Structure of proteins with single-site mutations: A minimum perturbation approach
(hemagglutinin/site-specific mutagenesis/protein folding/homology modeling)

H. H.-L. Shih, J. Brady*, and M. Karplus

Department of Chemistry, Harvard University, Cambridge, MA 02138

Contributed by M. Karplus, November 7, 1984

ABSTRACT A large number of mutant proteins with single amino acid substitutions are now being produced. The ability to predict the structural changes expected from such mutations would aid greatly in the efficient utilization of the mutagenic techniques and in the interpretation of the changes in stability and function that result. A minimum perturbation approach is suggested as a first step in such structural predictions and is tested by application to a recently isolated variant of the hemagglutinin glycoprotein. The agreement between the predicted structure and that inferred from the x-ray refinement is encouraging and provides support for the proposed modeling procedure.

The prediction of the native structure of a protein from its sequence is an objective that has been pursued by researchers for many years (1). Although the search for predictive methods was of primarily theoretical interest in the past, the powerful mutagenic techniques now available (2) are focusing renewed attention on the protein-folding problem. Since it is possible to make mutants almost at will, the essential question is which ones to make. A first step in answering this question is the prediction of the structure of such a mutant. For single amino acid substitutions, homology modeling of proteins, which has been widely attempted (3) but only rarely tested (4), is reduced to the simplest question: given two proteins that differ by a single amino acid, can one predict the structure of the mutant from the known (x-ray) structure of the parent? A procedure that is commonly used in homology modeling would introduce the new side chain in some arbitrary fashion, usually following the parent dihedral angles as far as possible, and then energy-minimize the entire structure to obtain a prediction for the mutant protein (4). We suggest an alternative approach that also makes use of the empirical protein energy but is closer in spirit to the model-building procedure that is used in structure determinations by x-ray crystallography. A search of the local conformational space of the substituted amino acid is made to find all structures that are of relatively low energy; each of these structures is then refined by constrained energy minimization, and the results are evaluated to determine the most likely structure(s). Such an approach is based on one assumption and one requirement. The assumption is that the conformation of a stable protein resulting from a single amino acid substitution resembles closely that of the parent protein, with any tertiary structural changes localized in the neighborhood of the substitution. Support for this assumption is provided, for example, by the x-ray structures for mutant hemoglobins (5), T4 lysozymes (6), and crambins (7), by the isomorphous crystallization of mutant proteins (8), and by solution data for phage λ-repressor mutants (9, 10). The requirement is that the approximate empirical energy functions now available be sufficiently accurate to determine the conformation of a side chain in the presence of the interactions with the known structure of the rest of the protein. This question has been examined in an earlier study of the bovine trypsin inhibitor (11). A rigid search of the full dihedral angle space for each individual side chain resulted in the correct positioning of the interior side chains (most χ angles within ±10°); exposed residues tended to be less well defined and, in some cases, could only be localized by including the crystal environment in the calculation (11). Further, although on a statistical basis side-chain dihedral angles tend to cluster in the positions suggested by isolated peptide calculations with standard geometries (12), individual residues were shown to have dihedral angles that deviate significantly from the statistical results, as a consequence of the local main-chain conformation and interactions with the surrounding protein (11).

The above considerations support the feasibility of a minimum perturbation approach in which energy minimization of a single structure is replaced by a thorough search of the local conformational space for the possible structures that are to be investigated. For a single amino acid substitution, such a search, which requires only limited computing time, avoids the local minimum problem of the usual energy function analysis. The details of the method used in the minimum perturbation approach are given in the next section. It is illustrated in the results section by an application to the mutation Gly-146 → Asp-146 in a recently isolated variant of the hemagglutinin (HA) glycoprotein (8). The predicted structure is compared with that inferred from the x-ray refinement. A concluding discussion evaluates the success of the present application and outlines the potential for future studies.

METHOD

The minimum perturbation approach makes use of an empirical energy function to search the conformational space and determine refined structures in the neighborhood of certain low-energy conformations found by the search. Details of the form of the empirical energy function and its use in examining protein structures have been given (13). The protein is represented by the extended atom model for nonpolar groups (e.g., a CH₃ group) and treats polar hydrogen explicitly. The total energy is a sum of local terms related to bond stretching, bond-angle bending, and dihedral and improper torsional rotation and of longer range terms due to van der Waals, electrostatic, and hydrogen-bonding interactions between nonbonded atoms. Given the structure of a protein, its energy can be evaluated directly. Alternatively, the lowest energy structure in the neighborhood of the initial structure can be found by various minimization algorithms. To take account of larger structural changes, molecular dynamics can be introduced as part of the minimization procedure (13).

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviation: HA, hemagglutinin.

*Present address: Department of Food Science, Cornell University, Ithaca, NY 14853.
The minimum perturbation approach has as its first step the verification that the original amino acid, which has been altered in the mutant, is in a minimum energy position in the isolated parent structure. This is done by the same dihedral search procedure as used in the bovine pancreatic trypsin inhibitor (11). If the correct minimum is not found, an investigation must be made to assess possible errors in the energy (e.g., neglect of contributions from the solvent or neighbors in the crystal; ref. 11) as well as possible errors in the structure. Assuming that the parent minimum is verified, the mutant amino acid is introduced into the parent sequence, and a grid search of its side-chain dihedral-angle space is made in the presence of the potential energy of interaction with the unaltered part of the parent structure. This initial search is restricted to varying the side-chain dihedral angles of the mutant amino acid, keeping all of its bond lengths and angles fixed and not moving any other protein atoms. Since only a small number of nonbonded interactions have to be calculated (i.e., those between the moving atoms of the amino acid and the rest of the protein), this search is very fast, even for a 10 Å grid involving the maximum of five dihedral angles, as for a lysine side chain.

The multidimensional potential energy surface generated in this way is examined for low-energy regions. If one or more regions have energies that suggest a satisfactory structure (i.e., energies similar to those of the parent protein), a search is restricted to them. If the search has yielded only high-energy structures, most likely due to unfavorable van der Waals contacts, more extended explorations of the energy surface, including variations in the surrounding side-chain dihedral angles and the positions of local backbone atoms, may be required to obtain satisfactory starting points for the next step. This consists of constrained energy minimization, the constraints being introduced to keep the portions of the protein that are assumed to be unaltered by the perturbation close to the known structure. (It has been found that the present energy functions, when used for energy minimization in vacuum, yield converged structures that have a root-mean-square shift in the atomic positions of about 1 Å relative to a well-refined x-ray structure). In the present study, atoms within 5 Å of the Cα atoms of the mutant amino acid were unrestrained, and atoms at successively larger distances were subjected to increasing restraints, with all atoms further than 15 Å kept fixed. Harmonic restraints were used, and adapted-basis Newton–Raphson (ABN) energy minimization was performed (13).

The search and minimization procedure may yield several reasonable structures. At best there will be one clearly defined lowest energy structure. Instead there may be several minima that are difficult to distinguish by the available methods; their existence may suggest that more than one side-chain conformation can occur.

RESULTS

The minimum perturbation approach has been tested in a blind study of an antigenic variant, V3a, of the HA glycoprotein of the influenza virus (10). In this variant an aspartic acid has replaced a surface glycine at position 146 of the parental (X31) sequence, whose x-ray structure has been determined at 3.0-Å resolution (14); residue 146 is part of an exposed loop region (positions 140–146) that has been implicated as an antigenic site (15). Since the parent sequence has a glycine in the position of interest, no preliminary search was made to verify its structure. The aspartic acid was introduced into a partially refined model of the parent structure, and the two dihedral angles (X1, X2) of the mutant side chain were varied systematically at 10° intervals. For each side-chain position, the energy was calculated; only the isolated protein was considered. The resulting potential energy map (see Fig. 1 Upper) has a well-defined low-energy region for −120° ≤ X1 ≤ −150° and 50° ≤ X2 ≤ 150°. The rest of the potential surface is much higher in energy with a secondary minimum region, more than 60 kcal above the primary region, in the neighborhood of X1 = 0° and X2 = 45°. This broad range of X2 values. Although the region accessible to X2 is characteristic of aspartic acid residues in proteins (11, 12), the low-energy region for X1 is displaced somewhat from the statistical norm (12). The extended high-energy ridge for −70° ≤ X1 ≤ 70° arises from repulsion between the carboxyl group of Asp-146 and its main-chain NH; that at X1 = 150° is due to repulsions from the main chain of Phe-147. Fig. 1 Lower makes clear that the conformational space is dominated by the van der Waals interactions, though electrostatic and hydrogen bonding contributions are of some significance.

The low-energy region in Fig. 1 has two local minima of very similar energy: the more stable one (I) is at X1 = −145° and X2 = 130° and the other one (II) is at X1 = −130° and X2 = 65°; in both conformations hydrogen-bonding interactions exist between the carboxyl oxygens of Asp-146 and two donor groups (the guanidinium group of Arg-141 and the amide H of Phe-147). Starting with the two structures, 400 steps of constrained-adapted-basis Newton–Raphson energy minimization were performed. The two starting structures yielded two final structures [I gave (I') X1 = −162° and X2 = 153°; and II gave (II') X1 = −155° and X2 = 98°]. Minimization of a third structure (X1 = −135° and X2 = 95°) on the ridge between the two minima (I and II) yielded a structure identical with II'. Fig. 2 Upper shows the rigid minimum structure (I) in stereo; the position of Asp-146 (X1 = −145° and X2 = 130°) is close to that obtained in the mutant x-ray refinement (X1 = −129° and X2 = 128°) (8). In Fig. 2 Lower the minimized structure (II') is introduced into a difference density map obtained by comparing the mutant and parent data (figure 2c of ref. 8). Although the orientation of Asp-146 is further from the refined x-ray result, the side chain fits into the difference density. It is of interest that in this case the amide group of the carboxyl oxygens is oriented so as to hydrogen bond with both Arg-141 and Phe-147.

From the calculations it is evident that the value of X1 is well defined and that of X2 is less so. This is in accord with analyses of aspartic acid residues in known protein structures (12). A more quantitative assessment of the theoretical results could be made by using the predicted structures as models in a refinement. The minimizations (I' and II') also suggest a small rotation of Arg-141 that improves the hydrogen bond with Asp-146. Such a displacement of Arg-141 was in fact found in the x-ray structure when it was reexamined to determine whether the predicted change was visible (8). No other significant changes are apparent in the experimental difference map; this is in accord with the minimum perturbation hypothesis.

In doing the calculation, a partly refined version of the parent structure was used. This had Ser-145 in a position with X1 = 0° (see Fig. 2) instead of the value obtained in the most recent refinement of the parent structure (X1 = 67°). It is of interest that this difference in structure of a neighboring residue, which points away from the substitution site, had no evident effect on the results of the search. This supports the idea of locality. Further, although Ser-145 was included in the minimization region, its final coordinates were similar to the initial values (X1 = −29°), as expected if it is in one, of several, local minima.

DISCUSSION

With an exact free-energy functional and infinite time, a complete calculation would yield the correct structure for a mutant protein that is at a global minimum. Since neither of
FIG. 1. Potential energy map for variation of $\chi_1$ and $\chi_2$ of Asp-146; the remainder of the protein was kept rigid at the parent x-ray structure. (Upper) The figure represents the dihedral-angle energy of the Asp-146 side chain plus its van der Waals, electrostatic, and H-bonding energy and interaction with the rest of the molecule; a 10-Å distance cut-off from C$^\alpha$ of Asp-146 was used for the calculation (for details of the energy function, see ref. 13). The energy contours are relative to the minimum energy geometry at zero only. Contours: ---, at 0.5, 1.5, 2.5, and 3.5 kcal (1 cal = 4.184 J); --, at 10, 20, 30, 40, and 50 kcal; ---, at 100, 200, and 300 kcal; and --, at 500, 2300, 4100, ... 9500 kcal. (Lower) Same as Upper for the van der Waals energy only. Contours: ---, from 5 to 20 kcal; --, from 100 to 300 kcal; ---, from 500 to 9500 kcal.

If these are available it is necessary to try to develop a method that can be applied within the limitations of the present theoretical methodology. We here suggest a conformational search as the first step, instead of the more usual direct energy-minimization approach. This suggestion is based on the assumption that the mutation, if stable, produces only a small local change in the parent structure (the minimum-perturbation hypothesis). If, unlike the example described above, no satisfactory starting structure is found in the initial search, more extended explorations must be made (e.g., the search could include neighboring side chains and the local backbone). However, with the necessity of enlarging the
search, the probability of a successful prediction is likely to decrease.

A possible corollary of the minimum perturbation hypothesis might be that the absence of a low-energy structure resulting from local changes is an indication that the mutant is not stable; as yet, this aspect of the problem has not been investigated. It also should be of interest to look for exceptions to the minimum-perturbation hypothesis, where more extensive correlated changes in a structure are induced by a single mutation. Some candidates might be residues in flexible loops, in hinges or at contacts between pairs of secondary structural elements, domain, or subunits.

The present approach does not address the question of the relative stability of the mutant versus the parent protein in cases where both yield stable folded structures. For any such evaluation, a knowledge of the structure is clearly important. However, since the thermodynamic stability of a folded protein is determined by the free-energy difference between the folded and denatured form, the effect of the mutation on the stability of both forms will have to be assessed before this problem can be solved (16).

We thank D. C. Wiley for calling our attention to the variant HA glycoprotein, for withholding all information concerning its structure until our calculations had been completed, and for discussions of the manuscript. The structure of the parent HA was kindly made available by D. C. Wiley. The assistance provided by B. Brooks and A. Brünger in applying CHARMM to this problem is gratefully acknowledged. We thank R. E. Brucoleri for discussions and calculations that have contributed to the search approach. We thank R. T. Sauer for helpful comments on the manuscript. The work was supported in part by a grant from the National Institutes of Health.