

Research article

Open Access

Cupin: A candidate molecular structure for the NepI-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.ibi.unicamp.br; Gonçalo AG Pereira - goncalo@unicamp.br; José CM Mombach - jcmombach@smail.ufsm.br

* Corresponding author

Published: 30 April 2008

Received: 20 November 2007

BMC Plant Biology 2008, 8:50 doi:10.1186/1471-2229-8-50

Accepted: 30 April 2008

This article is available from: <http://www.biomedcentral.com/1471-2229/8/50>

© 2008 Cechin et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Background: NEPI-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 GHRHDWE 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 $g(x)_5hxh(x)_{3,4}e(x)_6g$ and $g(x)_5pxg(x)_2h(x)_3n$, 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 Nürnberg 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).

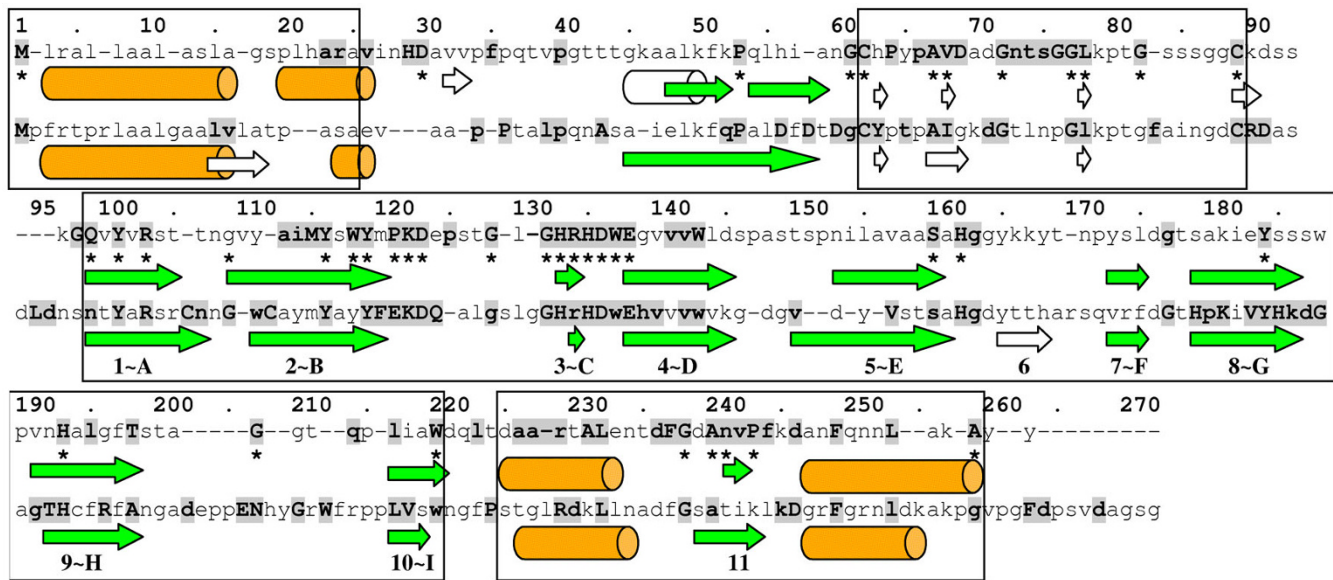


Figure 1
Consensus sequence and secondary structure prediction of NLPs. Consensus of all 27 type I NLPs (upper line) and 15 type II NLPs (lower line). Residues present in more than 85% of all sequences are in boldface and capitalized, residues present in more than 70% are in boldface and other residues are in more than 50% of all sequences. Cylinders represent α -helices and arrows β -strands. The 2d-structure predictions shown have a level of confidence greater than 33%. White cylinders and arrows represent low confidence level 2d-structures. Asterisks denote invariant residues (100% conserved) in type I NLPs with experimentally verified necrotic activity.

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 DxDxDgCY (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 a specific role (function and structure), and which of them may be substituted by other compatible ones (for example, same charge, hydrophobicity, 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 GHRHDWE 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^{13%}l^{7%}] [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.

We submitted all sequences in positions 132–138 and 132–139 in all the 42 NLPs to the search service of the Protein Data Bank (PDB) and obtained the following list of candidates ordered by e-value (see Methods): 1vj2 (e-value = 1.0), 1f51 (2.2), 1ixm (2.3), 2ftk (2.3), 1qtr (2.6),

1wm1 (2.6), 1x2b (2.6), 1x2e (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 rhshpwe is very similar to the pattern GHRHDWE of the NLPs. 1vj2 has four histidines acting as ligands for a manganese ion, what would explain the importance of this motif in the NLPs. The other candidates, 1f51, 1ixm and 2ftk present the sequence ghrhdwm in the middle of an α -helix and for that they were discarded. Further, all the proteins 1qtr, 1wm1, 1x2b, 1x2e, 2c0h and 2hi0 possess 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 cupins collected in a review by Dunwell [8] and we obtained the following histograms:

$$g^{56\%}xxxxh^{82\%}xh^{65\%}xxx[x-e]^{53\%}xxxxxg^{97\%}$$

$$g^{81\%}xxxxp^{90\%}xg^{68\%}xxh^{75\%}xxxn^{47\%}$$

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 (h^{82%}, h^{65%}, h^{75%}) and the glutamate (e^{53%}) to act as metal ligands, what might explain why these residues are highly conserved (see Table 1). Some cupins (called 3-residue) have three residues between the second and third ligands while others have four (4-residue cupins).

Comparing the residue histograms of NLPs and cupins in Table 1, we can see that the sequence hrhdxe is present in most NLPs and in most 3-residue cupins. 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 cupins. We concluded that the first motif in the cupins with its xhx-hxxx [x-e] pattern corresponds to the GHRHDWE [gh] pattern of the NLPs and the histidine h^{75%} 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 the protein, inducing the orientation g-h133-r[↑]-h135-d[↑]-w[↓]-e[↑]-[gh]-v-v-v-w[↓] (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 cupins, 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 cupins in [8], only the sequences from *Pyrococcus horikoshii* and *Arabidopsis thaliana* have a histidine at this position, ihqhdweh (GenBank gi 3256432, a hypothetical protein) and ahhtfg (gi 1169199, DNA-damage-repair/toleration protein),

Table 1: Statistics for NLPs and cupins.

Position	132	133 1 st lig- and	134	135 2 nd ligand	136	137	gap	138	139 3 rd ligand	193 4 th ligand
type I NLPs	g ⁸⁹ a ⁷ k ⁴	h ⁹⁶ y ⁴	r ¹⁰⁰	h ¹⁰⁰	d ⁹² f ⁴ y ⁴	w ¹⁰⁰	-	e ¹⁰⁰	g ³³ n ³⁰ h ¹⁹ a ¹¹ y ⁷	h ⁹² a ⁴ p ⁴
type II NLPs	g ⁸⁹ n ¹³	h ¹⁰⁰	r ⁷⁴ k ¹³ t ¹³	h ¹⁰⁰	d ¹⁰⁰	w ⁸⁰ f ¹³ j ⁷	-	e ¹⁰⁰	h ⁸⁰ n ²⁰	h ¹⁰⁰
3-residue cupins	e ³⁰ p ²⁴ a ⁹ i ⁹ x ²⁰	h ⁸⁵ q ⁶ x ⁹	r ¹⁸ i ¹⁸ h ¹⁵ y ¹² q ⁹ x ²⁷	h ⁷⁰ d ¹² x ¹⁸	d ²⁴ t ¹⁸ e ¹² p ⁹ x ³⁶	d ²⁷ a ¹⁸ y ¹² s ⁹ x ³³	-	e ²⁷ d ²⁴ a ¹⁵ x ³ 3	e ³⁹ a ¹⁸ v ¹⁵ n ⁶ h ⁶ x ¹⁵	h ⁷⁶ f ⁹ m ⁶ y ³ s ³
4-residue cupins	p ³⁴ i ³¹ x ³⁴	h ⁸⁰ q ¹⁴ x ⁶	y ³¹ w ¹⁷ i ⁹ k ⁹ r ⁹ x ²⁵	h ⁶⁰ n ¹⁴ x ²⁶	p ²³ s ¹⁷ q ⁹ x ⁵¹	h ¹⁷ n ¹¹ r ¹¹ d ⁶ q ⁹ x ⁴³	a ²⁹ r ¹⁷ s ¹⁷ q ¹ i ⁹ x ¹⁷	d ²³ t ¹⁷ s ¹¹ a ⁹ e ⁹ x ³¹	e ⁶¹ k ⁹ v ⁹ i ⁶ a ³ g ³ q ³ t ³	h ⁷⁴ f ⁹ v ⁹ q ⁶ m ³

Statistics for 27 type I NLPs, 15 type II NLPs, 33 3-residue cupins and 35 4-residue cupins. Values are given as percentage of these numbers. x means any other residue.

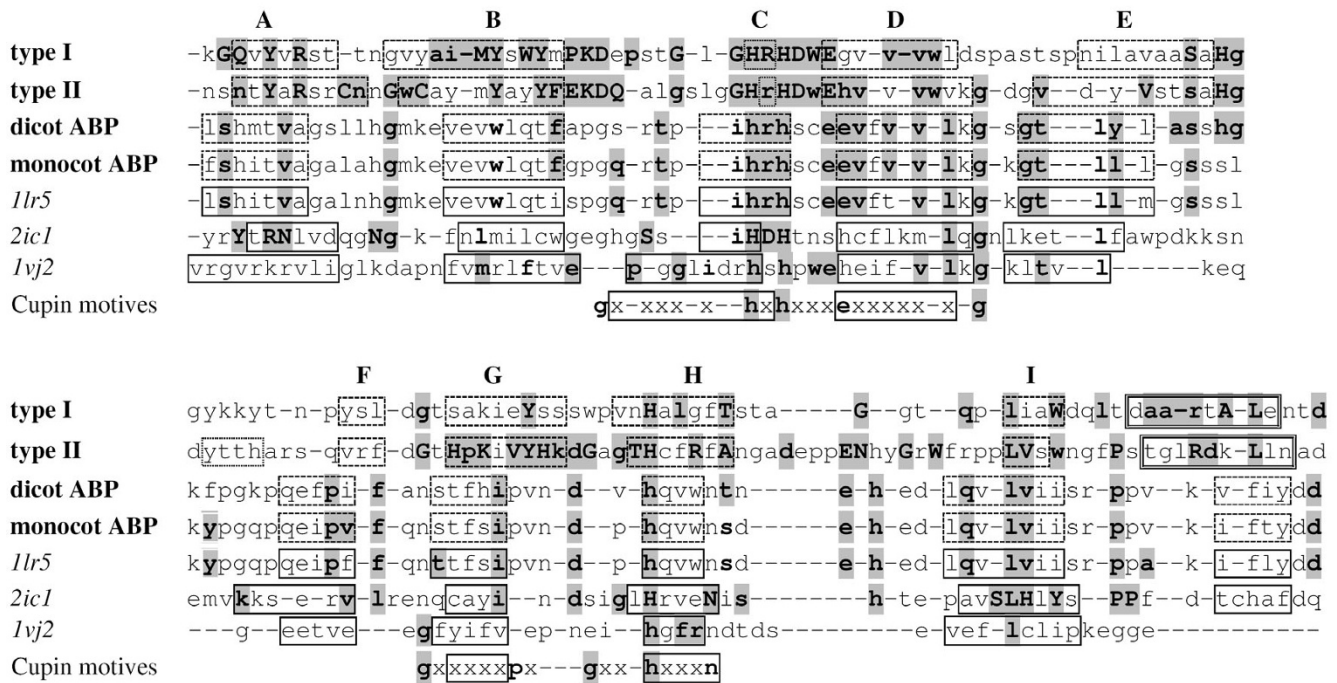


Figure 2
Sequence alignment of the β -barrel domain of NLPs, ABPs and some cupins. Alignment of the consensus sequence of all 27 type I NLPs (first line), 15 type II NLPs (second line), 32 dicot ABPs (third line), 9 monocot ABPs (fourth line), *Ilr5* (maize ABP), *2ic1* (cysteine dioxygenase type I, 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.

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) pkhadadn and *Bacillus subtilis* (gi 2636534) ahfdaytn. 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 (v^{8a}l^{3cg}) at this position showing that these cupins do not need residue 139 at all as an ion ligand. For example, pirin *1j11* (*dhpgrgfet...hae*) 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 *wheadvslit...h* and ion Fe³⁺); clavamate synthase *1ds1* (sequence *fhtemathr...h* and ion Fe²⁺); hypothetical protein *1jr7* (sequence *lhndgtyvee...h*, and ion Fe²⁺) and anthocyanidin synthase from *Arabidopsis thaliana 1gp4* (*ahtdvsaltf...h*, Fe³⁺). Even when present, e139^{53%} 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 acetylaceton 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 *ihdhtdshc...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.

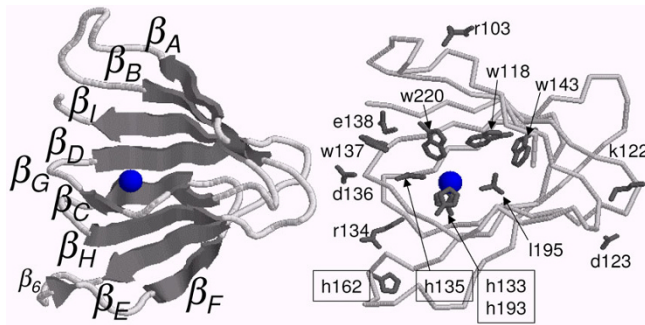


Figure 3
Three-dimensional structure prediction for the type I NLPs. Representation of the two β -sheets CHEF and ABIDG (left side) and the relative position of some of the conserved residues in the 3d-structure (right side). The sphere in the middle of the structure represents the putative metal ion.

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 G76G77 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 β_A and β_B 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 $\alpha_{4,5}\alpha_{22,23}\text{hxx}$, and type II NLPs have a similar pattern, $\alpha_4\alpha_{19}\text{hxx}$. 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

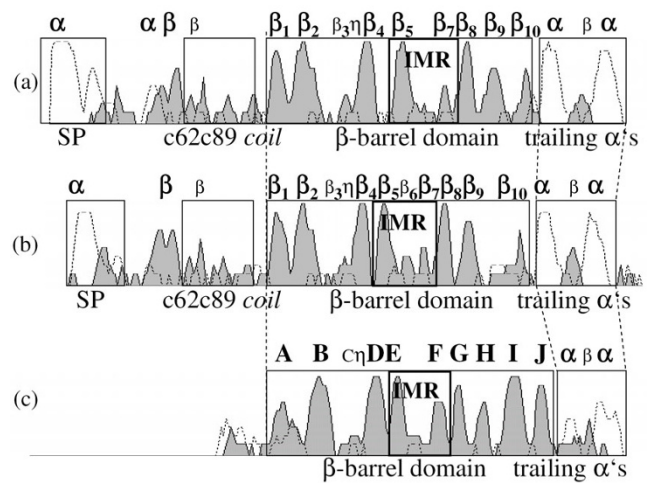


Figure 4
Confidence level of the PROF prediction. Representation of the level of confidence of the PROF 2d-structure prediction for type I NLP consensus sequence (a), type II NLP consensus sequence (b) and *Ilr5* cupin (c). The solid line (shaded gray) represents the confidence level for the β -strands, and the dashed line for the α -helices. SP = signal peptide. IMR = Inter Motif Region. η represents the GHRHDWE motif in type I and II NLP consensus sequences and ihrhscee in *Ilr5*, respectively.

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 *Ilr5* 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

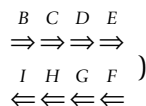
The RmlC-like cupin superfamily belongs to the SCOP double-stranded beta-helix fold. Cupins are double-stranded because they are composed of two sequences of antiparallel strands linked with short turns. If the NLPs are cupins, then there should be a correspondence between the 2d-structures of cupins and those predicted for NLPs. Cupins are formed by 8–10 β -strands called [A]BCDE-FGHI [J].

Table 2: Analyzed NLP sequences.

Plant pathogens organism		GenBank accession number (NLP type)	
Bacteria	<i>Enwinia carotovora atroseptica</i>	<u>CAG75986</u> ^{II} (NipEca) [1 0]	
Fungi	<i>Fusarium oxysporum</i>	<u>AAC97382</u> ^{**} (NEP1) [11] [12]	
	<i>Magnaporthe grisea</i>	<u>EDK02987</u> ^{II} , <u>EDJ98732</u> ^{II} , <u>EDJ96934</u> and <u>EDJ94825</u> [13]	
	<i>Verticillium dahliae</i>	<u>AAS45247</u> ^{**} (His_VdNEP) [14]	
	<i>Botrytis elliptica</i>	<u>CAJ98683</u> ^{**} (BeNEP1) and <u>CAJ98684</u> (BeNEP2) [15]	
	<i>Botrytis tulipae</i>	<u>ABB43261</u>	
	<i>Botrytis fabae</i>	<u>ABB43270</u>	
	<i>Gibberella zeae</i>	<u>XP 386193</u> , <u>XP 383570</u> ^{II} , <u>XP 387963</u> ^{II} and <u>XP 391669</u> ^{II} [16]	
	<i>Moniliophthora perniciosa</i>	<u>ABQ53551</u> ^{**} (Mp NEP1) and <u>ABQ32369</u> ^{**} (Mp NEP2) [5]	
	Oomycetes	<i>Phytophthora infestans</i>	<u>AAAY43363</u> ^{**} (NPP1), <u>AAAY43377</u> [°] (NPP1.2) and <u>AAAY43378</u> [°] (NPP1.3) [17]
		<i>Phytophthora megakarya</i>	<u>AAAX12401</u> and <u>AAAX12403</u> [18]
<i>Phytophthora parasitica</i>		<u>AAKI9753</u> ^{**} [6]	
<i>Phytophthora sojae</i>		<u>AAM48170</u> ^{**} (PsojNIP), <u>AAM48171</u> and <u>AAM48172</u> [19]	
<i>Pythium aphanidermatum</i>		<u>AAD53944</u> ^{**} (PaNie234) [20]	
Other organisms		GenBank accession number (NLP type)	
Bacteria	<i>Bacillus halodurans</i>	<u>BAB04114</u> [*] [19]	
	<i>Bacillus licheniformis</i>	<u>AAU23136</u>	
	<i>Vibrio pommerensis</i>	<u>CAC40975</u> ^{**} (causes hemolysis) [21]	
	<i>Streptomyces ambofaciens</i>	<u>CAJ89765</u> ^{II}	
	<i>Streptomyces coelicolor</i>	<u>CAB92890</u> ^{**} [19]	
	<i>Streptomyces griseus</i>	<u>BAF36639</u> ^{II}	
	<i>Streptomyces tsusimaensis</i>	<u>ABA59542</u> ^{II}	
	<i>Saccharopolyspora erythraea</i>	<u>YP 001105122</u> ^{II}	
	Fungi	<i>Aspergillus fumigatus</i>	<u>EAL86241</u> and <u>EAL86501</u> ^{II}
		<i>Aspergillus nidulans</i>	<u>EAA62936</u>
<i>Aspergillus niger</i>		<u>CAK46514</u>	
<i>Aspergillus oryzae</i>		<u>BAF63220</u>	
<i>Aspergillus terreus</i>		<u>EAU39525</u> ^{II}	
<i>Neurospora crassa</i>		<u>CAF05864</u> ^{II}	

Analyzed NLP sequences. Type II NLPs (15 sequences) are signed with the II symbol. NLPs signed with a single filled circle* and with double filled circles** cause a weak and a strong necrosis respectively. An empty circle° signs NLPs reported not to cause necrosis.

The formation of the β -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:



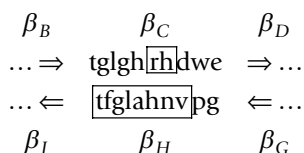
Finally, this double strand turns like an helix building up a β -barrel of two β -sheets: CHEF and BIDG with their hydrophobic residues aiming at the interior of the barrel. Inside the barrel, in the hydrophobic pocket, we find the metal ion bound to its ligand, next to the top of the barrel (the bottom is closed by the E and F β -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 cpsah g163g164 wdc 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 β -strands closing the bottom of the barrel. These β -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 consensus, and for the *1lr5* cupin, here we can see the correspondence between individual β -strands among these proteins. We observe that the β -barrel is built up by 7–9 high confidence level β -strands and 1–2 low confidence ones with a correspondence between strands in cupins and NLPs. For instance, β_1 in type I NLP consensus corresponds to β_A in cupins, β_2 to β_B , β_3 to β_C , β_4 to β_D , β_5 to β_E , β_7 to β_F , β_8 to β_C , β_9 to β_H and finally β_{10} corresponds to β_I . The most conserved pattern in NLPs, the GHRHDWE sequence (η in Figure 4), is between the putative C (β_3) and D (β_4) β -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 β -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 *1vj2*.

Two differences between the predictions obtained for type I and II NLP consensus and the cupin structure are worth mentioning: first, β_6 is present in type II NLPs but not in type I NLPs, and second, NLPs do not have β_7 . Certainly, the correspondence of β_6 to β_F 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 β -strand was just missed by the PROF program in type I NLP consensus and that β_8 , and not β_9 , should correspond to β_H in type I NLPs. Contrary to this idea, we propose that the conserved histidine H193^{95%} acts as metal ion ligand. In cupins, since the 4th ligand is in β_H , we propose H193 signs the position of the H β -strand in NLPs. More precisely, the 4th ligand (histidine) must be at the border of the β_H because the 1st and 2nd ligands are at the border of β_C (in the CD-coil) and C and H β -strands form an antiparallel β -sheet, as can be visualized by the following design (boxes represent β -strands).



Lastly, we could argue that the 3d-structure of type II NLPs includes 10 β -strands and not 9 as in type I NLPs and that β_8 is β_H with H179^{87%} (at the border of β_8) being the residue acting as 4th ligand. Besides the conserved histidines of type I NLPs, type II NLPs have two additional ones: H179^{87%} and H185^{100%}, 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 β_6 and β_7 (putative E and F β -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 S^{88%}a^{50%}H^{95%}g^{74%}. 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 a 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 β -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, Haberland 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 perniciosus* 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	...	β_E	aSaHggykkyt	β_F	...
dicot ABPs	...	β_E	asshgkfpqkp	β_F	...

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 ◊ AnvPmkdgnFlt ◊ kvgnayya). 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 wdedcfeaaak, the 15-residue *N. tabacum* Nt-abp1 C-terminal peptide (ywdeecyqttswkdel) 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 yWDEqCyqttxKDEL. 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 β_A , 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 nx_s are present in 64% of all necrotic type I NLPs (x = t^{6v}) 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	2d-Structure and glycosylation sites ([nx(st)])																
	C62c89-coli						β-barrel										
Type I NLP consensus	ααβαββ	Cβ	nts	L	Ck	β _A	β _B β _C	ηβ _D β _E	H	β _F	β _G	v	β _H	T	β _I	αβ	α
PpAAK19753	αββαββ	cβ	nts	l	ck	β _A	β _B β _C	ηβ _D β _E	h	β ₆ β _F	β _G	v	β _H	t	β _I	αβ	α
MpNEP1	αααβ	cβ	nws	l	ck	β _A	β _B β _C	ηβ _D β _E	h	β _F	β _G	v	β _H	t	β _I	α	α
Type II NLP consensus	αβαβ	Cβ	tln	l	Cr	β _A C	C β _B β _C	ηβ _D β _E	H	β ₆ β _F	β _G	g	β _H	A	β _I	α	α
Dicot ABP		cβ _A ⁿ	nis α	l	β _A ^r	β _A α	β _B β _C	ηβ _D β _E	h	β _F nst	β _G	v	β _H	t	β _I _β	αβ	cα
Monocot ABP		cβ _A ⁿ	dis α	m	β _A ⁱ	β _A α	β _B β _C	ηβ _D β _E	s	β _F nst	β _G	p	β _H	s	β _I _β	αβ	cα

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, PiNPP1.1 is 44%, PsojNIP 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 biding 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 3rd 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⁺]_{cyt} 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 ion channels. Correspondingly, the first plant reaction to the NLPs is an increase of ionic currents causing elevation of [Ca²⁺]_{cyt}. The elevation of H₂O₂ (and other ROS species) is upstream and downstream of the [Ca²⁺]_{cyt} elevation, and both [Ca²⁺]_{cyt} elevation and H₂O₂ are known by their roles in senescence and necrotic activities. Besides all these molecular and functional clues, for example, the fungus *M. pernicioso* causes the formation of witches' broom on *T. cacao* [5]. This reaction is typical of diverse plants in reaction to biotic stresses in its early phase and is related to the cytokinin/auxin balance.

There are several cupin candidates with a compatible IMR size, but only dicot ABPs seem to present some compatibility in relation to size and presence of the conserved histidine in the IMR of the NLPs and predicted glycosylation sites. Certainly, any auxin-like activity over the plant would be advantageous for the fungus (as discussed above). Last of all, as discussed by Fellbrich [6], NLPs are dependent on both, C and N-terminals of the protein for its activity, a feature shared by ABPs.

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).

Type I and II NLPs were aligned using ClustalW [44] with default parameters and *gapopen* = 0 (see alignment results in [45] and [46]).

These sequences are obtained computing the most frequent residues in the type I and II NLPs. If, in the alignment of the sequences, a gap is introduced in more than 50% of them, then the respective position is removed from the sequence. Therefore, type I and II NLP consensus are the consensus of ≥ 50% of the sequences. Furthermore, residues present in more than 85% of all sequences are in boldface and capitalized (see Figure 1). For example, the sequence GHRHDWE occurs in ≥ 85% of all type I NLPs, indicating its important role for the protein function [4]. Residues present in more than 70% of all sequences are in boldface (see Figure 1). For example, the same sequence GHrHDwE shows lower conservation in type II NLPs, as residues r and w occur in < 85% of type II NLP sequences.

We have performed an alignment (ClustalW with *gapopen* = 0) of type I NLPs for which we have experimental evidence of necrotic activity (denoted with one • and two filled circles **) as shown in Table 2 and we denote the residues conserved in all necrotic sequences with an asterisk in Figure 1 (see alignment results in [47]).

The secondary structure predictions were performed using the PROF program in the PredictProtein site [22].

Searches in the PDB

The 10 best candidates were obtained by submitting all sequences in positions 132–138 and 132–139 in the 42 NLPs to the search service of the Protein Data Bank (PDB) site [48,49] (Search Tool = *Fasta*). Each sequence generates a list of candidate proteins, which can be ordered by the e-value. The search for local sequences in the PDB was performed using Sequence features = *motif*.

Disulfide bonds and glycosylation analysis

The analysis of the bonding pattern among cysteines were performed with the DISULFIND program [50]. We also used the DiANNA 1.1 disulfide bond prediction program [29] for the analysis of the BeNEP2. In order to determine the probability of glycosylation of the sequons in the

Table 4: Residue histograms for NLPs and ABPs.

Position	158	159	160	161	162	163	164	165
Type I NLPs	a ⁴¹ s ⁴¹ c ¹⁸	a ⁴⁸ p ³⁰ y ⁷ l ⁴ m ⁴ t ⁴ v ⁴	s ⁹ n ⁴	a ⁵² g ²⁶ q ¹¹ .7 ⁴	h ⁹² a ⁴ y ⁴	g ⁷⁰ s ²⁶ n ⁴	g ⁵² k ¹⁵ d ¹¹ s ¹¹ n ⁴ r ⁴ t ⁴	y ⁵² w ²⁶ f ⁷ h ⁷ v ⁷
Type II NLPs	s ⁶⁰ a ³³ g ⁷	t ³³ a ²⁷ p ¹³ w ¹³ y ¹³	s ⁸⁰ t ²⁰	a ⁴⁷ c ¹³ n ¹³ q ¹³ e ⁷ k ⁷	h ¹⁰⁰	g ⁸⁰ k ⁷ r ⁷ s ⁷	d ³³ g ²⁰ k ²⁰ s ¹³ t ¹³	y ⁵³ f ²⁰ l ¹³ v ¹³
Dicot ABPs	a ⁴¹ l ³⁶ s ¹⁶ t ⁶	a ³⁸ s ³¹ p ³¹	e ³⁸ s ³¹ n ²⁵	s ⁵³ t ⁴⁷	h ¹⁰⁰	g ⁴⁴ a ¹⁶ s ¹⁶ l ¹³ e ¹³	k ⁵³ s ²⁵ e ⁹ t ⁶ n ¹³	f ⁴⁴ y ²⁵ t ¹⁹ s ⁹ h ³
Monocot ABPs	l ⁸⁹ m ¹¹	g ¹⁰⁰	s ¹⁰⁰	s ⁷⁸ t ²²	s ¹⁰⁰	l ⁵⁶ m ⁴⁴	k ⁸⁹ p ¹¹	y ¹⁰⁰

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 until the energy of 0.01 kJ/mol was reached. Then, we performed a simulated annealing from 0 to 30 K during 100 pS in vacuum. The force field used was *opls-aa/l*. Finally, a simulated 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. S. K. Braz for useful discussions and suggestions. Work partially supported by CNPq (research grants 506414/2004-3 and 474278/2006-9) and FAPESP (research grant 2007/02827-9).

References

- Bailey BA: Purification of a protein from culture filtrates of *Fusarium oxysporum* that induces ethylene and necrosis in leaves of *Erythroxylum coca*. *Phytopathology* 1995, **85**:1250-1255.
- Gijzen M, Nürnberger T: Nep1-like proteins from plants pathogens: Recruitment and diversification of the NPPI Domain Across Taxa. *Phytochemistry* 2006, **16(67)**:1800-1807.
- Tyler BM, Tripathy S, Zhang X, Dehal P, Jiang RH, Aerts A, Arredondo FD, Baxter L, Bensasson D, Beynon JL, Chapman J, Damasceno CMB, Dorrance AE, Dou D, Dickerman AW, Dubchak IL, Garbelotto M, Gijzen M, Gordon SC, Govers F, Grunwald NJ, Huang W, Ivors KL, Jones RW, Kamoun S, Kramps K, Lamour KH, Lee MK, McDonald WH, Medina M, Meijer HJG, Nordberg EK, Maclean DJ, Ospina-Giraldo MD, Morris PF, Phuntumart V, Putnam NH, Rash S, Rose JKC, Sakihama Y, Salamov AA, Savidor A, Scheuring CF, Smith BM, Sobral BWS, Terry A, Torto-Alalibo TA, Win J, Xu Z, Zhang H, Grigoriev IV, Rokhsar DS, Boore JL: **Phytophthora genome sequences uncover evolutionary origins and mechanisms of pathogenesis.** *Science* 2006, **313**:1261-1266.
- Pemberton CL, Salmond GPC: **The Nep1-like proteins – a growing family of microbial elicitors of plant necrosis.** *Mol Plant Pathol* 2004, **5(4)**:353-359.
- Garcia O, Macedo J, Tibúrcio R, Zapparoli G, Rincones J, Bittencourt L, Ceita G, Micheli F, Gesteira A, Mariano A, Schiavinato M, Medrano F, Meinhardt L, Pereira G, Cascardo J: **Characterization of necrosis and ethylene-inducing proteins (NEP) in the basidiomycete *Monilophthora perniciosa*, the causal agent of witches' broom in *Theobroma cacao*.** *Mycol Res* 2007, **111**:443-455.
- Fellbrich G, Romanski A, Varet A, Blume B, Brunner F, Engelhardt S, Felix G, Kemmerling B, Krzymowska M, Nürnberger T: **NPPI, a Phytophthora-associated trigger of plant defense in parsley and Arabidopsis.** *Plant J* 2002, **32(3)**:375-390.
- Dunwell JM: **Cupins: a new superfamily of functionally-diverse proteins that include germins and plant seed storage proteins.** *Biotechnol Genet Eng Rev* 1998, **15**:1-32.
- Dunwell JM, Khuri S, Gane PJ: **Microbial relatives of the seed storage proteins of higher plants: conservation of structure, and diversification of function during evolution of the cupin superfamily.** *Microbiol Mol Biol Rev* 2000, **64**:153-179.
- Dunwell JM, Culham A, Carter CE, Sosa-Aguirre CR, Goodenough PW: **Evolution of functional diversity in the cupin superfamily.** *Trends Biochem Sci* 2001, **26**:740-745.
- Pemberton CL, Whitehead NA, Sebahia M, Bell KS, Hyman LJ, Harris SJ, Matlin AJ, Robson ND, Birch PRJ, Carr JP, Toth IK, Salmond GPC: **Novel quorum-sensing-controlled genes in *Erwinia carotovora* subsp. *carotovora*: identification of a fungal elicitor homologue in a soft-rotting bacterium.** *Mol Plant Microbe Interact* 2005, **18**:343-353.
- Jennings JC, Apel-Birkhold PC, Bailey BA, Anderson JD: **Induction of ethylene biosynthesis and necrosis in weed leaves by a *Fusarium oxysporum* protein.** *Weed Sci* 2000, **48**:7-14.
- Bailey BA, Apel-Birkhold PC, Luster DG: **Expression of NEPI by *Fusarium oxysporum* f. sp. *erythroxyli* after gene replacement and overexpression using polyethylene glycol-mediated transformation.** *Phytopathology* 2002, **92(8)**:833-841.
- Dean RA, Talbot NJ, Ebbole DJ, Farman ML, Mitchell TK, Orbach MJ, Thon M, Kulkarni R, Xu JR, Pan H, Read ND, Lee YH, Carbone I, Brown D, Oh YY, Donofrio N, Jeong JS, Soanes DM, Djonovic S, Kolomiets E, Rehmeier C, Li W, Harding M, Kim S, Lebrun MH, Bohnert H, Coughlan S, Butler J, Calvo S, Ma LJ, Nicol R, Purcell S, Nusbaum C, Galagan JE, Birren BW: **The genome sequence of the rice blast fungus *Magnaporthe grisea*.** *Nature* 2005, **434(5)**:980-986.
- Wang JY, Cai Y, Gou JY, Mao YB, Xu YH, Jiang WH, Chen XY: **VdNEP, an Elicitor from *Verticillium dahliae*, induces cotton plant wilting.** *Appl Environ Microbiol* 2004, **70**:4989-4995.
- Staats M, van Baarlen P, Schouten A, van Kan JAL, Bakker FT: **Positive selection in phytotoxic protein-encoding genes of *Botrytis* species.** *Fungal Genet Biol* 2007, **44**:52-63.
- Han Y, Kim M, Lee S, Yun S, Lee Y: **A novel F-box protein involved in sexual development and pathogenesis in *Gibberella zeae*.** *Mol Microbiol* 2007, **63**:768-779.
- Kanneganti TD, Huitema E, Cakir C, Kamoun S: **Synergistic interactions of the plant cell death pathways induced by *Phytophthora infestans* Nep1-like protein PiNPPI.1 and INFI elicitor.** *Mol Plant Microbe Interact* 2006, **19**:854-863.
- Bae H, Bowers JH, Tooley PVV, Bailey BA: **NEPI orthologs encoding necrosis and ethylene inducing proteins exist as a multi-gene family in *Phytophthora megakarya*, causal agent of black pod disease on cacao.** *Mycol Res* 2005, **109(12)**:1373-1385.
- Qutob D, Kamoun S, Gijzen M: **Expression of a *Phytophthora sojae* necrosis-inducing protein occurs during transition from biotrophy to necrotrophy.** *Plant J* 2002, **32(3)**:361-373.
- Veit S, Worle JM, Nürnberger T, Koch W, Seitz HU: **A novel protein elicitor (PaNie) from *Pythium aphanidermatum* induces multiple defense responses in carrot, arabidopsis, and tobacco.** *Plant Physiol* 2001, **127**:832-841.

21. Jores J, Appel B, Lewin A: **Cloning and molecular characterization of a unique hemolysin gene of *Vibrio pommerensis* 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, Yachdav 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. Straganz GD, Eggera S, Aquinoc G, D'Auriac S, Nidetzky B: **Exploring the cupin-type metal-coordinating signature of acetylacetonate dioxygenase Dkel with site-directed mutagenesis: Catalytic reaction profile and Fe²⁺ binding stability of Glu-69 Gln mutant.** *Journal of Molecular Catalysis B, Enzymatic* 2006, **39**:171-178.
25. Eulgem T, Rushton PJ, Robatzek S, Somssich IE: **The WRKY superfamily of plant transcription factors.** *Trends Plant Sci* 2000, **5(5)**:1360-1385.
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* 2001, **28(2)**:123-133.
27. Qutob D, Kemmerling B, Brunner F, Kufner I, Engelhardt S, Gust AA, Luberacki B, Seitz HU, Stahl D, Rauhut T, Glawischig E, Schween G, Lacombe B, Watanabe N, Lam E, Schlichting R, Scheel D, Nau K, Dodt G, Hubert D, Gijzen M, Nurnberger T: **Phytotoxicity and Innate Immune Responses Induced by Nep I-Like Proteins.** *Plant Cell* 2006, **18(12)**:3721-3744.
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 EM: **KDEL-containing auxin-binding protein is secreted to the plasma membrane and cell wall.** *Plant Physiol* 1993, **101(2)**:595-606.
31. Woo E, Marshall J, Baully J, Chen J, Venis M, Napier RM, Pickersgill RW: **Crystal structure of auxin-binding protein I in complex with auxin.** *EMBO J* 2002, **21**:2877-2885.
32. Cleland RE, Prins HBA, Harper JR, Higinbotham N: **Rapid hormone-induced hyperpolarization of the oat coleoptile transmembrane potential.** *Plant Physiol* 1977, **59**:395-397.
33. Warwicker J: **Modelling of auxin-binding protein I 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. Baully JM, Sealy IM, Macdonald H, Brearley J, Dröge S, Hillmer S, Robinson 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 pernicioso*, causal agent of witches' broom disease of cocoa.** *Trans Br Mycol Soc* 1980, **74**:515-523.
36. Kilaru A, Bailey BA, Hasenstein KH: ***Monilophthora pernicioso* 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 Physiol* 1978, **62**:522-525.
38. Krupasager V, Sequeira L: **Auxin destruction by *Marasmius pernicioso*.** *Am J Bot* 1969, **56**:390-397.
39. David K, Carnero-Diaz E, Leblanc N, Monestiez M, Grosclaude J, Perrot-Rechenmann C: **Conformational dynamics underlie the activity of the auxin-binding protein, Nt-abpl.** *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 NepI from *Fusarium oxysporum*.** *Plant Sci* 2001, **161**:891-899.
41. Simmons CR, Liu Q, Huang Q, Hao Q, Begley TP, Karplus PA, Stipanuk MH: **Crystal structure of mammalian cysteine dioxygenase: A novel mononuclear iron center for cysteine thiol oxidation.** *J Biol Chem* 2006, **281(27)**:18723-18733.
42. Tian M, Win J, Hoorn R, Knaap E, Kamoun S: **A *Phytophthora infestans* cystatin-like protein targets a novel tomato papain-like apoplasmic protease.** *Plant Physiol* 2007, **143**:364-377.
43. Woo EJ, Dunwell JM, Goodenough PV, Marvier AC, Pickersgill RW: **Germin is a manganese containing homohexamer with oxalate oxidase and superoxide dismutase activities.** *Nat Struct Biol* 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.** *Nucl Acids Res* 1994, **22**:4673-4680.
45. **Type I NLPs alignment** [<http://adelmo.cechin.googlepages.com/typelg0.aln>]
46. **Type II NLPs alignment** [<http://adelmo.cechin.googlepages.com/typelg0.aln>]
47. **Necrotic sequences alignment** [<http://adelmo.cechin.googlepages.com/tlnecg0.aln>]
48. **RCSB 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. Ceroni 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**:14-23.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp

