Lee, A. W., Karplus, M., Poyart, C., & Bursaux, E. (1988) *Biochemistry* (in press).

Marden, M. C., Hazard, E. S., Kimble, C., & Gibson, Q. H. (1987) *Eur. J. Biochem. 169*, 611–615.

Minton, A. P., & Imai, K. (1974) *Proc. Natl. Acad. Sci. U.S.A. 71*, 1418–1421.

Monod, J., Wyman, J., & Changeux, J.-P. (1965) *J. Mol. Biol. 12*, 88–118.

Morris, R., & Gibson, Q. H. (1984) *J. Biol. Chem. 259*, 365–371.

Perutz, M. F., & Poyart, C. (1983) *Lancet*, 881–882.

Perutz, M. F., Fermi, G., Abraham, D. J., Poyart, C., & Bursaux, E. (1986) *J. Am. Chem. Soc. 108*, 1064–1078.

Perutz, M. F., Fermi, G., Luisi, B., Shaanan, B., & Liddington, R. C. (1987) *Acc. Chem. Res. 20*, 309–321.

Sawicki, C., & Gibson, Q. H. (1976) *J. Biol. Chem. 251*, 1533–1542.

Sawicki, C., & Gibson, Q. H. (1979) *J. Biol. Chem. 254*, 4058–4062.

Sawicki, C., & Gibson, Q. H. (1981) *Biophys. J. 35*, 265–270.

Shibayama, N., Morimoto, H., & Kitagawa, T. (1986) *J. Mol. Biol. 192*, 331–336.

Shulman, R. G., Hopfield, J. J., & Ogawa, S. (1975) *Q. Rev. Biophys. 8*, 325–420.

Viggiano, G., & Ho, C. (1979) *Proc. Natl. Acad. Sci. U.S.A. 76*, 3693–3677.

# Comparing the Polarities of the Amino Acids: Side-Chain Distribution Coefficients between the Vapor Phase, Cyclohexane, 1-Octanol, and Neutral Aqueous Solution[†]

Anna Radzicka and Richard Wolfenden*

*Department of Biochemistry, University of North Carolina, Chapel Hill, North Carolina 27514*

*Received May 26, 1987; Revised Manuscript Received September 29, 1987*

ABSTRACT: To obtain an indication of the tendencies of amino acids to leave water and enter a truly nonpolar condensed phase, distribution coefficients between dilute solution in water and dilute solution in wet cyclohexane have been determined for each of the common amino acid side chains at pH 7; they are found to be closely related to the inside–outside distributions of the side chains observed in globular proteins. There was no evidence that excess water enters cyclohexane in association with these solutes. Cyclohexane-to-water distribution coefficients can be combined with vapor-to-water distribution coefficients reported earlier to yield vapor-to-cyclohexane distribution coefficients. Vapor-to-cyclohexane distribution coefficients provide an experimental index of susceptibility to attraction by dispersion forces, and the corresponding free energies are found to be linearly related to side-chain surface areas. Observations using different solvents and variously substituted side chains suggest that alcohols such as 1-octanol exert a specific attraction on the side chain of tryptophan. When less polar phases are used as a reference, leucine, isoleucine, valine, phenylalanine, and methionine are found to be more hydrophobic than tryptophan.

When "molecular recognition" events occur in aqueous surroundings, solvent water must usually be displaced from the interacting groups. The solvation properties of amino acid side chains are of interest in attempting to understand the fidelity of activating enzymes in protein biosynthesis (Fersht & Kaethner, 1976) and to predict the biological activities of peptide analogues and mutant proteins. Solvation effects are also important in determining the structures of macromolecules, and the differential affinities of amino acid side chains for solvent water seem to constitute one of the stronger imperatives governing the structures of proteins in solution [for a comprehensive review, see Edsall and McKenzie (1983)]. For these reasons, there has been continuing interest in the differing solvation properties of amino acid side chains. Hydration potentials of amino acid side chains, representing their free energies of transfer from the vapor phase to neutral aqueous solution, have been found to be related to their outside–inside distributions in proteins, as indicated by earlier solvent accessibility calculations (Wolfenden et al., 1981). In a somewhat modified form, hydration potentials have been used to predict the portions of transmembrane proteins that are located within the lipid bilayer (Kyte & Doolittle, 1982). Distribution coefficients between water and hydroxylic solvents such as 1-octanol have also been found to be correlated with solvent accessibility in globular proteins (Fauchère & Pliska, 1983).

Different reference phases can be used to investigate the relative susceptibilities of amino acid side chains to different kinds of physical interaction. It is possible that any one of these phases might by coincidence approximate the environment experienced by a particular amino acid residue in some particular protein, but it also seems clear that no single reference phase is likely to serve as a general model for the interiors of globular proteins. For example, solutes in the dilute vapor phase are devoid of intermolecular contacts, whereas in the interior of a typical protein, there is little unoccupied volume, and dispersion and other attractive forces are at work. The solvation properties of hydroxylic organic solvents such as octanol are complicated by their ability to form specific hydrogen bonds to solutes and by the fact that they contain substantial concentrations of dissolved water. Thus, the

Scheme I
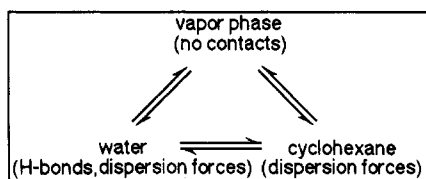


concentration of water in wet octanol, for example, is approximately 2.3 M (Leo & Elkins, 1971), whereas the interiors of globular proteins typically contain only a few highly localized water molecules (Edsall & McKenzie, 1983; Rashin et al., 1986).

To obtain an indication of the relative susceptibilities of amino acid side chains to attraction by dispersion forces alone, it seemed desirable to perform a different set of distribution measurements, using as a reference phase a solvent that is genuinely nonpolar and dissolves very little water.[1] Cyclohexane seemed appropriate for this purpose, because of its low dielectric constant and because it contains very little water at saturation ($2.5 \times 10^{-3}$ M, only slightly higher than that in the vapor phase over water at room temperature, $1.0 \times 10^{-3}$ M).[2] In this way, we hoped to obtain a measure of the hydrophobicity of amino acid side chains with respect to a truly nonpolar condensed reference phase, not necessarily representative of the interior of a protein. In addition, by comparing their free energies of transfer from water to cyclohexane with their free energies of transfer from vapor to water, it might be possible to obtain equilibrium constants for transfer of amino acid side chains from the vapor phase to cyclohexane, an experimental indication of their relative susceptibilities to attraction by dispersion forces uncomplicated by potential hydrogen bonding effects (Scheme I).

The behavior of compounds related to tryptophan and phenylalanine was of special interest because of earlier differences of opinion regarding which of these two amino acids is the more hydrophobic (Fauchère, 1985; Wolfenden & Radzicka, 1986; Fauchère & Pliska, 1986). It seemed desirable to determine whether these differences might depend on the nature of the reference phase chosen for distribution experiments and whether the presence of nearby peptide bonds in solutes might affect their relative distribution properties. The possibility of such interactions had been raised by earlier work showing that when two or more substituents are present within the same molecule, their combined influences on its free energy of solvation may be nonadditive (Fujita et al., 1964; Bone et al., 1983; Wolfenden et al., 1987). As the present results were to show, the relative "hydrophobicities" of certain amino acids depend on the nonpolar solvent that is chosen as a reference.

## MATERIALS AND METHODS

*Materials.* Ethyl methyl sulfide, propionic acid, acetamide, propionamide, 4-methylimidazole, and skatole were purchased from Aldrich Chemical Co. 4-Methylimidazole and toluene were redistilled, and acetamide and propionamide were recrystallized from water. *N*-Propylguanidine was synthesized from *O*-methylisourea and propylamine by the method of

Kimmel (1967). *N*-Methylhydrocinnamamide was synthesized by the method of Suzuki et al. (1960). 3-Indolepropionic acid *N*-methylamide was prepared by the mixed anhydride method of Boissonnas (1951). *N*-Acetyltryptamine was synthesized by acetylation of tryptamine with acetyl chloride, and *N*-acetylphenethylamine was prepared similarly by acetylation of phenethylamine with acetic anhydride (Ho et al., 1969). *N*-Methylhydrocinnamamide, 3-indolepropionic acid *N*-methylamide, and *N*-acetyltryptamine were purified by silicic acid chromatography, eluting with a mixture of methylene chloride and methanol or ethanol. The purity of each product was established at the outset by melting or boiling point, TLC and PMR, and UV if applicable. The spectroscopic properties of these materials were examined again in both phases after distribution experiments to establish the absence of impurities that might have affected the apparent distribution coefficients.

*Methods.* Distribution experiments were performed at 20 °C by using pairs of solvents that had been presaturated with each other. When solutes contained ionizable groups, HCl or KOH was added to suppress ionization. Distribution experiments were performed at several initial solute concentrations in order to establish that apparent distribution coefficients did not vary with solute concentration, as would have been the case if self-association had occurred in either phase.

Distribution coefficients of toluene, *N*-methylhydrocinnamamide, *N*-acetylphenethylamine, skatole, 3-indolepropionic acid *N*-methylamide, and *N*-acetyltryptamine were determined by introducing equal volumes of both phases into a Mixxor apparatus (Cole-Palmer Co.) of 10-mL capacity. After 100 strokes, phases were allowed to clear and centrifuged if necessary, and both layers were then analyzed by UV absorption, using extinction coefficients determined in each phase in separate experiments:

| solute | $H_2O$ | $C_6H_{12}$ | $CHCl_3$ | $C_8H_{18}O$ |
|---|---|---|---|---|
| toluene (262 nm) | 138 | 222 | 261 | |
| *N*-methylhydrocinnamamide (259 nm) | 178 | 166 | 219 | 207 |
| *N*-acetylphenethylamine (259 nm) | 184 | 156 | 206 | 169.5 |
| 3-methylindole (279 nm) | 7215 | 6252 | 6209 | |
| 3-indolepropionic acid *N*-methylamide (280 nm) | 5736 | 5076 | 5688 | 4875 |
| *N*-acetyltryptamine (280 nm) | 5210 | 5235 | 5717 | 6141 |

Distribution coefficients of 4-methylimidazole, acetamide, propionamide, *n*-propylguanidine, methyl ethyl sulfide, methanol, and ethanol were sufficiently one-sided that it was necessary to use a 400- or 500-fold excess of the less-favored phase. Extractions were performed overnight by using a wrist-action shaker, and material entering the less-favored phase was then recovered by back-extraction with a small volume of the more-favored solvent. In most of these cases, $D_2O$ was used as the more-favored solvent, and analysis was performed by integrating the intensities of proton magnetic resonances in the presence of known quantities of an added integration standard, pyrazine. In the exceptional case of methyl ethyl sulfide, the hydrophobic character of the compound made it necessary to use a 500-fold excess of $H_2O$ over cyclohexane-$d_{12}$, analyzing the nonpolar phase before and after extraction. NMR spectrometers used were a Varian EM-390, a Bruker WM-250, or a Varian XL-400FT, according to the requirements of the experiment.

After the concentrations of solutes had been determined before and after extraction, distribution coefficients were calculated by using eq 1, where $M$ is the volume of the

$$K_{M \rightarrow L} = (BM)/(LA - LB) \qquad (1)$$

more-favored solvent, used in both extractions, $L$ is the volume

---

[1] In an earlier study, Nagyvary and Fendler (1974) examined the tendencies of 14 amino acids to leave water and enter *n*-hexane, and the results were only loosely related to the present results. Because of their zwitterionic character in water, the amounts of the free amino acids that were found to enter hexane were very low. This may have led to some of the discrepancies between the two sets of results.

[2] Calculated from the vapor pressure of water of 20 °C (International Critical Tables).

Table I: Distribution Coefficients of Solutes Related to Amino Acid Side Chains, 20 °C[a]

| solute | method | initial concn (M) | chx → $H_2O$[b] | $CHCl_3$ → $H_2O$[c] | oct → $H_2O$[d] |
|---|---|---|---|---|---|
| toluene (Phe) | UV | 2–4 × $10^{-2}$ (org) | -2.19 ± 0.027 (-2.93)[e] | -2.28 ± 0.043 | -2.58[f] |
| N-methylhydrocinnamamide (Phe-NMe) | UV | 2–4 × $10^{-2}$ (org) | 1.19 ± 0.009 | -1.35 ± 0.02 | -1.17 ± 0.004 |
| N-acetylphenethylamine (ac-Phe) | UV | 4–40 × $10^{-3}$ (org) | 1.01 ± 0.016 | -1.41 ± 0.01 | -1.19 ± 0.018 |
| 3-methylindole (Trp) | UV | 3–30 × $10^{-4}$ (org) | -1.71 ± 0.002 | -2.24 ± 0.012 | -2.60[f] |
| 3-indolepropionic acid N-methylamide (Trp-NMe) | UV | 1–2 × $10^{-3}$ (org) | 1.60 ± 0.01 | -0.97 ± 0.01 | -1.23 ± 0.03 |
| N-acetyltryptamine (ac-Trp) | UV | 1–3 × $10^{-3}$ (org) | 1.93 ± 0.001 | -1.04 ± 0.004 | -1.41 ± 0.02 |
| 4-methylimidazole (His) | NMR | 0.5–1.0 (aq) | 3.34 ± 0.03 | | |
| acetamide (Asn) | NMR | 0.5–1.0 (aq) | 4.88 ± 0.02 | 2.00[g] (2.52)[h] | |
| propionamide (Gln) | NMR | 0.5–1.0 (aq) | 4.07 ± 0.01 | 1.40[j] | |
| acetic acid (Asp) | titration | 0.05–0.10 (aq) | 3.27 ± 0.09 (3.15,[k] 1.40[m]) | 1.47[f] | 0.25[f] (0.39)[m] |
| propionic acid (Glu) | titration | 0.05–0.10 (aq) | 2.25 ± 0.08 (2.54,[n] 2.46,[p] 0.80[m]) | 0.83[f] | -0.29[f] (0.26)[m] |
| N-propylguanidine (Arg) | NMR | 0.05–0.10 (aq) | 4.32 ± 0.02 | | |
| ethyl methyl sulfide (Met) | NMR | 0.5–1.0 (org) | -1.73 ± 0.06 | | |
| hydrogen (Gly) | solubilities | | -0.69[r] | | -0.41[f] |
| methane (Ala) | solubilities | | -1.33[s] | | -1.09[t] |
| propane (Val) | solubilities | | -2.97[u] | | -2.36[t] |
| n-butane (Ile) | solubilities | | -3.62[u] | | -2.89[t] |
| isobutane (Leu) | solubilities | | -3.62[v] | | -2.76[t] |
| n-butylamine (Lys) | solvent distributions | | 0.29[w] | -0.99[w] | -0.87[f] |
| methanethiol (Cys) | solubilities | | -0.93[x] | | |
| methanol (Ser) | NMR | 0.5–1.0 (aq) | 2.49 ± 0.03 (2.80,[y] 1.84[z]) | 1.26,[y] 1.36[aa] | |
| ethanol (Thr) | NMR | 0.5–1.0 (aq) | 1.89 ± 0.01 (2.13,[f] 0.85[f]) | | |
| p-cresol (Tyr) | UV | | 0.10[bb] | | |

[a] Distribution coefficients are expressed in logarithmic units, based on *un-ionized* compounds, with standard states taken as molar at infinite dilution in water and in the nonpolar solvent. Values determined in this investigation and values from the literature used in Table II are shown in bold type. [b] Distribution coefficient from cyclohexane to water. [c] Distribution coefficient from chloroform to water. [d] Distribution coefficient from 1-octanol to water. [e] Grunbauer et al. (1982). [f] Average of literature values collected by Hansch and Leo (1979). [g] Sandell (1960). [h] Pichugin et al. (1984). [j] Sandell (1958). [k] Davies (1951). [m] Korenman and Selmanshchuk (1982). [n] Kojima et al. (1970). [p] Rubio (1981). [r] Clever (1980). [s] Calculated from solubilities reported by Ben-Naim (1974). [t] P. Jow and C. Hansch, unpublished results cited in footnote f above. [u] T. Fujita and K. Nishioka, unpublished results cited in footnote f above. [v] Assumed identical with n-butane. [w] Sandell (1962). [x] Calculated from solubility in cyclohexane, 4.18 M, measured gravimetrically at 20 °C and atmospheric pressure in the present study; solubility in water as reported by Shih (1967). [y] Korenman and Chernorukova (1975). [z] Washburn and Spencer (1934). [aa] Sandell (1964). [bb] Golumbic et al. (1949).

of the less-favored solvent, $A$ is the initial solute concentration in the more-favored solvent before distribution, $B$ is the final solute concentration in the more-favored solvent after back-extraction, and $K_{M \to L}$ is the equilibrium constant for transfer from the less-favored to the more-favored solvent (Wolfenden & Williams, 1983).

RESULTS AND DISCUSSION

*Comparisons between Different Distribution Coefficients.* Table I shows distribution coefficients for *uncharged* solutes from dilute solution in nonpolar solvents to dilute solution in water at 20 °C, observed in the present experiments, along with results from earlier investigations. Values used in the following discussion are shown in bold type, along with the estimated experimental errors for values determined in this laboratory.

Table II lists free energies of transfer of amino acid side chains from cyclohexane to water at pH 7 in column a, and free energies of transfer from the vapor phase to water at pH 7 in column b. These were calculated as $-RT \ln K_D$, assuming that only the uncharged compounds enter the nonpolar phase, and correcting values obtained for the nonionized compounds for the fraction of each side chain that is ionized at pH 7, as described earlier (Wolfenden et al., 1981). These fractions are shown in a footnote to Table II. Column c shows mean free energies of transfer from octanol to water at pH 7, estimated by Guy (1985) from six sources in the literature and normalized to octanol. On the basis of these experimental values at 20 °C, free energies were calculated for transfer from the vapor phase to cyclohexane (column d), from wet octanol to cyclohexane (column e), and from the vapor phase to wet octanol (column f). Column g shows the accessible surface
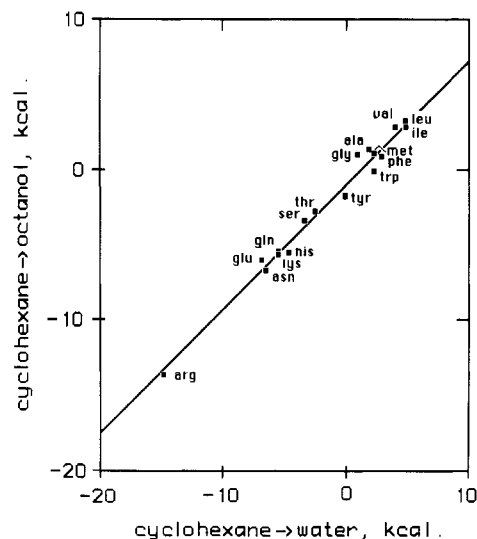


FIGURE 1: Side-chain free energies of transfer from cyclohexane to octanol plotted as a function of free energies of transfer from cyclohexane to water (from Table II).

area of each amino acid residue in a tripeptide (-Gly-R-Gly-, where R is the amino acid in question), calculated by Chothia (1976) from van der Waals radii.

These scales are compared with each other at the foot of Table II, where linear correlation coefficients are shown ($r$ values in excess of 0.9 are shown in bold type). There is a close relationship, for example, between the cyclohexane-to-octanol and cyclohexane-to-water scales ($r = 0.991$), shown in Figure 1. Table II shows that there is a similar relationship ($r =$

Table II: Nonpolar → Polar Distributions of Amino Acid Side Chains, pH 7 (kcal/mol)

| | (a) chx → wat[a] | (b) vap → wat[b] | (c) oct → wat[c] | (d) vap → chx[d] | (e) chx → oct[e] | (f) vap → oct[f] | (g) area[g] | (h) Chothia[h] | (j) Guy[j] |
|---|---|---|---|---|---|---|---|---|---|
| (most hydrophobic) | | | | | | | | | |
| leucine | **4.92** | 2.28 | 1.76 | -2.64 | 3.16 | 0.52 | 173.7 | -0.12 | 1.21 |
| isoleucine | **4.92** | 2.15 | 2.04 | -2.77 | 2.88 | 0.11 | 182.2 | 0.24 | 1.31 |
| valine | **4.04** | 1.99 | 1.18 | -2.05 | 2.86 | 0.81 | 157.2 | 0.09 | 1.09 |
| phenylalanine | **2.98** | -0.76 | 2.09 | -3.74 | 0.89 | -2.85 | 228.6 | 0.00 | 1.27 |
| methionine | **2.35** | -1.48 | 1.32 | -3.83 | 1.03 | -2.80 | 197.6 | -0.24 | 1.27 |
| tryptophan | **2.33** | -5.88 | 2.51 | -8.21 | -0.18 | -8.39 | 271.6 | -0.59 | 0.88 |
| alanine | **1.81** | 1.94 | 0.52 | 0.13 | 1.29 | 1.42 | 93.7 | -0.29 | -0.06 |
| cysteine | **1.28** | -1.24 | | -2.52 | | | 135.2 | 0.00 | 1.36 |
| glycine | **0.94** | 2.39 | 0.00 | 1.45 | 0.94 | 2.39 | 52.6 | -0.34 | -0.41 |
| tyrosine | **-0.14** | -6.11 | 1.63 | -5.97 | -1.77 | -7.74 | 239.9 | -1.02 | 0.33 |
| threonine | **-2.57** | -4.88 | 0.27 | -2.31 | -2.84 | -5.15 | 142.1 | -0.71 | -0.27 |
| serine | **-3.40** | -5.06 | 0.04 | -1.66 | -3.44 | -5.10 | 109.5 | -0.75 | -0.50 |
| histidine | **-4.66** | -10.27 | 0.95 | -5.61 | -5.61 | -11.22 | 188.1 | -0.94 | 0.49 |
| glutamine | **-5.54** | -9.38 | -0.07 | -3.84 | -5.47 | -9.31 | 177.7 | -1.53 | -0.73 |
| lysine | **-5.55** | -9.52 | 0.08 | -3.97 | -5.63 | -9.60 | 215.2 | -2.05 | -1.18 |
| asparagine | **-6.64** | -9.68 | -0.01 | -3.04 | -6.63 | -9.67 | 146.3 | -1.18 | -0.48 |
| glutamate | **-6.81** | -10.24 | -0.79 | -3.43 | -6.02 | -9.45 | 182.9 | -0.90 | -0.77 |
| aspartate | **-8.72** | -10.95 | | -2.23 | | | 142.6 | -1.02 | -0.80 |
| arginine | **-14.92** | -19.92 | -1.32 | -5.00 | -13.60 | -18.60 | 250.4 | -2.71 | -0.84 |
| (most hydrophilic) | | | | | | | | | |
| correlations | | | | | | | | | |
| chx → wat (a) | identity | | | | | | | | |
| vap → wat (b) | **0.939** | identity | | | | | | | |
| oct → wat (c) | 0.84 | 0.62 | identity | | | | | | |
| vap → chx (d) | 0.17 | 0.59 | 0.31 | identity | | | | | |
| chx → oct (e) | **0.991** | **0.968** | 0.76 | 0.33 | identity | | | | |
| vap → oct (f) | 0.87 | **0.987** | 0.48 | 0.66 | 0.927 | identity | | | |
| total area (g) | 0.13 | 0.43 | 0.31 | **0.919** | 0.27 | 0.58 | identity | | |
| Chothia (h) | 0.89 | **0.902** | 0.69 | 0.34 | **0.923** | 0.85 | 0.34 | identity | |
| Guy (j) | 0.81 | 0.66 | 0.85 | 0.16 | 0.75 | 0.46 | 0.20 | 0.77 | identity |

[a] Free energies (kcal/mol) of transfer from dilute solution in cyclohexane to dilute aqueous solution at pH 7, calculated as $-RT \ln K_D$ for distribution coefficients in Table I, after correction for the fraction of each side chain that is ionized at pH 7, as shown in Table I of Wolfenden et al. (1981). Apparent $pK_a$ values used for the side chains in these calculations were 13.65 for arginine, 3.86 for aspartate, 4.24 for glutamate, 6.00 for histidine, and 10.79 for lysine (Edsall & Wyman, 1978). [b] Free energies (kcal/mol) of transfer from the vapor phase to neutral aqueous solution [from Wolfenden et al. (1981)]. [c] Free energies (kcal/mol) of transfer from octanol to neutral aqueous solution [from Table 4 of Guy (1985)]. [d] Free energies (kcal/mol) of transfer from the vapor phase to cyclohexane, obtained by subtracting column a from column b. [e] Free energies (kcal/mol) of transfer from cyclohexane to octanol, obtained by subtracting column c from column a. [f] Free energies (kcal/mol) of transfer from the vapor phase to octanol, obtained by subtracting column c from column b. [g] Accessible surface area of side chain in the tripeptide (-Gly-R-Gly), where R is the amino acid in question, calculated from van der Waals radii by Chothia (1976). [h] Virtual free energy of transfer of residues, from the outside to the inside of a sample of globular proteins, based on the fraction of the total number of residues that are 95% buried according to Chothia's (1976) criterion. [j] Mean polarity, calculated as reported in Table 4 of Guy (1985) from layer analysis of distributions of residues in proteins. [k] Linear correlation coefficient of horizontal row with vertical column.

0.987) between vapor-to-octanol and vapor-to-water distribution coefficients. The strengths of these correlations and the similar absolute magnitudes of the cyclohexane-to-octanol and cyclohexane-to-water scales (the slope of the regression line in Figure 1 is 0.991) indicate that wet octanol resembles water as sampled by the affinities of typical amino acid side chains for wet octanol and for water, i.e., their free energies of transfer from cyclohexane to the solvent in question.

Cyclohexane, on the other hand, resembles the vapor phase in its lack of polarity, as indicated by the close relationship ($r = 0.94$) between the vapor-to-water and cyclohexane-to-water distribution coefficients (Figure 2). It is also evident that the larger nonpolar side chains such as those of Leu and Ile favor cyclohexane relative to the vapor phase, compared with the smaller nonpolar side chains such as those of Ala and Gly. This can be made more explicit. Differences between free energies of vapor-to-water and cyclohexane-to-water transfer are equivalent to the free energies of transfer of the side chains from the vapor phase to cyclohexane (Scheme I) and should provide an experimental indication of the relative susceptibilities of the side chains to attraction by dispersion forces acting between the side chains and cyclohexane. If that were the case, then free energies of vapor-to-cyclohexane transfer might be expected to be related to the surface area of each solute available for van der Waals contacts. Figure 3 shows the present free energies of transfer from vapor to
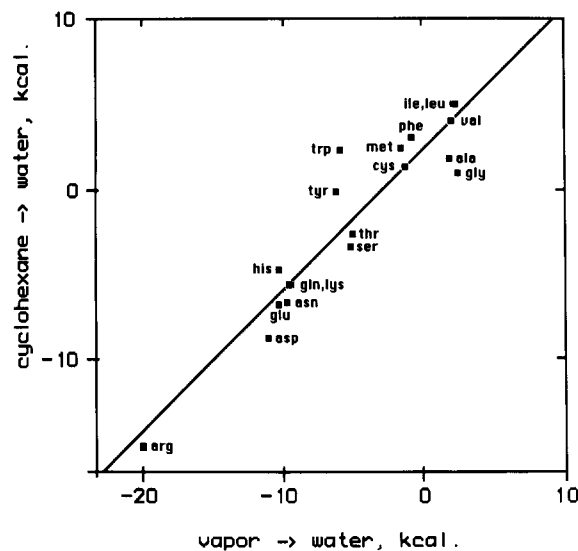


FIGURE 2: Side-chain free energies of transfer from cyclohexane to water plotted as a function of free energies of transfer from the vapor phase to water (from Table II).

cyclohexane (column d), plotted as a function of Chothia's accessible side-chain surface areas (column g). The two parameters are in fact closely related ($r = 0.92$), and variations in side-chain polarity do not disturb this relationship much.
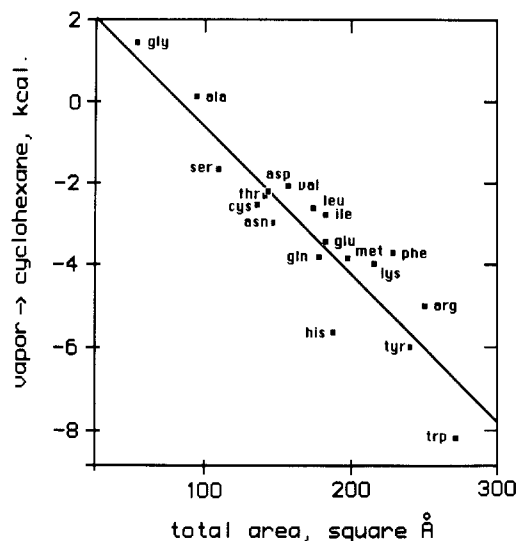
FIGURE 3: Side-chain free energies of transfer from the vapor phase to cyclohexane plotted as a function of accessible surface areas calculated by Chothia (1976) (from Table II).

It will be noted that the side chains of two amino acids, alanine and glycine, show a *positive* free energy of transfer from the vapor phase to cyclohexane. This is presumably related to the energetic cost of making a cavity in liquid cyclohexane, which is not fully compensated by the free energies associated with favorable contacts that are made between hydrogen (or methane) and cyclohexane.

*Comparison between Side-Chain Distribution Coefficients and Protein Folding.* As is evident from Table II and from earlier studies, the cost of extracting the more hydrophilic side chains from neutral aqueous solution into nonpolar environments is very great, leading one to question how closely the distribution properties of the side chains may be related to their typical dispositions in the three-dimensional structures of globular proteins. One would not, of course, expect that any real solvent could represent in detail the complex, anisotropic environment experienced by any individual amino acid residue in a protein. In wet octanol, for example, solutes may form specific hydrogen bonds to hydroxyl groups of octanol and to dissolved water that is present in abundance. Inside globular proteins, hydrogen bonds are often present, but water usually appears to be absent except occasionally at certain specific locations (Edsall & McKenzie, 1983; Rashin et al., 1986). In the dilute vapor phase, solutes exist as isolated molecules, and it is clear that this environment must differ from the environment of a typical amino acid residue in the interior of a globular protein, where there is little unoccupied volume and noncovalent forces of attraction are at work.

The present data supply distribution coefficients with respect to a third kind of reference phase, wet cyclohexane, that is capable of van der Waals attractions but not much else. Only about 1 water molecule is present for every 20 000 cyclohexane molecules, and NMR analysis has consistently shown that excess water molecules are not coextracted with solutes into this nonpolar phase. Thus, hydrogen-bonding partners are not available in wet cyclohexane, whereas hydrogen bonding is not entirely absent in typical globular proteins. X-ray structures suggest that when polar groups are buried, they tend to be matched by H-bonding partners. However, as a general tendency, it seems clear from this and earlier work that H-bonding partners are more easily found in water than in the interiors of proteins. Our purpose in this work was not to provide an ideal model for a protein interior but to provide a truly nonpolar condensed reference phase.

Several methods have been devised for calculating the relative accessibilities of amino acids in proteins, based on crystal structures. Chothia (1976) used the algorithm of Lee and Richards (1971), based on a rolling sphere with the approximate dimensions of a water molecule, to determine the relative accessibilities of residues that were considered to be buried if they were inaccessible to solvent water over 95% of their surface areas. Janin (1979) used a similar approach but considered residues to be buried if <20% of their surface areas were accessible to solvent. Column h of Table II shows virtual free energies of transfer of each amino acid residue, from the outside to the inside of a sample of globular proteins, based on the fraction of the total number of each kind of amino acid that is buried according to Chothia's criterion. Recognizing that certain side chains have both polar and nonpolar portions, other investigators have adopted a more detailed approach suggested by Wertz and Scheraga (1978). This approach seeks to avoid a "hard" binary classification of residues as buried or accessible by dividing each residue into sections or layers and expressing solvent accessibility in proteins as a function of the position of the section (Guy, 1985; Rose et al., 1985). For comparison with the present distribution coefficients, these values can be integrated to provide a "soft" view of each residue's solvation properties. The results of Guy's (1985) analysis are shown as virtual free energies in column j of Table II.

Correlation coefficients between these measures of accessibility in proteins and the side-chain distribution coefficients are shown at the foot of Table II. Chothia's scale is seen to be closely related to cyclohexane → water ($r = 0.89$), vapor → water ($r = 0.90$), cyclohexane → octanol ($r = 0.92$), and vapor → octanol ($r = 0.85$) distribution coefficients and less closely to octanol → water and vapor → cyclohexane distribution coefficients. This seems reasonable if Chothia's "hard" classification tends to reflect the simple presence or absence of polar interactions. These results are consistent with the earlier suggestion that wet octanol is water-like, whereas wet cyclohexane bears a closer resemblance to the vapor-phase extreme. Guy's "soft" scale, based on layer analysis, is better correlated with octanol → water ($r = 0.85$) and cyclohexane → water ($r = 0.81$) distribution coefficients than with the other distribution coefficients included in Table II. The quality of many of these correlations is better than those for physical distribution parameters that had been available until recently (Wolfenden et al., 1981, Table III), and the results are consistent with the well-recognized importance of polarity in determining the distributions of amino acids in proteins. On the other hand, none of these distribution parameters appears to be clearly superior to the others in its correlation with the overall accessibilities of amino acid side chains in protein structure. It would be naive to expect the environment in protein interiors to be well modeled by any simple solvent system, and the amphiphilic character of certain residues, such as lysine, interferes with the predictive value of two-phase distribution coefficients in those cases.

*Specific Interactions between Tryptophan and Octanol.* Individual amino acid residues undoubtedly feel the influence of specific forces within the protein interior that cannot be reproduced by any real solvent. Conversely, the present data furnish evidence that specific interactions sometimes occur between amino acid side chains and the organic solvent that is used as a reference. Early estimates of amino acid hydrophobicity, based largely on free energies of transfer of zwitterionic amino acids from water to alcohols, created the impression that tryptophan is the more hydrophilic of the two
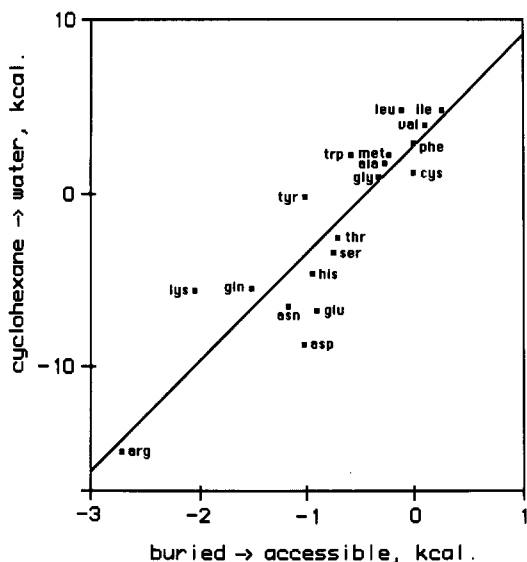
FIGURE 4: Side-chain free energies of transfer from cyclohexane to water plotted as a function of virtual free energies associated with the observed distributions of amino acids between buried and solvent-accessible positions in 12 proteins (Chothia, 1976) (from Table II).



FIGURE 5: Logarithms of distribution coefficients of phenylalanine- and tryptophan-related compounds, from dilute aqueous solution to 1-octanol, chloroform, and cyclohexane (Table I).

amino acids [for a review, see Wolfenden et al. (1981) and Kyte and Doolittle (1982)]. Table II shows, however, that when compounds containing the side chains of tryptophan and phenylalanine are compared, phenylalanine usually appears to be the more hydrophobic. Only when wet octanol is used as the nonpolar reference phase (column 4) does tryptophan appear to be the more hydrophobic. Upon encountering this anomaly, we suggested that it might be a function of the state of derivatization of the amino acid side chain (Wolfenden et al., 1981); other evidence has shown that when several substituents are present within the same molecule, their combined effects on its free energy of solvation sometimes depart markedly from additivity (Fujita et al., 1964; Bone et al., 1983; Wolfenden et al., 1987). The results presented in Table II and Figure 4 indicate, however, that the relationship between the distribution properties of the side chains of phenylalanine and tryptophan tends to be preserved regardless of the introduction of peptide bonds on either side of the α-carbon atom. Phenylalanine remains the more hydrophobic, unless the reference phase is a somewhat polar solvent such as an alcohol.

These results indicate that octanol exerts a specific attraction on the indole ring system, as compared with the benzene ring system. The structural basis of that attraction remains to be determined, but Leo (1972) has noted the tendency of octanol to act as a proton donor in forming hydrogen bonds. Consistent with their ability to act as proton acceptors, aromatic and other unsaturated systems are consistently more hydro-

philic than their saturated counterparts (Hine & Mookerjee, 1975). The ring system of tryptophan is more extensive than that of phenylalanine and might form hydrogen bonds to octanol more effectively for that reason. An alternative possibility is that the NH group of the indole ring, equipped with its lone pair of electrons, could serve as a proton acceptor from octanol. Whatever its structural origin may be, this reversal of the relative distribution coefficients of benzene and indole derivatives, depending on the reference used, indicates the need for caution in attempting to estimate relative polarities based on distributions between water and a single solvent such as octanol.

CONCLUSIONS

(1) Cyclohexane-to-water distributions of amino acid side chains, at pH 7, are related to their distributions between the interiors and exteriors of crystalline proteins, as measured by accessibility calculations.

(2) Free energies of transfer of amino acid side chains from vapor to cyclohexane, an experimental measure of their susceptibility to attraction by dispersion forces, are linearly related to their respective surface areas.

(3) The side chains of leucine, isoleucine, valine, phenylalanine, and methionine are more hydrophobic than that of tryptophan if chloroform, cyclohexane, or the vapor phase is used as a nonpolar reference. Alcohols such as 1-octanol appear to exert a specific attraction on indole derivatives.

REFERENCES

Ben-Naim, Y. M. (1974) J. Phys. Chem. 78, 170.
Boissonnas, R. A. (1951) Helv. Chim. Acta 34, 874.
Bone, R., Cullis, P., & Wolfenden, R. (1983) J. Am. Chem. Soc. 105, 1339.
Chothia, C. (1976) J. Mol. Biol. 105, 1.
Clever, H. L. (1980) Solubility Data Ser. 5–6, 148–152.
Davies, M. (1951) J. Chem. Soc., 1249.
Edsall, J. T., & McKenzie, H. A. (1983) Adv. Biophys. 16, 53.
Fauchère, J.-L. (1985) Trends Biochem. Sci. (Pers. Ed.) 10, 268.
Fauchère, J.-L., & Pliska, V. (1983) Eur. J. Med. Chem. 18, 369.
Fauchère, J.-L., & Pliska, V. (1986) Trends Biochem. Sci. (Pers. Ed.) 11, 69.
Fersht, A. R., & Kaethner, M. (1976) Biochemistry 15, 3342.
Golumbic, C., Orchin, M., & Weller, S. (1949) J. Am. Chem. Soc. 71, 2624.
Grunbauer, H. J. M., Bultsma, T., & Rekker, R. R. (1982) Eur. J. Med. Chem.–Chim. Ther. 17, 411.
Guy, H. R. (1985) J. Biophys. Soc. 47, 61.
Hansch, C., & Leo, A. J. (1979) in Substituent Constants for Correlation Analysis in Chemistry and Biology, Wiley, New York.
Hine, J., & Mookerjee, P. K. (1975) J. Org. Chem. 40, 292.
Ho, E. T., McIsaac, W. M., & Tansey, L. W. (1969) J. Pharm. Sci. 58, 563.
Janin, J. (1979) Nature (London) 277, 491.
Kimmel, J. R. (1967) Methods Enzymol. 11, 585.
Kojima, I., Yoshida, M., & Tanaka, M. (1970) J. Inorg. Nucl. Chem. 32, 987.
Korenman, Y. I., & Chernorukova, Z. (1975) Zh. Prikl. Khim. (Leningrad) 47, 2595.
Korenman, Y. I., & Selmanshchuk, N. N. (1982) Russ. J. Phys. Chem. 56, 558.
Kyte, J., & Doolittle, R. F. (1982) J. Mol. Biol. 157, 105.
Lee, B., & Richards, F. M. (1971) J. Mol. Biol. 55, 379.

Leo, A. J. (1972) *Adv. Chem. Ser. No. 114*, 51.

Leo, A. J., Hansch, C., & Elkins, D. (1971) *Chem. Rev. 71*, 525.

Pichugin, Y. I., Novikov, A. N., & Oleinik, S. T. (1984) *Kriobiol. Kriomed. 14*, 12.

Rashin, A. A., Iofin, M., & Honig, B. (1986) *Biochemistry 25*, 3619.

Rose, G. D., Geselowitz, A. R., Lesser, G. J., Lee, R. H., & Zehfus, M. H. (1985) *Science (Washington, D.C.) 229*, 834.

Rubio, F. C., Rodriguez, V. B., & Alameda, E. J. (1981) *Afinidad 7–8*, 303.

Sandell, K. (1958) *Mh. Chem. 89*, 36.

Sandell, K. (1960) *Naturwissenschaften 53*, 330.

Sandell, K. (1962) *Naturwissenschaften 49*, 12.

Sandell, K. (1964) *Naturwissenschaften 51*, 336.

Shih, T. T. C. (1967) *Tappi 50*, 634.

Suzuki, J., Tsuboi, M., & Shimanouchi, T. (1960) *Spectrochim. Acta 16*, 467.

Washburn, E., & Spencer, H. (1934) *J. Am. Chem. Soc. 56*, 361.

Wertz, D. H., & Scheraga, H. A. (1978) *Macromolecules 11*, 9.

Wolfenden, R., & Williams, R. (1983) *J. Am. Chem. Soc. 105*, 1028.

Wolfenden, R., & Radzicka, A. (1986) *Trends Biochem. Sci. (Pers. Ed.) 11*, 69.

Wolfenden, R., Andersson, L., Cullis, P. M., & Southgate, C. C. B. (1981) *Biochemistry 20*, 849.

Wolfenden, R., Liang, Y.-L., Matthews, M., & Williams, R. (1987) *J. Am. Chem. Soc. 109*, 463.

# Effects of Ammonium Sulfate on the Unfolding and Refolding of the Variable and Constant Fragments of an Immunoglobulin Light Chain[†]

Yuji Goto, Naoya Ichimura, and Kozo Hamaguchi*

*Department of Biology, Faculty of Science, Osaka University, Toyonaka, Osaka 560, Japan*

*Received July 10, 1987; Revised Manuscript Received November 13, 1987*

ABSTRACT: The equilibria and kinetics of unfolding and refolding by guanidine hydrochloride of the $V_L$ and $C_L$ fragments of a type $\kappa$ immunoglobulin light chain were studied in the presence of ammonium sulfate using circular dichroism and tryptophyl fluorescence at pH 7.5 and 25 °C. The unfolding equilibria of the $V_L$ and $C_L$ fragments were described in terms of the two-state transition. The midpoints of unfolding in the absence of ammonium sulfate were at 0.9 and 1.2 M guanidine hydrochloride for the $C_L$ and $V_L$ fragments, respectively. The transition curves were shifted to higher concentrations of guanidine hydrochloride by 1.4 and 1.6 M for the $C_L$ and $V_L$ fragments, respectively, per mole of ammonium sulfate. Unfolding reactions of the $V_L$ and $C_L$ fragments in 3 M guanidine hydrochloride followed first-order kinetics, and the rate constants for the two proteins were both greatly decreased by the presence of ammonium sulfate. The refolding reaction of the $C_L$ fragment in 0.3 M guanidine hydrochloride consisted of two phases, and the rate constants were increased a little by the presence of ammonium sulfate. The refolding reaction of the $V_L$ fragment in 0.3 M guanidine hydrochloride followed first-order kinetics, and the rate was not affected by the presence of ammonium sulfate. These results showed that ammonium sulfate stabilizes the $C_L$ and $V_L$ fragments mainly by decreasing the unfolding rate.

**S**ulfate ion has long been known to protect the protein structure from unfolding (von Hippel & Wong, 1964), but little is known about the quantitative nature of this effect. Recently, Mitchinson and Pain (1985) studied the effect of $(NH_4)_2SO_4$ on the stability of $\beta$-lactamase from *Staphylococcus aureus*. Their results showed that as the destabilizing effect of urea or guanidine hydrochloride (Gdn-HCl)[1] can be described quantitatively, so the stabilizing effect of $(NH_4)_2SO_4$ can also be described quantitatively. In the course of a series of studies on the unfolding and refolding of the $C_L$ and $V_L$ fragments obtained from the immunoglobulin light chain (Goto et al., 1979; Goto & Hamaguchi, 1979, 1981, 1982a,b, 1986a,b, 1987; Ashikari et al., 1984; Kikuchi et al., 1986; Tsunenaga et al., 1987), we found that the $C_L$ and $V_L$ fragments are stabilized by $(NH_4)_2SO_4$.

The immunoglobulin light chain consists of two independently folded structural units, the amino-terminal variable domain ($V_L$) and the carboxy-terminal constant domain ($C_L$), each with a molecular weight of about 12 000. Each domain consists of two $\beta$-sheets and has one intrachain disulfide bond buried in the interior hydrophobic region between the sheets (Beale & Feinstein, 1976; Amzel & Poljak, 1978). Recently, Tsunenaga et al. (1987) succeeded in isolating the $V_L$ and $C_L$ fragments from a type $\kappa$ light chain by limited proteolysis with clostripain (EC 3.4.22.8) and studied the unfolding and refolding of the light chain and its constituent $V_L$ and $C_L$ fragments. It was found that the unfolding of the type $\kappa$ light chain can be described by the independent folding of the two domains but that there are some interactions between the $V_L$ and $C_L$ domains in the refolding process. These results are very similar to those obtained for a type $\lambda$ light chain (Goto et al., 1979; Goto & Hamaguchi, 1982a). It was also found that the unfolding and refolding kinetics of the $V_L$ fragment are different from those of the $C_L$ fragment, in spite of their similar immunoglobulin folding. There might thus be a dif-

---

[1] Abbreviations: CD, circular dichroism; Gdn-HCl, guanidine hydrochloride; Tris, tris(hydroxymethyl)aminomethane.