Humberto Couto Fernandes

Structural and biochemical studies of plant PR-10 proteins from different sources

Thesis presented to the Scientific Council
of the Institute of Bioorganic Chemistry
Polish Academy of Sciences in Poznan
as a Ph.D. dissertation

Scientific adviser:
Prof. Mariusz Jaskolski

Poznan 2008

The X-ray diffraction experiments described in this work were carried out using synchrotron beamlines at the EMBL Unit at the DESY synchrotron in Hamburg, and at the BESSY synchrotron in Berlin.

Financial support for this work was provided by a Marie Curie Fellowship for Early Stage Researcher Training: MEST-CT-2004-504066 within the ADOPT program of the European Union.

Partial support was also provided by a grant from the Ministry of Science and Higher Education, number N N204 2584 33.

Acknowledgements

The work has been carried in the Center for Biocrystallographic Research, Institute of Bioorganic Chemistry, Polish Academy of Sciences in Poznan, within the framework of the ADOPT program of the European Union.

First, I would like to thank Professor Jaskolski for giving me the opportunity to work in his lab. Above all I acknowledge his willingness to give me a great deal of independence to pursue my interests.

To Professor Sikorski, my co-supervisor, I owe thanks for his continuous support, guidance and interest in my work and for all good advices during the past years.

My thanks also go to:

Professor Grzegorz Bujacz for very nice collaboration and advice on crystallization.

Dr. Filip Jelen and Professor Jacek Otlewski for guidance on ligand binding assays.

Dr. Michal Jasinski and Dr. Piotr Kachicki for help and assistance with antifungal tests.

Ms. Joanna Stępkowska for friendship and all the assistance with administrative matters.

All, present and former, colleagues and friends in the lab and in the Institute for the very good atmosphere and help, in and out of the lab. I specially thank Oliwia Pasternak for careful reading of this thesis.

I am grateful to the European Union for funding (Marie Curie Fellowship for Early Stage Researcher Training: MEST-CT-2004-504066) and my special thanks go to Dr Ralf Petri, scientific coordinator of the ADOPT program.

Above all I express my gratitude to my mother Piedade and my sister Inês.

This thesis is the result of three and half years of work. During this time, the experiments often took “five” minutes more then expected. I’m very grateful to Martyna for her tolerance and patience and, above all, for her optimism and sense of humor.

“Nothing in life is to be feared. It is only to be understood.”

Marie Curie (born Maria Skłodowska)

para ti Pai

The research described in this thesis has been carried out
at the Institute of Bioorganic Chemistry,
Polish Academy of Sciences in Poznan
in the Center for Biocrystallographic Research
under the supervision of Professor Mariusz Jaskolski

The project is part of the ADOPT program coordinated from the Max Planck
Institute in Cologne, Germany

Summary

The crystal structure of the *Lupinus luteus* LlPR-10.2B protein has been determined in complex with two different cytokinins. The two cytokinins represent the natural (zeatin) and the synthetic (diphenylurea) types of these plant hormones. The crystal structures have been determined at 1.35 and 1.95 Å resolution for the LlPR-10.2B complexes with zeatin and diphenylurea, respectively. Some adjustments of the protein structure have been observed, depending on the ligand and its binding mode. These rearrangements, as well as the binding patterns, have been investigated in detail. Comparisons with the apo structure of a close homologue, LlPR-10.2A, and with the PR-10-like CSBP protein in complex with zeatin have also been carried out. Ligand binding was further characterized using isothermal titration calorimetry. The same technique was also used for binding characterization of TcPR-10 (cacao) and HpHyp-1 (St John's wort) proteins.

Antifungal properties of three proteins (LlPR-10.2B, TcPR-10, HpHyp-1) have been investigated. None showed detectable antifungal activity.

Parts of this work have been published in:

Fernandes, H., Pasternak, O., Bujacz, G., Bujacz, A., Sikorski, M., Jaskolski, M. (2008), *Lupinus luteus* Pathogenesis-Related Protein as a Reservoir for Cytokinin. *J. Mol. Biol.* **378**, 1040-1051.

Fernandes, H., Konieczna, M., Kolodziejczyk, R., Bujacz, G., Sikorski, M., Jaskolski, M. (2008), Crystallization and preliminary crystallographic studies of Hyp-1, a St John's wort protein implicated in the biosynthesis of hypericin. *Acta Cryst.* F**64**, 405-408.

Abbreviations used:

2iP	6-(γ,γ-dimethylallylamino)purine
4-HPAA	4-hydroxyphenylacetaldehyde
ANS	8-anilino-1-naphthalenesulfonic acid
CCPU	*N*-(2-chloro-4-pyridyl)-*N'*-phenylurea
ADPs	Atomic Displacement Parameters
BIAs	Benzylisoquinoline Alkaloids
CCP4	Collaborative Computational Project No. 4
CD	Circular Dichroism
CSBP	Cytokinin-Specific Binding Protein
DLS	Dynamic Light Scattering
DMAB	Dimethylaminoborane
DPU	Diphenylurea
DSC	Differential Scanning Calorimetry
DSSP	Define Secondary Structure of Proteins
DTT	DL-Dithiothreitol
EDTA	Ethylene Diamine Tetraacetic Acid
FPLC	Fast Performance Liquid Chromatography
Hyp-1	Phenolic oxidative coupling protein from *Hypericum perforatum*
IMAC	Immobilized Metal ion Affinity Chromatography
IMM	Inner Mitochondrial Membrane
IPR	Intracellular Pathogenesis-Related
IPTG	IsoPropyl Thio-β-$_D$-Galactoside
ITC	Isothermal Titration Calorimetry
Kinetin	N^6-furfuryladenine
LB	Luria Bertani
MAD	Multiwavelength Anomalous Diffraction
MAP	Major Allergen Proteins
MIR	Multiple Isomorphous Replacement
MLP	Major Latex Proteins
MS	Mass Spectrometry
NCS	(S)-Norcoclaurine Synthase
NMR	Nuclear Magnetic Resonance

OD_{600}	Optical Density at 600 nm
OMM	Outer Mitochondrial Membrane
PAGE	Polyacrylamide Gel
PBS	Phosphate Buffered Saline
PCR	Polymerase Chain Reaction
PDA	Potato Dextrose Agar
PDB	Protein Data Bank
PEG	Polyethylene Glycol
Ppm	Parts per million
PR	Pathogenesis-Related
PR-1/17	seventeen classes of Pathogenesis-Related proteins
r.m.s.d.	root-mean-square deviation
SA	Simulated Annealing
SD	Standard Deviation
SDS	Sodium dodecyl sulfate
Se-Met	Selenomethionine
SIR	Single Isomorphous Replacement
SIRAS	Single Isomorphous Replacement with Anomalous Scattering
SPE-16	PR-10 homologue from *Pachyrrhizus erosus*
SR	Synchrotron Radiation
StAR	Steroidogenic Acute Regulary protein
START	StAR-Related lipid Transfer
STD	Saturation Transfer Difference
TEV	Tobacco Etch Virus
TLS	translation/libration/screw motion
T_m	midpoint temperature of a protein's unfolding transition
UV	Ultraviolet
WAP18	PR-10 homologue form *Morus bombycis* (mulberry)
Zeatin	N^6-(4-hydroxy-3-methyl-2-buten-1-yl)adenine

Pathogenesis-Related proteins are named using the first two letters of the species, followed by the letters PR and the number of the respective family (1/17). After a period, the class and subclass of the protein are given (e.g. LlPR-10.2B or PmPR-10.3.1).

Part I

Introduction

I.1. Pathogenesis-Related (PR) proteins

Plants, forced to grow in harsh conditions and exposed to pathogenic activity, have developed several means of defense. Besides establishing a physical barrier by strengthening their cell walls (Brisson *et al.*, 1994; Abramovitch and Martin, 2004), plants also produce antibiotic compounds called phytoalexins (Morrissey and Osbourn, 1999; Mysore and Ryu, 2004) and accelerate cell death (Dangl *et al.*, 1996; Greenberg, 1996; van Doorn and Woltering, 2005; Hofius *et al.*, 2007) to suppress the spread of infectious pathogens. Furthermore, the expression of a number of genes is induced by various types of pathogens (viruses, bacteria and fungi), or by chemicals such as ethylene and salicylic acid, which mimic the effect of pathogen infection and thus induce stress (Hammond-Kosack and Jones, 1996). These genes code the so-called PR (pathogenesis-related) proteins (Van Loon *et al.*, 1994).

PR proteins are a diverse group of proteins with various functions and activities which represent a generalized plant defense response against a range of pathogens (Markovic-Housley *et al.*, 2003). PR proteins are grouped into seventeen classes according to their biological activity or physicochemical properties and sequence homology (Van Loon *et al.*, 1994, 2006). According to accepted definition, they are synthesized *de novo* under stress conditions. Among them are chitinases, glucanases, enzymes of phenylpropanoid pathway, thionins, osmotines, proteases, proline-rich glycoproteins, and proteins of unknown function, PR-10, also known as intracellular pathogenesis-related (IPR) proteins (Van Loon *et al.*, 1994). They have been identified in more than 70 species belonging to both monocotyledonous and dicotyledonous plants (Colditz *et al.*, 2007). PR proteins do not constitute a superfamily of proteins but represent a collection of unrelated protein families which function as part of the plant defense system (Breiteneder, 2004).

I.2. PR-10 proteins

PR-10 proteins, first identified in cultured parsley cells (Somssich *et al.*, 1988), are small (155-163 amino acids), slightly acidic and resistant to proteases. They are usually encoded by multigene families, for instance in yellow lupine there are ten genes encoding PR-10 proteins (Handschuh *et al.*, 2007), suggesting that a number of different protein homologues can be expressed in various plant organs under different conditions. This makes the study of the function of PR-10 proteins very complex. Most PR proteins are either intracellular and localized in the vacuole or extracellular. In contrast, the PR-10 proteins are the only group that is cytosolic (Chadha and Das, 2006).

Based on sequence similarity (~50%), common allergens present in birch pollen grains (Breiteneder *et al.*, 1989), celery (Breiteneder *et al.*, 1995), apple (Vanek-Krebitz *et al.*, 1995) and other fruits are also included in the PR-10 group. Proteins of another PR-10 subfamily, called major latex proteins (MLP), are found in the latex of some plants, including the opium poppy (Nessler *et al.*, 1990; Osmark *et al.*, 1998) and bell pepper (Pozueta-Romero *et al.*, 1995). Interestingly, the classic PR-10 proteins and the MLPs have never been found to coexist in the same plant species (Sikorski *et al.*, 1999). Furthermore, based on distant sequence identity (<20%) and secondary structure prediction, a group of cytokinin-specific binding proteins (CSBP) was also postulated to be related to the PR-10 family (Fujimoto *et al.*, 1998) (Fig. 1). This tentative classification has been confirmed by X-ray crystallography (Pasternak *et al.*, 2006), which demonstrated structural similarity between CSBP and classic PR-10 proteins.

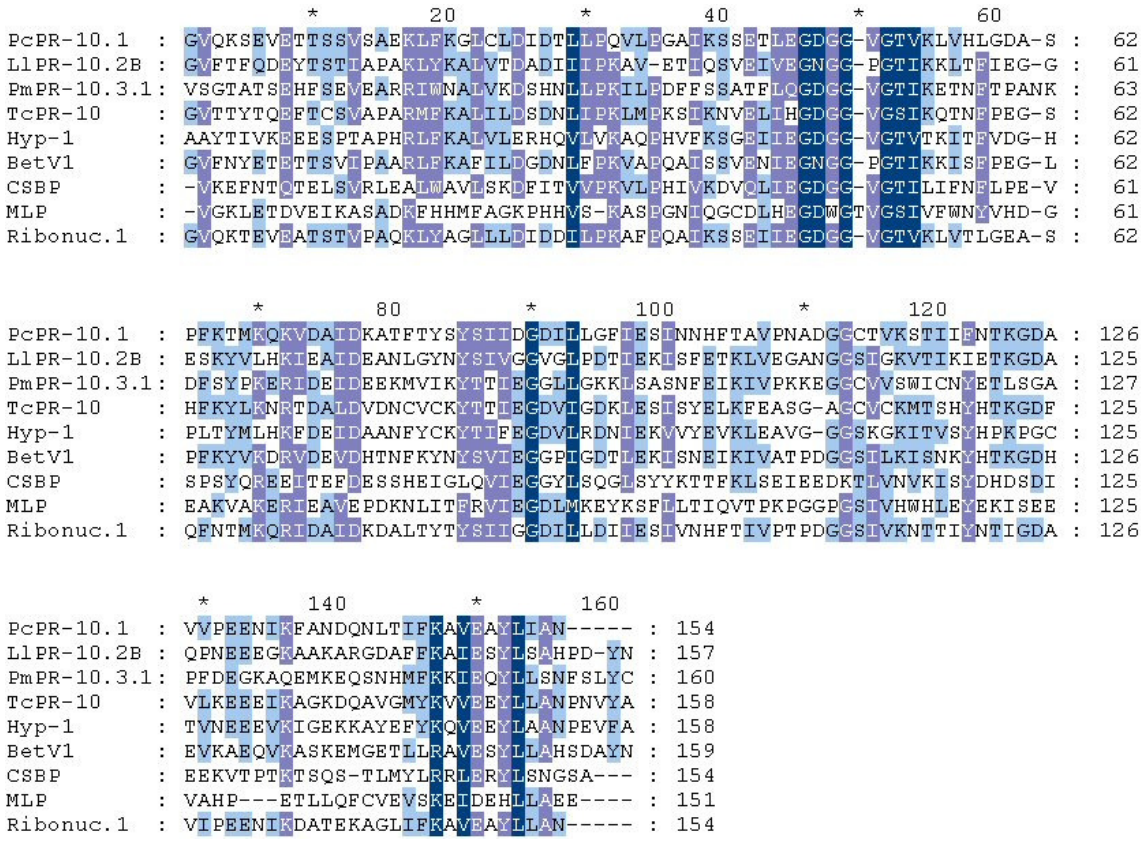


Figure 1. Amino acid sequence alignment of selected PR-10 proteins. Pc, Ll, Pm and Tc are *Petroselinum crispum* (parsley), *Lupinus luteus* (yellow lupine), *Pinus monticola* (white pine) and *Theobroma cacao* (cacao) PR-10 proteins, respectively. Hyp-1 is the *Hypericum perforatum* (St John's wort) phenolic oxidative coupling protein, Bet v 1 is a *Betula verrucosa* (white birch) pollen allergen, CSBP is the cytokinin-specific binding protein from *Vigna radiata* (mung bean), MLP is an *Arabidopsis thaliana* (thale cress) major latex protein, and Ribonuc. 1 is the *Panax ginseng* (ginseng) ribonuclease 1. The level of conservation is expressed by the darkness of the lettering background.

The sequence similarity and wide distribution of PR-10 proteins throughout the plant kingdom is an indication of an indispensable function in plants. However, it is difficult to assign a unique function to all PR-10 members because of lack of coordinated expression. They are involved in general defense mechanism, with some of the homologues found to accumulate around sites of pathogen invasion (Fristensky *et al.*, 1988; Somssich *et al.*, 1988; Matton and Brisson, 1989; Schmelzer *et al.*, 1989; Pinto and Ricardo, 1995; Breda *et al.*, 1996). It is also believed that PR-10 proteins have an important role in plant development since they have been found in dry seeds (Barratt and Clark, 1993; Warner *et al.*, 1994), and constitutive expression was observed in roots (Crowell *et al.*, 1992; Mylona *et al.*, 1994; Sikorski *et al.*, 1995; Legocki *et al.*, 1997;), stems (Warner *et al.*, 1994) and various parts of flowers (Breiteneder *et al.*, 1989; Warner *et al.*, 1993, 1994; Swoboda *et al.*, 1994; Constabel and Brisson *et al.*, 1995; Huang *et al.*, 1997). Moreover, senescent leaves often have increased levels of PR-10 protein expression (Crowell *et al.*, 1992). Also, some homologues appear to be induced by phytohormones like cytokinins (Carpin *et al.*, 1998), abscisic acid (Iturriaga *et al.*, 1994) or ethephon, an ethylene releasing compound (Broderick *et al.*, 1997).

Since 1988, when PR-10 proteins where first described in cultured parsley cells (Somssich *et al.*, 1988), many attempts have been made to establish the biochemical function and biological role of this protein family. Intriguingly, several biochemical functions have been proposed for this protein family but their true biological role is still obscure and a subject of debate. The PR-10 proteins are an atypical example in structural biology. Usually the role of a protein family is well established and only later structural models are elucidated to explain the function. Among PR-10 proteins, several crystal and solution models are known, but their role is still under debate and investigation.

I.2.1. The PR-10 fold

A number of PR-10 protein homologues have a known three dimensional structure (Table 1). The best structurally characterized PR-10 members are a group of yellow lupine proteins and the Bet v 1 family. These two groups together with the remaining structures create a large framework for structural comparisons. All the structures represent the three different branches of the PR-10 family, namely the classic PR-10 proteins, the cytokinin-specific binding proteins and the major latex proteins (Table 1).

Table 1. PR-10 protein structures in the PDB.

Source	Name	Method	PDB code	Reference
White Birch *B. verrucosa*	Bet v 1*a*	X-ray (2.0)*	1bv1	Gajhede *et al.*, 1996
	Bet v 1*a*	NMR (20)	1btv	Gajhede *et al.*, 1996
	Bet v 1*a* M139L mutant	NMR (23)	1b6f	Schweimer *et al.*, 1999
	Bet v 1*a* antibody complex	X-ray (2.9)	1fsk	Mirza *et al.*, 2000
	Bet v 1*l* deoxycholate complex	X-ray (1.9)	1fm4	Markovic-Housley *et al.*, 2003
	Bet v 1*a* E45S mutant	X-ray (3.1)	1llt	Spangfort *et al.*, 2003
	Bet v 1*a* N28T, K32Q, E45S P108G mutant	X-ray (2.15)	1qmr	Holm *et al.*, 2004
Cherry *P. avium*	Pru av 1	NMR (22)	1e09	Neudecker *et al.*, 2001
	Pru av 1 E45W mutant	NMR (24)	1h2o	Neudecker *et al.*, 2003
Celery *Apium graveolens*	Api G 1	X-ray (2.9)	2bk0	Schirmer *et al.*, 2005
Mung bean *V. radiata*	CSBP	X-ray (1.2)	2flh	Pasternak *et al.*, 2006
	CSBP zeatin and Ta_6Br_{12} complex	X-ray (1.8)	3c0v	Pasternak *et al.*, 2008
Thale cress *Arabidopsis thaliana*	MLP	NMR (20)	2i9y	Unpublished
Jicama *Pachyrhizus erosus*	SPE16	X-ray (2.2)	1tw0	Unpublished
	SPE16 ANS complex	X-ray (2.3)	1txc	Unpublished
Yellow lupine *L. luteus*	LlPR-10.1A	X-ray (1.95)	1icx	Biesiadka *et al.*, 2002
	LlPR-10.1B	X-ray (2.25)	1ifv	Biesiadka *et al.*, 2002
	LlPR-10.2A	X-ray (1.9)	1xdf	Pasternak *et al.*, 2005
	LlPR-10.2B zeatin complex	X-ray (1.35)	2qim	Fernandes *et al.*, 2008a
	LlPR-10.2B diphenylurea complex	X-ray (1.95)	3e85	Unpublished

* The numbers in parentheses indicate the resolution in Å for the X-ray structures, or the number of models for the NMR structures.

All the PR-10 structures have the same general fold consisting of a seven-stranded antiparallel β-sheet wrapped around a long C-terminal α-helix (Fig. 2A). These two structural elements, together with two additional short helices, create a hydrophobic cavity (Fig. 2B),

whose volume is unusually large for proteins of this size. The abnormal size of the cavity suggests its importance for physiological function, an obvious possibility being hydrophobic ligand binding.

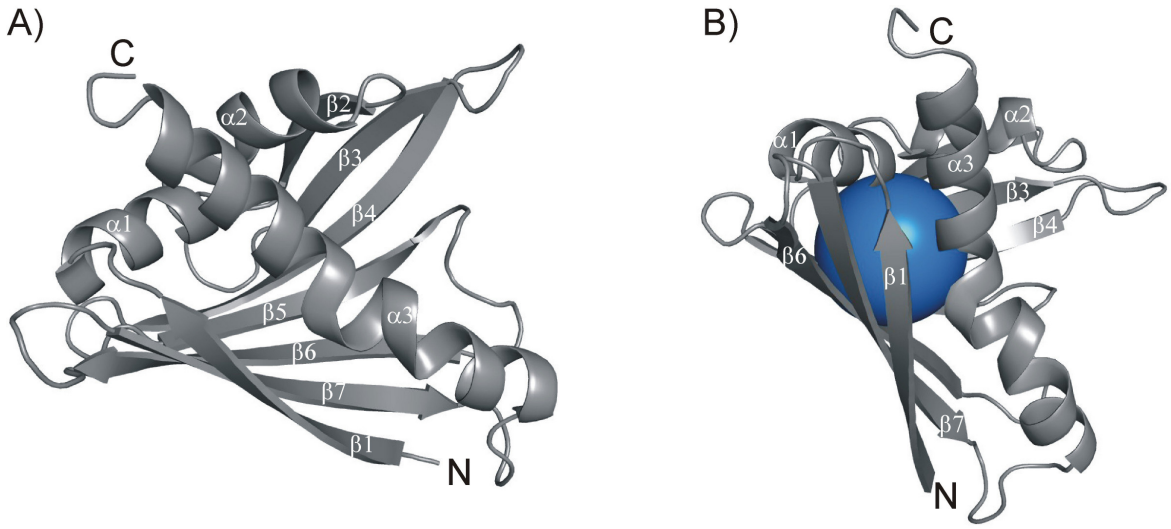


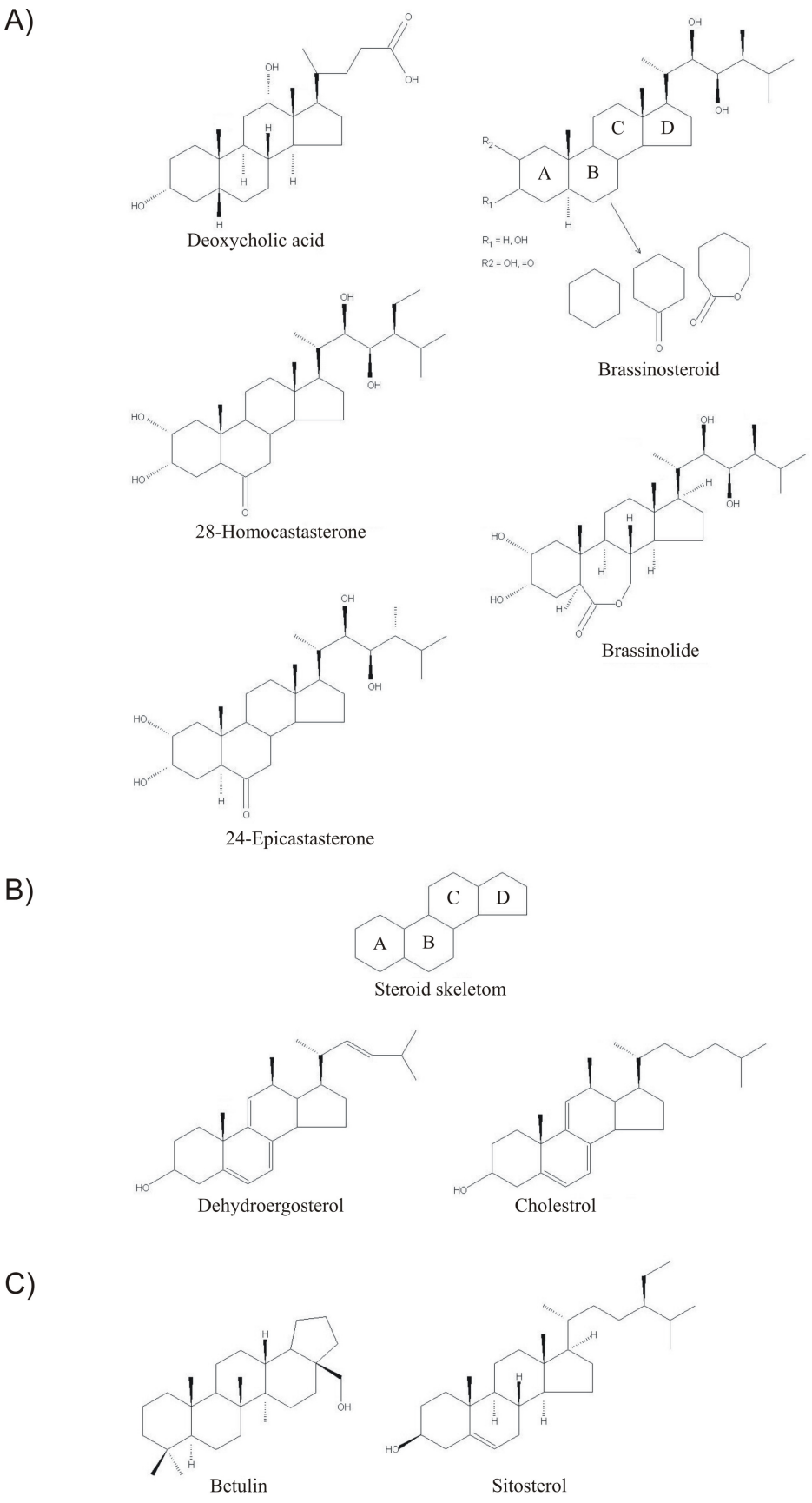
Figure 2. The general fold of PR-10 proteins. Here, the LlPR-10.1A protein (PDB code 1icx) is shown. (A) The secondary structure elements are labeled. (B) Another view, with the protein cavity represented as a blue sphere.
All the structural illustrations have been prepared using PyMol (DeLano, 2002).

I.2.2. Postulated biochemical functions

The following sections describe the potential biochemical functions which have been postulated for PR-10 proteins.

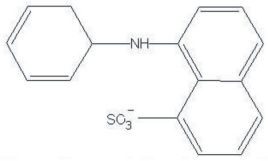
I.2.2.1. Ligand-binding

The size and lining of the internal cavity suggest that it might be a binding site for hydrophobic ligands. Additional evidence for that possible function comes from (i) the structural similarity between the PR-10 members and the START domain of human MLN64 protein, which is a steroid binding domain related to steroidogenic acute regulary protein (StAR) involved in cholesterol translocation in human placenta and brain (Tsujishita and Hurley, 2000; Markovic-Housley *et al.*, 2003; Koistinen *et al.*, 2005), (ii) the structural similarity with CSBP proteins (Pasternak *et al.*, 2006) and (iii) the crystal structure of a Bet v 1 homologue in complex with deoxycholate (Scheme 1) (Markovic-Housley *et al.*, 2003). The possibility of a ligand binding function of PR-10 proteins has prompted several groups to investigate this hypothesis in more detail. In 2001 Neudecker *et al.* determined the solution structure of Pru av 1, a major cherry allergen similar to Bet v 1*l*, by NMR spectroscopy. They showed that Pru av 1 can bind the plant brassinosteroid homocastasterone in the hydrophobic cavity (Scheme 1).

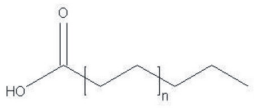


Scheme 1. Structure of (A) deoxycholic acid and selected brassinosteroids, (B) selected steroids and (C) the triterpene betulin and the phytosterol, sitosterol.

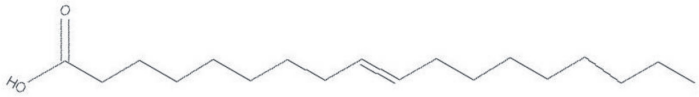
In 2002 after showing, by means of NMR spectroscopy, that Bet v 1 can bind the fluorescent probe ANS (Scheme 2) in the cavity, Mogensen and co-workers identified a range of physiologically relevant ligands of Bet v 1, using ANS displacement assay. The binding of non-fluorescent ligands was indirectly measured by the reduction of ANS fluorescence. The identified ligands included: fatty acids, flavonoids, and cytokinins (Scheme 2 and 3).



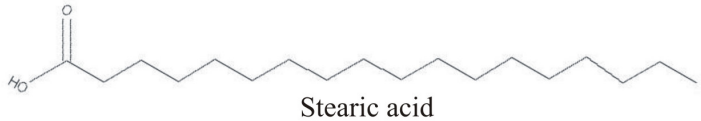
8-anilino-1-naphthalenesulfonic acid (ANS)



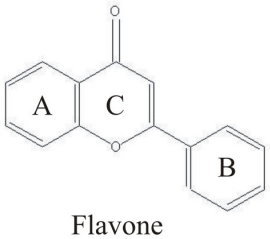
Fatty acids



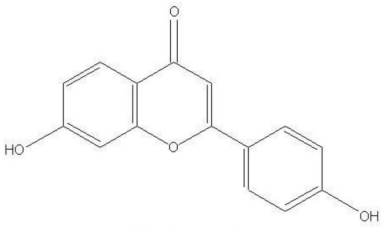
Oleic acid



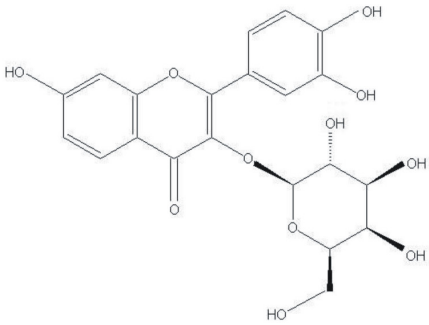
Stearic acid



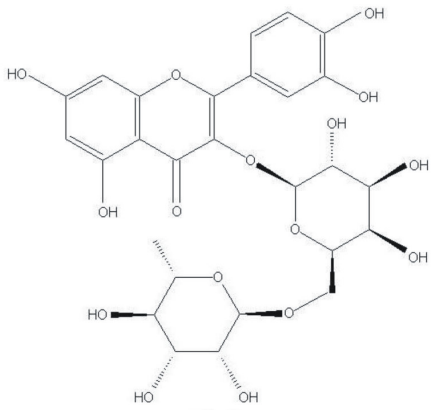
Flavone



Naringenin

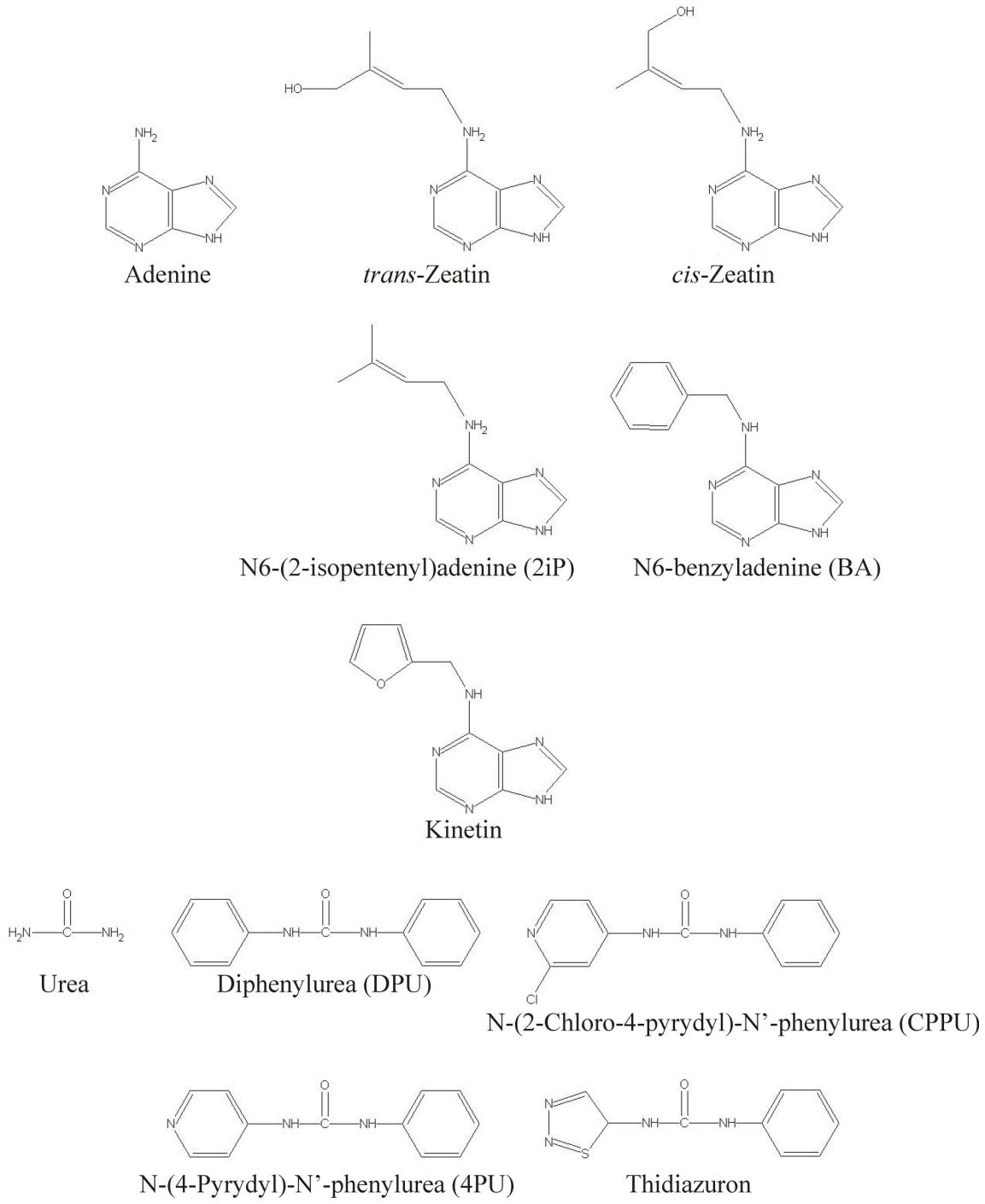


Hyperoside



Rutin

Scheme 2. Structures of the ANS fluorescent probe, selected fatty acids, and selected flavonoids.



Scheme 3. Structure of selected purine (adenine) and urea-type cytokinins.

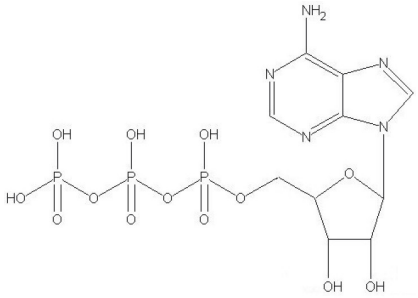
The ligands bound with low affinity, generally in the micromolar range (Mogensen *et al.*, 2002). ANS was also shown to bind to SPE16 (Wu *et al.*, 2003) and CSBP proteins (Pasternak *et al.*, 2006). Crystal structure of SPE16 (unpublished, PDB ID: 1txc) showed that the SPE16 protein accommodates the ANS molecules in the internal cavity. ANS is considered to bind to pre-existing hydrophobic (nonpolar) surfaces of proteins, primarily through its nonpolar

anilinonaphthalene group. Such binding is followed by an increase in ANS fluorescence intensity, and thus the molecule can be used as a "hydrophobic probe" to examine nonpolar character of proteins and membranes (Matulis *et al.*, 1999).

The affinity of fatty acids for Bet v 1 exhibits a parabolic dependence on fatty acids chain length. The highest affinities were observed for chains of 14-18 carbon atoms. Surprisingly, a kink caused by a double bond in the fatty acids chain did not affect the binding affinity. Oleic and stearic acid, both with 18 carbon atoms in the chain, but the former one with a kinked chain at position 9, have identical affinities (Mogensen *et al.*, 2002) (Scheme 2) for the protein. It was also shown that Bet v 1 can bind flavonoids, with two examples, flavone and naringenin (Scheme 2), reported by Mogensen *et al.* (2002). Two very potent, naturally occurring, cytokinins have been studied and shown to bind to Bet v 1, namely 2iP and kinetin (Scheme 3). Additionally, the authors discovered that Bet v 1 can bind dehydroergosterol, a naturally occurring sterol (Scheme 1). However, the Bet v 1 protein failed to bind some plant hormones such as indole-3-acetic acid, gibberellic acid and abscisic acid (Mogensen *et al.*, 2002). One year later (2003), the first crystal structure of Bet v 1 in complex with a ligand, deoxycholate, was published (Markovic-Housley *et al.*, 2003). Deoxycholate is not a plant metabolite, and consequently can not be a natural ligand for Bet v 1. However, it is structurally very similar to brassinosteroids (Scheme 1), which are plant steroid hormones involved in plant development and growth processes (Clouse, 1996, 2001). Unexpectedly deoxycholate, which was used as a reagent in the solubilization of the recombinant protein, remained bound to the protein throughout the purification process, indicating a very high affinity towards Bet v 1*l*. By means of mass spectrometry (MS), the authors showed that the protein can bind two brassinoloids, brassinolide and 24-epicastasterone (Scheme 1). Together, the crystal structure and the MS experiments indicate an involvement of Bet v 1 in the transport of plant steroids. This transport is rather general than specific since the protein binds not only brassinolide, the most active brasinosteroid, but also the less active 24-epicastasterone. Nevertheless, the binding constants are different. In 2005, Koistinen and coworkers used saturation transfer difference (STD) NMR to study PR-10 protein-ligand interactions. NMR spectroscopy offers several different techniques to monitor protein–ligand interactions (Carlomagno, 2005). One advantage of NMR is that it conserves the sample, enabling its (re)use for several consecutive measurements. Moreover it is possible to directly identify the binding component from a mixture of potential ligands, which is particularly important in screening ligand libraries. The major bottlenecks of the NMR methods for high-throughput screening are relatively long acquisition times and dependence on ligand excess. However, depending on the protein, concentration, and the information sought, the most efficient technique can be chosen. The STD

NMR technique which is based on the difference between a saturation transfer spectrum and a normal NMR spectrum (Mayer and Meyer, 1999) is a fast and informative way to explore ligand binding to proteins. One of the major advantages of this technique is a relatively small amount of protein needed compared to other NMR techniques. Additionally, the technique reveals the binding epitope of the ligand (Mayer and Meyer, 1999 and 2001). Since the saturation transfer is more efficient for those parts of the ligand that come into close contact with the protein, the relative intensities of ligand signals indicate its binding epitope(s). However, this method has also some drawbacks, as no information about the binding site of the protein can be gained using STD NMR and it is impossible to detect high-affinity ligands which undergo slow chemical exchange on the NMR time-scale (Mayer and Meyer, 2001). This last disadvantage can be overcome by combining NMR screening methods with competition binding experiments (Wang *et al.*, 2004). Alternatively, high-affinity ligands can be studied by structure-activity relationships (SAR) using NMR (Shuker *et al.*, 1996).

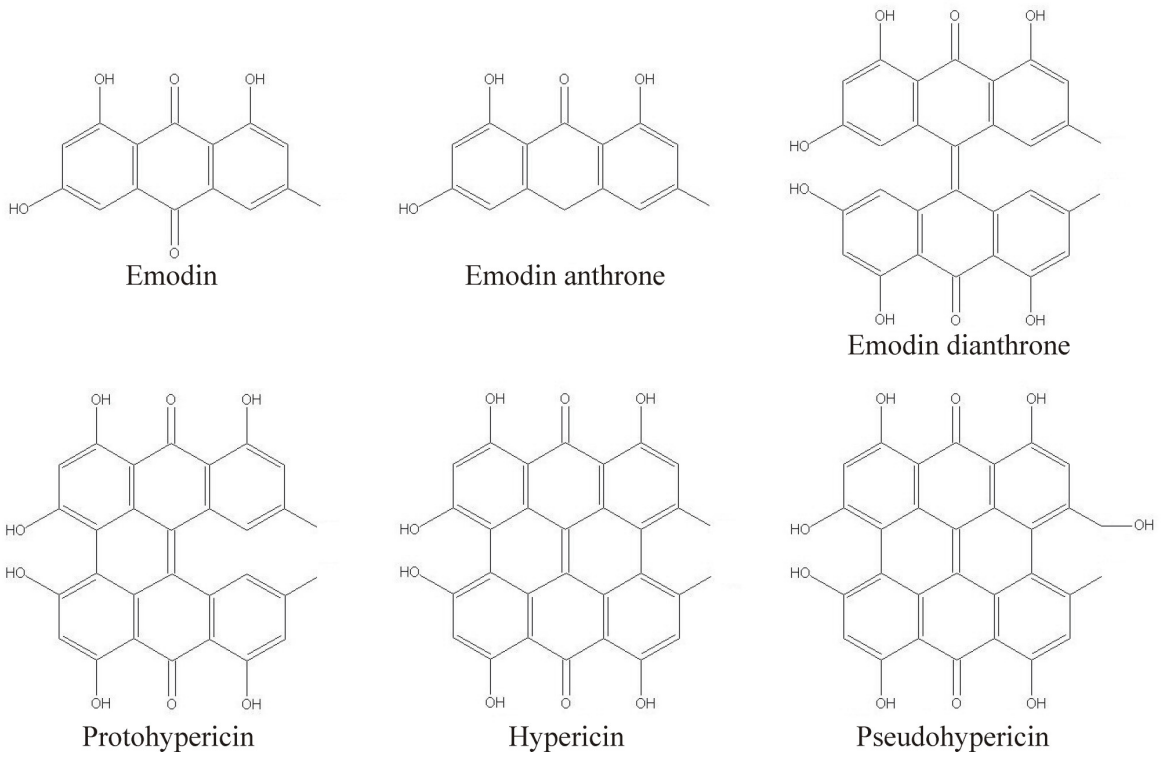
Koistinen and colleagues (2005) investigated another interesting PR-10 homologue form *Betula pendula* (birch), BpPR-10c, which shares 72% sequence identity with Bet v 1*l*, using STD NMR. BpPR-10c is not constitutively expressed in pollen, and it is the only homologue known that can be post-translationally modified by glutathione. The STD NMR results indicated that both reduced and glutathiolated forms of BpPR-10c interact with kinetin (Scheme 3), and that adenine, rather than the furan moiety of the kinetin molecule, interacts with the protein. No interaction of the protein with ATP (Scheme 4), another adenine derivative, has been observed despite the fact that the same method was used to characterize the binding of ATP to protein kinases (McCoy *et al.*, 2005). The binding of flavonoid glycosides by BpPR-10c, namely hyperoside and rutin (Scheme 2), was also reported. Hyperoside is one of the main phenolic compounds in birch (Laitinen *et al.*, 2000), and rutin contains a disaccharide moiety making it a relatively large molecule. The results clearly indicated that the interaction occurs



Adenosine triphosphate (ATP)

Scheme 4. Structure of adenosine triphosphate.

through the B ring of the quercetin moiety since no saturation transfer was observed for the glycoside and the A ring protons. Inspired by the results of Bais *et al.* (2003), reporting that Hyp-1 (phenolic oxidative coupling protein), a PR-10 homologue from St John's wort, forms hypericin from emodin (Scheme 5) (see "Phenolic oxidative coupling protein" subsection), Koistinen and colleagues (2005) investigated the possibility of an interaction between emodin and BpPR-10c. The authors demonstrated that emodin indeed interacts with BpPR-10c, however they did not observe any significant differences in STD factors for the aromatic protons and did not detect any formation of hypericin within the 72 hours of the NMR experiment. Additionally, the authors performed some docking experiments, which showed that the protein can accommodate two emodin molecules in its cavity. Docking simulations also showed that BpPR-10c can bind the triterpene betulin, sitosterol, and kinetin (Scheme 1 and 3). Two kinetin molecules could be placed in the protein cavity, and one extra kinetin molecule could be allocated close to the glycine-rich loop. This loop has been postulated as a possible cytokinin binding site in the yellow lupine LlPR-10.1A and LlPR-10.1B structures (Biesiadka *et al.*, 2002). Furthermore, Koistinen and colleagues showed that BpPR-10c interacted with deoxycholate, similarly to Bet v 1*l*, but not with indole-3-acetic acid or gibberellic, again resembling the behavior of the Bet v 1 protein (Mogensen *et al.*, 2002).



Scheme 5. Structure of emodin, hypericine and their derivatives.

The crystal structure of CSBP revealed that the protein has all the structural elements defining the PR-10 fold, including the seven-stranded β-sheet, the C-terminal α-helix, the loop system with a glycine-rich loop, and the internal cavity (Pasternak *et al.*, 2006). In the crystal structure, the CSBP molecule forms a complex with *trans*-zeatin, a natural plant hormone from the cytokinin family. Additionally, the authors showed that CSBP can bind other cytokinins, namely kinetin, 2iP and CPPU (Scheme 3). The authors reported that in solution the binding stoichiometry of zeatin to CSBP was 1:1 and that the K_d values cover the micromolar range. Cytokinins, together with other plant hormones are involved in the differentiation of the shoot meristem and root tissues, mitosis, leaf formation and senescence, and chloroplast development (Mok and Mok, 1994). Cytokinins are structurally diverse and can be divided into two main groups: adenine analogs (natural) and urea derivatives (synthetic) (Scheme 3). The adenine analogs contain an adenine ring with an N6-substitution of an isopentenyl backbone, an aromatic hydroxybenzyl group or a furfuryl group.

I.2.2.2. Enzymatic activities

I.2.2.2.1. RNase

In 1994 Moiseyev and co-workers isolated a protein with ribonuclease activity from a callus cell culture of *Panax ginseng*. They were surprised that this new ribonuclease did not share homology with other ribonucleases. It showed 60-70% sequence identity with two intracellular pathogenesis-related (IPR) proteins from parsley. Guided by these initial results, other groups also reported RNase activity of several other PR-10 proteins, including birch Bet v 1 (Bufe *et al.*, 1996; Swoboda *et al.*, 1996), white lupine LaPR-10 (Bantignies *et al.*, 2000), yellow lupine LlPR-10.1B (Biesiadka *et al.*, 2002), birch BpPR-10c (Koistinen *et al.*, 2002), cotton GaPR-10 (Zhou *et al.*, 2002), PePR-10 protein from jicama (Wu *et al.*, 2002), hot pepper CaPR-10 (Park *et al.*, 2004), yellow-fruit nightshade SsPR-10 (Liu *et al.*, 2006), peanut AhPR-10 (Chadha and Das, 2006), pea PsPR-10.1 (Srivastava *et al.*, 2006) and Chinese huang-qi (*Astragalus membranaceus*) AmPR-10 (Yan *et al.*, 2008). This activity is, however, controversial, since it is not universal throughout the PR-10 protein family.

In 1996, using the available PR-10 sequences, Walter and colleagues found 12 conserved residues that could be involved in the ribonuclease activity. One year later, Moiseyev and co-workers (1997), in the light of the appearance of a few more PR-10 homologues, reduced the number of conserved residues from twelve to nine. Most of them were glycine or proline residues that are probably essential for the polypeptide fold, as demonstrated by Gajhede *et al.* (1996) using the crystal and NMR structures of the birch Bet v 1 protein (Table 1). The

remaining conserved residues have side chains with functional groups which may be involved in the catalytic reaction: Glu96, Glu148 and Tyr150 (ginseng ribonuclease 1 sequence). The three-dimensional structure of Bet v 1 shows that Glu148 and Tyr150 are located at two opposing sides of the long C-terminal α-helix (Fig. 2), while Glu96 is located at the N-terminus of β6 which is relatively far from the other two residues. The three-dimensional structure of Bet v 1 also shows the structure of the glycine-rich loop, with the sequence GXGGXGXXK (positions 46-54 in ginseng ribonuclease 1). This loop shows remarkable sequence identity to P-loop, also known as phosphate binding loop. The P-loop motif was found in some nucleotide-binding proteins, in myosin head piece, in many kinases, elongation factors, as well as in several proteins involved in active transport (Saraste *et al.*, 1990). The conservation of such a sequence in PR-10 proteins may indicate the presence of a phosphate binding site, and therefore, a likely place for binding a phosphate group of RNA. However, it was observed (Biesiadka *et al.*, 2002) that the conformation of the PR-10 glycine-rich loop is very stable and different from the P-loop conformation.

For further understanding of the importance of the highlighted conserved residues in the ribonuclease-like activity of PR-10 proteins, Wu and colleagues (2003) performed site directed mutagenesis of three conserved amino acids (Glu95Ala, Glu147Ala and Tyr149Ala) of the SPE16 sequence, highlighted in the Moiseyev *et al.* (1997) studies. Additionally a version of the protein with a truncation of the glycine-rich loop was also constructed. The authors have demonstrated that recombinant SPE16 and the Glu95Ala mutant retain the ribonuclease activity, while the Glu147Ala, Tyr149Ala mutants and the glycine-rich-loop truncated versions show a noticeable decrease of the activity (Wu *et al.*, 2003). Similar studies have been reported for the PR-10 homologue from cotton (*Gossypium arboreum*, GaPR-10), where the mutational strategy involved the following replacements: Gly51Ala, Lys55Asn, Glu96Lys, Glu148Lys and Tyr150Phe (Zhou *et al.*, 2002). These amino acids were chosen because they were either in the region corresponding to the glycine-rich loop of Bet v 1 (Gly51 - Lys55), or were proposed to be involved in the catalytic reaction (Glu96, Glu148, and Tyr150). The five mutants displayed significantly lower RNase activity than wild type GaPR-10. The substitutions in the glycine-rich loop (Gly51Ala and Lys55Asn) resulted in about 50-60% decrease of RNase activity. It should be mentioned that the glycine-rich loop of GaPR-10 is atypical, with only two glycine residues instead of four. In the case of the Glu96Lys mutation, a comparable 55% decrease of activity was observed. The biggest impact on the activity was observed for the mutations at the C-terminus of the protein (Glu148Lys and Tyr150Phe), despite the fact that these two amino acids are located at two opposite sides of the long C-terminal α-helix of the molecule (Zhou *et al.*, 2002). More recently, another group has reported a similar mutational study, this time using a

peanut (*Arachis hypogaea*) PR-10 protein (AhPR-10). Chadha and Das (2006) prepared three mutated versions of the protein (Lys54Asn, Phe148Ser and His150Gln), and reported that the ribonuclease activity of the Lys54Asn mutant was almost completely abolished, whereas the Phe148Ser and His150Gln mutations resulted only in a decrease of the activity level. The AhPR-10 Phe148 and His150 residues correspond to the conserved Tyr148 and Glu150 residues present in all other PR-10 proteins with reported RNase activity.

Park and colleagues (2004) discovered that the hot pepper CaPR-10 protein could be phosphorylated and that such a modification stimulated the RNase activity. Also, the StAR protein (Arakane *et al.*, 1997), structurally related to the PR-10 class (see "Ligand-Binding" subsection), was shown to be phosphorylated, and its steroidogenic activity was modulated by this post-translational modification.

Despite the evidences of RNase activity for some PR-10 homologues, its biological significance is arguable. There are two reasons for that: (i) it is difficult to reconcile the low level of RNase activity with the high abundance of PR-10 proteins (unusual for enzymes) and (ii) it is puzzling that only one of two very close homologues shows ribonuclease activity, as in the case of the LlPR10.1A and LlPR10.1B homologues from yellow lupine (Biesiadka *et al.*, 2002). It should be remembered, however, that knowledge about cytosolic ribonucleases is rather limited, not only in plants but in eukaryotes in general. In addition, the information about plant RNases has so far been gathered outside of the plant pathology field. Consequently, there is no information available on specific cytosolic ribonucleases involved in plant defense (Walter *et al.*, 1996).

I.2.2.2.2. Phenolic oxidative coupling protein

An exciting relatively recent discovery in the PR-10 field is the implication, based on sequence homology (Fig. 1), of a PR-10 protein in the enzymatic reaction that is claimed to be responsible for the biosynthesis of hypericin from emodin (Scheme 5) in St John's wort (*Hypericum perforatum*) (Bais *et al.*, 2003).

St John's wort is an herbaceous perennial plant that has received considerable interest in North America and Europe due to its medicinal properties. In addition to its millennia-sanctioned applications for various ailments in folk medicine, several recent clinical studies have demonstrated the effectiveness of St John's wort as a phytomedicinal treatment for depression (Briskin and Gawienowski, 2001). Morphologically, St John's wort is characterized by the presence of several types of secretory structures, including translucent glands, secretory canals and, the most important structure for this plant, dark glands, which do not have a cavity and are mainly visible on leaves and flowers (Onelli *et al.*, 2002; Zobayed *et al.*, 2006). The occurrence

of these metabolically active multicellular structures is unique and characteristic for some species of the genus *Hypericum* (Kosuth *et al.*, 2007).

Secondary metabolites in medicinal plants have, in addition to diverse physiological functions during the plant's life cycle, an important role as a source of active pharmaceuticals (Kosuth *et al.*, 2007). The main bioactive compounds of *H. perforatum* belong to the hypericin group (hypericin and pseudohypericin, Scheme 5), thought to function as defensive allelochemicals in the plant (Zobayed *et al.*, 2006). Hypericin, a photosensitive, red-colored naphthodianthrone, is the major medicinal compound of *H. perforatum*. It has been successfully tested in numerous studies for the treatment of mild to moderate depression (Deltito and Beyer, 1998), as well as in antitumor and antiviral (HIV, hepatitis C virus) assays (Kubin *et al.*, 2005 and references therein). Screening studies for inhibitory effects of hypericin on various pharmaceutically important enzymes, such as MAO (monoaminoxidase), PKC (protein kinase C), dopamine-β-hydroxylase, reverse transcriptase, telomerase, and CYP (cytochrome P450), have yielded results supporting the therapeutic potential of hypericin. Studies of hypericin effects on GABA-activated (gamma amino butyric acid) currents and NMDA (N-methyl-D-aspartate) receptors also indicate therapeutic promise in stroke (apoplexy) (Kubin *et al.*, 2005).

Bais *et al.* (2003) hypothesized that in *H. perforatum* hypericin is synthesized from the emodin precursor in an enzymatic dimerization reaction. Emodin is an anthraquinone which has been suggested to act as feeding deterrent, to be an allelopathic compound, to have antimicrobial activity, and to participate in seed germination and dispersal (Izhaki, 2002). The final localization of hypericin in the dark glands is well accepted (Briskin and Gawienowski, 2001) but the site of its synthesis is unknown, the dark glands being a possibility (Onelli *et al.*, 2002) supported by the presence of emodin at a high concentration in the dark glands and its absence in other tissues (Zobayed *et al.*, 2006).

Bais and coworkers were able to isolate a gene, *hyp-1*, whose product seems to be responsible for the direct conversion of emodin to hypericin in dark-grown cell cultures of *H. perforatum*. The cDNA of *hyp-1* has 725 bp with an open reading frame of 480 nucleotides (including stop codon) and codes for a protein of 159 amino acid residues, with a molecular mass of 17.8 kDa and a pI of 5.54. Following the mRNA encoding the Hyp-1 protein, Kosuth and coworkers (2007) expected to find elevated transcript levels in the dark glands. The results were, however, negative. In addition, Kosuth and coworkers found the highest level of transcription of the *hyp-1* gene, but merely traces of hypericin, in roots, where there are no dark glands at all. These findings may have several explanations: (i) the sites of hypericin biosynthesis and accumulation are different; (ii) the level of steady state mRNA is not correlated with the level of the encoded polypeptide; (iii) the conditions of the *in vitro* experiment are

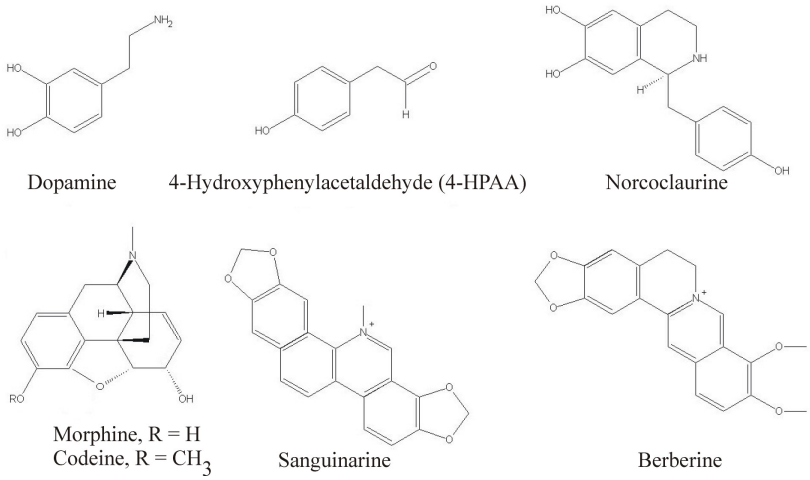
different than *in vivo*; (iv) it is questionable whether the *hyp-1* gene is really involved in hypericin biosynthesis (Kosuth *et al.*, 2007).

A BLAST search using the deduced amino-acid sequence of HpHyp-1 detected similarity to the PR-10 family (Bais *et al.*, 2003). Protein sequence comparisons revealed significant homology between HpHyp-1 and a major apple allergen (Mal d 1), with 45.3% sequence identity and 57.2% sequence similarity, followed by Pru av 1, a major cherry allergen (37.5% sequence identity and 50.6% sequence similarity).

Bais and colleagues (2003) hypothesized that Hyp-1 catalyzes an initial condensation reaction between emodin and emodin anthrone followed by dehydratation to form emodin dianthrone. Emodin dianthrone may subsequently undergo phenolic oxidation to protohypericin, which in turn yields hypericin (Scheme 5). Concluding, Hyp-1 may catalyze at least four serial reactions, possibly a condensation followed by dehydration and two phenolic oxidative coupling reactions.

I.2.2.2.3. (S)-Norcoclaurine synthase

Another notable point of interest in the PR-10 research field was the discovery of an enzymatic reaction involving a protein whose sequence is similar to that of PR-10 proteins (30-40% sequence identity). (S)-Norcoclaurine synthase (NCS) (EC 4.2.1.78) catalyzes the condensation of the amine moiety of dopamine and the aldehyde group of 4-HPAA as the first step in the biosynthesis of more than 2500 benzylisoquinoline alkaloids (BIAs) in plants, for instance morphine, sanguinarine, and berberine (Scheme 6). Benzylisoquinoline alkaloids are a large and diverse group of secondary metabolites found in several related plant families including Papaveraceae, Ranunculaceae, Berberidaceae, Fumariaceae, and Menispermaceae (Facchini, 2001). The general role of alkaloids in the chemical defense of plants against herbivores and pathogens suggests that BIAs contribute to the



Scheme 6. Structure of norcoclaurine, its precursors, and selected benzylisoquinoline alkaloids.

higher reproductive fitness of plants with the ability to produce these compounds (Schmeller *et al.*, 1997; Wink, 2003).

NCS was discovered in 1981 by Rueffer *et al.*, but not isolated or characterized at that time. In 2002, Samanani and Facchini purified and characterized the enzyme and, two years later, Samanani *et al.* (2004) isolated the gene from *Thalictrum flavum* (Tf*ncs*). The predicted amino acid sequence of TfNCS showed substantial homology with the PR-10 protein family. Weaker homology was also detected between NCS and major latex protein (MLP) homologues. Amino acid sequence alignment showed 28-38% identity and 50-60% similarity with PR-10 proteins. NCS also displayed 35% sequence identity with the Hyp-1 protein (Samanani *et al.*, 2004). The sequence of NCS is, however, unique because of oligopeptide extensions at the N- and C-termini of the predicted protein product. Analysis of the NCS amino acid sequence using the SignalP Server (http://www.cbs.dtu.dk/services/SignalP) (Emanuelsson *et al.*, 2000) revealed the presence of a putative 19-residue N-terminal signal peptide (Samanani *et al.*, 2004). The central biosynthetic role of (S)-norcoclaurine suggests an ancient recruitment of NCS from a PR-10 ancestor as a key evolutionary event, supported by molecular phylogenetic analysis, allowing certain plant taxa to produce numerous and structurally diverse benzylisoquinoline alkaloids with a large range of physiological and pharmacological functions (Samanani *et al.*, 2004). The duplication and subsequent insertion of an NCS ancestor into another gene encoding an N-terminal signal peptide could explain the unique occurrence of this sequence in NCS (Samanani *et al.*, 2004; Liscombe *et al.*, 2005).

Norcoclaurine synthase has sigmoidal substrate saturation kinetics, which is typical of enzymes with a regulatory role in the metabolism. NCS appears to assemble as a dimer (Samanani and Facchini, 2002) and shows a significant increase of activity in the presence of dopamine (Samanani *et al.*, 2004). The cellular abundance of dopamine is, however, surprisingly high, if the availability of this substrate is expected to play a role in the regulation of the metabolic flux into benzylisoquinoline alkaloid pathways. Nevertheless, the dopamine pool is localized within a vacuolar compartment suggesting that levels of regulation can be achieved by subcellular trafficking of dopamine and the abundance of NCS. On the other hand, the putative signal peptide suggests that, in contrast to classic PR-10 proteins, NCS is associated with a subcellular compartment other then cytosol, which could mediate accessibility to the cellular dopamine pool (Samanani *et al.*, 2004). Subcellular compartmentalization of NCS is consistent with the association of other enzymes of the benzylisoquinoline alkaloid biosynthetic pathway with the endoplasmic reticulum (Facchini, 2001).

In 2005, Liscombe *et al.* identified, isolated and functionally characterized two NCS homologues from opium poppy (*Papaver somniferum*, PsNCS). PsNCS1 and PsNCS2 also

possess the N- and C-terminal oligopeptide extensions; however, the TargetP Server did not predict any signal peptide for the opium poppy NCS proteins. The PsNCS C-terminus possesses a prominent C-terminal oligopeptide that extends 22 residues beyond the TfNCS sequence. Although the level of identity between the sequence of PsNCS1 and TfNCS is only 39% when compared by global alignment, it increases to 50% when a local alignment, excluding divergent N- and C-terminal domains, is performed. NCS activity was checked for six recombinant proteins that share the most extensive sequence identity with PsNCS1, namely PsPR-10.1, PsPR-10.2, OsPR-10, VrCSBP, HpHyp-1, BpMAP (Bet v 1*a*) and PmPR-10. None of the recombinant proteins showed any detectable NCS activity, in contrast to the high levels of activity displayed by recombinant PsNCS2 and TfNCS (Liscombe *et al.*, 2005). Recently, Minami and co-workers (2007) isolated and characterized an NCS protein from *Coptis japonica* (CjNCS1). CjNCS1 was reported to catalyzes the condensation of dopamine and 4-HPAA in a stereo-specific way. Interestingly the authors, in the same report, showed that CjPR-10A also catalyzes the NCS reaction.

I.2.2.3. Antimicrobial activities

Plants have evolved both general and highly specialized defense responses to prevent diseases caused by the majority of microbial pathogens they encounter. The defense mechanisms involve preformed and induced antimicrobial compounds to repel the pathogen attack, cell wall reinforcements to prevent pathogen entry, and programmed cell death to limit pathogen establishment and spread (Abramovitch and Martin, 2004).

I.2.2.3.1. Antifungal activity

Some PR-10 homologues have been shown to possess antifungal activity. For instance, AhPR-10 was reported to have antifungal activity against *Fusarium oxysporum* and *Rhizoctonia solani*. However, the protein does not show similar activity against *Sclerotium rolfsii*, *Aspergillus xavus, Aspergillus niger* and *Phytophthora infestans* (Chadha and Das, 2006). Remarkably, it was shown that the antifungal activity is connected with an internalization of the protein. The internalization occurred without disruption of the fungal membrane and seemed to go *via* active transport, since it is temperature dependent and inhibited by sodium azide. The authors linked this antifungal property with the ribonuclease activity of the protein (see “RNase” subsection). The authors showed that a mutant (Lys54Asn) of the protein with almost abolished RNase activity did not inhibit the growth of the fungus, despite being able to penetrate it. Apparently internalization is mandatory for the inhibition of the fungal growth since the protein does not enter *Sclerotium rolfsii*, which is resistant to AhPR-10 (Chadha and Das, 2006). Plant defense

proteins, like ribosome-inactivating proteins, have been previously shown to act after internalization into the fungal cell (Park *et al.*, 2002).

Ocatin, a PR-10 homologue (36% to 44% pairwise residue identity with PR-10/MLP family) has been also shown to have antifungal activity (Flores *et al.*, 2002). Ocatin is the most abundant soluble tuber protein from the Andean crop oca (*Oxalis tuberosa* Mol.). Flores *et al.* (2002) showed that ocatin inhibits the growth of *Phytophthora cinnamomi* (oomycetes), *Fusarium oxysporum* and *Nectria hematococcus* (ascomycetes) and *Rhizoctonia solani* (basidiomycetes). The growth of these fungi is inhibited in a dose-dependent manner. To inhibit a slow-growing fungus, *P. cinnamomi*, less ocatin is required in comparison to the fast-growing ones (*R. solani* and *F. oxysporum*) (Flores *et al., 2002*). Antifungal activity was also reported for a PR-10 protein from yellow-fruit nightshade (*Solanum surattense*) SsPR-10. Dialytically renatured SsPR10 recombinant protein inhibited the hyphal growth of *Pyricularia oryzae*, but did not inhibit the hyphal growth of *Sclerotinia sclerotiorum* and *Gibberella zeae* (Liu *et al.*, 2006). Previously, recombinant hot pepper (*Capsicum annuum*) CaPR-10 protein was reported to inhibit hyphal extension of *Phytophthora capsici* (Park *et al.*, 2004).

I.2.2.3.2. Antibacterial activity

Together with the antifungal activity of ocatin, Flores *et al.* (2002) also reported antibacterial activity of ocatin, which inhibits the growth of phytophatogenic bacteria, such as *Agrobacterium tumefaciens*, *Agrobacterium radiobacter*, *Serratia marcescens*, and *Pseudomonas aureofaciens*.

I.2.2.3.3. Antiviral activity

In 2004, Park *et al.*, demonstrated the capability of CaPR-10 to degrade viral RNA. Additionally, they showed a severe reduction of tobacco mosaic virus production in bugang leaves inoculated with recombinant CaPR-10, in contrast to infected control plants.

I.2.2.4. Storage protein function

In addition to its antimicrobial activities, ocatin also functions as a storage protein (Flores *et al.*, 2002). Storage proteins can be defined as proteins whose major role is to store nitrogen, sulfur and carbon, and which accumulate in substantial amounts under excess of nutrient supply (Shewry, 2003). They may enable plants to survive periods of adverse conditions or between growing seasons, and may also provide nutrients to support the growth of new plants as seedlings (from seeds) or shoots (from tubers).

Ocatin accounts for 40 to 60% of the total soluble oca tuber proteins. Ocatin expression seems to be highly regulated because it is tissue specific. It is confined first to the pith, and later

also to subepidermal and epidermal regions of the tuber. Its content progressively increase from 20-day-old to 3.5-month-old tubes, followed by a steady reduction to very low levels at older stages (Flores *et al.*, 2002). The tissue specificity of ocatin, its presence in high amounts in the oca tuber, and developmental regulation strongly suggest a major role as a tuber storage protein. Similar expression patterns have been reported previously for patatin and sporamin, the tuber storage proteins of potato and sweet potato, respectively (Hannapel, 1990). Patatin and sporamin are also synthesized in large amounts in tubers at precise stages of development and account for 40 to 45% and 70 to 80% of the total tuber-soluble proteins, respectively (Racussen and Foote, 1980; Maeshima *et al.*, 1985).

As stated before, ocatin acts not only as a storage protein but also shows *in vitro* antimicrobial activity (see "Antimicrobial activities" subsection). Also the major storage proteins of potato, sweet potato, yams and taro exhibit biological activities that are consistent with a role in protecting the tubers against pest, pathogens and perhaps also abiotic stress (Shewry, 2003) (see "Role in abiotic stress" subsection).

I.2.2.5. Role in abiotic stress

Some *pr-10* genes are expressed constitutively, suggesting a more general biological role of PR-10 proteins in plant development (Crowell *et al.*, 1992; Breda *et al.*, 1996; Sikorski *et al.*, 1999). The constitutive expression of a pea PR-10 gene in *Brassica napus* enhances its germination and growth under saline conditions (Srivastava *et al.*, 2004), suggesting a crucial role in plant stress tolerance. Stress responses are important aspects of environmental adaptation in plants. These responses enable plants to survive a range of environmental conditions by endowing them with a plastic developmental program (Crowell *et al.*, 1992) and suggest their importance in plant defense mechanism (Hashimoto *et al.*, 2004).

WAP18, a PR-10 protein from mulberry (*Morus bombycis* Koidz.), exhibits *in vitro* cryoprotective activity for a freeze-labile model enzyme (L-lactate dehydrogenase, LDH), suggesting that some PR-10 proteins might function in frost-tolerance mechanisms during winter (Ukaji *et al.*, 2004).

I.2.2.6. Membrane binding function

PR-10 proteins are structurally remarkably similar to the START domain of human MLN64 protein (Tsujishita and Hurley, 2000), which is a steroid binding domain related to steroidogenic acute regulary protein (StAR) involved in cholesterol (Scheme 1) translocation in human placenta and brain. The StAR protein is essential for normal adrenal and gonadal steroidogenesis in mammals, assisting in the import of cholesterol from the outer mitochondrial membrane

(OMM) to the inner mitochondrial membrane (IMM). At IMM, cholesterol is then converted to pregnenolone by the cholesterol side chain cleavage enzyme, P450scc (Stocco and Clark, 1996). The mechanism by which StAR moves cholesterol from the OMM to IMM is not completely understood. Interestingly, Bose *et al.* (1999) reported that the StAR protein undergoes structural transition to a molten globule state at low pH in solution or in association with synthetic membranes (Christensen *et al.*, 2001; Song *et al.*, 2001). The following observations: (i) that StAR bound to synthetic unilamellar vesicles has the C-terminal helix protected from limited proteolysis (Yaworsky *et al.*, 2005), (ii) that acidification of the medium promotes molten globule formation, also when only the C-terminal helix is protonated, and (iii) that disrupting the mitochondrial proton pump abolishes StAR activity; have led Baker and co-workers (2005) to suggest that the StAR protein interacts with protonated phospholipid head groups. Such interaction may promote a structural change in the protein that could lead to the binding/delivery of cholesterol molecules.

Till recently, no experimental data have been reported showing similar binding properties of PR-10 proteins. In 2007 Mogensen *et al.* provided some biophysical analysis of such interaction. They reported that Bet v 1 interacts with synthetic phospholipid vesicles, which constitute a model of biological membranes. The authors have shown that upon binding to membranes, Bet v 1 undergoes some structural changes, manifested as an increase of helix content. They reported that, in addition to the native state, the protein can exist in the so called A- and B-states. These two states diverge from the native state by a progressive (Native → A-state → B-state) increase in α-helicity and decrease of β-sheet structure, presumably due to the formation of amphipathic helices in the lipid interfacial region. The authors showed that binding to membranes occurs when the protein is in the B-state. Evidence from limited proteolysis indicates that the C-terminal part of the protein in the B-state is exposed, while the N-terminal part of the protein is generally shielded by the interaction with the membrane. This binding orientation is different from that observed for StAR, which has been shown to interact with the membrane *via* the C-terminal helix (Yaworsky *et al.*, 2005).

Goals of this work

Good structural characterization of a protein family combined with its elusive, but apparently indispensable, biological function is an atypical problem in structural biology, more characteristic of structural genomics than of classic biological approaches. Usually the role of a protein family is well established and only later structural models are elucidated to explain the physiological function. With pathogenesis-related class-10, or PR-10, proteins the situation is different. While several crystal and solution models are known, some even in complex with ligands, the role of PR-10 proteins is still under debate and investigation. This unusual situation has prompted me to take PR-10 protein characterisation as the topic of my dissertation.

The goals of this project include structural and biochemical studies of PR-10 proteins from different sources, namely from *Lupinus luteus* (LlPR-10.2B), *Pinus monticola* (PmPR-10.3.1), *Theobroma cacao* (TcPR-10) and *Hypericum perforatum* (HpHyp-1). The main experimental method applied has been protein X-ray crystallography. The structural studies have been complemented with ligand binding and antifungal assays.

Part II

Methods

II.1. DNA cloning

cDNA encoding a given protein was amplified by PCR (Polymerase Chain Reaction). The following PCR conditions were applied: an initial denaturation step at 94°C for 5 minutes, followed by 30 cycles of 1 minute at 94°C, 1 minute at 64°C and 1 minute at 72°C for denaturation, primer annealing and extension, respectively. After the cycles, a 10-minute elongation run at 72°C was performed. Gene-specific forward and reverse primers were used.

The PCR product was loaded onto a 1.0% agarose gel, subjected to electrophoresis, extracted from the gel, purified and sub-cloned in a pGEM plasmid (Promega). The plasmid was then used to transform *Escherichia coli* (DH5α) cells. After cultivation, the bacteria were transferred onto plates, and the colonies with the correct insert were selected using the galactosidase test (blue colonies in the presence of X-Gal).

After the plasmid was isolated and digested by two restriction enzymes (BamHI and NdeI), the digestion product was cloned into a similarly digested pET-3a vector. The cloned plasmid was then used to transform *E. coli* (BL21) cells. Insertion of the correct cDNA sequence was confirmed by PCR, and glycerol stocks were prepared.

II.2. Protein production and purification

Recombinant proteins were expressed and purified from *E. coli* cells using modified variants of an earlier protocol (Sikorski, 1997). The cloned expression vectors (pET-3a, pET-29c and pET151/D) were used to transform *E. coli* strains BL21 (DE3) pLysS and BL21 Star (DE3). The BL21 (DE3) pLysS strain is a high-stringency expression host, deficient in *lon* and *ompT* proteases. The BL21 Star (DE3) strain is an expression host with reduced mRNA degradation, achieved by a mutation in the RNase E gene (*rne*131), which is one of the major sources of mRNA degradation. The DE3 designation means that these strains contain the lambda DE3 lysogen, which carries the gene for T7 RNA polymerase under the control of the *lac*UV5 promoter (Studier *et al.*, 1990).

The transformed bacterial cells were grown at 30 or 37ºC in LB (Luria Bertani) culture supplemented with appropriate antibiotics. The protein expression was induced by IPTG (isopropyl thio-β-D-galactoside). After 4 or 5 hours of growth, the cells were harvested by centrifugation at 4300g and 4ºC for 20 minutes, and the pellet was stored at -20ºC.

Harvested cells were resuspended in lysis buffer (50 mM Tris-HCl, pH 8.0 with 10 mM EDTA) and lysed by sonication (Ultrasonic Processor, Model CV33, Fisher; 5 x 20 s bursts at 80% power). Cell debris was pelleted by centrifugation at 15555g and 4ºC for 1 hour.

When the protein of interest was found in the insoluble fraction, the inclusion bodies were treated with 7.2 M urea (Fischer, 1994) and later centrifuged at 15555g at 4°C.

The supernatant was fractionated with ammonium sulfate. The protein precipitate was resuspended in an appropriate buffer, followed by desalting dialysis.

When necessary, the preparations were clarified by centrifugation and subjected to S-carboxymethylation, followed by dialysis against 20 mM Tris buffer, pH 8.0, with 5% glycerol and 5 mM β-mercaptoethanol. The S-carboxymethylation of cysteine residues was performed by adding 10 µl of 1 M DTT per 1 mg of protein and incubating for 4 hours at room temperature. After covering an aliquot with aluminium foil, 40 µl of 1 M iodoacetic acid per 1 mg of protein were added and the reaction was kept in dark at room temperature for 30 minutes. To quench the iodoacetic acid reaction, 40 µl of 1 M DTT, per 1 mg of protein were added.

When appropriate, the proteins were passed through a DE52 cellulose column equilibrated with 20 mM Tris-HCl buffer, pH 8.5, with 5% glycerol and 5 mM β-mercaptoethanol or 20 mM sodium phosphate buffer, pH 8.0, with 5% glycerol and 5 mM β-mercaptoethanol. Fractionation was carried out by stepwise elution with NaCl.

When needed, protein preparations were clarified by centrifugation and subjected to methylation of lysine residues. The protein was concentrated to 10 mg/ml in PBS and divided into 1 ml aliquots. Each aliquot was supplemented with 20 µl of freshly prepared 1 M DMAB solution followed by the addition of 40 µl of freshly prepared 1 M formaldehyde solution. The tubes were shaken at 4ºC in dark for 2 hours. The addition of DMAB and formaldehyde and the shaking process were repeated twice. Finally, 10 µl of DMAB were added per aliquot followed by 12 hours of incubation a 4ºC in dark with shaking. The products of the reaction were submitted to size exclusion chromatography.

For purification, the proteins were submitted to FPLC separation (HiTrapQ, Superdex 75 and HiPrep S300). The purified fractions were pooled together and concentrated using Millipore Centricon 10 filters. This step was also used occasionally to change the protein buffer. If the volumes were to large, concentration under nitrogen pressure (Millipore, Ultracel Ultrafiltration Disc) was used.

Monitoring of all purification steps was done by gel electrophoresis in 15% polyacrylamide gel (PAGE) in denaturing (SDS) conditions.

II.3. Protein modification

X-ray crystallography is an excellent way of getting atomic-resolution knowledge of protein structure. However, crystallographers often struggle to obtain crystals with good diffraction properties, and the crystallization step is often a major bottleneck in protein crystallography. It is

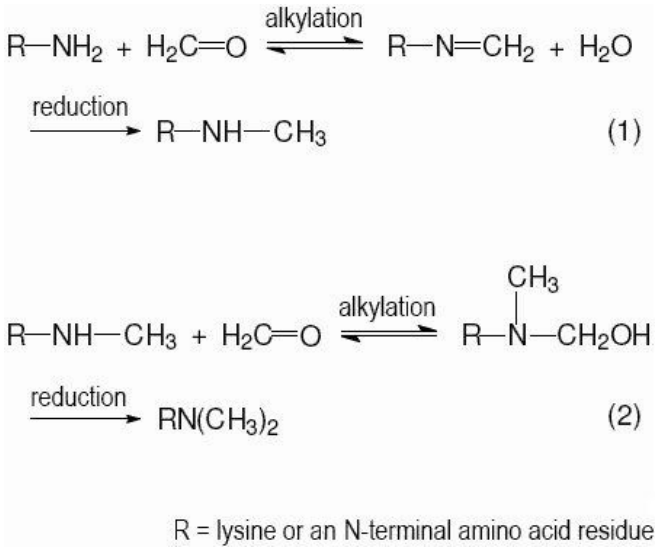
accepted that protein sample homogeneity, stability, and solubility are strongly correlated with the protein's probability to yield crystals. While this approach has proved successful in many cases, a significant number of proteins fail to crystallize despite being soluble and monodisperse in a DLS spectrum. It is, however, established that chemical modification can facilitate the crystallization of otherwise intractable proteins.

It is self-evident that molecular flexibility will tend to be detrimental for the formation of a highly ordered crystal lattice. At the other extreme, small motions, such as those due to flexible solvent-exposed amino acid side chains, can be equally disruptive to a potential well-ordered crystal lattice. The approaches in these cases can be the removal of disordered terminal extensions or selective mutation of flexible residues to smaller hydrophobic ones (Walter *et al.*, 2006). Despite the fact that the latter approach is already a well establish technique, "surface-entropy-reduction mutagenesis" (Derewenda, 2004 and references therein) remains a time- and labor-intensive procedure which requires the production of modified expression constructs. It should be also keep in mind that without previous structural knowledge, surface-entropy mutagenesis can potentially destroy key stabilizing salt bridges. On the other hand, if this type of surface engineering would be done by chemical modification, no such risk exists. Moreover, with the chemical approach, there is no need for recloning, expressing and purifying the new protein forms. Modification of an already mature protein also avoids the possibility of inducing misfolding of the nascent polypeptide chain (Rypniewski *et al.*, 1993). Unfortunately such a "shotgun" methods will indiscriminately label residues that may be critical for biological function.

II.3.1. Reductive methylation of lysine ε-amino groups

The most common approach to chemical modification is reductive methylation of free amino groups in which primary amines (i.e. lysine residues and the N terminus) are modified to tertiary amines. Originally used as a method of isotope labeling (Means and Feeney, 1968; Rice and Means, 1971), it proved to be valuable for protein crystallization, allowing in certain instances the formation of diffraction-quality crystals of proteins that had previously been refractory to crystallization (Rayment *et al.*, 1993; Kobayashi *et al.*, 1999; Kurinov *et al.*, 2000; Schubot and Waugh, 2004; Saxena *et al.*, 2006).

Reductive alkylation of lysine residues involves the initial formation of a Schiff base between the ε-amino group of a lysine residue and either a ketone or aldehyde, that is subsequently reduced to form a secondary or tertiary amine, as in Scheme 7 with formaldehyde as an example.



Scheme 7. Reductive methylation of primary $–NH_2$ groups with formaldehyde.

In the case of formaldehyde, the reaction rapidly proceeds to form dimethyllysine (Scheme 7), since the pK_a of the monomethyllysine residue is lower than that of lysine itself (Rypniewski *et al.*, 1993). The N-terminal residue of the protein backbone will be methylated as well. Several reducing agents have been successfully used to complete the methylation of lysines: sodium cyanoborohydride, sodium borohydride, and three different amine-borane complexes, namely a pyridine-borane complex, a dimethylamine-borane complex and a triethylamine-borane complex. Comparing the effectiveness of the reduction reaction, the production of homogeneous population of molecules and the possibility to work at near physiological pH, the dimethylamine-borane (DMAB) complex turned out to be the most effective compound (Rypniewski *et al.*, 1993). The protein to be methylated should be pure, homogenous and in a buffer free of primary amines. Phosphate buffered saline (PBS) is a good choice for this step. After the reaction, the protein should be separated from the reaction mixture by extensive dialysis, size exclusion chromatography, microfiltration, or precipitation. Size exclusion chromatography using Tris-based buffers helps to quench and remove excess chemicals (Walter *et al.*, 2006). In order to prevent any modification of cysteine (S-S bridges) or methionine residues, it is possible to add DTT (up to a total concentration of 1-5 mM).

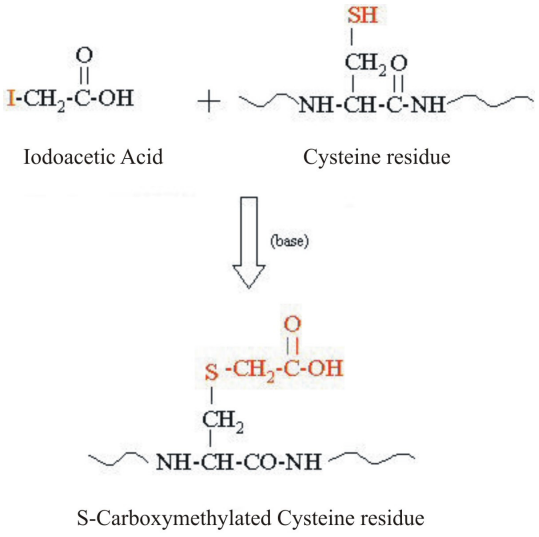
Laboratory procedure:

The protein solution was concentrated up to 10 mg/ml in PBS and divided into 1 ml aliquots. To each aliquot, first 20 µl of freshly prepared 1 M DMAB solution, and then 40 µl of freshly prepared 1 M formaldehyde solution were added. The tubes were shaken at 4ºC in dark for 2 hours. The addition of the DMAB and formaldehyde solutions and

the shaking procedures were repeated twice. Finally, 10 µl of DMAB were added per aliquot, followed by 12 hour incubation at 4ºC in dark while shaking. The products of the reaction were submitted to size exclusion chromatography.

II.3.2. S-carboxymethylation of cysteine residues

In addition to molecular flexibility and large or small motions that can hamper a crystallographic experiment, also free cysteine residues that have not been properly oxidized will often lead to difficulties. Cysteine residues in proteins are reactive and are readily oxidized to form disulfide bonds with other cysteine residues. When located on protein surface, such residues will result in intermolecular S-S bond formation and protein aggregation. One way of preventing this situation is to cap cysteine residues by means of reduction and S-carboxymethylation. S-carboxymethylation can be carried out using an alkylating agent, such as iodoacetic acid (Scheme 8) (Crestfield *et al.*, 1963).



Scheme 8. S-carboxymethylation of cysteine.

The reaction starts with reduction of all cysteine residues by the addition of DTT. The reduced protein is then incubated with iodoacetic acid, followed by addition of DTT to quench the reaction. The modified protein is later subjected to dialysis, spin columns, or fast reversed phase separation. Some proteins may require a denaturant to expose buried disulfides. In the S-carboxymethylation method, urea (1-8 M) or guanidine-HCl (1-6 M) can be used. However, if urea is used, care must be taken to avoid heating the reaction, since this can promote carbamylation.

Laboratory procedure:

S-carboxymethylation of cysteine residues was performed by adding 10 µl of 1 M DTT per 1 mg of protein and incubating the solution for 4 hours at room temperature, followed by the addition of 40 µl of 1 M iodoacetic acid per 1 mg of protein. The solution was left for 30 minutes at room temperature in dark. To quench the reaction, 40 µl of 1 M DTT per 1 mg of protein were added.

II.4. Thermofluor

Structural studies require relatively large amounts of protein. While enormous advances in molecular biology have allowed the production of proteins at high yields, the traditional techniques for checking the quality of protein preparations are not suitable for the high-throughput era. The major inconveniences of the traditional methods are the amount of protein needed and the long time required for the assays. Additionally, in the post-genomic era it is not always possible to assess the protein quality using a functional assay because often the function is not known.

Monodispersity of protein solution determined using dynamic light scattering (DLS) have been shown to correlates well with the protein's ability to crystallize in many cases, however, the method has not been used extensively with large buffer and ligand screens because it requires relatively large amounts of protein (Ericsson *et al.*, 2006). To overcome this, Ericsson and coworkers (2006) successfully applied a fluorescence-based thermal stability assay developed by Pantolano and coworkers (US patent 6,020,141) as a high-throughput screen for buffer optimization and ligand-induced stabilization of proteins. The method takes advantage of the fact that thermally-induced unfolding of proteins is an irreversible process, following a typical two-state model with a sharp transition between the folded and unfolded states. T_m is defined as the midpoint temperature of the protein-unfolding transition, and represents the temperature where the free energy of the native and non-native forms is equivalent. The folded and unfolded proteins can be distinguished through exposure to a hydrophobic fluoroprobe. The probe is quenched in aqueous solutions but will preferentially bind to the exposed hydrophobic interior of an unfolding protein, leading to a sharp decrease in quenching, so that a readily detected fluorescence emission can be studied as a function of temperature. For the experiments described in this work the fluoroprobe used was SYPRO Orange (patented by Invitrogen).

A more stable protein generally has a lower tendency to aggregate. However, it is likely that the correlation between intrinsic protein stability, on one hand, and intrinsic protein solubility, on the other, is not going to be absolute. On the contrary, there are factors that are likely to affect mainly solubility, such as ionic shield effects, hydrophobic surface effects, and

exposure of sulfhydryl groups, and these effects might not be observed in the thermal melting curves. Nevertheless, it was shown (Ericsson *et al.*, 2006) that the thermofluor method can be used as a fast means for checking the most stabilizing buffer for a protein. Additionally, the authors found a direct correspondence between the presence of additives that induce a positive shift in the transition midpoint in the thermofluor screen and increased likelihood of protein crystallization. They also found that, it is not so much an intrinsically high T_m, but rather the positive influence of some buffers/additives, that is related with a higher probability of obtaining crystals.

Melting temperatures obtained with the thermofluor method for several proteins have been shown to correlate well with values determined by other biophysical methods, such as circular dichroism (CD), turbidity measurements, and differential scanning calorimetry (DSC) (Ericsson *et al.*, 2006). Furthermore good agreement was also obtained for binding affinities estimated by the thermofluor method and by isothermal titration calorimetry (ITC) (Lo *et al.*, 2004; Matulis *et al.*, 2005). These features make the method well suited for high-throughput biophysical characterization, and for efficient search of optimal crystallization buffers and ligands. The method is general in the sense that no prior knowledge of the protein is required to screen for stabilizing conditions, it is fast, and requires only small amounts of protein.

Laboratory procedure:

The tests were carried out in a RT-PCR machine (RotorGene3000 from Corbett Research, Australia), with an integrated charge-coupled device (CCD) camera. The proteins were concentrated in stabilizing buffer (3 mM sodium citrate, pH 6.3), filtered (0.1 or 0.22 μm filters), and diluted just before the measurements. Solutions of 7.5 μl of 300x SYPRO Orange (Invitrogen), 12.5 μl of 2x test compound (different buffers), and 5 μl of 2.5 mg/ml protein were added to PCR tubes. The assays were performed with a temperature ramp from 32 to 99ºC with steps of 0.2ºC. The tubes were keep at 32ºC for 1 minute, and for 1 second at each step afterwards. The wavelengths for excitation and emission were 470 and 555 nm, respectively.

II.5. Circular dichroism

By the end of the 1980's, the number of structures in the Protein Data Bank (PDB) (Berman *et al.*, 2000) began to increase in an unprecedented manner. Despite the unquestionable value of these structures, structural biologists have realized that sometimes structural studies are needed under the conditions in which the proteins actually operate, as well as under other conditions.

Circular dichroism (CD) has become increasingly recognised as a valuable structural technique for addressing these issues (Kelly and Price, 2000; Kelly *et al.*, 2005).

CD is a form of spectroscopy based on the difference in absorption of the left- (L) and right- (R) circularly polarized light. This effect will occur when a chromophore is chiral (optically active) for one of the following reasons: (i) intrinsically by reason of its structure, or (ii) by being covalently linked to a chiral centre, or (iii) by being placed in an asymmetric environment. If, after passage through the sample, the two components of circularly polarized light are not absorbed, or they are absorbed to the same extent, no polarization will be observed. However, if one of the components is absorbed by the sample more strongly than the other, the resultant radiation, i.e. the combined components, would be now elliptically polarised, i.e. the resultant polarization vector would trace an ellipse.

CD instruments, spectropolarimeters, measure the difference in absorbance between the L- and R-circularly polarized components ($\Delta A = A_L - A_R$), which is generally reported in terms of ellipticity (θ) in degrees, with $\theta = \tan^{-1}$ (b/a), where b and a are the minor and major axes of the resulting ellipse.

One advantage of CD in studying proteins is that complementary structural information can be obtained from a number of spectral regions. There are three types of chromophores of interest. (i) The peptide bond (absorption bellow 240 nm), where there is a weak but broad $n \rightarrow \pi^*$ transition centered around 220 nm and a more intense $\pi \rightarrow \pi^*$ transition around 190 nm. (ii) The aromatic amino acid side chains (absorption in the range of 260 to 320 nm), with each of the aromatic amino acids tending to have a characteristic absorption profile. Actually the shape and magnitude of the near UV CD spectrum depends on the number and type of aromatic amino acid residues, their mobility, the nature of their environment and their spatial distribution. (iii) The disulfide bonds (weak broad absorption bands centred around 260 nm), with the intensity depending on the dihedral angle of the disulfide bond, the C-S-S bond angles and the effects of neighboring groups. In addition, non-protein cofactors can absorb over a wide spectral range, with the advantage that, generally, organic cofactors such as pyridoxal-5'-phosphate, heme, flavin, etc. show little, if any, CD signal when free in solution, but produce a measurable signal when bound to a protein partner. The CD signals in the appropriate spectral region are thus excellent indicators of the integrity of the cofactor-binding site (Kelly and Price, 2000).

CD spectra can give not only information on the secondary structure composition (% helix, sheet, turns, etc.) of proteins, but can also be used to confirm the tertiary structure fingerprint (which can be used to compare, for example, wild-type and mutant forms of proteins), the integrity of cofactor binding sites, to draw conclusions about the overall structural features of proteins and can provide a valuable check of conformational changes in proteins and

of protein folding in general, by measuring the rates of acquisition of secondary and tertiary structures. For example, CD is an excellent technique for determining the thermodynamics and kinetics of protein folding and denaturation and is unsurpassed for following the effects of mutations on protein folding and stability. In addition, it can be an excellent tool for monitoring protein-ligand interactions (Greenfield, 1999).

In contrast to X-ray crystallography and NMR spectroscopy, CD cannot give structural information at the atomic level. However, it can be a remarkable technique to give low resolution information, with the advantages of smaller amount of sample and time. On the other hand, special care must be taken to remove traces of imidazole, NaCl, DTT and EDTA, from the samples, since these compounds show high absorbance in the far UV region. Additionally, not all buffers can be used in the experiment because of the same reason. Although CD can be used to study proteins over a wide range of concentrations, care must be taken so that the total absorbance of the sample is within reasonable limits (typically below 1.0) in order to avoid excessive noise. To meet the absorbance criterion, one can adjust either the concentration of the protein, the cell pathlength, or both. Both, rectangular and cylindrical cells are available, with pathlengths ranging from 0.01-1 cm. Buffers for CD spectroscopy should not contain any materials that are optically active and should be as transparent as possible. Additionally, molecular oxygen absorbs light below 200 nm, and for optimum transparency, buffers should be prepared with glass-distilled water or the water should be degassed before use (Greenfield, 2006).

Laboratory procedure:

For the CD spectra of PR-10 proteins, 10 mM sodium phosphate buffer, pH 7.0, was used. The CD experiments were carried out by using a 1 mm path-length quartz cuvette in a JASCO J-715 Spectropolarimeter equipped with a JASCO PFD-350S temperature controlled cell holder. The proteins were concentrated in stabilizing buffers (3 mM citrate pH 6.3, 10 mM Tris pH 6.8 or 10 mM Tris pH 8.5), filtrated (0.1 or 0.22 μm filters) and diluted just before the measurements. The measurements were carried out at 20ºC.

II.6. Isothermal titration calorimetry

A fundamental principle of all biological processes is molecular organization and recognition. Biological macromolecules are able to interact with various small and large molecules with a fascinating high degree of specificity and affinity. A requirement for the understanding on the molecular level of such protein-ligand interaction is an accurate characterization and

quantification of the energies involved in these interactions. A direct method available to measure the heat change during complex formation, at constant temperature, is isothermal titration calorimetry (ITC). A quantitative description of the forces that govern molecular associations requires determination of changes of all thermodynamic parameters, including free energy of binding (ΔG), enthalpy (ΔH), and entropy (ΔS) of binding and the heat capacity change (ΔCp) (Perozzo *et al.*, 2004).

ITC measures the heat produced during molecular association. The direct thermodynamic observable is the heat associated with a binding event, i.e. a ligand is titrated into a solution containing the macromolecule of interest and the heat evolved or absorbed is detected. It allows the simultaneous determination of the equilibrium binding constant (K_B) and thus the standard Gibbs free energy change (ΔG), the enthalpy change (ΔH), the entropy change (ΔS), as well as the stoichiometry (n) of the association event. Moreover, experiments performed at different temperatures yield the heat capacity change (ΔCp) of the binding reaction. Care must be taken to obtain reliable data. For example, it is crucial that the ligand and macromolecule solutions are pure and exactly match with respect to pH, buffer capacity, and salt concentration. Preferably, the interacting components should be dissolved in the same buffer.

Laboratory procedure:

The titration curves were measured on a N-ITC III instrument from Calorimetry Science Corp. The experiments were performed using 750 µl or 1 ml of protein sample in the cell and 250 µl of ligand solution. Before the experiment, the proteins were extensively dialyzed against degassed final buffer at 4ºC. The protein concentrations were estimated using the molar extinction coefficient at 280 nm. The ligands were dissolved in the buffer used to dialyze the proteins. The experiments were performed at 20 or 25ºC. The ITC data were analyzed with the Origin software (MicroCal). The heats of mixing were subtracted.

II.7. Antifungal assay

Antifungal activity of the PR-10 proteins was determined by a radial growth inhibition assay adapted from the method of Schlumbaum *et al.* (1986). Conidial suspensions (50 µl, $2x10^6$ spores per ml) were placed in the center of potato dextrose agar (PDA) plates, and sterilized paper discs were placed around them. Subsequently, various amounts of filter sterilized proteins, ranging from 5 to 50 µg, were pipetted onto the discs. Proteins and the solubilization buffers were sterilized using Ultrafree-MC Sterile filters (0.22 µm GV Durapore; Millipore). As control, buffers and protein after thermal treatment (5 minutes at 100 ºC) were used. The plates were

incubated in dark at 28ºC and checked at 24 hours intervals. Antifungal activity is normally observed as a crescent-shaped zone of inhibition at the mycelial front.

II.8. Crystallization

Structural knowledge is an indispensable prerequisite for the understanding, at the chemical level, of the molecular mechanisms of protein function. Despite enormous technical advances in structural biology, some steps of the procedure are still tricky to overcome, especially the production of single crystals suitable for X-ray diffraction experiments. Actually, obtaining crystals that yield good diffraction data is currently the major bottleneck in structure biology. Concerning sample preparation, huge advances have been made, notably with the implementation of recombinant expression systems and the elucidation of complete genome sequences for a large number of higher and lower organisms. Also, new algorithms for structure solution associated with the increasing power of more and more accessible computers make the analysis of the diffraction data faster and more convenient. For example, the use of recombinant techniques has dramatically increased the yield of protein material, made the purification easier (by the use of affinity tags and fusion proteins), allowed protein modification (mutation and truncation) or the use of isolated domains, and allowed the labelling of proteins with Se-Met for phasing purposes. On the other hand, our understanding of the crystallization process, despite considerable progress, is still limited and different approaches are used to reach the goal. Typically, extensive screening searches lead to the identification of those variables that most significantly affect the crystal formation. These variables are the protein itself, the crystallization method, and several physical and chemical parameters. With respect to the protein, full-length and truncated versions, or versions with chemical modifications and mutations can be used. There are also several choices for setting the crystallization experiment, using vapor diffusion (in the hanging- or sitting-drop variants), dialysis, free interface diffusion, and batch crystallization under oil. Temperature, pH, type and concentration of buffer, additives, and most notably precipitants, are also important variables.

In addition to biotechnological methods, the most important factor for accelerating progress in structural biology is the development of extremely powerful X-ray sources. The only X-ray sources available in the early days of protein crystallography were sealed X-ray tubes. Unfortunately, for weakly diffracting crystals of macromolecules, their output was rather low and they were superseded by rotating-anode generators that deliver more intense X-ray beams. Nowadays the sources of the highest brilliance X-rays are synchrotron beamlines. These sources have evolved from exotic and hardly available, to the status of standard and most popular sites for collecting X-ray diffraction data from macromolecular crystals. Today, home laboratory

sources are often only used for preliminary characterization of crystal before a trip to a synchrotron (Dauter, 2006). Synchrotron radiation (SR) has had a profound impact on the field of protein crystallography. The properties of high brilliance and tunability have enable higher-resolution structure determinations, multiwavelength anomalous diffraction (MAD) experiments, study of much large macromolecular structures, the use of very small crystals, and dynamical time-resolved structural studies. Finely focused polychromatic beams can be used for ultra-fast data collection with the most advanced SR sources, where a single-bunch pulse of X-rays can be strong enough to yield a complete Laue diffraction dataset (Helliwell, 2001).

Due to the large number of variables, it is not uncommon that one cannot reproduce a crystallization experiment despite scrupulously following the protocol that had produce crystals earlier. To minimize such situations, all chemicals should be of reagent grade quality, and the protein should be at least 95% pure in overloaded Coomassie-stained SDS-polyacrylamide gel electrophoresis (SDS-PAGE). In fact, a 1 ppm contaminant in a typical 10-20 mg/ml protein solution amounts to approximately 10^9-10^{10} molecules per milliliter (Benvenuti and Mangani, 2007). Additionally, the sample should be homogeneous in terms of sequence integrity and conformation. The most common causes of micro-heterogeneities in purified samples are uncontrolled fragmentation and post-translational modification. Conformational heterogeneity may arise due to ligand binding, intrinsic flexibility of the macromolecular backbone, oxidation of cysteine or methionine residues, or partial denaturation. The difficulty of the crystallization process can be illustrated by the observation of Blow and colleagues (1994) that hemoglobin, lysozyme and xylose isomerase from a wide range of different species all crystallize easily, whereas other single-domain, monomeric, non-glycosylated, stable, and soluble proteins have never been crystallized from any species. Additionally, it is counterintuitive that myoglobins are more difficult to crystallize than hemoglobins. A critical point is to keep good track notes of all crystallization conditions, as well as of sample history. Even a negative result can be extremely important for the understanding of the protein behavior. It is not always the case that crystals will appear in the initial screens, and often the conditions have to be evaluated, selected and submitted to optimization.

Macromolecular crystals are, indeed, unique. On average they are composed of ~50% solvent, although this figure can range between 25 and 90% in particular cases (Matthews, 1985). Although solvent is the major contributor to the imperfect quality of many protein crystals, it is at the same time responsible for their value to biochemists. Because of the high solvent content, the individual macromolecules in a crystal are surrounded by hydration layers that maintain their structure virtually unchanged from that found in bulk solvent. The large size

of the solvent channels also allows free diffusion into or out of the crystal of chemical compounds, such as ions, substrates, inhibitors or other ligands.

Recent developments in robotic and nano-technology, have reduce considerably the amount of protein needed for crystallization trials, allowing for screening of a large number of condition using the same protein batch. The reproducibility of the robots is also improved in comparison with manual pipetting, especially for high-viscosity solutions, like those including PEGs.

Laboratory procedure:

For the crystallization of PR-10 proteins, the vapor diffusion method in its hanging drop variant was chosen. The drops were set using protein and well solutions, usually in a 1:1 ratio (typically 1.5 µl + 1.5 µl). The plates were stored at 19ºC (occasionally at 4 or 37ºC) and checked periodically for evaluation. The protein solutions were always filtered before use (0.45 µm filter or smaller). All other solutions were filtered as well (0.22 µm filters).

II.9. Basics of X-ray diffraction by crystals

Scattering by crystals is the result of interactions between X-rays as electromagnetic waves and the electrons of ordered molecules. The electromagnetic waves interact with the electrons present in the crystal, which oscillate with the same frequency as the incident wave. The oscillating electrons act therefore as radiation scatterer emitting radiation of the same frequency as the incident radiation, but since X-rays scattered by different electrons travel different distances, they recombine with different phases.

The diffraction spots are often called reflections, because the crystals can be viewed as being composed of numerous sets of planes that reflect the X-rays. These "mirrors" are called Bragg planes. As in optical reflection, the angle of incidence (angle at which the radiation strikes the plane of the mirror) is equal to the angle of reflection. The same is true of Bragg planes. A set of equivalent, equidistant, parallel planes with Miller indices (hkl) and interplanar distance d_{hkl} produces a reflected beam when X-rays of wavelength λ hit the planes at an angle θ, satisfying the Bragg condition:

$$2d_{hkl}\sin\theta = n\lambda$$

where *n* is an integer.

All geometrical considerations of diffraction can be rationalized using the concept of the Ewald sphere, which illustrates the Bragg's law of diffraction in three dimensions. The Ewald

construction is shown in Fig. 3. Radiation of wavelength λ is represented by a sphere of radius 1/λ centered on the diffracting crystal. The crystal is represented by its reciprocal lattice, with its origin at the point on the Ewald sphere where the direct beam emerges from it. The reciprocal lattice is related with the real lattice by the inverse relation $\mathbf{a_i} \cdot \mathbf{a_j}^* = \delta_{ij}$, where $\mathbf{a_i}$ are the real-lattice vectors, $\mathbf{a_j}^*$ are the corresponding reciprocal-lattice vectors, and δ_{ij} is the Kronecker symbol. The reciprocal lattice is spatially linked to the crystal because of the way the reciprocal lattice points are defined. In this way, when the crystal is rotated, the reciprocal lattice rotates with it. In the Ewald construction each reciprocal-lattice point lies at the end of a vector perpendicular to the corresponding set of lattice planes and with a length inversely proportional to the interplanar spacing *d*. If the reciprocal-lattice point lies on the surface of the Ewald sphere, the following trigonometric condition is fulfilled: $1/2d=(1/\lambda)\sin\theta$. This condition can be easily rearranged to take the form of Bragg's law:

$$\lambda = 2d \sin\theta$$

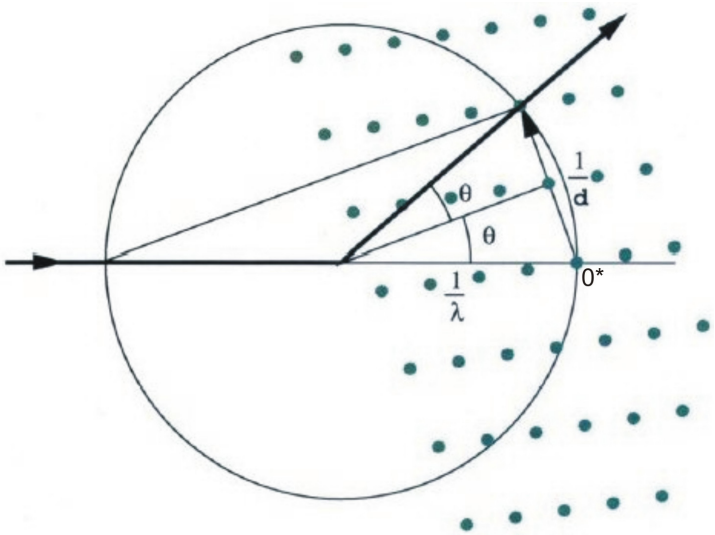


Figure 3. The Ewald construction. See text for details (Dauter, 1999).

Therefore, when a reciprocal-lattice point with indices *hkl* lies at the surface of the Ewald sphere, the interference condition for that particular set of planes is fulfilled and it gives rise to a diffracted beam directed along the line joining the sphere centre with the reciprocal-lattice point at the surface.

For any particular crystal orientation and monochromatic radiation, most of the reciprocal-lattice points will not lie on the surface of the Ewald sphere, and only a few of them can be in the diffracting position. The number depends on the density of the reciprocal lattice and hence on the unit-cell dimensions. A small-molecule crystal with short unit-cell dimensions and a sparsely populated reciprocal lattice may not give rise to any diffraction in some

orientations. Crystals of macromolecules have unit-cell dimensions much larger than the wavelength of the radiation used, and several reciprocal-lattice points (reflections) will lie on the surface of the Ewald sphere in any crystal orientation (Dauter, 1999). In order to bring all possible reciprocal-lattice points to the surface of the Ewald sphere, the crystal can be rotated, which is the most commonly used method, or a stationary crystal can be irradiated with a continuous-wavelength spectrum (white radiation) in a Laue experiment, which is less commonly used.

The intensity of a reflection with indices *hkl* is proportional to squared structure factor amplitude $|F(hkl)|^2$ where:

$$F(hkl) = \sum_{j=1}^{N} f_j \exp\left[2\pi i\left(hx_j + ky_j + lz_j\right)\right]$$

and f_j is the atomic scattering factor for X-rays for the j^{th} atom with co-ordinates (x_j,y_j,z_j) expressed as fractions of the unit cell lengths a, b, c, and N is the number of atoms in the unit cell. This is the structure factor equation in the absence of thermal motion. Due to the fact that atoms in crystal structures oscillate around equilibrium positions, the intensity of the diffracted beams is weakened. Thus, a correction factor is usually applied to the atomic scattering factor:

$$T(iso) = \exp(\frac{-B\sin^2\theta}{\lambda^2})$$

The thermal vibration is expressed as the temperature factor, or more correctly the atomic displacement parameter (B), which is proportional to the mean-square atomic displacement $<u^2>$ (which has units of $Å^2$) along a direction normal to the reflecting planes:

$$B = 8\pi^2 < u^2 >$$

II.10. The phase problem

The electrons of the ordered atoms that are present in a crystal scatter the X-ray in defined discrete directions. These different reflections are recorded as a diffraction pattern on a detector. In order to reconstruct the electron density, a Fourier transformation needs to be calculated. Here, the crystallographer faces the so-called phase problem. The phase problem results from the fact that the Fourier transformation requires two components for each reflection, its amplitude and phase. The electron density ρ(xyz) is the sum of the contributions to the point (xyz) of all waves scattered from planes (hkl), whose amplitude depends on the number of electrons in the plane, added with the correct relative phase relationship or, mathematically,

$$\rho(xyz) = 1/V\sum_{hkl} |F(hkl)| \exp(i\alpha_{hkl}) \exp[-2\pi i(hx + ky + lz)]$$

where V is the unit cell volume and α_{hkl} is the phase angle associated with the structure-factor amplitude |F(hkl)|.

On the diffraction pattern, only the intensities of the Bragg reflections are recorded. The amplitudes needed for the Fourier transformation are proportional to the square root of the intensities measured, but the phases are lost. This absence of phase information poses a serious problem to the otherwise straightforward reconstruction of the electron density $\rho(xyz)$ from the experimental structure factor amplitudes $|F(hkl)|$. To complicate it even more, the phase $\alpha(hkl)$ contains the most critical information (see the review Tayler, 2003). Without the phases the crystallographer cannot compute a picture of the molecule, so a large part of the crystallographic methodology is devoted to the solution of the phase problem. Fortunately, several methods can be used to estimate the missing phases, namely, isomorphous replacement, molecular replacement, anomalous diffraction, and direct methods.

II.10.1. Isomorphous replacement

Isomorphous replacement is the historically first method used for solving a protein structure. A variant of this method has been used in small-molecule crystallography from the mid XXth century. It was first applied to proteins by Max Perutz and John Kendrew in the 1950's. In the isomorphous replacement approach, protein crystals are soaked (or co-crystallized) in heavy-atom containing solutions (heavy in the sense that they have a large atomic number, *i.e.* many electrons) to create isomorphous heavy-atom derivatives (same unit cell, same space group, same orientation of the protein molecule in the cell), which give rise to measurable changes in reflection intensities that can be used to deduce the positions of the heavy atoms (Taylor, 2003). One reason for these intensity changes is that "heavy" atoms have a disproportionate contribution to the overall intensity. The contributions from the lighter atoms will tend to cancel out, because they scatter with different phase angles. In contrast, all of the electrons in a heavy atom will scatter essentially in phase. Because of this effect, different atoms contribute to the scattered intensity in proportion to the square of the number of electrons they contain. For example, a uranium atom contains over 15 times as many electrons as a carbon atom, so its contribution to the intensity will be equivalent to that of 235 carbon atoms. As a result, the change in intensity from the addition of 1 uranium atom to a protein of 20 kDa is easily measured.

After measuring the diffraction of the native and the derivative crystals, the differences in scattered intensities, that largely reflect the scattering contribution of the heavy atoms, can be estimated. Because of the presence of only a few heavy atoms, these differences can be used to calculate a Patterson map. Once one or more heavy atom sites have been found, approximate

protein phase angles can be estimated. This, however, leads to two possible solutions symmetrically distributed about the heavy-atom phase. As one derivative (single isomorphous replacement, SIR) leads to phase ambiguity, it is necessary to use more derivatives (multiple isomorphous replacement, MIR). Due to errors associated with the measurements of the structure factors and in the heavy-atom positions and occupancies, generally more than two derivatives need to be used (Taylor, 2003). Anomalous scattering can also be used to solve the phase ambiguity in a single isomorphous replacement experiment, leading to single isomorphous replacement with anomalous scattering (SIRAS) (see next subsection).

II.10.2. Molecular replacement

The term "molecular replacement" (MR) is generally applied to describe the use of a known molecular model to solve the unknown crystal structure of a related molecule. MR enables the solution of the crystallographic phase problem by providing initial estimates of the phases of the new structure using a model from a previously known structure. The use of MR is becoming more and more successful with the growth of the PDB, and consequently the number of possible models to use. The main advantages of this method are speed and high level of automation. The goal of MR is to orient and position the search model in the unit cell such that it coincides with the position of the unknown protein in the crystal. For that, in general three rotational and three translational parameters are needed, but the space group symmetry may limit these degrees of freedom. The requirement of six parameters per molecule leads to billions of combinations that, despite the increasing computer power, take too much time to calculate. The trick used in MR is to divide the problem, determining first the orientation (three parameters) and then the translation (also three parameters). This reduces considerably the time needed to solve the problem (Evans and McCoy, 2008).

Traditional MR methods are based on the properties of the Patterson function. This function was introduced by Arthur Lindo Patterson in 1934 and it is essentially the Fourier transform of the intensities rather than the structure factors, meaning that it can be computed without phase information. The Fourier synthesis is calculated assuming that the phases are all 0.0. This synthesis gives the distribution of the interatomic vectors in the unit cell in terms of Patterson density at each point (u,v,w) in the unit cell:

$$P(uvw) = 1/V \sum_{hkl} | F(hkl) |^2 \cos 2\pi(hu + hv + lw)$$

The unit cell of the Patterson function will contain N^2-N+1 peaks spread all over the cell, were N is the number of atoms in the unit cell. In the next MR step, a Patterson map from a trial atomic model is calculated and compared with the observed Patterson function. When the model

is oriented correctly and placed in the correct position in the unit cell, the two Patterson maps should be similar. Due to its characteristics, the success of MR to solve the phase problem strongly depends on the similarity between the model and the target structure. For a protein, at least 20% sequence identity is usually needed, but the higher this number, the better. Moreover, the use of a low homology model can lead to poor estimates of the true phase angles, leading to a strong bias towards the model, very difficult to overcome later.

The normal procedure of solving a structure by molecular replacement starts from collecting structure factor amplitudes for the unknown target structure. After that the structure factors of the known model must be calculated, which may require some optimization of the model, such as pruning different side-chains and removing flexible loops. Next, a rotation search is performed with the Patterson map of the known model over the unknown one. When a solution is found, the rotation function solution should be applied to the model coordinates to obtain a correctly oriented model. The correctly oriented model is then translated in the unit cell until the interatomic vectors generated from the model agree with those present in the experimental Patterson map. When a solution is found, the translation function should be applied to the model coordinates to obtain the correctly oriented and positioned model, followed by rigid body refinement to optimise the rotational and translational parameters.

II.10.3. Anomalous diffraction

The atomic scattering factor has three components: a normal λ-independent scattering term that is dependent on the Bragg angle, and two terms that are independent of the scattering angle, but dependent on the wavelength. These latter two terms represent the anomalous scattering that occurs at the absorption edge, when the energy of X-ray photons is sufficient to excite an electron from an inner shell of an atom in the crystal. This dispersive term (f') typically reduces the normal scattering factor, whereas the absorptive term (f'') is 90º advanced in phase, i.e. provides an imaginary phase component to atomic scattering. This leads to breakdown of Friedel's law, giving rise to anomalous differences that can be used to locate the anomalous scatterers. According to Friedel's law, in the absence of anomalous scattering, the diffraction pattern is centrosymmetric, i.e. I(h,k,l)=I(-h,-k,-l). The reflections with indices hkl and -h-k-l are know as Friedel pairs. In the absence of anomalous scattering, the structure factors of a Friedel pair have the same amplitude but opposite phase angles.

For a multiwavelength anomalous diffraction (MAD) (Hendrickson, 1991) experiment at least three datasets are needed, measured at wavelength corresponding to: (i) f' minimum (absorption edge), (ii) f'' maximum (absorption peak), and (iii) away from the absorption edge, usually at higher energy. Due to the small signal relative to the total diffraction intensity, a

multiwavelength anomalous diffraction experiment requires very careful data collection with high precision, high redundancy and good data coverage. Another drawback of the method is that it requires a tunable X-ray source, but this is nowadays easily achieved at synchrotron sources. A big advantage of the MAD method is that it uses for all the measurements only one single crystal, which completely eliminates isomorphism problems, often encountered between different crystals in the MIR method.

II.10.4. Direct methods

Direct methods use a sophisticated probability theory and the assumption of approximately equal, resolved atoms with non-negative electron density to estimate the reflection phases from the measured intensities (Sayre, 1952; Hauptman and Karle, 1953). They were first used in the field of protein crystallography by Sayre (1974), who reported the "direct-methods" structure of *C. pasteurianum* rubredoxin. Direct methods use mathematical relationships between reflection triplets (two-dimensional phase relationships) and quartets (three-dimensional phase relationships) to bootstrap a phase set for all reflections from known phases of a small or "seed" subset of reflections (Evans and McCoy, 2008). Direct methods are, however, of limited use in protein crystallography, because atomic resolution of 1.2 Å (Sheldrick, 1990) or better is needed and the number of independent non-H atoms should be less than 1000. Nevertheless, with the increasing computer power, the size of the structure is less important than the resolution of the data. A crucial breakthrough came in 1993, with the introduction of dual-space iteration algorithm, also know as "shake-and-bake" (Miller *et al.*, 1993). Its most distinctive feature is the repeated alternation of reciprocal-space phase refinement with a complementary real-space process that seeks to improve phases by applying constraints (Weeks *et al.*, 2001).

II.11. X-Ray data collection

Data collection is the last truly experimental step in X-ray crystallography. From this point all that is done are computations that can be repeated over and over again. Due to this last and sometimes unique chance, one has to be particularly cautious when collecting the diffraction data which will result in the required structural information about the molecule of interest (Dauter, 1999). Data collection is a step that requires a lot of decisions made by the crystallographer in order to get the best compromise possible. The alternatives start with the crystal choice, radiation source, rotation range and exposure time. One important point is also the decay of the crystals, which, when exposed to X-rays, suffer from radiation damage. Radiation damage is caused largely by 'primary' interactions between the molecules in the crystal and the X-ray photons. The energy is then dissipated in at least two ways. It produces

heat (thermal vibrations) that provides the energy necessary to break bonds between atoms in molecules. Moreover, the X-rays produce reactive radicals, at any thermal energy, which diffuse through the crystal causing further destruction ('secondary' damage). The use of cryo-temperatures (around 100 K) immobilizes the reactive products and prevents extensive secondary damage in areas of the crystal that are not exposed to the beam (Garman, 1999). It has been demonstrated that ionizing radiation causes expansion of the unit-cell dimensions and can induce specific structural changes (González, 2003). Nevertheless, radiation damage has been successfully used to solve the crystal structures of a protein and an oligonucleotide. The so called radiation-damage-induced phasing (RIP) technique was used (Ravelli *et al.*, 2003).

Laboratory procedure:

X-ray diffraction datasets were collected from single crystals using the oscillation method under cryogenic conditions. The crystals were flash-vitrified in a stream of cold nitrogen gas (Teng, 1990). To prevent ice formation, the crystals were equilibrated with a solution containing a cryoprotective agent. In all cases described in this work, monochromatic synchrotron radiation was used.

II.12. Data processing and scaling

The images were indexed, integrated and scaled using the DENZO and SCALEPACK programs from the HKL package (Otwinowski and Minor, 1997).

The fundamental stages of data analysis (Otwinowski and Minor, 2001) are:

- Visual inspection of the diffraction images;
- Auto-Indexing;
- Diffraction geometry refinement;
- Integration of the diffraction peaks;
- Conversion of the data to a common scale;
- Symmetry determination and merging of symmetry-related reflections;
- Statistical summary and estimation of errors.

II.13. Phasing

Molecular replacement calculations were performed using the AMoRe (Navaza, 1994), PHASER (McCoy *et al.*, 2007) and Molrep (Vagin and Teplyakov, 1997) programs distributed as part of the CCP4 package (Collaborative Computational Project, Number 4, 1994). In

addition, the evolutionary search algorithm implemented in EPMR (Kissinger *et al.*, 1999) was also used. Since AMoRe found the solution for the phase problem of the first case described in this work, it will be described in more detail.

II.13.1. AMoRe

AMoRe includes routines for a complete molecular replacement run. The program separates the rotation and translation searches, and follows them with a rigid-body refinement.

AMoRe steps:

- The observed data are extended to cover a hemisphere of reciprocal space;
- The structure factors for the model are calculated;
- The rotation function is calculated searching for Patterson correlation within a sphere centred on the origin. The correctness of the solution is checked by inspecting the correlation between the observed and model Patterson maps and the correlation coefficients and R factors between the observed and calculated structure factor amplitudes or intensities;
- The rotation solutions with the best scores are then submitted to translation function calculation;
- The solution with best correlation coefficient between the observed and calculated structure factor amplitudes is submitted to rigid-body refinement;
- If there is more than one independent molecule in the asymmetric unit, the known solution is fixed, structure factors are calculated and used for rotation and translation function as described above. The process is repeated till all molecules have been found;

II.14. Model building

Associated with reciprocal-space refinement, model building is performed in the coordinate space. This is done either automatically (ARP/wARP) (Perrakis *et al.*, 1999) or manually in such programs as Xfit (McRee, 1999), Coot (Emsley and Cowtan, 2004), or *O* (Kleywegt *et al.*, 2006).

II.14.1. Automatic builders

Model building in automatic mode was carried out with ARP/wARP (Perrakis *et al.*, 1999). This program links the automatic model building with structure refinement, and its main application

is the building of protein models starting from experimental phase angles. It is based on a unified approach to the structure solution process by combining electron density interpretation and maximum-likelihood model parameter refinement with REFMAC (Murshudov *et al.*, 1997). X-ray data to 2.7 Å resolution or higher are required although partial model building can sometimes be achieved at lower resolution. To complete the task, either the CCP4 application or the web-server versions have been used. At first, default values were used and later, when necessary, the parameters were adjusted to improve the results.

II.14.2. Manual builders

The Xfit program was used for manipulating the atomic coordinates (McRee, 1999). Xfit is part of the XtalView suite which is a complete package for solving a macromolecular structure by isomorphous replacement, including building the molecular model. One of its features is a built-in Fast Fourier Transform (FFT) capable of going in both directions from phases to maps and from models to phases, which allows flexibility in map calculations.

The Coot program can be used for model building, model completion and validation (Emsley and Cowtan, 2004). Coot displays maps and models and allows model manipulations such as idealization, real-space refinement, manual rotation/translation, rigid-body fitting, ligand search, solvent modeling, mutation, rotamer fitting, Ramachadran plots and density fit analysis.

II.15. Solvent modeling

Due to the high solvent content in macromolecular crystals, water molecules are in fact an important part of the structure. A well-ordered water molecule contributes more to the X-ray scattering than a poorly ordered part of the protein. The water molecules in the models were located automatically by the modelling programs (Xfit, Coot, ARP/wARP) or manually by inspection of electron density maps. Regardless of the way the water molecules were identified, they were all visually inspected to exclude the possibility that the density corresponds to an unexpected ligand, part of the protein (e.g. double conformation), or bound metal ions.

As a fast and systematic approach to ion search, the solvent models were submitted to the STAN webserver (STructure ANalysis server, http://xray.bmc.uu.se/cgi-bin/gerard/rama_server.pl). The algorithm screens all the water molecules in the structure and checks the possibility that some of them could be actually metal ions, such as Na^{+}, Li^{+}, Mg^{2+} or Ca^{2+} (Nayal and Di Cera, 1996). In principle, metal sites can be assigned unambiguously using anomalous differences measured at wavelengths on each side of the absorption edges of all the elements that come into question, but this is not always practicable (Muller *et al.*, 2003). To overcome this difficulty, careful inspection of the identity and architecture of the metal

coordination spheres and the bond-valence method (O'Keefe and Brese, 1991; Harding, 2001, 2004; Muller *et al.*, 2003) were used. The correct identification of metal ions in the model is not always easy but it is an important part of model building since many enzymes require their presence for optimal catalysis and also the ions can act as allosteric effectors that change the structure of the whole protein, or as cofactors binding to specific substrates (Nayal and Di Cera, 1996).

II.16. Refinement

After a structure solution has been obtained, regardless of the method used, the structure needs to be refined. Refinement is the optimization of a function of a set of observations calculated by changing the parameters of the model. The main goal of the refinement procedure is to produce new electron density maps for examination and for correcting errors in the model (manually or automatically). Maps after the refinement are typically better than those from experimental phases, and should improve as the model improves. In crystallographic refinement, three functions are commonly used: empirical energy function, least-squares residual, and maximum likelihood. The empirical energy function is based on the idea that the best model of a protein would be the one with the lowest energy. This function has two main problems, namely that till now it has not been possible to devise an empirical energy function that is accurate enough to reproduce experimental results, and secondly that there is no statistical theory underlying this function. The least-squares residual function is the simplest statistical method used in macromolecular refinement. This function calculates the sum of the squares of the differences between observed and calculated reflection intensities or amplitudes. The function is not particularly robust in cases where part(s) of the model is missing. This problem has been overcome by the use of a function that does not make the assumptions of least squares. Maximum-likelihood refinement generalizes the method and the model is adjusted to maximize the probability of given observations (Tronrud, 2004). Maximum likelihood does not assume a Gaussian error distribution and instead estimates the conditional distribution of experimental data when the model is known (Murshudov *et al.*, 1997).

The resolution of the protein native dataset defines the ratio of experimental data (reflections) to model parameters. Generally, in order to solve n unknowns, n equations are required. For example, for a relatively small protein (Mr 16757 Da, like LlPR-10.2B) with 1184 non-hydrogen atoms, a resolution between 2.5 and 2.0 Å is needed to have more data than parameters (four per atom: three x, y, z coordinates and one isotropic displacement parameter B). This is nevertheless a favorable example, because 100% completeness has been assumed and the protein is rather small. To create a better model, it is necessary to increase the data-to-

parameter ratio. Better diffraction data will increase the number of data. The number of parameters can be decreased, at least initially, by using constraints, which enforce strict relationships between the parameters. A better solution is the addition of stereochemical restraints, such as ideal (and thus expected) bond lengths, bond angles, dihedral, angles, chiral volumes or planarity, that have known values. Stereochemical restraints represent additional information that is not exact, but is subject to a probability distribution. The lengths of two chemically but not crystallographicaly equivalent bonds could be restrained to be approximately equal. A restraint is treated as an extra experimental observation with a standard uncertainty that determines its weight relative to the measured data. Stereochemical restraints for the ligand molecules and protein residues were used during the refinement of the crystal structures described in this work. For protein geometry, the updated stereochemical library compiled by Engh and Huber (2006) was used. For non-protein groups a dictionary of idealized stereochemical targets was prepared in REFMAC-5 (Murshudoc *et al.*, 1997), starting from a minimum description consisted of information about connectivity (i.e. which atom is bonded to what) and about the contents of the ligand (i.e. which element occupies which position).

At atomic or near-atomic resolution, structural models can be expanded to include anisotropic displacement parameters (Baniso) for individual atoms. The Baniso tensor describes an ellipsoid represented by six parameters per atom instead of one, as is the case with Biso. This increases the total number of model parameters from four per atom to nine per atom (three positional, plus six thermal parameters). The inclusion of anisotropic displacement parameters is only possible when the resolution is better then 1.4 Å. It yields a global improvement of the model, which can be confirmed by lower R and R_{free} residuals, and results in cleaner residual-density maps and lower uncertainties in atomic positions. Another way to improve the model, and with a much less expensive parameter expansion, is through the use of TLS (translation/libration/screw-motion) parameters. TLS models describe rigid-body vibrational motions of arbitrary objects. A single-group TLS model can be used to approximate the vibration of an entire protein molecule within a crystal lattice. At high resolution and with anisotropic vibrational displacement of each atom, TLS models are used to explain these individual atomic displacements in terms of concerted vibrations of larger groups of atoms. At more modest resolution, the TLS description allows a much simpler model to be derived and refined directly against the lower resolution data. In contrast to the extremely high increase of parameters for the anisotropic model, a single TLS group requires only 20 additional parameters (Painter and Merritt, 2005).

II.16.1. Electron density maps

Two main types of electron density maps are used in the refinement. One, "2Fo-Fc" type, shows the current best estimate of the electron density for the structure together with electron density for model errors. This is the map which should be occupied by the correct part of the model and should reveal the problematic points of the model. In case when maximum likelihood function is used for the refinement, a SigmaA-weighted map (2mFo-DFc) can be calculated, where m is the figure of merit and D is derived from σ_A. The other map type, Fo-Fc or mFo-DFc, called a difference map, is used to show the differences between the true structure and the current model. Ideally, positive density indicates were electrons should be added, while negative density shows where there is excess of electrons in the model, and a positive/negative pair indicates the way electrons should be moved towards the positive density. It is very useful during model rebuilding to display both the 2Fo-Fc map, and the positive and negative contours of the difference map. This will help not only in the rebuilding of the protein model but also in the location of water and other solvent molecules. The map contour is usually expressed in σ (r.m.s.d.) units which describe the distribution of map values around the mean value. 2Fo-Fc maps should be contoured at 1σ level, while for Fo-Fc maps $\pm$ 2.5-3σ contours are appropriate.

In an X-ray diffraction experiment, only the amplitudes are recorded and not the associated phases. Thus an essential part of the information required to generate an electron density map is derived, rather than observed. In practice, when a reasonable model is established, phases calculated from this tentative model are used to supply the missing phase information. Because these phases are derived from the existing model, such an electron density map is biased towards that model (Bhat and Cohen, 1984). SigmaA maps reduce this bias to the theoretically possible minimum (Read, 1986). Another type of maps that are relatively unbiased by the model are omit maps (Bhat and Cohen, 1984; Bhat, 1988; Hodel *et al.*, 1992). In the omit map, the part of the model to be checked is omitted and the phase calculation proceeds without this part. Since all parts of the model contribute to each reflection, by omitting successive model parts (usually 5% of the total model) it is possible to construct a full composite omit map of the model.

II.16.2. Refinement programs

For the refinement of macromolecular structures, the REFMAC program version 5.2.0019 was used (Murshudov *et al.*, 1997). REFMAC uses the maximum likelihood equations and it has been shown that this approach gives results that are consistently better than those derived using the least-squares methods (Tronrud, 2004). The refinements were carried out using the default settings, with the introduction of stereochemical restraint libraries for the ligands, and with TLS

parameters for the macromolecules (Painter and Merritt, 2005). The internet server http://skuld.bmsc.washington.edu/~tlsmd was used to establish the TLS model. The server produces 20 different possible divisions of the macromolecule into rigid groups. These descriptions were used in subsequent refinements and the one that produced best refinement statistics was used for the final refinement. For the LlPR-10.2B protein in complex with zeatin, a single TLS group was chosen. On the other hand, for the refinement of the LlPR-10.2B complex with DPU, two TLS groups, dividing the protein into an N-terminal (residues 1-86) and C-terminal parts (residues 87-154) were used. Another difference between the two complexes was the refinement of the B factors. For the zeatin complex, a full anisotropic model was used, but for the DPU complex due to the more modest resolution, isotropic B factors were used.

Occasionally, the structures were refined using the CNS program (Brünger *et al.*, 1998). This program was used because it applies simulated annealing (SA) during the refinement and also because it can be used to calculate composite omit maps. SA uses molecular dynamics to simulate the motion of the atoms at high temperatures. The algorithm is used to overcome local energy minima and, therefore, to correct larger errors (Korostelev *et al.*, 2002). SA is also a valuable tool to reduce data overfitting. For both procedures (SA and calculation of composite omit maps) the default parameters were used. During the SA refinement, the initial temperature was set to 3000 K and the temperature was decreased by 50 K in each cycle. For composite omit map calculation, 5% of the model was left out in each cycle. Due to the use of unusual ligands (zeatin, DPU), the parameter and topology files had to be created. The files were generated by the Dundee PRODRG2 web server (http://davapc1.bioch.dundee.ac.uk/prodrg) (Schuettelkopf and Aalten, 2004). A pdb file for each ligand was loaded and the default settings were chosen for the generation of the parameter and topology files.

II.17. Validation and PDB deposition

Since the process of building and refining a model of a biomacromolecule based on crystallographic data is to a certain degree subjective (Wlodawer *et al.*, 2008), quality-control techniques are required to assess its validity. This subjectivity is a result of the limited resolution and imperfect phase information that macromolecular crystallographers usually have to deal with. This means that errors in this process are almost inevitable, but it is the crystallographer's task to remove as many of them as possible prior to final analysis, deposition and publication of the results (Kleywegt, 2000).

In order to detect errors during model building and to validate the model before deposition, local and global statistics were used. For local statistics, Ramachadran plots, pep-flip test, rotamer side-chain fit test and hydrogen-bonding analysis were performed and inspected

(Kleywegt, 2000). The deviations pointed out by these independent tests were inspected by checking the local electron density and the structural context. Electron density maps, biased by the model as little as possible (composite omit maps), were used to verify the model. For global statistics analysis, the R and R_{free} values were used for monitoring the progress of the refinement, for detecting major problems with the model or diffraction data and for helping to reduce over-fitting of the data (Brünger, 1993). In addition, global quality indicators in the form of r.m.s.d. values were calculated for model parts with respect to idealized targets. The R and R_{free} are defined as:

$$R = \sum_{hkl} || F_{obs} | - | F_{calc} || / \sum_{hkl} | F_{obs} |$$

where $|F_{obs}|$ and $|F_{calc}|$ are the observed and calculated structure-factor amplitudes, respectively. R_{free} is calculated using a small, random subset of reflections (~1000) that are kept separate and are not used in the refinement. The PROCHECK program suite (Laskowski *et al.*, 1993), SFCHECK (Vaguine *et al.*, 1999) and MOLPROBITY (Davis *et al.*, 2007) were used throughout the refinement of the structure as well as for final validation and assessment of stereochemical correctness.

In order to investigate the correctness of the interpretation of the metal sites in the models, the bond-valence test (Muller *et al.*, 2003) was used. In this method, the valence (bond order) v_{ij} of a bond between two atoms *i* and *j* is assumed to be a function of the bond length d_{ij}:

$$v_{ij} = \exp[d_0 - d_{ij}]/b$$

where, d_0 is the so-called 'bond-valence parameter' that describes the expected bond length of an ideal single bond between atoms *i* and *j*. The v_{ij} values have been tabulated by Brese and O'Keeffe (1991). The factor *b* is usually taken to be a 'universal constant' equal to 0.37 Å. The bond valences of all bonds from an atom *i* sum up to the valency V_i of that atom, and for a metal cation this value is the same as the positive charge:

$$V_i = \sum_j v_{ij}$$

The protein crystal structures determined in this work, together with the corresponding structure factors, have been deposited in the Protein Data Bank (PDB), under accession codes 2qim (LlPR-10.2B/zeatin complex) and 3e85 (LlPR-10.2B/DPU complex). Prior to their deposition, the structures passed all the quality checks, as required by the PDB.

II.18. Molecular graphics and structure and sequence analysis

Molecular and electron-density illustrations were prepared using the program PyMOL (DeLano, 2002).

To calculate the volume of internal cavities in protein structures, the SURFNET program was used (Laskowski, 1995). The program VOIDOO (Kleywegt and Jones, 1994) was used to find and analyze the cavities within the protein while the FLOOD program (Kleywegt and Jones, 1994) was used to fill the cavities with water molecules for display in PyMOL.

To plot the protein-ligand interactions, the LIGPLOT program was used (Wallace *et al.*, 1995). The program automatically generates schematic diagrams of protein-ligand interactions, showing both the hydrogen bonds and hydrophobic contacts. The diagrams indicate the hydrogen bonds by dashed lines, while hydrophobic contacts are represented by an arc with spokes radiating towards the ligand atom.

Multiple sequence alignments were performed in ClustalW (Chenna *et al.*, 2003), and when residues needed to be highlighted, a text editor was used.

Part III

Results

III.1. *Lupinus luteus* PR-10.2B protein

III.1.1. Purification and characterization

III.1.1.1. Protein expression and purification

DNA coding for LlPR-10.2B was subcloned into pET-3a vector (Handschuh, 2004) and the recombinant protein was expressed in *E. coli* cells, strain BL21 (DE3) pLysS. The transformed cells were grown at 37ºC in LB culture supplemented with 100 µg/ml ampicillin and the protein expression was induced with 0.5 mM IPTG when OD_{600} reached 1.0. After 4 hours of growth at 37ºC, the cells were harvested by centrifugation at 4300g for 20 minutes at 4ºC, and the pellet was stored at -20ºC (Fig. 4). The cells were resuspended in lysis buffer (50 mM Tris-HCl, pH 8.0, with 10 mM EDTA) and lysed. Bacterial DNA was sheared by sonication (Ultrasonic Processor, Model CV33, Fisher; 5 x 20 s bursts at 80% power). Cell debris was pelleted by centrifugation at 15555g for 1 hour at 4ºC. The proteins were precipitated with 80% ammonium sulfate to separate them from non-protein cell components. The precipitate was resuspended and dialysed against 20 mM sodium phosphate buffer, pH 7.5, containing 5% glycerol and 5 mM β-mercaptoethanol. The dialysed fraction was clarified by centrifugation and passed through a DE52-cellulose column equilibrated with the same buffer. Fractionation was carried out by stepwise elution with NaCl (Fig. 4).

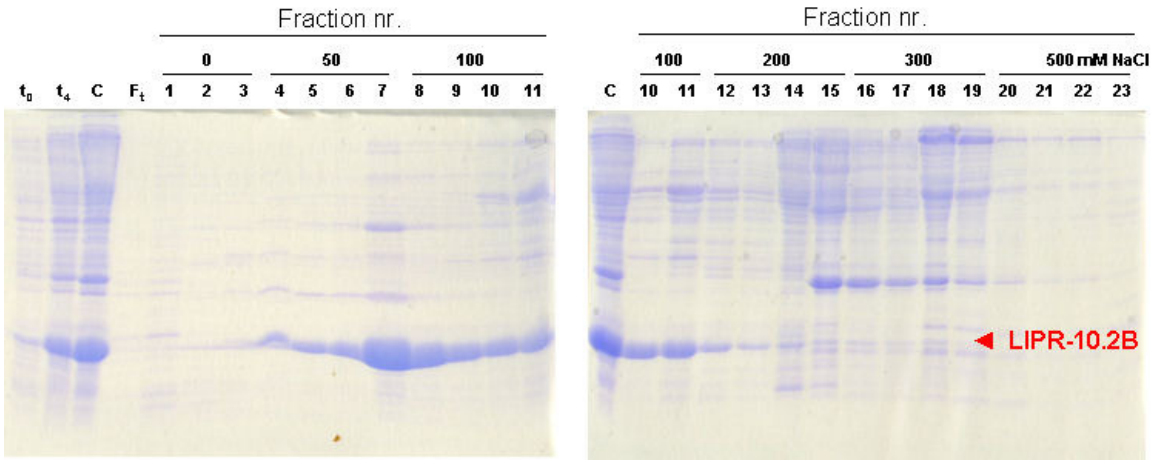


Figure 4. 15% SDS-PAGE analysis of stepwise fractionation of the recombinant LlPR-10.2B protein on DE52-cellulose in the presence of 20 mM sodium phosphate buffer pH 7.5. Lanes t_0 and t_4 represent the crude extract taken before and 4 hours after induction, respectively. C and Ft are the material loaded and the flow through of the DE52-cellulose column, respectively. The red arrow indicates the electrophoretic migration of the LlPR-10.2B protein.

The fractions eluted with 50, 100 mM NaCl and the first two 200 mM NaCl fractions were pooled together, concentrated and submitted for further purification on a HiPrep S300 (GE Healthcare) column (Fig. 5). The chromatography was performed in 20 mM sodium phosphate buffer, pH 7.5, with 5% glycerol and 5 mM β-mercaptoethanol, supplemented with 100 mM NaCl.

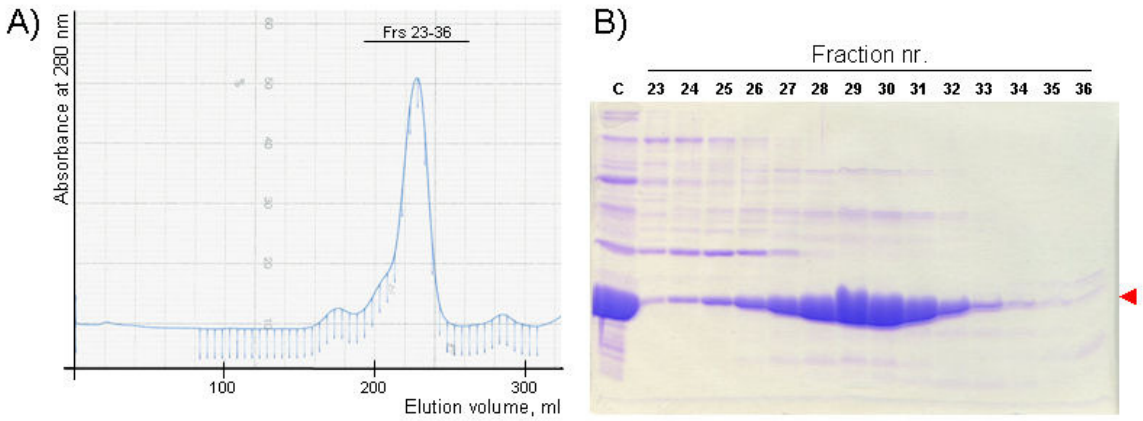


Figure 5. (A) Size exclusion chromatography on a HiPrep S300 (GE Healthcare) column of the DE52-cellulose fractions eluted with 50-200 mM NaCl, containing the LlPR-10.2B protein. (B) 15% SDS-PAGE analysis of the size-exclusion fractions. C is the material loaded on the column. The red arrow indicates the electrophoretic migration of the LlPR-10.2B protein.

The fractions 28 to 34 from the HiPrep S300 column were pooled, desalted and applied on a HiTrap Q (GE Healthcare) column. A linear NaCl gradient was used for the elution of the proteins (Fig. 6). The LlPR-10.2B protein shows a weak binding to the column and actually the protein eluted before the application of the NaCl gradient was taken for further purification.

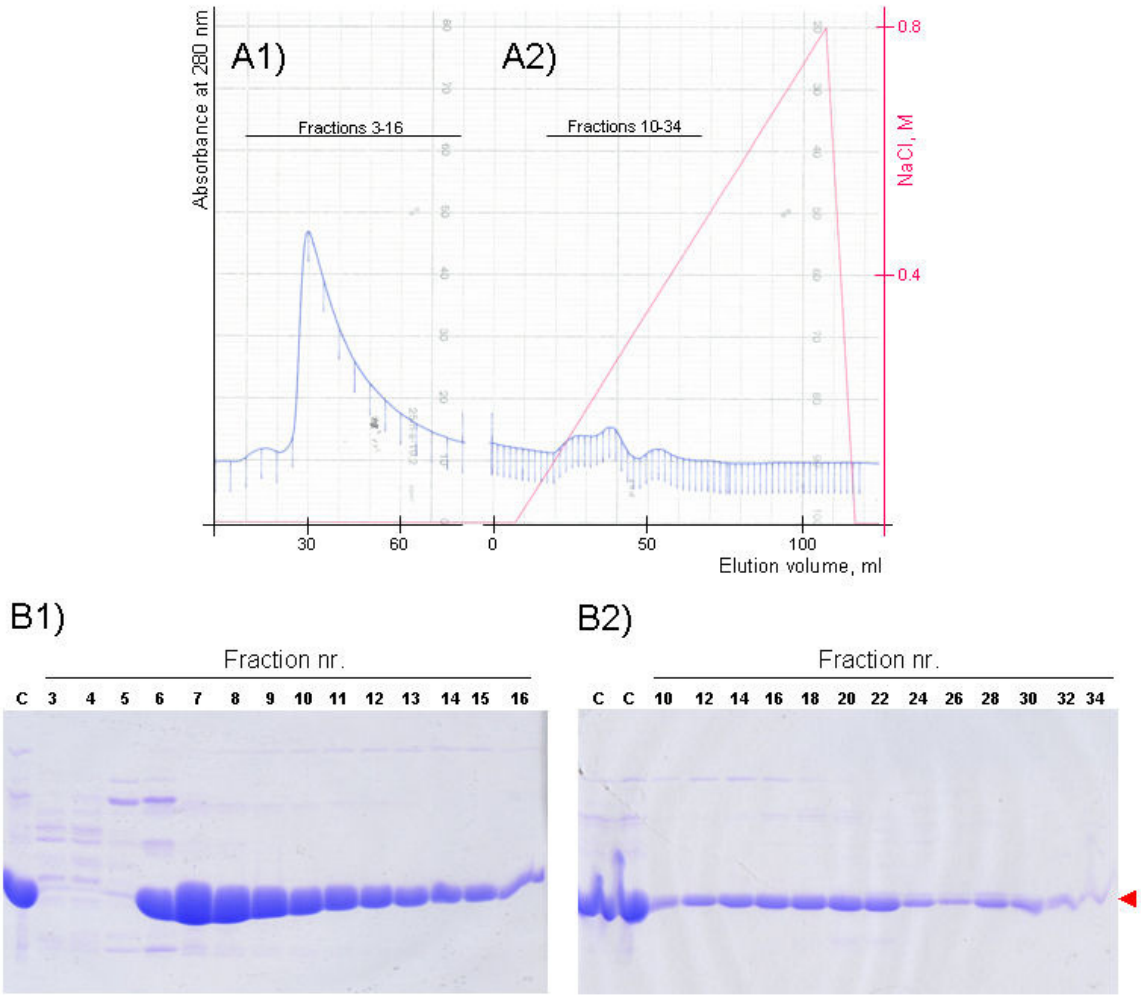


Figure 6. Ion-exchange chromatography on a 10 ml HiTrap Q (GE Healthcare) column. (A1) flow through; (A2) elution with NaCl gradient. (B1 and B2) 15% SDS-PAGE analysis of the eluted protein fractions. C is the material loaded on the column. The red arrow indicates the electrophoretic migration of the LlPR-10.2B protein.

Fractions 7 to 16 of the flow through were pooled together, concentrated and submitted for further purification on a Superdex 75 (GE Healthcare) column (Fig. 7). The chromatography was performed in 20 mM sodium phosphate buffer, pH 7.5, 5% glycerol, 5 mM β-mercaptoethanol supplemented with 50 mM NaCl. The yield of the recombinant protein after purification was about 47 mg per liter of liquid culture.

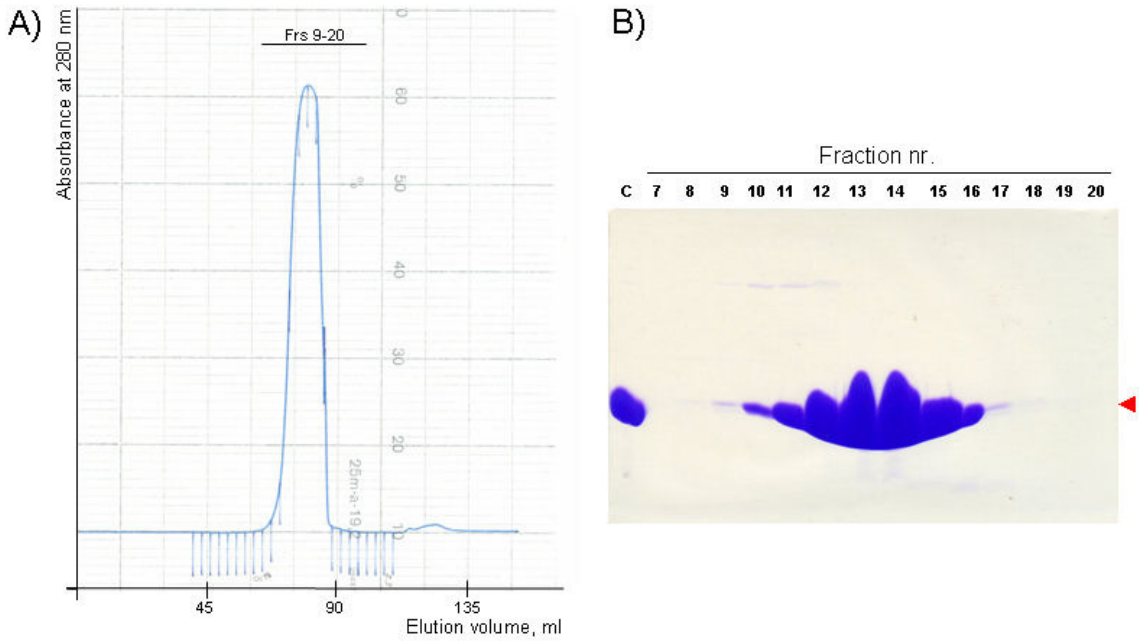


Figure 7. (A) Size exclusion chromatography on a Superdex 75 (GE Healthcare) column of the combined protein fractions eluted from a HiTrap Q (GE Healthcare) column. (B) The eluted protein fractions were analysed by 15% SDS-PAGE. C is the material loaded on the column. The red arrow indicates the electrophoretic migration of the LlPR-10.2B protein.

III.1.1.2. Circular dichroism

CD experiments were carried out at far-UV region (205-265 nm). The LlPR-10.2B protein was concentrated in 3 mM sodium citrate buffer, pH 6.3, and diluted just before measurement. The CD spectra were recorded at 3.39 µM protein concentration in 10 mM sodium phosphate buffer, pH 7.0. The scan speed was set to 100 nm/min and 4 scans were averaged. The spectra obtained at far-UV are presented without mathematical smoothing (Fig. 8).

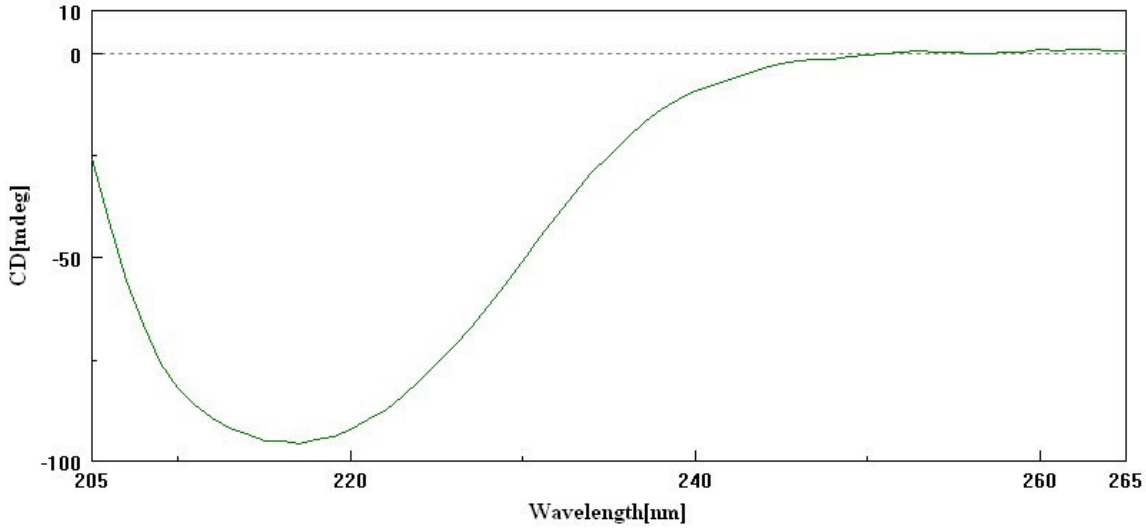


Figure 8. Unsmoothed far-UV CD spectrum of the LlPR-10.2B protein, recorded between 205 and 265 nm, at 20ºC.

III.1.1.3. Thermal stability shift assay

Thermal stability shift assay was performed using the termofluor approach. The LlPR-10.2B protein was concentrated in 3 mM sodium citrate buffer, pH 6.3, and diluted just before the measurements to approximately 2.5 mg/ml. A mixtures of 7.5 µl of 300x SYPRO Orange (Invitrogen), 12.5 µl of 2x the test compound (different buffers), and 5 µl of 2.5 mg/ml protein solution were used for the measurements. Water was added instead of the test compound to the control samples. A blanc tube with 12.5 µl 300x SYPRO Orange and 12.5 µl of water, was also added. The data are summarized in Fig. 9.

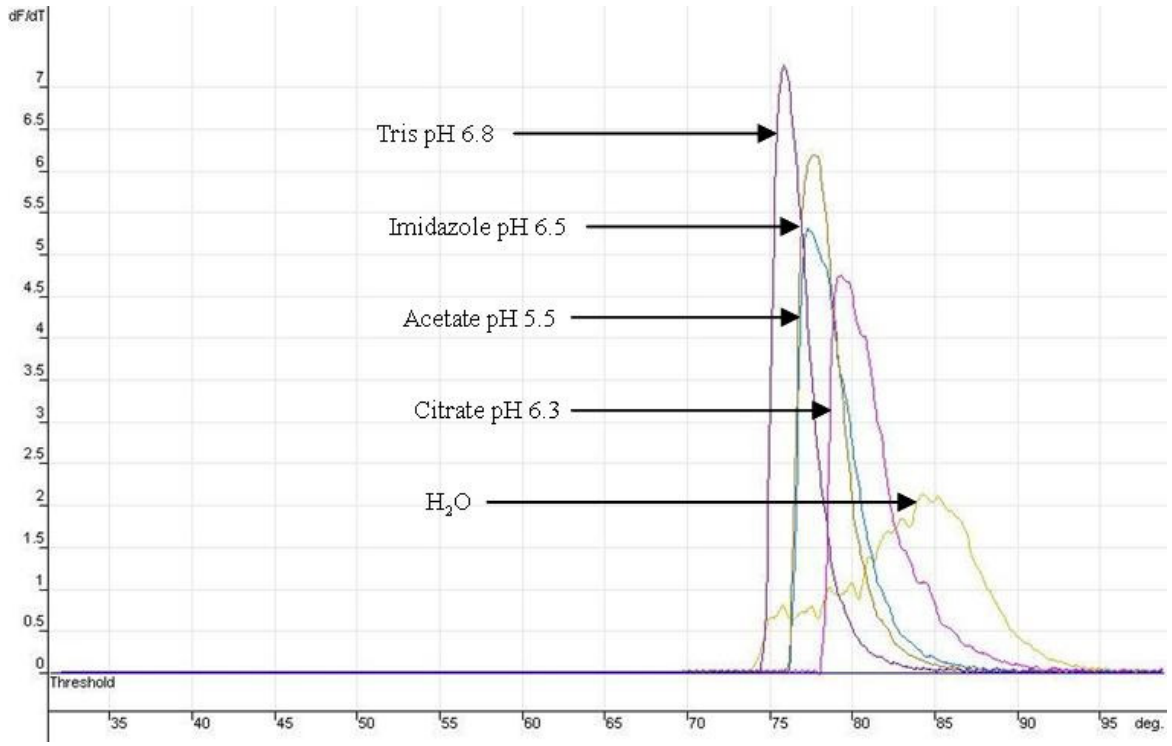


Figure 9. Thermal stability shift assays for the LlPR-10.2B protein, testing a selection of different buffers. Only a few sample runs are shown.

III.1.1.4. Isothermal titration calorimetry

Extensive attempts were made with different protein concentrations and experimental setups to assess LIPR-10.2B/cytokinin interaction characteristics. The ligand binding studies were performed using isothermal titration calorimetry as described in Methods. The data obtained are summarized in Table 2, and two runs shown in Fig. 10. Due to the low solubility of the ligands, for many ITC experiments, accurate ΔH and ΔS contributions to the free energy of binding could not be determined, therefore only stoichiometry and binding affinity were analyzed. Also, because of the low reaction heats and low values of the association constants, most of the ITC experiments were difficult to interpret and the calculated parameters have large experimental errors, and therefore should be used with caution.

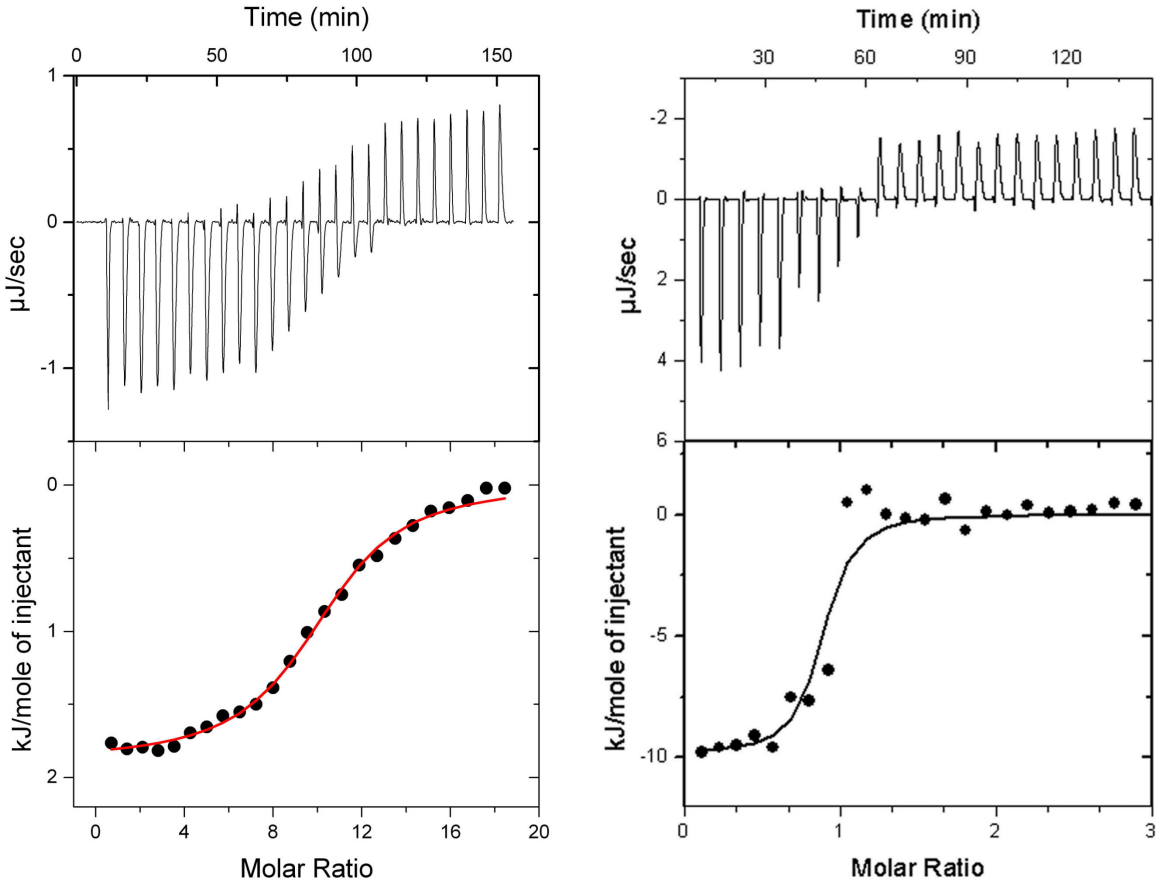


Figure 10. Calorimetric titration of LlPR-10.2B with *trans*-zeatin (Left) and DPU (Right). The top panels show raw heat data corrected for baseline drift obtained from 25 consecutive injections of 7.25 mM zeatin into the sample cell (1.0 ml) containing 0.1 mM LlPR-10.2B in 10 mM MOPS buffer, pH 7.0, at 25°C (Left) and 23 consecutive injections of 3.02 mM DPU into the sample cell (750 µl) containing 0.37 mM LlPR-10.2B in 3 mM citrate buffer, pH 6.3, at 20°C (Right). The bottom panels show the binding isotherm created by plotting the heat peak areas against the molar ratio of the ligands added to LlPR-10.2B present in the cell. The heats of mixing (dilution) were subtracted. The lines represent the best fit to a model of *n* independent sites. LlPR10.2B/zeatin binding is exothermic with 1:10 stoichiometry (n=10.2) with K_d of 36.4 ± 3.8 µM. LlPR-10.2B/DPU binding is exothermic with 1:1 stoichiometry (n=0.83) with K_d of 4.2 ± 1.6 µM.

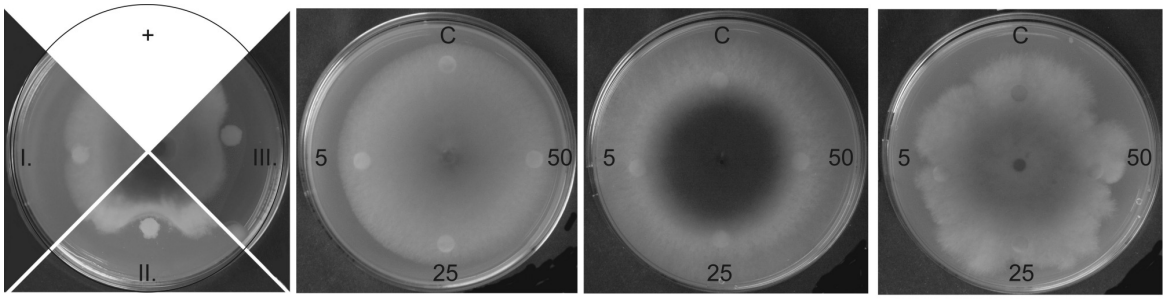
Table 2. Binding affinity of LlPR-10.2B and the studied ligands, estimated in multiple ITC runs.

Ligand	Run number	Temperature [ºC] and Buffer*	LlPR-10.2B concentration [mM]	Ligand concentration [mM]	n#	K_d [μM]
Zeatin	1	25 A	0.10	7.25	6.1	43.8 +/- 5.4
	2	25 A	0.10	7.25	10.2	36.4 +/- 3.8
	3	20 B	0.37	16.92	8.9	118 +/- 15
	4	20 B	0.37	16.92	1.3 6.0	12.8 +/- 4.4 144 +/- 25
	5	20 B	0.37	16.92	1.4 10.6	8.6 +/- 3.9 390 +/- 63
	6	20 B	0.37	13.94	2.0 6.8	15.4 +/- 7.1 428 +/- 150
Kinetin	1	20 B	0.37	0.33	0.1	0.3 +/- 0.2
	2	20 B	0.37	0.33	0.9 0.1	7.7 +/- 1.8 0.8 +/- 0.1
DPU	1	20 B	0.37	3.02	0.8	4.2 +/- 1.6
CPPU	1	20 B	0.37	23.99	8.5	18.7 +/- 4.8

* Buffer A is 10 mM MOPS pH 7.0 and buffer B is 3 mM citrate pH 6.3.
# SD values for n parameters were in each case below 10%.

III.1.1.5. Antifungal tests

The effect of the recombinant LlPR-10.2B protein on the *in vitro* growth of pathogenic fungi was tested. Purified LlPR-10.2B and controls (see methods) were added to sterile paper discs. The fungal mycelial front was allowed to grow until the edge of the plate, and growth inhibition zones were investigated. Purified LlPR-10.2B did not inhibit the growth of any of the fungi studied, namely *Colletotrichum lupini*, *Leptosphaeria maculans* and *Leptosphaeria biglobosa* (Fig. 11).



Positive control I. *Colletotrichum lupini* II. *Leptosphaeria maculans* III. *Leptosphaeria biglobosa*

Figure 11. Antifungal activity tests of recombinant LlPR-10.2B protein. Effect of filter discs containing 5, 25 or 50 μg of the protein, or buffer as a control (C), in *in vitro* growth of *Colletotrichum lupini*, *Leptosphaeria maculans* and *Leptosphaeria biglobosa*.

III.1.2. *Lupinus luteus* PR-10.2B/zeatin complex

III.1.2.1. Crystallization

The LlPR-10.2B/zeatin complex was formed by incubating 6 mg/ml protein solution in 3 mM Na-citrate buffer, pH 6.3, with *trans*-zeatin (100 mM, dissolved in 50 mM HCl) in 1:4 molar ratio. Single crystals of the complex were grown by Prof. Grzegorz Bujacz (Technical University of Lodz) by the hanging-drop vapor-diffusion method at room temperature, by equilibrating a mixture of 2 μl protein solution with 2 μl reservoir solution containing 1.2 M Na-citrate in 0.1 M MES, pH 6.5, against 0.5 ml of the reservoir solution (Fig. 12). The crystals grew within 2 weeks to dimensions suitable for X-ray diffraction analysis (0.25 x 0.25 x 1.00 mm).

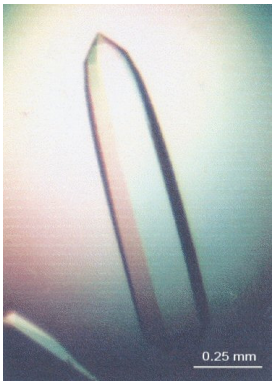


Figure 12. Single crystals of the LlPR-10.2B/zeatin complex.

III.1.2.2. X-Ray data collection

X-ray diffraction data extending to 1.35 Å resolution were collected at the EMBL X13 beamline at the DORIS ring of the DESY synchrotron. A single crystal was harvested into a cryoprotectant solution composed of the reservoir solution containing 30% (v/v) glycerol, and was flash-vitrified in a stream of N_2 gas at 100 K (Teng, 1990). The images were indexed, integrated and scaled using the HKL program package (Otwinowski and Minor, 1997), to yield a 100% complete data set characterized by an R_{int} of 0.042. The crystal belongs to space group $P6_5$ with unit cell dimensions a=73.8, c=67.2 Å (Table 3).

Table 3. Data collection statistics for the LlPR-10.2B/zeatin complex.

Space group	$P6_5$
Cell parameters [Å]	a=73.8, c=67.2
Resolution limits (last shell) [Å]	30.0-1.35 (1.40-1.35)
Radiation source	DESY, X13 (EMBL)
Wavelength [Å]	0.8030
Temperature [K]	100
No. of measured reflections	442110
No. of unique reflection	45644
Mosaicity [°]	0.36
R_{int} (last shell)	0.042 (0.670)
Completeness (last shell) [%]	100 (100)
Redundancy	9.7
$\langle I/\sigma(I)\rangle$ (last shell)	56.7 (2.5)

III.1.2.3. Structure solution, refinement and deposition

The structure was solved by molecular replacement, as implemented in AMoRe (Navaza, 1994), using diffraction data between 20.0-2.8 Å, and the crystal structure of LlPR-10.2F (R. Dolot, personal communication) as the search model. These proteins share 87% sequence identity.

The Matthews coefficient (Matthews, 1968) of 3.13 $Å^3\,Da^{-1}$ (60.6% solvent content) indicated one LlPR-10.2B molecule in the asymmetric unit, and therefore only one solution was sought. After rigid-body refinement, a solution in space group $P6_5$ was characterized by a correlation coefficient between observed and calculated structure factor amplitudes of 0.48 and an R factor of 0.50, and had a significantly higher correlation coefficient than the next solutions. Most of the residues were clearly visible in the electron density map phased by the molecular replacement model. The remaining side chains, zeatin and water molecules, as well as other

components of the solvent region (Ca^{2+}, glycerol) were added during interactive cycles of model building in XtalView (McRee, 1999), followed by structure factor refinement in REFMAC-5 (Murshudov *et al.*, 1997). When the isotropic refinement had converged, full anisotropic model was introduced resulting in a drop of R and R_{free} by 0.032 and 0.027, respectively. In the last stages of the refinement, the model was supplemented with geometrically positioned riding hydrogen atoms. The occupancies of residues in double conformation were manually adjusted based on electron density maps and temperature factors. The final model has an R-factor of 0.147 (R_{free} 0.194) and includes 157 protein residues, 233 water molecules, one glycerol molecule and a calcium ion (Table 4).

Atomic coordinates and structure factors for the LlPR-10.2B/zeatin complex have been deposited in the Protein Data Bank with the accession code 2qim.

Table 4. Refinement statistics for the LlPR-10.2B/zeatin complex.

Program used		REFMAC-5
Resolution limits [Å]		15.0-1.35
No. reflections		45644
No. reflections in test set		1163
No. of atoms	Protein	1211
	Ligand	64
	Metal	1 (Ca)
Glycerol molecules		1
Water molecules		233
R/R_{free}		0.147/0.194
⟨B⟩ [Å^2]	Proteins atoms	24.9
	Ligand atoms	21.9
	Water molecules	41.5

III.1.2.4. Model quality and overall fold

The refined 1.35 Å resolution crystallographic model of LlPR-10.2B has good overall geometry and Ramachandran statistics (Table 5 and Fig. 13).

Table 5. Stereochemical quality of the LlPR-10.2B/zeatin complex.

Rms deviation from ideal	Bond lengths [Å]	0.020
	Bond angles [°]	2.04
	Torsion angles [°]	5.8
	Chiral-center restraints [Å^3]	0.112
Ramachandran φ/ψ angles [%]	Most favored	93.9
	Additional allowed	6.1

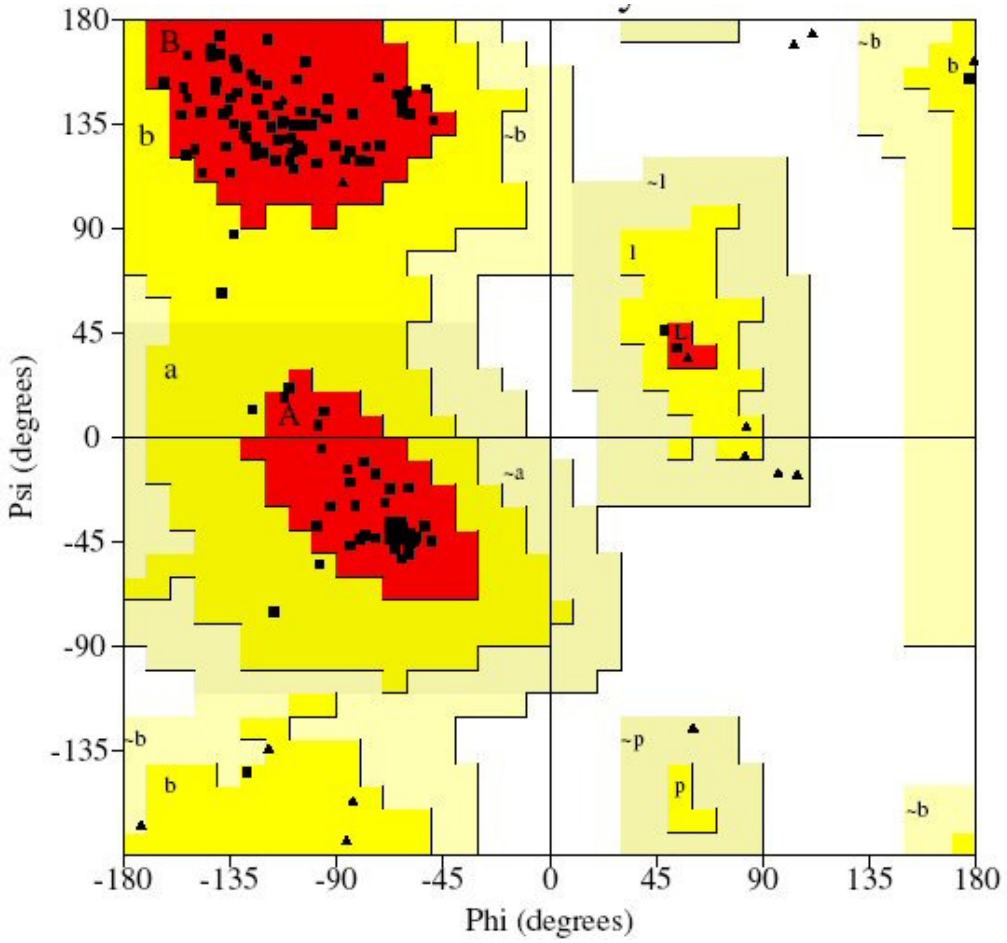


Figure 13. Ramachandran plot illustrating the φ and ψ backbone dihedral angles for the LlPR-10.2B protein in complex with zeatin. Glycine residues are identified by triangles as they are not restricted to the regions of the plot appropriate for other residue types (indicated by squares). The proline residues are not represented.

The quality of the electron-density maps is high and there are only a few less clear areas in the loop regions. The crystallographic asymmetric unit contains one LlPR-10.2B molecule with an overall fold consisting, as in other PR-10 members, of a seven-stranded antiparallel β-sheet wrapped around a long C-terminal helix α3. Two additional short helices (α1, α2), in a right-handed crossover, separate the strands β1 and β2, which form the two edges of the β-sheet.

The consecutive strands β2-β7 are connected by five hairpin loops bringing strand β7 back into hydrogen-bond contact with β1. The regularity of the β-sheet is distorted by six β-bulges, which are characteristic structural elements frequently utilized by protein folds to introduce curvature into β-sheets. The overall geometrical result is a highly curved β-sheet (Fig. 14).

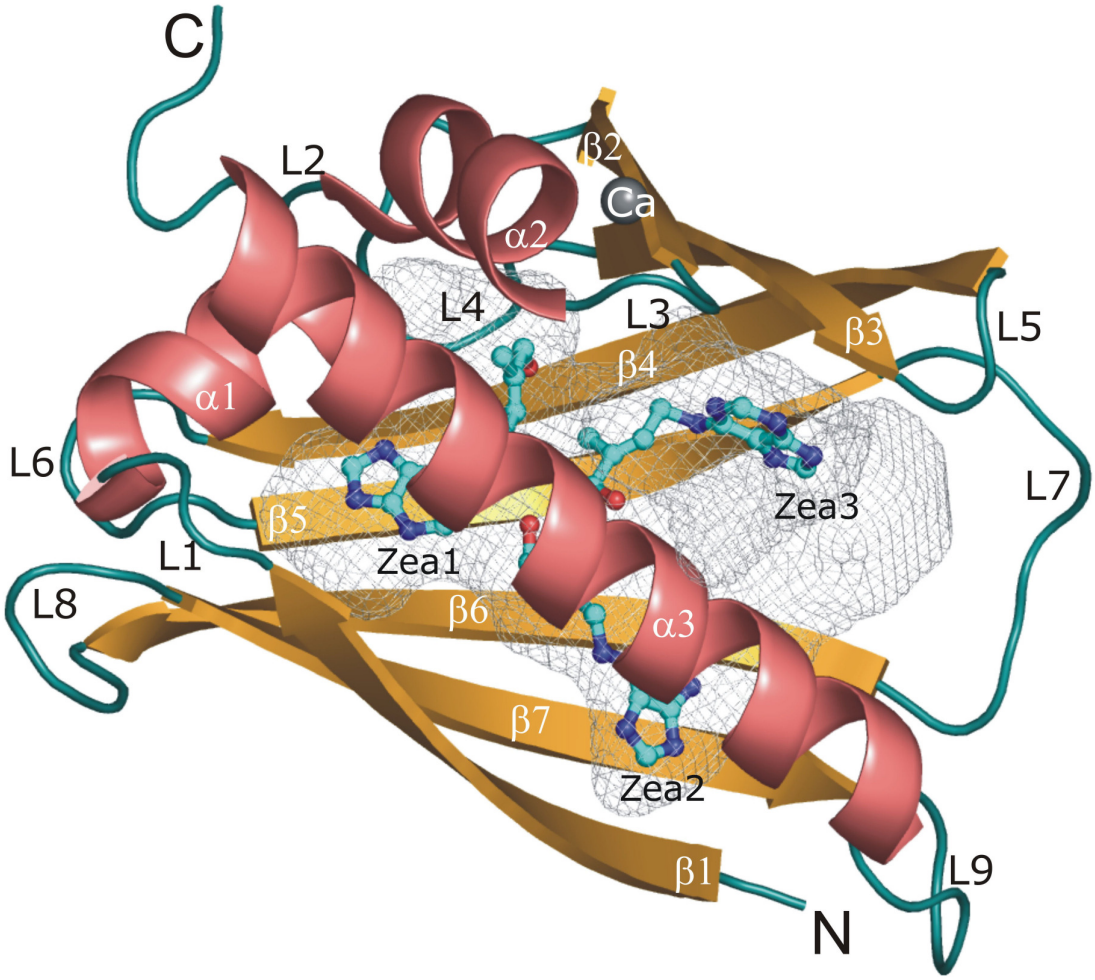


Figure 14. Overall structure of the LlPR-10.2B/zeatin complex. The three *trans*-zeatin molecules found inside the binding cavity are shown in ball-and-stick representation. The calcium cation is represented as a sphere. The orientation has been adjusted to visualize the O16 atoms of the zeatin molecules. The binding cavity is represented as a grey cast, calculated in VOIDOO (Kleywegt and Jones, 1994).

The β-bulges are structurally conserved in all PR-10 proteins, confirming their importance for maintaining the overall structure of the β-sheet. A large hydrophobic cavity is present within the protein core, formed between the residues lining the interior face of the β-sheet and the three α-helices (Fig. 14).

III.1.2.5. The glycine-rich loop

One of the loops, the glycine-rich loop (L4), shows extraordinary rigidity despite the presence of four glycine residues. It is characterized by excellent electron density (Fig. 15), low B-factors and the highest degree of PR-10 sequence conservation (Fig. 1). The sequence and structural conservation is observed even in distant homologues, such as CSBP. The rigidity of the loop is maintained by a pattern of three hydrogen bonds between the Oγ1 atom of Thr51 and the main-chain N-H groups of Asn46, Gly47 and Gly48.

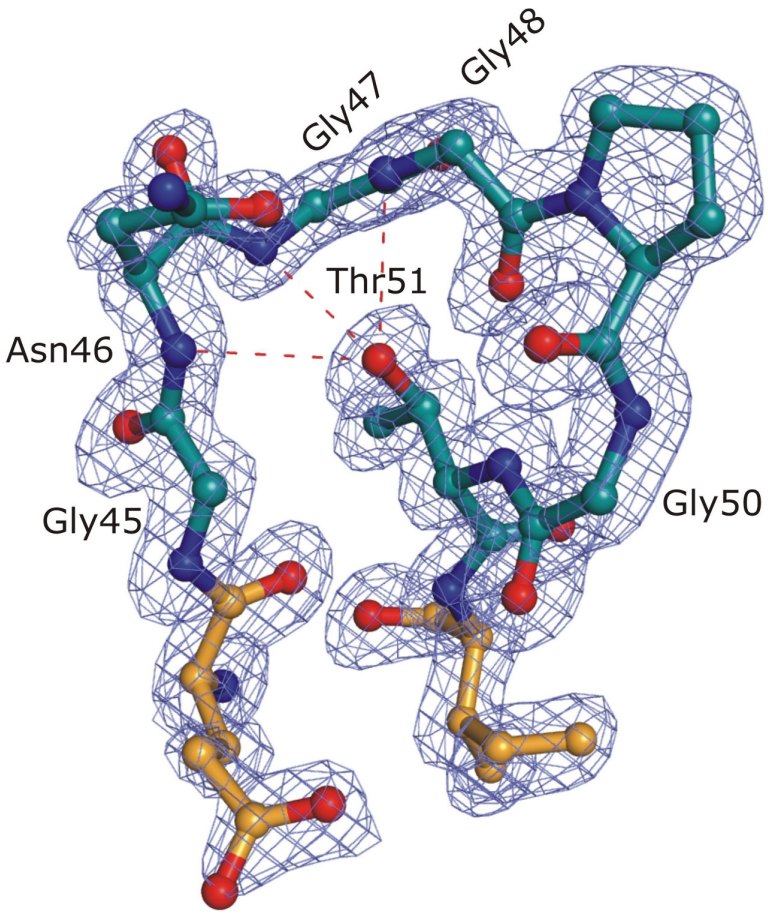


Figure 15. The glycine-rich loop (L4) with the three hydrogen bonds formed by the side chain of Thr51 and the main-chain amides of residues Asn46-Gly47-Gly48. The 2Fo-Fc electron density map is contoured at the 1.5σ level. The ball-and-stick illustrations of atomic models in the III.1.2 subsection are colored red for O atoms and blue for N atoms. C atoms display the color according to the secondary structure element they belong, as in Fig. 14.

III.1.2.6. Anisotropic displacement parameters

The anisotropy of an atom is defined by $A = E_{min}/E_{max}$, where E_{min} and E_{max} are, respectively, the minimum and maximum eigenvectors of the anisotropic displacement tensor and are connected with the lengths of the shortest and longest principal axes of the corresponding thermal ellipsoid. Analysis of the anisotropic displacement parameters was performed using the PARVATI and ANM tools (Merritt, 1999, Eyal *et al.*, 2006). Generally, both anisotropic and equivalent isotropic temperature factors (B_{eq}), increase together with the distance from the interior of the protein and are higher at the surface (Fig. 16). Residues exposed to solvent exhibit pronounced anisotropy, while surface residues engaged in crystal contacts are more isotropic and have lower B_{eq} values.

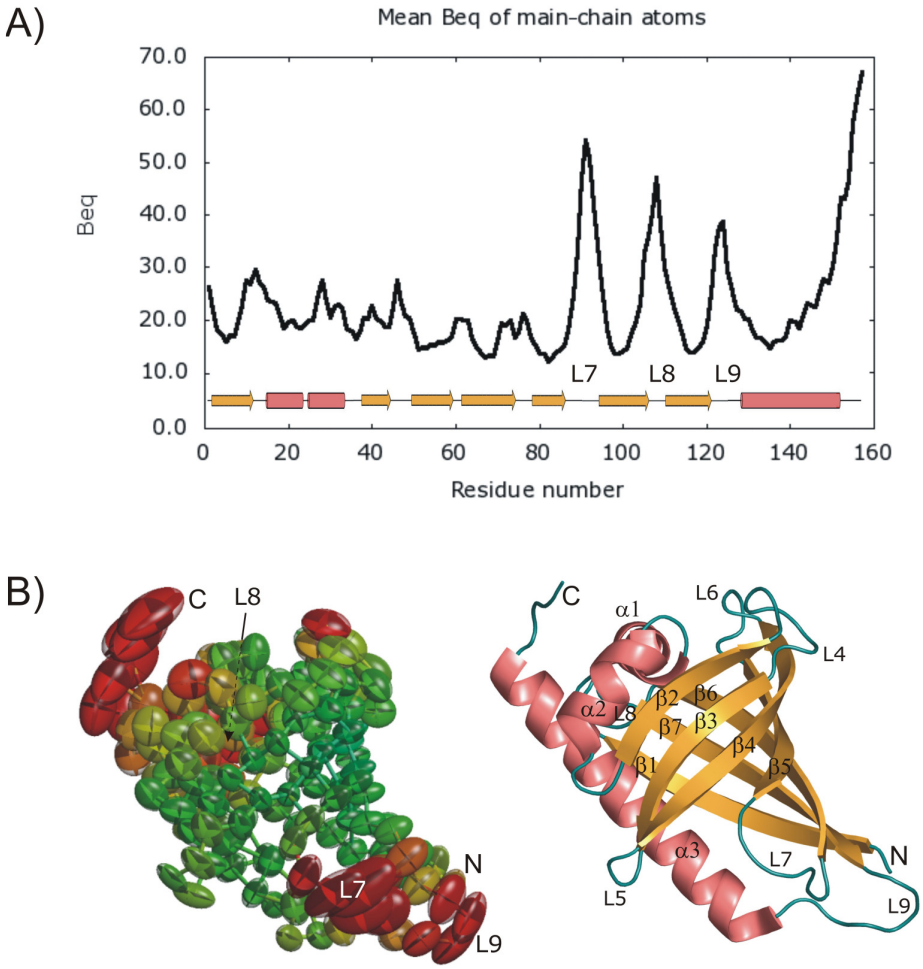


Figure 16. (A) Plot of B_{eq} of the main-chain atoms of the LlPR-10.2B molecule with the corresponding secondary structure elements. (B) The anisotropic displacement ellipsoids of the Cα atoms (left) and structure representation (right) in the same orientation. For clarity, not all secondary structure elements are labelled.

III.1.2.7. Asymmetric unit contents

Apart from the protein-zeatin complex, the asymmetric unit contains 233 modelled water molecules (three of them modelled with half occupancy), one glycerol molecule, and one calcium cation. The metal cation is octahedrally coordinated close to loop L3 by three carbonyl O atoms (Pro31, Val34, Ile37) and three water molecules (Fig. 17). The correctness of the interpretation of the metal site as calcium is confirmed by the satisfactory refinement of the B factor (29.0 $Å^2$), by the final $Ca^{2+}\cdots O$ distances (2.2-2.7 Å) and by the bond-valence test (Muller *et al.*, 2003), indicating a valency of 2.14 (expected: 2.0) for calcium against 1.36 (expected: 1.0) for sodium.

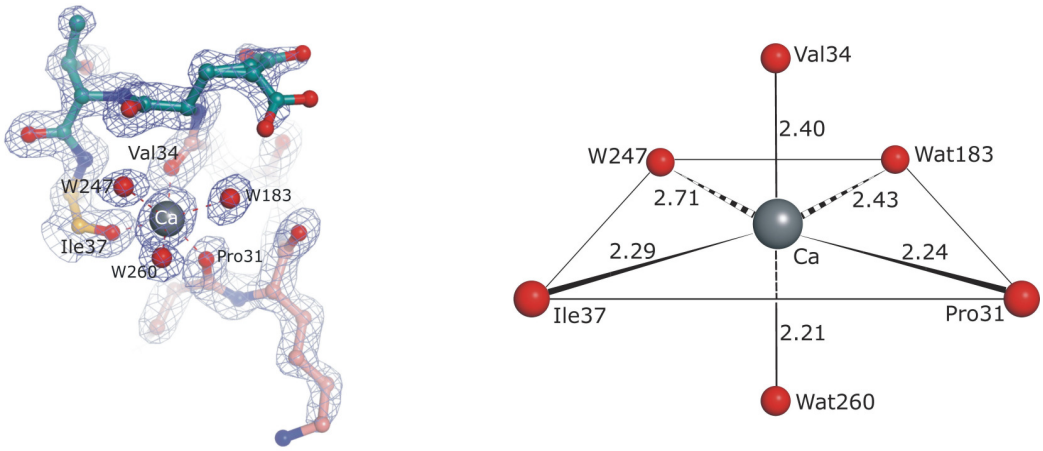


Figure 17. Coordination of the Ca^{2+} cation in the LlPR-10.2B/zeatin complex structure. On the left is shown the coordination geometry and the 2Fo-Fc electron density map (2.0 σ level) and on the right the coordination scheme with the $Ca\cdots O$ distances in Å.

III.1.2.8. The ligand binding cavity

The cavity enclosed within the LlPR-10.2B core has an unusually large volume, calculated as 4500 $Å^3$ by the SURFNET program (Laskowski, 1995). The residues shaping the cavity endow it with hydrophobic character. Nevertheless, a few polar residues are also pointing into its volume. Both the hydrophobic and polar residues are responsible for the binding of the ligand molecules. The cavity is not closed and can be accessed by a number of openings (Fig. 18). The largest, located between α3, L3, L5 and L7, is gated by a salt bridge between Arg138 and Glu59 (Fig. 18A and 18B). Close to and connected with the first opening is another one, between β5, L7 and β6 (Fig. 18C). This opening is clearly enlarged by a disruption of the antiparallel β-sheet by a water molecule, Wat195, two hydrogen bonds away from the end of the β-sheet (Fig. 19). An additional opening is found between β1 and α3 (Fig. 18D). A small opening is also present in the opposite wall of the cavity, between loops L2 and L4 (Fig. 18E).

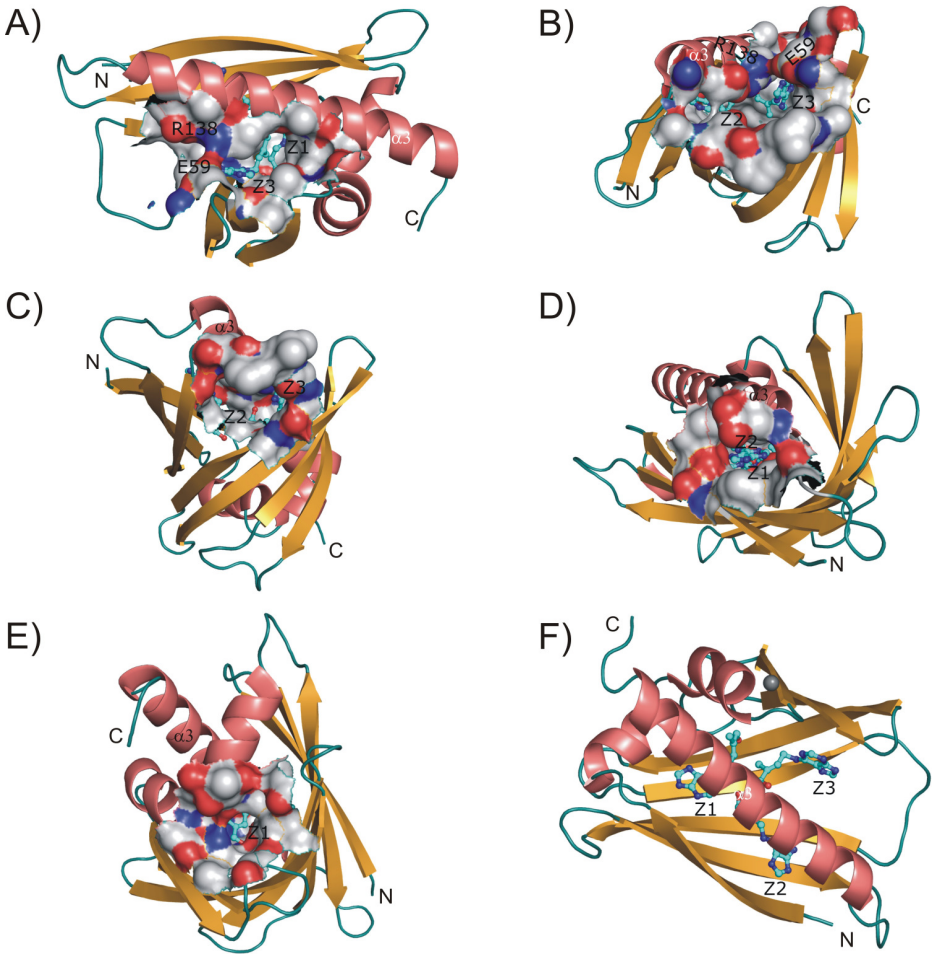


Figure 18. (A)-(E) Entrances to the LlPR-10.2B protein cavity in the protein-zeatin complex. See text for details. Panel (F) has been added for easier recognition of the orientation of the remaining panels.

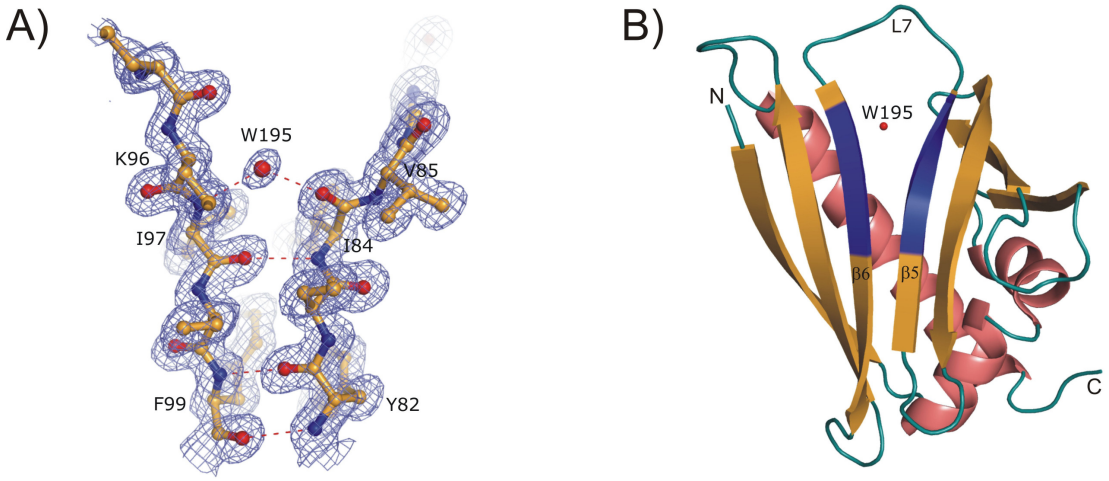


Figure 19. Disruption of the antiparallel β5-β6 β-sheet by a water molecule. (A) Ball-and-stick representation and 2Fo-Fc electron density map contoured at the 1.3σ level. (B) Cartoon representation of the protein with the regions represented in the left panel highlighted in blue.

The binding cavity is filled with three zeatin molecules and 25 water molecules (Fig. 20).

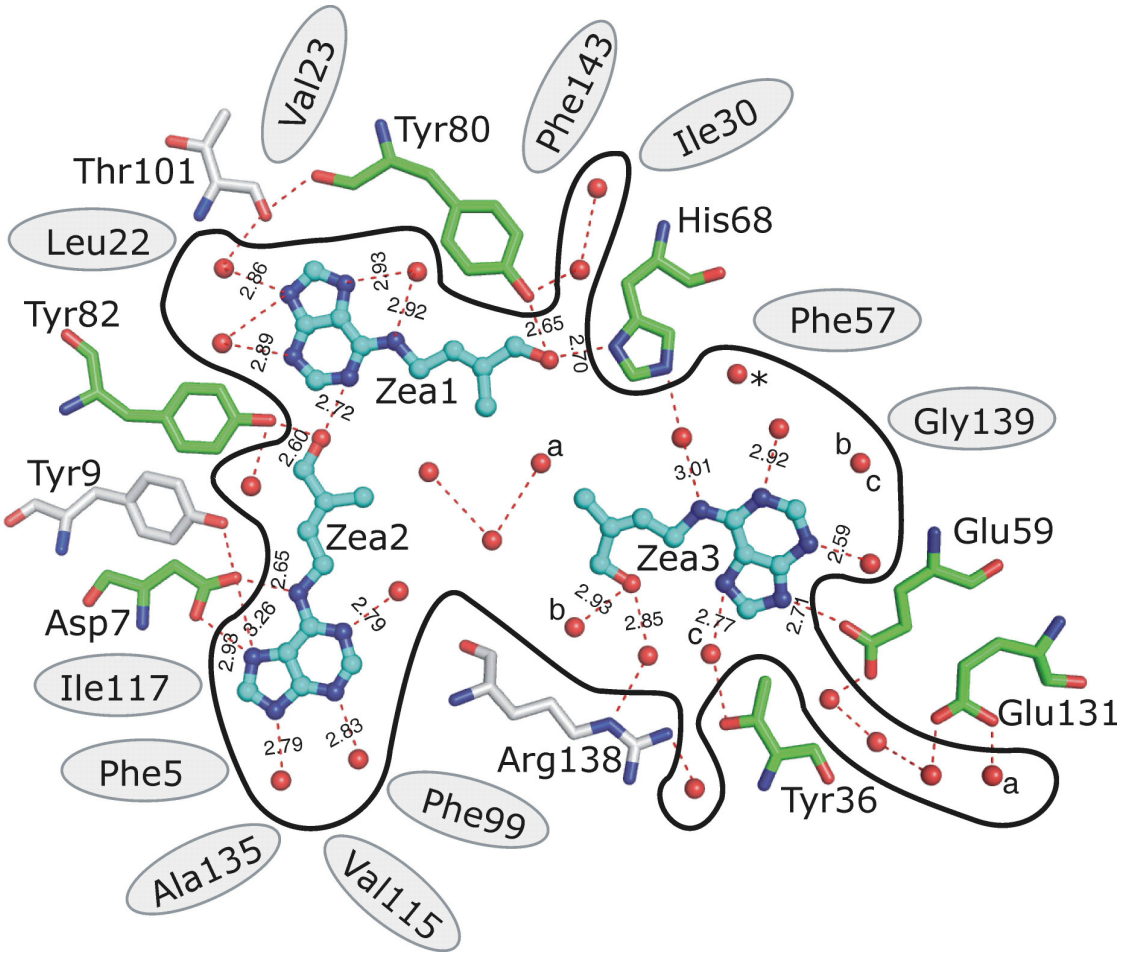


Figure 20. Representation of the three zeatin molecules found in the protein cavity (black outline) with their environment. Hydrogen bonds are represented by dash lines. The letters a-c mark water molecules that are linked by hydrogen bonds but could not be connected because of the distortion of this two-dimensional representation. The residues involved in hydrogen bonding are represented in stick mode (grey, if they participate in van der Waals interactions with the ligands). The ellipses indicate amino acid residues involved only in van der Waals contacts with the zeatin molecules. * Indicates a water molecule that has hydrogen-bonding interactions with the bulk solvent only.

III.1.2.9. Zeatin binding

The full occupancy and excellent electron density maps of the zeatin molecules (Zea1-Zea4), including the one outside of the binding cavity (Zea4), leave no doubt about the identity and conformation of the ligands (Fig. 21).

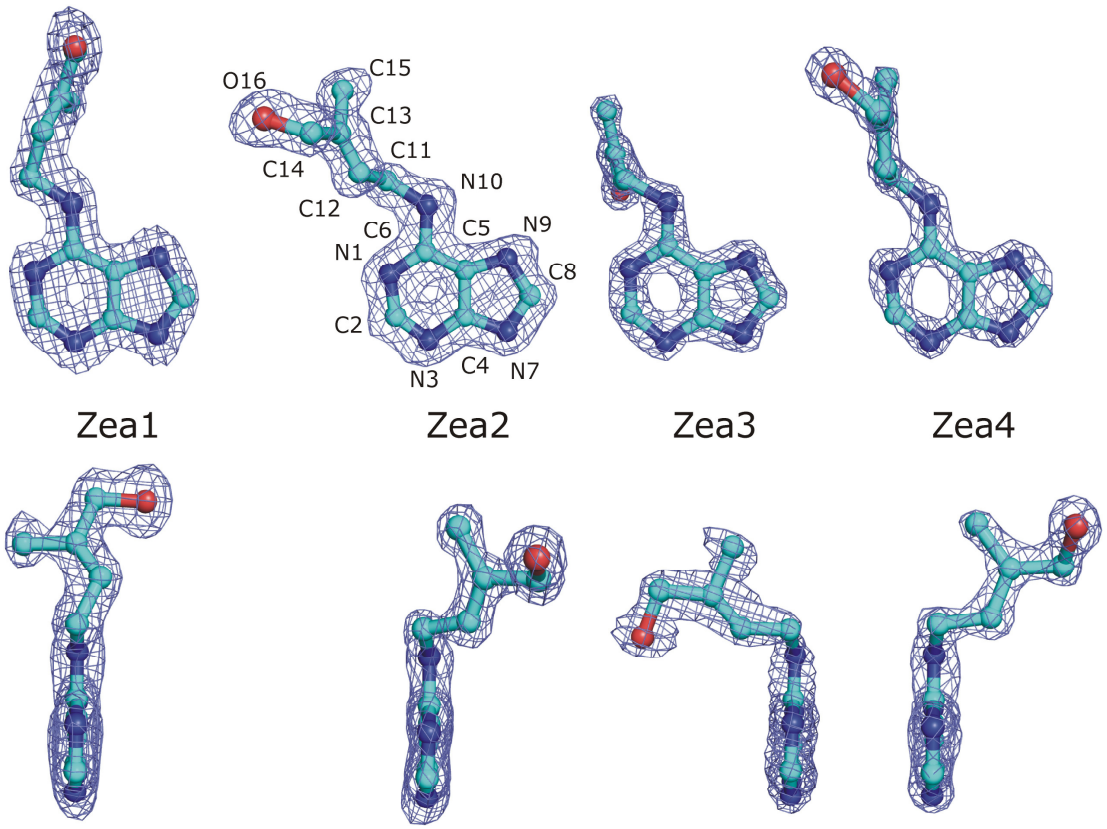


Figure 21. The four zeatin molecules present in the crystal structure of the LlPR-10.2B/zeatin complex, shown in two different views (rotated by 90º). The 2Fo-Fc electron density maps are contoured at the 1.5σ level. The zeatin atom numbering scheme used here (as defined in the PDB entry 2qim) is shown at Zea2.

All the zeatin molecules have evident *trans* configuration (Fig. 21 and Table 6). Each of them has at least one direct hydrogen-bond contact with the protein. Additionally, the ligands are stabilized by numerous van der Waals interactions (Fig. 20).

Table 6. Conformation of the isoprenoid chain of the zeatin molecules. Torsion angles in degrees.

	Zea1	Zea2	Zea3	Zea4
N1-C6-N10-C11	-1.3	-6.8	-8.9	5.7
C6-N10-C11-C12	154.8	83.2	-82.1	90.4
N10-C11-C12-C13	87.7	134.7	-131.1	126.5
C11-C12-C13-C14	-177.5	-177.5	178.6	-178.2
C12-C13-C14-O16	8.0	-124.2	3.8	-131.0

Zea1 is bound deep in the cavity close to the two short helices, with the adenine moiety closer to α1 and the aliphatic tail closer to α2 (Fig. 14). It is anchored to the protein only by two hydrogen bonds at the tip of the aliphatic tail (His68, Tyr80). The hydrogen bonds of the adenine moiety are with Zea2 (at N1), and with several water molecules (Table 7). Zea1 also forms several van der Waals contacts (Fig. 20).

Zea2 is aligned with helix α3 near the β1/α3 opening, with the adenine moiety closer to the N-terminal half and the aliphatic tail pointing in the C-terminal direction of the helix (Fig. 14). The terminal hydroxyl group of the aliphatic tail of Zea2 forms a hydrogen bond with Zea1 and with Tyr82. N9 and N10 donate three hydrogen bonds to the carboxylate group of Asp7 (N9···Oδ1; N9···Oδ2; N10···Oδ2). Zea2 is also stabilized by several van der Waals contacts (Fig. 20), and the remaining hydrogen-bonding centers of the adenine ring are satisfied by water molecules (Table 7).

Zea3 is placed near the L3/L5/L7/α3 opening of the cavity with the aliphatic tail pointing into the cavity (Fig. 14). The molecule is surrounded by six water molecules and forms only one hydrogen bond with the protein (N7···Oε2 (Glu59)) (table 7). Zea3 also participates in van der Waals contacts with Phe57, Arg138, Gly139 and zeatin molecules 1 and 2 (Fig. 20).

The terminal hydroxyl group of the aliphatic tail of Zea4, which is located between symmetry-related protein molecules, forms three hydrogen bonds, two with protein residues (Glu129 and Lys54') and one with a water molecule (Fig. 22). Two additional hydrogen bonds of Zea4 are formed between two water molecules and the adenine moiety (Table 7).

Table 7. Hydrogen bonding interactions of the zeatin molecules with corresponding bond lengths (Å) in parentheses. Atoms from a symmetry-related molecule are labelled with prime (').

Atom	Zea1	Zea2	Zea3	Zea4
N1	Zea2-O16 (2.72)	Wat164-O (2.79)	Wat211-O (2.92)	-
N3	Wat176-O (2.89)	Wat175-O (2.83)	Wat213-O (2.59)	Wat291'-O (2.84)
N7	Wat170-O (2.86)	Wat174-O (2.79)	Glu59-Oε2 (2.71)	Wat275-O (2.87)
N9	Wat181-O (2.93)	Asp7-Oδ1 (2.93) Asp7-Oδ2 (3.26)	Wat168-O (2.77)	- -
N10	Wat181-O (2.92)	Asp7-Oδ2 (2.65)	Wat167-O (3.01)	-
O16	His68-Nδ1 (2.70) Tyr80-Oη (2.65)	Tyr82-Oη (2.60) Zea1-N1 (2.72)	Wat173-O (2.93) Wat242-O (2.85)	Glu129-Oε1 (2.59) Wat281-O (3.00) Lys54'-Nζ (3.16)

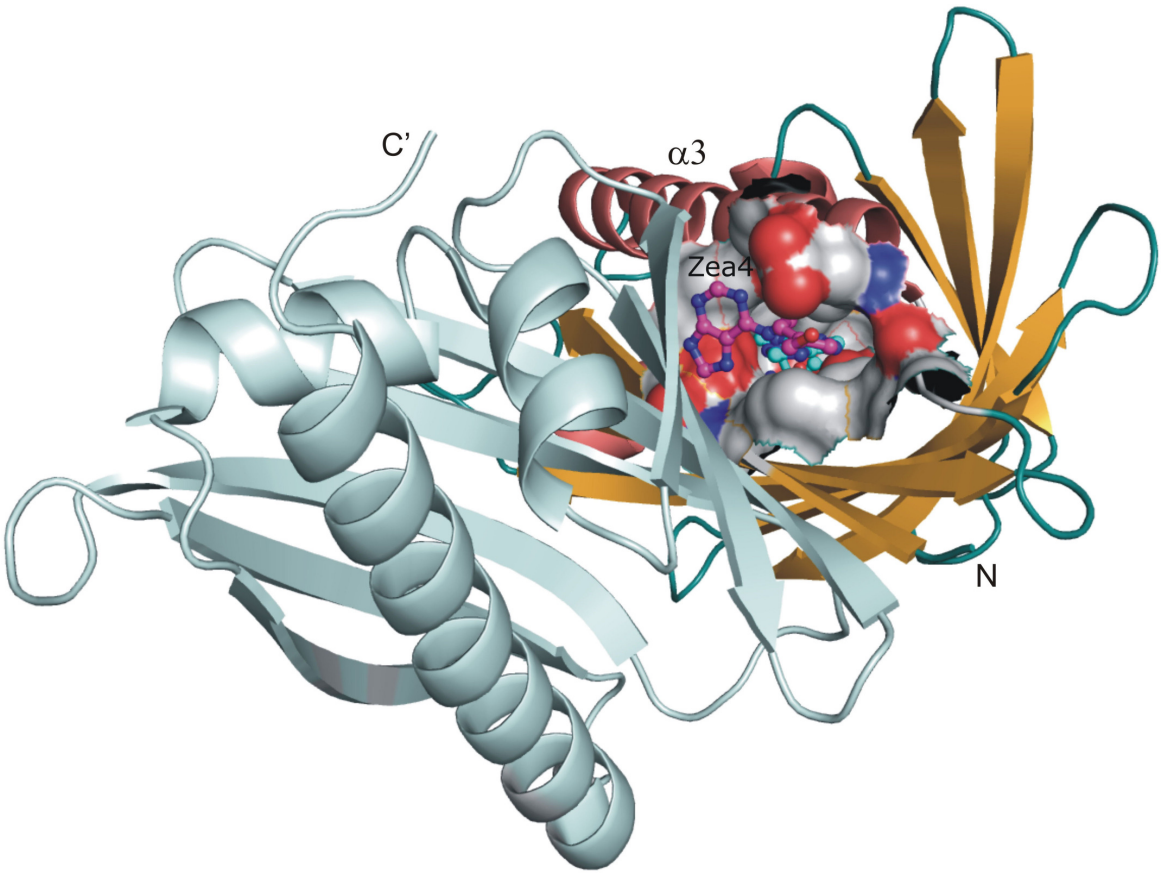


Figure 22. Location of Zea4 between two symmetry-related protein molecules. The protein molecules are related by a 60º rotation about the screw axis and the 1/6 translation of the repeat unit along the **c**-axis, characteristic of the space group $P6_5$.

III.1.2.10. The N- and C-termini

In all yellow lupine PR-10 crystal structures, the N-terminal $-NH_3^+$ group, in addition to being involved in the β1-β7 interaction, is also stabilized by conserved hydrogen bonds to residues in loop L9 (121ThrO, 123GlyO, 121ThrOγ1) (Biesiadka *et al.*, 2002; Pasternak *et al.*, 2005). In this new crystal structure of LlPR-10.2B in complex with zeatin, the same pattern is observed (Fig. 23).

The carboxyl terminus is less well ordered. Nevertheless, there is clear electron density till the last amino acid residue (Asn157). The stabilization of the C-terminus is mainly provided by intramolecular hydrogen bonds formed by Asn157 with Lys20 from helix α1 (Asn157O···W363···Lys20Nζ) and by Tyr156 with elements from helix α2 (Tyr156Oη···Asp25Oδ2). It is interesting to note that the former of these interactions is created with mediation of a water molecule.

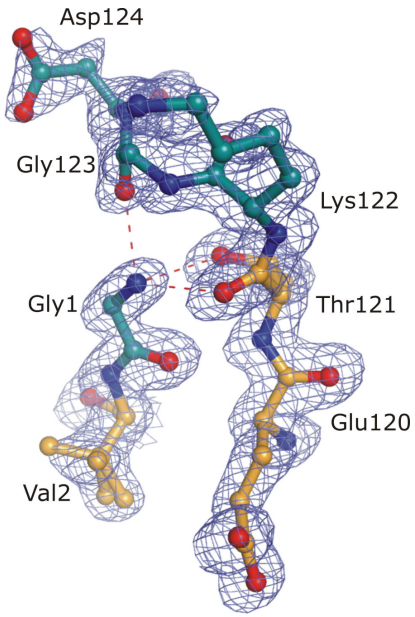


Figure 23. Ball-and-stick representation and 2Fo-Fc electron density map contoured at the 1.3σ level showing the stabilization of loop L9 by the N-terminal $-NH_3^+$ group of Gly1.

III.1.2.11. Crystal packing

The amino terminus of the LlPR-10.2B protein is highly ordered, forming the β1 strand of the molecular β-sheet. Apart from the intramolecular hydrogen bonding with strand β7, the β1 chain of the molecule forms antiparallel β-type interactions with strand β5' of a crystallographically related molecule (2ValO···89'GlyN, 4ThrN···87'GlyO, 4ThrO···87'GlyN, 6GlnN···85'ValO) (Fig. 24).

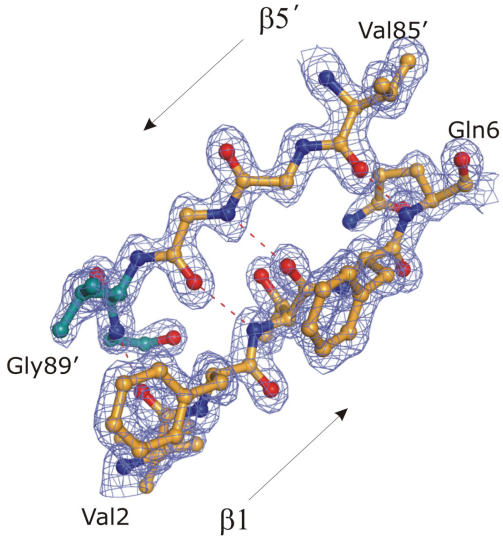


Figure 24. Ball-and-stick representation and 2Fo-Fc electron density map contoured at the 1.3σ level of the intermolecular β1-β5' interaction.

The intermolecular hydrogen bonding of β1 with β5' of a symmetry related molecule, generated by the 6_5 screw axis of the space group results in an infinite chain of connected protein molecules (Fig. 25), forming an endless antiparallel β-sheet (Fig. 26).

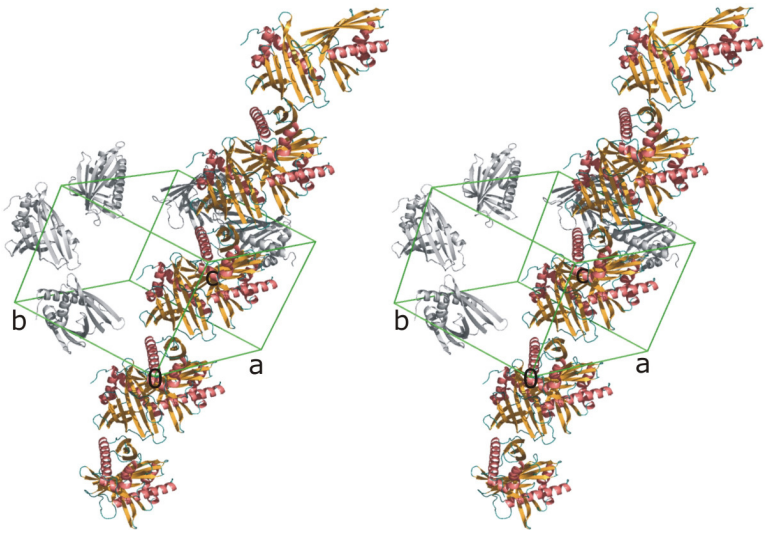


Figure 25. Stereoview of the infinite chain of LlPR-10.2B protein molecules connected via intermolecular hydrogen bonding of β1 with β5' of a symmetry-related molecule.

Figure 26. Arrangement of the infinite antiparallel β-sheet in the crystal structure of the LlPR-10.2B/zeatin complex. All non-β elements have been omitted for clarity. Residues involved in intermolecular hydrogen bonding are highlighted in blue.

III.1.3. *Lupinus luteus* PR-10.2B/diphenylurea complex

III.1.3.1. Crystallization

The LlPR-10.2B/DPU complex was formed by incubating 11 mg/ml protein (in 3 mM Na-citrate buffer, pH 6.3) with diphenylurea (100 mM, dissolved in acetone) in 1:5 molar ratio. The crystals were grown by Prof. Grzegorz Bujacz (Technical University of Lodz) by the hanging-drop vapor-diffusion method at room temperature, by equilibrating a mixture of 2 µl protein solution with 2 µl reservoir solution containing 1.4 M Na-citrate in 0.1 M citrate buffer, pH 6.0, against 0.5 ml of the reservoir solution.

III.1.3.2. X-Ray data collection

X-ray diffraction data extending to 1.95 Å resolution were collected at the EMBL X11 beamline at the DORIS ring of the DESY synchrotron. The crystal was flash-vitrified at 100 K in a stream of cold N_2 gas (Teng, 1990). The high salt content of the mother liquor provided sufficient cryoprotection. The images were indexed, integrated and scaled using the HKL program package (Otwinowski and Minor, 1997), to yield a 99.4% complete data set characterized by an R_{int} of 0.075. The crystal belongs to space group $C222_1$ with unit cell dimensions a=34.4, b=73.2, c=100.0 Å (Table 8).

Table 8. Data collection statistics for the LlPR-10.2B/DPU complex.

Space group	$C222_1$
Cell parameters [Å]	a=34.4, b=73.2, c=100.0
Resolution limits (last shell) [Å]	40.0-1.95 (2.02-1.95)
Radiation source	DESY, X11 (EMBL)
Wavelength [Å]	0.8100
Temperature [K]	100
No. of measured reflections	54070
No. of unique reflection	9491
R_{int} (last shell)	0.075 (0.398)
Completeness (last shell) [%]	99.4 (96.8)
Redundancy	5.7
$\langle I/\sigma(I)\rangle$ (last shell)	20.8 (3.8)

III.1.3.3. Structure solution, refinement and deposition

The structure was solved by molecular replacement with the program Molrep (Vagin and Teplyakov, 1997), using diffraction data between 20.0-2.0 Å and the LlPR-10.2B model from the zeatin complex. The Matthews coefficient (Matthews, 1968) of 1.86 $Å^3$ Da^{-1} (34.0% solvent content) indicated one LlPR-10.2B molecule in the asymmetric unit, therefore one solution was sought. After rigid-body refinement, an R factor of 0.47 characterized the best solution in the $C222_1$ space group. The top solution had a significantly higher correlation coefficient than the next solution.

Most of the protein residues were clearly visible in the electron density map phased by molecular replacement. The remaining side chains, the ligand molecules (DPU), water molecules and Na^+ ions were added during interactive cycles of model building in XtalView and Coot (McRee, 1999; Emsley and Cowtan, 2004), followed by structure refinement in REFMAC-5 (Murshudov *et al.*, 1997). The refinement was performed against all unique reflections, except 860 reflections (9.1%) randomly selected for R_{free} testing. Several steps of isotropic refinement were conducted, combined with introduction of alternative conformation for several side chains, and completion of the solvent model. Stereochemical restraints for the DPU molecules were imposed as for the protein residues. A dictionary of ideal bond lengths and angles for DPU was prepared in REFMAC-5 (Murshudov *et al.*, 1997). Water molecules were added if the corresponding Fo-Fc peaks were higher than 4σ, and they were retained if their B-factors refined to < 70 $Å^2$. The occupancies of residues in double conformation were manually adjusted based on electron density maps and temperature factors. The final model includes 154 protein residues, four DPU ligands, 134 water molecules and two sodium ions (Table 9).

Table 9. Refinement statistics for the LlPR-10.2B/DPU complex.

Program used		REFMAC-5
Resolution limits [Å]		15.0-1.95
No. reflections		8631
No. reflections in test set		860
No. of atoms	Protein	1169
	Ligand	64
	Metal	2 (Na)
Water molecules		134
R/R_{free}		0.193/0.246
⟨B⟩ [Å^2]	Proteins atoms	39.6
	Ligand atoms	36.0
	Water molecules	45.2

Atomic coordinates and structure factors for the LlPR-10.2B/DPU complex have been deposited in the Protein Data Bank with the accession code 3e85.

III.1.3.4. Model quality and overall fold

The refined 1.95 Å resolution crystallographic model of LlPR-10.2B has good overall geometry and Ramachandran statistics (Table 10 and Fig. 27).

Table 10. Stereochemical quality of the LlPR-10.2B/DPU complex.

Rms deviation from ideal	Bond lengths [Å]	0.018
	Bond angles [°]	1.88
	Torsion angles [°]	6.2
	Chiral-center restraints [Å^3]	0.106
Ramachandran φ/ψ angles [%]	Most favoured	93.8
	Additional allowed	6.2

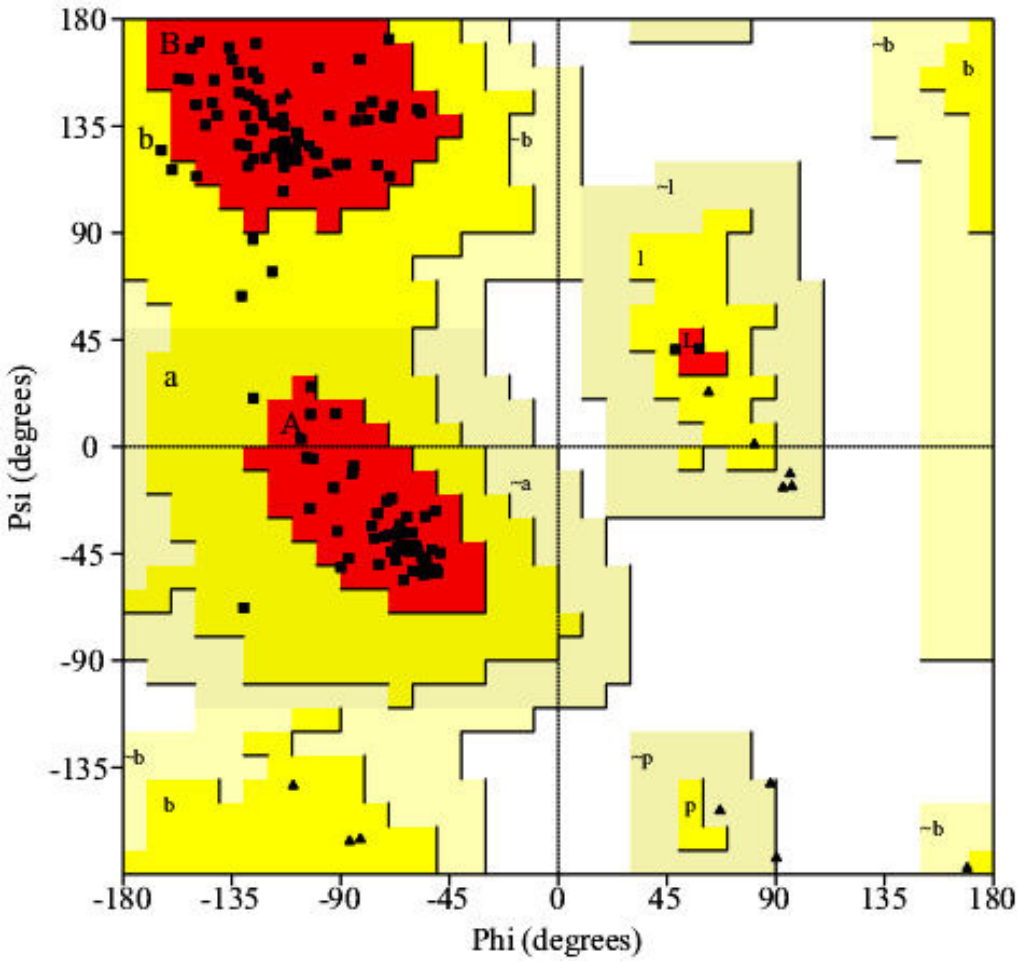


Figure 27. Ramachandran plot illustrating φ and ψ backbone dihedral angles for the LlPR10.2B protein in complex with DPU. Glycine residues are identified by triangles as they are not restricted to the regions of the plot appropriate for other residue types (indicated by squares). The proline residues are not represented.

The quality of the electron-density maps is high and there are only a few less clear areas in the loop regions. The last 3 amino acids at the C-terminus are not visible in the electron density. The crystallographic asymmetric unit contains one LlPR-10.2B molecule with an overall fold consisting, as in other PR-10 members, of three α-helices and seven β-strands giving rise to an α+β fold whose most pronounced feature are a β-grip over the C-terminal helix α3 and a vast cavity bound by all the helices and the concave face of the β-sheet. There are seven β-bulges, which distort the regularity of the β-sheet endowing it with a highly curved shape. In this complex, four DPU molecules are present in the cavity (Fig. 28).

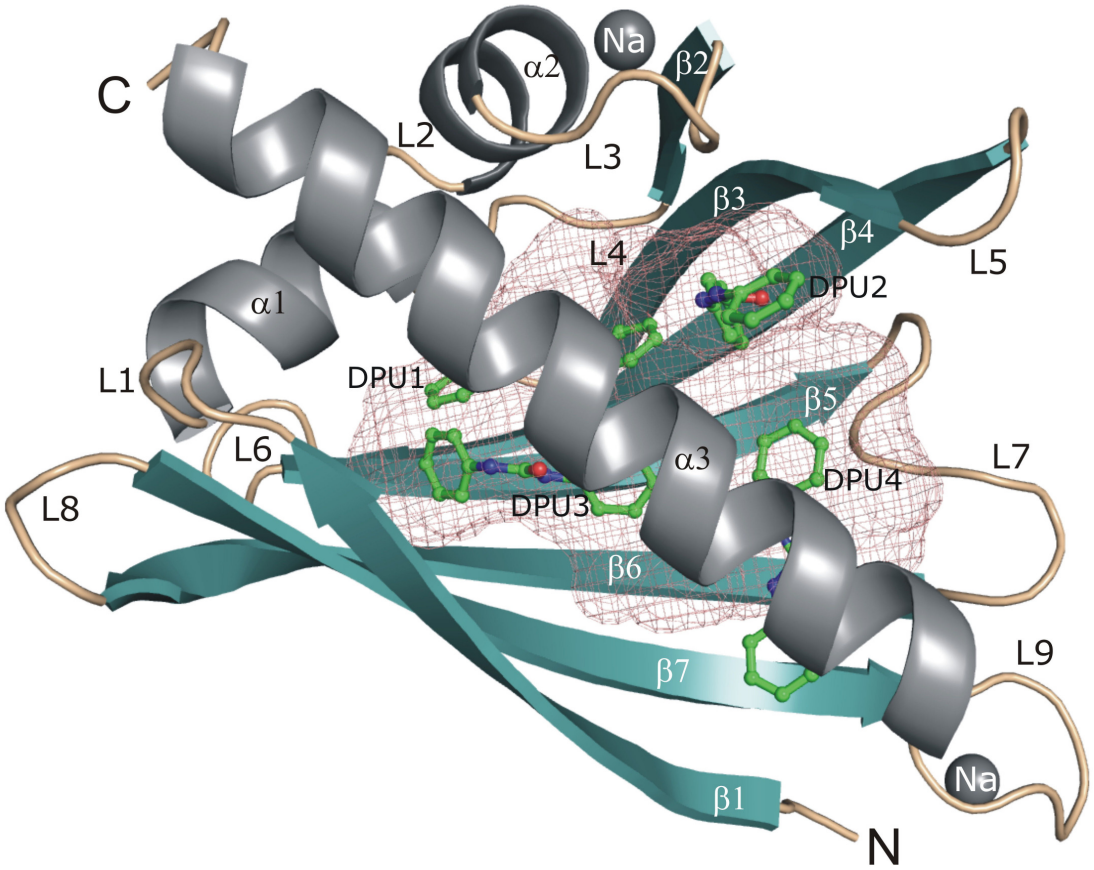


Figure 28. Overall fold of the LlPR-10.2B molecule with annotation of secondary structure elements. The binding cavity is represented as a rose cast, calculated in VOIDOO (Kleywegt and Jones, 1994).

III.1.3.5. The glycine-rich loop

The glycine-rich loop L4 is characterized by excellent electron density (Fig. 29) and low B-factors. The rigidity of the loop is maintained by a pattern of three hydrogen bonds between the Oγ1 atom of Thr51 and the main-chain N-H groups of Asn46, Gly47 and Gly48.

It is instructive to compare the electron density of the glycine-rich loop L4 in this LlPR-10.2B/DPU complex (Fig. 29) with the electron density obtained for the LlPR-10.2B/zeatin complex (Fig. 15) and note the difference of electron density maps obtained for crystals diffracting up to 1.95 Å (DPU complex) and 1.35 Å (zeatin complex).

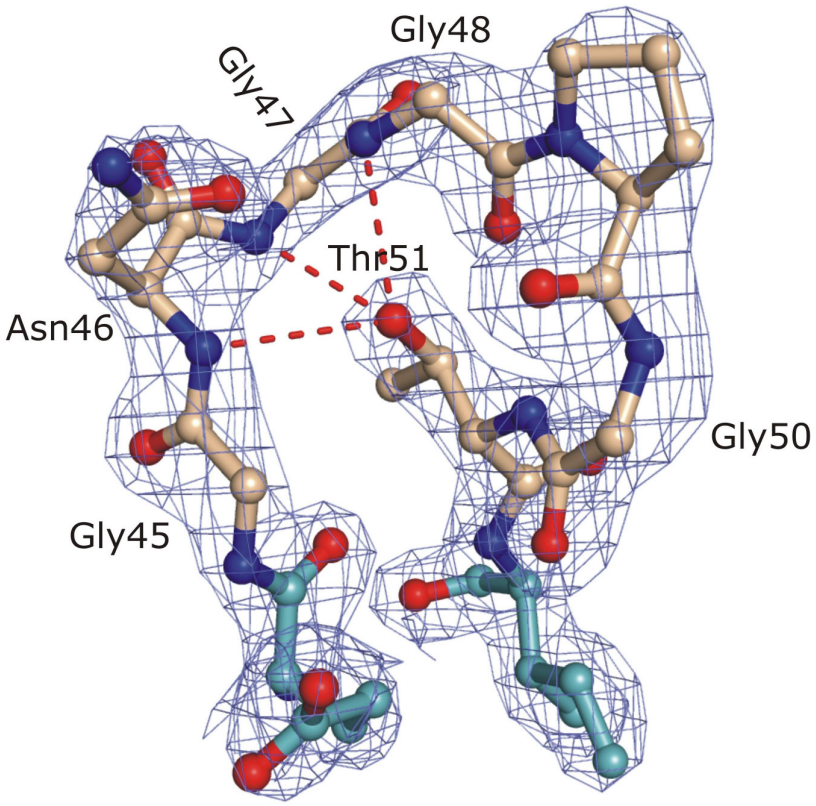


Figure 29. The glycine-rich loop (L4) with the three hydrogen bonds formed by the side chain of Thr51 and the main-chain amides of residues Asn46-Gly47-Gly48. The 2Fo-Fc electron density map is contoured at the 1.3σ level. The ball-and-stick illustrations of atomic models in the III.1.3 subsection are colored red for O atoms and blue for N atoms. C atoms display the color according to the secondary structure element they belong, as in Fig. 28.

III.1.3.6. Asymmetric unit contents

The asymmetric unit contains one LlPR-10.2B protein molecule in complex with four diphenylurea molecules, 134 modelled water molecules (one of them modelled with half occupancy) and two sodium cations. The metal cations have octahedral coordination close to loop L3 (Na1) where the ligands are three carbonyl O atoms (Pro31, Val34, Ile37) and three water molecules (Fig. 30A), and close to loop L9 (Na2) where the ligands are two carbonyl O atoms (Thr121, Gly123), the Oγ1 atom of Thr121 and three water molecules (Fig. 30B). The correctness of the interpretation of the metal sites as sodium is confirmed by the satisfactory refinement of the B factor (34.4 and 41.0 $Å^2$ for Na1 and Na2), by the final $Na^{+\cdots}O$ distances (2.1-2.9 Å) and by the bond-valence test (Muller *et al.*, 2003) indicating a valency of 1.17 and 1.10, for the first and second metal site, respectively, for sodium (expected: 1.0), versus 1.72 and 1.73, respectively, for calcium (expected: 2.0).

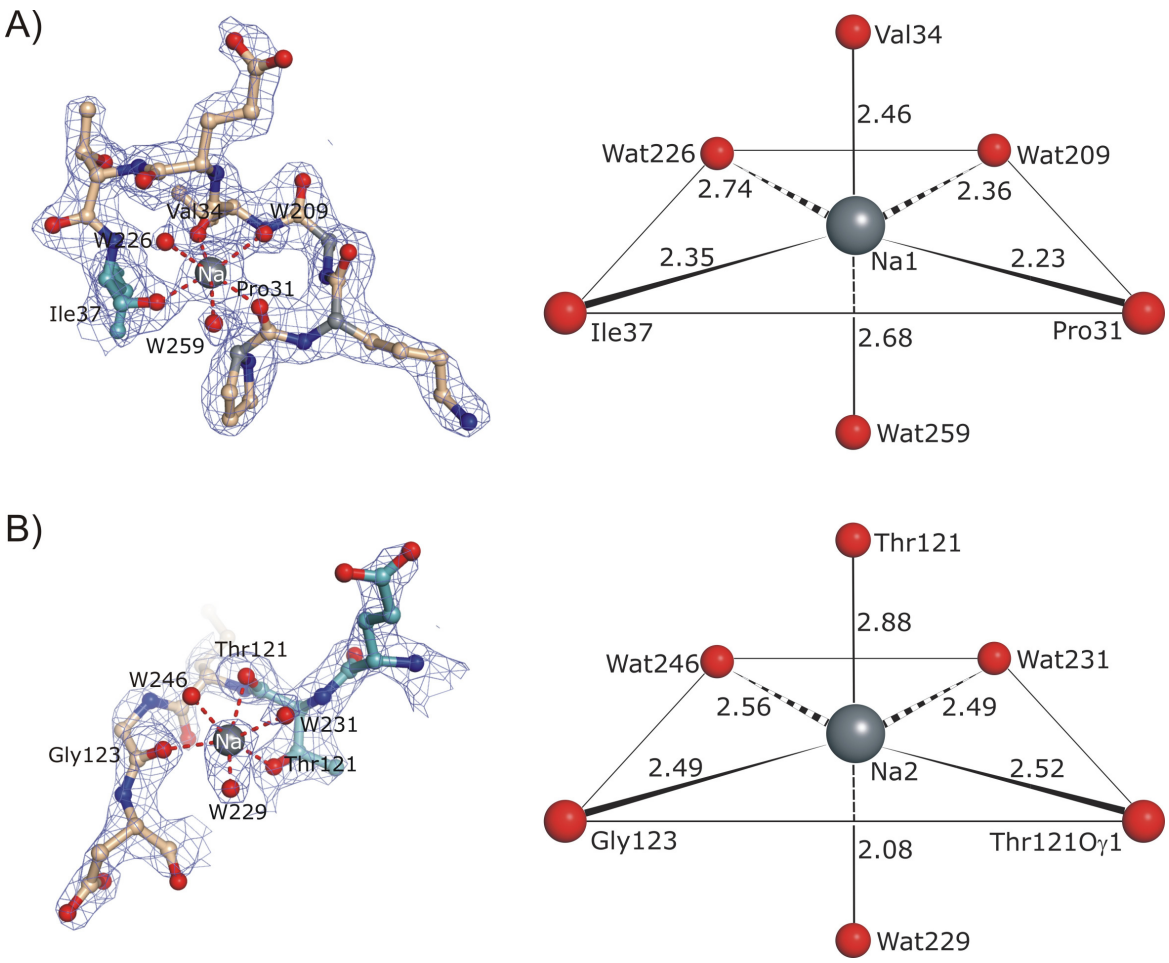


Figure 30. Coordination of the Na^+ cations in the LlPR-10.2B/DPU complex structure. On the left is shown the coordination geometry and the 2Fo-Fc electron density map (1.2 σ level) and on the right the coordination scheme with $Na^{\cdots}O$ distances in Å. (A) The Na^+ cation coordinated at loop L3. (B) The Na^+ cation coordinated at loop L9.

III.1.3.7. The ligand binding cavity

The cavity enclosed within the LlPR-10.2B protein core in the present DPU complex has a large volume, calculated as 3600 $Å^3$ by the SURFNET program (Laskowski, 1995). The cavity is not closed and can be accessed by two openings. The larger one is located between helix α3 and loops L3, L5, L7 and L9 and is gated by a salt bridge between Arg138 (helix α3) and Glu59 (β3) (Fig. 31A and 32) and by a water-bridged contact between the side chains of Glu131 (helix α3) and Thr93 (loop L7) (Fig. 31B and 32). The second opening is found between stand β1 and helix α3 (Fig. 31C).

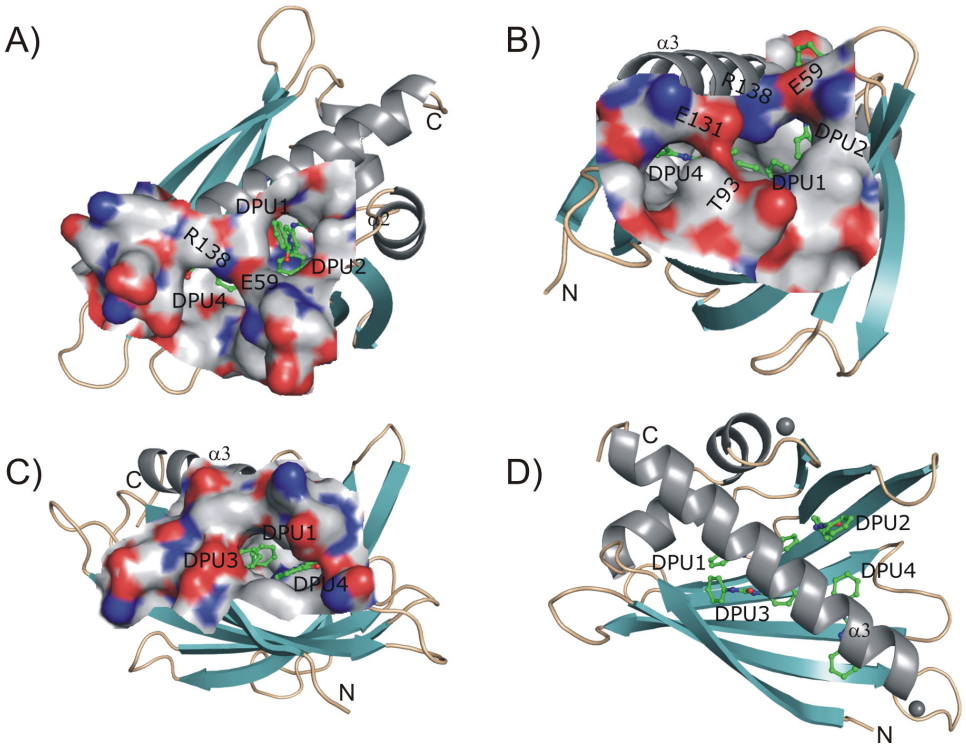


Figure 31. (A)-(C), Entrances to the cavity within the LlPR-10.2B protein in its DPU complex. See text for details. Panel (D) has been added for easier recognition of the orientation of the remaining panels.

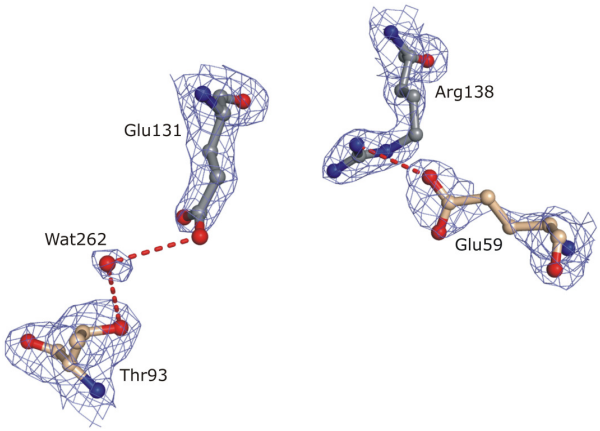


Figure 32. Close-up view of the gating interactions at the α3/L3/L5/L7/L9 entrance to the LlPR-10.2B/DPU cavity. The 2Fo-Fc electron density map is contoured at the 1.0 σ level.

The entrances seen near β5/L7/β6 and at L2/L4 in the crystal structure of the LlPR-10.2B/zeatin complex (Fig. 18) are closed. Despite the absence of the β5/L7/β6 opening, the LlPR-10.2B/DPU structure also contains a water molecule (W168) disrupting the β5-β6 sheet, two hydrogen bonds away from the ends of loop L7 (Fig. 33).

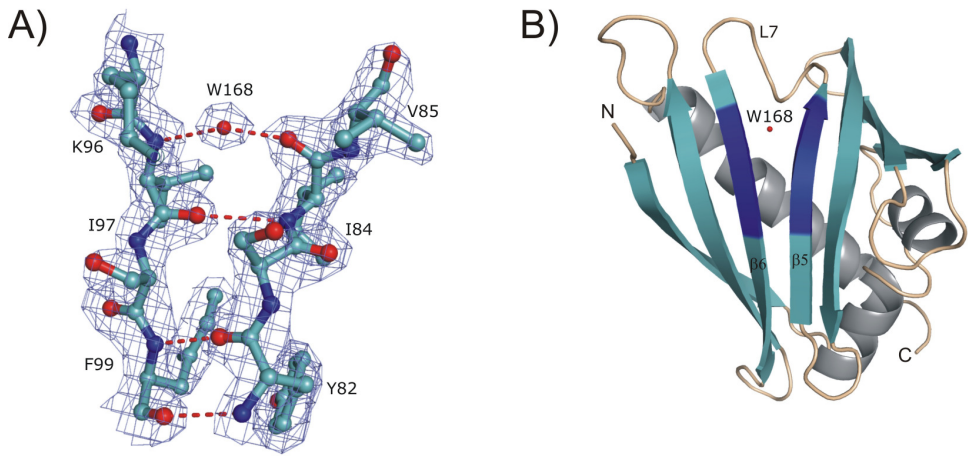


Figure 33. Disruption of the antiparallel β-sheet by a water molecule. (A) Ball-and-stick representation and 2Fo-Fc electron density map contoured at the 1.3σ level. (B) Cartoon representation of the protein with the regions represented in the left panel highlighted in dark blue.

In variance with the LlPR-10.2B/zeatin complex, where three zeatin molecules and 25 water molecules occupy the binding cavity, in the present complex the cavity is filled with four DPU molecules (DPU1-4) and eight water molecules (Fig. 34). One of the phenyl rings of the DPU4 ligand is outside of the cavity, pointing to the β1-α3 opening (Fig. 28 and 31C). A remarkable feature of the DPU binding in the hydrophobic cavity is the lack of any specific interatomic interactions, such as hydrogen bonds, between the protein and the DPU molecules.

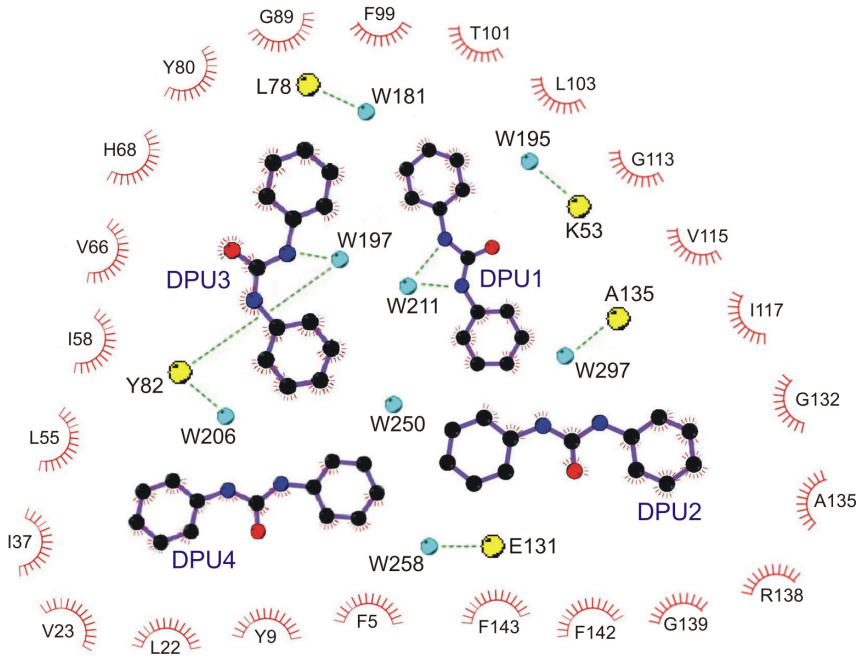


Figure 34. Representation of the four DPU molecules found in the protein cavity. Water molecules are shown as aquamarine spheres. Hydrogen bonds are represented by dash lines.

III.1.3.8. Diphenylurea binding

In this complex crystal structure four DPU molecules have been modelled in the electron density (Fig. 28 and 35) with full occupancies. The average B factors of the DPU molecules are 31.8, 33.1, 34.2 and 45.0 $Å^2$ for DPU 1, 2 3 and 4, respectively.

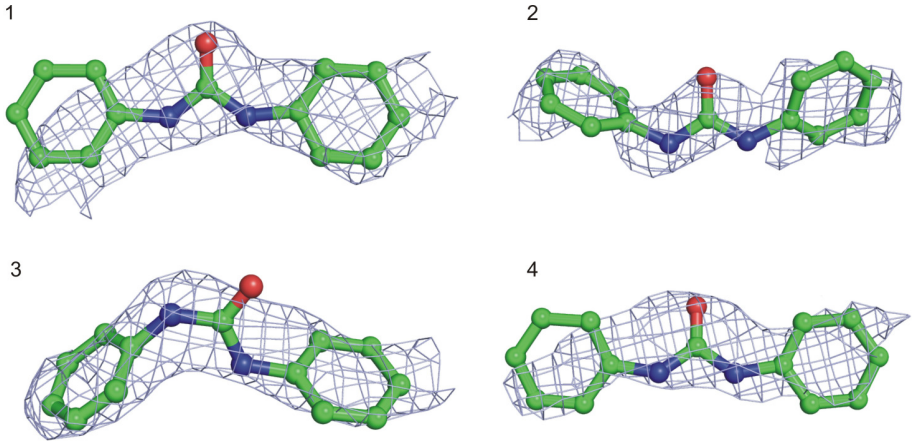


Figure 35. The four DPU molecules present in the crystal structure of the LlPR-10.2B/DPU complex. The 2Fo-Fc electron density maps are countered at the 1.0σ level.

DPU1 is found deep in the cavity close to helices α1 and α2 (Fig. 28). It is anchored to the protein by several van der Waals contacts involving Leu22, Val23, Leu55, His68, Tyr80 and Phe142. DPU2 is placed near the α3/L3/L5/L7/L9 opening of the cavity (Fig. 28). It establishes van der Waals contacts with Ile37, Phe57, Ile58 and Arg138. DPU3 is aligned with helix α3 near the β1/α3 opening (Fig. 28), in an orientation that is roughly parallel to the DPU1 molecule. It is anchored to the protein via van der Waals contacts with Tyr9, Leu22, Phe99, Thr101, Leu103, Gly113, Val115, Gly139 and Phe143. DPU3 makes also an indirect hydrogen bond contact with the protein (Tyr82), mediated by a water molecule (W197) (Fig. 34). DPU4 is placed at the N-terminal part of helix α3, with the two phenyl rings pointing at the two different entrances to the cavity, namely β1/α3 and α3/L3/L5/L7/L9 (Fig. 28). It is anchored to the protein via van der Waals contacts with Phe5, Gly89, Phe99, Ile117, Gly132, Ala135 and Arg138. The DPU ligands also establish van der Waals contacts between themselves, namely within the pairs DPU1-DPU2 and DPU1-DPU3.

The DPU molecules have a rigid and planar carbonyl diamide group with two lateral phenyl substitutes at different orientations (Fig. 36 and Table 11). In their elongated electron density, the DPU molecules can move along their long axes. This is consistent with the lack of direct hydrogen bond interactions with the protein scaffold.

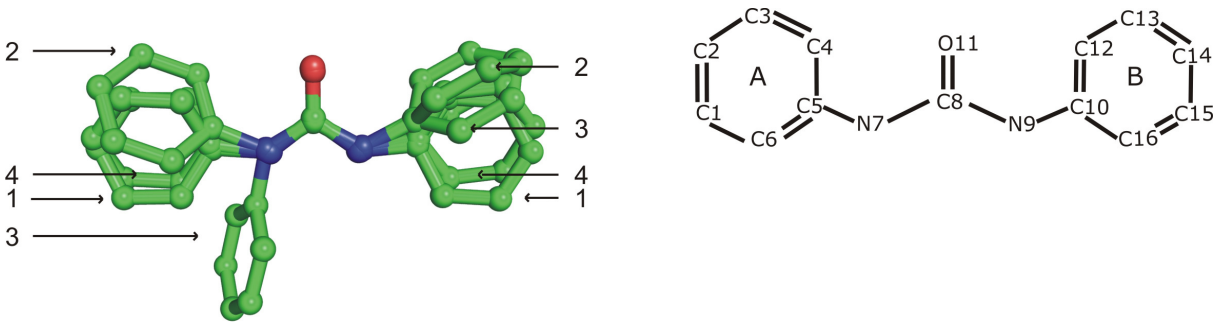


Figure 36. Conformation of the DPU molecules 1-4 shown in a superposition of their urea moieties. The atom numbering scheme of the diphenylurea molecule is also shown.

Table 11. Conformation of the DPU molecules. Torsion angles in degrees.

	DPU1	DPU2	DPU3	DPU4
C5-N7-C8-O11	1.0	-5.5	179.3	-1.5
C4-C5-N7-C8	-163.3	-133.0	-108.9	-25.2
C10-N9-C8-O11	3.5	3.1	-1.7	1.4
C12-C10-N9-C8	-13.9	65.2	117.6	17.3

III.1.3.9. The N- and C-termini

The amino terminus of the LlPR-10.2B protein is highly ordered, forming the β1 strand of the β-sheet. Due to the presence of the Na^+ cation coordinated in loop L9, the first two amino acid residues of β1 are pushed away from the β-sheet (Fig. 37). Because of the presence of the cation, the N-terminal $–NH_3^+$ group is not stabilized by the conserved hydrogen bonds to residues in loop L9 (121ThrO, 123GlyO, 121ThrOγ1) seen in other yellow lupine PR-10 structures (Biesiadka *et al.*, 2002; Pasternak *et al.*, 2005; Fernandes *et al.*, 2008 and the present work) (Fig. 23). Nevertheless, the electron density shows no disorder for the N-terminal part of the protein (Fig. 37).

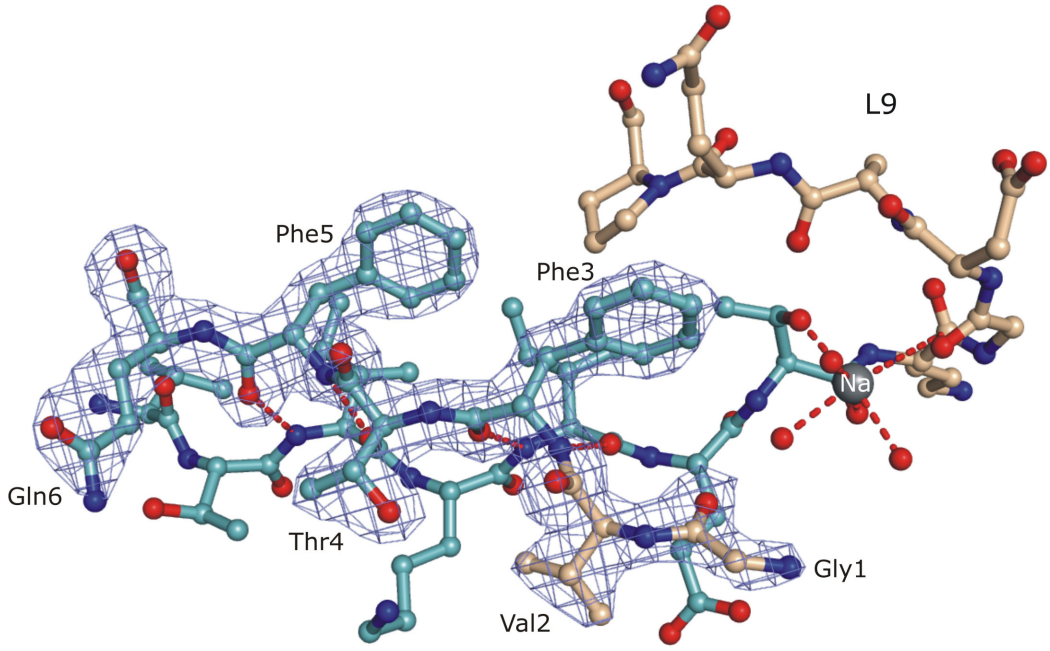


Figure 37. Ball-and-stick representation and 2Fo-Fc electron density map contoured at the 1.3σ level of the N-terminal part of the LlPR-10.2B protein. In the background, a part of β7 is shown. Loop L9 is also represented.

The carboxyl terminus is less ordered than the N-terminus and there is no visible electron density for the last three amino acid residues. The last amino acid modelled is Pro154 (Fig. 38).

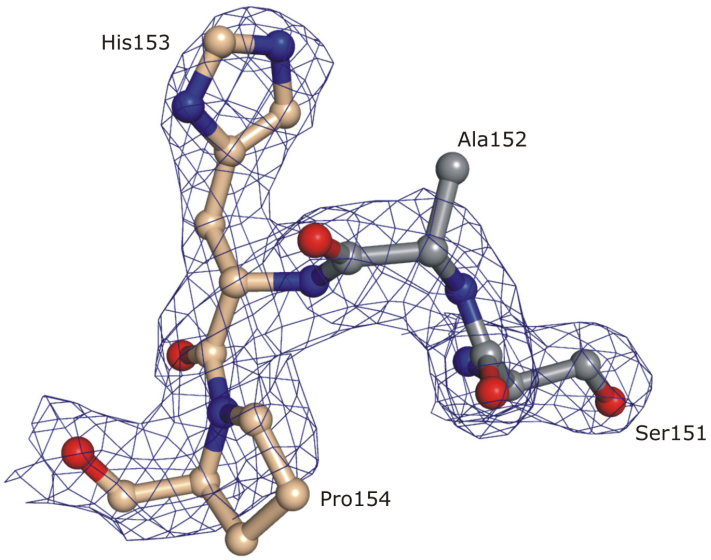


Figure 38. Ball-and-stick representation and 2Fo-Fc electron density map contoured at the 1.0σ level of the C-terminal part of the LlPR-10.2B protein.

III.2. *Pinus monticola* PR-10.3.1 protein

III.2.1. Cloning, protein expression and purification

The PmPR-10.3.1 coding sequence, kindly provided by Professor Abul K. M. Ekramoddoullah from Pacific Forest Center, Canada, was PCR-amplified from a pGEM vector and introduced into a pET-3a plasmid between. In more detail, the PCR conditions consisted of an initial 5 minutes denaturation step at 94°C, followed by 30 one-minute cycles at 94°C, 1 minute at 64°C and 1 minute at 72°C, for denaturation, primer annealing and extension, respectively. After the cycles, a 10 minute elongation run at 72°C was performed. A forward primer (5'-CGTCGCCATATGGTGA GTGGGACTGCAACAAT-3') and a reverse primer (5'-GCGGGATCCTTAGCAGTATAAGCTA AAGTTGGAGA-3') were used. The PCR product was loaded onto a 1.0% agarose gel, electrophoresed, extracted from the gel, purified and subcloned in the pGEM plasmid (Promega). The plasmid was then used to transform *E. coli* (DH5α) cells. The transformed bacteria were plated and selected by the galactosidase test (blue/white colonies in the presence of X-Gal). The colonies with the correct insert were taken for further subcloning into pET-3a expression plasmid, using the cloning site between the BamH I and Nde I restriction sites. The plasmid was then used to transform *E. coli* strain BL21. Insertion of the correct cDNA sequence was checked by PCR.

The sequence analysis reveal four mutations (Fig. 39) with respect to the expected sequence encoding the PmPR-10.3.1 protein (GenBank accession No. AAU00103), namely His10Asp, Leu35Phe, Pro69Ala, and Ile104Leu. An additional silent mutation was found in codon 57. Detailed analysis showed that the four mutation correspond to amino acid substitutions that can be found in the same positions in other proteins from the PmPR-10 family. For this reason, it can be expected that these mutations should not interfere with the folding of the protein. In addition, working with this construct has provided important information about the purification procedure.

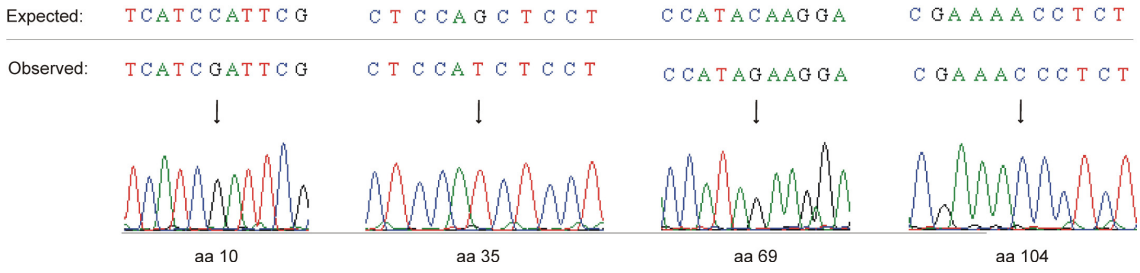


Figure 39. DNA sequencing chromatograms showing mutations found in the coding region of PmPR-10.3.1.

The expression vector was introduced in *E. coli* cells, strain BL21 (DE3) pLysS. The transformed cells were grown at 37ºC in LB culture supplemented with 100 µg/ml ampicillin and after they reached the logarithmic growth phase (OD_{600} near 1.0), protein expression was induced by 1.0 mM IPTG. After 4 hours of growth at 37ºC, the cells were harvested by centrifugation, resuspended in lysis buffer, and sonicated. Cell debris was pelleted by centrifugation at 15555g and 4ºC for 1 hour. Most of the PmPR-10.3.1 protein was found in the insoluble fraction; therefore the precipitate was treated with 7.2 M urea. Renaturation of the protein was conducted by dialysis against 20 mM Tris buffer, pH 8.0 containing 5% glycerol and 5 mM β-mercaptoethanol. The preparation was clarified by centrifugation and subjected to S-carboxymethylation, followed by dialysis against the initial buffer. The dialysed fraction was clarified by centrifugation and submitted to further purification on a HiPrep S300 (GE Healthcare) column (Fig. 40). The chromatography was performed in 20 mM Tris buffer, pH 8.0, 5% glycerol, and 5 mM β-mercaptoethanol.

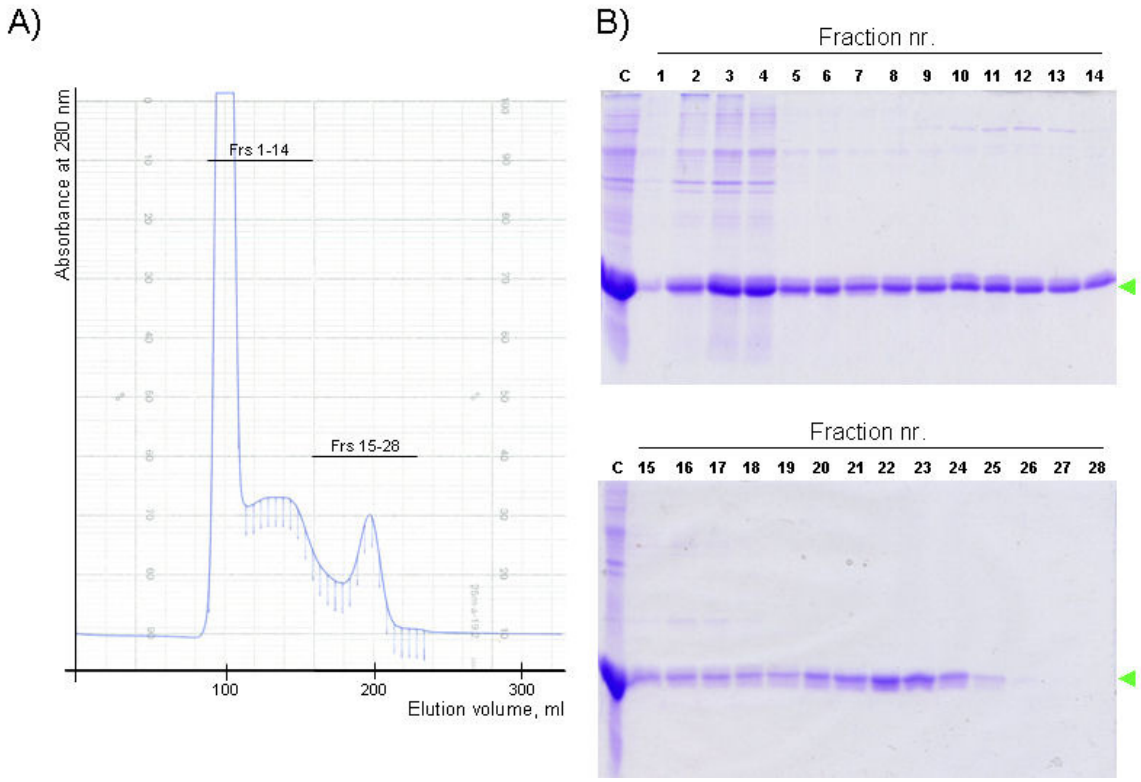


Figure 40. (A) Size exclusion chromatography on HiPrep S300 (GE Healthcare) column of dialysed PmPR-10.3.1 protein. (B) The fractions (1-28) of the protein peaks were analysed by 15% SDS-PAGE. C is the material loaded on the column. The green arrow indicates the electrophoretic migration of the PmPR-10.3.1 protein.

The elution volumes from the two peaks in the gel filtration chromatogram and the same migration in electrophoresis gels indicate that in solution the protein is present in the monomeric and homo-tetrameric forms. The monomeric fractions (20-25) and tetrameric (11-15) forms were separately pooled and submitted to size exclusion chromatography on a HiPrep S300 (GE Healthcare) column (Fig. 41) using the same buffer as the previous chromatographic step. The yield of the recombinant protein after purification was about 9 and 30 mg per liter of liquid culture, for the monomeric and tetrameric forms, respectively.

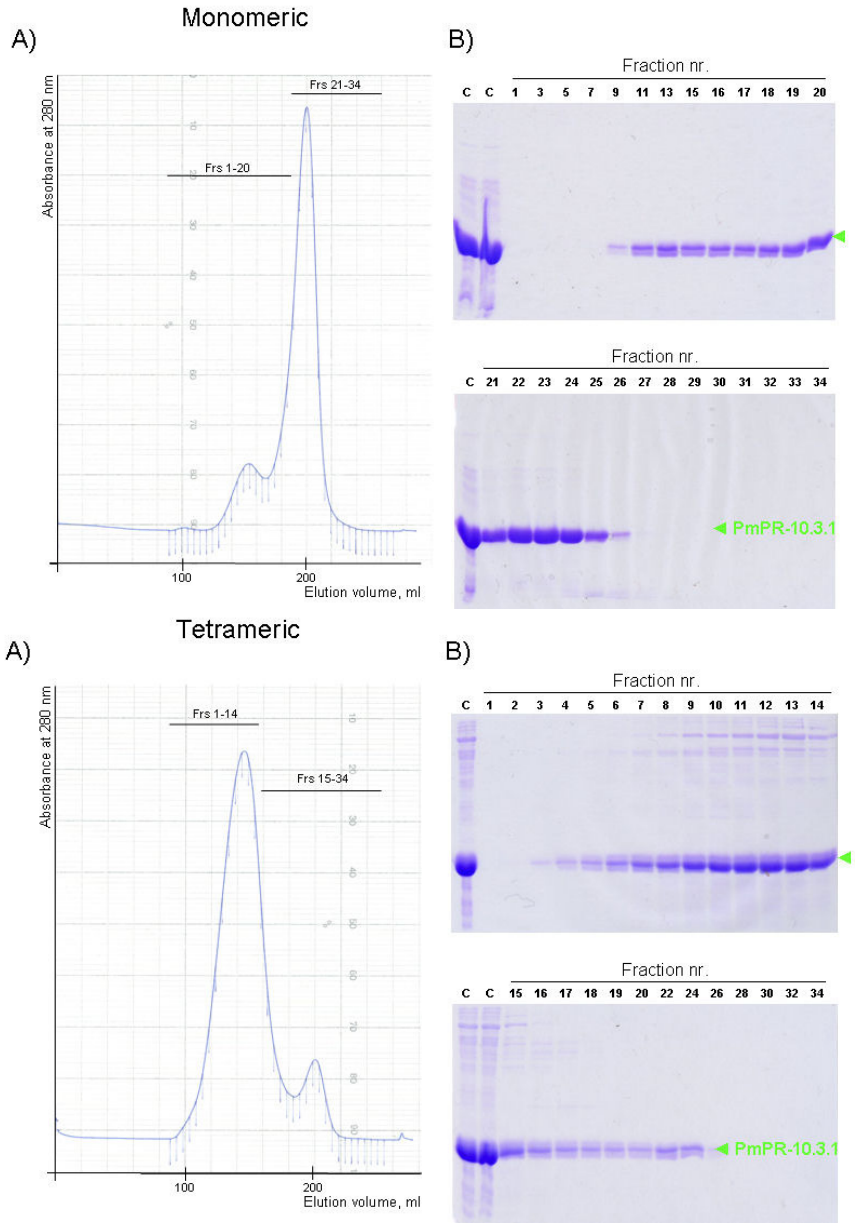


Figure 41. (A) Size exclusion chromatography on HiPrep S300 (GE Healthcare) column of the combined fractions of the peaks from the previous size exclusion chromatography of the PmPR-10.3.1 protein. (B) The fractions of the protein peaks were analysed by 15% SDS-PAGE. C is the material loaded on the column. The green arrow indicates the electrophoretic migration of the PmPR-10.3.1 protein.

III.2.2. Circular dichroism

CD experiments were carried out in the far-UV region (205-265 nm). The PmPR-10.3.1 protein was concentrated in stabilization buffer (10 mM Tris-HCl buffer, pH 8.5 or 6.8 for the monomeic or tetrameric forms, respectively) and diluted just before measurement. The CD spectra obtained were recorded at 6.66 (monomeric) and 4.98 µM (tetrameric form) protein concentration in 10 mM sodium phosphate buffer, pH 7.0. The scan speed was set to 100 nm/min and 5 scans were averaged. The spectra obtained are presented without mathematical smoothing (Fig. 42).

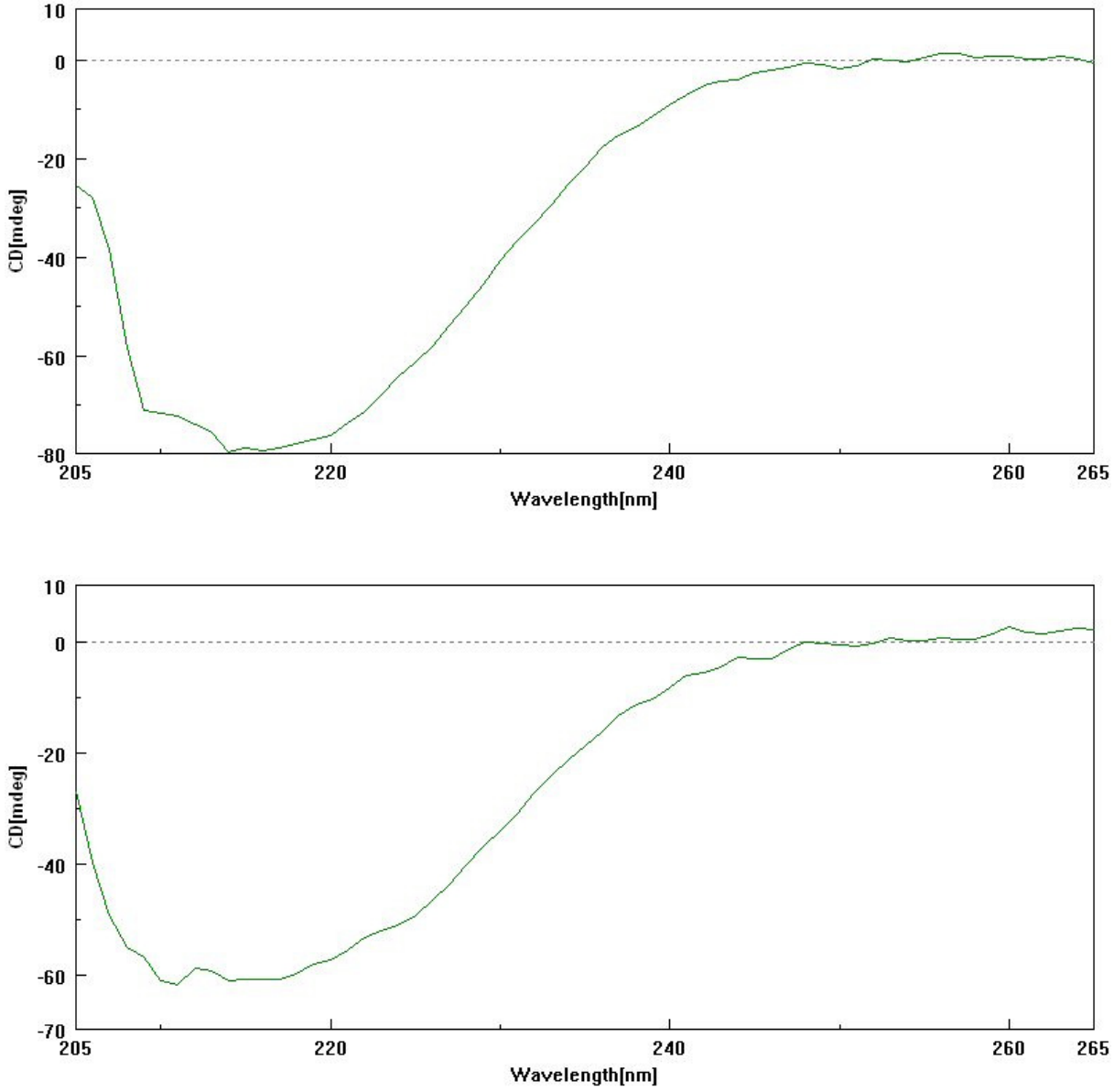


Figure 42. Unsmoothed far-UV CD spectra, recorded at 20ºC for the monomeric (top) and tetrameric form (bottom) of the PmPR-10.3.1 protein.

III.2.3. Crystallization

Prior to crystallization trials, a solution of PmPR10.3.1 protein was concentrated to 10 mg/ml in 10 mM Tris buffer, pH 8.5, for the monomeric form and in 10 mM Tris buffer, pH 6.8, for the tetrameric form. With these samples, several crystallization screening experiments were carried out. In parallel, some of the screens were carried out in the presence of plant hormones (zeatin, DPU and CPPU). The initial screens included Crystal Screens I and II and PEG/ion screen from Hampton Research as well as a wide range of salt and PEG conditions.

Slow-growing crystals appeared in the drop where the reservoir consisted of 10% PEG 4000 for the tetrameric form of the protein (Fig. 43). The crystals were grown by the hanging-drop vapor-diffusion method at room temperature. A mixture of 1.8 µl of protein solution was equilibrated with 1.8 µl of the reservoir solution containing 10% PEG 4000 in 0.1 M citrate buffer, pH 5.5, against 1.0 ml of the reservoir solution. The drop was supplemented with 0.2 µl 0.1 M $SrCl_2$.

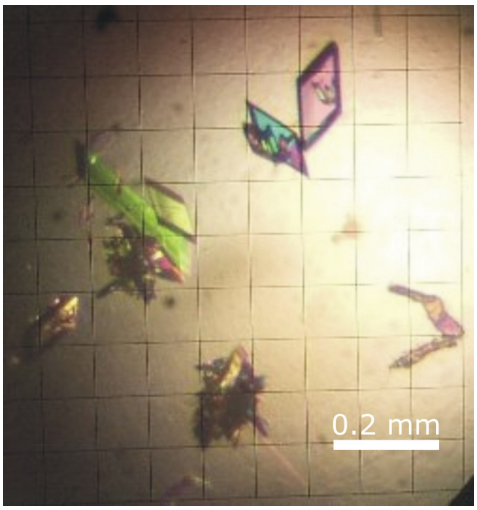


Figure 43. Crystals obtained from a preparation of the tetrameric form of the PmPR-10.3.1 protein.

III.3. *Theobroma cacao* PR-10 protein

III.3.1. Protein expression and purification

The expression vector with the coding sequence of the TcPR-10 protein was supplied by Dr. Abelmon S. Gesteira from UESC, Brazil. The TcPR-10 coding sequence was inserted into the pET-29c vector between the Nde I and Sal I restriction sites.

The expression vector was introduced into *E. coli* cells, strain BL21 (DE3) pLysS. The transformed cells were grown at 37ºC in LB culture supplemented with 100 µg/ml kanamycin and when they reached the logarithmic growth phase (OD_{600} near 1.0), protein expression was induced by the addition of 0.5 mM IPTG. After 4 hours of growth at 37ºC, the cells were harvested by centrifugation, resuspended in lysis buffer, and sonicated (Ultrasonic Processor, Model CV33, Fisher; 5 x 20 s bursts at 80% power). Cell debris was pelleted by centrifugation at 15555g and 4ºC for 1 hour. Most of the TcPR-10 protein was found in the insoluble fraction; therefore the precipitate was treated with 7.2 M urea. Renaturation of the protein was conducted by dialysis against 20 mM Tris buffer, pH 8.5 containing 5% glycerol and 5 mM β-mercaptoethanol. The protein fraction was precipitated with 80% ammonium sulfate to separate it from non-protein cell components. After the precipitate was resuspended, and dialysed against the initial buffer. The dialysed fraction was clarified by centrifugation and passed through a DE52-cellulose column equilibrated with the same buffer. Fractionation was carried out by stepwise elution with NaCl (Fig. 44).

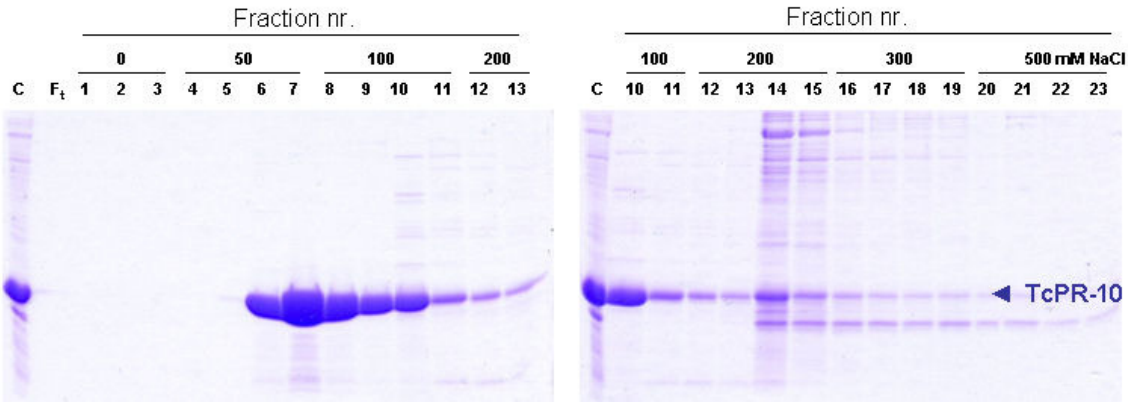


Figure 44. 15% SDS-PAGE analysis of stepwise fractionation of the recombinant TcPR-10 protein-containing fractions on DE52-cellulose in the presence of 20 mM Tris-HCl buffer, pH 8.5, 5% glycerol and 5 mM β-mercaptoethanol. C is the material loaded on the column. The blue arrow indicates the electrophoretic migration of the PmPR-10.3.1 protein.

The fractions eluted with the third and fourth portion of 50 mM NaCl, the 100 mM portions and the first two 200 mM NaCl portions were pooled, concentrated and submitted to further purification on a Superdex 75 (GE Healthcare) column. The chromatography was performed in the same buffer as the DE52-cellulose chromatography (Fig. 45). The yield of the recombinant protein after purification was about 30 mg per liter of liquid culture.

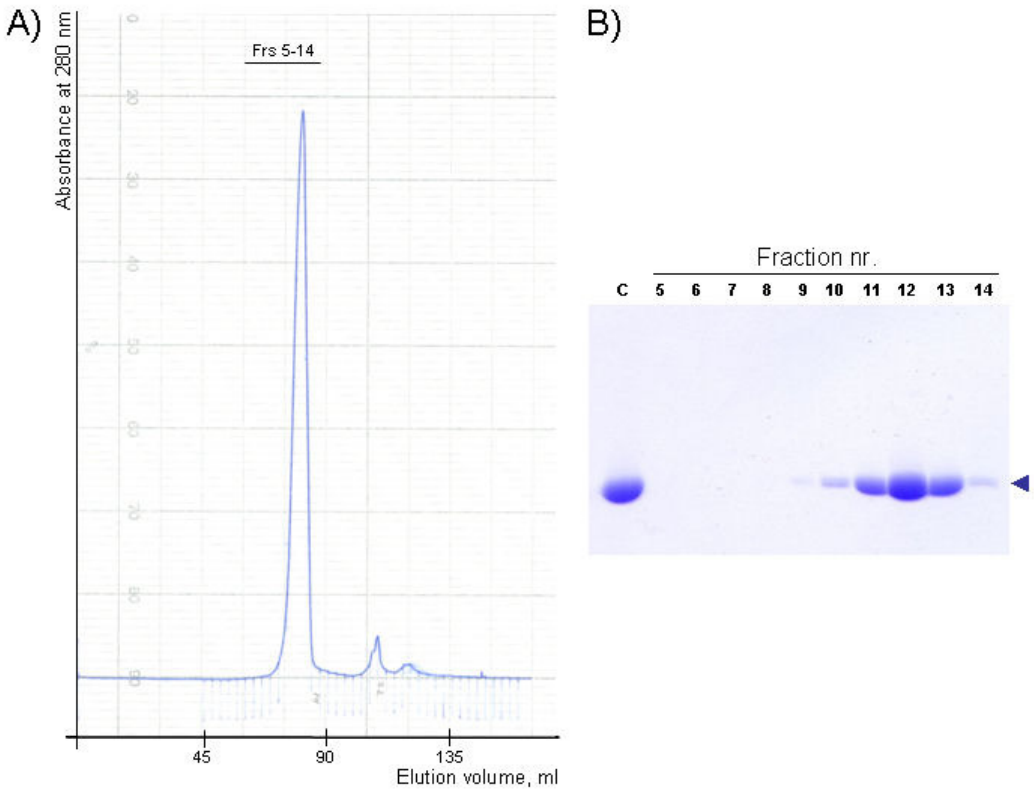


Figure 45. (A) Size exclusion chromatography on Superdex 75 (GE Healthcare) column of the combined fractions separated by ion-exchange chromatography on DE52 cellulose. (B) The eluted protein fractions were analysed by 15% SDS-PAGE. C is the material loaded on the column. The blue arrow indicates the electrophoretic migration of the TcPR-10 protein.

III.3.2. Circular dichroism

CD experiments were carried out at the far-UV region (205-265 nm). The TcPR-10 protein was concentrated in 3 mM sodium citrate buffer, pH 6.3, and diluted just before measurement. The CD spectra were recorded at 1.62 μM protein concentration in 10 mM sodium phosphate buffer, pH 7.0. The scan speed was set to 100 nm/min and 5 scans were averaged. The spectra are presented without mathematical smoothing (Fig. 46).

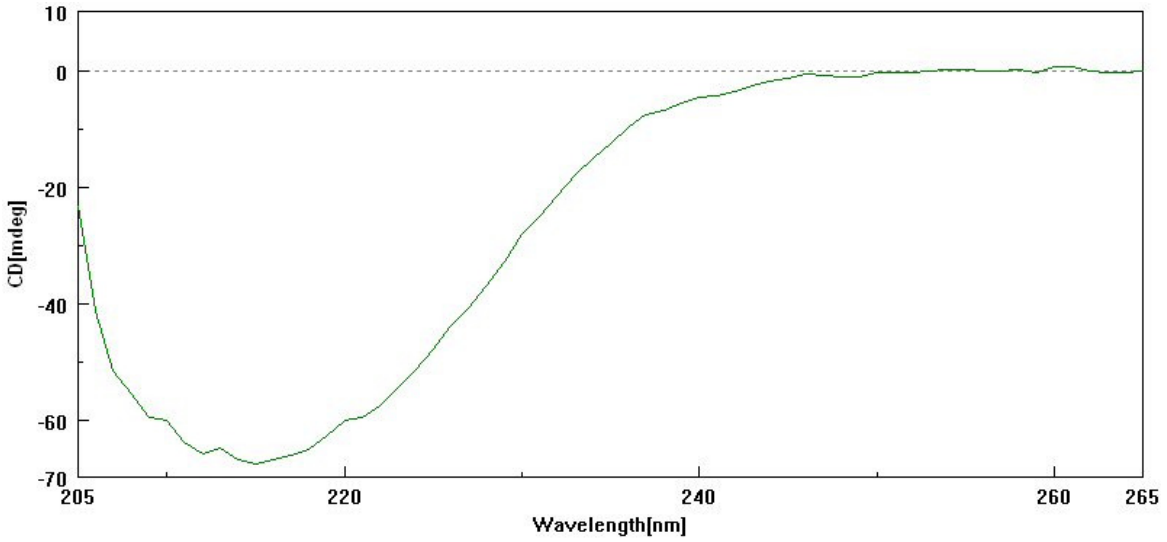


Figure 46. Unsmoothed far-UV CD spectrum of TcPR-10 protein, recorded at 20ºC between 205 and 265 nm.

III.3.3. Isothermal titration calorimetry

Ligand binding studies were performed using isothermal titration calorimetry (ITC) as described in Methods. The experimental data are summarized in Fig. 47. The only ligand investigated was zeatin. Due to the low solubility of the ligand, accurate ΔH and ΔS contributions to the binding process could not be determined precisely, therefore only stoichiometry and binding affinity were analyzed. The calorimetric measurements were analyzed by fitting a multiple-sites model to the raw data. The binding is characterized by two K_a values (3.7×10^6 and 4.2×10^4 M) and 1:1 stoichiometry (n = 1.3 and 1.2) for each of the two binding states.

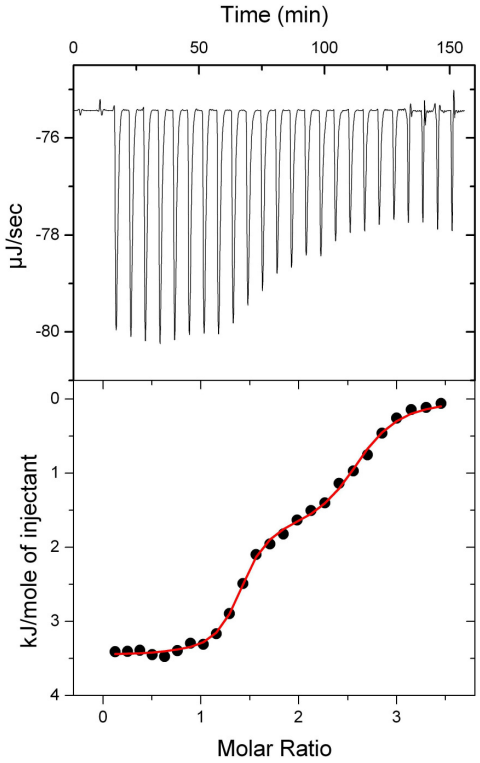


Figure 47. Isothermal calorimetric titration of TcPR-10 with *trans*-zeatin. The top panel shows raw heat data corrected for baseline drift obtained from 25 consecutive injections of 8.5 mM zeatin into the sample cell (1.0 ml) containing 0.66 mM TcPR-10 in 3 mM sodium citrate buffer, pH 6.3, at 20°C. The bottom panel shows the binding isotherm created by plotting the heat peak areas against the molar ratio of zeatin added to TcPR-10 present in the cell. The heats of mixing (dilution) were subtracted. The line represents the best fit to the model of multiple binding sites. The binding is characterized by two K_d values (0.27 ± 0.05 and 23.8 ± 3.5 µM) and 1:1 stoichiometry (n = 1.3 and 1.2) for each of the two binding states.

III.3.4. Antifungal tests

The effect of recombinant TcPR-10 protein on the *in vitro* growth of pathogenic fungi was determined. Purified TcPR-10 and controls (see Methods) were added to sterile paper discs. The fungal mycelial front was allowed to grow until the edge of the plate, and growth inhibition zones were investigated. Purified TcPR-10 did not inhibit the growth of any of the fungi studied, namely *Colletotrichum lupini*, *Leptosphaeria maculans* and *Leptosphaeria biglobosa* (Fig. 48).

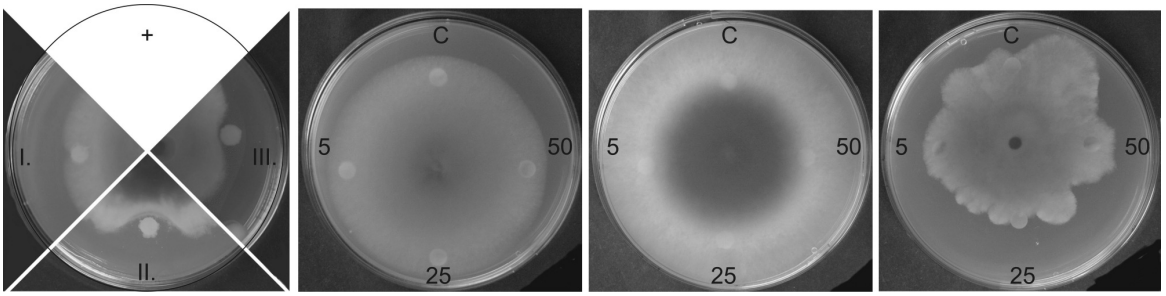


Figure 48. Antifungal activity test using recombinant TcPR-10 protein. Effect of filter discs containing 5, 25 or 50 µg of protein, or buffer as a control (C), in the *in vitro* growth of *Colletotrichum lupini*, *Leptosphaeria maculans* and *Leptosphaeria biglobosa*.

III.3.5. Crystallization

Prior to crystallization trials, the solution of TcPR-10 protein was concentrated to 20 mg/ml in 3 mM sodium citrate buffer, pH 6.3. With this sample, several crystallization screens were set up. In parallel, some of the screening experiments were carried out in the presence of plant hormones (zeatin, DPU and CPPU). Screening for crystallization conditions was performed using the hanging-drop vapor diffusion method at 20ºC by mixing 1.5 µl protein solution with 1.5 µl well solution. Initial screens included the Crystal Screens I and II and PEG/ion screen from Hampton Research, as well a wide range of salt conditions. In order to find additional crystallization conditions, the protein was also submitted to automated high-throughput crystallization screening at the EMBL laboratory in Hamburg, Germany (Fig. 49). 720 different conditions were set for each, the apo form and the protein in complex with zeatin and CPPU.

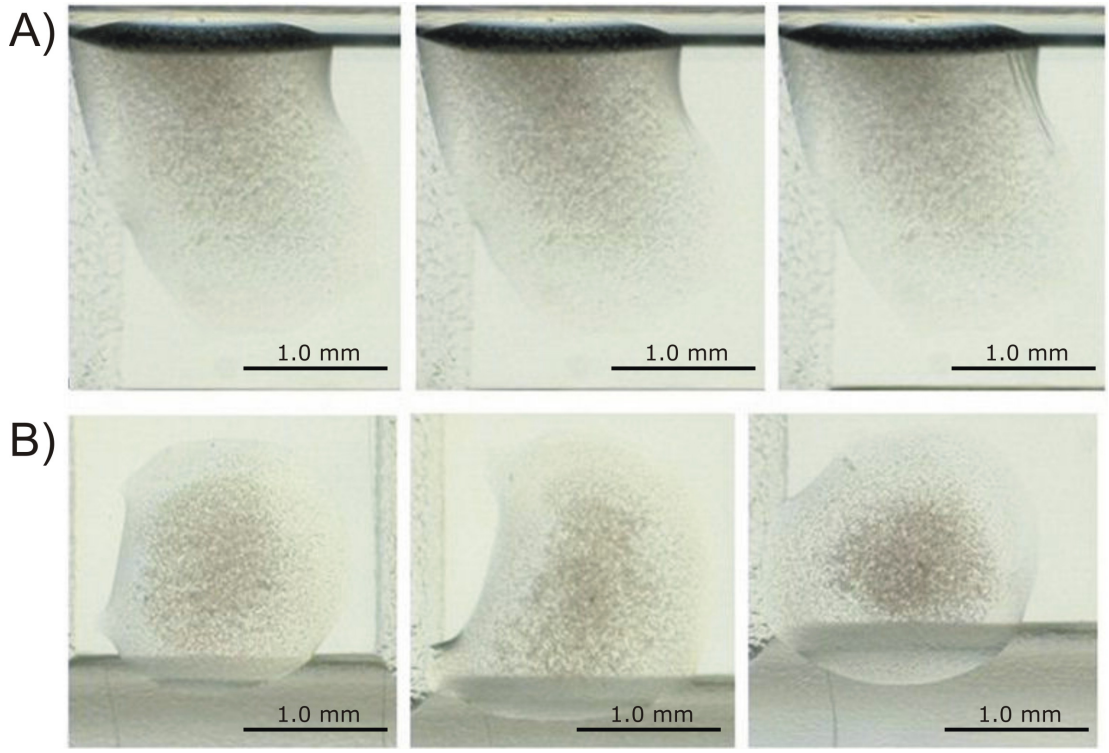


Figure 49. High-throughput crystallization of the TcPR-10 protein. (A) Time development of the crystallization experiment using 0.1 M Na-acetate, pH 4.6, 2.0 M NaCl for the TcPR-10/CPPU complex. From left to right: pictures taken 16, 79 and 164 days after setting the drop. (B) Comparison of the apo and holo TcPR-10 protein 164 days after the drops were set in 0.1 M HEPES pH 7.0, 0.8 M $(NH_4)_2SO_4$. From left to right: apo TcPR-10, TcPR-10/zeatin and TcPR-10/CPPU complexes.

One initial hit was optimized and some small single crystals were grown (Fig. 50). Crystals of TcPR-10 in complex with DPU were grown in 2.0 M NaCl, 0.1 M Na-acetate, pH 4.5, within two weeks.

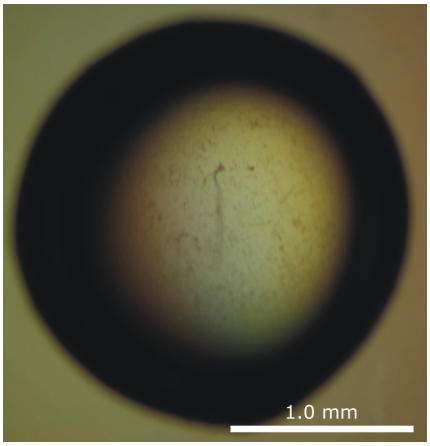


Figure 50. TcPR-10/DPU crystals. View of the crystallization drop.

When exposed to X-rays, the crystals produced a diffraction pattern of very low quality (Fig. 51). It was not possible to index the images and to determine the space group and cell dimensions. Nevertheless, it could be confirmed that the crystallized material was from protein and that a diffraction limit of 3.1 Å was achievable. Currently, the crystallization conditions are being optimized.

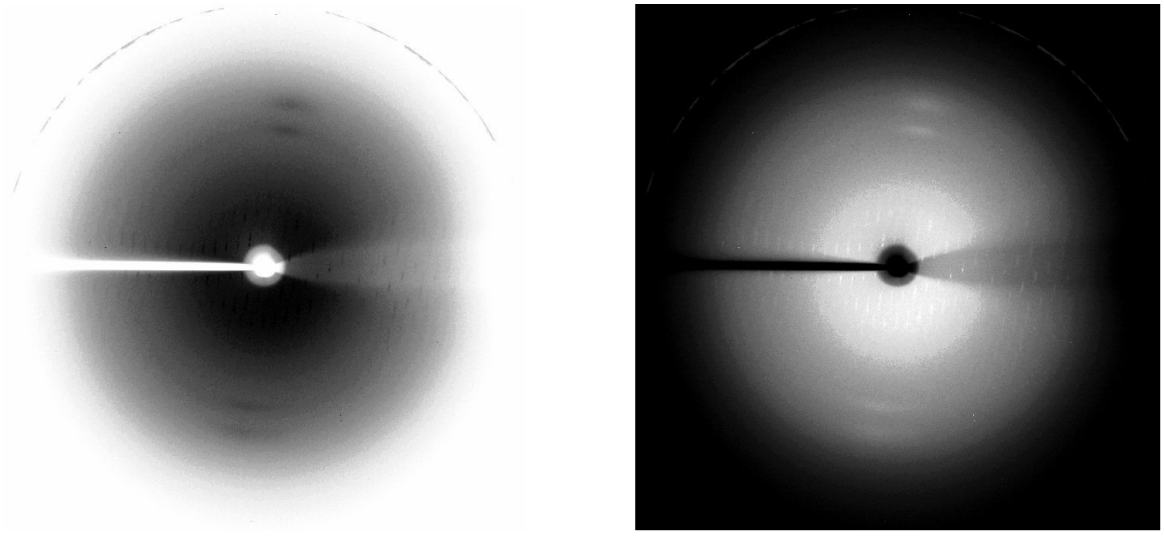


Figure 51. X-ray diffraction pattern (left), and its negative (right), recorded at 100 K for a single crystal of TcPR-10/DPU using synchrotron radiation. The edge of the detector corresponds to 2.3 Å resolution and the last visible spot correspond to the resolution of 3.1 Å.

III.4. *Hypericum perforatum* phenolic oxidative coupling protein (Hyp-1)

III.4.1. Protein expression and purification

Recombinant *H. perforatum* Hyp-1 protein was cloned in the pET151/D vector in fusion with a 6His tag (Konieczna, 2007). The vector was introduced into *E. coli* cells, strain BL21 Star (DE3). The transformed cells were grown at 30ºC in LB culture supplemented with 100 µg/ml ampicillin and when they reached the stationary growth phase (OD_{600} near 2.0), the protein expression was induced by 1 mM IPTG. After 5 hours of growth at 30ºC, the cells were harvested by centrifugation, resuspended in lysis buffer, and sonicated. The cell debris was pelleted by centrifugation at 15555g and 4ºC for 1 hour. The soluble fraction was passed through a DE52-cellulose column equilibrated with 20 mM sodium phosphate buffer, pH 7.5, containing 5% glycerol (Fig. 52). Fractionation was carried out by stepwise elution with NaCl.

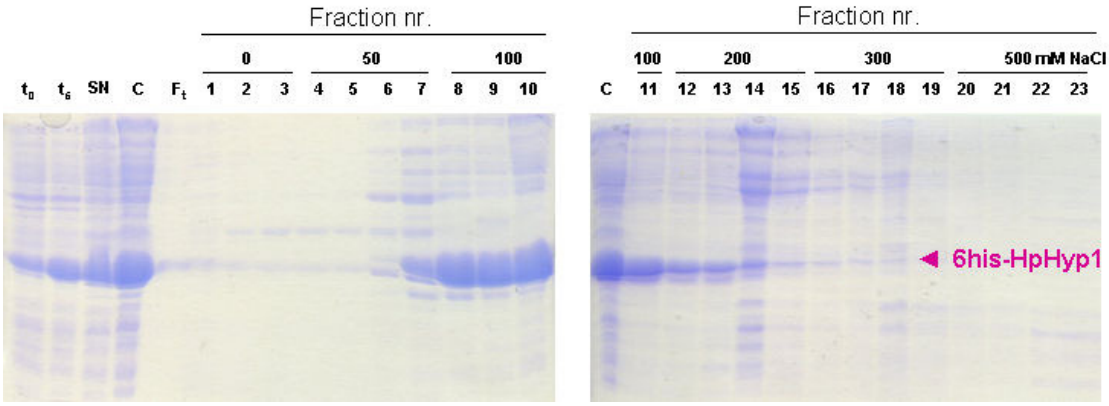


Figure 52. 15% SDS-PAGE analysis of NaCl fractionation of the recombinant 6His-HpHyp-1 protein on DE52-cellulose in the presence of 20 mM sodium phosphate buffer, pH 7.5. Lanes t_0 and t_5 are the crude extracts taken before and 5 hours after induction, respectively. SN is the soluble fraction after lysis. C and Ft are material loaded and the flow through of the DE52-cellulose column, respectively. The violet arrow indicates the electrophoretic migration of the 6His-HpHyp1 protein.

The fractions eluted with 100 and 200 mM NaCl were pooled and submitted to further purification on an IMAC Sepharose (GE Healthcare) column (Fig. 53). Fractionation was carried out by stepwise elution with imidazole.

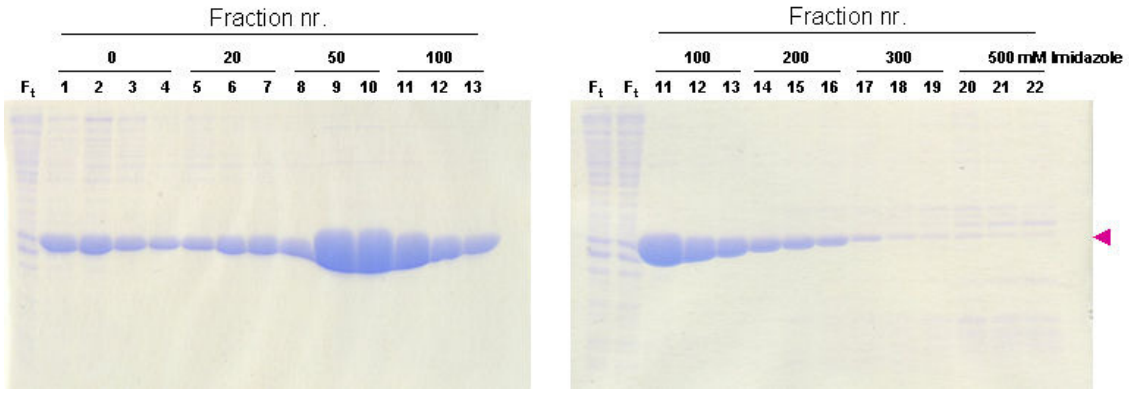


Figure 53. 15% SDS-PAGE analysis of the IMAC fractions of 6His-HpHyp-1 protein. Ft is the flow through of the IMAC Sepharose column. The violet arrow indicates the electrophoretic migration of the 6His-HpHyp1 protein.

The fractions eluted with 50 and 100 mM imidazole were pooled and the His tag was cleaved during an overnight digestion with His-tagged TEV protease at 4ºC in a dialysis bag, placed against 10 mM HEPES, pH 7.5. The TEV protease, the cleaved His tag and any uncleaved material were removed by a second round of purification on the IMAC Sepharose (GE Healthcare) column (Fig. 54). The flow through with the protein of interest was loaded again on the column for removal of any remaining unwanted material. The process was repeated and the flow through of all the steps was loaded on an SDS-PAGE gel for quality verification (labelled FT', FT'', and FT''', respectively). To assess the efficiency of the cleavage by the TEV protease, 500 mM imidazole were used to elute all material attached to the column. The yield of the recombinant protein after purification was about 37 mg per liter of liquid culture.

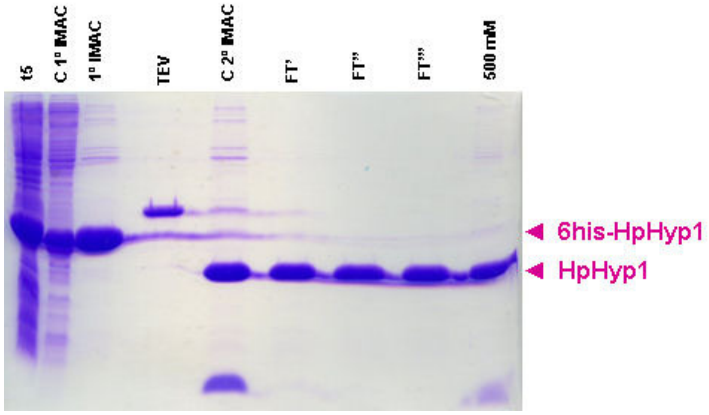


Figure 54. 15% SDS-PAGE analysis of TEV protease digestion, and a second IMAC chromatography fractions. The violet arrows indicates the electrophoretic migration of the 6His tagged HpHyp1 (top) and HpHyp1 protein (bottom).

III.4.2. Circular dichroism

CD experiments were carried out at far-UV region (205-265 nm). The HpHyp-1 protein was concentrated in 3 mM sodium citrate buffer, pH 6.3, and diluted just before measurement. The CD spectra were recorded at 6 µM protein concentration in 10 mM sodium phosphate buffer, pH 7.0. The scan speed was set to 100 nm/min and 5 scans were averaged. Spectra obtained at far-UV are presented without mathematical smoothing (Fig. 55).

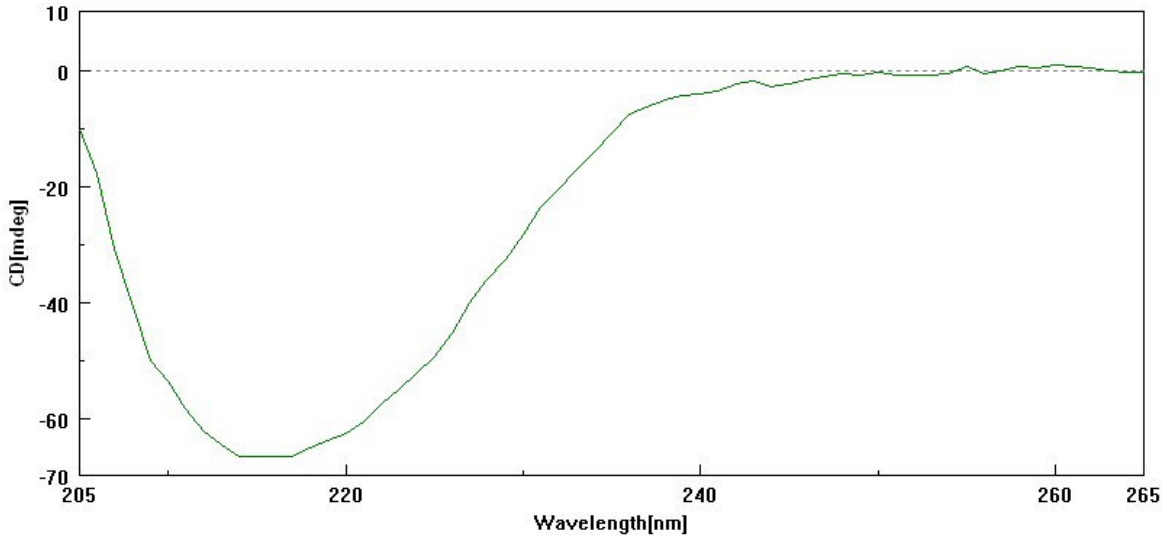


Figure 55. Unsmoothed far-UV CD spectrum of HpHyp-1 protein, recorded at 20ºC between 205 and 265 nm.

III.4.3. Thermal stability shift assay

Thermal stability shift assay was performed using the termofluor approach. The HpHyp-1 protein was concentrated in 3 mM sodium citrate buffer, pH 6.3, and diluted just before the measurements to approximately 2.5 mg/ml. The mixtures of 7.5 µl of 300x SYPRO Orange (Invitrogen), 12.5 µl of 2x test compound (different buffers), and 5 µl of 2.5 mg/ml protein were used for the measurements. Water was added instead of the test compound to the control samples. A blanc tube, 12.5 µl 300x Sypro Orange and 12.5 µl of water, was also added. The data are summarized in Fig. 56.

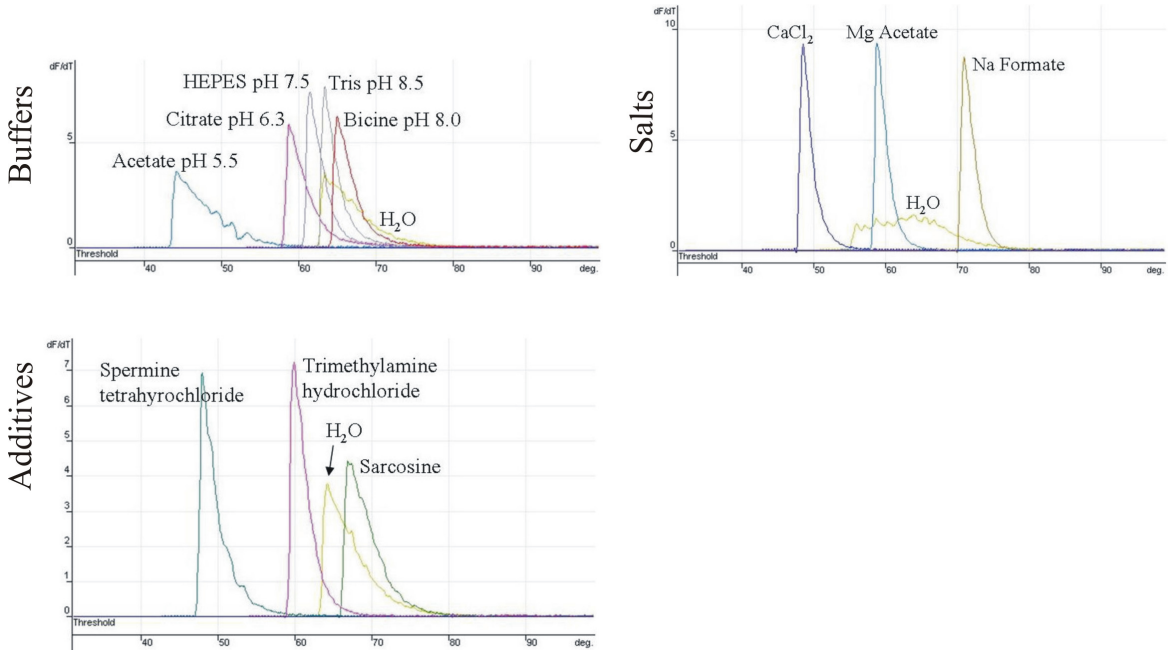


Figure 56. Thermal stability shift assay for the HpHyp-1 protein. Testing of different buffers, salts and additives. Only a small selection of the results is shown.

III.4.4. Isothermal titration calorimetry

The titration experiments were performed using isothermal titration calorimetry as described in Methods. Three cytokinins (kinetin, 2iP, CPPU) and rhein, an analogue of the HpHyp-1 protein substrate, emodin, were investigated (Table 12). The experiments were performed at 20ºC. Under the experimental conditions, HpHyp-1 did not show any binding of the ligands used.

Table 12. ITC runs for the HpHyp-1 protein with different ligands.

Ligand	HpHyp-1 concentration (mM)	Ligand concentration (mM)
Kinetin	0.19	0.33
Rhein	0.19	2.70
2iP	0.19	2.66
CPPU	0.19	23.99

III.4.5. Antifungal tests

The effect of the recombinant HpHyp-1 protein on the *in vitro* growth of pathogenic fungi was determined. Purified HpHyp-1 and controls (see Methods) were added to sterile paper discs. The fungal mycelial front was allowed to grow, and growth inhibition zones were investigated. Purified HpHyp-1 protein did not inhibit the growth of any of the fungi studied, namely *Colletotrichum lupini*, *Leptosphaeria maculans* and *Leptosphaeria biglobosa* (Fig. 57).

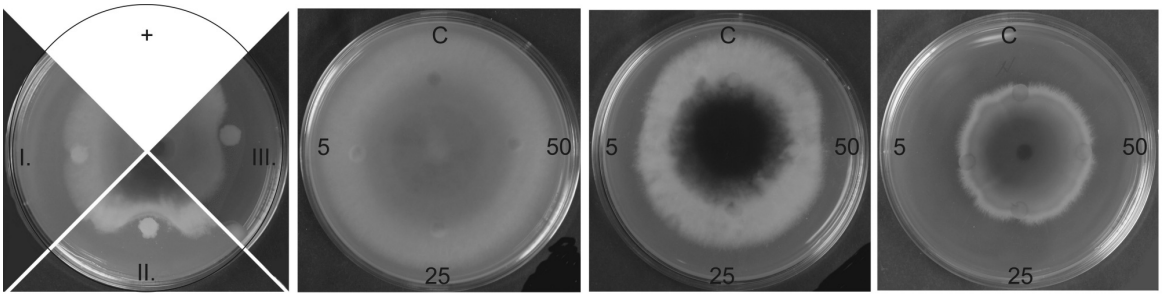


Figure 57. Antifungal activity of recombinant HpHyp-1. Effect of filter discs containing 5, 25 or 50 µg of protein, or buffer as a control (C), in the *in vitro* growth of *Colletotrichum lupini*, *Leptosphaeria maculans* and *Leptosphaeria biglobosa*.

III.4.6. Crystallization

Prior to crystallization, HpHyp-1 was concentrated to 6 mg/ml in 3 mM sodium citrate buffer, pH 6.3. Screening for crystallization conditions was performed using the hanging-drop vapor diffusion method at 20ºC by mixing 1.5 µl protein solution with 1.5 µl well solution. Initial screens included Crystal Screens I and II and PEG/ion screen from Hampton Research. Small crystals appeared after two days in the drop where the reservoir consisted of 0.2 M $MgCl_2$, 0.1 M Tris-HCl, pH 8.5, and 30% v/v polyethylene glycol 4000. Trials to improve the crystal size indicated the following composition of the well solution: 0.4 M NaCl, 0.1 M Tris-HCl, pH 8.5, and 30% v/v polyethylene glycol 4000, and the use of protein at very high concentration (55 mg/ml) (Fig. 58). Small crystals appeared after one day and reached their final dimensions of 0.10 x 0.02 x 0.02 mm in two weeks.

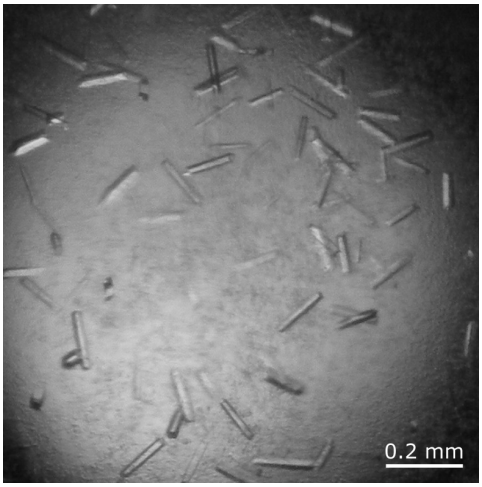


Figure 58. Single crystals of *H. perforatum* HpHyp-1 protein. The average dimensions of the crystals are 0.10 x 0.02 x 0.02 mm.

III.4.7. X-Ray data collection and processing of native data

Crystals were briefly soaked in a cryoprotectant solution obtained by mixing the well solution with 50% (v/v) polyethylene glycol 400 in a 1:1 ratio and were then flash-vitrified at 100 K under a stream of N_2 gas (Teng, 1990). X-Ray diffraction data for the crystals were measured at the 14.1 beamline of the BESSY synchrotron (Berlin, Germany) using a MAR 225 Mosaic CCD detector (Fig. 59). The crystal used for data collection was an uncut specimen. The dataset, collected to a maximum resolution of 1.73 Å, was measured in 360 0.5º oscillation steps (Dataset 1). The data were indexed, integrated, and scaled using HKL2000 (Otwinowski and Minor, 1997), producing a dataset that is 99.1% complete and characterized by an overall R_{int} of 0.070 (Table 13).

Another set of native X-ray diffraction data for the crystals was measured at the EMBL X13 beamline at the DORIS ring of the DESY synchrotron using a MARCCD 165 mm detector. The dataset, extending to a maximum resolution of 2.35 Å, was measured in 600 0.6º oscillation steps (Dataset 2). The data were indexed, integrated, and scaled using HKL2000 (Otwinowski and Minor, 1997), producing a dataset that is 95.0% complete and characterized by an overall R_{int} of 0.085 (Table 13).

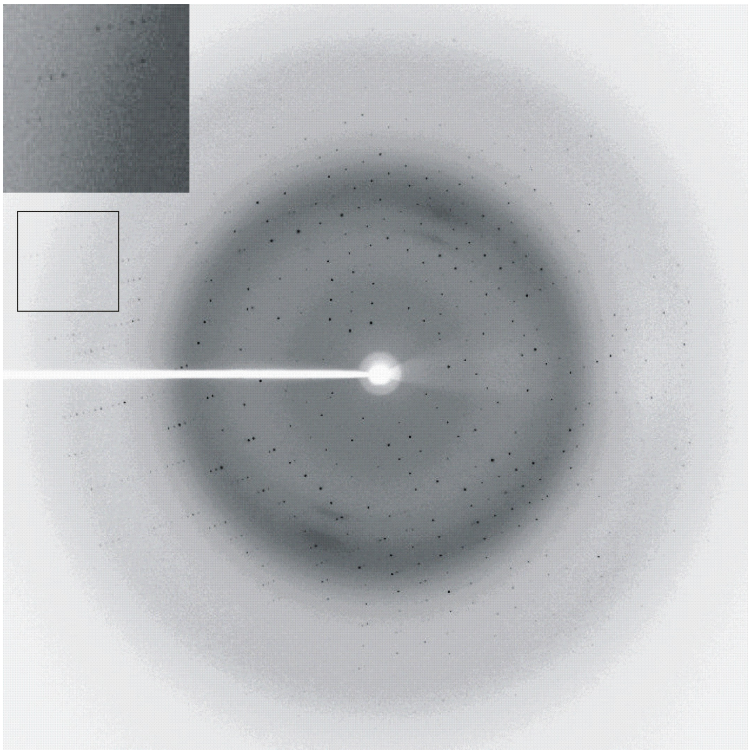


Figure 59. An X-ray diffraction pattern recorded for a single crystal of *H. perforatum* Hyp-1 protein using synchrotron radiation (0.5º oscilation). The edge of the detector (framed, inset) corresponds to a resolution of 1.73 Å.

III.4.8. Molecular replacement calculations

Attempts to solve the structure of the HpHyp-1 protein by molecular replacement using a number of programs, namely AMoRe (Navaza, 1994), PHASER (McCoy *et al.*, 2007), Molrep (Vagin and Teplyakov, 1997), and EPMR (Kissinger *et al.*, 1999), several search models from PR-10 coordinates available in the PDB (Berman *et al.*, 2000) (Table 1), and both native datasets, have been unsuccessful, although the target and probe molecules display amino acid sequence identity of up to 43% (Pru av 1). After the initial failure, truncated versions of the models with the removal of one or more loops, the helix α3, or side-chain truncation of non-conserved residues, were submitted to the MR process, although also without success. This failure may indicate that the HpHyp-1 protein does not have a canonical PR-10 fold or that its structural distortion is sufficiently high to prevent recognition by vector methods. The latter possibility is quite likely as it has precedent with other PR-10-like structures (Pasternak *et al.*, 2006, 2008). The problem lies in the C-terminal α-helix of PR-10 proteins, which shows very low sequence conservation and high degree of structural deformation.

Table 13. Summary of crystal data and data-collection statistics. Values in parentheses are for the last resolution shell.

	Dataset 1	Dataset 2
Space group	$P2_12_12_1$	$P2_12_12_1$
Unit-cell parameters (Å)	a = 37.5, b = 76.7, c = 119.8	a = 34.8, b = 75.6, c = 122.4
Resolution limits (Å)	30.0 – 1.73 (1.79 – 1.73)	50.0 – 2.35 (2.43 – 2.35)
Radiation source	BESSY 14.1 Berlin	X13 EMBL Hamburg
Wavelength (Å)	0.91841	0.80150
Temperature (K)	100	100
Mosaicity (°)	0.43	0.30
No. of measured reflections	247614	86392
No. of unique reflections	36647	24208
Redundancy	6.76	6.66
R_{int}	0.070 (0.496)	0.085 (0.327)
Completeness (%)	99.1 (91.6)	95.0 (79.4)
$\langle I/\sigma(I)\rangle$	24.9 (2.1)	26.2 (4.1)

III.4.9. X-Ray data collection and processing of MIR data

Several heavy atoms, listed in Table 14, were tried for an MIR approach. The crystals were tested for different soaking times in the mother liquor supplemented with 1-5 mM of a heavy-atom compound. The datasets were collected at the 14.1 beamline of the BESSY synchrotron (Berlin, Germany). All measurements were carried out under cryogenic conditions and, as for the native datasets, 50% (v/v) polyethylene glycol 400 mixed in a 1:1 ratio with the well solution was used as cryoprotectant. The quality of the crystals was initially tested by inspection of the diffraction images. The programs DENZO and SCALEPACK from the HKL2000 program package (Otwinowski and Minor, 1997) were used for indexing, integration and scaling of the diffraction data.

Table 14. Description of the heavy-atom derivatives used for MIR approach.

Heavy atom	Soaking conditions	Resolution [Å]	χ^2 #	Cell parameters [Å]	Anomalous signal
Au	75 min in 1mM $KAuCl_4$	2.29	30.5 (10.1)[2]	*a* = 34.8, *b* = 76.1, *c* = 121.3	-
Hg	4h in 1mM $HgCl_2$	2.37	31.0 (5.2)[1]	*a* = 37.5, *b* = 76.3, *c* = 119.4	-
Pt	8h in 5 mM K_2PtCl_4	2.19	8.6 (3.5)[1]	*a* = 37.6, *b* = 76.6, *c* = 119.9	-
Ta	2 weeks in 5mM $(Ta_6Br_{12})Br_2$	2.35	47.4 (2.3)[1]	*a* = 37.8, *b* = 76.5, *c* = 119.7	-

χ^2 of scaling between native ([1] dataset 1, and [2] dataset 2) and derivative datasets. The values in parentheses correspond to the last resolution shell.

Analysis of the cell parameters for the derivatives indicates that none of them diverges considerable from the native crystals. The two native data sets have different parameters allowing for a bigger range of cell parameter change for the derivatives. It turned out that the Au derivative has cell parameters more similar to the crystal used for Dataset 2 and all the others to Dataset 1. Nonetheless, the statistics proved that the crystals were non-isomorphous, as indicated by the large values of χ^2. The exception is the crystal soaked with Pt. The Pt derivative was tested to find the heavy-atom positions using the program SHELX (Sheldrick, 2008), but without success.

Part IV

Discussion

IV.1. DNA cloning

The DNA cloning process is not an error-free procedure. Each new construct should be carefully inspected, searching for DNA mutations, especially those that would result in a mutated protein. Some mutations may be desired and designed by the experimenter, either to study the importance of a particular amino acid for the protein function or for protein stability, but this is not the case in this particular project.

Sequence inspection revealed that the DNA construct encoding the PmPR-10.3.1 protein has five mutations, four of them resulting in alterations of the protein sequence. The protein mutations are at positions 10 (His10Asp), 35 (Leu35Phe), 69 (Pro69Ala) and 104 (Ile104Leu). Nevertheless, attempts to crystallize this protein have been continued because of the following reasons: (i) the same mutations are also found in the cDNA, (ii) the "new" amino acids are present in the same positions in some close homologues of the protein (Fig. 60) (Liu and Ekramoddoullah, 2004) and (iii) CD spectroscopy did not indicate any potential major problem with the protein fold (Fig. 42).

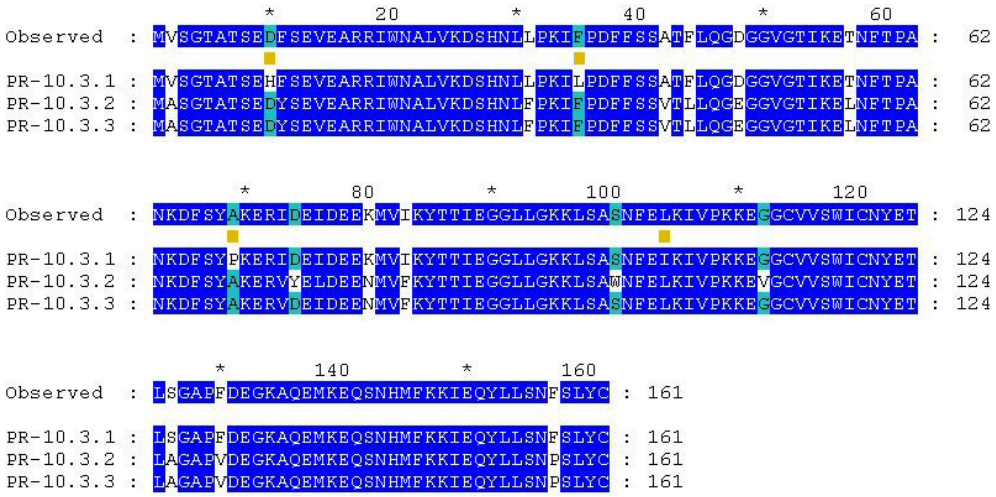


Figure 60. Sequence alignment of three members of the pine PR-10.3 subclass and the sequence identified for the new construct. Highlighted in yellow are the expected mutations of PmPR-10.3.1.

IV.2. Protein purification, modification and crystallization

Unfortunately, not all proteins included in this thesis could be produced in *E. coli* cells in soluble form. Two of them, namely the PmPR-10.3.1 and TcPR-10 proteins, were found in the insoluble fraction as inclusion bodies. Consequently, a refolding step was added to the purification protocol.

The *E. coli* expression system is one of the most widely used prokaryotic systems for genetic manipulations and for the industrial production of proteins of therapeutic or commercial interest. Compared with other established and emerging expression systems, *E. coli* offers

several advantages, including growth on inexpensive carbon sources, rapid biomass accumulation, amenability to high cell-density fermentations and a simple scale-up process. It is, however, not uncommon that overexpressed recombinant proteins fail to reach a correct conformation and undergo proteolytic degradation or associate with each other to form insoluble aggregates of non-native conformation, known as inclusion bodies (Baneyx and Mujacic, 2004). Inclusion bodies are formed intracellularly because of the aggregating characteristics of the protein or the inability of the cellular processes to ensure that the expressed polypeptide is soluble and correctly folded. Methods for preparation of denatured proteins involve isolation of the inclusion bodies, with some removal of contaminants, followed by solubilization using a concentrated solution of a chemical denaturant, typically urea or guanidinium chloride, which are also known as chaotropes because of their ability to disrupt the structure of liquid water. Refolding is then accomplished by controlled removal of excess denaturant. In most cases, this is allowed to occur in the presence of a suitable redox system and of other folding-promoting agents, according to one of the three principally different models: (i) dilution, (ii) solvent exchange by dialysis, diafiltration or size-exclusion chromatography, or (iii) reversible adsorption of the denatured proteins onto a solid support (Middelberg, 2002; Li *et al.*, 2004). Care must be taken to ensure that the protein resolubilize with the correct fold. This can be investigated either by CD studies (to confirm the tertiary structure fingerprint), biofunctional assays if any function is known for the protein, or ideally a combination of both methods.

For the PmPR-10.3.1 and TcPR-10 proteins recovered from inclusion bodies, CD spectra indicated that indeed the fold was correct. When the CD spectra for all the proteins under study (LlPR-10.2B, PmPR-10.3.1, TcPR-10, HpHyp-1) were superposed, it was easy to verify that the tertiary structure fingerprint was similar in all cases (Fig. 61).

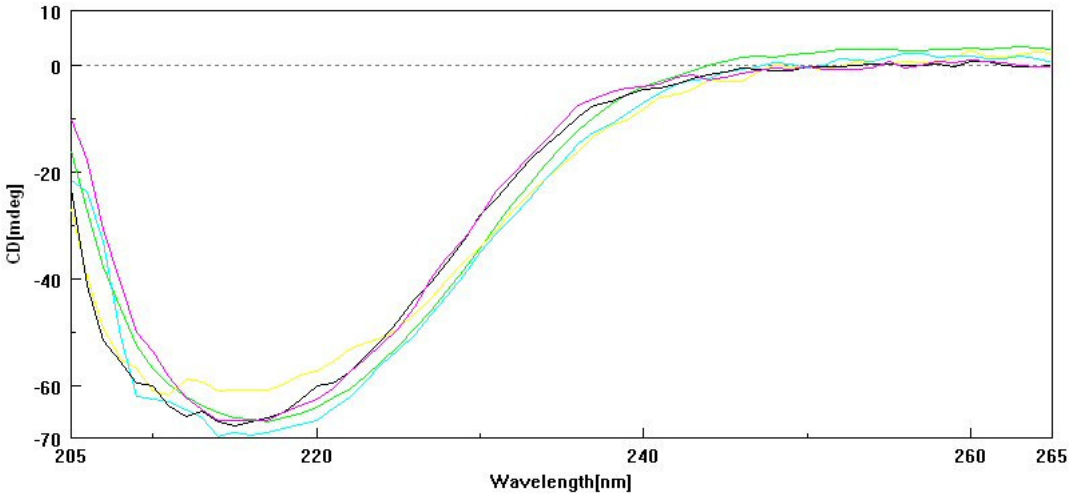


Figure 61. CD spectra for the proteins expressed in *E. coli* cells in soluble form: LlPR-10.2B (green), and HpHyp-1 (violet), and those expressed as inclusion bodies: TcPR-10 (black), and PmPR-10.3.1 in its monomeric (yellow) and tetrameric forms (light-blue).

In order to improve the purification of PmPR-10.3.1 and crystallization of TcPR-10, the proteins were subjected to chemical modifications. The amino acid sequence of PmPR-10.3.1 is exceptional within the PR-10 family as it contains three cysteine residues. Recombinant proteins with cysteine residues that have not been specifically oxidized, will invariably polymerize upon reconstitution. Without reduction, dimers and higher-order oligomers may form, and this covalent oligomerization will have an adverse effect on any crystallization attempt. By means of gel filtration, two different forms of the PmPR-10.3.1 protein could be isolated, namely a monomeric form and a tetrameric form. This fact can be related with possible intermolecular S-S bond formation, as treatment with DTT partially converts the tetrameric form to the monomeric one, whereas without DTT both forms interconvert into each other until an equilibrium is reached. One way of avoiding S-S related aggregation is to cap the cysteine residues by means of reduction and S-carboxymethylation. This protein modification did not prevent the tetramerization of PmPR-10.3.1, but it slowed down the equilibrium-forming process (Fig. 62). The CD spectra indicate that the two forms of the protein retain their tertiary structure fingerprint (Figs. 43 and 61).

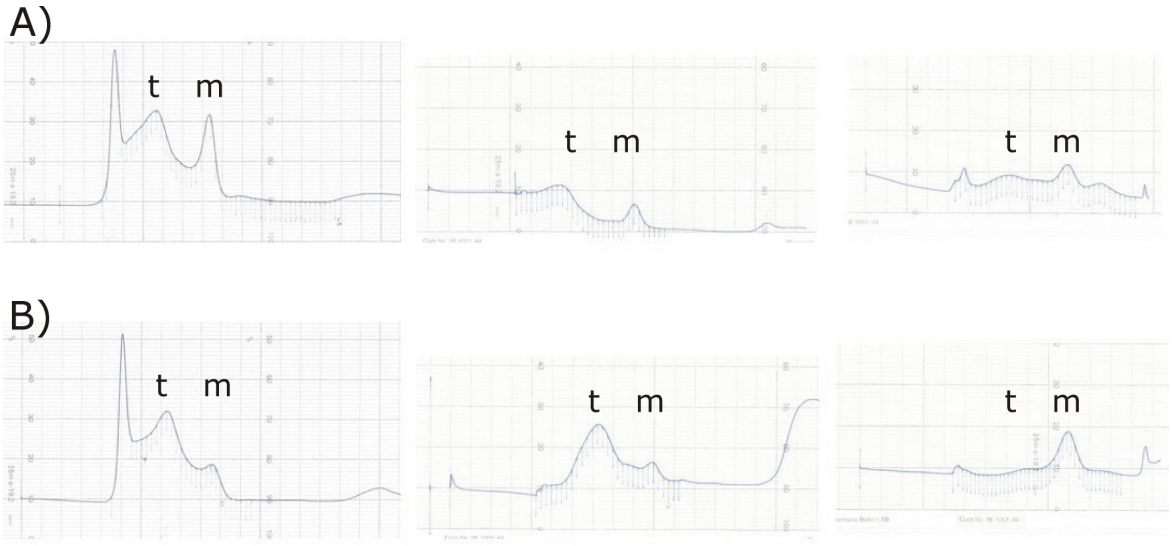


Figure 62. Size exclusion chromatography on a HiPrep S300 (GE Healthcare) column of the PmPR-10.3.1 protein. (A) Control. (B) After treatment with iodoacetic acid. Left to right: first run, second run of the tetrameric fraction, and second run of the monomeric fraction. The letters "t" and "m" indicate the tetrameric and monomeric forms, respectively.

In addition to manual screening for crystallization conditions, the TcPR-10 protein was submitted to an extensive high-throughput crystallization screen, performed automatically at the EMBL laboratory in Hamburg, Germany. The apo protein and its complexes with zeatin and CPPU were tested in 720 different crystallization conditions each, in the automatic approach

alone. Unfortunately, no well-diffracting crystals were obtained in any of the 2160 different conditions tested. The drops were individually inspected and some were selected for optimization. In a comprehensive optimization process, only one set of conditions (2.0 M NaCl, 0.1 M Na-Acetate pH 4.5) was found that produced diffracting protein crystals (Fig. 51). The diffraction quality was, however, very poor, precluding any attempt to solve the crystal structure of the protein. In order to overcome this bottleneck, the protein was subjected to reductive methylation, in which the primary amino groups of surface residues are modified to tertiary amines. The methylation of the protein was confirmed by mass spectroscopy (Fig. 63).

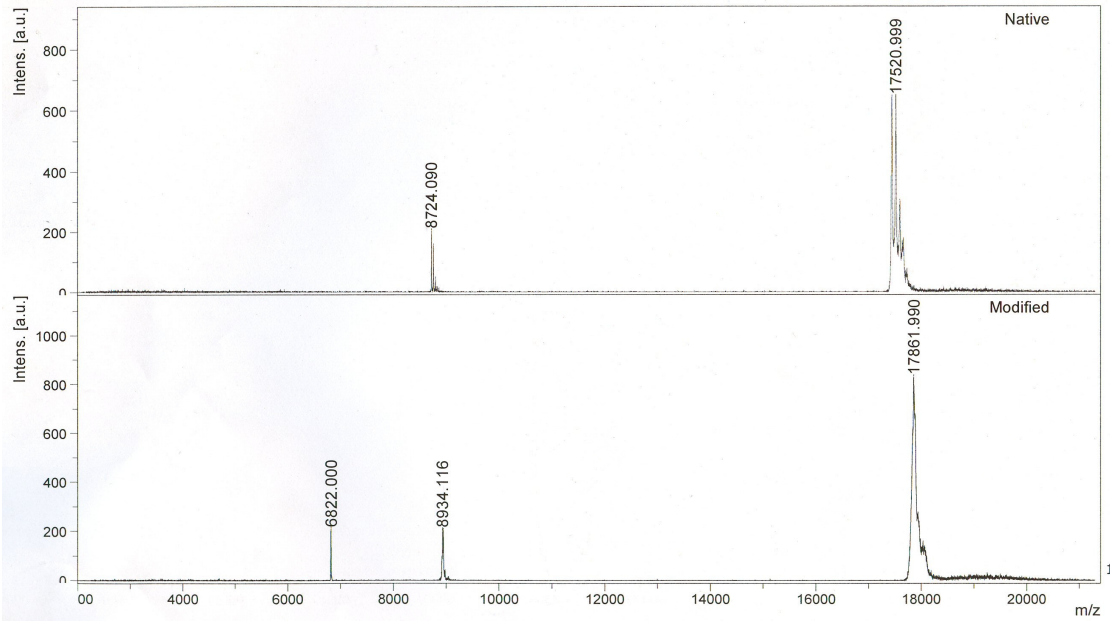


Figure 63. Mass spectra of the TcPR-10 protein; native (top) and reductively methylated (bottom).

The TcPR-10 protein has 16 lysine residues and obviously one terminal amino group (Fig. 64). The methylation reaction increased the protein mass by 341 Da, indicating that 12 amino groups were methylated (out of the total 17).

```
                    *         20          *         40          *         60
TcPR-10 : MGVTTYTQEFTCSVAPARMFKALILDSDNLIPKLMPKSIKNVELIHGDGGVGSIKQTNFPEG :  62

                   *         80          *        100          *        120
TcPR-10 : SHFKYLKNRTDALDVDNCVCKYTTIEGDVIGDKLESISYELKFEASGAGCVCKMTSHYHTKG : 124

                  *        140          *
TcPR-10 : DFVLKEEEIKAGKDQAVGMYKVVEEYLLANPNVYA : 159
```

Figure 64. Amino acid sequence of the TcPR-10 protein. The putative methylation sites are marked in yellow.

The modified protein crystallized at the same condition as the unmodified one. Inspection of the crystals under the polarizing microscope showed that the protein modification did not affect their appearance. Although the diffraction power of the newly obtained crystals has not yet been investigated, their similarity to the previous specimens may indicate that this approach did not improve crystal quality.

The four proteins under study (LlPR-10.2B, PmPR-10.3.1, TcPR-10, HpHyp-1) were extensively tested for crystallization conditions either in the apo forms or in complex with different ligands. In addition to the crystals described in the Results section, some other crystallization experiments also produced crystals (Fig. 65), however, their diffraction power has not yet been investigated. The crystals were grown by the hanging-drop vapor-diffusion method at room temperature. Each drop consisted of 1.5 µl protein solution and 1.5 µl reservoir solution equilibrated over 500 µl reservoir solution. Crystals of the LlPR-10.2B/2iP complex grew in two distinct crystallization conditions, namely in 1.2 Na-citrate, 0.1 M MES, pH 6.5 (Fig. 65A-i), and 1.6 M $CH_2(COOH)_2$, 0.1 M Tris-HCl, pH 8.5 (Fig. 65A-ii). The TcPR-10 protein was crystallized in its apo form (Fig. 65B-i), in complex with CPPU (Fig. 65B-ii), and in complex with zeatin (Fig. 65C-ii) in 0.2 M NaF and 20% w/v PEG 3350. In addition, crystals of the TcPR-10/zeatin complex appeared in 2.0 M NaCl, 0.1 M Na-acetate, pH 4.5, when the drop was supplemented with 0.2 µl 30% v/v ethanol (Fig. 65C-i).

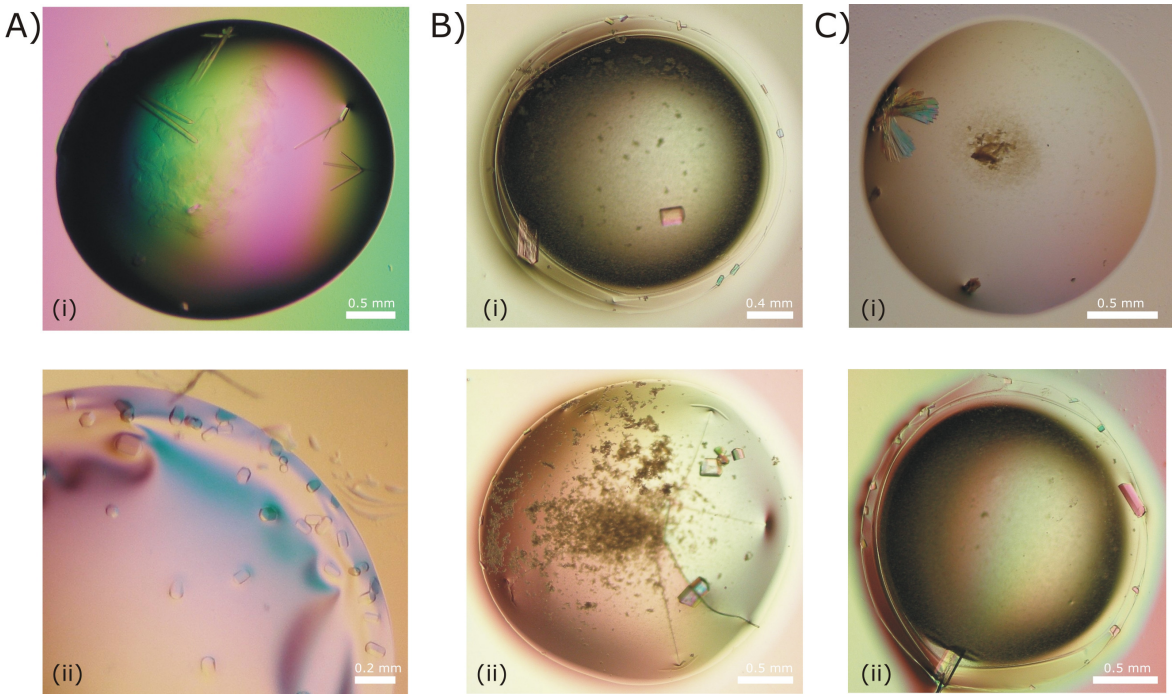


Figure 65. Crystallization of (A) LlPR-10.2B/2iP (i and ii), (B) TcPR-10, (i) apo, (ii) CPPU complex, and (C) TcPR-10/zeatin (i and ii).

IV.3. Thermofluor

Ericsson and co-workers (2006) have shown that the thermofluor method can be used as a fast tool for identifying the most stabilizing buffer for a protein. Additionally, the authors found a direct correspondence between the presence of additives that shift the transition midpoint in a thermofluor screen in the positive direction and an increased likelihood of protein crystallization. They also found that it is not so much an intrinsically high T_m, but rather the positive influence of some buffers/additives, that is related with a higher probability of obtaining crystals.

The crystals of the LlPR-10.2B protein in complex with zeatin and DPU had been obtained before the thermofluor assay was implemented. Nonetheless, it was interesting to observe that from a broad pH range tested, the protein was apparently more stable in buffers with pH between 5.5 and 6.8. This result is consistent with the condition in which the crystals of the complexes were obtained, namely in MES buffer, pH 6.5, for the LlPR-10.2B/zeatin complex, and citrate buffer, pH 6.0, for the LlPR-10.2B/DPU complex.

A similar correspondence was observed for the HpHyp-1 protein. In this case, the buffer screen did not indicate a narrow range of pH. Instead, the citrate, pH 6.3, HEPES, pH 7.5, Tris-HCl, pH 8.5, and bicine, pH 8.0, buffers gave the best results. The diffraction-quality crystals of this protein were obtained in the Tris-HCl buffer, pH 8.5.

Unfortunately, under similar condition the thermofluor method did not work for the PmPR-10.3.1 and TcPR-10 proteins.

IV.4. Ligand binding

Although the biological role of PR-10 proteins is still elusive, the abnormal size of the internal cavity and the structural similarities between PR-10 proteins, CSBP and the START domain strongly indicate a ligand binding role for these proteins. Additionally, several studies have been reported demonstrating the ability of PR-10 proteins to bind steroids, cytokinins, fatty acids and flavonoids (Mogensen *et al.*, 2002; Markovic-Housley *et al.*, 2003; Koistinen *et al.*, 2005).

Also the crystal structures presented in this work, show that LlPR-10.2B can serve as a ligand-binding protein. The structures provide evidence for the ability of the LlPR-10.2B protein to bind cytokinins, both from the natural (zeatin) and synthetic (DPU) groups. To better understand the interactions between the macromolecule and its small-molecule partners, binding studies using the ITC method have been performed. ITC measures the heat effects during molecular association, but since the thermal changes are very small, special care was taken to avoid side effects arising from dilution or simple mixing of the binding partners. To accomplish

this, the proteins were extensively dialysed and the ligands were dissolved in the dialysis buffer. This poses certain limitations on the experimental design, as cytokinins have very low solubility in polar solvents. Nevertheless, ITC was the method of choice for studying the binding characteristics, as it is the only direct method to determine binding stoichiometry. Under the experimental conditions with the limitation described above, data for four cytokinins were obtained (Table 2). Two of them represent the synthetic branch (DPU and CPPU) and the other two the natural branch, with one example of the isopentenyl N6-substitution (zeatin) and one of furfuryl N6-substitution (kinetin) at the adenine ring. The ITC data were analyzed by fitting independent- as well as cooperative multiple-site models to the raw data. Only the independent-site model provided meaningful results. The LlPR-10.2B protein has one binding site characterized by K_d in the micromolar range and different stoichiometry depending on the ligand molecule. Titration with *trans*-zeatin revealed a second binding site, characterized by K_d in the submillimolar range. CPPU and kinetin show clear interaction with the LlPR-10.2B protein; however, the thermodynamic parameters could not be determined accurately due to low solubility of both ligands.

The reported structures of LlPR-10.2B (Fernandes *et al.*, 2008 and the present work) and CSBP (Pasternak *et al.*, 2006) in complex with *trans*-zeatin, the present LlPR-10.2B/DPU structure, and the binding affinity studies for CSBP/cytokinin (Pasternak *et al.*, 2006) and LlPR-10.2B/cytokinin complexes (present work) give a unique chance to understand the interaction of cytokinin molecules with PR-10 and PR-10-like proteins. A comparison of the thermodynamic parameters of LlPR-10.2B with those obtained for CSBP by Pasternak and co-workers (2006) reveal interesting differences. Concerning *trans*-zeatin binding, the binding site characterized by K_d in the submillimolar range observed for LlPR-10.2B shows a similar K_d value (193 μM) as that obtained for CSBP (106 μM). However, the stoichiometry is very different, with ratios of 8:1 and 1:1 for LlPR-10.2B and CSBP, respectively. On the other hand, the binding site characterized by K_d in the micromolar range observed for LlPR-10.2B, was not observed for the CSBP complex with *trans*-zeatin. The variation of the binding characteristic observed for the LlPR-10.2B protein and CSBP may be reflecting large/significant dissimilarity of the binding cavity of the two proteins. The binding pocket of the LlPR-10.2B protein has a volume of 3600-4500 $Å^3$, while in CSBP it is 1100-1600 $Å^3$, as measured by the SURFNET program (Laskowski, 1995). The CSBP/DPU complex was not investigated, so no comparison can be made. Interestingly both proteins interact with kinetin and CPPU, suggesting that also a CSBP/DPU complex is feasible.

The crystal structure of the LlPR-10.2B/DPU complex show four DPU molecules inside the protein cavity, however, the ITC data show a 1:1 stoichiometry. This 1:1 ratio may be

explained by the crystal structure of the complex where only one DPU molecule forms a hydrogen bond, albeit through a water molecule, with the protein. All the other DPU molecules are attached to the protein via a weak network of van der Waals interactions. Probably the DPU molecule involved in the hydrogen bonding, DPU3 in the crystal structure, is the only one contributing to the observed enthalpy change of -9.8 kJ/mol upon titration of LlPR-10.2B.

The discrepancy in the number of bound ligand molecules observed in the crystallographic complexes and in the ITC experiment, arises most probably from the weak ligand binding of the studied molecules. Thus, buffer composition (mainly pH, ionic strength, cryoprotectant, organic solvents, etc.) and relative concentration of both partners used in the crystallization process and in *in vitro* binding studies, can influence the amount of the complexed species that could be observed. As the biological function of the PR-10 proteins is still not known, it is hard to speculate which technique is better for describing the real phenomenon. The main purpose of the binding studies, both through complex structure determination and ITC measurements, was to investigate what kind of ligands can be bound by the LlPR-10.2B protein and this goal has been successfully achieved.

ITC experiments were also performed for the TcPR-10 protein. Only one run was carried out and the ligand investigated was zeatin (Fig. 47). The results show that TcPR-10 also binds zeatin with affinity in the micromolar range. Since the structure of the TcPR-10 protein is not known yet, it is not possible to comment on the relation between the affinity values and the cavity volume.

Recently, some new catalytic functions for the PR-10 proteins have been proposed. According to these speculations the proteins could act as enzymes producing secondary metabolites such as (S)-norcoclaurine (Samanani and Facchine, 2002) and in the biosynthesis of hypericin from emodin (Bais *et al*., 2003). The three-dimensional structures for these proteins are still unknown, but recent reports showed that diffracting crystals could be obtained for both proteins (Pasquo *et al.*, 2008; Fernandes *et al.*, 2008b and present work). These proteins were classified as members of the PR-10 family based on sequence similarities. It may be expected that they have similar cavity as all the known structures of the whole family. In order to explore this point, the HpHyp-1/cytokinin binding ability has been studied (Table 12). Additionally, rhein, an analogue of the HpHyp-1 protein substrate emodin, was also investigated (Table 12). No binding constants could be determined for the HpHyp-1 protein. This indicates that the HpHyp-1 protein does not bind the studied ligands under the experimental conditions used.

IV.5. Antimicrobial activity

It has been shown that many PR proteins have antifungal or antibacterial activities (Hejgaard *et al.*, 1992; Yun *et al.*, 1996; Van Loon and Van Strien, 1999 and references therein), but till recently, the PR-10 family was not included into this group. PR-10 antifungal activity was shown for the first time in 2002 for ocatin, a PR-10 homologue from the Andean tuber crop, oca (*Oxalis tuberose* Mol.) (Flores *et al.*, 2002), later for the hot pepper (*Capsicum annuum*) CaPR-10 protein (Park *et al.*, 2004), and recently for the yellow-fruit nightshade (*Solanum surattense*) SsPR-10 (Liu *et al.*, 2006) and the peanut (*Arachis hypogaea*) AhPR-10 proteins (Chadha and Das, 2006).

The antifungal activity for three proteins included in this study, namely LlPR-10.2B, TcPR-10, and HpHyp-1, has been investigated. A selection of fungi was used, namely *Colletotrichum lupini*, *Leptosphaeria maculan* and *Leptosphaeria biglobosa*. The fungus *Colletotrichum lupini* is specific for lupine plants and infects the leaves. *Leptosphaeria maculan* and *Leptosphaeria biglobosa* are similar fungi that are specific for oilseed rape (*Brassica napus*) and also infect the leaves. *L. biglobosa* is, however, less aggressive then *L. maculan*. Antifungal activity of the proteins was also tested against another legume fungus, *Phoma medicaginis*, which is specific for *Medicago sativa* and also infects the leaves. Due to a very slow growth of this fungus and to the fact that all the antifungal tests have been performed in the same way, namely by putting protein-soaked filter papers 1 cm away from the fungus front (see Methods section), no conclusion could be drawn in this case, since the fungus growth did not reach the filter paper. None of the proteins investigated showed any inhibition effect on the *in vitro* growth of the fungi analyzed.

Apart from checking the antifungal activity of AhPR-10 for the native protein, Chadha and Das (2006) also investigated the influence of conserved residues in the protein sequence on the antifungal activity. They found that the mutations Phe148Ser and His150Gln caused partial loss of antifungal activity and that the Lys54Asn mutation almost abolished such activity. The authors also showed a link between the antifungal activity and ribonuclease activity, which was also detected in the AhPR-10 protein. Remarkably, the authors reported that the Lys54Asn mutant of the protein had almost no RNase activity and did not inhibit the growth of the fungus, despite being able to penetrate it. Apparently, internalization is mandatory but not sufficient for the inhibition of fungal growth.

Comparing the amino acid sequences of the proteins reported to have antifungal activity, and highlighting the residues implicated in such activity, one notes that they are also conserved in the LlPR-10.2B, TcPR-10, and HpHyp-1 proteins. Glu148 (LlPR-10.2B sequence numbering) is conserved in all sequences except in the AhPR-10 protein where it is replaced by

phenylalanine. Similarly, Tyr150 is replaced in the AhPR-10 and SsPR-10 proteins by histidine (Fig. 66).

The amino acid sequence alignment shows that the proteins under investigation (LlPR-10.2B, TcPR-10, and HpHyp-1) possess the residues highlighted to be responsible for the antifungal activity. However, no antifungal activity could be observed *in vitro* for them. A possible explanation could be connected with structural differences that would render some of the PR-10 members inactive as antifungal agents. Although this speculation cannot be verified directly by protein structure comparisons, as no structural data are available for PR-10 proteins with documented antifungal activity, it is remarkable that vector method are sometimes unable to "recognise" the PR-10 fold. This failure was observed for the CSBP protein (Pasternak *et al.*, 2006, 2008) and has been encountered in the present Hyp-1 case. It was already observed by Biesiadka *et al.* (2002) that the long and elastic C-terminal helix α3, located in the groove created by the twisted β-sheet of the PR-10 fold, displays various degrees of bending and variable fit in the groove as it can slide along its general direction, or undergo bending and other deformations.

```
                    *        20         *        40         *        60
Ocatin      : MGVFVFEDEITTTISPTRVFDSFVN-ADVVFLKVAAEHIKSVETLEGDGGVGTIRKIVFHEG-- :  61
CaPR10      : MGAYTFTDKSTASVAPSRLFKALVIDFNNLVSKLAP-DVKSIENVEGDGGAGTIKKMTFVDG-G :  62
SsPR10      : MGVNTYTHESTTSISPSRLFKALVLHFDNLVPQVVPHEVKNIETIEGDGGAGSIKQMNFVEG-G :  63
AhPR10      : MGVHTFEEESTSPVPPAKLFKATVVDGDELTPKLIP-AIQSIEIVEGNGGPGTVKKVTAVED-G :  62
LlPR-10.2B : MGVFTFQDEYTSTIAPAKLYKALVTDADIIIPKAVE-TIQSVEIVEGNGGPGTIKKLTFIEG-G :  62
PmPR-10.3. : MVSGTATSEDFSEVEARRIWNALVKDSHNLLPKIFPDFFSSATFLQGDGGVGTIKEINFTPANK :  64
TcPR-10     : MGVTTYTQEFTCSVAPARMFKALILDSDNLIPKLMPKSIKNVELIHGDGGVGSIKQINFPEG-S :  63
Hyp-1       : MAAYTIVKEEESPTAPHRLFKALVLERHQVLVKAQPHVFKSGEIIEGDGGVGTVKEITFVDG-H :  63

                         *        80         *       100         *       120
Ocatin      : HGGYIKQKIEVVDKENLQYNDSVIEGDAIVGSIEKILNENKIIPNADGGCTVKSKSTFYTSDES : 125
CaPR10      : PIKYMKHKIHVIDEKNLVTKYSLIESDVLENKAESVDYDGKFEASADGGCVCTTVTVYNTIGDY : 126
SsPR10      : PIKYLKHKIHVIDDKNLETKYSLIEGDVLGDKIESINYDIKFEASGDGGCVCKTTTEYHTKGDH : 127
AhPR10      : KTSYVLHKIDAIDEATYTYDYTISGGTGFQEIIEKVSFKTKL-EAADGGSKIKVSVTFHTKGDA : 125
LlPR-10.2B : ESKYVLHKIEAIDEANLGYNYSIVGGVGLPDTIEKISFETKLVEGANGGSIGKVTIKIETKGDA : 126
PmPR-10.3. : DFSYAKERIDEIDEEKMVIKYTTIEGGLLGKKLSASNFELKIVPKKEGGCVVSWICNYETLSGA : 128
TcPR-10     : HFKYLKNRTDALDVDNCVCKYTTIEGDVIGDKIESISYELKFEASG-AGCVCKMTSHYHTKGDF : 126
Hyp-1       : PLTYMLHKFDEIDAANFYCKYTIFEGDVLRDNIEKVVYEVKL-EAVGGGSKGKITVSYHPKPGC : 126

                      *       140         *       160
Ocatin      : EIPAQILAEG-KEKRLGMFKATEAYLLANPEAFN : 158
CaPR10      : VVTEEEHNVH-KEKANDLLKAIEAYLLANPSVYV : 159
SsPR10      : VITEEEDNVG-KDKAINLFKAVEAHLLANPSVYA : 160
AhPR10      : PLPDEFIKMSTKSQESHAMRVMFVH--------- : 150
LlPR-10.2B : QPNEEEGKAA-KARGDAFFKAIESYLSAHPDYN- : 158
PmPR-10.3. : PFDEGKAQEM-KEQSNHMFKKIEQYLLSNFSLYC : 161
TcPR-10     : VLKEEEIKAG-KDQAVGMYKVVEEYLLANPNVYA : 159
Hyp-1       : TVNEEEVKIGEKKAYEFY-KQVEEYLAANPEVFA : 159
```

Figure 66. Sequence alignment of PR-10 proteins with antifungal activity, and the proteins under investigation in this thesis.

IV.6. Comparison of the two *Lupinus luteus* PR-10.2B structures

A structural superposition of the LlPR-10.2B complexes studied in this thesis (with trans-zeatin and DPU) reveals that while they have the same overall fold, their Cα traces show small but significant differences (Fig. 67). The r.m.s.d. between their coordinates display a value of 0.77 Å. The major folding differences are seen in loops L7 (maximum deviation of 13.5 Å for Cγ2 atoms of Leu90; 9.3 Å for Cα of Gly89) and L9 (maximum deviation of 11.1 Å for Nζ atoms of

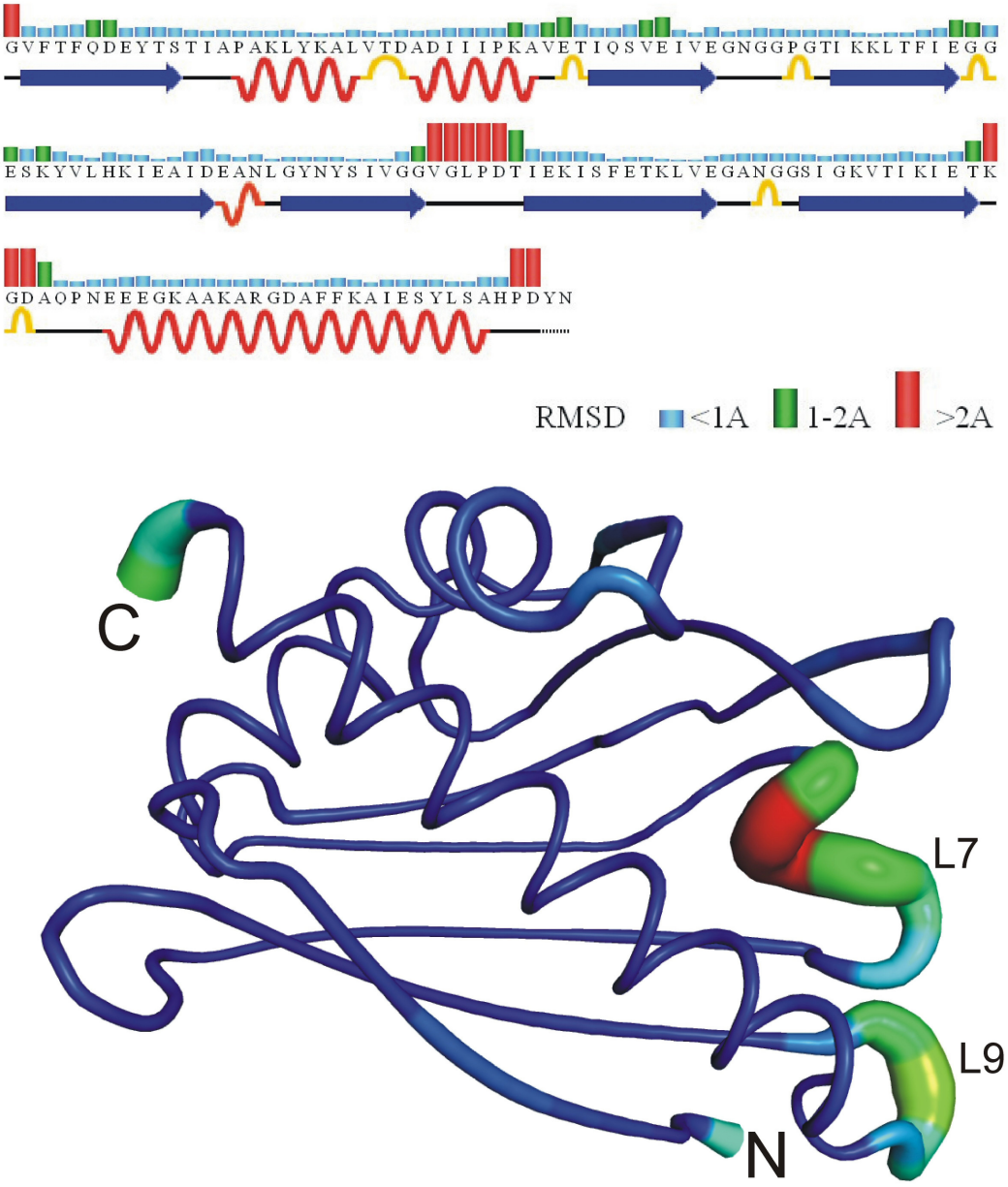


Figure 67. Representation of r.m.s. deviations between the Cα atoms of the LlPR-10.B complexes with zeatin and DPU. (top) The protein sequence with r.m.s.d. bars calculated with RMSDcalc (Claude *et al.*, 2004) and the DSSP-assigned secondary structure (Kabsch and Sander, 1983). (bottom) A cartoon "sausage" representation of the LlPR-10.2B protein colored according to the r.m.s.d. values as in the "bar" plot.

Lys122; 4.7 Å for Cα of Gly123). The conformational change of loop L9 is evidently caused by the sodium coordination in this area in the LlPR-10.2B/DPU structure (Na2). The same sodium cation is also responsible for pushing the positively charged N-terminus away from its typical β-sheet association with strand β7. Such repulsion, however, does not create any major disorder of the N-terminus as the residues have excellent definition in the electron density.

The large conformational re-shaping of loop L7 cannot be explained by the different packing modes of the two crystal structures (LlPR-10.2B/DPU, $C222_1$; LlPR-10.2B/zeatin, $P6_5$) since in both cases the lattice interactions of this loop are weak. In the LlPR-10.2B/DPU complex only one salt bridge is observed between Asp92-Oδ2 and Lys20-Nζ, and in the LlPR-10.2B/zeatin complex only one hydrogen bond is formed between Gly89-N and Val2-O. Therefore, it must be concluded that it is the ligand cargo in the binding cavity that influences the shape of loop L7.

A more revealing than a simple Cα alignment in this case of identical ligand-binding structures is an all-atom superposition of the protein scaffolds. Such a superposition, calculated in Lsqkab (Kabsch, 1976) is characterized by an r.m.s. deviation of 1.8 Å. The relatively large value of the r.m.s.d. for the LlPR-10.2B/zeatin-LlPR-10.2B/DPU pair is directly correlated with a significant change of the volume of the internal cavity. Calculations made in the SURFNET program (Laskowski, 1995) reveal that the cavity of the LlPR-10.2B/DPU complex is reduced by 900 $Å^3$ in comparison to the LlPR-10.2B/zeatin complex. Superposition of both complexes shows that the most affected part of the cavity is located close to loop L7 (Fig. 67).

The different stoichiometry, as well as orientation and binging mode of the ligand molecules in the LlPR-10.2B complexes result in a rearrangement of the side chains of the protein residues pointing into the interior of the cavity. Some of the side chains are only shifted in the same way as the Cα atoms, in concert with the rearrangement of the protein backbone, but some have a clearly different conformation in both complexes. The following residues: Asp7, Tyr9, Leu22, Phe57, Glu59, Lys64, Val66, Tyr80, Tyr82, Leu103, Ile119 and Phe142 show only minimal changes, while the residues Lys53, Ile84, Ile117, Glu131, Arg138 and Phe143 show major conformation changes (Fig. 68). The most interesting aspect of the cavity rearrangement is, however, the number and identity of the residues modeled with double conformation. Comparing only the residues with side chains pointing into the cavity, the LlPR-10.2B/zeatin complex has one residue with double conformation (Val66) and the LlPR-10.2B/DPU complex has two such residues: Leu55 and His68. The disorder of the cavity-forming residues is especially remarkable when one considers the significant ligand cargo of these complexes.

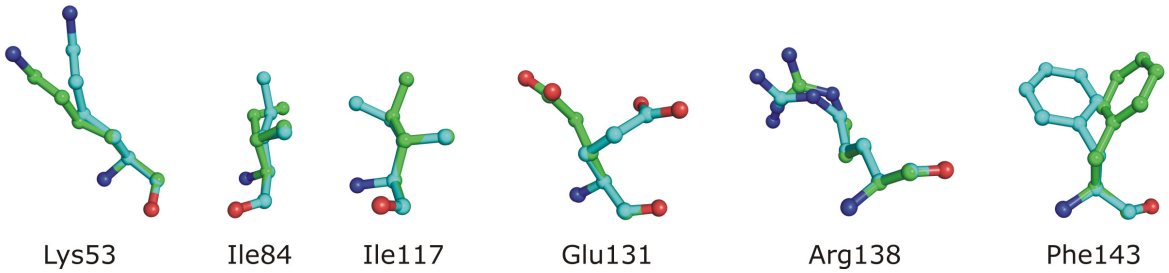


Figure 68. Superposition of the amino acids pointing into the cavity of the LlPR-10.2B molecule, with the most pronounced side chain variability. Residues from the LlPR-10.2B/zeatin complex are colored with aqua blue C atoms, and residues from the LlPR-10.2B/DPU complex with green C atoms. The structural superposition was calculated using the main-chain atoms, and the orientation shown in the figure is the one that best illustrates the conformational differences.

Another interesting observation comes from a comparison of the average volume "used" by individual atoms present in the cavity (ligand and water molecules) of the two complexes. In the LlPR-10.2B/zeatin and LlPR-10.2B/DPU complexes, those volumes are about 62 $Å^3$ and 54 $Å^3$ (excluding hydrogen atoms), respectively.

IV.7. Comparison of the LlPR-10.2B complexes with other PR-10 homologues

The amount of crystallographic information about PR-10 proteins has been accumulating rather quickly in the recent years. In addition to the present structures, three other yellow lupine homologues have been described, as well as several structures of the Bet v 1 protein, and one structure of the distant CSBP homologue from mung bean. Taking into account that in several cases more than one protein molecule was found in the asymmetric unit (e.g. LlPR-10.1B - 2, LlPR-10.2A - 2, CSBP - 4), the amount of structural information is quite significant. Although all the crystallographic models of PR-10 proteins share the same canonical fold, their superpositions reveal very significant structural differences (Table 15). When analyzed in detail, the differences are located mostly at the C-terminal helix α3, which displays different axial shifts as well as a variable degree of deformation in the central part and in its N-terminal connection with loop L9. Also, the volume of the internal cavity, formed with the participation of helix α3, shows remarkable variability.

The cavity volume is determined not only by the variability of the α3 helix sequence but also by its geometrical deformation, such as the strong inward kink reported for the LlPR-10.2A protein (Fig. 69). Until now, no report is available about structural characterization of the same protein with and without a natural ligand. The nearest such case is available for two homologous yellow lupine LlPR-10.2 proteins (91/96% sequence identity/similarity), one (2B) in complex

Table 15. R.m.s. deviations (Å) between superposed Cα atoms (number of pairs in parentheses) of PR-10 homologues, calculated in ALIGN (Cohen, 1997). 2qim and 3e85 are the PDB codes of the crystal structures of LlPR-10.2B in complex with zeatin and DPU, respectively. The codes 1A, 1B, 2A, correspond to yellow lupine homologues (respective PDB codes 1icx, 1ifv, 1xdf). 1bv1 and 1fm4 are the PDB codes of the crystal structures of the birch pollen protein Bet v 1*a* and Bet v 1*l*/deoxycholate complex, respectively. CSBP designates the cytokinin-specific binding protein from mung bean (2flh). If present, a letter after a hyphen designates the protein molecule in the asymmetric unit. The number below the protein code at each column head gives the volume of the internal cavity (Å^3), calculated in SURFNET (Laskowski, 1995).

	2qim 4509	3e85 3609	2A-A 2223	2A-B 1783	1A 3486	1B-A 3009	1B-B 3144	1bv1 3567	1fm4 3903	CSBP-A 1414	CSBP-B 1154	CSBP-C 1130	CSBP-D 1613
2A (A)	2.13 (145)	2.06 (143)	-	0.58 (147)	1.42 (145)	1.28 (149)	1.25 (150)	1.39 (133)	1.57 (134)	1.40 (118)	1.46 (126)	1.74 (132)	1.62 (126)
2A (B)	2.07 (144)	2.19 (145)		-	1.19 (138)	1.29 (149)	1.33 (148)	1.55 (135)	1.69 (136)	1.54 (122)	1.55 (129)	1.60 (125)	2.03 (136)
1A	2.07 (145)	1.97 (141)			-	1.18 (147)	1.41 (150)	1.40 (135)	1.64 (141)	1.66 (131)	1.67 (138)	1.76 (134)	1.73 (137)
1B (A)	1.77 (146)	1.88 (145)				-	0.36 (148)	1.55 (143)	1.72 (145)	1.40 (131)	1.39 (138)	1.53 (133)	1.53 (135)
1B (B)	1.91 (149)	1.96 (146)					-	1.50 (141)	1.73 (144)	1.47 (132)	1.37 (137)	1.53 (131)	1.54 (134)
1BV1	1.39 (146)	1.48 (145)						-	0.60 (140)	1.46 (126)	1.41 (141)	1.73 (134)	1.66 (133)
1FM4	1.35 (144)	1.28 (142)							-	1.66 (132)	1.61 (142)	1.70 (137)	1.70 (136)
CSBP (A)	1.76 (131)	1.86 (135)								-	0.53 (141)	0.62 (137)	0.31 (137)
CSBP (B)	1.68 (137)	1.68 (140)									-	0.43 (145)	0.42 (141)
CSBP (C)	1.74 (132)	1.74 (136)										-	0.54 (144)
CSBP (D)	1.73 (132)	1.72 (135)											-

with the plant hormone *trans*-zeatin (PDB code 2qim; described in this thesis), and one (2A) in its apo form (1xdf). These two almost identical protein sequences enclose very different cavities (Table 15), depending on the presence of the zeatin ligands. Also, the Cα r.m.s. deviation between these two homologues is the second highest (2.1 Å) in Table 15. This large deviation results not only from the kink of helix α3 in LlPR-10.2A (Fig. 69) but also from a rearrangement of the α2, β2, β3, β4, L3, L5, L7, L8 and L9 secondary structure elements. In consequence, the volume of the cavity has increased from 2000 Å^3 in LlPR-10.2A to 4500 Å^3 in the LlPR-10.2B complex with zeatin.

The residues involved in hydrogen bonds with zeatin in the LlPR-10.2B complex are all conserved in LlPR-10.2A (Fig. 70). Among the 14 residues involved in van der Waals contacts, only two are dissimilar, namely at position 10 (Tyr in LlPR-10.2B vs. Ser in LlPR-10.2A) and at position 57 (Phe vs. Leu). This strongly supports the hypothesis of helix α3 remodeling by the presence, number and identity of the ligands in the binding cavity. Additional support comes from multiple alignment of the PR-10 protein sequences, which shows that the most divergent

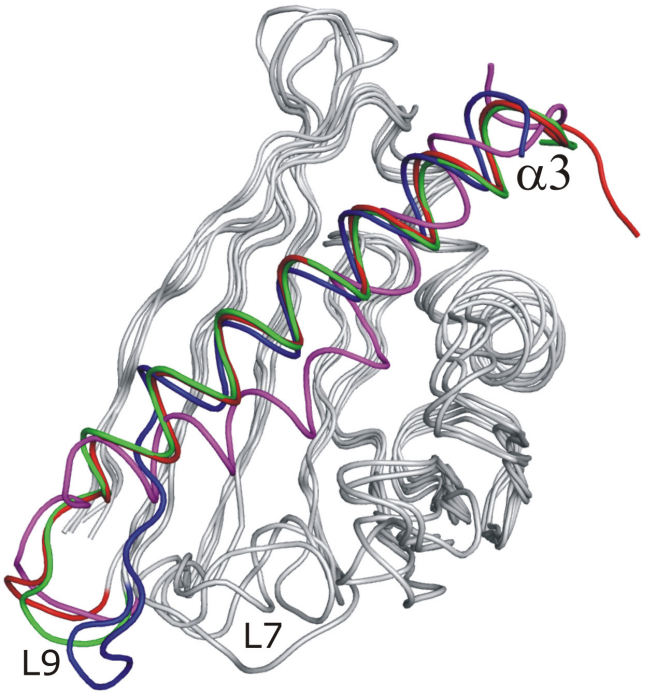


Figure 69. Superposition of PR-10 structures, calculated with ALIGN (Cohen, 1997). Color code: red, LlPR-10.2B/zeatin (PDB code 2qim); green, LlPR-10.2B/DPU (3e85); magenta, LlPR-10.2A molecule A (1xdf); blue, CSBP molecule B (2flh). The orientation emphasizes the differences in loops L7 and L9, and the C-terminal helix α3.

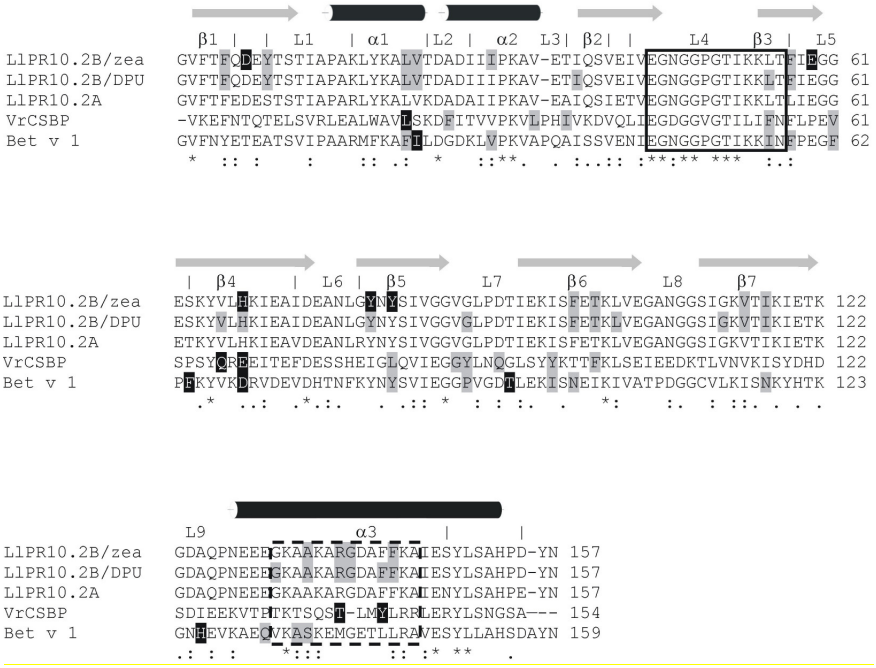


Figure 70. Sequence alignment of representative PR-10 proteins, calculated in ClustalW. Residues identical in all sequences are marked with *. The symbols : and . below the sequences indicate conservative and semi-conservative substitutions, respectively. The differences between the two members of the yellow lupine PR-10.2 subfamily are marked with |. The boxes outline regions that are most conserved (full line) or most divergent (dash line). The residues marked in gray interact with the ligands via van der Waals contacts, and those in black via hydrogen bonds. The secondary structure elements correspond to the LlPR-10.2B/zeatin structure.

element is located in the region of helix α3, especially in its N-terminal half. However, in the sequences of LlPR-10.2A and LlPR-10.2B this segment is perfectly conserved (Fig. 70), with only two conservative substitutions at the C-terminus of helix α3, namely at positions 149 (Ser in LlPR-10.2B vs. Asn in LlPR-10.2A) and 156 (Asp vs. Glu). It is important to note that the residues at these positions are not involved in ligand binding in the LlPR-10.2B/zeatin complex. The only substitution that could possibly disturb zeatin binding is at position 57, where a phenylalanine residue that (together with Arg138) forms stacking interactions with Zea3 is replaced in LlPR-10.2A by a leucine residue. A substitution that could indirectly affect ligand binding because of different side-chain electrostatics and size is Gly79→Arg, but the Arg residue is not pointing into the cavity. Since these substitutions are obviously not responsible for the difference in the cavity volume, and since the sequences around Phe142, where the α3-helix is kinked in the LlPR-10.2A structure, are exactly the same, it is logical to conclude that it is the presence and identity of the ligands in the binding cavity that modulate the conformation of the C-terminal helix α3.

The residues of LlPR-10.2B that are involved in *trans*-zeatin binding are not strictly conserved in all yellow lupine PR-10 homologues. Specifically, Asp7 is conserved in subclass 2 of LlPR-10 proteins but not in subclass 1, where it can be replaced by asparagine or glutamine. Similarly, Glu59 is conserved among subclass 2 homologues but in subclass 1 this position is occupied by a histidine residue. The remaining amino acids involved in hydrogen bonding with the ligands, His68, Tyr80 and Tyr82, are strictly conserved in the two subclasses. It is interesting to note that Zea1 interacts only with strictly conserved residues (His68, Tyr80), Zea2 with a conserved (Tyr82) and a variable (Asp7) residue, while Zea3 is attached at a variable site (Glu59).

Two other examples can be found of similar proteins where structures of the apo form and of a complex with a non-natural ligand exist: (i) the structures of the LlPR-10.2A protein (1xdf) and the second complex of the LlPR-10.2B protein described in this thesis where the protein was co-crystallized with DPU (3e85), and (ii) the apo form of Bet v 1*a* (1bv1) and its 94/97% identical/similar variant *l* in complex with deoxycholate (1fm4). Comparison (i) indicates that the Cα deviations between these two nearly identical homologues are among the highest (2.06 and 2.19 Å for molecule A and B, respectively) in Table 15. The difference in the volume of the cavity between the LlPR-10.2A molecule (2000 Å) and LlPR-10.2B in complex with DPU (3600 Å) is not as big as the one observed for the comparison between the apo LlPR-10.2A and the LlPR-10.2B/zeatin complex (2000 and 4500 Å, respectively). Nevertheless, the differences are remarkable, indicating that the presence and identity of the ligands in the binding

cavity modulate the structure of yellow lupine proteins of sub-class PR-10.2. This possible structural adaptation is reinforced by the comparison of the two complexes of LlPR-10.2B, with zeatin and DPU (see the previous sub-section of this discussion), were such adaptation is demonstrated. Comparison (ii), on the other hand, suggest that the Bet v 1 cavity with and without a ligand has similar volume (Table 15) and that, indeed, there is no major structural change upon deoxycholate binding.

Two additional comparisons can be made among the PR-10 structures. The first is between the two homologous yellow lupine protein of sub-class PR-10.1, namely LlPR-10.1A (1icx) and LlPR-10.1B (1ifv), which share 77/90% sequence identity/similarity. It indicates a PR-10 cavity with a rigid shape and volume, although, surprisingly, the Cα superposition reveals structural differences that are not insignificant (Table 15). The second comparison is among the four crystallographic copies of the CSBP/zeatin complex (2flh), which show similar conformations and cavity volumes (although significantly smaller than in the remaining, classic PR-10 molecules), despite the fact that only two of them bind their tandem zeatin molecules in the same way (in one case the "outer" zeatin ligand has a flipped orientation and in one complex it is absent altogether).

Another aspect in which the structures of PR-10 proteins show differences is in the presence, identity and coordination site of metal ions. The most interesting comparison is between the two crystal structures of the LlPR-10.2B protein in complex with cytokinins (this work). In the LlPR-10.2B/zeatin complex structure, there is only one metal cation, identified as calcium, coordinated close to loop L3 (Fig. 17), whereas in the LlPR-10.2B/DPU structure two metal cations have been modeled, close to loops L3 and L9 (Fig. 30). The metal cations in this complex have been identified as sodium. The difference in metal identity can be explained by the high sodium concentration in the crystallization buffer of the LlPR-10.2B/DPU crystals. On the other hand, calcium was not present in the crystallization buffer of the LlPR-10.2B/zeatin complex. Therefore, it must have been coordinated already at the protein expression stage, which indicates a relatively strong binding. Sodium was found in two other PR-10 structures, namely in LlPR-10.2A and in CSBP. In LlPR-10.2A, the sodium cation was found coordinated close to loop L3 (Pasternak *et al.*, 2005) and in CSBP close to loop L9 (Pasternak et al., 2006). In both structures the metal cation was found in only 50% of the molecules present in the respective asymmetric units, indicating that the binding constant is not very high. Moreover, the crystals where sodium cations were identified, are obtained in the presence of sodium citrate (1.2 – 1.4 M), which may bias the results since the citrate anion could chelate the true metal partner of the proteins.

IV.7. Cytokinin binding modes

A superposition of the two complexes of the LlPR-10.2B protein described in this thesis, with zeatin and with DPU, reveals that the ligand molecules occupy different positions in the protein cavity (Fig. 71A). There is no direct correspondence in the overlap between the ligands, but in general, DPU1-3 and zeatin 1-3 occupy similar spatial positions. DPU4 does not overlap with any zeatin molecule. An analysis of the LlPR-10.2B protein where the residues involved in interactions with the zeatin molecules in the LlPR-10.2B/zeatin complex, and with DPU molecules in the LlPR-10.2B/DPU complex, are highlighted, shows that these two different ligands have distinctive binding modes to the LlPR-10.2B protein (Fig. 70). As described in the Results, each zeatin molecule is bound to the protein by at least one direct hydrogen-bond and they are additionally stabilized by numerous van der Walls interactions (Fig. 20). The

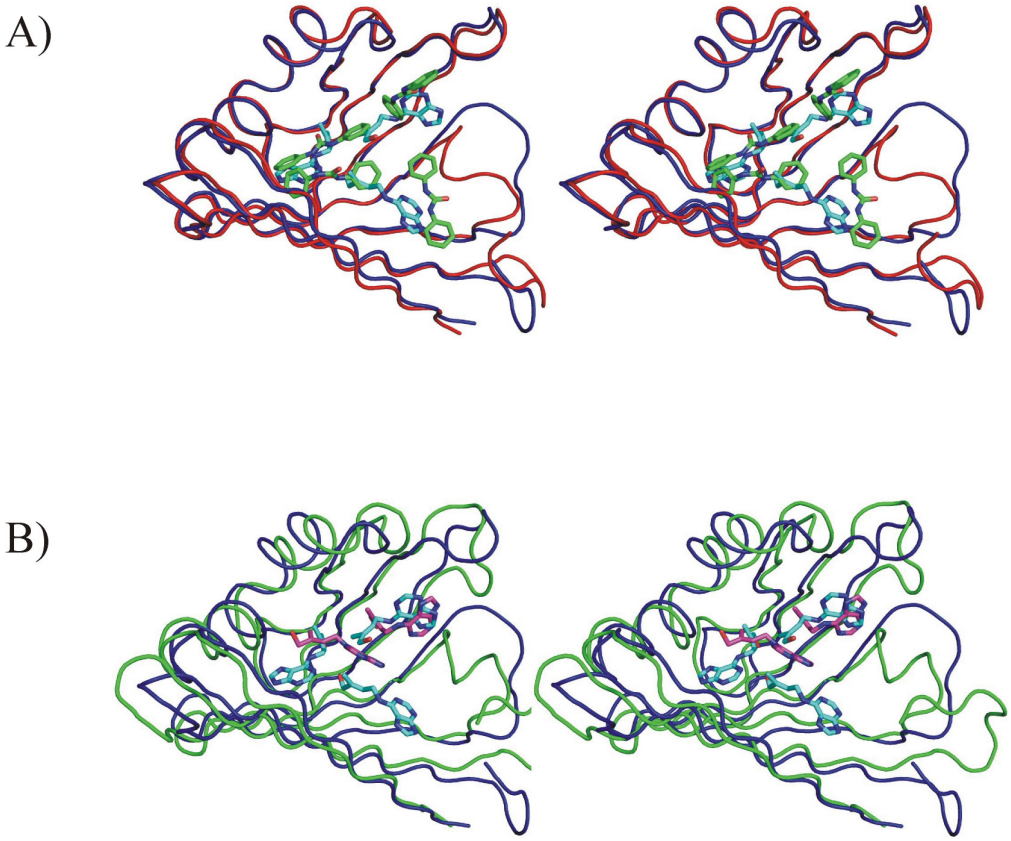


Figure 71. Stereoview of a superposition of PR-10 proteins, calculated with ALIGN (Cohen, 1997). (A) Superposition of the two LlPR-10.2B proteins described in this thesis. Color code: blue, LlPR-10.2B/zeatin complex, with C atoms of the zeatin molecules in blue; red, LlPR-10.2B/DPU complex, with C atoms of the DPU molecules in green. (B) Superposition of the LlPR-10.2B/zeatin and CSBP/zeatin (chain C) complexes. Color code: blue, LlPR-10.2B/zeatin, with C atoms of the zeatin molecules in blue; green, CSBP, with C atoms of the zeatin molecules in violet. The C-terminal helix α3 has been omitted for clarity.

DPU ligands on the other hand are anchored to the protein exclusively by van der Waals contacts (Fig. 34) with the exception of a single indirect hydrogen bond contact with the proteins (DPU3), which is mediated by a water molecule.

Remarkably, the binding mode of the zeatin ligands in the LlPR-10.2B complex is also different from the zeatin binding mode by CSBP (Fig. 71B). In contrast to the three ligand molecules found in the LlPR-10.2B/zeatin complex structure, in the CSBP/zeatin complex a maximum of two zeatin molecules ("inner" and "outer") were found in the protein cavity (Pasternak *et al.*, 2006). In addition to the surprisingly higher cytokinin binding ability of the LlPR-10.2B protein, the crystal structures show that the ligands are located differently in the cavity (Fig. 71B). When the Cα atoms of the LlPR-10.2B molecule are aligned with the four protein molecules found in the asymmetric unit of the CSBP structure, it is evident that the inner zeatin ligand in the CSBP complexes, which is conserved in all four CSBP copies, does not coincide with any ligand molecule in the LlPR-10.2B/zeatin structure. In particular, while the inner zeatin ligand in the CSBP complex is aligned along the α3 helix, the innermost zeatin molecule in the LlPR-10.2B complex (Zea1) is aligned with helices α1 and α2 (Fig. 71B). Considering the position of the outer zeatin molecule in the CSBP structure, which is present only in three out of the four copies of the protein molecule, it is noticeable that it corresponds to the Zea3 ligand in the LlPR-10.2B structure. However, a poor overlap is observed only with the outer zeatin ligand present in chain C of the CSBP protein (Fig. 71B). The other outer zeatin ligands present in the molecules A and D of the CSBP structure are rotated by 180º, pointing the aliphatic tail in the opposite direction.

Structural superposition of the LlPR-10.2B and CSBP molecules from their zeatin complexes reveals that their Cα traces differ by 1.7-1.8 Å. They also have different volumes of the internal cavity, 4500 $Å^3$ in LlPR-10.2B and 1100-1600 $Å^3$ in CSBP (Table 15). The high Cα deviations arise from the overall similar but in detail distinct arrangement of the secondary structure elements of these two proteins. In the CSBP protein, they adopt a conformation where the C-terminal helix α3 is less separated from the β-grip and the helices α1 and α2 are slightly closer to the center of the protein, leading to a significant reduction of the volume of the hydrophobic cavity when compared with the LlPR-10.2B protein. Although both proteins employ five residues for hydrogen bonding with the ligands, only one is positionally conserved. Nevertheless, this amino acid identity is not conserved as it is His68 in LlPR-10.2B and Glu69 in CSBP. It is interesting to note that the same position, but also occupied by a different amino acid residue (Asp), is used by Bet v 1*l* to form a hydrogen bond with the deoxycholate ligand (Markovic-Housley *et al.*, 2003). His68 in the LlPR-10.2B protein is also responsible for a van

der Waal contact with a DPU molecule. The remaining hydrogen bonds in the LlPR-10.2B and CSBP complexes are formed at different spatial positions. In the lupine protein they are formed by Asp7, Glu59, His68, Tyr80 and Tyr82, and in CSBP by Leu21, Gln66, Glu68, Thr138 and Thy141 (Fig. 70). The main forces responsible for the stabilization of the zeatin molecules in the LlPR-10.2B cavity are weak hydrogen bonds formed with mediation of water molecules and van der Waals interactions. Again, the amino acid residues responsible for the latter interactions are generally different in the LlPR-10.2B and CSBP proteins. The number of residues used for van der Waals contacts is also different, 14 in LlPR-10.2B vs. 11 in CSBP (Fig. 70).

Since ligand binding is a dynamic process and the side chains pointing towards the vast internal cavity have some conformational freedom, as already observed in a previous subsection where the residues with different conformation in the LlPR-10.2B/zeatin and LlPR-10.2B/DPU complexes where analyzed in detail, one can expect the PR-10 proteins to be rather general than specific ligand binders. This view is supported by the studies carried out to demonstrate the binding capability of the PR-10 proteins (Mogensen *et al.*, 2002; Markovic-Housley *et al.*, 2003; Koistinen *et al.*, 2005; Pasternak *et al.*, 2006), by the two crystal structures (with zeatin and DPU) and ITC data for LlPR-10.2B, and by the crystal structure of Bet v 1*l* in complex with deoxycholate, a non-plant metabolite (Markovic-Housley *et al.*, 2003). The observation that the residues involved in ligand binding are not conserved in the PR-10 family suggests that the proteins could be involved in storage of different water-insoluble hydrophobic ligands in an unspecific way.

IV.9. Tautomeric forms of zeatin in the LlPR-10.2B/zeatin complex

Although the quality and resolution of the present structure of the LlPR-10.2B/zeatin complex are high, and the definition of the electron density, especially for the ligand molecules, is excellent, the current 1.35 Å resolution X-ray diffraction data are not sufficient to locate the hydrogen atoms. Location of hydrogen atoms would be particularly important in the case of the zeatin molecules because their tautomeric state (N7 or N9 protonation, Fig. 24) or even the degree of protonation (neutral, or protonated at N1, N3, N7 or N9) is not known. However, some deduction based on the pattern of hydrogen bonds in the binding cavity is possible (Fig. 23). First of all, it should be noted that although zeatin was added to the crystallization buffer as a cation, the crystallization solution was buffered near neutral pH (6.5). Consequently, either protonation state of zeatin would be possible. The fact that among the residues surrounding the cavity there is a strong negative net charge (positive Arg138 vs. negative Asp7, Glu59, Glu131) indicates that the contents of the cavity have a positive net charge and that at least some of the zeatin ligands are protonated. With respect to Zea1, it is not possible to speculate about the

tautomeric state of the adenine ring as it interacts exclusively with O-H groups. Zea2, which forms a salt bridge with Asp7 through its N9/N10 atoms must be protonated at N9. The situation is opposite at Zea3, where the proton must reside at N7 in order to form a hydrogen bond with Glu59.

IV.10. Overview

Despite considerable progress in the characterization of plant PR-10 proteins, their biological function remains elusive. However, combining the results obtained by different groups, as well as by different techniques, it is possible to glean important new information.

One very interesting point is the purported PR-10 protein flexibility. Unfortunately, no structural data exists for the same protein in its apo form and in complex with a natural ligand to confirm/reject the proposition. Nevertheless, the increasing number of known PR-10 structures allows some speculations. Perhaps the most interesting conclusion can be drawn from a comparison of the *Lupinus luteus* proteins: PR-10.2A and its homologue PR-10.2B in complex with zeatin. Although these two proteins are very close homologues (91/96% sequence identity/similarity), the absence of a ligand is coupled with an extraordinary kink in helix α3, which reduces the volume of the internal cavity from 4500 Å in the LlPR-10.2B/zeatin complex to 2000 Å in the LlPR-10.2A protein. One could argue that since LlPR-10.2B and LlPR-10.2A are not the same protein, it is the nature of the proteins and not the presence of the ligand that causes this change. However, the interpretation that the differences in the shape and volume of the internal cavity are primarily due to the influence of the ligand bound is supported by the second complex structures of LlPR-10.2B, with the DPU ligand. The volume of the cavity in the DPU complex is reduced from 4500 Å to 3600 Å in comparison with the zeatin complex. Since this reduction cannot be explained by different crystal environment, it is most likely related with the ligand identity.

It is noteworthy that significant flexibility has been also reported for the structurally related StAR protein, which is involved in cholesterol translocation (Bose *et al*., 1999; Christensen *et al.*, 2001; Song *et al*., 2001). Recently, a similar flexibility has been shown for the Bet v 1 homologue (Mogensen *et al*., 2007), underlying the notion that this flexibility can be indeed an intrinsic property of proteins with PR-10 fold. It was shown in the same study that Bet v 1 can also bind to membranes, like the StAR protein (Mogensen *et al*., 2007).

The flexibility of the PR-10 proteins was postulated based on the ligand binding assays, were it was shown that the Bet v 1 protein is capable of binding different kinds of ligands (Mogensen *et al.*, 2002; Markovic-Housley *et al.*, 2003). Similar flexibility was also observed for CSBP, which can bind diverse cytokinins (Pasternak *et al*., 2006) and, within this work, for

the LlPR-10.2B protein, where the two crystal structures and the ITC studies show the ability of the protein to bind different cytokinin ligands.

The small molecule partners, such as cytokinins, in these interactions are also plastic because of their conformational freedom residing in the aliphatic fragment, as well as their unique adaptability to hydrogen-bonding situations, through different protonation states and tautomerization. Thus, the accumulating structural data on PR-10 proteins and their complexes may be revealing an exciting new mode discovered by nature for interactions between partners that are both capable of mutual adaptation. These interactions may not be very strong but are unusually versatile and, together with the relatively low binding constants, could guarantee release of the small-molecule signals to their final receptors.

Literature

Abramovitch R.B., Martin G.B. (2004), Strategies used by bacterial pathogens to suppress plant defenses. *Curr. Opin. Plant Biol.* **7**, 356-364.

Arakane F., King S.R., Du Y., Kallen C.B., Walsh L.P., Watari H., Stocco D.M., Strauss J.F. 3rd (1997), Phosphorylation of steroidogenic acute regulatory protein (StAR) modulates its steroidogenic activity. *J. Biol. Chem.* **272**, 32656-32662.

Bais H.P., Vepachedu R., Lawrence C.B., Stermitz F.R., Vivanco J.M. (2003), Molecular and biochemical characterization of an enzyme responsible for the formation of hypericin in St. John's wort (*Hypericum perforatum* L.). *J. Biol. Chem.* **278**, 32413-32422.

Baker B.Y., Yaworsky D.C., Miller W.L. (2005), A pH-dependent molten globule transition is required for activity of the steroidogenic acute regulatory protein, StAR. *J. Biol. Chem.* **280**, 41753-41760.

Baneyx F., Mujacic M. (2004), Recombinant protein folding and misfolding in Escherichia coli. *Nat. Biotechnol.* **22**, 1399-1408.

Bantignies B., Seguin J., Muzac I., Dedaldechamp F., Gulick P., Ibrahim R. (2000), Direct evidence for ribonucleolytic activity of a PR-10-like protein from white lupin roots. *Plant Mol. Biol.* **42**, 871-881.

Barratt D.H.P., Clark J.A. (1993), Proteins arising during the late stages of embryogenesis in Pisum sativum L. *Planta* **184**, 14–23.

Benvenuti M., Mangani S. (2007), Crystallization of soluble proteins in vapor diffusion for x-ray crystallography. *Nat. Protoc.* **2**, 1633-1651.

Berman H.M., Westbrook J., Feng Z., Gilliland G., Bhat T.N., Weissig H., Shindyalov I.N., Bourne P.E. (2000), The Protein Data Bank. *Nucleic Acids Res.* **28**, 235-242.

Bhat T.N., Cohen G.H. (1984), OMITMAP: an electron density map suitable for the examination of errors in a macromolecular model. *J. Appl. Cryst.* **17**, 244-248.

Bhat T.N. (1988), Calculation of an OMIT map. *J. Appl. Cryst.* **21**, 279-281.

Biesiadka J., Bujacz G., Sikorski M.M., Jaskolski M. (2002), Crystal structures of two homologous pathogenesis-related proteins from yellow lupine. *J. Mol. Biol.* **319**, 1223-1234.

Blow D.M., Chayen N.E., Lloyd L.F., Saridakis E. (1994), Control of nucleation of protein crystals. *Protein Sci.* **3**, 1638-1643.

Bose H.S., Whittal R.M., Baldwin M.A., Miller W.L. (1999), The active form of the steroidogenic acute regulatory protein, StAR, appears to be a molten globule. *Proc. Natl. Acad. Sci. U.S.A.* **96**, 7250-7255.

Breda C., Sallaud C., el-Turk J., Buffard D., de Kozak I., Esnault R., Kondorosi A. (1996), Defense reaction in Medicago sativa: a gene encoding a class 10 PR protein is expressed in vascular bundles. *Mol. Plant Microbe Interact.* **9**, 713-719.

Breiteneder H., Pettenburger K., Bito A., Valenta R., Kraft D., Rumpold H., Scheiner O., Breitenbach M. (1989), The gene coding for the major birch pollen allergen Betv1, is highly homologous to a pea disease resistance response gene. *Embo J.* **8**, 1935-1938.

Breiteneder H., Hoffmann-Sommergruber K., O'Riordain G., Susani M., Ahorn H., Ebner C., Kraft D., Scheiner O. (1995), Molecular characterisation of *Api g* 1, the major allergen of celery (*Apium graveolens*), and its immunological and structural relationships to a group of 17-kDa tree pollen allergens. *Eur. J. Biochem.* **223**, 484–489.

Breiteneder H. (2004), Thaumatin-like proteins -- a new family of pollen and fruit allergens. *Allergy* **59**, 479-481.

Briskin D.P., Gawienowski M.C. (2001), Differential effects of light and nitrogen on production of hypericins and leaf glands in Hypericum perforatum. *Plant Physiol. Biochem.* **39**, 1075-1081.

Brisson L.F., Tenhaken R., Lamb C. (1994), Function of Oxidative Cross-Linking of Cell Wall Structural Proteins in Plant Disease Resistance. *Plant Cell* **6**, 1703-1712.

Broderick K., Pittock C., Arioli T., Creaser E.H., Weinman J.J., Rolfe B.G. (1997), Pathogenesis related proteins in *Trifolium subterraneum*: a general survey and subsequent characterisation of a protein inducible by ethephon and redlegged earth mite attack. *Aust. J. Plant Physiol.* **24**, 819–829.

Brunger A.T. (1993), Assessment of phase accuracy by cross validation: the free R value. Methods and applications. *Acta Cryst.* D**49**, 24-36.

Brunger A.T., Adams P.D., Clore G.M., DeLano W.L., Gros P., Grosse-Kunstleve R.W., Jiang J.S., Kuszewski J., Nilges M., Pannu N.S., Read R.J., Rice L.M., Simonson T., Warren G.L. (1998), Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Cryst.* D**54,** 905-921.

Bufe A., Spangfort M.D., Kahlert H., Schlaak M., Becker W.M. (1996), The major birch pollen allergen, Bet v 1, shows ribonuclease activity. *Planta* **199**, 413-415.

Carlomagno T. (2005), Ligand-target interactions: what can we learn from NMR? *Annu. Rev. Biophys. Biomol. Struct.* **34**, 245-266.

Carpin S., Laffer S., Schoentgen F., Valenta R., Chenieux J.C., Rideau M., Hamdi S. (1998), Molecular characterization of a cytokinin-inducible periwinkle protein showing sequence homology with pathogenesis-related proteins and the Bet v 1 allergen family. *Plant Mol. Biol.* **36**, 791-798.

Chadha P., Das R.H. (2006), A pathogenesis related protein, AhPR10 from peanut: an insight of its mode of antifungal activity. *Planta* **225**, 213-222.

Chenna R., Sugawara H., Koike T., Lopez R., Gibson T.J., Higgins D.G., Thompson J.D. (2003), Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Res.* **31**, 3497-3500.

Christensen K., Bose H.S., Harris F.M., Miller W.L., Bell J.D. (2001), Binding of steroidogenic acute regulatory protein to synthetic membranes suggests an active molten globule. *J. Biol. Chem.* **276**, 17044-17051.

Claude J.B., Suhre K., Notredame C., Claverie J.M., Abergel C. (2004), CaspR: a web-server for automated molecular replacement using homology modelling. *Nucleic Acids Res.* **32**, W606-W609.

Clouse S.D. (1996), Molecular genetic studies confirm the role of brassinosteroids in plant growth and development. *Plant J.* **10**, 1-8.

Clouse S. (2001), Brassinosteroids. *Curr. Biol.* **11**, R904.

Cohen G.R. (1997), ALIGN: a program to superimpose protein coordinates, accounting for insertions and deletions. *J. Appl. Cryst.* **30**, 1160-1161.

Collaborative Computer Project No. 4 (1994), *Acta Cryst.* D**50**, 760-763.

Colditz F., Niehaus K., Krajinski F. (2007), Silencing of PR-10-like proteins in Medicago truncatula results in an antagonistic induction of other PR proteins and in an increased tolerance upon infection with the oomycete Aphanomyces euteiches. *Planta* **226**, 57-71.

Constabel C.P., Brisson N. (1995), Stigma- and vascular-specific expression of the PR10a gene of potato: A novel expression of a pathogenesis-related gene. *Mol. Plant Microbe Interact.* **8**, 104-113.

Crestfield A.M., Moore S., Stein W.H. (1963), The preparation and enzymatic hydrolysis of reduced and S-carboxymethylated proteins. *J. Biol. Chem.* **238**, 622-627.

Crowell D.N., John M.E., Russell D., Amasino R.M. (1992), Characterization of a stress-induced, developmentally regulated gene family from soybean. *Plant Mol. Biol.* **18**, 459-466.

Dangl J.L., Dietrich R.A., Richberg M.H. (1996), Death Don't Have No Mercy: Cell Death Programs in Plant-Microbe Interactions. *Plant Cell* **8**, 1793-1807.

Dauter Z. (1999), Data-collection strategies. *Acta Cryst.* D**55**, 1703-1717.

Dauter Z. (2006), Current state and prospects of macromolecular crystallography. *Acta Cryst.* D**62**, 1-11.

Davis I.W., Leaver-Fay A., Chen V.B., Block J.N., Kapral G.J., Wang X., Murray L.W., Arendall W.B. 3rd, Snoeyink J., Richardson J.S., Richardson D.C. (2007), MolProbity:

all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.* **35**, W375-W383.

DeLano W.L. (2002), The PyMOL Molecular Graphics System. DeLano Scientific, San Carlos, CA.

Deltito J., Beyer D. (1998), The scientific, quasi-scientific and popular literature on the use of St. John's Wort in the treatment of depression. *J. Affect. Disord.* **51**, 345-351.

Derewenda Z.S. (2004), Rational protein crystallization by mutational surface enginnering. *Structure* **12**, 529-535.

Emanuelsson O., Nielsen H., Brunak S., von Heijne G. (2000), Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *J. Mol. Biol.* **300**, 1005-1016.

Emsley P., Cowtan K. (2004), Coot: model-building tools for molecular graphics. *Acta Cryst.* D**60**, 2126-2132.

Engh R.A., Huber R. (2006), Structure quality and target parameters. *Intern. Tables Cryst.* F**18.3**, 382-392.

Ericsson U.B., Hallberg B.M., Detitta G.T., Dekker N., Nordlund P. (2006), Thermofluor-based high-throughput stability optimization of proteins for structural studies. *Anal. Biochem.* **357**, 289-298.

Evans P., McCoy A. (2008), An introduction to molecular replacement. *Acta Cryst.* D**64**, 1-10.

Eyal E., Yang L.W., Bahar I. (2006) Anisotropic network model: systematic evaluation and a new web interface. *Bioinformatics* **22**, 2619-2627.

Facchini P.J. (2001), ALKALOID BIOSYNTHESIS IN PLANTS: Biochemistry, Cell Biology, Molecular Regulation, and Metabolic Engineering Applications. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **52**, 29-66.

Fernandes H., Pasternak O., Bujacz G., Bujacz A., Sikorski M.M., Jaskolski M. (2008), Lupinus luteus pathogenesis-related protein as a reservoir for cytokinin. *J. Mol. Biol.* **378**, 1040-1051.

Fernandes H., Konieczna M., Kolodziejczyk R., Bujacz G., Sikorski M., Jaskolski M. (2008b), Crystallization and preliminary crystallographic studies of Hyp-1, a St John's wort protein implicated in the biosynthesis of hypericin. *Acta Cryst.* F**64**, 405-408.

Fischer B.E. (1994), Renaturation of recombinant proteins produced as inclusion bodies. *Biotechnol. Adv.* **12**, 89-101.

Flores T., Alape-Giron A., Flores-Diaz M., Flores H.E. (2002), Ocatin. A novel tuber storage protein from the andean tuber crop oca with antibacterial and antifungal activities. *Plant Physiol.* **128**, 1291-1302.

Fristensky B., Horovitz D., Hadwiger L.A. (1988), cDNA sequences for pea disease resistance response. *Plant Mol. Biol.* **11**, 713–715.

Fujimoto Y., Nagata R., Fukasawa H., Yano K., Azuma M., Iida A., Sugimoto S., Shudo K., Hashimoto Y. (1998), Purification and cDNA cloning of cytokinin-specific binding protein from mung bean (Vigna radiata). *Eur. J. Biochem.* **258**, 794-802.

Gajhede M., Osmark P., Poulsen F.M., Ipsen H., Larsen J.N., Joost van Neerven R.J., Schou C., Lowenstein H., Spangfort M.D. (1996), X-ray and NMR structure of Bet v 1, the origin of birch pollen allergy. *Nat. Struct. Biol.* **3**, 1040-1045.

Garman E. (1999), Cool data: quantity AND quality. *Acta Cryst.* D**55**, 1641-1653.

Gonzalez A. (2003), Optimizing data collection for structure determination. *Acta Cryst.* D**59**, 1935-1942.

Greenberg J.T. (1996), Programmed cell death: a way of life for plants. *Proc. Natl. Acad. Sci. U.S.A.* **93**, 12094-12097.

Greenfield N.J. (1999), Applications of circular dichroism in protein and peptide analysis. *Trends Analyt. Chem.* **18**, 236-244.

Hammond-Kosack K.E., Jones J.D. (1996), Resistance gene-dependent plant defense responses. *Plant Cell* **8**, 1773-1791.

Handschuh L. (2004), PhD Thesis. Institute of Bioorganic Chemistry, Polish Academy of Sciences, Poznan, Poland.

Handschuh L., Femiak I., Kasperska A., Figlerowicz M., Sikorski M.M. (2007), Structural and functional characteristics of two novel members of pathogensis-related multigene family of class 10 from yellow lupine+. *Acta Biochim. Pol.* **54**, 783-796.

Hannapel D.J. (1990), Differential Expression of Potato Tuber Protein Genes. *Plant Physiol.* **94**, 919-925.

Harding M.M. (2001), Geometry of metal-ligand interactions in proteins. *Acta Cryst.* D**57**, 401-411.

Hashimoto M., Kisseleva L., Sawa S., Furukawa T., Komatsu S., Koshiba T. (2004), A novel rice PR10 protein, RSOsPR10, specifically induced in roots by biotic and abiotic stresses, possibly via the jasmonic acid signaling pathway. *Plant Cell Physiol.* **45**, 550-559.

Hauptman H., Karle J. (1953), Solution of the Phase Problem I. The Centrosymmetric Crystal. ACA Monographs.

Hejgaard J., Jacobsen S., Bjorn S.E., Kragh K.M. (1992), Antifungal activity of chitin-binding PR-4 type proteins from barley grain and stressed leaf. *FEBS Lett.* **307**, 389-392.

Helliwell J.R. (2001), Synchrotron-radiation instrumentation, methods and scientific utilization. *Intern. Tables Cryst.* F**8.1**, 155-166.

Hendrickson W.A. (1991), Determination of macromolecular structures from anomalous diffraction of synchrotron radiation. *Science* **254**, 51-58.

Hodel A., Kim S.H., Brunger A.T. (1992), Model bias in macromolecular crystal structures. *Acta Cryst.* A**48**, 851-858.

Hofius D., Tsitsigiannis D.I., Jones J.D., Mundy J. (2007), Inducible cell death in plant immunity. *Semin. Cancer Biol.* **17**, 166-187.

Holm J., Gajhede M., Ferreras M., Henriksen A., Ipsen H., Larsen J.N., Lund L., Jacobi H., Millner A., Wurtzen P.A., Spangfort M.D. (2004), Allergy vaccine engineering: epitope modulation of recombinant Bet v 1 reduces IgE binding but retains protein folding pattern for induction of protective blocking-antibody responses. *J. Immunol.* **173**, 5258-5267.

Huang J.C., Chang F.C., Wang C.S. (1997), Characterization of a lily tapetal transcript that shares sequence similarity with a class of intracellular pathogenesis-related (IPR) proteins. *Plant Mol. Biol.* **34**, 681-686.

Iturriaga E.A., Leech M.J., Barrat D.H.P., Wang T.L. (1994), Two ABA-responsive proteins from pea (*Pisum sativum* L.) are closely related to intracellular pathogenesis-related proteins. *Plant Mol. Biol.* **24**, 235–240.

Izhaki I. (2002), Emodin - a secondary metabolite with multiple ecological functions in higher plants. *New Phytologist.* **155**, 205-217.

Kabsch W., Sander C. (1983), Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22**, 2577-2637.

Kelly S.M., Price N.C. (2000), The use of circular dichroism in the investigation of protein structure and function. *Curr. Protein Pept. Sci.* **1**, 349-384.

Kelly S.M., Jess T.J., Price N.C. (2005), How to study proteins by circular dichroism. *Biochim. Biophys. Acta* **1751**, 119-139.

Kissinger C.R., Gehlhaar D.K., Fogel D.B. (1999), Rapid automated molecular replacement by evolutionary search. *Acta Cryst.* D**55**, 484-491.

Kleywegt G.J., Jones T.A. (1994), Detection, delineation, measurement and display of cavities in macromolecular structures. *Acta Cryst.* D**50**, 178-185.

Kleywegt G.J. (2000), Validation of protein crystal structures. *Acta Cryst.* D**56**, 249-265.

Kleywegt G.J., Zou J.-Y., Kjeldgaard M., Jones M., Jones T.A. (2006), Around *O*. *Intern. Tables Cryst.* F**17.1**, 353-356.

Kobayashi M., Kubota M., Matsuura Y. (1999), Crystallization and improvement of crystal quality for x-ray diffraction of maltooligosyl trehalose synthase by reductive methylation of lysine residues. *Acta Cryst.* D**55**, 931-933.

Koistinen K.M., Kokko H.I., Hassinen V.H., Tervahauta A.I., Auriola S., Karenlampi S.O. (2002), Stress-related RNase PR-10c is post-translationally modified by glutathione in birch. *Plant Cell Env.* **25**, 707-715.

Koistinen K.M., Soininen P., Venalainen T.A., Hayrinen J., Laatikainen R., Perakyla M., Tervahauta A.I., Karenlampi S.O. (2005), Birch PR-10c interacts with several biologically important ligands. *Phytochemistry* **66**, 2524-2533.

Konieczna M. (2007), Master Thesis. Institute of Bioorganic Chemistry, Polish Academy of Sciences, Poznan, Poland.

Korostelev A., Bertram R., Chapman M.S. (2002), Simulated-annealing real-space refinement as a tool in model building. *Acta Cryst.* D**58**, 761-767.

Kosuth J., Katkovcinova Z., Olexova P., Cellarova E. (2007), Expression of the hyp-1 gene in early stages of development of Hypericum perforatum L. *Plant Cell Rep.* **26**, 211-217.

Kubin A., Wierrani F., Burner U., Alth G., Grunberger W. (2005), Hypericin--the facts about a controversial agent. *Curr. Pharm. Des.* **11**, 233-253.

Kurinov I.V., Mao C., Irvin J.D., Uckun F.M. (2000), X-ray crystallographic analysis of pokeweed antiviral protein-II after reductive methylation of lysine residues. *Biochem. Biophys. Res. Commun.* **275**, 549-552.

Laitinen M.L., Julkunen-Tiitto R., Rousi M. (2000), Variation in phenolic compounds within a birch (Betula pendula) population. *J. Chem. Ecol.* **26**, 1609-1622.

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

Laskowski R.A. (1995), SURFNET: A program for visualizing molecular surfaces, cavities and intermolecular interactions. *J. Mol. Graph.* **13**, 323-330.

Legocki A.B., Karlowski W.M., Podkowinski J., Sikorski M.M., Stepkowski T. (1997), Advances in molecular characterization of the Yellow lupin-*Bradyrhizibium* sp. (*Lupinus*) symbiotic model; in Biological Fixation of Nitrogen for Sustainable Agriculture (Legocki AB, Bothe H, Puchler A eds. NATO ASl Series, vol. G39, pp. 262-266, Springer Verlago).

Li M., Su Z.G., Janson J.C. (2004), In vitro protein refolding by chromatographic procedures. *Protein Expr. Purif.* **33**, 1-10.

Liscombe D.K., Macleod B.P., Loukanina N., Nandi O.I., Facchini P.J. (2005), Evidence for the monophyletic evolution of benzylisoquinoline alkaloid biosynthesis in angiosperms. *Phytochemistry* **66**, 1374-1393.

Liu J.J., Ekramoddoullah A.K.M., Yu X.S. (2003), Differential expression of multiple PR10 proteins in western white pine following wounding, fungal infection and cold-hardening. *Physiol. Plant.* **119**, 544-553.

Liu X., Huang B., Lin J., Fei J., Chen Z., Pang Y., Sun X., Tang K. (2006), A novel pathogenesis-related protein (SsPR10) from Solanum surattense with ribonucleolytic and antimicrobial activity is stress- and pathogen-inducible. *J. Plant Physiol.* **163**, 546-556.

Lo M.C., Aulabaugh A., Jin G., Cowling R., Bard J., Malamas M., Ellestad G. (2004), Evaluation of fluorescence-based thermal shift assays for hit identification in drug discovery. *Anal. Biochem.* **332**, 153-159.

Maeshima M., Sasaki T., Asahi T. (1985), Characterization of major proteins in sweet potato tuberous roots. *Phytochemistry* **34**, 1899-1902.

Markovic-Housley Z., Degano M., Lamba D., von Roepenack-Lahaye E., Clemens S., Susani M., Ferreira F., Scheiner O., Breiteneder H. (2003), Crystal structure of a hypoallergenic isoform of the major birch pollen allergen Bet v 1 and its likely biological function as a plant steroid carrier. *J. Mol. Biol.* **325**, 123-133.

Matthews B.W. (1968), Solvent content of protein crystals. *J. Mol. Biol.* **33**, 491-497.

Matthews B.W. (1985), Determination of protein molecular weight, hydratation, and packing from crystal density. *Methods Enzymol.* **114**, 176-187.

Matulis D., Baumann C.G., Bloomfield V.A., Lovrien R.E. (1999), 1-anilino-8-naphthalene sulfonate as a protein conformational tightening agent. *Biopolymers* **49**, 451-458.

Matulis D., Kranz J.K., Salemme F.R., Todd M.J. (2005), Thermodynamic stability of carbonic anhydrase: measurements of binding affinity and stoichiometry using ThermoFluor. *Biochemistry* **44**, 5258-5266.

Matton D.P., Brisson N. (1989), Cloning, expression, and sequence conservation of pathogenesis-related gene transcripts of potato. *Mol. Plant Microbe Interact.* **2**, 325–331.

Mayer M., Meyer B. (1999), Characterization of ligand binding by saturation transfer difference NMR spectroscopy. *Angewandte Chemie-International Edition* **38**, 1784-1788.

Mayer M., Meyer B. (2001), Group epitope mapping by saturation transfer difference NMR to identify segments of a ligand in direct contact with a protein receptor. *J. Am. Chem. Soc.* **123**, 6108-6117.

McCoy M.A., Senior M.M., Wyss D.F. (2005), Screening of protein kinases by ATP-STD NMR spectroscopy. *J. Am. Chem. Soc.* **127**, 7978-7979.

McCoy A.J., Grosse-Kunstleve R.W., Adams P.D., Winn M.D., Storoni L.C., Read R.J. (2007), Phaser crystallographic software. *J. Appl. Cryst.* **40**, 658-674.

McRee D.E. (1999), XtalView/Xfit--A versatile program for manipulating atomic coordinates and electron density. *J. Struct. Biol.* **125**, 156-165.

Means G.E., Feeney R.E. (1968), Reductive alkylation of amino groups in proteins. *Biochemistry* **7**, 2192-2201.

Merritt E.A. (1999), Expanding the model: anisotropic displacement parameters in protein structure refinement. *Acta Cryst.* D**55**, 1109-1117.

Middelberg A.P. (2002), Preparative protein refolding. *Trends Biotechnol.* **20**, 437-443.

Miller R., DeTitta G.T., Jones R., Langs D.A., Weeks C.M., Hauptman H.A. (1993), On the application of the minimal principle to solve unknown structures. *Science* **259**, 1430-1433.

Minami H., Dubouzet E., Iwasa K., Sato F. (2007), Functional analysis of norcoclaurine synthase in Coptis japonica. *J. Biol. Chem.* **282**, 6274-6282.

Mirza O., Henriksen A., Ipsen H., Larsen J.N., Wissenbach M., Spangfort M.D., Gajhede M. (2000), Dominant epitopes and allergic cross-reactivity: complex formation between a Fab fragment of a monoclonal murine IgG antibody and the major allergen from birch pollen Bet v 1. *J. Immunol.* **165**, 331-338.

Mogensen J.E., Wimmer R., Larsen J.N., Spangfort M.D., Otzen D.E. (2002), The major birch allergen, Bet v 1, shows affinity for a broad spectrum of physiological ligands. *J. Biol. Chem.* **277**, 23684-23692.

Mogensen J.E., Ferreras M., Wimmer R., Petersen S.V., Enghild J.J., Otzen D.E. (2007), The major allergen from birch tree pollen, Bet v 1, binds and permeabilizes membranes. *Biochemistry* **46**, 3356-3365.

Moiseyev G.P., Beintema J.J., Fedoreyeva L.I., Yakovlev G.I. (1994), High sequence similarity between a ribonuclease from ginseng calluses and fungus-elicited proteins from parsley indicates that intracellular pathogenesis-related proteins are ribonucleases. *Planta* **193**, 470-472.

Moiseyev G.P., Fedoreyeva L.I., Zhuravlev Y.N., Yasnetskaya E., Jekel P.A., Beintema J.J. (1997), Primary structures of two ribonucleases from ginseng calluses. New members of the PR-10 family of intracellular pathogenesis-related plant proteins. *FEBS Lett.* **407**, 207-210.

Mok, D.W.S., Mok, M.C. (1994), Cytokinins: Chemistry, Activity and Function. (Boca Raton, FL: CRC Press).

Morrissey J.P., Osbourn A.E. (1999), Fungal resistance to plant antibiotics as a mechanism of pathogenesis. *Microbiol. Mol. Biol. Rev.* **63**, 708-724.

Muller P., Kopke S., Sheldrick G.M. (2003), Is the bond-valence method able to identify metal atoms in protein structures? *Acta Cryst.* D**59**, 32-37.

Murshudov G.N., Vagin A.A., Dodson E.J. (1997), Refinement of macromolecular structures by the maximum-likelihood method. *Acta Cryst.* D**53**, 240-255.

Mylona P., Moerman M., Yang W.C., Gloudemans T., Van de Kerckhove J., van Kammen A., Bisseling T., Franssen H.J. (1994), The root epidermis-specific pea gene RH2 is homologous to a pathogenesis-related gene. *Plant Mol. Biol.* **26**, 39-50.

Mysore K.S., Ryu C.M. (2004), Nonhost resistance: how much do we know? *Trends Plant Sci.* **9**, 97-104.

Navaza J. (1994), AMoRe: an automated package for molecular replacement. *Acta Cryst.* A**50**, 157-163.

Nayal M., Di Cera E. (1996), Valence screening of water in protein crystals reveals potential Na+ binding sites. *J. Mol. Biol.* **256**, 228-234.

Nessler C.L., Kurz W.G., Pelcher L.E. (1990), Isolation and analysis of the major latex protein genes of opium poppy. *Plant Mol. Biol.* **15**, 951-953.

Neudecker P., Schweimer K., Nerkamp J., Scheurer S., Vieths S., Sticht H., Rosch P. (2001), Allergic cross-reactivity made visible: solution structure of the major cherry allergen Pru av 1. *J. Biol. Chem.* **276**, 22756-22763.

Neudecker P., Lehmann K., Nerkamp J., Haase T., Wangorsch A., Fotisch K., Hoffmann S., Rosch P., Vieths S., Scheurer S. (2003), Mutational epitope analysis of Pru av 1 and Api g 1, the major allergens of cherry (Prunus avium) and celery (Apium graveolens): correlating IgE reactivity with three-dimensional structure. *Biochem. J.* **376**, 97-107.

O'Keefe M., Brese N.E. (1991), Atom sizes and bond lengths in molecules and crystals. *J. Am. Chem. Soc.* **113**, 3226-3229.

Onelli E., Rivetta A., Giorgi A., Bignami M., Cocucci M., Patrignani G. (2002), Ultrastructural studies on the developing secretory nodules of Hypericum perforatum. *Flora* **197**, 92-102.

Osmark P., Boyle B., Brisson N. (1998), Sequential and structural homology between intracellular pathogenesis-related proteins and a group of latex proteins. *Plant Mol. Biol.* **38**, 1243-1246.

Otwinowski Z., Minor W. (1997), Processing of X-Ray Diffraction Data Collected in Oscillation Mode. *Methods Enzymol.* **276**, 307-326.

Otwinowski Z., Minor W. (2001), Denzo and Scalepack. *Intern. Tables Cryst.* F**11.4**, 226-235.

Painter J., Merritt E.A. (2005), A molecular viewer for the analysis of TLS rigid-body motion in macromolecules. *Acta Cryst.* D**61**, 465-471.

Park S.W., Stevens N.M., Vivanco J.M. (2002), Enzymatic specificity of three ribosome-inactivating proteins against fungal ribosomes, and correlation with antifungal activity. *Planta* **216**, 227-234.

Park C.J., Kim K.J., Shin R., Park J.M., Shin Y.C., Park K.H. (2004), Pathogenesis-related protein 10 isolated from hot pepper functions as a ribonuclease in an antiviral pathway. *Plant J.* **37**, 186-198.

Pasquo A., Bonamore A., Franceschini S., Macone A., Boffi A., Ilari A. (2008), Cloning, expression, crystallization and preliminary X-ray data analysis of norcoclaurine synthase from Thalictrum flavum. *Acta Cryst.* F**64**, 281-283.

Pasternak O., Biesiadka J., Dolot R., Handschuh L., Bujacz G., Sikorski M.M., Jaskolski M. (2005), Structure of a yellow lupin pathogenesis-related PR-10 protein belonging to a novel subclass. *Acta Cryst.* D**61**, 99-107.

Pasternak O., Bujacz G.D., Fujimoto Y., Hashimoto Y., Jelen F., Otlewski J., Sikorski M.M., Jaskolski M. (2006), Crystal Structure of Vigna radiata Cytokinin-Specific Binding Protein in Complex with Zeatin. *Plant Cell* **18**, 2622-2634.

Pasternak O., Bujacz A., Biesiadka J., Bujacz G., Sikorski M., Jaskolski M. (2008), MAD phasing using the (Ta(6)Br(12))(2+) cluster: a retrospective study. *Acta Cryst.* D**64**, 595-606.

Perozzo R., Folkers G., Scapozza L. (2004), Thermodynamics of protein-ligand interactions: history, presence, and future aspects. *J. Recept. Signal Transduct. Res.* **24**, 1-52.

Perrakis A., Morris R., Lamzin V.S. (1999), Automated protein model building combined with iterative structure refinement. *Nat. Struct. Biol.* **6**, 458-463.

Pinto M.P., Ricardo C.P. (1995), Lupinus albus L. pathogenesis-related proteins that show similarity to PR-10 proteins. *Plant Physiol.* **109**, 1345-1351.

Pozueta-Romero J., Klein M., Houlne G., Schantz M.L., Meyer B., Schantz R. (1995), Characterization of a family of genes encoding a fruit-specific wound-stimulated protein of bell pepper (Capsicum annuum): identification of a new family of transposable elements. *Plant Mol. Biol.* **28**, 1011-1025.

Racusssen D., Foote M. (1980), A major soluble glycoprotein of potato. *J. Food Biochem.* **4**, 43-52.

Ravelli R.B., Leiros H.K., Pan B., Caffrey M., McSweeney S. (2003), Specific radiation damage can be used to solve macromolecular crystal structures. *Structure* **11**, 217-224.

Rayment I., Rypniewski W.R., Schmidt-Base K., Smith R., Tomchick D.R., Benning M.M., Winkelmann D.A., Wesenberg G., Holden H.M. (1993), Three-dimensional structure of myosin subfragment-1: a molecular motor. *Science* **261**, 50-58.

Read R.J. (1986), Improved Fourier coefficients for maps using phases from partial structures with errors. *Acta Cryst.* A**42**, 140-149.

Rice R.H., Means G.E. (1971), Radioactive labeling of proteins in vitro. *J. Biol. Chem.* **246**, 831-832.

Rueffer M., El-Shagi H., Nagakura N., Zenk M.H. (1981), (S)-Norlaudanosoline Synthase: the first enzyme in the benzylisoquinoline biosynthetic pathway. *FEBS Lett.* **129**, 5-9.

Rypniewski W.R., Holden H.M., Rayment I. (1993), Structural consequences of reductive methylation of lysine residues in hen egg white lysozyme: an X-ray analysis at 1.8-A resolution. *Biochemistry* **32**, 9851-9858.

Samanani N., Facchini P.J. (2002), Purification and characterization of norcoclaurine synthase. The first committed enzyme in benzylisoquinoline alkaloid biosynthesis in plants. *J. Biol. Chem.* **277**, 33878-33883.

Samanani N., Liscombe D.K., Facchini P.J. (2004), Molecular cloning and characterization of norcoclaurine synthase, an enzyme catalyzing the first committed step in benzylisoquinoline alkaloid biosynthesis. *Plant J.* **40**, 302-313.

Saraste M., Sibbald P.R., Wittinghofer A. (1990), The P-loop--a common motif in ATP- and GTP-binding proteins. *Trends Biochem. Sci.* **15**, 430-434.

Saxena A.K., Singh K., Su H.P., Klein M.M., Stowers A.W., Saul A.J., Long C.A., Garboczi D.N. (2006), The essential mosquito-stage P25 and P28 proteins from Plasmodium form tile-like triangular prisms. *Nat. Struct. Mol. Biol.* **13**, 90-91.

Sayre D. (1952), The squaring method: a new method for phase determination. *Acta Cryst.* A**5**, 60-65.

Sayre D. (1974), Least-square phase refinement. II. High-resolution phasing of a small protein. *Acta Cryst.* A**30**, 180-184.

Schirmer T., Hoffimann-Sommergrube K., Susani M., Breiteneder H., Markovic-Housley Z. (2005), Crystal structure of the major celery allergen Api g 1: molecular analysis of cross-reactivity. *J. Mol. Biol.* **351**, 1101-1109.

Schlumbaum A., Mauch F., Vogeli U., Boller T. (1986), Plant chitinases are potent inhibitors of fungal growth. *Nature* **324**, 365-367.

Schmeller T., Latz-Bruning B., Wink M. (1997), Biochemical activities of berberine, palmatine and sanguinarine mediating chemical defence against microorganisms and herbivores. *Phytochemistry* **44**, 257-266.

Schmelzer E., Kruger-Lebus S., Hahlbrock K. (1989), Temporal and spatial patterns of gene expression around sites of attempted fungal infection in parsley leaves. *Plant Cell* **1**, 993–1001.

Schubot F.D., Waugh D.S. (2004), A pivotal role for reductive methylation in the de novo crystallization of a ternary complex composed of Yersinia pestis virulence factors YopN, SycN and YscB. *Acta Cryst.* D**60**, 1981-1986.

Schuttelkopf A.W., van Aalten D.M. (2004), PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. *Acta Cryst.* D**60**, 1355-1363.

Schweimer K., Sticht H., Boehm M., Roesch P. (1999), NMR Spectroscopy Reveals Common Structural Features of the Birch Pollen Allergen Bet v 1 and the cherry allergen Pru a 1. *Appl. Magn. Reson.* **17**, 449-464.

Sheldrick G.M. (1990), Phase annealing in SHELX-90: direct methods for larger structures. *Acta cryst.* A**46**, 467-473.

Sheldrick G.M. (2008), A short history of SHELX. *Acta Cryst.* A**64**, 112-122.

Shewry P.R. (2003), Tuber storage proteins. *Ann. Bot. (Lond.)* **91**, 755-769.

Shuker S.B., Hajduk P.J., Meadows R.P., Fesik S.W. (1996), Discovering high-affinity ligands for proteins: SAR by NMR. *Science* **274**, 1531-1534.

Sikorski M.M., Szlagowska A.E., Legocki A.B. (1995), cDNA sequences encoding for two homologues of *Lupinus luteus* (L.) IPR-like proteins (Accession Nos X79974 and X79975 for LlR10A and LIR18B mRNA's respectively). *Plant Physiol.* **110**, 335-347.

Sikorski M.M. (1997), Expression of Lupinus luteus cDNA coding for PR10 protein in Escherichia coli: purification of the recombinant protein for structural and functional studies. *Acta Biochim. Pol.* **44**, 565-578.

Sikorski M.M., Biesiadka J., Kasperska A.E., Kopcinska J., Lotocka B., Golinowski W., Legocki A.B. (1999), Expression of genes encoding PR10 class patho genesis-related proteins is inhibited in yellow lupine root nodules. *Plant Science* **149**, 125-137.

Somssich I.E., Schmelzer E., Kawalleck P., Hahlbrock K. (1988), Gene structure and in situ transcript localization of pathogenesis-related protein 1 in parsley. *Mol. Gen. Genet.* **213**, 93-98.

Song M., Shao H., Mujeeb A., James T.L., Miller W.L. (2001), Molten-globule structure and membrane binding of the N-terminal protease-resistant domain (63-193) of the steroidogenic acute regulatory protein (StAR). *Biochem. J.* **356**, 151-158.

Spangfort M.D., Mirza O., Ipsen H., Van Neerven R.J., Gajhede M., Larsen J.N. (2003), Dominating IgE-binding epitope of Bet v 1, the major allergen of birch pollen,

characterized by X-ray crystallography and site-directed mutagenesis. *J. Immunol.* **171**, 3084-3090.

Srivastava S., Fristensky B., Kav N.N. (2004), Constitutive expression of a PR10 protein enhances the germination of Brassica napus under saline conditions. *Plant Cell Physiol.* **45**, 1320-1324.

Srivastava S., Emery R.J.N., Kurepin L.V., Reid D.M., Fristensky B., Kav N.N.V. (2006), Pea PR 10.1 is a ribonuclease and its transgenic expression elevates cytokinin levels. *Plant Growth Regulation* **49**, 17-25.

Stocco D.M., Clark B.J. (1996), Role of the steroidogenic acute regulatory protein (StAR) in steroidogenesis. *Biochem. Pharmacol.* **51**, 197-205.

Studier F.W., Rosenberg A.H., Dunn J.J., Dubendorff J.W. (1990), Use of T7 RNA polymerase to direct expression of cloned genes. *Methods Enzymol.* **185**, 60-89.

Swoboda I., Scheiner O., Kraft D., Breitenbach M., Heberle-Bors E., Vicente O. (1994), A birch gene family encoding pollen allergens and pathogenesis-related proteins. *Biochim. Biophys. Acta* **1219**, 457-464.

Swoboda I., Hoffmann-Sommergruber K., ORiordain G., Scheiner O., HeberleBors E., Vicente O. (1996), Bet v 1 proteins, the major birch pollen allergens and members of a family of conserved pathogenesis-related proteins, show ribonuclease activity in vitro. *Physiol. Plant.* **96**, 433-438.

Taylor G. (2003), The phase problem. *Acta Cryst.* D**59**, 1881-1890.

Teng T.Y. (1990), Mounting of crystals for macromolecular crystallography in a free-standing thin film. *J. Appl. Cryst.* **23**, 387-391.

Tronrud D.E. (2004), Introduction to macromolecular refinement. *Acta Cryst.* D**60**, 2156-2168.

Tsujishita Y., Hurley J.H. (2000), Structure and lipid transport mechanism of a StAR-related domain. *Nat. Struct. Biol.* **7**, 408-414.

Ukaji N., Kuwabara C., Takezawa D., Arakawa K., Fujikawa S. (2004), Accumulation of pathogenesis-related (PR) 10/Bet v 1 protein homologues in mulberry (Morus bombycis Koidz.) tree during winter. *Plant Cell Env.* **27**, 1112-1121.

Vagin A., Teplyakov A. (1997), MOLREP: an automated program for molecular replacement. *J. Appl. Cryst.* **30**, 1022-1025.

Vaguine A.A., Richelle J., Wodak S.J. (1999), SFCHECK: a unified set of procedures for evaluating the quality of macromolecular structure-factor data and their agreement with the atomic model. *Acta Cryst.* D**55**, 191-205.

van Doorn W.G., Woltering E.J. (2005), Many ways to exit? Cell death categories in plants. *Trends Plant Sci.* **10**, 117-122.

van Loon L.C., Pierpoint W.S., Boller T., Conejero V. (1994), Recommendations for naming plant pathogenesis-related proteins. *Plant Mol. Biol. Rep.* **12**, 245–264.

Van Loon L.C., Van Strien E.A. (1999), The families of pathogenesis-related proteins, their activities, and comparative analysis of PR-1 type proteins. *Physiol. Mol. Plant Pathol.* **55**, 85-97.

van Loon L.C., Rep M., Pieterse C.M. (2006), Significance of inducible defense-related proteins in infected plants. *Annu. Rev. Phytopathol.* **44**, 135-162.

Vanek-Krebitz M., Hoffmann-Sommergruber K., Machado M.L.C., Susani M., Ebner C., Kraft D., Scheiner O., Breiteneder H. (1995), Cloning and sequencing of *Mal d* 1, the major allergen from apple (*Malus domestica*), and its immunological relationship to Bet v 1, the major birch pollen allergen. *Biochem. Biophys. Res. Commun.* **214**, 538–551.

Wallace A.C., Laskowski R.A., Thornton J.M. (1995), LIGPLOT: A program to generate schematic diagrams of protein-ligand interactions. *Prot. Eng.* **8**, 127-134.

Walter M.H., Liu J.W., Wunn J., Hess D. (1996), Bean ribonuclease-like pathogenesis-related protein genes (Ypr10) display complex patterns of developmental, dark-induced and exogenous-stimulus-dependent expression. *Eur. J. Biochem.* **239**, 281-293.

Walter T.S., Meier C., Assenberg R., Au K.F., Ren J., Verma A., Nettleship J.E., Owens R.J., Stuart D.I., Grimes J.M. (2006), Lysine methylation as a routine rescue strategy for protein crystallization. *Structure* **14**, 1617-1622.

Wang Y.S., Liu D., Wyss D.F. (2004), Competition STD NMR for the detection of high-affinity ligands and NMR-based screening. *Magn. Reson. Chem.* **42**, 485-489.

Warner S.A.J., Scott R., Draper J. (1993), Isolation of an asparagus intracellular PR gene (AoPR1) wound-responsive promoter by the inverse polymerase chain reaction and its characterization in transgenic tobacco. *Plant J.* **3**, 191-201.

Warner S.A., Gill A., Draper J. (1994), The developmental expression of the asparagus intracellular PR protein (AoPR1) gene correlates with sites of phenylpropanoid biosynthesis. *Plant J.* **6**, 31-43.

Weeks C.M., Sheldrick G.M., Miller R., Uson I., Hauptman H.A. (2001), Ab initio phasing by dual-space direct methods. In the proceedings of the 18th European Crystallographic Association Meeting: *Advances in Structure Analysis*, R. Kužel, J. Hašek (Eds.), *Czech & Slovak Cryst. Assn.* Prague, pp. 37-64.

Wink M. (2003), Evolution of secondary metabolites from an ecological and molecular phylogenetic perspective. *Phytochemistry* **64**, 3-19.

Wlodawer A., Minor W., Dauter Z., Jaskolski M. (2008), Protein crystallography for non-crystallographers, or how to get the best (but not more) from published macromolecular structures. *FEBS J.* **275**, 1-21.

Wu F., Li Y., Chang S., Zhou Z., Wang F., Song X., Lin Y., Gong W. (2002), Purification, characterization and preliminary crystallographic studies of a PR-10 protein from Pachyrrhizus erosus seeds. *Acta Cryst.* D**58**, 2165-2167.

Wu F., Yan M., Li Y., Chang S., Song X., Zhou Z., Gong W. (2003), cDNA cloning, expression, and mutagenesis of a PR-10 protein SPE-16 from the seeds of Pachyrrhizus erosus. *Biochem. Biophys. Res. Commun.* **312**, 761-766.

Yan Q., Qi X., Jiang Z., Yang S., Han L. (2008), Characterization of a pathogenesis-related class 10 protein (PR-10) from Astragalus mongholicus with ribonuclease activity. *Plant Physiol. Biochem.* **48**, 93-99.

Yaworsky D.C., Baker B.Y., Bose H.S., Best K.B., Jensen L.B., Bell J.D., Baldwin M.A., Miller W.L. (2005), pH-dependent Interactions of the carboxyl-terminal helix of steroidogenic acute regulatory protein with synthetic membranes. *J. Biol. Chem.* **280**, 2045-2054.

Yun D.J., D'Urzo M.P., Abad L., Takeda S., Salzman R., Chen Z., Lee H., Hasegawa P.M., Bressan R.A. (1996), Novel osmotically induced antifungal chitinases and bacterial expression of an active recombinant isoform. *Plant Physiol.* **111**, 1219-1225.

Zobayed S.M., Afreen F., Goto E., Kozai T. (2006), Plant-Environment Interactions: Accumulation of Hypericin in Dark Glands of Hypericum perforatum. *Ann. Bot. (Lond.)* **98**, 793-804.

Zhou X.J., Lu S., Xu Y.H., Wang J.W., Chen X.Y. (2002), A cotton cDNA (GaPR-10) encoding a pathogenesis-related 10 protein with in vitro ribonuclease activity. *Plant Science* **162**, 629-636.