MOLECULAR PHARMACOLOGY, 20:602-608

Measurement and Correlation of Partition Coefficients of Polar Amino Acids

LIBBY M. YUNGER AND RICHARD D. CRAMER, III

Departments of Pharmacology and Medicinal Chemistry, Smith Kline & French Laboratories, Philadelphia, Pennsylvania 19101

Received March 9, 1981; Accepted July 6, 1981

SUMMARY

The distribution coefficients of 12 commonly occurring amino acids were measured with high precision by a radiometric method. Values obtained for the logarithm of the ratio of distribution between octanol and 10 mM phosphate buffer, pH 7.0, are as follows: tryptophan, −1.11; histidine, −1.95; valine, −2.26; proline, −2.54; 3-carboxypiperidine, −2.66; alanine, −2.74; threonine, −2.94; serine, −3.07; γ-aminobutyric acid, −3.17; lysine, −3.05; glutamic acid, −3.69; and arginine, −4.08. Preliminary studies at other pH values show that, at pH 7, the charged amino acids partition into the octanol phase in their charged form. These and literature results can be reconciled with partition coefficient additivity rules most easily by assuming that (a) structural changes affect amino acid partitioning by only 0.6 of their effect on more lipophilic molecules, and (b) in octanol the NH_3^+ and COO^- moieties "self-solvate" polar side-chains as well as each other. The experimental values for the partition coefficients correlate moderately well with the hydrophobicity scale of Nozaki and Tanford [*J. Biol. Chem.* **246**:2211-2217 (1971)].

INTRODUCTION

Amino acids are necessary constituents of all living systems, primarily as building blocks for peptides, structural proteins, and enzymes, but also as neurotransmitters and neurotransmitter precursors. Despite the importance of aqueous/lipid partitioning equilibria to the conformation of macromolecules and the biological properties of molecules, reliable partition coefficients of many important polar amino acids seem to be missing from the literature. No doubt this gap is the result of the unusually hydrophilic character conferred on amino acids by their zwitterionic character, and the lack of any conventional chemical method of sufficient sensitivity for measuring the consequently small concentrations in octanol.

In this paper we describe a radiochemical method capable of measuring femtomoles (10^{-15} mole) of compound in small volumes, and the application of this method to the determination of octanol/water distribution coefficients of small hydrophilic molecules. Following convention, we refer to an experimentally measured value as a "distribution coefficient" and abbreviate its logarithm as logD,[1] reserving the term "partition coefficient" and the abbreviation logP for the behavior of any species having no net charge. However, as will be seen, our results do not support the implied assumption that a charged species does not partition into the octanol layer. We have sought correlations between our values and the "hydrophobicity scale" proposed for peptides by Nozaki and Tanford (1), as well as with "calculated" partition coefficients and experimental solubility ratios.

METHODS

Distribution coefficient measurements. L-[^{3}H]Alanine (82.7 Ci/mmole), L-[^{3}H]arginine (19.7 Ci/mmole), L-[^{3}H]GABA (28.2 Ci/mmole), L-[^{3}H]glutamic acid (40.3 Ci/mmole), L-[^{3}H]histidine (10.6 Ci/mmole), L-[^{3}H]lysine (74.3 Ci/mmole), [^{3}H]nipecotic acid(46.0 Ci/mmole), L-[^{3}H]proline (15.7 Ci/mmole), L-[^{3}H]serine (11.3 Ci/mmole), L-[^{3}H]threonine (1.4 Ci/mmole), L-[^{3}H]tryptophan (16.8 Ci/mmole), L-[^{3}H]valine (56.8 Ci/mmole), and L-[^{14}C]GABA (205 mCi/mmole) were obtained from New England Nuclear Corporation (Boston, Mass.). The radiochemical purity of all compounds was greater than 97%. A maximum of 3% impurity would not affect the observed logD measurements if the lipid solubility of the impurity were less than or equal to that of the authentic amino acid. The numerous extractions in our procedure would eliminate the effect of an impurity with greater lipid solubility.

Aliquots (40 μCi) of the tritiated amino acids or 10 μCi of [^{14}C]GABA were evaporated to dryness under nitrogen. The amino acids were redissolved in 4 ml of 10 mM sodium phosphate buffer (pH 7.0), and the specific activity (DPM per milliliter) was determined by liquid scintillation spectrometry. The aqueous material was ex-

[1] The abbreviations used are: logD, logarithm of the distribution coefficient; logP, logarithm of the partition coefficient; GABA, γ-aminobutyric acid; DPM, disintegrations per minute.

0026-895X/81/030602-07$02.00/0

tracted four times with octanol, as described below, to remove any labeled organic contaminants or exchanged tritium, and the final extraction was used to calculate the partition coefficient. All determinations were performed in triplicate.

For the first octanol extraction, 1-ml aliquots of the buffered amino acid solutions were added to plastic-stoppered centrifuge tubes containing 1 ml of octanol which had been saturated with water. The tubes were vigorously agitated for 30 sec using a Vortex mixer, and the layers were separated by centrifugation at 2000 × *g* for 15 min at 20°. Increasing or decreasing the mixing time did not significantly affect the extraction of amino acid into octanol. In a preliminary study, threonine yielded logD values of −2.89 if samples were agitated for 15 sec and −2.90 for 30- or 60-sec agitation. Radioactivity was determined in a 500-μl aliquot of the octanol layer, while the remaining octanol was removed by aspiration. Aliquots (1 ml) of octanol were added to all tubes and the extraction into octanol was repeated three more times. Radioactivity was determined in 750-μl aliquots of each octanol layer. After the fourth octanol extraction, radioactivity in the aqueous layer was also determined. The logD was calculated as log $[DPM_{octanol}]/[DPM_{buffer}]$ for the fourth octanol extraction. The DPM in the aqueous phase of earlier octanol extractions were estimated by adding the DPM in subsequent octanol phases to the DPM in the final aqueous phase, thus allowing us to obtain and compare distribution coefficients for the first three octanol extractions.

In a subsequent pilot study, the effect of pH on the observed logD for glutamic acid, arginine, and lysine was determined in the same manner except that the aqueous phase was either 10 mM sodium acetate buffer (pH 4.0) or 10 mM sodium borate buffer (pH 9.0). The pH of the aqueous layer after the four octanol extractions was estimated using indicator paper and compared with the color change produced by the initial aqueous solution (the presence of ^3H precluded the use of a pH electrode). Any pH change appeared to be less than 1 pH unit toward neutrality.

Calculations. LogP values were estimated for the amino acid protomer having no formal charge (i.e., glycine = NH_2CH_2COOH). Experimental values (2) used in these estimations were as follows: ethylbenzene, 3.15; 2-phenylpropionic acid, 1.84; phenethylamine, 1.41; benzene, 2.13; indole, 2.00; imidazole, −0.08; propylamine, 0.48; valeric acid, 1.39; pyrrolidine, 0.46; isopropanol, 0.05; ethanol, −0.30; acetic acid, −0.17; propionic acid, 0.33; diethyl sulfide, 1.95; butyric acid, 0.79; piperidine, 0.85; catechol, 0.88; phenol, 1.48; octylguanidinium sulfate, −0.88; and the standard methylene increment of 0.5. Estimates were performed by addition and subtraction of partial structures in the usual way (3). For example, tyrosine = $(1.48_{phenol}) + (1.84_{2\text{-phenylpropionic acid}} - 2.13_{benzene}) + (1.41_{phenethylamine} - 3.15_{ethylbenzene}) = 0.55$.

Regression studies were performed on the PROPHET computer system, using standard procedures. The intercepts and slopes of the calculated regression equations are followed by their respective 95% confidence intervals in parentheses. The statistical parameters r^2, *s*, and *F* have their conventional meanings. The population of the uncharged species, necessary in order to estimate its partition coefficient, was calculated by using the Henderson-Hasselbach equation.

RESULTS AND DISCUSSION

Experimental

Distribution coefficient values. Our distribution coefficient values are given in Table 1 along with literature distribution coefficients for many of the more lipophilic amino acids. The degree of agreement between literature results and ours for alanine and tryptophan is satisfactory. In the c e of histidine the discrepancy shown is not easily rationalized; however, our value seems in better agreement with correlation studies reported below.

It is worth emphasizing the experimental precautions necessary to obtain reliable logD values for these very hydrophilic substances. Less than 1% of a labeled lipophilic impurity would give erroneously high values. To avoid this concern, we extracted three times with octanol to remove any such trace contaminants before obtaining the values reported. That such a precaution was both desirable and effective is shown in Fig. 1, which records the apparent logD of seven representative amino acids as a function of extraction number. As was anticipated, the first extraction gave logD values which for some of the more hydrophilic acids were as much as 0.5 log unit too high. However, the third and fourth extractions in all cases yielded almost identical results.

A second concern involved possible loss of experimentally significant quantities of compound from the aqueous phase by adsorption to glassware. This problem was

TABLE 1

Distribution coefficients [log ($c_{octanol}/c_{water}$)] of some amino acids

Experimental values are the mean of three determinations of the fourth octanol extraction; uncertainties are standard deviations. Partitioning conditions were as follows: pH = 7.0 (phosphate); ionic strength (μ) = 0.01 M; aqueous amino acid concentrations (AA) = 0.076–12.6 μM.

Amino acid	LogD		Initial DPM
	Experimental	Literature	
			× 10^6
1. Tryptophan	−1.11 (± 0.01)	−1.06[a]	20.79
2. Phenylalanine	—	−1.52[a]; −1.35[b]	—
3. Isoleucine	—	−1.69[b]	—
4. Leucine	—	−2.06[a]; −1.52[b]	—
5. Methionine	—	−1.87[b]	—
6. Histidine	−1.95 (± 0.01)	−2.52[b]	23.23
7. Tyrosine	—	−2.05[c]	—
8. Valine	−2.26 (± 0.01)	—	18.14
9. Dopa (3,4-dihydroxyphenylalanine)	—	−2.38[c]	—
10. Proline	−2.54 (± 0.01)	—	20.75
11. Nipecotic acid (3-carboxypiperidine)	−2.66 (± 0.01)	—	20.87
12. Alanine	−2.74 (± 0.01)	−2.96[a]; −2.72[b]	13.73
13. Threonine	−2.94 (± 0.01)	—	23.45
14. Serine	−3.07 (± 0.01)	—	24.63
15. Glycine	—	−3.21[a]; −3.00[b]	—
16. GABA (^3H)	−3.18 (± 0.01)	—	23.59
GABA (^{14}C)	−3.16 (± 0.01)	—	5.70
17. Lysine	−3.05 (± 0.01)	—	22.18
18. Glutamic acid	−3.69 (± 0.01)	—	21.35
19. Arginine	−4.08 (± 0.01)	—	11.46

[a] Fauchere *et al.* (4); AA = 10^{-3} M; other conditions not stated.
[b] Klein *et al.* (5); pH = 7.0 (cacodylate); μ = 0.1; AA not stated.
[c] Mack and Bonisch (6); pH = 7.4; μ = 0.067 M; "very low" AA.

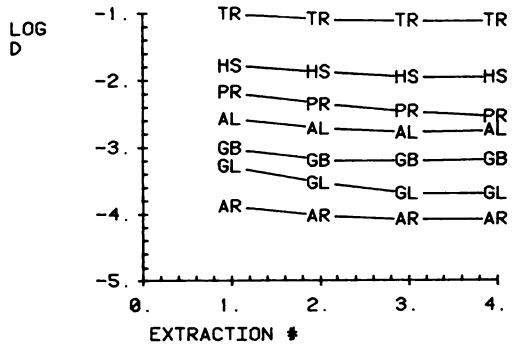

FIG. 1. *LogD for seven representative amino acids as a function of number of octanol extractions*

AL, alanine; *AR*, arginine; *GB*, tritiated GABA; *GL*, glutamic acid; *HS*, histidine; *PR*, proline; *TR*, tryptophan.

minimized by the use of (disposable) plastic vessels, and checked by obtaining a mass balance after all operations were completed. In all cases the initial DPM (last column of Table 1) and the DPM recovered from the four octanol phases plus the final aqueous phase agreed within 4%. Of course, low concentrations are helpful in interpreting the experimental values since they minimize the likelihood of amino acid aggregation in either phase. The actual concentrations of amino acids in the initial aqueous phase ranged from 0.076 μM for alanine to 12.6 μM for [^{14}C]-GABA.

We believe that the distribution coefficients in Table 1 constitute a firm basis for considering general problems in the transport of amino acids through biological membranes. Several recent investigations indicate that octanol is superior to many bulk solvents as a membrane model. In a comparison of the effects of mutual solvent solubility on the properties of partitioning system components, it was found that the dielectric constant and viscosity of octanol varied less upon equilibration with water than did those of cyclohexane or chloroform (7). The octanol/water system results correlated better with anesthetic potency than did oils or hydrocarbon solvent partitioning data (8). The latter finding has recently been extended to nerve blockade and erythrocyte stabilization,[2] both probably membrane-mediated phenomena. Experimental amino acid distribution coefficients may also be of value in considering the effects of hydrophilic or lipophilic side-chains on protein folding.

Effect of pH on observed logD of charged amino acids. Three of the amino acids have a side-chain which is ionized at physiological pH, thus conferring a net charge on the amino acid (positive for arginine and lysine, negative for glutamic acid). Most researchers assume that only the uncharged form of ionizable molecules can partition into the octanol phase. Table 2 gives the observed logD values for these amino acids at pH 4.0, 7.0, and 9.0. The results establish unmistakably that the logD values observed at pH 7 predominantly involve partitioning of the charged side-chain, in the cases of arginine and glutamic acid, to the virtual exclusion of the uncharged form. The underlying argument, using arginine as an example, runs as follows. The population of the uncharged form of arginine is 0.8% at pH 7 and 0.001% at pH 4. If only the uncharged form were partitioning, the logD observed at pH 4 for arginine should be 0.001/0.8, or 0.00125 of that observed at pH 7; i.e., at pH 4 the logD should be about −7. The observation that the logD is not affected by this 1000-fold increase in acidity is consistent only with the hypothesis that the positively charged form, which is the tautomer predominant at both pH values, is the major partitioning form. Therefore, the logP of protonated arginine must be about −4.08. From the observation of logD = −3.02 at pH 9, where one-half of the arginine is unprotonated, the logP of unprotonated arginine can be shown to be about −3.0. This difference of only 1 log unit between the logP values of protonated and unprotonated forms also contrasts with observations of 4.0 log unit differences between the charged and uncharged forms of more lipophilic molecules (3). An identical argument can be made for glutamic acid, using the pH 9 data. With lysine the argument would be more complex, because there is an 0.38 log unit decrease in logP from pH 7 to pH 4, or a halving of the concentration found in the octanol phase. Nevertheless, halving is much larger than the 1000-fold decrease in protonated lysine, and so it seems clear that, at pH 7, most, but not all, of the observed octanol concentration of lysine is protonated as well.

[2] R. D. Cramer, III. BC(DEF) parameters. IV. Correlation with biological data, in preparation.

TABLE 2

Effect of pH in the aqueous layer on the observed octanol/water distribution coefficients (logD) and calculated population of charged species, for three amino acids bearing a non-zero net charge at pH 7

Amino acid	pK$_a$[a]	pH 4		pH 7		pH 9	
		LogD	% Uncharged[b]	LogD	% Uncharged[b]	LogD	% Uncharged[b]
Arginine	9.09 (2.18, 13.2)	−4.08 (± 0.01)	0.001%	−4.08 (± 0.01)	0.8%	−3.02 (± 0.03)	44.0%
Glutamic acid	4.25 (2.19, 9.67)	−3.06 (± 0.01)	63.0%	−3.69 (± 0.01)	0.16%	−3.75 (± 0.01)	0.001%
Lysine	8.90 (2.20, 10.28)	−3.43 (± 0.01)	0.001%	−3.05 (± 0.01)	1.2%	−2.64 (± 0.02)	51%[c]

[a] pKa values from reference 9. Those in parentheses have little effect on the percentage uncharged amino acid present in the pH range studied.

[b] Based on a literature tabulation of solutions to the Henderson-Hasselbach equation (ref. 10). Only the pK$_a$ values not in parentheses were considered unless otherwise stated.

[c] Considering effects of both 8.90 and 10.28 pK$_a$ values.

Calculated

Partition coefficient values. The data set of Table 1 represents a good test of the assumptions familiar in partition coefficient work which have not been explored extensively for such strongly hydrophilic molecules. If these additivity assumptions hold for amino acids, there should be a good correlation between the observed logP and values estimated from additivity assumptions, and the slope of that correlation should equal to 1.0.

A major unknown for amino acids is the effect on partition coefficient of the equilibrium between the zwitterion and the tautomer having no formal charge; this equilibrium strongly favors the zwitterion even in the crystalline form. Because this feature is common to all compounds listed in Table 1, its effect on partition coefficient can be treated as an implicit dummy parameter, that is, as the intercept in the proposed correlation between observed distribution coefficient and calculated partition coefficient. We have already discussed a second complication, the net charge at pH 7 on three amino acids (No. 17–19).

Data associated with the various regression analyses to be discussed are assembled in Table 3. The first column gives the experimental logD from Table 1, discordant values from the literature being averaged for those amino acids not measured in our laboratory. In the second column appear estimates of logP for the amino acid, calculated for the protomer having no formal charges (see Methods for the list of model compounds). The third column contains the experimental logP, the logarithm of the partition coefficient necessary for the uncharged species assuming that only the uncharged species contributes to the observed distribution coefficient. (Of course, our data indicate that this assumption cannot be valid for the charged amino acids. Nevertheless, for the moment we adhere to convention.)

If the additivity rules and the assumption that the lipid phase contains only the uncharged form are valid, then correlation of Columns 2 and 3 of Table 3 should yield a straight line of unit slope. Instead, the actual regression equation is

$$\text{LogP}_{exp} = -2.00(\pm 0.44) + 0.100(\pm 0.22) \times \text{logP}_{calc} \quad (1)$$

$$r^2 = 0.045;\ s = 0.704;\ F(1,17) = 0.81\ (p = 0.38)$$

Evidently the correlation between calculated and experimental partition coefficients is not significant.

Examination of the residual plot (Fig. 2) shows the three points represented by *asterisks,* in particular the point for arginine, to be outliers. Not unexpectedly, these are the amino acids, Nos. 17–19, whose side-chains are charged at physiological pH. Therefore, Compounds 17–19 were excluded, and the correlation of Columns 2 and 3 of Table 3 was repeated. The equation was somewhat improved:

$$\text{LogP}_{exp} = -1.89(\pm 0.35) + 0.408(\pm 0.23) \times \text{logP}_{calc} \quad (2)$$

$$r^2 = 0.408;\ s = 0.504;\ F(1,14) = 9.66\ (p = 0.008)$$

However, the slope is less than 1.0 and the unexplained variance(s) is much greater than would be expected to result from uncertainty in estimates of logP.

Thus far the possibility of internal hydrogen bonding has not been considered. Bringing the charged amino acid moieties into the octanol phase requires the energetically costly process of stripping their water of solvation. Whenever polar side-chains exist having a conformation suitable for internal solvation, the process of octanol solvation will be facilitated. Such side-chains are present in serine and threonine (the hydroxyl groups), histidine (the imidazole moiety), methionine (the sulfur atom), and tryptophan (the indole-NH moiety), but, for conformational reasons, not in tyrosine or dopa. To allow

TABLE 3

Multiple regression study of amino acid partition coefficients

Amino acid	LogD, experimental	LogP		Total H-bond	Non-α-amino acid	Y-fitted	Residual
		Calculated	Experimental				
1. Tryptophan	−1.11	−0.03	−1.11	1.0	0.0	−0.90	−0.21
2. Phenylalanine	−1.43	0.1	−1.43	0.0	0.0	−1.74	0.31
3. Isoleucine	−1.69	0.13	−1.69	0.0	0.0	−1.72	0.03
4. Leucine	−1.79	0.15	−1.79	0.0	0.0	−1.70	−0.09
5. Methionine	−1.87	−1.6	−1.87	1.0	0.0	−1.91	0.04
6. Histidine	−1.95	−2.11	−1.91	1.0	0.0	−2.24	0.33
7. Tyrosine	−2.05	−0.55	−2.03	0.0	0.0	−2.15	0.12
8. Valine	−2.26	−0.35	−2.26	0.0	0.0	−2.03	−0.23
9. Dopa	−2.38	−1.15	−2.37	0.0	0.0	−2.54	0.17
10. Proline	−2.54	−0.85	−2.54	0.0	0.0	−2.35	−0.19
11. Nipecotic acid	−2.66	−0.46	−2.66	0.0	1.0	−2.76	0.10
12. Alanine	−2.74	−1.41	−2.74	0.0	0.0	−2.71	−0.03
13. Threonine	−2.94	−3.0	−2.94	1.0	0.0	−2.81	−0.13
14. Serine	−3.07	−3.35	−3.07	1.0	0.0	−3.04	−0.03
15. Glycine	−3.11	−1.91	−3.11	0.0	0.0	−3.03	−0.08
16. GABA	−3.17	−0.95	−3.17	0.0	1.0	−3.07	−0.10
17. Lysine	−3.05	−1.55	−1.15	0.0	0.0		
18. Glutamic acid	−3.69	−2.2	−0.94	0.0	0.0		
19. Arginine	−4.08	−5.93	−1.99	0.0	0.0		

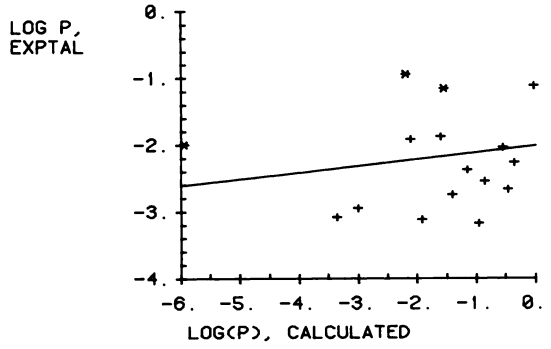

FIG. 2. *Relationship between logD and calculated logP*
The *x* axis shows logP as calculated for the amino acids (+) in the tautomer having no formal charge; the *y* axis shows experimental logD. The regression line (——) of Eq. 1 is also shown. *Asterisks* designate an amino acid having an ionizable side-chain.

regression involving this feature, the presence of one of the "side-chain-solvating" moieties is indicated by a "dummy" variable, equal to 0 or 1, in Column 4 of Table 3.

Such a "self-solvation" process might also explain differences in the behavior of α-amino acids from the other amino acids, No. 11 and No. 16 in Table 1. The carboxyl anion and ammonium cation of a zwitterionic α-amino acid are effectively solvating one another for parts of their radii, and this phenomenon reduces the energetic cost of moving the zwitterion into octanol. But GABA (11) and nipecotic acid (12) are known from NMR studies to adopt conformations in water in which the carboxyl and ammonium moieties are appreciably distant. Although there are self-solvating conformations available to GABA and nipecotic acid, their achievement in octanol would cost entropy for GABA and enthalpy (COO^- must become axial) for nipecotic acid. Consequently, a second dummy variable, distinguishing α-amino acids from others by the presence of a 1 for the latter, has been added as Column 5 of Table 3. To illustrate the hypothesized self-solvation interactions, a ball-and-stick representation of an internally solvated conformer of methionine is shown in Fig. 3.

Multiple regression of the experimental partition coefficient against these two dummy parameters as well as the calculated partition coefficient yields the following highly significant equation:

$$\text{LogP}_{exp} = -1.80 + 0.64(\pm 0.11) \times \text{logP}_{calc} + 0.918(\pm 0.26) \times D_{sc} - 0.66(\pm 0.29) \times D_{naa} \quad (3)$$

$$r^2 = 0.928;\ s = 0.191;\ F(3,12) = 51.2\ (p = 0.0001)$$

where D_{sc} indicates the presence of one of the five solvating side-chains and D_{naa} indicates a compound not an α-amino acid. The values calculated from this equation (*Y*-fitted) and the residuals are given in the last two columns of Table 3. Plotting *Y*-fitted against experimental logP (Fig. 4) does not suggest any marked individual outliers.

Equation 3 is probably as high a quality equation as one might expect for fitting calculated to experimental logP values in a situation where self-solvation is plainly

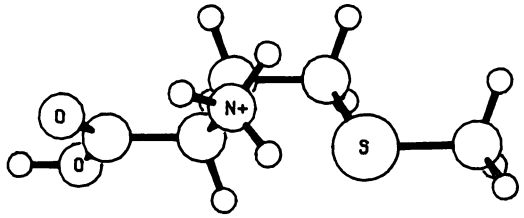

FIG. 3. *Possible conformer of methionine in which the —S—, $—NH_3^+$, and —COO— moieties partially solvate one another.*

of major importance. The equation also is easy to understand. As expected, amino acids with solvating side-chains are relatively stable in octanol, by an average of 0.92 log unit, whereas α-amino acids are more stable in octanol than other types by an average of 0.66 log unit. The intercept, −1.80, represents the additional energetic cost of transporting the zwitterion into octanol, as compared with the tautomer without formal charges for which the partition coefficients were estimated. Whether this energetic penalty is paid by transference of zwitterion directly into octanol, by self-neutralization to the uncharged form followed by transference, or by some combination of these events, cannot be established without a quantitative assay for one of these tautomeric forms in octanol.

A surprising feature of Eq. 3 is the fact that the coefficient relating calculated and experimental logP values is 0.64 (± 0.11), which is significantly less than 1.0 by approximately 7 SD. According to this equation, the $—CH_2—$ unit that increments the logP of familiar lipophilic molecules by 0.5 log unit increases the lipophilicity of an amino acid by only 0.64 × 0.5, or 0.32 unit. In essence, the equation suggests a "scale change" in the normally additive character of logP, only the order but not the magnitude of lipophilicity changes being preserved with these unusually hydrophilic molecules.

Of course, the validity of an equation such as Eq. 3 depends upon the model used for logP calculation. A logP model more recently proposed by Hansch and Leo (13) includes numerous group interaction corrections similar in rationale to our self-solvation terms. It yields logP values which are numerically closer together than are the calculated logP values in Table 3. Nevertheless, when the three charged amino acids are omitted, the

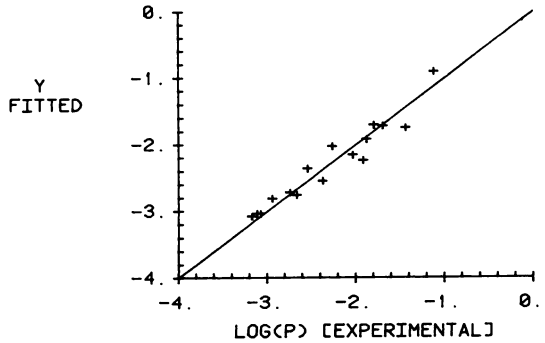

FIG. 4. *Multiple regression fit to amino acid data*
Comparison of the experimental logP values with the logP values computed using Eq. 3 (*Y*-fitted). +, amino acids; ——, regression-line.

equation comparable to Eq. 3 above also includes a slope which is significantly less than 1.0. Another relevant observation is the very small difference in log P which we found between the charged and uncharged protomers of amino acids such as arginine, as compared with that found with more lipophilic molecules, This, too, is a scale change.

We conclude that there is reason to suspect that familiar structure/logP relationships may not be directly applicable to more hydrophilic molecules. Indeed, the behavior of proteins themselves indicates that the additivity rules cannot be correct for extremely hydrophilic molecules. Because individual peptide units are always hydrophilic, strict additivity would predict that each peptide added to a protein should further increase its hydrophilicity, but it is well known that many large proteins are naturally embedded in membranes. The only real issue for those interested in structure/logP relationships seems to be, "under what circumstances do the deviations from additivity for more hydrophilic molecules become significant?"

Correlation of experimental logP values with other lipophilicity indices. Most other assessments of amino acid side-chain lipophilicity can be related to a classical compilation of amino acid solubility ratios by Edsall and Scatchard (14). A solubility ratio (*SR*) is the amount (weight or moles per volume) which can be dissolved in ethanol divided by the amount that can be dissolved in water. The correlation between the solubility ratios reported by Edsall and Scatchard (14) and our distribution coefficients for the six amino acids common to both data sets is outstanding:

$$\text{Log}(SR) = 0.30(\pm 0.27) + 1.33(\pm 0.10) \times \text{logP}_{exp} \quad (4)$$

$$r^2 = 0.994;\ s = 0.06;\ F(1,4) = 667.1\ (p = 0.001)$$

The values of log(*SR*) appear in the first column of Table 4, and the experimental logP values were given in Column 3 of Table 3. Glutamic acid was excluded from this correlation because of the ambiguities related to its ionizable side-chain. Given the differences between logP values and solubility ratios—octanol versus ethanol, mutual solvent saturation versus none, solute dilution versus saturation—Eq. 4 represents an astonishingly high correlation.

Based on a complicated thermodynamic treatment of these and additional amino acid solubilities obtained in mixed solvents, dioxane/water as well as ethanol/water, Nozaki and Tanford (1) proposed a "hydrophobicity scale" for "transfer of amino acid side chains from 100% organic solvent to water at 25°", and gave values for 12 of the most lipophilic amino acids. The 11 of these values for which "logP, experimental" values are also known comprise the second column of Table 4. The correlation between the Nozaki-Tanford hydrophobicity scale and logP values is only fair regardless of whether our experimental or calculated logP values are used:

$$\text{N-T hydrophobicity} = 4.4(\pm 1.3) + 1.36(\pm 0.59) \times \text{logP}_{exp} \quad (5)$$

$$r^2 = 0.685;\ s = 0.59;\ F(1,9) = 19.6\ (p = 0.001)$$

$$\text{N-T hydrophobicity} = 2.3(\pm 0.50) + 0.68(\pm 0.30) \times \text{logP}_{calc} \quad (6)$$

$$r^2 = 0.692;\ s = 0.59;\ F(1,9) = 20.2\ (p = 0.001)$$

Despite a similarity in statistical parameters, Eqs. 5 and 6 differ greatly. This is clearly indicated in the corresponding plots (Figs. 5 and 6) in which the outlying compounds for each equation are designated by abbreviations instead of + symbols. Histidine is a particularly poorly fitting amino acid in the more important correlation between the hydrophobicity scale and the experimental logP (Fig. 5). Histidine has a side-chain with a pK_a of 6.0, and this could affect its distribution if the pH was not controlled during the solubility ratio measurements. If histidine is therefore omitted from Eq. 5 as being incommensurate, a similar slope and intercept but a higher r^2 of 0.85 and a lower s of 0.41 are obtained.

We conclude that the Nozaki-Tanford hydrophobicity scale and the experimental logP values are at least comparable. Octanol/water partitioning is more appealing to us than are ethanol/water solubility ratios as a model of

TABLE 4

Ethanol:water solubility ratios and hydrophobicity scale values for some amino acids

Log(*SR*) are the logarithms of the ratios of the solubility in ethanol to the solubility in water (14). Hydrophobicity values are derived from solubility measurements (1).

Amino acid	Log(*SR*)	Hydrophobicity value
1. Tryptophan	—	3.4
2. Phenylalanine	—	2.5
4. Leucine	−2.13	1.8
5. Methionine	—	1.3
6. Histidine	—	0.5
7. Tyrosine	—	2.3
8. Valine	−2.65	1.5
9. Dopa	—	1.8
12. Alanine	−3.34	0.5
13. Threonine	−3.55	0.4
14. Serine	−3.85	0.3
15. Glycine	−3.87	—

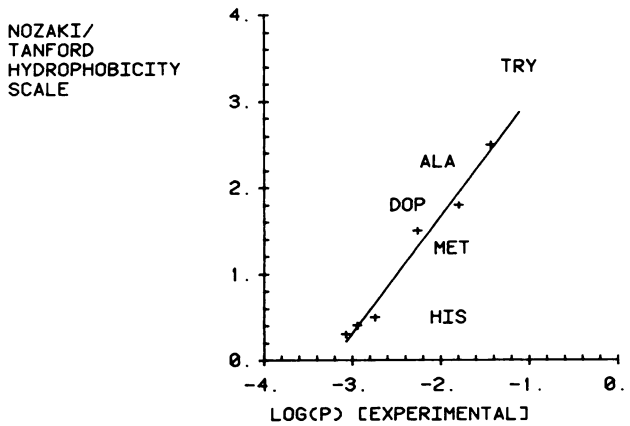

FIG. 5. *Relationship between the Nozaki-Tanford (1) hydrophobicity scale and experimental logP*

Points more than 0.5 log unit from the regression line (——) are indicated by their three-letter amino acid codes (centered at the experimental point); the other points are indicated by +.

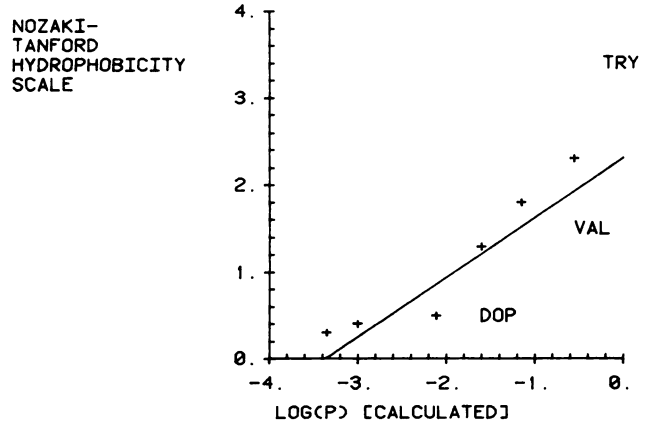

FIG. 6. *Relationship between the Nozaki-Tanford (1) hydrophobicity scale and calculated logP*

Points more than 0.5 log unit from the regression line (——) are indicated by their three-letter amino acid codes (centered at the experimental point); the other points are indicated by +.

solvation effects upon protein folding. Nevertheless, the Nozaki-Tanford hydrophobicity scale is well-established in the literature, and we note for those who prefer this existing model that Eq. 5 can be used with our data to estimate side-chain hydrophobicities of amino acids not studied by Nozaki and Tanford. However, no measurements to date indicate whether the side-chain hydrophobicities experimentally observed for zwitterionic amino acids are transferable to the uncharged —NH— and —CO— moieties adjoining the same residue in a peptide. Our "scale change" finding suggests that only the rank order, not the relative magnitudes or energetics, of the side-chain hydrophobicities is likely to be conserved.

CONCLUSIONS

Reliable experimental values for the distribution coefficients of amino acids would seem to be a necessary condition for rigorous discussion of issues such as the effects of lipophilicity or hydrophilicity on protein folding or amino acid transport. Table 1 contains values for 16 of the 21 "important" amino acids. The experimental procedure we describe should produce reliable results for any other suitably labeled compound.

Correlation analysis of these data suggests that such unusually hydrophilic molecules do, on the one hand, qualitatively follow the additive-constitutive rules worked out for the partitioning of much more lipophilic substrates, but, on the other hand, quantitatively respond more weakly to a change in substituent lipophilicity. We have also shown that three amino acids having an ionizable side-chain must partition into lipid as the protomer bearing a net charge. Finally, "internal hydrogen-bonding" or "self-solvating" seems to be as important for groups such as indole—NH or alkyl—S— as it is for alkyl—OH or imidazole.

A fair correlation between these distribution coefficients and a previously proposed "hydrophobicity" scale allows interconversion between the two scales.

ACKNOWLEDGMENT

We thank Dr. W. Bondinell for a valuable experimental suggestion.

Note added in proof. An r^2 of 0.012 was found between octanol/water partition coefficients and vapor/water partition coefficients (15) for the 12 amino acids tested in both systems.

REFERENCES

1. Nozaki, Y., and C. Tanford. The solubility of amino acids and two glycine peptides in aqueous ethanol and dioxane solutions. *J. Biol. Chem.* **246**:2211–2217 (1971).
2. Pomona College Medicinal Chemistry Project Data Bank, Claremont, Calif. (1981).
3. Leo, A., C. Hansch, and D. Elkins. Partition coefficients and their uses. *Chem. Rev.* **71**:525–616 (1971).
4. Fauchere, J. L., K. Q. Do, P. Y. C. Jow, and C. Hansch. Unusually strong lipophilicity of 'fat' or 'super' amino-acids, including a new reference value for glycine. *Experientia* **36**:1203–1204 (1980).
5. Klein, R., M. Moore, and M. Smith. Selective difffusion of neutral amino acids across lipid bilayers. *Biochim. Biophys. Acta* **233**:420–433 (1971).
6. Mack, F., and H. Bonisch. Dissociation constants and lipophilicity of catecholamines and related compounds. *Naun-Schmeideberg's Arch. Pharmacol.* **310**:1–9 (1979).
7. Smith, R. N., C. Hansch, and M. M. Ames. Selection of a reference partitioning system for drug design work. *J. Pharm. Sci.* **64**:599–606 (1975).
8. Franks, N. P., and W. R. Lieb. Where do general anesthetics act? *Nature (Lond.)* **274**:339–342 (1978).
9. *Merck Index*, Ed. 9. Merck & Company, Rahway, N. J. (1976).
10. Purcell, W. P., G. E. Bass, and J. M. Clayton, *Strategy of Drug Design: A Guide to Biological Activity*. Wiley-Interscience, New York (1973).
11. Ham, N. S. NMR Studies of Solution Conformations of Physiologically Active Amino Acids, in *Molecular and Quantum Pharmacology* (E.D. Bergmann and B. Pullman, eds.), D. Reidel, Boston, 261–268 (1974).
12. Brehm, L., P. Krogsgaard-Larsen, G. A. R. Johnston, and K. Schaumburg. X-Ray Crystallography and PMR Spectroscopic Investigations of Nipecotic Acid, a Potent Inhibitor of γ-Aminobutyric Acid Uptake, *Acta Chem. Scand.* B30:542–548 (1976).
13. Hansch, C., and A. Leo, *Substituent Constants for Correlation Analysis in Chemistry and Biology*. Wiley-Interscience, New York (1979).
14. Edsall, J. T., and G. Scatchard. Solubility of Amino Acids, Peptides and Related Substances in Water and Organic Solvents, in *Proteins, Amino Acids, and Peptides as Ions and Dipolar Ions* (E. J. Cohn and J. T. Edsall, eds.), Reinhold Publishing Corp., 177–196 (1943).
15. Wolfenden, R. V., P. M. Cullis, and C. C. F. Southgate. Water, protein folding, and the genetic code. *Science (Wash., D. C.)* **206**:575–577 (1979).

Send reprint requests to: Dr. Libby M. Yunger, F12, Department of Pharmacology, Smith Kline & French Laboratories, Philadelphia, Pa. 19101.