ELECTROSTATIC INTERACTIONS AND SECONDARY STRUCTURES IN PROTEINS

R. L. Jernigan, S. Miyazawa, and S. C. Szu, *Laboratory of Theortical Biology, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20205 U.S.A.*

In an attempt to understand the occurrences of secondary forms in proteins, we have calculated approximate free energies for a large set of regular secondary regions and have compared several methods for choosing a single molecular conformation. Approximate conformational free energies have been obtained by directly calculating electrostatic energies and approximating the nonelectrostatic energies in a simple manner, with parameters being related to the actual fractions of conformations occurring within the protein. Indirect evidence for the importance of electrostatic energies in determining secondary structures was offered by the success of the prediction method of Ptitsyn and Finkelstein (1970) and more directly by the observation of the occurrences of favorable side-chain charged pairs in α-helices (Maxfield and Scheraga, 1976).

For a protein of specified sequence, we calculate the energies of all possible regular regions of four states: α-helix, β-strand, α_L left-handed α-helix, and β_L left-handed β-strand, for maximum lengths of 26, 13, 3, and 3, respectively. We assume the side-chain atoms to be at fixed positions, relative to backbone atoms, for each type of residue. These sites are determined by averaging over the positions reported in the x-ray structures of six proteins. Point charges are assigned for ionized residues, from pK values at neutrality, and for polar bonds, from small molecule dipole moments. Backbone atomic charges and other structural parameters are the same as those used by Brant et al., (1967). Here the total energy of a region spanning residues i through j in a regular conformation of type t is approximated by

$$E_{t;i,j} = E^{bb-sc,es} + E^{sc-sc,es} + \sum_{k=i}^{j} E^{t}_{k}.$$

The first two terms on the right-hand side are the electrostatic energies for backbone-side chain and side chain-side chain interactions which are calculated explicitly in terms of point charge Coulombic interactions. All such interactions within each region are included. The dielectric constant is taken to be 3.5. There are only two types of parameters E^{t}_{k}, one for glycine and one for all other types of amino acids. Further simplifications of these parameters are $E_x^{\beta} = E^{\beta}_{gly} = E_{gly}^{\beta_L} = 0$, $E_x^{\alpha_L} = E_x^{\beta_L}$, $E^{\alpha}_{gly} = E_{gly}^{\alpha_L}$, and $E_{pro}^{\alpha_L} = E_{pro}^{\beta_L} = \infty$. These are based upon the simplest possible considerations of steric interactions. A special parameter E_{xpro} is used for the residue, except glycine, preceding a proline in an α-helix. A parameter E^{α}_{H} for the helical hydrogen bond energy is also utilized. The most important parameter is E_x^{α}. A simple curve, based on six proteins, has been obtained which relates E_x^{α} to the experimental fractions of α-helical and β-strand residues. In a similar way $E_x^{\alpha L}$ appears to be related to $-RT \ln(n_{\alpha L}/n_{\beta})$. The value of $E^{\alpha}_{H} = -0.4$ kcal/mol appears to give good results for most molecules.

An energy minimization scheme (Jernigan and Szu, 1979) has been developed that rigorously yields the minimum energy combination of independent regions. Here these are taken to be the regular secondary regions. This simple method is a dynamic programming algorithm. Its intention is to include a majority of possible interactions. With the present set of conformations, it would likely be most successful for proteins with a substantial portion of

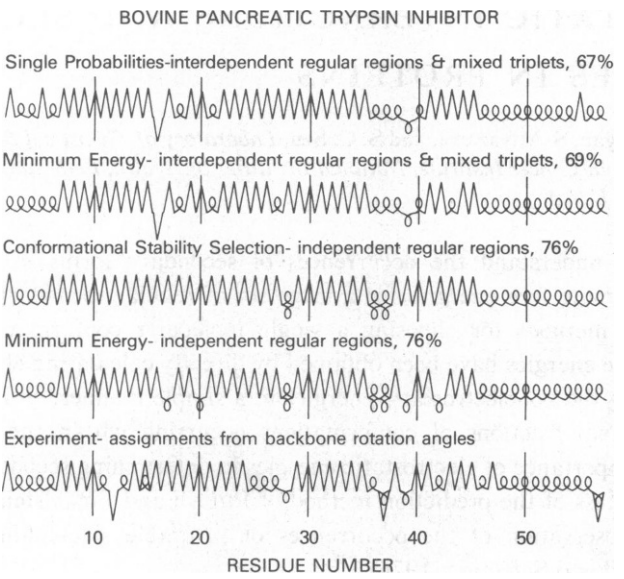

Figure 1 Secondary structures of trypsin inhibitor. Conformations of residues are specified by (α) as right-handed α-helix, (⋀) as right-handed β-stranded, (ʊ) as left-handed helix, and (⋁) as left-handed β-strand. Experimental cases which are intermediate in form are indicated by overlaps of two symbols; calculated cases which are indistinguishable are indicated by overlapping symbols. Experimental conformations are taken from Feldmann (1976). Parameters used for the top two methods were $E_x^{\alpha}/RT = 1.4$, $E_x^{\alpha_L}/RT = 3.2$ $E^{\alpha}_{gly}/RT = 0$, $E_{xpro}/RT = 10$ and $RT = 0.6$; for the bottom two methods these were $E^{\alpha}/RT = 1.5$, $E_x^{\alpha_L}/RT = 2$, $E^{\alpha}_{gly}/RT = 0.4$, and $E_{xpro}/RT = 5$, for 2,111 conformations. The percent of residues in agreement with the experimental conformation is indicated together with the name of the method.

their residues in regular conformations. Three other methods to determine a best set of conformations have been applied. The conformational stability selection method utilizes the same independent region energies and is an approximation to the minimum energy method. In it, the lowest energy conformer is chosen; regions of successively higher energy are chosen so that they do not overlap sequences whose conformations have already been assigned. The other two methods utilize energies for all regular regions as well as the energies of the 64 mixed triplets (four states) for each position within the sequence. Thus a limited range of interdependence between the regular regions is included. This set of energies is minimized in a method similar to that for independent regions. Also, probabilities at each position for each of the four states are calculated from a rigorously formulated partition function composed of all combinations of possible secondary regions (Jernigan et al., 1980). The most probable state is assigned for each residue. Results for all four of these methods applied to pancreatic trypsin inhibitor are shown in Fig.1. All results are similar in quality; however inclusion of the mixed triplets yields somewhat worse result. The dominant conformation is the helix near the carboxy- terminus; its electric stability is much greater than that of any other region. It is surprising that the results for the two methods given in the top two lines of the figure are so similar. Inspection of conformational probability profiles reveals that many residues exhibit a strong preference for single conformations; in these cases, the same conformations will be chosen by all selection methods. Differences among results for different methods may represent positions of relative flexibility and instability.

Received for publication 17 December 1979.

REFERENCES

Brant, D. A., W. G. Miller, and P. J. Flory. 1967. Conformational energy estimates for statistically coiling polypeptide chains. *J. Mol. Biol.* **23**:47–65.

Feldmann, R. J. 1976. Atlas of Macromolecular Structures on Microfiche (AMSOM). Tracor Jitco Inc., Rockville, MD. Coordinates taken from structure 3.6.1.1.1 originally supplied by R. Huber.

Jernigan, R. L., and S. C. Szu. 1979. Conformational energy minimization in the approximation of limited range interactions. *Macromolecules.* **12**:1156–1159.

Jernigan, R. L., S. Miyazawa, and S. C. Szu. 1980. Stabilization of regular conformational regions in proteins by intraregion electrostatic interactions. *Macromolecules.* In press.

Maxfield, F. R., and H. A. Scheraga. 1976. The effect of neighboring charges on the helix forming ability of charged amino acids in proteins. *Macromolecules.* **8**:491–493.

Ptitsyn, O. B., and A. V. Finkelstein. 1970. Connexion between the secondary and primary structures of globular proteins. *Biophysics (Engl. Transl. Biofizika).* **15**:785–795.

MEASUREMENTS OF THE ELECTRIC CHARGE AND ION-BINDING OF THE PROTEIN FILAMENTS IN INTACT MUSCLE AND CORNEA, WITH IMPLICATIONS FOR FILAMENT ASSEMBLY

G. F. Elliott, *Biophysics Research Group, Open University Oxford Research Unit, Foxcombe Hall, Boars Hill, Oxford, England*

Microelectrode techniques can be used to record a Donnan potential from muscle and corneal tissue, where the bounding membranes have been removed or inactivated. From this potential, due to the different concentrations of freely-diffusible ions inside and outside the tissue, the net internal fixed-charge concentration of the nondiffusible proteins can be calculated (e.g., Collins and Edwards, 1971). We have combined such observations with x-ray diffraction measurements of the distances between the protein filaments, and can deduce the charge per unit length of the filaments *in situ*. (Elliott et al., 1978, footnote 1).

For muscle (glycerinated rabbit muscle and mechanically and chemically skinned rat and frog fibers), the fact that the Donnan potential does not change as the sarcomere length is changed gives further information, and we can calculate separately the charges on the thick myosin-containing filaments and the thin actin-containing filaments. As the ionic strength of the bathing solution increases, the thin filament charge does not change, but the thick filament charge increases swiftly, becoming more negative (at physiological pH). The thin filament charge is ~ 1–2×10^4 e μm^{-1}, the thick filament charge is 4×10^4 e μm^{-1} in a 5 mM KCl solution and 12×10^4 e μm^{-1} in a 100 mM solution.[1]

This concentration dependence of the thick-filament charge probably means that the myosin-containing filaments bind negative ions, and we have therefore further investigated the differential binding of chloride and phosphate ions. Both of these have been shown to bind, and the maximum charge is achieved in a solution containing ~5 : 1 chloride : phosphate. Magnesium decreases the thick-filament charge (from $\sim 12 \times 10^4$ e μm^{-1} to $\sim 8 \times 10^4$ e μm^{-1} when the external magnesium is increased roughly three-fold from 1 mM) but will only do so in the presence of phosphate ions.[2] These ion-binding effects in muscle have a concentration

[1]G. R. S. Naylor and Elliott, G. F. Manuscript in preparation.

[2]Bridgman, T. D., E. M. Bartels, and G. F. Elliott. Manuscript in preparation.