74-18676 and BMS 71-01301), and the American Cancer Society (BC 182) and by research fellowships (to D.A.M.) from the Jane Coffin Childs Memorial Fund for Medical Research and the National Cancer Institute.

22 February 1977; revised 5 April 1977

Possible Surface Reactions on Mars: Implications for Viking Biology Results

Abstract. *The results of two of the three biology experiments carried out on the Viking Mars landers have been simulated. The mixture of organic compounds labeled with carbon-14 used on Mars released carbon dioxide containing carbon-14 when reacted with a simulated martian surface and atmosphere exposed to ultraviolet light (labeled release experiment). Oxygen was released when metal peroxides or superoxides were treated with water (gas exchange experiment). The simulations suggest that the results of these two Viking experiments can be explained on the basis of reactions of the martian surface and atmosphere.*

One of the principal objectives of the Viking mission was the search for life on Mars. The biological investigation designed to detect microorganisms in the martian soil—the labeled release (LR), the gas exchange (GEX), and the pyrolytic release (PR) experiments—provided positive responses (*1*, *2*). On the other hand, the analysis of martian soil by pyrolysis gas chromatography–mass spectrometry showed no organic compounds at concentrations as low as 10 parts per billion (*3*) at either Viking landing site (*4*). Although such a low concentration of organic carbon compounds is not, in itself, synonymous with the absence of life, this apparently anomalous observation led us to look for possible nonbiological explanations for the results of the biology experiments. The findings of our simulation studies may be relevant to the interpretation of the data from the Viking mission.

In the Viking LR experiment, when an equimolar mixture (total, $17.5 \times 10^{-4}M$) of glycine, DL-alanine, DL-sodium lactate, sodium formate, and calcium glycolate uniformly labeled with ^{14}C and adjusted to *p*H 7 with KOH was injected onto the surface of 0.5 ml of martian soil in the test chamber, $^{14}CO_2$ was released (*1*, *2*). In attempting to simulate this result in the laboratory by purely chemical means, we searched for possible conditions for the release of CO_2. In one procedure 3 ml of 30 percent H_2O_2 was placed in a glass tube, and 2 ml of $0.1M$ formic acid, adjusted to *p*H 7 with KOH, was introduced through a stopcock; the formic acid solution was degassed before being introduced into the glass tube. Upon mixing these two solutions, CO_2, at a concentration significantly above background, was detected by mass spectrometry. However, CO_2 release was not detected as a result of the interaction of the same proportions of 30 percent H_2O_2 and $2 \times 10^{-4}M$ formic acid, probably because of the low sensitivity of the mass spectrometer.