The Solubility of Amino Acids and Related Compounds in Aqueous Ethylene Glycol Solutions*

YASUHIKO NOZAKI AND CHARLES TANFORD

From the Department of Biochemistry, Duke University Medical Center, Durham, North Carolina

(Received for publication, March 25, 1965)

It has been shown in a previous paper (1) that the unfolding of a globular protein can be treated theoretically as a process resembling solution of a solid substance. The groups on the inside of the native protein are “dissolved” in the solvent medium when unfolding occurs. Reagents which increase the solubility of these groups will increase the stability of the unfolded form of the protein and, if present in sufficiently high concentration, will cause unfolding to occur.

This approach to protein unfolding can be put into quantitative form. If $\Delta F_{w^0}$ is the standard free energy difference between the completely unfolded and the native forms of a protein in any medium, and $\Delta F_{w^0,H_2O}$ represents the same quantity in water or a dilute salt solution,

$$\Delta F_w = \Delta F_{w^0,H_2O} + \sum \alpha_i n_i \Delta f_{i,i} + \delta \Delta W$$

(1)

where $\Delta f_{i,i}$ is the contribution which any constituent group of the protein may be expected to make to the free energy of transfer of the protein from water to any other solvent medium when the group is in free contact with the solvent, $n_i$ is the number of such groups which the protein contains, and $\alpha_i$ is the fraction of these which may be expected to be newly exposed to the solvent during the unfolding process; i.e. it represents the fraction of all groups of type $i$ which are in the interior of the native structure (and thus inaccessible to solvent), but exposed to solvent after unfolding. The reason for the inclusion of the parameter $\alpha_i$ is that only those groups which are newly exposed in the reaction can contribute to the difference between $\Delta F_{w^0}$ and $\Delta F_{w^0,H_2O}$. The term $\delta \Delta W$ in Equation 1 represents the change in the contribution which long range electrostatic interactions make to the free energy of unfolding. It is considered to be an unimportant term as long as the protein has a relatively small net charge.

The summation in Equation 1 extends over all constituent groups of the protein. However, if the reaction to which the equation applies is a transition from the typical native conformation of globular proteins (as exemplified by myoglobin (2)) to a more or less completely unfolded, randomly coiled form, then only hydrophobic groups and peptide groups make an important contribution. Charged groups lie largely on the surface of native proteins, so that $\alpha_i$ for such groups is close to zero. Polar side chains (e.g. serine and threonine) also tend to lie near the surface of the native structure, and in addition tend to be important in their interactions with solvent, i.e. $\Delta f_{i,i}$ values for them are small.

It is possible to calculate $\Delta f_{i,i}$ values for the constituent groups of proteins by studying the solubility of amino acids and related small molecules to which these same groups are attached (3). It has been shown (3) that the interaction of hydrophobic groups with solvent is, at least in ethanol and urea solutions, not greatly affected by the nature of the molecule to which the groups are attached; i.e. $\Delta f_{i,i}$ values obtained from the study of different model compounds are closely similar. Such values can then be confidently expected to apply to the same groups when they are attached to protein molecules, and they can thus be used directly in Equation 1. The same statement cannot be made about peptide groups, which do appear to be affected by vicinal groups on the same molecule. However, the values of $\Delta f_{i,i}$ for peptide groups on different model compounds do have the same order of magnitude, so that a rough estimate of the value of $\Delta f_{i,i}$ appropriate for the peptide groups of protein can be made.

The term $\Delta F_{w^0,H_2O}$ which appears in Equation 1 cannot at present be estimated theoretically or on the basis of studies of model compounds. It must of course be positive for all proteins which are globular in their native state. Its value can be estimated in various ways from experimental studies of protein denaturation. One such estimate (1) gives a value in the range of 100 to 200 cal per residue for complete unfolding of a number of proteins.

We have shown in two previous papers (1, 3) that Equation 1 can account for the ability of urea to cause unfolding of proteins. It was found that urea increases the solubility of virtually all groups of protein molecules except charged groups; i.e. the important $\Delta f_{i,i}$ values all become progressively more negative as the urea concentration is increased. At sufficiently high urea concentration, therefore, $\Delta F_{w^0}$ becomes negative, and transition to the unfolded form of the protein occurs.

If one is willing to draw conclusions from the very limited data available for estimation of $\Delta f_{i,i}$ values for other denaturing agents, one can at least qualitatively understand why organic solvents such as ethanol and dioxane are good denaturing agents, for they greatly increase the solubility of nonpolar substances and thus lead to large negative $\Delta f_{i,i}$ values for hydrophobic groups (4). The only available data on guanidine hydrochloride also confirm the applicability of the general approach embodied in Equation 1. $N$-Acetyltryptaglycine ethyl ester, a compound which contains both peptide and hydrophobic groups, is readily solubilized by guanidine hydrochloride, which appears to be 2 to 3 times as effective as urea for this purpose (5). This result...
is in accord with the frequently made observation that guanidine hydrochloride is 2 to 3 times as effective as urea as a denaturing agent for proteins.

To establish completely the validity of our approach to the problem of protein unfolding, we must not only be able to account for the unfolding action of good denaturating agents but also show that reagents which are poor unfolding agents would be expected to act in this way when the same theoretical treatment is applied to them. This is the purpose of the present paper. It has been shown that ethylene glycol is singularly ineffective as an unfolding agent for γ-globulin and β-lactoglobulin (6). We have therefore measured solubilities of amino acids and related compounds in aqueous ethylene glycol to determine whether the Δfᵢ values derived from such data would lead to a predicted result which agrees with the experimental result. It should be noted that the test can in principle be made quite stringent inasmuch as Δfᵢ,ₓ (which was an unknown parameter in the theoretical treatment of unfolding by urea) in order to be evaluated on the basis of the experimental data, must now be regarded as a fixed constant, since the value determined from the urea data must apply here also if the reaction which one is considering is the same in the two solvent systems.

The results to be presented below confirm the applicability of Equation 1 to the action of ethylene glycol, despite the fact that there are experimental difficulties in the exact evaluation of Δfᵢ,ₓ values. They also suggest an interesting new possibility, namely, that the unfolding of a protein which occurs at very high glycol concentrations may not lead to a completely unfolded conformation, but rather to a structure in which intramolecular hydrogen bonds between peptide groups persist.

**EXPERIMENTAL PROCEDURE**

The amino acids and peptides used in this study were mostly the best available grade from Mann, except that tyrosine was a product of Calbiochem. Carbobenzoxy and benzoyl derivatives were purchased from Cyclo. For optically active molecules, the L isomer was always used.

Solubility data themselves can be used as criteria for purity, as described in our earlier paper (3). Some of the compounds studied appeared pure by this criterion as received, and they were used without further purification. L-Histidine, diglycine, triglycine, carbobenzoxytyrosine, carbobenzoxytriglycine, benzoylglycine, and benzoyl-L-phenylalanine, however, had to be recrystallized before satisfactory solubility curves could be obtained.

Ethylene glycol was a J. T. Baker "analyzed" reagent and was used without further purification. Water was glass-distilled and was boiled to remove CO₂.

Solubility measurements were carried out as described previously (3). Analysis of the saturated solutions for solute content was by a dry weight method, except for tyrosine and the carbobenzoxy and benzoyl derivatives. The solutions were dried in a vacuum at 35° until a constant weight was obtained. They were then heated in a forced air circulating oven at 107° to see if further loss of weight occurred. Usually there was no additional loss of weight, or only an insignificant amount. L-Phenylalanine and L-asparagine, however, appeared to be hydrates at 35°, as previously reported. L-Glutamine also contained a small amount of residual solvent after heating at 35°, which was lost on heating to 107°. L-Tryptophan gave a residue with a slight yellow color when dried from ethylene glycol solutions.

The concentration of saturated solutions of L-tyrosine and of the carbobenzoxy and benzoyl derivatives was determined spectrophotometrically, with the use of peak wave lengths in the ultraviolet. The molar absorbances were determined from measurements on solutions of known concentrations. Both molar absorbances and peak wave lengths varied somewhat with ethylene glycol concentration.

All measurements were carried out at 25.1° in water and at three concentrations of ethylene glycol: 30, 60, and 90 ml/100 ml of solvent mixture. They will be designated as 30%, 60%, and 90% glycol, respectively. The solvent compositions in other units are 32.0, 61.8, and 90.5% by weight, or 5.36, 10.7, and 10.1 ml, respectively.

When solutions of benzoyl-L-tryptophan are shaken in 60 and 90% glycol, a purple color is slowly formed, corresponding to a new absorption band with a peak near 540 μ. Appearance of this color did not appear to affect the measured solubilities or the absorbance at 280 μ, the wave length at which the concentration was determined.

**RESULTS**

The solubilities determined in this study are summarized in Table I. As was shown in our previous paper (3), these data may be used to calculate the free energy of transfer (ΔFᵢ) of the solute from water to an ethylene glycol solution, at the same

<table>
<thead>
<tr>
<th>Solute</th>
<th>Solubilities at ethylene glycol concentration of 0%</th>
<th>30%</th>
<th>60%</th>
<th>90%</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>g/100 g solvents</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glycine</td>
<td>25.09</td>
<td>11.20</td>
<td>4.25</td>
<td>1.41</td>
</tr>
<tr>
<td>Alanine</td>
<td>16.67</td>
<td>7.95</td>
<td>3.48</td>
<td>1.32</td>
</tr>
<tr>
<td>Leucine</td>
<td>2.16</td>
<td>1.32</td>
<td>0.81</td>
<td>0.63</td>
</tr>
<tr>
<td>Phenylalanine</td>
<td>2.79</td>
<td>1.97</td>
<td>1.78</td>
<td>1.36</td>
</tr>
<tr>
<td>Tryptophan</td>
<td>1.38</td>
<td>1.60</td>
<td>2.50</td>
<td>4.90</td>
</tr>
<tr>
<td>Methionine</td>
<td>5.59</td>
<td>3.11</td>
<td>1.90</td>
<td>1.29</td>
</tr>
<tr>
<td>Threonine</td>
<td>9.80</td>
<td>4.41</td>
<td>1.76</td>
<td>0.65</td>
</tr>
<tr>
<td>Tyrosine</td>
<td>0.0451</td>
<td>0.0365</td>
<td>0.0246</td>
<td>0.0248</td>
</tr>
<tr>
<td>Histidine</td>
<td>4.30</td>
<td>2.59</td>
<td>1.48</td>
<td>0.73</td>
</tr>
<tr>
<td>Asparagine</td>
<td>2.51</td>
<td>1.36</td>
<td>1.06</td>
<td>0.708</td>
</tr>
<tr>
<td>Glutamime</td>
<td>4.15</td>
<td>2.00</td>
<td>0.79</td>
<td>0.27</td>
</tr>
<tr>
<td>Diglycine</td>
<td>22.75</td>
<td>8.60</td>
<td>2.65</td>
<td>0.73</td>
</tr>
<tr>
<td>Triglycine</td>
<td>6.40</td>
<td>2.17</td>
<td>0.65</td>
<td>0.15</td>
</tr>
<tr>
<td>Carbobenzoxyglu-</td>
<td>cine</td>
<td>0.450</td>
<td>0.795</td>
<td>2.08</td>
</tr>
<tr>
<td>Carbobenzoxytry-</td>
<td>sine</td>
<td>0.153</td>
<td>0.68</td>
<td>&gt;94</td>
</tr>
<tr>
<td>Carbobenzoxydig-</td>
<td>lyccine</td>
<td>0.065</td>
<td>0.075</td>
<td>0.16</td>
</tr>
<tr>
<td>Carbobenzoxytrig-</td>
<td>line</td>
<td>0.11</td>
<td>0.12</td>
<td>0.22</td>
</tr>
<tr>
<td>Benzyllglycine</td>
<td>0.367</td>
<td>0.537</td>
<td>1.175</td>
<td>2.99</td>
</tr>
<tr>
<td>Benzylophenylala-</td>
<td>nine</td>
<td>0.085</td>
<td>0.19</td>
<td>1.00</td>
</tr>
<tr>
<td>Benzoyltryptophan</td>
<td>0.056</td>
<td>0.21</td>
<td>2.42</td>
<td>~25</td>
</tr>
<tr>
<td>Benzylytrosine</td>
<td>0.268</td>
<td>1.19</td>
<td>11.2</td>
<td>&gt;42</td>
</tr>
<tr>
<td>Benzyldiglycine</td>
<td>0.33</td>
<td>0.39</td>
<td>0.68</td>
<td>1.30</td>
</tr>
</tbody>
</table>

Downloaded from http://www.jbc.org/ by guest on November 7, 2017
mole fraction of solute, at the limit of infinite dilution of the solute. The equation is

$$\Delta F_i = RT \ln \frac{N_{i,w}}{N_{i,EG}} + RT \ln \frac{\gamma_{i,w}}{\gamma_{i,EG}}$$

where \(N_{i,w}\) is the mole fraction of solute at saturation in water and \(N_{i,EG}\) the corresponding parameter in a glycol solution. The \(\gamma\) values are activity coefficients on a scale which makes \(\gamma_i = 1\) at infinite dilution of solute in each of the solvent media, and thus represent solely the effect of solute-solute interaction. The term \(RT \ln \frac{\gamma_{i,w}}{\gamma_{i,EG}}\) takes into account the fact that solubilities are affected when the solvent is changed, not only by interaction between solute and solvent (as represented by \(\Delta F_i\)) but also by changes in the interaction of solute molecules with each other. The activity coefficient term is unimportant when the solubility is low, but can become quite large when the solubility is high since experiments are then necessarily conducted under conditions in which solute-solute interaction can become important.

Activity coefficients for some of the amino acids and for diglycine and triglycine, in water solution, have been determined, so that \(\gamma_{i,w}\) for these solutes is known. Reasonable estimates can be made for those amino acids for which direct data are not available (1). The values of \(\gamma_{i,EG}\), however, have to be estimated without experimental data. We have made the same assumption as in our earlier paper (3), namely, that \(\gamma_{i,EG}\) has the same value of \(\gamma_{i,w}\) at the same molal concentration. For glycine, diglycine, and triglycine we have allowed for an additional factor. The solute-solute interactions for these substances are likely to be chiefly electrostatic. The free energy of interaction should thus depend inversely on the dielectric constant of the solution, and we have therefore multiplied log \(\gamma_{i,EG}\) (calculated on the basis of the equal molality assumption) by the ratio of dielectric constants, \(D_w/D_{EG}\), \(D_{EG}\) being the dielectric constant of the aqueous glycol solution for which \(\gamma_{i,EG}\) is being calculated. A similar correction was not applied to the activity coefficients of the other amino acids, because log \(\gamma_{i,w}\) for most of them is positive rather than negative, so that electrostatic forces are not likely to be a prime factor.

No activity coefficient data are available for the carbobenzyo and benzoyl derivatives. Since these are uncharged molecules, we have no basis for estimating even the sign of the activity coefficients, and we have thus been forced to neglect the activity coefficient term entirely.

All estimates of the activity coefficient term of Equation 2 are summarized in Table II. It is seen that much larger values are obtained (for glycine, diglycine, and triglycine) than were obtained in aqueous urea solutions. The reason for this is that these substances are very soluble in water, which leads to large numerical values of log \(\gamma_{i,w}\), but their solubilities decrease rapidly as glycol is added, so that log \(\gamma_{i,EG}\) values are smaller. In urea, on the other hand, these substances maintain their high solubilities. It should be noted that the major contribution to the data of Table II comes from the \(\gamma_{i,w}\) term, and that the data of Smith and Smith (7), which we have used, have recently been confirmed for several of the solutes (including glycine and diglycine) by Ellerton et al. (8).

Values of \(\Delta F_i\) calculated from the experimental data by Equation 2 with the aid of Table II are shown in Table IV.

In our previous paper (3), we defined the contribution which an amino acid side chain makes to \(\Delta F_i\) as the effect on \(\Delta F_i\) of substituting the side chain for a hydrogen atom. The side chain contribution (\(\Delta F_i\)) is thus obtained simply as the difference between \(\Delta F_i\) for the amino acid and \(\Delta F_i\) for glycine. We also defined the contribution which a peptide group makes to \(\Delta F_i\) as the effect of inserting a glycyl residue into another molecule, i.e. as a quantity which can be calculated as the difference between \(\Delta F_i\) values of diglycine and glycine or of triglycine and diglycine. \(\Delta F_i\) values calculated in this way are shown in Table IV.

Contributions to \(\Delta F_i\) of individual parts of the solute molecule, calculated in this way, have generally applicable physical significance only if the individual parts of the molecule interact with the solvent independently, i.e. if the effect of one part of the molecule on another, with respect to this property, is quite small. Whether or not this favorable situation exists is subject to experimental test, for, if it does exist, \(\Delta F_i\) values calculated for a particular group by substitution for a hydrogen atom of glycine should not differ greatly from \(\Delta F_i\) values calculated for the same group by substitution for a hydrogen atom of a different molecule. It has been shown (3) that such comparisons do indeed often lead to closely similar \(\Delta F_i\) values for amino acid side chains, but that unique values for peptide groups, and for amide groups on side chains, cannot be obtained, at least for aqueous urea solutions.

We have put this question to a severe test in the present studies by measuring solubilities for a few carbobenzyo and benzoyl derivatives of the amino acids and peptides. These derivatives not only contain very bulky added groups but also change the solute molecules from dipolar ions to uncharged molecules with very different solubility properties. The solubility of glycine in 90% glycol, for example, as shown in Table I, is only about 5% of the solubility in water. On the other hand, carbobenzyglycine is 26 times as soluble in 90% glycol as in water, and benzoylglycine 8 times as soluble. Nevertheless, the \(\Delta F_i\) values for two aromatic side chains calculated from the solubilities of these derivatives agree quite closely with those obtained from the amino acids themselves, as shown in Table V.

### Table II

**Values of RT ln (\(\gamma_{i,w}/\gamma_{i,EG}\)) for saturated solutions at 25°**

For carbobenzyo and benzoyl derivatives, all of which are uncharged, activity coefficients have been assumed close to unity at all glycol concentrations.

<table>
<thead>
<tr>
<th>Solute</th>
<th>Values at glycol concentrations of</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>30%</td>
</tr>
<tr>
<td>Glycine</td>
<td>35</td>
</tr>
<tr>
<td>Alanine</td>
<td>+25</td>
</tr>
<tr>
<td>Leucine</td>
<td>+5</td>
</tr>
<tr>
<td>Phenylalanine</td>
<td>+10</td>
</tr>
<tr>
<td>Tryptophan</td>
<td>0</td>
</tr>
<tr>
<td>Methionine</td>
<td>+15</td>
</tr>
<tr>
<td>Threonine</td>
<td>-5</td>
</tr>
<tr>
<td>Tyrosine</td>
<td>0</td>
</tr>
<tr>
<td>Histidine</td>
<td>+15</td>
</tr>
<tr>
<td>Asparagine</td>
<td>-10</td>
</tr>
<tr>
<td>Glutamine</td>
<td>+5</td>
</tr>
<tr>
<td>Diglycine</td>
<td>-65</td>
</tr>
<tr>
<td>Triglycine</td>
<td>-70</td>
</tr>
</tbody>
</table>
although all data agree in indicating that $\Delta f_{i}$ is large and negative for hydrophobic side chains. However, as is seen in Table V, there is qualitative agreement between the $\Delta f_{i}$ values obtained. They are always positive, whereas in urea $\Delta f_{i}$ values for the peptide group are always negative. It may be noted that $\Delta f_{i}$ value for the side chain of asparagine is shown in Table V.

Equally good agreement is not obtained for the tyrosyl side chain, benzoyl substituents than are other side chains. This is especially significant as aromatic side chains are more likely to be influenced by the aromatic groups on the carbobenzoxyl and benzoyl substituents than are other side chains. Equally good agreement is not obtained for the tyrosyl side chain, although all data agree in indicating that $\Delta f_{i}$ is large and negative in sign. Since we do not know the magnitude of the error introduced by neglect of the activity coefficient term of Equation 2 for carbobenzoxyl and benzoyl derivatives, and since we have chosen to carry out the test on compounds more likely than others to produce anomalous results, we take these data as a whole to indicate that no serious error is likely to result from assuming the chosen to carry out the test on compounds more likely than others to produce anomalous results, we take these data as a whole to indicate that no serious error is likely to result from assuming the constant molarity.

\begin{table}[h]
\centering
\caption{Free energies of transfer ($\Delta F_{i}$) from water to aqueous ethylene glycol solutions at 25\textdegree C.} \\
\begin{tabular}{|c|c|c|c|}
\hline
Solute & $\Delta F_{i}$ at glycol concentration of & 30\% & 60\% & 90\% \\
\hline
Glycine & 275 & 605 & 950 \\
Alanine & 305 & 910 & 990 \\
Leucine & 145 & 250 & 135 \\
Phenylalanine & 60 & -60 & -155 \\
Tryptophan & -240 & -705 & -1395 \\
Methionine & 210 & 315 & 280 \\
Threonine & 310 & 690 & 980 \\
Tyrosine & -30 & -185 & -460 \\
Histidine & 190 & 310 & 405 \\
Asparagine & 195 & 155 & 120 \\
Glutamine & 285 & 645 & 1010 \\
Diglycine & 350 & 775 & 1230 \\
Triglycerine & 415 & 890 & 1470 \\
Carbobenzoxyglycine & -480 & -1390 & -2540 \\
Carbobenzoxytyrosine & -1035 & & \\
Carbobenzoxydiglycine & -240 & -875 & -1600 \\
Carbobenzoxytryglycine & -210 & -760 & -1450 \\
Benzoylglycine & -300 & -1030 & -1850 \\
Benzoylphenylalanine & -630 & -1800 & -3220 \\
Benzoyltryptophan & -940 & -2570 & -4200 \\
Benzoyltyrosine & -850 & -2280 & \\
Benzoyldiglycine & -250 & -770 & -1420 \\
\hline
\end{tabular}
\end{table}

\begin{table}[h]
\centering
\caption{Contributions of amino acid side chains and of peptide groups ($\Delta f_{i}$) to free energy of transfer ($\Delta f$)} \\
\begin{tabular}{|c|c|c|}
\hline
Solute & $\Delta f$ at glycol concentrations of & 30\% & 60\% & 90\% \\
\hline
Alanine & -130 & -355 & -815 \\
Leucine & -215 & -665 & -1115 \\
Phenylalanine & -515 & -1310 & -2345 \\
Tryptophan & -65 & -200 & 670 \\
Tyrosine & -305 & -790 & -1410 \\
Histidine & -115 & -200 & -485 \\
Asparagine & -80 & -450 & -830 \\
Glutamine & 10 & 40 & 60 \\
Peptide group & 175 & 170 & +280 \\
Peptide group & 155 & 115 & +240 \\
\hline
\end{tabular}
\end{table}

\begin{table}[h]
\centering
\caption{Side chain or peptide contributions for different parent molecules} \\
\begin{tabular}{|c|c|c|c|}
\hline
Solute & $\Delta f$ at glycol concentrations of & 30\% & 60\% & 90\% \\
\hline
Phenylalanine & -215 & -665 & -1115 \\
Benzoylphenylalanine, benzoylglycerine & -950 & -770 & -1370 \\
Tryptophan, glycine & -515 & -1310 & -2345 \\
Benzoyltryptophan, benzoylglycerine & -560 & -1540 & -2350 \\
Tyrosine, glycine & -305 & -790 & -1410 \\
Benzoyltyrosine, benzoylglycerine & -470 & -1350 & \\
Carbobenzoxytyrosine, carboxyglycine & -555 & & \\
\hline
\end{tabular}
\end{table}

As was found in our studies with urea, solvent interactions of peptide groups are not independent of vicinal groups to the same extent as those of hydrophobic side chains. However, as is
(Table IV) is highly anomalous, since one expects the —\(\text{CH}_2\)CONH\(_2\) group to show some similarity to the peptide group. Asparagine also has a surprisingly large negative \(\Delta f_{i, i}\) value for transfer to urea. Experiments to investigate the possible reasons for these anomalies are in progress. Since asparagine side chains probably make little contribution to the free energy of unfolding of proteins (Equation 1), the question is not important for the present paper.

**DISCUSSION**

The results obtained in this study show that ethylene glycol is much less effective than urea for solubilizing the constituent groups of protein molecules. This is shown in Table VI, where \(\Delta f_{i, i}\) values for 6 M urea (32% by weight) are compared with those for 30% glycol (5.4 M = 32% by weight). The values for hydrophobic side chains are considerably smaller, and those for peptide groups are even of opposite sign. Since the ability of urea to cause unfolding of proteins, to yield a more or less random structure in which all or nearly all constituent groups are exposed to solvent, depends in about equal part on solubilization of hydrophobic side chains and peptide groups (1), it is evident that ethylene glycol cannot promote a similar unfolding process when its concentration is about the same as the concentration of urea which is required for unfolding. This result agrees with experiment. For example, \(\beta\)-lactoglobulin is unfolded by urea in a sharp transition which has its midpoint near 6 M (9). In ethylene glycol the native conformation is stable to well above 60% glycol concentration (6).

An attempt to calculate whether high concentrations of ethylene glycol should lead to unfolding must be regarded at the present state of our knowledge as speculation, rather than meaningful prediction. We do not in general know what fraction of any kind of group is buried in the interior of the native protein and newly exposed in the unfolding process; i.e. we do not know the \(\alpha_i\) values to be used in Equation 1. Furthermore, it is obvious from the data (Table V) that \(\Delta f_{i, i}\) values for peptide groups of proteins and for at least some amino acid side chains are still very uncertain. The values obtained from our solubility data on diglycine and triglycine, for example, are not necessarily applicable to the peptide groups of proteins. Nevertheless, we have made two such calculations in Table VII, to illustrate the influence which various factors have on the result. The calculations apply to the unfolding of \(\beta\)-lactoglobulin at pH 3, and we have followed exactly the same procedure as was used in the similar calculations applicable to the unfolding of \(\beta\)-lactoglobulin by urea (1). Two of the possible choices of the \(\alpha_i\) factors have been used, and for each one we have used the \(\Delta F_{\mu, \mathrm{H}_2\mathrm{O}}\) value which is calculated for that choice of the \(\alpha_i\) from the experimentally determined urea concentration at the midpoint of the transition as it occurs in aqueous urea solutions.

The results of the calculation show the influence of the positive values of \(\Delta f_{i, i}\) for peptide groups in opposing the unfolding process. (It should be noted that this effect would have been larger if we had used the even larger values of \(\Delta f_{i, i}\) derived in Table V from solubilities of carbobenzoxy- and benzoyldiglycine.) Because of this effect, one obtains a different answer to the question which was posed, depending on the choice of the \(\alpha_i\). If it is assumed that a relatively small fraction of the peptide groups must be newly exposed in the unfolding process (\(\alpha_i = 0.5\), Column 1), then transition to the unfolded form is predicted at a glycol concentration of about 80%. If a somewhat larger fraction of the peptide groups is required to be exposed (\(\alpha_i = 0.75\), Column II), the unfolded form would remain unstable even at 90% glycol.

**TABLE VI**

*Average value based on \(N\)-acetyl-l-tetraglycine ethyl ester. Average value based on glycine, diglycine, and triglycine is \(-170\) cal per mole.*

**TABLE VII**

Free energy of unfolding of \(\beta\)-lactoglobulin

**DISCUSSION**

The results obtained in this study show that ethylene glycol is much less effective than urea for solubilizing the constituent groups of protein molecules. This is shown in Table VI, where \(\Delta f_{i, i}\) values for 6 M urea (32% by weight) are compared with those for 30% glycol (5.4 M = 32% by weight). The values for hydrophobic side chains are considerably smaller, and those for peptide groups are even of opposite sign. Since the ability of urea to cause unfolding of proteins, to yield a more or less random structure in which all or nearly all constituent groups are exposed to solvent, depends in about equal part on solubilization of hydrophobic side chains and peptide groups (1), it is evident that ethylene glycol cannot promote a similar unfolding process when its concentration is about the same as the concentration of urea which is required for unfolding. This result agrees with experiment. For example, \(\beta\)-lactoglobulin is unfolded by urea in a sharp transition which has its midpoint near 6 M (9). In ethylene glycol the native conformation is stable to well above 60% glycol concentration (6).

An attempt to calculate whether high concentrations of ethylene glycol should lead to unfolding must be regarded at the present state of our knowledge as speculation, rather than meaningful prediction. We do not in general know what fraction of any kind of group is buried in the interior of the native protein and newly exposed in the unfolding process; i.e. we do not know the \(\alpha_i\) values to be used in Equation 1. Furthermore, it is obvious from the data (Table V) that \(\Delta f_{i, i}\) values for peptide groups of proteins and for at least some amino acid side chains are still very uncertain. The values obtained from our solubility data on diglycine and triglycine, for example, are not necessarily applicable to the peptide groups of proteins. Nevertheless, we have made two such calculations in Table VII, to illustrate the influence which various factors have on the result. The calculations apply to the unfolding of \(\beta\)-lactoglobulin at pH 3, and we have followed exactly the same procedure as was used in the similar calculations applicable to the unfolding of \(\beta\)-lactoglobulin by urea (1). Two of the possible choices of the \(\alpha_i\) factors have been used, and for each one we have used the \(\Delta F_{\mu, \mathrm{H}_2\mathrm{O}}\) value which is calculated for that choice of the \(\alpha_i\) from the experimentally determined urea concentration at the midpoint of the transition as it occurs in aqueous urea solutions.

The results of the calculation show the influence of the positive values of \(\Delta f_{i, i}\) for peptide groups in opposing the unfolding process. (It should be noted that this effect would have been larger if we had used the even larger values of \(\Delta f_{i, i}\) derived in Table V from solubilities of carbobenzoxy- and benzoyldiglycine.) Because of this effect, one obtains a different answer to the question which was posed, depending on the choice of the \(\alpha_i\). If it is assumed that a relatively small fraction of the peptide groups must be newly exposed in the unfolding process (\(\alpha_i = 0.5\), Column 1), then transition to the unfolded form is predicted at a glycol concentration of about 80%. If a somewhat larger fraction of the peptide groups is required to be exposed (\(\alpha_i = 0.75\), Column II), the unfolded form would remain unstable even at 90% glycol.
These results suggest that a partially unfolded form of β-lactoglobulin, in which hydrophobic groups are exposed to the solvent but peptide groups are not, might be more stable in concentrated ethylene glycol than a completely unfolded form. A conformation in which much of the polypeptide backbone is in the form of an α helix, but which has no other noncovalent intramolecular interactions, would be an example. Such a conformation would probably be unstable in water, relative to a completely unfolded molecule, because it would have a reduced configurational entropy (10) and because hydrogen bonds between peptide groups and water are likely to have a free energy which is lower rather than higher than the free energy of hydrogen bonds in a helical polypeptide chain (11, 12). In other words, ApU,Hzo groups and water are likely to have a free energy which is lower than a completely unfolded form. A conformation which one obtains from optical rotatory dispersion data (6).

Table VII. It is evident, however, that ΔF°,H2O could be as high as 50 Kcal per mole and the transition would still be favored at 90% ethylene glycol if 75% of the hydrophobic groups were to be exposed to solvent and no net change were to occur in the exposure of peptide groups. (The same value of ΔAW is likely to apply to any unfolding process in which appreciable separation of charges occurs.)

The experimental data on the effect of ethylene glycol on the conformation of β-lactoglobulin in fact suggest that a form with high helix content occurs at high glycol concentrations, as there is a sharp rise in the anomalous dispersion parameter (−b0) which one obtains from optical rotatory dispersion data (6). Sage and Singer (13) have come to a similar conclusion in their study of the unfolding of ribonuclease by ethylene glycol, and have suggested that peptide hydrogen bonds remain intact in this process.

It is of interest in this connection that ΔF°, values for the peptide group in aqueous ethanol and dioxane, as calculated from solubility studies now in progress, are even more positive than those obtained here. It would seem imperative, on the basis of these data, that proteins in these solvents acquire conformations in which contact between peptide groups and solvent is avoided. This result is in accord with the experimental observation that β-lactoglobulin, in ethanol and dioxane solutions, is transformed to a conformation with a very high helix content (9, 14). The high helix content of most proteins in 2-chloroethanol (15) presumably has the same underlying cause.

SUMMARY

The solubilities of several amino acids, peptides, and carbobenzoxy and benzoyl derivatives have been measured in water and at three concentrations of ethylene glycol. The data show that ethylene glycol is less effective than urea in reducing the free energy of hydrophobic interactions between nonpolar groups and water. Moreover, ethylene glycol increases the free energy of contact between peptide groups and the solvent, in contrast to urea, which decreases it.

These results, when incorporated into the equation for the free energy of unfolding of proteins, which was derived in an earlier paper, show that ethylene glycol must be a much less effective denaturing agent than urea. The data suggest that when denaturation does occur (at very high glycol concentrations), the product is likely to be an incompletely unfolded molecule in which hydrogen bonds between peptide groups are maintained. These conclusions agree with experimental data and serve to validate the thermodynamic approach to the understanding and prediction of protein conformation which is being developed in this and other laboratories.

Acknowledgment—We wish to acknowledge the patient and skillful technical assistance of Mrs. Pearlie M. Hudson.

REFERENCES

The Solubility of Amino Acids and Related Compounds in Aqueous Ethylene Glycol Solutions
Yasuhiro Nozaki and Charles Tanford


Access the most updated version of this article at http://www.jbc.org/content/240/9/3568.citation

Alerts:
- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

This article cites 0 references, 0 of which can be accessed free at http://www.jbc.org/content/240/9/3568.citation.full.html#ref-list-1