The Solubility of Amino Acids and Two Glycine Peptides in Aqueous Ethanol and Dioxane Solutions

ESTABLISHMENT OF A HYDROPHOBICITY SCALE*

YASUHIKO NOZAKI AND CHARLES TANFORD

From the Department of Biochemistry, Duke University Medical Center, Durham, North Carolina 27706

SUMMARY

The solubilities of amino acids, diglycine, and triglycine have been measured in water and aqueous ethanol as well as dioxane solutions. Free energies of transfer of amino acid side chains and backbone peptide units from water to ethanol and dioxane solutions have been calculated from these data. The results show the similarity between the effects of ethanol and dioxane on the stability of those side chains and peptide units. In particular, the free energies of transfer of hydrophobic side chains to 100% ethanol and dioxane are essentially identical, and have been used to establish a hydrophobicity scale for hydrophobic side chains.

The solubility study of amino acids and two glycine peptides reported in this paper represents the fourth of a series of such studies. Earlier papers have dealt with solubilities in aqueous urea (1, 2), ethylene glycol (3), and guanidine hydrochloride solutions (4). The present investigation covers solubilities in aqueous ethanol and dioxane solutions. The measurements in ethanol solutions partly duplicate earlier work by Dunn and Ross (5), Cohn et al. (6), and McMeekin, Cohn, and Weare (7, 8). (A comprehensive analysis of the earlier data is given by Cohn and Edsall (9).) No previous measurements in dioxane solutions have been reported.

A major purpose of all our previous studies has been to provide a rational explanation for the ability of various solvent media to promote protein denaturation. The process of denaturation is accompanied by the exposure of protein moieties that are in the "inside" of the native globular structure. If the freshly exposed moieties interact more favorably with a particular solvent medium than they do with water, that solvent medium will promote denaturation. Solubility measurements of model compounds that are similar except for the presence of a particular group in the one compound and its absence in the other can yield, after suitable thermodynamic manipulations and with simplifying assumptions, quantitative values for the free energy of interaction

of such a group with a particular solvent, relative to the free energy of interaction with water, and thus allow quantitative predictions of the denaturing ability of any solvent relative to the denaturing ability of water. No knowledge of the interaction free energies within the native protein is necessary. The theory and its application to denaturation equilibria have been discussed elsewhere (10, 11). Application of some of the present data to denaturation by ethanol and dioxane has already been briefly treated (11).

Solubility measurements in ethanol and dioxane serve a broader purpose, in that they provide a semiquantitative estimate of interactions inside the native protein molecule for hydrophobic moieties, relative to the interaction of the same moieties when exposed to water. The rationale here is that hydrocarbon groups and other hydrophobic moieties have a strongly unfavorable interaction with water, which is abolished when the group in question is removed from contact with water and placed inside the native structure, where such groups are not generally involved in any strong favorable or unfavorable interactions with their environment. It has been suggested that 100% ethanol can serve as a "model" for the inside of a protein molecule in this regard (12), i.e. that the difference in free energy between a hydrophobic moiety in 100% ethanol and in water is simply a measure of the unfavorable free energy of interaction with water, nonspecific (van der Waals) interactions having about the same free energy in 100% ethanol as in the interior of a protein molecule. Strong support for this idea comes from the finding that the free energy of solvent contacts of one amino acid side chain (norleucine) differs very little between methanol, ethanol, butanol, and acetone, in contrast to the large difference between any one of these solvents and water (9). There is also evidence that differences between free energies of hydrophobic moieties in 100% ethanol and the interior of detergent micelles are small (13). It is evident that the results of this paper, since they provide comparable data between ethanol and dioxane, will provide additional information on this question.

The foregoing principle cannot of course be applied to molecular moieties containing polar groups. Such groups are likely to have specific interactions with hydroxyl, carbonyl, and other polar groups of their environment, and, hence, unlikely to have similar contact free energies in, say, ethanol, dioxane, and the inside of a protein molecule. Our studies of amino acids with polar side chains will provide some information on this point.

For those amino acid side chains that appear to be predomi-

nantly hydrophobic (*i.e.* have similar contact free energies in a variety of media other than water, and a large positive contact free energy in water, relative to all such other media) the difference between solvation free energies in water and in 100% ethanol or other nonaqueous media may be considered as establishing a *hydrophobicity scale*, which can be regarded, for example, as a measure of the tendency of that side chain to be located in the interior of a native protein molecule (14). The present paper permits tryptophan and histidine side chains to be located on such a scale. Data for these were not previously available, although it has been guessed that the tryptophan side chain would turn out to be the most hydrophobic of all protein side chains, as indeed it will be shown to be.

EXPERIMENTAL PROCEDURE

All amino acid samples used are L isomers except for glycine and its peptides. L-Tyrosine and L-o-dihydroxyphenylalanine were purchased from Calbiochem, and triglycine from Cyclo Chemical Corporation. All other samples were purchased from Mann.

L-Histidine was prepared from its hydrochloride as described earlier (2). Diglycine was recrystallized twice from water. Triglycine was recrystallized twice from water and once from aqueous ethanol solution. Other samples were used without further purification.

Ethanol was purified by the method of Lund (15). p-Dioxane was Spectroquality Reagent by Matheson, Coleman and Bell and used without further purification. Water was tin and glass doubly distilled and freed from CO_2 by boiling.

The concentrations of ethanol and dioxane are expressed in terms of the volume (milliliters) of ethanol or dioxane in 100 ml of their aqueous mixtures.

Solubility Measurements—Samples and solvents were weighed directly in glass tubes with ground glass caps and Teflon sleeves. The air above the mixture was replaced by nitrogen. Five such tubes were prepared routinely, containing various amounts of sample, in such a way that two tubes would contain less than and

three others more than enough to saturate the solvent. The tubes were shaken normally for 24 hours, sometimes longer, when need arose, in a water bath maintained at 25.1 ± 0.05°. The solutions were separated from the saturating bodies by filtration through long stem glass filters, while the tubes were kept partially immersed in the bath. The supernatants thus obtained were assayed by a dry weight method, by evaporating in a vacuum desiccator at about 35° first and in a forced air-circulating oven at 107° until a constant weight was obtained. For glutamine, which undergoes a thermal decomposition on prolonged heating at 107°, the weights in a plateau region of drying curves obtained within several hours of heating at 107° were used for calculation. Tyrosine and dihydroxyphenylalanine were assayed spectrophotometrically at 275 and 280 mμ, respectively.

RESULTS

The solubilities determined in this study are listed in Tables I and II. With these data we have calculated the free energies of transfer of the solutes from water to aqueous ethanol and dioxane solutions. The free energy of transfer (ΔF_t) has been defined as the change in the chemical potential of solute i in going from water to any other solvent, at the same mole fraction, at the limit of infinite dilution. Since the chemical potential of solute i, at saturation, is the same in all solvent media, being equal to that of the saturating body,

$$\mu_{i,w}^0 + RT \ln N_{i,w} + RT \ln \gamma_{i,w}$$
$$= \mu_i^0 + RT \ln N_i + RT \ln \gamma_i \quad (1)$$

where $\mu_{i,w}^0$ and μ_i^0 are the standard chemical potentials in water and in the solvent to which transfer occurs, respectively; $N_{i,w}$ and N_i are the solubilities in terms of mole fraction; and $\gamma_{i,w}$ and γ_i are the activity coefficients at saturation, defined so as to represent the effect of solute-solute self-interaction only, solute-solvent interaction at infinite dilution being incorporated, by this definition, in μ_i^0 (2). A consequence of this formalism is that ln $\gamma_i \to 0$ for every solute in each solvent at infinite dilution, and that μ_i^0 for any solute differs in each solvent. From its defini-

TABLE I

Solubilities in aqueous ethanol solutions at 25.1°[a]

Solute	Solubilities at EtOH concentration of								
	0	20%	30%	40%	50%	60%	80%	90%	100%
	g/100 g solvent								
Glycine	25.16	11.30	6.83	4.25	2.57	1.40	0.243	0.050	
Leucine	2.17	1.32		0.853		0.628	0.302		
Phenylalanine	2.79	1.86	1.56	1.48	1.38	1.23	0.602	0.245	
Tryosine	0.0452	0.0318		0.0262		0.0192	0.0078	0.0029	0.00052
Dihydroxyphenylalanine	0.380	0.264		0.189		0.114	0.039		
Tryptophan	1.38[b]	1.13	1.11	1.25	1.42	1.40	0.78	0.33	
Histidine	4.30	2.22	1.55	1.10	0.76	0.50	0.108	0.025	
Asparagine	2.51	1.07		0.428		0.151			
Glutamine	4.15	1.87		0.78		0.26			
Triglycine	6.45	2.14		0.68		0.165			

[a] Where comparison is possible, the solubilities listed above agree quite well with the values given by Dunn and Ross (5), Cohn *et al.* (6), and Dunn, Ross, and Read (16) (see Fig. 1).

[b] Different samples of amino acid sometimes give slightly different solubilities in water, possibly indicating the existence of polymorphic crystalline forms, such as have been described for glycine by Iitaka (17). For example, the sample of tryptophan used for solubilities in ethanol had a solubility of 1.38 g/100 g H_2O, that used for the studies in dioxane (Table II) had a solubility of 1.28 g/100 g H_2O. Hade (18) reported a value of 1.36 g/100 g H_2O. Such differences do not affect the values of ΔF_t significantly (see text).

TABLE II

Solubilities in aqueous dioxane solutions at 25.1°

Solute	Solubilities at dioxane concentration of						
	0	20%	30%	40%	60%	80%	90%
	g/100 g solvent						
Glycine	25.16	10.9	6.35	3.30	0.64	0.062	
Leucine	2.17	1.35		0.71	0.31	0.056	
Phenylalanine	2.79	2.47	2.24	1.90	1.09	0.265	0.05
Tyrosine	0.0452	0.0457		0.0391	0.0209	0.0408	0.00065
Tryptophan	1.28	2.13	2.59	2.97	2.80	0.873	0.136
Histidine	4.28	2.47	1.63	1.00	0.255		
Asparagine	2.51	1.19		0.410	0.095		
Glutamine	4.15	1.96		0.671	0.164		
Diglycine	22.75	7.70		1.54	0.195		
Triglycine	6.5	2.0		0.42	0.045		

tion, ΔF_t is the free energy of transfer when $N_{i,w} = N_i$ and all $\ln \gamma_i$ vanish, *i.e.* $\Delta F_t = \mu_i{}^0 - \mu_{i,w}^0$. In terms of the solubility measurements,

$$\Delta F_t = RT \ln N_{i,w}/N_i + RT \ln \gamma_{i,w}/\gamma_i \qquad (2)$$

Experimental values for $\gamma_{i,w}$ are available for some of the solutes we have used, on the basis of the work of Smith and Smith [19] and of Ellerton *et al.* [20]. Reasonable estimates can be made for amino acids for which no actual data have been recorded, as described in our earlier papers [2, 3]. However, self-interaction in aqueous ethanol and dioxane solutions has not been measured, so that values of γ_i in such solutions have to be estimated by some arbitrary procedure. This same problem was encountered in our previous studies, and was dealt with in different ways. In guanidine hydrochloride [4] we felt that no reasonable basis for making an estimate exists. We therefore ignored the self-interaction term and reported values for the parameter $\Delta F'_t$, defined as

$$\Delta F'_t = RT \ln N_{i,w}/N_i \qquad (3)$$

In urea [2] we assumed that solute concentration is the major factor determining $RT \ln \gamma_i$ and assumed that the value of $RT \ln \gamma_i$ in a mixed solvent at saturation is the same as its value in water at the same concentration. In ethylene glycol [3] we used the same procedure, but, recognizing that self-interaction in glycine, diglycine, and triglycine is likely to be predominantly electrostatic in origin, we applied a correction for these three solutes to take into account the reduced dielectric constant of glycol-water mixtures. The correction consisted in multiplying the uncorrected value of $RT \ln \gamma_i$ by D_w/D where D is the dielectric constant of the solvent mixture, and D_w that of water.

The procedure followed here is the same as was used for ethylene glycol, and the values of $RT \ln \gamma_i$ obtained in this way are shown in Tables III and IV. The figures show that the self-interaction term becomes negligibly small at high concentrations of ethanol or dioxane, because all of the measured solubilities become very small under these conditions. Thus our rather arbitrary method for assigning values to $RT \ln \gamma_i$ is likely to represent an unimportant source of error for the data at high ethanol or dioxane concentration, which are in fact the data of greatest interest.

Values of ΔF_t, obtained according to Equation 2, with the self-interaction terms of Tables III and IV, are shown in Tables V

TABLE III

Values of RT ln γ_i for saturated solutions of amino acids, diglycine, and triglycine in aqueous ethanol solutions at 25.1°

Values of $RT \ln \gamma_i$ less than 10 cal per mole in magnitude are listed as zero.

Solute	Values at ethanol concentration of							
	0	20%	30%	40%	50%	60%	80%	90–100%
	cal/mole							
Glycine	−152	−116	−88	−66	−46	−32	0	0
Leucine	+22	+14	0			0	0	
Phenylalanine	+45	+30	+25	+24	+23	+20	+10	0
Tyrosine	0	0	0			0	0	0
Dihydroxyphenyl-alanine	0	0	0		0	0		
Tryptophan	+19	+14	+14	+17	+19	+19	+11	0
Histidine	+36	+18	+13	0	0	0	0	0
Asparagine	−18	0	0		0			
Glutamine	0	0	0		0			
Diglycine	−216	−136		−59		−20		
Triglycine	−138	−68		−31		−14		

TABLE IV

Values of RT ln γ_i for saturated solutions of amino acids, diglycine, and triglycine in aqueous dioxane solutions at 25.1°

Values of $RT \ln \gamma_i$ less than 10 cal per mole in magnitude are listed as zero.

Solute	Values at dioxane concentration of						
	0	20%	30%	40%	60%	80%	90–100%
	cal/mole						
Glycine	−152	−129	−104	−70	−26	0	
Leucine	+22	+14		0	0	0	
Phenylalanine	+45	+39	+36	+31	+18	0	0
Tyrosine	0	0		0	0	0	0
Tryptophan	+19	+27	+35	+39	+36	+12	0
Histidine	+36	+21	+14	0	0		
Asparagine	−18	0		0	0		
Glutamine	0	0		0	0		
Diglycine	−216	−159		−60	−14		
Triglycine	−138	−74		−29	0		

and VI. Values of $\Delta F'_t$ are given also. Listing these uncorrected free energies will facilitate future correction if accurate measurements of $RT \ln \gamma_i$ become available.

Fig. 1 shows a plot of ΔF_t *versus* solvent composition for several amino acids in aqueous ethanol solutions. As had already been observed by Cohn and Edsall [9], for similar plots of $\Delta F'_t$, there are considerably greater changes in slope than one obtains from solubility data in urea, or guanidine hydrochloride. The most prominent feature is a hump in the curves, near mole fraction 0.15 (40% ethanol by volume). This feature of the results, however, does not seem to stem from the irregularity in amino acid solubilities but is likely the result of the anomalous behavior of water-ethanol mixtures, as reflected in measurements of viscosity, density, vapor pressure, etc.

TABLE V

Free energies of transfer (ΔF_t) and apparent free energies of transfer ($\Delta F'_t$) from water to aqueous ethanol solutions at 25.1°

Solute	Values at ethanol concentration of							
	20%	30%	40%	50%	60%	80%	90%	100%
ΔF_t (cal/mole)								
Glycine	360	585	805	1035	1330	2215	3060	
Leucine	240		430		520	835		
Phenylalanine	190	270	265	260	285	590	1045	
Tyrosine	145		190		280	690	1190	2085
Dihydroxyphenyl-alanine	155		280		490	1005		
Tryptophan	60	40	−75	−195	−230	−5	425	
Histidine	355	530	705	880	1085	1865	2645	
Asparagine	435		900		1420			
Glutamine	415		860		1395			
Triglycine	520		1090		1820			
$\Delta F'_t$ (cal/mole)								
Glycine	395	650	890	1145	1450	2360	3240	
Leucine	230		415		505	815		
Phenylalanine	175	250	240	240	260	555	1005	
Tyrosine	145		190		280	690	1190	2085
Dihydroxyphenyl-alanine	155		280		490	1005		
Tryptophan	55	35	−75	−195	−230	−15	410	
Histidine	335	505	675	850	1055	1830	2610	
Asparagine	445		915		1435			
Glutamine	410		855		1410			
Triglycine	590		1195		1945			

TABLE VI

Free energies of transfer (ΔF_t) and apparent free energies of transfer ($\Delta F'_t$) from water to aqueous dioxane solutions at 25.1°

Solute	Values at dioxane concentration of					
	20%	30%	40%	60%	80%	90%
ΔF_t (cal/mole)						
Glycine	350	580	865	1635	2785	
Leucine	185		450	790	1595	
Phenylalanine	−25	−20	15	205	845	1790
Tryptophan	−415	−595	−745	−860	−360	605
Histidine	240	430	660	1320		
Tyrosine	−110		−140	75	835	1830
Asparagine	325		830	1535		
Glutamine	340		855	1540		
Diglycine	470		1195	2220		
Triglycine	525		1285	2425		
$\Delta F'_t$ (cal/mole)						
Glycine	375	630	945	1765	2930	
Leucine	175		435	770	1575	
Phenylalanine	−30	−30	0	175	800	1745
Tryptophan	−405	−580	−725	−840	−365	590
Histidine	220	410	630	1290		
Tyrosine	−110		−140	75	835	1830
Asparagine	335		845	1555		
Glutamine	340		850	1530		
Diglycine	525		1350	2420		
Triglycine	590		1395	2560		
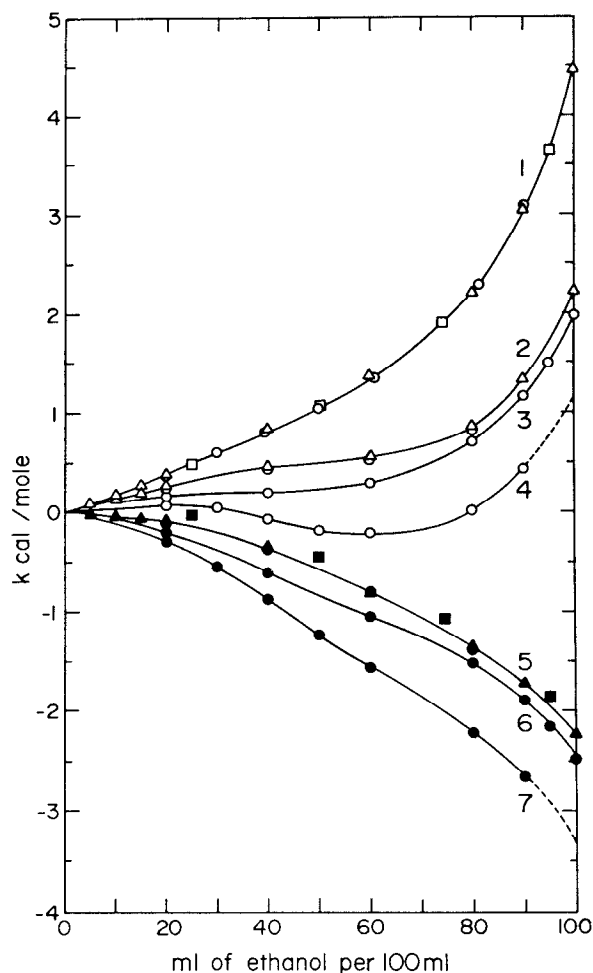
FIG. 1. Values of ΔF_t and Δf_t as a function of ethanol concentration. Curves *1*, ΔF_t for glycine; *2*, ΔF_t for leucine; *3*, ΔF_t for tyrosine; *4*, ΔF_t for tryptophan; *5*, Δf_t for leucine side chain; *6*, Δf_t for tyrosine side chain; and *7*, Δf_t for tryptophan side chain. O and ●, present study; △ and ▲ from solubility data in Reference 6; □ and ■, from solubility data in References 5 and 16. The curves for leucine represent three sets of data of which the two taken from Reference 6 have been obtained from solubilities of *l*-leucine and *d,l*-leucine (up to 80%) and those values are indistinguishable, therefore are represented by single points. The data calculated from solubilities of *d,l*-leucine in References 5 and 16 are located at about the same distance from the Δf_t curve representing the other three sets of data, indicating the possilble difference in crystalline samples used for water and ethanol soutions (see text).

Similar anomalies are observed when ΔF_t values in dioxane-water mixtures are plotted as a function of dioxane concentration. They can again be related to anomalous behavior of the solvent itself.

Calculation of Group Contributions—Group contributions (Δf_t) to ΔF_t have been calculated for the amino acid side chains and the backbone peptide unit, assuming additivity of the free energy of solvent interactions. The procedure is exactly the same as used in our previous papers (2–4), where such evidence as exists in support of the procedure has also been presented. The results are presented in Tables VII and VIII.

Table VII includes some Δf_t values calculated in the same way from solubilities found in the literature. Some of these represent results for racemic mixtures of the amino acids. The

TABLE VII

Contributions from amino acid side chains and peptide unit to free energies of transfer from water to aqueous ethanol solutions at 25.1°[a]

	Δf_t at ethanol concentration of						
Vol %: 20 Moles/liter: 3.4	30 5.1	40 6.8	60 10.2	80 13.6	90 15.3	100 17.0	
	cal/mole						
Side chain of							
Alanine[b]	−20	−40	−60	−130	−260	−365	−520
Valine[c]	−60		−195	−455	−790	−1075	−1475
Leucine	−120		−375	−810	−1385	−1735[d]	−2245[d]
Phenylalanine	−170	−320	−545	−1045	−1625	−2010	(−2600)
Tyrosine	−215		−615	−1050	−1530	−1865	−2390
Dihydroxyphenylalanine	−205		−525	−840	−1210	(−1470)	(−1850)
Tryptophan	−300	−550	−880	−1560	−2220	−2630	(−3220)
Histidine	−5	−60	−100	−245	−355	−410	(−450)
Serine[e]	(100)		(125)	(180)	(240)	(260)	(300)
Asparagine	75		90	90	(100)		150[f]
Glutamine	55		55	90	(150)		260[f]
Peptide unit as difference between							
Diglycine,[f] glycine	135		270	425	745	995	1270
Triglycine, diglycine	25		15	65	170[g]	495[g]	895[g]

[a] The values in parentheses have been obtained by graphical extrapolation or interpolation (see text).

[b] Based on the solubilities of d,l-alanine given in References 5, 6, 16, and 21.

[c] Based on the solubilities of d- and d,l-valine given in Reference 6 and d,l-valine given in References 5 and 16. All three sets of data give essentially identical results.

[d] Taken from Reference 6. There are no significant differences between our data and the data in this reference at lower ethanol concentrations (see Fig. 1).

[e] Based on the solubilities of d,l-serine given in References 5 and 22.

[f] Based on the solubilities given in Reference 7. The literature values for asparagine at lower ethanol concentrations are essentially identical with our values.

[g] Based on the solubilities given in Reference 8.

TABLE VIII

Contributions from amino acid side chains and peptide unit to free energies of transfer from water to aqueous dioxane solutions at 25.1°
The values in parentheses have been obtained by graphical extrapolation (see text)

	Δf_t at dioxane concentration of						
Vol %: 20 Moles/liter: 2.3	30 3.5	40 4.7	60 7.0	80 9.3	90 10.5	100 11.7	
	cal/mole						
Side chain of							
Leucine	−165		−415	−845	−1190	(−1310)	(−1400)
Phenylalanine	−375	−600	−845	−1430	−1940	(−2160)	(−2320)
Tyrosine	−465		−1055	−1560	−1955	(−2125)	(−2250)
Tryptophan	−765	−1175	−1610	−2495	−3145	(−3345)	(−3480)
Histidine	−115	−150	−205	−315			(−480)
Asparagine	−25		−35	−225			
Glutamine	−10		−10	−230			
Peptide unit as difference between							
Diglycine, glycine	120		330	585			
Triglycine, diglycine	55		90	205			

values of ΔF_t (properly corrected for self-interaction) should be the same for a racemic mixture as for a pure optical isomer (6, 9), although the solubilities themselves may differ greatly. The self-interaction term in a racemic mixture could differ from that for a pure isomer because two enantiomorphs can interact with each other in a manner not available to 2 molecules of identical configuration. The fact that no significant differences are observed between the values for these two types of isomers as shown

(Table VII and Fig. 1) further attests to the relative unimportance of the absence of precise values for the self-interaction term.

A phenomenon that may invalidate the calculation of ΔF_t from solubility data is the possible existence of different crystal modifications for a given solute, but even then the results will be valid if all measurements in a series refer to the same crystalline form. Equations 1 to 3 are valid if all solubility measurements represent saturation with respect to the *same* crystalline form. If one

starts with two different crystalline forms, and no transformation of one to the other occurs in the course of solubility measurements, two sets of solubilities will be obtained, which will, however, give consistent ΔF_t values. Comparison between racemic mixtures and pure optical isomers, in effect represents an example of this situation. Only if changes in the solvent medium induce a transformation in crystalline form will incorrect Δf_t values be obtained. There is no evidence that any such phenomenon significantly affected our results.

In Fig. 1 there is excellent agreement between Δf_t values for leucine based on our solubility data and those based on the work of Cohn *et al.* (6), the latter comprising results both for L-leucine and for the racemic mixture. There is a discrepancy, however, between these results and those based on the solubilities of d,l-leucine by Dunn and Ross (5) and Dunn, Ross, and Read (16), which may well be the result of saturation with respect to different crystalline forms, since the curve representing Dunn's data is parallel to the curve representing the rest of the data, *i.e.* it might indicate that solubility in water was measured on a different crystal form from that existing in all of the alcohol solutions by these authors.

Although it was not possible to extend all solubility measurements to 100% ethanol or dioxane, because of the low solubility in these solvents, Δf_t values for the hydrophobic side chains in 100% organic solvent are needed to generate a hydrophobicity scale. Where necessary, results have therefore been extrapolated to 100% ethanol or dioxane. The extrapolation can be done in different ways: by extrapolation of Δf_t itself, by extrapolation of ΔF_t values and subsequent calculation of Δf_t, or by extrapolation of $\Delta F'_t$ values and subsequent insertion of the self-interaction term and calculation of Δf_t. The data may be plotted as a function of volume per cent (as in Fig. 1), molarity, or mole fraction of solvent composition. We have found that the result depends relatively little on the method of extrapolation, and we believe the reliability of the extrapolated data to be

TABLE IX

Hydrophobicity scale: Δf_t for transfer of amino acid side chain from 100% organic solvent to water at 25°

Amino acid side chain[a]	Δf_t
	cal/mole
Tryptophan	3400[b]
Norleucine	2600[c]
Phenylalanine	2500[b]
Tyrosine	2300[b]
Dihydroxyphenylalanine	1800[d]
Leucine	1800[b]
Valine	1500[d]
Methionine	1300[d]
Histidine	500[b]
Alanine	500[d]
Threonine	400[d]
Serine	−300[d]

[a] No figure is given for proline. The rather large value given earlier is subject to greater uncertainty than any of the figures listed here.

[b] Average of values for ethanol and dioxane. For leucine and histidine the values for ethanol have been given double weight because of greater accuracy of extrapolation.

[c] Average of values for ethanol, butanol, and acetone.

[d] Values for ethanol only.

about ±50 cal per mole. The results for leucine and histidine in dioxane solution represent an exception, and the uncertainty for these amino acid side chains is estimated at ±100 cal per mole.

It is important to note that the probable error in the self-interaction correction for these results is very small. In determining Δf_t we subtract ΔF_t for glycine from ΔF_t for the amino acid in question, and the self-interaction correction thus enters into the result as

$$RT \ln \frac{\gamma_{i,w}}{\gamma_i} - RT \ln \frac{\gamma_{Gly,w}}{\gamma_{Gly}} \quad (4)$$

where i refers to the amino acid whose side chain Δf_t is being determined. It is evident from Tables III and IV that $RT \ln \gamma_i$ and $RT \ln \gamma_{Gly}$ are both negligibly small at the higher concentrations of organic solvent. Equation 4 thus reduces to $RT \ln \gamma_{i,w} - RT \ln \gamma_{Gly,w}$. The major contribution to this difference comes from $RT \ln \gamma_{Gly,w}$ and this is the most accurate of all the figures in Tables III and IV, direct experimental values determined from two laboratories (19, 20) being in excellent agreement.

DISCUSSION

Tables VII and VIII clearly confirm the very large stabilization of hydrophobic moieties in organic solvents, relative to water. The results obtained in ethanol and dioxane solutions are very similar, especially at the higher concentrations of the organic solvent.

With these results, and additional literature data for solubilities in ethanol, the amino acids can be arranged in a definite order with respect to their hydrophobicity, as shown in Table IX. The figures given in this table replace those cited in an earlier paper (12) as the best available estimate of hydrophobicity. The table includes data not previously available (for tryptophan, histidine, and dihydroxyphenylalanine). Other changes are small and are caused in large part by incorporation of the self-interaction term in Equation 2, which was neglected in the construction of the earlier table.

The results show that aromatic side chains are the most hydrophobic among the naturally occurring amino acid side chains. This is primarily caused by the larger size of the aromatic groups: if the results are expressed in terms of the contribution to Δf_t per constituent carbon atom, the aliphatic side chains actually are more hydrophobic (*e.g.* leucine 450, valine 500, and phenylalanine 360).

There is a striking difference between the two heterocyclic side chains, tryptophan and histidine. The former turns out to be the most hydrophobic of all the amino acid side chains, and the nitrogen atoms appear to contribute as much to the hydrophobicity as carbon atoms in their place would contribute. The value of Δf_t per constituent atom (carbon or nitrogen) is 340 cal per mole. The histidine side chain, despite the fact that the uncharged form was studied, receives only minimal stabilization in the hydrophobic environment: Δf_t is < 100 cal per mole per constituent atom. This discrepancy between tryptophan and histidine has been observed in all solvents we have studied.

Another result emerging from Table IX, that has possible theoretical significance is the effect of an OH group on hydrophobicity. The effect of an aliphatic OH group, by comparison of serine with alanine and threonine with α-aminobutyric acid (9), decreases hydrophobicity by about 800 cal per mole. The

effect of an aromatic OH group on the other hand is seen to be much smaller: the difference in Δf_t between tyrosine and phenylalanine is only 200 cal per mole. A second OH group (on dihydroxyphenylalanine) decreases the hydrophobicity by about 500 cal per mole.

It is interesting to note that the methionyl sulfur atom also confers hydrophobic properties on the side chain. Δf_t for methionine is about 740 cal per mole smaller than that for α-aminovaleric acid (9).

Among the entities that appear to be hydrophilic (*i.e.* contribute to diminished solubility in ethanol or dioxane), the backbone peptide unit stands out. Despite the possibility of hydrogen bonding to the OH group of ethanol or the ether oxygen atoms of dioxane, the backbone peptide unit has a strong preference for being in a purely aqueous environment. This result undoubtedly explains why proteins dissolved in organic solvents (*e.g.* 2-chloroethanol) tend to assume highly helical conformations (11, 23). Exposed peptide groups would clearly contribute a relatively large positive free energy. The contribution of a peptide unit as calculated from the comparison between diglycine and glycine was much smaller in magnitude than the similar contribution as calculated from triglycine and diglycine in aqueous urea and guanidine hydrochloride solutions, and they were about equal in aqueous glycol solutions. The trend observed in aqueous ethanol and dioxane solutions is just opposite to that in urea and guanidine hydrochloride solutions, as shown in Tables VII and VIII. However it can be seen that the values obtained by both ways become closer rather rapidly as ethanol concentration approaches 100%.

It has been observed in all our solubility studies that the results obtained for asparagine cannot be reconciled with those for glutamine. Asparagine consistently shows more negative ΔF_t values than those of glutamine, while one expects the opposite trend, since glutamine has an additional CH_2 group which in our solvent systems should show a negative contribution to ΔF_t. In our previous paper, we have shown that the free energy contributions from glutamine side chains for the transfer from water to aqueous guanidine hydrochloride, urea, and ethyleneglycol are essentially identical with the sum of independent contributions from the amide group CH_2CONH_2 and a methylene group (4). It is therefore likely that asparagine is the one which shows anomalous behavior and glutamine is not. Part of the explanation for this anomalous behavior may arise from the tendency of asparagine to form ring structures. X-ray diffraction studies of Kartha and de Vries (24) show that in asparagine monohydrate crystal a ring structure is formed, probably through a weak hydrogen bond between the α-amino nitrogen atom and

the oxygen atom of the amide group. On the other hand, no such ring structure was found in the study of glutamine crystals by Cochran and Penfold (25). This cannot, however, be the whole explanation, because the anomalous behavior is too irregular: for transfer from water to 90% ethylene glycol, for example, Δf_t is -830 cal per mole for the asparagine side chain and $+60$ cal per mole for glutamine. In ethanol and dioxane the results for the two side chains are nearly the same.

It seems probable, therefore, that both asparagine and glutamine can adopt more than one conformation, involving internal hydrogen bonds, as well as hydrogen bonds with solvent molecules. The values of Δf_t would then reflect alterations in the equilibria between the possible conformations, as well as the simple effect of environment on a molecule of unchanging conformation.

REFERENCES

1. WHITNEY, P. L., AND TANFORD, C., *J. Biol. Chem.*, **237**, PC1735 (1962).
2. NOZAKI, Y., AND TANFORD, C., *J. Biol. Chem.*, **238**, 4074 (1963).
3. NOZAKI, Y., AND TANFORD, C., *J. Biol. Chem.*, **240**, 3568 (1965).
4. NOZAKI, Y., AND TANFORD, C., *J. Biol. Chem.*, **245**, 1648 (1970).
5. DUNN, M. S., AND ROSS, F. J., *J. Biol. Chem.*, **125**, 309 (1938).
6. COHN, E. J., McMEEKIN, T. L., EDSALL, J. T., AND WEARE, J. H., *J. Amer. Chem. Soc.*, **56**, 2270 (1934).
7. McMEEKIN, T. L., COHN, E. J., AND WEARE, J. H., *J. Amer. Chem. Soc.*, **57**, 626 (1935).
8. McMEEKIN, T. L., COHN, E. J., AND WEARE, J. H., *J. Amer. Chem. Soc.*, **58**, 2173 (1936).
9. COHN, E. J., AND EDSALL, J. T., *Proteins, amino acids, and peptides as ions and dipolar ions*, Reinhold Publishing Corporation, New York, 1943, Chapter 9.
10. TANFORD, C., *J. Amer. Chem. Soc.*, **86**, 2050 (1964).
11. TANFORD, C., *Advan. Protein Chem.*, **24**, 1 (1970).
12. TANFORD, C., *J. Amer. Chem. Soc.*, **84**, 4240 (1962).
13. WISHNIA, A., *Proc. Nat. Acad. Sci. U. S. A.*, **48**, 2200 (1962).
14. BIGELOW, C. C., *J. Theor. Biol.*, **16**, 187 (1967).
15. LUND, H., *Ber. Deutsch. Chem. Ges.*, **67**, 935 (1934).
16. DUNN, M. S., ROSS, F. J., AND READ, L. S., *J. Biol. Chem.*, **103**, 579 (1933).
17. IITAKA, Y., *Acta Cryst.*, **14**, 1 (1961).
18. HADE, E. P. K., Ph.D. thesis, University of Chicago, 1962; in H. A. SOBER (Editor), *Handbook of biochemistry*, Chemical Rubber Company, Cleveland, 1968, p. B10.
19. SMITH, E. R. B., AND SMITH, P. K., *J. Biol. Chem.*, **117**, 209 (1937); **121**, 607 (1937); **132**, 47, 57 (1940).
20. ELLERTON, H. D., REINFELDS, G., MULCAHY, D. E., AND DUNLOP, P. J., *J. Phys. Chem.*, **68**, 398 (1964).
21. HOLLEMAN, A. F., AND ANTUSCH, A. C., *Rec. Trav. Chim. Pays-Bas*, **13**, 277 (1894).
22. DALTON, J. B., AND SCHMIDT, C. L. A., *J. Biol. Chem.*, **109**, 241 (1935).
23. IKAI, A., AND NODA, H., *J. Biol. Chem.*, **243**, 5028 (1968).
24. KARTHA, G., AND DE VRIES, A., *Nature*, **192**, 863 (1961).
25. COCHRAN, W., AND PENFOLD, B. R., *Acta Cryst.*, **5**, 653 (1952).