ELSEVIER

Journal of Molecular Catalysis B: Enzymatic 15 (2001) 177–189

JOURNAL OF MOLECULAR CATALYSIS B: ENZYMATIC

www.elsevier.com/locate/molcatb

Isolation and purification of cysteine peptidases from the latex of *Araujia hortorum* fruits Study of their esterase activities using partial least-squares (PLS) modeling

Nora Priolo [a], Cecilia M. Arribére [a], Néstor Caffini [a], Sonia Barberis [b,*], Rodolfo Nieto Vázquez [c], Juan M. Luco [d]

[a] *Laboratorio de Investigación de Proteínas Vegetales (LIPROVE), Departamento de Ciencias Biológicas, Facultad de Ciencias Exactas, Universidad Nacional de la Plata, CC 711, 1900 La Plata, Argentina*
[b] *Laboratorio de Bromatología, Departamento de Farmacia, Facultad de Química, Bioquímica y Farmacia, Universidad Nacional de San Luis, Chacabuco y Pedernera, 5700 San Luis, Argentina*
[c] *Área de Química Física, Facultad de Química, Bioquímica y Farmacia, Universidad Nacional de San Luis, Chacabuco y Pedernera, 5700 San Luis, Argentina*
[d] *Laboratorio de Alimentos, Facultad de Química, Bioquímica y Farmacia, Universidad Nacional de San Luis, Chacabuco y Pedernera, 5700 San Luis, Argentina*

Received 27 October 2000; received in revised form 23 March 2001; accepted 23 March 2001

Abstract

Three new cysteine peptidases (*araujiain h*-I, *araujiain h*-II and *araujiain h*-III) were isolated and purified to mass spectroscopy homogeneity from the latex of *Araujia hortorum* (*Asclepiadaceae*) fruits by ultracentrifugation and ion exchange chromatography. The enzymes have molecular masses of 24,030.87, 23,718 and 23,545.5 (mass spectrometry), respectively. The peptidases were activated by thiol compounds and inhibited by common thiol blocking reagents, particularly E-64 and $HgCl_2$, suggesting that the enzymes belong to the cysteine peptidases family. A quantitative structure–activity relationship study of their esterolytic activities was performed by means of partial least-squares regression and using the novel filtering technique called orthogonal signal correction. The numerical characterization of the variation in the physicochemical features of the *N*-α-carbobenzoxy-aminoacid-*p*-nitrophenyl esters used in the PLS regression modeling was accomplished by using a large number of descriptors extracted from the literature, based on various lipophilicity, polarity and steric scales of the amino acid side-chains in combination with a set of property descriptors derived from semiempirical calculations. From the obtained results it can be concluded that all hydrophobic, electronic, and steric factors are important in the esterase activity of the cysteine peptidases studied.

Keywords: *Araujia hortorum*; *Asclepiadaceae*; Plant peptidases; OSC–PLS; QSAR

1. Introduction

Proteolytic enzymes have been used in medicine and in industry for hundreds of years [1]. In recent years, they have become of vital importance due to

* Corresponding author. Fax: +54-2652-431301.
E-mail address: sbarberi@unsl.edu.ar (S. Barberis).

1381-1177/01/$ – see front matter
PII: S1381-1177(01)00022-4

the availability of standard, potent and active preparations with good solubility, stability and odor. These enzymes are used in industry to produce improved or novel products, to bypass long and involved chemical synthetic pathways and to separate and purify isomeric mixtures [2,3]. Furthermore, proteolytic enzymes of plant origin have received special attention in the field of medicine and industry due to their unique property of being active at a very wide ranges of temperature and pH. Thus, peptidases, which firmly maintain the first place in the world market of enzymes, play an important role in biotechnology, since proteolysis changes the chemical, physical, biological, and immunological properties of proteins. Besides, the enzymatic hydrolysis is strongly preferred over chemical methods, because it yields hydrolysates containing well-defined peptide mixtures and avoids the destruction of L-amino acids and the formation of toxic substances like lysino-alanine [4,5].

The growing importance of proteolytic enzymes, particularly peptidases, creates a need for discovering new plant sources of potent, more active and more specific proteolytic enzymes. Peptidases are frequently present in the latex of *Asclepiadaceae* (milkweed family). Nevertheless, studies about these peptidases are limited to a few species belonging to the genera *Asclepias* [6–12] and *Calotropis* [13–16]. In a recent paper [17], we reported the partial characterization of peptidases isolated from the latex of five species of *Asclepiadaceae* grown in Argentina, as well as the purification and characterization of the peptidases of one of them, *Morrenia brachystephana* Griseb. A further communication reported the presence of peptidases in the latex of *Morrenia odorata* (Hook. et Arn.) Lindley [18].

Five catalytic types of peptidases can now be recognized, in all of which serine, threonine, cysteine, aspartic or metallo groups play primary roles in enzyme catalysis. The serine, threonine and cysteine peptidases are catalytically very different from aspartic and metallopeptidases ones. In the three first types of peptidases the nucleophile of the catalytic site is part of an amino acid, whereas in the other two groups it is an activated water molecule. Cysteine peptidases are a class of proteolytic enzymes that take advantage of the nucleophilicity of a thiolate to achieve amide hydrolysis. This mechanism is similar to that of serine-type peptidases in that the proton donor is a His residue. Although there is evidence that in some cysteine peptidases a third residue is required to orientate the imidazolium ring of the His, there are a number of families in which only a catalytic dyad is necessary. To date, 41 families of cysteine peptidases are recognized. A family is a group in which every member shows a statistically significant amino acid sequence relationship with at least one other member of the family in the part of the molecule that is responsible for peptidase activity. Most plant cysteine peptidases belong to the papain family, including those of *Asclepiadaceae*, the milkweed family [19].

An often used approach to interpret the enzyme-catalyzed reactions is a qualitative and empirical one. As an alternative to this, the use of approaches employing computational chemistry have been useful in establishing quantitative structure–activity relationships (QSAR) applied to the study of biocatalytic activities [20]. Although simple or multiple linear regression (MLR) approaches are the most often adopted chemometric methods in QSAR studies, fortuitous or artifactual results may be obtained by the use of such techniques when there exists multicollinearity among predictor variables [21]. If this is the case, it should be possible to use multivariate projection methods such as partial least-squares (PLS) regression [22], which is insensitive to the collinearity among the variables and also offers the advantage of handling data sets where the number of independent variables is greater than the number of observations. In previous studies [23–28], multivariate statistical methods such as PLS have been successfully applied in establishing quantitative structure–activity and structure–property relationships (QSAR, QSPR).

In the present study, the isolation and purification of three cysteine peptidases obtained from the latex of *Araujia hortorum* fruits is informed. This species is a South American climbing plant that grows in the south of Brazil, Paraguay, Uruguay and Argentina [29]. In addition, the PLS technique is used for modeling the endoesterolytic activity determined with *N*-α-carbobenzoxy-*p*-nitrophenyl esters of several amino acids as substrates for the three isolated peptidases. Thus, the main purpose of this work was to understand in which way the different characteristics of substrates (molecular weight, bulkiness, hydrophobicity and several electronic structural features) affect the esterase activities of the cysteine peptidases under study.

2. Materials and methods

2.1. Plant material

Fruits of *A. hortorum* Fourn. were obtained from plants grown in Ringuelet, province of Buenos Aires, Argentina (Argentinian folk names: "tasi", "doca"). The plant is a vine, with egg-shaped, oblong leaves, 5–10 cm long, whitish and pubescent on its abaxial face; fruits are ovoid, green, smooth and 8–12 cm long [30]. Voucher specimens were deposited at the LPE herbarium (Facultad de Ciencias Exactas, Universidad Nacional de La Plata, Argentina).

2.2. Preparation of the crude extract

Latex obtained by superficial incisions of fruits, received on 0.1 M citrate–phosphate buffer (pH 6.5) containing 5 mM EDTA and cysteine, was first centrifuged at 16,000 × *g* for 30 min at 4°C. Gums and other insoluble materials were discarded, and the supernatant was ultracentrifuged at 100,000 × *g* for 60 min at 4°C. This new supernatant (crude extract), containing soluble proteins, was fractionated and conserved at −20°C for further studies.

2.3. Proteolytic (caseinolytic) activity assays

Proteolytic assays were made using casein (Hammarsten type; Research Organics, Cleveland, OH) as substrate. The reaction mixture was prepared by mixing 0.1 ml of enzyme extract with 1.1 ml of 1% casein containing 12 mM cysteine, in a 0.1 M Tris–HCl buffer (pH 8.0). The reaction was carried out at 45°C and stopped 2 min later by the addition of 1.8 ml of 5% trichloroacetic acid (TCA). Each test tube was centrifuged at 3000 × *g* for 30 min and the absorbance of the supernatant measured at 280 nm. An arbitrary enzyme unit (caseinolytic unit, U_{cas}) was defined as the amount of peptidase which produces an increment of one absorbance unit per min in the assay conditions [31].

2.4. Protein determination

Proteins present in the crude extract were determined by Bradford's method [32] using bovine albumin (Sigma Chem. Co., St. Louis, MO) as standard. During chromatographic separation, protein content of eluates was estimated by measuring the absorbance at 280 nm. Protein content of the active fractions was determined using the Lowry's method [33] with bovine albumin as standard.

2.5. Inhibitors effect

The action of different inhibitors of cysteine peptidases was evaluated by incubating the crude enzyme preparation for 10 and 30 min at 45°C with mercuric chloride and *trans*-epoxysuccinyl-L-leucylamido-(4-guanidino)butane (E-64). The residual caseinolytic activity after each incubation assay was measured as indicated above.

2.6. Purification of araujiain h-I, araujiain h-II, araujiain h-III

The purification of the main proteolytic components (*araujiain h*-I, *araujiain h*-II and *araujiain h*-III) was carried out by cation exchange chromatography (CM Sepharose CL-6B fast flow). Fourteen ml of the crude extract containing 75 mg of protein were loaded onto the column (Pharmacia K 15/30). The column was washed with 60 ml of 55 mM citrate–phosphate buffer (pH 6.4) and the bound material eluted with a linear gradient of sodium chloride (0–0.6 M) in the same buffer.

2.7. SDS-polyacrylamide gel electrophoresis

SDS-PAGE was performed according to Laemmli [34]. Samples were precipitated with 5 vol. of acetone and redissolved in sample buffer (62.5 mM Tris–HCl, pH 6.8 with 10% (v/v) glycerol and 0.25% (w/v) bromophenol blue) containing 10 μM E-64 to prevent autodigestion and boiled for 5 min; then, β-mercaptoethanol was added and the samples were boiled again. Samples were then loaded on a 14% polyacrylamide separating gel overlaid by a 5% polyacrylamide stacking gel. Electrophoresis was run in a vertical apparatus (Miniprotean II Cell, Bio-Rad, Hercules, CA) starting at 25 mA. When the samples left the stacking gel, the current was set at 50 mA until the bromophenol blue dye marker left the separating gel. The gels were stained with Coomassie Brilliant Blue R-250 and scanned for evaluation of

the molecular masses using the Scion Image Software (http://www. scioncorp.com).

2.8. Mass spectrometry

Matrix-assisted laser desorption ionization-time of flight mass spectrometry (MALDI-TOF MS) was used for the determination of the molecular masses, as well as the degree of purity of active chromatographic fractions. MALDI-TOF mass spectra were acquired on a BRUKER BIFLEX spectrometer equipped with a pulsed nitrogen laser (337 nm), in linear positive ion mode, using a 19 kV acceleration voltage. Samples were prepared by mixing equal volumes of a saturated solution of the matrix (3,5-dimethoxy-4-hydroxycinnamic acid–sinapic acid) in 0.1% TFA in water/acetonitrile 2:1, and a 1–10 mM protein solution. From this mixture, 1 μl was spotted on the sample slide and allowed to evaporate to dryness. Proteins of known molecular masses were used as standards for mass calibration.

2.9. Esterolytic activity determination

Measurement of endoesterolytic activity was performed according to Silverstein [35] with *N*-α-carbobenzoxy-*p*-nitrophenyl esters of some amino acids (Gln, Ala, Asp, Phe, Asn, Tyr, Trp, Gly, Leu, Ile, Val, and Pro). The synthetic substrates were obtained from Sigma Chem. Co. Assays were made at 40°C in 0.1 M Tris–HCl buffer (pH 8.0) containing 2 mM EDTA and 25 μM cysteine in the reaction mixture. Absorbance was measured at 405 nm every 10 s for the first minute, and then every 15 s. *Araujiain h*-I, *araujiain h*-II and *araujiain h*-III were used for this assay. Parallelly, blanks without enzymes and under the same reaction conditions were carried out. An arbitrary enzyme activity unit (U_{cbz}) was defined as the amount of peptidase that released 1.0 μM of *p*-nitrophenolate per min in the assay conditions. To determine the micromoles of *p*-nitrophenolate produced during the reaction, a standard curve (*p*-nitrophenol 15–70 μM) was carried out.

2.10. Structural descriptors

A large number of descriptors were extracted from the literature to characterize the amino acids under study. These can be divided into two major groups; those based on physicochemical or biochemical properties derived from experimental partition approaches, and a second group including nonempirical descriptors based on quantum chemical calculations.

The following experimental descriptors based on lipophilicity measurements were considered: the amino acid side-chain according to Levitt et al. (H_{LE}) [36], Fauchère et al. (H_F) [37], Cowan and Whittaker (H_{CW}) [38], Chmelik et al. (H_{CH}) [39], Tanford (H_T) [40], Radzicka and Wolfenden (H_{RW}) [41], and the scales of Manavalan et al. (Hf), (Gh), (Ha) [42,43]. In addition, several experimental parameters reported by Sandberg et al. [44], such as retention data obtained in seven thin-layer chromatography systems (TL1-TL7 here used as relative to Gly) were variables also screened in this study.

A further group of descriptors considered in this study included several variables compiled by Damborský [45]: local flexibility (Fr), flexibility index (Fb), chain flexibility (FO), molecular weight (MW), occupied volume by a residue buried (*V*) in globular protein, bulkiness (*B*), defined as the ratio of the side-chain volume to its length, and the dummy variable (*D*), which takes the value zero for amino acids without the aromatic side-chain and the unity for amino acids with an aromatic ring (Phe, Tyr and Trp).

The second group of variables included in the analysis that represent the whole amino acid were several nonempirical structural descriptors based on quantum chemical calculations. The calculations were performed using an AMBER force field with the HyperChem (Release 3.0 for Windows) for molecular mechanics and the MOPAC 6.0 [46] for semiempirical calculations. The structures of the amino acid derivatives were built in a fully extended conformation within the HyperChem package. To obtain reliable structures for the *N*-α-carbobenzoxy-aminoacid-*p*-nitrophenyl esters under study, a complete optimization of the geometrical parameters was carried out. The following quantum chemical indexes were considered: total energy (E_{total}), heat of formation (ΔH_f), energy of highest occupied molecular orbital (E_{HOMO}), energy of lowest unoccupied molecular orbital (E_{LUMO}), dipole moment (μ), the most positive partial charge on a hydrogen atom (qH^+), the most negative partial charge in the molecule (q^-) and the partial charges on the oxygen and carbon atoms (qC, qO) of the

carbonyl group which is attacked by the thiolate anion of cysteine peptidases.

2.11. Data analysis

The PLS projections in latent variables was the method here used to construct the QSAR models. PLS is based on the projection of the original multivariate data matrices down onto smaller matrices (T, U) with orthogonal columns, which relates the information in the response matrix Y to the systematic variance in the descriptor matrix X, as shown below.

$$X = \bar{X} + TP' + E$$

$$Y = \bar{Y} + UC' + F$$

$$U = T + H \text{ (the inner relation)}$$

where $\bar{X}$ and $\bar{Y}$ are the corresponding mean value matrices, T and U the matrices of scores that summarize the x and y variables, respectively, P the matrix of loadings showing the influence of the x in each component, C the matrix of weights expressing the correlation between Y and $T(X)$ and E, F, and H the corresponding residuals matrices. The PLS calculations also give an auxiliary matrix W (PLS weights), which expresses the correlation between U and X and is used to calculate T. Determinations of the significant number of model dimensions were done by crossvalidation [47]. For the QSAR analysis developed in the present study, the novel filtering technique called orthogonal signal correction (OSC) was used [48]. The OSC technique removes irrelevant systematic information from the data set leading to more powerful PLS models, because OSC uses the matrix Y to construct a filter of matrix X.

PLS analysis was carried out using the SIMCA 7.0 software package obtained from Umetri AB, Box 7960, 907 19 Umea, Sweden. All statistical and molecular orbital calculations were carried out on a PC-IBM computer.

3. Results and discussion

3.1. Isolation and purification of cysteine peptidases

Peptidases contained in the latex of *A. hortorum* Fourn. fruits were studied. The crude enzyme extract (supernatant of 100,000 × g) showed higher caseinolytic activity in the presence of 12 mM cysteine. Highest activity (four-fold) was recovered when the crude extract was obtained with 5 mM EDTA and 5 mM cysteine.

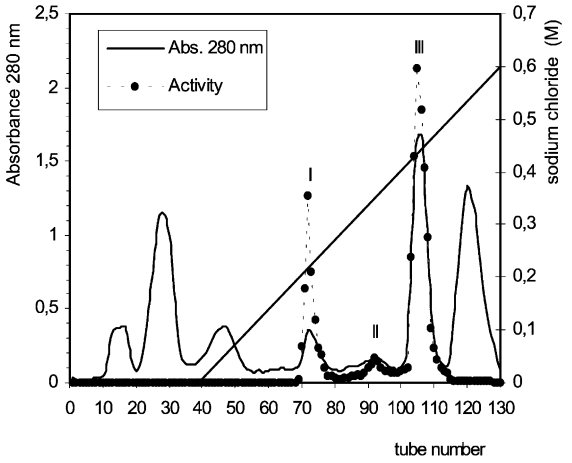


Fig. 1. Cation exchange chromatography (CM-Sepharose CL-6B fast flow, column Pharmacia K 15/30). Elution buffer 55 mM citric-phosphate (pH 6.4). Gradient: sodium chloride 0–0.6 M. Flow rate: 17 cm/h. Fraction volume: 1.6 ml. I: *araujiain h*-I, II: *araujiain h*-II, III: *araujiain h*-III.

Cation exchange chromatography of the crude extract yielded three main proteolytic peaks (Fig. 1): fraction I (*araujiain h*-I), fraction II (*araujiain h*-II) and fraction III (*araujiain h*-III), purified to mass spectroscopy homogeneity. The names *araujiain h*-I, *araujiain h*-II and *araujiain h*-III are proposed for these new peptidases, according to previous recommendations [11,12]. When the crude extract was previously inactivated by treatment with $HgCl_2$, the chromatographic pattern observed was identical to that of the untreated enzyme (data not shown), confirming that the obtained fractions are not autodigest artifacts.

Table 1 shows the purification scheme. A 4.75-fold purification degree (yield: 9.9%) was obtained for *araujiain h*-I, while for *araujiain h*-II and *araujiain h*-III, 1.65-fold (yield: 1.86%) and 2.6-fold (yield: 16.8%) purification degrees were obtained, respectively.

Inactivation assays with $HgCl_2$ and E-64 suggested the possible cysteinic nature of the peptidases present in the crude extract. When $HgCl_2$ was added,

Table 1
Purification of the proteolytic components present in the latex of *A. hortorum*

Sample	Volume (ml)	Protein (μg/ml)	Total proteins	UCAS/ (ml)	Total UCAS	Specific activity (UCAS)/(mg)	Purification (fold)	Yield (%)
Crude extract	7.5	1.067	8.0025	32.63	244.725	4.077	1	100
Araujiain h-I	19	0.066	1.254	1.280	24.320	19.40	4.75	9.93
Araujiain h-II	16.5	0.041	0.676	0.276	4.554	6.73	1.65	1.86
Araujiain h-III	21	0.185	3.885	1.960	41.160	10.59	2.60	16.8

proteolytic activity was almost completely reverted by adding cysteine to the incubation mixture. On the other hand when E-64 was added, the inhibition obtained for crude extract was complete and irreversible (Table 2). The activation, inhibition and reactivating results would indicate dependence of the proteolytic activity upon the presence of active SH groups in the enzymes present in the crude extract. The same behavior was previously shown by peptidases from other *Asclepiadaceae* species [10,12,14,17,18].

SDS-PAGE revealed that *araujiain h*-I, *araujiain h*-II and *araujiain h*-III appeared as unique bands, with relative masses of about 25 kDa, in good agreement with the values obtained by mass spectroscopy. The molecular masses of *araujiain h*-I, *araujiain h*-II and *araujiain h*-III obtained by mass spectroscopy were (24,030.87), (23,718) and (23,545.5), respectively (Fig. 2).

The endoesterolytic activities of *araujiain h*-I, *araujiain h*-II and *araujiain h*-III, can be analyzed at two levels. A qualitative analysis of the results shown in Table 3 suggests that the peptidases exhibit different patterns, in spite of the fact that all of them showed higher preference for the glutamine derivative. *Araujiain h*-I and *araujiain h*-II exhibited similar profiles, with the glutamine derivative followed by those of alanine, aspartic acid, glycine and leucine.

Table 2
Effect of inhibitors on proteolytic activity of crude extract

Inhibitors	Inhibitory concentration (mM)	Percentage activity
None		100
$HgCl_2$ (10 min)	0.1	0
$HgCl_2$ (10 min) + 12 mM Cys	0.1	96.25
E-64 (30 min)	0.01	0

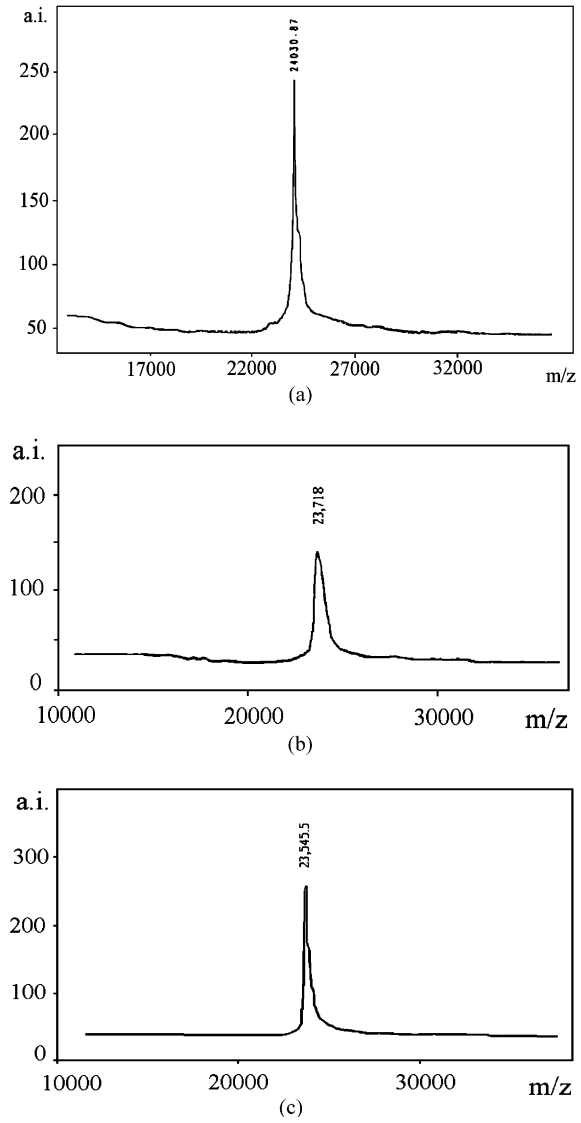


Fig. 2. Mass spectroscopy of (a) *araujiain h*-I; (b) *araujiain h*-II; (c) *araujiain h*-III.

Table 3
Endoesterolytic activity of *araujiain h*-I, *araujiain h*-II and *araujiain h*-III

N-CBZ-aminoacid-*p*-nitrophenyl ester	*Araujiain h*-I		*Araujiain h*-II		*Araujiain h*-III	
	U_{CBZ}	Percentage preference	U_{CBZ}	Percentage preference	U_{CBZ}	Percentage preference
Gln	59.02	100	6.06	100	72.38	100
Ala	27.50	47.61	2.84	57.89	29.45	41.66
Asp	8.50	16.13	2.84	57.89	71.61	98.96
Gly	5.10	10.45	1.08	31.58	56.34	78.12
Leu	4.85	10.06	1.47	36.84	25.62	36.46
Trp	2.57	6.28	0.70	26.32	29.45	41.66
Tyr	1.08	3.79	–	–	–	–
Asn	0.92	3.52	1.08	31.58	21.78	31.25
Phe	0.77	3.30	1.08	31.58	25.62	36.46
Val	0.32	2.54	0.00	0.00	6.45	10.42
Pro	0.00	0.00	0.00	0.00	0.00	0.00
Ile	–	–	0.00	0.00	0.00	0.00

Nevertheless, endoesterolytic activity of *araujiain h*-II on the alanine and aspartic derivatives is of the same order, while in *araujiain h*-I the alanine–aspartic relationship is 3:1. On the other hand, esterolytic behavior of *araujiain h*-III is quite different, as the enzyme showed similar preference for the glutamine and aspartic acid derivatives and also showed higher preference for the glycine derivative, followed by those of alanine and tryptophan.

Endoesterolytic activity can be also analyzed using the previously described QSAR methods, i.e. PLS regression, which provide information on fundamental intermolecular interactions that are important in governing the endoesterolytic activities of *araujiain h*-I, *araujiain h*-II and *araujiain h*-III.

3.2. PLS regression modeling of esterase activities

All variables used in the PLS calculations were initially autoscaled to zero mean and unit variance to give each descriptor equal importance in the PLS analysis. The statistical significance of the screened models was judged by the squared correlation coefficient (R^2) and the standard deviation (s). The predictive ability was evaluated by the squared crossvalidation coefficient (Q^2) which is based on the prediction error sum of squares (PRESS). Although the PLS method offers the advantage of handling data sets in which the number of independent variables is greater than the number of observations, considerably worse predictions are obtained if many irrelevant descriptors are included in the PLS model. Because of the large number of structural descriptors considered in this study, the VIP (variable importance for the projection) parameter [47] was used to unravel which descriptor variables were the most relevant to explain the esterolytic activity of isolated proteases. In the developed PLS models, the Y response matrix consisted of the enzyme activities showed in Table 3, while the X matrix consisted of several molecular parameters which were selected according to their VIP values (data not shown). A list of all descriptors used in the selected PLS models is given in Table 4.

The first PLS model (model 1) was performed on the $\log(U_{cbz} + 1)$ values of 11 *N*-α-carbobenzoxy-aminoacid-*p*-nitrophenyl esters for the *araujiain h*-I peptidase. As previously mentioned, the OSC–PLS methodology was applied in this study. Thus, the OSC analysis was carried out using the centered and scaled X matrix. After extracting four OSC-components, 30.6% of the original sum of squares remained in the corrected X matrix. The PLS analysis of this corrected data set resulted in a one-component PLS model with the following statistics: $R^2 = 0.998$, $Q^2 = 0.977$, $s = 0.028$. The second PLS model (model 2) was based on the same descriptor variables as used in model 1 and the activity data used were the $\log(U_{cbz} + 1)$ values of *araujiain h*-II peptidase. The OSC analysis resulted in 35.52% of the original sum of squares of matrix X after extracting four OSC-components. The

Table 4
Molecular parameters used in the developed PLS models[a]

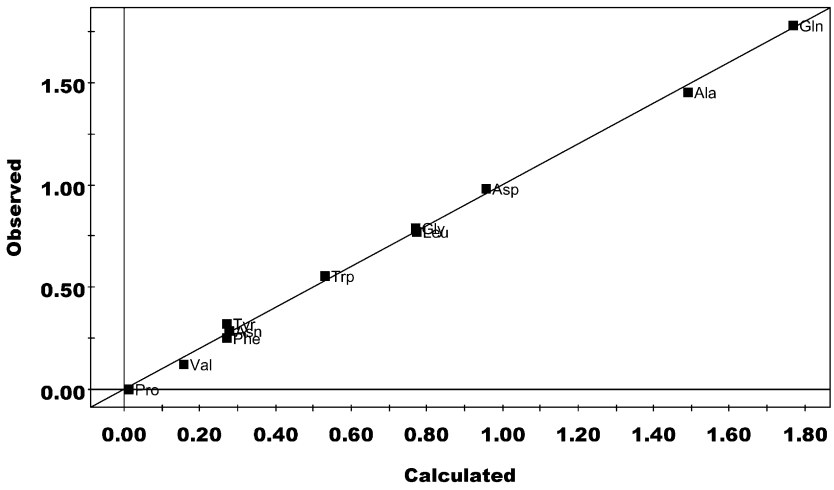
N-CBZ-aminoacid-*p*-nitrophenyl ester	H_{LE}	H_F	H_{RW}	H_{CW}	H_{CH}	B (Å^2)	Fr	Fb	FO	D
Gln	−0.20	−0.22	−5.54	−0.93	−0.30	14.50	1087.83	0.493	1.165	0
Ala	0.50	0.31	1.81	0.35	0.30	11.50	705.42	0.357	1.041	0
Asp	−2.50	−0.77	−8.72	−2.15	−0.55	11.70	34.96	0.511	1.033	0
Gly	0.00	0.00	0.94	0.00	0.00	3.40	33.18	0.544	1.142	0
Leu	1.80	1.70	4.92	1.80	1.50	21.40	4985.73	0.365	0.967	0
Trp	3.40	2.25	2.33	1.35	2.01	21.60	6374.07	0.305	0.925	1
Tyr	2.30	0.96	−0.14	0.39	0.85	18.00	4291.10	0.420	0.961	1
Asn	−0.20	−0.60	−6.64	−0.99	−0.48	12.80	513.46	0.463	1.117	0
Phe	2.50	1.79	2.98	1.69	1.91	19.80	5203.86	0.314	0.930	1
Val	1.50	1.22	4.04	1.32	1.07	21.60	4474.42	0.386	0.982	0
Pro	1.40	0.72	−0.52	0.84	0.53	17.40	431.96	0.509	1.055	0
Ile	1.80	1.80	4.92	1.83	1.53	21.4	5979.37	0.462	1.002	0
	TL2	TL3	TL6	TL7	E_{HOMO}[b]	E_{LUMO}[c]	μ[d]	qH^{+}[e]	q^{-}[e]	
Gln	9	−3	0	−2	−9.7284	−1.1731	6.149	0.2324	−0.3650	
Ala	16	5	20	1	−9.6715	−1.1788	6.682	0.1758	−0.3599	
Asp	22	−3	6	−2	−9.5070	−0.9740	8.249	0.2385	−0.4264	
Gly	0	0	0	0	−9.6668	−1.0959	6.469	0.1977	−0.3629	
Leu	38	33	41	31	−9.6699	−1.1093	7.908	0.1906	−0.3603	
Trp	42	35	17	29	−9.1184	−1.2203	8.293	0.2039	−0.3584	
Tyr	37	31	36	29	−9.4937	−1.0195	6.491	0.2226	−0.3601	
Asn	−5	−5	−11	−10	−9.5219	−0.9892	8.172	0.2385	−0.4265	
Phe	38	31	34	30	−9.5960	−1.1361	6.533	0.2056	−0.3611	
Val	30	23	36	21	−9.5455	−1.0732	6.306	0.1740	−0.3626	
Pro	8	1	25	9	−9.4838	−1.1115	6.272	0.1853	−0.3618	
Ile	36	31	40	31	−9.2719	−1.2724	6.228	0.1809	−0.4803	

[a] For explanation of the symbols of molecular parameters see text.
[b] Energy of the highest occupied molecular orbital (eV).
[c] Energy of the lowest empty molecular orbital (eV).
[d] Dipole moment calculated by AM1 (Debye).
[e] The charge (q) on the atoms were computed by AM1.

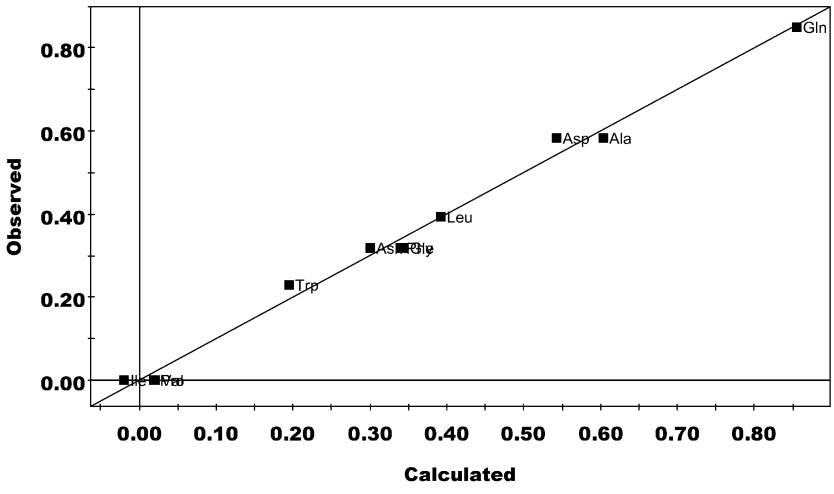
PLS analysis of this corrected data set resulted in a one-component PLS model with the following statistics: $R^2 = 0.992$, $Q^2 = 0.983$, $s = 0.026$. Finally, the third PLS model (model 3) was based on the $\log(U_{cbz} + 1)$ values of *araujiain h*-III peptidase. The OSC methodology was also applied in this case and again, after extracting four OSC-components, 31.1% of the original sum of squares remained in the corrected X matrix. The PLS analysis of this corrected data set resulted in a one-component PLS model with the following statistics: $R^2 = 0.996$, $Q^2 = 0.982$, $s = 0.042$. All obtained models are highly significant statistically and even though the number of data points used in the models was small, the high R^2 and Q^2 values provides confidence that the obtained results are not artifactual. The agreement between the observed and calculated values is very good, as shown in Fig. 3a–c.

In order to interpret the obtained PLS models, the weights of the descriptors variables on the OSC–PLS components were analyzed. From these values it can be seen how a single variable contributes in each OSC–PLS component to the modeling of the observed esterase activities. The PLS weights for the established one-dimensional QSAR models are listed in Table 5 . On the basis of these results, it appears that the variation of the amino acid side-chain has a strong effect on the activity of the compounds, indicating that the electronic and steric factors play an important role in the enzyme-ligand binding. On the other hand, on analyzing the PLS weights, the negative coefficients for all lipophilicity scales indicate a detrimental effect

of the hydrophobicity of substituents at the enzyme binding site. This suggests that, since hydrophobicity is related to bulkiness of the group, substituents would be producing steric hindrance instead of leading to any hydrophobic interaction with the enzyme. Further, it is possible, that the rigidity of the phenyl ring of Trp, Tyr and Phe does not allow the substituents to interact well with the hydrophobic space of the ligand binding pocket. This view is consistent with the large positive coefficient of FO, which reflects the importance of the chain flexibility in the interaction between the substituents and the active site. Contrary to this analysis, inspection of Table 5 reveals that the Fr flexibility descriptor exhibits a large negative coefficient in all OSC–PLS models. However, it should be noted that there is some collinearity between the lipophilicity scales and Fr (H_{LE} and Fr ($r = 0.82$), H_F and Fr ($r = 0.92$), H_{RW} and Fr ($r = 0.71$), H_{CW} and Fr ($r = 0.79$), H_{CH} and Fr ($r = 0.923$)), indicating that some component of hydrophobicity is

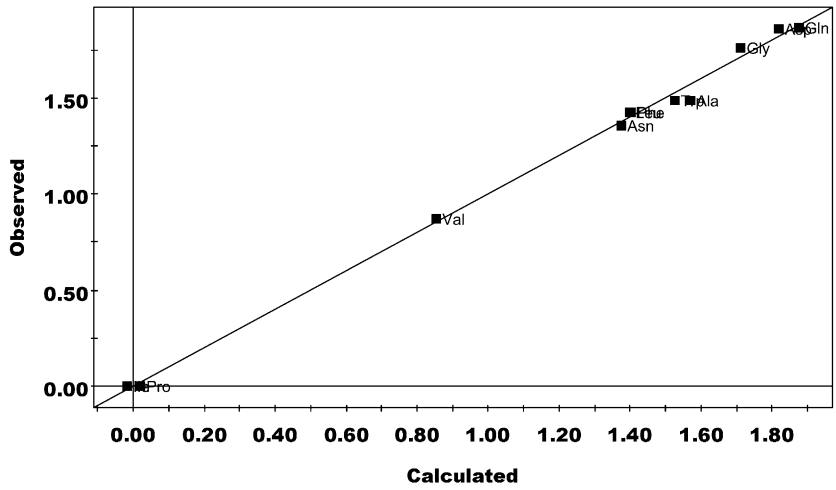


(a)



(b)

Fig. 3. Relationships between the experimental and calculated $\log(U_{cbz} + 1)$ values for *araujiain h*-I, *araujiain h*-II and *araujiain h*-III: (a) model 1; (b) model 2; (c) model 3.



(c)

Fig. 3 (*Continued*).

Table 5
PLS weights of QSAR models 1–3[a]

Descriptor	OSC-model 1 WC(1)	OSC-model 2 WC(1)	OSC-model 3 WC(1)
H_{LE}	−0.2937	−0.2828	−0.3209
H_F	−0.3460	−0.3363	−0.3489
H_{RW}	−0.1667	−0.2770	−0.3046
H_{CW}	−0.2580	−0.3113	−0.3482
H_{CH}	−0.3518	−0.3321	−0.3334
B	−0.2446	−0.2484	−0.2386
Fr	−0.2999	−0.2855	−0.2776
Fb	*0.1363*	*0.0608*	*−0.0633*
FO	*0.2856*	*0.2540*	*0.1401*
D	*−0.1568*	*−0.0472*	*0.1066*
TL2	−0.1635	−0.1980	–
TL3	−0.3346	−0.3174	–
TL6	−0.2178	−0.2547	–
TL7	−0.3395	−0.3232	–
E_{HOMO}	–	–	−0.2468
E_{LUMO}	–	–	0.1870
μ	–	–	0.2373
$q\mathrm{H}^+$	–	–	0.2949
q^-	–	–	0.2047

[a] Some descriptors weights have been shown in italics for clarity (see text). The dashes showed in the table indicate that these descriptors were not included in the OSC–PLS models. The number between parentheses in WC(1) indicates one-component of OSC–PLS model.

embodied in the Fr term. This explains the apparent contradiction between the signs of FO and Fr.

On the other hand, the large negative coefficient of term *B* (the ratio of side-chain volume to its lengths) provides additional support for a negative steric effect by comparatively bulkier groups, which limit the observed esterase activity. The values of $\log(U_{cbz} + 1)$, FO and *B* variables given in Tables 3 and 4 for Gln, Ala and Gly or Asp and Ile provide a good example of this. It is of interest here to note that the *N*-α-carbobenzoxy-*p*-nitrophenyl esters of Leu and Ile exhibit very different activity values even though their lipophilicity scales and FO and *B* values are quite similar. However, comparison of the Fr values for these compounds (see Table 4) shows that it is indeed the local flexibility contribution (calculated by multiplying the hydrophobicity scale, Hf, by the average volume of a buried residue in a globular protein, *V*) which accounts for the difference in the observed activity between these compounds.

Regarding electronic effects, it can be observed from Table 5 that the hydrogen-bonding interactions play an important role in the specific binding of the amino acid derivatives to the polar regions of *araujiain h*-I, *araujiain h*-II and *araujiain h*-III target site. With respect to *araujiain h*-I and *araujiain h*-II (models 1 and 2), the variables used were those based on the retention data in thin-layer chromatography (TL2,

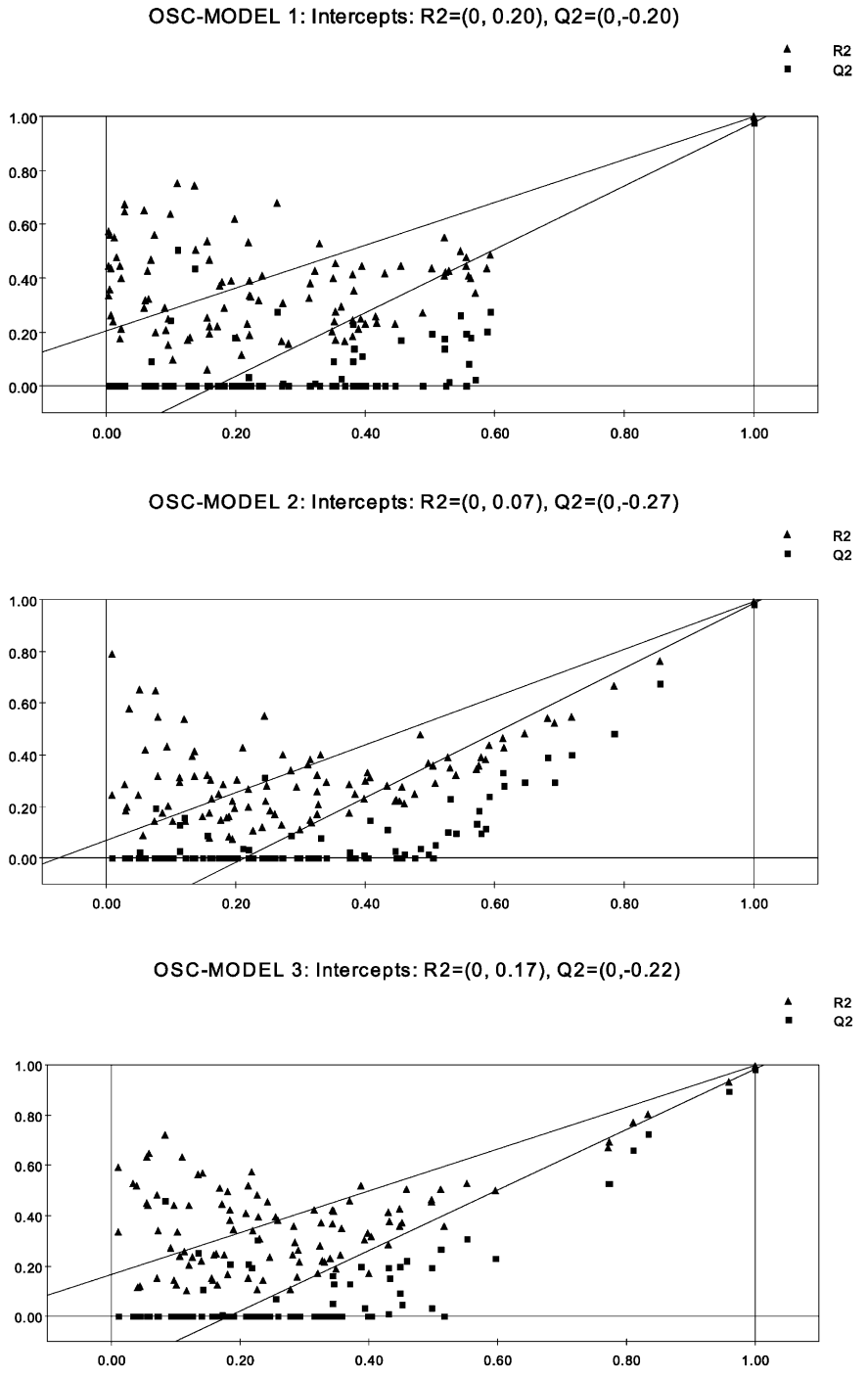


Fig. 4. Results of the permutation test. The R^2 and Q^2 values were obtained from 100 permutations and one OSC–PLS component.

TL3, TL6, TL7). Taking into account that these variables are based on normal-phase TLC systems, it may be speculated that they concentrate structural information related to specific polar interactions between the amino acid derivatives and the active site of enzymes. This conclusion is in agreement with findings by other authors with respect to the proposed retention mechanism for these stationary phases [49]. On the other hand, in contrast to the other models, for *araujiain h*-III (model 3) the TLC-variables make only a small contribution to the calculated activities. Thus, in order to reflect the ability of the compounds to participate in electronic interactions with the enzyme, various quantum chemical parameters were selected according to their VIP values. On analyzing the PLS weights, it may be concluded that these variables reflect a compromise between the charge-transfer phenomena (E_{HOMO}, E_{LUMO}) on the one hand and the electrostatic interactions (μ, qH^+, q^-) on the other.

3.3. Model validation

It is well known that the real predictive ability of any QSAR model cannot be judged solely by using internal validation techniques such as crossvalidation. Thus, the validity of the models was additionally tested by a permutation test [50]. Models were recalculated for randomly reordered response data (Y). These permuted y-values were related to intact predictor data by refitting the model and including crossvalidation. When plotting R^2 and Q^2 as a function of the correlation coefficient between the original values and the predicted values, the intercept with the y-axis expressed to which degree these values rely on chance. Fig. 4 shows the results obtained from 100 permutations for each one of the compounds under study. The intercepts of the two regression lines (for R^2 and Q^2) indicate the degree of overfit and overprediction. In general, intercept limits for $R^2 < 0.3$ and $Q^2 < 0.05$ indicate valid models, as is the case of the OSC–PLS models here developed.

4. Conclusions

Three new cysteine peptidases (*araujiain h*-I, *araujiain h*-II and *araujiain h*-III) were isolated and purified to mass spectroscopy homogeneity from the latex of *A. hortorum* (*Asclepiadaceae*) fruits by ultracentrifugation and ion exchange chromatography. The enzymes showed molecular masses of 24,030.87; 23,718 and 23,545.5 (mass spectrometry), respectively.

The developed QSAR models represent highly significant PLS regressions and show the principal factors that govern the esterolytic activities of the three isolated peptidases. Even though the number of data points used in each model is small, the high R^2 and Q^2 obtained values establish the importance of the hydrophobic, electronic, and steric characteristics of the substituents. Thus, from all these QSAR models, it appears that hydrophobic and steric factors reduce the enzymatic activity while the electronic interactions are of primary importance for an increase of the esterolytic activities. Finally, this study provides evidence for the great potential of PLS regression modeling based on the OSC methodology for the development of QSAR models.

Acknowledgements

Amino acid sequence and mass spectrometry determinations were done at the Institut de Biologia Fonamental, Universitat Autònoma de Barcelona, Spain. N.O.C. and M.C.A. are members of the CIC Researcher Career and Support Professional Career programmes, respectively. The present work was supported by grants from CIC, ANPCyT, University of San Luis and University of La Plata, Argentina.

References

[1] M.F. Chaplin, C. Bucke, Enzyme Technology, 1st Edition, Cambridge University Press, New York, USA, 1990, Chapter 4, pp. 142–145.
[2] S. Barberis, A. Illanes, Enzyme Biotechnology, 1st Edition, 35th Monograph, Biology Series, The Secretariat of the Organization of American States, Washington, DC, USA, 1994, Annex, pp. 250–253.
[3] W. Kullman, in: R.L. Schowen, A. Barth (Eds.), Advances in the Biosciences, Vol. 65, Peptides and Proteases: Recent Advances, Pergamon Press, UK, 1986, pp. 135–140.
[4] W.J. Lahl, S.D. Brown, Food Technol. 48 (1994) 68–71.
[5] M.I. Mahmoud, Food Technol. 48 (1994) 89–95.
[6] T. Winnick, A.R. Davis, D.M. Greenberg, J. Gen. Physiol. 23 (1940) 275–288.
[7] D.M. Greenberg, T. Winnick, J. Biol. Chem. 135 (1940) 775–780.

[8] D.C. Carpenter, F.E. Lovelace, J. Am. Chem. Soc. 65 (1943) 2364–2365.
[9] W.J. Brockbank, K.R. Lynn, Biochim. Biophys. Acta 578 (1979) 113–122.
[10] K.R. Lynn, W.J. Brockbank, N.A. Clevette-Radford, Biochim. Biophys. Acta 612 (1980a) 119–125.
[11] B.E. Barragán, M.T. Cruz, L.M. del Castillo, M. Castañeda-Agulló, Rev. Latinoamer. Quím. 16 (1985) 117–119.
[12] M. Tablero, R. Arreguín, B. Arreguín, M. Soriano, R.I. Sánchez, A. Rodríguez Romero, A. Hernández-Arana. Plant Sci. 74 (1991) 7–15.
[13] K.I. Abraham, P.N. Joshi, Biochim. Biophys. Acta 568 (1979a) 111–119.
[14] K.I. Abraham, P.N. Joshi, Biochim. Biophys. Acta 568 (1979) 120–126.
[15] G. Pal, N.K. Sinha, Arch. Biochem. Biophys. 202 (1980) 321–329.
[16] A. Sengupta, D. Bhattacharya, G. Pal, N.K. Sinha, Arch. Biochem. Biophys. 232 (1984) 17–25.
[17] M.C. Arribére, A.A. Cortadi, M.A. Gattuso, M.P. Bettiol, N.S. Priolo, N.O. Caffini, Phytochem. Anal. 9 (1998) 1–7.
[18] M.C. Arribére, S.E. Vairo Cavalli, N.S. Priolo, N.O. Caffini, M. Gattuso, A.A. Cortadi, Acta Horticulturae 501 (1999) 259–268.
[19] A.J. Barret, N.D. Rawlings, J.F. Woessner, Handbook of Proteolytic Enzymes, Academic Press, London, 1998, pp. 545–797.
[20] H. Kubinyi, Curr. Opin. Drug Design Discov. 1 (1998) 4–15.
[21] J.G. Topliss, R.P. Edwards, J. Med. Chem. 22 (1979) 1238–1244.
[22] H. Kubinyi, in: R. Mannhold, P. Krosgaard-Larsen, H. Timmerman (Eds.), QSAR: Hansch Analysis and Related Approaches, VCH, Weinheim, 1995, Chapter 5, pp. 91–107.
[23] J.M. Luco, M.E. Sosa, J.C. Cesco, C.E. Tonn, O.S. Giordano, Pestic. Sci. 41 (1994) 1–6.
[24] J.M. Luco, J.L. Yamin, H.F. Ferretti, J. Pharm. Sci. 84 (1995) 903–908.
[25] J.M. Luco, H.F. Ferretti, J. Chem. Inf. Comput. Sci. 37 (1997) 392–401.
[26] J.M. Luco, J. Chem. Inf. Comput. Sci. 39 (1999) 396–404.
[27] M.P. Montaña, N.B. Pappano, N.B. Debattista, J. Raba, J.M. Luco, Chromatographia 51 (2000) 727–735.
[28] M.F. Silva, L.F. Chipre, J. Raba, J.M. Luco, Chromatographia 53 (2001) 392–400.
[29] A. Burkart, Flora ilustrada de Entre Ríos (Argentina), Parte V, Dicotiledóneas Metaclamídeas (Gamopétalas), Col. Cient. del INTA, Buenos Aires, 1979, pp. 173–175.
[30] M.J. Dimitri, Enciclopedia Argentina de Agricultura y Jardinería, Acme SACI, Buenos Aires, 1972, pp. 774–780.
[31] N. Priolo, S. Morcelle del Valle, M. Arribére, L. López, N. Caffini, J. Protein Chem. 19 (2000) 39–49.
[32] M.M. Bradford, Anal. Biochem. 72 (1976) 248–254.
[33] O. H Lowry, N.J. Rosebrough, L. Farr, R.J. Randall, J. Biol. Chem. 193 (1951) 265–275.
[34] U.K. Laemmli, Nature 227 (1970) 680–685.
[35] R.M. Silverstein, Anal. Biochem. 62 (1974) 478–484.
[36] M. Levitt, J. Mol. Biol. 104 (1976) 59–107.
[37] J.L. Fauchère, M. Charton, L.B. Kier, A. Verloop, V. Pliska, Int. J. Pept. Prot. Res. 32 (1988) 269–278.
[38] R. Cowan, R.G. Wittaker, Peptide Res. 3 (1990) 75–80.
[39] J. Chmelik, J. Hudecek, K. Putyera, J. Makovicka, V. Kalous, J. Chmelikova, Collect. Czech. Chem. Commun. 56 (1991) 2030–2041.
[40] C. Tanford, J. Am. Chem. Soc. 84 (1962) 4240–4247.
[41] A. Radzicka, R. Wolfenden, Biochemistry 27 (1988) 1664–1670.
[42] P. Manavalan, P.K. Ponnuswamy, Nature 275 (1978) 673–674.
[43] P.K. Ponnuswamy, M. Prabhakaram, P. Manavalan, Biochim. Biophis. Acta 623 (1980) 301–316.
[44] M. Sandberg, L. Eriksson, J. Jonsson, M. Sjöström, S. Wold, J. Med. Chem. 41 (1998) 2481–2491.
[45] J. Damborský, Quant. Struct.-Act. Relat. 2 (1997) 126–135.
[46] M.S. Dewar, E.G. Zoebisch, E.F. Healy, J.J.P. Stewart, J. Am. Chem. Soc. 107 (1985) 3902–3909.
[47] S. Wold, in: H. van de Waterbeemd (Ed.), Chemometric Methods in Molecular Design, VCH, Weinheim, 1995, Chapter 4.4, pp. 195–218.
[48] S. Wold, H. Antii, F. Lindgren, J. Öhman, Chemom. Intell. Lab. Syst. 44 (1998) 175–185.
[49] R. Kaliszan, in: J.D. Winefordner (Ed.), Quantitative Structure–Chromatographic Retention Relationships, Wiley, New York, 1987 (Chapter 3).
[50] L. Eriksson, E. Johansson, M. Muller, S. Wold, Quant. Struct.-Act. Relat. 16 (1997) 383–390.