Octan-1-ol–Water Partition Coefficients of Zwitterionic α-Amino Acids. Determination by Centrifugal Partition Chromatography and Factorization into Steric/Hydrophobic and Polar Components

Nabil El Tayar, Ruey-Shiuan Tsai, Pierre-Alain Carrupt and Bernard Testa *
Institut de Chimie Thérapeutique, Ecole de Pharmacie, Université de Lausanne, BEP, CH-1015 Lausanne, Switzerland

The distribution coefficients of free α-amino acids have been measured around their isoelectric value by centrifugal partition chromatography (CPC), allowing the determination of the partition coefficient (log *P*) of the amino acids in their zwitterionic form. Good correlations with published distribution coefficients of amino acid derivatives allowed log *P* values to be calculated for the five amino acids (Arg, Asn, Asp, Glu, Lys) whose high polarity prevented direct measurements by CPC. The log *P* values were factorized into a steric component (molecular volume *V*, mainly accounting for inductive and hydrophobic forces) and a polarity factor (*Λ*, accounting for ion–dipole and dipole–dipole interactions and hydrogen-bonds) which correlates with other amino acid parameters expressing mainly polar interactions. Using two examples from the literature, we show that multiple linear regression analysis based on steric and polar parameters is able to afford a quantitative interpretation of factors influencing protein conformational stability.

The words lipophilicity and hydrophobicity, despite referring to distinct physical properties, have been used rather loosely in the literature. Lipophilicity, as expressed by the partition coefficient (log *P*), is a molecular parameter describing a partitioning equilibrium of solute molecules between water and an immiscible lipid-like organic solvent. Many studies have clearly and repeatedly demonstrated that lipophilicity encodes two major structural contributions, namely a steric or bulk term reflecting hydrophobic and dispersive forces, and polar terms reflecting dipole–dipole interactions and hydrogen bonds.[1–7] At equilibrium, a balance between all solute–solvent interactions is reached. In consistency with this analysis, the concept of hydrophobicity should be restricted to the balance of hydrophobic and dispersive forces, *i.e.* those elicited by non-polar groups. Although the driving force for hydrophobic interactions is still debated,[8–11] it is now well-established that the hydrophobicity of non-polar solutes is proportionally and solely related to a bulk term such as molar volume. To quantify this hydrophobic force, many scales of hydrophobicity for free amino acids and amino acid side-chains have been developed. These hydrophobicity scales were evaluated by theoretical calculations of the van der Waals volume[12] and the water-accessible surface area.[13–17]

Unlike hydrophobicity, lipophilicity provides additional structural information related to polar interactions. Several authors have measured indirectly the lipophilic character of amino acids. For example, relative solubilities of amino acids in water, ethanol and dioxane were used to derive, for the first time, a lipophilic scale (Δf_t) for free amino acids.[18] Fauchère and Pliska[19] have proposed another lipophilicity scale (π_R^*) for twenty naturally occurring amino acid side-chains based on the partitioning of N^α-acetyl-L-amino-acyl amides in the octan-1-ol–water solvent system. The main feature of these amino acid derivatives is that they mimic the situation in peptides and proteins in which the amino acid side-chain is flanked by two peptide bonds. Using analogues of amino acids, Wolfenden and co-workers[20,21] have determined the free energy of transfer of amino acid side-chains between water and the vapour phase (ΔG_{vap}) and between water and cyclohexane (ΔG_{cyc}). These values were considered as a measure of water affinity (and hence of polarity).

Lipophilicity is measured experimentally as a partition coefficient, which refers to a single, well-defined chemical form (*e.g.*, a neutral form of an ionizable compound, or a zwitterion). Partitioning resulting from the effect of more than one form of a compound is expressed as distribution coefficients (*i.e.* 'apparent' partition coefficients) which vary with pH. The lipophilicity of some free amino acids has been assessed by several experimental techniques such as the shake-flask technique, RP-HPLC and RP-TLC[22–26] but these methods suffer from a number of practical limitations and disadvantages due to various perturbating factors.[27] Recently, measurements of partition coefficients of various compounds by hydrodynamic equilibrium Centrifugal Partition Chromatography (CPC) which employs a purely liquid–liquid partitioning process have been successfully developed in our laboratory.[28,29] This method has several advantages over the classical technique for its precision, reproducibility, and decreased interference by impurities; it also allows the measurement of partition coefficients of very polar compounds (log *P* down to −3.3.).

In this study, we apply this CPC technique to determine the partition coefficient of zwitterionic free amino acids near their isoelectric points in an octan-1-ol–water system. Molecular volume (*V*) and water-accessible surface area (ASA) of free amino acids were also calculated. A comparison between these scales of free amino acids and currently available ones is the main purpose of our study. In addition, an interpretation of structural factors influencing the conformational stability of some water-soluble proteins is suggested.

Experimental

Materials.—The L-amino acids were purchased from Fluka (Buchs, Switzerland) or Sigma (St. Louis, Missouri, USA) and were of >99% purity. 2-Morpholinoethanesulfonic acid (MES, >99% purity) and 3-morpholinopropanesulfonic acid (MPS, >99% purity) were from Merck (Darmstadt, Germany), and octan-1-ol (purity *ca.* 98%) from Fluka. All compounds were used without further treatment.

Measurement of Distribution Coefficients (log *D*).—Measurements of distribution coefficients (*i.e.*, 'apparent' partition coefficients) in octan-1-ol–aqueous buffer systems were performed at three pH values by horizontal flow-through

centrifugal partition chromatography using a coil planet type centrifuge. The design principle of the instrument has been described elsewhere.[30,31] The apparatus from Pharma-Tech Research Corporation (Baltimore, Maryland, USA) used three columns, each of which was helically wound with five layers of PTFE tubing (3.00 mm ID, 3.94 mm OD). The three columns made a total volume capacity of 350 cm^3. A Kontron model 420 HPLC pump was used to propel the solvent, and a Kontron model 432 UV–VIS detector coupled with a Hewlett-Packard 3392A integrator to detect the solutes. A flowmeter from Deeside Industrial Estate (Queensferry, UK) was used for precise measurement of flow rates.

Measurements began by filling the columns with octan-1-ol pre-saturated with 0.05 mol dm^{-3} MES or MPS buffer depending upon the pH used. While the columns were revolving at a speed of 1000 rpm along the central axis, they were also rotating along their own axis in a mode of planetary motion. The mobile phase (aqueous phase) was then pumped into the columns in a 'head-to-tail' mode during the rotation. Depending upon expected distribution coefficients, the flow rate of the mobile phase was varied from 0.5–6.0 cm^3 min^{-1} (flow rates of 0.5 cm^3 min^{-1} for compounds with log $D < -2.3$; 1 cm^3 min^{-1} for those with log D between -2.3 and -1.3; and 6 cm^3 min^{-1} for those with log D between -1.3 and 0). A Merck injector was used to inject samples of 200 mm^3 (aqueous buffer, 1–50 mmol dm^{-3}). All the amino acids were detected at 200 nm, and all measurements were performed at 25 ± 0.1 °C in triplicate. Concentration effects were negligible, the difference in calculated log D values being <0.05 units.

Under flow rates of 0.5, 1 and 6 cm^3 min^{-1}, *ca.* 310, 305 and 270 cm^3 of the stationary phase were retained, respectively. The retention time of the solvent front (t_0) was measured by non-retained solutes (potassium dichromate or lysine). It follows that the distribution coefficient can be calculated by eqn. (1),

$$\log D = \log \frac{(t_R - t_0) \cdot U}{V_t - U \cdot t_0} \qquad (1)$$

where t_R is the retention time of the solute, U is the flow rate of the mobile phase, and V_t is the total capacity of the columns. The lower log D limit was -3.3.

Since the isoelectric points of most amino acids included in this study are within the range of 5.0–7.6, their distribution coefficients were measured at three different pH values, 5.0, 6.0 and 7.4. The highest log D value was taken to represent only the zwitterionic form and was equated with the partition coefficient (*i.e.* log P) of this form. For histidine, log D values at pH 7.2 and 7.8 were also measured. Our method failed to measure the distribution coefficients of glutamic acid, lysine, arginine, asparagine and aspartic acid owing to their very low distribution coefficients (log $D < -3.3$).

Theoretical Calculations and Statistics.—Molecular volume (V) and water-accessible surface area (ASA) of free amino acids were calculated by the program MOLSV (QCPE 509) running on a Silicon Graphics Personal Iris Workstation using van der Waals radii and interatomic distances of free amino acids in their zwitterionic form according to Gavezzoti,[32] except that the van der Waals radius of the water molecule was taken as 1.5 Å. Multiple linear regression analysis was performed using the QSAR Regression program (MEDCHEM Project, Pomona College, California, USA) running on a VAX 9000 computer. Cross-validation procedure was performed using the QSAR module of SYBYL (Tripos Associates, Inc., St. Louis, Missouri, USA) running on a Silicon Graphics Personal Iris Workstation.

In cross-validation procedures, the analysis is repeated omitting one compound at a time, and the resulting equation is used to predict the target property of the omitted compound. The standard errors of cross-validated predictions are calculated as the sum of squares of deviations of the observed *vs.* predicted values. This procedure is repeated until every compound in the data set has been predicted exactly once. A cross-validated r^2 is defined as eqn. (2), where SD is the sum of

$$r^2_{CV} = (SD - PRESS)/SD \qquad (2)$$

squares of deviations of the observed values from their mean and PRESS is the prediction error sum of squares. It is important to note the r^2_{CV} is considered as a real measure of the predictive power of the equation. For more detail, the reader is referred to the original study of Wold and co-workers.[33]

Results and Discussion

Distribution Coefficients of Amino Acids Near Their Isoelectric Point.—For zwitterionic molecules, distribution coefficients form a broad plateau when measured at pH values near this isoelectric point. For example, the distribution coefficient of zwitterionic peptides was shown to form a plateau over a few pH units and to drop outside this pH range due to the increased population of positively or negatively charged species.[34] In the present study, the distribution coefficients measured at three pH values showed little or no variation with pH (Table 1); the highest log D value, measured near the isoelectric point, was retained as log D^I = log P, the partition coefficient of the zwitterion. However, for histidine with an ionizable imidazole ring, log D values vary significantly in the low pH range. For this reason, the distribution coefficients of histidine at pH 7.2 and 7.8 were also measured. These log D values, including log D at pH 7.4, showed no variation with pH, and log D at 7.4 was retained as log P.

Correlations between our log P values and distribution coefficients measured at pH 7 and taken from the literature[22,23] are given in eqns. (3) and (4), respectively, where log D^7 is the

$$\log P = 0.96(\pm 0.20) \log D^7 - 0.23(\pm 0.44) \qquad (3)$$
$$n = 7 \quad r^2 = 0.969 \quad r^2_{CV} = 0.937 \quad s = 0.122 \quad F = 153.8$$

$$\log P = 0.94(\pm 0.12) \log D^7 - 0.13(\pm 0.30) \qquad (4)$$
$$n = 6 \quad r^2 = 0.992 \quad r^2_{CV} = 0.964 \quad s = 0.06 \quad F = 478$$

distribution coefficient at pH 7, n the number of compounds, r^2 the correlation coefficient, r^2_{CV} the cross-validated correlation coefficient, s the standard deviation of the regression and F the Fischer test of statistical significance of the equation. Values in parentheses are the 95% confidence limits. These equations indicate a good degree of agreement. Histidine is excluded from eqn. (4) due to an unreliably high log D^7 value reported by Younger and Cramer.[23] Indeed, Fauchère *et al.*,[19] have observed that radioactive histidine can be contaminated with some lipophilic radioactive material.

Obviously, scattered and incomplete literature data on the distribution coefficients of free amino acids in the octan-1-ol–water system (log D) reflect experimental difficulties due to their zwitterionic/dipolar nature and their very low solubility in organic solvents. Another problem with literature data is that all distribution coefficients were determined at a fixed pH; thus, the maximal population of the zwitterionic form is not assured as isoelectric points of amino acids vary with the ionization constants of the CO_2H and NH_2 groups.[35] It should be noted that our method failed to measure the distribution coefficients of glutamic acid, lysine, arginine, asparagine and aspartic acid due to their very low lipophilicity (log $D < -3.3$).

Comparison between Lipophilicity Scales of Amino Acid Side-chains.—To allow comparison, a lipophilicity scale (π_R) of

Table 1 Octan-1-ol–water distribution coefficients of free amino acids at various pH values

Amino acid	$\log D^5$	$\log D^6$	$\log D^{7.4}$	pI [a]	$\log D^{1}$ [b] ($\log P$)	$\log D^{7}$ [c]	$\log D^{7}$ [d]
Ala	−2.78	−2.77	−2.79	6.02	−2.77	−2.74	−2.72
Arg	*e*	—	—	10.76	(−3.79)	−4.08	—
Asn	—	—	—	5.43	(−3.48)	—	—
Asp	—	—	—	2.98	(−3.61)	—	—
Cys	−2.55	−2.57	−2.61	5.02	−2.55	—	—
Gln	−3.11	−3.15	−3.15	5.65	−3.11	—	—
Glu	—	—	—	3.08	(−3.51)	−3.69	—
Gly	−3.02	−3.00	−3.01	5.97	−3.00	—	−3.00
His	−3.33	−3.27	−2.85	7.59	−2.85 [f]	−1.95	−2.52
Ile	−1.82	−1.80	−1.82	6.02	−1.80	—	−1.69
Leu	−1.72	−1.72	−1.72	5.98	−1.72	—	−1.52
Lys	—	—	—	9.47	(−3.77)	−3.05	—
Met	−2.10	−2.10	−2.10	5.75	−2.10	—	−1.87
Phe	−1.44	−1.45	−1.45	5.48	−1.44	—	−1.35
Pro	−2.64	−2.62	−2.65	6.30	−2.62	−2.54	—
Ser	−3.08	−3.00	−3.02	5.68	−3.00	−3.07	—
Thr	−2.99	−2.83	−2.86	6.53	−2.83	−2.94	—
Trp	−1.15	−1.15	−1.16	5.89	−1.15	−1.11	—
Tyr	−2.11	−2.11	−2.11	5.66	−2.11	—	—
Val	−2.29	−2.29	−2.29	5.97	−2.29	−2.26	—

[a] Isoelectric points of amino acids taken from Jencks and Regenstein.[61] [b] Distribution coefficients at isoelectric point, taken as the highest $\log D$ value in a range of pH values covering the isoelectric point and equated with $\log P$ of the zwitterion. Data in parentheses are calculated from π_R values themselves calculated from eqn. (6). [c] Distribution coefficients at pH 7, reported by Younger and Cramer.[23] [d] Distribution coefficients at pH 7, reported by Klein *et al.*[22] [e] Not measurable by CPC method. [f] For His, distribution coefficients were also measured at pH 7.2 ($\log D = -2.84$) and at pH 7.8 ($\log D = -2.85$).

Table 2 Various lipophilicity and polarity scales of amino acid side-chains and steric parameter of free amino acids

Amino acid	π_R [a]	π_R^* [b]	Δf_t [c]/kcal mol^{-1}	ΔG_{vap} [d]/kcal mol^{-1}	ΔG_{cyc} [e]/kcal mol^{-1}	ΔG_R [f]/kcal mol^{-1}	ΔG_h [g]/kcal mol^{-1}	Λ [h]	V [k]/Å^3	ASA [m]/Å^3
Ala	0.23	0.31	−0.50	1.94	1.81	−0.55	−0.54	−0.02	82.20	254.20
Arg	−0.79 [j]	−1.01	—	−19.92	−14.92	2.00	−5.96	−2.56	163.00	363.40
Asn	−0.48 [j]	−0.60	—	−9.68	−6.64	0.51	−3.55	−1.24	112.30	303.60
Asp	−0.61 [j]	−0.77	—	−10.95	−8.72	1.20	−2.97	−1.08	103.70	287.90
Cys	0.45	1.54	—	−1.24	1.28	−1.40	−1.64	−0.11	99.10	282.90
Gln	−0.11	−0.22	—	−9.38	−5.54	0.29	−3.92	−1.19	127.50	335.00
Glu	−0.51 [j]	−0.64	—	−10.24	−6.81	0.76	−3.71	−1.43	120.50	311.60
Gly	0.00	0.00	0.00	2.39	0.94	0.00	−0.59	0.03	65.00	224.90
His	0.15	0.13	−0.50	−10.27	−4.66	−0.25	−3.38	−1.06	140.60	337.20
Ile	1.20	1.80	—	2.15	4.92	−2.10	0.32	0.04	131.70	322.60
Leu	1.28	1.70	−1.80	2.28	4.92	−2.00	0.27	0.12	131.50	324.00
Lys	−0.77 [j]	−0.99	—	−9.52	−5.55	0.78	−2.19	−2.26	144.30	336.60
Met	0.90	1.23	−1.30	−1.48	2.35	−1.60	−0.60	−0.33	132.30	336.30
Phe	1.56	1.79	−2.50	−0.76	2.98	−2.60	−1.06	−0.05	155.80	366.10
Pro	0.38	0.49	—	—	—	−1.50	0.32	−0.31	106.70	288.50
Ser	0.00	−0.04	0.30	−5.06	−3.40	0.09	−3.82	−0.40	88.50	266.70
Thr	0.17	0.26	−0.40	−4.88	−2.57	−0.58	−1.97	−0.53	105.30	283.90
Trp	1.85	2.25	−3.40	−5.88	2.33	−2.70	−3.80	−0.31	185.90	401.80
Tyr	0.89	0.96	−2.30	−6.11	−0.14	−1.70	−5.64	−0.84	162.70	377.80
Val	0.71	1.22	−1.50	1.99	4.04	−1.60	0.13	−0.13	115.60	295.10

[a] Lipophilicity constant of amino acid side-chain (R) as determined in this study [eqn. (5)]. [b] Lipophilicity constant of amino acid side-chain, taken from Fauchère and Pliska.[19] [c] Free energy of transfer of amino acid side-chain from pure organic solvent to water, relative to Gly as reported by Nozaki and Tanford.[18] [d] Free energy of transfer of amino acid side-chain from the vapour phase to water, taken from Radzicka and Wolfenden.[21] [e] Free energy of transfer of amino acid side-chain from cyclohexane to water, taken from Radzicka and Wolfenden.[21] [f] Free energy of transfer of amino acid side-chain R from protein interior to water, taken from Eisenberg and McLachlan.[40] [g] Hydration free energy of unfolding of amino acid side-chain, taken from Oobatake and Ooi.[42] [h] Polarity parameter calculated by eqn. (7). [j] Values estimated from eqn. (6). [k] Molecular volume of amino acids in their zwitterionic form, calculated by the MOLSV program. [l] Water-accessible surface area of amino acids in their zwitterionic form, calculated by the MOLSV program.

amino acid residues based on $\log P$ values of amino acids in their zwitterionic form has been determined as eqn. (5),[36]

$$\pi_R = \log P[^+H_3NCH(R)CO_2^-] - \log P[^+H_3NCH_2CO_2^-] \quad (5)$$

assuming that $\pi_H = 0$. This parameter is reported in Table 2 along with some lipophilic scales of amino acid side-chains. Fig. 1 displays the relationship between π_R and the π_R^* parameter reported by Fauchère and Pliska[19] and based on octan-1-ol–water partitioning of N^α-acetyl-L-amino-acyl amides. This figure clearly illustrates a deviant partitioning behaviour (*i.e.* higher than expected π_R^* values) of relatively non-polar bulky amino acid residues such as Val, Met, Ile, Leu, Phe and Trp, compared to the behaviour of polar amino acid residues. This discrepancy must reflect different intramolecular interactions of

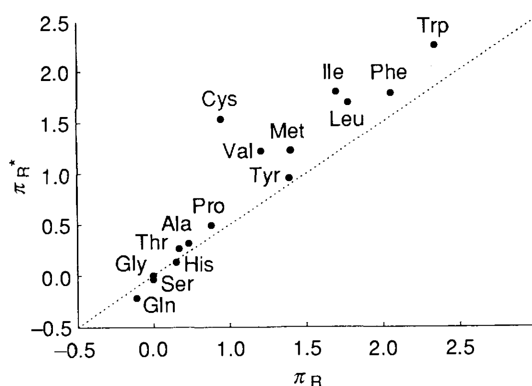

Fig. 1 Relationship between the lipophilic scale (π_R) from octan-1-ol–water distribution coefficient of zwitterionic amino acids and the lipophilic scale (π_R^*) from octan-1-ol–water distribution coefficient of N^α-acetyl-L-amino-acyl amides.[19] The dotted line represents $\pi_R = \pi_R^*$.

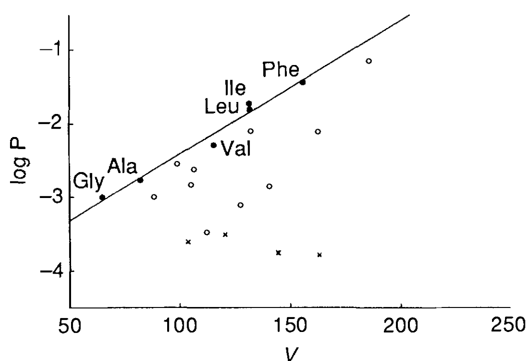

Fig. 2 The relationship between the log P and molecular volume (V) of zwitterionic amino acids. Filled circles (●) refer to amino acids with non-polar side-chain, open circles (○) to amino acids with polar side-chain, and crosses (×) to amino acids with ionizable side-chain, where log P was calculated as explained in Table 1.

the side-chain in zwitterionic free amino acids *versus* N^α-acetyl-L-amino-acyl amides. In fact, it is well-known that the zwitterionic $^-OOC–CH–NH_3^+$ moiety is more effective than the dipolar $CH_3CONHCHCONH_2$ moiety in exerting proximity effects which decrease the lipophilic character of non-polar side-chains.[9,37,38] Nevertheless, excluding cysteine, a global linear relationship between π_R and π_R^* is obtained, given in eqn. (6).

$$\pi_R = 0.76(\pm 0.09)\ \pi_R^* + 0.01(\pm 0.10) \quad (6)$$
$$n = 14 \quad r^2 = 0.968 \quad r^2_{CV} = 0.956 \quad s = 0.12 \quad F = 364.4$$

The anomalously high π_R^* value of cysteine presumably results from experimental difficulties during the lengthy shake-flask procedure (Fauchère, personal communication). Using eqn. (6) the π_R values of arginine, asparagine, aspartic acid, glutamic acid and lysine were estimated (Table 2) and from them log P values can also be calculated. In analogy with eqn. (6), π_R is found to be linearly related ($r^2 = 0.922$) to the free energy of transfer (ΔG_R) of amino acid side-chains from the interior of a protein to an aqueous environment[39–41] and to the free energy of transfer (Δf_t) of amino acid side-chains from pure organic solvent to water ($r^2 = 0.917$).[18] Conversely, no correlations between π_R and other lipophilic scales such as ΔG_{cyc} (the free energy of transfer of amino acid side-chain from cyclohexane to water), ΔG_{vap} (the free energy of transfer from vapour phase to water) and ΔG_h (the free energy of hydration) were obtained.[21,42,43] The absence of correlations may be related to the fact that ΔG_{cyc}, ΔG_{vap} and ΔG_h scales measure affinities of amino acid side-chains for water, and not a true partitioning between water and an amphiphilic polar organic solvent. A similar criticism was raised by Rose *et al.*[17] and Ooi *et al.*[43]

Polarity Scales.—*A priori*, only partial correlations are expected between lipophilicity and steric parameters. However, for non-polar solutes, several studies have clearly shown that lipophilicity, as expressed by partition coefficients (log P), correlates linearly with molar volume.[1,3,44] Polar solutes systematically deviate from such linearity due to polar interactions and these deviations have been taken as a measure of all polar interactions. In a previous study, we have designated as Λ (capital lambda, chosen simply because it is written as an upside-down V) the sum of these polar interactions,[3,45] *i.e.* eqn. (7).

$$\log P = aV + b\Lambda \quad (7)$$

An analysis of the physicochemical information content of this term Λ demonstrated that it mainly related to the hydrogen-bond acceptor basicity and polarizability of solutes.[45] In the present study, a similar approach was applied to amino acids as shown in Fig. 2. For the six amino acids with non-polar side-chains, a linear relationship was obtained between log P and the van der Waals' volume V given by eqn. (8).

$$\log P = 0.018(\pm 0.003)V - 4.22(\pm 0.41) \quad (8)$$
$$n = 6 \quad r^2 = 0.980 \quad r^2_{CV} = 0.967 \quad s = 0.10 \quad F = 207.8$$

From eqns. (7) and (8) and Fig. 2, the value of the polarity parameter Λ was calculated for the 14 amino acids with polar side-chains, its value for the 6 amino acids with non-polar side-chains being in the range −0.13 to 0.12 (Table 2). Calculated in such a manner, the Λ parameter gives a measure of the polarity of the side-chain and not of the whole molecule of amino acids. It is important to note that Λ values for Arg, Asp, Glu and Lys encode, in addition to hydrogen-bond acceptor basicity and polarizability, ion–dipole interactions as a result of the presence of an ionizable functional group in their side-chains.

This Λ parameter correlates well with the free energy of transfer from cyclohexane to water (ΔG_{cyc}) and the free energy of transfer from vapour phase to water (ΔG_{vap}) ($r^2 = 0.80$ and 0.84, respectively). It is therefore obvious that these parameters (Λ, ΔG_{vap} and ΔG_{cyc}) mainly characterize the polar interactions of amino acid side-chains with an aqueous environment. Interestingly, the Λ parameter is highly and negatively correlated ($r^2 = 0.89$) with the average water-accessible surface area of residues in folded globular proteins (ASA_f) as calculated from X-ray crystallographic data by Rose *et al.*[17] This can be taken to imply that polar side-chains have a tendency to be located at the surface of globular proteins and hence to be maximally exposed to water, the reverse applying to non-polar side-chains.[15–17]

An Interpretation of the Conformational Stability of some Water-Soluble Proteins.—The importance of hydrophobicity is repeatedly demonstrated in the folding of proteins and the self-organization of membrane receptors into cell membrane.[46,47] In an earlier study, Kauzmann[48] pointed out that the hydrophobicity of amino acid side-chains is the dominant force in the folding of some proteins. A number of more recent studies strongly support this view.[49–52] Site-directed mutagenesis experiments, *i.e.*, replacement of a given amino acid residue in a particular position of a protein with different amino acids, allow a comparison of structural factors influencing the stability of secondary and tertiary structures. The contributions of hydrophobic and electrostatic interactions and hydrogen bonding to the conformational stability of proteins are well documented in the literature.[46,53–59] However, no statistical

Table 3 Observed values of free energies of unfolding of mutant proteins

Amino acid	$\Delta\Delta G^{LY\,a}$/ kcal mol^{-1}	$\Delta_d G^{TS\,b}$/ kcal mol^{-1}
Ala	−1.1	8.5
Arg	*c*	—
Asn	—	8.2
Asp	−1.8	8.5
Cys	−0.4	11.0
Gln	—	6.3
Glu	−1.1	8.8
Gly	−1.8	7.1
His	—	10.1
Ile	0.0	16.8
Leu	0.9	15.0
Lys	—	7.9
Met	−0.3	13.3
Phe	−1.0	11.2
Pro	—	8.2
Ser	−1.9	7.4
Thr	−1.7	8.8
Trp	−3.6	9.9
Tyr	−2.7	8.8
Val	−0.6	12.0

[a] The difference between the free energy of unfolding of the mutant protein and that of wild type at the melting temperature of wild-type lysozyme, taken from Matsumura *et al.*[53] [b] The unfolding free energy in water of 19 proteins substituted by each of the 19 amino acids at the same position of tryptophan synthase α subunit, taken from Yutani *et al.*[58] [c] Not measured.

evaluation of the relative contributions of these interactions and their nature (*i.e.*, favourable or unfavourable) has been made.

Recently, the conformational stability of mutant forms of bacteriophage T4 lysozyme (substitution of isoleucine residue at position 3 with 13 different amino acids) and tryptophan synthase α subunit (substitution of amino acid residue at position 49 with 19 different amino acids) were measured thermodynamically by Matsumura *et al.*[53] and Yutani *et al.*,[58] respectively. The results showed that the conformational stability of these mutant proteins increases linearly with increasing lipophilicity of the side-chains, except for aromatic residues (Phe, Trp and Tyr). However, no mechanistic interpretation could be derived in terms of intermolecular forces.

In the present study, an attempt has been made to rationalize the relationship between structural properties of amino acid residues and their role in the conformational stability of the two proteins mentioned above.[53,58] Their stability data were used (Table 3) and correlated by multiple linear regression analysis with various lipophilic, steric and polar scales of amino acid residues. Some of the statistically most significant equations are as follows.

For mutant bacteriophage T4 *lysozyme.*

$$\Delta\Delta G^{LY} = 0.164(\pm 0.061)V - 0.066(\pm 0.024)V^2 + 0.145(\pm 0.067)\Delta G_{vap} - 10.30(\pm 3.76) \quad (9)$$

$n = 14 \quad r^2 = 0.859 \quad r^2_{CV} = 0.740 \quad s = 0.49 \quad F = 20.3$

Optimum $V = 123.7$ Å^3

$$\Delta\Delta G^{LY} = 0.181(\pm 0.067)V - 0.074(\pm 0.027)V^2 + 1.39(\pm 0.75)\Lambda - 11.04(\pm 4.21) \quad (10)$$

$n = 14 \quad r^2 = 0.828 \quad r^2_{CV} = 0.587 \quad s = 0.54 \quad F = 16.0$

Optimum $V = 122.5$ Å^3

A comparable equation is afforded by ΔG_{cyc}. Normalization of eqn. (9), *i.e.*, the calculation of the relative contribution of each parameter to the explained variance[60] shows that the relative contributions to $\Delta\Delta G^{LY}$ of V^2 is 47.9%, of V 46.4% and of ΔG_{vap} 5.7%. The same contributions were observed when normalizing the other equations.

For tryptophan synthase α unit.

$$\Delta_d G^{TS} = 0.334(\pm 0.165)V - 0.127(\pm 0.065)V^2 + 0.515(\pm 0.174)\Delta G_{cyc} - 10.31(\pm 10.11) \quad (11)$$

$n = 18 \quad r^2 = 0.788 \quad r^2_{CV} = 0.572 \quad s = 1.42 \quad F = 17.3$

Optimum $V = 131.7$ Å^3

Comparable equations are afforded by ΔG_{vap} and Λ. Normalization of eqn. (11) shows that the relative contributions to $\Delta_d G^{TS}$ of V^2 is 44%, of V 45.4% and of ΔG_{cyc} 10.6%. The same contributions were observed when normalizing the other equations. The ΔG_{cyc} and ΔG_{vap} values of proline were not available. It is important to note that using other molecular volume-related parameters such as ASA gave also statistically significant equations (results not shown).

Eqns. (9)–(11) clearly show that an optimum in molecular volume of amino acid side-chain has a determining influence on the conformational stabilities of the two proteins. Beyond this optimum value of volume (*ca.* 120 Å^3 for T4 lysozyme and 140 Å^3 for tryptophan synthase), a decrease in the conformational stability is related to an unfavourable steric hindrance, *i.e.* limited space available for the side-chain in the interior of the macromolecule. In addition to hydrophobic interactions expressed by the V term, polar interactions, as expressed either by ΔG_{cyc}, ΔG_{vap} or Λ, influence with weak relative contributions (*ca.* 6% and 11%, respectively) and negatively the conformational stability of T4 lysozymes and tryptophan synthase. Indeed, the relevant regression coefficients in eqns. (9)–(11) are positive for polar descriptors which decrease with increasing polarity (Λ, ΔG_{vap}, ΔG_{cyc}).

As discussed above, polar interactions favour the tendency of polar residues to be located at the surface of protein molecules. It is therefore not surprising that polar interactions, in position 3 of lysozyme and in position 49 of tryptophan synthase, are detrimental for the folded conformation. For example, the volume of glutamic acid ($V = 124.3$ Å^3), a highly polar amino acid, is very close to the optimum volume for T4 lysozyme but its $\Delta\Delta G^{LY}$ is two units less compared to that of leucine. However, it is important to keep in mind that polar interactions, in particular hydrogen bonding, would favour the folding of polar residue should strong inter-residue hydrogen-bonds occur in the interior of the protein.

Conclusions

This study shows how the lipophilicity of amino acids, as expressed by their partition coefficients, is indeed the resultant of steric hydrophobic and polar solute–solvent interactions. From measured log P values and calculated molecular volumes, a polarity factor can be obtained which correlates with amino acid parameters if and when these express mainly polar interactions. Multiple linear regression based on such steric and polar parameters could afford a quantitative interpretation of factors influencing protein conformational stability, as documented by two examples.

Acknowledgements

B. T., N. E. T. and P. A. C. are indebted to the Swiss National Science Foundation for research grant No. 31-27531.89.

References

1 I. Y. Moriguchi, Y. Kanada and K. Komatsu, *Chem. Pharm. Bull.*, 1976, **24**, 1799.

2 R. D. Cramer III, *J. Am. Chem. Soc.*, 1980, **102**, 1837.
3 B. Testa and P. Seiler, *Arzneim.-Forsch.*, 1981, **31**, 1053.
4 H. J. M. Grünbauer and E. Tomlinson, *Int. J. Pharmaceut.*, 1984, **21**, 61.
5 R. W. Taft, M. H. Abraham, G. R. Famini, R. M. Doherty, J.-L. M. Abdoud and M. J. Kamlet, *J. Pharm. Sci.*, 1985, **74**, 807.
6 H. van de Waterbeemd and B. Testa, in *Advances in Drug Research*, ed. B. Testa, Academic Press, London, 1987, vol. 16, p. 87.
7 N. El Tayar, R.-S. Tsai, B. Testa, P.-A. Carrupt and A. Leo, *J. Pharm. Sci.*, 1991, **80**, 590.
8 J. H. Hildebrand, J. M. Prausnitz and R. L. Scott, in *Regular and Related Solutions*, Van Nostrand Reinhold, New York, 1970, p. 85.
9 C. Tanford, in *The Hydrophobic Effect: Formation of Micelles and Biological Membrane*, Wiley-Interscience, New York, 1973.
10 P. L. Privalov and S. J. Gill, *Adv. Protein Chem.*, 1988, **39**, 191.
11 N. Muller, *Acc. Chem. Res.*, 1990, **23**, 23.
12 J.-L. Fauchère, in *Advances in Drug Research*, ed. B. Testa, Academic Press, London, 1986, vol. 15, p. 29.
13 B. Lee and F. M. Richards, *J. Mol. Biol.*, 1971, **55**, 379.
14 A. Shrake and J. A. Rupley, *J. Mol. Biol.*, 1973, **79**, 351.
15 C. Chothia, *J. Mol. Biol.*, 1976, **105**, 1.
16 J. Janin, *Nature*, 1979, **277**, 491.
17 G. D. Rose, A. R. Geselowitz, G. J. Lesser, R. H. Lee and M. H. Zehfus, *Science*, 1985, **229**, 834.
18 Y. Nozaki and C. J. Tanford, *J. Biol. Chem.*, 1971, **246**, 2211.
19 J.-L. Fauchère and V. Pliška, *Eur. J. Med. Chem.*, 1983, **18**, 369.
20 R. Wolfenden, L. Andersson, P. M. Cullis and C. C. B. Southgate, *Biochemistry*, 1981, **20**, 849.
21 A. Radzicka and R. Wolfenden, *Biochemistry*, 1988, **27**, 1664.
22 R. A. Klein, M. J. Moore and M. W. Smith, *Biochim. Biophys. Acta*, 1971, **233**, 420.
23 L. M. Younger and R. D. Cramer, *Mol. Pharmacol.*, 1981, **20**, 602.
24 V. Pliška, M. Schmidt and J.-L. Fauchère, *J. Chromatogr.*, 1981, **216**, 79.
25 J. M. R. Parker, D. Guo and R. S. Hodges, *Biochemistry*, 1986, **25**, 5425.
26 R. Cowan and R. G. Whittaker, *Peptide Res.*, 1990, **3**, 75.
27 J. C. Dearden and G. M. Bresnen, *Quant. Struct.-Act. Relat.*, 1988, **7**, 133.
28 P. Vallat, N. El Tayar, B. Testa, I. Slacanin, A. Marston and K. Hostettmann, *J. Chromatogr.*, 1990, **504**, 411.
29 N. El Tayar, R.-S. Tsai, P. Vallat, C. Altomare and B. Testa, *J. Chromatogr.*, 1991, **556**, 181.
30 Y. Ito, *J. Chromatogr.*, 1981, **207**, 161.
31 Y. Ito and H. Oka, *J. Chromatogr.*, 1988, **457**, 393.
32 A. Gavezzotti, *J. Am. Chem. Soc.*, 1983, **105**, 5220.
33 S. Hellberg, M. Sjöstrom, B. Skagerberg and S. Wold, *J. Med. Chem.*, 1987, **30**, 1126.
34 M. Akamatsu, Y. Yoshida, H. Nakamura, M. Asao, H. Iwamura and T. Fujita, *Quant. Struct.-Act. Relat.*, 1989, **8**, 195.
35 J. T. Edsall and J. Wyman, in *Biophysical Chemistry*, Academic Press, New York, 1958, vol. 1, p. 477.
36 T. Fujita, J. Iwasa and C. Hansch, *J. Am. Chem. Soc.*, 1964, **86**, 5175.
37 R. F. Rekker, *Il Farmaco Ed. Sci.*, 1979, **34**, 346.
38 R.-S. Tsai, B. Testa, N. El Tayar and P.-A. Carrupt, *J. Chem. Soc., Perkin Trans. 2*, 1991, 1797.
39 R. M. Sweet and D. Eisenberg, *J. Mol. Biol.*, 1983, **171**, 479.
40 D. Eisenberg and A. D. McLachlan, *Nature*, 1986, **319**, 199.
41 D. Eisenberg, M. Wesson and I. Yamashita, *Chem. Scr.*, 1989, **29A**, 217.
42 M. Oobatake and T. Ooi, *J. Biochem.*, 1988, **104**, 433.
43 T. Ooi, M. Oobatake, G. Némethy and H. A. Scheraga, *Proc. Natl. Acad. Sci. USA*, 1987, **84**, 3086.
44 I. Y. Moriguchi, *Chem. Pharm. Bull.*, 1975, **23**, 247.
45 N. El Tayar, B. Testa and P.-A. Carrupt, *J. Phys. Chem.*, in the press.
46 K. A. Dill, *Biochemistry*, 1990, **29**, 7133.
47 D. C. Rees, L. DeAntonio and D. Eisenberg, *Science*, 1989, **245**, 510.
48 W. Kauzmann, *Adv. Protein Chem.*, 1959, **14**, 1.
49 A. R. Fersht, *Nature*, 1985, **314**, 235.
50 J. T. Edsall and H. A. McKenzie, *Adv. Biophys.*, 1983, **161**, 53.
51 J. U. Bowie, J. F. Reidhaar-Olsen, W. A. Lim and R. T. Sauer, *Science*, 1990, **247**, 1306.
52 J. Novotny, R. E. Bruccoleri and M. Karplus, *J. Mol. Biol.*, 1984, **177**, 787.
53 M. Matsumura, W. J. Becktel and B. W. Matthews, *Nature*, 1988, **334**, 406.
54 M. Matsumura, B. W. Matthews, M. Levitt and W. J. Becktel, *Proc. Natl. Acad. Sci. USA*, 1989, **86**, 6562.
55 G. D. Fasman, in *Prediction of Protein Structure and the Principles of Protein Conformation*, ed. G. D. Fasman, Plenum Press, New York, 1989, p. 193.
56 J. S. Richardson and D. C. Richardson, in *Prediction of Protein Structure and the Principles of Protein Conformation*, ed. G. D. Fasman, Plenum Press, New York, 1989, p. 1.
57 P. C. Lyu, M. I. Liff, L. A. Marky and N. R. Kallenbach, *Science*, 1990, **250**, 669.
58 K. Yutani, K. Ogasahara, T. Tsujita and Y. Sugino, *Proc. Natl. Acad. Sci. USA*, 1987, **84**, 4441.
59 J. T. Kellis, K. Nyberg, D. Sali and A. L. Fersht, *Nature*, 1988, **333**, 784.
60 H. Mager and A. Barth, *Pharmazie*, 1979, **34**, 557.
61 W. P. Jencks and J. Regenstein, in *CRC Handbook of Biochemistry and Molecular Biology: Physical and Chemical Data*, ed. G. D. Fasman, 3rd edn., CRC Press, Boca Raton, Florida, 1976, vol. 1, p. 305.

Paper 1/04584I
Received 4th *September* 1991
Accepted 3rd *October* 1991