ELSEVIER

Journal of Chromatography A, 813 (1998) 255–265

JOURNAL OF CHROMATOGRAPHY A

High-performance size-exclusion chromatographic behaviour of substituted benzoylpoly-L-lysines by principal component analysis and molecular dynamics simulation

Ezio Bolzacchini, Viviana Consonni, Ruggero Lucini, Marco Orlandi, Bruno Rindone*

Dipartimento di Scienze dell'Ambiente e del Territorio, Universita' di Milano, Via Luigi Emanueli 15, I-20126 Milan, Italy

Received 8 January 1998; accepted 8 April 1998

Abstract

Conjugates containing benzoyl residues bound to poly-L-lysine (L-PLL) and L-lysine were prepared. The characterization was performed by UV spectrophotometry, ^{1}H NMR, size-exclusion chromatography (HPSEC) and circular dichroism. The principal component analysis method was used in order to evaluate the relationships among the retention time in HPSEC and several parameters of the substrates. The results of this approach suggest that the set of experimental variables give different significant information. A molecular dynamics calculation of a model of these conjugates shows folding of the polypeptide backbone upon derivatization with the 4-methylbenzoyl group. Multivariate analysis and molecular dynamics simulations are important tools for the interpretation of the physicochemical characteristics (including those involved in separation science) of hapten–polypeptide and hapten–protein conjugates used in clinical chemistry and in biotechnology.

Keywords: Principal component analysis; Molecular dynamics simulation; Chemometrics; Benzoylpolylysines; Peptides; Proteins; Haptens

1. Introduction

Hapten–protein conjugates are required both for 'in vitro' tests for the determination of human antibodies to the hapten [1,2] and for the preparation of antibodies to the hapten to be used in biosensors [3]. The chemical synthesis of hapten–protein conjugates requires the characterization of the conjugate, which is a family of molecules constituted from the protein and several hapten molecules located in different molecular regions [4]. Statistical methods are needed to describe this mixture.

The chromatographic behaviour of these materials is deeply influenced by their conformation, and this is important for any purification procedure. Moreover, the study of the conformation of the conjugate in water is a key step to understanding its action towards the immune system when the conjugate is used as an immunogen or as a reagent in clinical chemistry [4].

We report the preparation and the characterization of several conjugates derived from the reaction of benzoic acid derivatives with a synthetic polypep-

*Corresponding author.

0021-9673/98/$19.00
PII: S0021-9673(98)00305-7

tide, poly-L-lysine (L-PLL). The relationships among the experimental variables used to characterize the mixture were studied using principal-component analysis (PCA) [5–7]. A molecular dynamics (MD) experiment on a derivatized octalysine was also performed [8].

2. Experimental

2.1. Instrumentation

NMR spectra were measured using a Brucker AC300 instrument. HPLC analyses were performed with a Waters 600E HPLC. A Hewlett-Packard 1040 diode array detector was used. Circular dichroism (CD) were performed with a Jasco 500 instrument. IR spectra were recorded in KBr with a Jasco 5000 Fourier transform (FT) IR spectrometer.

2.2. Reagents

Poly-L-lysine (L-PLL) hydrobromide (n~145), *N*-(3-dimethylaminopropyl)-*N′*-ethylcarbodiimide (EC-DI) and L-lysine monohydrochloride were purchased from Sigma. Membranes for dialysis were Spectra/Por products. Benzoic acid derivatives were Fluka products.

2.3. Preparation of the conjugates

A volume of 0.09 mmol of substituted benzoic acid (2) was added to a solution of L-PLL (1) (0.0007 mmol) in 5 ml of 0.2 *M* aqueous buffer pH 8.5 (K_2HPO_4/KH_2PO_4) and the resulting solution was stirred at 4°C. After 5 min, 0.09 mmol of the carbodiimide (3) were added and this solution was stirred at 4°C for 12 h. Low-molecular-mass reagents and by-products were removed by dialysis for 120 h 4°C with an M_r 3500 cut-off membrane. The product (4) was then obtained by lyophilization.

2.4. HPLC analyses

High-performance size-exclusion chromatography (HPSEC) was performed by dissolving 2 mg of sample in 0.03 *M* phosphate buffer (K_2HPO_4/KH_2PO_4) pH 6.8–0.1 *M* potassium chloride and injecting it through a Rheodyne 100-μl loop. The instrument was equipped with a Synchropak GPC 100 (30 cm×7.8 mm I.D.) SEC column. The separation was achieved under isocratic conditions. The elution solvent was 0.03 *M* phosphate buffer (K_2HPO_4/KH_2PO_4) pH 6.8–potassium chloride 0.1 *M* at a flow-rate of 0.8 ml/min.

2.5. Preparation of ε-benzoyl-L-lysines (10)

Ten mmol of L-lysine monohydrochloride (6) were dissolved in 30–50 ml of water and boiled with a two-fold excess of $CuCO_3$ for 5 min. The dark blue solution was then filtered and cooled in ice. Benzoyl chlorides (8) (9 mmol) and 2 *M* NaOH were added with stirring and cooling in 10 portions in 30 min in such a way that the solution was kept slightly alkaline. The copper complex (9) separated as a blue solid which was filtered and washed with water, then suspended in 100 ml of water and a hydrogen sulphide stream was passed for 10 min. The water was then evaporated under reduced pressure and the residue was dissolved in 20 ml of ethanol–water which was filtered hot. The ε-benzoyl-L-lysines (10) crystallized on reducing to half volume and cooling.

3. Data

The variables (in brackets) used in this study were:

The retention time of the conjugates obtained by HPSEC (t_R).

The electronic contribution of the nuclear substituent [9–11] in the hapten (Hammett σ, SIG).

The lipophilicity parameter π (LOGP) of the hapten fragment expressed as the partition coefficient between *n*-octanol and water [12,13]. In this case, the calculated lipophilicity values were obtained by the package ChemPlus.

The hapten density obtained by ^{1}H NMR (HAPT).

The UV absorption maximum of the monoadducts (UV).

The IR absorption of the carbonyl in the monoadducts (IR).

The circular dichroism data for the conjugates at two pH values and the three wavelengths (208_1, 208_2, 220).

The ^{1}H NMR data of the protons of the lysine

residues both in the monoadducts (H1–H5) and the conjugates (H1A–H5A).

4. Results and discussion

Poly-L-lysine (L-PLL), (1), average M_r 20 000, was reacted with excess eighteen *meta*- and *para*-substituted benzoic acids (2) (haptens) (Fig. 1) in the presence of the water-soluble carbodiimide (3) to give, after extensive dialysis to eliminate excess (3) and the resulting urea (5), the conjugates (4) [4]. The conjugates were a family of poly-L-lysines derivatized at the primary amino group at C_ε with a substituted benzoyl group. Information about the structure of the conjugates in water were obtained using HPSEC. The analysis was performed eluting with aqueous phosphate buffer and monitoring the eluate at 220 nm (the absorption maximum of L-PLL) and at 260 nm (the absorption region of substituted benzoyl fragments). A single peak absorbing at 260 and at 220 nm was always observed. Table 1 reports the results thus obtained. The retention time for these conjugates was generally very similar to that of L-PLL. Exceptions were those containing the 4-methylbenzoyl-, the 3-methylbenzoyl- and the 4-fluorobenzoyl residue. These were eluted later, suggesting that these haptenized L-PLLs fitted the pores of the HPSEC material more efficiently than L-PLL. This occurs generally either when the molecular size is decreased (e.g. by a fragmentation reaction) or when folding of the polypeptide chain has occurred.

CD studies are often used for the study of the folding of proteins [14–16]. The CD spectra of the conjugates were then measured and Table 1 reports the values observed at pH 7 at 204 nm and at pH 11 at the two most significant wavelengths, 208 and 222 nm.

L-PLL is known to be an α sheet in anhydrous solid state, but even a few molecules of water give a mixture of α helix, β sheet and random conformation [17–23]. At pH 11, L-PLL is an α helix [24]. Most conjugates had CD spectra very different from that of L-PLL, suggesting that a conformational change had occurred. In particular, this fact was very evident in the comparison of the CD spectrum of 4-methylbenzoy-L-PLL, at pH 11 with that of L-PLL (Fig. 2). Here, the characteristic shape of the α-helix conformation of L-PLL at this pH was completely lost.

The conjugates are families of haptenized L-PLLs where the hapten is more or less randomly located along the L-PLL structure. Hence, a multivariate

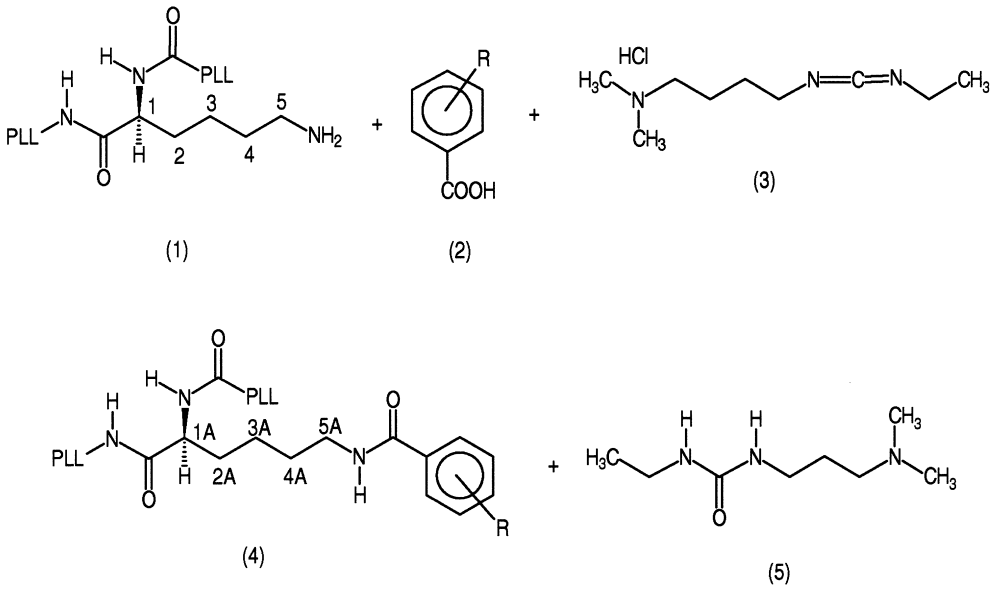

Fig. 1. Reaction of L-PLL with benzoic acids. See Section 4.

Table 1
Experimental data in the determination of the structure of ε-aroyl-L-PLL

R	HPSEC retention time (min)	CD pH 7 For aroyl-L-PLL conjugates	CD pH 11	CD pH 11	Hapten density	UV nm	IR cm^{-1}	^{1}H NMR For ε-aroyl-L-Lys conjugates				^{1}H NMR For ε-aroyl-L-PLL conjugates						Lipophilicity of R	Hammett sigma of R
		208 nm	208 nm	220 nm	%			ppm											
ID	t_R	208_1	208_2	220	HAPT	UV	IR	H1	H2	H3	H4	H5	H1A	H2A	H3A	H4A	H5A	LOGP	SIG
3-Br	11.64	−6.67	−3.84	−3.68	2.00	238.0	1638	4.15	1.81	1.51	1.81	3.03	4.83	1.81	1.51	1.81	3.35	2.54	0.391
Cl	11.24	−12.00	−4.19	−5.07	2.00	230.0	1638	4.35	1.73	1.46	1.73	2.99	4.95	1.73	1.46	1.74	3.10	2.26	0.259
3-CN	11.84	0	−3.31	−9.36	3.03	214.5	1657	4.30	1.76	1.50	1.76	3.00	4.80	1.71	1.47	1.71	3.45	1.78	0.43
3-F	11.13	6.80	−1.74	−2.16	5.06	218.7	1670	4.46	1.88	1.55	1.88	3.10	4.88	1.79	1.52	1.79	3.51	1.88	0.34
3-Me	13.30	−6.86	−4.07	−3.72	4.02	230.0	1640	4.28	1.70	1.45	1.71	2.99	4.76	1.71	1.45	1.71	3.28	2.21	−0.07
3-NO_2	11.84	−14.63	−7.44	−7.72	2.08	212.0	1640	4.39	1.85	1.55	1.81	3.12	4.95	1.85	1.55	1.85	3.12	1.70	0.71
3-OH	11.36	−2.3B	−1.23	−0.81	1.00	207.9	1680	4.32	1.80	1.60	1.80	3.07	4.81	1.64	1.43	1.64	3.48	1.46	0.12
3-OMe	11.56	−14.57	−6.83	−7.03	8.03	241.0	1636	4.25	1.69	1.39	1.69	2.98	4.85	1.69	1.39	1.69	3.40	1.49	0.062
4-Br	11.55	−6.15	−7.O8	−7.40	1.00	233.0	1636	4.39	1.73	1.42	1.73	2.95	4.80	1.73	1.42	1.73	2.85	2.54	0.232
4-Cl	11.37	−8.18	−6.48	−6.44	4.00	239.0	1636	4.15	1.75	1.46	1.75	2.97	4.65	1.75	1.46	1.75	2.95	2.26	0.227
4-CN	10.72	0	−8.79	−6.88	1.00	207.3	1631	4.30	1.79	1.48	1.79	3.30	4.77	1.61	1.42	1.61	3.37	1.78	0.66
4-F	12.44	0	−4.48	−2.17	0.05	2192	1630	4.20	1.69	1.40	1.69	2.90	4.75	1.63	1.41	1.63	3.44	1.88	−0.07
4-Me	13.31	−29.84	−1.51	−0.86	8.00	235.0	1636	4.28	1.72	1.41	1.72	3.00	4.78	1.71	1.44	1.71	3.35	2.21	−0.2
4-NO_2	11.72	−17.31	−4.11	−4.19	6.08	210.0	1640	4.45	1.81	1.50	1.81	3.10	4.90	1.81	1.50	1.81	3.15	1.70	0778
4-OH	11.28	25.57	−5.63	−4.61	4.04	208.7	1680	4.31	1.78	1.45	1.78	3.20	4.82	1.75	1.50	1.75	3.47	1.46	−0.37
4-*O*-Me	11.77	−12.12	−6.00	−5.96	12.00	247.0	1636	4.25	1.70	1.41	1.70	2.98	4.80	1.70	1.41	1.70	3.40	1.49	−0.27
4-*t*Bu	11.72	0	−7.20	−6.19	0	234.3	1686	4.30	1.75	1.50	1.75	3.00	4.81	1.62	1.44	1.62	3.48	3.37	−0.16
H	11.43	−16.86	−6.87	−6.96	4.00	223.0	1638	4.29	1.70	1.39	1.70	2.98	4.75	1.70	1.39	1.70	3.09	1.75	0

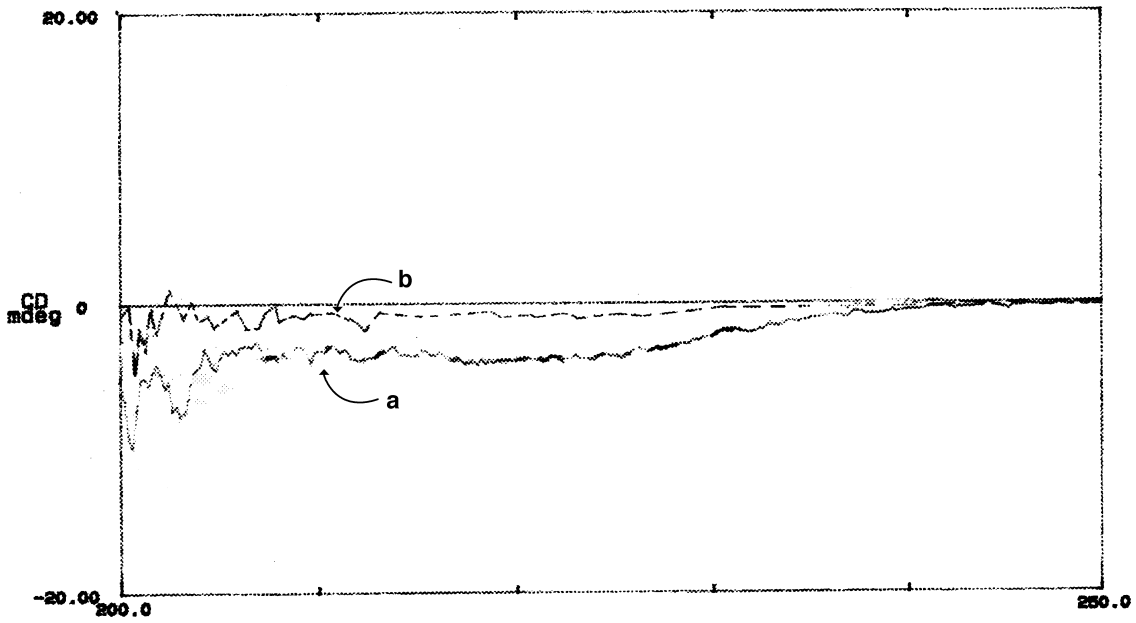

Fig. 2. CD spectra of L-PLL (a) and 4-methylbenzoy-L-PLL (b), at pH 11.

approach for the determination of similarity/dissimilarity of these was needed. Further data for this approach were then determined.

The number of molecules of hapten which were attached to L-PLL was measured using the ^{1}H NMR spectroscopy in $^{2}H_{2}O$, and the UV absorption of the conjugates at the absorption maximum of substituted ε-benzoyl-L-lysines (9) was also determined. Table 1 shows the data thus obtained. The integration of the aromatic protons and H_{α} allowed the hapten density % to be given.

For the measure of the UV absorption, ε-aroyl-L-lysines (10) were prepared by reacting L-lysine–copper(II) complex (7) with substituted benzoyl chlorides (8) (Fig. 3) [25]. Deprotection of the α-aminoacidic structure in compound (9) was then performed by treatment with hydrogen sulfide [26–28]. Both absorption maxima and molar extinction coefficients of these substituted ε-benzoyl-L-lysines were different from those of the hapten or the corresponding substituted ε-benzoyl poly-L-lysine. Hence, hapten density could not be calculated in this way.

Table 1 shows also the frequency of the IR absorption band of the amide carbonyl group in ε-aroyl-L-lysines and the ^{1}H NMR values for the hydrogens of the lysine residues both in ε-aroyl-L-lysines and in ε-aroyL-PLLs.

PCA was used to evaluate the information contained in the experimental data and the relationships among the variables. PCA was performed on a data matrix of 18 conjugates described by 19 variables and the 6 most significant components explained 85.7% of the total data variance. Table 2 shows the loading matrix of the six most significant principal components.

The meaning of the principal components and the relationships among the variables were evaluated from the loading plots.

It could be shown that:

1. The first component (34.7% of explained variance) mainly represents the difference between NMR and UV signals. This suggests that a bathochromic shift in the UV maxima is correlated with a downfield shift of the NMR signals of the lysine residues probably owing to the shielding properties of the π electrons generating the UV maximum.
2. The second component (16.9% of explained variance) represents the difference between IR

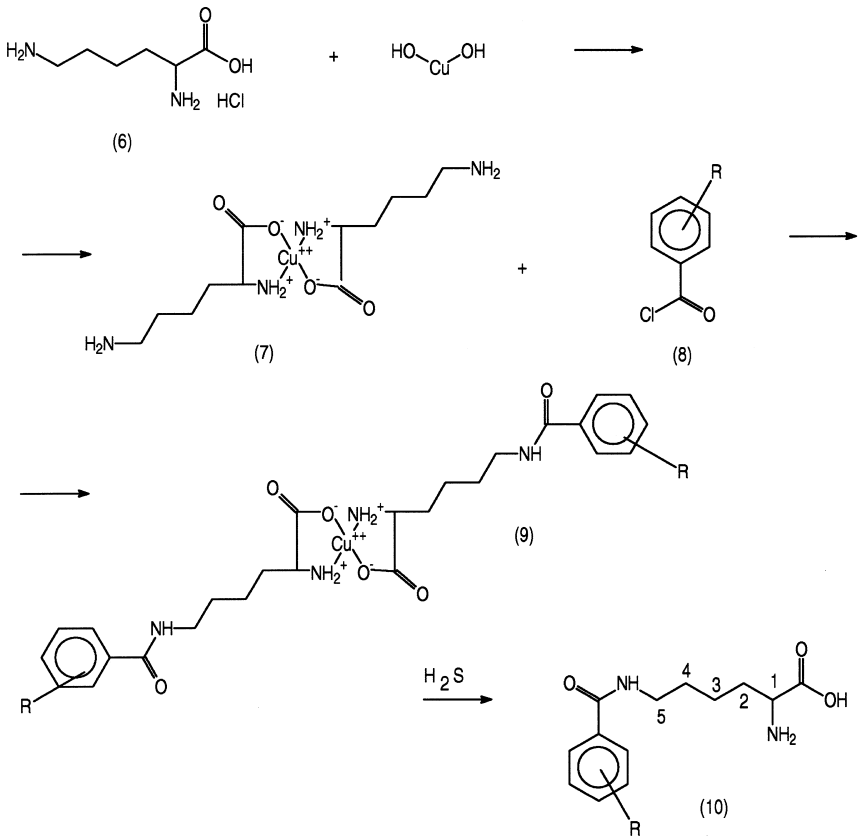

Fig. 3. Reaction of L-lysine–copper(II) complex with substituted benzoyl chlorides. See Section 4.

Table 2
Loading matrix of the six most significant principal components

	PC 1	PC 2	PC 3	PC 4	PC 5	PC 6
t_R	−0.418	0.193	0.623	−0.157	−0.229	−0.056
H1	0.614	0.069	0.006	0.229	−0.299	0.586
H2	0.961	−0.051	0.030	−0.072	0.023	−0.122
H3	0.890	−0.071	0.021	−0.297	−0.021	−0.038
H4	0.943	−0.122	0.068	−0.087	0.028	−0.141
H5	0.655	−0.306	−0.286	0.351	0.027	−0.179
UV	−0.628	0.403	0.175	−0.285	0.390	0.120
IR	0.392	−0.681	0.182	−0.201	0.271	0.350
H1A	0.608	0.246	0.118	0.168	−0.110	0.558
H2A	0.596	0.688	0.223	−0.022	0.274	−0.068
H3A	0.857	0.227	0.223	−0.209	0.198	−0.107
H4A	0.596	0.688	0.223	−0.022	0.274	−0.068
H5A	0.089	−0.722	0.397	0.188	0.153	0.006
HAPT	−0.241	0.346	0.414	0.596	0.390	0.078
208_1	0.332	−0.683	−0.179	−0.083	0.418	−0.048
208_2	0.189	−0.157	0.853	−0.027	−0.236	−0.070
220	0.044	−0.302	0.765	−0.033	−0.280	−0.189
SIG	0.645	0.380	−0.369	0.016	−0.396	−0.194
LOGP	−0.149	0.051	−0.058	−0.907	−0.025	0.203

signals, CD measures at pH 7 (208_1) and H5A with respect to H2A and H4A NMR signals. This component seems to be correlated with a conformational change.

3. The third component (13.2%) is mainly related to CD measures at pH 11 (208_2 and 220). This component reflects a conformational change that has occurred upon haptenization.
4. The fourth component (9.0%) is mainly related to the lipophilicity. This parameter suggests that the conformational change depends also from the lipophilicity of the hapten fragment.
5. The fifth component (6.4%) represents the difference between the Hammett σ electronic representation and the electronic information related to UV, HAPT and CD (208_1) measures. This component reflects the influence of the electronic density in the hapten on the haptenization reaction (a nucleophilic displacement).
6. The sixth component (5.3%) gives information about the peculiar role of H1 and H1A NMR signals. They are probably descriptors of the haptenization reaction.

From the loadings, as shown in Table 2, it can be observed that all the experimental variables give useful information in the most significant components, showing only local and low correlation among themselves. Thus, PCA allows description of the conjugates.

The projection of the 18 samples onto the PC space (scores plots) allows the analysis of the individual conjugates.

In Figs. 4–6, the 18 samples are shown in the space of PC1–PC2, PC3–PC4 and PC5–PC6, respectively. For example, the group of compounds on the left of PC1 is characterized by high UV and low NMR signals, in contrast with the compounds 3-F and 3-NO_2. Onto the second component, the four compounds 4-*tert*-Bu, 4-CN, 3-OH and 4-OH are mainly characterized by high values of H5A, IR and CD 208_1 and low values of H2A and H4A; their behaviour is opposite to that of the compounds 3-NO_2 and 4-NO_2 (Fig. 4). On the third component, the opposite behaviour of the compounds 4-Me and 4-CN is apparent, due to their different values of CD at pH 11 (208_2 and 220). The fourth PC highlights

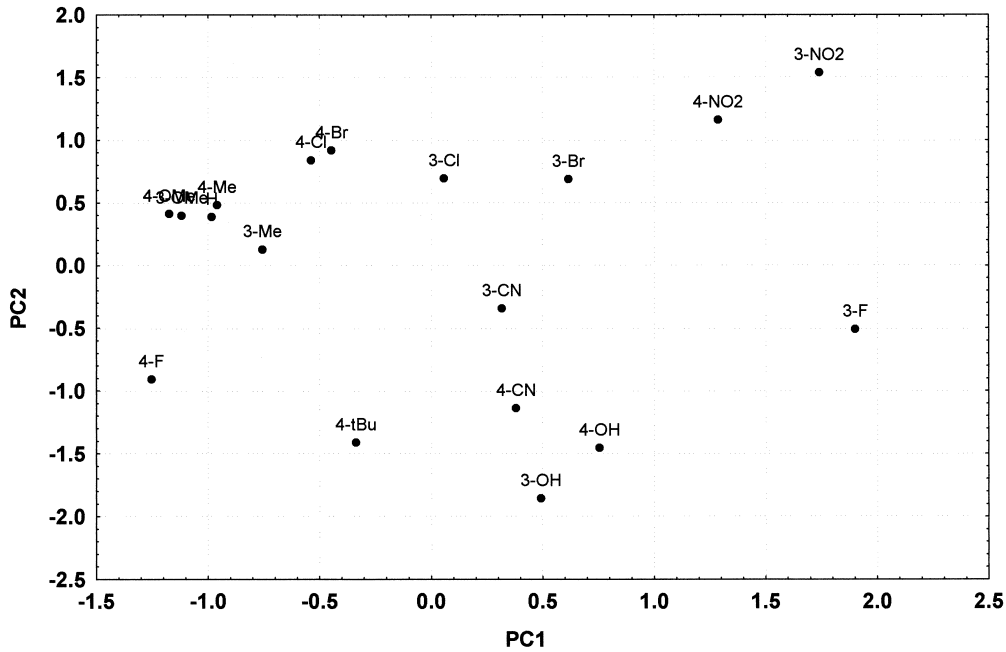

Fig. 4. Scores plot, PC2 vs. PC1. *t*Bu=*tert*-Butyl.

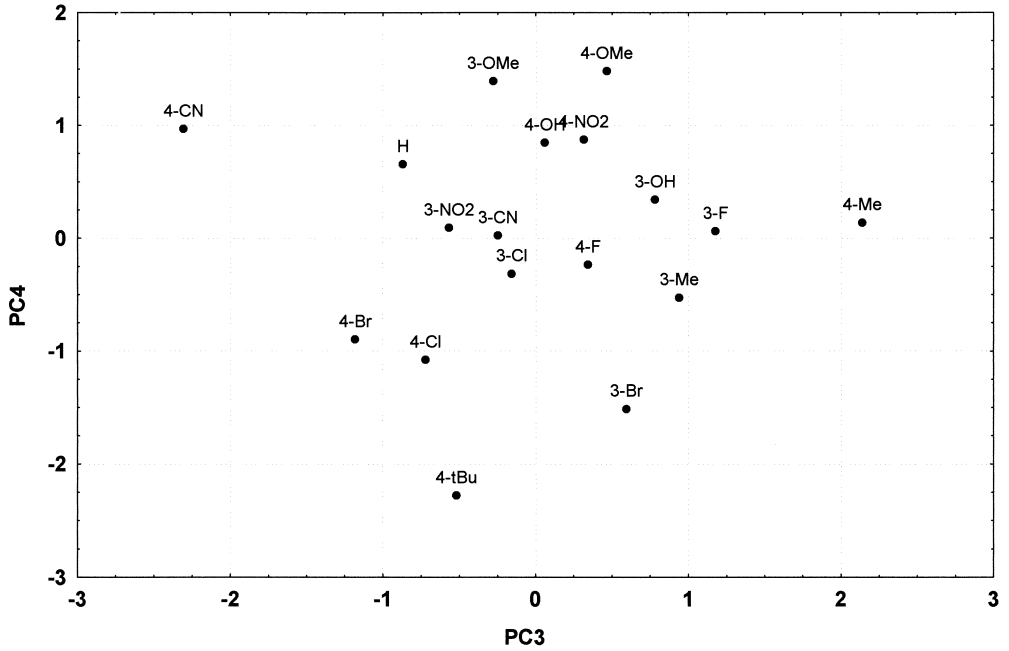

Fig. 5. Scores plot, PC4 vs. PC3.

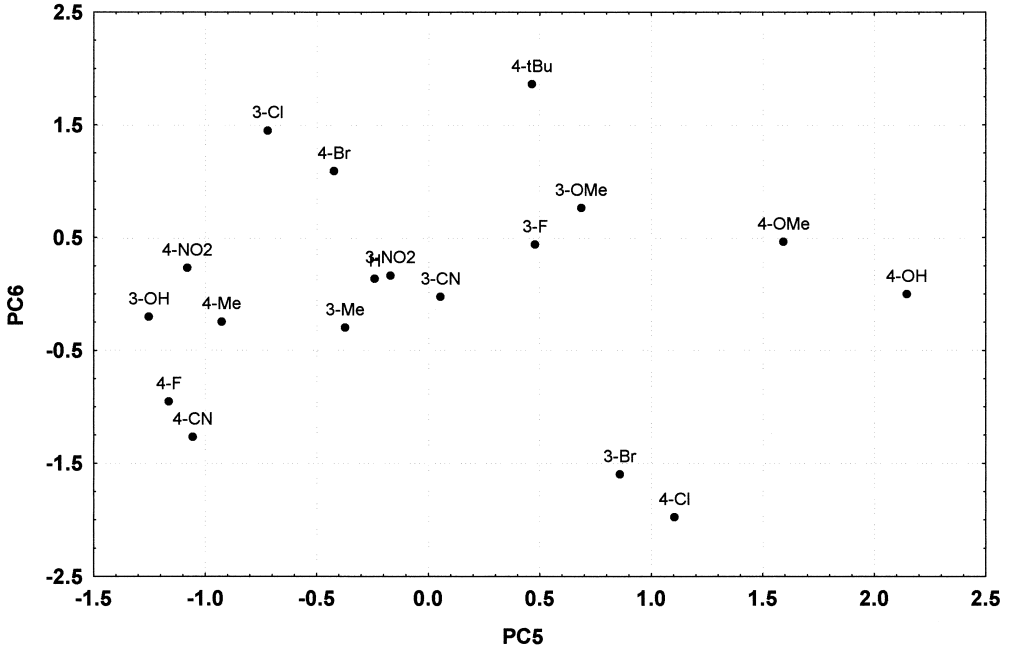

Fig. 6. Scores plot, PC6 vs. PC5.

the peculiar characteristics of the compounds and 4-*tert*-Bu and 3-Br, due to their high values of lipophilicity (Fig. 5).

Fig. 6 shows the scores plot PC6 vs. PC5.

Concerning the fact that three out of eighteen conjugates exhibited a higher retention time, a correlation is found between their retention times (t_R) and the values of CD at 220 and 208-2 (those measured at pH 11 and related to a α-helix conformation). This is not surprising since both a higher t_R and a higher CD at those wavelengths derive from the folding of the polypeptide upon haptenization.

This fact was further studied by performing a MD simulation with two models. A octalysine (11) was used for simulating L-PLL and a octalysine haptenized with two 4-methylbenzoyl groups at the ε-amino group of lysines 1 and 8 (12) was used for simulating 4-methylbenzoyl-L-PLL (Fig. 7). These molecules were put in a water box containing 500 water molecules randomly disposed and minimized. Fig. 8 shows the shape of the model of L-PLL (11). When two haptens were bound to this model to form (12), the MD experiment showed folding of the octalysine skeleton (Fig. 9).

A further parameter which was shown by the PCA analysis to be important at least in the folding of 4- and 3-methylbenzolyl-L-PLL is the high hapten density. This is not surprising since those liphophilic haptens, if present in large number in the conjugate, tend to be located in a lipophilic environment, as shown by the MD experiment, thus resulting in folding of the polypeptide.

In conclusion, PCA and MD simulations are important tools to give an interpretation to the physicochemical characteristics (including those involved in separation science) of hapten–polypeptide and hapten–protein conjugates used in clinical chemistry and in biotechnology.

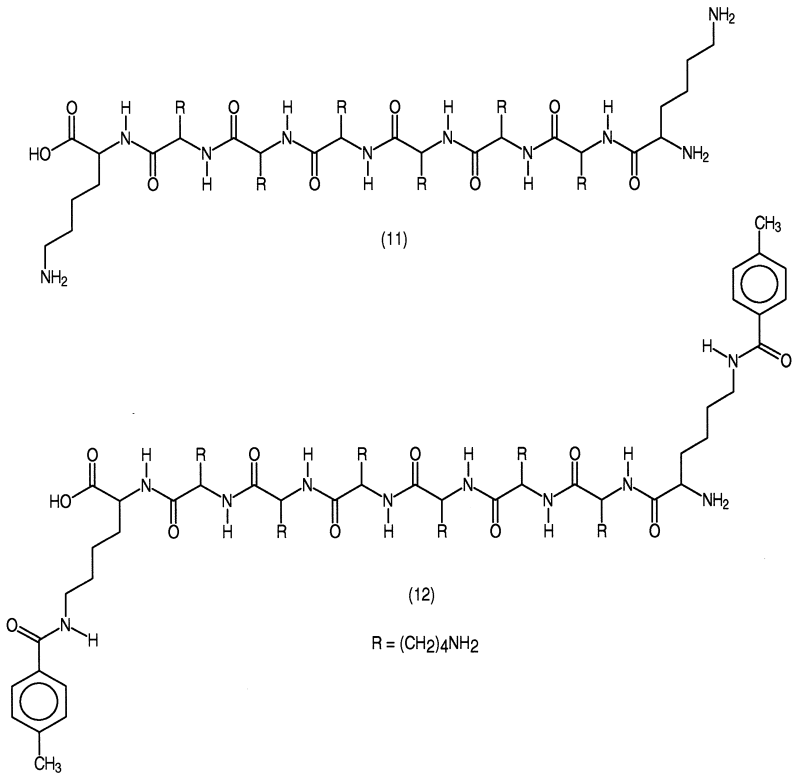

Fig. 7. Octalysine (11) and octalysine haptenized with two 4-methylbenzoyl groups (12) used for simulating L-PLL and 4-methylbenzoyl-L-PLL, respectively.

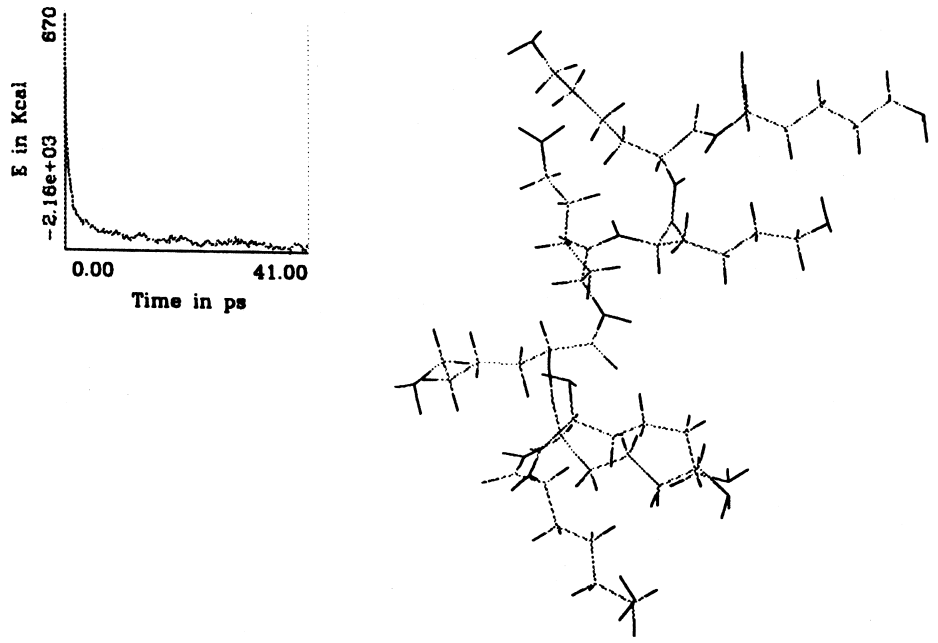

Fig. 8. Minimum energy conformation of the model of L-PLL. 1 cal=4.184 J.

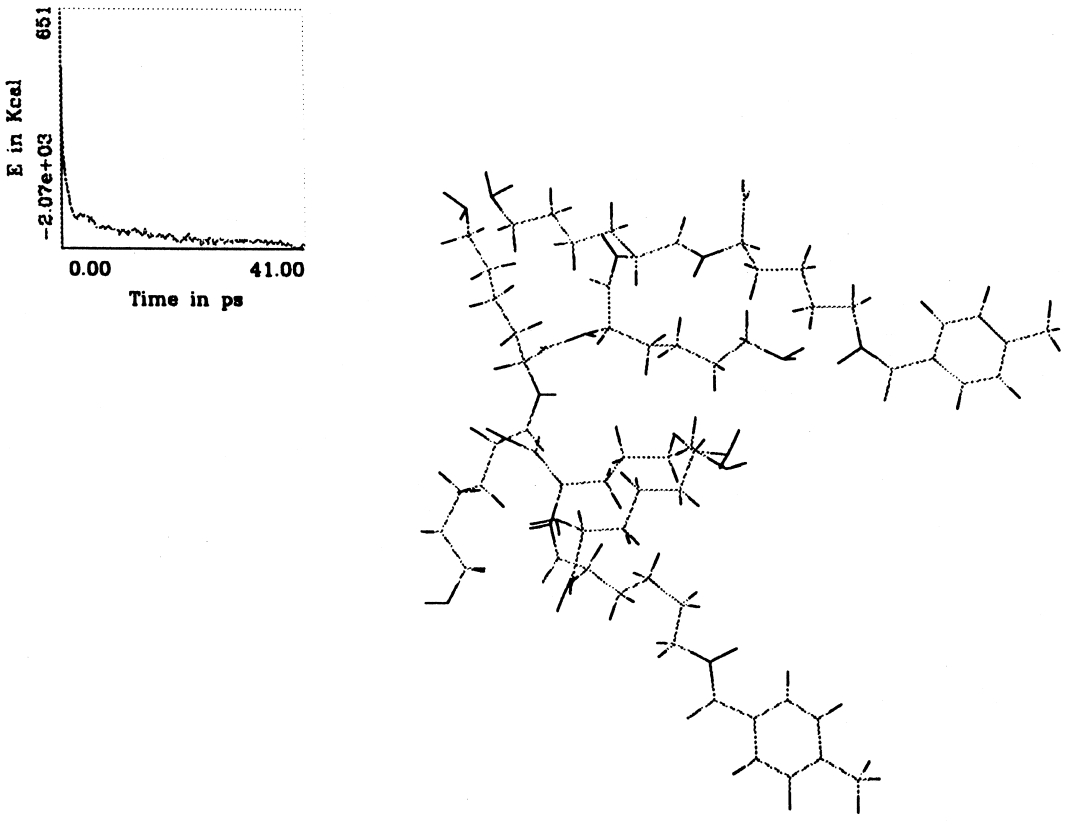

Fig. 9. Minimum energy conformation of the model of L-PLL haptenized with two 4-methylbenzoyl residues.

Acknowledgements

This work was supported by the Consiglio Nazionale delle Ricerche (Piano Finalizzato Chimica Fine) and by the Ministero dell'Universita' e della Ricerca Scientifica e Tecnologica (M.U.R.S.T). We thank Professor Roberto Todeschini for his suggestions in the multivariate analysis and our students Elisabetta Damaggio and Anna Brocca for their collaboration in the experimental work.

References

[1] C. Milstein, Angew. Chem. Int. Ed. Engl. 24 (1985) 816.
[2] B. Frangione, C. Milstein, J.R.L. Pink, Nature 221 (1969) 145.
[3] C.D. Watts, B. Hegarty, Pure Appl. Chem. 67 (1995) 1533.
[4] C. Caneva, P. Di Gennaro, F. Farina, M. Orlandi, B. Rindone, P. Falagiani, Bioconj. Chem. 4 (1993) 309.
[5] A. Lorber, L. Wangen, E. Kowalski, J. Chemometrics 1 (1987) 19.
[6] B. Efron, Society for Industrial and Applied Mathematics, Philadelphia, PA, 1982.
[7] R. Todeschini, I.E. Frank, G. Moro, U. Cosentino, Jerryll, Stanford, 1991, Rel 1.3.
[8] E. Clementi, Modern Techniques in Computational Chemistry, Escom, 1991.
[9] L.P. Hammett, Physical Organic Chemistry, McGraw-Hill, 1940.
[10] J. March, Advanced Organic Chemistry, Wiley, 1985, p. 244.
[11] P.R. Wells, Linear Free Energy Relationship, Academic Press, 1968.
[12] M.T. Bernabei, F. Forni, S. Bellei, R. Cameroni, Atti Soc. Nat. Mat. di Modena 111 (1980) 63.
[13] C. Hansch, A. Leo, S.H. Unger, K.H. Kim, D. Nikaitani, E.J. Lien, J. Med. Chem. 16 (1973) 1207.
[14] W.C. Johnson, Methods Biochem. Anal. 31 (1985) 62.
[15] E.H. Strickland, Crit. Rev. Biochem. 113 (1974) 175.
[16] F.D. Sonnichsen, J.E. Van Eyk, R.S. Hodges, B.D. Sykes 31 (1992) 8790.
[17] J.-H. Furhop, M. Krull, G. Buldt, Angew. Chem., Int. Ed. Engl. 26 (1987) 699.
[18] J. Lippert, J. Tyminski, P.J. Desmeules, J. Am. Chem. Soc. 98 (1976) 7075.
[19] N.I.T. Okubo, K. Yamamoto, H. Matsuoka, T. Hashimoto, M. Fujimura, J. Chem. Phys. 78 (1983) 541.
[20] H. Wakana, T. Shigaki, N. Saito, Biophys. Chem. 16 (1982) 275.
[21] A. Darke, G. Finer, Biopolymers 14 (1975) 441.
[22] U. Shmueli, W. Traub, J. Mol. Biol. 12 (1965) 205.
[23] F.J. Padden, H.D. Keith, G. Giannoni, Biopolymers 7 (1969) 793.
[24] U. Greenfield, G.D. Fasman, Biochemistry 8 (1969) 4108.
[25] M.T.L.S. Duarte, C.T. Carrondo, M.L.S. Simoes Goncalves, M.B. Hursthouse, N.P.C. Walker, H.M. Dawes, In. Chim. Acta 108 (1985) 11.
[26] A.C. Kurtz, J. Biol. Chem. 122 (1938) 477.
[27] A. Neuberger, F. Sanger, Biochem. J. 37 (1943) 515.
[28] A. Taurins, J. Res. 28 (1950) 762.