Cupin: A candidate molecular structure for the Nep1-like protein family

Adelmo L Cechin¹, Marialva Sinigaglia¹, Ney Lemke*², Sérgio Echeverrigaray³, Odalys G Cabrera⁴, Gonçalo AG Pereira⁴ and José CM Mombach⁵

Address: ¹Programa de Pós-Graduação em Computação Aplicada, Unisinos, Av. Unisinos – 950, São Leopoldo, Brasil, ²Departamento de Física e Biofísica, UNESP, Dist. Rubião Jr. sn, Botucatu, Brasil, ³Instituto de Biotecnologia, UCS, R. Francisco Getúlio Vargas 1130, Caxias do Sul, Brasil, ⁴Departamento de Genética e Evolução, IB/UNICAMP, Campinas, Brasil and ⁵Centro de Ciências Rurais, UFPampa/UFSM, São Gabriel, Brasil

Email: Adelmo L Cechin - acechin@unisinos.br; Marialva Sinigaglia - msinigaglia@gmail.com; Ney Lemke* - lemke@ibb.unesp.br; Sérgio Echeverrigaray - selaguna@ucs.br; Odalys G Cabrera - odalys@lge.ubi.unicamp.br; Gonçalo AG Pereira - goncalo@unicamp.br; José CM Mombach - jcmombach@smail.ufsm.br

* Corresponding author

Abstract

Background: NEP1-like proteins (NLPs) are a novel family of microbial elicitors of plant necrosis. Some NLPs induce a hypersensitive-like response in dicot plants though the basis for this response remains unclear. In addition, the spatial structure and the role of these highly conserved proteins are not known.

Results: We predict a 3d-structure for the β-rich section of the NLPs based on alignments, prediction tools and molecular dynamics. We calculated a consensus sequence from 42 NLPs proteins, predicted its secondary structure and obtained a high quality alignment of this structure and conserved residues with the two Cupin superfamily motifs. The conserved sequence GHRHDWE and several common residues, especially some conserved histidines, in NLPs match closely the two cupin motifs. Besides other common residues shared by dicot Auxin-Binding Proteins (ABPs) and NLPs, an additional conserved histidine found in all dicot ABPs was also found in all NLPs at the same position.

Conclusion: We propose that the necrosis inducing protein class belongs to the Cupin superfamily. Based on the 3d-structure, we are proposing some possible functions for the NLPs.

Background

More than 10 years ago, a 24-kD necrosis and ethylene inducing protein, named NEP1, capable of triggering plant cell death was purified from culture filtrates of Fusarium oxysporum. Since then, several other NEP1-like proteins (NLPs) have been identified in diverse microorganisms; including bacteria, fungi, and oomycetes [1]. In several cases, one species have more than one copy of NLPs and it is believed that several of these copies are pseudogenes [2,3]. NLPs constitute a family of phytotoxic proteins that contains a secretory signal sequence and are able to elicit cell death and defense responses in a large number of dicot plants (reviewed by [4] and [2]). Most species with NLPs are plant pathogens but there are excep-
tions, since genes encoding NLPs have been detected in fungal and bacterial species that are not known to be pathogenic.

A recently published study identified three copies of NLPs in the basidiomycete *Moniliophthora perniciosa* (MpNEPs). *M. perniciosa*, the causal agent of the witches' broom disease in *Theobroma cacao*, is responsible for major crop losses in the Americas. The authors observed that despite the high sequence similarity, MpNEP1 and MpNEP2 present different structural features, and MpNEP2 activity was resistant to high temperatures. They also demonstrated that these genes are differentially expressed in two different life stages of the fungus [5].

All NLPs contain a conserved domain called necrosis-inducing *Phytophthora* protein 1 (NPP1) [6]. The current lack of knowledge about functional domains, cellular targeting or protein binding motifs in this type of proteins complicates the unveiling of the actual function of NLPs [4]. There is an increasing interest in the determination of their function, role in plant-pathogen interactions and molecular structure [4]. The main conserved motif GHRHDKWE shows no significant similarity to any currently known protein sequence and so provides no clues to NLPs function.

The Cupin superfamily was identified by Dunwell in 1998 [7] and is among the most functionally diverse folding described to date, comprising both enzymatic and non-enzymatic members. These include helix-turn-helix transcription factor, AraC type transcription factor, oxalate decarboxylase, auxin-binding protein, globulins, etc. Many proteins on this superfamily have functions and chemical properties related to the NLPs: Auxin-Binding Proteins (ABPs) are hormone receptors and have a great influence on plant physiology; the related oxalate oxidase is involved in pathogen activities and germin-like proteins, apoplastic, glycoproteins are remarkably protease-resistant because of their cupin fold.

According to Dunwell et al. [8,9] the cupin domain comprises two conserved motifs, each corresponding to two β-strands, separated by a less conserved region composed of another two β-strands with an intervening variable loop. The total size of the inter motif region varies from 11 residues to ca. 50 residues. The characteristic conserved sequence in motif 1 and 2 is g(x)_3hxh(x)_3,4e(x)_6g and g(x)_3pxg(x)_2h(x)_3n, respectively.

**Introduction**

Homology searches using the NCBI-Blast produces no useful results in relation to the 3d-structure because the possible candidates have such a low score that they cannot be considered viable candidates. The result is a long list of necrosis and ethylene-inducing proteins, all of which are β-sheet-rich proteins, but none with useful information with associated 3d-structure. Any attempt to find other similar proteins based on their 1d-structure (sequence) to NLP protein results in other NLP proteins. In this article, we propose a 3d-structure for this protein family based on: (1) 1d-structure and conserved residues, (2) the supposed catalytic center, (3) the predicted signal sequences and target location, (4) cysteine and histidine conserved residues, and (5) the predicted 2d-structures. Our computational experiments in association with experimental clues point to the Cupin superfamily as the structure of the NLPs.

The article is divided as follows. In section Methods we present the sequences chosen for the analysis and the results of NLPs alignments concerning the conserved residues. Based on the pattern of conserved residues, we looked for candidate structures taking into account also the predicted 2d-structure. The candidates selected were those with the best agreement in 2d-structure and conserved residues with NLPs. With these we are proposing a 3d-structure for the core region (β-strand-rich region, positions 90–220 in Figure 1) of the type I NLPs. Based on this proposal; we analyze the most central part of the NLPs and discuss its relation to known proteins.

**Results and Discussion**

**Alignment Analysis**

Gijzen and Nürnberger classify NLPs into two groups: those containing two cysteines (type I NLPs) and those containing four (type II NLPs). Type I NLPs occur in fungi, oomycetes and bacteria while type II NLPs do not occur in oomycetes [2]. Sequence alignment differentiated the NLPs into these two main groups.

The statistical approach presented here parallels the different levels of phytopathogenicity shown by NLPs. They affect different species at different levels of intensity, being host specific.

The result of the alignment analysis of type I and II NLPs is shown in Figure 1. The first sequence represents all type I NLPs and will be called type I NLP consensus, the second sequence represents all type II NLPs and will be called type II NLP consensus. Because these consensus sequences statistically represent type I and II NLPs, we use them to obtain secondary structure predictions and to perform local alignments with proteins with known 3d-structure. Finally, we used the type I NLP consensus to build a 3d-structure. After obtaining type I and II consensus sequences, we submitted them to the PROF program in the PredictProtein site [22] to obtain a secondary (2d) structure prediction (see Figure 1).
We observed that the NLP 2d-structure may be divided into 5 parts or domains: (1) a signal peptide (positions 1–25) with an α-helix; (2) a start domain (positions 25–60) with 2–3 predicted β-strands and a predicted α-helix with low confidence level; (3) a coil flanked by two cysteines, C62 and C89, called c62c89-coil; (4) a β-strand-rich region (positions 90–220), composed by 9–10 β-strands and (5) an end domain (positions 220–270) with two predicted α-helices separated by a β-strand. The predicted central β-strand-rich region in NLPs included them in the all-β SCOP class of proteins.

In this article, we are proposing that the cupin fold is a suitable template for this region (residues 90–220) of the NLPs. The signal peptide is cut away from the sequence and the other regions play a secondary role in the main structure of the NLPs.

According to the literature, the difference between type I and II NLPs are cysteines C106 and C112, present only in type II NLPs [2]. However, we observed other differences, both in the conservation pattern of residues and in the 2d-structure. For example, while histidine H29 and aspartate D30 are conserved among type I NLPs, they are underrepresented in type II NLPs. Also, the conserved sequence DXDG (positions 56–63) and the conserved histidines H179 and H185 in type II NLPs (not present in type I NLPs) is intriguing. Concerning the 2d-structure, the main difference is β-strand 6, not predicted in type I NLPs. In order to investigate which residues are essential in NLPs, in the sense that only them (and no other) could play an specific role (function and structure), and which of them may be substituted by other compatible ones (for example, same charge, hydropathicity, α-helix or β-strand bias, etc.), we count the number of common residues in each position of the alignment and draw them in a succession of histograms. For instance, in type I NLPs the conserved motif GHRHDEG is described by: [g 89% a 7% k 4%] [h 96% y 4%] [r 100%] [h 100%] [d 92% f 4% y 4%] [w 100%] [e 100%], where the letter represent the one letter code and the number is the frequency of the aa at this position. For type II NLPs the sequence of histograms is [g 87% n 13%] [h 100%] [r 74% k 13% t 13%] [h 100%] [d 100%] [w 80% f 3% e 100%]. We can see above that the first glycine may be substituted by alanine in type I NLPs. This means that a small flexible residue in this position fulfills (though glycine is more suitable) the necessary role for the structure and function of the NLPs.
1um1 (2.6), 1x2b (2.6), 1xe (2.6), 2c0h (2.7) and 2hi0 (3.9).

The 1vj2 structure, a protein with unknown function from Thermotoga maritima (a thermophilic Eubacteria with an optimum growth temperature of 80°C) belongs to the RmlC-like cupin SCOP superfamily, and the Mainly Beta CATH class. From the 2d-structure analysis, NLPS were recognized as β-sheet-rich structures [19], possibly belonging to the all-β SCOP class of proteins of which the RmlC-like cupin is a superfamily. 1vj2 presents a compatible number of β-strands with those of NLPS, their position relative to conserved residues is the same and finally its sequence rshpwe is very similar to the pattern GHRHDWE of the NLPS. 1vj2 has four histidines acting as ligands for a manganese ion, what might explain the importance of this motif in the NLPS. The other candidates, 1fsi1, 1lsm and 2fik present the sequence gshrdwvm in the middle of an α-helix and for that they were discarded. Further, all the proteins 1trq, 1um1, 1x2b, 1xe, 2c0h and 2hi0 posses many α-helices intermixed by few β-strands, and were discarded too.

We investigated the degree of conservation of the residues in these two motifs in 68 cups collected in a review by Dunwell [8] and we obtained the following histograms:

\[
\begin{align*}
g & \ xhx \ xhx \ xhx & \ xhx \ xhx \ xhx \\
\text{g} & \ xhx \ xhx \ xhx & \ xhx \ xhx \ xhx \\
\text{g} & \ xhx \ xhx \ xhx & \ xhx \ xhx \ xhx \\
\text{g} & \ xhx \ xhx \ xhx & \ xhx \ xhx \ xhx \\
\text{g} & \ xhx \ xhx \ xhx & \ xhx \ xhx \ xhx \\
\end{align*}
\]

where we can see the typical positioning of the histidines enabling them to act as ion ligands [23]. For the positioning of these motifs in the 2d-structure or relative to the other β-strands, see Figure 2, last line. These two motifs (more exactly, all three histidines) are near each other in the 3d-structure, enabling the 3 histidines (h82%, h65%, h75%) and the glutamate (e53%) to act as metal ligands, what might explain why these residues are highly conserved (see Table 1). Some cups (called 3-residue) have three residues between the second and third ligands while others have four (4-residue cups).

Comparing the residue histograms of NLPS and cups in Table 1, we can see that the sequence hrhdxe is present in most NLPS and in most 3-residue cups. Also, 75% of all cupin sequences and 95% of all NLPS have a histidine (fourth ligand) in the second motif and at position 193, respectively. These correspond to the most important residues in the general cupin pattern and, the substitution of any of these residues will reduce the ability of the protein to hold the metal ion, as is the case in some cups. We concluded that the first motif in the cups with its xhxx [x-le pattern corresponds to the GHRHDWE [gh] pattern of the NLPS and the histidine h75% in the second cupin motif corresponds to H193.95% in the NLPS. Certainly this correspondence must be compatible with the 2d-structure, what we will see next.

The embedding of the sequence GHRHDWE in a β-strand imposes an alternate orientation (inwards and outwards) of the side chains. Furthermore, the hydrophobicity pattern must be compatible with that fact. Highly hydrophobic residues, such as tryptophan (w), extend their hydrophobic side chains toward the interior of the protein, inducing the orientation g-h133-r-t-h135-d-w-e-[gh]-v-v-v-w-t (a down arrow represents sidechain directed toward the interior of the protein and an up arrow the opposite). Histidines H133 and H135 obey this alternate pattern in the NLPS allowing them to act as ligands for metal ions (see Figure 3).

In cups, both positions 138 and 139 (see Table 1) typically contain negatively charged residues, such as aspartate (d) or glutamate (e). However, only E139 acts as a ligand for the metal ion. Therefore, although highly conserved in type I and II NLPS, E138 must be discarded as a viable ligand candidate. H139 could act as an ion ligand in type II NLPS, but only 19% of type I NLPS presents a histidine at this position. Among all 68 cups in [8], only the sequences from Pyrococcus horikoshii and Arabidopsis thaliana have a histidine at this position, ihqdeo3 (GenBank gi 3256432, a hypothetical protein) and ahhhtfgh (gi 1169199, DNA-damage-repair/tolerance protein).

Table 1: Statistics for NLPS and cups.

| Position | 132 | 133 l^4 ligand | 134 | 135 2nd ligand | 136 | 137 | gap | 138 | 139 3rd ligand | 193 4th ligand |
|----------|-----|----------------|-----|----------------|-----|-----|-----|-----|----------------|----------------|
| type I NLPS | g89a7k4 | h9ly4 | r100 | h100 | d2fhy4 | w100 | - | e100 | g237a19 | a11y7 |
| type II NLPS | g87n13 | h100 | r7f13x13 | h100 | d100 | w10137 | - | e100 | h80n20 | h100 |
| 3-residue cups | e8924909 | h8510 | r18h15 | y12q27 | h70d12x18 | d24tx12 | d76y12 | s3x33 | - | e27d15x3 | h80y15 |
| 4-residue cups | p3411x24 | h80q14x6 | - | h80n14x26 | d25t7q26 | 86h11d6 | h71f71 | p2701177q1 | 74h10 | e2b97v95y13 |
| Statistics for 27 type I NLPS, 15 type II NLPS, 33 3-residue cups and 35 4-residue cups. Values are given as percentage of these numbers. x means any other residue.
respectively. The previously obtained T. maritima 1vj2 with its sequence rhshpweh (ligands are italicized) is included in this group, too. It seems that the third histidine confers an increased stability to the binding of the metal ion necessary in the extreme temperature living conditions of P. horikoshii and T. maritima.

The second most frequent residue at position 139, asparagine (n), is found only in two cupins among the 68 in [8]: Arachis hypogaea (gi 1168390a) pkhadad n and Bacillus subtilis (gi 2636534) ahfdayt n. Because of the lack of histidines in positions 133 and 135, these cupins probably do not bind any metal ion.

Asparagine n139 is present in 26% of the NLPs and probably do not participate in the bind of any ion, too. Additionally, the fact that many NLPs (26%) have non-charged residues at position 139 raises the question if a charged residue is necessary at this position. Many cupins (29%) have uncharged residues (v8a7l3cg) at this position showing that these cupins do not need residue 139 at all as an ion ligand. For example, pirin 1j1l (dhphrgfet...) uses three histidines, (h133, h135, and h193) and a glutamate (e195) as ligands for Fe2+, but not e139, though it would be available to perform this function. Moreover, several cupins do not use any 3rd ligand at this position. Examples are isopenicillin N synthase from Aspergillus nidulans, PDB code 1bk0 (sequence whedvslit...h and ion Fe2+); clavaminate synthase 1ds1 (sequence fhtemathr...h and ion Fe2+); hypothetical protein 1jr7 (sequence lnhdgyvee...h, and ion Fe2+) and anthocyanidin synthase from Arabidopsis thaliana 1gp4 (ahdvsaltf...h, Fe3+). Even when present, e139 is not very conserved in cupins for a residue that should bind to a metal ion. Additionally, site-directed mutagenesis e139 → q139 in the cupin acetylacetone dioxygenase Dke1 results in increased loss of the Fe2+ ion and reduced thermal stability [24], but its functional characteristics remain practically unchanged. We conclude that asparagine n139 does not act as a ligand for the metal ion, resulting in 61% of all NLPs with no ligand at this position. Finally, human cysteine dioxygenase 2ic1 (see Figure 2) has just 3 histidine ligands for the Fe2+ ion (h133, h135 and h193). The third histidine in this sequence ihdhdsch...h does not act as a metal ligand and no other residue is necessary to hold the metal ion showing that NLPs could likewise, hold a metal ion at this site.

Figure 2
Sequence alignment of the β-barrel domain of NLPs, ABPs and some cupins. Alignment of the consensus sequence of all 27 type I NLPs (first line), 15 type II NLPs (second line), 32 dicot ABPs (third line), 9 monocot ABPs (fourth line), 1lr5 (maize ABP), 2ic1 (cysteine dioxygenase type 1, capitalized residues represent 100% conservation in 10 different organisms), 1vj2 (hypothetical protein) and the two main cupin motifs (last line). Solid line boxes represent real β-strands, dashed line boxes represent those predicted and dotted line boxes are predicted β-strands with low confidence level. Compatible residues are shown in boldface and those residues present in both type I/II NLPs and any of the other sequences are grey boxed. The first two lines follow the convention of Figure 1.
The Role of Cysteines

Fellbrich et al. [6] have shown that both cysteines C62 and C89 are conserved and necessary for the NLPs’ function. Also, the coil between them seems to encode a glycosylation site, which, for secreted proteins means protection against proteolysis, correct folding and thermal stability. Additionally, the highly conserved glycines G76-G77 seem to promote a fold exactly in the middle of this coil enabling the cysteines to come together.

The analysis of the bonding pattern among cysteines resulted in a 90% confidence level for C62 and C89 to be forming a disulfide bridge in type I and II NLPs. A search for the pattern GnxsGGL in the PDB rendered the protein 1eh6, which has a turn at s75G76, supporting the hypothesis that both cysteines are disulfide bonded.

In relation to the other two cysteines present only in type II NLPs, the program DISULFIND attributes a probability of just 30% for the bonding of C106 to any other cysteine and 0% for C112. However, from the position of these two cysteines, it is not difficult to infer that they are bonded if βA and βB form a β-sheet (see Figure 3). These two cysteines seem to enforce that these β-strands should present this conformation. It is also possible that these two cysteines might be bonded to the two conserved histidines H133 and H134 by a zinc ion, such as in the zinc finger of WRKY proteins. WRKY-proteins have a special zinc-finger motif characterized by the pattern cx4-cx2.23hxh, and type II NLPs have a similar pattern, cx3-cx1.9hxh. WRKY-proteins are transcript factors with up to 100 representatives in A. thaliana [25]. For instance, the protein AtWRKY6 is associated with both senescence- and defense-related processes [26]. The structure of the WRKY proteins may be shared by type II NLPs. However, it is less probable that they share the same function. [27] suggests that NLP-induced necrosis requires interaction with a target site at the extracytoplasmic side of dicot plant plasma membrane. They show that the ectopic expression of NLP in dicot plants resulted in cell death only when the protein was delivered to the apoplast. However, Bae et al. have shown that NEP1 in the plant was localized at the cell wall and cytosol. This result indicates that NEP1 can penetrate through the plasma membrane but may not be able to penetrate organelles [28]. It has been observed that NLPs are hydrophilic and not likely to cross the plasma membrane. Furthermore, our proposed model structure based on the ABP 1lr5 has many hydrophilic residues at the surface and the hydrophobic ones are buried supporting the hypothesis that NLPs are not able to cross the plasma membrane. Additionally, the rapid response of parsley protoplasts (approximately 150 seconds) to PpNPP1 (Phytophthora parasitica NPP1) is compatible with an interaction just at the plasma membrane level [6].

NLP 2d and 3d-Structure

The RmlC-like cupin superfamily belongs to the SCOP double-stranded beta-helix fold. Cupins are double-stranded because they are composed of two sequences of antiparallel strands linked with short turns. If the NLPs are cupins, then there should be a correspondence between the 2d-structures of cupins and those predicted for NLPs. Cupins are formed by 8–10 β-strands called [A]BCDEFGHI [1].
The formation of the \( \beta \)-barrel can be understood in the following way: It starts with E folding over F, then D over G, C over H, and eventually B folds over I:

\[
\begin{align*}
& B \quad C \quad D \quad E \\
& \Rightarrow \Rightarrow \Rightarrow \Rightarrow \\
& I \quad H \quad G \quad F \\
& \Leftarrow \Leftarrow \Leftarrow \Leftarrow
\end{align*}
\]

Finally, this double strand turns like an helix building up a \( \beta \)-barrel of two \( \beta \)-sheets: CHEF and BIDG with their hydrophobic residues aiming at the interior of the barrel. Inside the barrel, in the hydrophobic pocket, we find the metal ion bound to its ligand, next to the top of the barrel (the bottom is closed by the E and F \( \beta \)-strands, see Figure 3). Cupins presenting a catalytic activity bind their substrates on the top of the barrel close to the metal ion at the hydrophobic pocket.

The coil between E and F must be flexible enough to allow the folding of EDCB over FGHI. Glycine, as the most flexible residue, represents an excellent candidate to perform this role and we find two of them in the sequence of the putative EF-coil in the NLPs: g163g164. Moreover, g163g164 are 27 residues away from the 1st and 2nd histidine ligands and 26 residues away from the 4th histidine ligand in the type I NLPs (H133R134H135-x27-g163g164-x26-H193). The final result is that all three histidines are very close in the final structure (see Figure 3), exactly as they should be to act as ion ligands. Additionally, an interesting sequence is the necrotic type I NLP BeNEP2 cphas in the EF-coil, which is flanked by two cysteines. The DiANNA 1.1 disulfide bond prediction program [29] predicts these two cysteines are bonded with 82% confidence level supporting the above predictions for this coil with the E and F \( \beta \)-strands closing the bottom of the barrel. These \( \beta \)-strands and the loop in between form the so called Inter Motif Region (IMR), which contains 12 to 130 residues and showing no conserved pattern in the cupin. This highly variable region in the cupins and the low confidence level of the 2d-structure prediction for the NLPs in this region make difficult any 2d-structure alignment between cupins and NLPs. Figure 4 shows the confi-
dence levels for the 2d-structure prediction using the PROF program for type I and II NLP consensuses, and for the 1lr5 cupin, here we can see the correspondence between individual \( \beta \)-strands among these proteins. We observe that the \( \beta \)-barrel is built up by 7–9 high confidence level \( \beta \)-strands and 1–2 low confidence ones with a correspondence between strands in cupins and NLPs. For instance, \( \beta_{1} \) in type I NLP consensus corresponds to \( \beta_{3} \) in cupins, \( \beta_{2} \) to \( \beta_{6} \), \( \beta_{3} \) to \( \beta_{10} \), \( \beta_{3} \) to \( \beta_{12} \), \( \beta_{7} \) to \( \beta_{6} \), \( \beta_{9} \) to \( \beta_{6} \), \( \beta_{C} \), \( \beta_{A} \), and finally \( \beta_{10} \) corresponds to \( \beta_{6} \). The most conserved pattern in NLPs, the GHRHDWE sequence (\( \eta \) in Figure 4), is between the putative C (\( \beta_{1} \)) and D (\( \beta_{10} \)) \( \beta \)-strands. From the position of this pattern, despite the fact that C is a low confidence strand, its position can be easily determined. This correspondence is confirmed by the alignment of the \( \beta \)-barrels of some representative NLP and cupin sequences (see Figure 2). These are the consensus of 27 type I and 15 type II NLPs, 32 dicot ABPs (all cupins), 9 monocot ABPs (all cupins), and three other cupins discussed in this work: 1lr5 (ABP1), 2ic1 and 1uj2.

Two differences between the predictions obtained for type I and II NLP consensuses and the cupin structure are worth mentioning: first, \( \beta_{6} \) is present in type II NLPs but not in type I NLPs, and second, NLPs do not have \( \beta_{6} \). Certainly, the correspondence of \( \beta_{9} \) to \( \beta_{6} \) is a tempting assumption in type II NLPs, but this would not be compatible with the extremely good alignment between type I and II NLPs and with the alignment shown in Figure 2. We could argue that the corresponding \( \beta \)-strand was just missed by the PROF program in type I NLP consensus and that \( \beta_{9} \) and not \( \beta_{6} \) should correspond to \( \beta_{6} \) in type I NLPs. Contrary to this idea, we propose that the conserved histidine H193 is a metal ion ligand. In cupins, since the 4th ligand is in \( \beta_{4} \), we propose H193 signs the position of the H \( \beta \)-strand in NLPs. More precisely, the 4th ligand (histidine) must be at the border of the \( \beta_{4} \) because the 1st and 2nd ligands are at the border of \( \beta_{1} \) (in the CD-coil) and C and H \( \beta \)-strands form an antiparallel \( \beta \)-sheet, as can be visualized by the following design (boxes represent \( \beta \)-strands).

\[
\begin{array}{ccc}
\beta_{B} & \beta_{C} & \beta_{D} \\
\Rightarrow & \text{tfglahnve} & \Rightarrow \\
\text{…} & \text{…} & \text{…} \\
\beta_{1} & \beta_{H} & \beta_{C}
\end{array}
\]

Lastly, we could argue that the 3d-structure of type II NLPs includes 10 \( \beta \)-strands and not 9 as in type I NLPs and that \( \beta_{9} \) is \( \beta_{10} \) with \( H179 \) (at the border of \( \beta_{9} \)) being the residue acting as 4th ligand. Besides the conserved histidines of type I NLPs, type II NLPs have two additional ones: \( H179 \) and \( H185 \), which could act as ligands. First, the good alignment of type I NLPs and type II NLPs points to a common structure, second, most NLPs are type I, third, they include the most aggressive NLPs (necrotic ones), and fourth, type II NLPs do not occur in oomycetes (see Table 2). Therefore, type I NLPs represents the class of the NLPs and type II NLPs should be treated as an important but secondary source of information about the NLP structure.

**The Inter Motif Region (IMR)**

The previous analysis about the 2d-structure of NLPs and cupins shows that predictions for the region delimited by \( \beta_{6} \) and \( \beta_{8} \) (putative E and F \( \beta \)-strands), which corresponds to the low conserved IMR in cupins, is a difficult task. It contains 22–32 (22–28 in type II NLP consensus) residues with 11 (11 in type II NLP consensus) residues in the coil of the IMR in type I NLPs and also has the conserved sequence S88%a50%H95%g74%. Therefore, we have chosen among the 68 cupins in [8] those that are similar in size. The most similar cupin to the NLPs’ IMR is A. thaliana gi|461453, a possible receptor for the hormone auxin. The 3d-structure of maize Auxin Binding Protein (ABP) has been already determined (1lrh and 1lr5 in PDB, see Figure 2). It has one \( \beta \)-barrel domain, is a dimer in solution, has 21 residues in the IMR, and 11 in coil. ABPs are involved in cell expansion and are located in the ER lumen, the plasma membrane, and the cell wall [30]. It is ubiquitous amongst green plants [31].

It would be advantageous for the necrosis proteins to have control of the auxin-response in the host, for example changes in protoplast electrophysiology. Auxin induces \( H^{+} \) secretion into the cell wall causing hyperpolarization of the plasma membrane in *Avena* coleoptile cells [32], electrical response in tobacco protoplasts [33], and K⁺ currents in *Nicotiana tabacum* guard cells [34].

Auxin stimulates the growth of plant cells by regulating the activity of a \( H^{+} \)-ATPase in the plasma membrane. Proton secretion by this transport enzyme acidifies the cell walls increasing their extensibility. The internal hydrostatic pressure of the cell then extends the walls. In the interaction between *M. perniciosa* and *T. cacao*, it has been suggested an auxin-inducing phase that causes malformation and an auxin-depleting secondary phase that kills the host [35]. Kilaru et al. reported that the increase in auxin coincided with the phase transition of the fungus. This increase could be the effect of enhanced IAA (auxin) synthesis, suppression of IAA oxidase or secretion of IAA by the pathogen [36]. The possibility that the 3d-structure of NLPs is similar to that of the cupins and, in particular to ABPs, places the NLPs in control of plant auxin receptors and of the resulting ionic currents. Additionally, Haberlach et al. have shown that the balance of cytokinin and auxin was an important factor in maintaining or eliminat-
ing resistance in plant tissues. More specifically, they show that *P. parasitica* resistant *N. tabacum* became susceptible under high cytokinin/auxin levels [37]. It is conceivable that NLPs could compete with the natural ABPs resulting in an apparent increase of the cytokinin/auxin ratio in the plant. Other possibility is that NLPs could be auxin oxidases. Krupasager reported that in the dikaryotic stage, *Marasmius perniciosa* produces no significant amount of cytokinins or auxins but auxin-inactivating enzymes such as IAA-oxidase and laccase [38].

Differences between monocot and dicot ABPs (see Table 3) provide an additional important clue because *Phytophthora* species are primarily pathogens of dicotyledonous plants. Monocots are apparently not affected by NLPs [6]. For example, maize and barley do not show any cell death symptoms after infiltration with PpNPP1. For this reason, we investigated which are the differences between monocot and dicot ABPs [31], and if there is a relationship between these differences with conserved residues in the NLPs. These residues are:

| monocot ABPs | d | m | i | s | p | s |
| dicot ABPs | n | l | r | h | v | t |
| type I NLP | n73 | L78 | k90 | H162 | v191 | t200 |
| type II NLP | t73 | L78 | R90 | H162 | g191 | g200 |

We observe a high degree of correspondence between dicot ABP residues and NLP high conserved residues (n73, L78, H162, v191 and t200 in type I NLPs and L78, R90 and H162 in type II NLPs). Certainly, the histidine residue H162 in the EF-coil is the most striking difference between monocot and dicot ABPs shared by NLPs. The alignment between 32 dicot ABPs and type I NLPs in the EF-coil results in some common residues (see the sequences below) or residues with similar physicochemical characteristics (see figure 2 for the whole alignment):

| type I NLPs | βE | aSaHgykkyt | βF | ... |
| dicot ABPs | βE | aSshgkpgkp | βF | ... |

For a more detailed analysis, Table 4 shows for the residues of 9 ABP sequences of monocots and 32 ABP sequences of dicots in the EF-coil.

All 32 investigated dicot ABPs and 95% of all NLPs have a conserved histidine H162 at the EF-coil. The two NLP sequences which do not have it exactly at this position, but 4 positions downstream (PiNPP1.2 and PiNPP1.3), are inactive forms; probably originated by gene duplication of the most aggressive PiNPP1 from *Phytophthora infestans*. These two are expressed both in the biotrophic as in the necrotrophic phases, while PiNPP1 is expressed only in the latter phase [17].

Fellbrich [6] has shown that the last 8 residues in PiNPP1 may be deleted without loss of activity but not the last 20 residues (... ntdFGd &x25CA; AnVpmkdGnFlt &x25CA; kvngayya). The sequence AnVpmkdGnFlt coincides 100% with the conserved residues in necrotic type I NLPs GxAnxP. It is interesting to observe that the C-terminal seems to be important for NLPs as well as for ABPs. The synthetic peptide from the C-terminal of *Zea mays* ABP wdedcfak, the 15-residue *N. tabacum* Nt-abp1 C-terminal peptide (ywdecyqtwkdel) and the Nt-abp1 itself have been shown to induce hyperpolarization [39]. However, contrary to the effect of auxins, NLPs cause depolarization, alkalization of the surrounding media and K-efflux [40].

The histograms of 32 dicot ABPs in the C-terminal resulted in yWDEqCygxxKDEL. Although conserved, mutagenesis experiments have shown that the sequence kdel is not important for the activity of ABPs and may be deleted and it is related to the two negative residues DE. Since NLPs do not have such a sequence, it is conceivable that NLPs compete with ABPs causing the previous discussed effects.

Besides ABPs, another candidate similar in terms of IMR size is *Bacillus subtilis* gi|2635598, a hypothetical protein with no determined structure similar to the human cysteine dioxygenase 2ic1. It is a monomer with an IMR size of 23 residues and a coil of 11 residues. [41] has aligned 10 cysteine dioxygenases of different organisms and the 100% conserved and functionally important residues are capitalized in Figure 2 for the sake of comparison with those in NLPs. We observe that some of these residues are among the most conserved in the necrotic type I NLPs: For instance, Y101 and R103 in βN, H133, H135 and H193 (as ion ligands) and D136.

**Glycosylation**

Although an immunoassay study used for the detection of sugars in glycoconjugates did not reveal a carbohydrate moiety in PpNPP1 [6], glycosylation sites nxs are present in 64% of all necrotic type I NLPs (x = tN) in the c62c89-coil. The glycosylation occurs at asparagine (n) residues in the so called nx(st) sequon and the efficiency of this process depends on the residue x.

Glycosylation is important in most cell-surface and secreted proteins and is often critical for the interaction with other subunits at the cell surface (recognition), protection against proteolytic attack, protein solubility and thermostability. For instance, *P. infestans* has evolved an arsenal of protease inhibitors to overcome the action of plant proteases [42].
Table 3: Differences between monocot and dicot ABPs and NLPs. Differences between monocot and dicot ABPs and NLPs (shown in boldface).

| Protein          | C62c89-co | β-barrel |
|------------------|-----------|----------|
| Type I NLP consensus | αβαββ Cβ tns L Ck | βαα βαβ αβ βαβ |
| PpAAK19753       | αβαββ cb cβ tns l ck | βαα βαβ αβ βαβ |
| MpNEP1           | ααββ cb cb nts l ck | βαα βαβ αβ βαβ |
| Type II NLP consensus | αβαββ Cβ tln l Cr | βαα βαβ αβ βαβ |
| Dicot ABP        | αβαββ cis α l βαβ | βαα βαβ αβ βαβ |
| Monocot ABP      | αβαββ cis α m βαβ | βαα βαβ αβ βαβ |

The dicot glycosylation site nis sequence is not present in the monocot sequences (dis). nts, nst and nws are possible NLP glycosylation sites.

The probability of glycosylation (see section Methods) of the sequon ntsg varies between 44 and 62%. For instance, PinNPP1.1 is 44%, PsqNIP 50%, PpNPP1 54% and BeNEP1 62%. Furthermore, MpNEP1 and MpNEP2 are predicted not to be glycosylated because of the bulky tryptophan (w) in their sequon (nws). His VdNEP and BeNEP2 have no sequon.

Additionally, an important difference between monocot and dicot ABPs [31] pointing out that ABPs and NLPs share the same 3d-structure is that dicot ABPs have a glycosylation site nis next to the beginning of the protein, that is not present in monocot ABPs (dis). This site is also present in several type I NLPs (see Table 3).

Conclusion

The 3d-structure of the NLPs remains to be determined experimentally. However, in this paper we presented several evidences indicating that they belong to the Cupin superfamily. Using a cupin template and the type I NLP consensus we were able to calculate a prediction for the 3d-structure of the β-strand rich portion (positions 90–220 in Figure 1) which presented stability for 3nS under molecular dynamics. This structure presents the classical signature of a cupin protein. Furthermore, the prediction of the structure of the upstream coil bordered by two cysteines remains to be addressed. Cysteines in the upstream coil of type I NLPs are disulfide bonded, simplifying the problem by removing this sequence from the analysis. However, the right positioning of this free coil becomes a new problem, which is more complex than the first one. Is it free to move around or does it make part of the β-sheets? Its glycosylation site points to some role in anchoring the protein on the cell surface.

Several 2d-structure predictions software agree with a central β-strand rich portion anked by α-helices making highly probable that the central part of the protein belongs to the all-β SCOP Class.

The conserved pattern of cysteines points to two different NLP types: type I NLPs containing two cysteines upstream of the β-barrel and type II NLPs, containing two additional cysteines in the left border of the β-barrel. Predictions show that the first two are bonded together, but not the other two. The similarity between the pattern of cysteines in the WRKY proteins and the type II NLPs is so high, that would not be surprising if the cysteines participate in metal binding in a zinc-finger conformation. Also, WRKY proteins participate in the pathogen detection system of the plant, an interesting "coincidence" for a protein involved in phytopathogenic activities. Unfortunately, we have only part of the WRKY protein structure, the zinc-finger part (C-terminal).

Figure 1 shows clearly the biological role played by the residues of the GHRHDWE motif holding a putative ion. For this purpose, a necessary 3d-histidine is found downstream in the structure (H193) in the exact position according to the cupin structure and is also present in 95% of the NLPs analyzed. Only two NLPs do not have it, and they do not present necrotic activity. The relative position and number of histidines in the structure point to a metal ion containing protein. In addition, the exact metal ion remains to be determined but our study points to manganese or zinc.

The Cupin superfamily is very large with diverse functions. Many of these functions depend on a glutamate residue, which does not seem to be present in NLPs. When this residue is present, cupins are involved in enzymatic activities such as oxalate oxidase, decarboxylase, dioxygenases, etc. It remains to be determined experimentally if NLPs have some catalytic activity involving oxalate. In any case, the [Ca2+]nmax levels, ROS (H2O2) and oxalate are all intermixed in pathogen defense and sensing. Also, lignin processing is highly dependent on oxidases and peroxidases (cupins). It has been shown that germins function as oxalate oxidases (conversion of oxalate to CO2 and
H₂O₂ and superoxide dismutase (2O₂⁻ + 2H⁺ → H₂O₂ + O₂) [43]. Any interference in such activities would be advantageous for the fungus because of the correlation between H₂O₂ and [Ca²⁺]cyt. Even with no catalytic activity, the fold is resistant to oxidation, a characteristic necessary for oxidases, decarboxylases and peroxidases.

A related class of cupins is the auxin binding proteins, which do not show catalytic activity but work as signal transducers in plant cells. They have many common structural features and conserved residues in relation to the NLPs. In this respect, the most remarkable feature is the conservation of histidine H162, present in 95% of the NLPs and in almost all dicot ABPs and is related to the fact that NLPs attack only dicot plants. Also, the way ABPs transmit the information to the cell is intimately related to conservation of histidine H162, present in 95% of the NLPs. In this respect, the most remarkable feature is the structural features and conserved residues in relation to the transducers in plant cells. They have many common structural features and conserved residues in relation to the NLPs.

In order to obtain a non-redundant and representative set of NLPs, all sequences analyzed have a pairwise distance greater than 10%. The organisms and sequences include 42 NLPs (see Table 2).

### Methods

In order to obtain a non-redundant and representative set of NLPs, all sequences analyzed have a pairwise distance greater than 10%. The organisms and sequences include 42 NLPs (see Table 2).

#### Table 4: Residue histograms for NLPs and ABPs.

| Position | Type I NLPs | Type II NLPs | Dicot ABPs | Monocot ABPs |
|----------|-------------|--------------|------------|--------------|
| 158      | a11s41c18   | s100         | e100       | p101m11      |
| 159      | a8s20q7p13c7y4 | s6s4     | s170       | g100         |
| 160      | a22s6q17c7y4 | s160       | e100       | p101m11      |
| 161      | h100        | g100         | s170       | g100         |
| 162      | h100        | g100         | s170       | g100         |
| 163      | h100        | g100         | s170       | g100         |

Residue histograms for the 27 type I and 15 type II NLPs, 32 dicot ABP sequences and 9 monocot ABP sequences. Each column represents possible residues at that position. Values are given as percentage of these numbers. x means any other residue.
c62c89-coil, we submitted all type I NLP sequences to the NetNGlyc 1.0 Server [51].

**Molecular 3d-structure**

We used ClustalW with default parameters and gapopen = 0.0 to align the sequences shown in the Figure 2. Note that for the alignment, we used only the β-strand-rich region (β-barrel domain), positions 90–220. The 3d-structures were constructed with the molecular modelling program spdbv [52] using as template the monomer of the 1lr5 protein. First, we used Rasmol to cut the β-rich region of the 1lr5 protein. Then, we mutated the residues in the 1lr5 sequence to those residues in the type I NLP consensus, introduced the necessary residues where the alignment resulted in gaps for the 1lr5 sequence and deleted the residues in the cases the alignment produced gaps in the type I NLP consensus. After each step, we executed a minimization to fit the new residues to the old structure. We took the care to introduce and delete residues only in the coil regions of the protein. The resulting structure was minimized with GROMACS with a steepest descent algorithm.

Finally, we formed a simulated annealing from 0 to 30 K during 100 ns in vacuum. The force field used was opls-aa/l. Finally, we simulated a annealing of 100 ps from 0 to 10 K and 3 nS in vacuum. The force field used was opls-aa/l. Finally, we simulated a annealing of 100 ps from 0 to 10 K and 3 nS from 10 K to 273 K with molecular dynamics in explicit solvent (spc water model) was performed.

**Authors' contributions**

GAGP and OGC provided the protein sequences and the main motivation for the work. ALC, MS and JCMM have conceived the study relating the NLP structure and key residues to those of the cupin superfamily. ALC and MS performed the alignments and the analysis of the conserved sequences. ALC performed the 2d-structure predictions and alignments. ALC and NL calculated the proposed 3d-structure. SE and MS participated in the proposal of the NLP function. NL, JCMM, and MS made the major revisions of the work. All authors contributed to write the manuscript.

**Acknowledgements**

The authors thank Francisco Javier Medrano Martin for reviewing the manuscript. N. Lemke also would like to thank A. K. Brau for useful discussions and suggestions. Work partially supported by CNPq (research grants 506411/2004-3 and 474278/2006-9) and FAPESP (research grant 2007/ 02827-9).

**References**

1. Bailey BA: Purification of a protein from culture filtrates of *Fusarium oxysporum* that induces ethylene and necrosis in leaves of *Erythroxylum coca*. *Physiopathology* 1995, 85:1250-1255.

2. Gijzen M, Nürnberger T: *Nep*I-like proteins from plants pathogens: Recruitment and diversification of the NEP1 Domain. *Acad Trans. Phytopathology* 2006, 16(67):1800-1807.

3. Tyler BM, Tripathy S, Zhang X, Dehul P, Jiang RH, Aerts A, Arre-dondo FD, Baxter L, Benzasson D, Beynon JL, Chapman J, Damasceno CMB, Dorrance AE, Dou D, Dickerman AW, Dubchak IL, Galbetloto M, Gijzen M, Gordon SC, Gover F, Grunwald NJ, Huang W, Ivors KL, Jones RW, Kamoun S, Krampis K, Lamour KH, Lee MK, McDonald WH, Medina M, Meijer HG, Nordberg EK, Maclean DJ, Ospina-Giraldo MD, Morris PF, Phumthum V, Putnam NH, Rash S, Rose JKC, Salamun Y, Salamov A, Savador A, Scheuring CF, Smith BM, Sobral BWS, Terry A, Torto-Alabia TA, Win J, Xu Z, Zhang H, Grigoriev IV, Rokhsar DS, Boone JL: *Phytophthora* genome sequences uncover evolutionary origins and mechanisms of pathogene-

BMC Plant Biology 2008, 8:50  
http://www.biomedcentral.com/1471-2229/8/50

(page number not for citation purposes)
21. Jores J, Appel B, Lewin A: Cloning and molecular characterization of a unique hemolysin gene of Vibrio pomeruensis sp. nov.: development of a DNA probe for the detection of the hemolysin gene and its use in identification of related Vibrio spp. from the Baltic Sea. *FEMS Microbiol Lett* 2003, 229(2):223-229.

22. Rost B, Tschach G, Liu J: The PredictProtein Server. *Nucleic Acids Res* 2004, 32:321-326.

23. Dunwell JM, Purvis A, Khuri S: Cupins: the most functionally diverse protein superfamily? *Phytochemistry* 2004, 65:7-17.

24. Straaganza GD, Eggerta S, Aquinob G, D'Auriac S, Nidetzkya B: Exploring the cupin-type metal-coordinating signature of acetylacetone dioxygenase Dkel with site-directed mutagenesis: Catalytic reaction profile and Fe(II) binding stability of Glu-69 Gin mutant. *Journal of Molecular Catalysis B, Enzymatic* 2006, 39:171-178.

25. Eulgem T, Rusthon P, Robatzek S, Somssich IE: The WRKY superfamily of plant transcription factors. *Trends Plant Sci* 2000, 5(5):1360-1365.

26. Robatzek S, Somssich IE: A new member of the Arabidopsis WRKY transcription factor family, AtWRKY6, is associated with both senescence- and defence-related processes. *Plant J* 2000, 28(2):122-133.

27. Qutob D, Kemmerlinga S, Brunner F, Kunker W, Engelhardt S, Gust AA, Luberbicki B, Seitz HU, Stahl D, Rauhut T, Glawischning E, Sneewa G, Lacombe B, Watamab N, Lam E, Schlichthing R, Scheel D, Nau K, Dott G, Hubert D, Gippert M, Nurnberger T: Phytotoxicity and Insect Immune Responses Induced by Nep1-Like Proteins. *Plant Cell* 2006, 18(12):3771-3784.

28. Bae H, Kim MS, Sicher RC, Bae HJ, Bailey BA: Necrosis-and ethylene-inducing peptide from *Fusarium oxysporum* induces a complex cascade of transcripts associated with signal transduction and cell death in Arabidopsis. *Plant Physiol* 2006, 141:1056-1067.

29. Ferre F, Clote P: DIANNA: a web server for disulfide connectivity prediction. *Nucleic Acids Res* 2005, 33:230-232.

30. Jones AM, Herman EP: KDEL-containing auxin-binding protein is secreted to the plasma membrane and cell wall. *Plant Physiol* 1993, 101(2):595-606.

31. Woom E, Marshall J, Bauly J, Chen J, Venis M, Napier RM, Pickersgill RW: Crystal structure of auxin-binding protein 1 in complex with auxin. *EMBO J* 2002, 21:2877-2885.

32. Cleland RE, Prins HBA, Harper JR, Higginbotham N: Rapid hormone-induced hyperpolarization of the oat coleoptile membrane potential. *Plant Physiol* 1977, 59:395-397.

33. Warwicker J: Modelling of auxin-binding protein 1 suggests that its C-terminus and auxin could compete for a binding site that incorporates a metal ion and tryptophan residue 44. *Planta* 2001, 212:343-347.

34. Bauly JM, Sealy IM, Macdonald H, Brearley J, Dröge S, Hillmer S, Robinsion DG, Venis MA, Blatt MR, Lazarus CM, Napier RM: Overexpression of auxin-binding protein enhances the sensitivity of guard cells to auxin. *Plant Physiol* 2000, 124:1229-1238.

35. Evans HC: Pleomorphism in *Crinipellis perniciosa*, causal agent ofitches' broom disease of cocoa. *Trans Br Mycol Soc* 1980, 74:515-523.

36. Klaufr A, Bailey BA, Hasenstein KH: *Moniliphthora perniciosa* produces hormones and alters endogenous auxin and salicylic acid in infected cocoa leaves. *FEMS Microbiol Lett* 2007, 274:238-244.

37. Haberlach GT, Budde AD, Sequeira L, Helgeson JP: Modification of disease resistance of tobacco callus tissues by cytokinins. *Plant Physiology* 1978, 62:522-525.

38. Krupasager V, Sequeira L: Auxin destruction by *Marasmius perniciosus*. *Am J Bot* 1969, 56:390-397.

39. David K, Carnero-Diaz E, Leblanc N, Monetzi M, Groschlude J, Perrot-Rechenmann C: Conformational dynamics underlie the activity of the auxin-binding protein, Nt-abp1. *J Biol Chem* 2001, 276(37):34517-34523.

40. Jennings JC, Apel-Birkhold PC, Mock NM, C.J. Baker JDA, Bailey BA: Induction of defense responses in tobacco by the protein Nep1 from *Fusarium oxysporum*. *Plant Sci* 2001, 161:891-899.

41. Simmons CR, Liu Q, Huang Q, Hao Q, Blegly TP, Karplus PA, Stapanik MH: Crystal structure of mammalian cytochrome c oxidase: A novel mononuclear iron center for cytochrome thiol oxidation. *J Biol Chem* 2006, 281(27):18723-18733.

42. Tian M, Win J, Hoorn R, Knapp E, Kamoun S: A *Phytophthora infestans* cystatin-like protein targets a novel tomato papain-like aprotinin-like aprotinin-like proteinase. *Plant Physiol* 2007, 143:364-377.

43. Woon EJ, Dunwell JM, Goodenough PW, Marvier AC, Pickersgill RW: Germin is a manganese containing homohexamer with oxalate oxidase and superoxide dismutase activities. *Nat Struct Bio* 2000, 7(11):1036-1040.

44. Thompson JD, Higgins DG, Gibson TJ: CLUSTALW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 1994, 22:4673-4680.

45. Type I NLPs alignment [http://adelmo.cechin.googlepages.com/typeI.html]

46. Type II NLPs alignment [http://adelmo.cechin.googlepages.com/typell2.html]

47. Necrotic sequences alignment [http://adelmo.cechin.googlepages.com/necrotic alignment.html]

48. RCSI Protein Data Bank [http://www.rcsb.org]

49. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE: The Protein Data Bank. *Nucleic Acids Res* 2000, 28:235-242.

50. Cerioni A, Passerini A, Vullo A, Frasconi P: DISULFIND: a disulfide bonding state and cysteine connectivity prediction server. *Nucleic Acids Res* 2006, 34:177-181.

51. NetNGlyc 1.0 Server [http://www.cbs.dtu.dk/services/NetNGlyc]

52. Guex N, Peitsch M: SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling. *Electrophoresis* 1997, 18:14-21.