Amino Acids (1994) 7: 129–145

Amino
Acids

Lipophilicity of amino acids

Minireview Article

H. van de Waterbeemd[1,2], **H. Karajiannis**[1], and **N. El Tayar**[3]

[1] Pharmazeutisches Institut, Universität Bern, Bern, Switzerland
[2] F. Hoffmann-La Roche Ltd., Pharma Research New Technologies, Basel, Switzerland
[3] Ares-Sereno, Preclinical Pharmacology, Geneva, Switzerland

Accepted January 10, 1994

Summary. The lipophilicity (or hydrophobicity) of amino acids is an important property relevant for protein folding and therefore of great interest in protein engineering. For peptides or peptidomimetics of potential therapeutic interest, lipophilicity is related to absorption and distribution, and thus indirectly relates to their bioactivity. A rationalization of peptide lipophilicity requires basic knowledge of the lipophilicity of the constituting amino acids. In the present contribution we will review methods to measure or calculate the lipophilicities of amino acids, including unusual amino acids, and we will make a comparison between various lipophilicity scales.

Keywords: Amino acids – Lipophilicity – Hydrophobicity –Hydropathicity – Protein amino acids – Non-coded amino acids – Lipophilicity prediction – Unusual amino acids

Introduction

The 20 DNA-coded (L-)amino acids are the building blocks of peptides and proteins involved in essential processes in living organisms. However, many more amino acids have been found in nature, as seen in a compilation of ca 700 amino acids (Hunt, 1985). Many of these are secondary metabolites, others are of microbial and plant origin. Furthermore, many synthetic amino acids have been reported (Paderborn, 1988; Balaram, 1992; Nugent et al., 1993). Apart from their significance as peptide and protein building blocks, many amino acids have a biological activity themselves, e.g. GABA, Gly, Glu as inhibitory or excitatory neurotransmittors, or are important precursors for neurotransmittors, while Arg is converted to the smooth muscle relaxant nitric oxide or stimulates GH (growth hormone) secretion. Physicochemical and stereo-electronic properties of the

amino acids contribute to the stereochemical and electronic characteristics of peptides and proteins. The dominant factors in protein folding have been subdivided by Dill (1990) into contributions from electrostatics, hydrogen-bonding, van der Waals or hydrophobic interactions. Therefore, considerable efforts have been made to parametrize the key properties of amino acid residues and their side chains. In the present contribution we review some important aspects of the lipophilicity of coded and non-coded amino acids. In part II of this review we will focus on the lipophilicity of peptides (El Tayar et al., 1994). Properties of larger peptides and proteins often cannot simply be derived from the constituting amino acids, since the environment of an amino acid inside a macromolecule may differ considerably from that in aqueous solution. Ongoing efforts therefore are undertaken in our laboratories to understand conformation-lipophilicity relationships of peptides and peptidomimetics in more detail.

Lipophilicity, also referred to as hydrophobicity or hydropathicity, is an important physicochemical property that is often correlated to pharmacologically relevant aspects such as absorption and distribution, plasma protein and tissue binding. In structure-property correlation (SPC) studies and particularly in quantitative structure-activity relationship (QSAR) studies, lipophilicity appears to be a key parameter (Van de Waterbeemd and Testa, 1987; Van de Waterbeemd, 1992, 1993). There is however some confusion in the terminology of lipophilicity and hydrophobicity, as will be discussed below.

Information content of log P values

Lipophilicity, as expressed by the logarithm of partition coefficients (log P), is a physicochemical property of particular interest in molecular design because it encodes more structural information than previously thought (Van de Waterbeemd and Testa, 1987; El Tayar et al., 1992a, b). A great deal of interest in relating partition coefficients to chemical structures has been developed in the last three decades. For example, it has been shown that partition coefficients can be considered to result from two major structural contributions, namely a cavity or volume-related term (V) reflecting the energy needed to create a cavity in the solvent, i.e. an endoergic term, and an exoergic interactive term which results from solute-solvent interactions such as dipole-dipole, induced dipole-dipole and hydrogen bonds. Furthermore, it proved useful to define a parameter Λ representing the sum of all polar interactions affecting partition coefficients (Testa and Seiler, 1981; El Tayar et al., 1992b), such that

$$\log P = aV + \Lambda \tag{1}$$

The coefficient "a" depends on the solvent system for which partition coefficients are measured. Traditionally this is a 1-octanol/water system. A priori, linear correlations are to be expected between log P and the volume parameter for the nonpolar n-alkanes. Polar solutes systematically deviate from such linearity due to polar interactions and these deviations have been taken as a measure of the sum of all polar interactions. An analysis of the physicochemical information content of the Λ term, derived from octanol/water partition coefficients, demonstrated that it relates mainly to the hydrogen-bond acceptor basicity and

polarizability of the solutes (El Tayar et al., 1992b). Another consequence of the factorization of log P into a bulk and a polarity contribution is, that we can redefine some terms. The bulk-related expulsion from an aqueous environment is called hydrophobicity, while a partition process is related to the term lipophilicity. Hence, the above mentioned equation can be expressed as (Van de Waterbeemd and Testa, 1987; El Tayar et al., 1992b):

$$\text{lipophilicity} = \text{hydrophobicity} + \text{polarity (H-bonding)} \quad (2)$$

This approach was also applied to amino acids (El Tayar et al., 1992a). For the five amino acids with non-polar side-chains (Ala, Val, Leu, Ile and Phe) as well as for Gly, a linear relationship was obtained between log P and the van der Waals volume V (see Fig. 1). Also Cys, Pro, Ser, Met and Trp are close to this line. More polar are Thr, Asn, Gln, His and Tyr, as well as the charged amino acids Asp, Glu, Lys and Arg.

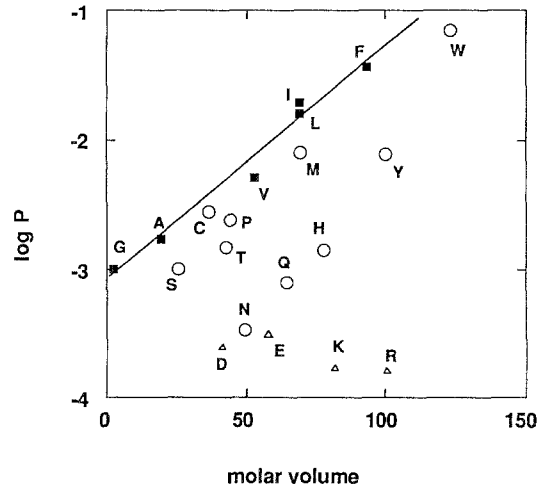

Fig. 1. Relationship between the distribution coefficient at the isoelectric pH (= log P of the zwitterion) and molar volume of the side chain of the natural amino acids. Filled squares are non-polar amino acids. Open circles and triangles represent polar amino acids (*S, C, P, T, M, Y, W, H, Q, N*), the triangles being those with ionizable side-chains (*D, E, K, R*)

In Fig. 1 the distribution coefficient at the isoelectric point is given, i.e. the log P value of the zwitterion. From these log P values, and using calculated molar volumes of the side chain, 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 1). 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. Λ has been further shown to be related to hydration energies (El Tayar et al., 1992b), hydration numbers (see Fig. 2) (Charton, 1990) and the degree of hydropobic polarization ($r = 0.937$) (Furet et al., 1988). Hydration numbers are properties of amino acids in proteins and denote the average number of water molecules associated with the residue.

Table 1. Physicochemical data (bulk and polarity) of coded amino acids

	Amino acid	Code		V_R[a]	Λ[b]	hydrophobic polarization[c]	hydration number[d]
1.	Alanine	Ala	A	19.6	-0.02	0	1.0
2.	Arginine	Arg	R	100.4	-2.56	30	2.3
3.	Asparagine	Asn	N	49.7	-1.24	8	2.2
4.	Aspartic acid	Asp	D	41.1	-1.08	14	6.5
5.	Cysteine	Cys	C	38.5	-0.11	6	
6.	Glutamine	Gln	Q	64.9	-1.19	18	2.1
7.	Glutamic acid	Glu	E	57.9	-1.43	20	6.2
8.	Glycine	Gly	G	2.4	0.03	0	1.1
9.	Histidine	His	H	78.0	-1.06	20	2.8
10.	Isoleucine	Ile	I	69.1	0.04	2	0.8
11.	Leucine	Leu	L	68.9	0.12	2	0.8
12.	Lysine	Lys	K	81.7	-2.26	32	5.3
13.	Methionine	Met	M	69.7	-0.33	4	0.7
14.	Phenylalanine	Phe	F	93.2	-0.05	4	1.4
15.	Proline	Pro	P	44.1	-0.31	12	
16.	Serine	Ser	S	25.9	-0.40	6	1.7
17.	Threonine	Thr	T	42.7	-0.53	10	1.5
18.	Tryptophan	Trp	W	123.3	-0.31	4	
19.	Tyrosine	Tyr	Y	100.1	-0.84	10	2.1
20.	Valine	Val	V	53.0	-0.13	0	0.9

[a] Molar volume in $Å^3$ of the side-chain calculated from the program MOLSV (QCPE 509); [b] Polarity parameter of the side-chains (El Tayar et al., 1992a); [c] Furet et al. (1988); [d] Charton (1990)

Therefore it is interesting to see that these correlate with Λ values derived from solvent/solvent partitioning.

Using the volume of the side-chain (V_R) instead of the entire volume of the amino acid (Table 1) yields the following linear free-energy relationship (LFER) (Vallat, 1992):

$$\log D = 1.86(\pm 0.11)V_R/100 + 1.00(\pm 0.05)\Lambda - 3.13(\pm 0.07)$$
$$n = 20; \quad r = 0.997; \quad s = 0.061; \quad F = 1521 \tag{3}$$

In this equation 95% confidence intervals are given in parentheses; n is the number of compounds, r is the correlation coefficient; s is the standard deviation of the regression (should be a value close to the mean experimental error in log

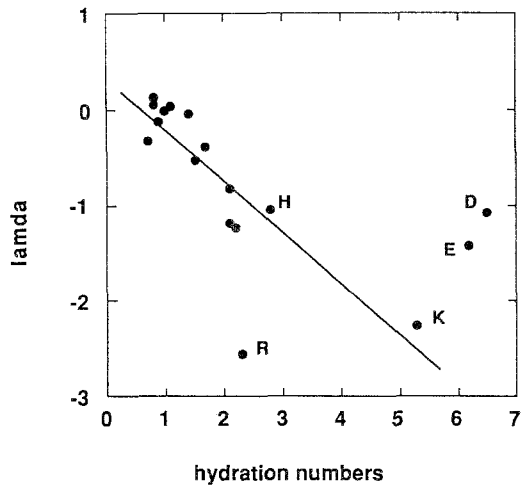

Fig. 2. Relationship between Λ (lamda) values (El Tayar et al., 1992a) and n_H (hydration numbers) (Charton, 1990), taking the side chain residues Asp (*D*), Glu (*E*), Arg (*R*), His (*H*), Lys (*K*) in their ionized form

D); F is the Fisher test for significance of the relationship. The side-chain molar volume (in $Å^3$) has been scaled by a factor 100. In this equation, the intercept is a measure of the distribution coefficient of the backbone. As indicated by Vallat (1992), the intercept, $-3.13\ (\pm 0.07)$, which represents the hydrophobic contribution of the $^{+}H_3N\text{-}CH\text{-}COO^{-}$ moiety, corresponds reasonably well to the value of -3.59 calculated by Abraham and Leo (1987).

Experimental procedures

Both experimental and theoretical approaches have proven useful for the characterization of amino acid lipophilicity. In principle the lipophilicity of amino acids can be estimated by:

- experimental methods:
  - shake flask
  - chromatography (HPLC, TLC, CPC)
  - pH-metric two-phase titration
- theoretical approaches using:
  - log P calculations
  - surface-related aspects of proteins

The classical technique of measuring partition coefficients (log P values) consists of the so-called shake flask procedure, in which a solute is distributed between two immiscible solvents by simple shaking. This approach has several disadvantages. First of all it is rather tedious and requires milligrams of compound. Small amounts of impurities may affect UV concentration measurements. Alternative methods have been developed. Most practical for routine measurements are chromatographic methods, such as thin-layer chromatography (TLC) (Dross et., 1993), reversed-phase high performance liquid chromatography (RP-HPLC)

(Braumann, 1986) and centrifugal partition chromatography (CPC) (Tsai et al., 1991; El Tayar et al., 1992a).

The experimental determination of the lipophilicity of amino acids is difficult because of their hydrophilic character. A means to circumvent this problem is to use more hydrophobic derivatives. Amino acids and small peptides are often zwitterions. In some cases a net positive or negative charge may prevail. In all these cases, pH and counterions play an important role on the measured lipophilicity.

Lipophilicity scales of proteinogenic amino acids

Many different approaches have been used to define lipophilicity or hydrophobicity scales of amino acids. Without pretending to be comprehensive we will discuss a selection of them here. A number of scales are derived from experimental partition experiments or closely related approaches. A second group of lipophilicity scales comes from theoretical considerations often based on comparative analysis of protein structures.

Experimental scales based on solubility and partitioning

The first lipophilicity scales have been developed by Tanford (Tanford, 1962; Nozaki and Tanford, 1971) and Levitt (1976), who used solubility data in water and ethanol. Other early efforts were based on surface tension of amino acid solutions (Bull and Breese, 1974). The experimental assessment of amino acid lipophilicity by partitioning measurements goes back to the early 1980s. Yunger and Cramer (1981) used a radiometric method to obtain distribution coefficients in octanol/water. Later attemps to measure log P values in octanol/water were more successful (Chmelik et al., 1991). Fauchère and Pliska have obtained side-chain lipophilicity values at pH = 7.1 from protected N-acetyl amino acid amides (Fauchère et al., 1980, 1988; Pliska et al., 1981; Fauchère and Pliska, 1983). Partitioning in an aqueous two-phase polymeric system buffered ficoll-dextran yielded a scale that appears to be dependent on the ionic strength and composition of the medium (Zaslavsky et al., 1982).

Various chromatographic techniques have been employed to measure empirical lipophilicity scales of amino acids. Paper chromatography was used already in the 1970s (Aboderin, 1971), later followed by thin-layer chromatography (Pliska et al., 1981; Skagerberg et al., 1987; Eriksson et al., 1988) as well as HPLC (Parker et al., 1986; Cowan and Whittaker, 1990). A scale correlated to the just-mentioned one of Fauchère-Pliska was obtained from retention time measurements of Z-protected amino acids and their methyl, ethyl and benzyl esters at pH = 3.0 and 7.5 (Cowan and Whittaker, 1990). Testa and co-workers measured log P values of amino acids at the isoelectric pH by CPC (centrifugal partition chromatography), a new technique for measuring lipophilicities in systems of immiscible solvents (Tsai et al., 1991; El Tayar et al., 1991).

In proteins, amino acids are sometimes modified by acylation, carboxylation, phosphorylation, sulfatation, glycosylation, etc. Although many hydrophobicity scales exist for the twenty DNA-coded amino acids, little is known about

post-translationally modified amino acyl residues. For such modified amino acids, some hydrophobicity parameters have been proposed by Black and Mould (1991).

Theoretical scales

Besides the above-mentioned experimental hydrophobicity scales, a number of scales are derived from these by ad hoc reasoning or by calculation from molecular surfaces or side-chain exposure in proteins. Before we discuss these, we also mention here the method of calculating log P values of organic molecules including amino acids and peptides.

The lipophilicity expressed as 1-octanol/water partition coefficient log P of amino acids can be calculated from the well-known addivity equation (Hansch and Leo, 1979; Leo, 1993; Rekker and Mannhold, 1992):

$$\log P = \sum a_i f_i + \sum b_j F_j \quad (4)$$

where f_i is the hydrophobic fragmental value of a particular molecular fragment and F_j are correction factors, e.g. due to the presence of a zwitterion ot the interaction between polar groups. The coefficients a_i and b_j are the number of occurrences of each.

For the prediction of protein antigenic determinants Hopp and Woods (1981) designed a hydrophobicity scale from an earlier one by Levitt (1976). A hydropathy scale has been derived from various experimental observations in the literature (Kyte and Doolittle, 1982). Accessible molecular surface areas have been used to derive free energies of hydration by the group of Scheraga (Ooi et al., 1987) and others (Oobatake and Ooi, 1988), following an approach by Eisenberg and McLachlan (1986). In order to bring some order in the various suggested hydrophobicity scales a so-called consensus scale has been proposed (Eisenberg et al., 1984). Log P values in cyclohexane/water of amino acid side chain analogs appear to be closely related to the inside-outside distributions of the side chains observed in globular proteins (Radzicka and Wolfenden, 1988). Charton has expressed amino acid lipophilicity scales as the sum of intermolecular forces between amino acid side chains and some surrounding medium (Charton and Charton, 1982; Charton, 1990). It was found that amino acid hydrophobicity shows a major dependence on the polarizability of the side chain and to a lesser extent on a steric factor. It was also concluded that there is no advantage over experimental scales in using surface or volume descriptors as a hydrophobicity scale.

A comparison among various amino acid lipophilicity scales is presented in Fig. 3. It is seen that considerable differences do occur. Using the hydrophobicity scales in Tables 2 and 3, clustering between scales and amino acids was examined, as illustrated in Fig. 4. Although these dendrograms reveal basic similarities among certain amino acids and lipophilicity scales, a full insight in these relationships is only obtained by also looking at the results from a correlation matrix and a principal component analysis of the same data (not shown here). The lipophilicity scale of Norinder seems not to be in line with the others, which is also seen in Fig. 3.

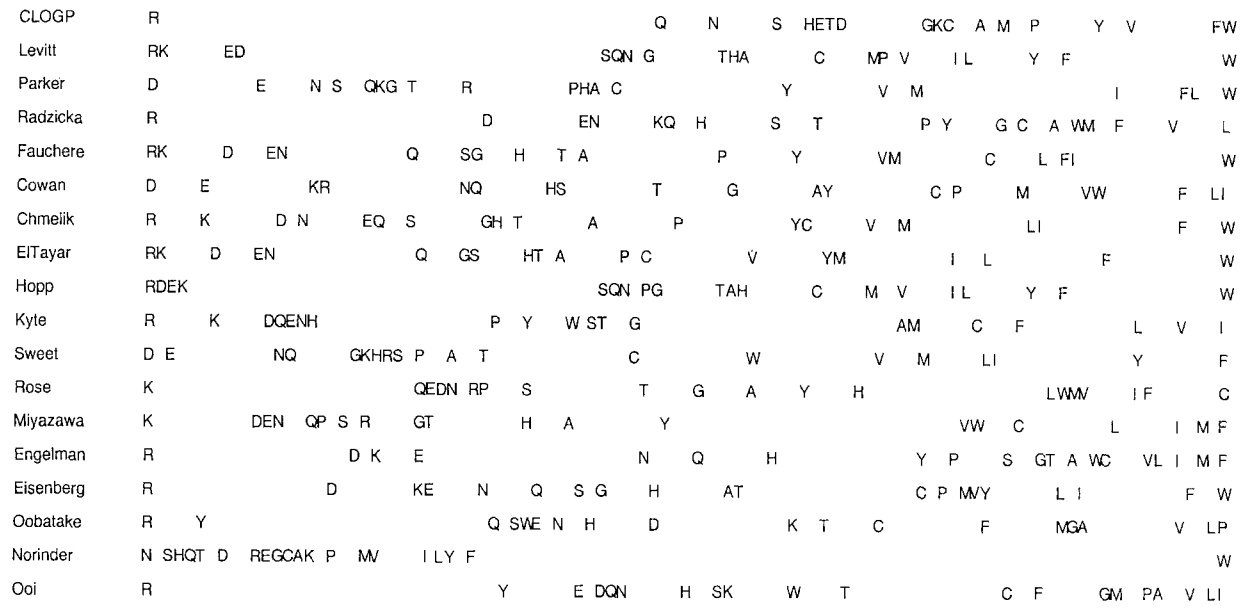

Fig. 3. Comparison of selected lipophilicity scales for amino acids. All scales have been autoscaled to make them comparable

Table 2. Lipophilicity scales of natural amino acids from experimental measurements

		CLOGP* v. 3.55	Levitt (1971)	Parker et al. (1986)	Radzicka Wolfenden (1988)	Fauchère et al (1988)	Cowan, Whittaker (1990)	Chmelik et al. (1991)	El Tayar et al. (1992)
1.	A	-2.999	0.5	-2.1	1.81	0.31	0.35	0.30	-2.77
2.	R	-7.938	-3.0	-4.2	-14.92	-1.01	-1.50	-0.91	-3.79
3.	N	-4.563	-0.2	-7.0	-6.64	-0.60	-0.99	-0.48	-3.48
4.	D	-3.823	-2.5	-10.0	-8.72	-0.77	-2.15	-0.55	-3.61
5.	C	-3.172	1.0	-1.4	1.28	1.54	0.76	0.86	-2.55
6.	Q	-4.927	-0.2	-6.0	-5.54	-0.22	-0.93	-0.30	-3.11
7.	E	-3.927	-2.5	-7.8	-6.81	-0.64	-1.95	-0.32	-3.51
8.	G	-3.308	0.0	-5.7	0.94	0.0	0.0	0.0	-3.00
9.	H	-4.000	0.5	-2.1	-4.66	0.13	-0.65	0.02	-2.85
10.	I	-1.542	1.8	8.0	4.92	1.80	1.83	1.53	-1.80
11.	L	-1.542	1.8	9.2	4.92	1.70	1.80	1.50	-1.72
12.	K	-3.299	-3.0	-5.7	-5.55	-0.99	-1.54	-0.74	-3.77
13.	M	-2.851	1.3	4.2	2.35	1.23	1.10	1.14	-2.10
14.	F	-1.581	2.5	9.2	2.98	1.79	1.69	1.91	-1.44
15	P	-2.655	1.4	-2.1	-0.52 **	0.72	0.84	0.53	-2.62
16.	S	-4.215	-0.3	-6.5	-3.40	-0.04	-0.63	-0.19	-3.00
17.	T	-3.906	0.4	-5.2	-2.57	0.26	-0.27	0.09	-2.83
18.	W	-1.581	3.4	10.0	2.33	2.25	1.35	2.01	-1.15
19.	Y	-2.248	2.3	1.9	-0.14	0.96	0.39	0.85	-2.11
20.	V	-2.071	1.5	3.7	4.04	1.22	1.32	1.07	-2.29

*CLOGP: calculated 1-octanol/water partition coefficients, included for comparison
** Estimated value.

Anomalous behaviour

The lipophilicity of some amino acids is controversial or unexpected. Proline often appears in solvent-exposed positions in proteins and is thus taken to be more hydrophilic than residues with noncyclic hydrocarbon side-chains of simi-

Table 3. Lipophilicity scales of natural amino acids from theoretical approaches

		Hopp, Woods (1981)	Kyte, Doolittle (1982)	Sweet, Eisenberg (1983)	Rose et al. (1985)	Miyazawa, Jernigan (1985)	Engelman et al. (1986)	Eisenberg, McLachlan (1986)*	Ooi et al. (1987)	Oobatake, Ooi (1988)	Norinder (1991)
1.	A	0.5	1.8	-0.40	0.74	5.33	1.6	0.55	2.50	-0.54	-33.19
2.	R	-3.0	-4.5	-0.59	0.64	4.18	-12.3	-2.00	-53.50	-5.96	-65.57
3.	N	-0.2	-3.5	-0.92	0.63	3.71	-4.8	-0.51	-27.75	-3.55	-155.49
4.	D	-3.0	-3.5	-1.31	0.62	3.59	-9.2	-1.20	-28.28	-2.97	-95.62
5.	C	1.0	2.5	0.17	0.91	7.93	2.0	1.40	-5.12	-1.64	-45.79
6.	Q	-0.2	-3.5	-0.91	0.62	3.87	-4.1	-0.29	-27.91	-3.92	-118.12
7.	E	-3.0	-3.5	-1.22	0.62	3.65	-8.2	-0.76	-28.98	-3.71	-59.25
8.	G	0.0	-0.4	-0.67	0.72	4.48	1.0	0.00	0.00	-0.59	-55.68
9.	H	0.5	-3.2	-0.64	0.78	5.10	-3.0	0.25	-23.27	-3.38	-126.76
10.	I	1.8	4.5	1.25	0.88	8.83	3.1	2.10	6.36	0.32	64.19
11.	L	1.8	3.8	1.22	0.85	8.47	2.8	2.00	5.91	0.27	65.29
12.	K	-3.0	-3.9	-0.67	0.52	2.95	-8.8	-0.78	-20.58	-2.19	-28.92
13.	M	1.3	1.9	1.02	0.85	8.95	3.4	1.60	0.40	-0.60	14.59
14.	F	2.5	2.8	1.92	0.88	9.03	3.7	2.60	-3.69	-1.06	103.72
15	P	0.0	-1.6	-0.49	0.64	3.87	-0.2	1.50	2.19	0.32	-5.10
16.	S	-0.3	-0.8	-0.55	0.66	4.09	0.6	-0.09	-21.52	-3.82	-135.36
17.	T	0.4	-0.7	-0.28	0.70	4.49	1.2	0.58	-14.25	-1.97	-114.03
18.	W	3.4	-0.9	0.50	0.85	7.66	1.9	2.70	-17.55	-3.80	683.54
19.	Y	2.3	-1.3	1.67	0.76	5.89	-0.7	1.70	-33.25	-5.64	78.78
20.	V	1.5	4.2	0.91	0.86	7.63	2.6	1.60	4.85	0.13	28.77

* changed sign

lar size (Gibbs, 1991). Trp, Tyr and Phe were found more hydrophobic than expected (Cowan and Whittaker, 1990). In the α-amino acids the positive and negative charge are in geminal position. In many non-coded amino acids, such as GABA, the distance between the charges is different. In an exploratory study it was shown that the distance between the charges in zwitterions appears to affect the overall molecular lipophilicity by ca 0.4 log P units per increase of 1 Å between the opposite charges (Tsai et al., 1991). However, this finding cannot be generalized. N-methylation does not increase the lipophilicity (Tsai et al., 1991), but may affect absorption properties (see part II). The hydrophobicity of tryptophan has been debated (Fauchère and Pliska, 1983). It appears that 1-octanol exerts a specific attraction on the side chain of tryptophan. With less polar organic phases, Leu, Ile, Val, Phe and Met are more hydrophobic than Trp (Radzicka and Wolfenden, 1988). A warning about the structure-addivity principle of side-chain lipophilicity in proteins has been given by Roseman (1988), by emphasizing that particularly for polar amino acids the hydrophilicity of side-chains are markedly reduced by flanking peptide bonds. Arg has a log P somewhat higher than calculated. Crystal structures show that the guanidino nitrogen and carboxylate oxygen atoms are bridged by two water molecules, increasing the log P by $3 \times (0.63) = 1.89$, i.e. the effect of three hydrogen bonds (Abraham and Leo, 1987). Tightly bound water might considerably alter the hydrophilic/hydrophobic nature of parts of a peptide or protein.

Hydropathicity profiles and antigenicity prediction

From the above it may be clear that the lipophilicity of an amino acid is highly context dependent, when it is embedded in larger structures such as peptides

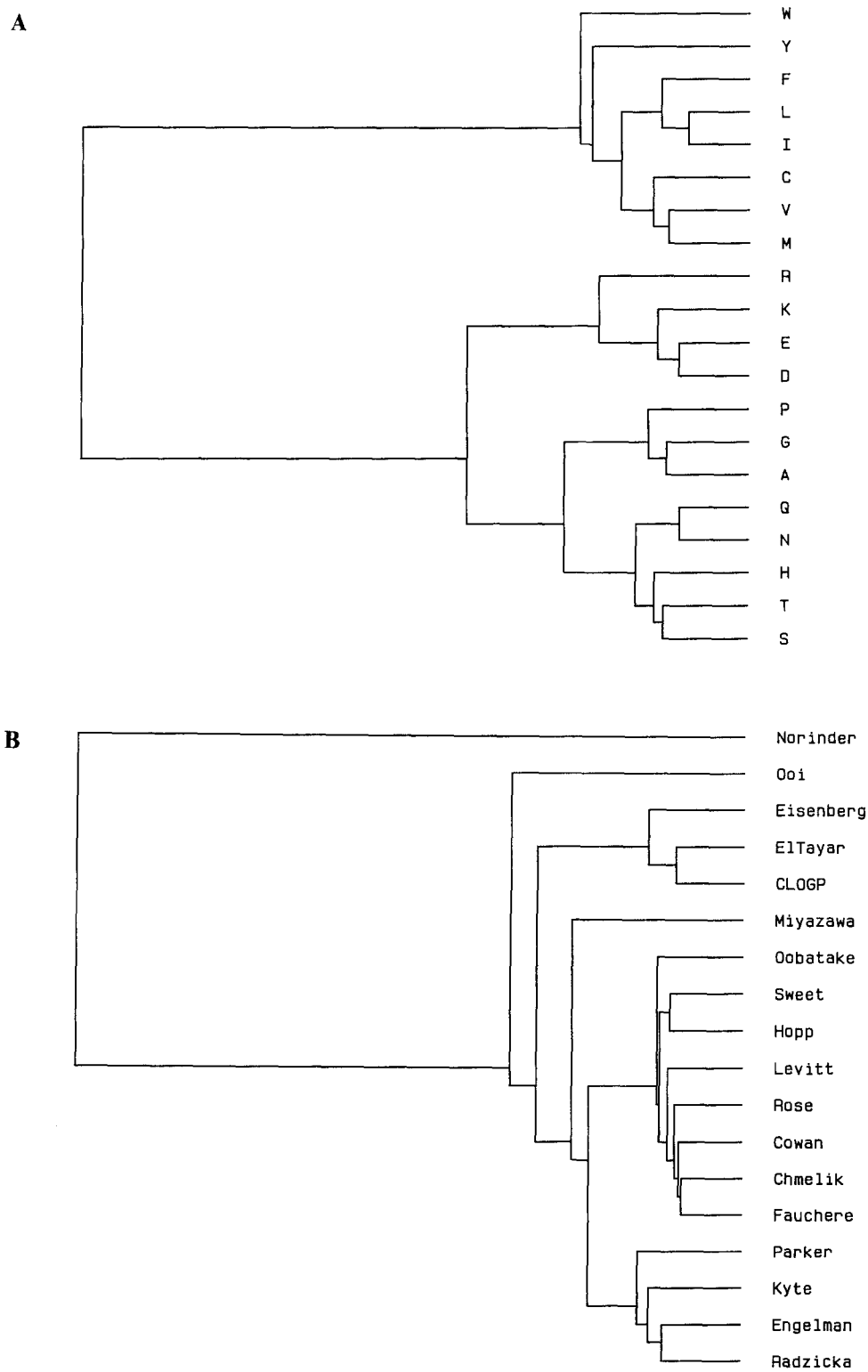

Fig. 4. Cluster analysis using Ward's method of the lipophilicity scales in Table 2 and 3. **A** Relationship between the 20 natural amino acids, **B** Relationship between lipophilicity scales

and proteins. Since they are derived in uncomparable ways, the various amino acid lipophilicity scales differ widely, as pointed out on several occasions (Cowan and Whittaker, 1990) and demonstrated in Fig. 4. This makes it a difficult task to subdivide amino acids into hydrophilic and hydrophobic ones, since for some of them this is fully arbitrary.

Nevertheless so-called hydropathicity profiles of proteins (Kyte and Doolittle, 1982) used to compare protein sequences or to detect potential antigenic or transmembranes segments are based on these scales. Antigenic sites are defined as mobile hydrophilic domains at the surface of a protein. These can be reasonably well predicted using an appropriate hydrophobicity scale (Parker et al., 1986). The effect of using different hydrophobicity scales on hydropathicity profiling is seen in Fig. 5 for two G-protein coupled receptors (endothelin-1 receptor (ETAr) and the follicle stimulating hormone receptor (FSHr)). The lipophilicity scales of El Tayar et al (1992a) and Oobatake and Ooi (1988) have been compared, since they are only poorly correlated to each other ($r = 0.41$). It is seen that both approaches do not give exactly the same prediction for residues involved in a transmembrane domain. For comparison, the domains as

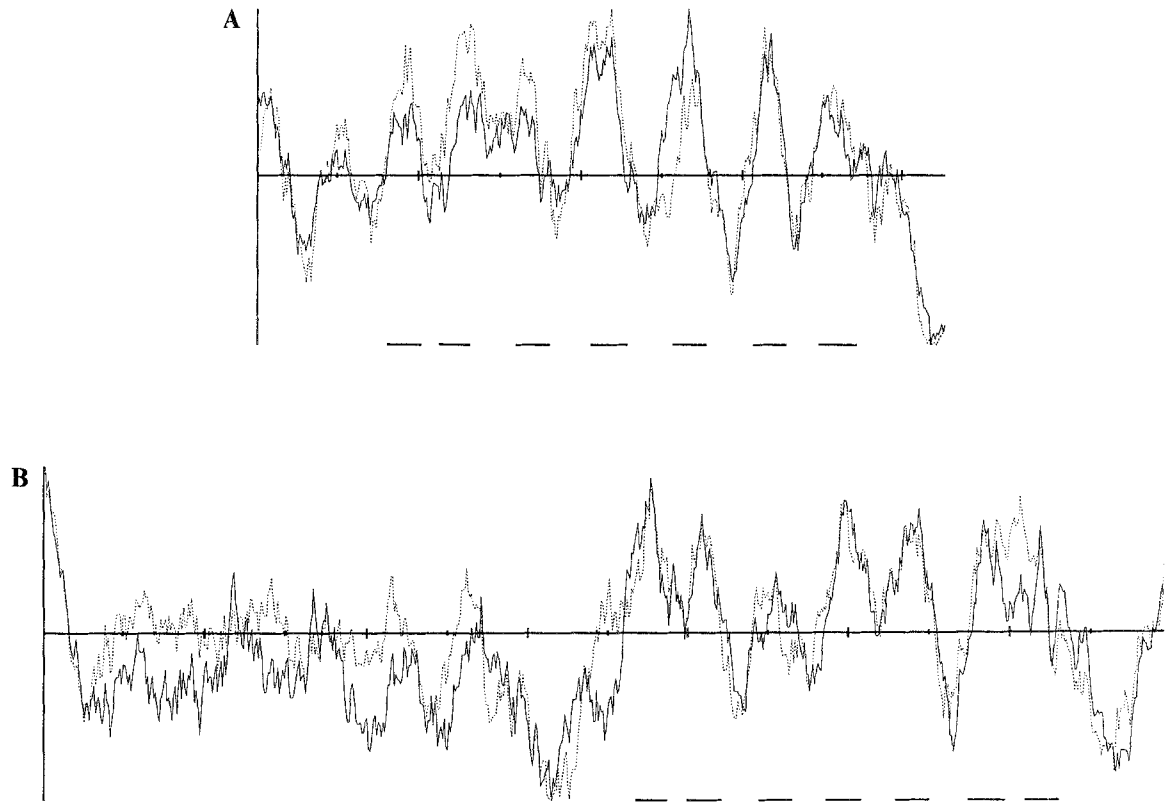

Fig. 5. Hydropathicity profiles using the lipophilicity scales of El Tayar et al (1992a) (solid line) and Oobatake et al. (1988) (broken line). **A** The endothelin ET-1 receptor, **B** The FSH receptor. The seven transmembrane domains as listed in the Swiss-Prot database are indicated

listed in the feature table of the Swiss-Prot database are indicated in Fig. 5. These are also estimated indications for the postions, since no X-ray crystal structures of G-protein coupled receptors are known at present. For making proper working models of these receptors, it seems therefore important to make a proper choice of lipophilicity scales used in hydropathicity profiling.

The preference for specific locations of amino acids at the end of α helices appears also partly related to the residue's lipophilicity (Richardson and Richardson, 1988). The C-cap position is often occupied by a Gly, while hydrophobic residues are found at N-cap $+4$ and C-cap -4.

Table 4. Principal properties (PPs) of amino acids

		z_1	z_2	z_3	F_2	F_4	t_1	t_2	t_3
1.	A	0.07	-1.73	0.09	-1.67	-0.27	-8.40	-8.01	-3.73
2.	R	2.88	2.52	-3.44	1.27	1.87	-2.27	18.90	-18.24
3.	N	3.22	1.45	0.84	-0.07	0.81	-3.11	-1.22	6.26
4.	D	3.64	1.13	2.36	-0.22	0.81	-6.84	-0.94	17.68
5.	C	0.71	-0.97	4.13	-0.89	-1.05	-2.44	-1.96	0.93
6.	Q	2.18	0.53	-1.14	0.24	1.10	-5.31	15.64	8.44
7	E	3.08	0.39	-0.07	0.19	1.17	-6.50	16.20	17.28
8.	G	2.23	-5.36	0.30	-1.96	-0.16	-8.48	-10.37	-5.14
9.	H	2.41	1.74	1.11	0.52	0.28	15.28	-3.67	6.72
10.	I	-4.44	-1.68	-1.03	-0.16	-0.77	-2.97	4.64	-0.77
11.	L	-4.19	-1.03	-0.98	0.00	-1.10	2.61	5.96	1.97
12.	K	2.84	1.41	-3.14	0.82	1.70	2.70	13.46	-14.03
13.	M	-2.49	-0.27	-0.41	0.18	-0.73	3.38	12.43	-4.77
14.	F	-4.92	1.30	0.45	0.98	-1.43	21.59	-5.73	1.03
15	P	-1.22	0.88	2.23	-0.33	-0.75	-5.35	-9.07	-1.52
16.	S	1.96	-1.63	0.57	-1.08	0.42	-15.88	-11.21	-2.44
17.	T	0.92	-2.09	-1.40	-0.70	0.63	-17.81	-13.64	-5.19
18.	W	-4.75	3.65	0.85	2.10	-1.57	21.68	-8.78	-2.53
19.	Y	-1.39	2.32	0.01	1.48	-0.56	23.90	-6.47	0.31
20.	V	-2.69	-2.53	-1.29	-0.71	-0.40	-5.80	-6.15	-2.26

z-scales (Hellberg et al., 1991): z_1 amino acid hydrophobicity, z_2 size, z_3 polarity
F-scales (Kidera et al., 1985): F_2 amino acid bulk, F_4 hydrophobicity
t-scales (Cocchi, 1993): t_1 size and polarizability, t_2 mixed properties, t_3 sum of the net charges of the atoms on the side-chains

Amino acid principal properties and experimental design of peptides

Large compilations of physicochemical properties of amino acids are available in the literature. Statistical methods using principal component analysis and cluster analysis have been used to describe physicochemical and conformational properties of the 20 naturally occuring amino acids by a minimum set of descriptors derived from a large data collection. Using such data reduction techniques, a set of ten orthogonal properties, called principal properties (PPs), was obtained from a data set containing originally 188 properties. These principal properties are therefore new calculated descriptors encoding several aspects of amino acids. Two of these ten are related to hydrophobicity indices for free amino acids, while one can be interpreted as a lipophilicity index for residues in a protein (Kidera et al, 1985). In an even more comprehensive study, 222 physicochemical amino acid indices were clustered by a minimum spanning tree into four regions (Nakai, 1988). One main region consisted of 82 descriptors related to hydrophobicity, which could be divided into several subclasses, namely preference for inside or outside locations in a protein, accessible surface area, surrounding hydrophobicity and various experimental scales, including partition coefficients and HPLC or TLC data.

The group of Wold and coworkers derived principal properties (PPs) of amino acids using PLS (Partial Least Squares Projection to Latent Structures), called z-scales, from retention times in HPLC (Skagerberg et al., 1987) and TLC (Eriksson, 1988; Jonsson, 1989). Similar scales based on the 3D-QSAR comparative molecular field analysis (CoMFA) method have been reported by others (Norinder, 1991; Cocchi and Johansson, 1993). Such scales may be used in the design of series of peptides with bioactive properties (Wold et al., 1987; Hellberg et al., 1987, 1991; Norinder, 1991; Cocchi and Johansson, 1993). The first two z-scales are related to the lipophilic and steric-bulk properties of the amino acids, respectively. A comparison of the corresponding PPs between the approaches of the Swedish (Hellberg et al., 1991) and Japanese (Kidera et al., 1985) groups

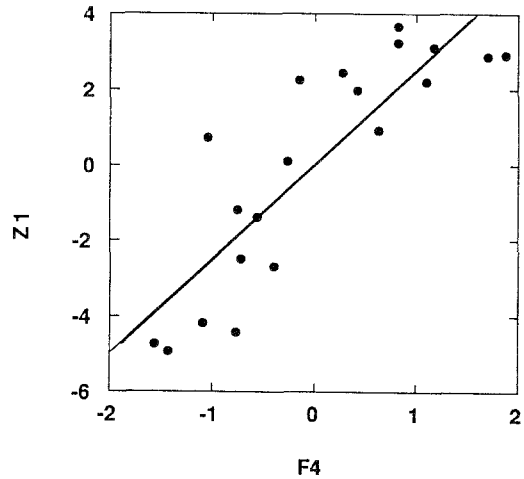

Fig. 6. A comparison of z-scales (Hellberg et al., 1991) to F-scales (Kidera et al., 1985); *z1* and *F4* both represent lipophilicity scales

is given in Fig. 6. This comparison demonstrates that principal properties obtained from different amino acid data collections are not unique numbers and that it is hard to decide which approach is best suited to be used in experimental design of new peptides. More experience and research is warranted here.

Lipophilicity of non-coded amino acids

Many hundreds of non-coded amino acids are known in nature. Physicochemical and conformational properties of peptides can be modulated by using natural or synthetic non-coded or unusual amino acid building blocks (Obrecht et al., 1992; Nugent et al., 1993). Despite this interest in unusual amino acids, the hydrophobic properties have been reported for only a small selection (Fauchère et al., 1980; Jonsson et al., 1989; Charton, 1990).

Using the same strategies as for proteinogenic amino acids, some non-coded amino acids are described by a set of parameters, allowing us again the calculation of more abstract principal properties (Wold et al., 1987; Jonsson et al., 1989; Larsson et al., 1993).

Conclusions

We have reviewed here a selection of the work on lipophilicity scales, some being controversial (Engelman et al., 1986). A situation best illustrated by a citation of Charton (Charton and Charton, 1982): "*it seems likely that no single hydrophobicity parameter can represent the complete range of amino acid behaviour*". The many lipophilicity scales published each reflect some aspect of the forces involved in peptide or protein folding and partioning between aqueous and lipidic environments. A better understanding of various lipophilicity scales may contribute to more rational protein engineering, as well as to the problem of protein folding. Amino acids have been characterized by many different physicochemical, structural (topologic and conformational) properties. Future studies on the effect of (unusual) amino acids on the conformation of a peptide are a great challenge for the definition of a new type of amino acid scales. The next years will undoubtly bring more information on non-coded amino acids, designed for their physicochemical as well as their structural effect on peptide-like compounds.

Acknowledgements

We thank Prof. K. Müller, Dr. D. Obrecht (F. Hoffmann-La Roche Ltd) and Prof. B. Testa (University of Lausanne) for fruitful discussions and valuable comments, and Dr. C. Broger (F. Hoffmann-La Roche Ltd.) for Figs. 3–5.

References

Aboderin AA (1971) An empirical hydrophobicity scale for α-amino-acids and some of its applications. Int J Biochem 2: 537–544

Abraham DJ, Leo A (1987) Extension of the fragment method to calculate amino acid zwitterion and side chain partition coefficients. Proteins Struct Funct Gen: 130–152

Balaram P (1992) Non-standard amino acids in peptide design and protein engineering. Curr Op Struct Biol 2: 845–851

Black SD, Mould DR (1991) Development of hydrophobicity parameters to analyze proteins which bear post- or cotranslational modifications. Anal Biochem 193: 72–82

Braumann T (1986) Determination of hydrophobic parameters by reversed-phase liquid chromatography: theory, experimental techniques and application in studies on quantitative structure-activity relationships. J Chromatogr 373: 191–225

Bull HB, Breese K (1974) Surface tension of amino acid solutions: a hydrophobicity scale of the amino acid residues. Arch Biochem Biophys 161: 665–670

Charton M (1990) The quantitative description of amino acid, peptide, and protein properties and bioactivities. Prog Phys Org Chem 18: 163–284

Charton M, Charton BI (1982) The structural dependence of amino acid hydrophobicity parameters. J Theor Biol 99: 629–644

Chmelik J, Hudecek J, Putyera K, Makovicka J, Kalous V, Chmelikova J (1991) Characterization of the hydrophobic properties of amino acids on the basis of their partition and distribution coefficients in the 1-octanol-water system. Collect Czech Chem Commun 56: 2030–2041

Cocchi M, Johansson E (1993) Amino acids characterization by GRID and multivariate data analysis. Quant Struct-Act Relat 12: 1–8

Cowan R, Whittaker RG (1990) Hydrophobicity indices for amino acid residues as determined by high-performance liquid chromatography. Peptide Res 3: 75–80

Dill KA (1990) Dominant forces in protein folding. Biochem 29: 7133–7155

Dross K, Sonntag C, Mannhold R (1993) On the precise estimation of R_M-values in reversed-phase TLC including aspects of pH-dependence. J Chromatogr 639: 287–294

El Tayar N, Tsai RS, Vallat P, Altomare C, Testa B (1991) The use of CPC for assessing lipophilicity: a comparative evaluation. J Chromatogr 556: 181–194

El Tayar N, Tsai RS, Carrupt PA, Testa B (1992a) Octan-1-ol-water partition coefficients of zwitterionic α-amino acids. Determination by centrifugal partition chromatography and factorization into steric/hydrophobic and polar components. J Chem Soc Perkin Trans 2: 79–84

El Tayar N, Testa B, Carrupt PA (1992b) Polar intermolecular interactions encoded in partition coefficients: an indirect estimation of hydrogen-bond parameters of polyfunctional solutes. J Phys Chem 96: 1455–1459

El Tayar N, Karajiannis H, Van de Waterbeemd H (1994) Structure-lipophilicity relationships of peptides and peptidomimetics. Amino Acids (in press)

Eisenberg D, McLachlan AD (1986) Solvation energy in protein folding and binding. Nature 319: 199–203

Engelman DM, Steitz TA, Goldman A (1986) Identifying nonpolar transbilayar helices in amino acid sequences of membrane proteins. Ann Rev Biophys Chem 15: 321–353

Eriksson L, Jonsson J, Sjöström M, Wold S (1988) Multivariate parametrization of amino acid properties by thin layer chromatography. Quant Struct-Act Relat 7: 144–150

Fauchère JL, Pliska V (1983) Hydrophobic parameters π of amino-acid side chains from the partitioning of N-acetyl-amino acid amides. Eur J Med Chem 18: 369–375

Fauchère JL, Do KQ, Jow PYC, Hansch C (1980) Unusually strong lipophilicity of 'fat' or 'super' amino-acids, including a new reference value for glycine. Experientia 36: 1203–1204

Fauchère JL, Charton M, Kier LB, Verloop A, Pliska V (1988) Amino acid side chain parameters for correlation studies in biology and pharmacology. Int J Pept Prot Res 32: 269–278

Furet P, Sele A, Cohen NC (1988) 3D molecular lipophilicity potential profiles: a new tool in molecular modeling. J Mol Graphics 6: 182–200

Gibbs PR, Radzicka A, Wolfenden R (1991) The anomalous hydrophilic character of proline. J Am Chem Soc 113: 4714–4715

Hansch C, Leo A (1979), Substituent constants for correlation analysis in chemistry and biology. Wiley, New York

Hellberg S, Sjöström M, Skagerberg B, Wold S (1987) Peptide quantitative structure-activity relationships, a multivariate approach. J Med Chem 30: 1126–1135
Hellberg S, Eriksson L, Jonsson J, Lindgren F, Sjöström M, Skagerberg B, Wold S, Andrews P (1991) Minimum analogue peptide sets (MAPS) for quantitative structure-activity relationships. Int J Pept Prot Res 37: 414–424
Hopp TP, Woods KR (1981) Prediction of protein antigenic determinants from amino acid sequences. Proc Natl Acad Sci USA 78: 3824–3828
Hunt S (1985) The non-protein amino acids. In: Barrett GC (ed) Chemistry and biochemistry of the amino acids. Chapman and Hall, London, pp 55–138
Jonsson J, Eriksson L, Hellberg S, Sjöström M, Wold S (1989) Multivariate parametrization of 55 coded and non-coded amino acids. Quant Struct-Act Relat 8: 204–209
Kidera, A, Konishi Y, Oka M, Ooi T, Scheraga HA (1985) Statistical analysis of the physical properties of the 20 naturally occuring amino acids. J Prot Chem 4: 23–55
Kyte J, Doolittle R (1982) A simple method for displaying the hydropathic character of a protein. J Mol Biol 157: 105–132
Larsson U, Carlson R, Leroy J (1993) Synthesis of amino acids with modified principal properties 1. Amino acids with fluorinated side chains. Acta Chem Scand 47: 380–390
Leo A (1993) Calculating log P_{oct} from structure. Chem Rev 93: 1281–1306
Levitt M (1976) A simplified representation of protein conformation for rapid simulations of protein folding. J Mol Biol 104: 59–107
Miyazawa S, Jernigan RL (1985) Estimation of effective interresidue contact energies from protein crystal structures: quasi-chemical appropimation. Macromol 18: 534–552
Nakai K, Kidera A, Kanehisa M (1988) Cluster analysis of amino acid indices for prediction of protein structure and function. Prot Eng 2: 93–100
Norinder U (1991) Theoretical amino acid descriptors. Application to bradykinin potentiating peptides. Peptides 12: 1223–1227
Nozaki Y, Tanford C (1971) The solubility of amino acids and two glycine peptides in aqueous ethanol and dioxane solutions. J Biol Chem 246: 2211–2217
Nugent WA, RajanBabu TV, Burk MJ (1993) Beyond nature's chiral pool: enantioselective catalysis in industry. Science 259: 479–483
Obrecht D, Spiegler C, Schönholzer P, Müller K, Heimgartner H, Stierli F (1992) A new general approach to enantiomerically pure cyclic and open-chain (R)- and (S)-α,α-disubstituted α-amino acids. Helv Chim Acta 75: 1666–1696
Oobatake M, Ooi T (1988) Characteristic thermodynamic properties of hydrated water for 20 amino acid residues in globular proteins. J Biochem 104: 433–439
Ooi T, Oobatake M, Némethy G, Scheraga HA (1987) Accessible surface areas as a measure of the thermodynamic parameters of hydration of peptides. Proc Natl Acad Sci USA 84: 3086–3090
Paderborn HJ (1988) Asymmetrische Synthese von α-Aminosäuren. Nachr Chem Tech Lab 36: 999–1002
Parker JMR, Guo D, Hodges RS (1986) New hydrophilicity scale derived from high-performance liquid chromatography peptide retention data: correlation of predicted surface residues with antigenicity and X-ray-derived accessible sites. Biochem 25: 5425–5432
Pietrzynski G, Rzeszotrask B (1993) Turn tendency of α, β-unsaturated peptides Ac-Pro-ΔX-$NHCH_3$ studied in solution by NMR, IR and CD spectroscopy. Amino Acids 5: 108
Pliska V, Schmidt M, Fauchère JL (1981) Partition coefficients of amino acids and hydrophobic parameters π of their side-chains as measured by thin-layer chromatography. J Chromatogr 216: 79–92
Radzicka A, Wolfenden R (1988) Comparing the polarities of the amino acids: side-chain distribution coefficients between the vapor phase, cyclohexane, 1-octanol, and neutral aqueous solution. Biochem 27: 1664–1670
Rekker RF, Mannhold R (1992) Calculation of drug lipophilicity. VCH, Weinheim
Richardson JS, Richardson DC (1988) Amino acid preferences for specific locations at the ends of α helices. Science 240: 1648–1652

Rose GD, Geselowitz AR, Lesser GJ, Zehfus MH (1985) Hydrophobicity of amino acid residues in globular proteins. Science 229: 834–838

Roseman MA (1988) Hydrophilicity of polar amino acid side-chains is markedly reduced by flanking peptide bonds. J Mol Biol 200: 513–522

Skagerberg B, Sjöström M, Wold S (1987) Multivariate characterization of amino acids by reversed phase high pressure liquid chromatography. Quant Struct-Act Relat 6: 158–164

Sweet RM, Eisenberg D (1983) Correlation of sequence hydrophobicities measures similarity in three-dimensional protein structure. J Mol Biol 171: 479–488

Tanford C (1962) Effect of solvent on the optical rotatory dispersion to the uncharged molecules containing the peptide group. J Am Chem Soc 84: 4240–4247

Testa B, Seiler P (1981) Steric and lipophilic components of the hydrophobic fragmental constant. Arzneimittelforschung 31: 1053–1058

Tsai RS, Testa B, El Tayar N, Carrupt PA (1991) Structure-lipophilicity relationships of Zwitterionic amino acids. J Chem Soc Perkin Tr 2: 1797–1802

Vallat Ph (1992) Lipophilicity and intramolecular interactions. Thesis University of Lausanne

Van de Waterbeemd H (1992) The history of drug research: from Hansch to the present. Quant Struct-Act Relat 11: 200–204

Van de Waterbeemd H (1993) Recent progress in QSAR-technology. Drug Des Discov 9: 277–285

Van de Waterbeemd H, Testa B (1987) The parametrization of lipophilicity and other structural properties in drug design, vol 16. In: Testa B (ed) Advances in drug research. Academic Press, London, pp 85–225

Wold S, Eriksson L, Hellberg S, Jonsson J, Sjöström M, Skagerberg B, Wikström C (1987) Principal property values for six non-natural amino acids and their application to a structure-activity relationship for oxytocin peptide analogues. Can J Chem 65: 1814–1820

Wolfenden R (1983) Waterlogged molecules. Science 222: 1087–1093

Yunger LM, Cramer RD (1981) Measurement and correlation of partition coefficients of polar amino acids. Mol Pharmacol 20: 602–608

Zaslavsky BYu, Mestechkina NM, Miheeva LM, Rogozhin SV (1982) Measurement of relative hydrophobicity of amino acid side-chain by partition in an aqueous two-phase polymeric system: hydrophobicity scale for non-polar and ionogenic side-chains. J Chromatogr 240: 21–28

Authors' address: PD Dr. H. van de Waterbeemd, F. Hoffmann-La Roche Ltd., Pharma Research New Technologies, PRTC 65/314, CH-4002 Basel, Switzerland.

Received August 23, 1993