Apart from the great reduction of the volume of the Hidder Zone, the principal differences between these and previous estimates are in the greater lateral extent of the middle portion of the intrusion from Lower Zone B through the Middle Zone. A further complication arises from the recent discovery that the phase layering of the Layered Series transgresses structural and stratigraphic horizons. The sequence of appearance of mafic phases is displaced upward in the eastern part of the intrusion relative to a section on the west side so that cumulus clinopyroxene and magnetite appear at olivine depleats at levels that rise toward the east relative to distinctive horizons, such as the conspicuous Triple Group. The fact that the composition of plagioclase does not seem to be affected by this phenomenon suggests that the liquidus temperatures of iron-bearing minerals reflected a difference in oxygen fugacity from one side of the intrusion to the other: This feature was only recognized very recently and requires more study.

This work was supported by grants from the NSF, the National Geographic Society and the Geological Society of America. Experimental work was carried out at the University of Edinburgh and the Carnegie Institution of Washington.


Computer simulation of protein folding
Michael Levitt* & Arieh Warshel

Department of Chemical Physics, Weizmann Institute of Science, Rehovot, Israel

A new and very simple representation of protein conformation has been used together with energy minimization and thermalization to simulate protein folding. Under certain conditions, the method succeeds in "remaking" bovine pancreatic trypsin inhibitor from an open-chain conformation into a folded conformation close to that of the native molecule.

PROTEIN molecules owe their enormous functional versatility to the fact that they spontaneously fold into complicated and unique conformations determined by the particular amino-acid sequence. Discovering the relationship between protein sequence and conformation is a fascinating theoretical problem of fundamental importance. Most previous theoretical work has used the concept of "local structure," in which the conformation of a short segment of polypeptide chain is supposed to depend almost entirely on the sequence of that segment. Although this approach has helped understand local secondary structure, it has not shown how residues distant along the chain can come together to form the overall conformation. The only promising attempt to study the tertiary folding of a protein, in this case myoglobin, was based on the package of cylinder "supposed to represent a helix." The method was not implemented on a computer and could be applied more generally to other proteins not built entirely from helices. Here we develop such a conformation. First, we simply the representation of a protein by averaging over the details. This is done both to make the calculations much more efficient and also to avoid having to distinguish between many conformations that differ only in these finer details. Second, we simulate the folding of this simple structure by the combination of energy minimization and normal mode analysis, which accelerate the process by avoiding all random fluctuations that occur in nature. Tests of the procedure on bovine pancreatic trypsin inhibitor (BPTI), show that under certain conditions it can rapidly reproduce the correct overall folding of this small protein molecule.

Simple representation of protein structure

Even the smallest protein (say 50 residues) is extremely complex, with about 750 atoms and a dozen or more degrees of freedom. In particular, interactions involving the rapidly moving solvent molecules and the protein surface are not easily treated by the many-body methods used so far. Here a simple effective potential is constructed, using a quasi-chemical analogy, which can be used for large proteins.
Fig. 1. Relationship between the simplified models of protein structure introduced here and the full all-atom structure of proteins. The top reference point for each residue in the simplified model correspond to the centers of the side chains and the Cα. Each residue is only allowed one degree of freedom: the torsion angle ψ between the N–Cα and Cα–Cβ bonds of residue i, and δ (in the same simplified geometry). The bond lengths, bond angles, and various angles used to define the geometry of the simplified molecule were taken as the average values found in eight globular conformations, though they could just as well have been taken from amino-acid model compounds.

... internal moieties of parts of the protein itself. Our method is designed to overcome these problems and is based on two assumptions: (1) that much of the protein's fine structure can be eliminated by averaging, and (2) that the overall chain conformation can be obtained by considering only the most effective variables (those that vary most slowly and yet cause the greatest changes in conformation). Averaging over groups of atoms in the full structure gives a simplified structure for each residue represented by only two atoms, the Cα atom, and the centroid of the side chain. Interaction is assumed to occur only between side chains, while the Cα positions define the chain path (Fig. 1). Each amino-acid residue only has one degree of freedom, the torsion angle ψ about the bond joining two adjacent Cα atoms. Although a simple form of this potential has been used before to study polyglycine random coils, it has never been applied to ordered globular proteins. This simplification reduces the degrees of freedom by a factor of four and the number of interaction curves by a factor of fifteen. One might hope that the reduced space used here to describe different conformations will have many fewer energy minima. The absence of lower dimensionality and the side chains are smooth with all the minor bumps of the all-atom structures. The effect of the fine details and more rapidly changing interactions is included implicitly in the effective time-averaged potential actions used. (By the ergodic theorem, this time averaging is equivalent to Boltzmann-weighted space-averaging over deformations generated by changing the fast variables.) For reasons about the torsion angle ψ, the effective potential is generated by averaging the energy over all those conformations of a dipeptide that have a particular value of ψ. As it was possible to study all 400 different dipeptides, calculations were done on six considered most representative: ala-al, gln-ala, gln-gly, gln-ala and pro-ala. This showed that the effective potential only depends in the nature of the bond amino acid, giving different potentials for the α-pet- 

Fig. 2. Stereo ribbon drawings of P1 in: a, its idealized native conformation; b, the minimum energy conformation generated starting at the idealized native conformation and c, the best conformation generated by folding an extended chain with a terminal hairpin (final conformation). In (c), the program used to rotate the molecules into the same orientation and then drew the ribbon between Cα's were provided by Dr. A. D. McLoon.
sphere. This effective potential between identical side chains was approximated by a Lennard-Jones type function, and potentials between pairs of different side chains were obtained by a geometric mean combining law. Because proteins fold in water not a vacuum, interactions with the solvent are included by assigning to each side chain a hydrophobic energy taken from the solubilities of amino acids in water and in ethanol. In the calculations, the energy of transfer of these two solvents was taken as the difference in energy of the side chain when isolated in water and when completely surrounded by other residues. When surrounded by an intermediate number of neighbors, the hydrophobic energy was varied according to a sigmoid function. More complicated models that include hydrogen bonds and S-S bridges will be described elsewhere, together with full details of the standard geometry and all energy parameters used.

The folding of this idealized protein can be simulated by solving the equations of molecular dynamics at sufficiently small time intervals. In a viscous medium like water, these equations of motion can be approximated by Langevin equations, where the change in the variables is dictated down the energy gradient with a random deflection due to Brownian motion. For greater computational efficiency we neglect these thermal fluctuations while the chain folds, and the end point of the trajectory is the potential energy minimum accessible from the starting conformation. We minimize the energy of the idealized protein chain with respect to all the side chain angles using a powerful quadratically convergent method (YAO9A, by R. Fletcher and taken from the Harwell Subroutine Library). After reaching a minimum, the thermal fluctuations are reintroduced and the conformation is considered to be vibrating about the minimum so that each normal mode has average kinetic energy kT/2 (where k is the Boltzmann constant and T the absolute temperature). A new starting conformation for the next pass of energy minimization is chosen by randomly stopping the simulation. At this time each normal-mode coordinate will be displaced randomly from the minimum by (RkT/2)^1/2, so that the associated energy becomes RkT/2. Here k is the eigenvalue of the energy second derivative matrix (corresponding to the particular normal mode, and R0) is a random magnitude uniformly distributed from 0 to 1. (An exponential distribution of random numbers between 0 and 1 would be more realistic) Normal-mode thermalization avoids non-productive changes in conformation for it knows which combinations of angle changes should cause the greatest change in energy increase.

Testing the simplified representation

The drastic simplifications used in the present representation of a protein conformation were tested by minimization from near the native folded conformation. Boivin parameter inhibitor was chosen for this test as it is the only small protein (less than 100 residues) of known conformation that has a single polyproline chain and no additional prosthetic group. As a first step, a simplified native PTI conformation was obtained by taking the Cα positions and side-chain conformation from the X-ray coordinates14 kindly supplied by Drs. Hasegawa and Steigemann. Next an idealized chain, based on the PTI sequence and having the same geometry for all side chains of the same type, was made to fit the simplified native coordinates by adjusting the α and ψ angles. This conformation, known as the idealized native structure, deviates by 1.1 Å from the simplified native structure. The r.m.s. deviation is

\[
\left( \frac{1}{N} \sum \Delta r_i^2 \right)^{1/2}
\]

where \( \Delta r_i \) is the difference, in the two structures, of the distance between side-chain centroids (α and ψ). Energy minimization from this starting conformation was then carried out to reproduce the stability of native PTI. After 558 cycles a perfect minimum is reached at an energy of -50.9 kcal/mole and a r.m.s. deviation of only 3.5 Å from the simplified native conformation. Thermal randomization about this minimum does not lead to further movement from the native molecule on subsequent minimization. Randomly disturbing the initial best-fit angles with a fluctuation between +15° and +15° has little effect on the conformation obtained by subsequent minimization. Figure 2 compares the minimum energy and native chain folding. In this plot, the chain was helix in the plane of the paper and is shown with bold lines, the chain to which the first termini have been eliminated, the terminal helix becomes distorted and consequently peaks too tightly against the B barrel center at residue 27.

Simulation of PTI folding

Having shown that so simple a model can represent the conformation of a folded protein, we tried to simulate the actual process of folding. Most tests were done with two procedures: starting conformation extended, except for the terminal helix, on the stability of the structure used to guide formation. Figure 4 shows the results of the successful runs. The collapse of the conformation, not to bend back, once formed, is then reached in a time that could be a reasonable time for the two terminal helices in the N-terminal folded conformation molecule (Fig. 4). The two terminal helices in the conformation, namely, lie essentially normal to the 10 Å. It is
Fig. 4. Simulation of P11 folding from an extended starting conformation with the terminal helix (a = 180°) for all except 48 to 58 where a = 45°. No knowledge whatever about native P11 is used during this simulation (apart from setting the terminal helix). The conformation was thermostatted at the end of each minimization except near cycles 490 and 570 when the energy rises sufficiently because the minimization was restarted after rounding the torsion angles to one degree. In the first two thermostattions, each normal mode was perturbed in the plus direction to raise the associated energy to 477.2 with T = 1,000 K. In the other three thermostattions, the perturbations were randomly in the plus and minus directions but always such as to raise the energy by 477.2 with T = 100 K. (Because the random numbers are distributed uniformly rather than exponentially, these temperatures do not correspond to the macroscopic temperature.) The 80 membrane diagrams, which show the C-chain path, refer from left to right to the 8 conformational states at the circled points on the r.m.s. deviation curve, respectively. The last five conformational states correspond to the terminal helix, which is approximately lower energy and are each a little closer to the end points at 46.9, 46.6 and 46.2 kcal mol⁻¹, respectively; r.m.s. deviations 0.62, 0.63, 0.64, 0.65 and 0.66 kcal mol⁻¹, respectively. The solid dots at the ends of a minimization indicate that a perfect minimum was reached. One cycle simulates about 0.6 s on an IBM 370/155 computer.

Starting conformations, one fully extended (all a = 180°), and an extended apart from the C-terminal helix (a = 180°, except for residues 48 to 58 where a = 45°). Restringing the terminal helix from the native structure is justified here as we are more concerned with the process of folding than with the reduction of the native conformation of an unknown protein. In the latter case, a statistical rule (see ref. 2) could be used to guess the position of the a helices in the starting conformation. Figure 4 shows the iteration history of minimisation with the second of these starting points, which was the most successful of those obtained to date. Thermal randomisation of the first minimisation, an irregular but extended conformation, raises the energy, and in this case causes the chain to bend back on itself decreasing the r.m.s. deviation. From this point, minimisation first opens the molecule again, but then reaches a new minimum where the chain now has kinks that could become the bends of a hairpin. After a second minimisation, the molecule begins to fold, bringing the terminal helix close to the b hairpins on residue 11. More minimisation and thermostattion brings together two top loops (residues 15 and 40) and then brings the N-terminal tail on to the rest of the molecule. The final folded conformation of Fig. 4 is remarkably like the native conformation (Figs 2 and 3). In both conformations the chain loops back on itself near residues 14, 27 and 40. In both conformations, the pairs of b-sheet-like residues that are experimentally known to form 3-5 bridges, are close together (1.5 Å). It is interesting that the C-terminal helix is the part of the native molecule that reproduces least well even though these residues had been set to a perfect helix in the starting conformation; this is due to the omission of peptide peptide hydrogen bonds which stabilise the helix and could now be introduced.

Repeating the folding simulation from the fully extended starting conformation also leads to a compact structure with many of the features of native P11, although after the same number of cycles the r.m.s. fit was a little worse (7.7 Å instead of 6.5 Å) and the energy was higher (47.7 kcal mol⁻¹ instead of 44.9 kcal mol⁻¹). Almost all the differences in conformation of these two folded structures involved the last 10 residues which remained extended if not preset to a helix and consequently failed to pack against the rest of the molecule.

Variation of folding conditions

Changing either set of starting torsion angles by a random value between 15° and 15° had little effect on the final conformation. Folding at a lower initial temperature (T = 300 K, rather than T = 1,000 K) failed to reach a contact conformation, as the thermal disturbances were too small to get out of the local minima corresponding to an extended chain.

Four additional runs of 600 cycles, under the same conditions as those used in Fig. 4 but based on different sequences of random numbers, gave rise to different folded shapes. In one of these, the folded molecule was close to the native structure (r.m.s. deviation of 6.3 Å), although the b sheet was formed between parallel rather than antiparallel chains (Fig. 3).
two others, the antiparallel β sheet centred near residue 27 was formed, but this hairpin did not subsequently fold on itself to give a compact shape (Fig. 3d and e). Conformations that deviated more from the native structure always had higher energies, which gives an independent criterion for choosing the best conformation and its stability. Minimisation leads to a conformation closer to the native one. Of the five runs using different random numbers for the thermodynamic start, one succeeded in getting to within 6.3 Å of the simplified native structure in less than 600 steps. That certain folding patterns are less successful is consistent with the experimental results of Creighton et al. who have analysed the predominant kinetic intermediates present at different times after starting PTI renaturation and found several with the wrong tertiary fold.

General model for protein folding?

It seems remarkable that so simple a model based on time-averaged forces can account for the stability and folding of a molecule as complicated as a protein. Looking at known protein conformations closely, one is struck by the precise geometry of the interatomic contacts that stabilize the molecule: all possible interior hydrogen bonds are well formed, and many of the nonpolar side chains interlock to form a close-packed interior. As the forces responsible for this precise geometry fall off rapidly with distance and improper orientation, it would seem that folding must depend on a very rare random fluctuation that happened to bring the right residues close together with sufficient precision for the short-range forces to take effect. It therefore seems unlikely that these short-range forces could 'direct' the folding from an open disordered structure. In view of the known results, however, the time average of these short-range forces may play an important role in directing protein folding. These effective forces, which are weak, few likely, and not at all dependent on orientation, restrict the number of low energy conformations severely; they cause the chain to fold into the approximate shape rapidly and without having to pass through many local minima.

Because this approximately folded molecule corresponds to a large region in the space of possible protein conformations, folding would not be so rare an event.

As a summary of the conclusions we propose that initially, when the chain has a flexible open structure, the effective time-averaged forces between the residues play a central role, folding the chain into a compact shape with groups close to their final positions (say within 5 Å). Once the chain becomes compact with less freedom of movement, the specific short-range interatomic forces become important, they form a precise conformation provided that the result gain in enthalpy overcomes the loss in entropy. The process would be rather like crystallisation, with the atoms simply falling into place from their nearby positions in the approximate folded conformation.

To simulate this second step one axisces over to a progressively more detailed models gradually incorporating more atoms and ending with the all-atom structures considered earlier work. Although calculating the energy of the all-atom model would be too time consuming, one would have a great advantage of starting close to the right conformation and could minimise successive overlapping zones of a residue at a time without having to search through many local minima.

The general concept of using a simple model based on effective time-averaged forces when the detailed forces are complicated has many potential applications; for example, the formation of protein quaternary structure and macromolecular complexes, virus assembly and so on. At such a level of complexity, forces would be time averaged over those substructures that are relatively fixed or seem to play an important role in the assembly. Such a hierarchical approach might eventually lead to a understanding and simulation of very complicated biological assembly processes.

We thank the Weizmann Institute Computer Center facilities and the European Molecular Biology Organization for supporting one of us (M.L.).

Received October 29, 1974; revised January 19, 1975.

letters to nature

Two kinds of stellar collapse

Here I shall show that the events which produce compact objects in a spherical symmetric manner and on a time scale much less than the orbital period, then the barycentre of the remaining system acquires the velocity

\[ v = \frac{M_M(n-1)-M_M}{M_M+M_{M-1}} \]  

where \( v \) is the initial relative orbital velocity of the two masses, \( M_M \) and \( M_{M-1} \) are the initial masses and \( q_{M,M-1} \) and \( M_M \) are the final masses. If more than one, then the initial mass of the binary is lost the system becomes unbound; if not, the system becomes eccentric. If the initial orbit is eccentric, the evolution is asymptotic, or it suffers another collision of the disk, with the companion star are included, then several parameters are needed to describe the event and to determine the new orbit. Conservation of angular momentum indicates that explosions producing less than a few tens of the binary mass, or giving the compact remnant a recoil velocity several tens of the binary will generally (except for fortuitous combinations of parameters) lead to orbits of substantial eccentricity. If \( M_M = M_{M-1} \) then we can not determine velocities of a tens of the binary will still result. Slightly more violent event produce hyperbolic orbits.

I first summarise the evidence for the expansion of an