UNIVERSITA' DEGLI STUDI DI PARMA

DOTTORATO DI RICERCA IN BIOLOGIA E PATOLOGIA MOLECOLARE

XXVIII CICLO

COMPARISON OF THE DEFENSIVE ELICITATION ACTIVITY OF CERATO-POPULIN AND CERATO-PLATANIN: A STRUCTURAL MODEL

Coordinatore:

Chiar.ma Prof.ssa Valeria Dall'Asta

Tutors:

Chiar.ma Prof.ssa Lorella Franzoni

Chiar.mo Prof. Alberto Spisni

Dottorando: Fabio Baroni

ABSTRACT

Plants can defend themselves from potential pathogenic microorganisms relying on a complex interplay of signaling pathways: activation of the MAPK cascade, transcription of defense related genes, production of reactive oxygen species, nitric oxide and synthesis of other defensive compounds such as phytoalexins. These events are triggered by the recognition of pathogen's effectors (effector-triggered immunity) or PAMPs (PAMP-triggered immunity).

The Cerato Platanin Family (CPF) members are Cys-rich proteins secreted and localized on fungal cell walls, involved in several aspects of fungal development and pathogen-host interactions. Although more than hundred genes of the CPF have been identified and analyzed, the structural and functional characterization of the expressed proteins has been restricted only to few members of the family. Interestingly, those proteins have been shown to bind chitin with diverse affinity and after foliar treatment they elicit defensive mechanisms in host and non-host plants. This property turns cerato platanins into interesting candidates, worth to be studied to develop new fungal elicitors with applications in sustainable agriculture.

This study focus on cerato-platanin (CP), core member of the family and on the orthologous cerato-populin (Pop1). The latter shows an identity of 62% and an overall homology of 73% with respect to CP. Both proteins are able to induce MAPKs phosphorylation, production of reactive oxygen species and nitric oxide, overexpression of defense's related genes, programmed cell death and synthesis of phytoalexins. CP, however, when compared to Pop1, induces a faster response and, in some cases, a stronger activity on plane leaves.

Aim of the present research is to verify if the dissimilarities observed in the defense elicitation activity of these proteins can be associated to their structural and dynamic features.

Taking advantage of the available CP NMR structure, Pop1's 3D one was obtained by homology modeling. Experimental residual dipolar couplings and ^{1}H, ^{15}N, ^{13}C resonance assignments were used to validate the model.

Previous works on CPF members, addressed the highly conserved random coil regions (loops β1-β2 and β2-β3) as sufficient and necessary to induce necrosis in plants' leaves: that region was investigated in both Pop1 and CP. In the two proteins

the loops differ, in their primary sequence, for few mutations and an insertion with a consequent diversification of the proteins' electrostatic surface. A set of 2D and 3D NMR experiments was performed to characterize both the spatial arrangement and the dynamic features of the loops. NOE data revealed a more extended network of interactions between the loops in Pop1 than in CP. In addition, in Pop1 we identified a salt bridge Lys^{25}/Asp^{52} and a strong hydrophobic interaction between Phe^{26}/Trp^{53}. These structural features were expected not only to affect the loops' spatial arrangement, but also to reduce the degree of their conformational freedom. Relaxation data and the order parameter S^2 indeed highlighted reduced flexibility, in particular for loop β1−β2 of Pop1. *In vitro* NMR experiments, where Pop1 and CP were titrated with oligosaccharides, supported the hypothesis that the loops structural and dynamic differences may be responsible for the different chitin-binding properties of the two proteins: CP selectively binds tetramers of chitin in a shallow groove on one side of the barrel defined by loops β1-β2, β2-β3 and β4-β5, Pop1, instead, interacts in a non-specific fashion with oligosaccharides.

Because the region involved in chitin-binding is also responsible for the defense elicitation activity, possibly being recognized by plant's receptors, it is reasonable to expect that those structural and dynamic modifications may also justify the different extent of defense elicitation. To test that hypothesis, the initial steps of a protocol aimed to the identify a receptor for CP, *in silico*, are presented.

The results presented in this thesis have been published:

1. F. Baroni, L. Pazzagli, S. Luti, A. Scala, F. Martellini, L.Franzoni, T. A. Pertinhez, A.Spisni.
^{1}H, ^{15}N and ^{13}C resonance assignment of cerato-populin, a fungal PAMP from Ceratocystis populicola.
Biomolecular NMR Assignments. 2014 Oct;8(2):405-8. (Epub 2013 Oct 4).

2. F. Baroni, L. Pazzagli, S. Luti, L. Mazzoli, A. Spisni, T. A. Pertinhez
A possible role of some structural and dynamic features of the fungal proteins cerato-populin and cerato-platanin during plant attack.
Plant Physiology and Biochemistry, January 2016, submitted.

INDEX

Introduction

1. THE CERATO PLATANIN FAMILY (CPF)

Plants can defend themselves from potential pathogenic microorganisms relying on a complex interplay of signaling pathways: activation of the MAPK (mitogen-activated protein kinase) cascade, transcription of defense related genes, production of reactive oxygen species (ROS), nitric oxide and synthesis of other defensive compounds such as phytoalexins [1, 2]. Their immune system is composed of surveillance molecular patterns able to detect several microbial elicitors. They are evolutionarily stable molecules, commonly known as pathogen- or microbe-associated molecular patterns (PAMPs or MAMPs) [3, 4], able to activate processes that lead the plant to switch from a developmental status to a defensive mode, functional to reject potentially harmful microbes.

Plant pathogenic fungi produce a large number of proteins devoid of catalytic activity, which are involved in various aspects of parasitism and in the development of disease. Most of them are recognized by the plant immunity system, triggering defenses activation. Currently, on the basis of their primary sequence, seven non-catalytic fungal protein families have been identified: the cerato-platanin family (PF07249), the class I hydrophobins (PF01185), the class II hydrophobins (PF06766), the elicitins (PF00964), the PcF family (PF09461), and the NIP-1 protein family (PF08995) [5].

The Cerato Platanin Family (CPF) members are small Cys-rich proteins (12-14 KDa) secreted into the culture filtrate or localized in fungal cell walls. Those proteins are only found in filamentous fungi (e.g. fungi producing hyphae as growth structures) or in fungi which have a pseudo-hyphal growth stage during their life cycle. Although an increasing number of experimental results suggest that CPF members do not always share the same biochemical and biological role, most of them have been shown to act either as virulence factors or as elicitors in host and non-host plants [6]. Examples are Snodprot1 from *Phaeosphaeria nodorumis,* required for virulence [7] and MpCP1from *Moniliophthora perniciosa* that exerts its necrosis-inducing ability

in tobacco and cacao leaves [8]. Several other examples are reviewed by Pazzagli et al., 2014 [9]. In addition, some members of the CP family play a major role in animal–fungus interactions, exhibiting allergenic properties and inducing strong immunological reactions [10-12].

More than hundred genes of the CPF have been identified and analyzed, but the structural and functional characterization of the expressed proteins has been restricted only to few members: cerato-platanin (CP), founder of the CPF, [13, 14], MgSM1, a CP-like protein from *M. oryzae*, and BcSpl1 from *B. cinerea* [15, 16].

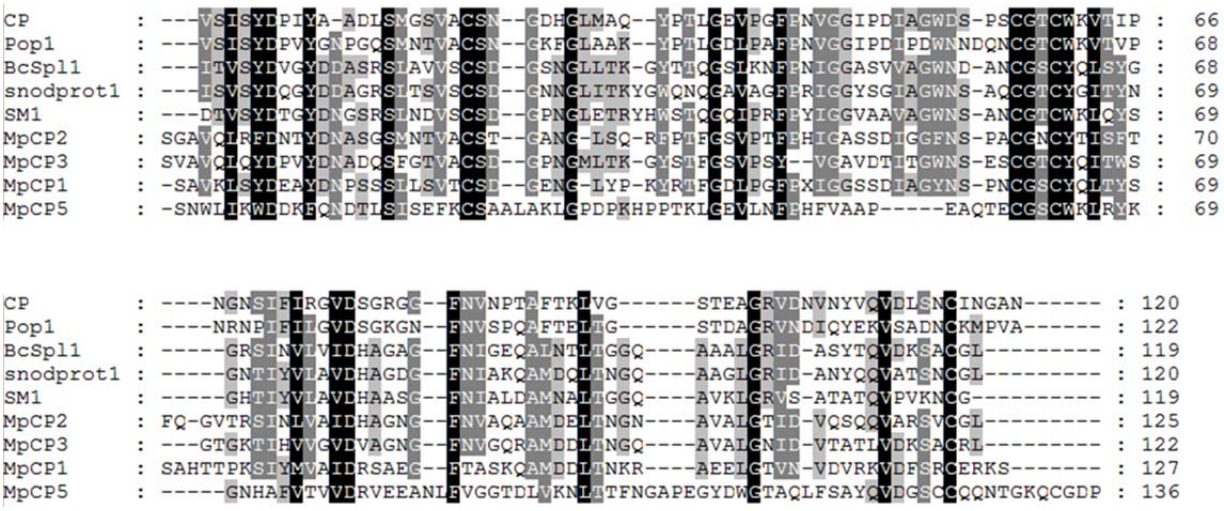

Figure 1. Alignment of 9 CPF proteins cited in this work, performed with ClustalW [29]. Conserved residues are shaded in colors ranging from black (high conservation) to light grey (low conservation).

1.1 CERATO-PLATANIN (CP)

Cerato-platanin (CP) has been the first member of this family whose gene was identified, cloned, expressed and functionally characterized [17-19]. Figure 1 shows the 120 amino acids sequence of CP defined by PFAM database as 'cerato-platanin domain', a domain that is found in more than 130 sequences of Ascomycetes and Basidiomycetes genomes [20]. As already highlighted by previously published alignments, the sequences of CPF members have a similarity about 40% and most of the conserved residues are hydrophobic.

CP is secreted by *Ceratocystis platani*, an ascomycete, causal agent of the canker stain disease in *Platanus acerifolia* [6, 18]. This fungus is a wound parasite that infects plane trees, causing interruption of water movement, cankers and eventually death. Cankers on the tree trunk are characterized by necrosis of inner bark and bluish-black to reddish-brown discoloration of sapwood. The disease, widely spread among Europe and North America, can lead to the death of the tree within 2-3 years.

CP is a 120 amino acids protein (12.4 KDa) found in the cell wall of ascospores, hyphae and conidia [21]. It has been shown that, after foliar treatment, it induces MAPKs phosphorylation, production of ROS and nitric oxide, overexpression of defense's related genes, programmed cell death and synthesis of phytoalexins. For this reason CP has been considered a PAMP [22, 23].

Although its eliciting activity on plants is well known, the biological role of this protein is still controversial.

The structure of CP, a double ψβ-barrel fold (figure 2A), is remarkably similar to that occurring in endoglucanases, in the plant defense protein Barwin and in domain I of expansins [13]. For this reason CP ability to hydrolyze cellulose to smaller oligosaccharides, typical function of endoglucanases, was tested. *In vitro*, the protein did not exhibit any hydrolytic ability. Interestingly, however, other experiments revealed that CP is able to bind with high affinity (K_d = 40.56 μM) tetramers of chitin [13, 14], thus sustaining the idea that CP and endoglucanases are connected not only by structural similarity. Interestingly, CP exhibits also an expansin-like activity (i.e. mechanical disruption of cell walls) on cellulose polymers [14]. The multifaceted biological activity of CP includes also some features typical of hydrophobins. Hydrophobins are self-assembling proteins that form amyloid-like fibrils, involved in the formation of aerial hyphae and in adhesion to surfaces. Also CP, *in vitro*, upon interaction with hydrophobic surfaces undergoes self-assembly: that interaction induces the rapid formation of unfolded and aggregated CP, which is a more active form of the protein [24]. These data were later confirmed *in vivo* observing that CP is able to interact with the leaves' hydrophobic cuticle. Following these observations it has been suggested that the active form of the protein has a partially unfolded structure and that the progression toward an unfolded form is probably induced by hydrophobic cuticular waxes [25].

CP besides being the first CPF protein isolated and investigated, it has been also the first one for which a 3D-structure was solved by NMR spectroscopy (PDB 2kqa) [13]. Susequently, using crystallography methods, the structures of Sm1 from *T. virens* (PDB 3m3g) and MpCP1, MpCP2, MpCP3 and MpCP5 from *M. perniciosa* (PDB 3suj, 3sul, 3sul and 3sum, respectively) were determined [26].

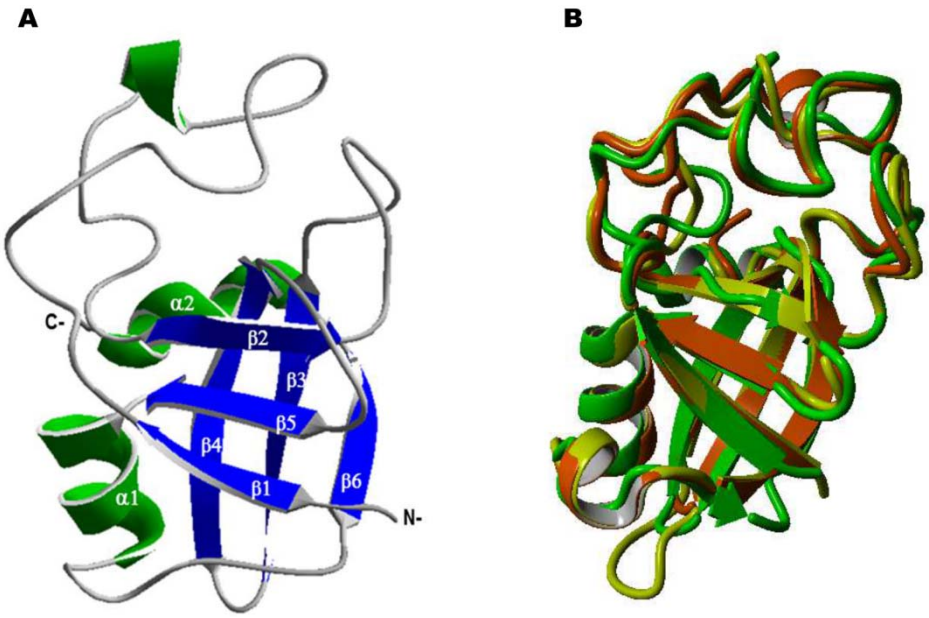

Figure 2. A) 3D structure of cerato-platanin (PDB: 2kqa) and **B)** structure alignment of CP (green), MpCP1 (orange, PDB: 3suj) and Sm1 (yellow, PDB: 3M3G). Alignment was done with MUSTANG [30] and structure manipulation performed with YASARA STRUCTURE 14.12.2 [31].

In figure 2 is reported the 3D-structure of CP (A) and the superposed structure of CP, MpCP1 and Sm1 (B). Cerato-platanin exhibits a globular fold containing 2 α-helices and 6 β-strands forming a double ψβ-barrel, and 2 disulfide bonds (Cys^{20}-Cys^{57}; Cys^{60}-Cys^{115}) conserved amongst the members of the CPF. Figure 2B, highlights that the high degree of homology observed between sequences of CPF members (figure 1) is conserved also at 3D level, strengthening the idea of a fold essential for cerato platanins activity.

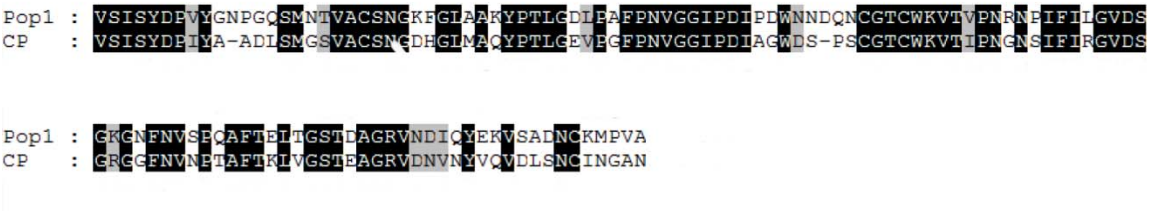

Figure 3. Alignment of Pop1 and CP performed with clustalW [29]. Identities are shaded in black; similarities in grey.

1.2 CERATO-POPULIN (Pop1)

Soon after the isolation of cerato-platanin, several other proteins orthologous to CP, and expressed in different *Ceratocystis* species, were identified and cloned [27]. Among them, we find cerato-populin (Pop1) from *Ceratocysitis populicola*, the causal agent of the canker stain disease in poplar trees.

Figure 3 reports an alignment between CP and Pop1. The latter shows an identity of 62% and an overall homology of 73% with respect to CP. Similarly to cerato-platanin, also Pop1 is able to induce the defense-related events previously described, showing also the same oligosaccharides-binding and expansin-like activity of the core member. However, CP induces a faster response and sometimes a stronger activity on plane leaves when compared to Pop1 [28]. A summary of the activity tested on Pop1 and CP is reported in Table 1.

	In CP, with respect to Pop1	References
Oligosaccharides-binding	tenfold higher affinity	14
Expansine-like activity	higher	14
Aggregation	comparable	26
MAPKs activation	faster	29
ROS production (NO and H_2O_2)	faster and fivefold higher	29
Synthesis of phytoalexins	faster	29
Overexpression of defense related genes	faster	29

Table1. The results of the available comparisons of CP' and Pop1's activities are reported, as well as the reference paper.

2. IMMUNITY AND PLANTS

The immune system of plants presents several differences compared to the well-studied one in vertebrates. These different features are mainly due to the absence of a circulatory system and of mobile immune cells, thus to the impossibility to use circulating immune receptors to detect the non-self. Accordingly, every plant cell is able to recognize a pathogen, establishing immune responses highly specific, that often generate lifelong memory. Every cell can counteract a pathogen invasion trough a complex interplay of preformed and/or infection-induced mechanisms. Among the first mechanisms we find physical barriers (the cuticle and the cell wall), antimicrobial chemicals (e.g. glucosides, saponins), enzymes and constitutively expressed receptors that can activate inducible plant defenses. Infections-induced mechanisms, instead, include cell wall reinforcements (callose, lignin and suberin), ROS production and expression of phytoalexins, antimicrobial enzymes (e.g. chitinisases, peroxidases) and antimicrobial proteins (e.g. defensins, thionins).

Depending on the pathogen molecule and the receptor involved in its recognition, plant immunity can be divided in two layers of surveillance mechanisms: PAMP-triggered immunity (PTI) and Effector-triggered immunity (ETI).

Molecules such as lipolysaccharides, peptidoglycans and flagellin have been long known as powerful PAMP. Recently, also chitin was included in this class. These compounds, conserved between fungi and bacteria, are recognized by pattern-recognition receptors (PRRs) which show structural and functional similarity to the animals' ones. As it happens in vertebrates, the recognition of a PAMP triggers a generic response, independent of the attacking pathogen: ion fluxes (mainly Ca^{2+}), oxidative burst, MAPKs activation and production of antimicrobial peptides.

Microbes (especially fungi), however, can avoid PTI via the production of effector proteins able to mask a PAMP, and thus the pathogen. In contrast, plants have co-evolved specific receptors for these effectors (induced or non-induced), known as resistance proteins (R-proteins). R-proteins mediated immunity (ETI) and PTI are linked to ROS accumulation and defense-related genes activation, though their response differs significantly both kinetically and quantitatively. ETI, in addition, leads to a hypersensitive response, typically associated with programmed cell death and release, in the surrounding tissue, of antimicrobial molecules (chitinases and

glucanases) that induce local resistance to the pathogen. This second layer of defense, whose pathways overlap the PTI ones, grants the plants with the possibility to achieve specificity, recognizing effectors highly polymorphic between different pathogen strains [32-35].

An important aspect to note is that both ETI and PTI responses show similar patterns, thus it is possible they involve each other's signaling pathways [36]. In this respect, to discriminate between activation of ETI or PTI may be difficult, if only based on the investigation of the signal-transduction pathway activated by a PAMP or an effector. The analysis of the structure, functions and eventually of the conformational dynamics of a molecule is essential to address the question: which receptor and which type of responses are involved?

2.1 RESISTANCE PROTEINS

Resistance proteins (R-proteins) are intracellular and extracellular receptors belonging to the NBS-LRR family, which is a subgroup within the STAND (Signal Transduction ATPases with Numerous Domains) [37]. All NBS-LRR proteins possess two main domains linked by a homologous region called ARC: a leucine-rich repeat (LRR) domain, involved in protein-protein and protein-ligand interactions and a nuclear binding site (NBS) domain with NTPase activity. The latter is suggested to play a crucial role, as molecular switch, to activate signal transduction.

The two domains are often associated to accessory ones, with different functions e.g. Toll/Interleukin-1 (TIR), Coiled-Coil (CC), WRKY and Bed-type Zinc-finger domains. A classification of NBS-LRR proteins based on the presence of these domains has been proposed [38].

Although they normally function as intracellular or extracellular receptors, certain NBS-LRR proteins, such as the ART protein from *Solanum lycopersicum* [39], have been found involved in processes active downstream the signaling cascades of other R-proteins or PRR. Moreover, cases of NBS-LRR proteins not associated with resistance are reported (e.g. CSA1 from *Arabidopsis thaliana*, involved in photomorphogenic processes) [40].

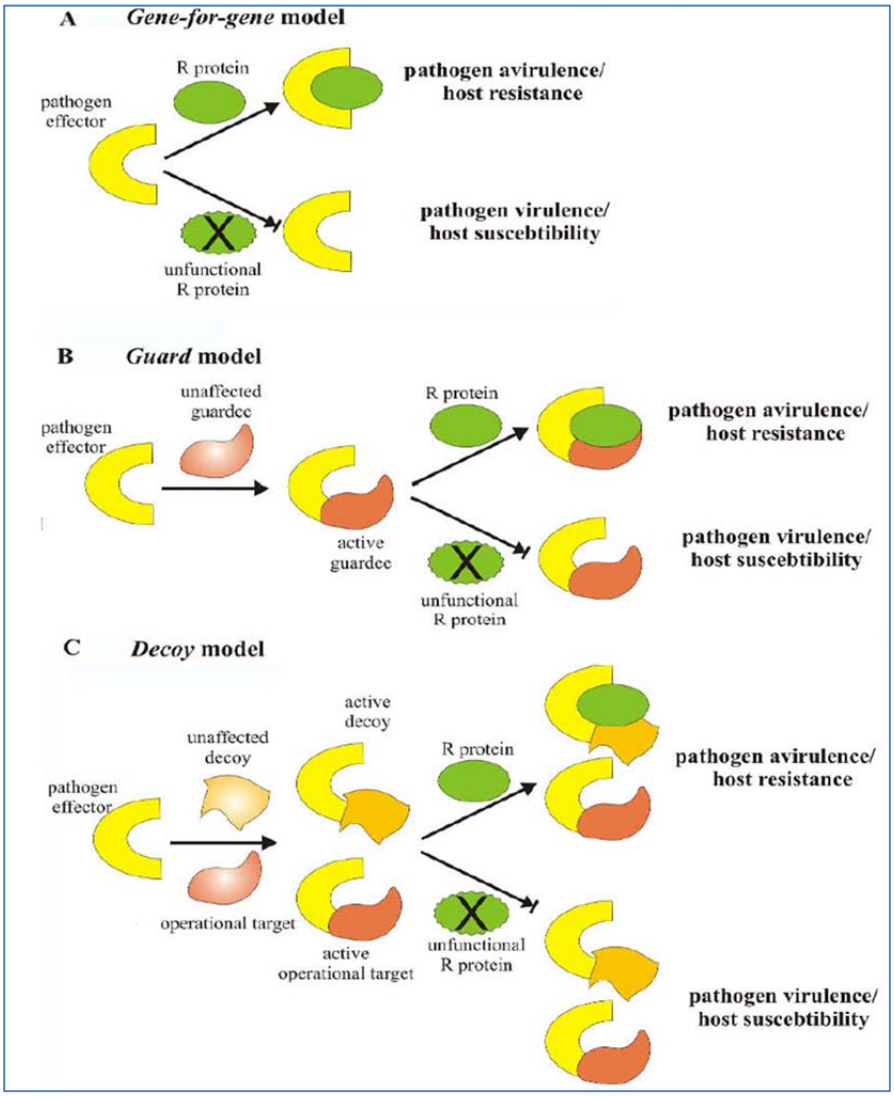

Figure 4. Models of effector recognition by R-proteins [43].

Three molecular models of effector recognition by R-proteins have been proposed during the last years [38 and 41-43], updating a highly variable panorama (Figure 4):

1) The gene-for-gene model: an R-protein interacts directly with an effector.
2) The guard model: the target protein (guardee) of an effector is guarded by an R-protein. If the effectors binds to its target, the R-protein will bind too.
3) The decoy model: specific proteins that are similar to those targeted by an effector are generated by the plant. Their function is to bind that effector, acting as mediators in the interaction with the R-proteins.

2.2 LEUCINE RICH REPEAT DOMAINS

Both R-proteins and PRR share, in most cases, the same domain involved in ligand recognition and protein-protein interactions: a leucine-rich repeat (LRR) domain.

The LRR domain was identified for the first time in the α2-glycoprotein of human serum, in 1985 [44]. Since then, this domain was found in several proteins involved in a variety of biological processes (ranging from signal transduction to cell adhesion, from DNA repair to the immune response), and its essential function appears to be to provide a structural framework for protein-protein interactions.

A LRR is a sequence of 20-40 amino acids containing a consensus sequence LxxLxLxxNxL (where L is a leucine residue or another aliphatic amino acid, N is asparagine, threonine, serine or cysteine, and x is any amino acid). This 11-residues motif is termed constant region and is common to all LRR: the remaining part of the repeat, termed variable region, identifies 7 different subtypes of LRR [45].

The tandem repeat, of at least 2 motifs, gives origin to the characteristic horseshoe-shaped super-helix of LRR domains (Figure 5), where each repeat forms a coil of the super-helix. The inner surface of the horseshoe structures is composed of parallel β-strands containing the hydrophobic and aliphatic residues of the consensus sequence. The outer part (residues from the variable region) is mostly helical and connected with β-strands by β-turns [46]

The concave face and the adjacent loops are the most common protein interaction surfaces on LRR proteins; two studies demonstrated that the involvement of the consensus region in recognizing PAMPs or effectors is conserved also in plants [48, 49]. However, in animals LRR, other regions of the domain are often involved, especially those corresponding to the variable region of the repeat: several examples are reported for Toll-like receptors, the most studied LRR-containing proteins. This aspect, together with the observation that the variable regions of plants' LRR are often longer and more variable than the animals ones, lead to hypothesize that the involvement of regions on the convex side of the structure cannot be excluded also in plants.

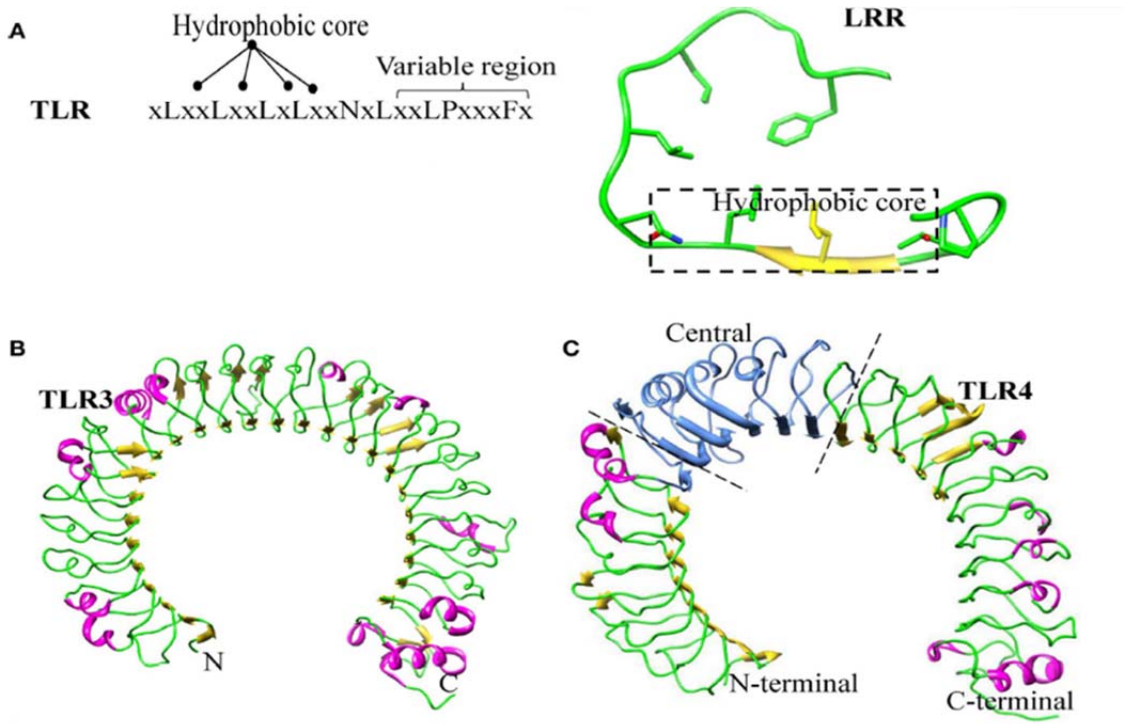

Figure 5. Structure of LRR domain of TLR3 and TLR4 [47]. **A)** Consensus sequence and variable region of a single repeat. Its secondary structure is highlighted on the right. **B)** and **C)** show the typical horseshoe superhelices formed by LRR repeats in tandem, with a parallel β-sheet on the concave side and mostly helical elements on the convex side.

Aim of the work

Microbial elicitors are key factors in activating plants' defense systems. Therefore, understanding the involved molecular mechanisms is a key step for the development of new compounds and protocols to avoid the use of chemical pesticides: a way to support sustainable agriculture and to reduce environment pollution.

CP and Pop1, are able to improve plant's defense also in non-host plants, such as the model plant *Arabidopsis*, and to limit the growth of *Botrytis cinerea*, a necrotrophic fungus that affects many plant species [50].

Because of the high homology the same fold of CP was expected for Pop1: hypothesis strengthened by preliminary circular dichroism data (see 7.2). In spite of these similarities, the two proteins showed different timing and extent of their activity (Table 1). These similarities and differences, render the two proteins optimal candidates to tackle the leading question of this work: which are the structural and molecular basis of the different defense elicitation activity of CP and Pop1?

Answering this question we expect to gain important novel insights needed to understand the mechanisms underlying activation of plants' defenses and strategic to engineer new fungal elicitors.

Recent works on the cerato platanin BcSpl1 from *B. cinerea*, highlighted two regions located on the surface-exposed loops β1-β2 and β2-β3 (Figure 2) as sufficient and necessary to induce necrosis in plants' leaves and that highly contribute to the fungus virulence. The reciprocal interactions of these loops, very conserved among CPF members, were alleged to be crucial for the protein activity [16].

Using as starting point the 3D structures of CP and Pop1, in this work we have investigated the structural and dynamic features of the two proteins with specific attention to their loops, aiming to characterize the molecular differences underlying their elicitation activity and to shed light on their still unclear biological function.

The second part of this work deals with the identification of a receptor for CP, in *Arabidopsis*. A protocol for *in silico* screening of candidate receptors is proposed and a discussion about the involvement of an effector-triggered immunity, instead that a PAMP-triggered one is presented.

Materials and Methods

3. EXPRESSION AND PURIFICATION OF CP AND POP1

Recombinant, ^{15}N-labelled and ^{13}C, ^{15}N-labelled CP was expressed according to previously published methods [51, 52].

Pop1 was expressed as a recombinant protein using *Pichia pastoris* (GS115-His strain, InVitrogen, San Diego, CA, USA) that allows to obtain a properly folded and biologically active protein in high yield [25]. Briefly, Pop1 gene was obtained from genomic DNA of *C. populicola* and the sequence corresponding to the mature protein was cloned in the pPIC9 vector to permit the secretion of Pop1. For large scale production of the ^{15}N uniformly labelled protein, a single colony of the transformed yeast was used. The ^{15}N labelled protein was obtained by growing the colony in 100mL BMG (Buffered Minimal Glycerol) containing $^{15}NH_4SO_4$, at 30 °C until an OD600nm = 6. After centrifugation at 3,000xg for 15 minutes the pellet was re-suspended in 600 mL of BMM (Buffered Minimal Methanol, containing 1% of $^{15}NH_4SO_4$), placed in a 2.0 L baffled flask and returned to the incubator for further growth. To maintain Pop1 induction, aliquots of 100% methanol were added every 24h for 6 days. The ^{15}N Pop1 was purified from culture filtrate by RP-HPLC (semi-preparative C4, 5μm, Dionex column). To obtain the ^{15}N-^{13}C uniformly double labelled proteins, also [^{13}C] glycerol and [^{13}C] methanol were used as previously described [52].

Proteins were purified by Reverse-Phase HPLC (10x250 mm, 5 μm, C4 column, Phenomenex) and were estimated to be 98% pure by SDS-PAGE. From 1 L of cultural filtrates about 12 mg of either CP or Pop1 were obtained. MALDI-TOF mass spectrometry (Ultraflex, Bruker, Bremen, Germany) confirmed that they were 100% ^{13}C- and ^{15}N-isotopically labelled.

Due to the cloning protocol, recombinant CP presented an additional 9-residues N-terminal tag (EEGVSLEKR) and recombinant Pop1 a 4-residues one (EAEA). Because both proteins showed the same biological activity as well as the same

circular dichroism and ^{1}H NMR spectra of native CP and Pop1, those residues have not been included in the structural analysis.

4. CIRCULAR DICHROISM EXPERIMENTS

3 μM of purified Pop1 solution in 20 mM potassium phosphate buffer, was used for the far-UV (190-250 nm) Circular Dichroism measurements.

Measurements have been recorded on a Jasco J-715 spectropolarimeter equipped with a Peltier system (PTC-348 WVI) for temperature control. Spectra have been collected at 20°C; parameters have been set as follows: 2mm optical path length, band width 1 nm, speed 50 nm/min, response 8 sec and data pitch 0.5 nm. The average of four spectra was taken to improve the signal to noise ratio and the spectrum was corrected by subtracting its blank. Ellipticity is reported as the mean residue molar ellipticity ($[\theta]$ degrees cm^2 $dmol^{-1}$).

Thermal denaturation curves were recorded over a 15-90 °C temperature interval, following the intensity of the CD signal at 200 nm. The temperature range was monitored at a temperature scan rate of 1 °C.min-1, using 2 mm optical path length and 1 nm bandwidth. Reversibility of denaturation was tested by an inverse temperature scanning, immediately after the end point temperature (90 °C) was reached and using the same temperature scan rate.

5. NMR EXPERIMENTS

All spectra were carried out at 293 K, on a 600 MHz Varian Inova AS NMR spectrometer, equipped with a z pulse field gradient unit and a triple resonance probe.

Sequence schemes employing pulsed field gradients were used to achieve suppression of solvent's signals and spectral artifacts. Selective pulses to cover aliphatic, aromatic, or carbonyl ^{13}C nuclear spectral regions were obtained by adiabatic pulse modulation. Quadrature detection in the indirectly detected dimensions was obtained by the hybrid States-TPPI (times proportional phase increment) method or by sensitivity-enhanced Echo-Antiecho combination. Linear prediction was applied to extend the indirect ^{13}C-detected dimension. Direct and

indirect dimensions were normally apodized using 90°-shifted squared sine-bell functions (for ^{13}C- and ^{15}N-edited dimensions) or Lorentzian-to-Gaussian functions (for the ^{1}H dimension). NMR data were processed using the NMRPipe software (53) and analyzed with NMRViewJ (54).

5.1 ^{1}H, ^{15}N and ^{13}C RESONANCE ASSIGNMENTS OF CERATO-POPULIN

The sample used for NMR measurements contained 280 µM of ^{13}C, ^{15}N enriched Pop1 in 20 mM phosphate buffer pH 5.8, 5% D_2O and 0.05% NaN_3.

The ^{15}N-HSQC and ^{15}N-NOESY-HSQC (mixing time 100 ms) spectra, the HNCA, HN(CO)CA, HNCO, HNCACB, HBHA(CO)NH, HNHA, CBCA(CO)NH 3D triple resonance experiments were used for backbone assignment.

For side-chainS assignments, were employed NMR data from ^{13}C-edited HCCH-TOCSY (mixing time 60 ms), HCCH-COSY (mixing time 30 ms) and ^{13}C-HSQC (CHSQC) experiments. Additionally, Two ^{13}C-NOESY-HSQC (mixing time 100 ms) spectra were also acquired: one of them aliphatic-edited, the other aromatic-edited. Tables 2 and 3 summarizes the parameters used for each spectra.

To note that in order to enhance sensitivity and resolution, the aromatic-edited ^{13}C-HSQC spectrum was acquired using the constant time (CT) technique [55].

	Spectral width (Hz)			Complex points		
	^{1}H	**^{15}N**	**^{13}C**	**t_1**	**t_2**	**t_3**
^{15}N-HSQC	4499.94	2430.0	*	1024	512	*
^{15}N-NOESY-HSQC	3603.47	2430.0	8999.88	913	129	129
HNCA	4499.94	2430.0	4524.56	551	129	154
HN(CO)CA	4499.94	2430.0	4524.56	1024	129	129
HNCO	4499.94	2429.98	1977.07	1024	65	65
HNCACB	3304.64	2429.98	12065.56	981	129	129
HBHA(CO)NH	3603.47	2429.98	3656.20	776	129	30
HNHA	4499.94	2429.98	9000.90	1024	65	129
CBCA(CO)NH	4499.94	2429.98	12064.05	1024	129	129

Table 2. Spectral widths and complex points, set for the acquisition of the experiments used for backbone assignments.

	Spectral width (Hz)			Complex points		
	^{1}H	^{15}N	^{13}C	t_1	t_2	t_3
^{13}C-HSQC (aliphatic)	8999.87	21110.89	*	1025	129	*
^{13}C-HSQC (aromatic)	4499.91	4499.94	*	1024	257	*
^{13}C-NOESY-HSQC (aliphatic)	4082.47	4112.77	12064.54	209	27	129
^{13}C-NOESY-HSQC (aromatic)	4499.91	12064.54	8998.87	1024	129	129
HCCH-TOCSY	8999.87	12062.72	8999.87	513	41	209
HCCH-COSY	8999.87	12064.54	8999.87	513	47	225

Table 3. Spectral widths and complex points, set for the acquisition of the experiments used for side-chains assignments.

5.2 RESIDUAL DIPOLAR COUPLING (RDCs)

$^{1}J_{HN-N}$ RDCs were derived from the difference in multiplet splittings of signals in two-dimensional (2D) IPAP ^{1}H-^{15}N-HSQC spectra [56] recorded in isotropic solution (containing 333 µM of ^{15}N labelled Pop1 in 20 mM phosphate buffer pH 5.8, 5% D_2O and 0.05% NaN_3) and after the addition of 11 mg/mL Pf1 phage (ASLA Biotech Ltd) [57].

The BestFit module of the software PALES [58] has been used to fit the experimental RDCs to the 3D model of Pop1. To evaluate the goodness of the fitting the quality factor Q described by Cornilescu was employed [59]:

$$Q = RMS\left(D_{HN}^{obs} - D_{HN}^{calc}\right) / RMS\left(D_{HN}^{obs}\right)$$

Where D_{HN}^{obs} and D_{HN}^{calc} are Pop1's residual dipolar couplings experimentally calculated and the model-based *in silico* calculated ones, respectively.

5.3 BACKBONE ^{15}N RELAXATION EXPERIMENTS

For the NMR ^{15}N relaxation experiments the samples contained a) 624 μM of ^{15}N labelled Pop1 in 20 mM phosphate buffer pH 5.8, 5% D_2O and 0.05% NaN_3; b) 700 μM of ^{13}C, ^{15}N enriched CP in 10 mM phosphate buffer pH 5.8, 5% D_20 and 0.05% NaN_3.

T_1, T_2 and ^{15}N–{^{1}H} NOE experiments were performed using standard pulse schemes [60] in an interleaved manner to collect eight points. T_1 delays were 0.01, 0.21, 0.41, 0.61, 0.81, 1.01, 1.41, 1.81 s, for Pop1, and 0.01, 0.21, 0.51, 0.71, 1.01, 1.31, 1.51, 1.71 s, for CP. T_2 experiments were collected with delays 0.01, 0.05, 0.09, 0.13, 0.17, 0.21, 0.25, 0.29 s for both Pop1 and CP. A repetition delay of 0.2 s was used in both T_1 and T_2 experiments. Heteronuclear NOE spectra were acquired with and without a 3 s saturation period and using a 2 s recycle delay in both saturated and non-saturated experiments.

The rate analysis routine of NMRView [54] was used to determine longitudinal (R_1) and transversal (R_2) relaxation rates by a non-linear two-parameters fit of the time dependence of peak intensities; errors were obtained from the covariance matrix. The heteronuclear NOE values for a given residue were calculated as the intensity ratio of the ^{15}N-^{1}H correlation peak in the presence and absence of proton saturation; errors were estimated from the baseline noise in the two spectra. Typically, errors were of the order of 5% for R_1 and R_2, and 3% for ^{15}N-{1H} NOE, for Pop1, and of 3% for R_1, R_2 and ^{15}N-{1H} NOE, for CP.

Model-free analysis of the relaxation data [61] was performed with the program TENSOR2 [62].

5.4 OLIGOSACCHARIDES TITRATION

According to de Oliveira et al. [13] β-(1,4)-tetramers of N-acetyl-D-glucosamine (Carbosynth Limited) were progressively added to 40 μM solution of ^{15}N-labelled CP, and to 103.5 μM ^{15}N-labelled Pop1 to achieve the protein:oligosaccharide molar ratios 1:2.5, 1:50, 1:100 and 1:200. The perturbation of the proteins' signals after addition of oligosaccharides was evaluated through the acquisition of ^{15}N-HSQC experiments at each step of the titration. The chemical shift perturbation (CSP) effect

was evaluated calculating, for each shifting residue, the total chemical shift variation, $\Delta\delta_{tot}$, according to the equation:

$\Delta\delta_{tot} = [(\Delta\delta_{HN})^2 + (\Delta\delta_N \times 0.154)^2]^{1/2}$

where 0.154 is the weight factor for ^{15}N shifts and $\Delta\delta_{HN}$ and $\Delta\delta_N$ are the chemical shift differences for the HN and ^{15}N respectively, between the free and the titrated proteins [63].

6. HOMOLOGY MODELING

6.1 POP1'S MODEL

The 3D model of Pop1 was obtained by homology modeling using as template the NMR-derived structure of cerato-platanin (PDB 2KQA) [13] with the server SWISS-MODEL [64]. Analysis of the structure quality was performed combining the results of PROCHECK [65], ProSA-web Z-scores [66], WHATIF [67, 68] and MODFOLD [69]. Residual dipolar couplings (see 5.2) were used for structure validation.

6.2 LRR-CONTAINING DOMAIN MODELS

The LRRSearch database [70] and the secondary structure prediction algorithm PSI-PRED [71] were used to identify possible leucine rich repeats in target sequences. Sequences have been then randomly fragmented in peptides containing at least 30 amino acids and 1 LRR.

Homology modeling of every fragment was performed with the Alignment routine of SwissModel [64]; one template per fragment was chosen evaluating percentage of identity, QMEAN4 factors [72, 73] and GMQE values [74]. The templates were used to build multiple template alignments. The obtained alignments were then used as starting point to build, with the dopehr_loopmodel protocol of MODELLER v.4 [75], multiple-templates-based homology models of the LRR-containing domain of every receptor.

Quality of the final models was evaluated with PROCHECK [65], ModFOLD [69] and ProQ [76].

Results and discussion

7. THE STRUCTURE OF CERATO POPULIN

7.1 RESONANCE ASSIGNMENTS OF POP1

Resonance assignments has been the starting point of all the subsequent NMR-based analysis. Through the analysis of ^{1}H-^{15}N HSQC, HNCA, HNCACB, CBCACONH, HNHA and HBHA(CO)HN experiments 93.6% of ^{1}H, 85.9% of ^{15}N and 91.2% of ^{13}C were assigned (including the four initial residues EAEA). The ^{1}H-^{15}N HSQC of Pop1 is shown in Figure 6.

Several overlapping peaks were found: Tyr^{36} and Asp^{110} (N 117.38, H 8.301); Asp^{60}, Phe^{78} and Val^{125} (N 119.58, H 8.268); Val^{48} and Ser^{102} (N 113.43, H 8.047); Phe^{45} and Gln^{112} (N 119.45, H 8.149). Moreover, many residues showed multiple peaks due to conformational exchange: noteworthy, 95% of these peaks are located in the region between Cys^{20} and Cys^{57}, predicted as unstructured.

Glu^{1}, Val^{12} and Asn^{59} were missing in the ^{1}H-^{15}N HSQC: their HN values were obtained analyzing a ^{15}N-edited NOESY HSQC experiment (except for Glu^{1}, of which amide resonances are missing), whereas their CA, CB, HA and HB through the analysis of CBCACONH and HBHAHN spectra. The chemical shift values of the backbone atoms obtained from the HNCACB experiment, were used as anchor points for the aliphatic side-chain resonance assignments, through the analysis of ^{1}H-^{13}C HSQC, HCCH-COSY, HCCH-TOCSY, ^{15}N-edited NOESY HSQC and ^{13}C-edited HSQC spectra. In order to assign aromatic rings, a 2D constant time ^{1}H-^{13}CHSQC and a 3D ^{13}C-edited NOESY-HSQC, filtered for the aromatic region, were analyzed. Finally carbonyls assignments were performed through the analysis of an HNCO spectrum.

The ^{1}H, ^{13}C and ^{15}N chemical shift assignments have been deposited in the BioMagResBank database (http://www.bmrb.wisc.edu) under the accession number 19321.

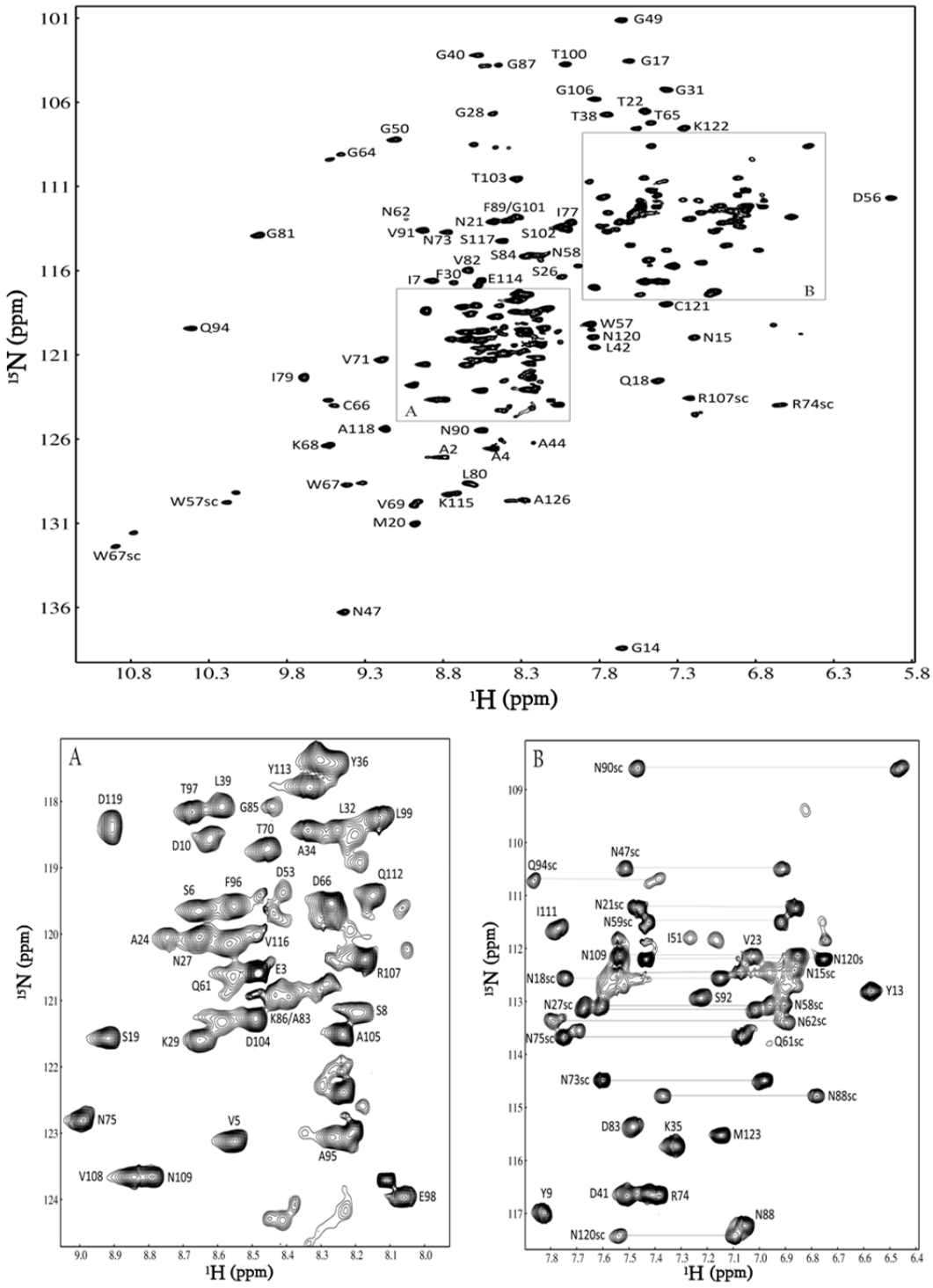

Figure 6. ^{1}H-^{15}N HSQC of Pop1, depicting the HN chemical shift of all residues (except for the missing ones, Glu^{1}, Val^{12} and Asn^{59}). In the case of multiple signals (conformational exchange) only the peak representing the major populated conformational state is labeled.

7.2 THE MODEL OF POP1

Taking advantage of the high sequence identity between Pop1 and CP (62%), figure 1A, the NMR-derived structure of CP (PDB 2KQA) [13] was used as template to build an homology model of Pop1 (Figure 7).

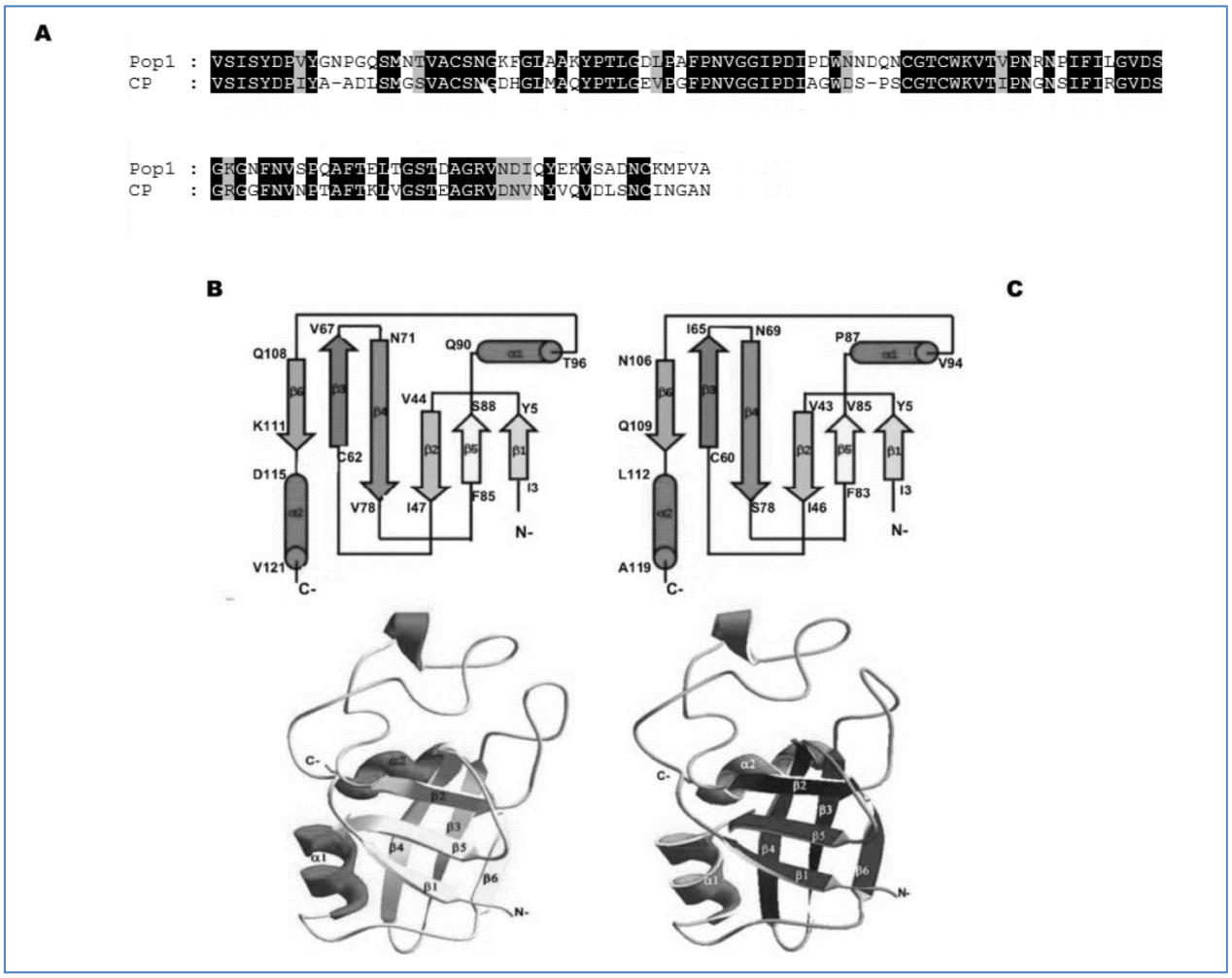

Figure 7. (A) Pop1 and CP's sequences alignment: in black and grey are reported identities and similarities, respectively. (B-C) Topological and ribbon representation of Pop1 (B) and CP's (C) 3D structures

The 3D model reveals a fold highly similar to the CP one (calculated backbone atoms RMSD is 0.61 Å): Pop1 exhibits a globular fold containing 2 α-helices and 6 β-strands forming a double ψβ-barrel, and the 2 disulfide bonds (Cys^{21}-Cys^{59}; Cys^{62}-Cys^{117}) conserved amongst the members of the CPF.

The model is consistent with the secondary structure predicted by the far-UV circular dichroism spectrum shown in figure 8B: β-sheet and random coil regions are predominant, with a lesser contribute of helical elements.

The Chemical Shift Index secondary structure prediction was used as further indication of agreement of the model to experimental data. Backbone protons (HA and HN) chemical shifts were used to predict the secondary structure of Pop1, according to the algorithm proposed by Wishart [77]: (Figure 8A)

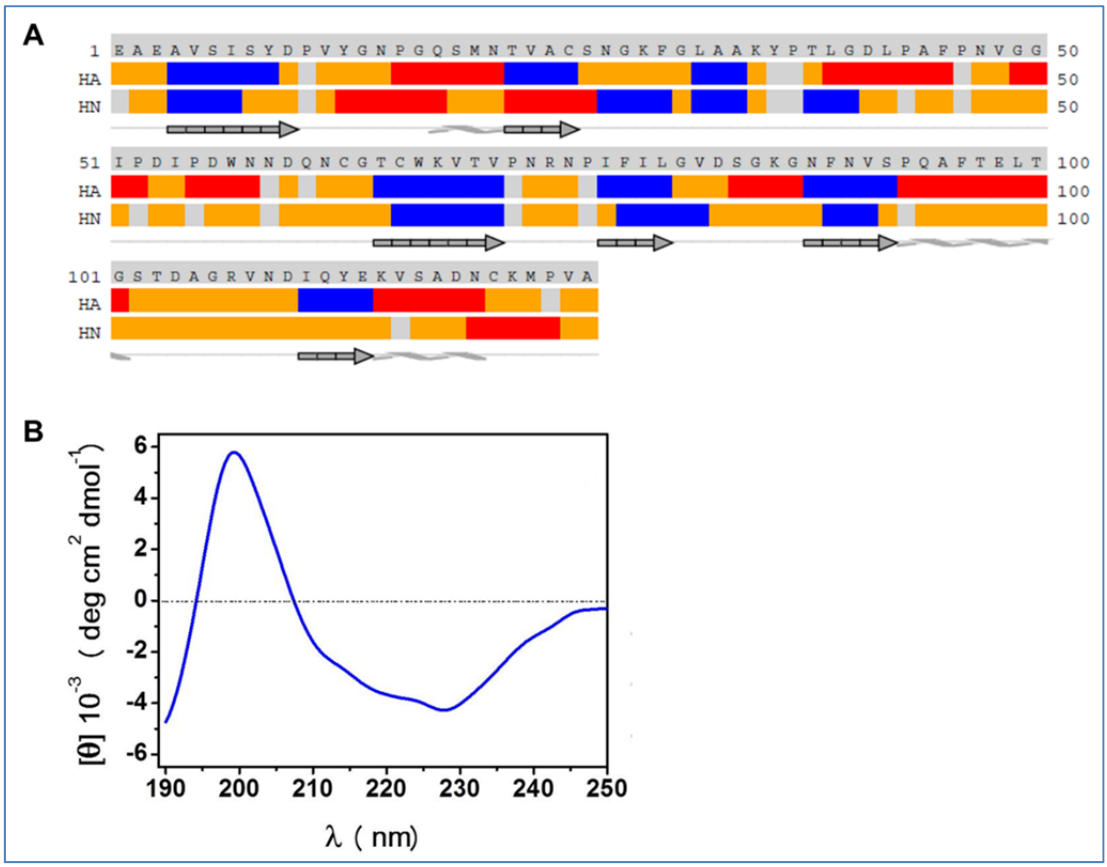

Figure 8. Chemical Shift Index prediction (A) and Circular Dichroism spectrum of Pop1 (B). The CD spectrum is typical of a well-structured α/β protein, with a positive peak at 200 nm and a single negative peak at 228 nm. Also, these values suggest an important contribution to the spectrum of random coil regions.

All the structural and stereochemical quality checks performed, addressed the model as good. ProSA-web Z-scores and PROCHECK Ramachandran plots are reported in figure 9. The Ramachandran plot showed 81.2% of the residues located in the most

favored regions, 11.5% in the additional allowed ones and 5.2% in the generously allowed ones, thus indicating good stereochemical quality of the model. Only Arg^{70} and Ala^{40} were found to be in disallowed regions (2.1%): the fact is justified observing that, in the alignment Pop1/CP, those residues correspond to two glycines in CP.

Z-score values are indicative of overall model quality, indicating whether the input structure is within the range of scores typically found for native proteins of similar size. The values reported for the homology model (-6.26) and for the template (-5.54), also confirmed good similarity between template and query structure.

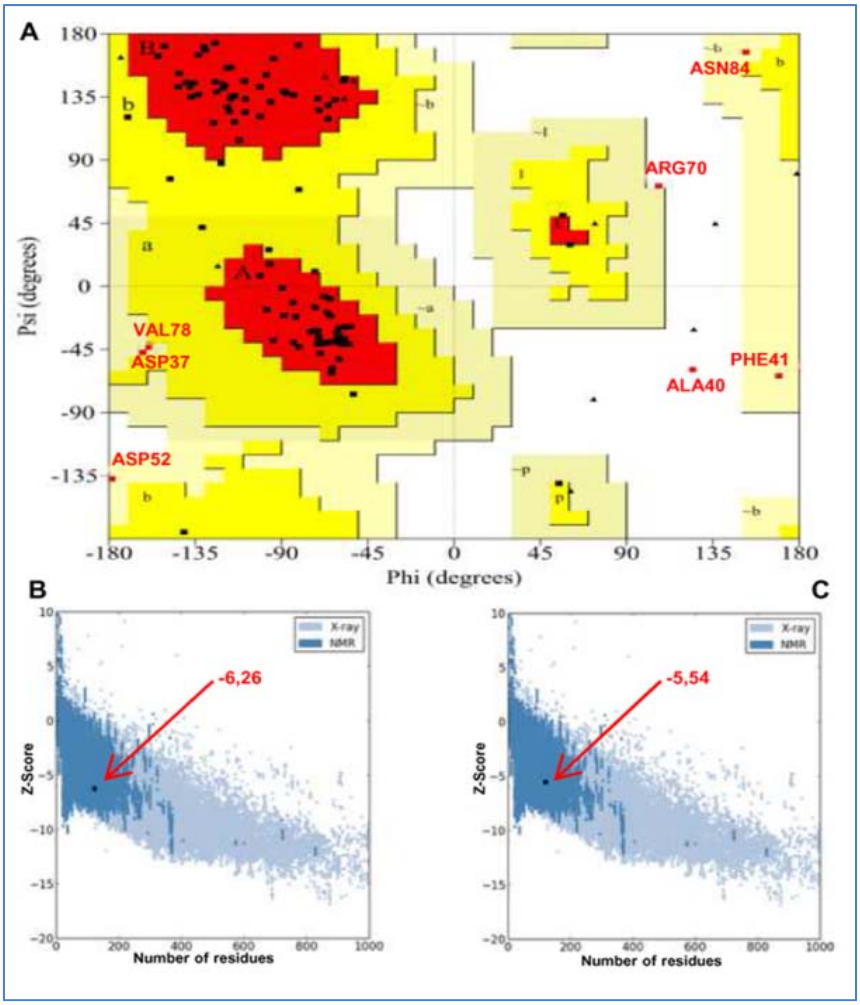

Figure 9. A) Ramachandran plot of Pop1, reporting the residues plotted in the most favored regions, 81.2% (A, B, L), in the additional allowed ones, 11.5% (a,b,l,p) and in the generously allowed ones, 5.2% (~a, ~b, ~l, ~p). Arg^{70} and Ala^{40} are plotted in disallowed regions. Glycines are shown as triangles. B) Z-score of Pop1's model and C) Z-score of CP's NMR-derived structure.

To validate Pop1 model a set of NMR-derived $^{1}J_{HN-N}$ Residual Dipolar Couplings (RDC) was fitted to the calculated structure.

Residual dipolar couplings are anisotropic magnetic interactions, originating from a partial alignment of the molecule with respect to a magnetic field. Proteins, in the presence of a magnetic field have preferred orientations due to their anisotropic magnetic susceptibility tensor. If the alignment is kept sufficiently low (partial), dipolar interactions, chemical shift anisotropy and electric quadrupole interactions are averaged. The resulting so-called *residual* anisotropic magnetic interactions can be used for a variety of applications, from dynamics to orientational restraints in structure determination.

In this studies Residual Dipolar Couplings have been used for structure validation, through the comparison of model-based RDCs calculated *in silico* with those obtained by inducing partial alignment of Pop1 in solution (the chosen alignment media was the filamentous phage pf1). The singular value decomposition method of PALES [78] was used to provide an initial guess of the magnitude and orientation, with respect to the external magnetic field, of the molecular alignment tensor. Subsequent optimization during the simulated annealing process led to the final values Da^{NH} = 7.265 Hz and R = 0.428, for the N-H normalized magnitude of the RDC tensor and rhombicity, respectively. Because of the overlap of peak multiplets and 13 missing residues in Pop1 ^{15}N-HSQC spectra, RDCs were obtained for 82 out of 122 residues (Appendix A).

The Q factor was used as indication of goodness of the fitting between model RDCs and NMR-calculated ones. The value 0.65 (figure 10A), is not satisfactory for a model assessed as very good by PROCHECK and experimental data (Q factor ranges from 0, indicative of perfect agreement, to 1, meaning bad fitting). This result, however, is not surprising considering that 63% of Pop1 residues are in random coil regions (Figure 7). Indeed, figure 10B shows that, if only the 32 RDCs associated to residues located in well-defined secondary structure elements are considered, the Q factor drops to 0.43 that can be considered a good value for a structure obtained by homology modeling.

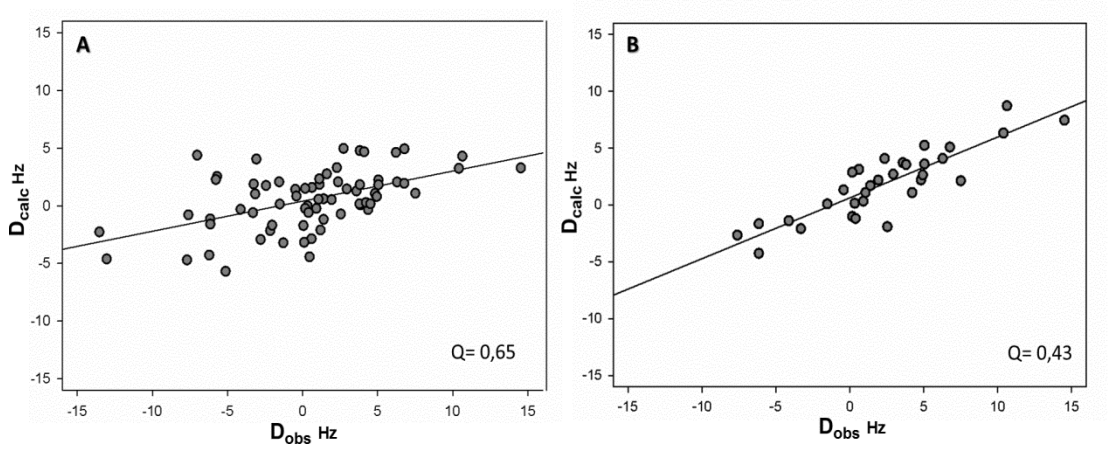

Figure 10. The correlation plots show the degree of agreement between experimentally measured RDCs and those calculated from the homology-built model. (A) shows the Q factor calculated when considering the entire set of RDCs obtained (82 $^{1}J_{HN-N}$); (B) reports data when only residues located in secondary structure are considered (32 $^{1}J_{HN-N}$).

8. STRUCTURAL BASIS OF THE DIFFERENT ELICITATION ACTIVITY OF POP1 AND CP

In spite of their similar fold, circular dichroism spectroscopy indicates that Pop1 and CP exhibit some differences in their secondary structure. In fact, CP, although it presents the same 200 nm positive peak observed in Pop1, exhibits two additional negative peaks at 210 and 230 nm. These spectral features suggest that CP secondary structure is characterized by longer helices, and strands, while Pop1 has a higher percentage of random coil regions (Figure 11A). Differences emerge also with respect to Pop1's thermal stability: monitored over a 15-90° C interval, the protein turns out to be resistant to heat denaturation up to 55 °C and completely unfolded at about 80 °C; the melting temperature (Tm) has been estimated at about 73.8 °C (Figure 11B). On the contrary, according to de Oliveira et al. [13] CP is slightly more resistant to thermal denaturation, with a Tm = 76° C and denaturation starting at 60° C. For both Pop1 and CP, the unfolding is not fully reversible as only 62% of the secondary structure can be recovered.

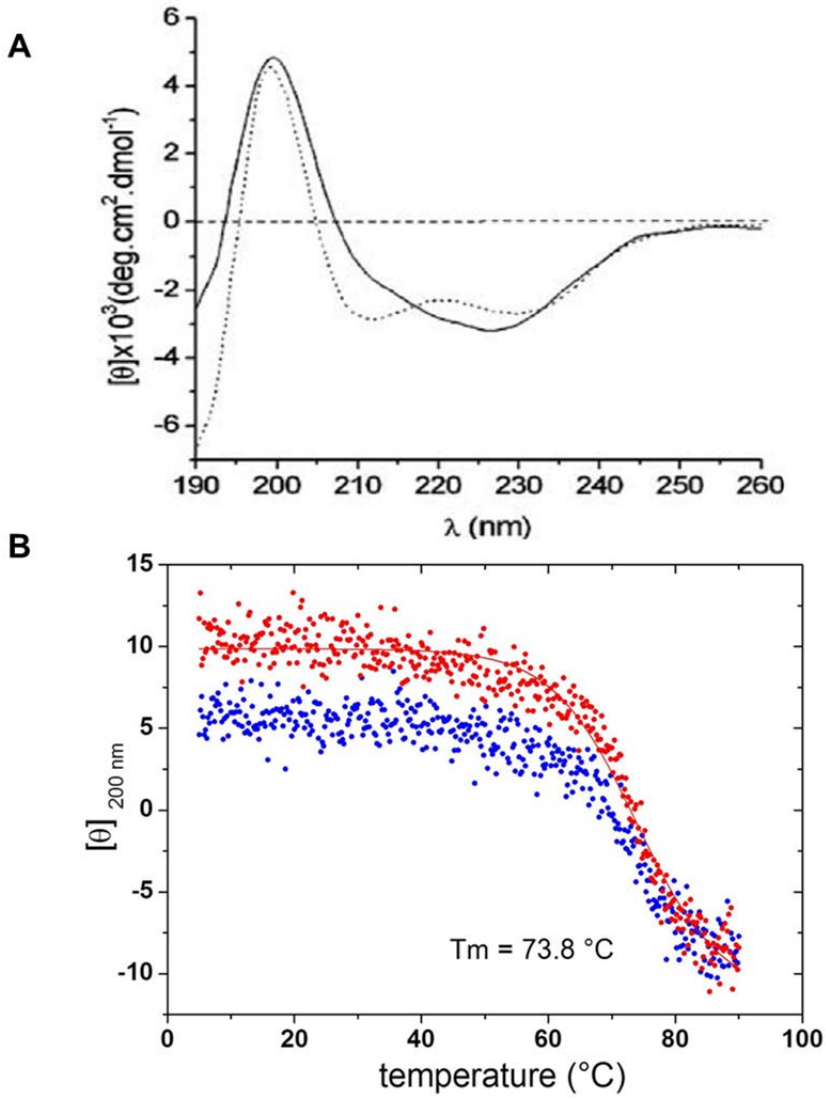

Figure 11. A) Far-UV CD spectra of Pop1 (black curve) and CP (dotted curve). CP spectrum was obtained by the data previously published [13]. B) Thermal denaturation (red) and renaturation (blue) curves of Pop1.

All together, these data indicate that the amino acid composition of the two proteins influences some structural features of Pop1 and CP. We envisage these dissimilarities may be the clue to understand the different timing of plant's defense response they induce.

Recently, Frias et al. [16], published results on the cerato platanin BcSpl1 from *B. Cinerea* indicating that two regions, located on the surface-exposed loops β1-β2 and β2-β3, were sufficient and necessary to induce necrosis in plants' leaves and to

express virulence ability. The authors pointed out, in addition, that the reciprocal interaction of these loops is crucial for the protein activity. Those two loops are highly conserved among the CPF (Figure 2B). The corresponding regions span the sequences Val[19]-Leu[28] and Ile[50]-Cys[59], for Pop1, and Val[18]-Leu[27] and Ile[49]-Cys[57] for CP, in loops β1-β2 and β2-β3, respectively (Figure 12A).

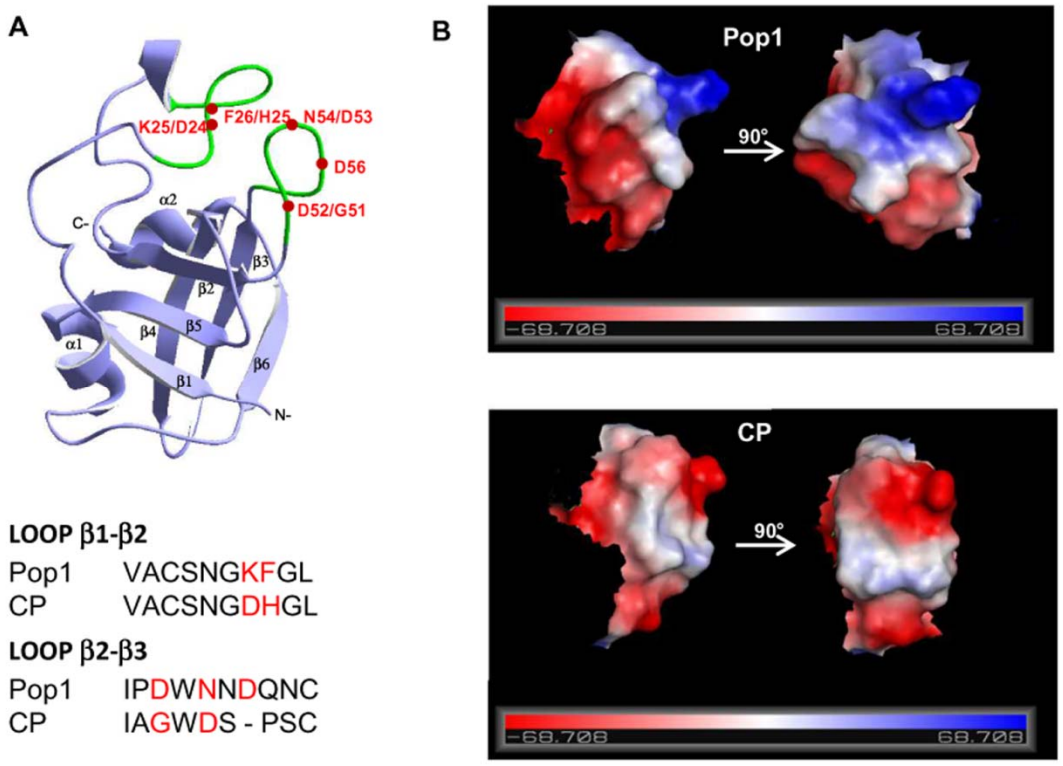

Figure 12. (A) 3D model of Pop1: in green are underlined loops β1-β2 and β2-β3, located on the edge of the ψβ-barrel (upper panel). In red are reported the substitutions and the insertion contributing to the different charge distribution (lower panel). (B) Loops' surface charge: in red are colored negative charges, while in blue positive ones. Neutral regions are reported in white. Electrostatic potential surfaces were calculated with PyMOL Molecular Graphics System, Version 1.2r3pre.

The analysis of the loops primary sequences shows, for the highly conserved loop β1-β2, the substitutions Lys[25]/Asp[24] and Phe[26]/His[25] (Pop1/CP) and, for the less conserved loop β2-β3, not only two mutations Asp[52]/Gly[51] and Asn[54]/Asp[53] but also an insertion in Pop1, residue Asp[56] (Figure 12A). The result of those mutations and insertion is a complete modification of the proteins' surface electric charge: CP

exhibits a negative and a neutral region of comparable size, while Pop1 on one side presents a negative surface and on the other, an extended positive region flanked by a small neutral one (Figure 12B).

Since we expect these features may affect *i)* the spatial arrangement of the loops, *ii)* their backbone dynamics, and *iii)* the interaction of the proteins with other molecules/receptors, those aspects were investigated.

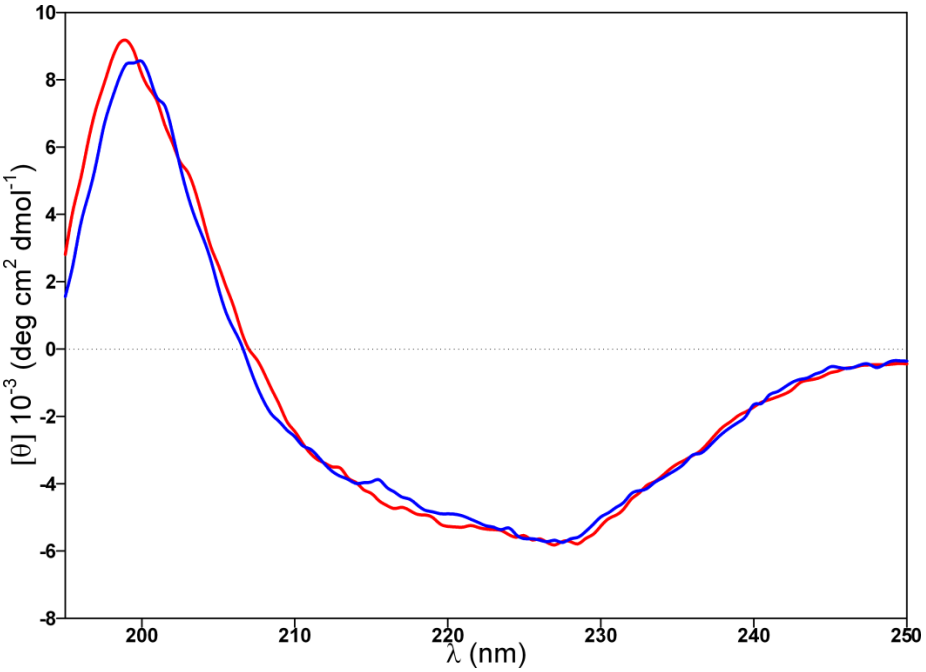

Figure 13. Far-UV (190-250 nm) circular dichroism spectra of Pop1 at pH 5.8 (blue curve) and pH 3 (red curve).

8.1 SPATIAL ARRANGEMENT OF THE LOOPS: ^{1}H-^{1}H NOES ANALYSIS

The study of the spatial arrangement and interactions of residues in disordered regions can be difficult due to the intrinsic impossibility to identify regular pattern in random coils. In the case of Pop1, studies were complicated by the lack of an experimental structure, accurate enough to measure distances with molecular visualization programs. Therefore, to highlight differences in the spatial organization of the loops in CP and Pop1 their ^{1}H-^{1}H NOEs networks were analyzed.

Nuclear Overhauser Effect (NOE) is the change in spin population of one nucleus when another magnetically active one, close in space (up to 5 Å), is saturated. The

signals originating by this interaction can be used to measure intra (and inter) molecular distances and to map the space around a certain residue. Moreover, considering that several significant interactions such as salt bridges, intramolecular hydrogen bonds and van der Waals interactions are normally in the range of 4 Å, NOEs can be also a useful tool to map interactions between residues.

Three 3D NOESY spectra (^{15}N-edited HSQC, aliphatic and aromatic ^{13}C-edited HSQC) were carried out for both proteins at pH 5.8. In the case of Pop1, because of the severe peaks overlap and poor resolution observed in the spectra at pH 5.8, aliphatic and aromatic ^{13}C-edited spectra were acquired also at pH 3. Resonance assignments of ^{13}C and ^{1}H and circular dichroism spectra (Figure 13) confirmed that the pH change did not alter the overall secondary and tertiary structure.

Pop1 loop β1-β2		**CP loop β1-β2**	
V19, HG21	D52, HA	**V18, HG21**	I49, HG12
A20, HN	D79, HA	**A19, HN**	D77, HA
C21, HB2	I50, HG13	**A19, HB1**	D77, HA
K25, HG3	N54, HB2	**H25, HB3**	D48, HB3
K25, HG3	W53, HB3	**L27, HD21**	I49, HD11
K25, HG3	N54, HB2		
K25, HG3	I50, HG13		
F26, HZ	P51, HA		

Table 4. NOEs analysis of Pop1 and CP's loops β1-β2. Only NOE between residues on the opposite loop or on different domains of the protein are reported.

Pop1 loop β2-β3					**CP loop β2-β3**	
I50, HN	G81, HA3		**W53, HZ3**	F26, HB2	**I49, HB**	L27, HD21
I50, HN	K82, HG3		**N54, HN**	G24, HN	**I49, HD11**	H25, HB3
D52, CA	K25, HD2		**N54, HN**	F26, HB2	**I49, HD11**	D57, HN
W53, HN	G27, HA3		**N54, HN**	F26, HB2	**A50, HN**	S58, HA
W53, HN	S20, HB3		**N55, CB**	F26, HB2	**G51, HN**	S58, HA
W53, HN	V19, HG11		**N55, CB**	L28, HD11	**W52, HE1**	S21, HG
W53, HD1	F26, HB2		**N58, ND2**	K111, HB2	**S56, HA**	A19, HB1
W53, HH2	F26, HA		**N58, ND2**	K111, HG3		
W53, HZ3	F26, HN		**C59, HN**	F26, HB3		
W53, HZ3	G24, HA2		**C59, HN**	K25, HG3		

Table 5. NOEs analysis of Pop1 and CP's loops β2-β3. Only NOE between residues on the opposite loop or on different domains of the protein are reported.

Analysis of the NOEs showed a greater number of interactions between loops β1-β2 and β2-β3 in Pop1 than in CP (Tables 4 and 5).

Based on NOEs detected between Pop1 Lys^{25}, loop β1-β2, and Asp^{52}, loop β2-β3, as well as with other residues nearby (Figure 14A), we hypothesize the formation of a salt bridge between Lys^{25} and Asp^{52}, not possible in CP: that salt bridge is expected to keep the two loops in close contact and possibly to restrict their conformational freedom. Moreover, the high number of NOEs between Phe^{26} and Trp^{53} in Pop1 suggests a strong hydrophobic interaction between the two aromatic rings, again not observed in CP between His^{25} and Trp^{52}.

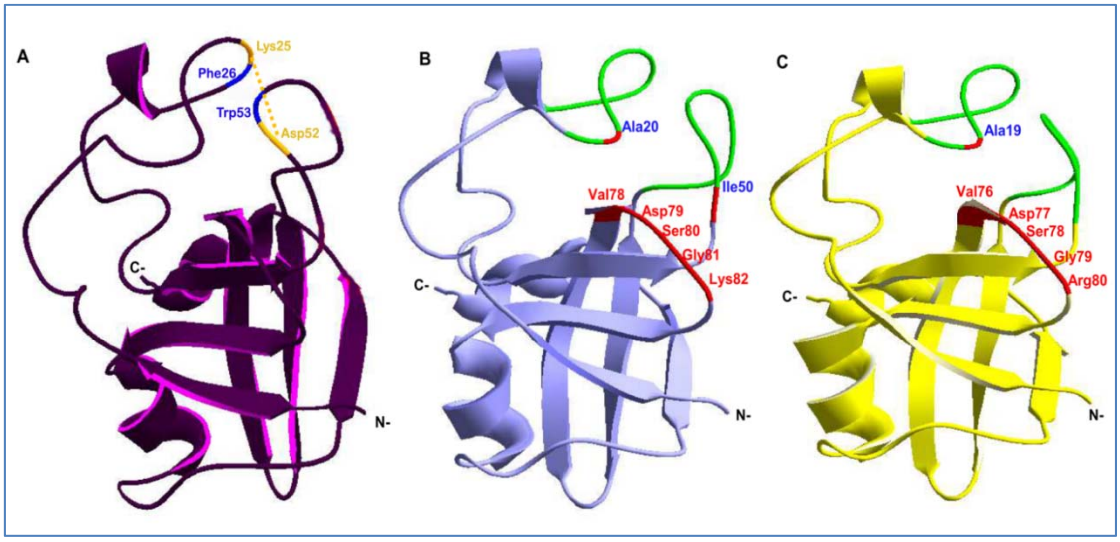

Figure 14. (A) 3D model of Pop1, reporting the hypothesized interactions between loops β1–β2 and β2–β3, based on NOEs data reported in Table 4. (B,C) Plot of the additional contacts between the loops (green) and the conserved motif on loop β4–β5 (red). In Pop1 (B) were found the interactions Ala^{20}/Asp^{79}, Ile^{50}/Gly^{81} and Ile^{50}/Lys^{82}; in CP (C) the only NOE observed was between Ala^{19} and Asp^{77}.

Interestingly, in both proteins we found additional contacts between these loops and a conserved five-residues motif in the short β4-β5 loop (Val^{78}, Asp^{79}, Ser^{80}, Gly^{81}, Lys^{82} in Pop1 and Val^{76}, Asp^{77}, Ser^{78}, Gly^{79}, Arg^{80} in CP). Also in this case, for Pop1 we detected a higher number of NOEs than for CP, especially between loops β4-β5 and β2-β3 (Figure 14B-C and Table 4).

This more extended network of interactions found in Pop1 points to a different arrangement of loops β1-β2 and β2-β3, not only with respect to each other, but also with respect to loop β4-β5.

8.2 BACKBONE DYNAMIC ANALYSIS

A complete set of ^{15}N relaxation experiments (T_1, T_2 and $^{15}N–\{^1H\}$ NOE) was carried out for both Pop1 and CP at 20°C and pH 5.8. (Figure 15).

Relaxation is the process that describes the mechanisms by which an excited nucleus returns to its equilibrium. The longitudinal relaxation time T_1 and the transverse relaxation time T_2 describe the time that a nucleus needs to fully relax by two different mechanisms: spin-lattice relaxation (T_1) and spin-spin relaxation (T_2). These two parameters can be complemented, eventually, by the analysis of the heteronuclear NOEs signals between ^{15}N and ^1H amide nuclei (NOE can be considered as another mechanism of relaxation when certain conditions are met, see 8.1).

The main advantage of describing backbone dynamics by NMR relaxation parameters (following the relaxation of ^{15}N nuclei), is the possibility to gain information about motions in the ps-ms timescale (overall tumbling, fast and slow loop motions, side chain rotation/reorientation, ecc.).

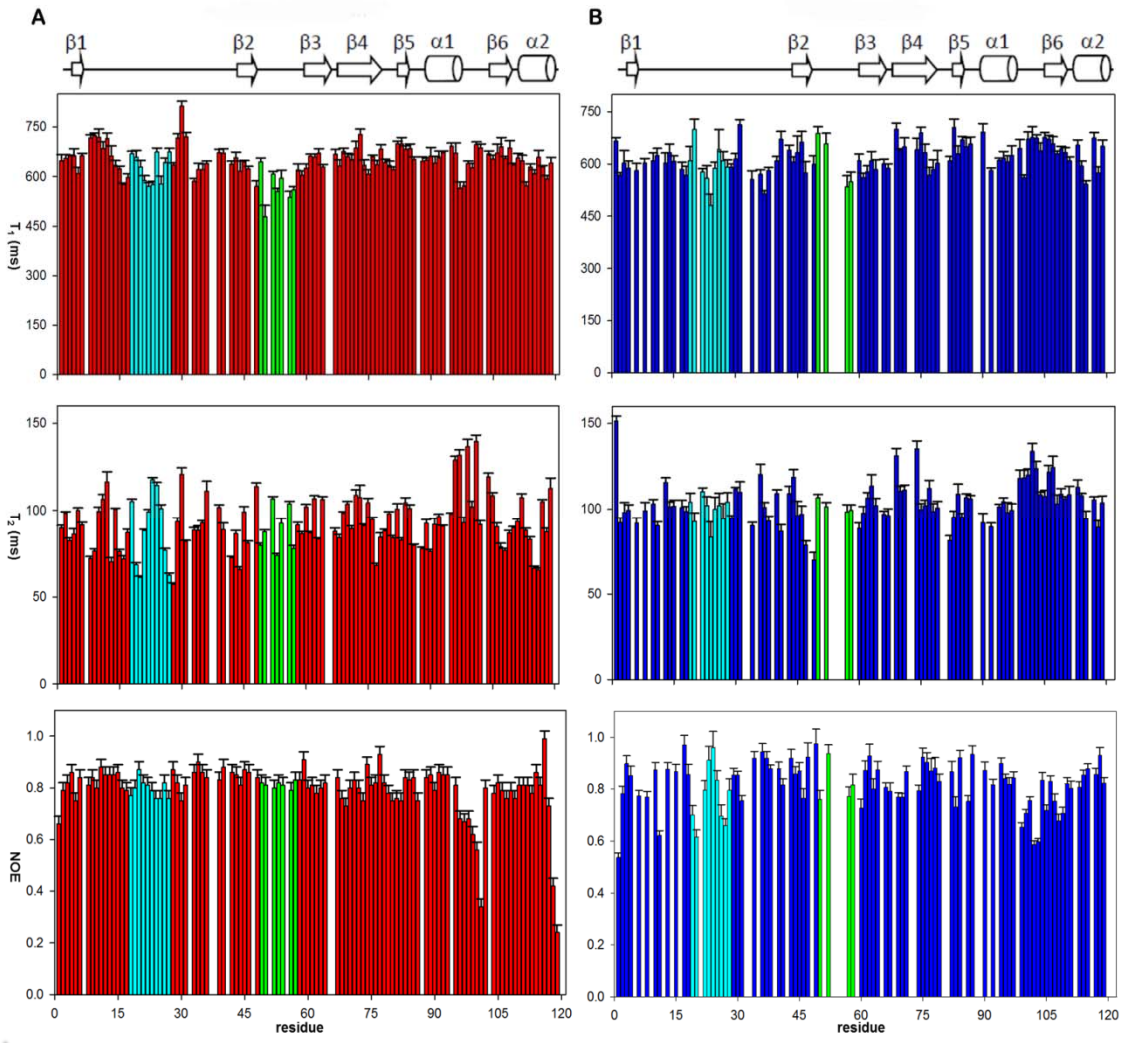

Figure 15. T_1, T_2 and heteronuclear NOE data for both CP (A) and Pop1 (B). Signals of residues 37, 51, 65, 103 (CP) and 5, 9, 16, 21, 32, 35, 53, 54, 55, 56, 62, 65, 73, 80, 88, 93, 98, 112, 116, 121 (Pop1) were either too weak or severely overlapping for a reliable analysis of peak intensities. Four residues in Pop1 (Gln^{14}, Gly^{81}, Phe^{89} and Ala^{91}) exhibited heteronuclear NOE values higher than the theoretical maximum of 0.97 indicating an underestimation of their errors or chemical exchange with the solvent [79]. These residues were not considered in the analysis of proteins' dynamics as long with all the prolines and the highly flexible C-terminus residues (Pop1's Ala^{122}, CP's Asn^{120}). In cyan are colored residues of loop β1–β2 and in green residues of loop β2–β3.

The relaxation data were similar for both proteins: average T_1 value 617 ms and 644 ms, average T_2 104 ms and 92 ms, average NOE 0.81 and 0.79 for Pop1 and CP, respectively.

Assuming Pop1 and CP overall motion to be isotropic, the program TENSOR2 was used to calculate from T_1/T_2 ratios (Figure 16) the overall correlation time (τc) to describe the rotational diffusion tensors of both proteins.

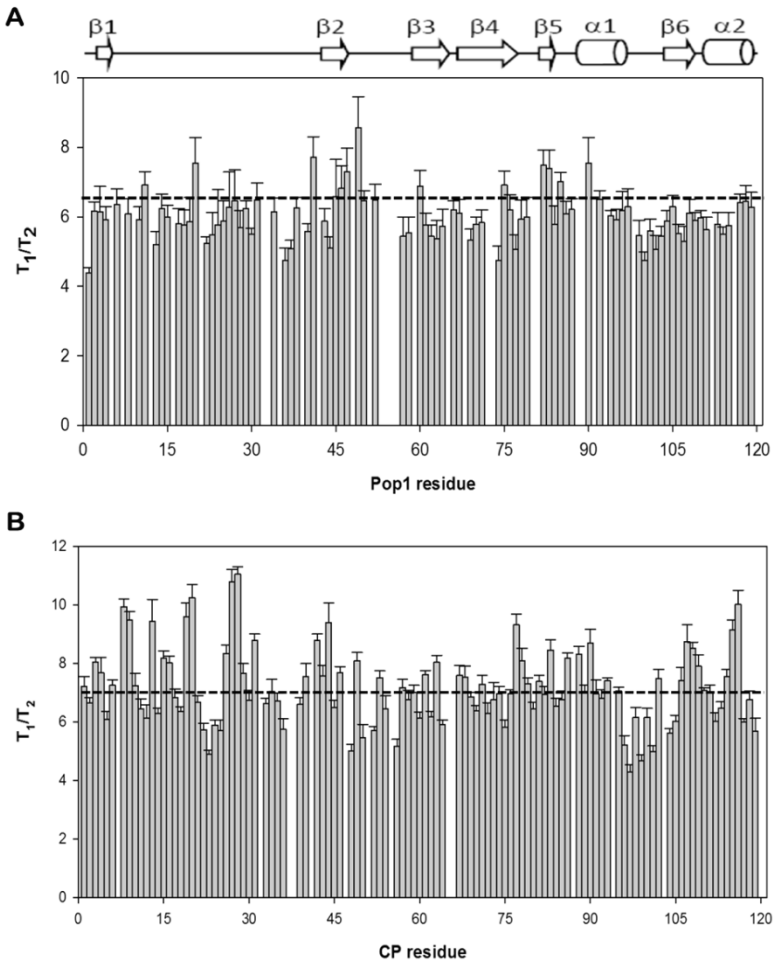

Figure 15. Plot of T_1/T_2 ratios for Pop1 (A) and CP (B). The dotted lines represent the populations' average T_1/T_2 ratios (6.50 for Pop1 and 7.15 for CP).

In the calculation of the tensor, according to previous works [80,81] those residues for which internal motions contributed significantly to the observed relaxation parameters were not included: of the 88 amide of Pop1, for which relaxation parameters were available, 18 were excluded, 5 because exhibited small $^{15}N–\{1H\}$ NOE values (<0.65) and 13 because exhibited T_1/T_2 ratios greater than one standard deviation from the mean value. For CP, of the 106 amide, 52 were excluded from the

calculation, 4 because NOE<0.65 and 48 because T_1/T_2 greater than one standard deviation from the mean value (Figure 14 and Figure 15). The estimated correlation times resulted τc = 6.98 ns for Pop1 and τc = 7.36 ns for CP, both in agreement with proteins in a monomeric state (according to Cavanagh [82], the theoretic correlation time is 6.99 ns for Pop1 and 7.02 ns for CP).

The obtained τc were used to calculate the order parameter S^2 for each residue, using the model-free approach as described by Lipari and Szabo [61].

In a model-free approach every residue motion can be represented by five different models, with respect to τc. Each model is described by different parameters: model 1 (S^2), 2 (S^2, T_e), 3 (S^2, K_{ex}), 4 (S^2, T_e, K_{ex}) and 5 (S^2_{slow}, T_e, S^2_{fast}). Model 6 is chosen when none of the five proposed fits can reproduce the relaxation data satisfactorily. In this analysis, the model which best describes the dynamics of the NH bond is chosen by the statistical approach outlined by Mandel et al. [83].

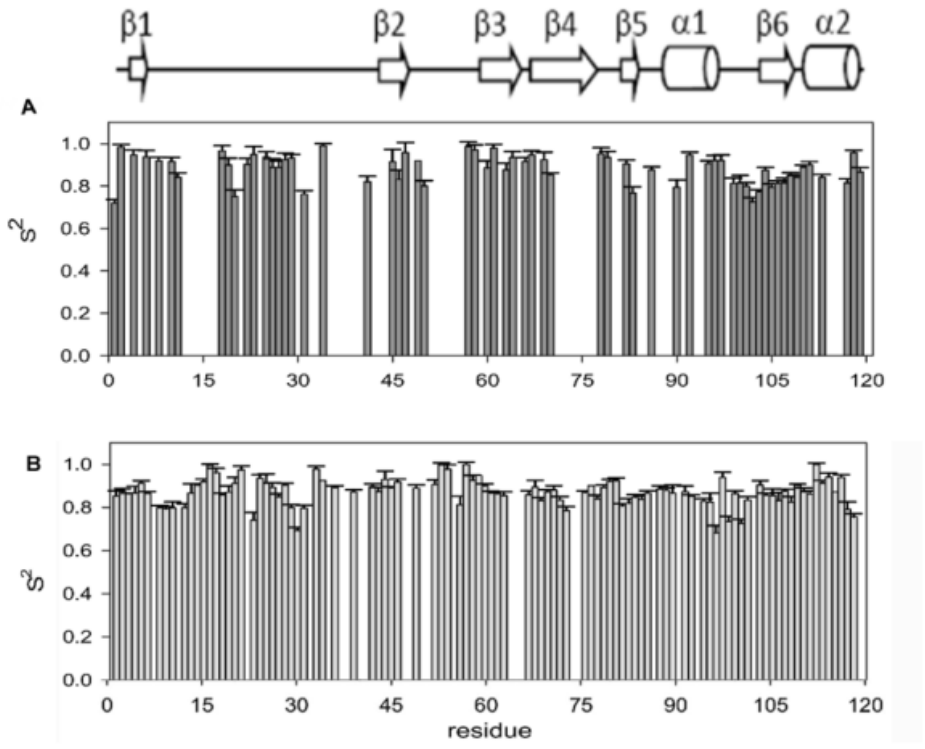

Figure 16. S^2 values are plotted against Pop1 (A) and CP (B) sequences. Residues that could not be fit with any model in the Lipari-Szabo analysis (model 6), are excluded from the graph (see Appendix C).

A detailed report of model-free parameters and chosen models for every residue can be found in Appendix B.

Figure 16 reports the calculated S^2 values plotted against Pop1 and CP's sequences. The order parameters S^2 reflects the motional restriction of the amide N-H bond vector on the ps–ns time scale, describing the fast internal motion relative to the overall rotational diffusion: it ranges from 0, complete flexibility, to 1, absence of flexibility. In both proteins it is consistent with an overall restricted structural flexibility (average S^2 value is 0.88 for both Pop1 and CP): only few residues present $S^2<0.75$ (Val^{1}, Ala^{20} and Gly^{102} in Pop1; Gly^{23}, Gln^{30}, Thr^{97}, Ala^{99} and Arg^{101} in CP). Interestingly, in both cases most of the residues belonging to the loops show values comparable to those of the structured regions.

Figure 17 shows TENSOR2 results for the loops of the two proteins: indeed their residues present similar S^2 values, averaging around 0.88. Exception are Ala^{20} of Pop1 and Gly^{23} and Ala^{50} of CP that exhibit a lower value.

It is worth noting that, for some residues, it has been necessary to introduce the parameter K_{ex}, a term reflecting the contribution of conformational exchange to the residues' dynamical models [84]: in particular residues Ala^{19}, Cys^{20}, Gly^{26} and Leu^{27} of loop β1-β2 of CP; Ala^{20} of loop β1-β2 of Pop1 and His^{49}, Ala^{50}, Cys^{53} and Cys^{57} of loop β2-β3 of CP required K_{ex} values >1.5. (values that normally identify residues experiencing significant conformational exchange on the μs–ms time scale).

Overall these values point to a higher conformational freedom of the CP's loops.

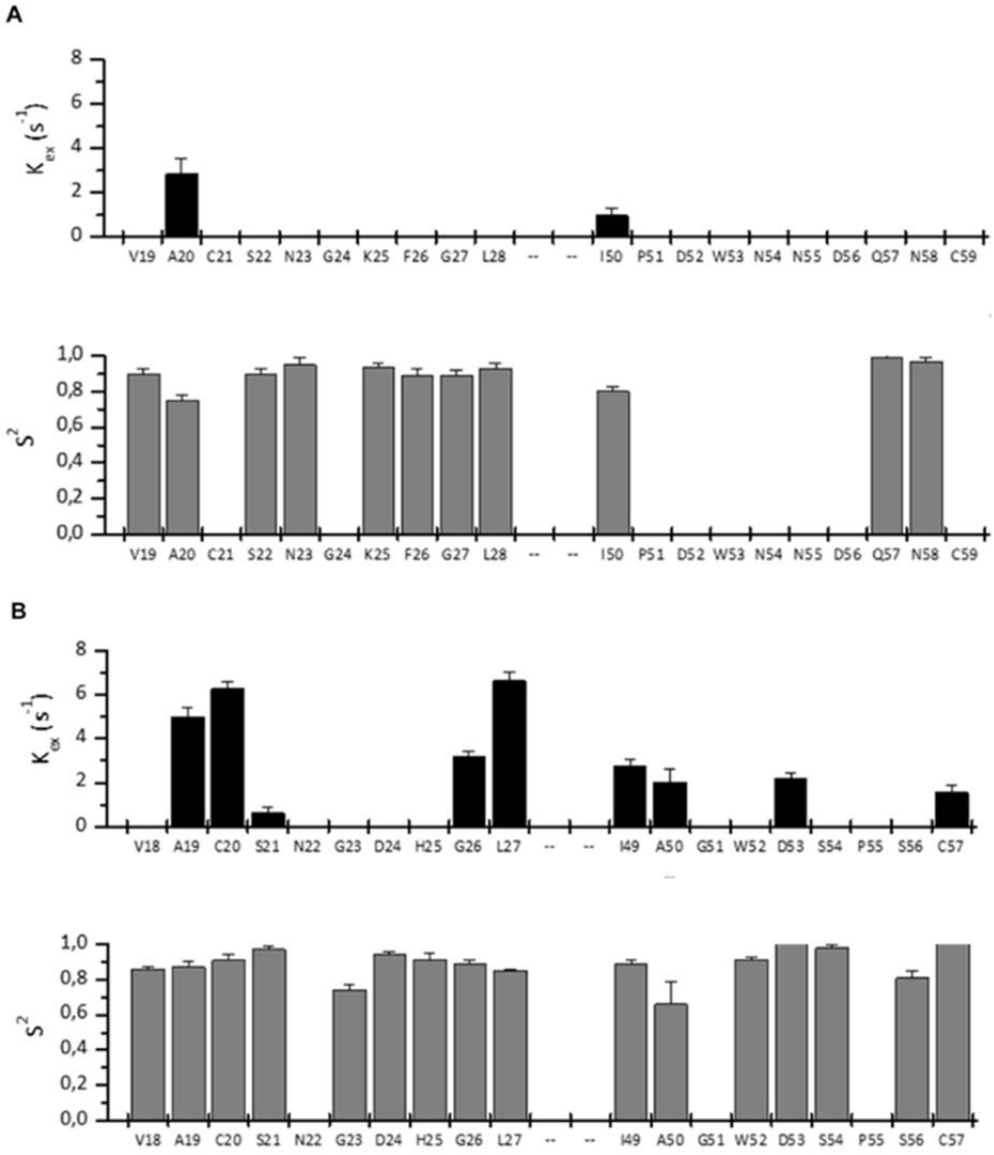

Figure 17. Model-free analysis of Pop1 (A) and CP's (B) loops. Model-free parameters and associated errors are plotted versus the proteins' residues: S^2 (order parameter), and K_{ex} (exchange rate). Residues that could not be fit with any model in the Lipari-Szabo analysis are excluded from the graph (see Appendix C).

8.3 INTERACTION OF CP AND POP1 WITH OLIGOSACCHARIDES

It was reported that loops β1-β2 and β2-β3 of CP are involved in the binding of oligosaccharides [13] and that CP binds to colloidal chitin very quickly with B_{max} = 2.07 ± 0.19 μmol/g+ and K_d = 40.56 ± 9.50 μM [14]. Pop1, instead, binds more slowly and to a lesser extent [14].

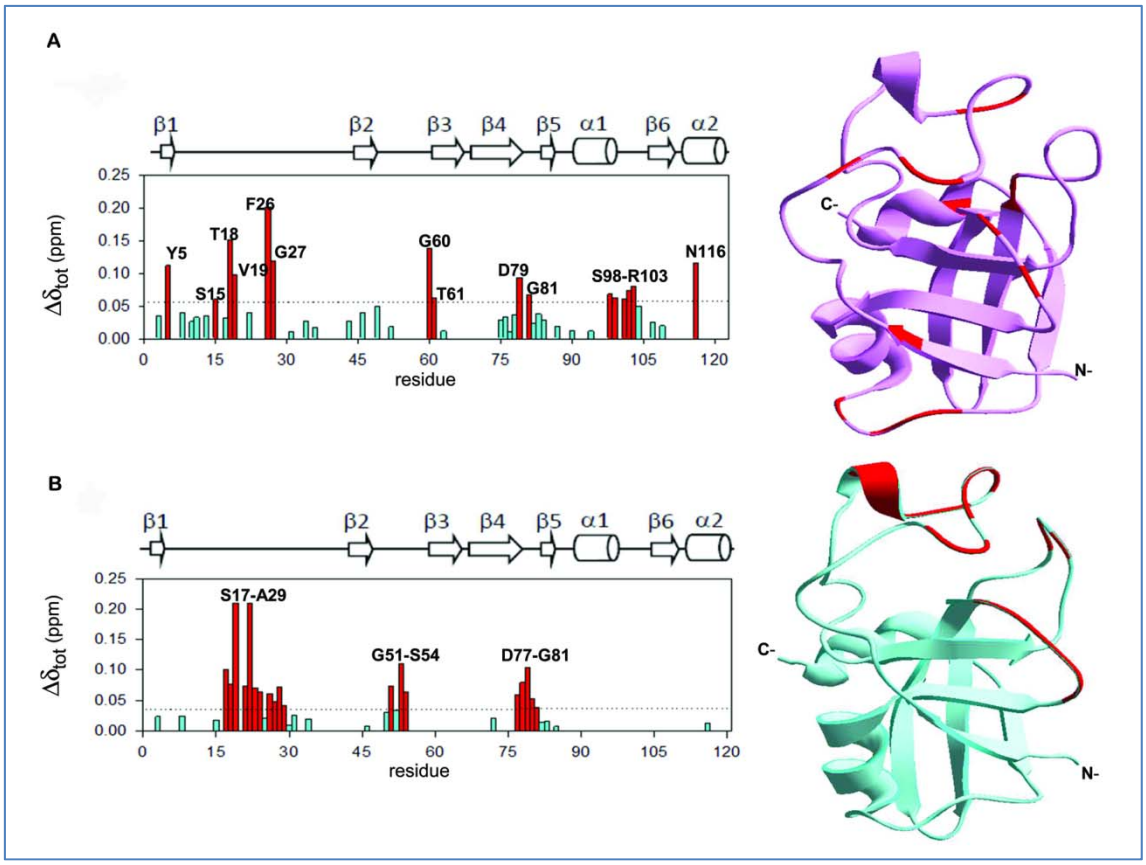

Figure 18. CSP effect on Pop1 (A) and CP (B) upon titration with oligosaccharides. The maximum shift for each amino acid at the last point of the titration with NAG (1:200) is plotted versus the residue number (left panels). Residues with $\Delta\delta > 0.058$ (Pop1) and $\Delta\delta > 0.039$ (CP) are plotted in red and mapped on proteins' structures (right panels).

We tested the ability of Pop1 to bind oligosaccharides, reproducing the experiment described by de Oliveira for CP [13], to verify whether the different flexibility of the loops and their spatial orientation would influence the protein binding capability.

Pop1 binding capability was investigated acquiring ^{1}H-^{15}N HSQC spectra at every step of a titration where we used tetramer of N-acetyl-D-glucosamine (NAG). The same titration was carried out also for CP.

Figure 18 shows the Chemical Shift Perturbation (CSP) observed upon titration of Pop1 and CP. Only residues with a chemical shift change ($\Delta\delta$) greater than the population standard deviation (0.058 in Pop1 and 0.039 in CP) were considered [85]. CSP (i.e. the shift of a resonance peak) originates from the sensitivity of nuclei

signals, in NMR, to the chemical environment and a shifted peak is indicative of a residue interacting with something else (in our titration, with NAG).

In the case of CP (Figure 18A), consistently with the previous data [13], NAG perturbed selectively a portion of loop β1-β2 (Ser^{17}-Ala^{29}), few residues on loop β2-β3 (Gly^{51}, Asp^{53}, Ser^{54}) and several residues on loop β4-β5 (Asp^{77}-Gly^{81}). The chemical shift changes were proportional to the quantity of added ligand, thus indicating selective binding and fast exchange between free- and bound- forms of the protein. For Pop1 (Figure 18B), CSP effects were observed for residues Ser^{15}, Thr^{18}, Val^{19}, Phe^{26}, Gly^{27} on loop β1-β2; Gly^{60} and Thr^{61} on loop β2-β3; Asp^{79} and Gly^{81} on loop β4-β5 and, with smaller Δδ, for five residues on loop α1-β6; Tyr^{5} on β1 and Asn^{116} on α2. This result, the lack of a direct correlation between CSP extent and NAG concentration and the observation that most of the residues located on loop α1-β6 showed a CSP only at high molar ratio (1:50 or 1:100), lead to the conclusion that Pop1's lack a well defined chitin binding site.

8.4 THE LOOP REGION OF CP AND POP1: OVERVIEW

To summarize, in CP and Pop1 the loops responsible for the defense elicitation activity differ, in their primary sequences, for few mutations and an insertion with consequent diversification of the proteins' electrostatic surfaces (Figure 12). NOE data reveal that these mutations create in the case of Pop1 a more extended network of interactions between the loops with respect to CP (Figure 14). The extended NOE pattern, the hypothesized salt bridge Lys^{25}/Asp^{52} and the strong hydrophobic interaction between Phe^{26}/Trp^{53} suggest a reduced degree of the loops conformational freedom in Pop1. The relaxation data (Figures 16 and 17) indeed confirmed reduced flexibility, in particular for loop β1–β2 of Pop1.

Overall, the presented data demonstrate that the detected structural and dynamic features influence the two proteins ability to bind oligosaccharides: CP selectively binds the tetramers of chitin (NAG) in a shallow groove, on one side of the barrel, defined by loops β1-β2, β2-β3 and β4-β5; Pop1, instead, does not present a specific binding site. To rationalize these findings, we propose that the reduced degree of conformational freedom of the loops, hampers the possibility to create the

appropriate binding site and/or to undergo the conformational reorganization, upon interaction with NAG, necessary for a selective and efficient interaction.

The fact that the protein region recognized by the plant as the defense elicitor sequence coincide with the protein chitin-binding sequence can be rationalized by the evolution of the plants' immune system toward the recognition of the most conserved regions of the protein, thus functionally relevant for the pathogen. Alternatively, a second explanation may be that the protein is used by the pathogen not only to attack the plant but also as an avirulence factor, able to remove the chitin fragments that would act as PAMPs and therefore activate the plant's defense system. Whatever the truth is, it is reasonable to think that the same differences in structural organization and conformational freedom that influence chitin binding, may also affect the ability to induce defenses.

To test this hypothesis, the first steps of the elicitation cascade triggered by CP are under investigation. The *in silico* protocol developed to search for a receptor is presented in chapter 9.

9. CERATO-PLATANIN RECEPTOR

The mechanism by which CP and Pop1 transduce the defense information to the plant still remains elusive. In fact, besides a direct interaction with the plant cuticle has been proposed [25], the most recent works [16, 86] suggest a more probable involvement of a "yet to be found" receptor able to recognize the structural motif created by loops β1-β2 and β2-β3.

Recent experiments in our research group [86] demonstrated that CP binds to a receptor on the plasma membrane of leaves' cells in *Arabidopsis* and is then internalized. Next step will be the identification and isolation of this receptor, with the aim to test whether the identified differences between CP and Pop1 are also responsible for the different elicitation activity.

A major problem arise due to the enormous number of proteins possibly involved in CP recognition, both among PAMP-recognizing receptor (PRR) or R-proteins. A 'wet lab' approach would require high experimental costs and a long time for the

analysis. We chose a protocol to screen, *in silico*, the genome of *Arabidopsis thaliana*, in order to reduce the number of candidates to be tested *in vitro* (Figure 19). The first question to answer was whether CP is involved in PAMP-triggered immunity (involvement of PRR) or in the Effector-triggered mechanism (involvement of R-proteins).

Recognizing the capacity of effectors like Ecp6 to bind chitin, thus preventing the onset of the plant's defenses [9, 87] and the ability reported for CP, Pop1 and other CPF members to bind, with variable affinity, chitin oligomers, could represent a similar mechanism of modulation of the plant immune responses. Chitin is an important structural component of fungal cell walls and its fragments are PAMPs very effective in eliciting plants' immune system [3]. It has been proposed that cerato platanins could have a role in binding and sequester chitin oligomers, released upon fungal attack, in order to prevent their detection by the plant [88]. Our studies, indeed, seem to confirm this hypothesis, showing the importance of chitin binding and its possible relationship to defense elicitation. Moreover, not to be discarded, is the fact that cerato platanins are proteins. PAMP triggered immunity rarely recognizes polypeptides, preferring the recognition of patterns common to several pathogens instead of species-specific molecules whose variability is relatively high (figure 1). ETI, instead, is a more specific defensive system, developed to recognize the specific proteins (effectors) secreted by a pathogen to overcome the first defensive layer.

All this considered, together with the observation that CP biochemical data (Table I) do not exclude the involvement of ETI instead of PTI, we started to screen *Arabidopsis* genes encoding for R-proteins.

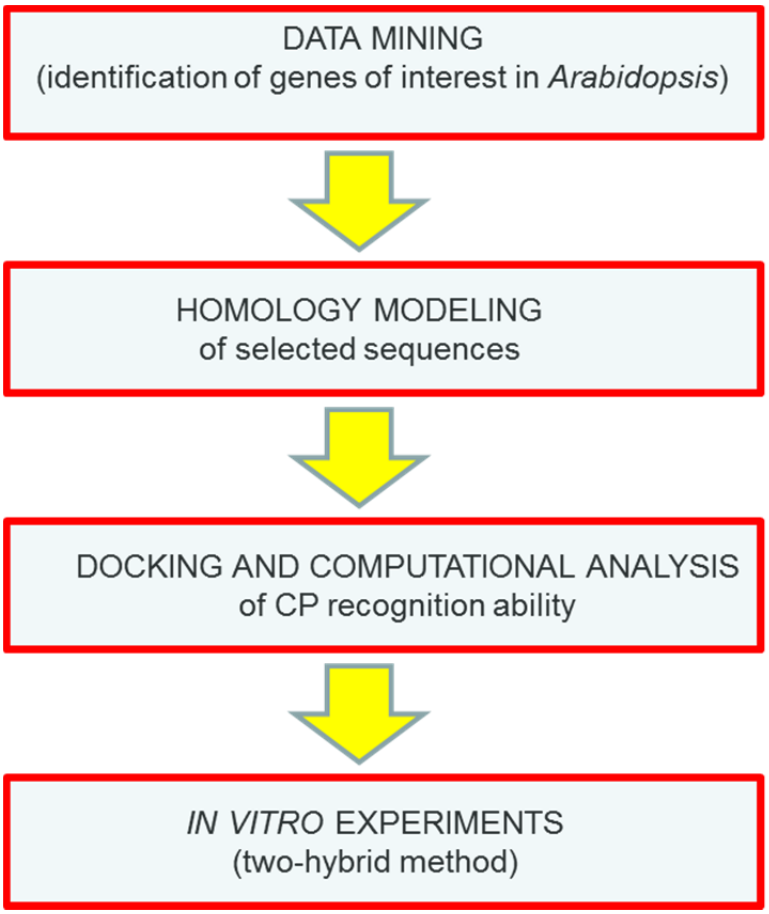

Figure 19 Strategy's pipeline for the identification of CP receptor.

9.1 R-PROTEINS IN *ARABIDOPSIS*: A FIRST SCREENING

The genome of *Arabidopsis thaliana* was the third multicellular organism's genome sequenced [89], whose genes have been intensively studied and characterized.

Meyers and coworkers [90] identified 207 R-genes expressed in *Arabidopsis* ecotype Columbia (*Col 0*) by sequence analysis and reannotation. Among these, 149 encoded for NBS-LRR proteins and 59 for NBS-proteins. Few years later, an analysis of the expression of these genes in different plant's tissues was performed, also highlighting that R-genes are constitutively expressed [91].

Recognizing that the ligand-recognition activity is associated to the LRR domain in most R-proteins, while the NBS one works as molecular switch, we investigated the first class of genes. Only NBS-LRR genes expressed in leaves at rate of at least 1 TPM (transcript per million) were selected: 78 showed these characteristic.

A further selection was based on two assumptions:

1. The gene must encode for a protein.
2. The encoded protein must be expressed on the cellular membrane. Genes with experimental evidence of different localization were excluded, as well as genes encoding for domains associated to DNA binding (e.g. Zinc-finger BED type)

51 genes were selected to be modelled and tested: 5 with experimental evidence of membrane localization, 46 of unknown cellular localization (or only inferred by homology or sequence analysis). A detailed summary of the genes selected for the analysis is reported in Appendix C.

9.2 MODELING R-PROTEINS

Homology modeling is currently the most accurate computational method for protein structure prediction. This approach build a three-dimensional model for a target protein sequence from a three-dimensional template structure of a homologous protein. Thus, the quality of the homology model will strongly depends on the sequence identity between target and template. Below 30% identity, serious errors may occur.

Homology model of LRR is highly impeded by the variability of these sequences: due to different repeat numbers and distinct arrangements of LRRs. In fact, a proper full-length template with sufficiently high sequence identity to the target is often missing. This problem is also exacerbated by a limited number of 3D structures available and by the higher variability of LRR in plants compared to those in animals. A variability that in plants can be also observed for constant regions.

In 2009 Tiandi and coworkers published a method to overcome this limitation, while modeling human and murine toll-like receptors [92]. We developed a variation of that protocol accounting for R-proteins features and new tools available.

After identification of the leucine rich repeats in a protein sequence, using both the program LRRSearch and a secondary structure prediction algorithm, each sequence was divided in fragments and for each fragment templates were searched (figure 20A).

A

MAFSSPSDFKRYHVFSSFHGPDVRSGFLSHLHNHFESKGITPFKDQEIERGHTIGPELIQAIRESRVSIVVLSEKYASSCWCLDE
LVEILKCKEASGQVVMTIFYKVDPSDVRKQRGDFGSTFKKTCEGKTWIVKQRWIKALEYIATVAGEHSLSWANEAELIQKIAT
DVSNKLNLTPSRDFEGMVGLEAHLTKLDSFLCLESDDVKMIGIWGPAGIGKTTIARALFNQLSTGFRLSCFMGTIDVNDYDS
KLCLQNKLLSKILNQKDMKIHHLGAIEEWLHNQRVLIVLDDVDDLEQLEVLAKESSWFGHGSRIIVSLNDRKILKAHGINDIY
DVDFPSEEEALEILCLSAFKQNSPQDGFEEVAKRVVELCGKLPLGLRVVGSSFYGESEDEWRIQLYGIETNLDRKIENVLRVGY
DKLSERHQSLFLHIACFFNHKSVDYVTTMLADSTLDVENGLKTLAAKSLVSTNGWITMHCLLQQLGRQVVVQQGDPGKRQ
FLVEAKEIRDVLANETGTESVIGISFDISKIETLSISKRAFNRMRNLKFLNFYNGSVSLLEDMEYLPRLRLLYWGSYPRKSLPLTF
KPECLVELYMGFSKLEKLWGGIQPLTNLKKINLGYSSNLKEIPNLSKATNLKTLTLTGCESLVEIPSSIWNLQKLEMLYASGCIK
LQVIPTNINLASLEEVNMSNCSRLRSFPDISSNIKRLYVAGTMIKEFPASIVGHWCRLDFLQIGSRSLKRLTHVPESVTHLDLR
NSDIKMIPDCVIGLPHLVSLLVENCTKLVSIQGHSPSLVTLFADHCISLKSVCCSFHGPISKLMFYNCLKLDKESKRGIIQQSGN
KSICLPGKEIPAEFTHQTIGNLITISLAPGCEEAYSTFSRFKACLLLSPIKNFAFNKINCFLRSKGVEISRTTESIYPFVSGGSLSEHL
FIFCGDLFPEENRSLMDVTPNEILFDFSSSDVEIVECGVKIFLSSGIEVGYCETGGNRNHHIDGEAEAFKVTQVENSKHTGHW
SWLRKLRLGKKKMKNNKTELSSRPVSGFSDLTRGIFIRYQEPEAVELTKDERITDLSSLLCIGSLVVCFFLSVFFGQPLVVSVLLL
FSCFFLQMQ

B

```
                *         20         *         40         *         60         *
query : NLKFLNFYNGSVSLLEDMEYLPRLRLLYWGSYPRKSLPLTF--KPECLVELYMGFSKLEKLW-GG-------IQ :  64
3rfs  : NVRYLALGGNKLHDISALKELTNLTYLILTGNQLQSLPNGVFDKLTNLKELVLVENQLQSLP-DGV------FD :  67
3w3k  : ----------------------------------------------LVELVFSGNRLDILWNDDDNRYISIFK :  27
4fcg  : ---------------------------------------------------------------------- :   -
4pq8  : ---------------------------------------------------------------------- :   -
4kng  : ---------------------------------------------------------------------- :   -

             80         *        100         *        120         *        140
query : PLTNLKKINLGYSSNLKEIPN-LSKATNLKTLTLTGCESLVEIPSS-IWNLQKLEMLYASGCIKLQVIPTNI-- : 134
3rfs  : KLTNLTYLNLAHN---------------------------------------------------------- :  80
3w3k  : GLKNLTRLDLSLN-RLKHIPN------------------------------------------------------ :  47
4fcg  : ----LEELDLRGCTALRNYPPIFGGRAPLKRLILKDCSNLLTLPLD-IHRLTQLEKLDLRGCVNLSRLPSLI-- :  67
4pq8  : ----------------------------LTTLNLSNN-QLTSLPQGVFERLTSLTTLNLSNN-QLTSLPQGVFE :  44
4kng  : ---------------------------------------------------------------------- :   -

          *        160         *        180         *        200         *        220
query : NLASLEEVNMSNCSRLRSFPDIS--SNIKRLYVAGTMIKEFPASIVGHWCRLDFLQ-IGSRSLKRLTHVP--ES : 203
3rfs  : ---------------------------------------------------------------------- :   -
3w3k  : ---------------------------------------------------------------------- :   -
4fcg  : ---------------------------------------------------------------------- :   -
4pq8  : RLTNLKTLNLSNNQ-L---------------------------------------------------------- :  59
4kng  : ----LRTLTLNGASQITEFPDLTGTANLESLTLTGAQISSLPQTVCNQLPNLQVLDLSYN-LLEDLPSFSVCQK :  69
```

Figure 20. The protein sequence encoded by the gene AT5G40910 (1104 amino acids, 121.44 KDa) is reported as example of the workflow used to create the models. **A)** In yellow are highlighted the constant regions of LRRs identified by the program LRRSearch. Using the confidence threshold for each LRR prediction and a secondary structure prediction, the LRR domain was predicted (sequence in green). **B)** Alignment of the five templates selected to cover the LRR domain, each of one spanning at least 2 LRR and overlapping each other.

The rationale behind fragmentation was i) to identify the best template to be used for a specific fragment, ii) to avoid having regions not covered by a template in the final alignment, iii) to use a reasonable number of templates in order to reduce computational time and iv) to obtain, when possible, templates covering at least two

LRRs and overlapping each other, in order to model a repeat 'in context'. Templates were selected only if two conditions were matched:

1. Identity between query and template > 30% and coverage > 80%.
2. The model obtained by using that template (SwissModel) present QMEAN4 and GMQE values acceptable, according to the ranking method used by SwissModel to assess quality of models.

The obtained templates were used to create a multiple alignment such as the one in figure 20B.

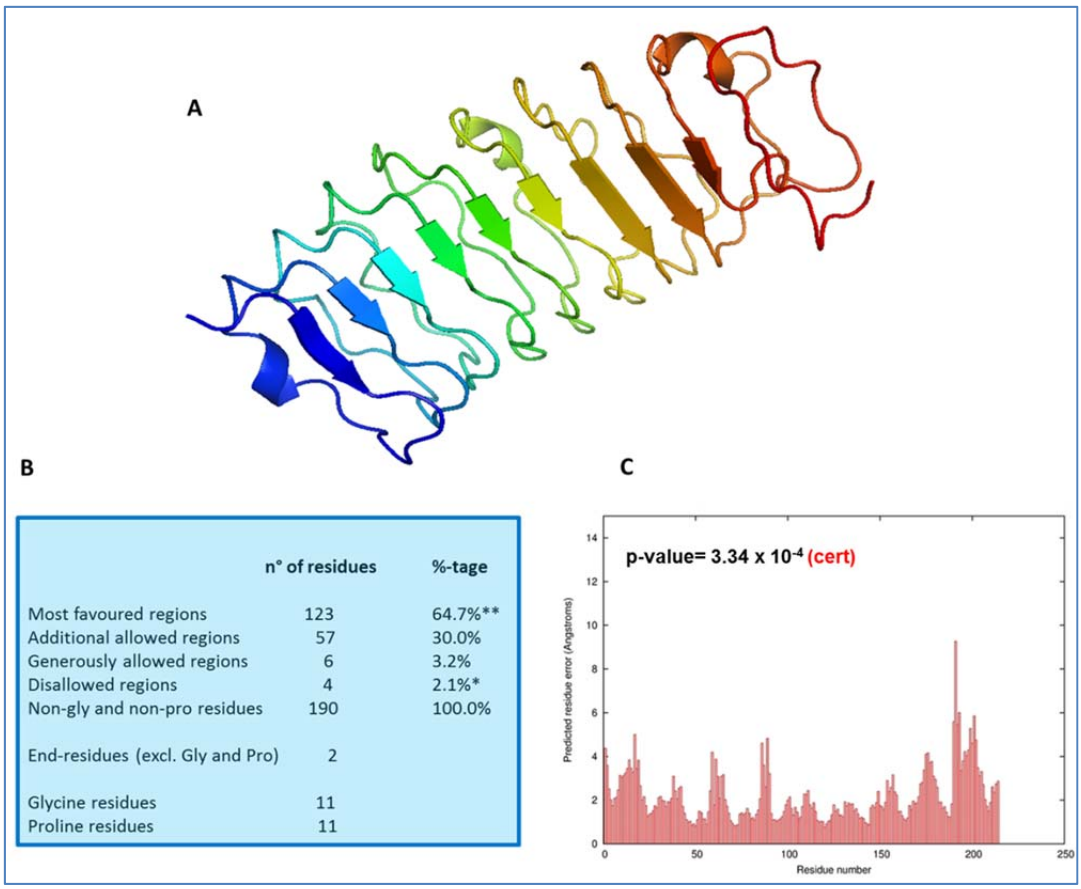

Figure 21. **A**) The model for AT5G40910 using the alignment in figure 20B. **B)** Estimation of the model quality through the analysis of its Ramachandran plot and **C)** the per-residue error plot, calculated by MODFOLD v. 4

The loop optimization protocol dopehr_loopmodel [93], available in MODELLER v.4, was chosen to build all our models in order to reduce errors arising from the

multiple gaps unavoidable in the template sequences. For each query ten models with different loop optimizations were obtained. Amongst the ten, the best one was chosen evaluating the LGscores predicted by PROQ Web server, a neural network that predict quality of structures, optimized to find correct models instead of native structures (LGScore > 1.5 correct; > 3 good; > 5 very good).

The chosen model quality was further evaluated analyzing its Ramachandran plot and MODFOLD server calculated p-value, a score indicating the probability that a model does not share any similarity with the native structure (p-value > 0.1 poor; < 0.05 medium; < 0.01 high; < 0.001 cert). Only if a model presented a p-value < 0.01 and a percentage of residues in most favored and additionally allowed regions > 90% was accepted. Figure 21 depicts an example of one of the models obtained and its quality evaluation.

This protocol allowed the modeling of 44 receptors out of 51 (Appendix C).

7 receptors presented regions longer than 10 residues for which a template was not identified. The loop modeling routine of MODELLER does not perform well on loops longer than 14 residues, and on loops longer than 8 residues severe errors may occur. In such cases the model presents errors too high to be accepted, thus the 9 candidates will be excluded from the future computational analysis, but directly tested *in vivo*.

The definition of a strategy to computationally test the ability of the 42 models to bind CP is still ongoing. A classical docking approach has been excluded due to the unstructured region of interest of CP. At present, homology- and structure-based methods are under investigation.

Conclusions

In conclusion, the data presented here identify structural and dynamic features that justify the diversity in biological activity of CP and Pop1. In particular, we show that variations in the spatial organization and conformational freedom of specific regions of the proteins may affect their ability to stimulate the plant's defense. To note that in a recent paper dealing with the decorin-binding protein A (DBPA) of the *Borrelia* spirochetes, causal agents of the Lyme disease in humans, that heavily depends on their ability to bind glycosaminoglycans (GAG) to colonize tissues, the authors derived similar conclusions: "*despite a similar topology, subtle changes in the primary sequence give rise to conformational modifications that significantly modify the ability of DBPA to bind GAG*" (94).

Although further biochemical and structural analysis of more CPF members are necessary to elucidate the role of the loops region, we envisage that engineered proteins may be designed with structural modifications in loops β1-β2, β2-β3 and β4-β5 to modulate their effector capability. Overall, the development of new fungal elicitors, designed on the basis of evolutionary conserved structural motifs and able to act as enhancers of crops and plants defense responses appear to be the winning strategy to lower the usage of chemical pesticides, in favor of a sustainable agriculture. Proteins induced resistance, in particular, offers the expectation to overcome fungicide insensitivity and breakdown of host resistance, problems that are worsened by the specter of global climate change and by the ever-increasing human population.

For this reason further studies are ongoing in order to identify a receptor for cerato-platanin using as starting point the results of this work.

REFERENCES

[1] Dodds, P. N. and Rathjen, J. P. (2010) *Plant immunity: towards an integrated view of plant-pathogen interactions*. Nat. Rev. Genet. 11, 539-548.

[2] Buscaill, P. and Rivas, S. (2014) *Transcriptional control of plant defence responses*. Curr. Opin. Plant Biol. 20, 35-46.

[3] Boller, T. and He, S. Y. (2009) *Innate immunity in plants: and arms race between pattern recognition*. Science 324, 742-744.

[4] Jones, J. D. G. and Dangl, J. (2006) *The plant immune system*. Nature 444, 323-329.

[5] Finn, R.D., Tate, J., Mistry, J. et al., Coggill, P. C., Sammut, J. S., Hotz, H. R., Ceric, G., Forslund, K., Eddy, S. R., Sonnhammer, E. L., Bateman, A. (2008) *The Pfam protein families database*. Nucleic Acids Research 36, 281–288.

[6] Pazzagli, L., Pantera, B., Carresi, L., Zoppi, C., Pertinhez, T. A., Spisni, A., Tegli, S., Scala, A., Cappugi, G. (2006) *Cerato-platanin, the first member of a new fungal protein family: cloning, expression, and characterization*. Cell Biochem. Biophys. 44, 512–521.

[7] Jeong, J. S., Mitchell, T. K., Dean, R. A. (2007) *The Magnaporthe grisea snodprot1 homolog,MSP1, is required for virulence*. FEMS Microbiol. Lett. 273, 157–165.

[8] Zaparoli, G., Cabrera, O. G., Medrano, F. J., Tiburcio, R., Lacerda, G., Pereira, G. G. (2009) *Identification of a second family of genes in Moniliophthora perniciosa, the causal agent of witches' broom disease in cacao, encoding necrosis-inducing proteins similar to cerato-platanins*. Mycol. Res. 113, 61–72.

[9] Pazzagli, L., Seidl-Seiboth, V., Barsottini, M., Vargas, W. A., Scala, A. and Mukherjee, P. K. (2014) *Cerato-platanins: elicitors and effectors*. Plant Sci. 228, 79-87.

[10] Rementeria, A., López-Molina, N., Ludwig, A., Vivanco, A. B., Bikandi, J., Pontón, J., Garaizar, J. (2005) *Genes and molecules involved in Aspergillus fumigatus virulence*. Rev. Iberoam. Micol. 22, 1–23.

[11] Sheu, F., Chien, P. J., Hsieh, K. Y., Chin, K. L., Huang, W. T., Tsao, T. Y., Chen, Y. F., Cheng, H. C., Chang, H. H. (2009) *Purification, cloning, and functional characterization of a novel immunomodulatory protein from Antrodia camphorate (bitter mushroom) that*

exhibits TLR2-dependent NF-B activation and M1polarization within murine macrophages. J. Agric. Food. Chem. 57, 4130–4141.

[12] Pan, S. and Cole, G. T. (1995) *Molecular and biochemical characterization of a Coccidioidesimmitis-specific antigen*. Infect. Immun. 63, 3994–4002.

[13] de Oliveira, A. L., Gallo, M., Pazzagli, L., Benedetti, C. E., Cappugi, G., Scala, A., Pantera, B., Spisni, A., Pertinhez, T. A. and Cicero, D. O. (2011) *The structure of the elicitor Cerato-platanin (CP), the first member of the CP fungal protein family, reveals a double psi beta-barrel fold and carbohydrate binding*. J. Biol. Chem. 286, 17560-17568.

[14] Baccelli, I., Luti, S., Bernardi, R., Scala, A. and Pazzagli, L. (2014) *Cerato-platanin shows expansin-like activity on cellulosic materials*. Appl. Microbiol. Biotechnol. 98, 175-184.

[15] Yang, Y., Zhang, H., Li, G., Li, W., Wang, X. and Song, F. (2009) *Ectopic expression of MgSM1, a Cerato-platanin family protein from Magnaporthe grisea, confers broad-spectrum disease resistance in Arabidopsis*. Plant Biotechnol. J. 7, 763-777.

[16] Frias, M., Brito, N., Gonzalèz, M. and Gonzalèz, C. (2014) *The phytotoxic activity of the cerato platanin BcSpl1 resides in a two-peptide motif on the protein surface*. Mol. Plant. Pathol. 15, 342-351.

[17] Hall, N., Keon, J. P. R., Hargreaves, J. A. (1999) *A homologue of a gene implicated in the virulence of human fungal diseases is present in a plant fungal pathogen and is expressed during infection*. Physiol. Mol. Plant Pathol. 55, 69–73.

[18] Pazzagli, L., Cappugi, G., Manao, G., Camici, G., Santini, A., Scala, A. (1999) *Purification ,characterization, and amino acid sequence of cerato-platanin, a new phytotoxic protein from Ceratocystis fimbriata f. sp. Platani*. J. Biol. Chem. 274, 24959–24964.

[19] Bennici, A., Calamassi, R., Pazzagli, L., Comparini, C, Schiff, S., Bovelli, R., Mori, B., Tani, C., Scala, A. (2005) *Cytological and ultrastructural responses of Platanus acerifolia(Ait.) Willd. leaves to cerato-platanin, a protein from Ceratocystis fimbriata f. sp. Platani*. Phytopath. Medit. 44, 153–161.

[20] Chen, H., Kovalchuk, A., Keriö, S., Asiegbu, F. O. (2013) *Distribution and bioinformatics analysis of cerato-platanin protein family in Dikarya*. Mycologia 105, 1479–1488.

[21] Boddi, S., Comparini, C., Calamassi, R., Pazzagli, L., Cappugi, G., Scala, A. (2004) *Cerato-platanin protein is located in the cell walls of ascospores, conidia and hyphae of Ceratocystis fimbriata f. sp. Platani*. FEMS Microbiol. Lett. 233, 341–346.

[22] Scala, A., Pazzagli, L., Comparini, C., Santini, A., Tegli, S., and Cappugi, G. (2004) *Cerato-platanin, an early produced protein by Ceratocystis fimbriata f. sp. Platani, elicits phytoalexins synthesis in host and non-host plants.* J. Plant Pathol. 86, 23–29.

[23] Fontana, F., Santini, A., Salvini, M., Pazzagli, L., Cappugi, G., Scala, A.,Durante, M., and Bernardi, R. (2008) *Cerato-platanin treated plane leaves restrict Ceratocystis platani growth and overexpress defence-related genes.* J. Plant Pathol. 90, 295–306.

[24] Pazzagli, L., Zoppi, C., Carresi, L., Tiribilli, B., Sbrana, F.,Schiff, S., et al. (2009). *Characterization of ordered aggregates of cerato-platanin and their involvement in fungus–host interactions*. Biochimica et Biophysica Acta. 1790, 1334–1344.

[25] Martellini, F., Faoro, F., Carresi, L., Pantera, B., Baccelli, I., Maffi, D., Tiribilli, B., Sbrana, F., Luti, S., Comparini, C., Bernardi, R., Cappugi, G., Scala, A. and Pazzagli, L. (2013). *Cerato-populin and cerato-platanin, two non-catalytic proteins from phytopathogenic fungi, interact with hydrophobic inanimate surfaces and leaves*. Mol. Biotechnol. 55, 27-42.

[26] de Barsottini, O., Mario, R., de Oliveira, J. F., Adamoski, D, Teixeira, P. J., do Prado, P. F. .V, Tiezzi, H. O. et al., (2013) *Functional diversification of cerato-platanins in Moniliophthora perniciosa as seen by differential expression and protein function specialization.* Mol. Plant–Microbe Interact. 26, 1281–1293.

[27] Comparini, C., Carresi, L., Pagni, E., Sbrana, F., Sebastiani, F., Luchi, N., Santini, A., Capretti, P., Tiribilli, B., Pazzagli, L., Cappugi, G. and Scala, A. (2009) *New proteins orthologous to cerato-platanin in various Ceratocystis species and the purification and characterization of cerato-populin from Ceratocystis populicola*. Appl Microbiol Biotechnol. 84, 309-322.

[28] Lombardi, L., Faoro, F., Luti, S., Baccelli, I., Martellini, F., Bernardi, R., Picciarelli, P., Scala, A. and Pazzagli, L. (2013) *Differential timing of defence-related responses induced by cerato-platanin and cerato-populin, two non-catalytic fungal elicitors*. Physiol. Plant. 149, 408-21.

[29] Higgins, D., Thompson, J., Gibson, T., Thompson, D., Higgins, D. G., Gibson, T. J. (1994) *CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice*. Nucleic Acids Res. 22, 4673–4680.

[30] Konagurthu, A. S., Whisstock, J. C., Stuckey, P. J., Lesk, A. M. (2006) *MUSTANG: A multiple structural alignment algorithm*. Proteins 64, 559-574.

[31] Krieger, E., Koraimann, G., Vriend, G. (2002) *Increasing the precision of comparative models with YASARA NOVA, a self-parameterizing force field.* Proteins. 47, 393–402.

[32] Spoel, S. H. and Xinnian, D. (2012) *How do plants achieve immunity? Defence without specialized immune cells*. Nature Reviews Immunology 12.2, 89-100.

[33] Newman, M. A., Sundelin, T., Nielsen, J. T. and Erbs, G. (2013) *MAMP (microbe-associated molecular pattern) triggered immunity in plants.* Front. Plant Sci. 4, 139.

[34] Caldo, R. A., Nettleton, D., Wise, R. P. (2004) *Interaction-dependent gene expression in Mla-specified response to barley powdery mildew*. Plant Cell 16, 2514-2528.

[35] Tao, Y., et al. (2003) *Quantitative nature of Arabidopsis responses during compatible and incompatible interactions with the bacterial pathogen Pseudomonas syringae.* The Plant Cell 15, 317-330.

[36] Shen, Q.H. and Schulze-Lefert, P. (2007) *Ruble in the jungle: compartmentalization, trafficking, and nuclear action of plant immune receptors*. EMBO J. 26, 4293-4301.

[37] Lukasik, E. and Takken, F. (2009) *STANDing strong, resistance proteins instigators of plant defence.* Curr. Opin. Plant Biol. 12, 427-436.

[38] Głowacki, S., Macioszek, V., Kononowicz, A. (2011). *R proteins as fundamentals of plant innate immunity*. Cellular and Molecular Biology Letters, 16, 1-24.

[39] Gabriëls, S. H. E. J., Vossen, J. H., Ekengren, S. K., van Ooijen, G., Abd-El-Haliem, A. M., van den Berg, G. C. M., Rainey, D. Y., Martin, G. B., Takken, F. L. W., de Wit, P. J. G. M. and Joosten, M. H. A. J. (2007) *An NB-LRR protein required for HR signalling mediated by both extra and intracellular resistance proteins*. Plant J. 50, 14-28.

[40] Faigón-Soverna, A., Harmon, F. G., Storani, L., Karayekov, E., Staneloni, R. J., Gassmann, W., Más, P., Casal, J. J., Kay, S. A. and Yanovsky, M. J. (2006) *A constitutive*

shade-avoidance mutant implicates TIR-NBS-LRR proteins in Arabidopsis photomorphogenic development. Plant Cell 18, 2919-2928.

[41] Flor, H. H. (1971) *Current status of the gene-for-gene concept*. Annu. Rev. Phytopathol. 9, 275-298.

[42] Tameling, W. I. L. and Baulcombe, D. C. (2007) *Physical association of the NB-LRR resistance protein Rx with a Ran GTPase-activating protein is required for extreme resistance to potato virus X*. Plant Cell 19, 1682-1694.

[43] van der Hoorn, R. A. L. and Kamoun, S. (2008) *From guard to decoy: A new model for perception of plant pathogen effectors*. Plant Cell 20, 2009-2017.

[44] Takahashi, N., Takahashi, Y. and Putnam, F. W. (1985) *Periodicity of leucine and tandem repetition of a 24-amino acid segment in the primary structure of leucine-rich α2-glycoprotein of human serum*. Proc. Natl. Acad. Sci. USA 82, 1906-1910.

[45] Kajava, A., V. (1998) *Structural diversity of leucine-rich repeat proteins*. J. Mol. Biol. 277, 519–527.

[46] Bella, J., Hindle, K. L., McEwan, P. A. and Lovell, S. C. (2008) *The leucine-rich repeat structure*. Cell. Mol. Life Sci. 65, 2307-2333.

[47] Manavalan, B., Basith, S., and Choi, S. (2011) *Similar structures but different roles–an updated perspective on TLR structures*. Frontiers in physiology, 2.

[48] Dodds, P. N., Lawrence, G. J. and Ellis, J. G. (2001) *Six amino acid changes confined to the leucine-rich repeat β-strand/β-turn motif determine the difference between the P and P2 rust resistance specificities in flax*. Plant Cell 13, 163-178.

[49] Wang, C. I. A., Gunčar, G., Forwood, J. K., Teh, T., Catanzariti, A. M., Lawrence, G. J., Loughlin, F. E., Mackay, J. P., Schirra, H. J., Anderson, P. A., Ellis, J. G., Dodds, P. N. and Kobe, B. (2007) *Crystal structures of flax rust avirulence proteins AvrL567-A and -D reveal details of the structural basis for flax disease resistance specificity*. Plant Cell 19, 2898-2912.

[50] Baccelli, I., Lombardi, L., Luti, S., Bernardi, R., Picciarelli, P., Scala, A. and Pazzagli, L. (2014) *Cerato-platanin induces resistance in Arabidopsis leaves through stomatal perception, overexpression of salicylic acid- and ethylene-signaling genes and camalexin biosynthesis*. PLoS One 10.1371/journal.pone.0100959.

[51] Carresi, L., Pantera, B., Zoppi, C., Cappugi, G., de Oliveira, A. L., Pertinhez, T. A., Spisni, A., Scala, A. and Pazzagli, L. (2006) *Cerato-platanin, a phytotoxic protein from Ceratocystis fimbriata: expression in Pichia pastoris, purification and characterization*. Protein Expr. Purif. 49, 159-167.

[52] Oliveira, A. L., Pazzagli, L., Pantera, B., Cappugi, G., Benedetti, C. E., Spisni, A., and Pertinhez, T. A. (2006) J. Biomol. NMR 36, 50.

[53] Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J. and Bax, A. (1995) *NMRPipe: a multidimensional spectral processing system based on UNIX pipes*. J. Biomol. NMR 6, 277-293.

[54] Johnson, B. A. (2004) *Using NMRView to visualize and analyze the NMR spectra of macromolecules*. Methods Mol. Biol. 278, 313–352.

[55] Vuister, G. and Bax, A. (1992) *Resolution enhancement and spectral editing of uniformly 13 C-enriched proteins by homonuclear broadband 13 C decoupling.* J. Magn. Reson., 98, 428-435

[56] Ottiger, M., Delaglio, F. and Bax, A. (1998) *Measurement of J and dipolar couplings from simplified two-dimensional NMR spectra*. J. Magn. Reson. 131, 373-378.

[57] Hansen, M. R., Mueller, L. and Pardi, A. (1998) *Tunable alignment of macromolecules by filamentous phage yields dipolar coupling interactions*. Nat. Struct. Biol. 5, 1065-1074.

[58] Zweckstetter, M. (2008) *NMR: prediction of molecular alignment from structure using the PALES software*. Nat. Protoc. 3, 679-690.

[59] Cornilescu, G., Marquardt, J. L., Ottiger, M. and Bax, A. (1998) *Validation of protein structure from anisotropic carbonyl chemical shifts in a dilute liquid crystalline phase*. J. Am. Chem. Soc. 120, 6836-6837.

[60] Kay, L. E., Torchia, D.A. and Bax, A. (1989) *Backbone dynamics of proteins as studied by 15N inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease.* Biochemistry 28, 8972–8979.

[61] Lipari, G. and Szabo, A. (1982) *Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules*. J. Am. Chem. Soc. 104, 4546-4570.

[62] Dosset, P., Hus, J. C., Blackledge, M. and Marion, D. (2000) *Efficient analysis of Macromolecular Rotational Diffusion from Heteronuclear Relaxation Data*. J. Biomol. NMR 16, 23-28.

[63] Ayed, A., Mulder, F. A., Yi, G. S., Lu, Y., Kay, L. E. and Arrowsmith, C. H. (2001) *Latent and active p53 are identical in conformation*. Nat. Struct. Biol. 8, 756-760.

[64] Arnold K, Bordoli L, Kopp J, and Schwede T (2006) *The SWISS-MODEL Workspace: A web-based environment for protein structure homology modelling*. Bioinformatics 22: 195-201.

[65] Laskowsky, R. A., MacArthur, M. W., Moss, D. S. and Thornton, J. M. (1993) *PROCHECK: A program to check the stereochemical quality of protein structures*. J. Appl. Cryst. 26, 283-291.

[66] Wiederstein, M. and Sippl, M. J. (2007) *ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins*. Nucleic Acids Res. 35, 407-410.

[67] Vriend, G. (1990) *WHAT IF: a molecular modeling and drug design program.* Journal of molecular graphics 8, 52-56.

[68] Chinea, G., Padron, G., Hooft, R. W. W., Sander, C. and Vriend, G. (1995) *The use of position specific rotamers in model building by homology*. Proteins 23, 415–421.

[69] McGuffin, L. J., Buenavista, M. T., and Roche, D. B. (2013) *The ModFOLD4 server for the quality assessment of 3D protein models*. Nucleic acids research, gkt294.

[70] Bej, A., Sahoo, B. R., Swain, B., Basu, M., Jayasankar, P., and Samanta, M. (2014) *LRRsearch: an asynchronous server-based application for the prediction of leucine-rich repeat motifs and an integrative database of NOD-like receptors*. Computers in biology and medicine 53, 164-170.

[71] Jones, D. T. (1999) *Protein secondary structure prediction based on position-specific scoring matrices*. J. Mol. Biol. 292, 195-202.

[72] Benkert, P., Tosatto, S. C., Schomburg, D. (2008) *QMEAN: A comprehensive scoring function for model quality assessment*. Proteins 71, 261-77.

[73] Benkert, P., Biasini, M., Schwede, T. (2011) *Toward the estimation of the absolute quality of individual protein structure models*. Bioinformatics. 27, 343-50.

[74] Biasini, M., Bienert, S., Waterhouse, A., Arnold, K., Studer, G., Schmidt, T., Kiefer, F., Galloo Cassarino, T., Bertoni, M., Bordoli, L. and Schwede, T. (2014) *SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information*. Nucl. Acids Res. gku340.

[75] Eswar, N., Marti-Renom, M. A., Webb, B., Madhusudhan, M. S., Eramian, D., Shen, M., Pieper, U. and Sali, A. (2006) *Comparative protein structure modeling with MODELLER*. Current Protocols in Bioinformatics 15, 5.6.1-5.6.30.

[76] Wallner, B., Elofsson, A. (2003) *Can correct protein models be identified?* Protein Sci. 12, 1073-1086.

[77] Wishart, D. S., Bigam, C. G., Holm, A., Hodges, R. S. and Sykes, B. D. (1995) *1H, 13C and 15N random coil NMR chemical shifts of the common amino acids. Investigations of nearest-neighbor effects*. J Biomol NMR 5, 67–81**.**

[78] Losonczi, J. A., Andrec, M., Fischer, M. W. and Prestegard, J. H. (1999) *Order matrix analysis of residual dipolar couplings using singular value decomposition.* J. Magn. Reson. 138, 334-342.

[79] Grzesiek, S., Bax, A. (1993) *The importance of not saturating water in protein NMR. Application to sensitivity enhancement and NOE measurements*. J. Am. Chem. Soc. 115, 12593-12594.

[80] Campos-Olivas, R. and Summers, M. F. (1999) *Backbone Dynamics of the N-Terminal Domain of the HIV-1 Capsid Protein and Comparison with the G94D Mutant Conferring Cyclosporin Resistance/Dependence*. Biochemistry 38, 10262-10271.

[81] Farrow, N. A., Muhandiram, R., Singer, A. U., Pascal, S. M., Kay, C. M., Gish, G., Shoelson, S. E., Pawson, T., Forman-Kay, J. D., and Kay, L. E. (1994) *Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by 15N NMR relaxation.* Biochemistry 33, 5984-6003.

[82] Cavanagh, J., Fairbrother, W. J., Palmer, A. G., Rance, M. and Skelton, N. J. (2007) *Protein NMR Spectroscopy: Principles and Practice*, 2nd edition, Elsevier Academic Press.

[83] Mandel, A. M., Akke, M., Palmer, A. G. (1995) *Backbone dynamics of Escherichia coli ribonuclease HI: correlations with structure and function in an active enzyme*. J. Mol. Biol. 246, 144-163.

[84] Luginbuhl, P., Pervushin, K. V., Iwai, H. and Wüthrich, K. (1997) *Anisotropic molecular rotational diffusion in 15N spin relaxation studies of protein mobility*. Biochemistry 36, 7305-7312.

[85] Williamson, M. P. (2013) *Using chemical shift perturbation to characterize ligand binding*. Prog. Nucl. Magn. Reson. Spec. 73, 1-16.

[86] Luti, S., (2015) Personal communication.

[87] de Jonge, R., van Esse, H. P., Kombrink, A., Shinya, T., Desaki, Y., Bours, R., van der Krol, S., Shibuya, N., Joosten, M. H. and Thomma, B. P. (2010) *Conserved fungal LysM effector Ecp6 prevents chitin-triggered immunity in plants*. Science 329, 953-955.

[88] Boller, T. (2004) *Chemoperception of microbial signals in plant cells*. Annu. Rev. Plant Physiol. Plant Mol. Biol. 46, 189-214.

[89] Arabidopsis Genome Initiative (2000) *Analysis of the genome sequence of the flowering plant Arabidopsis thaliana*. Nature 408, 796.

[90] Meyers, B. C., Kozik, A., Griego, A., Kuang, H., & Michelmore, R. W. (2003) *Genome-wide analysis of NBS-LRR–encoding genes in Arabidopsis*. The Plant Cell 15, 809-834.

[91] Tan, X., Meyers, B. C., Kozik, A., West, M. A., Morgante, M., St Clair, D. A. et al., (2007) *Global expression analysis of nucleotide binding site-leucine rich repeat-encoding and related genes in Arabidopsis*. BMC plant biology 7, 56.

[92] Wei, T., Gong, J., Jamitzky, F., Heckl, W. M., Stark, R. W., and Rössle, S. C. (2009). *Homology modeling of human Toll-like receptors TLR7, 8, and 9 ligand-binding domains.* Protein Science 18, 1684-1691.

[93] Fiser, A., Do, R. K. G., and Sali, A. (2000) *Modeling of loops in protein structure*. Protein Sci. 9, 1–21.

[94] Ashli, M. M. and Xu, W. (2015) *Structural mechanisms underlying sequence-dependent variations in GAG affinities of decorin binding protein A, a Borrelia burgdorferi adhesion*. Biochemical J. 467, 439-451.

APPENDIX A

Residual Dipolar Couplings

Experimental RDCs (D_{HN}^{obs} column) versus calculated RDCs (D_{HN}^{calc} column). The third column reports D_{HN}^{obs} - D_{HN}^{calc}. Only the 82 residues for which was possible to calculate a D_{HN}^{obs} are reported.

	D_{HN}^{obs}	D_{HN}^{calc}	ΔD_{HN}
1 VAL	6.78	4.9346	1.8454
2 SER	10.64	4.2965	6.3435
3 ILE	-10.8	0.0513	-10.8513
4 SER	1.4	0.6011	0.7989
6 ASP	3.59	1.2711	2.3189
8 VAL	-7.7	-4.6949	-3.0051
10 GLY	1.93	0.5342	1.3958
11 ASN	5.05	2.2364	2.8136
13 GLY	5.26	-2.3079	7.5679
14 GLN	-2.44	1.7625	-4.2025
15 SER	1.12	1.8457	-0.7257
18 THR	2.28	3.3022	-1.0222
19 VAL	-1.29	-3.2026	1.9126
20 ALA	-9.146	-1.063	-8.083
22 SER	-1.52	0.149	-1.669
23 ASN	-7.02	4.3832	-11.4032
26 PHE	14.53	3.2791	11.2509
27 GLY	11.32	3.5809	7.7391
28 LEU	-10.45	0.7258	-11.1758
29 ALA	-4.48	0.5467	-5.0267
30 ALA	0.46	-4.4219	4.8819
31 LYS	-13.05	-4.5997	-8.4503
34 THR	-5.7	2.5406	-8.2406

36 GLY	3.85	0.0816	3.7684
37 ASP	0.33	-0.0064	0.3364
38 LEU	0.62	1.5789	-0.9589
40 ALA	6.29	2.0644	4.2256
41 PHE	-5.13	-5.6768	0.5468
43 ASN	-6.22	-4.2725	-1.9475
44 VAL	12.95	1.0119	11.9381
46 GLY	4.82	1.0612	3.7588
47 ILE	7.51	1.0858	6.4242
49 ASP	10.4	3.2322	7.1678
50 ILE	4.36	-0.329	4.689
52 ASP	6.77	1.9409	4.8291
58 ASN	-5.78	2.2715	-8.0515
60 GLY	-3.16	1.0223	-4.1823
61 THR	-3.26	1.8771	-5.1371
62 CYS	-8.61	1.0639	-9.6739
63 TRP	0.59	-2.8562	3.4462
64 LYS	-2.16	-2.1486	-0.0114
66 THR	-1.56	2.071	-3.631
67 VAL	1.61	2.7679	-1.1579
69 ASN	2.361	2.0694	0.2916
70 ARG	3.81	4.7766	-0.9666
71 ASN	2.94	1.455	1.485
75 ILE	1.4	-4.6839	6.0839
76 LEU	-2.03	-1.672	-0.358
77 GLY	3.79	0.1653	3.6247
78 VAL	-11.01	-3.8417	-7.1683
79 ASP	-6.08	1.5677	-7.6477
81 GLY	0.17	-0.2297	0.3997
83 GLY	-11.59	-1.1864	-10.4036
84 ASN	4.22	0.2738	3.9462

85 PHE	-13.533	-2.2656	-11.2674
86 ASN	-0.49	1.4218	-1.9118
87 VAL	-6.16	-1.1393	-5.0207
88 SER	4.94	0.808	4.132
92 PHE	2.54	-0.7176	3.2576
94 GLU	1.06	0.5629	0.4971
95 LEU	3.81	1.8237	1.9863
96 THR	1.19	-2.0987	3.2887
97 GLY	0.39	-0.5829	0.9729
99 THR	4.52	0.1764	4.3436
100 ASP	4.1	4.6805	-0.5805
102 GLY	1.4	-1.1795	2.5795
103 ARG	0.05	-1.7104	1.7604
104 VAL	8.16	-2.4795	10.6395
105 ASN	0.92	-0.2345	1.1545
106 ASP	2.72	4.9672	-2.2472
107 ILE	6.22	4.6153	1.6047
108 GLN	-3.08	4.0425	-7.1225
109 TYR	-2.8	-2.9252	0.1252
110 GLU	-6.15	-1.5813	-4.5687
111 LYS	0.1	-3.1682	3.2682
113 SER	0.167	1.507	-1.34
114 ALA	5.05	1.8213	3.2287
115 ASP	-4.13	-0.3056	-3.8244
117 CYS	1.11	2.3281	-1.2181
118 LYS	-0.418	0.8401	-1.2581
119 MET	-7.61	-0.7918	-6.8182
122 ALA	-3.31	-0.5988	-2.7112

APPENDIX B

Model-Free Analysis

The tables report the models selected by TENSOR2 for each residue with associated parameters and errors: models 1 (S^2), 2 (S^2, T_e), 3 (S^2, K_{ex}), 4 (S^2, T_e, K_{ex}) and 5 (S^2_{slow}, T_e, S^2_{fast}). Model 6 is chosen when none of the five proposed fits can reproduce data satisfactorily . The statistical approach to the selection of models has been outlined by Mandel et al. [83].

Model-free analysis of Pop1 (τ_c = 6.98 ns).

Residue Pop1	Model	S^{2slow}	error	T_e (ns)	error (ns)	K_{ex} (s^{-1})	Error (s^{-1})	$S^{2\ fast}$	error
1	5	0.72	0.02	1.13	0.07	0	0	0.8	0.01
2	1	0.99	0.01	0	0	0	0	1	0
3	6 (5)	1.06	0.05	6.98	2.88	0	0	0.93	0.04
4	1	0.95	0.02	0	0	0	0	1	0
6	4	0.94	0.03	0.04	0.42	1.05	0.53	1	0
8	1	0.92	0.02	0	0	0	0	1	0
10	1	0.92	0.02	0	0	0	0	1	0
11	4	0.84	0.02	0.08	0.02	2.16	0.32	1	0
13	6 (5)	0.92	0.06	6.98	2.69	0	0	0.86	0.02
15	6 (5)	1.04	0.03	6.98	2.87	0	0	0.93	0.02
17	6 (5)	1.05	0.04	6.98	3.01	0	0	0.95	0.03
18	1	0.97	0.02	0	0	0	0	1	0
19	2	0.9	0.03	0.08	0.41	0	0	1	0
20	4	0.75	0.03	0.05	0.01	2.84	0.7	1	0
22	5	0.9	0.03	4.67	2.37	0	0	0.91	0.01
23	1	0.95	0.04	0	0	0	0	1	0
24	6 (4)	0.78	0.13	6.98	2.35	2.43	1.36	1	0
25	1	0.94	0.02	0	0	0	0	1	0
26	2	0.89	0.04	0.07	0.48	0	0	1	0
27	2	0.89	0.03	0.1	0.24	0	0	1	0
28	1	0.93	0.03	0	0	0	0	1	0
29	3	0.93	0.02	0	0	0.78	0.29	1	0
30	6 (5)	0.97	0.03	6.98	2.75	0	0	0.86	0.01
31	4	0.76	0.02	0.01	0	1.07	0.48	1	0
34	3	0.99	0	0	0	0.65	0	1	0
36	6 (5)	0.82	0.05	6.98	1.99	0	0	0.88	0.02
37	6 (5)	0.87	0.04	6.98	1.97	0	0	1	0.02

38	6 (5)	1.04	0.02	6.98	3.01	0	0	0.96	0.01
40	6 (5)	1.03	0.05	0.18	1.39	0	0	0.87	0.04
41	3	0.82	0.03	0	0	2.88	0.59	1	0
43	6 (5)	1.03	0.05	6.98	2.87	0	0	0.87	0.02
44	6 (5)	0.88	0.05	6.98	2.08	0	0	0.85	0.02
45	1	0.91	0.06	0	0	0	0	1	0
46	3	0.83	0.04	0	0	1.56	0.65	1	0
47	3	0.96	0.05	0	0	2.64	0.7	1	0
49	3	0.92	0	0	0	4.64	0	1	0
50	3	0.8	0.03	0	0	0.97	0.33	1	0
52	6 (5)	1.14	0.04	6.98	2.27	0	0	0.87	0.02
57	1	0.99	0.02	0	0	0	0	1	0
58	1	0.97	0.02	0	0	0	0	1	0
60	4	0.89	0.03	0.05	0.03	1.95	0.56	1	0
61	1	0.98	0.02	0	0	0	0	1	0
62	6 (5)	0.94	0.05	6.98	2.91	0	0	0.92	0.03
63	1	0.88	0.03	0	0	0	0	1	0
64	1	0.94	0.03	0	0	0	0	1	0
66	3	0.92	0.02	0	0	0.69	0.26	1	0
67	1	0.95	0.02	0	0	0	0	1	0
69	5	0.92	0.04	1.9	1.18	0	0	0.76	0.02
70	2	0.85	0.01	0.02	0.01	0	0	1	0
71	6 (5)	1.04	0.04	6.98	2.93	0	0	0.85	0.02
74	6 (5)	0.78	0.08	6.98	1.82	0	0	0.78	0.03
75	6 (5)	1.15	0.03	6.98	2	0	0	0.87	0.02
76	6 (5)	1.09	0.04	6.98	2.59	0	0	0.9	0.02
77	6 (5)	0.86	0.06	6.98	2.46	0	0	0.9	0.03
78	1	0.95	0.03	0	0	0	0	1	0
79	1	0.94	0.03	0	0	0	0	1	0
82	3	0.9	0.02	0	0	2.8	0.45	1	0
83	4	0.77	0.03	0.02	0.01	2.41	0.49	1	0
84	6 (5)	1.06	0.06	6.98	2.92	0	0	0.89	0.03
86	2	0.88	0.02	0.03	0.01	0	0	1	0
87	6 (5)	1.11	0.03	6.98	2.28	0	0	0.87	0.02
90	3	0.8	0.03	0	0	2.54	0.8	1	0
92	3	0.95	0.01	0	0	1.24	0.31	1	0
94	6 (5)	1.06	0.02	6.98	2.22	0	0	0.93	0.01
95	1	0.91	0.01	0	0	0	0	1	0
96	1	0.92	0.02	0	0	0	0	1	0
97	1	0.92	0.03	0	0	0	0	1	0
99	2	0.81	0.03	0.05	0.01	0	0	1	0
100	5	0.82	0.03	1.84	0.28	0	0	0.92	0.02
101	2	0.8	0.02	0.02	0.01	0	0	1	0

102	2	0.73	0.02	0.05	0.01	0	0	1	0
103	2	0.77	0.01	0.06	0	0	0	1	0
104	1	0.88	0.01	0	0	0	0	1	0
105	4	0.8	0.02	0.03	0.01	0.91	0.32	1	0
106	1	0.81	0.01	0	0	0	0	1	0
107	1	0.82	0.02	0	0	0	0	1	0
108	2	0.85	0.01	0.06	0.01	0	0	1	0
109	2	0.84	0.01	0.04	0.01	0	0	1	0
110	1	0.89	0.01	0	0	0	0	1	0
111	1	0.9	0.01	0	0	0	0	1	0
113	1	0.84	0.01	0	0	0	0	1	0
114	6 (5)	0.96	0.03	6.98	2.59	0	0	0.9	0.02
115	6 (5)	0.99	0.03	6.98	2.42	0	0	1.01	0.02
117	3	0.82	0.02	0	0	0.92	0.24	1	0
118	3	0.96	0	0	0	1.14	0	1	0
119	1	0.87	0.02	0	0	0	0	1	0

Model-free analysis of CP (τ_c =7.36).

Residue CP	**Model**	**$S^{2\ slow}$**	**error**	**T_e (ns)**	**error (ns)**	**K_{ex} (s^{-1})**	**error (s^{-1})**	**$S^{2\ fast}$**	**error**
1	4	0.85	0.02	0.07	0.02	1.71	0.35	1	0
2	3	0.88	0.01	0	0	0.53	0.17	1	0
3	3	0.86	0	0	0	2.62	0.21	1	0
4	3	0.87	0.03	0	0	2.07	0.58	1	0
5	1	0.91	0.01	0	0	0	0	1	0
6	3	0.86	0.01	0	0	1.43	0.21	1	0
8	3	0.8	0.01	0	0	5.07	0.22	1	0
9	3	0.79	0.01	0	0	4.38	0.27	1	0
10	3	0.8	0.03	0	0	1.29	0.44	1	0
11	6 (5)	1.04	0.04	7.35	3.03	0	0	0.85	0.02
12	1	0.8	0.02	0	0	0	0	1	0
13	3	0.87	0.04	0	0	4.72	0.69	1	0
14	1	0.9	0.01	0	0	0	0	1	0
15	3	0.92	0.01	0	0	3	0.22	1	0
16	3	1	0.01	0	0	2.97	0.32	1	0
17	3	0.96	0.02	0	0	0.89	0.35	1	0
18	1	0.86	0.01	0	0	0	0	1	0
19	3	0.87	0.03	0	0	5	0.4	1	0
20	3	0.91	0.03	0	0	6.27	0.34	1	0
21	3	0.97	0.02	0	0	0.63	0.23	1	0
22	6 (5)	0.9	0.03	7.35	2.38	0	0	0.96	0.02
23	5	0.74	0.03	6.55	1.91	0	0	0.87	0.01

24	5	0.94	0.02	1.39	1.67	0	0	0.83	0.01
25	5	0.91	0.04	1.77	1.62	0	0	0.96	0.03
26	3	0.89	0.02	0	0	3.17	0.26	1	0
27	3	0.85	0.01	0	0	6.65	0.37	1	0
28	3	0.9	0.01	0	0	7.47	0.23	1	0
29	3	0.8	0.01	0	0	1.89	0.3	1	0
30	4	0.69	0.01	0.01	0	0.63	0.31	1	0
31	3	0.8	0.01	0	0	3.44	0.16	1	0
33	3	0.98	0.01	0	0	0.54	0.22	1	0
34	3	0.92	0	0	0	1.16	0	1	0
35	6 (1)	0.97	0.01	0	0	0	0	1	0
36	1	0.89	0.01	0	0	0	0	1	0
39	1	0.87	0.01	0	0	0	0	1	0
40	6 (3)	0.85	0.02	0	0	1.85	0.47	1	0
42	3	0.9	0.01	0	0	3.9	0.19	1	0
43	3	0.87	0.02	0	0	1.94	0.33	1	0
44	3	0.93	0.04	0	0	4.99	0.59	1	0
45	1	0.89	0.01	0	0	0	0	1	0
46	3	0.92	0.01	0	0	2.22	0.21	1	0
48	6 (5)	0.75	0.04	7.35	1.91	0	0	0.89	0.02
49	3	0.89	0.02	0	0	2.77	0.27	1	0
50	6 (4)	0.66	0.13	7.35	2.03	2.01	0.62	1	0
52	5	0.91	0.02	4.95	2.38	0	0	0.89	0.01
53	3	1	0.01	0	0	2.18	0.24	1	0
54	1	0.98	0.02	0	0	0	0	1	0
56	5	0.81	0.04	5.36	2.06	0	0	0.96	0.02
57	3	1	0.01	0	0	1.56	0.3	1	0
58	3	0.93	0.02	0	0	0.76	0.29	1	0
59	3	0.95	0	0	0	1.15	0	1	0
60	1	0.9	0.01	0	0	0	0	1	0
61	3	0.87	0.01	0	0	1.98	0.12	1	0
62	1	0.86	0.01	0	0	0	0	1	0
63	3	0.86	0.02	0	0	2.58	0.22	1	0
64	6 (5)	0.93	0.02	7.35	2.69	0	0	0.88	0.01
67	3	0.86	0.02	0	0	1.94	0.32	1	0
68	4	0.9	0.03	0.03	0.02	2.03	0.46	1	0
69	4	0.83	0.01	0.03	0.01	0.97	0.18	1	0
70	1	0.87	0.01	0	0	0	0	1	0
71	3	0.88	0.02	0	0	1.51	0.3	1	0
72	1	0.84	0.02	0	0	0	0	1	0
73	2	0.78	0.02	0.02	0.01	0	0	1	0
74	6 (3)	0.9	0.02	0	0	1.04	0.27	1	0
75	6 (5)	0.91	0.03	7.35	2.38	0	0	0.9	0.02

76	3	0.87	0.01	0	0	1	0.18	1	0
77	3	0.9	0	0	0	4.74	0	1	0
78	3	0.84	0.02	0	0	2.62	0.45	1	0
79	3	0.89	0.01	0	0	1.56	0.2	1	0
80	1	0.92	0.01	0	0	0	0	1	0
81	3	0.92	0.01	0	0	1.73	0.22	1	0
82	4	0.81	0.01	0.02	0.01	1.01	0.29	1	0
83	3	0.82	0.02	0	0	3.08	0.3	1	0
84	1	0.85	0.01	0	0	0	0	1	0
85	3	0.84	0.02	0	0	0.68	0.25	1	0
86	4	0.87	0.01	0.03	0.01	3	0.17	1	0
88	3	0.89	0.01	0	0	3.12	0.22	1	0
89	3	0.88	0.01	0	0	1.11	0.27	1	0
90	3	0.87	0.04	0	0	3.59	0.49	1	0
91	6 (1)	0.93	0.02	0	0	0	0	1	0
92	3	0.87	0.03	0	0	0.78	0.34	1	0
93	3	0.85	0.01	0	0	1.64	0.08	1	0
95	3	0.83	0.01	0	0	1.09	0.13	1	0
96	5	0.83	0.04	1.47	0.47	0	0	0.81	0.02
97	5	0.68	0.03	2.12	0.39	0	0	0.88	0.02
98	2	0.94	0.02	0.47	0.36	0	0	1	0
99	5	0.74	0.02	1.49	0.19	0	0	0.83	0.02
100	2	0.86	0.01	0.15	0.06	0	0	1	0
101	5	0.73	0.02	0.63	0.07	0	0	0.85	0.01
102	3	0.83	0.02	0	0	1.73	0.29	1	0
104	5	0.9	0.02	2.75	2.11	0	0	0.81	0.01
105	1	0.86	0.01	0	0	0	0	1	0
106	3	0.85	0.03	0	0	1.64	0.41	1	0
107	3	0.83	0.04	0	0	3.53	0.49	1	0
108	3	0.88	0.01	0	0	3.38	0.23	1	0
109	4	0.83	0.03	0.02	0.01	2.44	0.36	1	0
110	3	0.9	0.01	0	0	1.21	0.11	1	0
111	3	0.87	0.02	0	0	1.08	0.27	1	0
112	1	0.86	0.01	0	0	0	0	1	0
113	1	1	0.01	0	0	0	0	1	0
114	3	0.91	0.01	0	0	1.99	0.25	1	0
115	3	0.94	0.02	0	0	4.66	0.37	1	0
116	3	0.87	0	0	0	5.63	0	1	0
117	5	0.94	0.01	0.82	0.8	0	0	0.91	0.01
118	4	0.79	0.04	0.55	0.14	2.39	0.44	1	0
119	2	0.76	0.01	0.25	0.08	0	0	1	0

APPENDIX C

R-proteins

The table reports gene accession number (NCBI gene database), localization (membrane or experimentally not defined), transcription level (TPM rate) [91] and if a model has been created with the protocol presented in this thesis.

Accession number	Localization	TPM rate	Model
AT5G40100	Membrane	1	No
AT5G45510	Membrane	62	No
AT4G26090	Membrane	51	Yes
AT1G12220	Membrane	6	Yes
AT1G10920	Membrane	1	Yes
AT1G56520	Not Defined	10	Yes
AT4G14370	Not Defined	21	Yes
AT5G38850	Not Defined	1	Yes
AT1G63740	Not Defined	1	Yes
AT5G58120	Not Defined	31	Yes
AT1G63860	Not Defined	7	Yes
AT1G63880	Not Defined	26	Yes
AT2G16870	Not Defined	2	Yes
AT4G19050	Not Defined	3	No
AT5G66910	Not Defined	17	Yes
AT1G15890	Not Defined	3	No
AT5G43740	Not Defined	1	Yes
AT1G62630	Not Defined	6	Yes
AT5G46470	Not Defined	3	Yes
AT5G46510	Not Defined	9	Yes
AT5G46260	Not Defined	4	Yes
AT5G46270	Not Defined	25	Yes
AT3G44670	Not Defined	3	Yes
AT3G44400	Not Defined	1	Yes
AT1G69550	Not Defined	3	Yes
AT5G44510	Not Defined	11	Yes
AT2G14080	Not Defined	29	Yes
AT5G38340	Not Defined	1	Yes
AT5G38350	Not Defined	1	Yes
AT5G40910	Not Defined	22	Yes
AT4G19500	Not Defined	9	Yes
AT4G16860	Not Defined	26	No
AT4G16900	Not Defined	8	Yes
AT4G16950	Not Defined	58	No
AT5G22690	Not Defined	9	Yes
AT5G40060	Not Defined	130	Yes
AT5G46450	Not Defined	42	Yes
AT4G12010	Not Defined	32	Yes

AT4G19510	Not Defined	52	Yes
AT4G36150	Not Defined	11	Yes
AT5G44870	Not Defined	58	Yes
AT5G45060	Not Defined	41	Yes
AT4G19530	Not Defined	105	Yes
AT1G27180	Not Defined	2	Yes
AT5G36930	Not Defined	3	Yes
AT1G17600	Not Defined	14	Yes
AT5G63020	Not Defined	28	Yes
AT3G50950	Not Defined	194	Yes
AT1G53350	Not Defined	3	Yes
AT1G59620	Not Defined	3	No
AT1G58602	Not Defined	130	Yes

ACKNOWLEDGEMENTS

This work was possible thanks to the help and collaboration of several people and institutions.

Professor Luigia Pazzagli and Dr. Simone Luti (Department of Biomedical Experimental and Clinical Sciences, University of Florence, Italy) expressed and purified cerato-populin and cerato-platanin, while CD and NMR spectra were acquired with the instruments available at CIM (Centro Interdipartimentale di Misure, University of Parma, Italy).

Another crucial contribute was that of Professor Gert Vriend and Dr. Hanka Venselaar (CMBI, Radboud University, Nijmegen, The Netherlands), whose advices on homology modeling of leucine rich repeats were precious.

Dr Thelma A. Pertinhez and professor Alberto Spisni tutored me for the entire PhD course: their suggestions, critics and teachings not only contributed to the preparation of this thesis, but also to my professional growth.

I must thank also Professor Lorella Franzoni, the tutor of this PhD.

Finally, all my colleagues deserve to be thanked for the several advices they granted me during these three years: Drs. Elena Ferrari, Emanuela Casali and Giulia Rizzatello.