Peptides 38 (2012) 127–136

Peptides

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

In silico identification of novel hevein-like peptide precursors

William F. Porto[a], Valéria A. Souza[a], Diego O. Nolasco[a,b], Octávio L. Franco[a,*]

[a] *Centro de Análises Proteômicas e Bioquímicas, Programa de Pós-Graduação em Ciências Genômicas e Biotecnologia, Universidade Católica de Brasília, Brasília – DF, Brazil*
[b] *Curso de Física, UCB, Brasília, Brazil*

ARTICLE INFO

Article history:
Received 25 April 2012
Received in revised form 27 July 2012
Accepted 27 July 2012
Available online 4 September 2012

Keywords:
Hevein domain
Regular expression
Data mining
Molecular modeling
Molecular dynamics simulation
Fungal hevein-like peptide

ABSTRACT

Lectins are proteins with ability to bind reversibly and non-enzymatically to a specific carbohydrate. They are involved in numerous biological processes and show enormous biotechnological potential. Among plant lectins, the hevein domain is extremely common, being observed in several kinds of lectins. Moreover, this domain is also observed in an important class of antimicrobial peptides named hevein-like peptides. Due to higher cysteine residues conservation, hevein-like peptides could be mined among the sequence databases. By using the pattern CX(4,5)CC[GS]X(2)GXCGX[GST]X(2,3)[FWY]C[GS]X[AGS] novel hevein-like peptide precursors were found from three different plants: *Oryza sativa*, *Vitis vinifera* and *Selaginella moellendorffii*. In addition, an hevein-like peptide precursor from the phytopathogenic fungus *Phaeosphaeria nodorum* was also identified. The molecular models indicate that they have the same scaffold as others, composed of an antiparallel β-sheet and short helices. Nonetheless, the fungal hevein-like peptide probably has a different disulfide bond pattern. Despite this difference, the complexes between peptide and N,N,N-triacetylglucosamine are stable, according to molecular dynamics simulations. This is the first report of an hevein-like peptide from an organism outside the plant kingdom. The exact role of an hevein-like peptide in the fungal biology must be clarified, while in plants they are clearly involved in plant defense. In summary, data here reported clear shows that an *in silico* strategy could lead to the identification of novel hevein-like peptides that could be used as biotechnological tools in the fields of health and agribusiness.

1. Introduction

Lectins are peptides or proteins that have at least one domain with ability to bind reversibly and non-enzymatically to a specific carbohydrate, which could be mono- or oligosaccharide [12,45]. Therefore, lectin protein families can be found in plants [6], animals [53], fungi [25] and bacteria [32], being observed at subcellular levels in membranes and cell secretions [12].

Lectins can be divided into three subtypes: merolectins, hololectins and chimerolectins. This division is related to the number of carbohydrate binding-domains. Merolectins are composed of a single carbohydrate binding domain and are unable to precipitate glycoconjugates or cause cell agglutination [45]. Hololectins are exclusively composed of carbohydrate binding domains, containing two or more domains and being able to agglutinate cells and/or precipitate glycoconjugates [45]. Finally, chimerolectins are characterized by one or more carbohydrate binding domains and an additional domain responsible for another biological activity (*e.g.* chitinase activity) [45].

Due to lectins' functional plasticity, they are involved in numerous biological processes including defense against pathogens, symbiosis and cell signaling [12]. Among pathogen defense functions, lectins can perform bactericidal [40], fungicidal [40,64] and antiviral activities [39]. Indeed, lectins have an enormous potential for developing novel drugs, pesticides and/or transgenic organisms, since they can bind specifically to carbohydrates normally absent in vertebrates and plants, such as chitin or the bacterial cell wall peptidoglycan carbohydrate. Furthermore, by chitin targeting, several pests can be tackled, since chitin is the main component of the fungal cell wall and also of the exoskeleton of invertebrates, such as nematodes and insects [20,43].

Among the lectins, the hevein domain is extremely common, being found in chimerolectins, hololectins and merolectins (Fig. 1) [7]. The name 'hevein' was proposed by Archer in 1960 [4], when the first peptide with this domain was isolated from the latex of the rubber tree (*Hevea brasiliensis*). This domain is rarely found in proteins that do not belong to the plant kingdom. As an exception, a protein containing an hevein domain from the phytopathogenic

Abbreviations: NR, NCBI's non-redundant protein database; SVM, support vector machine; CAMP, Collection of Antimicrobial Peptides; DOPE, discrete optimized protein energy; $(GlcNAc)_3$, N,N,N-triacetylglucosamine; HEV32, hevein-32; MD, molecular dynamics simulation; RMSD, root-mean-square deviation; RMSF, root-mean-square fluctuation.

* Corresponding author. Tel.: +55 61 34487167/+55 61 34487220; fax: +55 61 33474797.
E-mail address: ocfranco@gmail.com (O.L. Franco).
URL: http://www.capb.com.br (O.L. Franco).

0196-9781/
http://dx.doi.org/10.1016/j.peptides.2012.07.025

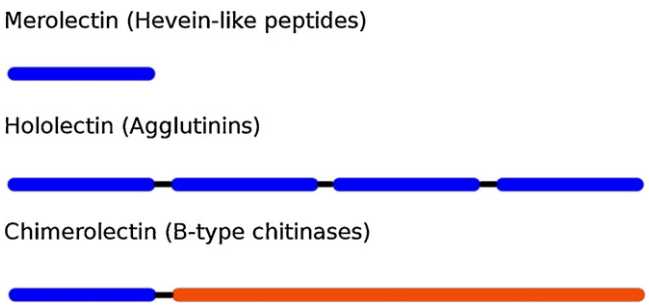

Fig. 1. General architecture of lectins containing the hevein domain [7]. The blue boxes indicate the hevein-domain and the orange box a chitinase domain. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

fungus *Magnaporthe grisea* was identified by Kamakura et al. [29] (GenBank ID: BAB79692.1). The overall hevein domain structure is composed of an anti-parallel β-sheet and occasional short helices; the scaffold is stabilized by three to five disulfide bonds [40,64]. This structural framework exposes four amino acid residues (one serine and three aromatic) involved in chitin-binding and related oligomers [40,64]. The amino acid residues involved are arranged as follows, X_i-X_{i+2}-X_{i+4}-X_{i+11}, where the residue X_i is normally a serine residue and X_{i+11}, a tyrosine residue; X_{i+2} and X_{i+4} are normally an aromatic residue. The residues in X_{i+2} and X_{i+4} stabilize the complex through CH-π stacking. If the residues in X_{i+2} and X_{i+4} are tyrosine or tryptophan, they also contribute through hydrogen bonds [10].

Although the mechanism of action of hevein-like peptides has not been completely elucidated, it is known that hevein-like peptides are able to inhibit the development of chitin-containing fungi [37,40]. This fungicidal activity has been related to its chitin-binding domain, where the cell wall elongation is retarded or stopped after the chitin-binding step. In addition, some peptides are also able to inhibit chitin-free fungal and/or bacterial development, as these activities are related to their cationic and amphipathic surface, which could interact with negatively charged membranes of target pathogens [40,64]. In order to improve plant resistance to phytopathogenic fungi, hevein-like peptides have been expressed in tobacco [33,52], tomato [31] and *Arabidopsis* plants [51,52].

These peptides can therefore be included in the selective class of promiscuous peptides, where a peptide or a peptide family can have multiple activities under different environmental conditions [16]. In the case of family promiscuity, the multiple functions are related to different exposed residues in the same scaffold, which in turn are stabilized by their disulfide bonds [16]. Due to the conservation of disulfide bonds, these classes are good targets for mining protein databases. This kind of approach has been applied to cyclotides [42] and defensins [65] and has revealed novel aspects about them. Identification of novel hevein-like peptides may bring to light new possibilities for their use as well as knowledge about their functions. To this end, this work reports the identification of novel hevein-like peptide precursors through computational methods. Sequences from plants and also from a phytopathogenic fungus were identified and their structures and possible functions were predicted. The results presented here may also suggest new prospects for hevein domain interactions that are applicable to chitin studies.

2. Materials and methods

2.1. Data set construction

The data set of hevein-like peptides was constructed by using an automatic search system. Briefly, the system here proposed runs the Blast software [2], reads its output, gets the retrieved sequences and subsequently runs Blast once more with these retrieved sequences. This process was repeated until no novel sequences were obtained, as described by Zhu [65] with minor modifications. Additionally, the system was set to filter fragments and sequences larger than 130 amino acid residues. The initial sequence used for searching was the Ac-AMP2's precursor, identified from *Amaranthus caudatus* [9] (UniProt ID: P27275), since it has antimicrobial and antifungal activities. The search was performed in SwissProt database [56]. The final data set was manually curated, selecting only the sequences annotated as fungicidal.

2.2. Pattern definition and regular expression search

The software Pratt 2.1 [27] was used for pattern identification into the hevein-like data set, using the default parameters (number of consecutive wild cards, maximum number of flexible spacers and maximum number of consecutive wild cards set to five, two and two, respectively). The pattern with the highest fitness value was used for searching against NCBI's non-redundant protein database (NR), through regular expressions and PERL scripts. The script was set to select sequences annotated as hypothetical, unnamed and/or unknown proteins, restricting the maximum size to 130 amino acid residues. However, no sequences were obtained. Therefore wild cards were introduced into the pattern, mixing it with the chitin-binding motif from Prosite (Prosite ID: PS00026), generating a more generalized arrangement, and the search through regular expression was done again.

2.3. Sequence analysis

The sequences found by regular expression search were further submitted to Phobius [28] and SignalP 4.0 [44] for identification of signal peptides. Subsequently the signals were removed and the mature sequences were submitted to InterProScan [47] for domain identification, the largest domain signature was chosen as the actual domain. The antimicrobial activity was predicted by a support vector machine (SVM) specific to cysteine stabilized peptides [46] and also by Collection of Antimicrobial Peptides (CAMP) algorithms [57]. In addition, a multiple alignment was constructed by ClustalW [58], for verifying the similarities among the sequences.

2.4. Molecular modeling

The LOMETS server [63] was used to find the best template for comparative modeling. In addition to the template indicated by LOMETS, the hevein-32 structure (HEV32, PDB ID: 1T0W) [1] was also used as a template, since it was solved in complex to N,N,N-triacetylglucosamine ($(GlcNAc)_3$). The inclusion of this additional structure allows to identify the binding position of $(GlcNAc)_3$ without docking experiments. Therefore, two thousand theoretical three-dimensional models were constructed through Modeller 9.10 [14]. The $(GlcNAc)_3$'s atoms were imported by setting as true the property io.hetatm from the class environ from Modeller 9.10. The final model was selected according to the discrete optimized protein energy (DOPE) scores. This score assesses the energy of the model and indicates the best probable structures. If necessary, an additional energy minimization with two thousand cycles of steepest descent using the GROMOS96 implementation of Swiss-PdbViewer [17] was performed. The model with the best DOPE score was evaluated through PROSA II [61] and PROCHECK [35]. PROCHECK checks the stereochemical quality of a protein structure, through the Ramachandran plot, where good quality models are expected to have more than 90% of amino acid residues in most favored and additional allowed regions. PROCHECK also gives the G-factor, a measurement of how unusual the model is, where values below −0.5 are unusual, while PROSA II indicates the fold quality.

The electrostatic surface was calculated through APBS [5]. Surface potentials were set to $\pm 5\ kT\ e^{-1}$ (133.56 mV). Structure and surface visualization were done in PyMOL (The PyMOL Molecular Graphics System, Version 1.4.1, Schrödinger, LLC). Additionally, structural alignments were performed for verifying the structure similarities among the identified sequences through Dali Lite [18] and for verifying the similarities to structures deposited on PDB through Dali Server [23]. The assessment of structural alignments was done through *Z*-Score. A structural alignment with *Z*-Score higher than 2 may be considered significant.

2.5. Molecular dynamics

The molecular dynamics simulations (MD) of the peptide-$(GlcNAc)_3$ complexes were carried out in water environment, using the Single Point Charge water model [8]. The analyses were performed by using the computational package GROMACS 4 [22]. The dynamics utilized the tridimensional models of the peptide-$(GlcNAc)_3$ complexes as initial structures, immersed in water molecules in cubic boxes with a minimum distance of 0.7 nm between the complexes and the boxes frontiers. Chlorine ions were also inserted at the complexes with positive charges in order to neutralize the system charge. Geometry of water molecules was constrained by using the SETTLE algorithm [41]. All atom bond lengths were linked by using the LINCS algorithm [21]. Electrostatic corrections were made by Particle Mesh Ewald algorithm [11], with a cut off radius of 1.4 nm in order to minimize the computational time. The same cut off radius was also used for van der Waals interactions. The list of neighbors of each atom was updated every 10 simulation steps of 2 fs. The conjugate gradient and the steepest descent algorithms – 2 ns each – were implemented for energy minimization. After that, the system underwent into a normalization of pressure and temperature, using the integrator stochastic dynamics – 2 ns each. The systems with minimized energy, balanced temperature and pressure were carried out using a step of position restraint, using the integrator molecular dynamics – 2 ns. The simulations were carried out at 300 K *in silico*. The total time for each ensemble simulation was 50 ns. The MD simulations were analyzed by means of root-mean-square deviation (RMSD), root-mean-square fluctuation (RMSF) and number of hydrogen bonds that kept the complex stable along the simulation.

3. Results

Initially, by using the automatic search system, thirteen sequences were retrieved from SwissProt database. Due to the presence of hevein domains in other lectins which are not hevein-like peptides, the automatic search system was set to avoid sequences longer than 130 amino acid residues, ensuring the selection of hevein-like peptides. However, from the thirteen sequences, ten sequences showed the hevein domain. The other three sequences were removed from further analysis. Among the sequences containing the hevein domain, nine showed similarities to merolectins and only one was similar to hololectin. Among the merolectins, eight sequences were annotated as fungicidal peptides. These data are summarized in Table 1.

The eight fungicidal sequences were used for pattern recognition. The best generated pattern was C[GNP][ANS]X[LM] CC[GS]X[FWY]G[FWY]CGX[GST][ADNP]XYC[GS]X[AGS] with a fitness of 61.5531, where an amino acid between brackets indicates that the position can be filled up by one of them; 'X' indicates a wild card, which can be filled up by any of 20 natural amino acid residues. The other generated patterns were redundant or did not have the cysteine residues in conserved positions (data not shown). However, no sequences were retrieved from the regular expression search for uncharacterized sequences into NR. From this information, a more generalized pattern was proposed, by adding some wild cards and mixing it with the chitin-binding motif from Prosite (ID PS00026), generating the pattern CX(4,5)CC[GS]X(2)GXCGX[GST]X(2,3)[FWY]C[GS]X[AGS], where the numbers between brackets indicate the number of repetitions of the prior character (*i.e.*, 'X(4,5)' means that 'X' can be repeated four to five times). Using this new pattern, five uncharacterized sequences were obtained from NR. Due to the typical cysteine pattern and the presence of conserved residues, probably involved in chitin interactions, these sequences fall into the class of hevein-like peptides.

Initially, the sequence CBI18789 (GenBank ID: CBI18789) obtained from *Vitis vinifera* was found. This sequence shows 104 amino acid residues, of which the first 50 compose a signal peptide, according to Phobius and SignalP. The mature peptide has 54 amino acids. InterProScan indicates that the chitin-binding domain is present between residues 1 and 44 from the mature sequence. The ten remaining amino acids compose a short hydrophobic tail. The LOMETS server indicates that the best template for this sequence is the structure of pokeweed lectin-C from *Phytolacca americana* (PDB ID: 1ULK) [19], which shares 44.44% of identity with CBI18789. Theoretical models were constructed by using the structures 1ULK and 1T0W (see Table 2 for validation parameters). The overall structure is composed of an anti-parallel β-sheet and two short α-helices, being stabilized by four disulfide bridges (Fig. 2A). The rigid model structure suggests that four residues are responsible for binding on $(GlcNAc)_3$: SER^{20}, TYR^{22}, TYR^{24} and TYR^{31} (Fig. 2A). The complex stability was evaluated through MD, where the affinity between the peptide and the $(GlcNAc)_3$ molecule was observed. During the simulations, the complex was maintained by at least one hydrogen bond, varying to two, three of four hydrogen bonds along the simulation (Fig. S1A). This peptide undergoes a secondary structure loss (Fig. 3A), with great structural modification indicated by the backbone's RMSD of 8 Å (Fig. 4). The great RMSD variation is related to the first 17 residues of N-terminal and to the last 9 from C-terminal, according to the RMS fluctuation (Figs. 4A and S2A). However, as the

Table 1
Summary of hevein-like data set. The peptide's names are abbreviated as follows: *Amaranthus caudatus* Antimicrobial Peptide 2 (Ac-AMP2), *Ipomoea nil* Antimicrobial Peptide 2 (Pn-AMP1), *Eucommia ulmoides* Antifungal Peptide 1 (EAFP1), *Eucommia ulmoides* Antifungal Peptide 2 (EAFP2), Wheat Antimcrobial Peptide 1a (WAMP-1a), *Leymus arenarius* Antimicrobial Peptide 1a (LAMP-1a) and *Amaranthus retroflexus* Antimicrobial Peptide (Ar-AMP).

SwissProt ID	Name	Sequence	Annotated as antifungal?	Ref.
P27275	Ac-AMP2	VGECVRGRCPSGMCCSQFGYCGKGPKYCGR	Yes	[9]
P80359	Pseudo-Hevein	EQCGRQAGGKLCPNNLCCSQYGWCGSSDDYCSPSKNCQSNCKGGG	Yes	[55]
P81591	Pn-AMP1	QQCGRQASGRLCGNRLCCSQWGYCGSTASYCGAGCQSQCRS	Yes	[34]
P81859	Chitin-Binding Lectin	IDHRCGREATPPGKLCNDGRCCSQWGWCGTTQAYCSGKCQSQCDCNRDL	No	[60]
P83596	EAFP1	QTCASRCPRPCNAGLCCSIYGYCGSGNAYCGAGNCRCQCRG	Yes	[24]
P83597	EAFP2	QTCASRCPRPCNAGLCCSIYGYCGSGAAYCGAGNCRCQCRG	Yes	[19]
P85966	WAMP-1a	AQRCGDQARGAKCPNCLCCGKYGFCGSGDAYCGAGSCQSQCRGCR	Yes	[3]
P86521	LAMP-1a	AQKCGEQGRGAKCPNCLCCGRYGFCGSTPDYCGVGCQSQCRGC	Yes	n/a[a]
Q5I2B2	Ar-AMP	AGECVQGRCPSGMCCSQFGYCGRGPKYCGR	Yes	[8]

[a] Only deposited in SwissProt database.

Table 2
Summary of validation parameters of structural three-dimensional models.

Peptide id	DOPE score	Most favored regions (%)	Allowed regions (%)	*G*-Factor	*Z*-Score (PROSA II)
CBI18789	−3769.92578	86	9.3	−0.21	−4.77
EEE61250	−2001.24658	82.6	13	−0.41	−8.18
XP_002973523	−2576.74854	75	25	−0.41	−7.06
XP_001804616	−2142.82056	80	16.7	−0.28	−7.13

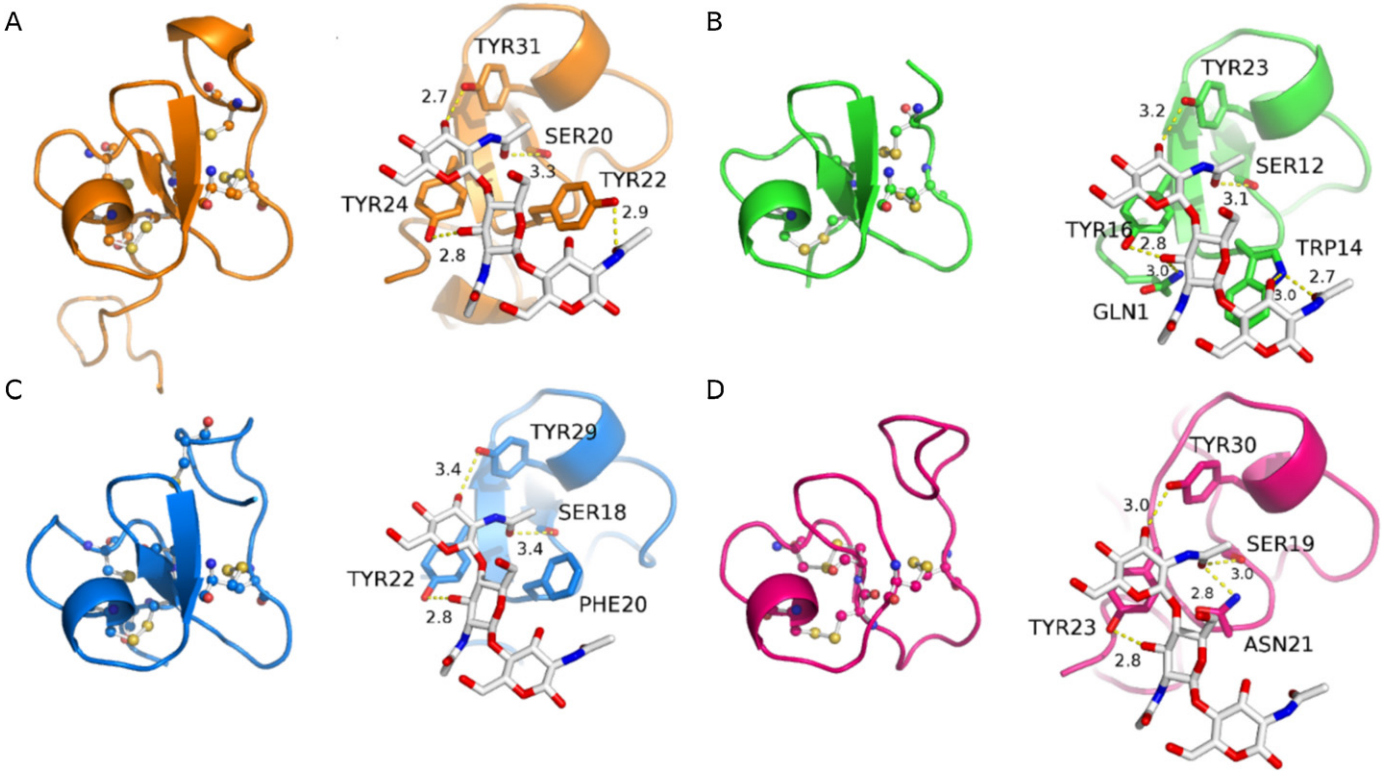

Fig. 2. Molecular models of hevein-like peptides (left panel) and their amino acid residues probably responsible for hevein-like peptide interactions with $(GlnNac)_3$. (A) CBI18789 from *V. vinifera*; (B) EEE61250 from *O. sativa*; (C) XP_002973523 from *S. moellendorffii*; and (D) XP_001804616 from *P. nodorum*. Disulfide bonds are represented as ball and stick. Distances are measured in angstroms.

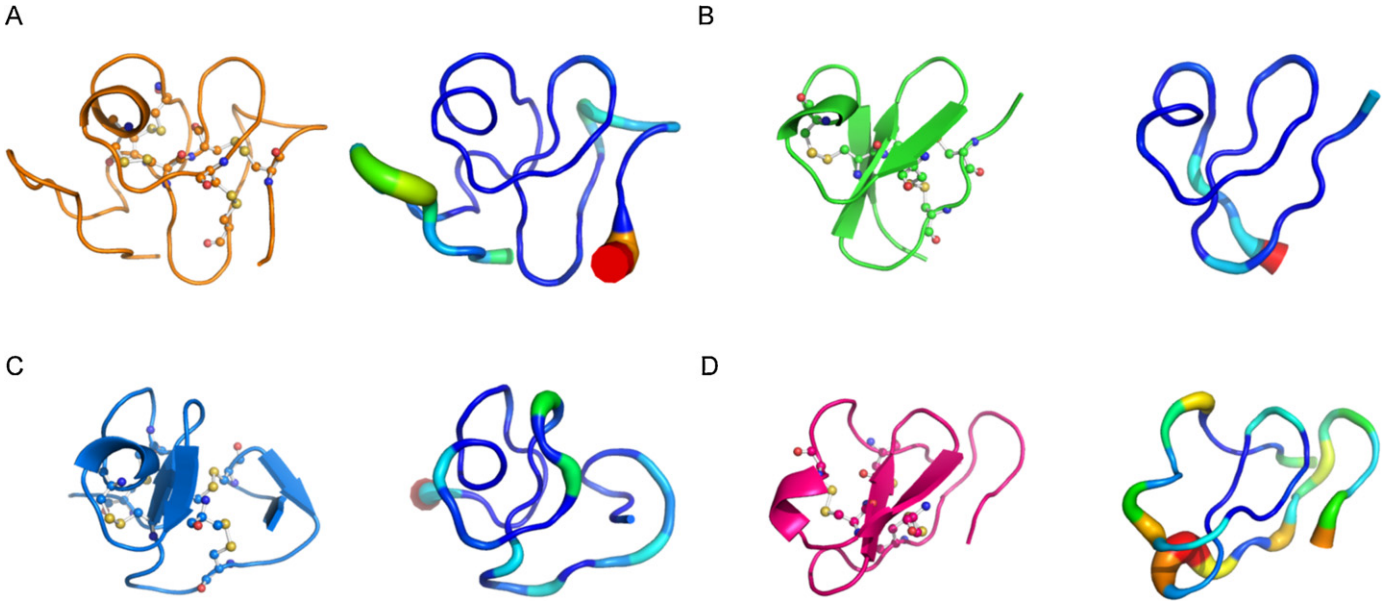

Fig. 3. Final three-dimensional model structures of hevein-like peptides after 50 ns of MD (left panel) and relative B-factors. (A) CBI18789 from *V. vinifera*; (B) EEE61250 from *O. sativa*; (C) XP_002973523 from *S. moellendorffii*; and (D) XP_001804616 from *P. nodorum*.

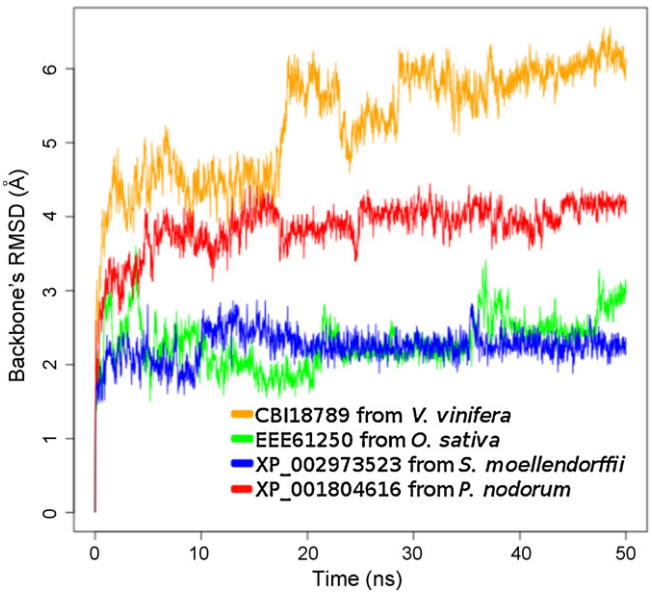

Fig. 4. RMSD backbones of hevein-like peptides along 50 ns of molecular dynamics.

structure is knotted by four disulfide bridges, the exposed residues are kept in positions where they can interact with $(GlcNAc)_3$.

Furthermore, the sequence EEE61250 (GenBank ID: EEE61250) was found from *Oryza sativa*. This sequence shows 122 amino acid residues and has a signal peptide comprising the first 23 residues according to Phobius and SignalP. The resulting mature peptide shows 99 amino acids. Additionally, this sequence shows a precursor organization similar to that observed for Ac-AMP2 and Ar-AMP (Fig. S3), which have a propeptide after the hevein domain. Assuming that this sequence has the same post-translational modifications as Ac-AMP2 and Ar-AMP, the final mature sequence has 32 amino acid residues. InterProScan indicates that the chitin-binding domain covers the whole mature sequence. Their three-dimensional model is composed of an anti-parallel β-sheet and one short α-helix, is stabilized by three disulfide bridges (Fig. 2B), and was constructed using the structure indicated by LOMETS 1ULK (71.88% of identity) in addition to 1T0W. Validation parameters are summarized in Table 2. The rigid model structure suggests that five residues are responsible for binding on $(GlcNAc)_3$: GLN^{1}, SER^{12}, TRP^{14}, TYR^{16} and TYR^{23} (Fig. 2B). As observed for the grape's peptide, the MD also indicates that this is a stable complex, being maintained by at least one hydrogen bond, and varying from one to six hydrogen bonds (Fig. S1B). The peptide structure shows the backbone's RSMD of 3 Å (Fig. 4) and does not lose the secondary structure, gaining instead an additional β-strand (Fig. 3B). This assumption was confirmed by a slight RMS fluctuation at the C-terminal (Figs. 4 and S2B).

Two sequences from *Selaginella moellendorffii* were retrieved, XP_002962191 (GenBank ID: XP_002962191) and XP_002973523 (GenBank ID: XP_002973523). XP_002962191 is 125 amino acids long, while XP_002973523 showed a length of 64 residues. A signal peptide was predicted in XP_002962191 covering the first 28 residues, resulting in a mature peptide with 97 amino acid residues. As well as the rice's peptide, this sequence may have a precursor organization similar to that of Ac-AMP2 and Ar-AMP. However, no similar cleavage sites have been observed among them. Therefore, XP_002962191 was removed from analysis, avoiding wrong conclusions. In contrast, the sequence XP_002973523 probably belongs to the hevein-like class. This sequence has a predicted signal peptide comprising the first 23 residues, resulting in a 41 amino acid long mature peptide. InterProScan indicates that the chitin-binding domain covers the whole mature sequence. The LOMETS server indicates that the best template for this sequence is the structure of class I chitinase from *O. sativa* (PDB ID: 2DKV) [30], that shares 46.34% of identity with XP_002973523. This model was submitted to an additional energy minimization, in order to stabilize the disulfide bond between CYS^{5} and CYS^{17}, since their sulfur atoms were distant by 2.2 Å, being the 2 Å correct distance for disulfide bond formation. The overall structure is composed by an anti-parallel β-sheet and one short α-helix, being stabilized by four disulfide bonds (Fig. 2C). Table 2 summarizes the validation data of the three-dimensional model. The rigid model structure suggests that three residues are responsible for binding on $(GlcNAc)_3$: SER^{18}, PHE^{20}, TYR^{22} and TYR^{29} (Fig. 2C). The complex is stabilized by three hydrogen bonds during the most of MD time, varying to zero to six hydrogen bonds (Fig. S1C). In this case, when the complex has no hydrogen bonds, it is stabilized by through CH-π stacking. This structure has a low backbone's RMSD variation, only 2.2 Å, indicating that it is very stable (Fig. 4). In the final structure, a short β-hairpin is observed (Fig. 3). The RMS fluctuation indicates a major fluctuation of two active residues PHE^{20} and TYR^{22} (Figs. 4 and S2C).

From the phytopathogenic fungus *Phaeosphaeria nodorum* the sequence XP_001804616 (GenBank ID: XP_001804616) was retrieved. This sequence is 58 amino acids long and the first 20 residues are predicted as a signal peptide. InterProScan indicates that the chitin-binding domain covers the whole mature sequence, which has 38 amino acid residues. Interestingly, XP_001804616 lacks two cysteine residues that are involved in different disulfide bond formation (Fig. 5). Thus, only two disulfide bridges would be correctly formed. However, in preliminary molecular models, the free cysteine residues are close to each other, indicating that an additional disulfide connection could be constructed (data not shown). Therefore, the molecular models were constructed including the third disulfide bridge, using the structures 1ULK (indicated by the LOMETS server, 44.74% of identity) and 1T0W. Due to the different disulfide bonding pattern, the model of the XP_001804616 mature sequence seems to be more unstable than the previous models, showing only one short α-helix, lacking the anti-parallel β-sheet (Fig. 2D). Despite these differences, the rigid molecular model suggests that four residues are responsible for binding $(GlcNAc)_3$: SER^{19}, ASN^{21}, TYR^{23} and TYR^{30} (Fig. 2D). Even with these differences, the validation parameters are similar to the other three models (Table 2). This complex was also stable during the MD, being stabilized by one, two or three hydrogen bonds, in the major

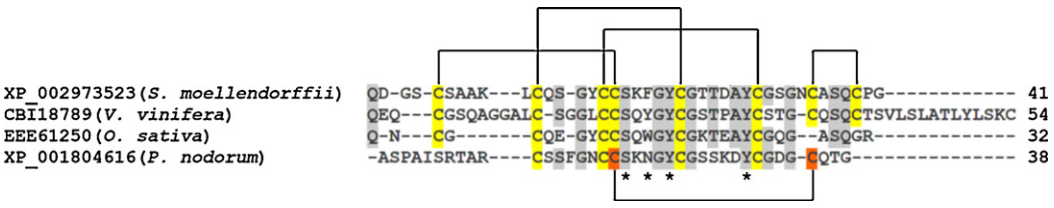

Fig. 5. Multiple alignment of hevein-like peptides here proposed. Conserved and semi-conserved residues are gray highlighted; the cysteine residues are in yellow and the cysteine residues that make the unusual bond of XP_001804616 (*P. nodorum*) are highlighted in orange. The residues responsible for interaction with $(GlnNac)_3$ are star marked. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

part the time. However, the absence of hydrogen bonds can be observed several times in the interval of 4.5 and 10 ns (Fig. S1D), where, actually, the hydrogen bond is made and undone, until the complex reach to stabilization. For this complex, the backbone's RMSD had increased in 4.1 Å (Fig. 4). A gain of secondary structure was observed, since the β-sheet that was lacked in the rigid model is formed (Fig. 3D). The RMS fluctuation indicates that the RMSD variation is caused mainly by the N-terminal loop (Figs. 4D and S2D), which is more unstable, due to the absence of a disulfide bridge.

Multiple sequence alignment (Fig. 5) shows that the residues that interact with chitin in the models are in the same position within the alignment. The alignment also shows that there is a size variation before the second cysteine residue. Moreover, it shows that the sequences from plants are more similar among themselves than in relation to the sequence from *P. nodorum*. In

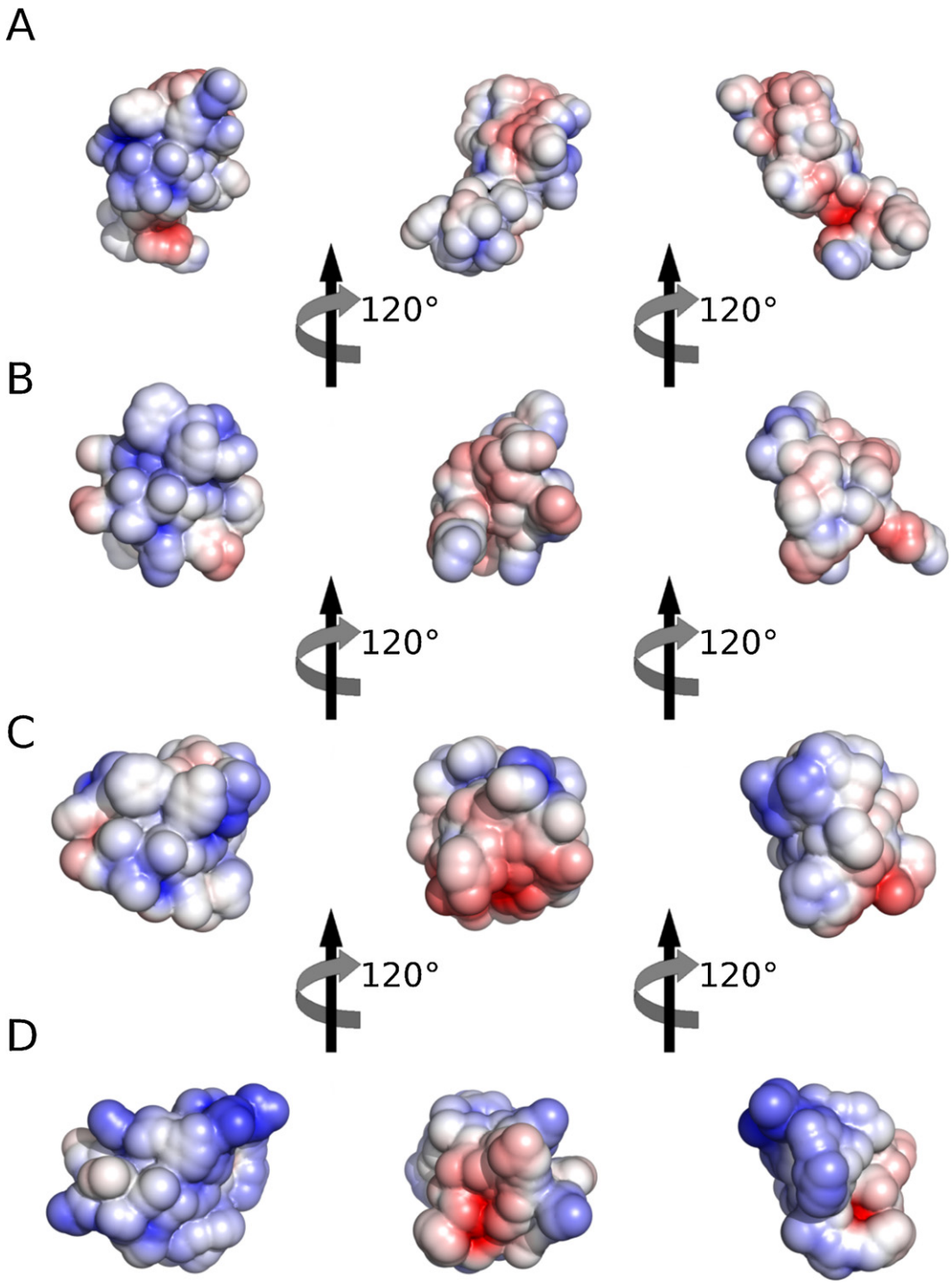

Fig. 6. Electrostatic potential of hevein-like peptides predicted structures. (A) CBI18789 from *V. vinifera*; (B) EEE61250 from *O. sativa*; (C) XP_002973523 from *S. moellendorffii*; and (D) XP_001804616 from *P. nodorum*. Cationic regions are blue colored, hydrophobic regions are white and anionic regions are red. Surface potentials were set to ±5 kT e^{-1} (133.56 mV). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

Table 3
Dali pairwise alignments. The results are arranged in a 4 × 4 matrix, where the upper triangle indicates the *Z*-Score values and the lower triangle, the backbone's RMSD values (in angstroms).

	EEE61250	XP_001804616	XP_002973523	CBI18789
EEE61250	–	3.7	3.6	3.7
XP_001804616	1.5	–	5.1	5.7
XP_002973523	1.0	1.4	–	7.2
CBI18789	1.3	1.0	0.8	–

addition to sequence alignments, structural pairwise alignments were also carried out. The most similar three-dimensional models were CBI18789 (*V. vinifera*) and XP_002973523 (*S. moellendorffii*) with a *Z*-Score on Dali of 7.2. The most different ones were EEE61250 (*O. sativa*) and XP_002973523 (*S. moellendorffii*) with a *Z*-Score of 3.6. The structural pairwise alignment results are summarized in Table 3. The structural alignments against the whole Protein Data Bank indicate that the four sequences here reported are related to other lectins with the hevein domain (Fig. S4). The models of CBI18789 (*V. vinifera*) and XP_002973523 (*S. moellendorffii*) are more similar to their own templates, the lectin PDB 1ULK and the chitinase PDB 2DKV, respectively. In the case of XP_001804616 (*P. nodorum*), agglutinin isolectin 1 was the most similar structure (PDB ID: 2UVO) [49]. Furthermore, in the case of EEE61250 (*O. sativa*), the hevein (PDB ID: 1Q9B) shows higher similarity [48].

Despite these sequence and structure differences, the four peptides were predicted to be antimicrobial peptides by the machine learning methods, both in the specific SVM for cysteine stabilized peptides and in the general methods from CAMP. However, by using CAMP's discriminant analysis, the mature sequence from EEE61250 (*O. sativa*) was negatively predicted, indicating that this peptide may not show antimicrobial activity. In addition, the electrostatic surfaces for each theoretical model were also calculated (Fig. 6). An amphipathic surface can be observed in all peptides here reported. Taking into account that the amphipathic surfaces are required for membrane interactions, it seems that they probably could interact with anionic membranes.

4. Discussion

By means of high throughput genome sequencing methods, the use of sequence databases emerges as a novel source for identifying biologically active molecules [54]. The availability of genome databases and their translations offers a remarkable information resource, revealing novel aspects about several classes of peptides and proteins. The data mining methods allow several sequences to be found simultaneously in diverse organisms. Both nucleotide and protein sequence databases are undeniably a source of biologically active molecules. Therefore, several methods have been proposed for exploring it, including artificial intelligence [15,36,46,57] and similarity search methods [42,54,65]. The similarity search methods are more restricted for a determined class than the artificial intelligence ones. Nevertheless, similarity search methods can bring to light novel aspects about the distribution and/or evolution of an antimicrobial class.

The use of patterns for searching novel sequences is more useful for cysteine stabilized classes, since their structures are stabilized by disulfide bonds, which typify the class [54]. Thus, this method is appropriate in the search for novel hevein-like peptides in protein databases. However, a pattern first needs to be defined. Hence, the automatic search system was used for retrieving the hevein-like sequences, and subsequently these sequences were used for pattern recognition through Pratt 2.1. By means of the first generated pattern no novel sequences were retrieved by regular expression search. This first attempt failed due to the high pattern specificity, since it was constructed based on only eight sequences. Therefore, a more generalized pattern was needed, in order to find the uncharacterized hevein-like peptides precursors in NR.

Among the five originally identified precursors, three of them (CBI18789 from *V. vinifera*, EEE61250 from *O. sativa* and XP_002962191 from *S. moellendorffii*) have sequences larger than the hevein domain. This feature had already been observed for the hevein-like precursors of Ac-AMP2 [9], Ar-AMP [37] and WAMP-1 [3]. In fact, these sequences after the hevein domain are propeptides and are posteriorly cleaved, leaving the mature peptide. In the case of hevein-like peptides, the propeptides could be related to the evolution process which originated this class. Andreev et al. [3] have proposed that the WAMP-1 peptide emerges from a deletion of the catalytic domain of a chimerolectin (class I chitinase from plants). These three peptides with sequences after the hevein domain also show similarities to chimerolectins (data not shown), indicating that these sequences may be originated from a similar evolutionary process of WAMP-1. There are two other hypotheses involving the evolution of lectins with the hevein domain, which propose duplication or transposition of hevein domains, contrasting with Andreev et al. [3]. Wright et al. [62] have proposed that the consecutive duplication of hevein domains in the same gene generated the hololectins, and Shinshi et al. [50] have suggested that the transposition of an hevein domain into the gene of a chitinase generated the chimerolectins.

Presumably, these sequences after the hevein domain are remnants of the evolutionary process. In CBI18789 (*V. vinifera*), this sequence corresponds to a short hydrophobic tail. This tail does not generate great changes in structure and, probably, neither in protein function. In fact a transition of coil-to-β-sheet was observed in MD, however, without influences in the binding to $(GlcNAc)_3$ (data not shown). In addition, there is no clear evidence that this tail is cleaved. Similarly, XP_002962191 from *S. moellendorffii* also shows an additional sequence after the hevein domain. Nonetheless, it is longer than CBI18789's tail and, in this case, there may be a structural or functional change if it is not cleaved. Hence, it was removed from the analysis, since clear evidence of cleavage was not observed. In contrast, EEE61250 (*O. sativa*) has a similar cleavage site to Ac-AMP2 and Ar-AMP, indicating that the additional sequence may be a propeptide. In this last case, besides the cleavage site, these sequences also share the same number of cysteine residues. The other two retrieved sequences have only the signal peptide and the hevein domain, without additional sequences after the hevein domain.

Although the biological activities have been not characterized yet, there are some indications of what they might be. The antimicrobial activity predictions were almost all positive, except in the case of EEE61250 (*O. sativa*), only negatively predicted by CAMP discriminant analysis. These predictions show that overall the properties of these sequences are similar to those from well-known antimicrobial peptides, such as hydrophobicity, net charge and secondary structure [46,57]. Loose et al. [38] proposed that AMPs work as a formal language, analogous to a grammatical structure composed of several rules (patterns and motifs) and a vocabulary (amino acids). In this view, the positive predictions are probably due to the grammatical structure of chitin-binding motif, present in all putative mature sequences here reported.

Other evidence of their biological activities was drawn from molecular models in complex to $(GlcNAc)_3$ (Fig. 2) in addition to the molecular dynamics simulations. The proposed mechanism of action of fungicidal activity in hevein-like peptides is related to the inhibition of cell wall elongation. The molecular dynamics show that the four hevein-like peptides here reported can bind to $(GlcNAc)_3$ (Fig. S1). Among the sequences here reported, the sequence EEE61250 (*O. sativa*) seems to have the strongest fungicidal activity against chitin-containing fungi. The molecular model indicates that it interacts with chitin through five amino

acid residues making six hydrogen bonds (Fig. 2B). Besides, this sequence has aromatic residues identical to Pn-AMP2 [33], one of the strongest hevein-like peptides already reported, which requires concentrations of 0.6–75 $\mu g\,ml^{-1}$ for 50% of inhibition of fungal growth. Following the same reasoning, the activity of CBI18789 (*V. vinifera*) would be similar to EAFP2 [24], since their aromatic residues are identical. And for XP_002973523 (*S. moellendorffii*), the activity would be similar to Ac-AMP2 [9]. Nonetheless, for the peptide XP_001804616 (*P. nodorum*), there are no peptides with identical active residues. Otherwise, this peptide can also make four hydrogen bonds (Fig. 2D). Moreover its hydrophobic interactions are reduced, since it lacks an aromatic residue. Taking into account the electrostatic surface, all peptides might interact with anionic membranes from chitin-free fungi and/or bacteria, since they have an amphipathic electrostatic surface (Fig. 6). However, despite these indications, only *in vitro* tests can reveal their actual activities.

In fact, the most intriguing sequence is XP_001804616 (*P. nodorum*). Although the hevein domain was previously identified in the chimerolectin CPB1 from *M. grisea* and also the fact that this domain appears in other chimerolectins in databases [29], XP_001804616 is the first report of a fungal hevein-like peptide, a merolectin. This peptide has two notorious differences when compared with plant hevein-like peptides. The first one is the absence of two conserved cysteine residues involved in two different disulfide bonds. Due to this absence, their pairs may make an unusual disulfide bridge (Fig. 5). Jayanthi et al. [26] show, through molecular dynamics simulations, that the first and the last disulfide bridges in EAFP2 do not generate clear structural modifications. The molecular model from XP_001804616 also indicates similar structural process, since the overall folding is the same as others.

The second remarkable difference relies on the key residues involved in chitin-binding. Instead of one serine and three aromatic residues, fungal peptide has one serine, one asparagine and two aromatic residues. Interestingly, in CBP1 from *M. grisea* an aromatic residue is also lacking in the position X_{i+2}. Changes in the key residues may result in changes in the chitin affinity, the serine to aspartate mutation in HEV32 reduces the affinity to $(GlcNAc)_3$ by almost half [10]. This mutation in X_{i+2} is probably involved in its actual function. Although the predictions indicate that XP_001804616 (*P. nodorum*) has antimicrobial activity, it may be a virulence factor. The chitin-binding protein Avr4 form *Cladosporium fulvum* protects the chitin wall against hydrolysis by plant chitinases through its chitin-binding ability [59]. Chitin protection has also been proposed for several predicted proteins from *M. grisea*, with the pattern CX(5)CCX(7)C, related to chitin-binding proteins [13]. In addition, the CBP1 protein from *M. grisea* plays a crucial role in appressorium differentiation, a prerequisite for penetration into host plants [29].

Naturally, a novel question about the evolution of hevein domains emerges, since the current propositions were made based only on hevein domains from plants. Do hevein domains have a common ancestor or did they arise by co-evolution? This question will remain unclear at least until more fungal proteins with hevein domains have been discovered. As novel fungal sequences are found, they may shed some light on the evolution and the mechanisms of chitin binding of the hevein domain. At least molecular models indicate that there are no clear differences among the structures of fungal and plant hevein-like peptides. The rigid model structures from the peptides here reported are very similar to each other (Table 3) and to other lectins with the hevein domain with solved structures (Fig. S4). They also have a similar behavior in molecular dynamics simulations. Independently of the peptide, the hevein domain keeps its fold since the structure is knotted by at least three disulfide bonds. The hevein domain is so stable that when an elevated variation in the backbone's RMSD is observed (Fig. 4), the RMS fluctuation is only improved on loop regions (Fig. S2). Even when there is losing of secondary structure, as observed in CBI18789 (*V. vinifera*), the peptides keep interacting with $(GlcNAc)_3$ (Fig. S1). On the other hand, the greater difference between the hevein-like peptide from *P. nodorum* and the hevein-like peptide from plants is, in fact, the amino acid residues involved in chitin-binding: while in plants the framework X_i-X_{i+2}-X_{i+4}-X_{i+11} is filled up by a serine followed by three aromatic residues, in *P. nodorum* it is filled up by a different amino acid residue in position X_{i+2}. The influence of this mutation in chitin binding ability is unclear.

5. Conclusions

In conclusion, data here reported indicate that searching for patterns in databases can produce new information about certain classes of proteins, in this case the hevein-like peptides. The NR database is almost a metaproteomics resource due to the diversity of sources of protein sequences deposited in it. Similarity search methods, including local alignment and regular expression search, are pivotal tools for exploring this source. Through these methods, novel hevein-like peptide precursors were identified, including one from a fungal source, a surprising result, since the hevein-like peptides were until now restricted to plants. This discovery was only possible because the search was made directly in NR. The peptides here identified can be used for construction of novel transgenic organisms with resistance to phytopathogenic fungi, as soon as their activities have been tested. Moreover, the discovery of a fungal hevein-like peptide in the present work raises a novel question about the hevein domain's evolution: did it arise from a common ancestor or by co-evolution? In addition, contrasting with the other three hevein-like peptides identified from plants, the function of the hevein-like peptide from *P. nodorum* is notoriously a mystery. Although the hevein-like peptides are involved in plant defense against microbes, what is their exact role in fungal biology? In fact, more hevein-like peptides from fungi need to be identified to allow further *in vitro* and *in vivo* characterization.

Acknowledgements

The authors are grateful to Center for Scientific Computing (NCC/GridUNESP) of the São Paulo State University (UNESP), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) and Universidade Católica de Brasília (UCB) for the support.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.peptides.2012.07.025.

References

[1] Aboitiz N, Vila-Perelló M, Groves P, Asensio JL, Andreu D, Cañada FJ, et al. NMR and modeling studies of protein–carbohydrate interactions: synthesis, three-dimensional structure, and recognition properties of a minimum hevein domain with binding affinity for chitooligosaccharides. Chembiochem 2004;5(9):1245–55.

[2] Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 1997;25(17):3389–402.

[3] Andreev YA, Korostyleva TV, Slavokhotova AA, Rogozhin EA, Utkina LL, Vassilevski AA, et al. Genes encoding hevein-like defense peptides in wheat: distribution, evolution, and role in stress response. Biochimie 2012;94:1009–16, http://dx.doi.org/10.1016/j.biochi.2011.12.023.

[4] Archer BL. The proteins of *Hevea brasiliensis* Latex. 4. Isolation and characterization of crystalline hevein. Biochem J 1960;75:236–40.

[5] Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA. Electrostatics of nanosystems: application to microtubules and the ribosome. Proc Natl Acad Sci USA 2001;98(18):10037–41.

[6] Beintema JJ, Peumans WJ. The primary structure of stinging nettle (*Urtica dioica*) agglutinin. A two-domain member of the hevein family. FEBS Lett 1992;299(2):131–4.

[7] Beintema JJ. Structural features of plant chitinases and chitin-binding proteins. FEBS Lett 1994;350:159–63, http://dx.doi.org/10.1016/0014-5793(94)00753-5.

[8] Berendsen HJC, Postma JPM, van Gunsteren WF, Hermans J. Interaction models for water in relation to protein hydration. In: Pullman B, editor. Intermolecular Force. Reidel: Dordrecht; 1981. p. 331–42.

[9] Broekaert WF, Marien W, Terras FRG, De Bolle MFC, Proost P, Van Damme J, et al. Antimicrobial peptides from *Amaranthus caudatus* seeds with sequence homology to the cysteine/glycine-rich domain of chitin-binding proteins. Biochemistry 1992;31(17):4308–14.

[10] Chávez MI, Vila-Perelló M, Cañada FJ, Andreu D, Jiménez-Barbero J. Effect of a serine-to-aspartate replacement on the recognition of chitin oligosaccharides by truncated hevein. A 3D view by using NMR. Carbohydr Res 2010;345(10):1461–8, http://dx.doi.org/10.1016/j.carres.2010.02.019.

[11] Darden T, York D, Pedersen L. Particle mesh Ewald: An $N\log(N)$ method for Ewald sums in large systems. J Chem Phys 1993;98:10089–92.

[12] De Holf PL, Brill LM, Hirsch AM. Plant lectins: the ties that bind in root symbiosis and plant defense. Mol Genet Genomics 2009;282:1–15, http://dx.doi.org/10.1007/s00438-009-0460-8.

[13] Dean RA, Talbot NJ, Ebbole DJ, Farman ML, Mitchell TK, Orbach MJ, et al. The genome sequence of the rice blast fungus *Magnaporthe grisea*. Nature 2005;434(7036):980–6.

[14] Eswar N, Marti-Renomet MA, Webb B, Madhusudhan MS, Eramian D, Shen M, et al. Comparative protein structure modeling with MODELLER. Curr Protoc Bioinform 2007;(Suppl. 15):5.6.1–30.

[15] Fernandes FC, Porto WF, Franco OL. A wide antimicrobial peptides search method using fuzzy modeling. Advances in bioinformatics and computational biology. LNCS 2009;5676:147–50, http://dx.doi.org/10.1007/978-3-642-03223-3_14.

[16] Franco OL. Peptide promiscuity: an evolutionary concept for plant defense. FEBS Lett 2011;585(7):995–1000.

[17] Guex N, Peitsch MC. SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. Electrophoresis 1997;18(15):2714–23.

[18] Hasegawa H, Holm L. Advances and pitfalls of protein structural alignment. Curr Opin Struct Biol 2009;19:341–8.

[19] Hayashida M, Fujii T, Hamasu M, Ishiguro M, Hata Y. Similarity between protein–protein and protein–carbohydrate interactions, revealed by two crystal structures of lectins from the roots of pokeweed. J Mol Biol 2003;334(3):551–65.

[20] Hernández-Gay JJ, Ardá A, Eller S, Mezzato S, Leeflang BR, Unverzagt C, et al. Insights into the dynamics and molecular recognition features of glycopeptides by protein receptors: the 3D solution structure of hevein bound to the trisaccharide core of *N*-glycoproteins. Chem Eur J 2010;16:10715–26, http://dx.doi.org/10.1002/chem.201000939.

[21] Hess B, Bekker H, Berendsen HJC, Fraaije JGEM. LINCS. A linear constant solver for molecular simulations. J Comp Chem 1997;18(12):1463–72.

[22] Hess B, Kutzner C, van der Spoel D, LIndahl E. GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. J Chem Theory Comput 2008;4:435–47.

[23] Holm L, Rosenström P. Dali server: conservation mapping in 3D. Nucleic Acids Res 2010;38:W545–9.

[24] Huang RH, Xiang Y, Tu GZ, Zhang Y, Wang DC. Solution structure of Eucommia antifungal peptide: a novel structural model distinct with a five-disulfide motif. Biochemistry 2002;43(20):6005–12.

[25] Inbar J, Chet I. A newly isolated lectin from the plant pathogenic fungus *Sclerotium roltsii*: purification, characterization and role in mycoparasitism. Microbiology 1994;140:651–7, http://dx.doi.org/10.1099/00221287-140-3-651.

[26] Jayanthi M, Kangueane P, Prakash NAU, Jigisha DVA, Sekar K. Influence of disulfide bonds on the conformation of the antifungal protein from *Eucommia ulmoides* Oliver. Open Struct Biol J 2009;3:94–103.

[27] Jonassen I, Collins JF, Higgins DG. Finding flexible patterns in unaligned protein sequences. Protein Sci 1995;4(8):1587–95.

[28] Käll L, Krogh A, Sonnhammer EL. Advantages of combined transmembrane topology and signal peptide prediction—the Phobius web server. Nucleic Acids Res 2007;35:W429–32 [Web Server issue].

[29] Kamakura T, Yamaguchi S, Saitoh K, Teraoka T, Yamaguchi I. A novel gene, CBP1, encoding a putative extracellular chitin-binding protein, may play an important role in the hydrophobic surface sensing of *Magnaporthe grisea* during appressorium differentiation. Mol Plant Microbe Interact 2002;15(5):437–44.

[30] Kezuka Y, Kojima M, Mizuno R, Suzuki K, Watanabe T, Nonaka T. Structure of full-length class I chitinase from rice revealed by X-ray crystallography and small-angle X-ray scattering. Proteins 2010;78(10):2295–305.

[31] Khaliluev MR, Mamonov AG, Smirnov AN, Kharchenko PN, Dolgov SV. Expression of genes encoding chitin binding proteins (PR 4) and hevein like antimicrobial peptides in transgenic tomato plants enhanced resistance to *Phytophthora infestance*. Russ Agric Sci 2011;37(4):297–302, http://dx.doi.org/10.3103/S1068367411040082.

[32] Kolbe S, Fischer S, Becirevic A, Hinz P, Schrempf H. The *Streptornyces reticuli* α-chitin-binding protein CHBZ and its gene. Microbiology 1998;144:1291–7, http://dx.doi.org/10.1099/00221287-144-5-1291.

[33] Koo JC, Chun HJ, Park HC, Kim MC, Koo YD, Koo SC, et al. Over-expression of a seed specific hevein-like antimicrobial peptide from *Pharbitis nil* enhances resistance to a fungal pathogen in transgenic tobacco plants. Plant Mol Biol 2002;50(3):441–52.

[34] Koo JC, Lee SY, Chun HJ, Cheong YH, Choi JS, Kawabata S, et al. Two hevein homologs isolated from the seed of *Pharbitis nil* L. exhibit potent antifungal activity. Biochim Biophys Acta 1998;1382:80–90.

[35] Laskowski RA, Rullmannn JA, MacArthur MW, Kaptein R, Thornton JM. AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. J Biomol NMR 2007;8(4):477–86.

[36] Lata S, Sharma BK, Raghava GP. Analysis and prediction of antibacterial peptides. BMC Bioinform 2007;8:263, http://dx.doi.org/10.1186/1471-2105-8-263.

[37] Lipkin A, Anisimova V, Nikonorova A, Babakov A, Krause E, Bienert M, et al. An antimicrobial peptide Ar-AMP from amaranth (*Amaranthus retroflexus* L.) seeds. Phytochemistry 2005;66:2426–31, http://dx.doi.org/10.1016/j.phytochem.2005.07.015.

[38] Loose C, Jensen K, Rigoutsos I, Stephanopoulos G. A linguistic model for the rational design of antimicrobial peptides. Nature 2006;443(7113):867–9.

[39] Luo Y, Xu X, Liu J, Li J, Sun Y, Liu Z, et al. A novel mannose-binding tuber lectin from *Typhonium divaricatum* (L.) Decne (family Araceae) with antiviral activity against HSV-II and anti-proliferative effect on human cancer cell lines. J Biochem Mol Biol 2007;40(3):358–67.

[40] Martins JC, Maes D, Loris R, Pepermans HA, Wyns L, Willem R, et al. H NMR study of the solution structure of Ac-AMP2, a sugar binding antimicrobial protein isolated from *Amaranthus caudatus*. J Mol Biol 1996;258(2):322–33.

[41] Miyamoto S, Kollman PA. SETTLE. An analytical version of the SHAKE and RATTLE algorithm for rigid water models. J Comp Chem 1992;13(8):952–62.

[42] Mulvenna JP, Mylne JS, Bharathi R, Burton RA, Shirley NJ, Fincher GB, et al. Discovery of cyclotide-like protein sequences in graminaceous crop plants: ancestral precursors of circular proteins? Plant Cell 2006;18(9):2134–44.

[43] Odintsova TI, Vassilevski AA, Slavokhotova AA, Musolyamov AK, Finkina EI, Khadeeva NV, et al. A novel antifungal hevein-type peptide from *Triticum kiharae* seeds with a unique 10-cysteine motif. FEBS J 2009;276(15):4266–75.

[44] Petersen TN, Brunak S, von Heijne G, Nielsen H. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Methods 2011;8(10):785–6, http://dx.doi.org/10.1038/nmeth.1701.

[45] Peumans WJ, Van Damme JM. Lectins as piant defense proteins. Plant Physiol 1995;109:347–52.

[46] Porto WF, Fernandes FC, Franco OL. An SVM model based on physicochemical properties to predict antimicrobial activity from protein sequences with cysteine knot motifs. LNCS 2010;6268:59–62, http://dx.doi.org/10.1007/978-3-642-15060-9_6.

[47] Quevillon E, Silventoinen V, Pillai S, Harte N, Mulder N, Apweiler R, et al. InterProScan: protein domains identifier. Nucleic Acids Res 2005;33:W116–20 [Web Server issue].

[48] Reyes-López CA, Hernández-Santoyo A, Pedraza-Escalona M, Mendoza G, Hernández-Arana A, Rodríguez-Romero A. Insights into a conformational epitope of Hev b 6.02 (hevein). Biochem Biophys Res Commun 2004;314:123–30.

[49] Schwefel D, Maierhofer C, Beck JG, Seeberger S, Diederichs K, Möller HM, et al. Structural basis of multivalent binding to wheat germ agglutinin. J Am Chem Soc 2010;132(25):8704–19.

[50] Shinshi H, Neuhas JM, Ryals J, Meins Jr F. Structure of a tobacco endochitinase gene: evidence that different chitinase genes can arise by transposition of sequences encoding a cysteine-rich domain. Plant Mol Biol 1990;14(3):357–68.

[51] Shukurov KK, Voblikoba VD, Nikonorova AK, Egorov TA, Grishin EV, Babakov AV. Increase of resistance of *Arabidopsis thaliana* plants to phytopathogenic fungi expressing hevein like peptides from weed plant *Stellaria media*. Russ Agric Sci 2010;36(4):265–7, http://dx.doi.org/10.3103/S1068367410040117.

[52] Shukurov KK, Voblikova VD, Nikonorova AK, Komakhin RA, Komakhina VV, Egorov TA, et al. Transformation of tobacco and *Arabidopsis* plants with *Stellaria media* genes encoding novel hevein-like peptides increases their resistance to fungal pathogens. Transgenic Res 2011, http://dx.doi.org/10.1007/s11248-011-9534-6.

[53] Siebert HC, Lu SY, Wechselberger R, Born K, Eckert T, Liang S, et al. A lectin from the Chinese bird-hunting spider binds sialic acids. Carbohydr Res 2009;344(12):1515–25, http://dx.doi.org/10.1016/j.carres.2009.06.002.

[54] Silverstein KAT, Moskal WA, Wu HC, Underwood BA, Graham MA, Town CD, et al. Small cysteine-rich peptides resembling antimicrobial peptides have been under-predicted in plants. Plant J 2007;51(2):262–80, http://dx.doi.org/10.1111/j.1365-313X.2007.03136.x.

[55] Soedjanaatmadja UM, Hofsteenge J, Jeronimus-Stratingh CM, Bruins AP, Beintema JJ. Demonstration by mass spectrometry that pseudo-hevein and hevein have ragged C-terminal sequences. Biochim Biophys Acta 1994;1209:144–8.

[56] The UniProt Consortium. Reorganizing the protein space at the Universal Protein Resource (UniProt). Nucleic Acids Res 2012;40:D71–5.

[57] Thomas S, Karnik S, Barai RS, Jayaraman VK, Idicula-Thomas S. CAMP: a useful resource for research on antimicrobial peptides. Nucleic Acids Res 2010;38:D774–8 [Database issue].
[58] Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 1994;22(22):4673–80.
[59] van Esse HP, Bolton MD, Stergiopoulos I, de Wit PJGM, Thomma BPHJ. The chitin-binding *Cladosporium fulvum* effector protein Avr4 is a virulence factor. Mol Plant Microbe Interact 2007;20(9):1092–101, http://dx.doi.org/10.1094/MPMI-20-9-1092.
[60] Voelter W, Wacker R, Franz M, Maier T, Stoeva S. Complete structural characterization of a chitin-binding lectin from mistletoe extracts. J Prakt Chem 2000;342:812–8, http://dx.doi.org/10.1002/1521-3897(200010)342:8<812::AID-PRAC812>3.0.CO;2-J.
[61] Wiederstein M, Sippl MJ. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucleic Acids Res 2007;35:W407–10 [Web Server issue].
[62] Wright HT, Sandrasegaram G, Wright CS. Evolution of a family of *N*-acetylglucosamine binding proteins containing the disulfide-rich domain of wheat germ agglutinin. J Mol Evol 1991;33(3):283–94, http://dx.doi.org/10.1007/BF02100680.
[63] Wu S, Zhang Y. LOMETS: a local meta-threading-server for protein structure prediction. Nucleic Acids Res 2007;35(10):3375–82.
[64] Xiang Y, Huang RH, Liu XZ, Zhang Y, Wang DC. Crystal structure of a novel antifungal protein distinct with five disulfide bridges from *Eucommia ulmoides* Oliver at an atomic resolution. J Struct Biol 2004;148:86–97.
[65] Zhu S. Discovery of six families of fungal defensin-like peptides provides insights into origin and evolution of the CSαβ defensins. Mol Immunol 2008;45(3):828–38.