Peptides 32 (2011) 1924–1933

Contents lists available at ScienceDirect

Peptides

journal homepage: www.elsevier.com/locate/peptides

Chemometric analysis of Hymenoptera toxins and defensins: A model for predicting the biological activity of novel peptides from venoms and hemolymph

Daniel M. Saidemberg[a,b], Nicoli B. Baptista-Saidemberg[a,b], Mario S. Palma[a,b,*]

[a] Center of Study of Social Insects (CEIS)/Dept. Biology, Institute of Biosciences of Rio Claro, São Paulo State University (UNESP), Rio Claro, SP 13506-900, Brazil
[b] Instituto Nacional de Ciência e Tecnologia (INCT) em Imunologia/iii, São Paulo, SP, Brazil

ARTICLE INFO

Article history:
Received 10 June 2011
Received in revised form 29 July 2011
Accepted 1 August 2011
Available online 7 August 2011

Keywords:
Polycationic peptides
Insect venoms
PCA
System biology
Toxins

ABSTRACT

When searching for prospective novel peptides, it is difficult to determine the biological activity of a peptide based only on its sequence. The "trial and error" approach is generally laborious, expensive and time consuming due to the large number of different experimental setups required to cover a reasonable number of biological assays. To simulate a virtual model for Hymenoptera insects, 166 peptides were selected from the venoms and hemolymphs of wasps, bees and ants and applied to a mathematical model of multivariate analysis, with nine different chemometric components: GRAVY, aliphaticity index, number of disulfide bonds, total residues, net charge, pI value, Boman index, percentage of alpha helix, and flexibility prediction. Principal component analysis (PCA) with non-linear iterative projections by alternating least-squares (NIPALS) algorithm was performed, without including any information about the biological activity of the peptides. This analysis permitted the grouping of peptides in a way that strongly correlated to the biological function of the peptides. Six different groupings were observed, which seemed to correspond to the following groups: chemotactic peptides, mastoparans, tachykinins, kinins, antibiotic peptides, and a group of long peptides with one or two disulfide bonds and with biological activities that are not yet clearly defined. The partial overlap between the mastoparans group and the chemotactic peptides, tachykinins, kinins and antibiotic peptides in the PCA score plot may be used to explain the frequent reports in the literature about the multifunctionality of some of these peptides. The mathematical model used in the present investigation can be used to predict the biological activities of novel peptides in this system, and it may also be easily applied to other biological systems.

1. Introduction

Peptides may be constituents of larger proteins, in which case they are responsible for molecular recognition and biological activities, or they may be biosynthesized for important roles in many physiological processes, acting as neurotransmitters, hormones, toxins, antibiotics, and defensins [43]. Peptides in general target a wide variety of protein receptors at the level of biological membranes and may interact with the phospholipids of the plasma/organelle membranes and/or with cytosolic proteins, which may regulate their activities. Peptides are used as toxins in animal venom as part of the chemical weapons arsenal for predation and/or defense purposes, and they can even be used to protect the host from infections by pathogens [42]. These peptides are directed against a wide range of pharmacological targets, and they can induce pain, inflammation, blood pressure changes, heart arrhythmia, and neurotoxicity, among other toxic actions [12]. Many of the peptides from animal toxic secretions seem to have evolved convergently with their cellular and molecular targets to optimize their effects, making them highly selective ligands for specific types of receptors [56]. However, peptides are metabolically unstable due to protease cleavage of their backbone chain, and they also have poor bioavailability, mainly due to the low membrane transport characteristics of the peptide's amide backbone structure.

The interaction of some peptides with their biological targets may occur through the direct binding of their linear sequences in a potentially large number of conformations that are accessible to these peptides. The pressure for conservation of the primary structures of the peptide toxins/defensins from animal venoms/hemolymph during evolution for each group of venomous animals has been non-uniform among these groups [21]. Apparently, the major factor determining the level of conservation/modification of amino acid sequences during evolution was probably the necessity of obtaining high affinity binding to one or more specific receptors [43]. The venoms/hemolymph of many wandering Arthropods evolved to contain structurally compact

* Corresponding author at: CEIS/Dept. Biology, IBRC-UNESP, Rio Claro, SP, Brazil. Tel.: +55 19 352653 65; fax: +55 19 352653 65.
E-mail address: mspalma@rc.unesp.br (M.S. Palma).

0196-9781/$ – see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.peptides.2011.08.001

peptides due to the presence of disulfide bonds, which stabilize the tertiary structure of these peptides. This stabilization is necessary to make the peptides active such that they can suitably perform their biological functions. These peptides are characterized both by their compact tertiary structures and by their high affinity for their specific receptors [18,52]. Thus, for different groups of venomous organisms, nature has adopted a different strategy to create and evolve the peptide toxins based on the biology, life history, longevity, and foraging/feeding behavior of the organisms, among other parameters [43].

Snake venom evolved to present linear peptides acting at the level of receptors localized on the endothelium surface, which causes a decrease in the blood pressure of the victims [19,20]. These peptides usually define their secondary structures during their interaction with the targeted receptors. The evolution of the toxins from the venoms/hemolymph of spiders and scorpions resulted in many peptides with compact tertiary structures, which bind with high affinity to nervous receptors, modulating ion flux through the cellular membranes [21]. The skin secretions of frogs evolved to create a wide variety of linear, antimicrobial peptides [53]. Meanwhile, the action of evolution in the venoms/hemolymph from Hymenoptera insects resulted in a series of short, linear, polycationic peptides with multifunctional activities, which cause pain [6], antimicrobial actions [11,16], and inflammation processes characterized by mast cell degranulation [42], chemotaxis of polymorphonucleated leukocytes, and cytolysis [10].

Many studies focusing on structure/activity relationship (SAR) have been conducted with specific groups of peptides to understand their mechanisms of action, and to create a rationale for the development of novel peptides with the potential to become drugs for therapeutic applications [46,48]. There are many literature data on *in silico* prediction of peptides biological activities, such as: the use of prediction of activity spectra for substances (PASS) computer system, which showed a high predictive power for nine known biological activities [37]; the use of molecular electronegativity topological distance vector (METDV), which is based on Pauling electronegativity and topological distance between each non-hydrogen atom for the determination of structure activity relationship of bradykinin-potentiating pentapeptides for angiotensin converting enzyme (ACE) inhibition activity [59]; These ACE inhibitors were also studied by quantitative structure–activity relationships (QSARs) studies using a set of amino acid descriptors including hydrophobic, stereo and electrical characteristics on angiotensin-converting enzyme (ACE) inhibitor dipeptides, bactericidal peptides and oxytocin peptides, showing that structural properties are important to understand and to predict bioactivities [57]. The variables most frequently investigated are peptide hydrophobicity and amphipathicity index, net positive/negative charge [13], pI value [31,34], content of α-helix and/or β-strands, and number of disulfide bond-forming cysteine residues [9,32,40,51], among others [49]. However, it is known that the presentation of specific biological activities and their modulation results from the interaction of all these parameters simultaneously, not from a single factor individually. Thus, experimentally, it is necessary to synthesize huge libraries of peptide sequences, reflecting all the combinatorial possibilities of the parameters mentioned above, and to evaluate the biological activity of each specific sequence to potentially obtain a single sequence or a few leader sequences, making this task quite challenging.

However, nature provides some interesting biological systems that can be used as models for virtual studies of this type, such as the large number of components from the venom and hemolymph of Hymenoptera insects. These components consist of a large variety of linear, α-helical peptides and some disulfide bond-containing peptides, each with different types of biological activities [12,42,43]. There is constant selection for the peptide sequences that are successful in specific target binding, while those that are not efficient are discarded in each different physiological/pharmacological situation. The richness of peptides from Hymenoptera venoms and hemolymphs constitutes a suitable biological system for searching through different sequences, providing a large number of different combinations of the physicochemical parameters mentioned above, which affect specific functional roles. Thus, we used this biological system to investigate the simultaneous combination of six different chemometric components in a mathematical model of multivariate analysis that was blinded from information about the biological activity of each peptide. We calculated the values of GRAVY (an index of hydrophobicity), aliphaticity index of the side chain of the amino acid residues of each peptide chain, number of disulfide bonds, total residues, net charge, isoelectric point (pI), Boman index, percentage of alpha helix, and Karplus & Schulz Flexibility Prediction, and included these values in a PCA protocol, with partial least squares regression (NIPALS), to determine whether these parameters contribute to a mathematical grouping of the sequences. In addition, we tried to correlate the observed grouping with the biological activity of each group. This model was validated with a series of other polycationic peptides from other animal origins.

2. Materials and methods

The amino acid sequences of 166 peptides from the venoms and hemolymph of Hymenoptera insects (bees, wasps and ants) were obtained from UNIPROT (http://www.uniprot.org) and NCBI (http://www.ncbi.nlm.nih.gov), and their sequences, numbering and names are shown in Supplemental Table 1 (supplementary content). The physico-chemical properties were calculated by Protparam (http://ca.expasy.org/tools/protparam.html), Peptide Property Calculator (http://www.peptideresource.com/software.html), Boman index (http://aps.unmc.edu/AP/prediction/prediction_main.php), alpha helix (%) by Consensus Data Mining secondary structure prediction (CDM) (http://gor.bb.iastate.edu/cdm/), and Karplus & Schulz Flexibility Prediction (http://tools.immuneepitope.org/tools/bcell/iedb_input).

To validate the model constructed for the Hymenoptera peptides, 80 peptides from other organisms were used, and their sequences, numbering, names and the supporting literature are shown in Supplemental Table 2 (supplementary content).

The physicochemical parameters calculated for each peptide sequence were grand average of hydropathicity (GRAVY), aliphatic index, isoelectric point (pI), net charges, number of amino acid residues, number of disulfide bonds, flexibility, alpha helix (%), and Boman index (kcal/mol). The aliphatic index of a protein is calculated according to the formula [24]:

$$\text{Aliphatic index} = X(\text{Ala}) + aX(\text{Val}) + b[X(\text{Ile}) + X(\text{Leu})]$$

- X(Ala), X(Val), X(Ile), and X(Leu) are mole percent ($100 \times$ mole fraction) of alanine, valine, isoleucine, and leucine, respectively.
- a and b are the relative volumes of valine side chain ($a = 2.9$) and of Leu/Ile side chains ($b = 3.9$), respectively, in relation to the side chain volume of alanine.

Boman index is an estimate of the potential of peptides/proteins to bind to other proteins and is the sum of the free energies of the amino acid residue side chains, divided by the total number of amino acid residues; this index is expressed as kcal/mol [5]. Among all the peptides, a lower index value indicates that the peptide likely has more antibacterial activity without many side effects, whereas a higher index value indicates that the peptide is multifunctional with hormone-like activities. The index values for the defensins

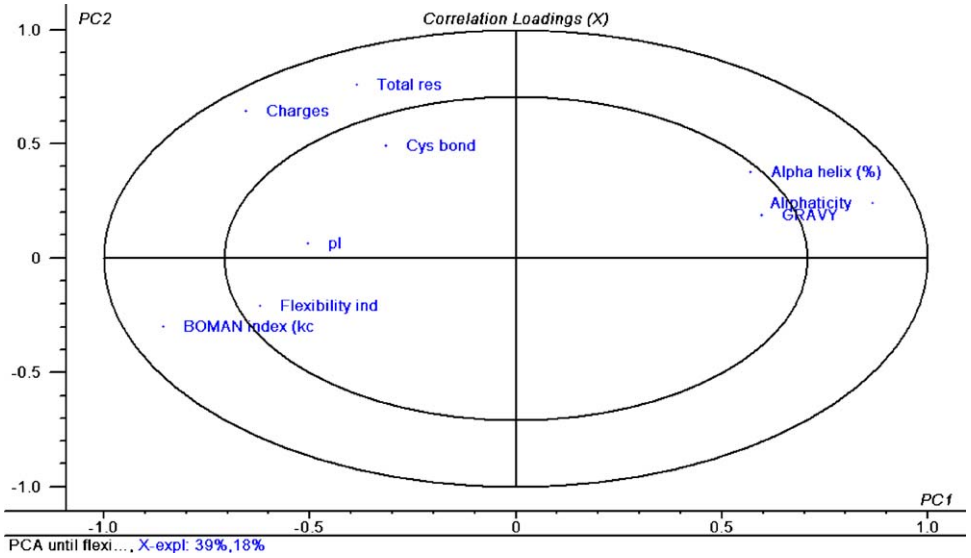

Fig. 1. PCA X-loadings plot showing the correlation between the variables.

are in the intermediate range [5]. The Karplus & Schulz Flexibility Prediction is a tool for the selection of peptide antigens [26]. For the estimation of alpha helix percentage we used the CDM prediction. The basic premise of CDM is that the combination of two complementary methods can enhance the performance of the overall secondary structure prediction by harnessing the distinct advantages for both methods [8,29,54,55]:

- Structural Fragment Database Mining (FDM) exploits the availability of sequentially similar fragments in the PDB, which leads to highly accurate (much better than GOR V) prediction of structure for such fragments, but such fragments are not available for many cases.
- GOR V is an algorithm that predicts the secondary structure of less similar fragments fairly accurately, where usually the FDM method does not find suitable structures [8,29,54,55].
- The PCA was performed using The Unscrambler software, version 9.8 (CAMO A/S, Trondheim, Norway), and the model was validated by full cross validation as reported by Esbensen [17]. The outliers were manually removed to derive a better prediction model. The general PCA formula is: $X = TP^{T} + E$

As the PCA model is centered, it gives:

$$X = 1 \cdot x_{\text{mean}} + T_{(A)} \cdot P_{(A)}^{T} + E_{(A)}$$

where: X – the x value; $T_{(A)}$ – the score of the (A) component; P – the X-loading; and $E_{(A)}$ – x-residuals for a model using (A) PCs.

The algorithms used in The Unscrambler for PCA are described in Martens and Næs [36]. The software uses the NIPALS algorithm, which extracts one variable at a time. Each factor is obtained iteratively on the "T" scores to obtain a better score. The current version of the software permits use of a stop criteria based on: $||\text{told} - t|| < 1e-12$, which gives more strict orthogonality in scores and loadings; the maximum number of iterations was 100.

Later, the individual position of each point (peptide) is identified and verified if the points with similar biological activity are grouped neighbor to each other, forming a group; this is done manually, using the help of the algorithm, which automatically identifies each peptide.

3. Results

The PCA grouping of peptide classes was mathematically determined by the physicochemical parameters (grand average hydrophobicity index (GRAVY), aliphaticity index, number of disulfide bonds, total number of residues, net charge, and isoelectric point (pI)), flexibility index, percentage of alpha helix, and Boman index without any use of alignment of sequences; i.e., the peptides were classified only according to their intrinsic properties without including any influence from their biological activity. Positive values of GRAVY are indicative of hydrophobicity, while negative values are indicative of hydrophilicity [30]. The aliphatic index of a peptide is considered to be the relative volume occupied by aliphatic side chains (alanine, valine, isoleucine, and leucine). Positive values for this index are related to an increase in the stability of the peptides [24], but this observation can be extended to peptides in general.

Fig. 1 reports the PCA X-loadings plot, showing the correlation between the nine variables, while the individual peptides are identified by numbers, as shown in Table S1 (supplementary information). This figure shows that the first two PCs basically describe the hydrophobicity of the peptides (GRAVY and aliphaticity) and percentage of α-helix, which are negatively correlated to flexibility and Boman index, and also to net charge, pI, total number of residues, and number of disulfide bonds. The second PC basically discriminates between the total number of amino acid residues and net charge, against the other variables (Figs. 1 and 2).

Fig. 2 shows the PCA score plot for the first two PCs and indicates that they explain about 57% of the model variance. When considering the first five PCs, the model explains about 75% of the variance observed in Fig. 2, indicating that these parameters are enough to explain practically all the variance of the model. However, the two first PCs better characterize the relationship between the physicochemical/biophysical properties and the groupings observed in Fig. 2.

The third PC (correlated with number of disulfide bonds) does not add any new information in relation to the two first PCs. However, the fourth PC discriminates the groups as a function of GRAVY and percentage of alpha helix (data not shown).

To better understand the correlation between variables and objects described in Figs. 1 and 2, the same data were also shown in Figs. 3 and 4, emphasizing the three dimensional

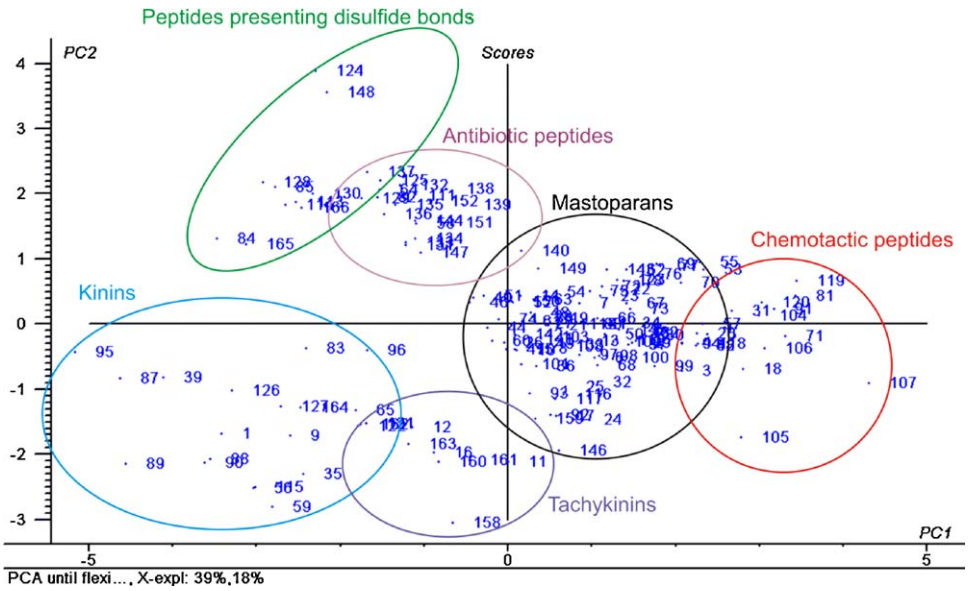

Fig. 2. PCA score plot showing the grouping of the peptides according to their physico-chemical properties. The two first components explain 57% of the *X* variance.

representations of the correlations between the samples and the variables: aliphaticity (Fig. 3A), GRAVY (Fig. 3B), net charge (Fig. 3C), alpha helix (%) (Fig. 4A), and Boman index (Fig. 4B).

Fig. 5 shows the residual variance of the model used in the present study; it shows a step-like representation of the calibration variance and the validation variance for different numbers of PCs. There is a tendency for these values to decrease as a function of the increase in the number of PCs, indicating that the present model is valid, because a higher number of PCs gives a smaller error in the model. In fact, the calibration variance and the validation variance tend to zero after a few PCs.

The purpose of multivariate calibration is to construct a predictive model based on multiple predictor variables. Multivariate calibration is in fact a two-stage procedure: (i) the model is build using training samples, for which the predictor and predictand variables are known or measured, and (ii) the model is then validated by comparing the predictions against reference values for samples that were not used for the model building [36].

To validate the model used to predict the activities of Hymenoptera venom peptides, another series of 80 peptides from other organisms (Table S2 in supplementary information) presenting the same types of activities as those presented by the Hymenoptera peptides were analyzed and compared against the Hymenoptera model. After the calculation of predictor and predictand variables for these peptides, their distribution in the PCA score plot (Fig. 6) and PCA X-loadings plot (Fig. 7) gave a very similar pattern as that observed for the Hymenoptera peptides (Fig. 2). In both cases, the grouping pattern was the same; i.e., those peptides described in the literature as mast cell degranulators were distributed within the same coordinates already occupied by the mastoparans, while a similar distribution was also observed for the other groups (chemotactic peptides, kinins, tachykinins, linear antibiotic peptides and the group of peptides presenting disulfide bridges).

Among the peptides used for validation of the model that predicts the biological activities of Hymenoptera peptides, the group of mast cell degranulators included, Temporin L (n° 167), Aurein 1.1 (n° 171), Aurein 1.2 (n° 172), Alloferon 1 (n° 173), Phyloxin (n° 196), Pyrrhocoricin (n° 197), Metchnikowin (n° 198), Laminin alpha peptide α5f (n° 212), Laminin alpha peptide α5,2 (n° 213), Laminin alpha peptide α5b-sc (n° 214), Vasoactive Intestinal Peptide (n° 216), Bombesin (n° 218), Canine BLI-II (n° 220), Granuliberin-R (n° 221), and Kassinakinin S (n° 222). In addition to these peptides, others that also had chemotactic activity were positioned in the group of chemotactic peptides, including, β-casochemotide-1 (n° 205), Laminin beta peptide β1 (n° 207), Elastin derived peptide (n° 208), Bombesin like peptide (BLI) (n° 219), and Pev Kinin-2 (n° 224).

The tachykinins group included peptides such as LomTK I, LomTK II, LomTK III, LomTK IV (n° 231–234), CusTK III (n° 236), Substance P (n° 215), Vasoactive intestinal peptide (n° 216), NRP11 (n° 228), Peptide P7 (n° 229), Pev-tachykinin (n° 230), CusTK II (n° 235), UruTK II (n° 238), Pev Kinin-1 (n° 223), Laminin alpha peptide α3 (n° 210), Laminin alpha peptide α5,2 (n° 213), and Laminin alpha peptide α5b-sc (n° 214). The kinin group included Laminin α peptide α1 (n° 209), Laminin α 5-1 (n° 211), NRP 11 (n° 228), and a series of non-named peptides – RPPGFSPFR (n° 239), RPKPQQFFGLM (n° 240), PPGFSPFRR (n° 241), GPPDPNKFYPVM (n° 242), MKRPPGFSPFRSSRIG (n° 243), MKRSRGPSPRR (n° 244), RAPVPPGFTPFR (n° 245), and DLPKINRKGPRPPGFSPFR (n° 246).

The group of antimicrobial peptides included, Dermaseptin B2 (n° 168), Apidaecin IA (n° 169), Apidaecin IB (n° 170), Lactoferricin B (n° 174), Cecropin A (n° 175), Bombinin (n° 176), Bombinin-like peptide 1 (n° 177), Maximin 1 (n° 178), Brevinin 1 (n° 179), Esculentin 2A (n° 180), Gaegurin-1 (n° 181), Brevinin 1EMa (n° 182), Rugosin A (n° 183), Ranatuerin 1 T (n° 186), Ranatuerin 1 (n° 187), Ranatuerin 2P (n° 189), Cecropin (n° 189), Cecropin B (n° 190), Cryptidin-1 (n° 191), Androctonin (n° 192), Dermaseptin-S1 (n° 193), Dermaseptin S3 (n° 194), Drosocin (n° 199), Gomesin (n° 200), Protegrin 2 (n° 201), Protegrin 3 (n° 202), Caerin 1.8 (n° 203), and Apidaecin II (n° 205). These antibiotic peptides have higher values of alpha helix percentage than Hymenoptera venom antibiotic peptides, inducing the model to have some rotation in relation to the model of Hymenoptera venom peptides, but not changing the whole distribution of the peptides, which remained exactly the same. Meanwhile, the group of peptides presenting disulfide bridges was composed of Esculentin 2A (n° 180), Rugosin A (n° 183), Thanatin (n° 185), Cryptidin-1 (n° 191), Androctonin (n° 192), Ranatuerin 2P (n° 188), Gomesin (n° 200), Protegrin 2 (n° 201), and Protegrin 3 (n° 202).

It is important to emphasize that this analysis also revealed some overlap among some groups of peptides, such as that between the chemotactic peptides and mast cell degranulator groups, as observed for the peptides Bombesin like peptide I (n° 218), Canine BLI-II (n° 220), and Fibroblast Kinin I and II (n° 226 and 227,

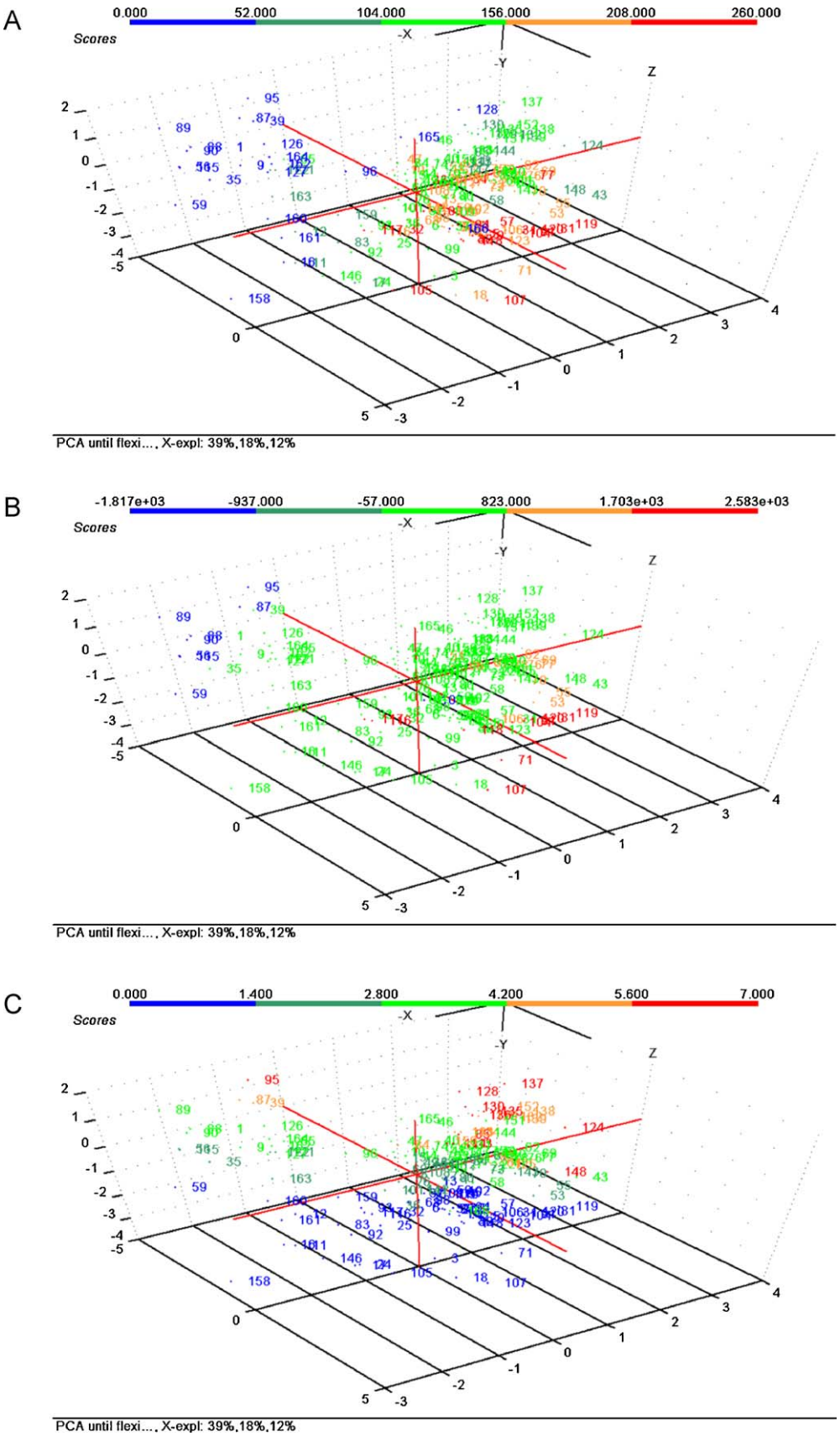

Fig. 3. Three-dimensional correlation between the peptides and the grouping of samples by (A) aliphaticity, (B) GRAVY and (C) net charge.

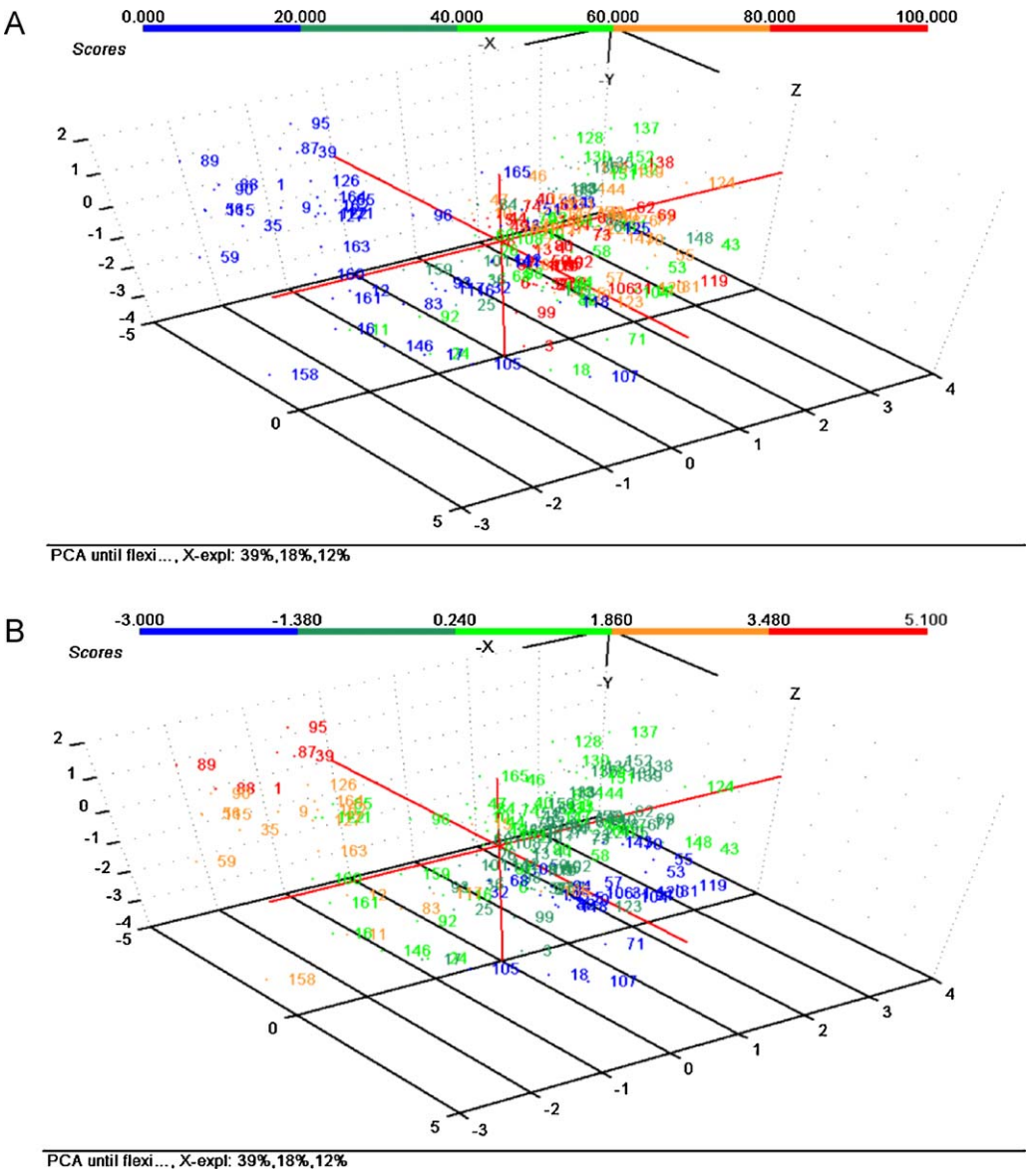

Fig. 4. Three-dimensional correlation between the peptides and the grouping of samples by (A) alpha helix (%) and (B) Boman index (kcal/mol).

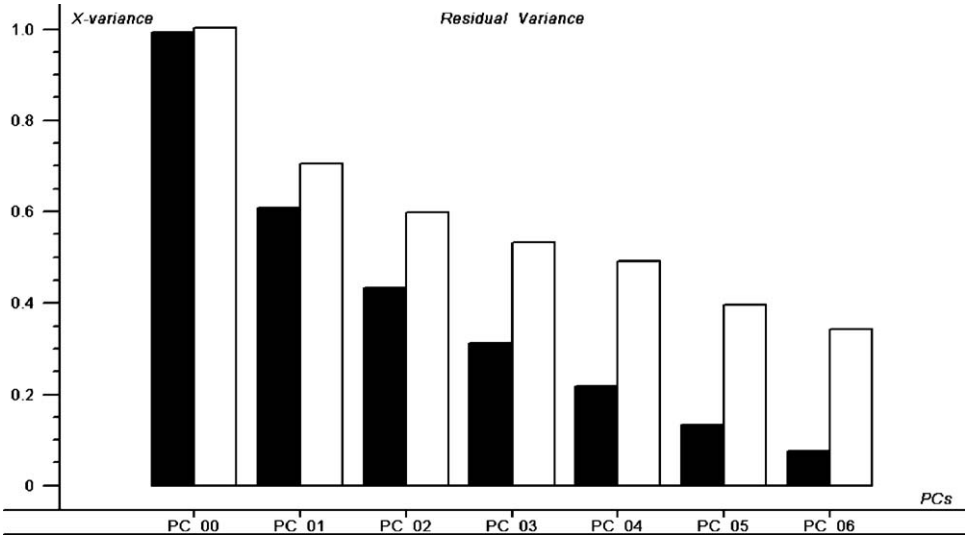

Fig. 5. Residual variance of the model used in the present study in a step-like representation of the calibration variance (black) and the validation variance (white) for different numbers of PCs.

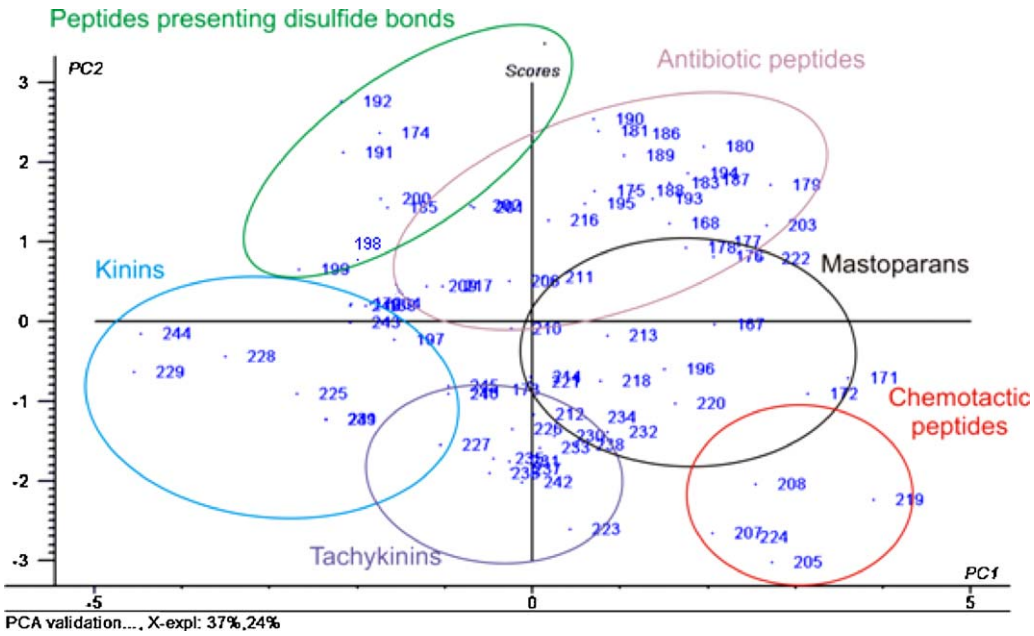

Fig. 6. PCA score plot showing the grouping of the peptides of the validation group according to their physicochemical properties. The first two components explain 61% of the *X* variance.

respectively), and LomTKII (n° 232). Another overlap observed was in the intersection of the kinins and tachykinins groups, for the peptides Catestatin (n° 217), Laminin alpha peptide α5 β1γ1 (n° 206), Laminin alpha peptide α1 (n° 209), Laminin alpha peptide α5-1 (n° 211), and a non-named kinin, DLPKINRKGPRPPGFSPFR (n° 246).

4. Discussion

The model developed to predict the biological activities of the Hymenoptera peptides was validated both through the determination of the residual variance for different numbers of PCs (Fig. 5) and with a sample of 80 peptides not belonging to the Hymenoptera model, which resulted in the same grouping pattern (Fig. 6).

The representation of the score plot for the Hymenopteran model (Fig. 2) shows six groupings, which will be discussed in terms of the function of their potential biological activities. The group of chemotactic peptides can be seen in the right corner of this figure, presenting the highest GRAVY and aliphaticity index values and the lowest pI values, as well flexibility and Boman indexes, tending to be neutral in relation to the net charge (Fig. 2). This indicates the importance of the hydrophobicity of these peptides for the chemotaxis of polymorphonucleated leukocytes (PMNLs). This activity generally requires binding of the peptides to a G-protein coupled receptor (GPCR), initiating a cascade of actions that result in chemoattraction of the target cells toward the source of the stimulus (presence of the peptide) [38]. Interestingly, it is well known that the peptide ligands of GPCRs-related chemotaxis are short, linear and relatively hydrophobic, assuming their final conformations during interaction with the receptors, and tending to present α-helical conformations (Fig. 4A) [38]. This profile fits well with the position of the group of chemotactic peptides

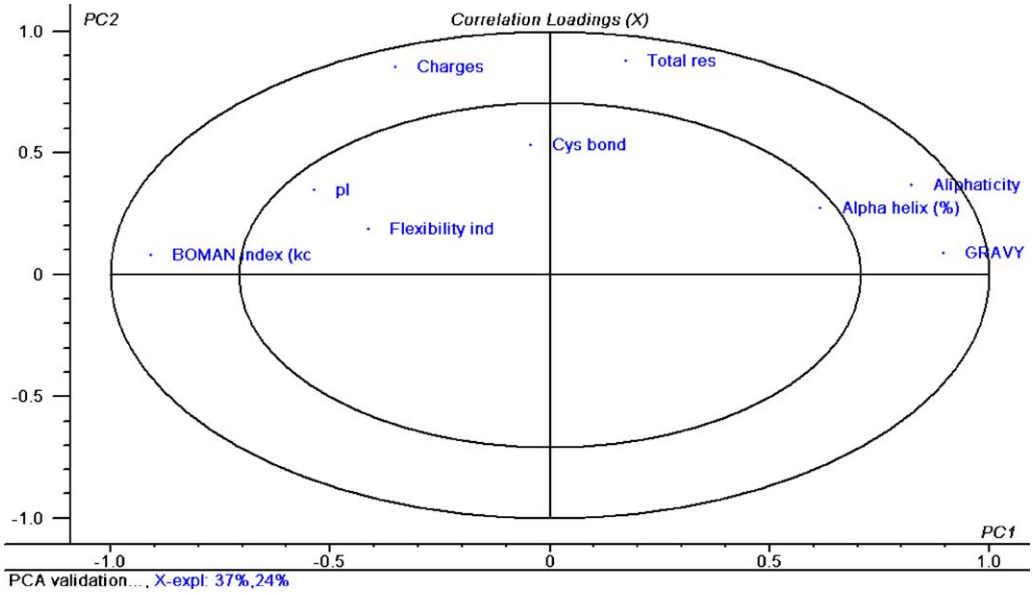

Fig. 7. PCA X-loadings plot showing the correlation between the variables for the validation model.

observed in Fig. 2; thus, the sequence of peptide 71 (Icaria-CP) could be used as reference for this activity. A typical profile of physicochemical parameters for peptides presenting chemotactic activity for PMNLs is high GRAVY and aliphaticity index values (Fig. 3A and B, respectively) and reduced net charges (Fig. 3C).

Intersecting partially with the group of chemotactic peptides is the group of mastoparan peptide, as shown in the score plot (Fig. 2). Mastoparans are described as amphipathic peptides that interact directly with specific GPCRs related to mast cell degranulation [26,33]. Peptides such as Polybia MP-III (n° 99), mastoparan-1 (n° 28) and crabrolin (n° 57) are positioned in the mentioned intersection, suggesting that these peptides also may present some chemotactic activity. These molecules are amphiphilic, presenting α-helix conformations under hydrophobic conditions, like the mastoparans [1,10,13–16]. This observation is not surprising because some mastoparans, such as Polybia-MP-II and -III (n° 98 and 99) have been reported to present chemotactic activity for polymorphonucleated leukocytes (PMNLs) [14]. Bombolitin-III (n° 53) is reported to be an amphipathic peptide, presenting similar functions of mastoparans, since they also interact with cell membranes, causing some mast cell degranulation [1,45]. The reciprocal situation also occurs, in which some chemotactic peptides also present a reduced mast cell degranulation, as previously reported for Protonectin (1–6) (n° 107) [3].

Some mastoparans also present antimicrobial action against Gram-positive and Gram-negative bacteria [11,44], which may explain a partial overlapping of this group with the antimicrobial peptides (Fig. 2). The mastoparan group is the most diversified one in the score plot (Fig. 2), and some of these peptides can be spotted close to virtually all of the other groups.

Some peptides from ant venom, such as the ponericins-G6, -G7, and -W6 (n° 141–143), one poneratoxin (n° 123), and two dinoponeratoxins (n° 140 and 145), were previously reported to be antimicrobial peptides [41]; however, according their position in the score plot (Fig. 2), they were grouped as mastoparans in this study. Considering that some mastoparan-like peptides may also interact with the bacterial membrane, causing disruption of the membrane both in Gram-positive and Gram-negative bacteria because of their amphipathicity [12], it is possible that the ponericins, poneratoxin and dinoponeratoxins and osmin (n° 149) would also present antimicrobial activity.

In the lower left corner of the score plot (Fig. 2), it is possible to identify the group of wasp kinins; these peptides are structurally related to bradykinins and cause local vasodilation, smooth muscle contraction, and hypotensive action, in addition to relaxing the duodenum of rats [4,39,47]. Other poorly characterized peptides from ant venoms are also positioned within this group, such as Formaecin-1 and -2 (n° 126 and 127). This observation indicates that these peptides should also be assayed for typical kinin activities; these peptides have high pI values and Boman indexes, high flexibility, reduced aliphaticity and GRAVY values (Fig. 3A and B).

In the lower left corner of the score plot (Fig. 2), the group corresponding to the tachykinins also can be seen; this group is part of a large family of neuropeptides commonly found in amphibians and mammals [27], in addition to the venoms of some species of social wasps [58]. These peptides were so named because of their ability to rapidly induce the contraction of gut tissue; they also excite neurons, evoke behavioral responses, are potent vasodilators and contract (directly or indirectly) many smooth muscles [22,35]. The tachykinins present intermediate values of GRAVY and aliphaticity (Fig. 3A and B), in addition to reduced net charges (Fig. 3C). This group also have intermediate percentages of α-helix and Boman indexes (Fig. 4A and B).

In the upper left corner of the score plot (Fig. 2), two groups can be found: the antibiotic peptides and the peptides with disulfide bonds in their structures. The group of antibiotic peptides is characterized by linear molecules, following the distribution of intermediary values of aliphaticity (Fig. 3A) and GRAVY (Fig. 3B). Apparently, the actions of these peptides in bacterial systems occur by direct interaction with the microbial membranes, which in turn seems to be dependent on the amphipathicity of the peptides [16]. The intermediate values of GRAVY and aliphaticity, associated with the relatively high values of the net charge of these peptides, seem to favor the necessary amphipathicity for direct interaction with the bacterial membranes. Despite not being characterized as having antimicrobial actions, some large linear peptides like mellitin (n° 152) are located in this group, indicating that they may potentially present antimicrobial activity. This group includes some peptides that have not been well characterized up to now, such as Abaecin (n° 165), which is not a venom toxin, but a polycationic and linear peptide from honeybee hemolymph, presenting high antimicrobial activity [7]; the peptides Ponericins and Dinoponeratoxins (n° 123–147), are ant venom components, characterized by large number of amino acid residues in their linear chain, also presenting antimicrobial activity [25].

In the upper left corner of the score plot (Fig. 2), is located a group of wasp and bee venom peptides presenting long backbone chains, rich in positive charges and with one or two disulfide bonds. Certainly, the presence of disulfide bonds plays a strong role in the formation of this group. These peptides are poorly characterized regarding their functionality. Peptides such as Paulistine (n° 111), Seduline (n° 113) and Sylverin (n° 114) are reported as inflammatory components, which apparently do not present antimicrobial activity [12,15,42]. Apamin (n° 166) is described as a neurotoxin, acting by blocking the slow conductance of Ca^{2+}-dependent K^+ channels in the central nervous system of mammals, specifically at low concentrations [50,51]. Secapine (n° 168) is a neurotoxic agent causing piloerection, smooth sedation, and hypothermia [2]. The MCD peptide (n° 167) and Tertiapine (n° 148) have two disulfide bonds; the first is reported to cause mast cell degranulation, while the second is a potent blocker of voltage-sensitive K^+ channels [4,28]. Furthermore, it has been suggested that bee venom peptides share the same folding pattern, which is centered around a β-turn covalently bound to the α-helix segment by a disulfide bond, suggesting that Apamine, Tertiapine, and MCD form a unique molecular class [23]. The formation of a disulfide bond is considered to be a post-translational modification that stabilizes the tertiary structure of some peptides and is necessary to make them active, so they can adequately perform their biological function, thus making disulfide bonds important structural components of many peptide neurotoxins of animal venoms [43]. It is possible, therefore, that this group of peptides may be functionally related to neurotoxicity during the envenoming process; however, determining their precise role will require much more experimentation. These peptides also have high values of Boman free energy index, indicating possible interaction with proteins (receptors) (Fig. 4B).

5. Conclusions

In general, it is difficult to determine the specific biological activity of novel peptides based solely on their amino acid sequences, especially in the venoms from the social Hymenoptera, in which novel peptides are being described frequently, with a complex panel of biological activities, characteristically of polyfunctional nature. Using the “trial and error” approach may be laborious, expensive and time consuming due to the potentially enormous number of different experimental setups of pharmacological/physiological assays that are required to minimally cover a reasonable number of biological assays. However, nature offers some interesting systems of biologically active peptides that are structurally and functionally well characterized,

and reliably documented in the literature, which can be used as "models" to investigate the relationship between a series of intrinsic physicochemical parameters of these peptides and their biological activities. Thus, we chose 166 peptides from the venoms and hemolymphs of Hymenoptera insects as a virtual library (biological model) and applied a mathematical model of multivariate analysis with nine different chemometric components (corresponding to the most investigated physicochemical descriptors of peptides): GRAVY, aliphaticity of the side chain of the amino acid residues of each peptide chain, number of disulfide bonds, total residues, net charge, pI value, CDM prediction of alpha helix, flexibility, and Boman index. PCA with Partial Least Squares Regression was performed with these data.

While being constructed, this virtual system was blinded from any information about the biological activity of the peptides; however, this analysis permitted the grouping of peptides in a way strongly correlated to their biological function. Six different groupings were observed, which seemed to correspond to the following classes: chemotactic peptides, mastoparans, tachykinins, kinins, antibiotic peptides and a group of long peptides with one or two disulfide bonds and biological activities that are not yet clearly defined. The partial overlapping between the groups of mastoparans with chemotactic peptides, tachykinins, kinins and antibiotic peptides in the PCA score plot (Fig. 2) may be used to explain frequent reports in the literature about the multifunctionality of some of these peptides.

The groupings reported above seems to be characterized by well-defined ranges of values for some basic physicochemical parameters that are intrinsic to each peptide sequence, which may be used in future studies as part of a strategy for selecting the most informative sequences in a series of homologs peptides. This could allow for more efficient determination of biological activities such as chemotaxis of PMNLs, mast cell degranulation, antibiosis, and even more potent analogs of kinins. The mathematical model used in the present investigation may also be applied to other biological systems that involve peptide components, and other different and physicochemical parameters may be included in the analysis in addition to, or as a substitute for the more common parameters used here.

Acknowledgements

This research was supported by grants from FAPESP (BIOprospecTA Proc. 04/07942-2, 06/57122-6) and INCT-Imunologia. M.S.P. is a researcher for the Brazilian Council for Scientific and Technological Development (CNPq).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.peptides.2011.08.001.

References

[1] Argiolas A, Pisano JJ. Bombolitins, a new class of mast cell degranulating peptides from the venom of the bumblebee *Megabombus pennsylvanicus*. J Biol Chem 1984;260(3):1437–44.
[2] Banks BC, Shipolini RA. Chemistry and pharmacology of honey-bee venom. In: Piek T, editor. Venoms of the Hymenoptera. Biochemical, pharmacological and behavioural aspects. London: Academic Press; 1986. p. 330–424.
[3] Baptista-Saidemberg NB, Saidemberg DM, de Souza BM, César-Tognoli LM, Ferreira VM, Mendes MA, et al. Toxicon 2010;56(6):880–9.
[4] Bessone R, Martin-Eauclaire MF, Crest M, Mourre C. Heterogeneous competition of Kv1 channel toxins with kaliotoxin for binding in rat brain: autoradiographic analysis. Neurochem Int 2004;45(7):1039–47.
[5] Boman HG. Antibacterial peptides: basic facts and emerging concepts. J Intern Med 2003;254:197–215.
[6] Brigatte P, Cury Y, De Souza BM, Baptista-Saidemberg NB, Saidemberg DM, Gutierrez VP, et al. Hyperalgesic and edematogenic effects of peptides isolated from the venoms of honeybee (*Apis mellifera*) and neotropical social wasps (*Polybia paulista* and *Protonectarina sylveirae*). Amino Acids 2011;40(1):101–11.
[7] Casteels P, Ampe C, Riviere L, Van Damme J, Elicone C, Fleming M, et al. Isolation and characterization of abaecin: a major antibacterial response peptide in the honeybee (*Apis mellifera*). Eur J Biochem 1990;187(2):381–6.
[8] Cheng H, Sen TZ, Kloczkowski A, Margaritis D, Jernigan RL. Prediction of protein secondary structure by mining structural fragment database. Polymer 2005;46:4314–21.
[9] Cohen L, Moran Y, Sharon A, Segal D, Gordon D, Gordon D, et al. Drosomycin, an Innate immunity peptide of *Drosophila melanogaster*, interacts with the fly voltage-gated sodium channel. J Biol Chem 2009;284:23558–63.
[10] De Souza BM, Mendes MA, Santos LD, Marques MR, César LMM, Almeida RNA, et al. Structural and functional characterization of two novel peptide toxins isolated from the venom of the social wasp *Polybia paulista*. Peptides 2005;26:2157–64.
[11] De Souza BM, Marques MR, Tomazela DM, Eberlin MN, Mendes MA, Palma MS. Mass spectrometric characterization of two novel inflammatory peptides from the venom of the social wasp *Polybia paulista*. Rapid Commun Mass Spectrom 2004;18:1095–102.
[12] De Souza BM, Palma MS. Peptides from Hymenoptera venoms: chemistry, biochemistry and potential applications in health and biotechnology. In: De Lima ME, Pimenta AMC, Martin-Euclaire MF, Zingali RB, editors. Animal toxins: the state of art. Belo Horizonte: UFMG Press; 2009. p. 345–69.
[13] De Souza MS, Dos Santos Cabrera MP, Ruggiero Neto J, Palma MS. Investigating the effect of different positioning of lysine residues along the peptide chain of mastoparans for their secondary structures and biological activities. Amino Acids 2011;40(1):77–90.
[14] De Souza BM, Silva AVR, Resende VMF, Arcuri HA, Dos Santos Cabrera MP, Ruggiero-Neto J, et al. Characterization of two novel polyfunctional mastoparan peptides from the venom of the social wasp *Polybia paulista*. Peptides 2009;30:1387–95.
[15] Dohtsu K, Okumura K, Hagiwara K, Palma MS, Nakajima T. Isolation and sequence analysis of peptides from the venom of *Protonectarina sylveirae* (Hymenoptera-Vespidae). Nat Toxins 1993;1(5):271–6.
[16] Dos Santos Cabrera MP, De Souza BM, Ruggiero Neto J, Palma MS. Interactions of mast cell degranulating peptides with model membranes: a comparative biophysical study. Arch Biochem Biophys 2009;486:1–11.
[17] Esbensen KH. Multivariate data analysis – in practice. 5th ed. Oslo: Camo Software AS; 2001.
[18] Escoubas P. Molecular diversification in spider venoms: a web of combinatorial peptide libraries. Mol Divers 2006;10:545–54.
[19] Fry BG, Wuster W, Kini RM, Brusic V, Khan A, Venkataraman AD, et al. Molecular evolution and phylogeny of elapid snake venom three-finger toxins. J Mol Evol 2003;5:110–29.
[20] Fry BG. From genome to "venome": molecular origin and evolution of the snake venom proteome inferred from phylogenetic analysis of toxin sequences and related body proteins. Genome Res 2005;15:403–20.
[21] Fry BG, Roelants K, Champagne DE, Scheib H, Tyndall JDA, King GF, et al. The toxicogenomic multiverse: convergent recruitment of proteins into animal venoms. Annu Rev Genom Hum Genet 2009;10:483–511.
[22] Helke CJ, Krause JE, Mantyh PW, Couture R, Bannon MJ. Diversity in mammalian tachykinin peptidergic neurons: multiple peptides, receptors, and regulatory mechanisms. FASEB J 1990;4(6):1606–15.
[23] Hidera RC, Ragnarsson U. A comparative structural study of apamin and related bee venom peptides. Biochim Biophys Acta – Prot Struct 1981;667(1):197–8.
[24] Ikai AJ. Thermostability and aliphatic index of globular proteins. J Biochem 1980;88:1895–8.
[25] Johnson SR, Copello JA, Evans MS, Suarez AV. A biochemical characterization of the major peptides from the venom of the giant neotropical hunting ant *Dinoponera australis*. Toxicon 2010;55:702–10.
[26] Karplus PA, Schulz GE. Prediction of chain flexibility in proteins – a tool for the selection of peptide antigens. Naturwissenschafren 1985;72:212–3.
[27] Khawaja AM, Rogers DF. Tachykinins: receptor to effector. Int J Biochem Cell Biol 1996;28(7):721–38.
[28] Kitamura H, Yokoyama M, Akita H, Matsushita K, Kurachi Y, Yamada M. Tertiapin potently and selectively blocks muscarinic K^+ channels in rabbit cardiac myocytes. J Pharmacol Exp Ther 2000;293(1):196–205.
[29] Kloczkowski A, Ting KL, Jernigan RL, Garnier J. Combining the GOR V algorithm with evolutionary information for protein secondary structure prediction from amino acid sequence. Proteins 2002;49:154–66.
[30] Kyte J, Doolittle RF. A simple method for displaying the hydropathic character of a protein. J Mol Biol 1982;157(1):105–32.
[31] Lai R, Liu H, Lee WH, Zhang Y. An anionic antimicrobial peptide from toad *Bombina maxima*. Biochem Biophys Res Commun 2002;295:796–9.
[32] Langham AA, Khandelia H, Schuster B, Waring AJ, Lehrer RI, Kaznessisa YN. Correlation between simulated physicochemical properties and hemolycity of protegrin-like antimicrobial peptides: predicting experimental toxicity. Peptides 2008;29(7):1085–93.
[33] Lapinsh M, Gutcaits A, Prusis P, Post C, Lundstedt T, Wikberg JES. Classification of G-protein coupled receptors by alignment-independent extraction of principal chemical properties of primary amino acid sequences. Prot Sci 2002;11:795–805.
[34] Leite NB, da Costa LC, Alvares DS, dos Santos Cabrera MP, De Souza BM, Palma MS, et al. The effect of acidic residues and amphipathicity on the lytic activities

of mastoparan peptides studied by fluorescence and CD spectroscopy. Amino Acids 2011;40(1):91–100.
[35] Maggio JE. Tachykinins. Annu Rev Neurosci 1988;11:13–28.
[36] Martens H, Næs T. Multivariate calibration. John Wiley & Sons Ltd.; 1991. p. 438.
[37] Martynova NB, Filimonov DA, Poroikov VV. Computer prediction of biological activity spectra for low-molecular peptides and peptidomimetics. Russ J Bioorg Chem 2000;26(5):297–305.
[38] Mendes MA, Souza BM, Palma MS. Structural and biological characterization of three novel mastoparan peptides from the venom of the neotropical social wasp *Protopalybia exigua* (Saussure). Toxicon 2005;45:101–6.
[39] Mendes MA, Palma MS. Two new bradykinin-related peptides from the venom of the social wasp *Protopolybia exigua* (Saussure). Peptides 2006;27:2632–9.
[40] Muta T, Fujimoto T, Nakajima H, Iwanaga S. Tachyplesins isolated from hemocytes of Southeast Asian horseshoe crabs (*Carcinoscorpius rotundicauda* and *Tachypleus gigas*). Identification of a new tachyplesin, tachyplesin III, and a processing intermediate of its precursor. J Biochem 1990;108:261–6.
[41] Orivel J, Redeker V, Le Caer JP, Krier F, Revol-Junelles AM, Longeon A, et al. Ponericins: new antibacterial and insecticidal peptides from the:ant *Pachycondyla goeldii*. J Biol Chem 2001;276(21):17823–9.
[42] Palma MS. Insect venom peptides. In: Kastin A, editor. Handbook of biologically active peptides. San Diego: Academic Press; 2006. p. 409–17.
[43] Palma MS. Peptides as toxins/defensins. Amino Acids 2011;40:1–4.
[44] Park NG, Yamato Y, Lee S, Sugihara G. Interaction of mastoparan-B from venom of a hornet in Tawian with phospholipid bilayers and its antimicrobial activity. Biopolymers 1995;36:793–801.
[45] Peggion E, Mammi S, Schievano E. Conformation and interactions of bioactive peptides from insect venoms: the bombolitins. Biopolymers 1997;43(6):419–31.
[46] Pérez-Faginas P, Aranda MT, García-López MT, Snoeck R, Andrei G, Balzarini B, et al. Synthesis and SAR studies on azetidine-containing dipeptides as HCMV inhibitors. Bioorg Med Chem 2011;2011(19):1155–61.
[47] Pisano JJ. Kinins of non-mammalian origin. In: Erdos EG, editor. Handbook of experimental pharmacology XXV: bradykinin, kallidin and kallikrein. Berlin: Springer Verlag; 1970. p. 659–64.
[48] Qua H, Magottia P, Ricklina D, Wub EL, Kourtzelisa J, Wua YQ, et al. Novel analogues of the therapeutic complement inhibitor compstatin with significantly improved affinity and potency. Mol Immunol 2011;48: 481–9.
[49] Ray S, Kepler TB. Amino acid biophysical properties in the statistical prediction of petide-MHC class I binding. Immunome Res 2007;3:9.
[50] Romey G, Hugues M, Schmid-Antomarchi H, Lazdunski M. Apamin: a specific toxin to study a class of Ca2+ -dependent K+ channels. J Physiol 1984;79(4):259–64.
[51] Schneider JJ, Unholzer A, Schaller M, Schäfer-Korting M, Korting HC. Human defensins. J Mol Med 2005;83:587–95.
[52] Sollod B, Wilson D, Zhaxybayeva O, Gogarten JP, Drinkwater R, King GF. Were arachnids the first to use combinatorial peptide libraries? Peptides 2005;26:131–9.
[53] Uccelletti D, Zanni E, Marcellini L, Palleschi P, Barra D, Mangoni ML. Anti-*Pseudomonas* activity of frog skin antimicrobial peptides in a *Caenorhabditis elegans* infection model: a plausible mode of action *in vitro* and *in vivo*. Antimicrob Agents Chemother 2010;54(9):3853–60.
[54] Sen TZ, Jernigan RL, Garnier J, Kloczkowski A. GOR V server for protein secondary structure prediction. Bioinformatics 2005;21(11):2787–8.
[55] Sen TZ, Cheng H, Kloczkowski A, Jernigan RL. A consensus data mining secondary structure prediction by combining GOR V and fragment database mining. Protein Sci 2006;15(11):2499–506.
[56] Vetter I, Davis JL, Rash LD, Anangi R, Mobli M, Alewood PF, et al. Venomics: a new paradigm for natural products-based drug discovery. Amino Acids 2011;40:15–28.
[57] Wang X-Y, Wang J, Hu Y, Lin Y, Shu M, Wang L, et al. Predicting the activity of peptides based on amino acid information. J Chin Chem Soc 2011;58(5):1–7.
[58] Yshii LM, Souza GHMF, Camargo EA, Eberlin MN, Ribela MTCP, Muscara MN, et al. Characterization of the mechanisms underlying the inflammatory response to *Polistes lanio lanio* (paper wasp) venom in mouse dorsal skin. Toxicon 2009;53(1):42–52.
[59] Zhihua L, Yuzhang W, Xuejun Q, Yuegang Z, Bing N, Ying W. Use of a novel electrotopological descriptor for the prediction of biological activity of peptide analogues. Lett Pept Sci 2002;9:273–81.