Journal of Chromatography, 536 (1991) 165–183
Elsevier Science Publishers B.V., Amsterdam

CHROMSYMP. 2033

High-performance liquid chromatography of amino acids, peptides and proteins

CVII.[a] Analysis of group retention contributions for peptides separated with a range of mobile and stationary phases by reversed-phase high-performance liquid chromatography

M. C. J. WILCE, M. I. AGUILAR and M. T. W. HEARN*
Department of Biochemistry and Centre for Bioprocess and Technology, Monash University, Clayton, Victoria 3168 (Australia)

ABSTRACT

An extensive data base which comprises the retention data of a total of 2106 peptides has been established and used to derive individual amino acid group retention coefficients. A multiple linear regression matrix approach was employed for solving the numerical value of the coefficients from the multivariate structure–retention dependencies. Statistical analysis of the retention data revealed that a minimum of 100 peptides is required to provide consistent values of the amino acid coefficients. Categorisation of all peptides allowed the influence of various chromatographic parameters on the coefficients to be evaluated. In particular, a decrease in the alkyl chain length of the chemically modified *n*-alkylsilica from octadecyl to butyl did not generally coincide with a decrease in the value of the group retention coefficients of individual amino acids. This study has established a detailed computational basis for characterising peptide retention behaviour and provides further insight into the mechanism of the interaction of peptides with immobilised hydrocarbonaceous ligands.

INTRODUCTION

Reversed-phase high-performance liquid chromatography (RP-HPLC) is now the most widely used technique for the analytical, micro-, semi-preparative separation of peptides and proteins. This is due to a number of important reasons including: (i) the excellent resolution that can be achieved for homologous, as well as structurally disparate compounds under a wide range of chromatographic conditions; (ii) the high recovery of solutes (even at ultra microanalytical levels); (iii) the reproducibility of the separation; (iv) the ability to utilize RP-HPLC to evaluate various physicochemical

[a] For Part CVI, see ref. 29.

0021-9673/91/$03.50

parameters associated with peptide and protein surface interactions and folding (conformational) hierarchies.

Despite the ever increasing usage of RP-HPLC for the separation and analysis of peptides and proteins, the molecular processes that control the interaction between the non-polar stationary phase, the mobile phase, and the solute are not yet fully understood. The development of models to describe the mechanisms by which peptides and proteins interact with reversed-phase chromatographic systems would provide the basis for practical optimisation protocols concomitant with the elucidation of peptide and protein binding mechanisms.

No fully developed mechanistic, thermodynamic or extrathermodynamic models are yet available which adequately accommodate all the structural (primary, secondary, tertiary and higher order) properties and kinetic vagaries of peptide or protein retention behaviour with porous, chemically modified *n*-alkylsilicas. Because of this, most investigators have relied upon empirical non-mechanistic models. The most fully evolved and accessible of these models is based on the linear solvent strength (LSS) gradient elution concepts as originally developed by Snyder (see refs. 1–4). This model provides a quantitative basis for the evaluation of peptide and protein retention behaviour under ideal reversed-phase conditions, and allows a more rational selection of chromatographic parameters to achieve a set of optimal chromatographic conditions, particularly for those peptides that do not succumb to sorbent-induced conformational effects.

Several studies have characterised the retention behaviour of closely related structural analogues in terms of the LSS model and have demonstrated that the interaction of peptides with the hydrocarbonaceous stationary phase is intimately dependent on the relative topographic arrangement of primary hydrophobic and hydrophilic amino acid residues [2,4,5] coded within the peptide sequences. While the LSS retention model provides a useful basis for the optimisation of peptide separations, a number of chromatographic experiments are required for the procedure. An alternative experimental approach to the optimisation of peptide separations, based on a more complete understanding of the interaction processes of peptides and proteins in RP-HPLC, is to use chromatographically derived hydrophobicity coefficients to predict and confirm the retention time of peptides of known amino acid composition. Knowledge of reliable sets of amino acid coefficients would also be useful in validating LSS retention predictions with peptides which show regular retention behaviour. Several sets of retention coefficients have been reported which have been derived by a number of different methods [6–17]. A commonly used procedure for the measurement of the influence of individual amino acids residues on peptide retention times involves the use of synthetic peptide analogues in which designated positions are systematically changed [16,17].

We have previously reported an alternative approach for the derivation of hydrophobicity coefficients in which retention coefficients were determined using iterative linear regression analysis of retention data derived from peptides with significantly different amino acid sequences [6]. The results of these and several other studies generally indicate that the hydrophobic contribution of amino acid residues in small peptides results in an essentially additive effect on peptide retention to alkylsilicas [6–17]. These observations are in accordance with the linear free energy relationships and associated predictions of the Martin[64], Hansch[65] and Hammet[66]

equations. With larger peptides, where secondary and tertiary structural features become more important, greater deviations of experimental retention times from those calculated from the summated coefficients have typically been observed. These differences arise as a consequence of the solute interacting with the sorbent through only a proportion of its total molecular surface. In an attempt to allow for the deviations in predicted retention time from experimentally observed retention behaviour of large polypeptides and proteins, Mant *et al.* [18], used a linearisation approach, in which correction factors were introduced to derive the predicted retention time based on the use of the natural logarithm of the length of the particular peptide or protein. This numerical method follows the observation that there is an exponential-like relationship between predicted and observed retention time with an increase in peptide chain length. Although such numerical correction methods give improved correlation between the predicted and experimental retention times for the test set of solutes, they unfortunately shed no light on the mechanism of binding for these larger molecules nor do they represent a *de novo* procedure for the prediction of retention of peptides of similar length and composition but difference sequence.

It is now well established that peptides and proteins interact with the stationary phase surface in an orientation specific manner [19–22]. Thus, their chromatographic retention behaviour is determined by the molecular composition of a specific contact region. For small peptides comprising between *ca.* 3–15 amino acid residues, the contact region may represent a large proportion of the molecular surface. However, for large polypeptides and protein molecules, the chromatographic contact region will be a relatively small portion of the entire solute surface. Because the molecular composition of the contact region largely determines the retention properties of a particular solute, chromatographic retention parameters contain a vast amount of information regarding the interactive segment of the peptidic species.

The present investigation therefore extends our previous studies by deriving coefficients from a database of over 2000 peptides. In particular this study permitted a detailed analysis of the influence of several experimental parameters such as stationary phase and mobile phase composition on the individual coefficients for each amino acid. These data are particularly relevant to further understanding of the mechanistic basis of the interaction of peptides and proteins in RP-HPLC.

MATERIALS AND METHODS

An extensive literature search was carried out to establish the chromatographic data base for this study. The information required included peptide sequence, retention characteristics, and chromatographic conditions. Hydrophobic group retention coefficients were calculated from the literature peptides and their retention data using multiple linear regression analyses.

The multiple linear regression was carried out using a matrix approach for solving sets of simultaneous equations. For the multiple linear regression technique used in this study the group retention contribution of each amino acid is considered as an unknown X_i. The percentage mole fraction of organic modifier present at the time of elution of each peptide, MF_k, and the corresponding amino acid composition are considered as a set of simultaneous equations as follows;

$$a_{11}X_1 + a_{12}X_2 + \ldots a_{1n}X_n = MF_1$$
$$a_{21}X_1 + a_{22}X_2 + \ldots a_{2n}X_n = MF_2$$
$$a_{m1}X_1 + a_{m2}X_2 + \ldots a_{mn}X_n = MF_m$$

which can be rewritten in matrix form,

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \cdot & \cdot & & \cdot \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{bmatrix}$$

$$\mathbf{X} = \begin{bmatrix} X_1 \\ X_2 \\ \cdot \\ X_n \end{bmatrix} \qquad \mathbf{b} = \begin{bmatrix} MF_1 \\ MF_2 \\ \cdot \\ MF_m \end{bmatrix}$$

The vector **X**, of group retention coefficients, can be solved as $\mathbf{X} = (\mathbf{A}^T\mathbf{b})(\mathbf{A}^T\mathbf{A})^{-1}$. These equations were then solved with partial pivoting. The theoretical *MF* for each peptide was then calculated by addition of the appropriate retention coefficients. Thus, in order to compare the retention data derived from the range of chromatographic conditions listed in Table I, all elution times were converted to *MF* according to the known rate of change of organic solvent concentration. The degree of correlation between the calculated group retention coefficients and the retention data was assessed by the correlation coefficient, R^2.

The multiple linear regression was performed on one of two computers. For small data sets, it was most convenient to use a multiple linear regression program written by the authors in Pascal for the IBM PC/AT or compatible, but for large data sets the group retention coefficients were calculated using the multiple linear regression routine from the SPSS^X package on the Monash University VAX computer. Both programs produced identical results. A simple database program was also written in Pascal by the authors for the IBM PC/AT to store the peptide data and to allow the retrieval of particular peptides and their chromatographic data from criteria specified by the user.

RESULTS AND DISCUSSION

Derivation of amino acid coefficients

Peptidic solutes are retained in RP-HPLC by the expulsion of the solute from the polar mobile phase with concomitant adsorption onto the non-polar stationary phase. The differential retardation of the solute species is dependent upon its intrinsic hydrophobicity, the eluotropicity of the mobile phase and the nature of the hydrocarbonaceous stationary phase.

The basis of the calculations to derive the amino acid group retention coefficients assumes that peptide retention can be described solely in terms of ideal reversed-phase behaviour, and that there is a first order dependency of peptide retention on the mole fraction of organic modifier. Thus in the absence of electrostatic or hydrogen bonding effects, the solute retention will be determined solely by the nature of the solvophobic

solute–ligand interaction and is given in terms of the capacity factor k' according to

$$k'_{\text{hydrophobic}} = \varphi K_{\text{hydrophobic}}$$

where K is the equilibrium association constant and φ is the phase ratio (volume stationary phase/volume mobile phase). The selectivity between two peptides, P_i and P_j separated under a defined set of chromatographic conditions can then be expressed as

$$\ln \alpha_{i,j} = \ln (k'_j/k'_i)$$

If we consider two peptides of similar sequence only differing by one amino acid residue, then the group retention coefficient due to the different amino acid can be defined as

$$\tau = \ln \alpha_{i,j} = \ln (k'_{\text{peptide } i}) - \ln (k'_{\text{peptide } j})$$

The τ contribution is thus a function of the differences in the overall standard unitary free energy changes and can be formally associated with the transfer of peptide solute i from the mobile phase to the stationary phase relative to the transfer of peptide j of identical residue number. According to the solvophobic theory [23] the surface area of the solute molecule which is in contact with the non-polar stationary phase plays a significant role in determining the magnitude of hydrophobic interaction. Since linear free energy relationships are anticipated between bulk phase partition parameters and functional group contributions, linear relationships should also exist between retention behaviour, as expressed by ln k' values and the surface area of the solute contact region with the stationary phase. Furthermore, as these hydrophobicity coefficients are derived from chromatographic retention data, they are physicochemically related to the binding energy of each amino acid. Derivation of the coefficients in this manner therefore provides a general approach to quantitating the relative propensity of each residue in a particular amino acid sequence to interact with a surface of defined ligand structure and density. As discussed earlier, the optimisation of solute retention can be achieved by varying the organic modifier, or through manipulation of secondary chemical equilibria such as ion-pairing, ionisation and solvation effects, or by selecting different stationary phases [1–4]. This paper considers, through the use of group retention coefficients, changes in the interactive nature of amino acid residues under a range of chromatographic conditions such as changes in the chain length of the stationary phase ligand, and different organic modifiers in the mobile phase.

To derive amino acid group retention coefficients for a large number of peptides from a range of chromatographic conditions, and to explore the effect of varied chromatographic conditions on the group retention coefficients, a database of 2106 peptides was established as described in the Materials and Methods section. A total of 44 different sets of peptides consisting of 1337 peptides and 14 726 amino acids were selected from this database including those from our previous studies. Table I lists the chromatographic conditions which include both the stationary phase and mobile phase characteristics, and the literature reference for each peptide data set. Over the

TABLE I

PARAMETERS AND LITERATURE SOURCES OF CHROMATOGRAPHIC DATA

NR = Not recorded in literature reference; DL = data collected in this laboratory.

Data set number	Stationary phase[a]	Mobile phase[b]	Flow-rate	Column length	Column type	Ref.
1	C18	TA	1.00	250	Ultrasphere ODS	30
2	C18	TA	1.50	300	μBondapak	31
4	C18	TA	1.00	NR	Chromegabond MC-18	32
5	C18	TA	3.00	250	SynChropak RP-P	33
6	C18	TA	0.70	NR	RPC18	33
7	C18	TPA	0.70	NR	SynChropak RP-P	33
8	C4	TA	1.00	250	RP304	34
10	C18	TPA	1.00	NR	Unknown	35
11	C18	TA	3.00	250	RP300	36
12	C18	TA	1.50	250	Vydac TP RP	37
13	C18	TA	0.50	300	TSK LS-410	38
14	C18	TA	0.80	300	TSK-Gel LS-410AK	38
16	C8	TA	1.00	250	Altex Ultrasphere	39
18	C8	TA	1.00	250	Altex Ultrasphere	39
19	C8	TA	1.20	NR	Aquapore RP300	40
20	C18	TA	0.50	250	SynChropak	41
21	C4	TA	0.80	250	Vydac	41
22	C18	TA	1.00	250	Vydac	42
23	C4	TA	1.00	250	NR	43
24	C18	TA	1.00	250	Nucleosil	44
25	C18	TA	1.00	250	Developmental column	DL
26	C4	TA	1.00	250	Bakerbond wide pore	DL
27	C18	TA	1.00	250	Bakerbond wide pore	DL
28	C18	TA	1.00	250	Bakerbond wide pore	DL
31	C18	TA	1.00	300	μBondapak	45
35	C18	TA	2.00	250	Aquapore RP-300	46
37	C18	TA	1.00	150	Cosmosil 5C18-P	47
60	C8	TPA	1.00	250	LiChrosorb RP-8	48
61	C18	TA	0.80	300	Cosmosil 5C18	49
62	C4	TPA	1.00	250	Bakerbond	50
63	C18	TPA	1.00	300	Bakerbond	50
65	C18	TA	1.00	159	NR	51
69	C18	TA	1.20	250	μBondapak	52
70a	C18	TA	1.20	250	μBondapak	53
70b	C18	TA	1.00	250	μBondapak	53
73	C18	TA	1.00	NR	Ultropac TSK ODS-120	54
74	C18	TA	1.00	250	μBondapak	55
75	C18	TA	1.00	250	RP-P SynChropak	56
76	C18	TA	1.00	250	RP-P SynChropak	57
77	C18	TA	1.00	250	Brownlee RP-300	58
79	C18	TA	1.00	150	Cosmosil 5C18	59
81	C18	TA	1.00	NR	Cosmosil 5C18 P	60
83	C18	TA	1.00	NR	TSK Gel 410AK	61
86	C18	TPA	1.00	300	TSK Gel LS 410 A	62
87	C18	TPA	1.00	300	TSK Gel LS 410 A	63

[a] C18 = Reversed-phase octadecyl stationary phase. C8 = reversed-phase octyl stationary phase. C4 = reversed-phase butyl stationary phase.

[b] TA = Trifluoroacetic acid–acetonitrile–water; TPA = trifluoroacetic acid–1-propanol–acetonitrile–water. All peptides were separated with gradient elution between mobile phases composed of 0.1% trifluoroacetic acid–water and 0.1% trifluoroacetic acid–acetonitrile or (1-propanol–acetonitrile)–water mixtures.

range of conditions examined it has been observed that there is a direct linear relationship between retention time and the percentage organic modifier present at the time of elution of a peptidic solute [12].

The group retention coefficients presented here were generated with a multiple linear regression that solves sets of simultaneous equations in a matrix format. An alternative computational approach for the derivation of group retention coefficients is multiple linear analysis with forcing [6], which in some circumstances has been shown to generate comparable results. Although the programming of forcing routines is less complex than the matrix method and also has a much smaller memory overhead, the matrix method is considered to be superior because the group retention coefficients are derived by statistical means. This matrix approach thus provides significantly more information about the individual group retention coefficients. In particular the interrelationship between the variability of the group retention coefficients for a particular amino acid, and the co-correlation of the group retention coefficients between different amino acid residues, allow independent or synergistic effects on solute retention to be explored.

The establishment of a large peptide database allowed the selection of groups of peptides depending on their chromatographic characteristics. Before multiple linear regression analysis of these groups could be examined the influence of various parameters such as sample size and peptide length on the computational results were addressed.

The effect of sample size

The statistical significance of any mathematical procedure is closely dependent on the number of data points available. The availability of several hundred peptides for the derivation of coefficients provided the opportunity for statistical evaluation of the influence of the sample size on the group retention coefficients. Seven different sample sizes were generated which consisted of 25, 30, 50, 70, 100, 200 and 500 peptides. The peptides for the sample size study were all randomly selected from 971 peptides. They were all separated with an octadecylsilica stationary phase with gradient elution between mobile phases of 0.1% trifluoroacetic acid (TFA)–water and 0.1% TFA–acetonitrile (ACN)–water mixtures. The random selection for each sample size was repeated six times to obtain six randomly selected groups of peptides for comparison. Each randomly selected group was then subjected to multiple linear regression to determine the group retention coefficients. Table II lists the six correlation coefficients, (R^2), for the comparison of the observed *MF versus* predicted *MF* within each randomly selected peptide set, the average of these values and the sum of variance for each of the seven sample size groups. The average correlation coefficient, R^2, and the sum of variance for each sample size are plotted in Fig. 1. Examination of the correlation coefficients in Table II indicate that the correlation coefficients increase with decreasing sample sizes, which suggests that sample sizes of less than 30 peptides provide the most accurate group retention coefficients solely for that specific set of peptides. However, the influence of sample size on the final value of each retention coefficient was further assessed through the determination of the sum of variances, listed in Table II and plotted in Fig. 1. The variance of the group retention coefficients for the amino acids between the six data sets within each sample size were calculated according to

$$\text{variance} = \frac{1}{6}\sum_{j=1}^{6}(X_{i,j} - M)^2$$

where M is equal to the mean of the group retention coefficients X_1 to X_6. The variance for each amino acid within each sample size were summed to give the sum of variance, listed in Table II and plotted in Fig. 1. As demonstrated in Fig. 1 both the sum of variances, and R^2 of the group retention coefficients reach an asymptotic plateau when the sample size is greater than 100 peptides. The sum of variances indicates that the small randomly selected peptide data sets produce group retention coefficients that vary considerably for each amino acid between data sets. Thus smaller sample sizes, such as a data set of 10–25 peptides, result in group retention coefficients that are accurate descriptors only for the particular set of peptides from which they were derived. Interestingly, except for a remarkably few exceptions most hydrophobic coefficients have been previously derived using data sets of this sample size and this would account for their lower utility for predicting retention behaviour with unrelated peptides. As is evident from the present study, larger data sets result in lower correlation coefficients but also lower sums of variances and provide amino acid group retention coefficients which are more universally applicable.

This variation between group retention coefficients calculated from small data sets, *i.e.*, less than 100 peptides, may be derived from the fact that the calculated retention coefficient of a particular amino acid in a peptide is not simply a result of its interaction with the chromatographic system, *i.e.*, the non-polar stationary phase and the aquo-organic–ionic modifier elution system. Rather it is a measure of the interaction of a particular amino acid with the chromatographic system in a discrete environment within the peptide. If it is assumed that only those amino acids that are sequentially adjacent have a nearest neighbour effect, or take part in creating a Nearest Neighbour Environment (NNE) for a particular amino acid, then that amino acid can be found in a total of 400 different NNEs. The group retention coefficient of any amino

TABLE II

REPLICATE AND AVERAGE CORRELATION COEFFICIENTS (OBSERVED *MF VERSUS* PREDICTED *MF*), AND THE SUM OF VARIANCE FOR RANDOMLY SELECTED PEPTIDE DATA SETS OF VARIED SAMPLE SIZE

Number of peptides	Replicate number						Average	Sum of variance
	1	2	3	4	5	6		
	Correlation coefficient, R^2							
500	0.49	0.39	0.43	0.40	0.48	0.49	0.46	7.2
200	0.48	0.48	0.47	0.50	0.51	0.49	0.49	16.5
100	0.58	0.52	0.70	0.46	0.47	0.47	0.53	44.7
70	0.65	0.69	0.57	0.47	0.58	0.59	0.59	61.3
50	0.65	0.72	0.60	0.63	0.87	0.66	0.68	157.2
30	0.89	0.85	0.69	0.91	0.63	0.88	0.81	456.2
25	0.97	0.98	0.94	0.92	0.97	0.95	0.96	1504.3

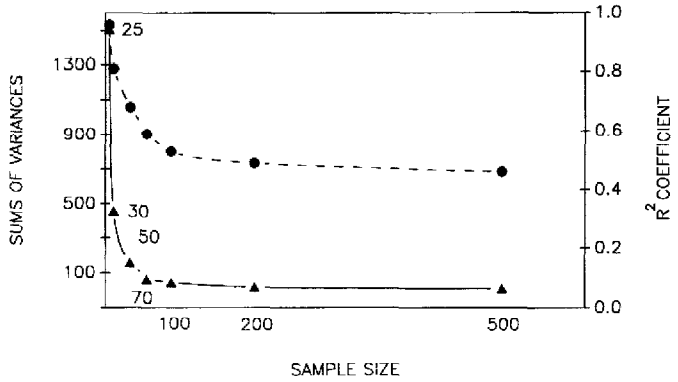

Fig. 1. Dependence of the (▲) sums of variances and (●) correlation coefficient, R^2, for calculated group retention coefficients with varying sample size.

acid will therefore be affected by its NNE. With a small sample size only a very limited number of these NNE will exist for a certain amino acid. As a consequence, it would be anticipated that a smaller variation in the calculated group retention coefficients, *i.e.*, higher correlation coefficient, R^2, would be observed. Conversely, with a large sample, as observed with the 500 peptide sample size data set a particular amino acid will be found in a far greater range of NNE of which the retention coefficient will represent an average. Similar conclusions have been made by Houghten and DeGraw [17] who systematically substituted every residue position of a 13-amino acid peptide with all of the 20 naturally occurring amino acids, thus generating 13 sets of 20 peptides.

Furthermore, if the secondary structure of a peptide is considered, spatially adjacent amino acids will also form part of the NNE. If a particular amino acid has a penchant for specific regions of secondary structure as proposed by Chou and Fasman [24], then the group retention coefficient will include information concerning secondary structure preferences for each amino acid.

The effect of peptide length

As stated earlier, for larger peptides there is generally poorer correlation between the observed retention time and the theoretical retention time based on the summation of the group retention coefficients than for small peptides. From temperature studies on the retention behaviour of a range of peptides we have deduced that these reduced correlations are due to stabilized secondary structures that effectively control the orientation of those amino acid residues that will interact with the stationary phase [19,20]. To study the effect of peptide length on the group retention coefficients, and to establish an optimum length for the generation of hydrophobic retention coefficients, a group of coefficients were calculated from sets of peptides selected from those peptides eluted from an octadecyl stationary phase with a gradient mobile phase of 0.1% TFA–ACN–water. The variation in peptide length ranged from 4–15 residues. In order to generate group retention coefficients which cover a reasonable number of NNEs, each sample size data set comprised more than 100 peptides. Table III lists the maximum peptide length, the number of peptides selected in each group, the

TABLE III

CORRELATION COEFFICIENTS FOR PEPTIDE DATA SETS SELECTED ACCORDING TO PEPTIDE LENGTH

Maximum peptide length	Number of peptides	Correlation coefficient, R^2
15	778	0.60
10	611	0.58
8	478	0.58
5	248	0.66
4	165	0.63

correlation coefficients, R^2, and the sample size. The relative constancy of the correlation coefficients of between 0.58 to 0.66 demonstrates that there is no appreciable effect of peptide length on the group retention coefficients within the range of peptide length considered in this study. From Fig. 2, which shows the frequency distribution of chain length of all the peptides used for this study, it can be seen that by far the greater majority of the peptides are less than 15 residues in length.

Collective grouping of the peptides

As noted in the earlier section, *The effect of sample size*, the minimum size for a set of peptides required to generate consistent group retention coefficients is approximately 100 record entries. In order to satisfy this criterion, the peptide literature data sets were combined according to the nature of either the stationary and/or the mobile phase to produce data sets of greater than 100 peptides. The range of chromatographic conditions was examined (Table I) and 12 groups were chosen as shown in Table IV. The mormalised individual amino acid coefficients for each of the 12 groups are plotted in histograms shown in Fig. 3.

The first 6 groupings in the histograms, MF to MFCTF1C, examine the effect of particular chromatographic parameters on the group retention coefficients. Group 1 represents the group retention coefficients calculated using all the available peptides and their MF. The next 5 groups incorporate a stepwise refinement of the selection criteria, thereby reducing the number of variable parameters. This allows the impact of a particular parameter to be assessed. Group 2 represents the group retention coefficients for all peptides eluted with a gradient of aqueous TFA–ACN as the mobile phase. Group 3 shows the group retention coefficients for peptides eluted from octadecylsilica sorbents. Group 4, designated MFC18TA, corresponds to a chromatographic system incorporating a octadecylsilica column and an aqueous TFA–ACN gradient mobile phase. Group 5 and group 6 represent further refinement of the selection criteria of group 4. Group 5 (MFTF1) contains only data obtained with a flow-rate of 1.0 ml/min, while group 6 (MFTF1C) represents data obtained with a flow-rate of 1.0 ml/min and a column length of 250 nm.

Group 7 and 8, MFTPA and MFC18TPA, contain retention data of peptides eluted with a gradient mobile phase of aqueous 0.1% TFA–1-propanol–ACN (33:77). The group retention coefficients in group 7 are generated from peptides eluted on all stationary phases, while group 8, MFC18TPA, is taken from those peptides eluted

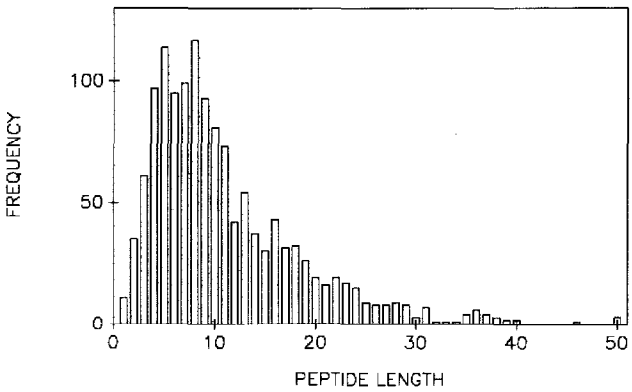

Fig. 2. Frequency distribution histogram of the length of peptides used in this study.

with the same mobile phase but only with octadecylsilica columns. The final four groups, MFC4 to MFC8TA, represent a set of chromatographic conditions in which the nature of the chemically bonded alkyl ligand is changed. Group 9, MFC4, are those group retention coefficients generated from peptides eluted on a *n*-butylsilica bonded stationary phase, and MFC4TA, group 10, those peptides chromatographed on a *n*-butylsilica column and with a gradient mobile phase of 0.1% TFA–ACN–water. The final two groups 11 and 12 contain peptides eluted from an octylsilica column and with 0.1% TFA–ACN–water.

The classification used to generate groups 1–6, MF to MFCTF1C, allows the assessment of the influence of different chromatographic parameters on the group retention coefficients for peptides eluted in a gradient of TFA–ACN–water and with an octadecylsilica column. Table IV shows the correlation coefficients, R^2, for all the

TABLE IV

EXPLANATION OF PEPTIDE RETENTION DATA GROUPINGS

The chromatographic selection criteria are shown in the second column, followed by the number of peptides in each group and the correlation coefficient (observed *versus* expected *MF*).

Database codes	Explanation	Number of peptides	Correlation coefficient, R^2
MF	All peptides	2016	0.29
MFTA	TFA–ACN	1258	0.62
MFC18	RP-C18	1244	0.57
MFC18TA	RP-C18 and TFA–ACN	971	0.66
MFCTF1	RP-C18, TFA–ACN and flow-rate 1.0 ml/min	494	0.69
MFCTF1C	RP-C18, TFA–ACN, flow-rate 1.0 ml/min and column length 250 mm	325	0.69
MFTPA	TFA–propanol–ACN	220	0.69
MFC18TPA	RP-C18 and TFA–propanol–ACN	196	0.69
MFC4	RP-C4	136	0.77
MFC4TA	RP-C4 and TFA–ACN	96	0.80
MFC8	RP-C8	104	0.87
MFC8TA	RP-C8 and TFA–ACN	104	0.87

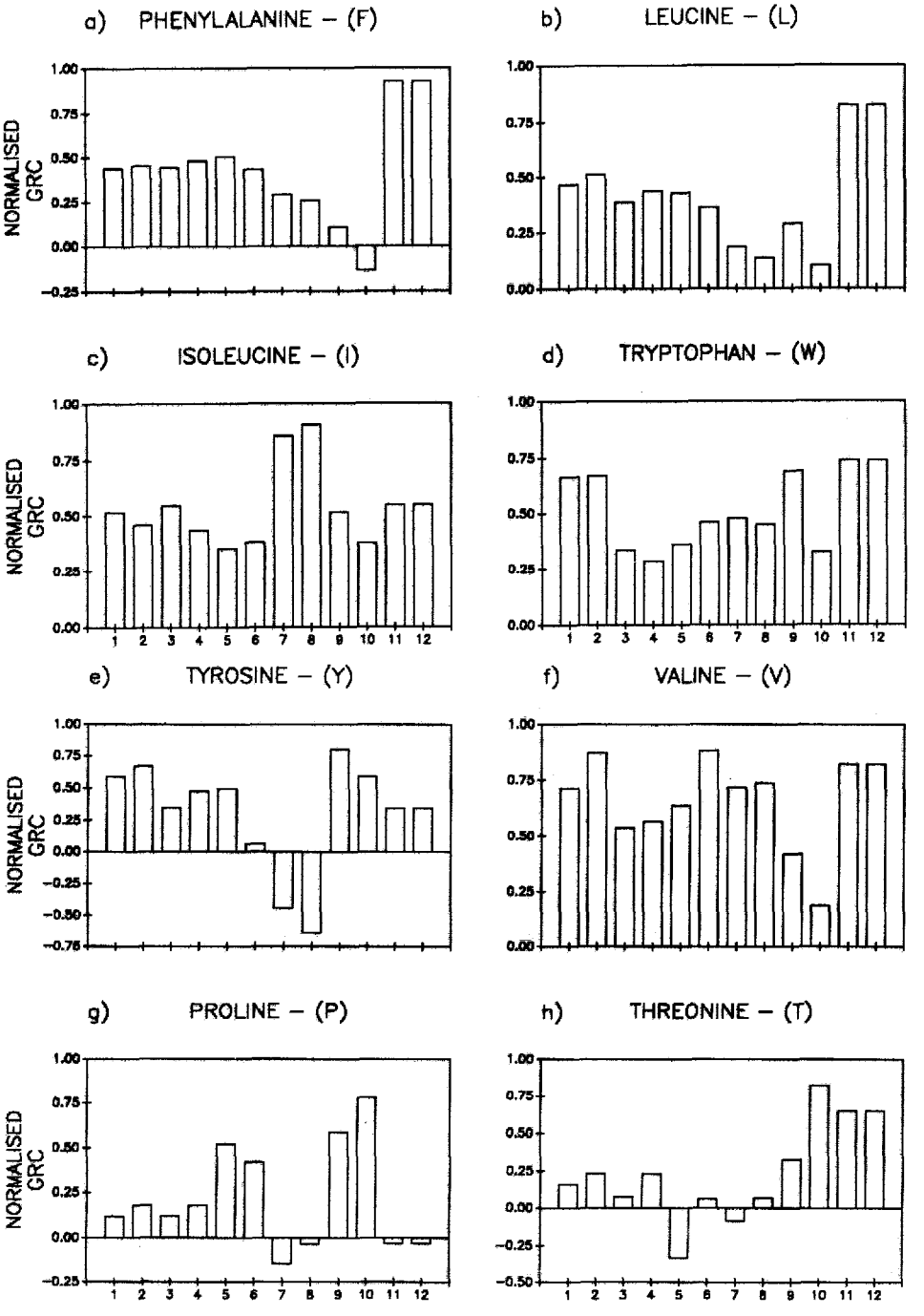
a) PHENYLALANINE – (F)
b) LEUCINE – (L)
c) ISOLEUCINE – (I)
d) TRYPTOPHAN – (W)
e) TYROSINE – (Y)
f) VALINE – (V)
g) PROLINE – (P)
h) THREONINE – (T)
NORMALISED GRC

Fig. 3.

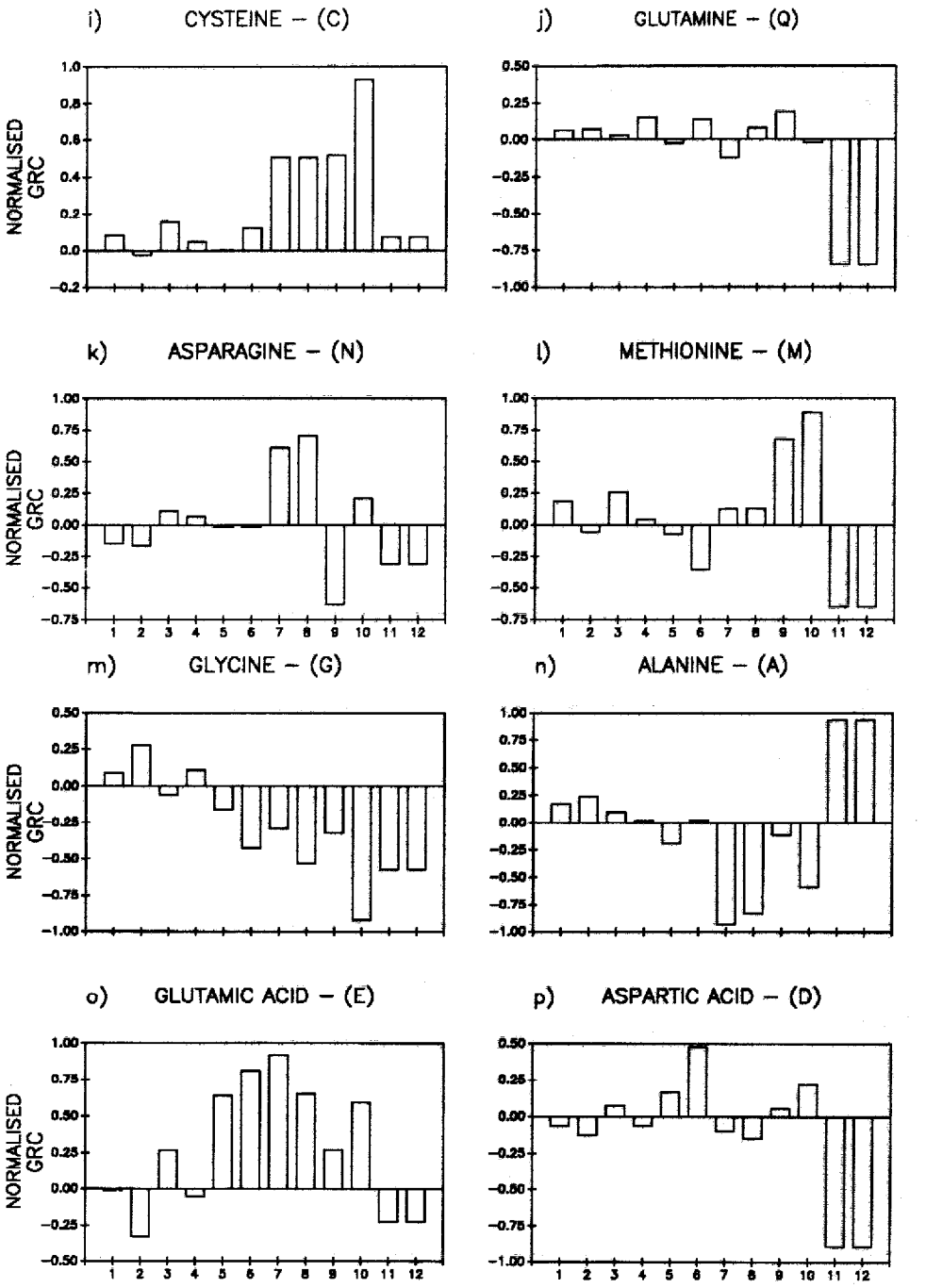

Fig. 3.

(Continued on p. 178)

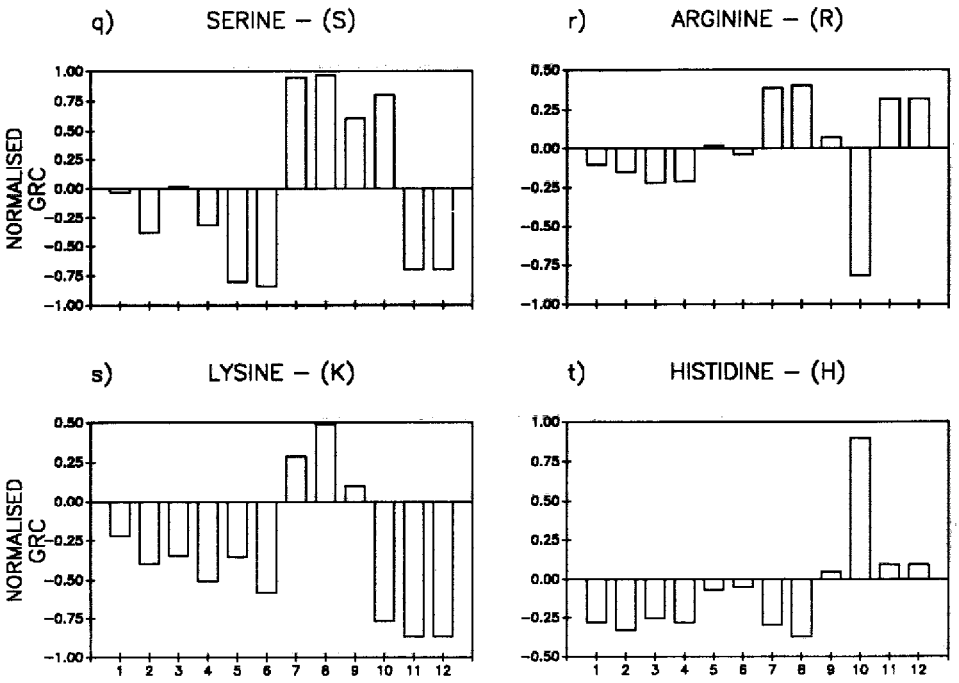

Fig. 3. Amino acid group retention coefficients (GRC) for each group. Groupings: 1 = MF; 2 = MFC18; 3 = MFTA; 4 = MFC18TA; 5 = MFCTF1; 6 = MFCTF1C; 7 = MFTPA; 8 = MFC18TPA; 9 = MFC4; 10 = MFC4TA; 11 = MFC8; 12 = MFC8TA. Table IV provides an explanation of these groupings.

groups of group retention coefficients. Examination of the histograms shown in Fig. 3 demonstrate that for each amino acid, a reasonably constant value for the group retention coefficients is obtained for groups 1–6. The range of R^2 values, as shown in Table III, indicate that all of the selection criteria have an effect on the group retention coefficients. However, once the stationary phase ligand and the mobile phase had been selected the other chromatographic parameters such as flow-rate and column length have only a marginal effect. The general agreement of the group retention coefficients amongst these columns therefore again confirms that the nature of the stationary and mobile phase and not the column length, gradient slope, flow-rate etc., are the dominant variables to be considered, and is consistent with the approach of data grouping used in this study.

The effect of mobile phase composition and alkyl chain length

Numerous investigations have examined various factors involved in the development and application of non-polar bonded stationary phases (for a compendium see refs. 25 and 26), but the underlying separation mechanisms for peptides and proteins are still poorly understood. One of the major factors which control the retention of peptide solutes in RP-HPLC is the relative eluotropicity of the organic modifier. Furthermore, a number of long range Van der Waals, London and Lifshitz intermolecular interations, as well as, dispersion forces, dipole polarisation and

hydrogen-bonding, can also occur between the sample and mobile phase molecules which further influence peptide retention behaviour. The present study considers the effect of two mobile phases aqueous, TFA–ACN and TFA–1-propanol–ACN, on the derived amino acid group retention coefficients. A statistical measure was used to estimate the relative interactive potential of each amino acid with aqueous TFA–ACN and TFA–1-propanol–ACN. This approach involved taking the difference between the coefficients for each amino acid for the pair of mobile phases. The mean and standard deviation of the difference coefficients were then determined according to:

$$\text{mean} = \frac{1}{20} \sum_{i=1}^{20} (XI_i - XII_i)$$

$$\text{standard deviation} = \frac{1}{20} \left\{ \sum_{i=1}^{20} [(XI_i - XII_i) - \text{mean}]^2 \right\}$$

where XI_i is mobile phase I coefficient of amino acid i, XII_i is mobile phase II coefficient and $XI_i - XII_i$ is difference coefficient ΔGRC. Those amino acids which have a difference coefficient, ΔGRC, further than 1 standard deviation from the mean are considered to have significantly different group retention coefficients in the two mobile phases. Table V lists the ΔGRC, the mean and standard deviations.

In changing the mobile phase from TFA–ACN to TFA–1-propanol–ACN the amino acids F, L, I, Y, C and A all have significant difference coefficients. Furthermore, when present within the peptide contact area F, L, Y and A interact more strongly with the octadecylsilica ligand when the mobile phase is TFA–ACN, while I and C interact more strongly when the mobile phase is TFA–1-propanol–ACN. This observation therefore provides an explanation for the experimental observation of selectivity changes or reversals which can occur with a group of peptides when the organic modifier is changed.

The relative interactive potential of the 20 naturally occurring amino acids was also examined through variation in the chain length of the bonded alkylsilica in particular, octadecylsilica, octylsilica and butylsilica. The same statistical measure used to analyse the change in mobile phase was also applied to the comparison of the three stationary phases.

A decrease in the alkyl chain length from 18 to 8 carbon atoms resulted in statistically significant ΔGRCs for the amino acids F, L, W, Q, M, A and D, as listed in Table V. Furthermore, the strongly hydrophobic amino acids F, L and W, all exhibit enhanced values when interacting with the octylsilica column rather than the octadecylsilica as shown in Fig. 3. These amino acids therefore, when present in the peptide contact area, all manifest an increased apparent hydrophobicity. Conversely, the hydrophobicity coefficients for the amino acids Q, M and D were found to be larger with the 18 carbon alkyl chain. Those amino acids that have significant ΔGRCs when comparing a 18- and 4-carbon alkyl chain are F, L, C and H. The two hydrophobic amino acids F and L again exhibited enhanced hydrophobicity with the octadecylsilica column, whereas the amino acids C and H show apparent hydrophobicity increases with the butylsilica. The final comparison between the octylsilica and butylsilica revealed statistically significant ΔGRCs for the amino acids F, L, C, M, A, R and H. In

TABLE V

DIFFERENCE COEFFICIENTS

The italicized numbers indicate those difference coefficients, ΔGRC, further than one standard deviation from the mean. The mean and standard deviation for each set are also given.

Amino acid[a]	C18 – C8	C18 – C4	C8 – C4	C18 – TPA[b]
F	*−4.44*	*6.16*	*10.16*	*2.24*
L	*−3.07*	*2.67*	*5.75*	*2.41*
I	−0.90	0.46	1.35	−3.78
W	*−3.62*	−0.32	3.30	−1.30
Y	0.50	−0.48	−0.98	*4.47*
V	−0.71	1.06	1.77	−0.48
P	−0.84	−2.41	−3.25	0.86
T	−1.17	−1.66	−0.50	0.46
C	−0.24	*−8.81*	*−8.57*	*−4.54*
Q	*2.00*	0.35	−1.66	0.14
N	1.52	−0.58	−2.10	−2.57
M	*3.34*	−4.05	*−7.38*	−0.41
G	1.36	2.06	0.70	1.27
A	*−2.56*	1.70	*4.26*	*2.40*
E	0.35	−1.29	−1.63	−1.40
D	*2.64*	−0.90	−3.54	0.28
S	0.78	−2.21	−2.99	−2.54
R	−2.11	2.44	*4.55*	−2.44
K	1.16	0.84	−0.32	*−3.19*
H	*−2.99*	*−9.41*	*−6.42*	0.75
Mean	−0.41	−0.84	−0.42	−0.40
S.D.	2.10	3.38	4.43	2.24

[a] For abbreviations see Fig. 3.
[b] C18 – TPA is RP-C18 and TFA–1-propanol–ACN.

particular, Fig. 3 shows that C, M and H increased GRC values with the *n*-butylsilica sorbent.

Currently, a full molecular rationalisation of the varying specificities of the different alkyl ligands in terms of specific ligand–solute interactions is not feasible. This difficulty is primarily due to limited amount of information available on the exact nature of the sorbent surface. The main physical and chemical parameters used to characterise chemically modified, microparticulate *n*-alkylsilicas are specific surface area, mean pore diameter, specific pore volume, mean particle size, alkyl chain length and alkyl ligand density. However, knowledge of these bulk parameters gives very little information in terms of how the ligand interacts with a solute. In particular, the exact nature of the dynamic structure of the bonded phase has not been fully characterised. However it has been demonstrated by NMR studies that under conditions typical of chromatographic separations, the alkyl chains may assume different conformations [27]. These depend on the interaction of the chains with themselves, with neighbouring chains, with the "capping" groups, *e.g.*, trimethylsilyl groups, with the mobile phase and finally the solute. It has also been demonstrated that increasing the polarity of the mobile phase increases the mobility of the bonded alkyl groups [28]. All of these factors

will therefore influence the affinity of the solute for the non-polar sorbent which is ultimately manifested as the experimental retention time. The range of values of the group retention coefficients generated with different stationary phases indicates that the chain length exerts a dramatic effect on the way in which the surface of a peptide solute is probed by the alkyl ligand during the retention process. In accord with the concepts of the solvophobic theory, these results confirm that the surface area of the peptide solute involved with the interaction with the non-polar sorbent will vary significantly with a change in the characteristics of either the stationary or the mobile phase.

The procedure of group retention contribution mapping also allows the interactive specificities for particular amino acids to be evaluated. For instance, F, L, I and W all have relatively similar group retention coefficients when their corresponding peptides are eluted in TFA–ACN–water, but the values of the group retention coefficients vary noticeably when the mobile phase is changed to TFA–1-propanol–ACN–water (Table V and Fig. 3, columns 4 and 8). Clearly the hydrophobic nature of an amino acid within a peptide is dependent on the solvated microenvironment in which it is located. For instance, within a peptide structure, the amino acid Y is manifestly more hydrophobic when eluted with TFA–ACN–water than with TFA–1-propanol–ACN–water. Many other amino acids show this same phenomenon as can be seen in Fig. 3. Changes in the stationary phase also have a comparable effect. The retention coefficient changes for amino acid M indicate that within a peptide structure, this amino acid is quite hydrophobic when interacting with a butylsilica column but becomes less so with an octylsilica sorbent.

One point of interest is that the amino acids A, C, F, H and L have significant ΔGRCs for at least three of the four comparisons. What feature is it that these amino acids have that make their retention on RP-HPLC so susceptible to a change in either stationary or mobile phase composition? Because of the atomic features of these amino acids, it would appear that these differences cannot be attributed to hydrogen bond interactions. Similarly, these differences cannot be ascribed to sensitivity to side chain ionisation phenomena. Since the dynamic structure of the stationary phase is not known at the requisite level of molecular definition the precise mechanism by which these amino acids within a peptide structure interdigitate with the hydrocarbonaceous ligand remains to be elucidated. Until this is studied in greater detail with even more comprehensive data bases than those used in the present study, fully mechanistic interaction models will be difficult to produce. However the emergence of such data bases and descriptive molecular interaction models would provide valuable additional insight into the physicochemical origin of the hydrophobic effect, which even after 30 years of intensive research by many groups of investigators still remains to be resolved.

CONCLUSION

This study presents and validates a new mathematical approach to the calculation of group retention coefficients. The establishment of a large database of peptide retention data for the generation of the group retention coefficients allowed the influence of a number of parameters on the group retention coefficients to be assessed. In particular it was found that to generate consistent group retention coefficients the retention data of at least 100 peptides is required. The database

contained peptide retention data from peptides eluted from a range of chemically modified *n*-alkylsilicas and aquo-organic mobile phases. This allowed the examination of the variation of the interaction of the individual amino acids with the alkyl ligands under a range of chromatographic conditions.

ACKNOWLEDGEMENTS

These investigations were supported by the Australian Research Council and National Health and Medical Research Council of Australia (M.T.W.H.). The support of the Potter Foundation and the Buckland Foundation is also gratefully acknowledged.

REFERENCES

1 M. A. Stadalius, H. S. Gold and L. R. Snyder, *J. Chromatogr.*, 296 (1984) 31.
2 M. I Aguilar, A. N. Hodder and M. T. W. Hearn, *J. Chromatogr.*, 327 (1985) 115.
3 M. T. W. Hearn and M. I. Aguilar, *J. Chromatogr.*, 359 (1986) 31.
4 M. T. W. Hearn and M. I. Aguilar, *J. Chromatogr.*, 392 (1987) 33.
5 M. T. W. Hearn and M. I. Aguilar, *J. Chromatogr.*, 352 (1986) 35.
6 S. J. Su, B. Grego, B. Niven and M. T. W. Hearn, *J. Liq. Chromatogr.*, 4 (1981) 1745.
7 R. F. Rekker and H. M. De Kart, *Eur. J. Med. Chem.*, 14 (1979) 479.
8 J. L. Meek and Z. L. Rosetti, *J. Chromatogr.*, 211 (1981) 15.
9 V. Pliska and M. Fauchere, in E. Gross (Editor), *Peptides, Structure and Biological Function*, Pierce, Rockford, IL, 1979, p. 249.
10 A. Segrest and R. Feldmann, *J. Mol. Biol.*, 87 (1974) 853.
11 K. J. Wilson, A. Honegger, R. P. Stotzel and G. J. Hughes, *Biochem. J.*, 199 (1981) 31.
12 C. A. Browne, H. P. J. Bennett and S. Solomon, in M. T. W. Hearn, F. E. Regnier and C. T. Wehr (Editors), *High Performance Liquid Chromatography*, Academic Press, New York, 1983, p. 65.
13 Y. Nazaki and C. Tanford, *J. Biol. Chem.*, 246 (1971) 2211.
14 T. Sasagawa, T. Okuyama and D. C. Teller, *J. Chromatogr.*, 240 (1982) 329.
15 D. Guo, C. T. Mant, A. K. Taneja, J. M. R. Parker and R. S. Hodges, *J. Chromatogr.*, 359 (1986) 499.
16 J. M. R. Parker, D. Guo and R. S. Hodges, *Biochem.*, 25 (1986) 5425.
17 V. Houghten and S. T. DeGraw, *J. Chromatogr.*, 386 (1987) 223.
18 C. T. Mant, N. E. Zhou and R. S. Hodges, *J. Chromatogr.*, 476 (1989) 363.
19 A. W. Purcell, M. I. Aguilar and M. T. W. Hearn, *J. Chromatogr.*, 476 (1989) 113.
20 A. W. Purcell, M. I. Aguilar and M. T. W. Hearn, *J. Chromatogr.*, 476 (1989) 125.
21 F. E. Regnier, *Science (Washington, D.C.)*, 228 (1987) 319.
22 M. L. Heinitz, E. Flanigan, R. C. Orlowsi and F. E. Regnier, *J. Chromatogr.*, 443 (1988) 229.
23 Cs. Horváth, W. R. Melander and I. Molnar, *J. Chromatogr.*, 125 (1976) 129.
24 P. Y. Chou and G. D. Fasman, *Ann. Rev. Biochem.*, 47 (1978) 251.
25 K. K. Unger, *Porous Silica*, Elsevier, Amsterdam, 1979.
26 K. K. Unger, *Packings and Stationary Phases in Chromatographic Techniques*, Marcel Dekker, New York, 1989.
27 E. Bayer, A. Paulus, B. Peters, G. Laupp, J. Reniers and K. Albert, *J. Chromatogr.*, 364 (1986) 25.
28 J. Nawrocki and B. Buszewski, *J. Chromatogr.*, 449 (1988) 1.
29 M. T. W. Hearn, *High Performance Liquid Chromatography of Peptide and Proteins: Separation, Analysis and Conformation*, CRC Press, Boca Raton, FL, 1990, in press.
30 G. E. Deibler, L. F. Boyd, R. E. Martenson and M. W. Kies, *J. Chromatogr.*, 326 (1985) 433.
31 S. James and H. P. I. Bennett, *J. Chromatogr.*, 326 (1985) 329.
32 H. W. Lahm and S. Stein, *J. Chromatogr.*, 326 (1985) 357.
33 M. D. Moore, P. Q. Behrens and A. F. Riggs, *J. Biol. Chem.*, 261 (1986) 10511.
34 I. Correas, D. W. Speicher and V. T. Marches, *J. Biol. Chem.*, 261 (1986) 13362.
35 J. A. Cromlish, N. G. Seidah and M. Chretien, *J. Biol. Chem.*, 261 (1986) 10859.

36 D. Barra, M. E. Schinina, W. H. Bannister, J. V. Bannister and F. Bossa, *J. Biol. Chem.*, 262 (1987) 1001.
37 H. C. Ludwig, F. Lottspeich, A. Henschen, R. Ladenstein and A. Bachen, *J. Biol. Chem.*, 262 (1987) 1016.
38 T. Sueyoshi, T. Miyata, N. Hashimoto, H. Kato, H. Hayashida, T. Miyata and S. Iwanaga, *J. Biol. Chem.*, 262 (1987) 2768.
39 H. S. Lu, P. J. Yuan and R. W. Gracy, *J. Biol. Chem.*, 259 (1984) 11958.
40 D. Barra, M. E. Schinina, M. Simmaco, J. V. Bannister, W. H. Bannister, G. Rotilio and F. Bossa, *J. Biol. Chem.*, 259 (1984) 12595.
41 P. H. Lai, R. Everett, F. F. Wang, T. Arakawa and E. Goldwasser, *J. Biol. Chem.*, 261 (1986) 3116.
42 P. J. Neame, J. E. Christner and J. R. Baker, *J. Biol. Chem.*, 261 (1986) 3519.
43 K. M. Zsebo, H. S. Lu, J. C. Fieschko, L. Goldstein, J. Davis, K. Duker, S. V. Suggs, P. H. Lai and G. A. Bitter, *J. Biol. Chem.*, 261 (1986) 5858.
44 H. Hirano, C. Fukazawa and K. Harada, *J. Biol. Chem.*, 259 (1984) 14371.
45 B. Grego, F. Lambrou and M. T. W. Hearn, *J. Chromatogr.*, 266 (1983) 89.
46 S. T. Davrat-Larrogue, K. Brew and R. E. Fenna, *J. Biol. Chem.*, 261 (1986) 3607.
47 T. Suzuki, *J. Biol. Chem.*, 261 (1986) 3692.
48 K. G. Linden and W. F. Benisek, *J. Biol. Chem.*, 261 (1986) 6454.
49 T. Yubisui, T. Miyata, S. Iwanaga, M. Tamura and M. Takeshita, *J. Biochem.*, 99 (1986) 407.
50 Y. Kitagawa, S. Tsunasawa, N. Tanaka and Y. Katsube, *J. Biochem.*, 99 (1986) 1289.
51 J. Hirabayashi, H. Kawasaki, K. Suzuki and K. Kasai, *J. Biochem.*, 101 (1987) 775.
52 T. Takagi and K. Konishi, *J. Biochem.*, 95 (1984) 1603.
53 M. Singhofer-Wowra, L. Clayton, P. Dawson, K. Gull and M. Little, *Eur. J. Biochem.*, 161 (1986) 669.
54 O. Markovic and H. Jornvall, *Eur. J. Biochem.*, 158 (1986) 455.
55 O. V. Beg, H. von Bahr-Lindstrom, Z. H. Zaidi and H. Jornvall, *Eur. J. Biochem.*, 159 (1986) 195.
56 H. Wada, T. Takai and T. Tanabe, *Eur. J. Biochem.*, 167 (1987) 13.
57 F. Tokunaga, T. Miyata, T. Nakamura, T. Morita, K. Kuma, T. Miyata and S. Iwanaga, *Eur. J. Biochem.*, 167 (1987) 405.
58 F. Schoentgen, F. Saccoccio, J. Jolles, I. Bernier and P. Jolles, *Eur. J. Biochem.*, 166 (1987) 333.
59 T. Miyata, K. Usui and S. Iwanaga, *J. Biochem.*, 95 (1984) 1793.
60 T. Nakamura, T. Hirai, F. Tokunaga, S. Kawabata and S. Iwanaga, *J. Biochem.*, 101 (1987) 1297.
61 T. Hirao and K. Takahashi, *J. Biochem.*, 96 (1984) 775.
62 Y. Hamazume, T. Mega and T. Ikenaka, *J. Biochem.*, 95 (1984) 1633.
63 S.-H. Kim, S. Hara, S. Hase, T. Ikenaka, H. Tada, K. Kitamura and N. Kaizuma, *J. Biochem.*, 98 (1985) 435.
64 A. J. P. Martin and R. L. M. Synge, *Biochem. J.*, 35 (1941) 1358.
65 T. Fujita, I. Iwaja and C. Hansch, *J. Am. Chem. Soc.*, 86 (1964) 5175.
66 L. P. Hammett and H. L. Pflunger, *J. Am. Chem. Soc.*, 55 (1933) 4079.