Veatch, W. R., & Stryer, L. (1977) *J. Mol. Biol. 117*, 1109–1113.
Watts, A., Volotovski, I. D., & Marsh, D. (1979) *Biochemistry 18*, 5006–5013.
Wey, C. L., Ahl, P. L., Cone, R. A., & Gaffney, B. J. (1979) *Biophys. J. 25*, 169a.
Wu, E.-S., Jacobson, K., Szoka, F., & Portis, A., Jr. (1978) *Biochemistry 17*, 5543–5550.

Affinities of Amino Acid Side Chains for Solvent Water†

R. Wolfenden,* L. Andersson, P. M. Cullis, and C. C. B. Southgate

ABSTRACT: Equilibria of distribution of amino acid side chains, between their dilute aqueous solutions and the vapor phase at 25 °C, have been determined by dynamic vapor pressure measurements. After correction to pH 7, the resulting scale of "hydration potentials", or free energies of transfer from the vapor phase to neutral aqueous solution, spans a range of ~22 kcal/mol. The side chain of arginine is much more hydrophilic than those of the other common amino acids, with an equilibrium constant of $\sim 10^{15}$ for transfer from the vapor phase to neutral aqueous solution. Hydration potentials are more closely correlated with the relative tendencies of the various amino acids to appear at the surface of globular proteins than had been evident from earlier distribution studies on the free amino acids. Both properties are associated with a pronounced bias in the genetic code.

In biological systems, "chemical recognition" usually depends on the structural complementarity of different compounds or functional groups that are attracted to each other by noncovalent forces. When these interactions occur in an aqueous environment, solvent water must usually be stripped away from the interacting groups. If information were available about the absolute tendencies of these compounds to leave water and enter an empty cavity that neither attracts nor repeals solutes, then it might be possible to draw inferences about specific forces of attraction or repulsion that may be at work in specific cases.

The absolute affinity of a compound for an aqueous environment can be evaluated by determining its vapor pressure over dilute aqueous solutions. From the results, it is a simple matter to calculate a dimensionless equilibrium constant for its transfer from water to a featureless "solvent" of unit dielectric constant, the dilute vapor phase. Measurements of this kind, performed on a variety of organic compounds, suggest that the free energy of interaction between complex molecules and water can usually be approximated as an additive function of their constituent groups (Butler, 1937; Hine & Mookerjee, 1975). Earlier measurements were confined to relatively volatile solutes that exhibit substantial vapor pressures over water. More sensitive techniques have allowed the recent extension of these measurements to include polar molecules bearing functional groups of biological interest such as the peptide bond (Wolfenden, 1976, 1978).

Differences between amino acid residues, in their strength of solvation by water, are likely to be significant in determining the configurations of proteins in solution (Kauzmann, 1959; Tanford, 1962; Perutz, 1965). It would therefore be of interest to have information about the relative affinities of amino acid side chains for solvent water. Even using material of very high specific activity, efforts in this laboratory to detect glycine in the vapor phase over concentrated aqueous solutions have been unavailing. This is hardly surprising, since glycine in neutral aqueous solution is present as the uncharged species only to the extent of ~1 part in 200 000 (Edsall & Wyman, 1958). Free energies of solvation of charged ammonium and carboxylate groups are each in the neighborhood of −70 to −80 kcal/mol [Kebarle, 1976; see also Tse et al. (1978)], so that the zwitterionic species of glycine in water can be considered totally nonvolatile. The rare, uncharged species of glycine can be expected, from bond contributions based on correlations of data from the literature (Hine & Moorkerjee, 1975), to exhibit an equilibrium constant of $\sim 8 \times 10^{-10}$ for transfer from dilute aqueous solution to the vapor phase. Accordingly, the concentration of glycine, at equilibrium in the vapor phase over an aqueous solution containing 1 M glycine, is expected to be no higher than 10^{-14} M; much lower values are expected for more polar amino acids.

Even if methods more sensitive than those presently available should make it possible in the future to determine free energies for removal of amino acids from water to the vapor phase, it is far from clear that they would serve as good models (even in a relative sense) for the behavior of the various amino acid residues in proteins. It has even been suggested that there may be no significant relationship between free energies of transfer of amino acids from water to organic solvents and their tendencies to appear in internal peptide linkage in the interior rather than at the surface of globular proteins (Janin, 1979). This would seem to indicate either that solvation effects may be less important than originally suspected or that free amino acids are poor models for the relative solvation behavior of amino acids in proteins. Side chains of amino acids in proteins are flanked by peptide bonds, associated with a free energy of solvation of about −10 kcal/mol (Wolfenden, 1978), modest in comparison with the large negative free energies of solvation of ammonium and carboxylate groups mentioned above. The solvent-organizing power of these charged groups is very great and might be expected to affect the relative distribution properties of nearby substituents (Nemethy, 1967; Nandi, 1976).

To obtain results that might be more closely comparable with the solvation behavior expected of amino acids in polypeptides, we decided to examine the behavior of the amino acid

† From the Department of Biochemistry, University of North Carolina, Chapel Hill, North Carolina 27514. *Received June 17, 1980.* Supported by grants from the National Science Foundation (PCM-7823016) and the National Institutes of Health (GM-18325) and by fellowships from the American-Scandinavian Foundation (to L.A.) and from the Jane Coffin Childs Foundation (to P.M.C.).

0006-2960/81/0420-0849$01.00/0

Table I: Water Affinity of Amino Acid Side Chains

$NH_2(R)CHCOOH$	RH	log (RH_{vap}/RH_{H_2O})	pK_a[a]	log α,[b] pH 7	hydration[c] potential
alanine	methane	1.43[d]			1.94
arginine	*n*-propylguanidine	−8.00[e]	13.65	−6.65	−19.92
asparagine	acetamide	−7.12[f]			−9.68
aspartate	acetic acid	−4.91[f]	3.86	−3.14	−10.95
cysteine	methanethiol	−0.91[g]			−1.24
glutamine	propionamide	−6.90[h]			−9.38
glutamate	propionic acid	−4.74[h]	4.24	−2.76	−10.20
glycine	hydrogen	1.76[j]			2.39
histidine	4-methylimidazole	−7.51[h]	6.00	−0.04	−10.27
isoleucine	1-butane	1.58[d]			2.15
leucine	isobutane	1.68[d]			2.28
lysine	*n*-butylamine	−3.21[k]	10.79	−3.79	−9.52
methionine	methyl ethyl sulfide	−1.09[l]			−1.48
phenylalanine	toluene	−0.56[d]			−0.76
serine	methanol	−3.72[g]			−5.06
threonine	ethanol	−3.59[g]			−4.88
tryptophan	3-methylindole	−4.33[e]			−5.88
tyrosine	4-cresol	−4.49[m]			−6.11
valine	propane	1.46[d]			1.99

[a] Apparent pK_a values of the side chains in free amino acids, from Edsall & Wyman (1958). pK_a values for compounds of structure RH yielded somewhat different values for α, but the resulting effects on the present correlations were not significant. [b] Fraction of the side chain that is not ionized at pH 7.0, calculated from these apparent pK_a values. [c] Free energy of transfer of RH from the vapor phase to water buffered at pH 7 (see Discussion). [d] McAuliffe (1966). [e] Calculated from a value of −8.20, obtained for methylguanidine in the present study (see text). Two additional methylene groups are expected to increase the equilibrium constant for transfer of propylguanidine, from water to the vapor phase, by a factor of ~1.59 (Wolfenden & Lewis, 1976) or 0.20 logarithmic units. [f] Wolfenden (1976). [g] Hine & Weimar (1965). [h] This study. [j] Weast (1972). Hydrogen is slightly less soluble than might be predicted by extrapolation of values for the normal hydrocarbons (compare vapor pressures of hydrogen, methane, and higher alkanes). The extrapolated value would place glycine nearer the regression line of Figure 2, but the effect is minor (~0.5 kcal) in terms of the scale of the vapor to water axis of this figure. [k] Butler & Ramchandani (1935). [l] Interpolated between reported values of −1.13 for dimethyl sulfide (Hine & Weimar, 1965) and −1.05 for diethyl sulfide (Deno & Berkheimer, 1960). [m] Parsons et al. (1972).

side chains by themselves. In this paper we report distribution coefficients, from neutral aqueous solution to the vapor phase, of the side chains of α-amino acids that are commonly found in proteins. These "hydration potentials", some of which were included in a preliminary report (Wolfenden et al., 1979), support the view that solvation effects serve as a dominant principle in organizing the tertiary structure of proteins.

Materials and Methods

N-Methylguanidinium sulfate was purchased from Eastman Chemical Co. 3-Methylindole (skatole) was obtained from Sigma Chemical Co. and purified by sublimation. 4-Methylimidazole-*2*-^{14}C was prepared from acetol acetate formaldehyde-^{14}C (New England Nuclear Co.) by the method of Weidenhagen & Herrmann (1935). Propionic acid-*1*-^{14}C was obtained from California Bionuclear Corp., and its amide was prepared by ammonolysis of the ethyl ester, generated by treatment of propionic acid with ethanolic HCl. Last traces of radioactive propionic acid were removed from propionamide-*1*-^{14}C by codistillation with added nonradioactive propionic acid.

Distribution experiments were performed as described earlier (Wolfenden, 1978) at 25 °C by using dilute (0.01 M) solutions with pH adjusted with HCl or KOH if necessary to maintain solutes in an uncharged state. For confirmation of the identity of material transferred to the traps, 1-octanol/water distributions were determined and compared with those observed for the authentic solutes. From 0.1 N HCl, observed distribution coefficients to 1-octanol were acetic acid, 0.51; acetamide, 0.070; propionic acid, 2.14; propionamide, 0.22; and 4-methylimidazole, 0.003. From 0.1 N KOH, similar values were observed for the amides; values for the acids were reduced to <0.001; and the value for 4-methylimidazole rose to 0.75. A colorimetric method proved suitable for determination of *N*-methylguanidine (Eggleton et al., 1943), and 3-methylindole was determined directly in the vapor phase as described below.

Results

Of the side chains of amino acids that occur naturally in proteins, many are identical with or closely similar to compounds whose distribution coefficients between water and the vapor phase have been determined previously. Values for the less hydrophilic side chains, including those of glycine, leucine, isoleucine, valine, alanine, phenylalanine, cysteine, threonine, serine, and tyrosine, and their sources in the literature, are shown in Table I. Methyl ethyl sulfide, the side chain of methionine, has not been examined directly, but values reported for dimethyl sulfide (Hine & Weimar, 1965) and diethyl sulfide (Deno & Berkheimer, 1960), 0.074 and 0.089, respectively, are sufficiently similar that is seemed reasonable to interpolate a value of 0.083 for the mixed disulfide.

Most of the remaining amino acid side chains were examined by a dynamic method described recently, using water-saturated nitrogen as the carrier gas (Wolfenden, 1978). In every case except arginine, equilibria of transfer were determined over a range of solute concentrations varying by at least a factor of 5, at a total ionic strength of 0.3 maintained with KCl in the aqueous phase, in the presence of buffers if necessary to maintain the solute in an uncharged state. Determinations were made with pots and traps immersed in a water bath maintained at 25.0 ± 0.1 °C. Material transferred to the traps was determined colorimetrically (in the case of guanidine derivatives) or from its radioactivity. The nature of radioactive material accumulating in the traps was determined by comparing its equilibrium distribution between 1-octanol and 0.1 M aqueous HCl or KOH.

In the exceptional case of the side chain of arginine, extreme measures were needed in order to detect material in the vapor phase. Best estimates place the pK_a of guanidine and simple alkyl derivatives in the neighborhood of 13.65 for the protonated species (Davis & Elderfield, 1932; Hall & Sprinkle, 1932). Preliminary experiments indicated that the vapor pressures of simple guanidines as free bases are extremely low

and that there would be no chance of detecting them in the vapor phase unless the underlying aqueous solutions were extremely concentrated and alkaline.

We decided to examine the simpler homologue, methylguanidine, shorter than the side chain of arginine by two methylene groups and possible to prepare at very high concentrations in water. Methylguanidinium sulfate was mixed with 1 molar equivalent of barium hydroxide, and after removal of barium sulfate by filtration, solutions containing 0.68 M and 1.1 M methylguanidine were introduced into the apparatus described previously (Wolfenden, 1978). Material accumulating in the traps was analyzed colorimetrically, and it proved possible to obtain sufficient transfer in 3 days for a useful estimate of the rate of transfer by using a flow rate of 100 mL/min of carrier gas and a volume of 100 mL of water in the primary trap. In the pot containing 0.66 M methylguanidine, the rate of decomposition of the solute was shown to follow satisfactory first-order kinetics, with a half-life of 7.3 ± 0.7 days under the alkaline conditions of the experiment. This experiment was repeated once at 0.66 M and once at 1.1 M with consistent results. Our best estimate of the equilibrium constant for transfer of uncharged methylguanidine from water to the vapor phase, based on the observed rate of its accumulation in the traps, was 5.0×10^{-9} for the free base. Correcting for decomposition of solute in the pot during the course of the experiment, an approximate equilibrium constant of 6.3×10^{-9} was obtained.

Distribution coefficients of normal carboxylic acids and amides were determined as described earlier (Wolfenden, 1978). To obtain information about histidine, we determined the water-vapor distribution of 4-methylimidazole-2-^{14}C in dilute aqueous solution at pH 10 where this base is uncharged by using the same method. The chemical identity of material transferred to the traps was confirmed by determining its distribution coefficient to 1-octanol from 0.01 M HCl in water (0.75) and from 0.1 M KOH in water (0.003). Identical values were observed for the authentic compound under investigation. At 25 °C the rate of transfer of radioactive 4-methylimidazole, from a solution adjusted to a total concentration of 0.01 M, corresponded to a distribution coefficient of 3.1 (±0.4) $\times 10^{-8}$ for transfer of 4-methylimidazole water to the vapor phase.

Of the side chains included in the present study, only that of tryptophan proved too volatile to be suitable for vapor pressure determination by the dynamic method. Instead, 3-methylindole was examined by measuring its ultraviolet absorbance in the vapor phase over aqueous solutions of known concentration. A brass jacket, thermostated at 25 °C, designed to accommodate a cylindrical cuvette of 10-cm path length, was equipped with a port through which an aqueous solution of 3-methylindole could be introduced without disturbing the cuvette or interrupting the path of light through the vapor space between its windows. The cell compartment outside the jacket was maintained at 30 °C to prevent condensation on the end windows of the cuvette. Injection of a small volume of aqueous 3-methylindole (0.5 mL) led to rapid development, complete in 45 s, of a stable vapor-phase spectrum, with a sharp peak at 288 nm as reported for indole in cyclohexane (Friedel & Orchin, 1951). Since the absorption spectrum of 3-methylindole in the vapor phase was found to be virtually identical with the spectrum of the authentic compound in cyclohexane, its concentration in the vapor phase was estimated from an extinction coefficient determined in cyclohexane in separate experiments, $\epsilon_{267} = 6.74 \times 10^3$. Experiments with solutions containing methylindole at concentrations of 0.25, 0.50, and 1.0×10^{-2} M yielded consistent values averaging

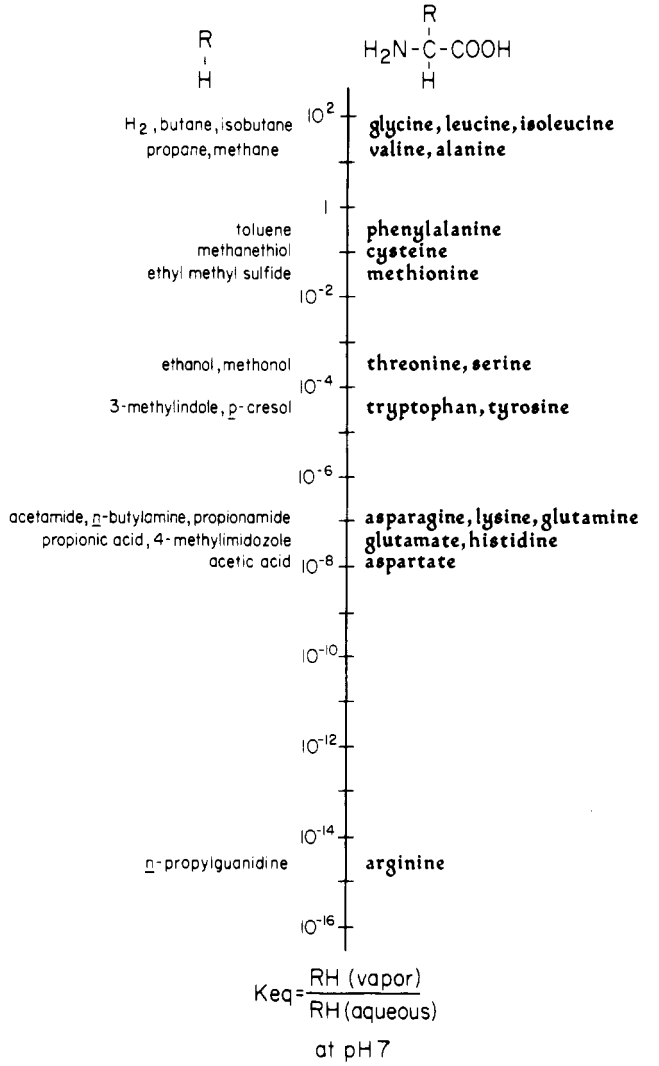


FIGURE 1: Equilibrium constants for transfer of amino acid side chains from water, buffered at pH 7.0, to the vapor phase.

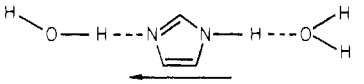
4.7×10^{-5} for the equilibrium constant of transfer of methylindole from water to the vapor phase.

Discussion

Influence of Amino Acid Structure on Hydration Potential. Table I shows the amino acid side chains of the present study, along with values for their relative hydration potential or effective free energy of transfer from the vapor phase to dilute aqueous solution buffered at pH 7.[1] Values were calculated by assuming that the vapor pressure of molecules bearing charged groups is negligible (see introduction). Equlibrium constants for transfer of uncharged side chains from vapor to water (Table I, column 3) were divided by the fraction of each

[1] It should be noted that these values refer to molecules of the structure R–H, where R is the side chain of each amino acid. In view of the additivity of free energies of transfer from water to vapor, as a function of the constituent groups of complex molecules (Butler, 1937; Hine & Mookerjee, 1975), a value for the *total* hydration potential of an internal amino acid residue in a polypeptide can be roughly estimated by adding the hydration potential of the side chain to the hydration potential of a glycine residue. The latter can be regarded as a "frame-shifted" analogue, as shown below, of acetamide or *N*-methylacetamide, both of which exhibit hydration potentials of approximately −10 kcal/mol (Wolfenden, 1978): glycine residue in polymer, ...CH_2CO/$NHCH_2CO$/$NHCH_2CO$...; "frame-shifted" glycine residue, ...CH_2CO/NH/CH_2CONH/CH_2CO...; *N*-methylacetamide, $CH_3CONHCH_3$; acetamide, CH_3CONH_2.

Chart I: Inductive Shift of Electrons



side chain that is present in uncharged form in aqueous solution at pH 7 (Table I, column 5), based on literature pK_a values of the free amino acids. The resulting scale of equilibrium constants, spanning a range of ~16 orders of magnitude, is displayed in Figure 1.

The side chain of arginine is located at the hydrophilic extreme. Even the neutral species of propylguanidine surpasses other side chains in its hydrophilic character (Table I, column 3). Hydrogen-bonding interactions with solvent water are presumably strong and numerous for this side chain, which contains a larger number of electronegative atoms than do the side chains of other amino acids. The extremely high pK_a values of guanidine derivatives, associated with the presence of positive charge at pH values near neutrality, make an important additional contribution to the hydration potential of the arginine side chain. The influence of ionization effects is likewise evident when the side chains of glutamic and aspartic acids are compared with those of glutamine and asparagine. Whereas the amides are considerably more hydrophilic than the parent acids in their uncharged state, these differences are almost eliminated (and in fact reversed) when the aqueous phase is buffered at pH 7 (compare columns 3 and 5 of Table I). Ionization effects also contribute greatly to the hydration potential of the side chain of lysine: under alkaline conditions the side chain of lysine is actually less hydrophilic than that of serine.

The very hydrophilic character of the side chain of histidine, in contrast with that of lysine, persists even at pH values where no charge is present. This can be ascribed to the presence of more than one electronegative atom in the side chain. Unlike imidazole, pyrrole and cyclopentadiene have relatively low boiling points and show very limited solubility in water. Distant polar interactions, of the type noted previously for *p*-nitrophenol (Hine & Mookerjee, 1975), may account for the special affinity of neutral imidazole derivatives for aqueous surroundings. The –NH– group presumably interacts with water by acting as an acid in hydrogen bonding, the =N– atom acting as a base in hydrogen bonding. These interactions are expected to reinforce each other inductively: the –NH– group would then act as an electron donor, increasing the basicity of the =N– atom, and the =N– atom would tend to withdraw electrons from the –NH– group, increasing its acidity (Chart I). As a result, imidazole may interact more strongly with water than would be expected on the basis of the relatively modest affinities of pyrrole and cyclopentadiene for solvent water.

Most aromatic compounds are appreciably soluble in water, and it is therefore not surprising to find amino acids with aromatic side chains scattered through the midrange of values in Figure 1. The side chains of tryptophan and tyrosine are equipped with heteroatoms capable of forming hydrogen bonds with water. Indole is a weak base in aqueous solution, becoming protonated at H_0 values in the neighborhood of –3.5 (Hinman & Lang, 1964), and neither benzene nor naphthalene approaches 3-methylindole in its affinity for water. Equilibria for their transfer from the vapor phase to water are 4.5 and 59, respectively (McAulife, 1956), as compared with a value of 21 000 for 3-methylindole.

Threonine and serine resemble tryptophan and tyrosine in their hydrophilic character. The sulfur-containing side chains are much less strongly attracted to solvent water than are those containing oxygen. The remaining side chains, all aliphatic hydrocarbons, exhibit distribution coefficients from water to the vapor phase that are somewhat greater than unity and closely similar to each other in magnitude.

Table II: ΔG (kcal/mol) for Transfer of Amino Acids from the Interior to the Surface of Proteins (See Text)

	Chothia (1976)	Janin (1979)		Chothia (1976)	Janin (1979)
alanine	−0.29	0.3	leucine	−0.12	0.5
arginine	−2.71	−1.4	lysine	−2.05	−1.8
asparagine	−1.18	−0.5	methionine	−0.24	0.4
aspartate	−1.02	−0.6	phenylalanine	0	0.5
cysteine	0	0.9	proline	−0.90	−0.3
glutamine	−1.53	−0.7	serine	−0.75	−0.1
glutamate	−0.90	−0.7	threonine	−0.71	−0.2
glycine	−0.34	0.3	tryptophan	−0.59	0.3
histidine	−0.94	−0.1	tyrosine	−1.02	−0.4
isoleucine	0.24	0.7	valine	0.09	0.6

Relationship of Hydration Potentials to Protein Folding. Since the early observations of Perutz (1965) on hemoglobin, there has been a continuing effort to relate the observed structures of proteins to the physical properties of their constituent amino acids. In attempting to analyze the wealth of information contained in the structures of proteins whose crystal structures have been determined, it was especially important to have some quantitative index of the accessibility of the various parts of the protein to the surrounding solvent. Lee & Richards (1971) devised an index that uses a rolling sphere model to represent a molecule of solvent and allows a computer to search out the accessible regions of the protein. This model has been applied to proteins in several ways (Lee & Richards, 1971; Shrake & Rupley, 1973; Chothia, 1976; Janin, 1979). It has been used, for example, to generate a table of "residue accessibilities", that displays the number of individual amino acid residues of each kind that are found to be inaccessible to solvent over 95% of their respective surface areas, in the structures of globular proteins (Chothia, 1976). These residues, considered "buried", can then be compared with the number of residues of that amino acid that are found to be accessible and used to generate a kind of partition coefficient that describes the statistical tendency of each amino acid to be found on the surface, rather than in the interior of a large sample of proteins.

In Table II (column 2), the resulting partition coefficients, based on Chothia's sample of 2493 amino acid residues in 12 proteins, are expressed in units of free energy associated with transfer of residues of each kind from the interior to the exterior of globular proteins. Column 3 reproduces a table prepared by Janin (1979), who performed similar calculations on a somewhat larger sample than Chothia's and considered buried residues to be those with <20 Å^2 of accessible surface area.

In Figure 2, free energies associated with access to solvent of amino acid residues in 12 proteins (Table II, column 2) are plotted as a function of the hydration potential of the 19 common α-amino acid side chains determined in the present study. Relatively few lysine and arginine residues are present in the interiors of known proteins, so that considerable uncertainty is associated with the vertical position of these residues in Figure 2. It is immediately evident, however, that there is a relationship between the present scale of hydration potentials and the tendencies of residues to be found at the surface of globular proteins. On the basis of the observed linear correlation coefficient ($r = 0.902$), the probability that the observed values are randomly related [estimated from the tables of Federighi (1959)] is less than 0.0000001. Similar

Table III: Measurements of Amino Acid Affinities for Watery Surroundings

compds examined	transfer process	method	correlations with protein folding[a] no. of compds examined (n)	correlation coeff (r)	probability that null hypothesis is true (P)
amino acids	water to ethanol, ΔG	solubilities[b]	11	0.18	>0.25
amino acids	water to surface, ΔG	surface tension lowering[c]	20	0.53	0.01
amino acids	water to hexane, ΔG	distributions[d]	14	0.67	0.05
amino acids	water to vapor, ΔH	sublimation rates,[e] solubilities	13	0.57	0.02
side chains	water to vapor, ΔG	vapor pressures[f]	19	0.90	<0.0000001

[a] Based on linear regression of plots similar to that of Figure 2. [b] Nozaki & Tanford (1971). [c] Bull & Breese (1974). [d] Fendler et al. (1975). [e] Gaffney et al. (1979). [f] This study.

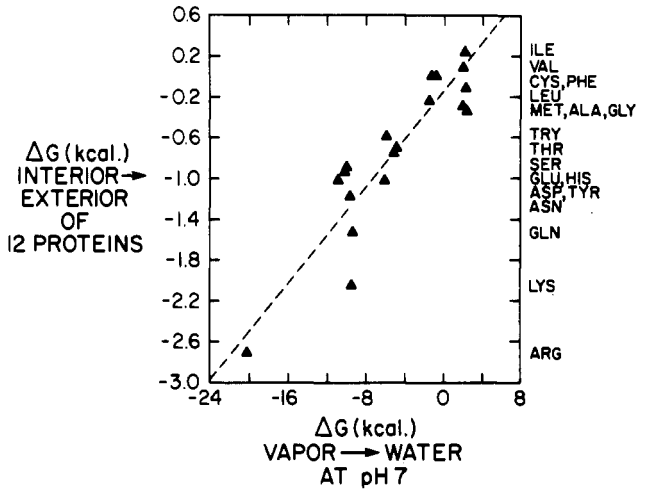


FIGURE 2: Free energies associated with access of amino acid residues to solvent in 12 proteins, plotted as a function of hydration potentials of the amino acid side chains.

results are obtained if Janin's values (Janin, 1979) are plotted against the present scale of hydration potentials.

It is of interest to compare the present scale of hydration potentials with the results of earlier measurements that were related in character. Free energies of transfer of amino acids near their isoelectric points in water, to ethanol or dioxane, were determined by comparing amino acid solubilities in two solvents (Nozaki & Tanford, 1971). Bull & Breese (1974) determined the tendency of amino acids near their isoelectric points to be transferred from the bulk to the surface of aqueous solutions, from surface tension lowering at various concentrations. Solubility measurements were used by Fendler et al. (1975) to determine free energies of transfer of amino acids from water to hexane at unspecified pH. A recent mass spectrometric study by Gaffney et al. (1979) yielded enthalpies of sublimation of some solid amino acids, which could be combined with heats of solution to yield enthalpies of transfer of amino acids (apparently without control of pH) from water to the vapor phase. As is shown in Table III, these studies yielded results that show significant correlation with protein folding when they are used as a basis for plots similar to that of Figure 2. Nevertheless, the confidence level of the present correlation is several orders of magnitude more favorable than any that is obtained by using the results of earlier studies (Table III).

Discrepancies between these scales, in their ordering of the amino acids and correlation with protein structures, probably arise in part from differences between conditions chosen for distribution experiments. Unrecognized self-association of amino acids, near their limits of solubility in organic solvents, may have tended in some cases to interfere with the determination of free energies of transfer of monomeric species between phases. Nor do most previous scales, in comparing the different amino acids, refer to conditions of constant pH. The nature of the nonpolar solvent chosen as a reference for distribution studies may also have contributed to these differences. Tryptophan, for example, is more hydrophobic than phenylalanine as judged from its distribution between water and ethanol (Nozaki & Tanford, 1971), but this order is reversed when the surface of solutions is chosen as the nonpolar reference condition (Bull & Breese, 1974).

In addition, the nature of the compounds chosen for comparison appears to be important. Tryptophan, for example, is found to surpass phenylalanine and leucine in its equilibrium constant for transfer from water to ethanol. This is unexpected in view of the solubility properties of indole, which is sparingly soluble in water and ethanol. Benzene is sparingly soluble in water but miscible with ethanol in all proportions, and aliphatic hydrocarbons exhibit the same tendency. Nor can this be simply a matter of the relative resistance of these solvents to the formation of cavities of different sizes for different solutes. Naphthalene, with space-filling requirements that resemble those of indole, shows behavior comparable with that of benzene (Windholz et al., 1976). Evidently, the relative hydrophobic character of indole and benzene becomes inverted when they are incorporated into the side chains of zwitterionic amino acids. It seems reasonable to suppose that the presence of so disruptive a local influence as a pair of electrostatic charges on the α-ammonium and α-carboxylate groups may tend to give a misleading impression of the relative solvation properties of amino acid residues as they occur in proteins.

Specific short-range interactions are undoubtedly operative on individual residues packed in the interior of globular proteins, and it would be naive to suppose that any solvent could provide even an approximate view of the microenvironment experienced by any particular residue buried in the interior of a globular protein. But in the larger view provided by the statistical approach to solvent accessibility, it is of interest that "representative" amino acids are ordered in a manner that resembles their relative tendencies to pass from the vapor phase or from hexane to solvent water. Specific forces of attraction and repulsion may be at work in alcohol, dioxane, or water surfaces, leading to relatively weak correlations with protein folding.

Relationship between Hydration Potentials and the Genetic Code. Following the initial observations of Nirenberg et al. (1963) suggesting that amino acids with similar properties have similar codons, Woese et al. (1966) noted the existence of a relationship between the sensitivity of amino acid R_f values to solvent water content in paper chromatographic systems and the composition of their coding triplets. The meaning of chromatographic mobility is not easily interpreted in thermodynamic terms, nor is it a simple matter to determine or

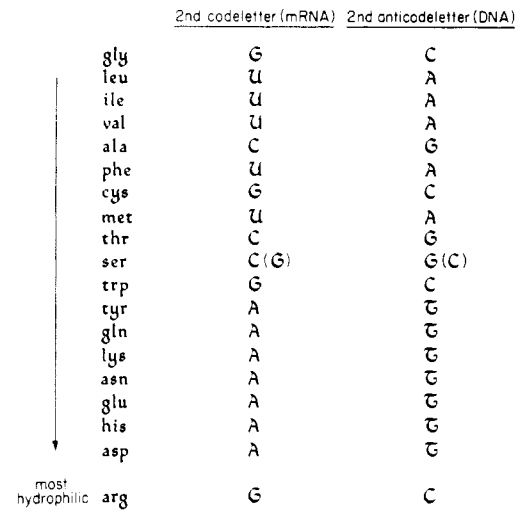


FIGURE 3: Second code letters for amino acids, arranged in order of increasingly negative hydration potentials of their side chains.

control pH or ionic strength in these systems. The present scale of hydration potentials provides another opportunity to inquire whether there is any systematic relationship between the composition of the genetic code and the affinities of amino acid side-chains for solvent water. The code is moderately degenerate at the first position and highly degenerate at the third position of code words for the various amino acids, but every amino acid except serine is invariably associated with the presence of one and the same base at the second position. Comparing the natural code with computer-generated random codes, Alff-Steinberger (1969) noted that the second position of the natural codons appears to play a dominant role in determining various physical properties of amino acids. Somewhat similar conclusions were drawn by Salemme et al. (1977).

When the amino acids of the present study are arranged in order of increasing affinity for water (Figure 3), the corresponding second code letters show an evident bias. The more hydrophilic amino acids all have purines in their codons. Also conspicuous is the fact that seven amino acids have A as their second code letter. All of them are hydrophilic, and they are clustered in a similar region on both axes of Figure 2. At the first and third positions of the code (not shown) no such bias is evident.

It seems natural to suppose, with Sonneborn (1965), that coding similarities between amino acids with similar physical properties might have been helpful during evolution in tending to maintain the structural stability of globular proteins. Using the conformation parameters of Chou & Fasman (1974), Goodman & Moore (1977) examined the differences between the conformational preferences of pairs of amino acids and concluded that there is little indication, at the level of secondary structure, that the genetic code was designed to minimize the effects of mutation on protein conformation. At the level of tertiary structure, on the other hand, it seems reasonable to suppose that the large differences between free energies of solvation of the various side chains, so closely associated with their distributions in real proteins, may have left their mark on the code. Figure 2 shows that the energetic cost of removing hydrophilic side chains from water is much greater than the cost of pulling hydrophobic side chains into water, and, indeed, it has been observed that hydrophobic residues occur rather often at the surfaces of proteins (Klotz, 1970; Lee & Richards, 1971). In contrast, hydrophilic residues are rarely found in the interior. Figure 2 suggest that their spontaneous appearance there would be seriously disruptive to the overall stability of a protein. It is therefore of interest that the code seems to have evolved in such a way that hydrophilic amino acids are grouped together and presumably tend to arise, by spontaneous mutation, from other hydrophilic amino acids.

Notwithstanding the strength of the apparent three-way relationship between hydration potential, physical distribution in globular proteins, and second code letter, we do not believe that the present findings lead to any obvious conclusions regarding mechanisms that might have determined the evolution of the code. Crick (1968) has offered the reasonable suggestion that the code may have evolved in an accidental manner from a primitive system in which a small number of amino acids and nucleotides interacted with each other, directly or indirectly, before the advent of an adapter system. It has sometimes been suggested that codon or anticodon assignments might be based on their selective affinities for the individual amino acids that might be identified by appropriate analysis of the modern code. A chemical basis for direct interaction, involving specific forces of attraction, is not readily apparent in view of the extreme structural diversity of the hydrophilic amino acids (Figure 3), all of which have A at the second position of their codons. An alternative possibility, suggested by Nagyvary & Fendler (1974), would have involved cocompartmentation of amino acids and nucleotides with similar distribution properties. Some recent discussions of this hypothesis (Weber & Lacey, 1978; Reuben & Polk, 1980) have been concerned with possible correlations based on the "hydrophilicities" of amino acids, defined in terms of their mobilities in paper chromatographic systems containing high concentrations of ammonium sulfate in water. The physical significance of these mobilities does not appear obvious, nor does their order bear much resemblance to that of the present scale of water affinities of the amino acid side chains. For further examination of the Nagyvary–Fendler hypothesis, it would be of interest to have detailed information about the solvation properties of the bases commonly found in nucleic acids.

Conclusions

The present findings provide strong support for the view that interactions of amino acid side chains with surrounding water have exerted an important influence on the folding of proteins and the evolution of the genetic code. These interactions are also doubtless significant in determining free energy changes associated with the development of very large changes in binding affinities associated with enzyme catalysis. Hydration potentials reflect the aggregate of many interactions between water and a solute in dilute aqueous solution, and it would be an oversimplification to suppose that all of these are broken when two solutes come together in water. Individual interactions between polar solutes and water show marked directional preferences; for example, the carbonyl group of peptides appears to bind water very much more strongly than does the NH group [Wolfenden (1978) and references cited therein]. A thorough understanding of specific interactions in a watery environment will require detailed information about these directional preferences, which should become available as experimental and theoretical approaches to solvation are refined.

References

Alff-Steinberger, C. (1969) *Proc. Natl. Acad. Sci. U.S.A. 64*, 584–591.

Bull, H. B., & Breese, K. (1974) *Arch. Biochem. Biophys. 161*, 665–670.

Butler, J. A. V. (1937) *Trans. Faraday Soc. 33*, 229–236.

Butler, J. A. V., & Ramchandani, C. N. (1935) *J. Chem. Soc.*, 952–960.
Chothia, C. (1976) *J. Mol. Biol. 105*, 1–14.
Chou, P. Y., & Fasman, G. D. (1974) *Biochemistry 13*, 211–245.
Crick, F. H. C. (1968) *J. Mol. Biol. 38*, 367–369.
Davis, T. L., & Elderfield, R. C. (1932) *J. Am. Chem. Soc. 54*, 1499–1503.
Deno, N. C., & Berkheimer, H. E. (1960) *J. Chem. Eng. Data 1*, 1–8.
Edsall, J. T., & Wyman, J. (1958) *Biophysical Chemistry*, pp 485–487, Academic Press, New York.
Eggleton, P., Elsden, S. R., & Gough, N. (1943) *Biochem. J. 37*, 526–534.
Federighi, E. T. (1959) *J. Am. Stat. Assoc. 54*, 683–688.
Fendler, J., Nome, F., & Nagyvary, J. (1975) *J. Mol. Evol. 6*, 215–232.
Friedel, J., & Orchin, M. (1951) *Ultraviolet Spectra of Aromatic Compounds*, Wiley, New York.
Gaffney, J. F., Pierce, R. C., & Friedman, L. (1977) *J. Am. Chem. Soc. 99*, 4293–4300.
Goodman, M., & Moore, G. W. (1977) *J. Mol. Evol. 10*, 7–47.
Hall, N. F., & Sprinkle, M. R. (1932) *J. Am. Chem. Soc. 54*, 3469–3485.
Hine, J., & Weimar, R. D. (1965) *J. Am. Chem. Soc. 87*, 3387–3396.
Hine, J., & Mookerjee, P. K. (1975) *J. Org. Chem. 40*, 292–298.
Hinman, R. L., & Lang, J. (1964) *J. Am. Chem. Soc. 86*, 3796–3806.
Janin, J. (1979) *Nature (London) 277*, 491–492.
Kauzmann, W. (1959) *Adv. Protein Chem. 16*, 1–63.
Kebarle, P. (1976) *Jerusalem Symp. Quantum Chem. Biochem. (1975) 8*, 81–94.
Klotz, I. M. (1970) *Arch. Biochem. Biophys. 138*, 704–720.
Lee, B., & Richards, F. M. (1971) *J. Mol. Biol. 55*, 379–400.
McAuliffe, C. (1966) *J. Phys. Chem. 70*, 1767–1782.
Nagyvary, J., & Fendler, J. H. (1974) *Origins Life 5*, 357–362.
Nandi, P. K. (1976) *Int. J. Pept. Protein Res. 8*, 253–264.
Nemethy, G. (1967) *Angew. Chem., Int. Ed. Engl. 30*, 1107–1120.
Nirenberg, M. W., Jones, O. W., Leder, P., Clark, F. C., Sly, S., & Petska, S. (1963) *Cold Spring Harbor Symp. Quant. Biol. 28*, 549–557.
Nozaki, Y., & Tanford, C. (1971) *J. Biol. Chem. 246*, 2211–2217.
Parsons, G. H., Rochester, C. H., Rostron, A., & Sykes, P. C. (1972) *J. Chem. Soc., Perkin Trans. 2*, 136–140.
Perutz, M. (1965) *J. Mol. Biol. 13*, 646–668.
Reuben, J., & Polk, F. E. (1980) *J. Mol. Evol. 15*, 103–112.
Salemme, F. R., Miller, M. D., & Jordan, S. R. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 2820–2824.
Shrake, A., & Rupley, J. A. (1973) *J. Mol. Biol. 79*, 351–372.
Sonneborn, R. M. (1965) *Evol. Genes Proteins Symp., 1964*, 377–397.
Tanford, C. (1962) *J. Am. Chem. Soc. 84*, 4240–4247.
Tse, Y.-C., Newton, M. D., Vishveshwara, S., & Pople, J. A. (1978) *J. Am. Chem. Soc. 100*, 4329–4331.
Weast, R. C., Ed. (1972) *CRC Handbook of Chemistry and Physics*, Chemical Rubber Co., Cleveland, OH.
Weber, A. L., & Lacey, J. C. (1978) *J. Mol. Evol. 11*, 199–210.
Weidenhagen, R., & Herrmann, R. (1935) *Chem. Ber. 68*, 1953–1961.
Windholz, M., Budavari, S., Stroumtsos, L. Y., & Fertig, M. N., Eds. (1976) *The Merck Index*, 9th ed., Merck & Co., Rahway, NJ.
Woese, C., Dugre, D. H., Saxinger, W. C., & Dugre, S. A. (1966) *Proc. Natl. Acad. Sci. U.S.A. 55*, 966–974.
Wolfenden, R. (1976) *J. Am. Chem. Soc. 98*, 1987–1988.
Wolfenden, R. (1978) *Biochemistry 17*, 201–204.
Wolfenden, R., & Lewis, C. A. (1976) *J. Theor. Biol. 59*, 231–235.
Wolfenden, R., Cullis, P. M., & Southgate, C. C. F. (1979) *Science (Washington, D.C.) 206*, 575–577.