[CONTRIBUTION FROM THE DEPARTMENT OF BIOCHEMISTRY, DUKE UNIVERSITY, MEDICAL CENTER, DURHAM, NORTH CAROLINA]

Contribution of Hydrophobic Interactions to the Stability of the Globular Conformation of Proteins

BY CHARLES TANFORD

RECEIVED APRIL 9, 1962

A simple model is developed for calculation of the difference in free energy (ΔF) between the native and unfolded forms of a protein molecule in solution. A major term in the expression for ΔF arises from the increase in entropy which accompanies unfolding. This term is negative, *i.e.*, it favors the unfolded form. In water, therefore, where a compact globular conformation is stable, local interactions must exist which make a large positive contribution to ΔF. One such interaction in the hydrophobic interaction, which results from the unfavorable arrangement of water molecules which takes place whenever there is contact between water and a non-polar portion of a protein molecule. There are many such contacts when the protein molecule is unfolded, but relatively few in the native state, so that a positive contribution to ΔF results. When amino acids with non-polar side chains are dissolved in water, the same interactions must occur. The magnitude of these interactions can then be estimated from relative solubilities of appropriate amino acids in water and other solvents. Such estimates are made in this paper, and the conclusion is that these hydrophobic interactions alone may be able to account for the instability of an unfolded protein, relative to a suitable globular conformation, in aqueous solution. The model used cannot predict the structure which will be adopted by a given protein molecule in its native state. General considerations suggest, however, that the hydrophobic interactions are compatible with a large variety of structures and that specificity of structure is at least partly due to hydrogen bonds between peptide groups (as well as other polar groups) trapped within the hydrophobic interior.

It has been established for a number of proteins that "denaturation" or unfolding of the native structure in solution is a reversible process. This has led to the conclusion that the conformation of at least the smaller proteins, in solution, is always the thermodynamically stable conformation, determined completely by the nature and sequence of the constituent amino acids and by their interactions with each other and the surrounding solvent.[1] This conclusion permits consideration of the stability of protein conformations in terms of thermodynamic properties, and the first objective of this paper is to present a simple theoretical treatment of this problem.

In a sufficiently approximate theory of the thermodynamic properties of polymer solutions,[2,3] the energy and entropy contributions to the free energy are calculated separately, and the macromolecular nature of the solute enters into the theory only in the calculation of the entropy, into which it enters in two ways, (1) in the form of a potentially large volume of exclusion, which leads to a potentially large effect of solute concentration on the free energy, and (2) in the form of a potentially large configurational entropy of the polymer chain. The energy contribution to the free energy, in a theory at this level of approximation, is computed solely as the sum of nearest neighbor interaction energies, with different values assigned to interactions due to polymer segment/polymer segment, polymer segment/solvent and solvent/solvent contacts. For this part of the calculation, only the immediate environment of any part of the polymer chain need be specified. The result obtained would be the same if the individual parts of the polymer chain were actually independent molecules.

When solutions in water are considered, it is necessary to take into account a nearest neighbor interaction entropy, as well as interaction energy, because of the existence of organized structures in the solvent which may be perturbed by contact with parts of the polymer chain.[4] A calculation of this effect is clearly also independent of the macromolecular nature of the solute. Even for relatively small molecules the effects of water structure on partial molal heat capacities, chemical potentials (solubility), etc., are known to be the sum of the effects on individual parts of the molecule.[5]

In the present paper we shall estimate, at this level of approximation, the free energy change attending a change in protein conformation. More specifically, we shall consider protein *denaturation*, *i.e.*, the change from a unique folded "native" conformation to a completely unfolded conformation. We shall consider the change occurring essentially at infinite dilution, so that the concentration-dependent terms of the free energy do not enter into the calculation. The only part of the calculation which specifically involves the knowledge that we deal with macromolecules is then the calculation of the configurational entropy of the polymer chain. Vicinal interaction effects may be computed as if the interacting parts were on small molecules rather than protein molecules.

That the effects of interaction on the parts of large molecules can be expected to be similar to such effects on similar small molecules does not simplify the theoretical calculation of such effects. It does permit us, however, to treat these effects in a protein molecule without theoretical calculation at all, if we can find suitable small molecules (such as amino acids) where similar interactions occur and where their effect on thermodynamic properties has been measured. This is the technique which will be employed in this paper. We shall make use of solubilities of amino acids and related substances in various solvents. These data give us the effect of local environment on the chemical potential in solution (a separation into energy and entropy will not be attempted) and allow us to

(1) See for example C. B. Anfinsen, *J. Polymer Sci.*, **49**, 31 (1961).

(2) M. L. Huggins, *J. Phys. Chem.*, **46**, 151 (1942); *Ann. N. Y. Acad. Sci.*, **41**, 1 (1942); *J. Am. Chem. Soc.*, **64**, 1712 (1942).

(3) P. J. Flory, *J. Chem. Phys.*, **10**, 51 (1942).

(4) H. S. Frank and M. W. Evans, *ibid.*, **13**, 507 (1945).

(5) An example is provided by aqueous solutions of normal saturated alcohols. See Frank and Evans, ref. 4, Fig. 3. See also footnote 26, below.

estimate the changes in local interaction free energy which appropriate parts of a protein molecule undergo when the protein molecule is unfolded.

Many of the parts of the calculation which will be required have already been reported in two papers by Kauzmann.[6,7] It should in fact be emphasized that this paper can claim no novelty in conceptualization of the problem of protein denaturation, as the method of approach rests entirely on principles implicit in Kauzmann's papers. Valuable contributions to this problem have also been made by Klotz,[8] who was one of the first to recognize the importance of solvent interactions to any theory of protein structure.

The Free Energy of Unfolding in Water.—Figure 1 is a schematic diagram of the two conformations which we have chosen as models for the native and unfolded states. The native conformation is taken to be a uniquely folded one with no flexibility at all. It is assumed that the interior of the folded molecule is devoid of and inaccessible to the solvent. Reasons for believing this to be generally true for globular proteins in aqueous solution were given in an earlier paper.[9] The statement is certainly true for the one protein (myoglobin) whose three-dimensional structure, in hydrated crystals, has been almost completely determined.[10]

The unfolded conformation is taken to be a random flexible coil, sufficiently extended so that all of its parts are in intimate contact with the surrounding solvent.[11]

There are clearly two major differences between the two conformations. The first is the difference in flexibility, which will enter into the calculation of the change in configurational entropy. The second difference is that many parts of the protein molecule, in the native conformation, have an environment consisting entirely of other parts of the protein molecule; whereas in the unfolded form all parts are in intimate contact with the solvent. This will introduce large changes into the contribution of vicinal interactions to the free energy. To calculate these changes it is necessary to be somewhat specific about the actual structure of the native conformation, and we shall make the following assumptions concerning it:

(1) The charged groups[12] of the protein molecule will be assumed to be at the protein/solvent interface of the native structure, in as intimate contact with the solvent as the same groups would have when attached to a small molecule in aqueous solutions. (This means that these groups will have the same environment in the native and unfolded conformations.) This assumption is dictated by experimental and theoretical studies of the hydrogen ion titration curves of proteins. Theoretically, both the intrinsic pK's of titratable groups and the coulombic interaction between charges depend crucially on a parameter d, which represents the closeness of approach of the solvent to the charged group.[13,14] Experimental values, both for intrinsic pK's and for electrostatic interaction, require that this parameter, for titratable groups of protein molecules, have the same value as for appropriate model compounds. Where appreciable titration anomalies are experimentally observed, they are nearly always in a direction which indicates that the uncharged form is removed from the protein/solvent interface in the native conformation, the titration anomaly being due to the conformational change which must occur (as an accompaniment of titration) to bring the charged form to its normal position in intimate contact with solvent.[15]

It is important to note that only those parts of the ionic protein side chains which are actually charged need be in contact with the solvent. The series of amines, $CH_3(CH_2)_nNH_3^+$, with n ranging from 0 to 16 have identical pK values,[16] and even tertiary amines often have pK's which do not differ appreciably from those for the primary amines. Thus an incomplete hydration shell suffices to account for the similarity in electrochemical properties between charged groups on proteins and similar groups on small molecules, and no need exists for the uncharged part of the side chain to be at the protein/water surface.

(2) We shall assume that most of the non-polar or hydrophobic parts of the molecule will be in the interior of the native structure. (The principal result of this paper will be to show that it is impossible to account for the stability of the globular structure in water unless this assumption is made.)

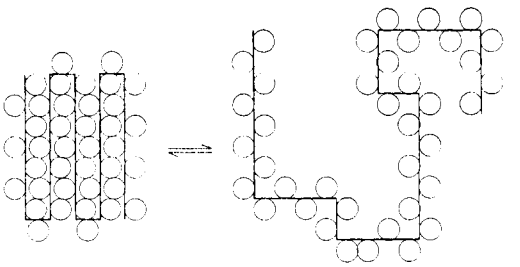

Fig. 1.—Schematic diagrams of the native and unfolded states of a protein molecule. The circles represent side chains, which in a real protein would of course vary in size. The important point which the figure illustrates is that many of the side chains, in the native state, must be in contact with each other but removed from contact with solvent. In the completely unfolded state all side chains project into the surrounding solvent.

(6) W. Kauzmann, in "The Mechanism of Enzyme Action" (W. D. McElroy and B. Glass, eds.), Johns Hopkins University Press, Baltimore, 1954, p. 40.

(7) W. Kauzmann, *Advances in Protein Chem.*, **14**, 1 (1959).

(8) I. M. Klotz, *Science*, **128**, 815 (1958).

(9) C. Tanford, P. K. De and V. G. Taggart, *J. Am. Chem. Soc.*, **82**, 6028 (1960).

(10) J. C. Kendrew, H. C. Watson, B. E. Strandberg, R. E. Dickerson, D. C. Phillips and V. C. Shore, *Nature*, **190**, 666 (1961). The author is indebted to Dr. Kendrew for making the three-dimensional model of the protein available to him for examination.

(11) The possible differences between a real unfolding process and this idealized model, and their effect on the calculations, will be discussed later, in connection with the results of Table III.

(12) Our calculation will initially be for neutral pH, and under these conditions most carboxyl, amino and guanidine groups will be charged and so will some of the imidazole groups of histidine side chains.

(13) C. Tanford and J. G. Kirkwood, *J. Am. Chem. Soc.*, **79**, 5333 (1957). See especially footnote 14.

(14) C. Tanford, *ibid.*, **79**, 5340, 5348 (1957).

(15) C. Tanford, *ibid.*, **83**, 1268 (1961).

(16) H. C. Brown, D. H. McDaniel and O. Häfliger, in "Determination of Organic Structures by Physical Methods," E. A. Braude and F. C. Nachod, eds., Academic Press, Inc., New York, N. Y., 1955, p. 573.

The hydrophobic parts of the protein molecule must be taken to include not only the fully non-polar side chains, such as leucine, but also the non-polar parts of side chains which contain charged or polar groups, *e.g.* the four methylene groups of lysine side chains.

It should be pointed out that what is important in the distinction between being inside the protein molecule and being at the protein/solvent interface is the sum of all nearest-neighbor contacts. Thus a leucine side chain could actually extend to the interface (provided it did not bulge outward from the interface), where it would make contact with, say, a single water molecule. This side chain would still be "inside" by our definition because the majority of nearest neighbor contacts would not be with the solvent.

(3) In the primary structure of a protein molecule there is no neat spatial separation between polar and non-polar parts. Thus the collapse of the molecule into a globular structure with a well-defined interior cannot occur in such a way as to exclude polar or hydrogen bond-forming groups from the interior. We shall assume then that polar groups may be either inside or at the surface. If they are inside we shall assume that they form hydrogen bonds to other polar groups. (This assumption, like the preceding one about the location of non-polar groups, will be seen to be a necessary one if we are to account for the stability of the globular structure in water.) If polar groups are at the protein/solvent interface, they will of course also be hydrogen-bonded, in this case to water molecules.

It is of interest to refer again to the structure of myoglobin in hydrated myoglobin crystals.[10] In this structure most of the peptide groups are "inside," and they are hydrogen-bonded to each other, predominantly so as to form α-helices. Indirect evidence[9,17] suggests that many globular proteins do not contain such helices. Nevertheless we would anticipate that peptide groups and other polar groups will still participate in some form of hydrogen bonding and will assume this in our model.

With these assumptions, the free energy of unfolding may be calculated as

$$\Delta F = -T\Delta S_{\text{conf}} + \Sigma\Delta f_{\text{u}} \qquad (1)$$

where ΔS_{conf} is the change in conformational entropy of the polypeptide chain, and Δf_{u} is the free energy change for transfer of the small component groups of the molecule, from the environment they have in the native form, to the environment they have in the unfolded form.

Calculation of ΔS_{conf}.—The conformational entropy per mole of protein is computed in the usual way[2,3,6] as $R \ln z^x$ where x is the number of points of flexibility per molecule, and z is the number of possible orientations of equal energy at each such point. In the native conformation, which we have assumed completely inflexible, the conformational entropy is thus zero.

The unfolded form, on the other hand, is a flexible coil, but there is not much evidence on which to base an estimate of x and z. Doty, Bradbury and Holtzer[18] have found that poly-γ-benzyl-L-glutamate, in solvents in which it is randomly coiled, has an intrinsic viscosity proportional to the 0.87 power of the molecular weight. This indicates that the polypeptide chain in this molecule has relatively low flexibility.[19] Low flexibility would also result from planarity of the peptide group as a whole. Kauzmann[6] has suggested that randomly coiled proteins may have on the average three positions of flexibility per residue with $z = 2$ at each position. This leads to an entropy of $4.1n$ e.u. per mole, where n is the number of residues. We shall not attempt to seek a better figure, but use Kauzmann's figure with the understanding that it can represent only an order of magnitude estimate. Kauzmann's figure leads to a free energy contribution ($-T\Delta S_{\text{conf}}$) at 25° equal to −1200 calories per residue. This entropy factor alone thus stabilizes the unfolded form of the protein molecule by a very large amount.

An important factor omitted in this calculation of ΔS_{conf} is the influence of disulfide cross links. The existence of such cross links must to some extent reduce the number of actual configurations available to the unfolded chain and hence must reduce ΔS_{conf} below the value which would obtain in the absence of cross-links. Methods for calculation of this effect have also been developed by Kauzmann.[20]

Calculation of $\Sigma\Delta f_{\text{u}}$.—Since the entropy factor calculated above stabilizes the unfolded form of the protein by a large amount, there must be correspondingly large terms in the factor $\Sigma\Delta f_{\text{u}}$ which stabilize the globular form, *i.e.*, there must be Δf_{u} terms which are large and positive. It is evident that the charged groups of the protein molecule cannot be the source of these terms, for we have already noted that the environment of these groups must remain essentially unchanged in the denaturation process, and Δf_{u} for these groups must therefore be zero.[21] The major possibilities are then[7]:

(1) That the polar uncharged groups, especially the peptide groups, have a lower free energy when they participate in intramolecular hydrogen bonds within the native structure than when hydrogen-bonded to water in the unfolded state.

(2) That the hydrophobic interaction between water and non-polar parts of the protein molecule makes a large contribution to $\Sigma\Delta f_{\text{u}}$. This interaction leads to an increased free energy whenever there is contact between water and a non-polar part of the protein. Since there are many such contacts in the unfolded form, but relatively few in the native form (actually none in the idealized model here used), the contribution to $\Sigma\Delta f_{\text{u}}$ will be positive.

On the basis of studies with model compounds, it is considered improbable that intramolecular hydrogen bonds are appreciably stronger than hydrogen bonds to water. The data of Schellman[22]

(17) B. Jirgensons, *Tetrahedron*, **13**, 166 (1961).

(18) P. Doty, J. H. Bradbury and A. M. Holtzer, *J. Am. Chem. Soc.*, **78**, 947 (1956).

(19) C. Tanford, "Physical Chemistry of Macromolecules," John Wiley and Sons, Inc., New York, N. Y., 1961, p. 409.

(20) W. Kauzmann, in "Sulfur in Proteins," R. Benesch, *et al.*, eds., Academic Press, Inc., New York, N. Y., 1959, p. 39.

(21) The electrostatic interaction between these groups is a relatively unimportant factor, as will be shown below.

and of Gill, *et al.*,[23] suggest that inter-peptide hydrogen bonds do lead to a slightly positive value of Δf_u for the peptide group. The more direct data of Klotz and Franzen[24] however would give the opposite result, suggesting a moderately negative value for Δf_u for the peptide group. We shall not attempt to resolve this discrepancy but confine our attention to the second possibility. Our objective will be to estimate as quantitatively as possible the contribution which hydrophobic interactions alone are likely to make to $\Sigma\Delta f_u$.

It is possible to approach this problem by making use of the solubilities of amino acids in ethanol and water which have been measured in several laboratories.[25] Apart from an activity coefficient term, which involves only the effect of amino acid concentration on the free energy, the relative solubilities in these solvents are a measure of the free energy change ΔF_t for transfer of one mole of amino acid, at the same concentration, from ethanol to water, *i.e.*

$$\Delta F_t = RT \ln (N_{EtOH}/N_{H_2O}) \qquad (2)$$

where N_{H_2O} and N_{EtOH} represent solubilities in mole fraction units in the two solvents. Moreover, we may expect that ΔF_t will be primarily a measure of those parts of the chemical potential of the amino acid molecules which arise from interaction with the solvent, just the part in which we are interested.

Cohn and Edsall[25] made the significant discovery that ΔF_t for the various amino acids could be closely computed as the sum of *constant* contributions from the various parts of the amino acid molecule, *i.e.*, we can write

$$\Delta F_t = \Sigma\Delta f_t \qquad (3)$$

where each Δf_t is a constant. For example, a CH_2 group always increases ΔF_t by about 700 cal., rearrangement of H_2N–CO–CHR–OH to ^+H_3N–CHR–COO^- (R being any side chain) always decreases ΔF_t by about 3700 cal., and so forth.[26] In the present application we can thus use these data to calculate the contribution of amino acid side chains (particularly the non-polar ones) to ΔF_t, each such contribution being simply the difference between ΔF_t for an amino acid and the value for glycine. The calculations are shown in Table I. Table I also contains ΔF_t values for ethane and methane, to show the similarity between the hydrocarbon side chains of amino acids and hydrocarbon molecules themselves. The table also contains figures for some polar side chains which will be discussed below.

It has been shown by Cohn and Edsall[25] that the Δf_t values which one obtains for transfer of a non-polar amino acid side chain from ethanol to water are not very different from the Δf_t one obtains for transfer of these side chains from several other solvents to water, as is illustrated for norleucine by Table II. This result suggests that the Δf_t values of Table II, and those for non-polar side chains in Table I, reflect primarily the hydrophobic interactions which take place when water is the solvent. The solvent from which transfer to water is carried out is less important,

TABLE I[a]

FREE ENERGY CHANGE IN CALORIES PER MOLE FOR TRANSFER FROM ETHANOL TO WATER AT 25°

	ΔF_t, whole molecule	Δf_t, side chain contribution
Non-polar side chains		
Glycine	−4630	0
Alanine	−3900	+ 730
Valine	−2940	+1690
Leucine	−2210	+2420
Isoleucine	−1690[b]	+2970[b]
Phenylalanine	−1980	+2650
Proline	−2060[c]	+2600[c]
Other side chains		
Methionine	−3330	+1300
Tyrosine	− 930[d]	+2870[d]
Threonine	−4190	+ 440
Serine	−4590	+ 40
Asparagine	−4640	− 10
Glutamine	−4730	− 100
Aspartic acid[e]	−4090	+ 540
Glutamic acid[e]	−4080	+ 550
Contribution of a CH_2 group		
Ethane	+3020[f]	....
Methane	+2260[f]	....
Ethane–methane	....	+ 760
Alanine–glycine	....	+ 730
Leucine–valine	....	+ 730

[a] Except where indicated otherwise, data are taken from the compilations cited in ref. 25. [b] At 20° Δf_t has been calculated relative to glycine at 20°. The data used were of relatively low precision. [c] At 19° Δf_t has been calculated relative to glycine at 19°. The data used were of relatively low precision. [d] The data for tyrosine are for 95% ethanol, and Δf_t has been calculated relative to glycine in 95% rather than 100% ethanol. One would expect Δf_t in 100% ethanol to be slightly larger than the value given. [e] The side chain carboxyl groups were uncharged in these experiments. [f] Data for ethane and methane are from A. Seidell, "Solubilities of Organic Compounds," 3rd Ed., Vol. II, D. Van Nostrand Co., New York, N. Y., 1941.

TABLE II[a]

FREE ENERGY CHANGE FOR TRANSFER OF NORLEUCINE SIDE CHAINS FROM SEVERAL SOLVENTS TO WATER AT 25°

Solvent	ΔF_t (cal./mole) Norleucine	Glycine	Δf_t (cal./mole) Norleucine side chain
Methanol	900	3430	−2530
Ethanol	1930	4630	−2700
Butanol	2330	5190	−2860
Acetone	3320	6000	−2680

[a] Based on data compiled by Cohn and Edsall, ref. 25.

(22) J. A. Schellman, *Compt. rend. trav. lab. Carlsberg*, **29**, 223 (1955). The calculation of Δf_u from these data is discussed by Kauzmann, ref. 7, p. 36.

(23) S. J. Gill, J. Hutson, J. R. Clopton and M. Downing, *J. Phys. Chem.*, **65**, 1432 (1961).

(24) I. M. Klotz and J. S. Franzen, *J. Am. Chem. Soc.*, **82**, 5241 (1960); *ibid.*, in press.

(25) Summaries of these data, with references to the original literature, are given by E. J. Cohn and J. T. Edsall, "Proteins, Amino Acids and Peptides," Reinhold Publishing Corp., New York, N. Y., 1943, Chapter 9; and by M. S. Dunn, F. J. Ross and M. P. Stoddard, in "Handbook of Chemistry and Physics," Chemical Rubber Publishing Co., Cleveland, Ohio.

(26) This additivity is in fact the justification for the entire treatment of this paper. If solvent interactions of amino acids can be taken as the sum of contributions from the various parts of the molecule, then the local interactions of protein molecules can surely be taken as the sum of contributions from the peptide groups and the amino acid side chains.

We can therefore assume with considerable confidence that the free energy of transfer of non-polar side chains from the inside of a protein molecule to an aqueous medium will also have the same order of magnitude as these Δf_t values. In other words, we can equate the Δf_t values for these side chains with the Δf_u values required in equation 1.

It was pointed out earlier that some amino acid side chains which carry charged or polar groups also include non-polar portions which may be expected to be hydrophobic in nature. One of these is the tyrosine side chain. It is seen from Table I that Δf_t for this side chain is essentially the same as Δf_t for the phenylalanine side chain. We have obtained the same result from solubility measurements in several other solvents. For example, preliminary data in dioxane–water mixtures yield Δf_t = 1800 cal./mole for transfer of a phenylalanine side chain from 60% dioxane to water, and Δf_t = 1760 cal./mole for the corresponding figure for a tyrosine side chain. It can therefore be concluded that the hydroxyl group of tyrosine does not appreciably affect the adaptability of the aromatic group as a whole to an aqueous environment, and the Δf_t value for the tyrosine side chain may thus also be equated with the Δf_u value needed in equation 1. In the absence of direct information, it is probable that the tryptophan side chain will behave similarly, with Δf_u of the order of 3000 cal./mole or larger.

On the other hand, it is likely that the four methylene groups of a lysine side chain are not as hydrophobic as a terminal $-(CH_2)_3-CH_3$ group (*i.e.*, a norleucine side chain) would be. McMeekin, *et al.*,[27] have measured the solubilities of the hydantoic acids of norleucine and of α-aminocaproic acid in several solvents. Calculation of Δf_t for the norleucine side chain from these data yields 2700 cal./mole for transfer from ethanol to water and 2900 cal./mole for transfer from butanol to water, essentially the same result as is obtained from solubilities of the parent amino acid in Table II. For the $-(CH_2)_4-$ group, however, one obtains Δf_t = 1500 cal./mole for transfer from ethanol to water, and 1800 cal./mole for transfer from butanol to water. While the $-(CH_2)_4-$ group in this compound is not bounded by the same groups as it would be on a lysine side chain, the result suggests that Δf_u for the non-polar part of that side chain would also be well below the value for a norleucine side chain. The same conclusion presumably applies to the three methylene groups of arginine. The Δf_u value for these groups is likely to be of the magnitude of Δf_u for an alanine side chain.

Table I indicates that the sulfur atom of methionine is less effective than the more strongly polar uramido group (of the hydantoic acids) in reducing hydrophobic interaction of neighboring methylene groups with water. While it would be desirable to have Δf_t values for transfer from other solvents to water, so as to exclude the possibility that the value for ethanol may partly represent a special effect of that solvent, it is not unreasonable, in the absence of such data, to take Δf_t as equal to Δf_u.

(27) T. L. McMeekin, E. J. Cohn and J. H. Weare, *J. Am. Chem. Soc.*, **58**, 2173 (1936).

It is also reasonable to use the value of Δf_t of Table I for threonine as equivalent to Δf_u, since the threonine side chain contains a terminal methyl group.

All the remaining side chains contain both polar groups and non-polar groups, but the non-polar parts are small in extent and they are terminated by polar or charged groups. It is likely that these side chains will show much more specificity in their interactions with their immediate environment than do the purely non-polar side chains. (This is indicated for instance by the fact that the usual effect of a CH_2 group is absent when we compare glutamic acid with aspartic acid in Table I. On the other hand, the two acids differ markedly from the corresponding amides.) For some of these partly polar side chains, Table I lists Δf_t values for transfer from ethanol to water. We cannot safely assume that Δf_t for transfer from, say, butanol to water would necessarily have the same value and cannot assume that the Δf_t values can be equated with Δf_u. For five side chains, those of cysteine, cystine, histidine, ionized glutamate and ionized aspartate, we have no data at all on which an estimate can be based.

We are now in a position to calculate the major part of the contribution of hydrophobic interactions of the non-polar parts of a protein molecule to $\Sigma\Delta f_u$. The result will depend on the amino acid composition, and the calculation for three well-known proteins is shown in Table III. The

TABLE III

CONTRIBUTION OF THE MOST IMPORTANT HYDROPHOBIC INTERACTIONS TO THE FREE ENERGY OF UNFOLDING AT 25°

Side chain	Δf_u per side chain, cal./mole	Number present in myoglobin[a]	Number present in β-lactoglobulin[b]	Number present in ribonuclease[c]
Tryptophan	3000	2	2	0
Isoleucine	2970	9	10	3
Tyrosine	2870	3	4	6
Phenylalanine	2650	6	4	3
Proline	2600	4	8	4
Leucine	2420	18	22	2
Valine	1690	8	10	9
Lysine	1500	19	15	10
Methionine	1300	2	4	4
Alanine	730	17	14	11
Arginine	730	4	3	4
Threonine	440	5	8	10
Total number of residues		153	162	124
$-T\Delta S_{conf}$, kcal./mole		−184	−194	−149
$\Sigma\Delta f_u$, kcal./mole		+173	+192	+100

[a] A. B. Edmundson and C. H. W. Hirs, *Nature*, **190**, 663 (1961). [b] W. G. Gordon, J. J. Basch and E. B. Kalan, *J. Biol. Chem.*, **236**, 2908 (1961); K. A. Piez, E. W. Davie, J. E. Folk and J. A. Gladner, *ibid.*, **236**, 2912 (1961). [c] C. H. W. Hirs, W. H. Stein and S. Moore, *ibid.*, **235**, 633 (1960).

table compares the value of $\Sigma\Delta f_u$ to the value of $-T\Delta S_{conf}$, estimated on the basis of −1200 cal./residue, as discussed earlier. In comparing these figures, it should be recalled that the value for $T\Delta S_{conf}$ represents no more than an intelligent guess. It could well be in error by as much as a factor of two.[28] The value of $\Sigma\Delta f_u$ is less subject

to error, but it might be increased by perhaps 10% if we were to equate Δf_u values with Δf_t values for transfer from butanol to water rather than these for transfer from ethanol to water (*cf.* Table II) and if we were to include possible minor contributions from histidine, glutamate and perhaps other side chains.

It should also be kept in mind that the model we have used is an idealized one. It is not likely that all non-polar parts of the molecule are shielded from the solvent in an actual native structure, nor has it been established that the unfolded forms of globular proteins (as experimentally studied in denaturing solvents) are so completely unfolded as to permit free contact of all parts of the molecule with the solvent. The value of $\Sigma\Delta f_u$ for a real unfolding process might thus be appreciably less than the values obtained here. On the other hand, incomplete unfolding in the denatured state also implies a reduced value for the conformational entropy. Moreover, there may be some flexibility in the native conformation. Thus there are factors which would reduce $T\Delta S_{conf}$ also, and the idealization of the unfolding process which is inherent in the calculation is likely to introduce less error into the relative magnitudes of $\Sigma\Delta f_u$ and $T\Delta S_{conf}$, than into the magnitude of each factor alone.

Table III shows that $\Sigma\Delta f_u$ is indeed of the same order of magnitude as $T\Delta S_{conf}$ for myoglobin and β-lactoglobulin. In these proteins the stability of the native conformation in water can be explained, within the uncertainty inherent in the calculation, entirely on the basis of the hydrophobic interactions of the non-polar parts of the molecule.

Table III also shows, however, that $\Sigma\Delta f_u$ falls short of compensating for $T\Delta S_{conf}$ for unfolding of ribonuclease. This result is at least partly an artifact, arising from our failure to consider the restrictive influence of disulfide bonds (four of which are present in ribonuclease) on the randomness of the unfolded form. In other words, it is incorrect to compute ΔS_{conf} for this protein on the same basis as is used for the other two proteins. It is of interest in this connection that ribonuclease molecules with cleaved disulfide bonds (both oxidized and reduced) have in fact been shown to exist in an unfolded form in aqueous solution.[29,30]

Hydrogen Bonds and the Specificity of Protein Structure.—The foregoing calculation has shown that hydrophobic interactions alone can account for the stability in water of a protein conformation in which most non-polar groups avoid contact with the solvent. However, the model used for the calculation includes a provision that polar groups which find themselves in the interior of the native molecule are hydrogen-bonded to other such groups. No direct use has been made of this assumption, but it has been implicit in the calculation of Table III, where it was assumed that only the term $T\Delta S_{conf}$ of equation 1 was a major term favoring the unfolded form. If the native conformation had included polar groups which were not hydrogen-bonded to each other, then Δf_u for such groups would have been large and negative owing to the formation of hydrogen bonds to water in the unfolded form. An appreciable number of isolated polar groups within the globular structure would clearly make such a structure unstable.

It is thus evident that not just any globular structure with non-polar groups on the inside, and charged groups on the outside, will be stable. It is also necessary that most of the polar groups which find themselves on the inside of the globular structure form suitable hydrogen bonds. Since formation of hydrogen bonds is possible only for highly specific relative orientations of the participating groups, whereas formation of hydrophobic regions carries no such requirement, it is likely that the necessity for forming such hydrogen bonds is the crucial factor which limits the number of possible structures which will satisfy all of the conditions needed for stability of a compact conformation.

The question of what particular structure a given protein molecule will assume is therefore not answered by the analysis of this paper. What the paper does provide is an explanation for the instability of an unfolded conformation in water, and perhaps also an explanation for the prevalence of globular structures (rather than rod-shaped ones). It is clearly easiest to keep non-polar groups from contact with solvent if the area of the protein/solvent interface is minimized.

Electrostatic Interaction between Charged Groups.—The influence of electrostatic interaction between charged groups has been neglected in the foregoing treatment. It can be included by supposing that all of the preceding calculations apply to a hypothetical discharged state of the protein molecule, in which charges have been removed, but in which all other interactions remain the same. The value of ΔF for unfolding of the charged protein molecule then becomes

$$\Delta F = -T\Delta S_{conf} + \Sigma\Delta f_u + \Delta W_{el} \qquad (4)$$

where ΔW_{el} is the difference between the work of charging in the unfolded form and the work of charging in the native conformation. The latter is numerically much larger, so that $\Delta W_{el} \simeq -W_{el}$ for the native conformation.

The calculation of W_{el} has been discussed before in considerable detail.[31,32] It was shown that it is ordinarily a small quantity near the isoelectric *p*H of the protein, so that it has little effect on the calculations of this paper, as long as these are thought of as applying to protein molecules of relatively low charge.

Protein Denaturation.—We have shown that the crude theoretical treatment of the unfolding of proteins which was presented at the beginning of this paper can be used to account for the stability of a compact globular structure for proteins in aqueous solutions, when they are at room temperature, approximately in the isoelectric state. The underlying model must however be considered in part speculative, and it is therefore desirable to

(28) Dr. W. Kauzmann (personal communication) suggests that the value we have used, based on his original estimate, is more likely to be too large than too small.

(29) W. F. Harrington and J. A. Schellman, *Compt. rend. trav. lab. Carlsberg*, **30**, 21 (1956).

(30) C. B. Anfinsen, ref. 1, discussion on p. 47.

(31) C. Tanford in "Symposium on Protein Structure," A. Neuberger, ed., Methuen and Co., London, 1958.

(32) C. Tanford, ref. 19, chapters 7 and 8.

indicate that it can account not only for the stability of the native conformation, where it is experimentally found to be stable, but that it can also account for its instability under conditions where protein denaturation normally occurs.

To achieve this we must first decide on a reasonable figure for ΔF for the unfolding process in water at room temperature.

The calculations presented in this paper have been much too uncertain to permit a theoretical estimate of ΔF. Experimental estimates of this quantity are however available,[33] and they all suggest that ΔF is not very large, perhaps only 10,000 or 20,000 cal./mole for the smaller proteins at room temperature. Thus all changes in external conditions which decrease any of the terms of ΔF (equation 4) by as much as 10,000 or 20,000 cal. should lead to denaturation.

Denaturation by Acids and Bases.—This subject has been discussed elsewhere.[6,31,32] Presumably the major effect is on ΔW_{el} of equation 4. This factor will attain a magnitude of the order of $-20{,}000$ cal. when small proteins such as those of Table III are titrated to extreme acid or basic pH. Many experimental observations on denaturation of this kind are however still largely unexplained, such as the fact that unfolding of serum albumin by acid begins at a pH which is much closer to the isoelectric point than one would anticipate.

Denaturation by Urea, Guanidine and Other Organic Substances.—The addition of organic substances to approximately isoelectric aqueous protein solutions at room temperature must have its principal influence through an effect on the Δf_u terms of equation 1 or 4. When the protein is removed from its isoelectric region, however, an increase in W_{el} as the dielectric constant of the solvent is decreased may also be expected to be important.

To illustrate the effect of an organic additive on Δf_u, we can use the solubilities of amino acids in ethanol–water mixtures determined by Cohn, *et al.*[36] We have found that an ethanol concentration of 40% by volume is sufficient to drive the equilibrium between native and unfolded β-lactoglobulin completely towards the unfolded form at 25°.[37] Solubilities of leucine and valine, in the same medium and at the same temperature, indicate that Δf_t for transfer of leucine and valine side chains, from pure ethanol to 40% ethanol, is 1960 and 1340 cal./mole, respectively, compared with the values of 2420 and 1690 cal./mole for transfer from pure ethanol to water. The hydrophobic interaction free energy in 40% ethanol is thus clearly less than in water (about 20% less for both leucine and valine side chains). If a similar reduction applies to all hydrophobic interactions, then $\Sigma\Delta f_u$ will be reduced by about 20%, which, for β-lactoglobulin, makes ΔF more negative by about 40,000 cal./mole, which is more than enough to account for the unfolding of the protein in this solvent.

We are in the process of measuring amino acid solubilities in a number of denaturing solvents. Preliminary results indicate that 30% dioxane, which is about equally as effective as 40% ethanol in unfolding β-lactoglobulin, also changes Δf_t of non-polar amino acid side chains to about the same extent as 40% ethanol. Eight molar aqueous urea, which is also about as effective as 40% ethanol in unfolding β-lactoglobulin, also solubilizes non-polar side chains to nearly the same extent.[38] Our data indicate, however, that it is likely that a part of the stabilization of the unfolded protein in 8 *M* urea may be ascribed to favorable interaction between urea and some of the polar groups of the protein.

High Concentrations of Organic Solvent.—At high concentrations of organic solvents which form relatively few or relatively poor hydrogen bonds (*e.g.*, ethanol and dioxane), the free energy for unfolding becomes increasingly negative, and the unfolded form becomes increasingly stable relative to the native conformation. At the same time, however, internal hydrogen bonds become stronger,[39] and ion pair formation between charged groups should become favored. There is thus created the possibility for forming an entirely new ordered structure, in which the position of non-polar groups is unimportant, since Δf_u for these groups will now be small, but in which formation of intramolecular hydrogen bonds and intramolecular ion pairs is important. It is perhaps a structure of this kind which is formed in β-lactoglobulin[9,37] and ribonuclease[40] at high concentrations of organic solvents.

Thermal Denaturation in Water.—The denaturation of proteins in water by an increase in temperature implies that the enthalpy portion of ΔF is positive. This result is contrary to what the present theory would predict, as the enthalpy change associated with the Δf_u terms of equations 1 or 4 is zero or slightly negative. A possible explanation is that it may be erroneous to consider the configurational free energy change which results from the increased flexibility of unfolded polypeptide chains as simply an entropy term: the existence of potential energy barriers to rotation may introduce an enthalpy term also. An alternative explanation is that the positive enthalpy of unfolding may arise from the rearrangement of hydrogen bonds which accompanies unfolding.[41]

Our inability to explain thermal denaturation serves to indicate the limitations of the present paper. We have computed accurate values of Δf_u for the hydrophobic groups, but have been

(33) Such an estimate can be made for instance by observing ΔS and ΔH values for unfolding at elevated temperatures and then using these values to obtain ΔF at 25°. Another way is to use the pH at which acid or base denaturation first sets in to calculate the value of the electrostatic work of charging which is needed to reduce ΔF to zero. Estimates have been made in this way, for example, for chymotrypsinogen[34] and ribonuclease.[29,35]

(34) M. A. Eisenberg and G. W. Schwert, *J. Gen. Physiol.*, **34**, 583 (1951).

(35) Ref. 19, p. 516.

(36) E. J. Cohn, T. L. McMeekin, J. T. Edsall and J. H. Weare, *J. Am. Chem. Soc.*, **56**, 2270 (1934).

(37) C. Tanford and P. K. De, *J. Biol. Chem.*, **236**, 1711 (1961).

(38) P. L. Whitney and C. Tanford, *J. Biol. Chem.*, **237**, PC 1735 (1962).

(39) For example, Klotz and Franzen[24] find that the tendency for formation of intramolecular hydrogen bonds in dioxane is considerably stronger than it is in water.

(40) R. E. Weber and C. Tanford, *J. Am. Chem. Soc.*, **81**, 3255 (1959).

(41) The data of refs. 22 and 24 lead to opposite conclusions on this point.

hampered by the absence of quantitative knowledge of the other terms which affect ΔF for the unfolding process. The general theory, in the context of which our Δf_u terms have been placed, is useful as a framework for assessing the importance of hydrophobic interactions, but it is too crude to permit analysis of those aspects of protein denaturation which do not directly arise from hydrophobic interactions.

Acknowledgments.—The author is greatly indebted to Drs. J. T. Edsall, W. Kauzmann and I. M. Klotz for reading and criticizing an earlier draft of this paper. He also acknowledges the support of this work by research grant G-17477, from the National Science Foundation, and by research grant A-4576, from the National Institute of Arthritis and Metabolic Diseases, United States Public Health Service.

[CONTRIBUTION FROM THE COATES CHEMICAL LABORATORY, LOUISIANA STATE UNIVERSITY, BATON ROUGE 3, LOUISIANA]

Rate Equation for Adsorption of a Neutral Substance at a Metal–Electrolyte Interface

BY PAUL DELAHAY AND DAVID M. MOHILNER[1]

RECEIVED APRIL 9, 1962

An equation is derived for the rate of adsorption of a neutral substance at a metal–electrolyte interface for processes obeying the logarithmic Temkin isotherm. A recent investigation in this Laboratory has indicated that this isotherm is obeyed over a rather wide range of concentrations by a number of organic substances of varied structure. Adsorption kinetics is characterized by an *exchange rate* v^0 which is expressed in terms of a *standard rate constant* k^0, the activity of adsorbed species in solution a, the charge-dependent part ΔG^q of the standard free energy of adsorption ΔG^0, a *coverage parameter* (not the coverage) λ $(0 < \lambda < 1)$ and a *charge parameter* ρ $(0 < \rho < 1)$. Variations $\delta(\Delta G^q)$ with the charge density q on the electrode can be determined experimentally from the dependence of ΔG^0 on q. More conveniently $\delta(\Delta G^q)$ is obtained from the variations of the potential E with $\ln a$ at constant q. $(\partial E/\partial \ln a$ at constant q is the Esin and Markov coefficient.) The parameters λ and ρ are determined from $(\partial \ln v^0/\partial \ln a)$ at constant q and $(\partial \ln v^0/\partial \Delta G^q)$ at constant a. The adsorption rate is expressed in terms of v^0, the parameter b characterizing the isotherm, λ, ρ, the variation $\delta(\Delta G^q)$ of ΔG^q and the variation $\delta\Gamma$ of the surface concentration Γ which result from a change of q. Correction for mass transfer and for the double layer structure is indicated. The principle of a new *coulostatic* method for the measurement of v^0 is discussed, and correlation of the present theory with other methods for v^0 determination is outlined. The basic equation for adsorption rate reminds one of the Butler, Erdey-Gruz, Volmer equation for electrode kinetics; v^0 is the counterpart of the exchange current, and the charge parameter ρ is analogous to the transfer coefficient in electrode kinetics.

The thermodynamics of adsorption of a neutral substance at a metal–electrolyte interface, which is based on the Gibbs adsorption isotherm, is well understood but hardly anything is known about adsorption kinetics, except for purely diffusion controlled processes.[2,3] The kinetic problem is attacked here for adsorption of a neutral substance obeying the logarithmic Temkin isotherm,[4] and a basic equation is derived which expresses the rate of adsorption as a function of experimental quantities. This equation appears to be quite general as will be shown below and may serve as a basis for the development of adsorption kinetics at metal–electrolyte interfaces. The key ideas are: (i) In the thermodynamic analysis of adsorption at a metal–electrolyte interface, it is often convenient to choose the charge density q as the independent electrical variable rather than the electrode potential E.[5] The charge density q, rather than E, is the "natural" electrical parameter in the treatment of adsorption, whereas the opposite holds for electrode processes. (ii) It is inferred from recent work in this Laboratory that the logarithmic Temkin isotherm is approximately obeyed for a number of organic neutral substances of varied structure over a fairly wide concentration range. (iii) The equations given by Temkin[4] for the rates of adsorption and desorption of a gas on a solid apply (see below). (iv) The influence of the charge density on rates of adsorption and desorption can be expressed in terms of a charge parameter defined below and the charge-dependent part of the standard free energy of adsorption. This idea is novel, to our knowledge, and is the key to the following treatment. Any surface process subsequent to adsorption will be neglected or assumed to be sufficiently fast and consequently not rate-determining. A more general treatment of adsorption processes followed by a slow surface process is now being considered.

Rate Equation for the Adsorption–Desorption Process

Activities and Isotherm.—We represent the adsorption of a neutral substance O at a metal–electrolyte interface by

$$O + S = A \qquad (1)$$

where S is a site and A is the adsorbed species. This formalism is inspired from the application of order-disorder theory to adsorption.[6] We assume that the equations derived by Temkin[4] for the rates of adsorption $(\vec{v})$ and desorption $(\overleftarrow{v})$ apply, and we first derive expressions for the activities of the species of eq. 1. We shall then verify that the logarithmic Temkin isotherm is obeyed. The

(1) Postdoctoral fellow, 1960–1962.

(2) For a review, *cf.* R. Parsons, Chap. 1 in "Advances in Electrochemistry and Electrochemical Engineering," Vol. I, edited by P. Delahay, Interscience-Wiley, New York, N. Y., 1961, pp. 1–64.

(3) Adsorption kinetics with diffusion control has been worked out for linear and Langmuir isotherms. The main significance of this work resided in showing that diffusion-controlled adsorption can be slow. This point had been overlooked in a number of investigations, especially with the dropping mercury electrode.

(4) M. I. Temkin, *Zhur. fiz. Khim.*, **15**, 296 (1941); translation available. The logarithmic Temkin isotherm, often referred to as a "Temkin isotherm," is only a particular form of the more general equation derived by Temkin.

(5) R. Parsons, *Trans. Faraday Soc.*, **51**, 1518 (1955).

(6) For a review, *cf.*, *e.g.*, J. M. Honig, *J. Chem. Educ.*, **38**, 538 (1961).