Quantitative Analysis of Hydrophobicity of Oligopeptides Using Physicochemical Amino Acid Side Chain Parameters and Submolecular Structural Descriptors

1990

Miki Akamatsu

Quantitative Analysis of Hydrophobicity of Oligopeptides Using Physicochemical Amino Acid Side Chain Parameters and Submolecular Structural Descriptors

1990

Miki Akamatsu

Contents

General Introduction

Peptides and their analogs have been attracting interests as potential therapeutic drugs and agrochemicals, recently (Hadži and Jerman-Blažič, 1987; Claassen, 1990; Menn, 1989). The hydrophobicity of drugs is regarded as being a highly important parameter to control transport behaviors from the site of administration to their site of action through a number of biomembranes as well as their binding with hydrophobic receptor sites (Hansch and Fujita, 1964). In addition, the hydrophobicity of component amino acids and oligopeptides is believed to govern not only three-dimensional structure of proteins such as folding and conformational patterns determining their biological functions (Kauzman, 1959), but also affiliation properties of their partial domains into hydrophobic biomembraneous phases (Kyte and Doolittle, 1982).

The log P value, P being the partition coefficient of neutral molecules measured with the 1-octanol/water system, has been widely used to represent molecular hydrophobicity (Hansch and Fujita, 1964). Fauchère and Pliska (Fauchère and Pliška, 1983) have measured the partition ratio, P', of a number of N-acetylamino acid amides with a system of a 1-octanol/pH 7.1 aqueous solution. They estimated the side chain hydrophobicity parameter, π, as the difference of log P' from that of N-acetylglycine amide where the "side chain" is H. They used these π values to represent variations in molecular hydrophobicity in examining quantitative structure-activity relationships for such peptide series as LHRH, angiotensin II, enkephalin (Fauchère, 1984), and pepstatin (Nisato *et al.*, 1987) and their respective analogs, where only one or two of the amino acids at certain positions on the peptide backbone were varied.

Although these studies and some other quantitative approaches (Hadži and Jerman-Bražič, 1987) have been challenging to establish suitable models for designing more active analogs, their success seems to be limited. As was pointed out by Fauchère (Fauchère, 1984), this is mainly due to a lack of reliable sets of structural

descriptors for amino acids and peptides.

A number of hydrophobicity indices of amino acid side chains have been proposed besides π values of Fauchère and Pliška in the field of protein research. Most of them are either defined from partition or phase transfer parameters of unit amino acids (Nozaki and Tanford, 1971) or their analogs (Wolfenden *et al.*, 1981), or statistically calculated from the solvent-accessible surface area of proteins (Kyte and Doolittle, 1982).

This thesis is directed to correlate the variations in the log P or log P' value among peptides and protected peptides having unionizable side chains with free-energy-related physicochemical parameters for the side chain substituents and substructures. Using the results of the correlaiton analyses, a new "effective" hydrophobic parameter for amino acid side chains was defined. The parameter will not only be able to provide the quantitative structure-activity relationship of bioactive peptides with invaluable information, but also possess a potentiality to predict the secondary structure of proteins and peptides.

In chapter 1, a standard procedure to measure P' of zwitterions was described and the log P' value of di- and tripeptides was correlated with substituent and structural parameters. In chapter 2, hydrophobicity of N-acetyl-di- and tripeptide amides in terms of log P were analyzed. In chapter 3, the zwitterionized di- to pentapeptides were brought together and correlation with substituent and structural parameters was discussed. In chapter 4, the quantitative structure-activity relationship of the bitter thresholds of peptides was examined using measured and calculated log P' values. In addition, the new hydrophobicity index of amino acid side chains was defined from the empirical equation obtained for log P' values of oligopeptides and its possible application was discussed.

Chapter 1 Hydrophobicity of Zwitterionized Di- and Tripeptides Having Unionizable Side Chains and Correlation with Substituent and Structural Parameters

1-1 Introduction

There are a number of biologically active peptides, for instance, kyotorphin (Takagi *et al.*, 1979) and MSH (melanocyte-stimulating hormone) release-inhibiting hormone (Nair *et al.*, 1971) which are di- and tripeptides, respectively. In spite of the hydrophobicity of peptides are very important, it has not been examined extensively and systematically. This is probably due to difficulties in establishing experimental conditions for measuring the partition coefficient. Since peptides are zwitterions, their partition ratio is generally very low and dependent not only on the pH of the aqueous phase but also on the concentration and species of the counter ions contained in the aqueous buffer solution.

In this section, we examined and established experimental conditions to obtain reliable partition ratios, P', and measured the P' values of a number of zwitterionized di- and tripeptides composed of amino acids having unionizable side chains in the system 1-octanol/pH 7.0 aqueous buffer. Then, we formulated an equation correlating the variations in the log P'(pH 7) value among peptides with free-energy-related physicochemical parameters for the side chain substituents and substructures.

1-2 Experimental procedure

1-2-1 Measurement of partition ratio (P') of zwitterionized di- and tripeptides

The partition ratio, P', was measured by the flask-shaking method (Fujita, *et al.*, 1964) at 25±3°C in 1-octanol and aqueous buffer. The two phases were mutually saturated in advance. As the aqueous buffer, we used 0.1 M sodium phosphate/phosphoric acid, the ionic strength of which was kept at 0.1 unless otherwise noted. For some compounds, the P' value was measured using an aqueous phase the pH of which was varied from 3 to 8. For the majority of the other compounds, P' was measured with the aqueous phase of pH 7. After the partitioning equilibrium was established, the concentration of peptides in the aqueous phase was measured by their UV absorbance either directly for those containing aromatic amino acids or after HPLC to eliminate the contamination of 1-octanol dissolved in the aqueous phase. The measurement after HPLC was made at 220 nm using a column of Shim-pack CLC-ODS (6.0 mm x 15 cm) or YMCA-303(ODS) (4.6 mm x 25 cm) with 0.01~0.02 M sodium phosphate (pH 7)/acetonitrile for elution. The concentration in the 1-octanol phase was taken as the difference. The log P' value measured in this study was from about -2.5 to 0.5. For the measurement of the lowest values, 25 ml of 1-octanol and 2~2.5 ml of the aqueous buffer were used as the partitioning system with 0.2~0.5 mg of each sample. The highest log P' values in the above range were measured with 2 ml of 1-octanol and 3~5 ml of the aqueous phases and 0.2~0.5 mg of the samples. The measurement was repeated at least three times for each P' value. For the lowest values from -2.0 to -2.5, the individually measured log P' values were taken as reliable enough when they were found in a range of ±0.05. For such peptides, the standard deviation of the log P' averaged over repeats was ±0.05. As the log P' value increased from -2.0 to 0.5, the standard deviation tended to decrease from ±0.05 to ±0.02.

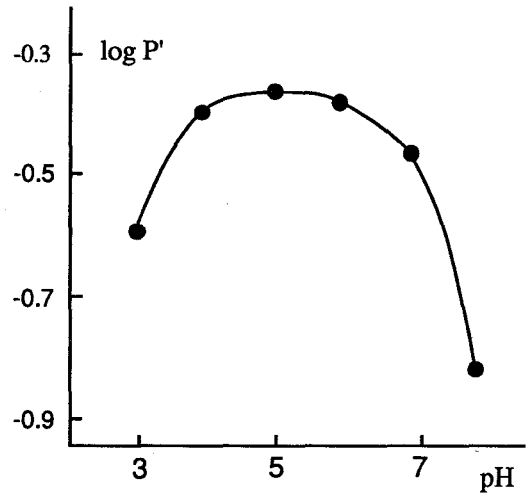


Figure 1-1. pH-log P' profile for Trp-Phe

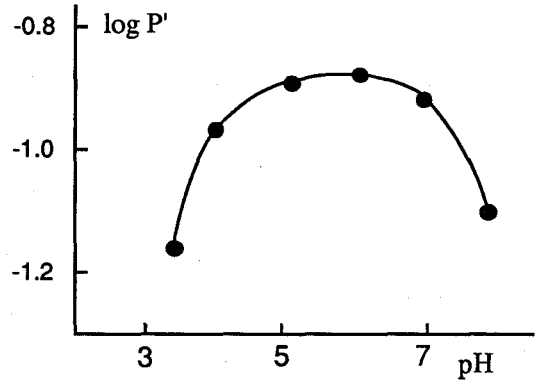


Figure 1-2. pH-log P' profile for Trp-Phe-Ala

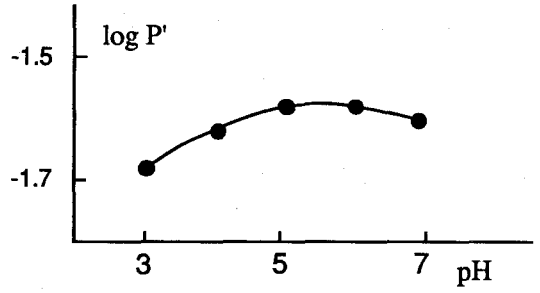


Figure 1-3. pH-log P' profile for Pro-Leu-Leu

1-2-2 Effect of pH on the partition ratio

For compounds where the P' value was measured at various pH's, we attempted to estimate from the pH-log P' profile, the true hydrophobic parameter, log P, for molecular species where the electric charges are canceled. As shown in Figs. 1-1~1-3 for Trp-Phe, Trp-Phe-Ala, and Pro-Leu-Leu, the pH-log P' profile takes a flat parabolic form. The maximum log P' value expected to be observed at the isoelectric point should be the log P value. Unfortunately, the isoelectric point of most peptides in this study was difficult to measure because of their limited solubility. For such di- and tripeptides with sufficient aqueous solubility such as those containing glycine, alanine, and tyrosine, it has been reported to be between 5.5 and 5.7 (Ellenbogen, 1952; Perrin, 1965). For those containing phenylalanine and serine, it is lowered to 5.2~5.5 (Perrin, 1965). For other peptides with amino acids having polar side chains such as threonine, methionine, and tryptophan, the isoelectric point seems to lie between 5.3 and 5.6, considering that the isoelectric point of these "polar" amino acids, 5.6~5.9 (Perrin, 1965), is higher than that of phenylalanine, 5.5 (Perrin, 1965), but lower than that of glycine and alanine, 6.0 (Perrin, 1965). Thus, the isoelectric points of di- and tripeptides studied in this work are considered to exist between 5.2 and 5.7. We assumed that, for other peptides in this work, the log P' value varies with pH following a flat parabola similar to those in Figs. 1-1~1-3. Because of the flat form, the pH-profile of the log P' value is almost horizontal within the pH range between 5 and 6, thus covering sufficiently the range of isoelectric points of di- and tripeptides. Thus, the log P' value measured with the aqueous buffer at pH 6 should be very close to log P.

For 22 peptides in Table 1-1, we measured log P' values both at pH 6 and 7. The log P'(pH 6) which approximates the log P value is correlated very well with log P'(pH 7) as shown in eq. 1-1.

Table 1-1. Log P' Values of Peptides at pH 6 and pH 7

No.	Compounds	log P'(pH 7)	log P'(pH 6)	
			Obsd.	Calcd.(eq. 1-1)
1	FL	-1.17	-1.07	-1.09
2	LF	-1.15	-1.12	-1.07
3	FF	-0.85	-0.82	-0.75
4	LL	-1.46	-1.47	-1.41
5	WW	-0.27	-0.12	-0.14
6	WF	-0.47	-0.40	-0.35
7	WA	-1.98	-1.85	-1.95
8	WL	-0.73	-0.63	-0.63
9	FP	-1.36	-1.23	-1.29
10	PF	-2.07	-2.24	-2.05
11	FFF	-0.02	0.13	0.13
12	GFF	-1.33	-1.21	-1.26
13	FVG	-2.33	-2.47	-2.32
14	FVF	-0.76	-0.62	-0.66
15	FVA	-2.19	-2.01	-2.18
16	LLL	-0.94	-0.92	-0.85
17	WGG	-2.72	-2.65	-2.74
18	WFA	-1.00	-0.90	-0.91
19	WWL	0.36	0.60	0.53
20	PLL	-1.64	-1.62	-1.59
21	LPL	-1.56	-1.52	-1.50
22	LLP	-1.58	-1.47	-1.52

$$\log P'(\text{pH } 6) = \underset{(0.050)}{1.062}\ \log P'(\text{pH } 7) + \underset{(0.073)}{0.149} \qquad (1\text{-}1)$$

$$n = 22 \quad s = 0.085 \quad r = 0.995 \quad F_{1,20} = 1949$$

Table 1-2. Effects of Counter Anions on log P' of Peptides at pH 1

No.	Compound	log P'(pH 1)		
		ClO_4^- [a]	NO_3^- [a]	Cl^- [a]
1	LF	0.14	-0.44	-0.66
2	FF	0.37	-0.17	-0.43
3	WF	0.77	0.17	0.02
4	WA	-0.68	-1.22	-1.42
5	WW	1.10	0.50	0.21
6	WL	0.68	0.08	-0.18

[a] The hydrophobicity of anions decreases as ClO_4^->NO_3^->Cl^-. While the free energy of transfer of ClO_4^-, NO_3^-, and Cl^- ions from water to nitrobenzene is 8.7, 24.4, and 30.5 KJ/mol (Koryta, 1984), the log of the distribution constant of $HClO_4$, HNO_3, and HCl between propylene carbonate and water is -0.05, -1.05, and -1.60, respectively (Cabon *et al.*, 1978).

In eq. 1-1 and the following correlation equations, n is the number of compounds, s is the standard deviation, r is the correlation coefficient, F is the ratio of variances between calculated and observed values, and the figures in parentheses are the 95% confidence intervals. Equation 1-1 indicates that the log P' value measured with the pH 7 buffer parallels that measured at pH 6. The constant term indicates that log P'(pH 6) is slightly but significantly higher than log P'(pH 7). Since the measurement was made only with the pH 7 buffer for the majority of peptides, we used the log P'(pH 7) value for the formulation of empirical correlation

equations.

In preliminary examinations for the experimental conditions, we observed unexpectedly high log P' values with the acidic buffer (pH<4), when NaCl was used to adjust the ionic strength of the buffer solutions. The theoretically reasonable flat parabola was not obtained until the buffer solution was prepared only from sodium hydrogen phosphate and dihydrogen phosphate. This was considered to be due to the partitioning of the ion-pair with counter anions. The phosphate anion, probably existing as a mixture of mono- and divalent species under the acidic pH conditions, is perhaps much less hydrophobic than chloride. We attempted to confirm this by using 0.1N perchloric, nitric, or hydrochloric acid as the aqueous phase (pH 1) of the partitioning system. As shown in Table 1-2, the log P' value of some peptides clearly increases with the hydrophobicity of the counter anions in the aqueous phase at pH 1 (Koryta, 1984; Cabon *et al.*, 1978). The log P' value was also not constant in the basic region depending upon the counter cation in the aqueous phase. For instance, we observed that log P'(pH 13) of Trp-Phe using 0.1N Et_4NOH, -0.68, was much higher than that with 0.1N NaOH, -1.27.

The above results are a warning against measuring the partition ratio under conditions where it depends on the hydrophobicity of counter ions since the latter form ion-pairs with ionized peptides and partition into the organic phase.

1-2-3 Physicochemical substituent parameters

In the course of preliminary analyses, we found that the variations in the log P'(pH 7) value are governed at least by the hydrophobic and steric effects of side chain substituents of component amino acids. As the hydrophobic parameter of side chain substituents, we chose to use the π value which was shown to be of general utility for aliphatic substituents free from components such as intramolecular steric and hydrogen-bonding interactions. The π

Table 1-3. Hydrophobicity Scales of Amino Acid Side Chains

Amino Acid	π [a]	π(Fauchère) [b]
Gly	0.00	0.00
Ala	0.32 [c]	0.31
Val	1.27 [c]	1.22
Leu	1.81 [c]	1.70
Ile	1.81 [c]	1.80
Phe	1.95 [d]	1.79
Tyr	1.20 [e]	0.96
Trp	1.92 [f]	2.25
Met	0.61 [g]	1.23
Ser	-1.49 [h]	-0.04
Thr	-1.18 [i]	0.26
Asn	-1.95 [j]	-0.60
Gln	-1.41 [k]	-0.22
Pro	0.86 [l]	0.72

a Except for log P values of γ-indole-3-butyric acid at pH 1 for Trp residue and $Ph(CH_2)_2CH(OH)CH_3$ for Thr residue which were newly measured here, the log P, π and factor values of related compounds, substituents, and structural features were taken from the literature (Iwasa *et al.*, 1965).

b From ref. (Fauchère and Pliska, 1983).

c Estimated from the π value of alkyl groups, the group branch factor(F_{gBr} = -0.22), and the chain branch factor (F_{cBr} = -0.13).

d $LogP[Ph(CH_2)_3OH] - \log P[CH_3(CH_2)_2OH] + \pi(Me) + F_{gBr} = 1.88 - 0.25 + 0.54 + (-0.22)$

e $\pi(Phe) + \log P[p\text{-}CH_3C_6H_4OH] - \log P[CH_3C_6H_5] = 1.95 + 1.94 - 2.69$

f Log P[γ-Indole-3-butyric acid, pH 1] - $\log P[CH_3CH_2COOH] + F_{gBr} = 2.47 - 0.33 + (-0.22)$

g $\pi(C_2H_5) + \pi(SMe) + F_{gBr} = 1.08 + (-0.25) + (-0.22)$

Table 1-3. continued

h Log P[$Ph(CH_2)_3OH$] - log P[PhC_2H_5] + F_{gBr} = 1.88 - 3.15 + (-0.22)

i Log P[$Ph(CH_2)_2CH(OH)CH_3$] - log P[PhC_2H_5] + F_{gBr} = 2.19 - 3.15 + (-0.22)

j Log P[$Ph(CH_2)_3CONH_2$] - log P[PhC_3H_7] + π(Me) + F_{gBr} = 1.41 - 3.68 + 0.54 + (-0.22), F_{gBr} being the factor correcting the effect of branching of the side chain from the peptide skeleton according to ref. (Hansch and Leo, 1979).

k π(Asn) + π(Me) = -1.95 + 0.54

l Log P[*cyc*-HexOH] - log P[*i*-PrOH] + F_{gBr} + F_b^{ring}(the ring-bond factor) = 1.17 + (-0.22) - 0.09 = 0.86

value of some polar side chains was estimated from experimentally measured log P values of related compounds. The measurements were based on the fact that the "intrinsic" aliphatic π value of polar side chain substituents should be evaluated from log P value of related compounds in which the polar substituents in question are separated by at least two methylene units from another functional group or a chromophore (Iwasa *et al.*, 1965). As shown in Table 1-3, our value for alkyl side chains, equivalent with that listed by Hansch and Leo (Hansch and Leo, 1979), is close to that of Fauchère and Pliška (Fauchère and Pliška, 1983). Our π value for polar side chains in serine, threonine, methionine, and tryptophan is, however, more negative than the corresponding value of Fauchère and Pliška. Being experimentally derived from log P' values of N-acetylamino acid amides, their π value for polar side chains is composed not only of the "intrinsic" hydrophobicity but also of components attributable to various intramolecular factors inherent to the peptide-type structure of N-acetylamino acid amides.

For the steric effects of side chain substituents except for "that" of proline, we used the $E'_S{}^c$ parameter as shown in Table 1-4.

Table 1-4. Steric Parameters of Amino Acid Side Chains

Amino Acid	E'_S		υ [b]	$E'_S{}^c$ [c]
	Obsd. [a]	Calcd.(eq. 1-3)		
Gly	0.00	0.00	0.00	0.00
Ala	-1.12	-1.08	0.52	-0.20
Val	-1.60	-1.59	0.76	-1.29
Leu	-2.05	-2.06	0.98	-1.44
Ile	-2.12	-2.13	1.02	-1.81
Phe	-1.51	-1.46	0.70	-0.90
Tyr	d	-1.46	0.70	-0.90 [e]
Trp	d	-1.46	0.70	-0.86 [f]
Met	-1.64	-1.63	0.78	-1.03
Ser	-1.09	-1.10	0.53	-0.48
Thr	-1.04	-1.04	0.50	-0.73
Asn	d	-1.59	0.76	-0.98 [f]
Gln	d	-1.42	0.68	-0.82 [f]

a From ref. (MacPhee *et al.*, 1978), unless noted. The reference point is shifted so that the value for H in glycine is zero.

b From ref. (Charton, 1983).

c From ref. (Takayama *et al.*, 1985). The reference point is shifted so that the value for H in glycine is zero.

d Not available in ref. (MacPhee, *et al.*, 1978).

e Taken as that of Phe.

f From eq. 1-3.

The $E'_S{}^c$ is the "corrected" Dubois steric parameter related to the original Dubois E'_S (MacPhee *et al.*, 1978) by eq. 1-2, where n is the number of α-hydrogen atoms in aliphatic substituents.

$$E'_S{}^c = E'_S - 0.306(3 - n) \qquad (1\text{-}2)$$

The E'_S value is the steric parameter defined as being an "improved" Taft E_S value (Taft, 1965). In eq. 1-2, the correction term takes the same form as that in the correction for the Taft E_S value of alkyl substituents made by Hancock and coworkers (Hancock *et al.*, 1961). By the correction, they tried to eliminate a component attributable to possible hyperconjugation effects on such reference reactions as acid-catalyzed hydrolysis of esters in defining the E_S value. As indicated previously (Takayama *et al.*, 1985), however, the "corrected" $E'_S{}^c$ value is not the parameter corrected for the hyperconjugation effect attributable to the α-hydrogen atoms of substituents, but the parameter representing not only the steric bulk but also the effect of α-branching. The coefficient of the correction term is fixed as -0.306 in eq. 1-2, but values between -0.25 and -0.35 were shown to work as well. The relevance of the use of $E'_S{}^c$ for the steric effect of aliphatic substituents was discussed in detail in our previous analyses of the log P value of aliphatic amines and the ion-pair formation-partition constant of aliphatic ammonium ions (Takayama *et al.*, 1985). The reference point of $E'_S{}^c$ was shifted so that $E'_S{}^c(H)$ is zero for glycine.

For most side chain substituents dealt with in this work, the E'_S value was available. The E'_S value of indole-3-methyl was estimated by use of a highly linear relationship between E'_S value and Charton's υ steric (Charton, 1983) parameter expressed by eq. 1-3.

$$E'_S = \underset{(0.061)}{-2.090}\ \upsilon\ \underset{(0.043)}{-\ 0.008} \qquad (1\text{-}3)$$

$$n = 9 \quad s = 0.022 \quad r = 0.9995 \quad F_{1,7} = 6591$$

The υ value represents an "effective" width of substituents in Å-units relative to that of H and is estimable from geometry using empirical rules. The $E'_S{}^c$ value of 4-hydroxybenzyl was taken to be equivalent to that of benzyl in phenylalanine.

1-3 Results

First, we examined the log P'(pH 7) value of peptides composed of nonpolar amino acids; glycine, alanine, valine, leucine, isoleucine, and phenylalanine, in terms of the summation of the side chain p value of component amino acids and derived eq. 1-4 with the indicator variable I_{tri}.

$$\log P'(\text{pH } 7) = \underset{(0.217)}{0.804}\ \Sigma\pi - \underset{(0.417)}{0.689}\ I_{tri} - \underset{(0.752)}{4.425} \qquad (1\text{-}4)$$

$$n = 20 \quad s = 0.333 \quad r = 0.892 \quad F_{2,17} = 32.9$$

The I_{tri} is zero for dipeptides and one for tripeptides. In trying to improve the correlation, we noticed that the log P' value of peptides containing β-branched amino acids with α-branching on the side chain, such as valine and isoleucine were more negative than the value calculated by eq. 1-4 (data not shown). Since the steric effect of the "crowded" structure of the branched side chain on the relative solvation of the NHCO moiety and/or terminal NH_3^+ and COO^- groups with partitioning solvents was anticipated to contribute to these deviations, we introduced steric term(s) into eq. 4. Among various steric parameters such as Verloop STERIMOL (Verloop *et al.*, 1976), Charton's υ (Charton, 1983), and the "uncorrected" Dubois E'_S (MacPhee, 1978), the $E'_S{}^c$ parameter worked best, yielding eq. 1-5.

$$\log P'(\text{pH } 7) = \underset{(0.081)}{1.029}\ \Sigma\pi - \underset{(0.225)}{0.842}\ I_{tri} + \underset{(0.131)}{0.519}\ E'_S{}^c(R_N) + \underset{(0.188)}{0.337}\ E'_S{}^c(R_M) + \underset{(0.128)}{0.335}\ E'_S{}^c(R_C) - \underset{(0.264)}{4.239} \qquad (1\text{-}5)$$

$$n = 20 \quad s = 0.109 \quad r = 0.990 \quad F_{5,14} = 153$$

R_N and R_C represent the side chain of amino acids at the N- and C-termini, respectively. R_M is for the side chain of the central amino acid for tripeptides. Equation 1-5 indicates that the steric effect of the side chains on the relative solvation with partitioning solvents depends upon their location in the molecule.

When including peptides with polar amino acids, no relevant correlation equation was derived unless indicator variable terms for the presence of respective polar amino acids were introduced to give eq. 1-6 (Y, W, M, S, and T are letter notation for tyrosine, tryptophan, methionine, serine, and threonine, respectively):

$$\begin{aligned}\log P'(\text{pH } 7) = {} & \underset{(0.075)}{0.960}\ \Sigma\pi - \underset{(0.136)}{0.635}\ I_{tri} + \underset{(0.096)}{0.561}\ E'_S{}^{c}(R_N) \\ & + \underset{(0.123)}{0.337}\ E'_S{}^{c}(R_M) + \underset{(0.097)}{0.255}\ E'_S{}^{c}(R_C) + \underset{(0.079)}{0.165}\ I_Y \\ & + \underset{(0.096)}{0.352}\ I_W + \underset{(0.149)}{0.637}\ I_M + \underset{(0.219)}{1.665}\ (I_S + I_T) - \underset{(0.210)}{3.912} \qquad (1\text{-}6) \\ & n = 59 \quad s = 0.138 \quad r = 0.982 \quad F_{9,49} = 148\end{aligned}$$

Since the slope values for I_S and I_T terms were very close in the preliminary calculation, they are combined in eq. 1-6. Proline and its derivatives were excluded from the analysis, since the E'_S value for the cyclic "side chain" was not available. In Table 1-5, the log P' values calculated by eq. 1-6 were compared with those experimentally measured. Table 1-6 shows the degree of independence of variables used in eq. 1-6. In Table 1-7, the development of eq. 1-6 was listed.

Table 1-5. Log P'(pH 7) and Physicochemical Parameters of Di- and Tripeptides

No.	Compounds	$\Sigma\pi$	$E'_s{}^c(R_N)$	$E'_s{}^c(R_M)$	$E'_s{}^c(R_C)$	log P'(pH 7)	
						Obsd.	Calcd.(eq. 1-6)
1	FL	3.76	-0.90	0.00	-1.44	-1.17	-1.18
2	LF	3.76	-1.44	0.00	-0.90	-1.15	-1.34
3	FF	3.90	-0.90	0.00	-0.90	-0.85	-0.90
4	LL	3.62	-1.44	0.00	-1.44	-1.46	-1.61
5	LV	3.08	-1.44	0.00	-1.29	-2.05	-2.09
6	VL	3.08	-1.29	0.00	-1.44	-2.07	-2.05
7	AI	2.13	-0.20	0.00	-1.81	-2.60	-2.44
8	II	3.62	-1.81	0.00	-1.81	-1.82	-1.91
9	LI	3.62	-1.44	0.00	-1.81	-1.64	-1.71
10	VV	2.54	-1.29	0.00	-1.29	-2.82	-2.53
11	WW	3.84	-0.86	0.00	-0.86	-0.27	-0.22
12	WF	3.87	-0.86	0.00	-0.90	-0.47	-0.56
13	WA	2.24	-0.86	0.00	-0.20	-1.98	-1.94
14	WL	3.73	-0.86	0.00	-1.44	-0.73	-0.83
15	LY	3.01	-1.44	0.00	-0.90	-1.94	-1.90
16	YL	3.01	-0.90	0.00	-1.44	-1.75	-1.73
17	VY	2.47	-1.29	0.00	-0.90	-2.52	-2.33
18	WY	3.12	-0.86	0.00	-0.90	-1.13	-1.11
19	FY	3.15	-0.90	0.00	-0.90	-1.68	-1.46
20	YY	2.40	-0.90	0.00	-0.90	-1.87	-2.01
21	LM	3.12	-1.44	0.00	-1.02	-1.87	-2.02
22	ML	3.12	-1.02	0.00	-1.44	-1.84	-1.89
23	MV	2.58	-1.02	0.00	-1.29	-2.53	-2.37
24	FM	3.26	-0.90	0.00	-1.02	-1.59	-1.58
25	SL	0.32	-0.48	0.00	-1.44	-2.49	-2.58
26	PF	2.81	-	0.00	-0.90	-2.07	-
27	PL	2.67	-	0.00	-1.44	-2.41	-
28	PI	2.67	-	0.00	-1.81	-2.56	-
29	FP	2.81	-0.90	0.00	-	-1.36	-
30	LP	2.67	-1.44	0.00	-	-1.76	-
31	IP	2.67	-1.81	0.00	-	-1.79	-
32	FFF	5.85	-0.90	-0.90	-0.90	-0.02	0.03

Table 1-5. continued

33	GFF	3.90	0.00	-0.90	-0.90	-1.33	-1.34
34	FVG	3.22	-0.90	-1.29	0.00	-2.33	-2.40
35	FVF	5.17	-0.90	-1.29	-0.90	-0.76	-0.75
36	FVA	3.54	-0.90	-1.29	-0.20	-2.19	-2.14
37	LVV	4.35	-1.44	-1.29	-1.29	-2.10	-1.94
38	LII	5.43	-1.44	-1.81	-1.81	-1.11	-1.21
39	LVL	4.89	-1.44	-1.29	-1.44	-1.57	-1.46
40	LAL	3.94	-1.44	-0.20	-1.44	-2.03	-2.01
41	LLL	5.43	-1.44	-1.44	-1.44	-0.94	-1.00
42	WGG	1.92	-0.86	0.00	0.00	-2.72	-2.83
43	WFA	4.19	-0.86	-0.90	-0.20	-1.00	-1.01
44	WWL	5.65	-0.86	-0.86	-1.44	0.36	0.44
45	LLY	4.82	-1.44	-1.44	-0.90	-1.34	-1.28
46	VFY	4.42	-1.29	-0.90	-0.90	-1.50	-1.40
47	GFY	3.15	0.00	-0.90	-0.90	-1.96	-1.89
48	YLV	4.28	-0.90	-1.44	-1.29	-1.45	-1.59
49	YVF	4.42	-0.90	-1.29	-0.90	-1.37	-1.31
50	YGF	3.15	-0.90	0.00	-0.90	-1.86	-2.09
51	YYL	4.21	-0.90	-0.90	-1.44	-1.38	-1.35
52	AYI	3.33	-0.20	-0.90	-1.81	-2.04	-2.06
53	IYV	4.28	-1.81	-0.90	-1.29	-1.77	-1.92
54	MLF	5.07	-1.02	-1.44	-0.90	-1.03	-1.00
55	LSL	2.13	-1.44	-0.48	-1.44	-2.35	-2.17
56	ISL	2.13	-1.81	-0.48	-1.44	-2.28	-2.38
57	ISI	2.13	-1.81	-0.48	-1.81	-2.64	-2.48
58	SLI	2.13	-0.48	-1.44	-1.81	-1.99	-2.05
59	SLL	2.13	-0.48	-1.44	-1.44	-2.03	-1.96
60	FIT	2.58	-0.90	-1.81	-0.73	-1.95	-1.71
61	LIT	2.44	-1.44	-1.81	-0.73	-2.14	-2.14
62	IIT	2.44	-1.81	-1.81	-0.73	-2.23	-2.35
63	LTI	2.44	-1.44	-0.73	-1.81	-2.30	-2.06
64	TLI	2.44	-0.73	-1.44	-1.81	-1.66	-1.90
65	TVL	1.90	-0.73	-1.29	-1.44	-1.97	-2.27
66	PLL	4.48	-	-1.44	-1.44	-1.64	-
67	LPL	4.48	-1.44	-	-1.44	-1.56	-
68	LLP	4.48	-1.44	-1.44	-	-1.58	-
69	IPI	4.48	-1.81	-	-1.81	-1.65	-

Table 1-6. Squared Correlation (r^2) Matrix for Variables Used in Eq. 1-6.

	$\Sigma\pi$	I_{pep}	$E'_S{}^c(R_N)$	$E'_S{}^c(R_M)$	$E'_S{}^c(R_C)$	I_Y	I_W	I_M
I_{pep}	0.089							
$E'_S{}^c(R_N)$	0.027	0.000						
$E'_S{}^c(R_M)$	0.100	0.584	0.000					
$E'_S{}^c(R_C)$	0.001	0.004	0.015	0.002				
I_Y	0.009	0.000	0.015	0.004	0.005			
I_W	0.026	0.017	0.030	0.045	0.030	0.018		
I_M	0.027	0.054	0.000	0.019	0.000	0.028	0.013	
$(I_S + I_T)$	0.312	0.121	0.007	0.132	0.066	0.078	0.035	0.024

Table 1-7. Development of Eq. 1-6.

const.	$\Sigma\pi$	I_W	I_{pep}	$I_S + I_T$	$E'_s{}^c(R_N)$	I_M	$E'_s{}^c(R_M)$	$E'_s{}^c(R_C)$	I_Y	r	s	$F_{X,Y}$ [a]
-3.191	0.455									0.756	0.442	$F_{1,57}$=77.1
-3.164	0.419	0.561								0.845	0.364	$F_{1,56}$=28.6
-3.134	0.452	0.517	-0.235							0.861	0.349	$F_{1,55}$=5.93
-3.727	0.642	0.517	-0.562	0.724						0.904	0.296	$F_{1,54}$=22.9
-3.594	0.734	0.427	-0.700	0.949	0.458					0.944	0.230	$F_{1,53}$=36.8
-3.787	0.779	0.456	-0.705	1.076	0.482	0.453				0.960	0.197	$F_{1,52}$=21.1
-3.938	0.841	0.413	-0.590	1.242	0.528	0.508	0.228			0.966	0.184	$F_{1,51}$=8.73
-3.912	0.927	0.320	-0.657	1.513	0.567	0.547	0.327	0.243		0.976	0.158	$F_{1,50}$=19.6
-4.075	0.960	0.351	-0.703	1.666	0.561	0.637	0.338	0.255	0.164	0.982	0.136	$F_{1,49}$=18.2

[a] F statistic for significance of the addition of variables.
X: The number of independent variables added at each step of the development, Y: n-m-1,
m being the total number of independent variables in the developed equation. $F_{1,40,0.05} = 4.09$

1-4 Discussion

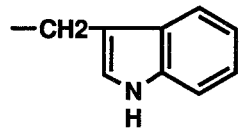
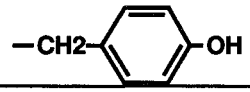
Equation 1-6 shows that the "true" hydrophobicity of peptides approximated by log P'(pH 7) is in fact governed by not only the intrinsic hydrophobic factor of the side chain of component amino acids but also by factors attributed to the steric effects of the side chain and to structural features of polar side chains expressed by specific indicator variables.

The fact that the slope of the $\Sigma\pi$ term is very close to 1, as it theoretically should be, shows that the intrinsic hydrophobic factor of side chains of constituent amino acids indeed contributes to the total hydrophobicity of peptides almost as such after factors attributed to steric and other structural effects are separated.

The coefficient of the $E'_S{}^c$ term of the N-terminal side chain is considerably larger than that of the C-terminal side chain. For the side chain of the central moiety in tripeptides, it lies in-between. At the N-terminus, the NH_3^+ group, working as a hydrogen donor, is solvated by the hydrogen bond with partitioning solvents as hydrogen acceptors. 1-Octanol, being more basic than water, would effectively compete in this type of solvation. At the C-terminal, the COO^- group could be solvated more favorably with the more acidic water as the hydrogen donor. The more crowded the side chain substituents, the less favorable would be the solvation of terminal charged groups with the bulkier 1-octanol. The solvation with 1-octanol is more favorable at the N-teminal than at the C-terminal end and its relative solvation more sensitively controlled by the steric congestion of the side chain. The relative solvation of the backbone CONH moiety, in particular, that at the uncharged NH group with hydrogen accepting solvents could be less sensitive to the steric effect of the side chain than that at the terminal charged NH_3^+. The variations in the size of the $E'_S{}^c$ term depending upon the location of side chain substituents could be rationalized by these discussions. Setting as $E'_S{}^c(H) = 0$ for dipeptides in eq. 1-6 does not mean that the "phantom" central amino acid is regarded as being glycine. The

effect attributed to the "phantom" glycine is taken care of by the I_{tri} term.

Table 1-8. Regression Coefficient of Indicator Variable Terms

Amino Acid	Side Chain	Regression Coefficient [a]	n [b]
Ser	$-CH_2OH$	1.665(0.8~1.1) [c]	2
Thr	$-CH(CH_3)OH$	1.665(0.8~1.1) [c]	2
Met	$-CH_2CH_2SCH_3$	0.637	3
Trp		0.352	4
Tyr		0.165	6

[a] The regression coefficient value of indicator variable terms in eq. 1-6.

[b] The number of bonds separating the polar hetero atom in the polar group from the a-carbon of the peptides.

[c] The value in parentheses is "corrected" by subtracting the intramolecular bridging-solvation factor.

The indicator variable terms for amino acids having the polar side chain are always positive. The log P' value of peptides with these side chains is higher than that predicted by intrinsic hydrophobicity of the peptide skeleton and side chains as well as factors attributed to the steric effect on the relative solvation. In Table 1-8, the regression coefficient value of indicator variable terms is

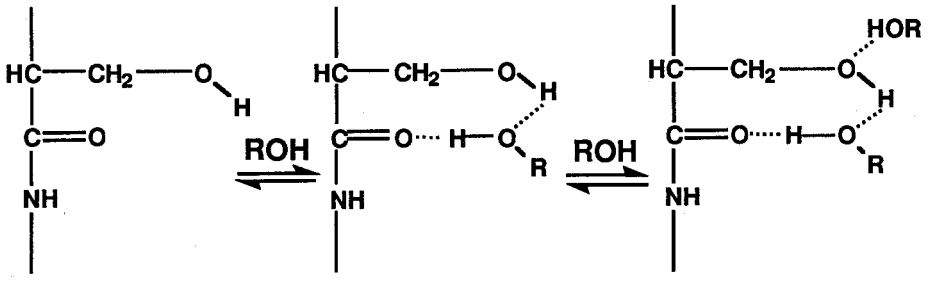


Figure 1-4. Intramolecular bridging-solvation of the hydroxyl group of Ser residue and the carbonyl group on the backbone. R is either 1-Oct or H.

compared with the location of the polar group on the side chain. The more distant the polar group from the peptide backbone, the lower the regression coefficient. This effect may correspond to "the polar proximity factor" termed by Hansch and Leo (Hansch and Leo, 1979) for the reduction of hydrophilicity observed when polar groups are crowded together.

For the side chain of serine and threonine, the size of the coefficient is remarkably high. This is probably due to the fact that the hydroxyl group in the side chain and the carbonyl group in the backbone are well positioned for an intramolecular bridging-solvation, as shown in Fig. 1-4. This type of bridging hydration has been observed in glutamine in the crystal structure of human deoxyhaemoglobin (Fermi *et al.*, 1984). Abraham and Leo (Abraham and Leo, 1987) discussed the possibility for serine and threonine to take this type of bridging-solvation structure in rationalizing the side chain π value of Fauchère and Pliška (Fauchère and Pliška, 1983) which is significantly higher than the value usually used for aliphatic systems (See Table 1-3). This type of solvation was estimated to make the log P' value 0.6~0.9 unit higher than that of the structure without such "intramolecular" solvation (Fujita, 1983; Leo, 1983). Subtracting the value attributed to the bridging solvation, the size of

the regression coefficient for Ser and Thr residues is about 0.8~1.1. The "corrected" regression coefficient value seems to decrease with the number of bonds separating the polar hetero atom on the side chain from the backbone more regularly than the uncorrected value, as shown in Table 1-8. When the number of bonds increases, the net inductive electron-withdrawing effect of the side chain polar groups on the backbone functional group is gradually reduced. The electron-withdrawing effect of substituents works to raise the partition coefficient in a series of substituted compounds regardless of whether the functional group is hydrogen-donating or hydrogen-accepting (Fujita *et al.*, 1964). Thus, the greater the number of bonds, the lower should be the increment in the log P' value assigned as the polar proximity factor (Hansch and Leo, 1979).

The coefficient of the I_{tri} term means that the log P' value decreases by about 0.6 log unit by introduction of one more peptide bond into the dipeptide backbone, other factors being equal. It also corresponds with π value of the CH_3CONH group. Since it contains components assignable to electronic effects of neighboring polar groups such as NH and CO on the relative solvation, it could be taken as being the parameter intrinsic only to the peptide backbone. This will be discussed more in detail in chapter 2. The intercept of eq. 1-6 should correspond with the log P' value of Gly-Gly.

We did not include peptides containing proline in eq. 1-6, since the $E'_S{}^C$ value of the cyclic "side chain" was not available. In fact, variations of log P' values depending upon the position of Pro residue are more pronounced than for peptides composed of other amino acids. Log P' of peptides including proline will be discussed in chapter 3.

Equation 1-7 was derived for the same set of compounds as that in eq. 1-6 using side chain π values proposed by Fauchère and Pliška.

$$\log P'(\text{pH } 7) = \underset{(0.104)}{1.067}\ \Sigma\pi - \underset{(0.150)}{0.670}\ I_{tri} + \underset{(0.122)}{0.647}\ E'_S{}^{c}(R_N) + \underset{(0.159)}{0.454}\ E'_S{}^{c}(R_M) + \underset{(0.122)}{0.322}\ E'_S{}^{c}(R_C) + \underset{(0.098)}{0.345}\ I_Y - \underset{(0.143)}{0.159}\ I_W + \underset{(0.162)}{0.231}\ (I_S + I_T) - \underset{(0.250)}{4.074} \qquad (1\text{-}7)$$

$$n = 59 \quad s = 0.168 \quad r = 0.973 \quad F_{9,49} = 109$$

The quality of the correlation is slightly lower than that of eq. 1-6. The $\Sigma\pi$ and I_{tri} terms and the intercept are similar to those in eq. 1-6. Since their π values were experimentally derived from the log P' values of N-acetylamino acid amides, they were expected to inherently contain components assignable to various intramolecular structural factors as mentioned above. Thus, the size of the regression coefficient of indicator variable terms for tryptophan, methionine, serine, and threonine is much less important than that of the corresponding terms in eq. 1-6. Most of the components assignable to the polar proximity factor and the intramolecular hydrogen-bond formation in the polar side chain of these amino acids seem to be incorporated into the π value of Fauchère and Pliška. The term for tyrosine is more important in eq. 1-7, perhaps due to their unreasonably lower π value of the tyrosine side chain. The coefficient of the $E'_S{}^{c}$ terms in eq. 1-7 is slightly larger than that of the corresponding $E'_S{}^{c}$ terms in eq. 1-6. The steric effect of side chains on the relative solvation of the backbone CONH and $CONH_2$ groupings in the N-acetylamino acid amides is apparently much less important than that on the solvation of zwitterionic NH_3^+ and COO^- groups. In fact, the content of the component assignable to this type of steric factor was not counted in π values of Fauchère and Pliška. Thus, even with π values of Fauchère and Pliška, only our procedure taking care of steric effects of side chains in terms of $E'_S{}^{c}$ resulted in a successful correlation analysis.

Chapter 2 Hydrophobicity of N-Acetyl-Di- and Tripeptide Amides Having Unionizable Side Chains and Correlation with Substituent and Structural Parameters

2-1 Introduction

In this chapter, the P values of N-acetyl-peptide amides of a number of di- and tripeptides with unionized amino acid side chains were analyzed. There are a number of biologically active peptides in which both of N- and C-termini are protected. LHRH (luteinizing hormone-releasing hormone) (Schally *et al.*, 1971), α-MSH (α-melanocyte-stimulating hormone) (Harris and Lerner, 1957), and AKH (adipokinetic hormone) (Stone *et al.*, 1976) are examples where the N-terminus is acylated and the C-terminus is amidated. Moreover, the N-acyl-peptide amides, carrying no electric charge except for that on the ionizable side chains, simulate better than non-protected oligopeptides the partial domain of proteins where each amino acid residue is flanked by two peptide bonds. The relative solubility of protected peptides in 1-octanol was greatly increased compared to the corresponding free peptides. So, it enabled us to measure the P value of peptides containing glutamine and asparagine with fairly hydrophilic side chains more easily and precisely. Then, an empirical correlation equation rationalizing physicochemical factors governing the variations in the log P for the set of protected peptides was formulated by a procedure similar to that used for zwitterionized peptides (Akamatsu *et al.*, 1989).

2-2 Experimental procedure

2-2-1 Measurement of partition coefficient (P) N-acetyl-di- and tripeptide amides

Although N-acetyl-peptide amides used here do not have ionizable groups, the partition coefficient, P, was measured by the flask-shaking method (Fujita *et al.*, 1964) at 25±3°C in 1-octanol and aqueous buffer under identical experimental conditions as those for zwitterionized peptides. The procedure was similar to those in chapter 1. The measurement after HPLC was made at 220 nm using a column of Cosmosil 5C18 (4.6 mm x 25 cm) or YMCA-303 (ODS) (4.6 mm x 25 cm) with water/acetonitrile for elution. The log P value measured in this chapter was from about -2.0 to 0.9. For the measurement of the lowest values, 25 ml of 1-octanol and 2~2.5 ml of the aqueous buffer were used as the partitioning system with 0.2~0.5 mg of each sample. The highest log P values in the above range were measured with 1 ml of 1-octanol and 7~8 ml of the aqueous phases and 0.2~0.5 mg of the samples. The measurement was repeated at least three times for each P value. For the lowest log P values around -2.0, the standard deviation averaged over repeats was ±0.05. As the log P value increased from -2.0 to 0.9, the standard deviation tended to decrease to ±0.01.

2-2-2 Physicochemical substituent parameters

As described in chapter 1, we preliminarily found that the variations in the log P value are also governed at least by the hydrophobic and steric effects of side chain substituents of component amino acids. As the hydrophobic parameter of side chain substituents, the π value for the aliphatic system as that used in chapter 1 was selected (Iwasa *et al.*, 1965). Differing from that defined by Fauchère and Pliška, our π value is "intrinsic" and believed to be free from components such as intramolecular steric

and hydrogen-bonding interactions. The π value of the side chain group of glutamine and asparagine was estimated from experimentally measured log P values of related alkanoyl amides.

For the steric effects of side chain substituents, the $E'_S{}^c$ parameter, the "corrected" Dubois steric parameter was used. The E'_S values of $-CH_2CONH_2$ in glutamine and $-CH_2CH_2CONH_2$ in asparagine were estimated using a highly linear relationship between the E'_S value and Charton's υ steric (Charton, 1983) parameter in a manner similar to that for the indole-3-methyl in tryptophan in chapter 1. The relevant π and $E'_S{}^c$ values are listed in Tables 1-3 and 1-4.

2-3 Results

First, we examined the log P value of protected peptides composed of nonpolar amino acids; glycine, alanine, valine, leucine, isoleucine, and phenylalanine, in terms of the summation of the side chain π value of component amino acids and derived eq. 2-1 with the indicator variable I_{tri}.

$$\log P = 0.834\ \Sigma\pi - 0.551\ I_{tri} - 2.347 \qquad (2\text{-}1)$$
$$\qquad\quad (0.073) \qquad (0.139) \qquad (0.183)$$

$$n = 27 \quad s = 0.154 \quad r = 0.979 \quad F_{2,24} = 283$$

The I_{tri} is zero for dipeptides and one for tripeptides. In chapter 1, we observed (Akamatsu *et al.*, 1989) that "crowded" structures of the α-branched side chain such as those in valine and isoleucine lower the log P value of zwitterionized peptides. This effect was best simulated by the $E'_S{}^c$ value among other steric parameters. In trying to improve the present correlation for protected peptides, we also introduced the steric parameter, $E'_S{}^c$, into eq. 2-1, yielding eq. 2-2, which is of an excellent quality.

$$\log P = 1.077\ \Sigma\pi - 0.538\ I_{tri} + 0.294\ E'_S{}^C(R_N) \quad (0.081)\ (0.097)\ (0.093)$$

$$+ 0.288\ E'_S{}^C(R_M) + 0.243\ E'_S{}^C(R_C) - 2.403 \quad (0.100)\ (0.085)\ (0.111) \qquad (2\text{-}2)$$

$$n = 27 \quad s = 0.088 \quad r = 0.994 \quad F_{5,21} = 361$$

R_N and R_C stand for the side chains of amino acids at the N- and C-termini, respectively. R_M represents the side chain of the central amino acid for tripeptides. For dipeptides lacking R_M, the $E'_S{}^C(R_M)$ term was not counted. Since the slope values for $E'_S{}^C(R_N)$, $E'_S{}^C(R_M)$, and $E'_S{}^C(R_C)$ terms are very close in eq. 2-2, they are combined as $\Sigma E'_S{}^C$ term leading to eq. 2-3.

$$\log P = 1.068\ \Sigma\pi - 0.557\ I_{tri} + 0.268\ \Sigma E'_S{}^C - 2.378 \quad (0.080)\ (0.080)\ (0.078)\ (0.105) \qquad (2\text{-}3)$$

$$n = 27 \quad s = 0.088 \quad r = 0.994 \quad F_{3,23} = 593$$

Equations 2-2 and 2-3 indicate that the steric effect of the side chains on the relative solvation with partitioning solvents is approximately independent from their location in the molecule. The more positive the $E'_S{}^C$ value, i.e., the lower the steric bulk and α-branching effects, the higher the log P value.

When including peptides with polar amino acids, indicator variable terms for the presence of respective polar amino acids were required for rationalizing the variations in log P to give eq. 2-4 of high quality (Y, W, S, T, Q, and N are the one-letter notation for tyrosine, tryptophan, serine, threonine, glutamine, and asparagine, respectively).

Table 2-1. Log P and Physicochemical Parameters of N-Acetyl-di- and tripeptide-amides

No.	Compounds		E'_S[c]			log P	
	Ac-pep-NH_2	$\Sigma\pi$	R_N	R_M	R_C	Obsd.	Calcd.(eq. 2-4)
1	GV	1.27	0.00	0.00	-1.29	-1.33	-1.36
2	AV	1.59	-0.20	0.00	-1.29	-1.13	-1.07
3	LV	3.08	-1.44	0.00	-1.29	0.26	0.19
4	GF	1.95	0.00	0.00	-0.90	-0.56	-0.55
5	IV	3.08	-1.81	0.00	-1.29	0.16	0.11
6	VV	2.54	-1.29	0.00	-1.29	-0.32	-0.33
7	FV	3.22	-0.90	0.00	-1.29	0.43	0.47
8	AL	2.13	-0.20	0.00	-1.44	-0.54	-0.54
9	AA	0.64	-0.20	0.00	-0.20	-2.00	-1.80
10	GL	1.81	0.00	0.00	-1.44	-0.78	-0.83
11	LI	3.62	-1.44	0.00	-1.81	0.68	0.63
12	FG	1.95	-0.90	0.00	0.00	-0.50	-0.55
13	VA	1.59	-1.29	0.00	-0.20	-1.14	-1.07
14	WV	3.19	-0.86	0.00	-1.29	0.73	0.70
15	SV	-0.22	-0.48	0.00	-1.29	-1.53	-1.55
16	SF	0.46	-0.48	0.00	-0.90	-0.79	-0.75
17	TV	0.09	-0.73	0.00	-1.29	-1.25	-1.28
18	TI	0.63	-0.73	0.00	-1.81	-0.86	-0.84
19	YV	2.47	-0.90	0.00	-1.29	-0.20	-0.24
20	YL	3.01	-0.90	0.00	-1.44	0.32	0.29
21	YF	3.15	-0.90	0.00	-0.90	0.54	0.56
22	QV	-0.15	-0.82	0.00	-1.29	-1.85	-1.86
23	QL	0.39	-0.82	0.00	-1.44	-1.32	-1.33
24	QF	0.53	-0.82	0.00	-0.90	-1.14	-1.06
25	FQ	0.53	-0.90	0.00	-0.82	-1.03	-1.06
26	VQ	-0.15	-1.29	0.00	-0.82	-1.82	-1.86
27	IN	-0.15	-1.81	0.00	-0.98	-1.41	-1.43
28	NV	-0.69	-0.98	0.00	-1.29	-1.85	-1.87
29	NI	-0.15	-0.98	0.00	-1.81	-1.43	-1.43
30	NF	-0.01	-0.98	0.00	-0.90	-1.14	-1.07
31	LN	-0.15	-1.44	0.00	-0.98	-1.30	-1.34
32	VAA	1.91	-1.29	-0.20	-0.20	-1.40	-1.35
33	VAV	2.86	-1.29	-0.20	-1.29	-0.67	-0.62

Table 2-1. continued

34	VIG	3.08	-1.29	-1.81	0.00	-0.45	-0.46
35	ALV	3.40	-0.20	-1.44	-1.29	-0.14	-0.09
36	VFA	3.54	-1.29	-0.90	-0.20	0.06	0.18
37	AVI	3.40	-0.20	-1.29	-1.81	-0.20	-0.18
38	IFA	4.08	-1.81	-0.90	-0.20	0.52	0.63
39	GAV	1.59	0.00	-0.20	-1.29	-1.56	-1.64
40	AGF	2.27	-0.20	0.00	-0.90	-0.71	-0.84
41	IAV	3.40	-1.81	-0.20	-1.29	-0.21	-0.18
42	FGL	3.76	-0.90	0.00	-1.44	0.60	0.43
43	FIG	3.76	-0.90	-1.81	0.00	0.34	0.34
44	VVI	4.35	-1.29	-1.29	-1.81	0.49	0.56
45	GLG	1.81	0.00	-1.44	0.00	-1.23	-1.40
46	WAA	2.56	-0.86	-0.20	-0.20	-0.38	-0.31
47	WIG	3.73	-0.86	-1.81	0.00	0.62	0.58
48	WGF	3.87	-0.86	0.00	-0.90	0.99	0.94
49	WAV	3.51	-0.86	-0.20	-1.29	0.36	0.42
50	LSF	2.27	-1.44	-0.48	-0.90	0.23	0.23
51	LTL	2.44	-1.44	-0.73	-1.44	0.24	0.22
52	AYL	3.33	-0.20	-0.90	-1.44	-0.04	0.00
53	AYF	3.47	-0.20	-0.90	-0.90	0.26	0.28

Table 2-2. Squared Correlation (r^2) Matrix for Variables Used in Eq. 2-4.

	$\Sigma\pi$	I_{pep}	$\Sigma E'_s{}^c$	I_Y	I_W	$(I_S + I_T)$	I_Q
I_{pep}	0.380						
$\Sigma E'_s{}^c$	0.159	0.108					
I_Y	0.053	0.000	0.000				
I_W	0.088	0.064	0.005	0.011			
$(I_S + I_T)$	0.080	0.004	0.006	0.013	0.013		
I_Q	0.174	0.074	0.008	0.011	0.011	0.013	
I_N	0.272	0.074	0.010	0.011	0.011	0.013	0.011

Table 2-3. Development of Eq. 2-4.

const.	$\Sigma\pi$	$(I_S + I_T)$	I_{pep}	I_N	I_Q	$\Sigma E'_s{}^c$	I_W	I_Y	r	s	$F_{X,Y}$ [a]
-1.552	0.510								0.889	0.379	$F_{1,51} = 193$
-1.675	0.541	0.503							0.909	0.349	$F_{1,50} = 10.2$
-1.709	0.641	0.589	-0.441						0.932	0.306	$F_{1,49} = 16.3$
-1.990	0.745	0.793	-0.500	0.728					0.956	0.252	$F_{1,48} = 24.5$
-2.437	0.895	1.114	-0.549	1.208	0.791				0.978	0.180	$F_{1,47} = 46.5$
-2.382	1.110	1.566	-0.568	1.926	1.277	0.299			0.994	0.099	$F_{1,46} = 110$
-2.375	1.058	1.490	-0.580	1.782	1.180	0.247	0.239		0.997	0.074	$F_{1,45} = 36.2$
-2.375	1.044	1.476	-0.570	1.753	1.162	0.237	0.258	0.073	0.997	0.072	$F_{1,44} = 3.91$

[a] F statistic for significance of the addition of variables.
X : The number of independent variables added at each step of the development, Y: n-m-1,
m being the total number of independent variables in the developed equation. $F_{1,40,0.05} = 4.085$

$$\log P = 1.044\ \Sigma\pi - 0.570\ I_{tri} + 0.237\ \Sigma E'_s{}^{c}$$
$$(0.047) \qquad (0.054) \qquad (0.046)$$

$$+ 0.073\ I_Y + 0.258\ I_W + 1.476\ (I_S + I_T)$$
$$(0.075) \qquad (0.080) \qquad (0.106)$$

$$+ 1.162\ I_Q + 1.753\ I_N - 2.375 \qquad (2\text{-}4)$$
$$(0.121) \qquad (0.154) \qquad (0.074)$$

$$n = 53 \quad s = 0.072 \quad r = 0.997 \quad F_{8,44} = 840$$

Since the slope values for I_S and I_T terms were very close in the preliminary calculation, they are combined in eq. 2-4. The slope value of I_Y was justified by the t-test at better than the 94.5% level of significance. The corresponding terms and the intercept are practically identical between eqs. 2-3 and 2-4, showing that the separation of indicator variable terms is nearly complete. In Table 2-1, the log P values calculated by eq. 2-4 are compared with those experimentally measured. Table 2-2 shows the degree of independence of variables used in eq. 2-4, and Table 2-3 lists the development of eq. 2-4.

2-4 Discussion

Equation 2-4 shows that the hydrophobicity of protected peptides in terms of log P is governed by factors similar to those of zwitterionized peptides.

In chapter 1, we examined the log P value of zwitterionized peptides at pH 7. The coefficient of the I_{tri} term, -0.570 in eq. 2-4, is very close to that for zwitterionized peptides, -0.635 in eq. 1-6. This means that the log P value decreases by about 0.6 log unit upon introducing one more peptide bond into the dipeptide backbone irrespective of whether the termini are protected or free, other factors being equal. The coefficient of the I_{tri} term corresponds with the increment of the log P value from glycylglycine (GG) to

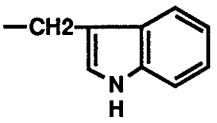
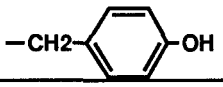
glycylglycylglycine (GGG) for which all independent variable parameters in eqs. 1-6 and 2-4 are zero except for I_{tri} of GGG (= 1). This, in turn, is equivalent with the π value of the CH_3CONH in the peptide backbone.

The coefficient of the I_{tri} term was considerably higher than the "intrinsic" $\pi(CH_3CONH)$ value [-2.17 (Iwasa *et al.*, 1965)] under conditions without any significant stereoelectronic effects. The increase of the $\pi(CH_3CONH)$ value could be rationalized by the "polar proximity" factors (Hansch and Leo, 1979) for the enhancement of hydrophobicity observed when polar groups get closer. The factors could be approximately estimated by counting the interaction between two CONH groups separated by CH_2 as about 1.73 (Hansch and Leo, 1979) which is close to the real situation, the increase being about 1.6 (-0.6 + 2.17). The intercept of eq. 2-4 should correspond with the log P value of Ac-Gly-Gly-NH_2.

In eq. 1-6 for zwitterionized peptides, the steric effect of side chains differs according to their locations. That of the N-terminus is highest. As rationalized in chapter 1, the NH_3^+ group as the strongest hydrogen donor in the molecule is solvated most effectively among other solvation sites. The solvation with 1-octanol of the NH_3^+ group favorable to enhancing the log P value would suffer steric hindrance of the N-terminal side chain substituent to the highest extent. In N-acetyl-peptide amides, it is invariably the CONH group toward which the side chain substituents exert the steric effect on the relative solvation. The fact that the coefficient of the $E'_S{}^c$ terms in eq. 2-2 did not vary much depending upon the side chain location can be understood on this basis.

As observed in eq. 1-6 for zwitterionized peptides, the indicator variable terms were also required to represent specific effects of the side chains of polar amino acid components after their hydrophobic and steric effects were separated in eq. 2-4 for protected peptides. In Table 2-4, the regression coefficient values in eq. 2-4 are compared with those in eq. 1-6. Although the values for protected peptides were slightly lower than those for the free

Table 2-4. Regression Coefficient of Indicator Variable Terms

Amino Acid	Side Chain	Regression Coefficient		n[a]
		Eq. 5	Eq. 6	
Ser	$-CH_2OH$	1.476(0.6~0.9)[b]	1.665(0.8~1.1)[b]	2
Thr	$-CH(CH_3)OH$	1.476(0.6~0.9)[b]	1.665(0.8~1.1)[b]	2
Met	$-CH_2CH_2SCH_3$	-	0.637	3
Asn	$-CH_2CONH_2$	1.753(0.9~1.2)[b]	c	3
Gln	$-CH_2CH_2CONH_2$	1.162(0.3~0.6)[b]	c	4
Trp		0.258	0.352	4
Tyr		0.073	0.165	6

[a] The number of bonds separating the polar hetero atom in the polar group from the a-carbon of the peptides.

[b] The value in parentheses is "corrected" by subtracting the intramolecular bridging-solvation factor.

[c] Not measured because of very low log P values of peptides containing these amino acids.

peptides, the correspondence was very good, the simple correlation coefficient (r) being 0.9999 (n = 4) with the slope of 1.07. This should indicate that factors governing variations in the log P values other than the hydrophobic and steric effects of side chain substituents are almost identical between the two series. In chapter 1, we suggested that the remarkably high coefficient values for Ser and Thr in eq. 1-6 are due to a possible intramolecular bridging-type solvation, the solvent being chelated by the side chain OH and the backbone CONH (Fujita, 1983; Leo, 1983). For Asn and Gln side chains in the protected peptides, the coefficient was also

very high, suggesting that the same type of bridging-solvation is also made between the side chain amide and the backbone CONH. In fact, bridging-hydration has been suggested for the glutamine side chain in the crystal structure of human deoxyhaemoglobin (Fermi *et al.*, 1984; Abraham and Leo, 1987).

The size of the "corrected" regression coefficient subtracted by the value attributed to the bridging solvation (0.6~0.9) tended to decrease with the number of bonds separating the polar heteroatom on the side chain (n in Table 2-4) from the backbone. As in previous discussions for the free-peptides and the $\pi(CH_3CONH)$ value, this effect could be attributed to the polar proximity factor being varied according to the distance (n) between the polar group and the peptide backbone (Hansch and Leo, 1979).

The "corrected" regression coefficient for the Asn and Gln side chains was higher than that expected from those for Ser, Thr and Trp simply on the basis of the distance (n). This indicates that their effect is dependent not only on the proximity polar factor according to the distance between bidirectionally interacting partners but also the stereoelectronic characteristics of functional groups governing the ease of the bridging-type solvation.

Equation 2-5 was derived for the same set of protected peptides as that in eq. 2-4 using side chain π values proposed by Fauchère and Pliška (Fauchère and Pliška, 1983).

$$\begin{aligned} \log P = {} & 1.167\ \Sigma\pi - 0.570\ I_{tri} + 0.332\ \Sigma E'_s{}^{c} + 0.289\ I_Y \\ & (0.054) \quad (0.063) \quad (0.052) \quad (0.084) \\ & - 0.312\ I_W + 0.482\ I_N - 2.340 \\ & (0.107) \quad (0.107) \quad (0.078) \end{aligned} \qquad (2\text{-}5)$$

$$n = 53 \quad s = 0.085 \quad r = 0.995 \quad F_{6,46} = 804$$

The quality of the correlation was slightly lower than that of eq. 2-4. The $\Sigma\pi$ and I_{tri} terms and the intercept were similar to those in eq. 2-4. Since their π values were experimentally derived

from the log P values of N-acetylamino acid amides, they were expected to inherently contain components assignable to various intra-molecular structural factors for polar side chains. The size of indicator variables in eq. 2-5 was, however, not well rationalized physicochemically.

The quality of the correlation for protected peptides of eq. 2-4 was better than that for the free peptides of eq. 1-6. This was probably because the protected peptides are neutral. Without taking care of ionization effects, the accuracy of log P value of protected peptides should be much higher than that of free peptides. In addition, the terminal ionized groups, NH_3^+ and COO^-, in free di- and tripeptides may distort slightly but significantly the electronic environment of the peptide bond and the polar groups of the side chains in a way not completely rationalized by the independent variables considered in eq. 2-4.

Chapter 3 Hydrophobicity of Zwitterionized Di- to Pentapeptides Having Unionizable Side Chains and Correlation with Substituent and Structural Parameters

3-1 Introduction

This chapter deals with attempts made to extend the analysis for di- and tripeptides performed in chapter 1 to tetra- and pentapeptides. There are a number of bioactive tetra- and pentapeptides, for instance, tuftsin (Nishioka *et al.*, 1972), proctoline (Starratt and Brown, 1975), laminin (Graf *et al.*, 1987), and enkephalin (Hughes *et al.*, 1975). In these molecules, such conformational factors as the β-turn formation are expected to contribute to the net molecular hydrophobicity in addition to those taken into account for di- and tripeptides. Introducing a parameter derived from the β-turn "potential" index proposed by Chou and Fasman (Chou and Fasman, 1977), we formulated an empirical equation correlating the log P' value of over 100 di- to pentapeptides on the same standards. The correlation equation should be able to predict the non-measured log P' value of peptides at least up to pentapeptides consisting of amino acids with the unionizable side chain.

3-2 Experimental procedure

3-2-1 Measurement of partition ratio (P') of zwitterionized di- to pentapeptides

The log P' value was measured by the procedure described in chapter 1. The measurement was repeated at least three times for each P' value. For peptides having the most negative log P' values between -2.8 and -2.0, the standard deviation of the log P' value

averaged over repeats was ±0.05. The standard deviation decreased to ±0.02 as the log P' became more positive than -1.0.

3-2-2 Measurement of circular dichroism

A Jasco model J-20 spectropolarimeter equipped with an Ithaco Dynatrac 391A Lock-in Amplifier was used to obtain the CD spectra of the peptides. The ellipticity was calibrated first, using *n*-propylammonium (+)-camphor-10-sulfonate (Gillen and Williams, 1975). The peptide concentration ranged from 0.5 to 1.0 mg/ml and the cell length from 0.2 to 0.5 mm. Molar ellipticity [θ] was expressed in deg·cm^2·dmol^{-1}·residue number^{-1}.

3-2-3 Physicochemical substituent parameters

To represent the hydrophobic and steric effects of side chain substituents, the same parameters as those in chapter 1 and 2 were used. To deal with the conformational effect arisen from the possible β-turn structure for tetra- and pentapeptides, we adopted the β-turn "potential" index for component amino acids proposed by Chou and Fasman (Chou and Fasman, 1977). Their β-turn index is defined for each amino acid in each of the four consecutive positions statistically from the data for 457 β-turned backbone substructures found in the 29 proteins of known sequence and crystallographic structure. As will be shown later, the logarithm of the β-turn index, f, for the i-th amino acid in the four consecutive positions, log f_i, was regarded as being a free-energy-related β-turn potential parameter of each amino acid.

For the inductive electronic parameter of side chain substituents, the Charton σ_I value was used (Charton and Charton, 1983). The relevant parameter sets are listed in Table 3-1. Factors governing the value of log P'(pH 7) were analyzed by the multiple regression technique in terms of the above-mentioned physico-chemical free-energy-related parameters for the side chain

substituents and indicator variables for particular substructures.

Table 3-1. Electronic Parameter and β-turn Potential Indices of Amino Acid Side Chains

Amino Acid	σ_I [a]	log f_i [b]	log f_{i+1} [b]	log f_{i+2} [b]	log f_{i+3} [b]	log f_t [b]
Gly	0.00	0.09	0.00	0.31	0.25	0.19
Ala	-0.01	-0.19	-0.02	-0.37	-0.17	-0.19
Val	0.01	-0.19	-0.26	-0.46	-0.13	-0.28
Leu	-0.01	-0.18	-0.54	-0.40	-0.09	-0.24
Ile	-0.01	-0.17	-0.39	-0.57	-0.14	-0.27
Phe	0.03	-0.07	-0.37	-0.17	-0.07	-0.19
Tyr	0.03	0.03	-0.10	0.09	0.18	0.05
Trp	0.00	-0.10	-0.89	-0.10	0.30	0.00
Met	0.04	-0.07	-0.07	-0.85[c]	-0.14	-0.19
Ser	-0.11	0.14	0.17	0.12	0.03	0.13
Thr	-0.04	0.02	0.09	-0.09	-0.03	-0.00
Pro	-	0.13	0.53	-0.23	-0.11	0.19

a From ref. (Charton and Charton, 1983).

b Calculated from ref. (Chou and Fasman, 1977).

c Not reliable. The corrected value -0.33 was used in eqs. 3-13 and 3-14.

3-3 Results and Discussion

The observed log P'(pH 7) values of tetra- and pentapeptides with physicochemical parameters are shown in Table 3-2, along with those for di- and tripeptides reported in Chapter 1.

Hydrophobicity of Tetra- and Pentapeptides

With the results for di- and tripeptides in mind, we first analyzed the log P'(pH 7) values of tetra- and pentapeptides using

parameter terms corresponding to those used in eq. 1-6, and formulated eq. 3-1.

$$\log P' = \underset{(0.157)}{1.025\ \Sigma\pi} - \underset{(0.226)}{0.262\ I_{pent}} + \underset{(0.205)}{0.575\ E'_S{}^c(R_N)}$$
$$+ \underset{(0.137)}{0.491\ [\Sigma E'_S{}^c(R_M) + E'_S{}^c(R_C)]} + \underset{(0.335)}{0.329\ I_W} + \underset{(0.432)}{0.887\ I_M}$$
$$+ \underset{(0.476)}{1.772\ (I_S + I_T)} - \underset{(0.670)}{4.544} \qquad (3\text{-}1)$$
$$n = 46 \quad s = 0.335 \quad r = 0.926 \quad F_{7,38} = 32.6$$

I_{pent} is an indicator variable taking zero for tetra- and unity for pentapeptides. $\Sigma E'_S{}^c(R_M)$ means the sum of $E'_S{}^c$ parameters for side chains except for those of two terminal amino acids. Preliminary examinations indicated that their steric effect is almost position-independent, so their $E'_S{}^c$ values were added together. The coefficients of $\Sigma E'_S{}^c(R_M)$ and $E'_S{}^c(R_C)$ terms were also so close that these two were combined. Equation 3-1 seems to be acceptable, but the quality of the correlation in terms of r and s is considerably poorer than eq. 1-6. The I_Y term for the tyrosine side chain is insignificant at the 95% level in eq. 3-1. The I_W term for the tryptophan is also only justified over the 94.5% level. Moreover, the coefficient of I_{pent} corresponding to $\Delta\log P$ with introduction of one more peptide unit is significantly more positive than that of I_{tri} in eq. 1-6. The intercept is about 0.6 unit more negative than that in eq. 1-6, reflecting the difference in the reference peptide series between eqs. 1-6 and 3-1: dipeptides and tetrapeptides.

Since physicochemical factors governing the log P' value of lower peptides could at least be involved as factors for tetra- and pentapeptides on the same standards, these discrepancies should indicate that variables other than those used in eqs. 1-6 and 3-1 are required for log P' of tetra- and pentapeptides. We considered that the specific conformational feature such as β-turns could be the factor required for tetra- and penta- but not for di- and tripeptides.

Table 3-2. Log P' and Physicochemical Parameters of Di- to Pentapeptides

No.	Compounds	$\Sigma\pi$	$E'_S{}^c(R_N)$	$\Sigma E'_S{}^c(R_M)$	$E'_S{}^c(R_C)$	log F_{turn}	log f_{i+2}	log P'		
								Obsd.	Calcd.	
									(eq. 3-15)	(eq. 3-16)
1	FL	3.76	-0.90	0.00	-1.44	0.00	0.00	-1.17	-1.23	-1.24
2	LF	3.76	-1.44	0.00	-0.90	0.00	0.00	-1.15	-1.36	-1.38
3	FF	3.90	-0.90	0.00	-0.90	0.00	0.00	-0.85	-0.93	-0.95
4	LL	3.62	-1.44	0.00	-1.44	0.00	0.00	-1.46	-1.66	-1.67
5	LV	3.08	-1.44	0.00	-1.29	0.00	0.00	-2.05	-2.12	-2.13
6	VL	3.08	-1.29	0.00	-1.44	0.00	0.00	-2.07	-2.09	-2.09
7	AI	2.13	-0.20	0.00	-1.81	0.00	0.00	-2.60	-2.50	-2.50
8	II	3.62	-1.81	0.00	-1.81	0.00	0.00	-1.82	-1.98	-1.98
9	LI	3.62	-1.44	0.00	-1.81	0.00	0.00	-1.64	-1.77	-1.78
10	VV	2.54	-1.29	0.00	-1.29	0.00	0.00	-2.82	-2.55	-2.56
11	WW	3.84	-0.86	0.00	-0.86	0.00	0.00	-0.27	-0.21	-0.22
12	WF	3.87	-0.86	0.00	-0.90	0.00	0.00	-0.47	-0.56	-0.58
13	WA	2.24	-0.86	0.00	-0.20	0.00	0.00	-1.98	-1.89	-1.91
14	WL	3.73	-0.86	0.00	-1.44	0.00	0.00	-0.73	-0.86	-0.87
15	WY	3.12	-0.86	0.00	-0.90	0.00	0.00	-1.13	-1.14	-1.14
16	LY	3.01	-1.44	0.00	-0.90	0.00	0.00	-1.94	-1.93	-1.94
17	YL	3.01	-0.90	0.00	-1.44	0.00	0.00	-1.75	-1.80	-1.80
18	VY	2.47	-1.29	0.00	-0.90	0.00	0.00	-2.52	-2.36	-2.37
19	FY	3.15	-0.90	0.00	-0.90	0.00	0.00	-1.68	-1.51	-1.51

Table 3-2. continued

20	YY	2.40	-0.90	0.00	-0.90	0.00	0.00	-1.87	-2.08	-2.08
21	LM	3.12	-1.44	0.00	-1.02	0.00	0.00	-1.87	-2.01	-2.02
22	ML	3.12	-1.02	0.00	-1.44	0.00	0.00	-1.84	-1.91	-1.91
23	MV	2.58	-1.02	0.00	-1.29	0.00	0.00	-2.53	-2.37	-2.37
24	FM	3.26	-0.90	0.00	-1.02	0.00	0.00	-1.59	-1.58	-1.59
25	SL	0.32	-0.48	0.00	-1.44	0.00	0.00	-2.49	-2.66	-2.67
26	PF	2.81	-	0.00	-0.90	0.00	0.00	-2.07	-	-1.95
27	PL	2.67	-	0.00	-1.44	0.00	0.00	-2.41	-	-2.24
28	PI	2.67	-	0.00	-1.81	0.00	0.00	-2.56	-	-2.35
29	FP	2.81	-0.90	0.00	-	0.00	0.00	-1.36	-	-1.37
30	LP	2.67	-1.44	0.00	-	0.00	0.00	-1.76	-	-1.79
31	IP	2.67	-1.81	0.00	-	0.00	0.00	-1.79	-	-1.99
32	FFF	5.85	-0.90	-0.90	-0.90	0.00	0.00	-0.02	0.05	0.04
33	GFF	3.90	0.00	-0.90	-0.90	0.00	0.00	-1.33	-1.29	-1.31
34	FVG	3.22	-0.90	-1.29	0.00	0.00	0.00	-2.33	-2.27	-2.29
35	FVF	5.17	-0.90	-1.29	-0.90	0.00	0.00	-0.76	-0.71	-0.72
36	FVA	3.54	-0.90	-1.29	-0.20	0.00	0.00	-2.19	-2.03	-2.05
37	LVV	4.35	-1.44	-1.29	-1.29	0.00	0.00	-2.10	-1.90	-1.90
38	LII	5.43	-1.44	-1.81	-1.81	0.00	0.00	-1.11	-1.20	-1.19
39	LVL	4.89	-1.44	-1.29	-1.44	0.00	0.00	-1.57	-1.44	-1.43
40	LAL	3.94	-1.44	-0.20	-1.44	0.00	0.00	-2.03	-2.00	-2.01
41	LLL	5.43	-1.44	-1.44	-1.44	0.00	0.00	-0.94	-0.97	-0.97
42	WGG	1.92	-0.86	0.00	0.00	0.00	0.00	-2.72	-2.71	-2.73

Table 3-2. continued

43	WFA	4.19	-0.86	-0.90	-0.20	0.00	0.00	-1.00	-0.90	-0.92
44	WWL	5.65	-0.86	-0.86	-1.44	0.00	0.00	0.36	0.48	0.48
45	LLY	4.82	-1.44	-1.44	-0.90	0.00	0.00	-1.34	-1.25	-1.24
46	VFY	4.42	-1.29	-0.90	-0.90	0.00	0.00	-1.50	-1.38	-1.38
47	GFY	3.15	0.00	-0.90	-0.90	0.00	0.00	-1.96	-1.87	-1.87
48	YLV	4.28	-0.90	-1.44	-1.29	0.00	0.00	-1.45	-1.58	-1.57
49	YVF	4.42	-0.90	-1.29	-0.90	0.00	0.00	-1.37	-1.28	-1.28
50	YGF	3.15	-0.90	0.00	-0.90	0.00	0.00	-1.86	-2.08	-2.09
51	YYL	4.21	-0.90	-0.90	-1.44	0.00	0.00	-1.38	-1.39	-1.38
52	AYI	3.33	-0.20	-0.90	-1.81	0.00	0.00	-2.04	-2.09	-2.08
53	IYV	4.28	-1.81	-0.90	-1.29	0.00	0.00	-1.77	-1.92	-1.91
54	MLF	5.07	-1.02	-1.44	-0.90	0.00	0.00	-1.03	-0.92	-0.92
55	LSL	2.13	-1.44	-0.48	-1.44	0.00	0.00	-2.35	-2.21	-2.21
56	ISL	2.13	-1.81	-0.48	-1.44	0.00	0.00	-2.28	-2.41	-2.41
57	ISI	2.13	-1.81	-0.48	-1.81	0.00	0.00	-2.64	-2.52	-2.52
58	SLI	2.13	-0.48	-1.44	-1.81	0.00	0.00	-1.99	-2.09	-2.08
59	SLL	2.13	-0.48	-1.44	-1.44	0.00	0.00	-2.03	-1.97	-1.97
60	FIT	2.58	-0.90	-1.81	-0.73	0.00	0.00	-1.95	-1.68	-1.68
61	LIT	2.44	-1.44	-1.81	-0.73	0.00	0.00	-2.14	-2.10	-2.10
62	IIT	2.44	-1.81	-1.81	-0.73	0.00	0.00	-2.23	-2.31	-2.31
63	LTI	2.44	-1.44	-0.73	-1.81	0.00	0.00	-2.30	-2.10	-2.10
64	TLI	2.44	-0.73	-1.44	-1.81	0.00	0.00	-1.66	-1.93	-1.93
65	TVL	1.90	-0.73	-1.29	-1.44	0.00	0.00	-1.97	-2.28	-2.28

Table 3-2. continued

66	PLL	4.48	-	-1.44	-1.44	0.00	0.00	-1.64	-	-1.89
67	LPL	4.48	-1.44	-	-1.44	0.00	0.00	-1.56	-	-1.44
68	LLP	4.48	-1.44	-1.44	-	0.00	0.00	-1.58	-	-1.44
69	IPI	4.48	-1.81	-	-1.81	0.00	0.00	-1.65	-	-1.75
70	FGGF	3.90	-0.90	0.00	-0.90	0.18	0.31	-1.51	-1.34	-1.36
71	VAAF	3.86	-1.29	-0.40	-0.90	-0.65	-0.37	-1.91	-2.22	-2.25
72	LLVF	6.84	-1.44	-2.73	-0.90	-1.25	-0.46	-0.25	-0.27	-0.28
73	LLLV	6.70	-1.44	-2.88	-1.29	-1.24	-0.40	-0.51	-0.53	-0.52
74	VGFF	5.17	-1.29	-0.90	-0.90	-0.43	-0.17	-0.51	-0.99	-1.01
75	AVLL	5.21	-0.20	-2.73	-1.44	-0.94	-0.40	-1.74	-1.25	-1.25
76	IAGF	4.08	-1.81	-0.20	-0.90	0.06	0.31	-1.78	-1.73	-1.75
77	FFFF	7.80	-0.90	-1.80	-0.90	-0.68	-0.17	1.63	1.43	1.41
78	LLGF	5.57	-1.44	-1.44	-0.90	-0.48	0.31	-0.42	-0.50	-0.51
79	LLAF	5.89	-1.44	-1.64	-0.90	-1.16	-0.37	-1.00	-0.77	-0.78
80	LLLF	7.38	-1.44	-2.88	-0.90	-1.19	-0.40	0.24	0.23	0.23
81	IIVV	6.16	-1.81	-3.10	-1.29	-1.14	-0.46	-1.41	-1.35	-1.34
82	IIGF	5.57	-1.81	-1.81	-0.90	-0.31	0.31	-0.99	-0.82	-0.82
83	IAAI	4.26	-1.81	-0.40	-1.81	-0.69	-0.37	-2.82	-2.41	-2.42
84	FFGF	5.85	-0.90	-0.90	-0.90	-0.20	0.31	0.17	0.23	0.21
85	VLVL	6.16	-1.29	-2.73	-1.44	-1.28	-0.46	-1.23	-1.00	-1.00
86	WLLV	6.81	-0.86	-2.88	-1.29	-1.16	-0.40	0.23	0.27	0.27
87	WGLL	5.54	-0.86	-1.44	-1.44	-0.58	-0.40	0.06	-0.53	-0.54
88	YILG	4.82	-0.90	-3.25	0.00	-0.51	-0.40	-1.49	-1.59	-1.58

Table 3-2. continued

89	FVYF	6.37	-0.90	-2.19	-0.90	-0.31	0.09	-0.32	0.29	0.30
90	IYIV	6.09	-1.81	-2.71	-1.29	-0.96	-0.57	-1.09	-1.25	-1.24
91	VFLT	3.85	-1.29	-2.34	-0.73	-0.99	-0.40	-1.32	-1.21	-1.22
92	MILI	6.04	-1.02	-3.25	-1.81	-0.99	-0.40	-0.49	-0.54	-0.52
93	VMFI	5.64	-1.29	-1.92	-1.81	-0.58	-0.17	-0.63	-0.49	-0.48
94	PLLL	6.29	-	-2.88	-1.44	-0.89	-0.40	-1.06	-	-1.32
95	LPLL	6.29	-1.44	-	-1.44	-0.14	-0.40	-0.92	-	-0.88
96	LLPL	6.29	-1.44	-	-1.44	-1.04	-0.23	-1.00	-	-0.75
97	LLLP	6.29	-1.44	-2.88	- .	0.00[a]	0.00[a]	-1.18	-	-1.09
98	IPGI	4.48	-1.81	-	-1.81	0.54	0.31	-1.69	-	-1.92
99	VPVL	5.21	-1.29	-	-1.44	-0.21	-0.46	-1.91	-	-1.81
100	VPGV	3.40	-1.29	-	-1.29	0.53	0.31	-2.83	-	-2.50
101	YPGW	3.98	-0.90	-	-0.86	1.17	0.31	-1.25	-	-1.10
102	YPGI	3.87	-0.90	-	-1.81	0.73	0.31	-1.65	-	-1.86
103	GGFVF	5.17	0.00	-2.19	-0.90	-0.21	-0.17	-1.40	-1.26	-1.27
104	VFVGL	6.30	-1.29	-2.19	-1.44	-0.11	0.31	-0.97	-0.70	-0.70
105	VGFVF	6.44	-1.29	-2.19	-0.90	-0.49	-0.17	-0.50	-0.77	-0.78
106	GAALL	4.26	0.00	-1.84	-1.44	-0.39	-0.37	-2.55	-2.32	-2.33
107	AFGVF	5.49	-0.20	-2.19	-0.90	-0.37	0.31	-0.59	-0.70	-0.71
108	AGFVF	5.49	-0.20	-2.19	-0.90	-0.48	-0.17	-1.10	-1.07	-1.08
109	LIIGA	5.75	-1.44	-3.62	-0.20	-0.42	0.31	-1.65	-1.36	-1.36
110	GLLGF	5.57	0.00	-2.88	-0.90	-0.48	0.31	-0.18	-0.73	-0.73
111	ALLGF	5.89	-0.20	-2.88	-0.90	-0.48	0.31	-0.63	-0.54	-0.53

Table 3-2. continued

112	IIIIG	7.24	-1.81	-5.43	0.00	-0.87	-0.57	-0.97	-1.31	-1.30
113	IVVVI	7.43	-1.81	-3.87	-1.81	-1.01	-0.46	-0.89	-1.13	-1.11
114	FGAGI	4.08	-0.90	-0.20	-1.81	0.24	0.31	-1.87	-2.09	-2.10
115	FAAAL	4.72	-0.90	-0.60	-1.44	-0.63	-0.37	-2.23	-2.00	-2.02
116	WGGFV	5.14	-0.86	-0.90	-1.29	0.15	0.31	-0.44	-0.75	-0.75
117	WLFAA	6.32	-0.86	-2.54	-0.20	-0.98	-0.17	-0.32	-0.16	-0.18
118	IAYWG	5.25	-1.81	-1.96	0.00	0.21	0.09	-1.47	-1.12	-1.12
119	GLSVL	3.40	0.00	-3.21	-1.44	-0.45	0.12	-1.64	-1.60	-1.59
120	SLAIV	3.72	-0.48	-3.45	-1.29	-0.89	-0.37	-1.94	-1.96	-1.95
121	YTGFL	3.78	-0.90	-1.63	-1.44	0.36	0.31	-1.18	-0.98	-0.97
122	LVGTF	3.85	-1.44	-2.02	-0.90	-0.16	0.31	-1.18	-1.30	-1.30
123	L-enk[b]	4.96	-0.90	-0.90	-1.44	0.28	0.31	-0.80	-1.22	-1.22
124	M-enk[c]	3.76	-0.90	-0.90	-1.02	0.28	0.31	-1.39	-1.57	-1.57

a These parameter terms were not counted, see text.

b [Leu]enkephalin(YGGFL)

c [Met]enkephalin(YGGFM)

The β-turns, categorized into at least three types, have been observed as regular conformational patterns in regions of backbone chain reversals of globular proteins (Venkatachalam, 1968). The β-turned substructures consist of four consecutive amino acid residues, mostly with hydrogen-bonding formation between the CO-oxygen of the residue at the position i and theNH-hydrogen of the residue at the position (i + 3). One of the β-turn structural types (type I named by Venkatachalam) is shown in Fig. 3-1 (Venkatachalam, 1968).

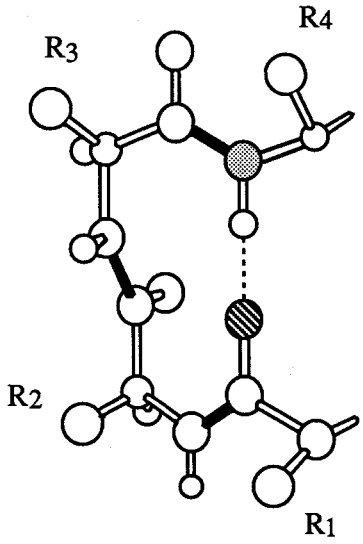


Figure 3-1. β-turn Structure(Type I)
(Dickerson *et al.*, 1971)

We assumed that tetra- and pentapeptides exist as an equilibrium mixture of random and β-turned structures depending upon the β-turn potential of component amino acids in partitioning solvents so that the partition of peptides can be depicted as shown in Fig. 3-2. The net P' value is expressible by eq. 3-2.

$$P' = \frac{[C_{oct}]_R + [C_{oct}]_\beta}{[Cw]_R + [C_w]_\beta} = \frac{[C_{oct}]_R}{[Cw]_R} \left[\frac{1 + K_{oct}}{1 + K_w} \right] \qquad (3\text{-}2)$$

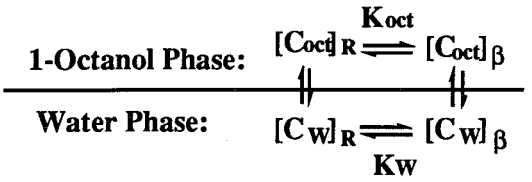


Figure 3-2. Partition and Conformational Equilibria of Peptides; [C] represents the concentration, and suffixes, R and β, express the random and β-turn structure, respectively.

In Fig. 3-2 and eq. 3-2, K_{oct} and K_W are the conformational equilibrium constants in 1-octanol and water phases, respectively, being reflected by the β-turn potential of four consecutive amino acids. Di- and tripeptides are unable to take the β-turn structure so that $K_{oct}=K_W=0$ in eq. 3-2. Thus, eq. 3-3 holds for these lower peptides.

$$\log P' = \log \frac{[C_{oct}]_R}{[C_W]_R} = \log P'_R \qquad (3\text{-}3)$$

P'_R is the P' value for molecules with random structures. It has been shown that, the more hydrophobic the environment, the easier is the intramolecular hydrogen-bond formation (Fujita, 1983) For oligopeptides, the intramolecular hydrogen-bonding could lead to the formation of conformationally fixed structures such as β-turns and α-helices. As will be shown later, the tetra- and pentapeptides studied here were considered to exist almost as the random conformers in the aqueous phase, but to take the β-turn structure in aliphatic alcohols to various extents according to the β-turn potential of component amino acids. Thus, such conditions as $1>>K_W$ and $1<<K_{oct}$ could hold for tetra- and pentapeptides, so that eq. 3-2 reduces to eq. 3-4 from which eq. 3-5 is formulated.

$$P' = P'_R \cdot K_{oct} \qquad (3\text{-}4)$$

$$\log P' = \log P'_R + \log K_{oct} \qquad (3\text{-}5)$$

We examined whether or not the conformational equilibrium constant in the 1-octanol phase, K_{oct}, would be expressible in terms of the Chou-Fasman β-turn parameter, f, of component amino acids. Chou and Fasman have proposed their parameter so as to reflect the "potential" of certain amino acids for the β-turn formation at the four consecutive positions. For each amino acid involved in the β-turn substructures, the f parameter is defined in terms of the relative frequency of occurrence at each bend position among the residues i, i+1, i+2, and i+3. It is normalized by dividing the average frequency of occurence of the amino acid in question involved in 29 proteins. We first tried to use the product of f_i~f_{i+3} values, F_{turn}, as the net potential for the β-turn formation similar to that dealt with by Lewis and coworkers (Lewis *et al.*, 1971). The K_{oct} value was then assumed to be expressible by a linear free-energy-relationship with the β-turn parameters as shown in eq. 3-6.

$$\log K_{oct} = a \log F_{turn} + c \qquad (3\text{-}6)$$

In eq. 3-6, $\log F_{turn} = \Sigma \log f_i (i = i \sim i+3)$, "a" (>0) is the slope and c is the intercept. For pentapeptides where the β-turn formation is possible either with residues 1-4 or with 2-5, we took the greater one of the two F_{turn} values. The original β-turn parameter is proposed in terms of the relative probability of each amino acid participating in the β-turn formation. The use of the logarithm of the Chou-Fasman parameter as being free-energy-related could be justified on this basis. For tetra- and pentapeptides, eq. 3-7 is the counterpart of eq. 3-3.

$$\log P' = \log P'_R + a \log F_{turn} + c \qquad (3\text{-}7)$$

Using the $\log F_{turn}$ as the additional independent variable, eq. 3-8 was formulated for tetra- and pentapeptides.

$$\log P' = 1.056\Sigma\pi - 0.515\ I_{pent} + 0.580\ E'_s{}^{c}(R_N)$$
$$(0.142) \qquad (0.256) \qquad (0.183)$$
$$+ 0.350\ [\Sigma E'_s{}^{c}(R_M) + E'_s{}^{c}(R_C)] + 0.541 \log F_{turn}$$
$$(0.150) \qquad\qquad\qquad (0.335)$$
$$+ 0.363\ I_W + 0.742\ I_M + 1.771\ (I_S + I_T) - 4.740 \qquad (3\text{-}8)$$
$$(0.300) \qquad (0.396) \qquad (0.425) \qquad\quad (0.610)$$
$$n = 46 \quad s = 0.299 \quad r = 0.943 \quad F_{8,37} = 37.2$$

The log F_{turn} term was indeed significant and the statistical quality was much improved from that of eq. 3-1. The corresponding terms between eqs. 1-6 and 3-8 were much closer than those between eqs. 1-6 and 3-1. The log F_{turn} term was positive, showing that the higher the β-turn propensity of component amino acids, the higher the net hydrophobicity as expected from eq. 3-7 where "a" is positive.

The use of the log F_{turn} term was to adopt a model where the β-turn potential of each amino acid at each of the four positions are considered to contribute to the β-turn formation with an equivalent significance *a priori*. We next tested whether or not this model was best by using individual log f_i values as independent variables singly or in various combinations. Interestingly, the single use of log f_{i+2} for the third amino acid residue in place of log F_{turn} was found to be enough, as shown in eq. 3-9. For pentapeptides where there are two choices for the "third" amino acid residues, the higher log f value was used.

$$\log P' = 0.980\Sigma\pi - 0.459\ I_{pent} + 0.539\ E'_s{}^{c}(R_N)$$
$$(0.136) \qquad (0.219) \qquad (0.176)$$
$$+ 0.350\ [\Sigma E'_s{}^{c}(R_M) + E'_s{}^{c}(R_C)] + 0.677 \log f_{i+2}$$
$$(0.137) \qquad\qquad\qquad (0.345)$$
$$+ 0.422\ I_W + 0.769\ I_M + 1.619\ (I_S + I_T) - 4.609 \qquad (3\text{-}9)$$
$$(0.291) \qquad (0.375) \qquad (0.414) \qquad (0.573)$$
$$n = 46 \quad s = 0.286 \quad r = 0.948 \quad F_{8,37} = 41.1$$

Not only the corresponding terms, except for the pair of log F_{turn} and log f_{i+2} terms, but also the correlation quality is practically equivalent between eqs. 3-8 and 3-9. This was thought to be due to a high collinearity (r = 0.812) between log F_{turn} and log f_{i+2} values for 46 tetra- and pentapeptides. Although the r and s terms are nearly alike in both, eq. 3-9 is preferred over eq. 3-8, because the conformational parameter in the latter, log F, actually consists of four terms as opposed to the single term log f_{i+2}, in eq. 3-9. Equation 3-9 indicates that the ease of the β-turn formation is most significantly governed by the β-turn potential of the third residue among four consecutively linked amino acids.

Besides the $f_i(i = i\sim i+3)$ parameters for each amino acid at each of the four bend positions, Chou and Fasman have estimated the relative frequency of occurrence of each amino acid in all four bend positions, f_t, based on 457 β-turns in 29 proteins (Chou and Fasman, 1977). We examined correlations of log f_t derived from their study (Table 3-1) with each of the log $f_i(i = i\sim i+3)$ values. For the set of ten component amino acids (omitting methionine) in peptides included in eqs. 3-8 and 3-9, eq. 3-10 formulated for the log f_{i+2} value showed the best quality.

$$\log f_t = \underset{(0.137)}{0.588} \log f_{i+2} + \underset{(0.043)}{0.015} \qquad (3\text{-}10)$$

$$n = 10 \quad s = 0.051 \quad r = 0.962 \quad F_{1,8} = 97.9$$

Equation 3-11, formulated for the log f_i value for the first residue followed eq. 3-10.

$$\log f_t = \underset{(0.445)}{1.319} \log f_i + \underset{(0.059)}{0.006} \qquad (3\text{-}11)$$

$$n = 10 \quad s = 0.071 \quad r = 0.924 \quad F_{1,8} = 46.6$$

Neither log f_{i+1} nor log f_{i+3} value was able to explain the variance in log f_t over 50% (100 x r^2). The f_{i+2} value of methionine is estimated in the original work (Chou and Fasman, 1977), based on only a single occurrence at the bend position, i+2, so that it is not as reliable as that for other residues. Taking the cyclic structure, the conformational effect of the proline residue could differ from that of the other amino acids. For the set of eighteen amino acid residues deleting proline and methionine from the original data of Chou and Fasman (Chou and Fasman, 1977), eq. 3-12 was obtained.

$$\log f_t = \underset{(0.123)}{0.504}\ \log f_{i+2} + \underset{(0.034)}{0.003} \qquad (3\text{-}12)$$

$$n = 18 \quad s = 0.066 \quad r = 0.909 \quad F_{1,16} = 75.7$$

The f_t value was reasonably considered to represent the ease of participation of a certain amino acid residue in the β-turn formation within conformations of natural globular proteins. "Linear free-energy-relationships" as shown in eqs. 3-10 and 3-12 for the f_{i+2} with the f_t value, reflecting an overall "standard" potential for the β-turn formation, were considered to be a background for the formulation of eq. 3-9 where only the log f_{i+2} term suffices for rationalizing the log P' value of tetra- and pentapeptides.

As could be understood from Fig. 3-1, the side chain of the residue, i+2, would exert a significant effect on the torsion angle of the adjacent CONH plane sterically. In fact, we formulated eq. 3-13 for 10 amino acid side chains. The alanine side chain (Me) was deleted but the methionine side chain was included after its log f_{i+2} value was corrected by eq. 3-12.

$$\log f_{i+2} = \underset{(0.163)}{0.522}\ E'_S{}^{c} + \underset{(0.172)}{0.334} \qquad (3\text{-}13)$$

$$n = 10 \quad s = 0.107 \quad r = 0.934 \quad F_{1,8} = 54.7$$

While the β-turn propensity was understood to be governed by the steric effect of side chains in terms of $E'_S{}^c$, no reasonable explanation was given why the alanine side chain is an outlier. Including the alanine side chain and using E'_S instead of $E'_S{}^c$, eq. 3-14 was formulated.

$$\log f_{i+2} = \underset{(0.213)}{0.345}\ E'_S + \underset{(3.441)}{2.461}\ \sigma_I + \underset{(0.334)}{0.246} \qquad (3\text{-}14)$$

$$n = 11 \quad s = 0.166 \quad r = 0.842 \quad F_{2,8} = 9.75$$

The σ_I is a parameter for the electron-withdrawing property of aliphatic substituents defined by Charton and Charton (Charton and Charton, 1983). The σ_I term is significant only at the 85% level. In spite of this, eqs. 3-13 and 3-14 were considered to support the above view. In eq. 3-14, the E'_S worked much better than $E'_S{}^c$. This could mean that the steric effect operating here is similar to that in the reference aliphatic ester system from which Taft E_S is defined (Taft, 1965; see chapter 1-2-3). The physicochemical significance of the $\log f_{i+2}$ term in eq. 3-9 is perhaps to represent the steric effect of the side chain of the residue i+2 on the twisting of the adjacent CONH group. The bulkier the side chain substituent, the greater would be the twisting so that the direction of the NH group of the residue i+3 as the hydrogen-donor toward the CO group of the residue i as the acceptor is distorted more severely. The positive σ_I term would indicate that the higher the electron-attracting ability of the side chain, the greater the acidity of the NH hydrogen leading to a higher hydrogen-donating property.

The most significant driving force for the β-turn formation could be a gain of stabilization energy by intramolecular hydrogen-bond formation. The significant correlation of $\log f_t$ with $\log f_i$ as shown in eq. 3-11 is also taken to support the above discussion. The carbonyl group of the first residue is the hydrogen-bond acceptor and the relative probability of each residue being found at this

position should be related to that in finding among β-turn substructures. Eqs. 3-10~3-14 showing that the stabilization of the β-turned structure is largely dependent on the steric effect of side chains of amino acids involved are in accord with the result of Charton and Charton (Charton and Charton, 1983) analyzed from somewhat different points of view.

Hydrophobicity of Di- to Pentapeptides

To correlate the log P' values for di- to pentapeptides as a set, eqs. 1-6 and 3-9 were combined together using two indicator variables to give eq. 3-15. The one is I_{turn} which takes zero for di- and tripeptides and unity for tetra- and pentapeptides. The addition of the I_{turn} term corresponds with the incorporation of the intercept "c" needed only for tetra- and pentapeptides in eq. 3-7 after log F_{turn} is replaced by log f_{i+2}. The other, I_{pep}, is the combined parameter of I_{tri} and I_{pent} which takes zero for dipeptides and one, two, and three with ascending numbers of peptide bonds.

$$\begin{aligned}\log P' = {} & 0.943\ \Sigma\pi - 0.579\ I_{pep} + 0.550\ E'_S{}^{c}(R_N) \\ & (0.069) \quad (0.105) \quad (0.095) \\ & + 0.307\ [\Sigma E'_S{}^{c}(R_M) + E'_S{}^{c}(R_C)] + 0.521\ I_{turn} \\ & (0.077) \quad (0.206) \\ & + 0.747\ \log f_{i+2} + 0.135\ I_Y + 0.375\ I_W + 0.654\ I_M \\ & (0.231) \quad (0.094) \quad (0.113) \quad (0.170) \\ & + 1.584\ (I_S + I_T) - 3.838 \qquad (3\text{-}15) \\ & (0.207) \quad (0.204)\end{aligned}$$

$$n = 105 \quad s = 0.212 \quad r = 0.969 \quad F_{10,94} = 144$$

The correspondence of eq. 3-15 for 105 peptides with eq. 1-6 for lower as well as with eq. 3-9 for higher peptides is very good, supporting the procedure with assumptions made for eqs. 3-2, 3-3, 3-5, 3-6, and 3-7 using the Chou and Fasman β-turn parameter for

Table 3-3. Squared Correlation (r^2) Matrix for Variables Used in Eq. 3-15.

	$\Sigma\pi$	I_{pep}	$E'_S{}^C(R_N)$	$\Sigma E'_S{}^C(R_M, R_C)$	I_{turn}	log f_{i+2}	I_Y	I_W	I_M
I_{pep}	0.395								
$E'_S{}^C(R_N)$	0.020	0.011							
$\Sigma E'_S{}^C(R_M, R_C)$	0.376	0.438	0.000						
I_{turn}	0.465	0.783	0.000	0.353					
log f_{i+2}	0.093	0.005	0.020	0.167	0.043				
I_Y	0.009	0.013	0.001	0.028	0.021	0.011			
I_W	0.003	0.009	0.009	0.051	0.006	0.000	0.005		
I_M	0.013	0.012	0.000	0.001	0.001	0.000	0.004	0.010	
$(I_S + I_T)$	0.231	0.001	0.000	0.030	0.016	0.006	0.026	0.025	0.016

Table 3-4. Development of Eq. 3-15.

const.	$\Sigma\pi$	I_W	log f_{i+2}	I_{pep}	$I_S + I_T$	$E'_s{}^c(R_N)$	$\Sigma E'_s{}^c(R_M, R_C)$	I_M	I_{turn}	I_Y	r	s	$F_{X,Y}$ [a]
-3.109	0.409										0.771	0.521	$F_{1,103}$=151
-3.156	0.401	0.560									0.819	0.472	$F_{1,102}$=23.7
-3.257	0.432	0.552	0.689								0.838	0.450	$F_{1,101}$=11.0
-3.352	0.517	0.490	0.810	-0.181							0.858	0.427	$F_{1,100}$=12.3
-3.891	0.663	0.527	0.981	-0.314	0.688						0.890	0.381	$F_{1,99}$=26.4
-3.657	0.734	0.462	0.960	-0.401	0.819	0.404					0.918	0.333	$F_{1,98}$=31.7
-3.666	0.876	0.324	0.635	-0.358	1.232	0.490	0.263				0.935	0.300	$F_{1,97}$=25.4
-3.883	0.958	0.338	0.590	-0.364	1.501	0.532	0.337	0.684			0.958	0.242	$F_{1,96}$=51.6
-3.754	0.930	0.357	0.757	-0.564	1.516	0.546	0.304	0.621	0.490		0.966	0.220	$F_{1,95}$=21.1
-3.838	0.943	0.375	0.747	-0.579	1.584	0.550	0.307	0.654	0.521	0.135	0.969	0.212	$F_{1,94}$=8.05

[a] F statistic for significance of the addition of variables.
X : The number of independent variables added at each step of the development, Y: n-m-1,
m being the total number of independent variables in the developed equation. $F_{1,120,0.05} = 3.92$

the conformational effect in higher peptides. The log K_{oct} value was estimated by substituting values for the log f_{i+2} and I_{turn} terms into the corrected eq. 3-6. It ranged between zero and 0.75, however. The value was found not entirely in accord with the conditions of $K_{oct}>>1$ for eq. 3-4 but the procedure was admissible at least as a first approximation. In Table 3-2, the calculated log P' value using eq. 3-15 is shown for 105 peptides. In Tables 3-3 and 3-4, the degree of independence of variables used in eq. 3-15 and the development of eq. 3-15, are listed, respectively.

Hydrophobicity of Peptides Including Pro Residue

We did not include peptides containing proline so far, since the $E'_S{}^c$ value for the "side chain" of proline was not easily estimated. By substituting the values of available parameters for peptides including proline such as $\Sigma\pi$, I_{pep}, log f_{i+2}, and I_{turn} into eq. 3-15, we calculated the summation of these parameter terms and examined the difference, Δlog P', from the observed value. The Δlog P' value should correspond with the component of log P' value attributable to the steric effect together with that specific to the Pro residue.

As shown in Table 3-5, the effects seem dependent not only on the location but also on the number of residues involved. When the Pro residue is at the N-terminus, the Δlog P' value is invariably negative, being -0.5 ~ -0.9. At the C-terminus, however, it shows the reverse effect only in dipeptides. For tripeptides without N-terminal proline, the Δlog P' is nearly zero. For tetrapeptides, the Δlog P' is always negative. We considered that the effect of the Pro residue at positions other than the N-terminus is to lower the log P' value almost regularly with increasing numbers of total residues from dipeptides regardless of its location.

Although the variation patterns of the Δlog P' value looked rather complex, we assumed that they are represented by two indicator variables. The one is for the effect when at the N-terminus, $I_P(N)$, and the other is for the effect of the number of

Table 3-5. Δlog P' and Indicator Variables of Peptides Including Pro Residue

Compounds	Δlog P'	$I_P(N)$	I_P(#pep)
PI	-0.683	1	-1
PL	-0.648	1	-1
PF	-0.605	1	-1
FP	0.325	0	-1
IP	0.531	0	-1
LP	0.354	0	-1
IPI	0.090	0	0
PLL	-0.562	1	0
LPL	-0.128	0	0
LLP	-0.148	0	0
PLLL	-0.888	1	1
LPLL	-0.407	0	1
LLPL	-0.607	0	1
LLLP	-0.437	0	1
IPGI	-0.123	0	1
VPGV	-0.688	0	1
VPVL	-0.457	0	1
YPGW	-0.509	0	1
YPGI	-0.140	0	1

residues, I_P(#pep). The values of these indicator variables were set as zero for tripeptides without the N-terminal proline, since their Δlog P' value is closest to zero. The values of indicator variables are also shown in Table 3-5. With these two additional indicator variable terms for the Pro residue, eq. 3-16 was finally formulated for 124 peptides without any decrease in the correlation quality.

$$\log P' = 0.942\ \Sigma\pi - 0.582\ I_{pep} + 0.546\ E'_S{}^{c}(R_N)$$
$$(0.064) \qquad (0.096) \qquad (0.089)$$

$$+ 0.295\ [\Sigma E'_S{}^{c}(R_M) + E'_S{}^{c}(R_C)] + 0.516\ I_{turn}$$
$$(0.071) \qquad\qquad (0.172)$$

$$+ 0.764 \log f_{i+2} + 0.144\ I_Y + 0.378\ I_W + 0.659\ I_M$$
$$(0.211) \qquad (0.089) \qquad (0.106) \qquad (0.165)$$

$$+ 1.581\ (I_S + I_T) - 0.807\ I_P(N) - 0.346\ I_P(\#pep)$$
$$(0.197) \qquad (0.225) \qquad (0.118)$$

$$- 3.866 \qquad\qquad (3\text{-}16)$$
$$(0.190)$$

$$n = 124 \quad s = 0.209 \quad r = 0.967 \quad F_{12,111} = 134$$

In Table 3-2, the calculated log P' with eq. 3-16 is listed. For Leu-Leu-Leu-Pro, where no β-turn formation with intramolecular hydrogen bonding is possible, the I_{turn} and log f_{i+2} terms were not counted to give the calculated log P' value. Tables 3-6 and 3-7 shows the degree of independence of variables used in eq. 3-16 and the development of eq. 3-16, respectively.

At the protonated amino group of the N-terminus working as the hydrogen donor, the solvation with the more basic 1-octanol could effectively compete with that with the less basic water. Since the number of polarized N^+-H bonds in peptides including proline is lower by unity than that in others without cyclic amino acids at the N-terminus, the solvation with 1-octanol is less significant in peptides including proline than that in other regular peptides, leading to the lower log P' value. The slope of the $I_P(N)$ term, -0.81, was in the same order as that previously observed (-0.52) for the effect of the decrease in the number of N^+-H bonds on the ion-pair formation-partition equilibrium for various aliphatic ammonium ions and picrate in the 1-octanol/water system (Takayama *et al.*, 1985).

At positions other than the N-terminus, one of amide NH sites working as the hydrogen-donor is reduced by replacing the regular primary amino acid residue with proline. By the same token as that

Table 3-6. Squared Correlation (r^2) Matrix for Variables Used in Eq. 3-16.

	$\Sigma\pi$	I_{pep}	$E'_S{}^c(R_N)$	$\Sigma E'_S{}^c(R_M, R_C)$	I_{turn}	log f_{i+2}	I_Y	I_W	I_M	$(I_S + I_T)$	$I_P(N)$
I_{pep}	0.415										
$E'_S{}^c(R_N)$	0.024	0.001									
$\Sigma E'_S{}^c(R_M, R_C)$	0.398	0.437	0.002								
I_{turn}	0.428	0.762	0.000	0.316							
log f_{i+2}	0.111	0.005	0.009	0.183	0.034						
I_Y	0.009	0.004	0.001	0.019	0.007	0.020					
I_W	0.002	0.004	0.007	0.039	0.002	0.001	0.000				
I_M	0.011	0.009	0.000	0.000	0.001	0.000	0.002	0.008			
$(I_S + I_T)$	0.197	0.001	0.000	0.036	0.013	0.004	0.018	0.018	0.011		
$I_P(N)$	0.004	0.023	0.173	0.000	0.009	0.002	0.009	0.005	0.003	0.007	
$I_P(\#pep)$	0.070	0.096	0.021	0.043	0.098	0.000	0.005	0.001	0.000	0.001	0.063

Table 3-7. Development of Eq. 3-16.

const.	$\Sigma\pi$	I_W	log f_{i+2}	I_{pep}	$I_S + I_T$	$E'_s{}^c(R_N)$	$\Sigma E'_s{}^c(R_M, R_C)$	I_M	$I_P(N)$	I_{turn}	$I_P(\#pep)$	I_Y	r	s	$F_{X,Y}$ [a]
-3.086	0.393												0.754	0.515	$F_{1,122}$=161
-3.132	0.386	0.582											0.806	0.466	$F_{1,121}$=27.9
-3.230	0.414	0.567	0.565										0.820	0.453	$F_{1,120}$=8.51
-3.348	0.503	0.519	0.708	-0.182									0.840	0.431	$F_{1,119}$=13.3
-3.865	0.648	0.553	0.916	-0.320	0.709								0.874	0.388	$F_{1,118}$=28.9
-3.715	0.696	0.515	0.940	-0.371	0.798	0.278							0.892	0.363	$F_{1,117}$=17.7
-3.738	0.837	0.388	0.640	-0.326	1.222	0.381	0.255						0.910	0.334	$F_{1,116}$=22.1
-3.933	0.920	0.394	0.604	-0.333	1.489	0.430	0.334	0.725					0.935	0.287	$F_{1,115}$=41.8
-3.783	0.930	0.349	0.593	-0.372	1.463	0.545	0.324	0.680	-0.677				0.946	0.262	$F_{1,114}$=24.0
-3.704	0.920	0.355	0.743	-0.572	1.510	0.563	0.296	0.630	-0.739	0.449			0.955	0.243	$F_{1,113}$=20.3
-3.769	0.924	0.370	0.788	-0.573	1.506	0.549	0.286	0.624	-0.851	0.500	-0.328		0.964	0.218	$F_{1,112}$=28.4
-3.866	0.942	0.378	0.764	-0.582	1.581	0.546	0.295	0.659	-0.807	0.516	-0.346	0.144	0.967	0.209	$F_{1,111}$=10.4

a F statistic for significance of the addition of variables.
X : The number of independent variables added at each step of the development, Y: n-m-1,
m being the total number of independent variables in the developed equation. $F_{1,120,0.05} = 3.92$

for the N-terminal N^+-H sites, the reduction of the NH sites would induce the reduction of log P'. On the other hand, the steric inhibition effect of the "side chain" of the Pro residue on the hydrogen-bonding solvation of neighboring CONH or COO^- group could be lowered by the cyclization. This steric effect would be favorable to the solvation of the bulkier 1-octanol leading to the augmentation of log P'. For tripeptides, these two oppositely operating factors may be balanced. The positive effect is predominant for dipeptides, but the negative effect overweights for higher peptides gradually with increases in the number of residues. No theoretical rationalization for variations in the balance between two opposite factors is available at the moment. The measurement of the log P' value for more peptides including proline at various positions is needed before drawing definite conclusions.

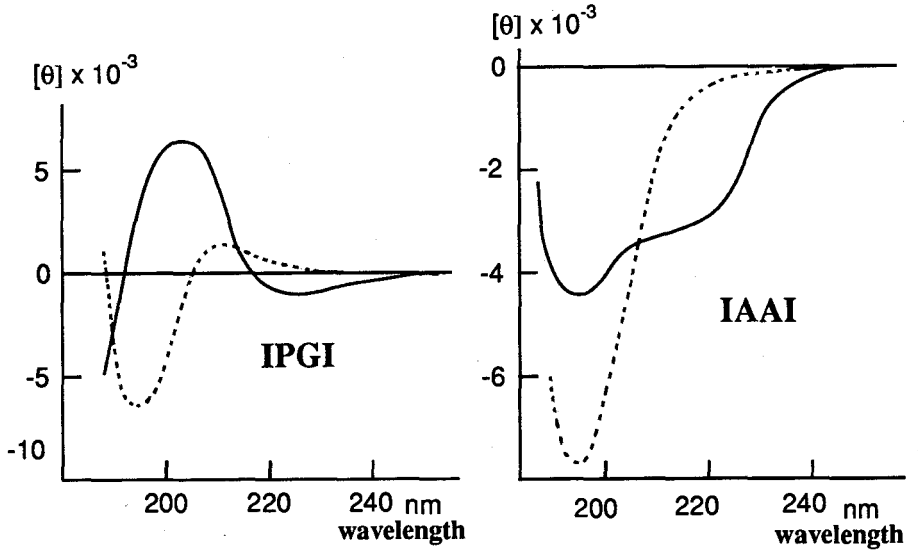


Figure 4. CD Spectra of Ile-Pro-Gly-Ile and Ile-Ala-Ala-Ile.
(——) in TFE, (---) in aqueous 0.1 M sodium phosphate/phosphoric acid

CD Spectra of Higher Peptides

To examine the β-turn formation potential, we measured CD spectra of some tetra- and pentapeptides in aqueous 0.1 M sodium phosphate/phosphoric acid buffer (pH 7) and TFE. The spectra of Ile-Pro-Gly-Ile and Ile-Ala-Ala-Ile in these two phases are shown in Fig. 3-3. In the spectrum of Ile-Pro-Gly-Ile in TFE, a positive dichroic band occurs at about 203 nm. This shows the characteristic pattern for the type II β-turn structure (Bush *et al.*, 1978). In the aqueous phase, however, a negative band shows up at about 187 nm which is attributed to the disordered random structure (Townend *et al.*, 1966). For Ile-Ala-Ala-Ile, a negative band is observed in both phases at 196 nm for the disordered structure, but the spectrum in TFE differs from that in aqueous solution showing a shoulder at 213 nm. This could be due to the fact that Ile-Ala-Ala-Ile exists as a mixture of disordered and β-turned conformers. The log f_{i+2} (for Gly: 0.31) and log F_{turn} (0.54) values for Ile-Pro-Gly-Ile are much higher than the log f_{i+2} (for Ala: -0.37) and log F_{turn} (-0.69) values for Ile-Ala-Ala-Ile, respectively. This explains the difference in CD spectral patterns between the two tetrapeptides, and gives the experimental support to this procedure for analysis.

Chapter 4 Application

4-1 Quantitative structure-activity relationships of the bitter thresholds of peptides

4-1-1 Introduction

Many dipeptides and tripeptides are bitter. There is no simple correspondence for the tastes of component amino acids; for example, peptides D-Leu-Gly and D-Leu-D-Leu that contain sweet amino acids are bitter (Wieser and Belitz, 1976). These complex features have made it difficult to obtain an overall view of their structure-activity relationships.

The state of structure-activity relationship studies of bitter compounds has been summarized by Belitz *et al* (Belitz *et al.*, 1983). Based on data already reported, what we can say about the structural characteristics of bitter compounds is that there is always a polar function and a hydrophobic group within the molecule, the former probably affecting taste quality and the latter affecting taste intensity. Since the hydrophobic moieties are sterically various, the participation of steric factors has been also suggested. To obtain more information, a quantitative approach may be of use. For derivatives of amino acids and peptides, Gardner has investigated the relationship between the bitter thresholds and molecular connectivity, finding a significant correlation with the first-order-valence correlated index $^{1}\chi^{v}$ (Gardner, 1980). This correlates with the partition coefficients of a wide range of compounds, so he suggested that the result is a reflection of the influence of hydrophobicity on bitterness.

In this chapter, we applied our measured and calculated log P' values to analyze the structure-bitterness relationships of di- and tripeptides. The log P' value and a molecular-dimensional parameter were found to be of significance in determining the bitter

threshold of the peptides.

4-1-2 Experimental procedure

Test compounds

Compounds were purchased from the Sigma Chemical Co., Wako Pure Chemical Industries, Ltd., and Tokyo Kasei Kogyo Co. Ltd.

Bitter thresholds

The preparation of the test solution and the methods of evaluation were essentially the same as those reported by Wieser and Belitz (Wieser and Belitz, 1975; Wieser and Belitz, 1976). Briefly, a series of solutions of increasing concentration in which one solution was twice as strong as the preceding one was prepared, and a 2- to 3 ml portion of each was tested by each person. The panel consisted of six persons and the threshold concentration, T (in M), was recorded. When the compound was a hydrochloride, the solution was neutralized with 0.1 N NaOH. The standard error of the determination was within ±15%.

4-1-3 Analysis of bitter thresholds

The threshold data of compounds were mainly from literature reported by Wieser and Belitz (Wieser and Belitz, 1976). Our own data for compounds Leu-Phe and Ala-Phe, the bitterness of which was not found by previous workers, were included in the analysis. The correspondence between the reported and our values for other peptides were excellent. The T value in the dependent variable, log (1/T), is either the center value of the reported range or the average of our determinations of the threshold concentration The log (1/T) data are summarized in Table 4-1.

Table 4-1. Bitterness Threshold and Physicochemical Parameters of Peptides

No.	Compounds	log P' [a] (pH 7)	D/10	log (1/T)	
				Obsd.	Calcd.(eq. 4-1)
1	FL	-1.17*	1.61	2.87	2.78
2	LF	-1.15*	1.61	2.75	2.79
3	FF	-0.85*	1.72	3.10	3.07
4	LL	-1.46*	1.49	2.35	2.51
5	VL	-2.07*	1.37	2.00	2.04
6	II	-1.82*	1.49	2.26	2.29
7	VV	-2.82*	1.24	1.71	1.48
8	AL	-2.35	1.24	1.70	1.77
9	AF	-2.08	1.36	1.72	2.03
10	AV	-2.83	1.12	1.16	1.37
11	GI	-2.64	1.23	1.70	1.58
12	GL	-2.54	1.23	1.68	1.64
13	GF	-2.27	1.35	1.80	1.90
14	GV	-3.02	1.10	1.13	1.25
15	IA	-2.93	1.30	1.68	1.46
16	IG	-3.19	1.30	1.68	1.31
17	IL	-1.82	1.49	2.26	2.29
18	IV	-2.30	1.37	2.05	1.90
19	LA	-2.73	1.30	1.72	1.59
20	LG	-2.98	1.30	1.72	1.43
21	FG	-2.55	1.41	1.77	1.79
22	VA	-3.16	1.18	1.16	1.22
23	VG	-3.42	1.18	1.19	1.07
24	WW	-0.27*	2.08	3.60	3.71
25	GW	-1.94	1.52	1.89	2.24
26	IW	-1.21	1.78	3.05	2.90
27	LW	-1.01	1.78	3.40	3.02
28	YL	-1.75*	1.68	2.40	2.49

Table 4-1. continued

29	LY	-1.94*	1.71	2.46	2.40
30	GY	-2.83	1.45	1.77	1.65
31	FY	-1.68*	1.82	3.13	2.66
32	IS	-3.08	1.35	1.49	1.42
33	SL	-2.49*	1.35	1.49	1.77
34	IT	-2.84	1.37	1.49	1.57
35	LVL	-1.57*	1.86	2.70	2.75
36	LLL	-0.94*	1.86	2.87	3.12
37	GGL	-3.18	1.59	1.13	1.57
38	LGG	-3.62	1.66	1.13	1.36
39	LGL	-2.25	1.86	2.26	2.34
40	GLY	-2.21	1.81	2.52	2.33

[a] Asterisked values were experimentally measured and the others were calculated from eq. 1-6.

To best express the steric features of the molecule, we first defined the length parameter D (Nakayama *et al.*, 1984), which corresponds to the maximum length in Å of a molecule in the fully extended conformation in which the zigzag peptide backbone extends straight. It is measured as the length along the D-axis that runs in the middle of the two straight lines; one passes through the two amido nitrogen atoms five atoms apart from each other in peptides with three or more amino acids and the other passes through the corresponding amide carbon atoms (Fig. 4-1A). The main chain was constructed so as to give the longest D. It follows that the side chains rather than the amino or hydroxycarbonyl groups of the terminal amino acids are incorporated into the main chain. When the C-terminal is glycine or alanine, however, the carboxylic acid moiety constitutes the main chain, since it is longer than the side groups H and Me. Similarly, when glycine is at the N-end, the main

groups H and Me. Similarly, when glycine is at the N-end, the main chain comes to include the amino group. By this definition, the angle between the D-axis and the bond-axis that links the terminal group at the C-end to their connecting atom becomes 39.2°, and that at the N-terminal becomes 31.3°. For compounds smaller or shorter than tripeptides, i.e. dipeptide and amino acid derivatives, the D-axis was drawn according to this criterion. These situations are explained schematically in Fig. 4-1. The steric parameter values were calculated based on the CPK model by use of a computer program modified from that originally made by Verloop *et al* (Verloop *et al.*, 1976) for the estimation of the STERIMOL parameters.

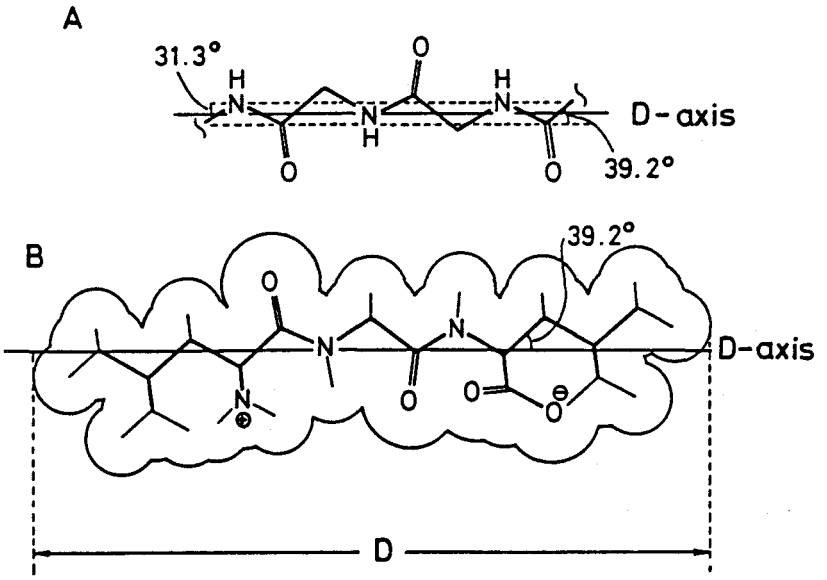


Figure 4-1. Definition of the length parameter D: A, schematic representation of the D axis that runs at the middle of the two dotted lines, one passing through the nitrogen atoms and the other passing through the carbonyl carbon atoms of the two peptie bonds that are one amino acid unit apart from each other; B, the D parmeters of Leu-Gly-Leu.

4-1-4 Results

We analyzed the bitterness threshold of 40 di- and tripeptides listed in Table 4-1 quantitatively using a dimensional parameter D defined to represent the maximum length of the molecule and the log P'(pH 7) value either experimentally measured or estimated by eq. 1-6, yielding eq. 4-1.

$$\log (1/T) = \underset{(0.121)}{0.597} \log P'(pH\ 7) + \underset{(0.394)}{0.845} (D/10) + \underset{(0.803)}{2.117} \qquad (4\text{-}1)$$

$$n = 40 \quad s = 0.208 \quad r = 0.954 \quad F_{2,37} = 185$$

The D value was scaled by 0.1 in the regressions to make the size comparable to that of the log P'. Equation 4-1 shows that the more hydrophobic and the longer the molecule, the lower the bitterness threshold. The log P'(pH 7) value was measured for less than half the total compounds, but was calculated by eq. 1-6 for the rest. The fact that the log P'(pH 7) values calculated by eq. 1-6 for unmeasured peptides gave a good quality correlation in eq. 4-1 seems to indicate a highly reliable predictability of eq. 1-6 and to support the procedures used for its development.

4-1-5 Discussion

The total length D of the molecule in the extended conformation was found to be an important factor that governs the bitterness of peptides. The conformation at the site of action, or the active conformation, may, however, not necessarily be the extended one for compounds with flexible skeletal structures like those studied here. When the coefficient of D is positive, the molecules with a longer total length may be able to take on a conformation closer to the one that can fit the receptor best (the optimal shape). The compound best in terms of D appears to be longer in the extended form than any of

the compounds studied here. The case where the optimal D is apparent has been previously documented in the analysis of insect juvenile hormone mimics, compounds with a long zigzag aliphatic chain (Nakayama *et al.*, 1984). The positive sign of the log P'(pH 7) term in eq. 4-1 may reflect the partitioning from a polar aqueous medium, saliva, onto the hydrophobic receptor cavity of the tongue, rather than the transport process to the site.

4-2 A new hydrophobicity scale of side chains

4-2-1 Definition

From eq. 3-16 we can propose a new effective hydrophobicity scale, π_α, for unionizable amino acid side chains. The π_α value was defined as the summation of factors attributable to each residue as the component of the total log P' value as shown in eq. 4-2, where δ takes 0.55 for N-terminal residues and 0.30 for others. The conformational factors were not included since they are not applicable to di- and tripeptides as well as to those higher than pentapeptides where other conformational effects such as the α-helix formation should be considered. For proline, the π_α value was varied depending upon the situations.

$$\pi_\alpha = [\text{intrinsic } \pi] + \delta\, E'_S{}^{c} + [\text{coefficient of I for each polar side chain and proline}] \quad (4\text{-}2)$$

The newly defined π_α value is listed in Table 4-2 along with some other hydrophobicity scale publicized earlier (Kyte and Doolittle, 1982; Fauchère and Pliška, 1983; Nozaki and Tanford, 1971; Wolfenden *et al.*, 1981). In Fig. 4-2, relationships between π_α(MC) and other hydrophobicity parameters are shown. The patterns of variations in π value of Fauchère and Pliška and $\Delta\mu^0$

Table 4-2. Hydrophobicity Scales of Amino Acid Side Chains

Amino Acid	π_α [a] N [b]	π_α [a] MC [c]	π(F.P.) [d]	$\Delta\mu^0$ [e] (N.T.)	H.I. [f] (K.D.)	H.P. [g] (W.)
Gly	0	0	0	0	0	0
Ala	0.19	0.24	0.31	0.5	2.2	-0.45
Val	0.49	0.82	1.22	1.5	4.6	-0.40
Leu	0.92	1.28	1.70	1.8	4.2	-0.11
Ile	0.72	1.17	1.80	-	4.9	-0.24
Phe	1.35	1.57	1.79	2.5	3.2	-3.15
Tyr	0.78	1.01	0.96	2.3	-0.9	-8.50
Trp	1.72	1.93	2.25	3.4	-0.5	-8.27
Met	0.67	0.93	1.23	1.3	2.3	-3.87
Ser	-0.08	0.04	-0.04	-0.3	-0.4	-7.45
Thr	0.07	0.25	0.26	0.4	-0.3	-7.27
Pro	h	h	0.72	-	-1.2	-

a From this work, calculated from eq. 4-2.

b For N-terminal residues.

c For central and C-terminal residues.

d π value of Fauchère-Pliška (Fauchère and Pliška, 1983) from log P values of N-acetylamino acid amides.

e Relative free energy of transfer from either ethanol or dioxane to water (kcal/mole) of Nozaki-Tanford (Nozaki and Tanford, 1971).

f Hydropathy index of Kyte-Doolittle (Kyte and Doolittle, 1982). Reference was shifted to Gly.

g Hydration potential (kcal/mole) of Wolfenden *et al.* (Wolfenden *et al.*, 1981). Reference was shifted to Gly.

h π_α(location, number of residues) of proline;
π_α(N, 2): 0.35, π_α(MC, 2): 1.16, π_α(N, 3): 0.00, π_α(MC, 3): 0.81, π_α(N, 4): -0.34, π_α(MC, 4): 0.46.

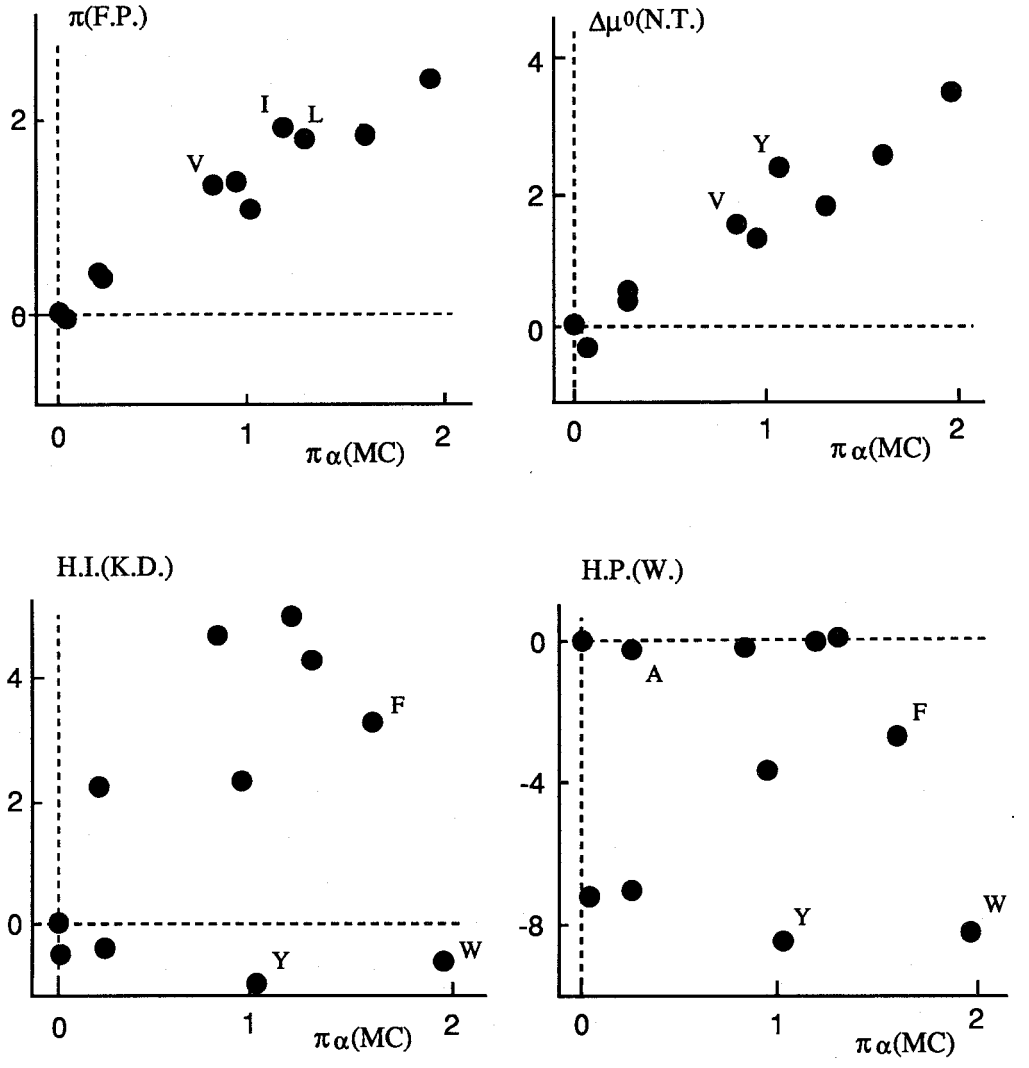


Figure 4-2. Relationships between π_α (MC) and other hydrophobicity parameters

value of Nozaki and Tanford are similar to that of our π_α value. The π(F.P.) values for those of isoleucine (I), valine (V), and leucine (L) the side chains of which are bulky, and $\Delta\mu^0$ values of valine (V) and tyrosine (Y) are, however, slightly but significantly larger than ours. The correspondence of π_α with other parameters was not

good, because the definitions of these parameters included various types of factors relating to protein structures. Being not straightforward, these parameters are considered to deal with carefully.

4-2-2 Possible application in predicting secondary structures of peptide backbone

Comparison of the value calculated by eq. 4-3 as the non-conformational component supposed for the "imaginary" random form with the experimentally observed log P' should be useful to obtain information on the component attributable to the effect of the conformation.

$$\log P'(\text{random}) = \Sigma\pi_\alpha - 0.582\ I_{pep} - 3.866 \qquad (4\text{-}3)$$

For tetra- and pentapeptides, the conformational effect was represented by I_{turn} and $\log f_{i+2}$ terms. Thus, examinations of the difference between experimental log P' and calculated log P'(random) should allow us to predict the secondary structural factors of tetra- and pentapeptides of which the conformational parameters are unknown. This procedure may be extended to higher peptides where secondary structural factors differ from those included in tetra- and pentapeptides. To estimate the log P'(random) value for partial domain of proteins, each π_α is calculated using the recommended $\delta = 0.30$. Most of the hydrophobicity indices of amino acid side chains so far publicized are either defined from partition or phase transfer parameters of unit amino acids or their analogs (Fauchère and Pliška, 1983; Nozaki and Tanford, 1971), or statistically calculated from the solvent-accessible surface area of proteins (Kyte and Doolittle, 1982). Our new index, π_α, differs from others in that it was defined from the experimentally measured net hydrophobicity of oligopeptides existing in solutions as such.

Chapter 5 Synthesis of Compounds

Zwitterionized di- to pentapeptides

Dipeptides not containing proline were obtained commercially except for Leu-Ile which was prepared by the reaction of Z-Leu-OH with H-Ile-OBzl by the DCC (dicyclohexylcarbodiimide)/HOBt (1-hydroxybenzotriazole) coupling method (König and Geiger, 1970) and subsequent hydrogenation in the presence of palladium catalyst on activated charcoal (Bergmann and Zerras, 1932).

Some tripeptides were obtained commercially and others lacking proline were synthesized by the hydrogenation of Z-tripeptide benzyl esters which were prepared by coupling Z-amino acids with dipeptide benzyl ester hydrochlorides. Dipeptide benzyl ester hydrochlorides were obtained by the removal of the Boc-group from Boc-dipeptide benzyl esters with hydrogen chloride in dioxane (Anderson and McGregor, 1957), which were prepared by the coupling of Boc-amino acids with amino acid benzyl esters by the DCC/HOBt method.

Tetra- and pentapeptides other than enkephalin analogs were newly synthesized either by such standard solution phase techniques as the N-hydroxysuccinimide ester method (Anderson *et al.*, 1964), or by the solid-phase method on 4-(hydroxymethyl)-phenylacetamidomethyl (PAM) resins with an automated peptide synthesizer (Applied Biosystems 430A) (Merrifield, 1963; Mitchell *et al.*, 1976). Enkephalin derivatives were purchased from Sigma Chemical Company.

Peptides containing proline were synthesized using an automated peptide synthesizer, (Applied Biosystems 430A), by the solid-phase method PAM-resin (Merrifield, 1963; Mitchell *et al.*, 1976). Boc-amino acids were coupled successively with DCC or by the DCC/HOBt method. Peptides were deprotected by the

anisole/hydrogen fluoride method (Wang and Merrifield, 1980) and precipitated with ether.

Z- and Boc-amino acids were commercially available and amino acid benzyl esters were synthesized by the esterification of amino acids with benzyl alcohol (Gibian and Schröder, 1961).

Acetyl-di- and tripeptide amides

Most of the Boc-amino acids, *O*-benzylated hydroxy Boc-amino acids, and amino acid amides, and Z-tryptophan were commercially available.

The amides of alanine, isoleucine, phenylalanine, and glutamine as the trifluoroacetate were synthesized by the mixed carbonic anhydride method (Vaughan and Osato, 1952). The carboxyl group in Boc-amino acids was activated by the addition of *iso*-butyl chloroformate leading to mixed anhydrides at -30°C. Then, aqueous ammonia was added to the solution, forming Boc-amino acid amides. The amino acid amide trifluoroacetates were obtained by the removal of the Boc-group from Boc-amino acid amides by treating with trifluoroacetic acid.

The Boc- and Z-amino acid N-hydroxysuccinimide esters were synthesized by the esterification of Boc- and Z-amino acids with N-hydroxysuccinimide using DCC as the condensing agent (Anderson *et al.*, 1964). Boc-glutamine and Boc-asparagine *p*-nitrophenyl esters were prepared by the esterification of Boc-glutamine and Boc-asparagine with *p*-nitrophenol with the aid of DCC (Bodanszky and du Vigneaud, 1959).

Boc- and Z-dipeptide amides were prepared by coupling Boc- and Z-amino acid N-hydroxysuccinimide esters or *p*-nitrophenyl esters with amino acid amide hydrochlorides. Dipeptide amide trifluoroacetates were obtained by the removal of the Boc-group from Boc-dipeptide amides with trifluoroacetic acid (Kappeler and Schwyzer, 1961) or hydrogen chloride in dioxane (Anderson and

McGregor, 1957). For Z-dipeptide amides including tryptophan, the Z-group was removed by hydrogenation in the presence of palladium catalyst on activated charcoal (Bergmann and Zerras, 1932). Tripeptide amide trifluoroacetates were prepared by coupling activated esters of Boc-amino acids with dipeptide amide trifluoroacetates then removing the Boc-group. Acetylation of di- and tripeptide amide trifluoroacetates was done with acetic anhydride (Siegel and Awad, 1973). The *O*-benzyl group of side chains of serine, threonine and tyrosine was removed by hydrogenation similarly to the removal of Z-group.

Purification was done by reversed-phase liquid chromatography with a column of Shim-Pack PREP-ODS (20.0mm x 25cm) or Cosmosil-ODS (10.0mm x 25cm) and water/acetonitrile or 0.1% aqueous trifluoroacetic acid/acetonitrile containing 0.1% trifluoroacetic acid as the eluent. The purity and composition of peptides were confirmed by the HPLC, amino acid analysis, and/or elemental analyses for C, H, and N.

General Conclusion

The present results are believed to indicate a great advantage in analyzing the log P' and log P value of zwitterionized and protected peptides by using various physicochemically well defined side chain parameters such as the intrinsic π constant, the $E'_S{}^c$ parameter for the steric effect on the relative solvation of functional groups on the backbone, and the Chou-Fasman conformational potential index along with the (sub)structural indicator variables for polar side chain features, the internal hydrogen bond formation and the polar proximity factor. The size of each of the regression coefficients was physicochemically well rationalized except for those for the Pro residue.

Recently, a computerized estimation system of the log P and log P' of amino acids, peptides, and related compounds has been explored by Abraham and Leo (Abraham and Leo, 1987). Our experimental log P' and log P values and the empirical correlation such as eq. 1-6, 2-4 and 3-16 should serve for elaborating the computerized system. For instance, the steric effect of side chain substituents, which makes the log P (or log P') value lower and the β-turn effect, which makes the value higher have not been considered as factors to be summed up with those assignable to other structural features in their present version of the system. Thus, if the steric and conformational factors could be incorporated in the system, its precision would be much improved.

It is hoped that the present study could substantiate the approach to predictions of not only the log P' and log P value itself of unmeasured compounds but also the secondary structures of higher peptides in the hydrophobic environments.

References

Abraham, D.J. and Leo, A., *PROTEINS: Structure, Function, and Genetics*, 130-152 (1987).

Akamatsu, M., Yoshida, Y., Nakamura, H., Asao, M., Iwamura, H. and Fujita, T., *Quant. Struct.-Act. Relat.* **8**, 195-203 (1989).

Akamatsu, M., Okutani, S., Nakao, K., Hong, N.J. and Fujita, T., *Quant. Struct.-Act. Relat.*, in press (1990).

Anderson, G.W. and McGregor, A.C., *J. Am. Chem. Soc.* **79**, 6180-6183 (1957).

Anderson, G.W., Zimmerman, J.E. and Callahan, F.M., *J. Am. Chem. Soc.* **86**, 1839-1842 (1964).

Asao, M., Iwamura, H., Akamatsu, M. and Fujita, T., *J. Med. Chem.* **30**, 1873-1879 (1987).

Belitz, H.-D., Chen, W., Jugel, H., Stempfl, H., Treleano, R. and Wieser, H., *Chem. Ind. (London)*, 23-26 (1983).

Bergmann, M. and Zerras, L., *Chem. Ber.* **65**, 1192-1201 (1932).

Bodanszky, M. and du Vigneaud, V., *J. Am. Chem. Soc.* **81**, 5688-5691 (1959).

Bush, C.A., Sarkar, S.K. and Kopple, K.D., *Biochemistry* **17**, 4951-4954 (1978).

Cabon, J.-Y. and L'Her, M. and Courtot-Coupez, J., *J. Chem. Res.*, 342-343 (1978).

Charton, M., *Topics Curr. Chem.* **114**, 57-91 (1983).

Charton, M. and Charton, B.I., *J. Theor. Biol.* **102**, 121-134 (1983).

Chou, P.Y. and Fasman, G.D., *J. Mol. Biol.* **115**, 135-175 (1977).

Claassen, V. (Ed), *Trends in Drug Research*, pp. 73-108, Elsevier Science Publishers, Amsterdam (1990).

Dickerson, R.E., Takano, T., Eisenberg, D., Kallai, O.B., Samson, L., Cooper, A. and Margoliash, E., *J. Biol. Chem.* **246**, 1511-1535 (1971).

Ellenbogen, E., *J. Am. Chem. Soc.* **74**, 5198-5201 (1952).

Fauchère, J.-L. and Pliška, V., *Eur. J. Med. Chem. - Chim. Ther.* **18**, 369-375 (1983).

Fauchère, J.-L. *QSAR in Design of Bioactive Compounds. Proceedings of the 1st Telesymposium on Medicinal Chemistry*, J.R. Prous, Barcelona 1984, pp. 135-144.

Fermi, G., Perutz, M.F., Shaanan, B. and Fourme, R., *J. Mol. Biol.* **175**, 159-174 (1984).

Fujita, T., Iwasa, J. and Hansch, C., *J. Am. Chem. Soc.* **86**, 5175-5180 (1964).

Fujita, T., Nishioka, T. and Nakajima, M., *J. Med. Chem*, **20**, 1071-1081 (1977).

Fujita, T., *Prog. Phys. Org. Chem.* **14**, 75-113 (1983).

Gardner, R.J., *J. Sci. Food Agric.* **31**, 23-30 (1980).

Gibian, von H. and Schroder, E., *Liebigs Ann. Chem.* **642**, 145-162 (1961).

Gillen, M.F. and Williams, R.E., *Can. J. Chem.* **53**, 2351-2353 (1975).

Graf, J., Ogle, R.C., Robey, F.A., Sasaki, M., Martin, G.R., Yamada, Y. and Kleinman, H.K., *Biochemistry* **26**, 6896-6900 (1987).

Graham, L.L. and Chang, C.Y., *J. Phys. Chem.*, **75**, 784-788 (1971).

Hadži, D. and Jerman-Blažič, B. (Eds.), *QSAR in Drug Design and Toxicology*, Elsevier Science Publishers, Amsterdam 1987, pp. 221-297.

Hancock, C.K., Meyers, F.A. and Yager, B.J., *J. Am. Chem. Soc.* **83**, 4211-4213 (1961).

Hansch, C. and Fujita, T., *J. Am. Chem. Soc.* **86**, 1616-1626 (1964).

Hansch, C. and Leo, A.J. (Ed.), *Substituent Constants for Correlation Analysis in Chemistry and Biology*, pp.17-43, John Wiley and Sons, New York (1979).

Harris, J.I. and Lerner, A.B., *Nature* **179**, 1346-1347 (1957).

Hughes, J., Smith, T.W., Kosterlitz, H.W., Forthergill, L.A., Morgan, B.A. and Morris, H.R., *Nature* **258**, 577-579 (1975).

Iwamura, H., *J. Med. Chem.* **23**, 308-312 (1980).

Iwasa, J., Fujita, T. and Hansch, C., *J. Med. Chem.* **8**, 150-153 (1965).

Kappeler, H. and Schwyzer, R., *Helv. Chim. Acta* **44**, 1136-1141 (1961).

Karle, I.L., Karle, J., Wieland, T., Burgermeister, W., Faulstich, H. and Witkop, B., *Proc. Nat. Acad. Sci. USA* **70**, 1836-1840 (1973).

Kauzman, W., *Adv. Protein Chem.* **14**, 1-63 (1959).

König, W. and Geiger, R., Chem. Ber. 103, 788-798 (1970).

Koryta, J., *Electrochimica Acta* **29**, 445-452 (1984).

Kyte, J. and Doolittle, R.F., *J. Mol. Biol.* **157**, 105-132 (1982).

Leo, A., *J. Chem. Soc. PERKIN TRANS. II*, 825-838 (1983).

Lewis, P.N., Momany, F.A. and Scheraga, H.A., *Proc. Nat. Acad. Sci. USA* **68**, 2293-2297 (1971).

MacPhee, J.A., Panaye, A. and Dubois, J.-E., *Tetrahedron* **34**, 3553-3562 (1978).

Menn, J.J. and Borkovec, A.B., *J. Agric. Food Chem.* **37**, 271-278 (1989).

Merrifield, R.B., *J. Am. Chem. Soc.* **85**, 2149-2154 (1963).

Mitchell, A.R., Erickson, B.W., Ryabtsev, M.N., Hodges, R.S. and Merrifield, R.B., *J. Am. Chem. Soc.* **98**, 7357-7362 (1976).

Nair, R.M.G., Kastin, A.J. and Schally, A.V., *Biochem. Biophys. Res. Commun.* **43**, 1376-1381 (1971).

Nakai, K., Kidera, A. and Kanehisa, M., *Protein Engineering* **2**, 93-100 (1988).

Nakayama, A., Iwamura, H. and Fujita, T., *J. Med. Chem.* **27**, 1493-1502 (1984).

Nisato, D., Wagnon, J., Callet, G., Mettefeu, D., Assens, J.-L., Plouzane, C., Tonnerre, B., Pliška, V. and Fauchère, J.-L., *J. Med. Chem.* **30**, 2287-2291 (1987).

Nishioka, K., Constantopoulos, A., Satoh, P.S. and Najjar, V.A., *Biochem. Biophys. Res. Commun.* **47**, 172-179 (1972).

Nozaki, Y. and Tanford, C., *J. Biol. Chem.* **7**, 2211-2217 (1971).

Perrin, D.D., *Dissociation Constants of Organic Bases in Aqueous Solution*, Butterworths, London 1965.

Schally, A.V., Arimura, A., Kastin, A.J., Matsuo, H., Baba, Y., Redding, T.W, Nair, R.M.G., Debeljuk, L. and White, W.F., *Science* **173**, 1036-1038 (1971).

Siegel, S. and Awad, W.M., Jr., *J. Biol. Chem.* **248**, 3233-3240 (1973).

Starratt, A.N. and Brown, B.E., *Life Sci.* **17**, 1253-1256 (1975).

Stone, J.V., Mordue, W., Batley, K.E. and Morris, H.R., *Nature* **263**, 207-211 (1976).

Taft, R.W., Jr., in: Newman, M.S. (Ed.), *Steric Effects in Organic Chemistry*, John Wiley & sons, Inc., New York, N.Y. 1965, pp. 556-675.

Takagi, H., Shiomi, H., Ueda, H. and Amano, H., *European J. Pharmacol.* **55**, 109-111 (1979).

Takayama, C., Akamatsu, M. and Fujita, T., *Quant. Struct. -Act. Relat.* **4**, 149-160 (1985).

Townend, R., Kumosinski, T.F., Timasheff, S.N., Fasman, G.D. and Davidson, B., *Biochem. Biophys. Res. Comm.* **23**, 163-169 (1966).

Vaughan, J.R., Jr. and Osato, R.L., *J. Am. Chem. Soc.* **74**, 676-678 (1952).

Venkatachalam, C.M., *Biopolymers* **6**, 1425-1436 (1968).

Verloop, A., Hoogenstraaten, W. and Tipker, J., in: Ariens, E.J. (Ed.), *Drug Design Vol. VII*, Academic Press, New York , pp.165-207.

Wang, S.S. and Merrifield, R.B., *Biochemistry* **19**, 3233-3238 (1980).

Wieser, H. and Belitz, H.-D., *Z. Lebensm. Unters.-Forsch.* **159**, 65-72 (1975).

Wieser, H. and Belitz, H.-D., *Z. Lebensm. Unters.-Forsch.* **160**, 383-392 (1976).

Wolfenden, R., Andersson, L., Cullis, P.M. and Southgate, C.C.B., *Biochemistry* **20**, 849-855 (1981).

Original Papers

Akamatsu, M., Yoshida, Y., Nakamura, H., Asao, M., Iwamura, H. and Fujita, T., "Hydrophobicity of zwitterionized di- and tripeptides having unionizable side chains and correlation with substituent and structural parameters", *Quant. Struct.-Act. Relat.* **8**, 195-203 (1989).

Akamatsu, M., Okutani, S., Nakao, K., Hong, N.J. and Fujita, T., "Hydrophobicity of N-acetyl-di- and tripeptides having unionizable side chains and correlation with substituent and structural parameters", *Quant. Struct.-Act. Relat.*, in press.

Akamatsu, M. and Fujita, T., "Hydrophobicity of zwitterionized di- to pentapeptides having unionizable side chains and correlation with substituent and structural parameters", in preparation.

Asao, M., Iwamura, H., Akamatsu, M. and Fujita, T., "Quantitative structure-activity relationships of the bitter thresholds of peptides", *J. Med. Chem.* **30**, 1873-1879 (1987).

Related Papers

Takayama, C., Akamatsu, M. and Fujita, T., "Effects of structure on 1-octanol/water partitioning behavior of aliphatic amines and ammonium ions", *Quant. Struct.-Act. Relat.* **4**, 149-160 (1985).

Takayama, C., Akamatsu, M. and Fujita, T., "Quantitative structure-activity study of the inhibition of acetylcholinesterase with aliphatic ammonium ions", *Quant. Struct.-Act. Relat.* **8**, 90-97 (1989).

Acknowledgments

The present thesis is based on the studies performed at the Department of Agricultural Chemistry, Kyoto University. The author would like to express her sincere gratitude to Professor Toshio Fujita for his expert supervision and stimulating discussions. She also wishes to express her acknowledgment to Dr. Hajime Iwamura, Dr. Keiichiro Nishimura, Dr. Takaaki Nishioka, Dr. Yoshiaki Nakagawa and Dr. Hideto Miyoshi. She is deeply indebted to Dr. Nam Joo Hong, Messrs. Masaaki Asao, Yohji Yoshida, Hideaki Nakamura, Shin-ichi Okutani, Kazuya Nakao, Akira Oku and Tadashi Katayama for their skillful and active collaborations in this work. The author thanks Dr. Motoyoshi Nomizu of the Kirin Brewery Co., Ltd. and Dr. Takaharu Tanaka and Mr. Norio Iwasawa of Suntory Ltd. for synthesizing some peptides and derivatives using the automated peptide synthesizer. She also appreciates the help of Professor Sho Takahashi in the measurement and analysis of CD.

Finally, the author sincerely thanks to Mrs. Terumi Nishiguchi, Mr. Yukio Koizumi and Mrs. Fumie Fujita for their continuous encouragement and support.