

Extension of the Fragment Method to Calculate Amino Acid Zwitterion and Side Chain Partition Coefficients

Donald J. Abraham¹ and Albert J. Leo²

¹Department of Medicinal Chemistry, University of Pittsburgh, Pittsburgh, Pennsylvania 15261; ²Pomona College Medicinal Chemistry Project, Seaver Chemistry Laboratory, Claremont, California 91711

ABSTRACT The fragment method of calculating partition coefficients (P) has been extended to include the common amino acids (AAs). The results indicate that polar and charged side chains influence the hydrophobicity of atoms in the side chain in a predictable manner. Field effects, as evidenced through polar proximity factors and bond factors, need to be considered for accurate estimation of transfer phenomena. The calculated log P and ΔG° values of the 20 AAs agree well with the observed values. Pro calculates to be more hydrophilic than the observed log P. Hydrophobicity scales for peptide side chain residues are compared and evaluated in terms of suitability. Calculated π values for nonpolar side chain residues agree well with the observed values; calculated values for uncharged polar side chain residues deviate by about 0.6 log units except for Gln and Cys; and polar side chain residues with charged side chains calculate as too hydrophilic. Reasons for the differences are explored. We also suggest that tightly bound water to polar moieties in amino acids and peptides may be transferred into the octanol phase during partitioning experiments. A quantitative methodology is presented which characterizes the thermodynamic partitioning of groups and individual atoms in amino acids and proteins.

Key words: amino acid hydrophobicity scale, QSAR, amino acids, peptide side chains

INTRODUCTION

The hydrophobic nature of amino acids (AAs) has been of long-standing interest to molecular biologists. The classification of hydrophilic vs. hydrophobic AAs in proteins varies from study to study.¹⁻¹¹ Hydrophobicity scales are based upon solution measurements, empirical calculations, and partitioning between degree of solvent accessible surface area and buried areas in proteins; for example, see Nozaki and Tanford⁴, Tanford,^{5,6} Eisenberg,⁷ Eisenberg and McLachlan,⁸ Wolfenden,¹¹ Wolfenden et al.,¹⁰ and Rose et al.¹

We now demonstrate that it is possible to calculate the partition coefficients ($\log P_{\text{octanol/water}}$) of AAs by using standard methodology developed by Hansch and Leo,¹² Rekker,¹³ and Nys and Rekker.¹⁴ Comparison of calculated vs. experimentally determined values of log P for all 20 AAs in octanol-water¹⁵ verify the ability to assign hydrophobic parameters to individ-

ual atoms and small components of the AA side chains. Therefore, this study produces an alternative approach to generalizing AAs as hydrophobic or hydrophilic. It will become clear that AA side chain hydrophobicities vary from atom to atom and are influenced by the proximity of adjacent polar groups and charges.

Only one other study, by Eisenberg and McLachlan,⁸ has attempted to assign thermodynamic parameters to individual atoms or groups of atoms in protein side chains. Their system estimates atomic free energies of transfer from one complete set of experimental measurements on substituted amino acid analogs. Our method utilizes *fundamental fragment values* obtained from partitioning experiments performed on thousands of compounds.¹² The fundamental fragment constants are defined as nonambiguous entities that can be utilized in computer programs or algorithms so that the partition coefficient of any organic molecule can be estimated, whether known or contemplated for synthesis.

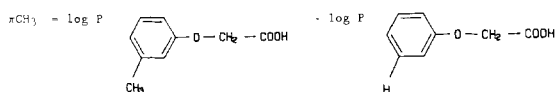
BACKGROUND

The initial reports of Meyer¹⁶ and Overton¹⁷ on the narcotic activity of many simple organic molecules in relation to their oil-water partition coefficients illustrated for the first time a means of relating hydrophobicity to biological activity. Since that time there has been an intense effort, primarily by medicinal chemists, to quantify hydrophobicity with biological activity [QSAR (quantitative structure activity relationships)]. Corwin Hansch and co-workers¹⁸⁻²¹ have pioneered this approach with the development of the hydrophobic π constants for substituent groups which can be used to estimate biological activity or to calculate new partition coefficients. For a comprehensive review of this work see Hansch and Leo.¹²

The π constants are defined as the change in log P when a non-hydrogen-bonding hydrogen atom is replaced by the atom or group in question. They are derived from the log of the partition coefficients (P) in a standard set of compounds. For example:

Received December 16, 1986; accepted May 15, 1987.

Address reprint requests to Donald J. Abraham, Department of Medicinal Chemistry, 728 Salk Hall, University of Pittsburgh, Pittsburgh, PA 15261, or to Albert J. Leo, Pomona College Medicinal Chemistry Project, Seaver Chemistry Laboratory, Claremont, CA 91711.



A newer parameter, the fragment constant f , has been introduced by Rekker¹³ and Nys and Rekker.¹⁴ The fragment method allows calculations to be made when no "parent" value has been measured. The π and f values for any group are related as follows:

$$f_{\text{CH}_3} = \log P_{\text{CH}_4} - f_{\text{H}}^*$$

$$f_{\text{CH}_2} = f_{\text{CH}_3} - f_{\text{H}} \text{ and } f_{\text{CH}} = f_{\text{CH}_2} - f_{\text{H}}$$

$$f(x) = \pi(x) + f_{\text{H}}, \text{ e.g., } f_{\text{CH}_3} = \pi\text{CH}_3 + f_{\text{H}};$$

$$f_{\text{CH}_3} = 0.66 + 0.23.$$

We will use fragment constants to calculate the partition coefficients of AAs by simple addition of f with other factors (F) that affect the partitioning equilibrium in the more complex solutes where summation of fragments alone leads to spurious results. The following equation summarizes this approach:

$$\log P = \sum_{n=1}^N a_n f_n + \sum_{m=1}^M b_m F_m$$

where a is the number of occurrences of fragment f of structural type n and b is the number of occurrences of factor F of structural type m . The F factors are empirically derived quantities that indicate increases (+) or decreases (−) in hydrophobicity that arise from chain or group branching [F_{cBr} (−) or F_{gBr} (−)]; bond types [F_b (−)]; and proximity effects of polar groups with n carbon separation [F_{Pn} (+)].

Chain or Group Branching Factors

It is well known that branched alkyl chains are more water soluble than their straight-chain isomers. Since there is no such compensating factor in the lipid phase, chain branching leads to a lower partition coefficient. The increased water solubility and decreased partition coefficient are usually ascribed to the smaller cavity needed to encompass the branched solute. Branching of a polar group also exhibits the same phenomenon. For example, isomeric pentanols with the hydroxyl group at a branch point also demonstrate an increased water solubility.²²

Alkyl branching is observed to decrease the partition coefficient by -0.13 log units ($F_{\text{cBr}} = -0.13$) and polar group branching [for example, as with a hydroxyl group] by -0.22 log units ($F_{\text{gBr}} = -0.22$).¹² The branching factors are necessary for calculation of $\log P$ of AAs that have branched groups (Val, Leu, Ile, Thr).

Bond Factors

It can be shown that it takes more energy to create a given-sized cavity in water (due to disruption of the hydrogen bonding network in water) than to create a

similar-sized cavity in octanol.^{23,24} In alkane solutes, where there are no polar interactions to confuse matters, the $\log P$ increases linearly with the volume and surface area of the cavity needed to accommodate each additional CH_2 unit from C_3 to C_8 . However, there is a much larger increase in $\log P$ in going from C_1 to C_2 . An explanation for this behavior is that the flexibility of larger chains reduces the degree of order in the solvation shell. The bond factor F_b with value of -0.12 log units expresses this decrease in $\log P$ for each subsequent bond beyond the first C_1 -to- C_2 bond. Therefore F_b is taken $(n-1)$ times where n equals the number of bonds between nonhydrogen atoms (see ref. 12 for details).† F_b will be used routinely in calculating the $\log P$ of the AAs as shown below. F_b for aliphatic rings (Pro) reduces the partition coefficient by -0.09 log units per bond.

Charged nitrogens (e.g., Lys, Arg, and the α -amino of the zwitterion) behave in a more dynamic way in lowering the partition coefficient. In effect, cationic charges are transmitted through the chain of adjacent carbon atoms or methylene groups, with the greatest decrease in hydrophobicity occurring nearest the protonated nitrogen. Bond factors that express this graded reduction in hydrophobicity begin to level off at the fifth bond to the normal F_b value of -0.12 log units, with the first bond, F_{b+1} , reducing hydrophobicity by -0.78 log units, the second, F_{b+2} , by -0.40 , the third, F_{b+3} , by -0.26 , and the fourth, F_{b+4} , by -0.19 . These values were obtained by comparison of the partition coefficients for several series of charged amino compounds (for details see ref. 12: pp. 37–43).

The cationic charge distribution in the side chains of Lys and Arg has not been demonstrated unequivocally. Earlier workers report the charge distribution of ammonium cations to adjacent nonpolar atoms.²⁵ However, more recent calculations do not demonstrate much charge delocalization.²⁶ A recent crystal structure of a drug bound to hemoglobin shows the methylenes in Lys 99 β to be in close contact with the amide oxygen of the antilipidemic agent bezafibrate²⁷ (see figures 7a,b). Other similar close contacts of Lys methylenes with surrounding polar groups in proteins occur. Whether such contacts arise from Van der Waals or dipolar interactions due to cationic charge distribution to adjacent methylenes, as we suggest here, needs to be explored. (See note added in proof.)

Polar Proximity Factors

Polar proximity factors arise from field effects. Polar atoms are most hydrophilic when isolated from

*In the Hansch and Leo system¹² the fragment value of hydrogen is taken as one-half the measured $\log P$ (octanol/water) of hydrogen gas; i.e., $0.5(0.45) \sim 0.23$.

†Bonds within fragments are not counted since they are included in the fragment value.

one another. When crowded together, some hydrophilicity is lost. This "proximity effect" is taken into account as a percentage of the negative fragment sum which is added back. When neither polar fragment carries a charge, this proximity effect drops off rapidly with distance: i.e., 42% loss for one intervening alkyl carbon, 26% for two, and 10% for three.¹² When either fragment contains a charged atom, the effect acts over greater distances¹² (Table I). Even though the "hybrid" amino acid fragment (see next section) is electrically neutral, the "charged" coefficient scale applies. Note in the examples that follow, the α -carbon is counted to determine the intervening distance and the sign of the overall correction is positive.

CALCULATIONS

AA Fragment Constants

The fragment method of calculating the log P (octanol/water) from structure¹²⁻¹⁴ has been applied to neutral solutes with some success and has been incorporated in a computer program called CLOGP-3.²⁸ The method also has been extended to estimate what the distribution ratio of ion pairs would be under standard conditions (i.e., very low solute concentration, 0.1 M of low molecular weight counter ion).¹² In many applications these standardized distribution ratios can be treated as if they were true partition coefficients.

In order to define a fragment constant for the α -AA zwitterion, two problems unique to AAs require an extension of the values and procedures previously published for formulating the fragment from its component parts.¹²⁻¹⁴ The first problem concerns the nature of the structure of the zwitterion and the second involves the nature of the propagation of a zwitterion charge throughout the molecule.

The standard fragment values for neutral and charged carboxyl fragments are -1.11 (f_{COOH}) and -5.10 (f_{COO^-}),¹² respectively. The corresponding values for the amino fragment are -1.54 (f_{NH_2}) and -3.40 ($f_{\text{NH}_3^+}$),¹² respectively. However, the protonated amino fragment would be lower if it had not been measured with a rather large anion, Cl. Furthermore, the value for $f_{\text{NH}_3^+}$ of -3.40 was arrived

at by allowing for the ability of the ammonium ion to delocalize its positive charge along any attached alkyl chain.¹²⁻¹⁴ One would expect the new alpha α -amino zwitterion fragment, termed a "hybrid" fragment (f_{Hyb}), to have a value more negative than the average of the two ionic values (-5.1 and -3.4) which equals -4.25 . The idea for defining the "hybrid fragment" in a new manner is suggested from X-ray crystal structures of AAs that show that the α -amino zwitterion exhibits a close contact between the nitrogen and its hydrogen with one of the carboxyl oxygens.³⁰⁻³² The negative charge on carboxyl groups is usually distributed equally on each oxygen. However, in some crystal structures, Gln³³ and Asn,³⁰ one of the zwitterion carboxyl oxygens appears to be in the keto form (for a view of this close contact in Asn, see Fig. 1). In these cases, the oxygen in close contact with the ammonium ion hydrogen has more double-bond character. The close proximity of the amino proton to the carboxylate anion found in AAs may be the reason for the increased hydrophobicity (increase in log P) which is best represented by a hybrid fragment value of $f_{\text{Hyb}} = -4.51$. In the original fragment definition, only ordinary covalent bonds were considered in linking a multiatom fragment together. The new hybrid fragment representing two polar groups separated by a carbon differs in this respect (see structure of hybrid fragment below). It is probably fortuitous that this value is the sum of a protonated acid and a charged amine, i.e., $f_{\text{COOH}} (-1.11) + f_{\text{NH}_3^+} (-3.40)$, since the formulation of large fragments directly by summing smaller fragments has not been successful. Accurate evaluation of the hybrid values as -4.51 comes from and is justified by the AA experimental data.

Since AAs in this study all contain a common unit that includes f_{Hyb} and the α -carbon atom, it is convenient to define this moiety as a "superfragment" (f_{Sup}) which has the value -3.59 . The "superfragment" is calculated by using the hybrid fragment, the fragment for the α -carbon (CH), the factors for group branching, delocalized first and second bonds, and the polar proximity effect. Note that we consider the positive charge on the nitrogen to be preferentially

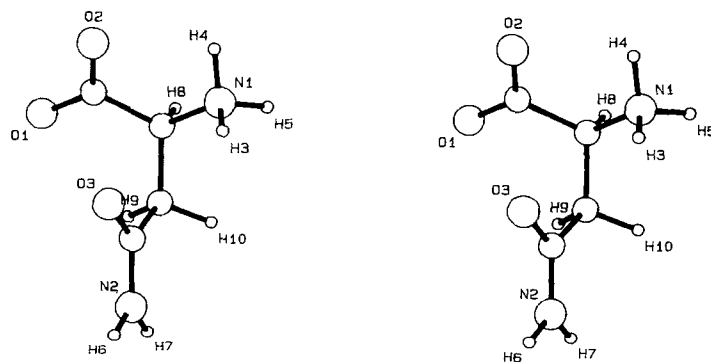
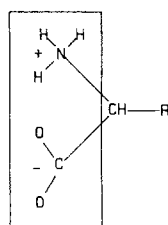


Fig. 1. Stereo diagram of L-asparagine³⁰ showing the close contact of the zwitterion amino proton H4 with the carboxyl O2 that has more double-bond character than O1.

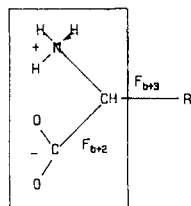
TABLE I. Fragment Constants, Bond Factors, Branching Factors, and Polar Proximity Factors Needed to Calculate the Partition Coefficients of the Amino Acids or Side Chains [log P ref.: both Fauchere and Pliska's¹⁵ and Yungster and Cramer's²⁹ data used]

Fragment Constants <i>f</i> and corresponding $\Delta G^\circ = 2.303RT(f)$											
<i>f</i>	CH ₃	CH ₂	CH	H	NH ₃ ⁺	NH ₂	COOH	COO ⁻	-S-	SH	CONH
	0.89	0.66	0.43	0.23	-3.40	-1.54	-1.11	-5.10	-0.79	-0.23	-2.71
ΔG°	1.21	0.90	0.59	0.31	-4.64	-2.10	-1.51	-6.93	-1.07	-0.31	-3.68
	aliphatic		phenyl	HOphenyl	imidazole		indole	guanide		carboxamide	
	OH		C ₆ H ₅	C ₆ H ₄ OH	C ₃ H ₃ N ₂		C ₈ H ₆ N	CH ₅ N ₃		CONH ₂	
<i>f</i>	-1.64	1.90	1.23	1.23	-0.31	1.91	2.60	-5.32		-2.11	
ΔG°	-2.23	2.58	1.67	1.67	-0.42	2.60		-7.23		-2.87	
Super Fragment $f_{\text{sup}} = -3.59$ Hybrid Fragment $f_{\text{Hyb}} = -4.51$											
For the details of the fragment calculations for the aromatic rings see Hansch and Leo ¹² and Table IV.											
Alkyl Chain Branching											
F_{Br}	-0.13										
ΔG°	-0.18										
Polar Group Branching											
F_{gBr}	-0.22										
ΔG°	-0.30										
Bond Factors											
Normal Chain	Ring										
	F_{Br}	F_{b+1}	F_{b+2}	F_{b+3}	F_{b+4}						

transmitted toward the carboxylate ion and secondarily toward the side chain. Therefore, the bond factor F_{b+2} is designated on the bond between the α -carbon and the carboxylate anion and the F_{b+3} bond factor on the bond between the α - and β -carbons.



Hybrid Fragment



Super Fragment

$$f_{\text{Sup}} = f_{\text{Hyb}} + f_{\text{CH}} + F_{\text{gBr}} + F_{b+1} + F_{b+2} + FP_1$$

$$= -4.51 + 0.43 - 0.22 - 0.78 - 0.40 - 0.42(f_{\text{Hyb}})$$

$$= -3.59$$

Calculation of the AA Partition Coefficients

The following examples will clarify the application of the established rules¹² with the new fragment values for the common amino acids. All f 's and F 's needed for these calculations are listed in Table I. All polar proximity factors for interactions of the zwitterion with polar groups on the side chains will use f_{Hyb} + $f_{\text{polar group}}$. Aromatic side chain residues, except His, do not need polar proximity factors indicating little interaction with the zwitterion. All calculated log P for AAs will be compared with the carefully measured values obtained by Fauchere and Pliska.¹⁵ The calculations for the AAs are divided into three categories: AAs that have no polar proximity effects; AAs that have polar proximity effects and uncharged side chain residues; and AAs that have polar proximity effects and charged side chain residues.

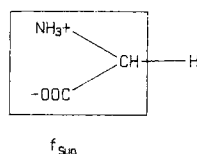
AAs That Have No Polar Proximity Effects

Glycine.

$$\text{Log } P_{\text{glycine}} = f_{\text{Sup}} - F_{\text{gBr}}^* + f_{\text{H}} = -3.14$$

$$= -3.59 - (-0.22) + 0.23$$

$$\text{Log } P_{\text{glycine}} (\text{obs}^{15}) = -3.25$$

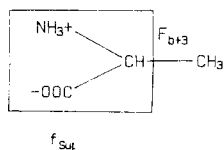


Alanine.

$$\text{Log } P_{\text{alanine}} = f_{\text{Sup}} + f_{\text{CH}_3} + F_{b+3} = -2.96$$

$$= -3.59 + 0.89 + (-0.26)$$

$$\text{Log } P_{\text{alanine}} (\text{obs}^{15}) = -2.89$$



Valine.

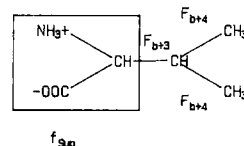
$$\text{log } P_{\text{valine}} = f_{\text{Sup}} + 2f_{\text{CH}_3} + f_{\text{CH}} + F_{b+3} + 2F_{b+4}^\dagger + F_{\text{cBr}}$$

$$= -2.15$$

$$= -3.59 + 2(0.89) + 0.43 + (-0.26)$$

$$+ 2(-0.19) + (-0.13)$$

$$\text{Log } P_{\text{valine}} (\text{obs}^{15}) = -2.08$$

 f_{Sup}

Leucine.

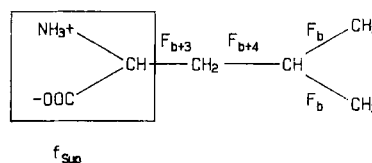
$$\text{Log } P_{\text{leucine}} = f_{\text{Sup}} + 2f_{\text{CH}_3} + f_{\text{CH}_2} + f_{\text{CH}} + F_{b+3}$$

$$+ F_{b+4} + 2F_b + F_{\text{cBr}} = -1.54$$

$$= -3.59 + 2(0.89) + 0.66 + 0.43 + (-0.26)$$

$$+ (-0.19) + 2(-0.12) + (-0.13)$$

$$\text{Log } P_{\text{leucine}} (\text{obs}^{15}) = -1.61$$

 f_{Sup}

Isoleucine.

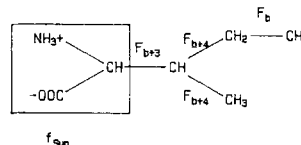
$$\text{Log } P_{\text{isoleucine}} = f_{\text{Sup}} + 2f_{\text{CH}_3} + f_{\text{CH}_2} + F_{\text{CH}} + F_{b+3}$$

$$+ 2F_{b+4} + F_b + F_{\text{cBr}} = -1.61$$

$$= -3.59 + 2(0.89) + 0.66 + 0.43 + (-0.26)$$

$$+ 2(-0.19) + (-0.12) + (-0.13)$$

$$\text{Log } P_{\text{isoleucine}} (\text{obs}^{15}) = -1.72$$

 f_{Sup}

Proline.

$$\text{Log } P_{\text{proline}} = [f_{\text{Sup}} - f_{\text{H}}] + 3f_{\text{CH}_2} + F_{b+1}^\ddagger + F_{b+2}^\ddagger +$$

$$2F_{b+3}^\ddagger = -3.42$$

$$[-3.59 - 0.23] + 3(0.66) + (-0.75)$$

$$+ (-0.37) + 2(-0.23)$$

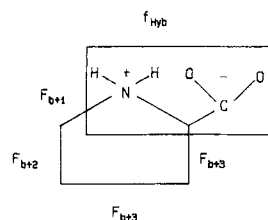
$$\text{Log } P_{\text{proline}} (\text{obs}^{15}) = -2.50$$

$$(\text{obs}^{29}) = -2.54$$

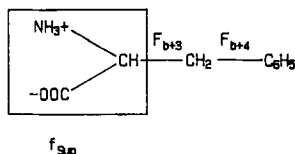
*Since glycine is the only amino acid without a branching beta carbon, we must subtract the group branch factor F_{gBr} from the super fragment.

†At a symmetrical branch point, the bond factors have an equal effect in each direction; i.e., both bonds to the terminal methyls are evaluated as F_{b+4} .

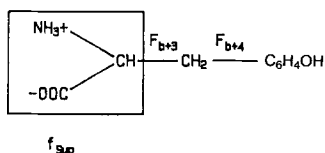
‡Indicates a bond factor for rings which is +0.03 log units more hydrophobic than F_{b+1} (-0.78 + 0.03), F_{b+2} (-0.40 + 0.03), or F_{b+3} (-0.26 + 0.03).

**Phenylalanine.**

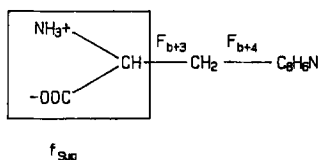
$$\begin{aligned}\text{Log } P_{\text{phenylalanine}} &= f_{\text{Sup}} + F_{\text{CH}_2} + f_{\text{C}_6\text{H}_5\#} + F_{b+3} \\ &\quad + F_{b+4} = -1.48 \\ &= -3.59 + 0.66 + 1.90 + (-0.26) + (-0.19) \\ \text{Log } P_{\text{phenylalanine}} (\text{obs}^{15}) &= -1.63\end{aligned}$$

**Tyrosine.**

$$\begin{aligned}\text{Log } P_{\text{tyrosine}} &= f_{\text{Sup}} + f_{\text{CH}_2} + [f_{\text{C}_6\text{H}_4} + f_{\text{OH}_{\text{aromatic}}}]\# \\ &\quad + F_{b+3} + F_{b+4} = -2.15 \\ &= -3.59 + 0.66 + [1.67 + (-0.44)] \\ &\quad + (-0.26) + (-0.19) \\ \text{Log } P_{\text{tyrosine}} (\text{obs}^{15}) &= -2.42 \\ (\text{obs}^{15,34}) &= -2.05\end{aligned}$$

**Tryptophane.**

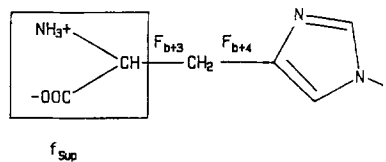
$$\begin{aligned}\text{Log } P_{\text{tryptophane}} &= f_{\text{Sup}} + f_{\text{CH}_2} + f_{\text{C}_8\text{H}_6\text{N}} + F_{b+3} + F_{b+4} = -1.47 \\ &= -3.59 + 0.66 + 1.91 + (-0.26) + (-0.19) \\ \text{Log } P_{\text{tryptophane}} (\text{obs}^{15}) &= -1.75 \\ (\text{obs}^{29}) &= -1.11 \\ (\text{obs}^{15,35}) &= -1.06\end{aligned}$$

Indole = C₈H₆N

See Table IV for another method of calculating log

 P_{indole} Log $P_{\text{indole}} = 2.14$ $f_{\text{indole}} = 2.14 - f_{\text{H}} = 1.91$ **Histidine.**

$$\begin{aligned}\text{Log } P_{\text{histidine}} &= f_{\text{Sup}} + f_{\text{CH}_2} + f_{\text{C}_3\text{H}_3\text{N}_2} + F_{b+3} + F_{b+4} \\ &= -3.69 \\ &= -3.59 + 0.66 + (-0.31) + (-0.26) + (-0.19) \\ \text{Log } P_{\text{histidine}} (\text{obs}^{15} \text{ pH } 7.0) &= -3.56; -2.90 \\ (\text{obs}^{37} \text{ charged}) &= -4.15 \\ (\text{obs}^{37} \text{ uncharged}) &= -2.84\end{aligned}$$



$$\begin{aligned}\text{Log } P_{\text{imidazole}} &= -0.08 \\ f_{\text{imidazole}} &= -0.08 - f_{\text{H}} = -0.31 \\ \text{imidazole} &= \text{C}_3\text{H}_3\text{N}_2\end{aligned}$$

His can be considered as an AA with a polar side chain since one of the ring nitrogens on the imidazole ring is separated by three carbons from the zwitterion and is capable of protonation at that site. The crystal structure of His does show a hydrogen bond between the zwitterion nitrogen and the δ nitrogen of the imidazole ring³⁶ (Fig. 2). Therefore, we have also placed His in the next category (AAs with polar proximity effects), which gives a closer agreement with the observed value. Also, see Table IV for another method of calculating aromatic fragment constants from the sum of the fundamental fragment constants of the atoms.

AA That Have Polar Proximity Effects and Uncharged Side Chain Residues

We will now explore the effects on log P of two aliphatic polar groups in the same molecule. Specifically, the degree that the zwitterion and the side chain heteroatoms interact with each other depends upon the number of alkyl carbons that separate them.

The reduction in hydrophilicity (+ log P) which results from crowding polar fragments together has been discussed in an earlier section. The percentage of this "polarity loss" is expressed by the appropriate coefficient taken from Table I. To be consistent with ref. 12, chapter IV, one starts counting "normal"

$$\begin{aligned}\#f_{\text{C}_6\text{H}_5} &= \log P_{\text{C}_6\text{H}_5} - f_{\text{H}} \\ &= 2.13 - 0.23 = 1.90\end{aligned}$$

See Table IV for another method of calculating aromatic fragment constants from the sum of the fundamental fragment constants of the atoms.

$$\text{**Or } f_{\text{C}_6\text{H}_4\text{OH}} = \log P_{\text{C}_6\text{H}_5\text{OH}} - f_{\text{H}}$$

$$1.46 - 0.23 = 1.23$$

or see Table IV for another method of calculating aromatic fragment constants from atoms. All three methods for estimating the value of the aromatic fragment give the same results.

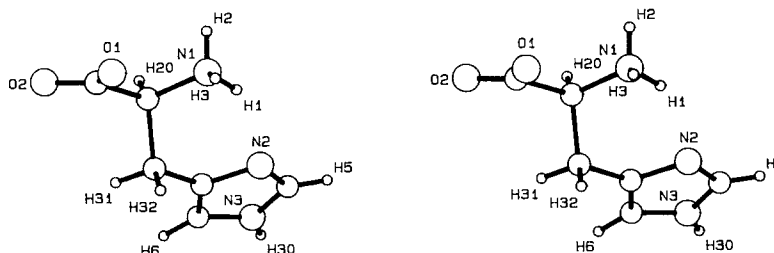
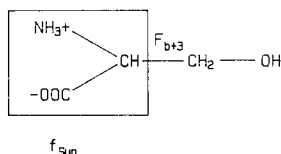


Fig. 2. Stereo diagram of L-histidine,³⁶ which illustrates the hydrogen bond between the proton (H1) of the zwitterion amine with the δ nitrogen (N2) of the imidazole ring. The close contact within the zwitterion of amino proton H3 and the carboxyl oxygen O1 is also easily visualized.

bonds (for $n-1$ term) with the first bond to a polar fragment. For threonine, serine, asparagine, glutamine, and cysteine the net ($n-1$) is zero; for methionine it is one.

Serine.

$$\begin{aligned}\text{Log } P_{\text{serine}} &= f_{\text{Sup}} + f_{\text{CH}_2} + f_{\text{OH}} + F_{b+3} + F_{P2} = -3.17 \\ &= -3.59 + 0.66 + (-1.64) + (-0.26) + (1.66) \\ \text{Log } P_{\text{serine}} (\text{obs}^{15}) &= -3.30\end{aligned}$$



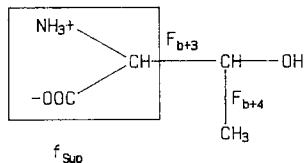
Note that charge is not propagated to the alcohol bond because the OH is referred to as an H-Polar group in QSAR work. See methionine for details.

$$F_{P2} = -0.27(f_{\text{Hyb}} + f_{\text{OH}}) = -0.27(-4.51 + -1.64)$$

f_{OH} is an aliphatic hydroxyl

Threonine.

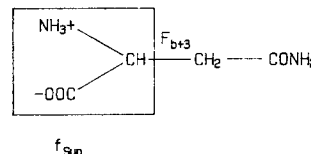
$$\begin{aligned}\text{Log } P_{\text{threonine}} &= f_{\text{Sup}} + f_{\text{CH}} + f_{\text{CH}_3} + f_{\text{OH}} + F_{b+3} + F_{b+4} \\ &+ F_{\text{gBr}} + F_{P2} = -2.92 \\ &= -3.59 + 0.43 + 0.89 + (-1.64) + (-0.26) \\ &+ (-0.19) + (-0.22) + (1.66) \\ \text{Log } P_{\text{threonine}} (\text{obs}^{15}) &= -2.91\end{aligned}$$



$$F_{P2} = -0.27(f_{\text{Hyb}} + f_{\text{OH}}) = -0.27(-4.51 + -1.64)$$

Asparagine.

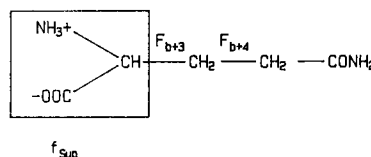
$$\begin{aligned}\text{Log } P_{\text{asparagine}} &= f_{\text{Sup}} + f_{\text{CH}_2} + f_{\text{CONH}_2} + F_{b+3} + F_{P2} = -3.51 \\ &= -3.59 + 0.66 + (-2.11) + (-0.26) + (1.79) \\ \text{Log } P_{\text{asparagine}} (\text{obs}^{15}) &= -3.41\end{aligned}$$



$$F_{P2} = -0.27(f_{\text{Hyb}} + f_{\text{CONH}_2}) = -0.27(-4.51 + -2.11)$$

Glutamine.

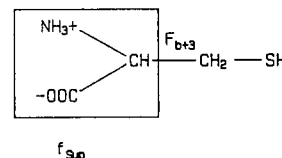
$$\begin{aligned}\text{Log } P_{\text{glutamine}} &= f_{\text{Sup}} + 2f_{\text{CH}_2} + f_{\text{CONH}_2} + F_{b+3} + F_{b+4} \\ &+ F_{P3} = -3.24 \\ &= -3.59 + 2(0.66) + (-2.11) + (-0.26) + (-0.19) \\ &+ (1.59) \\ \text{Log } P_{\text{glutamine}} (\text{obs}^{15}) &= -3.15\end{aligned}$$



$$F_{P3} = -0.24(f_{\text{Hyb}} + f_{\text{CONH}_2}) = -0.24(-4.51 + -2.11)$$

Cysteine.

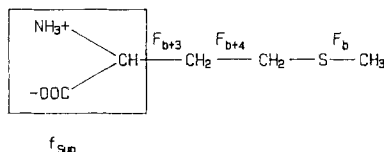
$$\begin{aligned}\text{Log } P_{\text{cysteine}} &= f_{\text{Sup}} + f_{\text{CH}_2} + f_{\text{SH}} + F_{b+3} + F_{P2} = -2.14 \\ &= -3.59 + 0.66 + (-0.23) + (-0.26) + 1.28 \\ \text{Log } P_{\text{cysteine}} (\text{obs}^{15}) &\leq -2.49\end{aligned}$$



$$F_{P2} = -0.27(f_{\text{Hyb}} + f_{\text{SH}}) = -0.27(-4.51 - 0.23)$$

Methionine.

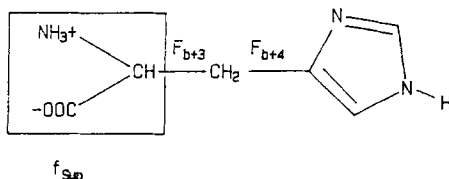
$$\begin{aligned}
 \text{Log } P_{\text{methionine}} &= f_{\text{Sup}} + 2f_{\text{CH}_2} + f_{\text{S}} + f_{\text{CH}_3} + F_{b+3} + F_{b+4} \\
 &+ F_{\text{P}}^* + F_{\text{P}3} = -1.47 \\
 &-3.59 + 2(0.66) - 0.79 + 0.89 - 0.26 - 0.19 \\
 &- 0.12 + 1.27 \\
 \text{Log } P_{\text{methionine}} (\text{obs}^{15}) &= -1.84
 \end{aligned}$$



$$F_{\text{P}3} = -0.24(f_{\text{Hyb}} + f_{\text{S}}) = -0.24(-4.51 - 0.79)$$

Histidine.

$$\begin{aligned}
 \text{Log } P_{\text{histidine}} &(\text{uncharged}) \\
 &= f_{\text{Sup}} + f_{\text{CH}_2} + f_{\text{C}_3\text{H}_3\text{N}_2} + F_{b+3} + F_{b+4} \\
 &+ F_{\text{P}3} = -2.34 \\
 &-3.59 + 0.66 + (-0.31) + (-0.26) + (-0.19) \\
 &+ 1.35 \\
 \text{Log } P_{\text{histidine}} (\text{obs}^{15} \text{ pH } 7.0) &= -3.56; -2.90 \\
 (\text{obs}^{37} \text{ charged}) &= -4.15 \\
 (\text{obs}^{37} \text{ uncharged}) &= -2.84
 \end{aligned}$$



$$F_{\text{P}3} = -0.24(f_{\text{Hyb}} + f_{\text{-N=}}) = -0.24(-4.51 + -1.12)$$

$$\begin{aligned}
 \text{Log } P_{\text{imidazole}} &= -0.08 \\
 f_{\text{imidazole}} &= -0.08 - f_{\text{H}} = -0.31 \\
 \text{imidazole} &= \text{C}_3\text{H}_3\text{N}_2
 \end{aligned}$$

AA That Have Polar Proximity Effects and Charged Side Chain Residues

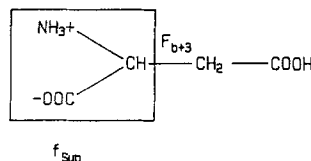
There are obvious difficulties with calculating and measuring partition coefficients of charged species that vary with pH. We are most interested in physiological conditions around pH 7.4 where the α -carbon zwitterion (neutral) has the hybrid nature described above. Also, the measurement of the distribution ratio of extremely hydrophilic solutes is very difficult, and with current techniques it is doubtful that any measurement below $\log -4.5$ can be reliably reproduced. Nonetheless, this group of AAs gives us the opportunity to compare experimental and calculated distributions at different pHs.

The $\log P$ of the amino acids with carboxyl side chains has been calculated in three ways: assuming a net charge of zero for the hybrid fragment and where the side chain carboxyl is protonated (low pH), completely ionized, or equilibrated between the neu-

tral and ionized species at a pH of 5 or 6. The calculated values for the protonated side chain carboxyls agree well with the measured values (± 0.5 log units); and the calculated values for the ionized chain carboxyls also agree well when partitioning of both the neutral and ionized species are accounted for.

Aspartic acid.**a. Side chain carboxyl protonated.**

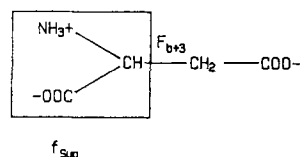
$$\begin{aligned}
 \text{Log } P_{\text{aspartic acid}} &= f_{\text{Sup}} + f_{\text{CH}_2} + f_{\text{COOH}} + F_{b+3} + F_{\text{P}2} = -2.78 \\
 &-3.59 + 0.66 - 1.11 + (-0.26) + 1.52 \\
 \text{Log } P_{\text{aspartic acid}} (\text{obs}^{37}) &= -3.38
 \end{aligned}$$



$$F_{\text{P}2} = -0.27(f_{\text{Hyb}} + f_{\text{COOH}}) = -0.27(-4.51 - 1.11)$$

b. Side chain carboxyl 100% ionized.

$$\begin{aligned}
 \text{Log } P_{\text{aspartic acid}} &= f_{\text{Sup}} + f_{\text{CH}_2} + f_{\text{COO}^-} + F_{b+3} + F_{\text{P}2} = -5.69 \\
 &-3.59 + 0.66 + (-5.10) + (-0.26) + 2.60 \\
 \text{Log } P_{\text{aspartic acid}} (\text{obs}^{15}) &\geq -4.25 \\
 (\text{pH} = 7.0, \text{ according to footnote f in Table II of} \\
 \text{ref. 15}) \text{ Note that } \geq \text{ is a misprint in ref. 15 and} \\
 \text{in all examples here and below, should be } \leq.
 \end{aligned}$$

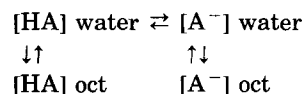


$$F_{\text{P}2} = -0.27(f_{\text{Hyb}} + f_{\text{COO}^-}) = -0.27(-4.51 - 5.10)$$

c. At pH 5-6, $pK_a = 3.86$.

$$\begin{aligned}
 \text{Log } P_{\text{aspartic acid}} &= -3.94 \text{ to } -4.86 \\
 \text{Log } P_{\text{aspartic acid}} (\text{obs}^{15}) &\geq -4.25
 \end{aligned}$$

The results for $\log P$ at various pH values are obtained from the following equilibrium equations, 1-4



$$(1) \text{ pKa} = \text{pH} - \log([\text{A}^-]_{\text{w}}/[\text{HA}]_{\text{w}})$$

$$(2) \log P_{\text{asp}^-} = \log([\text{A}^-]_{\text{o}}/[\text{A}^-]_{\text{w}}) = -5.69$$

*Sulfur and hydroxyl are two of several moieties that are referred to as H-Polar groups in QSAR work (see ref. 12). Any polar fragment in the chain is considered to "stop" charge propagation and the F_b + count stops at the preceding isolating carbon (I.C.). The bond correction for the balance of the chain is calculated with the normal $(n-1)$ factor. This methodology also applies for Ser, Thr, Glu, and Asp.

$$(3) \log P_{\text{asp}} = \log ([\text{HA}]_o / [\text{HA}]_w) = -2.78$$

$$(4) \log P_{\text{asp cal}} = \log \frac{([\text{A}^-]_o + [\text{HA}]_o)}{([\text{A}^-]_w + [\text{HA}]_w)}$$

if $[\text{A}^-]_w = 1$; $[\text{A}^-]_o$, $[\text{HA}]_w$, and $[\text{HA}]_o$ can be obtained from eqs. 1-3.

pH	$\log P_{\text{asp cal}}$
10	-5.69
7	-5.49
6	-4.86
5.33*	-4.25
5	-3.94
4	-3.16
2	-2.79

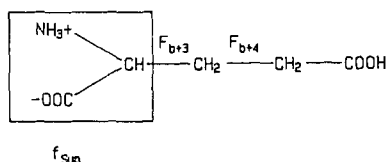
*At this pH, $\log P_{\text{cal}} = \log P_{\text{obs}}$.

Glutamic Acid.

a. Side chain carboxyl protonated.

$$\begin{aligned} \log P_{\text{glutamic acid}} &= f_{\text{Sup}} + 2f_{\text{CH}_2} + f_{\text{COOH}} + F_{b+3} + F_{b+4} \\ &+ F_{P3} = -2.48 \\ &-3.59 + 2(0.66) + (-1.11) + (-0.26) + (-0.19) \\ &+ 1.35 \end{aligned}$$

$$\log P_{\text{glutamic acid}} (\text{obs}^{37}) = -2.94$$



$$F_{P3} = -0.24(f_{\text{Hyb}} + f_{\text{COOH}}) = -0.24(-4.51 - 1.11)$$

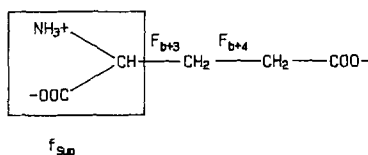
b. Side chain carboxyl 100% ionized.

$$\begin{aligned} \log P_{\text{glutamic acid}} &= f_{\text{Sup}} + 2f_{\text{CH}_2} + f_{\text{COO}^-} + F_{b+3} + F_{b+4} \\ &+ F_{P3} = -5.51 \\ &-3.59 + 2(0.66) + (-5.10) + (-0.26) + (-0.19) \\ &+ 2.31 \end{aligned}$$

$$\log P_{\text{glutamic acid}} (\text{obs}^{15}) \geq -4.19$$

(pH 7.0, according to footnote f in Table II of ref. 15)

Note: \geq is a misprint in ref. 15 and should be reversed, \leq .



$$F_{P3} = -0.24(f_{\text{Hyb}} + f_{\text{COO}^-}) = -0.24(-4.51 - 5.10)$$

c. At pH 2-10, $pK_a = 4.24$. See Asp for details of the calculations using eqs. 1-4.

pH	$\log P_{\text{glu cal}}$
10	-5.51
7	-5.05

6	-4.22
5.96**	-4.19
5	-3.31
4	-2.68
2	-2.48

**At this pH, $\log P_{\text{cal}} = \log P_{\text{obs}}$.

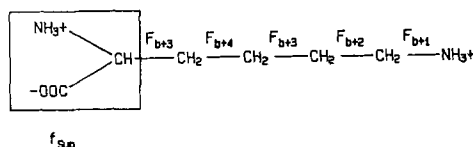
For the amino acids with basic side chains, Lys and Arg, in the pH range 4-8, the side chain amino or guanidine group would be protonated. The same difficulties mentioned above for accurately measuring the $\log P$ of the acidic AAs also exist here. There is, however, a very good agreement with the calculated and observed values for Lys. The deviation in $\log P$ for charged Arg is -1.47.

Lysine.

$$\begin{aligned} \log P_{\text{lysine}} &= f_{\text{Sup}} + 4f_{\text{CH}_2} + f_{\text{NH}_3} + F_{b+1} + F_{b+2} \\ &+ 2F_{b+3} + F_{b+4} + F_{P+5} = -4.66 \\ &-3.59 + 4(0.66) + (-3.40) + (-0.78) + (-0.40) \\ &+ 2(-0.26) + (-0.19) + 1.58 \end{aligned}$$

$$\log P_{\text{lysine}} (\text{obs}^{15}) \geq -4.44$$

Note: \geq is a misprint in ref. 15 and should be reversed, \leq .

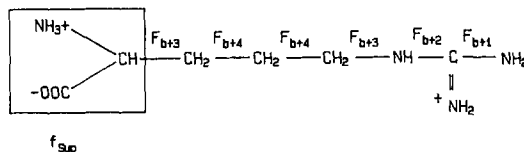


$$F_{P5} = -0.20(f_{\text{Hyb}} + f_{\text{NH}_3^+}) = -0.20(-4.51 - 3.40)$$

Arginine.

$$\begin{aligned} \log P_{\text{arginine}} &= f_{\text{Sup}} + 3f_{\text{CH}_2} + f_{\text{CN}_3\text{H}_5} + 2F_{b+3} \\ &+ 2F_{b+4} + F_{P+4} = -5.67 \\ &-3.59 + 3(0.66) + (-5.32) + 2(-0.26) \\ &+ 2(-0.19) + 2.16 \end{aligned}$$

$$\begin{aligned} \log P_{\text{arginine}} (\text{obs}^{15}) &= -4.20 \\ (\text{obs}^{29}) &= -4.08 \end{aligned}$$



$$\begin{aligned} F_{P4} &= -0.22(f_{\text{Hyb}} + f_{\text{CN}_3\text{H}_5}^\dagger) \\ &= -0.22(-4.51 - 5.32) \end{aligned}$$

The crystal structure of Arg reveals the bridging of a carboxyl oxygen with a guanidine nitrogen by two water molecules^{38,39} (Fig. 3). The bridging waters, if tightly bound during transfer experiments, add two

[†]The new fragment value for the guanide residue is -5.32 rather than that reported in ref. 12. The new value takes into account the delocalization of the protonated nitrogen, i.e., f_{b+1} and F_{b+2} .

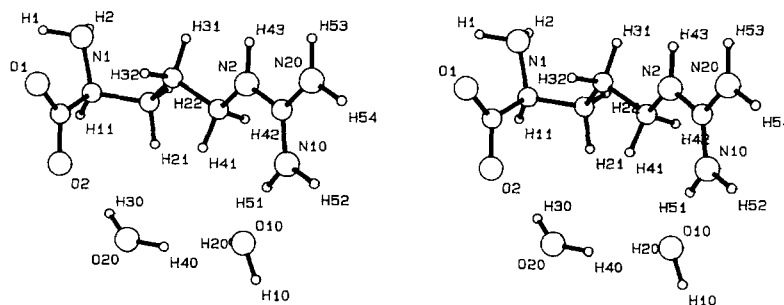
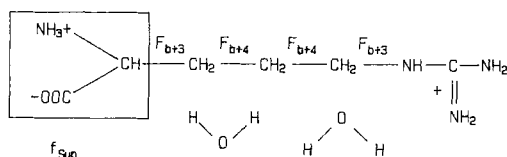


Fig. 3. Stereo diagram of arginine dihydrate, a non- α -carbon zwitterion. The carboxyl oxygen O2 is hydrogen bonded to proton H30 of water O20. The proton H51 of the guanidinium nitrogen N10 is hydrogen bonded to water O10. The water-to-water hydrogen bond bridge is between O20H40 and O10.

intramolecular hydrogen bonds to the AA. Such intramolecular hydrogen bonds add about 0.63 log units/hydrogen bond, increasing the hydrophobicity (higher log P).¹² This could explain the observed increase in hydrophobicity for Arg since the log P calculated, including the hydrogen bond factor, $\log P = -5.67 + 2(0.63) = -4.41$, is closer to the observed log P (-4.08).



The distance between the guanidine nitrogen and carboxylate that is bridged by two waters is 6.2 Å. A distinct energy minimum around 6.0 Å has been observed in the solution of methane molecules in water.⁴⁰ Perhaps this distance is ideal for interaction of two molecules of water in solvation processes. In the Arg dihydrate crystal structure, the waters appear to mediate the charges between the protonated amino cation and the carboxylate anion. A search of the literature demonstrates several hydrated AAs in which water bridges cation and anion: Lys,⁴¹ Arg-Glu complex,⁴² Asn,³⁰ Ser,⁴³ and TyrGlyGly.⁴⁴

To gain further evidence for water-mediated charge stabilization, bound waters in hemoglobin were studied. Several cation-anion interactions are mediated by a water molecule [Lys 61 $\alpha 2$ - $\alpha 2$ Heme COO^- ; Asp 6 $\alpha 1$ -Lys127 $\alpha 1$; Lys 16 $\alpha 1$ -Glu 116 $\alpha 1$; and Glu 27 $\alpha 2$ -Arg 31 $\alpha 2$]. Considering the above, it seems reasonable that arginine could transfer to octanol as the dihydrate.

Summary of AA Partition Coefficient Calculations

The calculated and observed values for 19 AAs agree quite well when pH effects and hydration phenomena are considered. Only Pro calculates too hydrophilic. This difference probably arises from the definition of the hybrid fragment for AAs which is not structurally

consistent with the zwitterion of Pro that contains a secondary cyclic amine. The results from the calculations above are summarized in Table II, column a and Table III, column a.

AA Side Chain Hydrophobicities

Most studies are aimed at scaling the hydrophobicities of the AA residue side chains in polypeptides and not the zwitterionic AA as we have calculated here. Since protein backbones except for Pro and the first and last residues are structurally constant, (NH-CH-CO), the side chain differences are the most interesting facets in investigating protein folding and protein substrate interactions. Because the polar side chains in a peptide are no longer influenced by the zwitterion charges on the α -carbons of AAs but by the peptide linkages, the relative hydrophobicities of the side chains may fall into a different order than that found for the AA side chain values (Table III, columns a and b).

To estimate the relative hydrophobicities of the AA residue side chains in peptides and proteins, Fauchere and Pliska¹⁵ made N-acetyl amides of all the natural amino acids and measured their octanol/water partition coefficients. Using π as defined above they subtracted the carefully measured value for the acetyl amide derivative of glycine from each of the other AAs. These π values appear to us to be the most dependable measured values currently available for estimating nascent side chain hydrophobicity.

We have calculated the side chain hydrophobicities (π) using standard fragment, bond, and polar proximity values (Table I). The calculated fragment values are converted to π values by subtraction of H (0.23) from f fragment and compared to the π values obtained from the observed data.

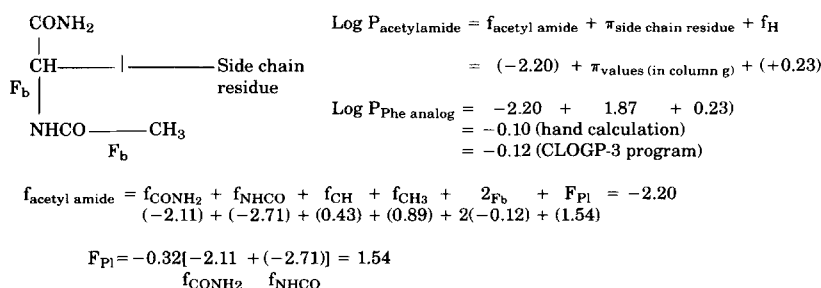
We have also calculated (using a computer program,²⁸ CLOGP-3.33) the log P of the acetyl amide analogs; the results are listed in Table II column d. For details of this calculation see the legend of Table II.

TABLE II. Calculated: Log P for AA; Log P Acetylamide Analogs; π of Side Chain From Experimental Measurements; Side Chain Fragments and Side Chain Fragments With Full Polar Proximity Participation of the Peptide Backbone

AA	Amino acids		Acetylamide analog			π of side chain residues CLOGP methodology			π of side chain residues with F_{FP}	
	Log P cal (a)	Log P obs (b)	Meas. ref. 15 (c)	CLOGP* (d)	Dev. ref. 15 (e)	Ref. 15 π (f)	CLOGP Cal. π (g)	Dev. ref. 15 (h)	CLOGP Cal† $\pi + F_{FP}$ (i)	Dev. Ref. 15 (j)
1. Phe	-1.48	-1.63	-0.04	-0.12	-0.08	1.79	1.87	+0.08	1.87	+0.08
2. Ile	-1.61	-1.72	-0.03	-0.23	-0.20	1.80	1.81	+0.01	1.81	+0.01
3. Leu	-1.54	-1.61	-0.13	-0.23	-0.10	1.70	1.81	+0.11	1.81	+0.11
4. Val	-2.15	-2.08	-0.61	-0.76	-0.15	1.22	1.27	+0.05	1.27	+0.05
5. Tyr	-2.15	-2.42, -2.05	-0.87	-0.78	+0.09	0.96	1.20	+0.24	1.20	+0.24
6. Ala	-2.96	-2.89	-1.52	-1.68	-0.16	0.31	0.32	+0.01	0.32	+0.01
7. Gly	-3.14	-3.25	-1.83	-1.99	-0.17	0.00	0.00	0.00	0.00	0.00
8. Pro	-3.42	-2.50	-1.34	-1.03	+0.31	0.72	0.95	+0.23	0.95	+0.23
9. Trp	-1.47	-1.75, -1.11	0.42	-0.12	-0.54	2.25	1.88	-0.37	1.88	-0.37
10. Ser	-3.17	-3.30	-1.87	-2.62	-0.75	-0.04	-0.62	-0.58	0.01	+0.05
11. Thr	-2.92	-2.91	-1.57	-2.31	-0.74	0.26	-0.30	-0.56	0.33	+0.07
12. Met	-1.47	-1.84	-0.60	-1.21‡	-0.61	1.23	0.81	-0.42	1.05	-0.18
13. His	-3.69	-3.56, -2.90	-1.70	-2.31	-0.61	0.13	-0.34	-0.47	0.34	+0.47
His**	-2.34	-2.84	-1.70	-1.96‡	-0.26	0.13	0.01	-0.12	0.25	+0.12
14. Cys	-2.14	-2.49	-0.29	-1.57	-1.28	1.54	0.43	-1.11	1.05	-0.49
15. Gln	-3.24	-3.15	-2.05	-3.16‡	-1.11	-0.22	-1.15	-0.93	-0.91	-0.69
Gln+H ₂ O						-0.22	-0.52	-0.30	-0.28	-0.06
16. Asn	-3.51	-3.41	-2.41	-2.97	-0.56	-0.60	-0.97	-0.37	-0.34	+0.26
17. Glu COOH	-2.48	-2.94	-2.47	-2.26‡	+0.14	-0.64	-0.25	+0.39	-0.01	+0.63
Glu COO-	-5.51	-4.19	-2.47		-3.34	-0.64	-3.84	-3.20	-3.60	-2.96
18. Asp COOH	-2.78	-3.38	-2.60		+0.37	-0.77	-0.23	+0.54	0.40	+1.17
Asp COO-	-5.69	-4.25	-2.60		-2.55	-0.77	-3.18	-2.41	-2.55	-1.78
19. Lys NH ₂							0.05			
Lys NH ₃ ⁺	-4.66	-4.44	-2.82	-3.77‡	-0.95	-0.99	-1.80	-0.81	-1.32	-0.33
20. Arg	-5.67	-4.20, -4.08	-2.84	-5.01‡	-2.17	-1.01	-3.04	-2.03	-2.51	-1.50
Arg+2H ₂ O	-4.41	-4.20, -4.08	-2.84	-3.75‡	-0.91	-1.01	-1.78	-0.77	-1.25	-0.24

*These values are obtained using the computer program CLOGP-3. Subtraction of the CLOGP values for the acetylamide of glycine from the other acetylamide analogs in column d will give the value of π for the side chain residues listed under CLOGP methodology column g. Any small differences observed between the values obtained by [$f_{AA\text{ analog}} - f_{Gly\text{ analog}}$] and those listed in column g (that come from the text) arise from the rounding off of numbers used in the hand calculations. For example, for f_H , the computer program CLOGP-3 uses 0.227 and Table I value which we use throughout is 0.23.

To calculate the CLOGP results for the acetylammides listed in Table II, column d, by hand, simply add the fragments and factors of the amide backbone (-2.20) to the π of the side chain residue (column g) and the f for H (+0.23 originally subtracted to produce the π values). The following example should clarify the process.



†For peptides with polar side chains, CLOGP calculates the proximity effect between the polar side chain and the peptide backbone by averaging the amide fragment values on either side of the α -carbon. Better agreement is obtained if the amide fragment values are added, as seen in the last two columns where "FP" stands for "full proximity."

‡Calculated by hand since the present version of CLOGP does not calculate $F_{Pn} > 2$. The f for the imidazole ring contains an electronic contribution from the unprotonated nitrogen; see ref. 49.

**His is calculated as the imidazole group being polar.

TABLE III. Hydrophobicity Scales for Amino Acid Side Chains From Various Studies*

ΔG° cal AA zwitterion	ΔG° obs AA (b)	ΔG° cal AA analog side chain (c)	ΔG° obs AA analog side chain (d)	ΔG° cal AA analog F _{FP} (e)	ΔG_R (f)	ΔG° cal AA side chain (g)	Hydro- phobic value (h)	Contact energy $-0.6^*q_i^*e_i/2$ side chain (i)	$\langle A^{\circ} - \langle A \rangle / A^{\circ}$ (j)
TRP	-2.00(1)	-2.38(4)	2.56(1)	3.06(1)	2.6(1)	3.15(1)	3.4(1)	3.18(6)	0.85(5)
PHE	-2.01(3)	-2.22(2)	2.54(2)	2.43(3)	2.3(3)	3.05(2)	2.5(2)	4.55(1)	0.88(2)
LEU	-2.09 (4)	-2.19(1)	2.46(3)	2.31(4)	1.9(4)	2.71(3)	1.8(4)	3.99(4)	0.85(6)
ILE	-2.19(5)	-2.34(3)	2.46(4)	2.45(2)	1.9(5)	2.71(4)		4.35(3)	0.88(3)
VAL	-2.92(7)	-2.83(5)	1.73(5)	1.66(7)	1.5(7)	1.99(6)	1.5(5)	3.15(7)	0.86(4)
TYR	-2.92(8)	-3.29(6)	1.63(6)	1.31(8)	1.6(6)	2.32(5)	2.3(3)	1.41(8)	0.76(9)
PRO	-4.65(14)	-3.40(8)	1.29(7)	0.98(9)	1.2(8)	1.37(8)		-0.61(15)	0.64(14)
MET	-2.00(2)	-3.66(9)	1.10(8)	1.67(6)	2.4(2)	1.47(7)	1.3(6)	4.47(2)	0.85(7)
CYS	-2.91(6)	-3.38(7)	0.58(9)	2.09(5)	0.38(12)	1.27(9)		3.45(5)	0.91(1)
ALA	-4.02(10)	-3.93(10)	0.44(10)	0.42(10)	0.67(9)	0.72(10)	0.5(8)	0.85(9)	0.74(10)
GLY	-4.27(11)	-4.42(13)	0.00(11)	0.00(13)	0.00(14)	0.00(11)		0.00(12)	0.72(11)
THR	-3.97(9)	-3.95(11)	-0.41(13)	0.35(11)	0.52(11)	-0.36(13)	0.4(9)	0.01(11)	0.70(12)
HIS	-5.01(16)	-4.84(15)	-0.01(12) [†]	0.18(12)	0.34(12) [‡]	-0.31(12)	0.5(7)	0.62(10)	0.78(8)
SER	-4.31(12)	-4.48(14)	-0.84(15)	-0.05(14)	0.01(13)	0.01(13)	0.3(10)	-0.39(14)	0.66(13)
ASN	-4.77(15)	-4.63(16)	-1.32(16)	-0.82(16)	-0.46(16)	-0.60(17)		-0.77(17)	0.63(16)
GLN	-4.40(13)	-4.28(12)	-0.71(14) [‡]	-0.30(15)	-0.38(15) [‡]	-0.22(15)	-1.43(15)	-0.61(16)	0.62(17)
LYS	-6.33(18)	-6.03(20)	-2.45(18)	-1.35(19)	-1.79(18)	-0.57(16)		-1.53(20)	0.52(20)
GLU	-3.37**	-4.00	-0.34**	-0.01**	-0.10**				
COO ⁻	-7.49(19)	-5.69(17)	-5.22(20)	-0.87(17)	-0.76(18)			-0.83(18)	0.62(18)
ARG	-5.99(17) [†]	-5.70(18)	-2.42(17)	-1.37(20)	-2.10(20)			-0.30(13)	0.64(15)
ASP	-3.78**	-4.59	-0.31**	0.54**	-0.03**				
COO ⁻	-7.73(20)	-5.78(19)	-4.32(19)	-1.05(18)	-1.20(19)			-0.89(19)	0.62(19)

*The order of rank (highest to lowest hydrophobicity) of the AA side chains appears in () after the value reported. (a) values from the text. (b) values from ref. 15. (c) ΔG° calculated from 2.303RT π at 298°K (peptide side chain) from CLOGP π values in Table II, column g; see text for details. (d) ΔG° calculated from π values of Fauchere and Pliska.¹⁵ (e) These values are calculated from the π values with F_{FP} listed in Table II, column i, which represent a better agreement with the observed side chain π values in ref. 15. (f) Hydrophobic energies of AA side chains by Eisenberg and McLachlin.⁸ (g) Fragment constants f for AA side chains calculated by Rekker.¹³ (h) Nozaki and Tanford hydrophobicity scales for transfer of AA side chains from 100% organic solvent to water.⁴ (i) These values from Miyazawa and Jernigan are $(-0.6^*q_i^*e_i/2) - (-0.6^*q_{gly}^*e_{gly}/2)$; where e_i is an average contact area of the type-i residue and q_i is the average number of contacts for a residue of type-i, (see ref. 50). (j) The mean fractional area loss, denoted f , where $f = 1 - (\langle A \rangle / A^{\circ})$ in ref. 1, p. 835. Note that $\Delta G^{\circ} = 2.303RT \log P [2.303 \times 0.00198 \times 298] = 1.35886 \log P$; $\Delta G^{\circ} = RT \ln P$ for columns a and b.

[†]His is calculated with imidazole as a polar side chain with an F_{P3} factor.

[‡]These values are calculated from the log P values that include one water bound to Gln.

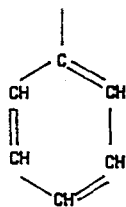
**These values for ASP and Glu are derived from the log P or fragment values with the side chain carboxyl protonated (COOH).

^{††}The Arg values are for the dihydrate.

TABLE IV. Fragment Constants and Free Energy Values for Individual Aromatic Atoms

	atom type	f	$\Delta G^{\circ'}$	
aromatic,	CH	0.355	0.48	
aromatic,	C	0.13	0.18	
aromatic,	OH	-0.44	-0.60	
aromatic,	C'	0.225	0.31	ring fusion carbon; not hetero
aromatic,	C*	0.44	0.60	ring fusion carbon; is hetero
aromatic,	-NH-	-0.67	-0.91	fused in ring
aromatic,	-N=	-1.12	-1.52	fused in ring

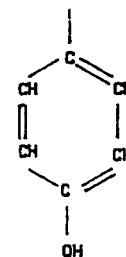
Also needed for His is a sigma rho correction for the interaction of the two nitrogens. The electronic sigma rho factor is $F_{\text{elect}} = 0.67$ for ring interactions.



PHE ring

$$f_{\text{C}_6\text{H}_5} = 5f_{\text{CH}} + f_{\text{C}} = 1.91$$

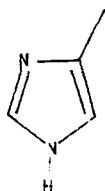
$$\Delta G^{\circ'} = 2.60$$



TYR ring

$$f_{\text{C}_6\text{H}_5\text{O}} = 4f_{\text{CH}} + 2f_{\text{C}} + f_{\text{OH}} = 1.24$$

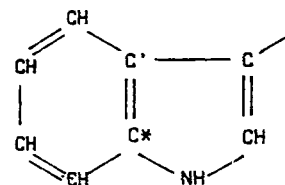
$$\Delta G^{\circ'} = 1.69$$



HIS ring

$$f_{\text{C}_3\text{H}_3\text{N}_2} = f_{\text{N}} + f_{\text{NH}} + 2f_{\text{CH}} + f_{\text{C}} + F_{\text{elect}} = -0.28$$

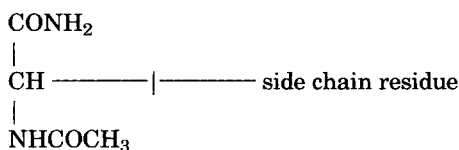
$$\Delta G^{\circ'} = -0.38$$



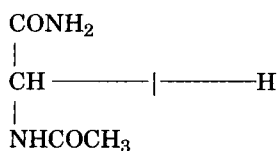
TRP ring

$$f_{\text{C}_8\text{H}_6\text{N}} = 5f_{\text{CH}} + f_{\text{C}^*} + f_{\text{C}'} + f_{\text{C}} + f_{\text{NH}} = 1.90$$

$$\Delta G^{\circ'} = 2.58$$

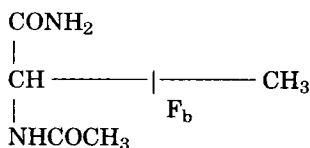
Acetyl Amide π Calculations for Side Chains Having No Polar Interaction With α -Diamide


1. Gly $f_H - f_H$
0.23 - 0.23



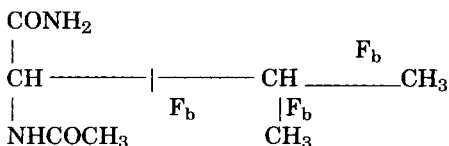
Cal π 0.00 OBS¹⁵ 0.00 DIFF 0.0

2. Ala $f_{\text{CH}_3} + F_b + F_{\text{gBr}} - f_H$
0.89 + (-0.12) + (-0.22) - 0.23



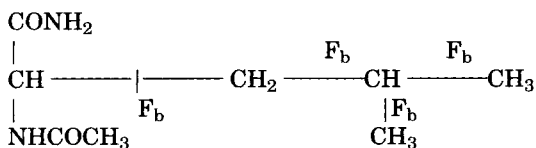
Cal π 0.32 OBS¹⁵ 0.31 DIFF +0.01

3. Val $f_{\text{CH}} + 2f_{\text{CH}_3} + F_{\text{cBr}} + 3F_b + F_{\text{gBr}} - f_H$
0.43 + 2(0.89) + -0.13 + 3(-0.12) + (-0.22) - 0.23



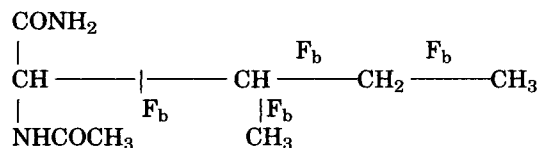
Cal π 1.27 OBS¹⁵ 1.22 DIFF +0.05

4. Leu $f_{\text{CH}_2} + f_{\text{CH}} + 2f_{\text{CH}_3} + F_{\text{cBr}} + 4F_b + F_{\text{gBr}} - f_H$
0.66 + 0.43 + 2(0.89) + -0.13 + 4(-0.12) + (-0.22) - 0.23



Cal π 1.81 OBS¹⁵ 1.70 DIFF +0.11

5. Ile $f_{\text{CH}_2} + f_{\text{CH}} + 2f_{\text{CH}_3} + F_{\text{cBr}} + 4F_b + F_{\text{gBr}} - f_H$
0.66 + 0.43 + 2(0.89) + -0.13 + 4(-0.12) + (-0.22) - 0.23

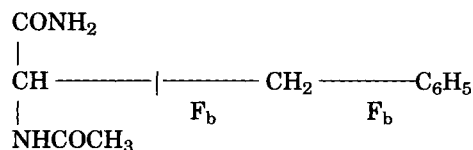


Cal π 1.81 OBS¹⁵ 1.80 DIFF +0.01

6. Pro $3f_{\text{CH}_2} + 4F_b \text{ ring} + 2F_{\text{gBr}} - f_H$
3(0.66) + 4(-0.09) + 2(-0.22) - 0.23

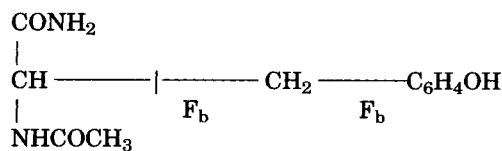
Cal π 0.95 OBS¹⁵ 0.72 DIFF 0.23

7. Phe $f_{\text{CH}_2} + f_{\text{C}_6\text{H}_5} + 2F_b + F_{\text{gBr}} - f_H$
0.66 + 1.90 + 2(-0.12) + (-0.22) - 0.23



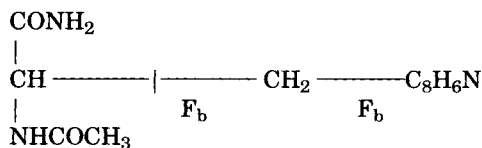
Cal π 1.87 OBS¹⁵ 1.79 DIFF +0.08

8. Tyr $f_{\text{CH}_2} + f_{\text{C}_6\text{H}_4\text{OH}} + 2F_b + F_{\text{gBr}} - f_H$
0.66 + 1.23 + 2(-0.12) + (-0.22) - 0.23



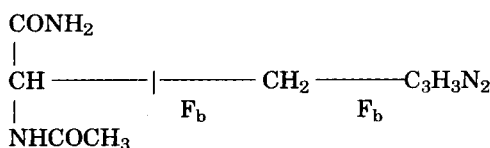
Cal π 1.20 OBS¹⁵ 0.96 DIFF 0.24

9. Trp $f_{\text{CH}_2} + f_{\text{C}_8\text{H}_6\text{N}} + 2F_b + F_{\text{gBr}} - f_H$
0.66 + 1.91 + 2(-0.12) + (-0.22) - 0.23



Cal π 1.88 OBS¹⁵ 2.25 DIFF -0.37

10. His $f_{\text{CH}_2} + f_{\text{C}_3\text{H}_3\text{N}_2} + 2F_b + F_{\text{gBr}} - f_H$
0.66 + -0.31 + 2(-0.12) + (-0.22) - 0.23



Cal π -0.34 OBS¹⁵ 0.13 DIFF -0.47

Here His is *not* considered as a polar side chain. As described for His AA above, His forms an internal

H-bond in the crystal and since one ring nitrogen is separated by three carbons from the polar backbone, the calculation of His might best be considered as a polar side chain.

π Calculations for Acetyl Amide AA Analogs With Polar Proximity Effects

There is a fairly consistent deviation of -0.6 to -0.8 log units between calculated and observed log P values for the uncharged side chain residue acetyl amide analogs Ser, Thr, Met, and Asn. The Gln and Cys acetyl amide analogs deviate by approximately twice this amount (-1.11 and -1.29). The α -diamide polar proximity effect seems adequately accounted for in the ten previous examples and is further confirmed by hundreds of examples of other polar groups separated by a single isolating carbon.^{12,35} The polar interactions of the side chain fragment with the α -diamide moiety seem to increase hydrophobicity to a greater degree than has been allowed for previously.

This deviation between the observed and calculated values prompted us to look for other causes for this discrepancy. Three possibilities are

1. that the terminal ends of the blocked amino acids (CONH_2) or acetamide (CH_3CONH) could assume a folded conformation to interact with the polar side chains and/or

2. the polar side chain residues could interact and tightly bind water molecules as depicted for Arg and/or

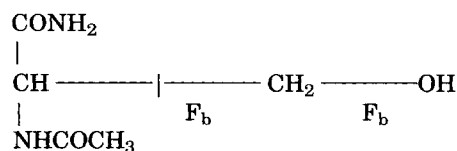
3. the polar proximity correction should *not average* but *sum* the fragment constants for both amide groups attached to the alpha carbon, i.e., the sum of $f_{\text{NH}_2\text{CO}}$ and f_{NHCO} .

The following calculations show the results for both the CLOGP methodology and the same calculation utilizing full participation of both amide groups (Full) in the polar proximity factor.

1. Ser $f_{\text{CH}_2} + f_{\text{OH}} + 2F_b + F_{\text{P2}} + F_{\text{gBr}} - f_{\text{H}}$

$$\text{CLOGP} = 0.66 + (-1.64) + 2(-0.12) + 0.26(2.41 + 1.64) + (-0.22) - 0.23$$

$$F_{\text{FP2}} = 0.26(4.82 + 1.64)^*$$



The value of 2.41 is the average of f_{NHCO} and f_{CONH_2} . For the normal CLOGP calculation as shown above, the polar proximity factor is the average of the peptide linkage on each side of the alpha carbon. This is designated as F_{Pn} where n is the number of carbons between the polar side chain group and the polar peptide. For serine $n = 2$.

$$\begin{aligned} F_{\text{P2}} &= -0.26 [1/2(f_{\text{NHCO}} + f_{\text{CONH}_2}) + f_{\text{OH}}] \\ &= -0.26 [1/2(2.71 + 2.11) + 1.64] = 1.05 \end{aligned}$$

for the full effect of the backbone amide linkages

$F_{\text{FP}} = \text{full amide participation}$

$$\begin{aligned} F_{\text{FP2}} &= -0.26 [f_{\text{NHCO}} + f_{\text{CONH}_2} + f_{\text{OH}}] \\ &= -0.26 [2.71 + 2.11 + 1.64] = 1.68 \end{aligned}$$

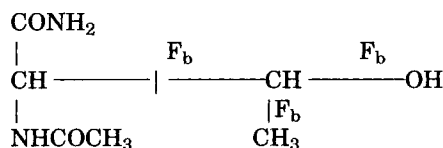
Cal π with F_{P} -0.62 OBS¹⁵ -0.04 DIFF -0.58

Cal π with F_{FP} 0.01 -0.04 $+0.05$

2. Thr $f_{\text{CH}} + f_{\text{CH}_2} + f_{\text{OH}} + F_{\text{gBr}} + 3F_b$
+ $F_{\text{P2}} + F_{\text{gBr}} - f_{\text{H}}$

$$\begin{aligned} \text{CLOGP} &= 0.43 + 0.89 + (-1.64) + (-0.22) + \\ &\quad 3(-0.12) + 0.26(2.41 + 1.64) + (-0.22) - \\ &\quad 0.23 \end{aligned}$$

$$F_{\text{FP2}} = 0.26(4.82 + 1.64)$$



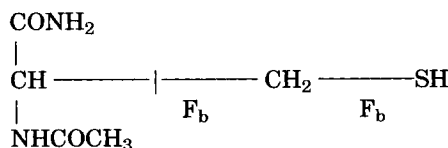
Cal π with F_{P} -0.30 OBS¹⁵ 0.26 DIFF -0.56

Cal π with F_{FP} 0.33 0.26 $+0.07$

3. Cys $f_{\text{CH}_2} + f_{\text{SH}} + 2F_b + F_{\text{P2}} + F_{\text{gBr}} - f_{\text{H}}$

$$\begin{aligned} \text{CLOGP} &= 0.66 + (-0.23) + 2(-0.12) + \\ &\quad 0.26(2.41 + 0.23) + (-0.22) - 0.23 \end{aligned}$$

$$F_{\text{FP2}} = 0.26(4.81 + 0.23)$$



Cal π with F_{P} 0.43 OBS¹⁵ 1.54 DIFF -1.11

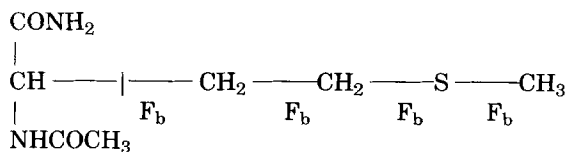
Cal π with F_{FP} 1.05 1.54 -0.49

4. Met $2f_{\text{CH}_2} + f_{\text{CH}_3} + f_{\text{S}} + 4F_b + F_{\text{P3}} + F_{\text{gBr}} - f_{\text{H}}$

$$\begin{aligned} \text{CLOGP} &= 2(0.66) + (0.89) + (-0.79) + 4(-0.12) + \\ &\quad 0.1(2.41 + 0.79) + (-0.22) - 0.23 \end{aligned}$$

$$F_{\text{FP3}} = 0.1(4.82 + 0.79)$$

*Note that for simplicity in all examples in this section, the minus signs have been eliminated mathematically and F_{FP2} can be calculated as shown.

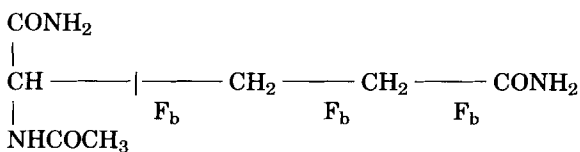


$$\begin{array}{llll}
 \text{Cal } \pi \text{ with } F_p & 0.81 & \text{OBS}^{15} & 1.23 \quad \text{DIFF} \quad -0.42 \\
 \text{Cal } \pi \text{ with } F_{FP} & 1.05 & & 1.23 \quad -0.18
 \end{array}$$

$$5. \text{Gln } 2f_{\text{CH}_2} + f_{\text{CONH}_2} + 3F_b + F_{P3} + F_{gBr} - f_H$$

$$\text{CLOGP} = 2(0.66) + (-2.11) + 3(-0.12) + 0.1(2.41 + 2.11) + (-0.22) - 0.23$$

$$F_{FP3} = 0.1(4.82 + 2.11)$$



$$\begin{array}{llll}
 \text{Cal } \pi \text{ with } F_p & -1.15 & \text{OBS}^{15} & -0.22 \quad \text{DIFF} \quad -0.93 \\
 \text{Cal } \pi \text{ with } F_{FP} & -0.91 & & -0.22 \quad -0.69
 \end{array}$$

With one water add +0.63 to each cal. value above.

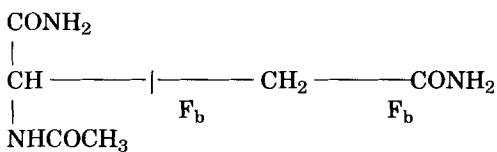
$$\text{Cal } \pi \text{ with } F_p \quad -0.52 \quad \text{OBS} \quad -0.22 \quad \text{DIFF} \quad -0.30$$

$$\text{Cal } \pi \text{ with } F_{FP} \quad -0.28 \quad -0.22 \quad -0.06$$

$$6. \text{Asn } f_{\text{CH}_2} + f_{\text{CONH}_2} + 2F_b + F_{P2} + F_{gBr} - f_H$$

$$\text{CLOGP} = 0.66 + -2.11 + 2(-0.12) + 0.26(2.41 + 2.11) + (-0.22) - 0.23$$

$$F_{FP2} = 0.26(4.82 + 2.11)$$



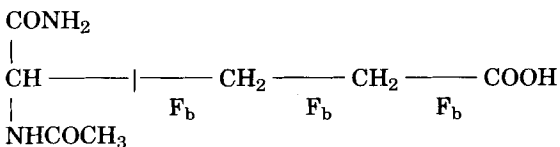
$$\begin{array}{llll}
 \text{Cal } \pi \text{ with } F_p & -0.97 & \text{OBS}^{15} & -0.60 \quad \text{DIFF} \quad -0.37 \\
 \text{Cal } \pi \text{ with } F_{FP} & -0.34 & & -0.60 \quad +0.26
 \end{array}$$

$$7. \text{Glu } 2f_{\text{CH}_2} + f_{\text{COOH}} + 3F_b + F_{P3} + F_{gBr} - f_H$$

$$\text{CLOGP} = 2(0.66) + (-1.11) + 3(-0.12) + 0.1(2.41 + 1.11) + (-0.22) - 0.23$$

$$F_{FP3} = 0.1(4.82 + 1.11)$$

Protonated Acid



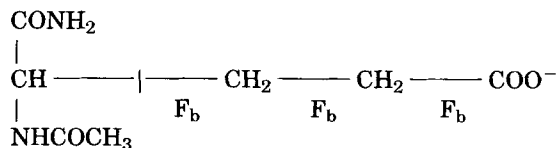
$$\begin{array}{llll}
 \text{Cal } \pi \text{ with } F_p & -0.25 & \text{OBS}^{15} & -0.64 \quad \text{DIFF} \quad +0.39 \\
 \text{Cal } \pi \text{ with } F_{FP} & -0.01 & & -0.64 \quad +0.63
 \end{array}$$

$$\text{Glu } 2f_{\text{CH}_2} + f_{\text{COO}^-} + 3F_b + F_{P3} + F_{gBr} - f_H$$

$$\text{CLOGP} = 2(0.66) + (-5.10) + 3(-0.12) + 0.1(2.41 + 5.10) + (-0.22) - 0.23$$

$$F_{FP3} = 0.1(4.82 + 5.10)$$

Ionized Acid



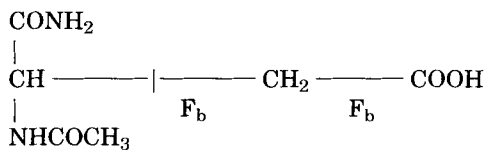
$$\begin{array}{llll}
 \text{Cal } \pi \text{ with } F_p & -3.84 & \text{OBS}^{15} & -0.64 \quad \text{DIFF} \quad -3.20 \\
 \text{Cal } \pi \text{ with } F_{FP} & -3.60 & & -0.64 \quad -2.96
 \end{array}$$

$$8. \text{Asp } f_{\text{CH}_2} + f_{\text{COOH}} + 2F_b + F_{P2} + F_{gBr} - f_H$$

$$\text{CLOGP} = (0.66) + (-1.11) + 2(-0.12) + 0.26(2.41 + 1.11) + (-0.22) - 0.23$$

$$F_{FP2} = 0.26(4.82 + 1.11)$$

Protonated Acid



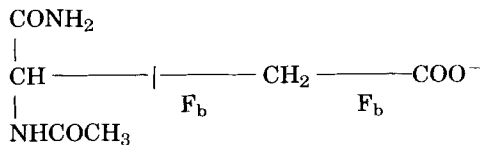
$$\begin{array}{llll}
 \text{Cal } \pi \text{ with } F_p & -0.23 & \text{OBS}^{15} & -0.77 \quad \text{DIFF} \quad +0.54 \\
 \text{Cal } \pi \text{ with } F_{FP} & 0.40 & & -0.77 \quad +1.17
 \end{array}$$

$$\text{Asp } f_{\text{CH}_2} + f_{\text{COO}^-} + 2F_b + F_{P2} + F_{gBr} - f_H$$

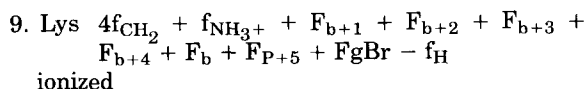
$$\text{CLOGP} = (0.66) + (-5.10) + 2(-0.12) + 0.26(2.41 + 5.10) + (-0.22) - 0.23$$

$$F_{FP2} = 0.26(4.82 + 5.10)$$

Ionized Acid

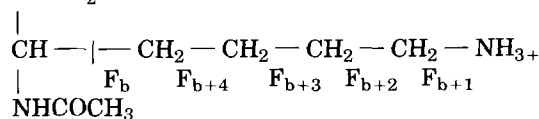


$$\begin{array}{llll} \text{Cal } \pi \text{ with } F_p & -3.18 & \text{OBS}^{15} & -0.77 \text{ DIFF } -2.41 \\ \text{Cal } \pi \text{ with } F_{FP} & -2.55 & & -0.77 \quad -1.78 \end{array}$$



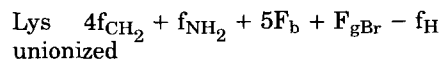
$$\begin{aligned} \text{CLOGP} &= 4(0.66) + (-3.40) + (-0.78) + (-0.40) \\ &\quad + (-0.26) + (-0.19) + (-0.12) + \\ &\quad 0.2(2.41 + 3.40) + (-0.22) - 0.23 \end{aligned}$$

$$F_{FP+5} = 0.2(4.82 + 3.40)$$

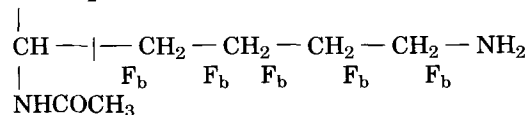


$$\text{Cal } \pi \text{ with } F_p \quad -1.80 \text{ OBS}^{15} \quad -0.99 \text{ DIFF } -0.81$$

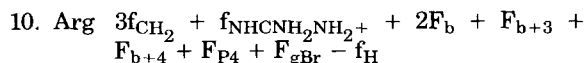
$$\text{Cal } \pi \text{ with } F_{FP} \quad -1.32 \quad -0.99 \quad -0.33$$



$$\begin{aligned} \text{CLOGP} &= 4(0.66) + -1.54 + 5(-0.12) + (-0.22) \\ &\quad - 0.23 \end{aligned}$$

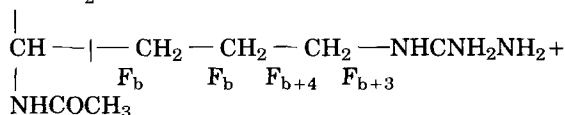


$$\text{Cal } \pi \quad 0.05$$



$$\begin{aligned} \text{CLOGP} &= 3(0.66) + (-5.65)^* + 2(-0.12) \\ &\quad + (-0.26) + (-0.19) + \\ &\quad 0.22(2.41 + 5.65) + (-0.22) - 0.23 \end{aligned}$$

$$F_{FP+4} = 0.22(4.82 + 5.65)$$



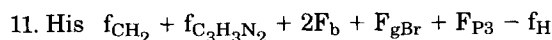
$$\text{Cal } \pi \text{ with } F_p \quad -3.04 \text{ OBS}^{15} \quad -1.01 \text{ DIFF } -2.03$$

$$\text{Cal } \pi \text{ with } F_{FP} \quad -2.51 \quad -1.01 \quad -1.50$$

with 2 waters

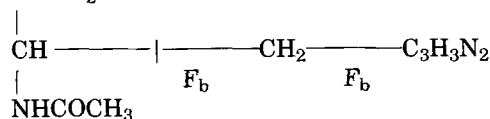
$$\begin{array}{llll} \text{Cal } \pi \text{ with } F_p & & \text{CAL} & \text{OBS} & \text{DIFF} \\ -3.04 + (2 \times +0.63) & = & -1.78 & -1.01 & -0.77 \end{array}$$

$$\begin{array}{llll} \text{Cal } \pi \text{ with } F_{FP} & & \text{CAL} & \text{OBS} & \text{DIFF} \\ -2.51 + (2 \times +0.63) & = & -1.25 & -1.01 & -0.24 \end{array}$$



$$\begin{aligned} \text{CLOGP} &= 0.66 + -0.31 + 2(-0.12) + (-0.22) + \\ &\quad 0.1(2.41 + 1.12) - 0.23 \end{aligned}$$

$$F_{FP3} = 0.1(4.82 + 1.12)$$



AVE. BACKBONE AND CONSIDER HIS POLAR

$$\text{Cal } \pi \text{ with } F_p \quad 0.01 \quad \text{OBS}^{15} \quad 0.13 \quad \text{DIFF } -0.12$$

FULL BACKBONE AND CONSIDER HIS POLAR

$$\text{Cal } \pi \text{ with } F_{FP} \quad 0.25 \quad \text{OBS} \quad 0.13 \quad \text{DIFF } +0.12$$

Summary of CLOGP Calculations for Acetyl Amide Analogs and π Calculations for Side Chain Residues

For nonpolar side chain residues, the method described in this paper satisfactorily calculates log P(o/w) values of the acetyl amide derivatives and the π values for the side chain residues. Although the Trp side chain π value differs by only -0.37 log units from the observed value, it is the only AA side chain residue in the nonpolar side chain category to calculate more hydrophilic than the observed. A possible explanation for this reversal is addressed later.

It is clear that the CLOGP calculations for log P of the polar side chain acetylamide analogs and for their respective π values do not agree as well with the observed values as those for the AA. Both postulates 1 and 2 listed above (to account for the differences between CLOGP and the observed values) would raise the calculated log P. If a folded conformation between the acetylamide (CH_3CONH) or terminal amide (CONH_2) and the polar side chain is the cause of the observed differences, then it is at least a possibility that some of the side chain values predicted by CLOGP could be more useful than those of Fauchere and Pliska¹⁵ in cases where the side chain residue of a peptide or protein may not possess sufficient flexibility to interact with the backbone.

Folded-back conformations of peptides have been observed in X-ray crystal structures. For example, in *p*-bromocarbobenzoxy-glycyl-prolyl-leucyl-glycine, the peptide is folded back at Pro-Leu via an intramolecular hydrogen bond.⁴⁵

Perhaps the most obvious case in point is cystine. The measured value for the diacetylamide of cystine (-1.70) is much more hydrophobic than predicted (-2.87). CPK models show that cystine diacetylamide

*The f value for the guanidinium group in Arg contains the bond factors F_{b+1} and F_{b+2} .

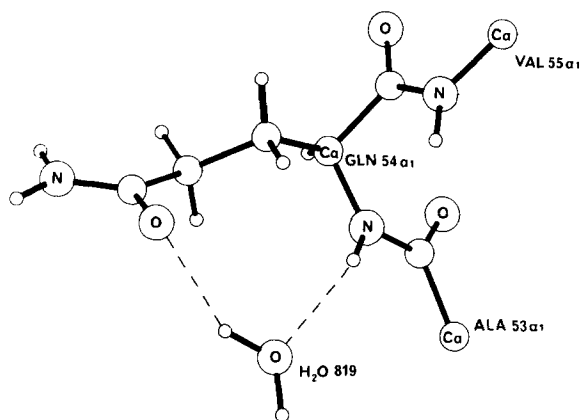


Fig. 4. Type 1 water-bridging structure for glutamine. The coordinates for this drawing were taken from deoxy HbA⁴⁶ GLN 54 α 1. The hydrogen atoms have been calculated. The water bridges the carbonyl oxygen of the side chain amide and the NH of the peptide backbone, either in the same or adjacent residue.

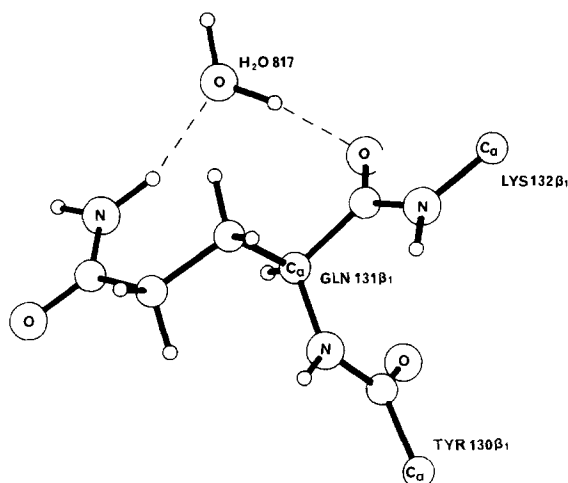
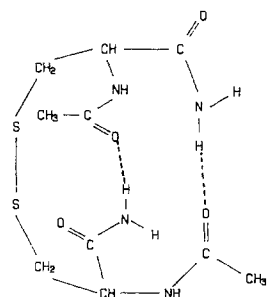


Fig. 5. Type 2 water-bridging structure for glutamine. The coordinates for this drawing were taken from deoxy HbA⁴⁶ GLN 131 β 1. The hydrogen atoms have been calculated. The water bridges the NH_2 of the side chain amide and the carbonyl oxygen of the peptide backbone.

could assume a conformation where the NH_2 of each terminal amide acts as a hydrogen bonding donor to the carbonyl of each acetyl moiety (shown schematically below). This conformation results in one side of the molecule exposing only the hydrophobic sulfur and hydrocarbon, while the other, although polar, has much of its hydrogen bonding capacity tied up internally.



Fauchere and Pliska¹⁵ partitioned the acetylamide derivatives at pH 7.1, and one would expect the side chains in aspartic and glutamic acid analogs to be completely ionized. CLOGP calculates both the Asp and Glu analogs at least two log units too hydrophilic. Even using corrections for the distributions in octanol and water of ionized and unionized species failed to better approximate calculated to the observed values. A folded conformation of the " α " amide (CONH_2) in the glutamic acid analog back toward the side chain carboxyl (similar to that discussed above) could stabilize the latter's proton to discourage its loss, raise its pK_a , and increase its hydrophobicity. This appears possible in the case of the aspartic acid analog also, but only if a rather stable hydrate is formed. Almost all Glu and Asp residues in Hb are associated with waters.

The hydroxyl group in serine and threonine does not seem well positioned to form a strong hydrogen

bond with either carbonyl in the acetylamide analog. However, it is well positioned for a bridging-hydrate and in other cases this calls for an additional factor of +0.7 to +0.8 log units (see above discussion of arginine). Bridging-hydration would bring the π value in line with that measured. This explanation will not serve for the sulfur analog, Cys, but thiols are rather unstable and either the fragment value used in CLOGP or the measured log P may be in error.

The deviation of the CLOGP calculations for asparagine (-0.56) and glutamine (-1.56) could be explained by a conformationally induced interaction between the α amide $-\text{NH}_2$ and the side chain amide $\text{C}=\text{O}$, but in all probability it would require "bridge-hydration." In the case of the glutamine analog, the longer chain would allow an additional interaction of the side chain $\text{C}=\text{O}$ with the α N-acetyl moiety using a hydration bridge. This would be necessary to rationalize the greater deviation of the glutamine analog. We find that such water bridging is predominant with glutamine in deoxy HbA.⁴⁶ Most of the glutamines with bridged waters of hydration in deoxy HbA consist of two structural types which we have designated as type 1 and type 2 (see Figs. 4, 5). In the type 1 structure, the water bridges the carbonyl oxygen of the amide side chain and the NH of the peptide backbone. Figure 4 shows an example of this type of bridging with GLN 54 α 1 of deoxy HbA. In the type 2 structure, the water bridges the amide NH_2 of the side chain and the carbonyl oxygen of the backbone. Figure 5 shows an example of the type 2 bridging with GLN 131 β 1 of deoxy HbA. These hydrogen-bonded structures appear to be good candidates for carrying waters of hydration into octanol.

There are also numerous examples of waters of hydration buried in hydrophobic pockets of proteins.⁴⁷ A good example of a water buried in a hydrophobic cavity can be seen in Figure 6, which depicts

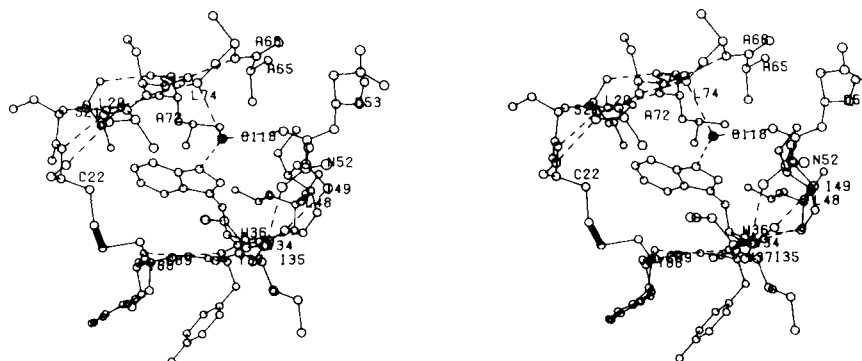


Fig. 6. Stereo diagram of water 0118 buried deep in the protein core of Bence-Jones protein (Rhe) surrounded by hydrophobic residues. The water is hydrogen bonded to Trp36.⁴⁸

water 0118 hydrogen bonded to Trp 36 in the structure of Bence-Jones protein (Rhe).⁴⁸ It seems reasonable that the reversal of the Trp hydrophobicity described above might be due to tryptophan hydrogen bonding to octanol or to a tightly bound water of hydration that would exist in the cavity that encompasses the solute. The variance in the position of Trp in different hydrophobicity scales may be due to hydrogen bonding phenomena.

Considering the above facts we suggest that polar (charged and noncharged) side chain acetyl amide analogs measured by Fauchere and Pliska,¹⁵ except the sulfur analogs, may have waters of hydration or folded acetyl amide structures associated with their transfer into octanol. This requires a positive correction factor of + 0.75, except for Gln which is + 1.50 due to a tightly bridged hydrate structure.*

CONCLUSION

We have found good to excellent agreement for calculating the partition coefficients of the 20 amino acids (Table II, columns a and b). The large difference between the calculated and observed values for Pro may be due to the fact that the superfragment is not structurally similar to the amino carboxylate moieties in Pro. The difference observed for Arg may be due to changes in the polar proximity factor due to hydration phenomena or to difficulty in measuring the log P of highly water-soluble materials. Good agreement has also been found between calculated and observed ranking of side chain hydrophobicity (Tables II, III) if hydration phenomena are considered. Table III also lists hydrophobicity scales from other studies for comparison.

However, the calculated values for side chain hydrophobicity that agree best when compared with the calculated values obtained from the observed transfer data are found by assigning a field effect for both peptide linkages to the α -carbon (and not averaging as found in CLOGP methodology, see point 3 on page 145). This full-effect approach (F_{FP}) of the polar side chains with the polar backbone may in part take into effect the hydration phenomena and/or folded conformations which would increase the hydrophobicity of

the side chains as discussed above. Although this full-effect approach is not included in the CLOGP program, it may be the best way to estimate the hydrophobicities of polar peptide side chain residues.

Future workers in this area might benefit from the following points:

1. The partitioning data of Fauchere and Pliska for AAs and their peptidlike analogs set a standard for this field. Their data are obtained for all 20 natural AAs and peptidlike analogs under carefully controlled experimental conditions using octanol-water as the partitioning system.

2. Octanol/water partition coefficients have been successfully used in a large number of diverse studies to quantitatively correlate drug structure with pharmacological activity, protein binding, membrane transport, and enzyme inhibition and binding. The fragment methodology for calculation of the octanol-water partition coefficients is general in its application to any class of molecules. It is useful for comparison with known measurements and can be used to estimate and accurately predict with reasonable assurance the hydrophobicity of molecules when experimental data are not available or the compounds are not yet synthesized. The octanol-water system probably best mimics the hydrophobicity and hydrophobicity interfaces of the exterior and interior of regions of proteins and membranes; 1-octanol has gained acceptance as the standard solvent for transfer experiments.

3. The hydrophobicity varies from atom to atom in the AA or peptide side chains so that general classifications as hydrophobic or hydrophilic can be misleading. The hydrophobicity of individual atoms can vary with their proximity to other polar or charged atoms.

4. The use of statistical data (which average the number of AA side chains buried, etc., from a large

*It has also been suggested to us by one of the reviewers of this manuscript that another explanation for the increase in hydrophobicity of Gln may be due to the participation of the hydroxyl group of octanol that might replace water in the hydrate structures we proposed in Figures 4 and 5. We plan to make other physical measurements to confirm or disprove these hypotheses.

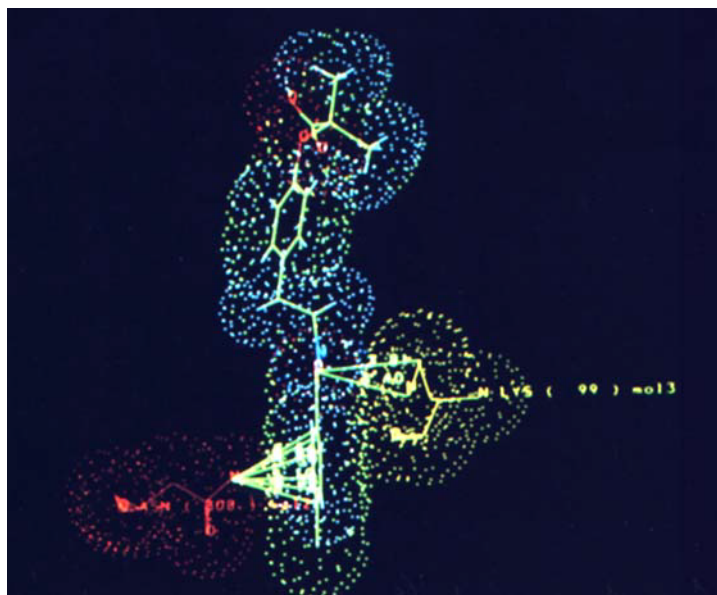


Fig. 7a. The close contact of the $\text{CH}_2\delta$ and $\text{CH}_2\epsilon$ hydrogens of Lys 99 α (in yellow) with the oxygen of the amide group of bezafibrate is indicated by the two green lines on the right side of the photograph. The six green lines on the left of the photograph from the amide nitrogen of Asn 108 β to the chlorophenyl ring of the drug is similar to the Lys $\text{CH}_2\epsilon$ proton in close contact with the center of the π electrons of the ring of His 143 β 1 as shown in Figure 7c. This picture was produced using the program MOGLI from Evans & Sutherland.



Fig. 7b. The close contact of the $\text{CH}_2\delta$ and $\text{CH}_2\epsilon$ methylenes of Lys 99 α (shown with dot surfaces) with the amide group of bezafibrate is on the left side of the photograph. The six pink lines on the right of the photograph, from the amide nitrogen of Asn 108 β to the chlorophenyl ring of the drug is reminiscent of the Lys $\text{CH}_2\epsilon$ hydrogen close contacts with the center of the π electrons of the His 143 β 1 ring shown in Figure 7c. This picture was produced using the program FRODO written by Alwyn Jones.

TABLE V. Close Contacts of Lys Methylene Groups With Oxygen Atoms in Hemoglobin*

$\text{NH}_3^+ - \underset{\epsilon}{\text{CH}_2} - \underset{\delta}{\text{CH}_2} - \underset{\gamma}{\text{CH}_2} - \underset{\beta}{\text{CH}_2} - \underset{\alpha}{\text{CH}} -$						
Lys	CH ₂ ε	Å	CH ₂ δ	Å	CH ₂ γ	Å
7α1	O BB LEU 2α1	3.2	Oδ1 ASP 74α1	3.7	O BB SER 3α1	3.3
7α2	O BB LEU 2α2	3.2			O BB SER 3α2	3.3
11α1	O BB VAL 70α1	3.6	O BB LYS 7α1	3.5	O BB LYS 7α1	3.2
11α2					O BB LYS 7α2	3.2
16α1			Oε2 GLU 116α1	3.5	Oε2 GLU 116α1	3.6
16α2	Oε2 GLU 116α2	3.6				
40α1	OH HIS 146β2	3.4			O BB PRO 37α1	3.6
40α2	O BB PHE 33α2	3.6	O BB PRO 37α2	3.5	O BB PRO 37α2	3.6
	OH HIS 146β1	3.7				
90α1			O BB HIS 89α1	3.5		
127α1	OH ARG 141α2	3.3	Oδ2 ASP 6α1	3.6		
	O BB VAL 1α1	3.1				
	Oδ2 ASP 6α1	3.2				
127α2	Oδ2 ASP 6α2	3.3	Oδ2 ASP 6α2	3.7	O BB ALA 123α2	3.5
	O BB VAL 1α2	3.3				
	OH ARG 141α1	3.5				
139α1			Oδ1 ASP 85α1	3.3		
139α2			Oδ1 ASP 85α2	3.6	O BB SER 138α2	3.5
17β2	Oε1 GLU 121β2	3.0				
61β1					O BB ASN 57β1	3.5
61β2	O BB MET 55β2	3.3			O BB ASN 57β2	3.5
65β1			Oβ1 ASP 22β1	3.4	Oδ1 ASP 21β1	3.1
65β2			Oδ1 ASP 21β2	3.4	Oδ1 ASP 21β2	3.1
			OH ₂ 986	3.6		
82β1	Nε2 HIS 143β1	3.2				
	Cδ2	3.6				
	Cγ	3.7				
	Nδ1	3.6				
	Cε1	3.2				
95β2			O BB ASP 94β2	3.6		
132β2	Oε1 GLU 7β2	3.2			O BB ALA 128β2	3.5
					OH ₂ 890	3.2

*The abbreviation O BB stands for the oxygen of the backbone peptide linkage. The OH in the His 146β subunits and the OH in the Arg 141α subunits represents the C terminal oxygen anion. A 2FO-FC electron density map revealed most of the Lys side chains to be in well defined densities. The distances are from carbon to oxygen.

number of protein structures) does not reveal the underlying basic chemical or physical mechanisms involved with atom-to-atom interactions. The local environment can change the hydrophobicity characteristics of side chain and backbone atoms and such information is lost with purely statistical methods.

5. Tightly bound water to the polar moieties of AAs or peptide-side chain residues might be carried into the octanol phase in two-phase partitioning. Tightly bound water may play an important role in other transfer phenomena for which log P (oct/water) is just a model.

In an extension of this work we are refining the fragment constants for atom types in the AA side

chain residues to include bond factors and proximity factors to give numerical entities that we can program to evaluate docking and protein-folding phenomena.

NOTE ADDED IN PROOF

We decided to survey the close contacts of Lys methylenes in the 1.74 Å structure of deoxyhemoglobin⁴⁶ for an interaction similar to the interaction of Lys 99β with the amide oxygen of bezafibrate²⁷ (discussed on page 131; see Fig. 7a & b). To our surprise we found a large number of Lys methylene close contacts with backbone amide oxygens as well as with carboxylate oxygen anions (see Table V). In another interesting interaction, one of the CH₂ε hydrogens of Lys 82β1 points to the center of the π electrons of the His

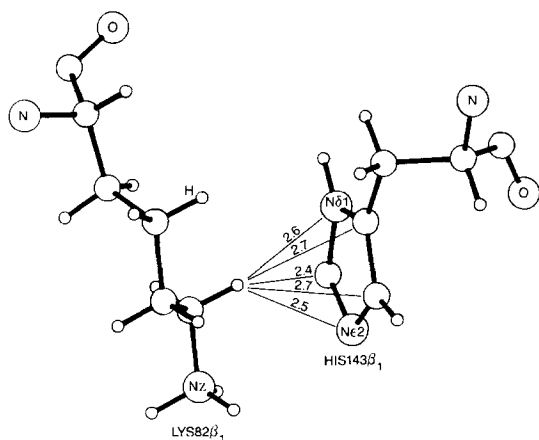


Fig. 7c. One of the hydrogens of $\text{CH}_2\epsilon$ of Lys 82 β 1 centers the π electrons of the His 143 β 1 ring. These distances are from the hydrogen to the ring. The distances in the table above are from the carbon of Lys 82 β 1 to the His 143 β 1 ring.

143 β 1 ring (see Fig. 7c). This is reminiscent of the interaction of NH_2 of Asn 108 β 1 with the π electrons of the chlorophenyl ring of bezafibrate²⁷ (see Fig. 7a & b). These close contacts ($\text{C}\epsilon$, $\text{C}\gamma$, and $\text{C}\delta$ with polar oxygen atoms) add support to the conclusions drawn from the fragmentation calculations vs. solubility measurements for quaternary ammonium compounds, ie, the attached methylenes are not uniformly hydrophobic and the graded charge distribution may account for the close contacts observed.

ACKNOWLEDGMENTS

The authors thank Drs. Nick Franks, David Eisenberg, Corwin Hansch, Richard McClure, and Max Perutz for helpful suggestions and discussions. The authors are indebted to both reviewers of this paper for their careful evaluation and suggestions. In particular, reviewer #2 suggested the use of the equilibrium equations for correcting the partition coefficient calculations for Glu and Asp; the need for a correction in our equation for Gly; and the inclusion of the work of Miyazawa and Jernigan,⁵⁰ which best approximates hydrophobicity by using a statistical method. We also thank Ms. Alicia Delserone and Ms. Franki Williams for their help in preparing this manuscript. This work was supported in part by NIH grant HLBI-RO1-HL-32793, to D.J.A.

REFERENCES

- Rose, G.D., Geselowitz, A.R., Lesser, G.-J., Lee, R.H., Zehfus, M.H.: Hydrophobicity of amino acid residues in globular proteins. *Science* 229:834-838, 1985.
- Chothia, C.J.: The nature of the accessible and buried surfaces in proteins. *Mol. Biol.* 105:1-14, 1976.
- Chothia, C.: Principles that determine the structure of proteins. *Annu. Rev. Biochem.* 53:537-572, 1985.
- Nozaki, N., Tanford, C.: The solubility of amino acids and two glycine peptides in aqueous ethanol and dioxane solutions: Establishment of a hydrophobicity scale. *J. Biol. Chem.* 246:2211-2217, 1971.
- Tanford, C.: The hydrophobic effect and the organization of living matter. *Science* 200:1012-1018, 1978.
- Tanford, C.: *The Hydrophobic Effect: "Formation of Micelles and Biological Membranes."* New York: Wiley, 1980.
- Eisenberg, D.: Three-dimensional structure of membrane and surface proteins. *Annu. Rev. Biochem.* 53:595-623, 1984.
- Eisenberg, D., McLachlan, A.: Solvation energy in protein folding and binding. *Nature* 319:199-203, 1986.
- Janin, J.: Surface and inside volumes in globular proteins. *Nature (Lond.)* 277:491-492, 1979.
- Wolfenden, R., Anderson, L., Cullis, P.M., Southgate, C.B.: Affinities of amino acid side chains for solvent water. *Biochemistry* 20:849-855, 1981.
- Wolfenden, R.: Waterlogged molecules. *Science* 222:1087-1093, 1983.
- Hansch, C., Leo, A.J.: "Substituent Constants for Correlation Analysis in Chemistry and Biology." New York: J. Wiley and Sons, 1979:1-339.
- Rekker, R.F.: "The Hydrophobic Fragmental Constant." New York: Elsevier, 1977.
- Nys, G.C., Rekker, R.F.: Statistical analysis of a series of partition coefficients with special reference to the predictability of folding drug molecules: Introduction of hydrophobic fragmental constants. *Chem. Ther.* 8:521-535, 1973.
- Fauchere, J.-L., Pliska, V.E.: Hydrophobic parameters π of amino-acid side chains from the partitioning of N-acetyl-amino-acid amides. *Eur. J. Med. Chem.* 18:369-375, 1983.
- Meyer, H.: Zur theorie der alkoholnarkose; Erste mittheilung: Welche eigenschaft der anasthetica bedingt ihre markotische wirkung? *Arch. Exp. Pathol. Pharmacol.* 42:109-118, 1899.
- Overton, E.: Über die allgemeinen osmotischen eigenschaften der zelle, ihre vermutlichen ursaden und ihre bedeutung fur die physiologie. *Vierteljahrsschr. Naturforsch. Ges. Zurich* 44:88-135, 1899.
- Hansch, C., Fujita, T.: ρ - σ - π analysis: A method for the correlation of biological activity and chemical structure. *J. Am. Chem. Soc.* 86:1616-1626, 1964.
- Hansch, C., Deutsch, E.W., Smith, R.N.: The use of substituent constants and regression analysis in the study of enzymatic reaction mechanisms. *J. Am. Chem. Soc.* 87:2738-2742, 1965.
- Fujita, T., Iwasa, J., Hansch, C.: A new substituent constant, π , derived from partition coefficients. *J. Am. Chem. Soc.* 86:5175-5180, 1964.
- Leo, A., Hansch, C., Elkins, D.: Partition coefficients and their uses. *Chem. Rev.* 71:525-616, 1971.
- Seidel, A.: "Solubilities of Organic Compounds." New York: Van Nostrand, 1941:313.
- Pearlman, R.S.: Molecular surface area and volume: Their calculation and use in predicting solubilities and free energies of desolvation. In: "Partition Coefficient Determination and Estimation." Dunn W., Block J., Pearlman R., eds. New York: Pergamon, 1986:3.
- Leahy, D.E.: Intrinsic molecular volume as a measure of the cavity term in linear solvation energy relationships: Octanol-water partition coefficients and aqueous solutions. *J. Pharm. Sci.* 75:629-636, 1986.
- Momany, F.A., McGuire, R.F., Burgess, A.W., Scheraga, H.A.: Energy parameters in polypeptides: VII. Geometric parameters, partial atomic charges, nonbonded interactions, hydrogen bond interactions, and intrinsic torsional potentials for the naturally occurring amino acids. *J. Phys. Chem.* 79:2361-2381, 1975.
- Weiner, S.J., Kollman, P.A., Case, D.A., Singh, U.C., Ghio, C., Alagona, G., Profeta, S., Jr., Weiner, P.: A new force field for molecular mechanical simulation of nucleic acids and proteins. *J. Am. Chem. Soc.* 106:765-776, 1984.
- Perutz, M.F., Fermi, G., Abraham, D.J., Poyart, C.J., Bursaux, E.: Hemoglobin as a receptor of drugs and peptides: X-ray studies of the stereochemistry of binding. *J. Am. Chem. Soc.* 108:1064-1078, 1986.
- Licensed by Pomona Medicinal Chemistry Project, Chemistry Department, Pomona College, Claremont, CA 91711.
- Yunger, L.M., Cramer, R.D.: Measurement and correlation of partition coefficients of polar amino acids. *Mol. Pharmacol.* 20:602-608, 1981.
- Verbist, J.J., Lehmann, M.S., Koetzle, T.F., Hamilton, W.C.: Precision neutron diffraction structure determination of protein and nucleic acid components: VI. The crystal and molecular structure of the amino acid L-asparagine monohydrate. *Acta Cryst.* B28:3006-3013, 1972.
- Parthasarathy, R.: Crystal structure of glycylglycine hydrochloride. *Acta Cryst.* B25:509-518, 1969.
- Jeffrey, G.A., Maluszynska, H.: A survey of hydrogen bond

- geometries in the crystal structures of amino acids. *Int. J. Macromol.* 4:173-185, 1982.
33. Koetzle, T.F., Frey, M.N., Lehmann, M.S., Hamilton, W.C.: Precision neutron diffraction structure determination of protein and nucleic acid components: XIII. Molecular and crystal structure of the amino acid L-glutamine. *Acta Cryst.* B29:2571-2575, 1972.
 34. Mack, F., Bonisch, H.N.-S.: Dissociation constants and lipophilicity of catecholamines and related compounds. *Arch. Pharmacol.* 310:1-9, 1979.
 35. Unpublished results, Pomona Medicinal Chemistry Project, Chemistry Department, Pomona College, Claremont, CA 91711.
 36. Madden, J.J., McGandy, E.L., Seaman, N.C., Harding, M.M., Hoy, A.: The crystal structure of the monoclinic form of L-histidine. *Acta Cryst.* B28:2382-2389, 1972.
 37. Pliska, V.E., Schmidt, M., Fauchere, J.L.: Partition coefficients of A.A. and hydrophobic parameters, π , of their side chains as measured by thin layer chromatography. *J. Chromatogr.* 216:79-92, 1981.
 38. Karle, I., Karle, J.: An application of the symbolic addition method to the structure of L-arginine dihydrate. *Acta Cryst.* 17:835-841, 1964.
 39. Lehmann, M.S., Verbist, J.J., Hamilton, W.C., Koetzle, T.F.: Precision neutron diffraction structure determination of protein and nucleic acid components: Part V. Crystal and molecular structure of the amino-acid L-arginine dihydrate. *J. Chem. Soc. Perkin II.* pp. 133-137, 1973.
 40. Beveridge, D.L., Ravishanker, G., Mezei, M.: "Structure and Dynamics: Nucleic Acids and Proteins." Clementi, E., Sarma, R.H., eds. New York: Adenine Press, 1983:477-483.
 41. Koetzle, T.F., Lehmann, M.S., Verbist, J.J., Hamilton, W.C.: Precision neutron diffraction structure determination of protein and nucleic acid components: VII. The crystal and molecular structure of the amino acid L-lysine monohydrochloride dihydrate. *Acta Cryst.* B28:3207-3214, 1972.
 42. Bhat, T.N., Vijayan, M.: X-ray studies of crystalline complexes involving amino acids: II. The crystal structure of L-arginine L-glutamate. *Acta Cryst.* B33:1754-1759, 1977.
 43. Frey, M.N., Lehmann, M.S., Koetzle, T.F., Hamilton, W.C.: Precision neutron diffraction structure determination of protein and nucleic acid components: XI. Molecular configuration and hydrogen bonding of serine in the crystalline amino acids L-serine monohydrate and DL-serine. *Acta Cryst.* B29:876-884, 1973.
 44. Carson, W.M., Hackert, M.L.: Crystal structure of L-tyrosyl-glycyl-glycine monohydrate, the N-terminal tripeptide of the enkephalins. *Acta Cryst.* B34:1275-1280, 1978.
 45. Ueki, T., Kakudo, M., Sasada, Y., Katsube, Y.: Structure of *p*-bromocarbobenzoxy-glycyl-prolyl-leucyl-glycine. *Acta Cryst.* B25:1840-1849, 1969.
 46. Fermi, G., Perutz, M.F., Shaanan, B., Fourme, R.: The crystal structure of human deoxyhaemoglobin at 1.74 Å resolution. *J. Mol. Biol.* 175:159-174, 1984.
 47. Sax, M., Swaminathan, S., Furey, W. Jr.: The conservation of hydration sites and the common fold of the main chain in a family of proteins. Personal Communication.
 48. Furey, W., Jr., Wang, B.C., Yoo, C.S., Sax, M.: Structure of a novel Bence-Jones Protein (Rhe) fragment at 1.6 Å resolution. *J. Mol. Biol.* 167:661-692, 1983.
 49. Leo, A.J. The octanol-water coefficient of aromatic solutes: The effect of electronic interactions on alkyl chains, hydrogen bonds, and ortho substituents. *J. Chem. Soc. Perkins Trans. II.* pp. 825-838, 1983.
 50. Miyazawa, S., Jernigan, R.L.: Estimation of effective inter-residue contact energies from protein crystal structures: Quasi-chemical approximation. *Macromolecules* 18:532-552, 1985.