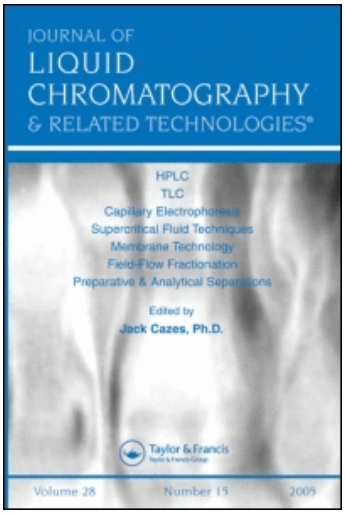

Journal of Liquid Chromatography & Related Technologies
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713597273

Comparative Study of Different Predictive Methods of Peptides Retention Time on Chromatographic Reversed-Phase Columns

B. De Collongue[a]; N. M. Gosselet[a]; B. Sébille[a]; B. Schoot[b]
[a] Laboratoire de Physico-Chimie de Biopolymères, CNRS-Université Paris XII, Thiais, France [b] Roussel-Uclaf, Romainville, France

To cite this Article De Collongue, B. , Gosselet, N. M. , Sébille, B. and Schoot, B.(1994) 'Comparative Study of Different Predictive Methods of Peptides Retention Time on Chromatographic Reversed-Phase Columns', Journal of Liquid Chromatography & Related Technologies, 17: 20, 4349 — 4364
To link to this Article: DOI: 10.1080/10826079408013622
URL: http://dx.doi.org/10.1080/10826079408013622

JOURNAL OF LIQUID CHROMATOGRAPHY, 17(20), 4349-4364 (1994)

COMPARATIVE STUDY OF DIFFERENT PREDICTIVE METHODS OF PEPTIDES RETENTION TIME ON CHROMATOGRAPHIC REVERSED-PHASE COLUMNS

B. DE COLLONGUE[1], N. M. GOSSELET[1], B. SÉBILLE[1], AND B. SCHOOT[2]

[1]Laboratoire de Physico-Chimie de Biopolymères
CNRS-Université Paris XII, Unité mixte 27
2 rue Henri Dunant, 94320 Thiais, France
[2]Roussel-Uclaf
102 route de Noisy, Romainville, France

ABSTRACT

Two peptides samples were studied in order to compare the validity of some predictive retention times. We examined the predictive retention times stated by Guo and based on the relative hydrophobic contribution of each amino acid residue, by Mant who takes into account the length of the peptides and by Chabanet who discriminates three groups of amino acids. Although Chabanet's model provided good results, we found that a more general correlation was obtained with Mant's retention prediction especially in the case of peptides containing a high fraction of basic residues.

INTRODUCTION

Peptide mapping is a commonly used technique to analyse the structure of a protein. This technique involves the enzymatic hydrolysis of the protein, followed by a reversed-phase high performance liquid chromatography (RP-HPLC) of the digest. RP-HPLC allows a good separation of a large range of peptides on the basis of small changes in polarity and length when a A-B gradient eluent is applied (A : H_2O ; 0,1% trifluoro-acetic acid (T.F.A) / B : acetonitrile (CH_3CN) ; 0,1 % T.F.A).

Furthermore, RP-HPLC provides information about the amino-acid composition of peptides : thus in 1981, Meek [1, 2] could predict the retention times of small peptides by summing the relative hydrophobic contribution (Rc) of every constitutive amino-acid residues and terminal groups (terminal amino group and carboxylic function).

$$T_{retention} = \Sigma Rc \quad (1)$$

Since then, several researchers [3-5] have determined different values of Rc, for specific chromatographic systems. All of them were stated on experimental results : retention times of peptides of known composition were treated by a multi-regression computer program. Later, Guo [6, 7] employed synthetic octapeptides as models to determine Rc values which are commonly used.

However, experimental retention times of peptides larger than 9 or 10 amino-acid residues are lower than the ones calculated from equation 1 [8,9]. Mant et al. [10, 11] evidenced that the difference between experimental and calculated retention times varied with the length and hydrophobicity of the residues. Consequently these authors suggested to use a linear relationship in which the cross-product (Σ Rc ln N) appears :

$$\Sigma Rc - T_{retention} = A \Sigma Rc \, Ln \, N + B \quad (2)$$

A and B are constants, N is the total number of amino-acid residues.

Nevertheless, this modified expression of the predicted elution times of peptides is not sufficient to explain deviations observed for large peptides.

Recently Chabanet [13] developed a new prediction model based on Chothia's conclusions and equation 1. Chothia [12] pointed out that 50% of the surface of the charged residues in folded proteins remains accessible . Chabanet divided amino-acid residues into three groups : charged, polar and non polar and assigned a pair of constants (k, f) to each

group. The k values depend on amino-acids residues accessibility in proteins, the f and rc_i values are determined by multilinear regression. The contribution of each residue to peptide retention is a decreasing function ($R'c_i$) of the peptide length, thus it appears that the $R'c_i$ of these three distinct groups are differently affected by the peptide length according to :

$$T_{retention} = \Sigma\, n_i\, R'c_i + b' \qquad (3)$$

$$R'c_i(rc_i,N) = (rc_i - rc_i/k)\exp(-fN^2) + rc_i/k$$
$$\lim_{N\to 0} R'c_i = rc_i$$

To each amino-acid residue corresponds a constant rc_i, b' represents the retention contribution of the terminal amino group and carboxylic function, N is the total number of amino-acid residues.

Yet, these empirical models are not sufficient to predict and to explain the retention of peptides, of any length. In many examples peptides with the same residue composition have different retention times [14, 15] because of non similar residue sequences. Other factors could indeed influence the peptide retention. Conformational and nearest neighbour effects can play a important role. Zhou [16] showed that the existence of amphipatic α-helices could explain the deviations observed in comparison with the predicted retention times. Sereda et al [17] evidenced that the influence of a α amino group on the hydrophobicity of the N terminal residue depends of its nature.

In order to appreciate their validity for peaks identification peptide maps we compared three prediction methods : Meek's (with Guo's Rc), Mant's and Chabanet's ones.

Consequently, we studied here on one hand a serie of synthetic peptide containing 9 or 10 amino-acid residues and on the other hand an enzymatic digest of r-hu IFNγ (Scheme 1) containing peptide of 4 to 31 residues.

These two sets of experiments were carried out on two different columns, with different gradient shapes, so we first examined the influence of these factors on peptide retention.

SCHEME 1: Primary structure of r-hu. IFN γ (144 amino-acids)

```
0                                                  10
MET GLN ASP PRO TYR VAL LYS GLU ALA GLU ASN LEU LYS LYS TYR PHE ASN ALA GLY
      20                                           30
HIS SER ASP VAL ALA ASP ASN GLY THR LEU PHE LEU GLY ILE LEU LYS ASN TRP LYS GLU
      40                                           50
GLU SER ASP ARG LYS ILE MET GLN SER GLN ILE VAL SER PHE TYR PHE LYS LEU PHE LYS
      60                                           70
ASN PHE LYS ASP ASP GLN SER ILE GLN LYS SER VAL GLU THR ILE LYS GLU ASP MET ASN
      80                                           90
VAL LYS PHE PHE ASN SER ASN LYS LYS LYS ARG ASP ASP PHE GLU LYS LEU THR ASN
     100                                          110
TYR SER VAL THR ASP LEU ASN VAL GLN ARG LYS ALA ILE HIS GLU LEU ILE GLN VAL
     120                                          130
MET ALA GLU LEU SER PRO ALA ALA LYS THR GLY LYS ARG LYS ARG SER GLN MET LEU
                140
PHE ARG GLY ARG ARG ALA SER GLN
```

EXPERIMENTAL SECTION

Materials:

Synthetic peptides tested are reported in table I. Peptides 1 and 2 were obtained from Neosystem (Strasbourg, France), peptides 3 to 15 were from Interchim (Montluçon, France). Recombinant human interferon gamma (r-hu. IFNγ) is extracted from E. Coli, into which a plasmid coding for this protein has been transfected.

r-hu. IFNγ contains the sequence of 143 amino-acid residues as for the natural protein, plus an additional N-terminal methionine. This recombinant protein is not glycosylated.

r-hu. IFNγ (1mg/ml) was incubated in 50mM tris-HCl buffer (pH=7.0) during 15 hours at 37°C at a 2% ratio (weight:weight) with endoprotease from Staphylococcus Aureus strain V8 (Böhringer. Mannheim).

Chromatographic measurements

The analysis of the synthetic peptides and of the r-hu IFNγ hydrolysate was performed on the HPLC system 1 : two pumps (model 420, Kontron, Massachussetts, U.S.A.) connected to a Rheodyne sample injector (model 7125, Berkeley, CA, U.S.A.) equipped with a 20μl sample loop. A Spectra-Physic UV detector (model 100) was used at 220nm or 280nm. The column containing RP C18 (5μm spherical particles, 130 x 4.6mm I.D.) was obtained from Brownlee-Labs (Santa-Clara, C.A., U.S.A.).

TABLE I : Synthetic peptides composition and retention times on chromatographic system 1

n°	SYNTHETIC PEPTIDES	observed retention times($\%CH_3CN$)
1	ALA ALA ALA LYS LYS LYS ARG ALA ALA ALA	13.5
2	ALA LYS ALA ALA LYS ALA ARG ALA LYS ALA	14.3
3*	pGLU ALA LYS SER GLN GLY GLY SER ASN	9.6
4	TYR SER ARG VAL SER ARG ARG SER ARG	17.1
5	ARG ARG LYS ALA SER GLY PRO PRO VAL	20.2
6	TRP ALA GLY GLY ASP ALA SER GLY GLU	17.7
7	ALA PRO LEU LYS PRO ALA LYS SER ALA	17.5
8	TYR GLY ALA VAL GLY VAL GLY LYS SER	19
9	MET GLN MET LYS LYS VAL LEU ASP SER	24
10	GLU LEU ALA GLY ALA PRO PRO GLU PRO ALA	20.9
11	TYR GLY GLY PHE LEU ARG LYS TYR PRO	29.2
12	pGLU TRP PRO ARG PRO GLN ILE PRO PRO	29.6
13	TYR GLY GLY PHE LEU ARG ARG ILE ARG	27.6
14	SER PHE PRO TRP MET GLN SER ASP VAL THR	33.1
15	ALA PRO ARG LEU ARG PHE TYR SER LEU	33.4

* according to Snider's [19] results, we made the assumption that the Rc for pGLU is the same as that for GLN

Gradient linear elution was carried out with eluent A (0.1% T.F.A., H_2O) and eluent B (CH_3CN-H_2O (60 : 40) (v : v), 0.1% T.F.A.). The gradient slope was 1% CH_3CN per minute and the flow-rate 1ml/min.

The peptides of the r-hu. IFNγ hydrolysate were separated on the HPLC system 2 : a Waters (Millipore Corp., model 600) low pressure gradient HPLC pump system, a sample injector (model 7010/ 7012, Rheodyne, Berkeley, C.A., U.S.A) a Beckman ultrasphere C-18 column (750 x 4mm I.D., 3μm spherical particles), a Waters

990+ diode array detector system (the detector was equipped with a special high-pressure-resistant cell in order to enable the on line coupling with a thermospray mass spectrometer). 200μg of the hydrolysed protein were injected directly onto the HPLC column.

The peptides were eluted with a A-B gradient (A : H_2O; 0,1%T.F.A.) (B : CH_3CN; 0,1% T.F.A.) as follows :

0 - 40min : linear gradient 0% to 20% (v:v) B in A

40 -60min : linear gradient 20% to 40% (v:v) B in A

60 - 65min : linear gradient 40% to 50% (v:v) B in A

65 - 70min : linear gradient 50% to 0% (v:v) B in A

Flow-rate = 1.2ml/min

U.V. spectra of the eluate were recorded at given intervals (about 1s) at a wavelength range between 220nm and 290nm.

Retention times were expressed as the CH_3CN concentration in the solvent at the time of elution. This was calculated by subtracting the gradient elapsed time (as defined by Guo [6]) from the peak elution time and multiplying the result by the percentage of CH_3CN per minute in the linear gradient.

All chemicals were of analytical purity and special chromatographic grade.

RESULTS AND DISCUSSION

Comparison of peptide retention times predicted and observed on the two chromatographic systems.

The two sets of peptides were analysed with two different chromatographic systems which differed mainly by the chromatographic columns and the flow-rate chosen.

In order to compare the results obtained for the two sets of peptides, we studied the retention of 6 well identified IFNγ hydrolysate peptides (Table II) on both chromatographic systems. We observed that peptide retention times, obtained on each one, differed only by a shift (10% CH_3CN) (Fig. 1). The relationship between the retention times on the two systems was linear with a correlation coefficient of 0.997.

TABLE II : r-hu I.F.N. γ hydrolysate peptides and their retention times on chromatographic system 1 and 2

peptides	Observed retention(%CH_3CN) system 1	system 2
[103-112]	21	11.3
[0-7]	22	11.7
[94-102]	22	12.1
[76-93]	29	18.4
[76-102]	31	22.0
[25-39]	39	30.2

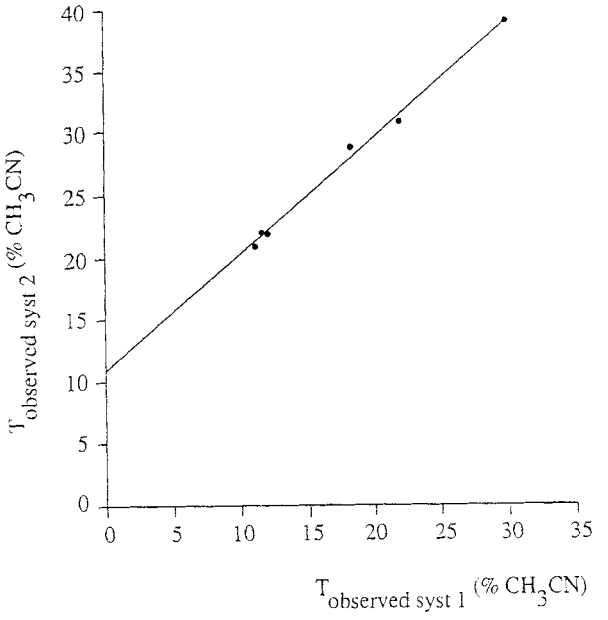

Figure 1 : Correlation between observed retention times obtained on system 1 and system 2 for 6 peptides (●) of the I.F.N. γ hydrolysate.
$T_{system\ 1} = 0.94\ T_{system\ 2} + 10.8$, the correlation coefficient is 0.997

This confirms the results obtained by Guo [6, 7] : he observed that peptides retention times obtained on various R.P. chromatographic columns exhibit a constant time difference and he showed that the influence of the flow-rate (in the range we used) and the support was negligible.

Consequently, we calculated the theoretical retention times of those 6 peptides according to Guo, Mant and Chabanet's rules. The correlation coefficients between these predicted retention times and the experimental ones (observed on each chromatographic system) are reported on table III. These linear functions were similar with a shift of 10% CH_3CN.

Analysis of sample 1 containing peptides of 9 or 10 amino acid residues.

Meek's model :

We compared the experimental peptide retention times (Table I) to the ones calculated by summing Guo's Rc according to the equation n° 1. and obtained with experimental conditions [6] similar to ours. The results in figure 2 show a linear correlation with a correlation coefficient of 0.94. However the slope is less than unity. The difference in length between Guo's octapeptides and the peptides studied here may already affect the retention prediction times.

Mant's model :

On figure 3, we also fitted the results according to equation n°2 as reported by Mant [10].

A linear correlation was observed and allowed us to obtain the A and B values of equation n°2 (A=0.15; B=-15.9). With these values, we calculated the theoretical retention times for comparison with the others models the correlation coefficient obtained between the predicted and the observed retention times was found equal to 0.94 (fig 4).

Chabanet's model :

Chabanet's prediction rules were applied to this sample of peptides. The comparison between the predicted and the observed retention times is presented on figure 5. The

TABLE III : Comparison of the results obtained on both systems for 6 peptides

Theoretical relationships of peptide retention times	Correlation coefficients (r) obtained on system 1	on system 2
$T = \Sigma Rc$ (1)	$Tobs(\%CH_3CN) = 0.50 \Sigma Rc + 17.6$ $r = 0.94$	$Tobs(\%CH3CN) = 0.55\Sigma Rc + 7.0$ $r = 0.96$
$\Sigma Rc - T = A\Sigma RcLnN + B$ (2)	$A = -16,39$ $B = 0.16$ $r = 0.97$	$A = -6.01$ $B = 0.15$ $r = 0.97$
$T = \Sigma n_i R'c_i(rc_i, N) + b'$ (3)	$Tobs(\%CH3CN) = 0.81\Sigma n_i R'c_i + 14.4$ $r = 0.99$	$Tobs(\%CH3CN) = 0.86\Sigma n_i R'c_i + 3.9$ $r = 0.98$

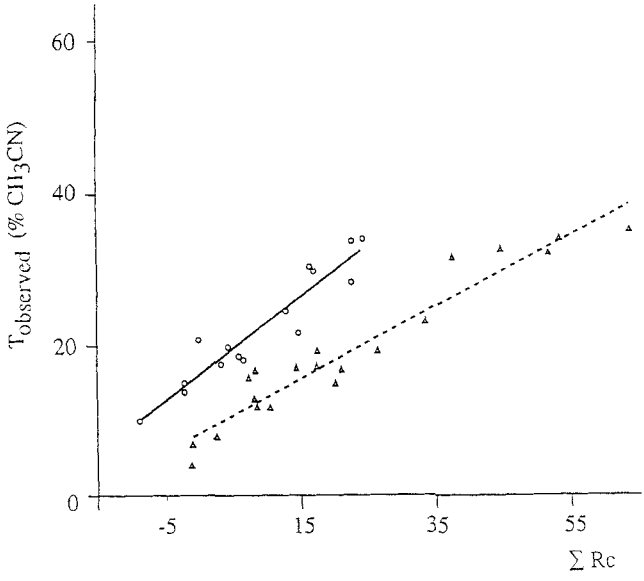

Figure 2 : Correlation between peptide retention times and the sum of residue contributions according to Guo's results (6).
-o- linear regression obtained with the first peptide sample obtained on chromatographic syst. 1
$T_{real} = 0.68\Sigma n_i Rc_i + 15.9$, the correlation coefficient is 0.94
- -Δ - linear regression obtained with the second peptide sample obtained on the chromatographic syst. 2
$T_{real} = 0.45\Sigma n_i Rc_i + 8.1$, the correlation coefficient is 0.96

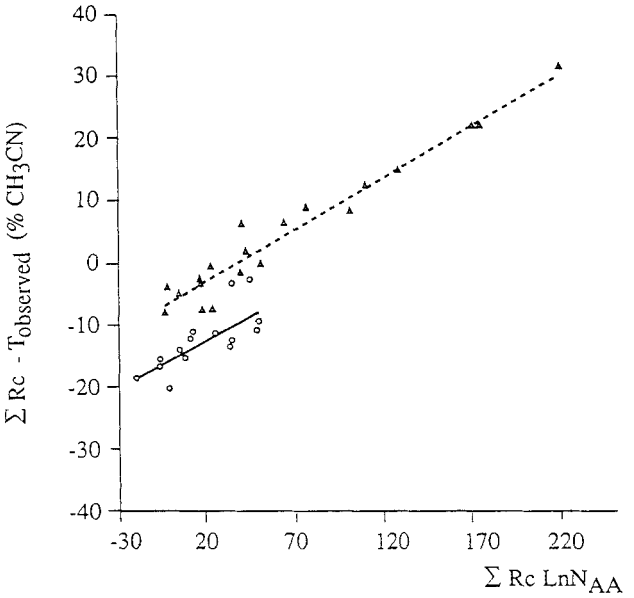

Figure 3 : Determination of A and B values of from the correlation between the observed retention time, length and hydrophoby of peptides according to equation n° 2.
-o- linear regression obtained with the first peptide sample obtained on the chromatographic syst. 1
A = 0.15, B = -15.9, the correlation coefficient is 0.80
- -Δ- linear regression obtained with the second peptide sample obtained on the chromatographic syst. 2
A = 0.16, B = -6.6, the correlation coefficient is 0.98

correlation factor between predicted and actual retention times was 0.95 but the slope remained different from unity (0.73). However if the linear regression is carried out without taking into account any of the three peptides with more than four positives charges, then the slope increases (0.85) and the correlation factor is improved (r = 0.97). It means that Chabanet's model may not be sufficient to provide accurate retention prediction for very basic peptides of small length.

Analysis of sample 2 : a r-hu IFNγ peptide map

We applied the same three prediction methods to calculate the retention of 20 peptides of a r-hu. IFNγ hydrolysate (table IV). The peptide map of this protein has been

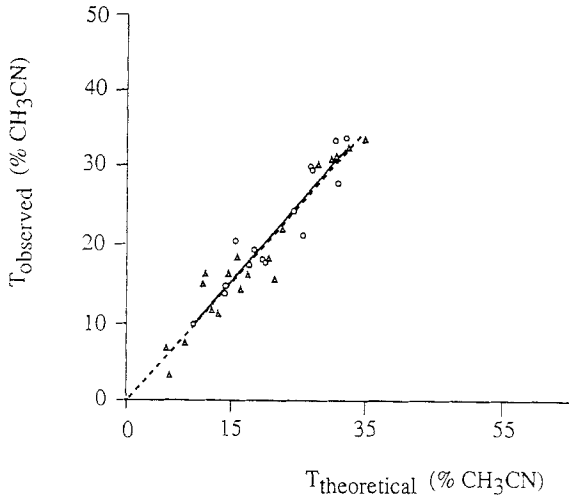

Figure 4 : Correlation between the calculated retention times according to the A and B values given on figure 3.
—o— linear regression obtained with the first peptide sample obtained on the chromatographic syst. 1
$T_{real} = T_{theoretical}$, the correlation coefficient is 0.94
- -Δ- linear regression obtained with the second peptide sample obtained on the chromatographic syst. 2
$T_{real} = 0.98T_{theoretical} + 0.39$, the correlation coefficient is 0.96

characterised by different techniques involving diode array absorbance detection, mass spectrometry and chemical sequencing [20].

In the chromatographic map of these peptides obtained with system 2, a few peaks could not be identified. For this study, we analysed only on the peaks, the nature of which was certain after identification with thermospray mass spectrometry.

Meek's model :

Peptide retention times were calculated by summing Guo's Rc. The relationship between these theoretical values and the experimental ones was linear as shown on figure n°2 with a correlation coefficient of 0.958. Nevertheless the slope was even lower than the one obtained for the previous peptides sample. It should be noted that most of the peptides of the hydrolysate contained a higher number of amino-acid residues than the synthetic

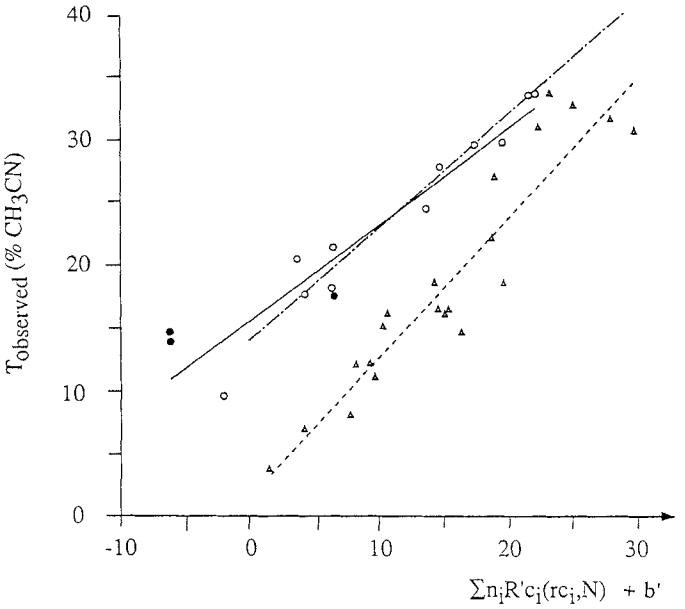

Figure 5 : Correlation between the observed and predicted retention times, calculated with equation n°3.
-o-●- linear regression obtained with the first peptide sample obtained on the chromatographic syst. 1
T_{real} = 0.73 $\Sigma n_i R'c_i(rc_i,N)$ + 15.5, the correlation coefficient is 0.95
-o· — linear regression obtained with the first peptide sample without taking into account the three peptides (●) with more than four positive charges
T_{real} = 0.85 $\Sigma n_i R'c_i(rc_i,N)$ + 14, the correlation coefficient is 0.97
- -Δ- linear regression obtained with the second peptide sample obtained on the chromatographic system 2
T_{real} = 1.1$\Sigma n_i R'c_i(rc_i,N)$ + 1.5, the correlation coefficient is 0.94

peptides we studied. As noticed by various authors [8-13], equation 1 overestimates the retention of large peptides. The formation of stabilised secondary and tertiary structure may perturb the interaction of certain residues with the hydrophobic stationary phase.

Mant's model :

For each peptide, we correlated the difference between the sum of the retention coefficients and the observed retention time to the cross-product ΣRcLnN (fig.3). We obtained a correlation coefficient of 0.98 for a linear relationship. The value of the slope A

TABLE IV : r-hu I.F.N. γ hydrolysate peptides and their retention times on chromatographic system 2

PEPTIDES	observed retention times (%CH3CN)
[0-7]	11.7
[94-102]	12.1
[120-143]	14.3
[10-21]	15.0
[72-75]	3.5
[120-133]	6.9
[40-46]	7.7
[103-112]	11.3
[113-119]	15.7
[92-102]	16.2
[76-91]	16.2
[10-24]	16.3
[94-112)	18.2
[76-93]	18.4
[76-102]	22.0
[25-39]	30.2
[10.39]	30.7
[22.39]	31.0
[47-71)	32.3
[40-71]	33.4

(A=0.16) was similar to the one we got for the other sample of peptides. Furthermore, the shift of peptide retention times observed between the two chromatographic systems corresponds to the difference between the B values obtained for each sample of peptides ($\Delta B = 10\%CH_3CN$). Therefore we conclude that Mant's model is valuable whatever the length or composition of peptides. The knowledge of A and B values allowed us to calculate the theoretical retention times of the peptides of sample 2. The correlation coefficient was found equal to 0.96 (Fig.4). This peptide sample contains more items than the one previously studied, what can explain why the correlation coefficient is higher.

Chabanet's model :

The figure 5 represents the correlation between the predicted and the observed retention times. The correlation coefficient is 0.94. Predicted and observed retention times differ only by a shift ($1.5\%CH_3CN$) due to the difference between our column and the one used by Chabanet.

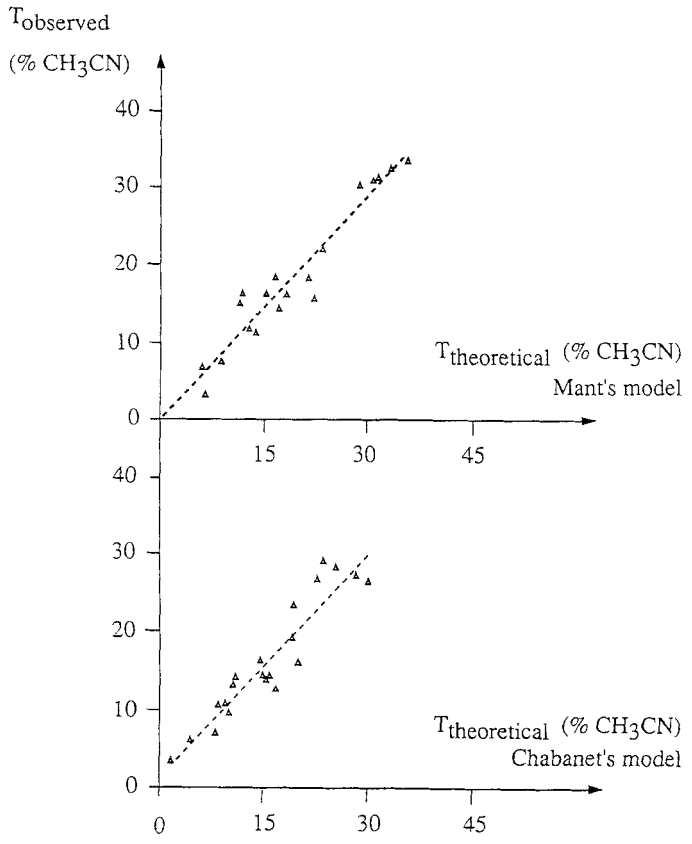

Figure 6 : Correlation between the observed and calculated retention times with the second peptide sample obtained on chromatographic system 2 with Mant's model, r=0.96 ; with Chabanet's model, r=0.94

Comparison between Mant's and Chabanet's models :

We compared on the figure 6 the predicted retention times obtained with Mant's and Chabanet's models. Both ones can be used indifferently.

CONCLUSION

For the comparison of three models of prediction of peptide retention we used two peptide samples:

a/ A sample of synthetic peptides of identical length but differing in composition (mainly basic and neutral amino acids).

b/ A hydrolysate of r-hu IFNγ : this hydrolysate was obtained using S. Aureus Strain V8 endoprotease. Thus it consists mainly of neutral and basic peptides

The application of Meek's model confirms the observation that the differences in peptide length neglected by Meek influence the retention of the peptide on a RP HPLC.

Mant's model fit's well the retention times of the peptides in RP HPLC. We have shown that it could be applied to a peptide map containing mainly neutral and basic peptides. We were able to compared separation obtained on different columns.

The validity of Chabanet's model is confirmed by the analysis of a peptide map containing essentially peptides of various length. However, we pointed out that for extremely very basic peptides (over 50 % basic amino acids) then this model is not valid. Further it should be noted that R. Cowan and R.G. Whittaker[18] have outlined the fact that different scales of hydrophobicity of amino-acid residues were stated but they differed widely particularly in the values of the polar ones.

REFERENCES

1. J.L. Meek, Proc. Natl. Acad. Sci.,U.S.A. 77 : 1632 (1980)

2. J.L. Meek, Z.L. Rossetti. J. Chromatogr. 211 : 15-28 (1981)

3. C.A. Browne, H.P. Bennett, S. Solomon, Anal. Biochem. 124 : 201-208 (1982)

4. S.J. Su, B. Grego, B. Niven , M.T.W. Hearn, J. Liq. Chromatogr. 4 : 1745-1764 (1981)

5. K.J. Wilson, A. Honegger, R.P. Stözel, G.J. Hughes, Biochem. J. 199 : 31-41 (1981)

6. D. Guo, C.T. Mant, A.K. Taneja, J. M. R. Parker, R. S. Hodges, J. Chromatogr. 359 : 499 (1986)

7. D. Guo, C.T. Mant, A.K. Taneja, R. S. Hodges, J. Chromatogr. 359 : 519-531 (1986)

8. J. Sasagawa, T. Okuyama, D.C. Teller, J. Chromatogr. 240 : 329-340 (1982)

9. Y. Sakamoto, N. Kawakami, T. Sasagawa, J. Chromatogr. 442 : 69-79 (1988)

10. C.T. Mant, T.W.L. Burke, A. Black, R.S. Hodges, J. Chromatogr. 458 : 193-205 (1988)

11. C.T. Mant, N.E. Zhou, R.S. Hodges, J. Chromatogr. 476 : 363-375 (1989)

12. C. Chotia, J. Mol. Biol. 105 : 1 (1976)

13. C. Chabanet, M. Yvon, J. Chromatogr. 599 : 211-225 (1992)

14. R.A. Houghten, S.T. Degraw, J. Chromatogr. 386 : 223-228 (1987)

15. K. Jinno, E. Tannigawa, Chromatographia 25 : 613-617 (1988)

16. N.E. Zhou, C.T. Mant, R.S. Hodges, Pept. Res. 19 : 8-20 (1990)

17. T.J. Sereda, C.T. Mant, A.M. Quinn, R.S. Hodges, J. Chromatogr. 646 : 17-30 (1993)

18. R. Cowan, R.G. Whittaker, Pept. Res. 3 : 75-80 (1990)

19. R.H. Snider, C.F. Moore, E.S. Nylén, K.L. Becker, Biochromatography 3(3) : 100-111 (1988)

20. R. Legrand, J.B. Falconnet, D. Prevost, B. Schoot, P. Devaux, J. Chromatogr 647 : 3-12 .(1993)

Received: May 9, 1994
Accepted: May 24, 1994