Journal of Chromatography, 386 (1987) 223–228
Elsevier Science Publishers B.V., Amsterdam — Printed in The Netherlands

CHROMSYMP. 1015

EFFECT OF POSITIONAL ENVIRONMENTAL DOMAINS ON THE VARIATION OF HIGH-PERFORMANCE LIQUID CHROMATOGRAPHIC PEPTIDE RETENTION COEFFICIENTS

RICHARD A. HOUGHTEN and STEPHEN T. DeGRAW
Research Institute of Scripps Clinic, Department of Molecular Biology, 10666 North Torrey Pines Road, La Jolla, CA 92037 (U.S.A.)

SUMMARY

While attempting to derive a set of high-performance liquid chromatographic amino acid retention coefficients from a set of 298 related peptide analogues of a 13-amino acid peptide, we found that each position within the peptide would require a different set of retention coefficients to accurately predict peptide retention times. Furthermore, our results show that peptides having the same amino composition but slightly different sequences can have very different retention times. We believe that individual sequence domains and the resulting differences in the solid phase–mobile phase interactions must be taken into account for the accurate prediction of peptide retention times.

INTRODUCTION

There are a number of reports in the literature which correlate high-performance liquid chromatography (HPLC) peptide elution times with specific retention coefficients[1–4] for the twenty naturally occuring amino acids. These retention coefficients have been used in various formulas[1–4] to give approximations of peptide elution times with correlations as high as 0.98. However, substantial variations exist which have been explained as due to ill defined anomalous conformational effects or stationary-phase interactions[5]. Since retention coefficients are normally derived from large numbers of unrelated peptides having a wide range of lengths, net charges and hydrophobicities, the determination of the reasons for deviations from theory have been obscured. In order to derive a set of retention coefficients for our own use, 298 related peptide analogues of a 13-amino acid sequence from the influenza virus hemagglutinin (HA 1:98–110), originally synthesized for use in antigen–antibody interaction studies[6], were analyzed at pH 2.1 by reversed-phase (RP) HPLC. The peptides were prepared by the method of simultaneous multiple peptide synthesis[6,7]. The 298 peptide analogues of the original sequence, YPYDVPDYASLRS*, consisted of 260

* Abbreviations for the amino acid residues are: A (Ala) = alanine, R (Arg) = arginine, N (Asn) = asparagine, D (Asp) = aspartic acid, C (Cys) = cysteine; Q (Gln) = glutamine, E (Glu) = glutamic acid, G (Gly) = glycine, H (His) = histidine, I (Ile) = isoleucine, L (Leu) = leucine, K (Lys) = lysine, M (Met) = methionine, F (Phe) = phenylalanine, P (Pro) = proline, S (Ser) = serine, T (Thr) = threonine, W (Trp) = tryptophan, Y (Tyr) = tyrosine, V (Val) = valine.

0021-9673/87/$03.50

TABLE I

RELATIVE RP-HPLC AMINO ACID RETENTION COEFFICIENTS FOR 260 PEPTIDE ANALOGUES OF THE SAME SEQUENCE (YP-YDVPDYASLRS), WITH GLYCINE ARBITRARILY SET TO ZERO

Note should be made of the difference in range of variation of the coefficients at each position and the minimum and maximum retention coefficient for individual amino acids. The effect of substitution of D-amino acids are shown at the bottom of the table. Min. = minimum, Max. = maximum, Ave. = average.

Amino acid	*Y (Tyr)*	*P (Pro)*	*Y (Tyr)*	*D (Asp)*	*V (Val)*	*P (Pro)*	*D (Asp)*	*Y (Tyr)*	*A (Ala)*	*S (Ser)*	*L (Leu)*	*R (Arg)*	*S (Ser)*	*Avg.*	*Min.*	*Max.*
Lys	−3.21	−0.82	−2.70	−3.51	−3.11	−2.24	−3.20	−3.51	−1.41	−1.21	−1.71	−1.63	−1.51	−2.29	−3.51	−0.82
His	−2.81	−1.10	−3.50	−2.50	−2.87	−2.19	−1.80	−3.27	−1.60	−1.10	−1.35	−1.52	−2.22	−2.14	−3.50	−1.10
Gln	−1.20	−0.45	−1.10	−2.62	−1.21	0.08	−0.20	−1.96	0.10	−0.04	−0.83	−1.01	−0.42	−1.69	−2.62	0.10
Arg	−1.21	−1.80	−1.50	−0.80	−1.13	1.05	−2.60	−1.45	−1.50	−1.01	−0.58	−0.75	−1.10	−1.11	−2.60	1.05
Asn	−1.10	−0.35	−0.60	−1.60	−1.10	−1.31	0.10	−1.42	−0.30	−0.32	−0.42	−0.74	−0.58	−0.75	−1.60	0.10
Ser	0.60	−1.00	−1.00	−0.01	−0.32	−0.52	0.50	−0.86	−0.20	−0.81	−0.23	−0.24	−1.05	−0.40	−1.05	0.60
Asp	−0.80	0.00	0.20	0.10	0.55	−1.03	1.80	−0.81	−0.81	0.23	0.64	1.00	−0.10	−0.07	−1.03	1.80
Gly	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Glu	−0.70	0.70	−0.40	0.55	1.60	1.48	1.00	−1.01	1.50	1.74	1.23	1.02	−0.36	0.64	−1.01	1.74
Ala	−0.30	0.30	1.80	0.42	1.34	−1.28	2.40	1.59	1.50	1.05	1.88	1.28	2.03	1.08	−1.28	2.40
Thr	−1.00	0.20	0.53	0.21	2.20	2.01	0.70	0.51	1.40	0.89	1.01	1.43	2.23	0.94	−1.00	2.23
Pro	0.66	2.10	3.75	1.42	3.18	0.97	2.30	3.39	1.70	2.21	2.32	2.65	3.42	2.31	0.66	3.75
Cys	1.00	2.00	3.50	1.85	4.22	3.68	3.70	4.34	2.48	2.71	3.21	3.98	3.74	3.10	1.00	4.34
Val	2.41	3.20	6.02	1.52	5.98	2.81	4.90	7.77	3.01	3.71	4.37	5.16	3.68	4.20	1.52	7.77
Tyr	3.20	3.90	6.30	2.75	6.42	5.33	5.30	7.79	3.52	3.76	4.78	5.67	7.33	5.08	2.75	7.79
Met	7.78	4.30	6.30	3.25	7.41	6.72	5.90	8.67	6.81	4.87	5.44	7.22	7.43	6.32	3.25	8.67
Ile	4.10	3.94	8.35	5.10	8.24	8.01	6.71	7.34	7.72	5.41	7.72	9.01	9.92	7.04	3.94	9.92
Leu	5.52	6.40	10.95	5.05	9.24	7.30	7.30	5.47	8.20	7.21	8.34	8.82	9.01	7.60	5.05	10.95
Phe	8.90	6.12	11.21	7.60	11.11	9.71	9.30	8.74	10.41	7.68	10.41	12.68	10.96	9.11	6.12	12.68
Trp	11.70	8.71	12.71	9.15	12.01	12.02	10.82	10.96	11.22	8.71	11.71	13.74	11.67	11.17	8.71	13.74
Min.	−3.21	−1.80	−3.50	−3.51	−3.11	−2.24	−3.20	−3.51	−1.60	−1.21	−1.71	−1.63	−2.22	−2.29		
Max.	11.70	8.71	12.71	9.15	12.01	12.02	10.82	10.96	11.22	8.71	11.71	13.74	11.67	11.17		
Range	14.91	10.51	16.21	12.66	15.12	14.26	14.02	14.47	12.82	9.92	13.42	15.37	13.89	13.46		
Ave.	1.68	1.82	3.04	1.40	3.19	2.73	2.75	2.61	2.77	2.28	2.90	3.41	3.20	2.53		
D-Amino acid	0.30	−1.04	−0.78	0.71	1.08	1.50	0.51	−0.37	−1.10	−0.67	1.36	0.10	–	0.13	−1.10	1.50

substitution analogues in which the original amino acid at each position of the peptide was replaced with each of the 20 natural amino acids, 13 analogues in which original amino acid was replaced individually by its D-amino acid counterpart, 12 analogues which were single-position omission analogues, and 13 analogues which were single-position glycine insertion analogues. From the HPLC analysis, retention coefficients were determined empirically, by comparing the retention times of the peptide substitution analogues relative to each other and to the original peptide. While we found that predicted and experimentally determined retention times on average agreed, we determined that, in fact, each position within a peptide requires a different set of amino acid retention coefficients to accurately predict peptide retention times. The purpose of this paper is to illustrate that, as currently defined, one general set of amino acid retention coefficients cannot usefully predict elution times for peptides.

EXPERIMENTAL

Peptide resins were synthesized 100 at a time by the method of simultaneous multiple peptide synthesis[5] and the peptides were cleaved from the resins with a multiple hydrogen fluoride cleavage apparatus (Multiple Peptide Systems, La Jolla, CA, U.S.A.), capable of cleaving up to 24 peptide resins simultaneously[7]. The average purity of these peptides by analytical HPLC immediately following hydrogen fluoride cleavage was 85% (65%–95%). The peptides were not further purified. Amino acid composition, as determined by amino acid analysis, was ± 10% of theory.

The HPLC system consisted of two Beckman (Beckman Instruments, Anaheim, CA, U.S.A.) 110A pumps, a Beckman 421 controller, a Hitachi (supplied by Cole Scientific, Calabasas, CA, U.S.A.) 100-20 variable-wavelength detector with an Altex (Beckman Instruments) 20-μl flow-cell and a BioRad (Richmond, CA, U.S.A.) As-48 autosampler with a 20-μl injection loop. Peptide retention times were determined at room temperature on a Vydac (Hesperia, CA, U.S.A.) peptide and protein C_{18} column and a Vydac pH-stable C_8 column (both were 25 cm × 4.6 mm I.D., 5 μm particle size). The linear gradient was 15% acetonitrile in deionized distilled water at zero time to 40% acetonitrile in deionized distilled water in 50 min. The flow-rate was 1.0 ml/min with detection at 215 nm[8]. All mobile phases contained 0.1% trifluoroacetic acid (pH 2.1). Mobile phases were filtered and simultaneously degassed by passage through a 0.45-μm PTFE filter (Lazar Scientific, Los Angeles, CA, U.S.A.). Since all of the peptides in this study were single-point replacement, insertion, or omission analogues of the same original sequence, it was possible to use a shallow gradient (0.5%/min) for all peptide sequences, thus ensuring high resolution with short analysis time.

The peptide analogues were initially chromatographed individually, with the original peptide present in the injected sample as an internal standard, until all twenty amino acid replacement analogues at a given position had been analyzed (columns, Table I). This was done for all 260 analogues (20 analogues × 13 positions) under the same chromatographic conditions. To confirm the results obtained from the individual analysis, groups of six of the substitution analogues from each position were analyzed as a mixture. This mixture was selected to include analogues having two of the most hydrophilic amino acid replacements (*i.e.*, lysine and arginine), two of the

most hydrophobic replacements (*i.e.* phenylalanine and isoleucine), one neutral amino acid substitution (glycine), and the original peptide as an internal standard. The RP-HPLC of this mixture confirmed the range of variance that was evident when the peptides were analyzed individually, and demonstrated that the presence of other peptides in the sample had no effect on the rate of elution. The peptides were also analyzed as they appear in a row across Table I, in which the same amino acid (*i.e.* Lys, row 1) had been substituted at each of the positions in the peptide.

RESULTS AND DISCUSSION

To compare the retention times of the twenty substitution analogues at each position with one another, the retention time for the glycine analogue in each set was subtracted from the retention times of the other analogues to yield a set of retention coefficients normalized to glycine (Gly = 0, Table I, Fig. 1). These retention coefficients are listed in Table I for each of the twenty amino acid replacements at every position in the peptide. The current concept and use of amino acid retention coefficients should, theoretically, give an identical coefficient for the same amino acid at each position. Likewise, the range of amino acid retention coefficients at each position should be the same. Table I demonstrates that the same amino acid at different positions within the peptide sequence often has very different retention coefficients. The actual range of coefficients for each position is represented graphically in Fig. 1. This bar graph dramatically illustrates that the ranges of coefficients are quite variable from position to position within the sequence. The data in Table I indicate that thirteen different sets of retention coefficients, one for each position in the pep-

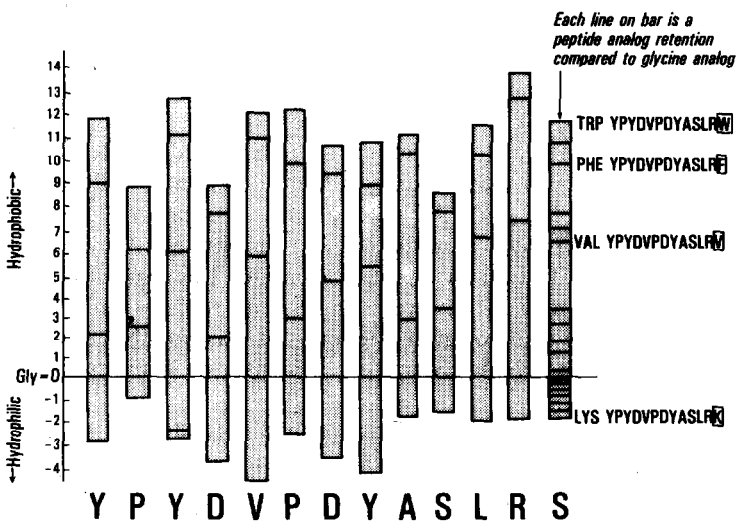

Fig. 1. Bar graph illustrating Table I. Each vertical bar represents the range of retention coefficients normalized to glycine, from the most hydrophobic to the most hydrophilic amino acid, for the twenty peptide analogues at each position with the most hydrophilic retention coefficient set to zero. The lines indicated on the far right bar represent, respectively, the relative retention coefficients of W, F, I, L, M, Y, V, C, P, T, A, E, G, D, S, N, R, Q, H, and K.

tide, would be required to accurately predict the retention behaviour of these 260 related peptides. While we found changes from the expected order of elution, this was normally consistent with that reported by other investigators[1–4]. The possibility that our results were due to ion exchange effects of the peptide interacting with the unreacted silica of the C_{18} column was considered[9]. This possibility was excluded after a column warranted by the manufacturer to be exhaustively endcapped (Vydac pH-stable C_8) was found to give results identical to those found using the C_{18} columns.

Peptides with the same amino acid composition but slightly different sequences (so called "SCDS analogues") were then examined. One set of these SCDS peptides were glycine insertion analogues, in which one glycine was inserted into the original sequence between each position making a set of fourteen residue peptides (*i.e.* YPYGDVPDYASLRS, YPGYDVPDYASLRS, etc.). These glycine insertion analogues had different elution times on RP-HPLC, both when chromatographed individually or together as a mixture (Fig. 2). Since the compositions of these peptides were identical, and the sequences varied only by the insertion of a single glycine, the elution times would have been identical if one general set of retention coefficients were sufficient for prediction of elution times. The actual elution times for these SCDS peptides varied significantly depending on the location of the glycine insertion.

To examine the importance of small changes in chirality on retention behavior, single-position D-amino acid replacements at each position of the original peptide were examined. The D-amino acid analogues were found to have different retention times relative not only to the L-amino acid control peptide but also to one another (bottom, Table I). While it is known that peptide RP-HPLC retention interactions are complex, we have attempted to reduce the variables to a workable number by

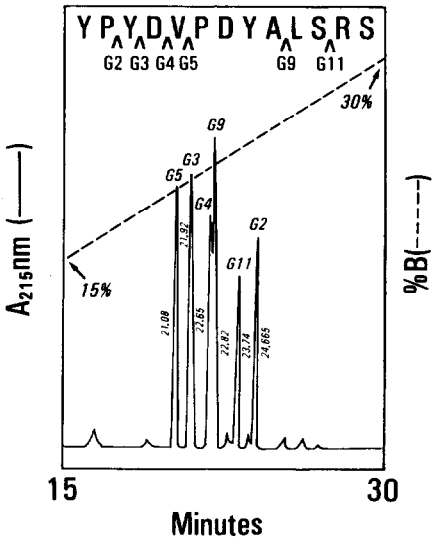

Fig. 2. Variation of RP-HPLC retention times for glycine insertion analogues. Analogues were synthesized in which a glycine residue was inserted between adjacent residues in YPYDVPDYALSRS (*i.e.* G5 = YPYDVGPDDYALSRS).

keeping the peptide population to a similar size (twelve to fourteen amino acids) and by analyzing peptides having only small changes in composition or sequence. Our results show that peptides having the same overall amino acid composition but different sequence can have very different retention times. This indicates that as currently defined, one general set of amino acid retention coefficients cannot usefully predict elution times for peptides. We believe that conformations induced by stationary phase interactions, mobile phase interactions or individual sequence domains[10,11] are significant factors in the retention behaviour of peptides and must be taken into account for the prediction of peptide retention times. Interestingly, these interactions appear, in some respects, to model the binding of a peptide antigen with its antibody[6]. All aspects of the above are being investigated further.

ACKNOWLEDGEMENTS

We are grateful to Sarah Hoffmann and John Ostresh for assisting in the editing and to Susan Null for the preparation of this manuscript.

REFERENCES

1 J. L. Meek, *Proc. Natl. Acad. Sci. U.S.A.*, 77 (1980) 1632–1635.
2 M. J. O'Hare and E. C. Nice, *J. Chromatogr.*, 171 (1979) 209–226.
3 T. Sasagawa, T. Okuyama and D. C. Teller, *J. Chromatogr.*, 240 (1982) 329–340.
4 D. Guo, C. T. Mant, J. M. R. Parker and R. S. Hodges, in C. M. Deber, V. T. Hruby and K. D. Kopple (Editors), *Proc. of the Ninth Amer. Pept. Symposium*, Pierce Chemical Co., 1985, pp. 23–26.
5 M. T. W. Hearn and M. I. Aguilar, *J. Chromatogr.*, 352 (1986) 35–66.
6 R. A. Houghten, *Proc. Natl. Acad. Sci. U.S.A.*, 82 (1985) 5131–5135.
7 R. A. Houghten, M. K. Bray, S. T. DeGraw, C. J. Kirby, *Int. J. Pept. Prot. Res.*, 27 (1986) 675–680.
8 G. Winkler, P. Wolschann, P. Briza, F. Heinz and C. Kunz, *J. Chromatogr.*, 347 (1985) 83–88.
9 H. Gaertner and A. Puigserver, *J. Chromatogr.*, 350 (1985) 279–284.
10 P. Manavalan and P. K. Ponnuswamy, *Nature (London)*, 275 (1978) 673–674.
11 J. Janin, *Nature (London)*, 277 (1979) 491–492.