to the two rings is 6.8 Å (*18*). The layer expansion of the MoS_2-ferrocene system is 5.6 Å (Table 1). Hence, the ferrocene molecules lie between the layers of MoS_2 with the planes of the cyclopentadienyl rings most probably perpendicular to the basal planes. In heptane and other higher alkanes, the van der Waals diameter perpendicular to the molecular axis is about 4 Å and the axial dimension is larger than 11.1 Å (*23*). The increase in c of systems of heptane (or higher alkanes, $n = 7$ to 12) with MoS_2 is 3.9 Å, so the molecular axis is probably parallel to the basal planes of the MoS_2.

In the new materials we have described, the included organic molecules need not be electron donors; thus a large class of exotic systems of transition-metal dichalcogenides containing polymers, organic dyes, liquid crystals, or organic semiconductors can be realized. It has recently been suggested (*24*) that the best candidates for new kinds of high-temperature superconductors are organic superconductors and inorganic layered compounds, particularly intercalated ones. In this context, this new class of restacked materials may provide an interesting combination of layered materials and organic superconductors.

REFERENCES AND NOTES

1. F. R. Gamble *et al.*, *Science* **174**, 493 (1971).
2. A. J. Jacobson, in *Intercalation Chemistry*, W. S. Whittingham and A. J. Jacobson, Eds. (Academic Press, New York, 1982), chap. 7.
3. R. H. Friend and A. D. Yoffe, *Adv. Phys.* **36**, 1 (1987).
4. D. W. Murphy and G. W. Hull, *J. Chem. Phys.* **62**, 973 (1975).
5. C. Liu *et al.*, *Thin Solid Films* **113**, 165 (1984).
6. A. Lerf and R. Schöllhorn, *Inorg. Chem.* **16**, 2950 (1977).
7. L. F. Nazar and A. J. Jacobson, *Chem. Commun.* **26**, 570 (1986).
8. R. Clement *et al.*, *Inorg. Chem.* **25**, 1404 (1986).
9. M. B. Dines, *Mater. Res. Bull.* **10**, 287 (1975).
10. P. Joensen *et al.*, *ibid.* **21**, 457 (1986).
11. R. Schöllhorn and A. Weiss, *J. Less-Common Met.* **36**, 229 (1974).
12. W. M. R. Divigalpitiya, S. R. Morrison, R. F. Frindt, *Thin Solid Films*, in press.
13. C. B. Roxlo *et al.*, *J. Catal.* **100**, 176 (1986).
14. C. B. Roxlo, H. W. Deckman, J. Gland, S. D. Cameron, R. R. Chianelli, *Science* **235**, 1629 (1987).
15. L. E. Scriven and C. V. Sterling, *Nature* **187**, 186 (1960).
16. G. L. Gaines, *Insoluble Monolayers at Liquid-Gas Interfaces* (Interscience, New York, 1966).
17. W. M. R. Divigalpitiya, S. R. Morrison, R. F. Frindt, U.S. Patent Application No. 07/297,464.
18. M. B. Dines, *Science* **188**, 1210 (1975).
19. W. B. Davies, M. L. H. Green, A. J. Jacobson, *Chem. Commun.* **19**, 781 (1976).
20. R. P. Clement *et al.*, *Inorg. Chem.* **17**, 2754 (1978).
21. T. J. Pinnavaia, *Science* **220**, 365 (1983).
22. S. A. Solin, *J. Mol. Catal.* **27**, 293 (1984).
23. L. C. Pauling, *Nature of the Chemical Bond* (Cornell Univ. Press, Ithaca, NY, 1960).
24. V. L. Ginzburg, *Phys. Today* **42**, 9 (March 1989).

25. Supported under a grant from 3M Canada. We thank P. Joensen for x-ray rocking curve measurements and O. Rajora for the scanning electron microscope analysis.

2 June 1989; accepted 24 August 1989

Toward Protein Tertiary Structure Recognition by Means of Associative Memory Hamiltonians

MARK S. FRIEDRICHS AND PETER G. WOLYNES

The statistical mechanics of associative memories and spin glasses suggests ways to design Hamiltonians for protein folding. An associative memory Hamiltonian based on hydrophobicity patterns is shown to have a large capacity for recall and to be capable of recognizing tertiary structure for moderately variant sequences.

HOW THE SEQUENCE OF A PROTEIN determines its three-dimensional structure is a major unsolved question at the nexus of condensed matter physics, chemistry, and biology. It is widely (but not universally) believed that protein structure can be found by minimizing the free energy of the protein chain; numerous models using a reduced description of polymer conformation have been devised for estimating this free energy (*1*). These models give insight into the essential physics, but a systematic approach seems necessary to solve the structure prediction problem. One way to develop simplified models is to use inverse statistical mechanics: a great deal of information on protein structure is available, so that we ask, "Is there a way to infer a Hamiltonian that will give the known structures?"

The theory of associative memories as developed by Hopfield and others (*2, 3*) is an application of inverse statistical mechanics that draws upon analogies with spin systems and spin glasses. Many aspects of protein thermodynamics and dynamics have also been related to spin systems and their phase transitions (*4–7*). The oldest of these connections is the description of secondary structure formation that makes use of one-dimensional Ising models for helix-coil transitions (*5*). Researchers have used the direct connection to spin models to develop secondary structure prediction schemes (*8, 9*) based on associative memories. It is clear, however, that secondary structure is determined in part by influences distant in protein sequence and, therefore, it is appropriate to address the question of tertiary structure directly. Abstract statistical mechanical models have already been used to investigate aspects of tertiary structure formation with notions from spin glasses and associative memories (*7*). These studies have, however, not presented structural inference schemes. In this report, we propose a simple version of an associative memory Hamiltonian appropriate for predicting tertiary structure; the framework is easily extended. The applicability of the Hamiltonian to the recall of structures from a given database, and, to a modest extent, its ability to generalize from variant sequences are assessed.

We describe tertiary structure by the α-carbon (Cα) coordinates of the protein. Sequence information is encoded in "charges," which are measures of hydrophobicity, size, electrical charge, and other amino acid properties. Folding is the arrangement of these charges into particular patterns. Our viewpoint is that folding is essentially crystallization in a finite system, thereby suggesting an analogy between proteins and associative memory spin systems that undergo phase transitions.

In spin associative memories, the interactions are given by the spin correlation function over the memory set (*2*). Similarly, the interaction potential for protein residues can be chosen as the "charge" density correlation function over the database. Thus the Hamiltonian in terms of the Cα coordinates is

$$H = -\sum_{i \neq j} \sum_{m,n} \sum_{\alpha} \lambda_{m,n}\, q_{m,i}\, q_{n,j}\, q^{\alpha}_{m,i}\, q^{\alpha}_{n,j}\, \theta_{sw}(r_{ij} - r^{\alpha}_{ij})$$

where $\lambda_{m,n}$ is the weight of terms containing charges of the type m and n, $q_{m,i}$ is the mth charge of the ith residue of the protein to be folded, $q^{\alpha}_{m,i}$ is the corresponding charge for the αth protein in the memory database, θ_{sw} is a square-well function centered about zero, r_{ij} is the distance between the Cα's of residues i and j, and r^{α}_{ij} is the analogous quantity for the memory proteins.

This Hamiltonian is a function only of scalar distances and is therefore translationally and rotationally invariant. With a single memory, it may be used to determine structures consistent with distance constraints derived from nuclear magnetic resonance (NMR). The Hamiltonian's interactions are quite long range and capture the nonlocality of folding. Mean-field theories based on this Hamiltonian should be reasonably accurate. They suggest that recall should be essentially perfect for small databases. If the database is uncorrelated, then as the number of memo-

School of Chemical Sciences and Beckman Institute, University of Illinois, Urbana, IL 61801.

Table 1. Summary of results.

Target protein	Protein memories (no.)	Nucleated residues (no.)	Energy as fraction of target energy	Energy-annealed x-ray structure as fraction of target energy	Root mean square (Å)	
					Non-nucleated portion	Full protein
C. pasteurianium rubredoxin	40	10	0.92	1.01	0.56	0.64
	40	20	0.56	1.02	2.12	1.77
	80	10	0.63	1.05	5.55	6.63
	80	20	0.65	1.05	2.65	2.29
BPTI	80	20	0.69	1.05	3.36	2.88
D. vulgaris rubredoxin	80	20	1.06	1.20	3.12	2.52

ries approaches a critical value proportional to the residue number (the "capacity"), recall will degrade, and a spin glass transition will result. For recall with few errors, the phase transition will be first order; hence after a critical nucleus has formed, the dynamics will proceed (mostly) downhill to give a minimum free energy structure. When a spin glass forms, the transition will resemble the Potts glass (*10*) or random energy model transition (*11*). In the spin glass phase, there are a number of nearly degenerate, largely unrelated low free energy structures. As this phase is approached, the dynamics slows; an evermore exhaustive search through the phase space is required as in the Levinthal paradox (*12*). For uncorrelated memories, this network's capacity should be larger than that of the Ising system. Correlations among the protein's charge patterns could also act to reinforce each other so that the spin glass phase would only occur for a very large number of memories. For a sufficiently large database, low-resolution folded structures for sequences not explicitly contained in the database should be obtainable, if the capacity is not exceeded.

Initially, we characterize the sequence by a binary-valued hydrophobic charge given by the sign in the Eisenberg consensus scale (*13*). This choice of charge is motivated by the observations that α helices and β sheets (*14*) have characteristic hydrophobicity patterns. In addition, the protein interior is dominated by hydrophobic residues, and reverse turns often contain hydrophilic clusters (*15*).

If one uses just the binary-valued hydrophobicity as a charge, there is an unwanted symmetry in H involving sign inversion of all charges. This symmetry is broken by using both the hydrophobic charge and a charge +1 for every residue, giving a Hamiltonian

$$H = -\sum_{i \neq j} \sum_{\alpha} (q_{h,i}\, q_{h,j}\, q^{\alpha}_{h,i}\, q^{\alpha}_{h,j} + q_{h,i}\, q^{\alpha}_{h,i} + q_{h,j}\, q^{\alpha}_{h,j} - \gamma)\theta_{sw}(r_{ij} - r^{\alpha}_{ij})$$

where $q_{h,i}$ $(= \pm 1)$ is the hydrophobic charge of residue i and γ is a small constant (≈ 0.01).

Our memory database was a subset of the one used by Qian and Sejnowski (*8*); the homologies within the set as measured by diagon plots (*16*) were minimal. The square well widths were set to $0.5|i - j - 1|$Å, where $|i - j|$ is the sequence distance, and the depths were set so that the short-range $(|i - j| \leq 10)$ and long-range interactions $(|i - j| > 10)$ contributed equally to H. To minimize H, a Monte Carlo–simulated annealing scheme was used (*17*). The chain conformation was represented in the virtual and torsional angle scheme of Srinivasan *et al.* (*18*) with fixed bond lengths. Each attempted move consisted of changing the angles associated with a randomly chosen residue; the new angles were randomly selected from a residue-dependent database. The use of this type of database is an importance sampling scheme that accelerates the minimization. Because folding is a first-order phase transition requiring nucleation, the calculations begin with the first 10 or 20 Cα's in their crystallographic positions. The next 20 residues are placed in a random coil configuration and annealed. Annealing schedules are fixed so that the probability of accepting a move giving a positive change in H decreases uniformly from 0.5 to 0. The remaining residues are then appended and annealed, again in cassettes of length 20 (or less for the last segment). In some cases the resulting structure is reannealed across the boundary separating the cassettes.

Our studies are primarily directed toward estimating the Hamiltonian's capacity. Three proteins were studied: two rubredoxins and bovine pancreatic trypsin inhibitor (BPTI). Table 1 shows the results for some of the lowest energy structures obtained for recall (the targeted protein included in the memory database) of *Clostridium pasteurianium* rubredoxin with 40 and 80 protein memories in the database and BPTI with 80 database proteins. The overlap between the x-ray and calculated structures for the 40 memory rubredoxin and BPTI is shown in Fig. 1, A and B, respectively, with the nucleating seed consisting of the first 20 residues.

The seed's role seems to be primarily kinetic. For the 40 memory rubredoxin calculations, the results for nuclei of lengths 10 and 20 are similar; the 10-residue nucleus calculation, however, required a much longer and slower annealing than was used for the 20-residue nucleus structure. In the 80 memory rubredoxin calculations, the structures obtained with the different seed lengths differ significantly in their rms (root-mean-square) values, although their energies are comparable. Since the energy for the ten-residue nucleus calculation is still only a fraction of the target energy, we believe a better understanding of the system's thermodynamic phase diagram would allow the

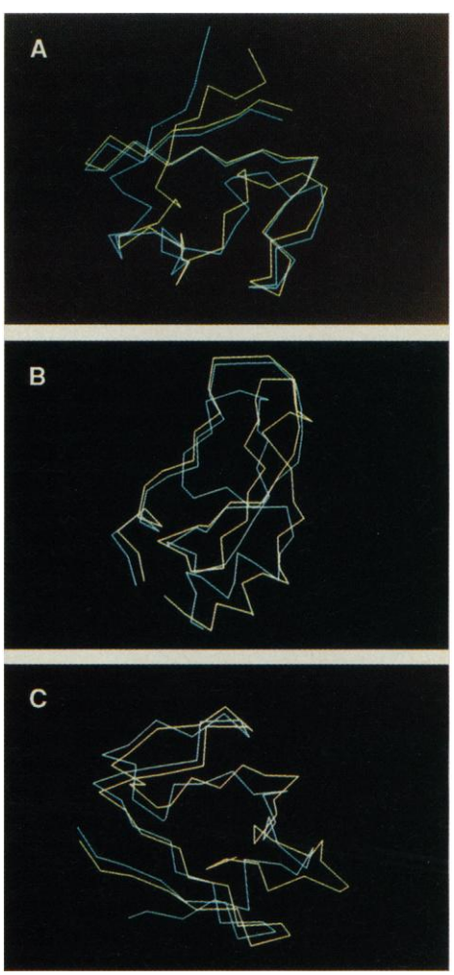

Fig. 1. (**A**) Overlap of crystallographic *Clostridium pasteurianium* rubredoxin (cyan) and structure calculated with 40 memories in the database (yellow). (**B**) Overlap of crystallographic BPTI (cyan) and structure calculated with 80 memories in the database (yellow). (**C**) Overlap of crystallographic *Desulfovibrio vulgaris* rubredoxin (cyan) and structure calculated with 80 memories (yellow) (including *Clostridium pasteurianium* rubredoxin only in the database).

design of a better annealing schedule and, hence, more accurate structures.

Evidently the capacity of these Hamiltonians per residue is significantly larger than it is for Ising neural networks per spin (*2, 3*). The energies of the final annealed structures are sufficiently close to those obtained by annealing the x-ray structure of the target protein that rms values less than 3.0 Å are secured for most calculations. The number of protein families being small—of order 20 to 40 (*19*)—one has hope that a cunningly chosen database would have sufficient capacity to classify proteins into these families.

The Hamiltonian can also recognize variant sequences as demonstrated in the last entry of Table 1 and in Fig. 1C. The *Desulfovibrio vulgaris* rubredoxin differs from the *Clostridium pasteurianium* form included in the database at 50% of the residues. Of these, six are not synonymous in terms of our simple hydrophobicity scale. The 2.5 Å rms value demonstrates the Hamiltonian is able to generalize to this degree of substitutional mutation.

The large capacity of this simple associative memory Hamiltonian and its modest ability to generalize with respect to site mutations suggest that this approach offers a fruitful perspective on tertiary structure recognition. As it stands, the associative memory approach should be considered as a framework (as opposed to a method) for predicting structures. The recall of structure is, however, comparable or better than earlier studies that used only hydrophobicity statistics (*20*), which give rms values of 4 to 8 Å, although this is a somewhat unfair comparison. Further features must be incorporated in a fully predictive associative-memory Hamiltonian. Structures which have been modified by insertions and deletions must also be recognized. This requires a consideration of the invariances of Hamiltonians to these sequence transformations. The role of vector charges, many-body interactions, and modifications of the interaction network, such as dilution or changing the range of the potentials, are also of interest.

REFERENCES AND NOTES

1. M. Levitt and A. Warshel, *Nature* **253**, 694 (1975); J. Skolnick, A. Kolinski, R. Yaris, *Proc. Natl. Acad. Sci. U.S.A.* **85**, 5057 (1987); N. Go, *Annu. Rev. Biophys. Bioeng.* **12**, 183 (1983); H. Wako and H. A. Sheraga, *J. Protein Chem.* **1**, 85 (1982).
2. J. J. Hopfield, *Proc. Natl. Acad. Sci. U.S.A.* **79**, 2554 (1982).
3. D. J. Amit, H. Gutfreund, H. Sompolinsky, *Phys. Rev. Lett.* **55**, 1530 (1985).
4. D. L. Stein, *Proc. Natl. Sci. Acad. U.S.A.* **82**, 3670 (1985).
5. D. Poland and H. A. Scheraga, *Theory of Helix-Coil Transitions in Biopolymers* (Academic Press, New York, 1970).
6. J. D. Bryngelson and P. G. Wolynes, *Proc. Natl. Sci. Acad.* **84**, 7524 (1987).
7. T. Garel and H. Orland, *Europhys. Lett.* **6**, 307 (1988); *ibid.*, p. 597; E. I. Shakhnovich and A. Gutin, *ibid.* **8**, 3271 (1988); *J. Phys. A* **22**, 1647 (1989).
8. N. Qian and T. J. Sejnowski, *J. Mol. Biol.* **202**, 865 (1988).
9. L. H. Holley and M. Karplus, *Proc. Natl. Sci. Acad. U.S.A.* **86**, 152 (1989).
10. T. R. Kirkpatrick and P. G. Wolynes, *Phys. Rev. B* **36**, 8552 (1987).
11. B. Derrida, *Phys. Rev. Lett.* **45**, 79 (1980).
12. J. D. Bryngelson and P. G. Wolynes, *J. Phys. Chem.*, in press; C. Levinthal, *J. Chim. Phys.* **65**, 44 (1968).
13. D. Eisenberg, R. M. Weiss, T. C. Terwilliger, W. Wilcox, *Faraday Symp. Chem. Soc.* **17**, 109 (1982).
14. D. Eisenberg, R. M. Weiss, T. C. Terwilliger, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 140 (1984); M. Shiffer and A. B. Edmundson, *Biophys. J.* **7**, 121 (1967); J. Palan and P. Puigdomenech, *J. Mol. Biol.* **88**, 457 (1974); V. I. Lim, *ibid.*, p. 857; *ibid.*, p. 873.
15. I. D. Kuntz, *J. Am. Chem. Soc.* **94**, 4009 (1972).
16. R. Staden, *Nucleic Acids Res.* **10**, 2951 (1982).
17. S. Kirkpatrick, C. D. Gelatt, Jr., M. P. Vecchi, *Science* **220**, 671 (1983).
18. R. Srinivasan *et al.*, *J. Mol. Biol.* **98**, 739 (1975).
19. J. S. Richardson, *Adv. Protein Chem.* **34**, 167 (1981).
20. M. Ycas, N. S. Goel, J. W. Jacobsen, *J. Theor. Biol.* **72**, 443 (1978); N. S. Goel and M. Ycas, *ibid.* 77, 253 (1979); I. D. Kuntz, G. M. Crippen, P. A. Kollman, *Biopolymers* **18**, 939 (1979).

21. We thank J. Bryngelson, H. Frauenfelder, J. Onuchic, K. Schulten, Z. Schulten, and J. Widom for helpful discussions, and D. Evensky and Z. Schulten for help with the graphics. Supported by NSF grant CHE 84-18619; the computations were done at the National Center for Supercomputer Applications, Urbana, Illinois.

9 June 1989; accepted 13 September 1989

Styles of Volcanism on Venus: New Arecibo High Resolution Radar Data

DONALD B. CAMPBELL, JAMES W. HEAD, ALICE A. HINE, JOHN K. HARMON, DAVID A. SENSKE, PAUL C. FISHER

Arecibo high-resolution (1.5 to 2 km) radar data of Venus for the area extending from Beta Regio to western Eisila Regio provide strong evidence that the mountains in Beta and Eisila Regiones and plains in and adjacent to Guinevere Planitia are of volcanic origin. Recognized styles of volcanism include large volcanic edifices on the Beta and Eisila rises related to regional structural trends, plains with multiple source vents and a mottled appearance due to the ponding of volcanic flows, and plains with bright features surrounded by extensive quasi-circular radar-dark halos. The high density of volcanic vents in the plains suggests that heat loss by abundant and widely distributed plains volcanism may be more significant than previously recognized. The low density of impact craters greater than 15 km in diameter in this region compared to the average density for the higher northern latitudes suggests that the plains have a younger age.

VOLCANISM IS ONE OF THE FUNDAMENTAL processes of heat transfer from planetary interiors (*1*). The location of volcanic deposits and edifices, their volumes, and their sequence provide evidence for quantitative assessments of heat transfer in space and time. The nature of volcanic deposits provides clues to the style of volcanism, which is related to composition, volatile content, interaction with the crust during ascent, and the structure of the crust and lithosphere (*2*). New data for about 32×10^6 km^2 of the equatorial region of Venus (7% of the surface area of the planet) (Fig. 1) provide higher quality images than previously available for this region because of significant improvement in sensitivity and increased resolution by a factor of 5 to 10 for more than 50% of the region. They provide information about the nature of volcanic deposits and permit comparison to other parts of Venus previously imaged at both high and low resolution. These new data show that volcanism is an extremely widespread process in this part of Venus and that volcanic deposits cover most of the surface area and occur in a variety of environments and styles.

Observations of Venus were made during the summer of 1988 with the 12.6-cm wavelength Arecibo radar facility, and data were obtained with resolutions between 1.5 and 2 km. A circularly polarized signal was transmitted, and both senses of received circular polarization were recorded. The equatorial region was viewed at incidence angles from about 12° to 60° (the extremes encompass only a small fraction of the area, and the incidence angle for most of the coverage was between 20° and 50°, similar to the range expected for the Magellan mission), and the signal-to-noise ratio decreased with increasing incidence angle because of both the

D. B. Campbell, National Astronomy and Ionosphere Center, Cornell University, Ithaca, NY 14853.
J. W. Head, D. A. Senske, P. C. Fisher, Department of Geological Sciences, Brown University, Providence, RI 02912.
A. A. Hine and J. K. Harmon, National Astronomy and Ionosphere Center, Arecibo Observatory, Arecibo, Puerto Rico 00612.