

Complexes of PR-10 proteins with plant hormones: crystallographic and biophysical studies

Joanna Śliwiak

Doctoral Thesis



Institute of Bioorganic Chemistry

Polish Academy of Science

Scientific adviser:

Prof. Mariusz Jaskólski

Poznań 2017



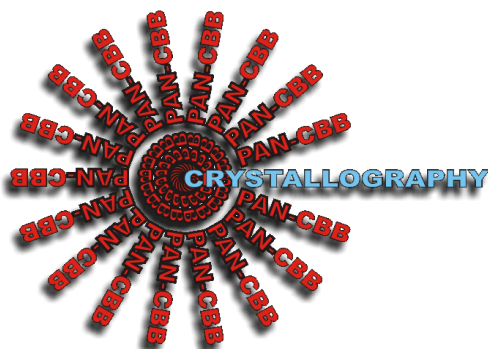
Foundation for Polish Science



**INNOVATIVE
ECONOMY**

NATIONAL COHESION STRATEGY

EUROPEAN UNION
EUROPEAN REGIONAL
DEVELOPMENT FUND



Financial support for this work was provided by the European Union within the European Regional Development Fund. The International PhD Programme was administered by the Foundation for Polish Science.

There is no such thing as a failed experiment, only experiments with unexpected outcomes.

Richard Buckminster Fuller

Contents

1. Abstract	4
1. Streszczenie	5
2. List of original publications submitted as doctoral dissertation and list of structures deposited in the Protein Data Bank	6
3. Introduction	8
3.1. Plant defense.....	8
3.2. PR-10 proteins.....	9
3.3. PR-10 fold.....	11
3.4. Plant hormones as PR-10 ligands	13
4. Aims of the thesis.....	16
5. Methodological approaches for ligand binding verification and characterization.....	17
5.1. Co-crystallization trials.....	17
5.2. Fluorescence measurements and ANS displacement assays.....	19
5.3. Isothermal Titration Calorimetry.....	19
5.4. Circular dichroism.....	20
6. Main results and discussion.....	22
6.1. Change of PR-10 protein conformation upon ligand binding (Publication I).....	22
6.2. Modulated superstructure of Hyp-1 protein in complex with ANS as a case for testing new maximum likelihood methods (Publications II and III)	23
6.3. Hyp-1 protein from St John's wort as a PR-10 protein with novel type of tripartite cavity (Publications III, IV and V)	26
6.4. Melatonin, a new ligand for PR-10 proteins (Publication V)	29
6.5. Calorimetric characterization and determination of binding specificity of Phytohormone Binding Proteins (Publication VI).....	31
7. Summary.....	33
8. Acknowledgments.....	34
9. References	35

1. Abstract

This doctoral thesis presents studies of pathogenesis related proteins of class 10 (PR-10). So far no specific function has been assigned to this group. The PR-10 family consists of small, intracellular biomolecules of acidic character, whose expression is elevated in response to biotic and abiotic stress. The characteristic fold of PR-10 proteins, which consists of seven antiparallel β strands wrapped around an extended, C-terminal α helix, creates a globular protein with a large, hydrophobic cavity, which is capable of accommodating small-molecule ligands. Although many constitutively expressed plant proteins exhibit the PR-10 fold, they are not typically related to pathogenesis. A variety of different ligand classes, including cytokinins, flavonoids and steroids, have been demonstrated so far (via crystallographic and biophysical methods) to occupy the PR-10 cavity.

In the present work, new ligands for PR-10 proteins from St. John's wort and yellow lupine plants have been screened, mainly via the utilization of co-crystallization screens. The identified complexes were additionally characterized by calorimetry, fluorescence spectroscopy and circular dichroism. These studies resulted in the determination of the mechanism of *trans*-zeatin binding by the lupine isoform LIPR-10.1A. Additionally, a modulated superstructure of the Hyp-1 protein in complex with a fluorescent dye was obtained, with as many as 28 protein copies in the asymmetric unit, which served as a case study for developing new statistical methods by the authors of *Phaser* – a program for molecular replacement. Moreover, the studies presented in this dissertation contributed to the discovery of binding of melatonin, yet another biomolecule, whose levels are elevated during unfavorable environmental conditions, by the Hyp-1 protein from St. John's wort and the lupine isoform LIPR-10.2B. Additionally, the thermodynamic parameters of the interaction between PR-10-fold proteins known as "Cytokinin-Specific Binding Proteins", and *trans*-zeatin and gibberellin, were determined by the use of calorimetry. The calorimetric titrations revealed stronger and more specific binding of gibberellin than of *trans*-zeatin by these proteins. Based on the analysis of PR-10 and PR-10-related protein structures deposited in the Protein Data Bank, the PR-10 cavities have been classified into three categories with regard to their shape, volume and binding specificity. Such a classification could be used for predicting the behavior of PR-10 proteins towards different ligands from the available structural information. The presented work significantly widens our knowledge about the PR-10 family, providing novel information about their new natural ligands, their binding mechanisms, and their binding specificity.

1. Streszczenie

Prezentowana praca doktorska przedstawia badania nad białkami związanymi z patogenezą roślin z grupy 10 (PR-10). Białkom tym jak dotąd nie została przypisana jedna, konkretna funkcja. Są to małe, wewnątrzkomórkowe biomolekuły o kwaśnym charakterze, których ekspresja jest podwyższona w odpowiedzi na stres biotyczny i abiotyczny. Ich charakterystyczny zwój składający się z antyrównoległego arkusza β zawiniętego wokół wydłużonej, C-terminalnej helisy α , tworzy globularne białko z dużą hydrofobową wnęką zdolną do wiązania małowcząsteczkowych ligandów. Zwój ten spotykany jest również w roślinnych białkach produkowanych konstytutywnie i nie związanych z patogenezą. Dotychczasowo udokumentowano wiązanie ligandów różnych klas (cytokinin, flowonoidów czy steroidów) wewnątrz PR-10 przy pomocy metod krystalograficznych i biofizycznych.

W niniejszej pracy poszukiwano nowych ligandów dla białek PR-10 z dziurawca zwyczajnego i łubinu żółtego, głównie przy pomocy krystalograficznych testów przesiewowych. Powstałe kompleksy były dodatkowo charakteryzowane przy pomocy kalymetrii, fluorescencji i dichroizmu kołowego. Efektem tych badań było ustalenie kolejności i mechanizmu wiązania *trans*-zeatyny przez łubinową izoformę LIPR-10.1A, czy też modulowana struktura krystaliczna białka Hyp-1 w kompleksie ze znacznikiem fluorescencyjnym, posiadająca aż 28 cząsteczek białka w części asymetrycznej, która posłużyła za poligon doświadczalny do opracowania nowych metod statystycznych przez twórców programu do podstawienia cząsteczkowego "Phaser". Dodatkowo badania w ramach niniejszej rozprawy doktorskiej doprowadziły do odkrycia wiązania melatoniny (której stężenie jest podwyższone w roślinie w czasie niekorzystnych warunków środowiskowych) przez białko Hyp-1 z dziurawca i izoformę łubinową LIPR-10.2B. Metodą mikrokalymetrii zostały także określone parametry termodynamiczne oddziaływań białek o zwoju PR-10 zwanych dotąd "specyficznie wiążącymi cytokininy" (CSBP), z zeatyną i gibereliną, które to oddziaływania scharakteryzowane były dotąd jedynie strukturalnie. Wyniki w/w badań kalymetrycznych potwierdziły silniejsze i bardziej specyficzne wiązanie gibereliny niż zeatyny przez białka CSBP. Dodatkowo praca niniejsza na podstawie zdeponowanych w bazie PDB struktur roślinnych białek PR-10 i im pokrewnych, dokonuje podziału tych białek pod względem kształtu wnęki wiążącej, wyodrębniając 3 typy różniące się wielkością i kształtem wnęki oraz specyficznością wiązania ligandów. Taki podział może pomóc w przewidywaniu zachowania się białek PR-10 wobec różnych ligandów w oparciu o informację strukturalną. Przedstawione badania znacznie poszerzają naszą wiedzę o rodzinie białek PR-10, dostarczając informacji o ich nowych naturalnych ligandach, mechanizmie, oraz specyficzności wiązania.

2. List of original publications submitted as doctoral dissertation and list of structures deposited in the Protein Data Bank

Original publications included in this dissertation:

- I. **Sliwiak, J.**, Dolot, R., Michalska, K., Szpotkowski, K., Bujacz, G., Sikorski, M. & Jaskolski, M. (2016). Crystallographic and CD probing of ligand-induced conformational changes in a plant PR-10 protein. *J. Struct. Biol.* 193, 55-66.
- II. **Sliwiak, J.**, Jaskolski, M., Dauter, Z., McCoy, A. J. & Read, R. J. (2014). Likelihood-based molecular-replacement solution for a highly pathological crystal with tetartohedral twinning and sevenfold translational noncrystallographic symmetry. *Acta Crystallogr.* D70, 471-480.
- III. **Sliwiak, J.**, Dauter, Z., Kowiel, M., McCoy, A. J., Read, R. J. & Jaskolski, M. (2015). ANS complex of St John's wort PR-10 protein with 28 copies in the asymmetric unit: a fiendish combination of pseudosymmetry with tetartohedral twinning. *Acta Crystallogr.* D71, 829-843.
- IV. **Sliwiak, J.**, Dauter, Z. & Jaskolski, M. (2013). Hyp-1 protein from St John's wort as a PR-10 protein. *BioTechnologia* 94, 47-50.
- V. **Sliwiak, J.**, Dauter, Z. & Jaskolski, M. (2016). Crystal structure of Hyp-1, a *Hypericum perforatum* PR-10 protein, in complex with melatonin. *Front. Plant Sci.* 7, a.668.
- VI. Ruszkowski, **M.**, **Sliwiak, J.**, Ciesielska, A., Barciszewski, J., Sikorski, M. & Jaskolski, M. (2014). Specific binding of gibberellic acid by Cytokinin-Specific Binding Proteins: a new aspect of plant hormone-binding proteins with the PR-10 fold. *Acta Crystallogr.* D70, 2032-2041.

Protein structures deposited in the PDB as a result of this research:

1. **5I8F** - Crystal structure of St. John's wort Hyp-1 protein in complex with melatonin
2. **5C9Y** - Crystal structure of yellow lupine LLPR-10.1A protein partially saturated with *trans*-zeatin
3. **4Y31** - Crystal structure of yellow lupine LLPR-10.1A protein in ligand-free form
4. **4N3E** - Crystal structure of Hyp-1, a St John's wort PR-10 protein, in complex with 8-anilino-1-naphthalene sulfonate (ANS)
5. **4RYV** - Crystal structure of yellow lupin LLPR-10.1A protein in complex with *trans*-zeatin

The X-ray diffraction experiments presented in this dissertation were carried out in the following synchrotron centers:

- Advanced Photon Source (APS) Argonne National Laboratory, USA
- BESSY, HZB Berlin
- Petra III, EMBL/DESY, Hamburg
- MAX-lab, Lund

3. Introduction

3.1. Plant defense

Plants are constantly subjected to numerous environmental biotic and abiotic factors, some of which can lead to stress. In order to protect themselves from the damage coming from pathogen attack or environmental challenges, they have developed a wide range of constitutive as well as inducible defense strategies. The former include mechanical barriers such as cell wall, waxes, lignin or epidermis with its specialized guard cells, as well as chemical barriers - phytoanticipins (VanEtten *et al.*, 1994). The latter strategy includes: accumulation of toxic phytochemicals (phytoalexins), a heterogeneous group of low molecular mass secondary metabolites with antimicrobial activity (e.g. 3-deoxyanthocyanins, the flavonoid luteolin, phenolic compound resveratrol or indolic phytoalexins like camalexin); localized cell wall reinforcement; oxidative burst; deliberate cell suicide; as well as expression of so-called pathogenesis-related proteins (Ahuja *et al.*, 2012; Bindschelder *et al.*, 2006; Gilchrist, 1998). Most of the pathogenesis-related (PR) proteins, also known as inducible defense-related proteins, can be induced by salicylic acid, jasmonic acid, ethylene or by wounding, and they possess antimicrobial activities, related to the hydrolysis of bacterial cell wall, contact toxicity or possibly to involvement in defense signaling, among others. Proteins homologous to members of the PR families are found to be developmentally regulated. The divergence of PR proteins suggests that they may play essential but variable roles in plants, both in defense and under normal conditions (Van Loon *et al.*, 2006). The PR proteins have been classified into 17 subclasses (Table 1., modified from Van Loon *et al.*, 2006) according to their biological activity or physicochemical properties and sequence homology.

Family	Properties	Size (kDa)	Member
PR-1	Antifungal	15	Tobacco PR-1a
PR-2	β -1,3-glucanase	30	Tobacco PR-2
PR-3	Chitinase type I, II, IV, V, VI, VII	25-30	Tobacco P, Q
PR-4	Chitinase type I, II	15-20	Tobacco "R"
PR-5	Osmotin	25	Tobacco S
PR-6	Proteinase-inhibitor	8	Tomato Inhibitor I
PR-7	Endoproteinase	75	Tomato P ₆₉
PR-8	Chitinase type III	28	Cucumber chitinase
PR-9	Peroxidase	35	Tobacco lignin-forming peroxidase
PR-10	Unknown	17	Parsley "PR1"
PR-11	Chitinase, type I	40	Tobacco class "V" chitinase
PR-12	Defensin	5	Radish Rs-AFP3
PR-13	Thionin	5	<i>Arabidopsis</i> THI2.1
PR-14	Lipid-transfer protein	9	Barley LTP4
PR-15	Oxalate oxidase	20	Barley OxOa (germin)
PR-16	Oxalate-oxidase-like	20	Barley OxOLP
PR-17	Unknown	27	Tobacco PRp27

Table 1. Families of pathogenesis-related proteins with representative members.

3.2. PR-10 proteins

Among the 17 families of defense-related proteins, there are only two, namely PR-10 and PR-17, whose function is still obscure (Table 1). The PR-10 members are homologous to ribonucleases (Moiseyev *et al.*, 1994) and some have reported, although rather dubious, ribonuclease activity (Wu *et al.*, 2003; Xie *et al.*, 2010; Krishnaswamy *et al.*, 2011; Fernandes *et al.*, 2013), which some authors have associated with their antiviral properties (Park *et al.*, 2004). The PR-10 family is the only subclass of cytoplasmic pathogenesis-related proteins, whereas other PR proteins are secreted to xylem fluid or deposited in vacuoles (Van Loon *et al.*, 2006). Other example of PR proteins not excreted to the lumen include (S)-norcoclaurine synthases (NCS), enzymes that catalyze the condensation of 3,4-dihydroxyphenylethylamine (dopamine) and 4-hydroxyphenylacetaldehyde in the biosynthesis of benzylisoquinoline alkaloids. NCSs are located in the subcellular compartment and show 28-38% sequence identity with other PR-10 members (Samanani *et al.*, 2004). Accordingly, the PR-10 family has been divided into two subfamilies: intracellular pathogenesis-related proteins (IPR) and NCSs (Liu & Ekramoddoullach, 2006). The first PR-10 gene was identified in parsley, after treatment of cell suspension with fungal elicitor (Somssich *et al.*, 1988) and so far more than 100 PR-10 homologs have been identified in more than 70 species of mono- and dicotyledonous plants (Liu & Ekramoddoullah *et al.*, 2006).

The IPR proteins are acidic (pI 4.5-5.5), small (~16-18 kDa) and protease resistant. Their open reading frames are usually interrupted by one intron of 76-359 base pairs, at position which is highly conserved in seed plants (Hoffmann-Sommergruber *et al.*, 1997). The IPR proteins are encoded by multigene families, which could be related to their multifunctional features. They acquired different functions through mutations, which have occurred during evolution as a promiscuity process (Lebel *et al.*, 2010; Franco, 2011). For instance, there are as many as 19 PR-10 homologs in *Malus domestica*, 17 in *Vitis vinifera*, 10 in *Lupinus luteus* or 10 in *Bethula pendula* (Lebel *et al.*, 2010, Handschuh *et al.*, 2007). Multiple copies of PR-10 members, which are present in the form of chromosome clusters, are the result of gene duplications (Liu & Ekramoddoullach, 2006). Diversification of function as a consequence of the existence of multiple PR-10 isoforms is the main reason why no specific common function has been assigned to this group. Specific isoforms within one organism can be induced by different factors, in an organ-specific manner, which suggests their different roles in plants (Agarwal & Agarwal, 2014).

The PR-10 proteins are primarily involved in defense as they are induced by pathogen invasion and abiotic factors such as draught, cold, salinity, copper, oxidative stress, UV radiation or wounding. PR-10 expression can also be triggered by phytohormones, including salicylic acid (SA), jasmonic acid (JA) and its ester, abscisic acid (ABA), ethylene or auxin (AUX) (Hwang *et al.*, 2003; Park *et al.*, 2004; Xie *et al.*, 2010; Jain *et al.*, 2012; Agarwal & Agarwal, 2014). Nevertheless, PR-10 homologs can be also produced constitutively in particular organs or during a specific developmental stage, e.g. in pollen, fruits, dry seeds, roots or senescent leaves (Mogensen *et al.*, 2002; Liu *et al.*, 2006; Barrat & Clark, 1991; Yamamoto *et al.*, 1997; Sikorski *et al.*, 1999).

In addition to the alleged ribonuclease activity, some PR-10 members have been reported to show antimicrobial activity (Park *et al.*, 2004; Liu *et al.* 2006; Andrade *et al.*, 2010), can inhibit papain activity (Andrade *et al.*, 2010) or carry out storage (Goulas *et al.*, 2007) and antifreeze functions (Ukaji *et al.*, 2010). Recent studies have demonstrated that there are PR-10 homologs with aldo/keto reductase activity, capable of neutralizing toxic aldehydes derived from lipid

peroxidation (Jain *et al.*, 2016). Moreover, some members of the PR-10 family show strong allergenic properties, being common pollen and food allergens (Breiteneder & Ebner, 2000).

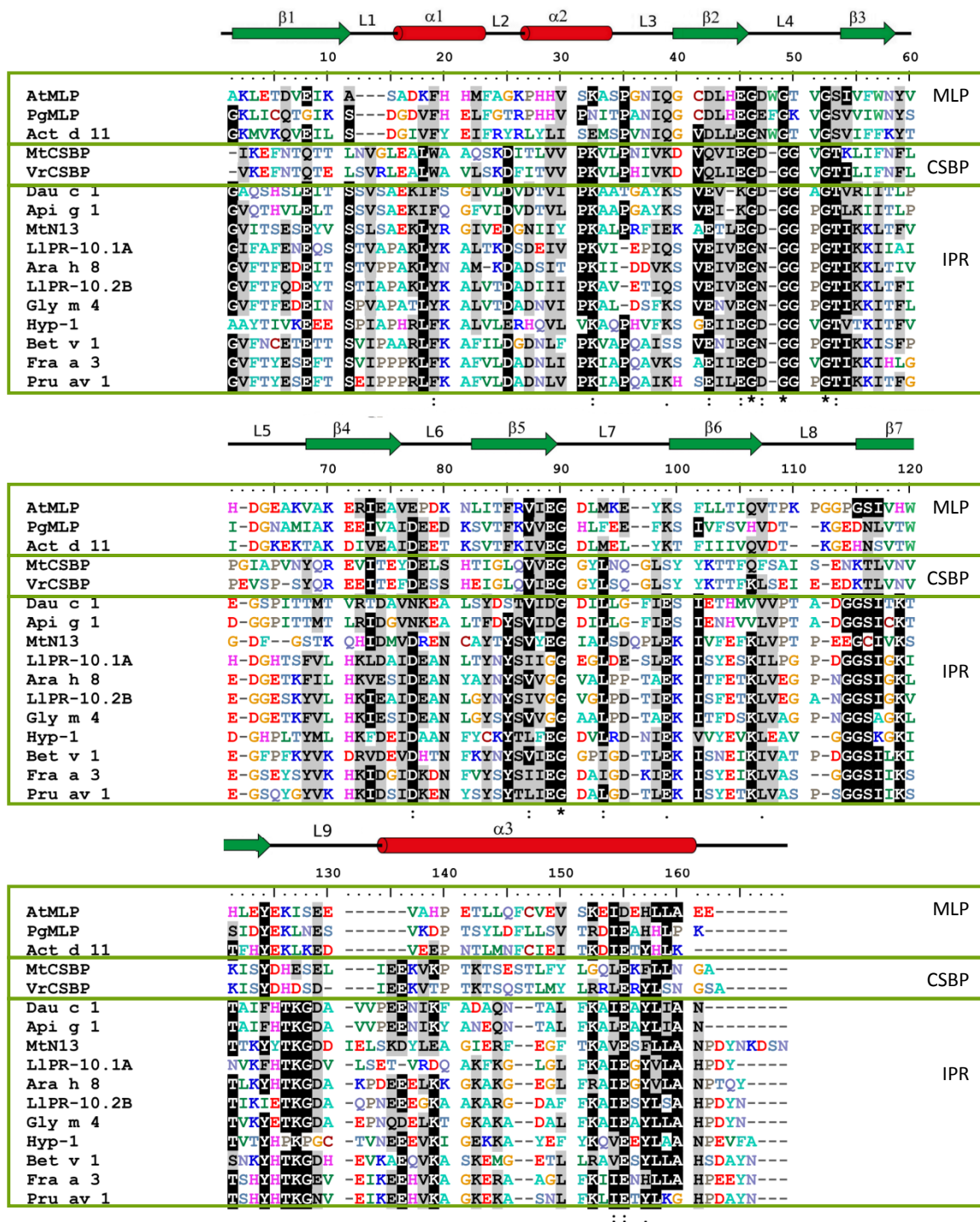


Figure 1. Amino acid sequence alignment of selected IPR, MLP and CSBP proteins of known 3D structure deposited in the Protein Data Bank. The origin of the proteins and PDB codes of their selected representatives are as follows: *Arabidopsis thaliana* Major Latex Protein (2i9y), *Panax ginseng* MLP (4rei), *Acnidia deliciosa* allergen, “kirola” (4igx), *Medicago truncatula* CSBP (4q0k), *Vigna radiata* CSBP (2flh), *Daucus carota* major allergen (2wql), *Apium graveolens* major allergen (2bk0), *M. truncatula* noduline (4jhh), *Lupinus luteus* PR-10 isoform 1A (4ryv), *Arachis hypogaea* panallergen (4ma6), *L. luteus* PR-10 isoform 2B (2qim), *Glycine max* allergen (2k7h), PR-10 protein from *Hypericum perforatum* (5i8f), *Betula verrucosa* major pollen allergen (4a80), PR-10 protein from *Fragaria ananassa* (4c94), *Prunus avium* major allergen (1e09).

Analysis of the PR-10 sequences (Figure 1) shows that alongside the “classic” PR-10 family, there are also families related to PR-10 proteins, including major latex proteins (MLPs) or cytokinin specific binding proteins (CSBPs).

MLPs, first detected in opium poppy latex, were found e.g. in tobacco, thale cress or bell pepper (Osmark *et al.*, 1998). They are closely related to ripening-related proteins (RRPs), e.g. kiwi fruit allergen Act d 11, so called “kirola” (Chruszcz *et al.*, 2013). The MLP/RRP subfamily shares only low sequence identity with the proper PR-10 family (<20%), and the MLP/RRP members are expressed mainly in fruits and flowers. Although the exact function of MLP/RRP proteins is unknown, a potential role in stress response is suggested by their acidic pI similar to that of IPR proteins. Additionally, the MLP protein from bell pepper was found to be induced by wounding (Osmark *et al.*, 1998). Moreover, MLP/RRP genes are also intervallic with an intron.

CSBPs are found in legume plants, such as *Glycine max*, *Lupinus luteus*, *Medicago truncatula* or *Vigna radiata*, and share <20% sequence identity with the proper PR-10 family. This group of proteins was first identified by Fujimoto *et al.* (1998) as strong cytokinin binders with a reported K_d in the nanomolar range. It was later shown by Pasternak *et al.* (2006), however, that VrCSBP indeed binds cytokinins but in a less specific way and with much lower affinity. Thus the name originally assigned to the CSBP proteins turned out to be inadequate. This aspect is explained further with a proposal of a revised terminology in one of the publications (VI) presented here (Chapter 6.5.).

Regardless of the differences in sequence or function, the main common feature of all of PR-10 and PR-10 related proteins is the three dimensional structure of a conserved fold, which is known in the literature as the “Bet v 1 fold” or, more generally, the “PR-10 fold”.

3.3. PR-10 fold

The number of different protein folds is much lower than the number of sequences that can be found in the Protein Data Bank. The reason for that phenomenon is that folds are more conserved than the sequences, which can diverge during the evolution, even below the limit of similarity detection, while retaining the same shape of the protein molecule (Orengo *et al.*, 1993, Holm & Sander, 1995). Moreover, convergent evolution can also lead to similar protein folds since a particular structural motif can be a stable solution for fulfilling an evolutionarily challenged function (Radauer *et al.* 2006). In particular, domains that are specialized in small ligand binding have evolved to accommodate different ligands and still share structural features within their superfamilies (Iyer *et al.*, 2001). It is possible that the same domain, for example a double-stranded β -helix, can be used for both, catalysis and ligand binding (Gane *et al.*, 1998).

The proteins with PR-10 fold have an α/β structure consisting of a helix-grip fold, which is typical of a large, incompletely described superfamily that includes archaeal, bacterial and eukaryotic ligand binding proteins with diverse roles in metabolism and signal transduction (Iyer *et al.*, 2001). Therefore, the PR-10-like topology can be found in many distant homologs e.g. in the START domain, which is widely distributed in bacteria and eukaryotes and is specialized in lipid and steroid transport; in polyketide cyclases (aromatases), i.e. enzymes that bind linear polyketides and sterically induce their cyclization; in hydroxynitrile liases, catalyzing cyanide release and identified recently in ferns (PDB code: 5e4m, Lanfranchi *et al.*, unpublished data); in the plant abscisic acid receptor PYL1; or many other poorly characterized bacterial proteins (Iyer *et al.*, 2001; Radauer *et al.*, 2008).

For instance, when one submits the PDB entry 2qim (the structure of *L. luteus* isoform LIPR-10.2B) for 3D alignment at the PDBeFold service (Krissinel & Henrick, 2004) against the whole PDB archive with the lowest acceptable match of 70%, one obtains hits with 257 polypeptide chains from 133 PDB entries of 74 proteins with unique sequence, 17 of which belong to the PR-10 family, 4 are MLPs and 2 are CSBPs. Moreover, there are also 6 abscisic acid receptors, 3 lipid transporting proteins, 2 polyketide cyclases, 1 hydroxynitrile liase and as many as 39 uncharacterized bacterial proteins in that list.

The PR-10 fold is formed by a seven-stranded antiparallel β sheet wrapped around a long C-terminal helix α_3 helix, which, together with two additional short α -helices (α_1 and α_2), which support the C-terminal end of helix α_3 , create a globular protein with a hydrophobic, empty interior that can be accessed by two main entrances: E1 and E2 (Fig. 2a, b). The shape of a typical PR-10 fold can be compared to a gripping baseball glove. Four odd-numbered (L3, L5, L7, L9) of the numerous loops form the fingertips of the gripping glove. Of particular importance is loop L9, which is the point of entry for helix α_3 .

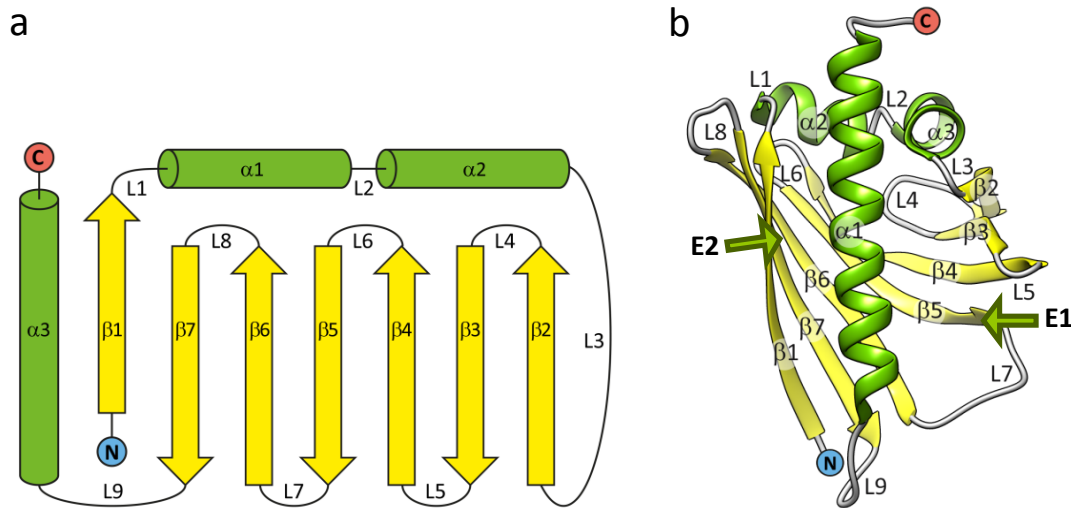


Figure 2. (a) PR-10 topology diagram, the β -strands (yellow arrows) and α helices (green cylinders) are connected by structurally important loops. (b) The general PR-10 fold; the secondary structure elements and the two cavity entrances E1 and E2 are labeled.

As revealed by sequence alignments of PR-10 members (Fig. 1), the middle part of helix α_3 seems to be the least conserved structural element of the PR-fold. Interestingly, it also has the most variable conformation (distortions of regular helical structure) and is responsible for shaping the interior of the internal cavity (Fernandes *et al.*, 2013). Additionally, loops L3, L5 and L7 also show low sequence conservation. Intriguingly, these variable structural elements surround the two cavity entrances, E1 (L3, L5, L7 and the middle part of helix α_3) and E2 (the middle part of helix α_3 together with strand β_1). The possibility thus suggests itself that the sequence variability of these structural elements may play a role in the recognition of different ligands.

The most conserved sequence, even in distant PR-10 homologs, like CSBPs, is that of loop L4, located between strands β_2 and β_3 . It is termed the “glycine rich loop” and has the following sequence: EG(D/N)GG(V/P)GT. In MLP/RRP this loop is less conserved and has slightly different

sequence: EGx(W/F)G(T/S)VGS. Although loop L4 shows high sequence similarity to the phosphate binding loop (P-loop), which is found in nucleotide binding proteins (Saraste *et al.*, 1990), PR-10 proteins do not show affinity for ATP (Koistinen *et al.*, 2005) and the conformation of the glycine-rich loop is different from that of typical P-loops (Biesiadka *et al.*, 2002). Site directed mutagenesis at the glycine-rich loop in a PR-10 protein from pea (Krishnaswamy *et al.*, 2011) affected the ribonuclease and antifungal activity suggesting a role of the L4 loop in e.g. RNA binding. Recent studies (Jain *et al.*, 2016) show that some PR-10 proteins, e.g. chickpea CaARP, have AKR motifs on the protein surface near loop L4, responsible for their aldo/keto reductase activity. The above authors postulate that the glycine-rich loop could be the binding site for the reduced form of nicotinamide adenine dinucleotide phosphate (NADPH). This cofactor could be utilized for reducing cytotoxic aldehydes, such as methylglyoxal, which is derived from lipid peroxidation. Such an activity would be a benefit during abiotic stress.

3.4. Plant hormones as PR-10 ligands

The Bet v 1 or PR-10 fold is an ideal structural solution for binding/transporting small, hydrophobic molecules, for instance, as it is utilized in lipid and steroid transport by proteins like STAR domains, in poliketide synthases, which bind linear polyketides before their cyclization or in NCSs, which accommodate dopamine and 4-hydroxyphenylacetaldehyde before the catalysis (Iyer *et al.*, 2001; Radauer *et al.*, 2008). Indeed, numerous biochemical and crystallographic studies carried out for the Bet v 1 allergen from birch pollen revealed its ability to bind fatty acids, flavonoids, steroids and cytokinins (Mogensen *et al.*, 2002; Kofler *et al.*, 2012). The main question, which is difficult to answer by *in vitro* studies, regards the biological relevance of binding of a particular molecule. The PR-10 homologs occur in plant organisms in multiple copies/isoforms and this is clearly related to their diversification. Expression of different homologs in different organs and under different conditions means that the moderate differences in sequence may have consequences in developing sufficient affinities for binding of particular ligands in specific situations (for example, at particular phytohormones ratios). Therefore, *in vitro* ligand screening, ligand competitive assays, and structural characterization of ligand complexes of PR-10 isoforms can together provide information about the preferential binding partners and thus hint at possible physiological roles of particular PR-10 homologs.

It has been discovered recently that PYL-1, which is a component of the receptor of ABA (Fig. 3), a plant hormone that mediates adaptation to abiotic stress and regulates developmental signals such as seed maturation, shares its fold with the PR-10 family (Miyazono *et al.*, 2009, 3kdi). This discovery has raised the question whether other PR-10 and PR-10 related proteins could serve as phytohormone receptors and be involved in plant hormonal regulation.

Cytokinins are plant hormones that promote cell division and differentiation in various developmental processes and play crucial role in defense. Additionally, they also regulate the nodulation process in legumes (Hwang *et al.*, 2012). Naturally occurring cytokinins are N⁶-substituted adenine derivatives. Interestingly, some urea derivatives, such as diphenylurea (DPU), act as artificial cytokinins (Fig. 3, Ricci & Bertolotti, 2008).

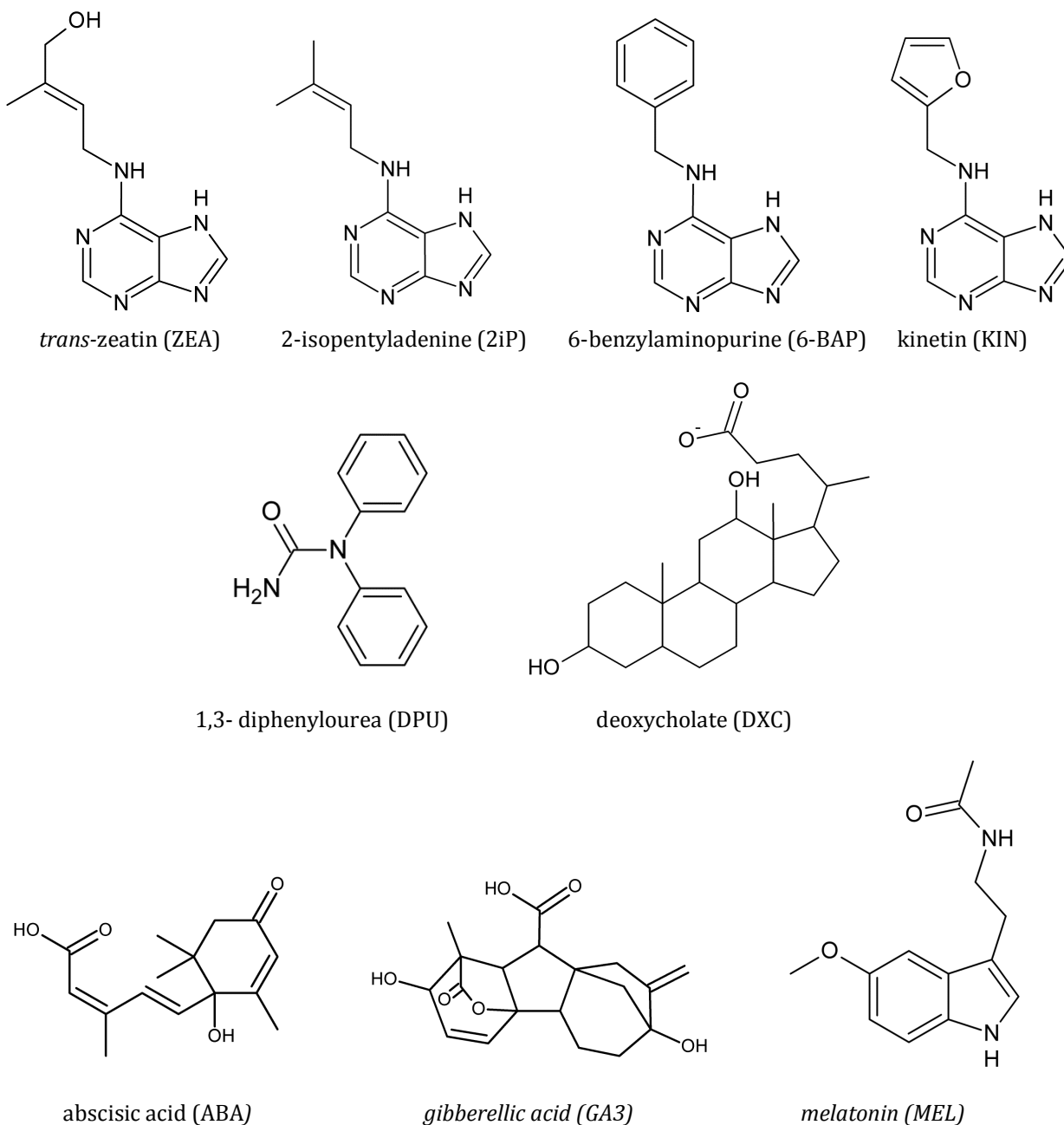


Figure 3. Plant hormones and their analogs reported to form crystalline complexes with proteins of PR-10 fold. GA3 and MEL (*in italics*) have been added to this list as a result of the present work, as described in Chapters 6.4. and 6.5.

MtN13, an *M. truncatula* noduline, which is expressed in nodule cortex in early stages of nodulation and belongs to the PR-10 family based on ~40% identity, appeared to bind cytokinins (*trans*-zeatin, kinetin, 6-benzylaminopurine and 2-isopentyladenine, Fig. 3) in a highly specific manner, as demonstrated in a crystallographic study presented by Ruszkowski *et al.* (2013; PDB IDs 4jhg, 4jhh, 4jhi, 4gy9). Cytokinins however had been identified as PR-10 ligands before. The CSBP proteins, described in Chapter 3.2., were the first proteins reported as strong and specific cytokinin binders, although later research showed that they bind cytokinins with low affinity (Pasternak *et al.*, 2006; 2flh). Crystallographic studies of the yellow lupine isoform LIPR-10.2B resulted in a high resolution structure of its *trans*-zeatin complex (ZEA) (Fernandes *et al.*, 2008; 2qim), although the multiple copies of the ZEA molecules in the binding

cavity indicated cytokinin storage rather than receptor function of this PR-10 homolog. Additionally, this isoform also binds the artificial cytokinin DPU, but in a different mode than in the case of ZEA (Fernandes *et al.*, 2009; 3e85). Bet v 1 also binds cytokinins, as shown by biochemical assays. However, the structure of Bet v 1 in complex with kinetin (Kofler *et al.*, 2012, 4a85) rather contradicts its relevance in cytokinin signaling, as the ligand electron density is very poor and direct contacts between the ligand and the protein are tenuous and unconvincing.

Brassinosteroids (BRs) promote cell expansion and elongation (Clouse & Sasse, 1998) and are necessary for pollen tube formation (Hewitt *et al.*, 1985). There are no structural reports about PR-10 complexes with proper BRs. However, the major birch pollen allergen Bet v 1 has been studied in complex with deoxycholate (DXC, Fig. 3), which shares the chemical scaffold with brassinosteroids. The crystal structure of this complex shows that the DXC backbone fits very well in the cavity of the Bet v 1 protein (Kofler *et al.*, 2012, 4a83), strongly suggesting the possibility of interaction with BRs.

4. Aims of the thesis

Plant intracellular pathogenesis-related proteins of class 10 are a large group of multi-gene proteins without a well defined function. The growing number of structural and biochemical studies seem to suggest that they form a group of homologs with diversified functions as well as with different tissue- and condition-dependent expression profiles. Their common feature is a characteristic fold with an internal cavity, that seems ideal for binding/transporting of small-molecule ligands. Some subgroups of PR-10 related proteins seem to have evolved to perform very specific function as phytohormone receptors or as enzymes. This suggests that the best approach to investigating the PR-10 family would be to study each representative subgroup of homologs separately, by screening them against new ligand libraries and by analyzing the complexes structurally.

The goal of my thesis has been the structural and biophysical characterization of PR-10 phytohormone complexes in an effort to elucidate the binding mechanism and to search for new physiologically relevant ligands. PR-10 proteins from several plants have been studied, namely from *Hypericum perforatum* (Hyp-1), *Lupinus luteus* (LIPR-10.1A, LIPR-10.1B and LIPR-10.2B), *Medicago truncatula* (MtCSBP) and *Vigna radiata* (VrCSBP). Ligand binding capacity was screened using conditions of commercial crystal screens and by co-crystallization with different phytohormones and natural ligands. The diffraction experiments on crystalline PR-10 complexes were carried out using synchrotron X-ray radiation, and the crystal structures were deposited in the PDB. The complexes were also characterized using several biophysical methods, such as calorimetric titration, circular dichroism and fluorescence spectroscopy.

5. Methodological approaches for ligand binding verification and characterization

5.1. Co-crystallization trials

In my work, the method of choice used for searching for new ligands of the PR-10 isoforms was co-crystallization screening with selected phytohormones and plant mediators related to stress response. The tests were carried out for the following recombinant PR-10 proteins purified from bacterial cell cultures: Hyp-1, LIPR-10.1A, LIPR-10.1B and LIPR-10.2B. The protein preparations were incubated with selected ligands prior to setting up crystallization trials. Nearly all classes of plant hormones were tested except for strigolactone, because of its instability, and brassinolide, because of solubility problems. Ethylene was also excluded from the list because of the technical challenges of forming a protein complex with a gaseous ligand. Moreover, melatonin was added to this list because apart from antioxidant properties it also exhibits the properties of a plant hormone, as further discussed in Chapter 6.4. Additionally, plant mediators of special significance in stress conditions were tested, namely flavonoids. In the case of Hyp-1, the substrate (emodin) and product (hypericin) of the hypothetical hypericin biosynthetic reaction (Bais *et al.*, 2003) were also taken into account. It is important to note that not all selected ligands were co-crystallized with every studied PR-10 protein so far, as summarized in Table 2.

Crystalline complexes of the LIPR-10.2B protein with 8-anilino-naphthalene-1-sulfonic acid (ANS), N⁶-(2-isopentenyl)adenine (2iP), N-(2-chloro-4-pyridyl)-N'-phenylurea (CPPU) and melatonin (MEL) as well as with the flavonoids naringenin (NAR) and quercetin (QUE) and with apigenin glucoside (vitexin, VIT) are not described in this work. However, they were included in Table 2 to highlight the differences in the binding capacities of the apparently similar yellow lupine PR-10 isoforms. The X-ray diffraction experiments of LIPR-10.2B crystalline complexes with ANS, 2iP and CPPU resulted in structures of low resolution and with poor ligand electron density, thus the growth conditions of these crystals need further optimization. The structures of the LIPR-10.2B complexes with flavonoids and melatonin, which have well defined ligand electron densities, allowing the determination of the ligand position and interactions with protein residues, will be the topic of future studies. Flavonoids have no hormonal role in plants thus complexes of the lupine isoform with these molecules require a separate discussion. The structure of the LIPR-10.2B/MEL complex on the other hand revealed, alongside two MEL molecules, an unambiguous presence of a MEL derivative in the protein cavity, whose nature remains to be determined. Also, the role of the protein in the processing of the ligand needs further investigation.

In all available literature reports about ligands bound by PR-10 proteins, *trans*-zeatin (ZEA) is the most frequently reported binding partner. Therefore, I used ZEA in competitive crystallization assays to test its ability to perturb MEL complex formation of Hyp-1 and LIPR-10.2B. These proteins were incubated with equimolar solutions of both ligands before submitting them for crystallization in the final growth conditions established for the crystals of the protein/MEL complexes.

Ligand	Hyp-1	LIPR-10.1A	LIPR-10.1B	LIPR-10.2B
<i>trans</i> -zeatin (ZEA)	-	+	-	+
N6-(2-isopentenyl)adenine (2iP)	-	-	-	+*
kinetin (KIN)	-	-	-	-
1,3-diphenylurea (DPU)		-	-	+
N-(2-chloro-4-pyridyl)-N'-phenylurea (CCPU)		-	-	+*
salicylic acid (SAL)	-	-	-	-
abscisic acid (ABA)	-	-	-	-
indole-3-acetic acid (AUX)	-	-	-	-
gibberellic acid (GA3)	-	-	-	-
methyl jasmonate (MeJA)	-			-
melatonin (MEL)	+	-	-	+
2-hydroxymelatonin (2HM)				-
serotonin (SER)	-			-
naringenin (NAR)		-	-	+
quercetin (QUE)	-	-	-	+
apigenin glucoside - vitexin (VIT)				+
emodin (EMO)	-			
hypericin (HYP)	-			
8-anilinonaphthalene-1-sulfonic acid (ANS)	+	-	-	+*

Table 2. Results of co-crystallization screening experiments of four PR-10 proteins with phytohormones and other ligands with the use of Crystal Screen I and II and Peg Ion Screen I and II (Hampton Res.). +/- denote presence/absence of crystalline complex. The shaded areas mark protein/ligand combinations that were not tested. * Asterisk marks crystal structures with poor ligand electron density.

Other crystallization-oriented approaches, such as protein purification in the presence of the ligand, or soaking the crystals of ligand-free protein in a ligand-containing buffer, were not applied. The simple reason for resigning from the former method is that large amounts of the ligands (to be used as additives of purification buffers) are usually not available. The problem that has led to rejecting the latter approach was the fact that some of the studied proteins (e.g. LIPR-10.2B or Hyp-1) do not form crystals in ligand-free form. Moreover, the soaking procedure often dramatically reduces the resolution limit of the crystals.

After diffraction experiments (with the use of synchrotron radiation), structure solution (through molecular replacement) and refinement, the structures of the obtained complexes were carefully analyzed and compared.

5.2. Fluorescence measurements and ANS displacement assays

8-Anilino-naphthalene-1-sulfonic acid (ANS) is a fluorescent dye, whose fluorescence emission is highly sensitive to the polarity of the environment. A blue shift of its fluorescence together with dramatic intensity increase are observed when the environment is changed from polar to non-polar. ANS molecule interacts with proteins primarily by ion pairing (salt bridges) between its negatively charged sulfonate group and positively charged amino acid residues, Lys, His or Arg. The ion pairs must be stabilized by additional complementary interactions, usually by van der Waals forces (Gasymov & Glasgow, 2007). These properties of ANS were utilized to measure the affinity of ANS for Hyp-1 in solution by monitoring the increase of ANS fluorescence in the presence of increasing concentration of the protein. ANS can also be used in so-called ANS displacement assays, or ADA, where ANS displacement by a ligand of choice is coupled with the monitoring of ANS fluorescence decay. ADA can serve as a powerful method for ligand affinity measurements, however, it is crucial to obtain accurate structural data of its protein complex for proper interpretation of the ADA results and to verify that the ANS binding site is the same as that of the studied ligand.

Therefore, I started my studies of Hyp-1 and yellow lupine PR-10 isoforms by preparing their crystalline complexes with ANS. Two lupine isoforms from subclass 1 gave no crystals in ANS co-crystallization screens, while LIPR-10.2B co-crystallized with ANS producing crystals that diffracted X-rays to 2.0 Å resolution. However, the electron density indicated that the ligand is highly disordered in the crystal structure (J. Sliwiak, unpublished results). Only the Hyp-1/ANS crystal structure provided accurate structural information about the location and character of ANS binding. Briefly, the structure (deposited in the PDB as 4n3e) revealed that apart from three internal binding sites, there are also external, interstitial ANS binding sites, as discussed in Chapter 6.3. Such external binding sites were not detected in the case of the structure of the Hyp-1 protein in complex with the natural ligand melatonin (Chapter 6.4.). These additional ANS sites could explain the negative ADA results, where attempts to displace ANS from its Hyp-1/ANS complex with melatonin or *trans*-zeatin resulted in a fluorescence change that was too low to obtain the binding curve and affinity parameters. Although the ANS displacement assay carried out for Hyp-1 was unsuccessful, the crystal structure of the Hyp-1/ANS complex turned out to be fiendishly complicated, as it represents a rare case of modulated superstructure in macromolecular crystallography. The translational non-crystallographic symmetry (tNCS) of those crystals, interpreted as commensurate modulation, became the testing ground for developing new statistical methods for molecular replacement by the authors of the *Phaser* (McCoy *et al.*, 2007) program (Chapter 6.2.).

5.3. Isothermal Titration Calorimetry

Isothermal Titration Calorimetry (ITC) is a method that allows obtaining the thermodynamic parameters of bi-molecular interactions in aqueous solutions. By measuring the input of power required to maintain the same temperature in the reference and the sample cells

after each injection of a ligand aliquot, one can obtain the binding affinity (K_a), stoichiometry (N) as well as enthalpy and entropy changes (ΔH and ΔS respectively) of a molecular interaction.

In addition to crystallization screening, ITC was also used by me for the verification of binding between selected phytohormone ligands (from Table 2) of sufficient solubility and PR-10 proteins from lupine and St John's wort. The experiments yielded a measurable heat effect only for the interactions of LIPR-10.1A and LIPR-10.2B with ZEA. Even in those cases, however, the effect was too weak for unambiguous affinity determination. A sigmoidal titration curve was obtained, though, for the titration of Hyp-1 with ANS, which allowed the determination of the binding parameters for this interaction.

The method of ITC requires a measurable enthalpy change upon complexation, which may come, e.g. from the formation of hydrogen bonds. Thus, in the case of entropy-driven binding, based mainly on hydrophobic interactions, it is often difficult to observe a heat effect that would be sufficient for the determination of the binding parameters. Therefore, yellow lupine PR-10 isoforms with type II cavity (see section 6.3), where ligands are accommodated mainly by hydrophobic interactions, appeared to be very difficult to study using the ITC method. In the case of the Hyp-1/ANS complex, where hydrophobic interactions play an important role, it was still possible to use the ITC method thanks to the salt-bridge interaction between the ANS sulfonate group and Lys8, Arg27 and Lys33.

In the case of CSBP's (now Phytohormone Binding Proteins), the goal of my work was to establish the thermodynamic parameters of the interaction of these proteins with GA3 and ZEA by ITC calorimetry and to correlate the results with structural information. The small type I cavities (see section 6.3) of the VrPhBP and MtPhBP proteins accommodate the hydrophilic GA3 molecule by creating numerous hydrogen bonds and water bridges, as shown by the respective crystal structures determined at high resolution (4psb and 4q0k, respectively). It was, therefore, possible to determine the binding parameters of these mostly enthalpy-driven interactions. The details of these findings are further discussed in Chapter 6.5.

5.4. Circular dichroism

Circular dichroism, or CD, is a very sensitive and rapid method for detecting changes in the content of secondary structure of proteins that occur in the presence of denaturing agents or cofactors (Whitmore & Wallace, 2008). Different secondary structure elements of the protein chain give rise to different CD spectra. The α -helix, for example, gives rise to two negative bands at 222 and 208 nm and a positive band at 193 nm, whereas β sheets are manifested by one negative band at 218 nm and one positive band at 195 nm. A change in secondary structure content is typically reflected by a corresponding change of the respective band intensity. I used this method to monitor the ordering of the extended C-terminal helix α_3 of the LIPR-10.1A protein in the presence of increasing concentration of *trans*-zeatin. Such ordering was noted in a set of three crystal structures of this isoform, consisting of the free form, a form partially saturated with ZEA, and a form fully saturated with five molecules of this ligand, as discussed in Chapter 6.1. Inspection of the CD spectra of LIPR-10.1A recorded in the presence of increasing molar excess of ZEA revealed an increase of the intensity of the band at 222 nm, suggesting an increase of α -helix content. The increase reached a plateau at five-fold molar excess of ZEA, consistent with the crystal structure of LIPR-10.1A fully saturated with ZEA.

All the briefly summarized biophysical methods above were used to characterize the crystalline PR-10/ligand complexes obtained in this work. These methods could not be used as prescreening before the crystallization trials, owing to the fact that not all ligands had suitable spectral properties, solubility or interactions with the proteins producing sufficiently large enthalpy change. Therefore, crystallization screening after protein incubation with a set of ligands was the method of choice for identifying the binding partners. A similar approach has been reported recently independently by Schiebel *et al.* (2016), who demonstrated that six prescreen biophysical methods missed as many as 44% of crystallographically discovered small ligands.

The ability of crystal formation by a protein in the presence of a ligand may be treated as a sign of a stabilizing effect of the ligand on the protein fold, as well as an indicator of mutual affinity. Moreover, crystal structures provide accurate information about protein-ligand interactions, which can be used to elucidate the binding mechanism and for comparative studies.

6. Main results and discussion

6.1. Change of PR-10 protein conformation upon ligand binding

(Publication I)

The work presented in paper I is the first comparative crystallographic study of the structures of the same PR-10 protein in ligand-free form and in complex with a physiologically relevant natural ligand, with emphasis on the structural adaptation of the protein for ligand binding and on elucidation of the binding mechanism. It is noteworthy that the capacity for natural ligand binding was screened using three lupine PR-10 isoforms: LILPR-10.1A, LILPR-10.1B and LIPR-10.2B. The screening resulted in the observation that representatives of subclass 1 of the lupine isoforms are much less promiscuous in ligand selection than members of subclass 2, as LIPR-10.2B was the only protein capable of creating crystalline complexes with natural cytokinins (ZEA, 2iP), their analogs (DPU, CPPU), flavonoids, flavonoid glucoside and melatonin, whereas LIPR-10.1A formed a crystalline complex only with *trans*-zeatin and LIPR-10.1B with none of the tested molecules.

The protein analyzed in paper I, the yellow lupine isoform LIPR-10.1A, was crystallized in free form and in two stoichiometric complexes with *trans*-zeatin, leading to three high resolution structures, namely: of the free (F) form, unsaturated with *trans*-zeatin (U), and in *trans*-zeatin-saturated (Z) state.

The first sign of the impact of the ligand presence was noticed in crystal packing. The highest ligand:protein ratio used for the crystallization of the Z complex changed the intermolecular interactions by creating additional superficial binding sites in a crystal structure that is different from that of the F and U forms. The state of ligand saturation is also revealed by excellent electron density of the three cavity-docked ZEA molecules. In contrast, in the U form the electron density of two ZEA binding sites is poor. The F/U/Z set of three crystal structures allowed me to track the conformational changes occurring upon ligand binding. They include: (i) gradual ordering of helix $\alpha 3$; (ii) change of curvature of loops L3, L5 and L7, which tighten their grip around the E1 entrance after ligand binding; and (iii) gradual cambering of the β -sheet padding the bottom of the cavity. Moreover, the availability of the U form allowed me to elucidate the mechanism of ZEA binding and the cooperation of the LIPR-10.1A binding sites. This was possible because I noticed that in the panorama of the F-U-Z structures the shape of the cavity clearly demonstrates that the ZEA3 binding site is formed only after docking of the ZEA1 and ZEA2 molecules (Fig. 4a). Moreover, thanks to the high resolution of all three structures, I could also track the cooperative movements of some particular residues belonging to the $\alpha 3$ helix that prepare the ZEA3 binding site and which contribute to the ordering of the $\alpha 3$ helix (Fig. 4b). The ordering of helix $\alpha 3$ seems to control and drive the binding process, whereas the final $\alpha 3$ conformation effectively blocks the routes of ligand escape. This structural element of the PR-10 fold, which has the most variable sequence of the protein, has been thereby demonstrated to be crucial for natural ligand recognition. In agreement with these structural conclusions, an increase of helical content was also observed by circular dichroism during LIPR-10.1A titration with *trans*-zeatin.

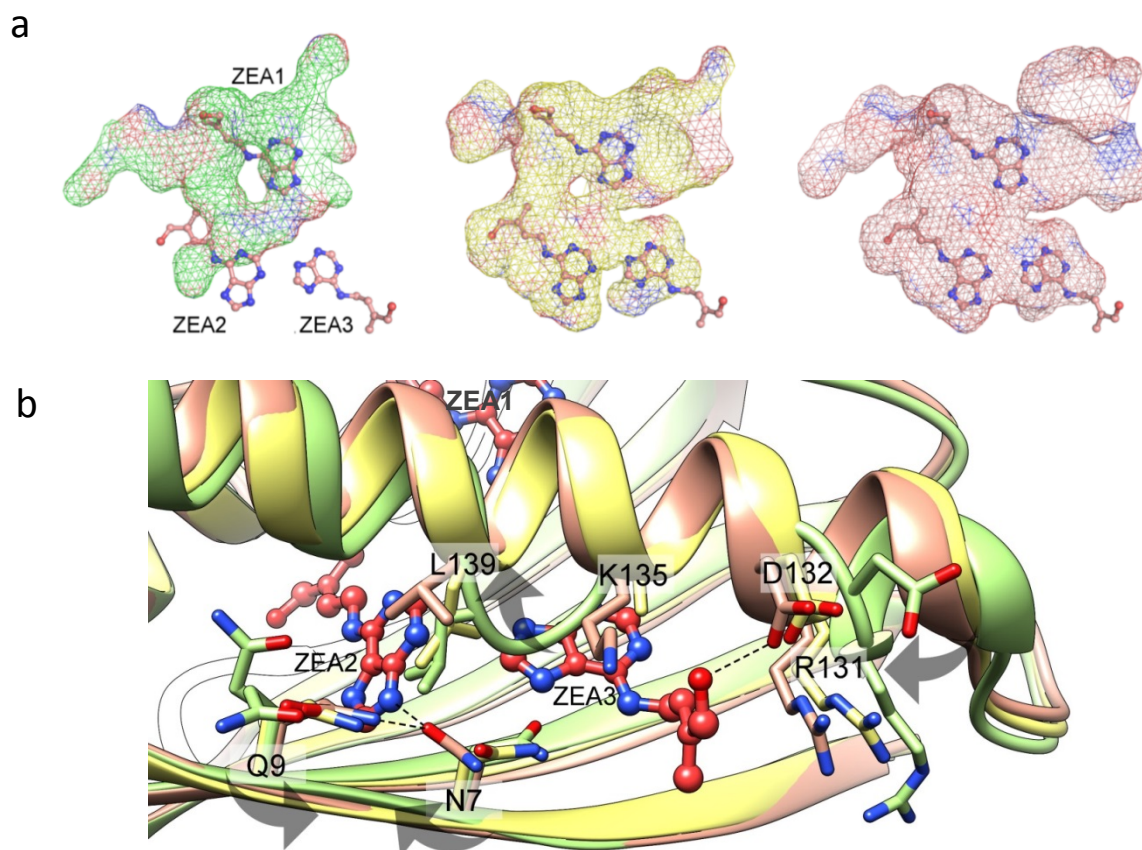


Figure 4. (a) The shape of the internal cavity (van der Waals surface, mesh) of the LIPR-10.1A protein at three *trans*-zeatin saturation states: free F (green), unsaturated U (yellow), and saturated Z (red). The ZEA molecules have been superposed from the saturated structure Z to illustrate the gradual formation of the binding sites. (b) Cooperativity of residues forming the ZEA2 and ZEA3 binding sites; the gray arrows indicate the sequence of side-chain movements that create the ZEA3 binding site and order helix $\alpha 3$ (green – free form F, yellow – unsaturated form U, salmon saturated form Z). (from Sliwiak *et al.*, 2016; Publication I)

This study, which presents the first in-depth discussion of structural transformations of a plant protein with PR-10 fold upon ligand binding, sheds new light on the binding mechanism, and opens new possibilities for comparative analyses of PR-10 homologs from yellow lupine and from other organisms. Moreover, the results of the co-crystallization screening experiments carried out for representatives of two subclasses of lupine PR-10 isoforms and the wide range of natural ligands tested, provide new information about different ligand binding capabilities of these two subclasses of PR-10 proteins.

6.2. Modulated superstructure of Hyp-1 protein in complex with ANS as a case for testing new maximum likelihood methods (Publications II and III)

A “side effect” of the numerous co-crystallization screens with PR-10 proteins was the crystallization of a very interesting but highly complicated crystal form, which turned out to be an excellent test case for the developers of new algorithms for protein crystal structure solution by molecular replacement, as implemented in the *Phaser* program (McCoy *et al.*, 2007). The crystals of Hyp-1 in a complex with ANS diffract X-rays in a highly unusual way, with strong sevenfold repetitive modulation of the reflection intensities along the c^* direction, manifested as strong main reflections separated by much weaker satellites (Fig. 5a and b). This phenomenon

was diagnosed as resulting from a commensurate structure modulation, equivalent to sevenfold extension of the crystal unit cell in the longest c direction. Since the modulation could be interpreted as commensurate (Lovelace *et al.*, 2008), it was possible to describe the structure with sevenfold non-crystallographic translation of the basic packing unit (consisting of four Hyp-1 molecules) in the longest unit cell dimension. In other words, the crystal packing could be interpreted as an extreme case of translational non-crystallographic symmetry (tNCS).

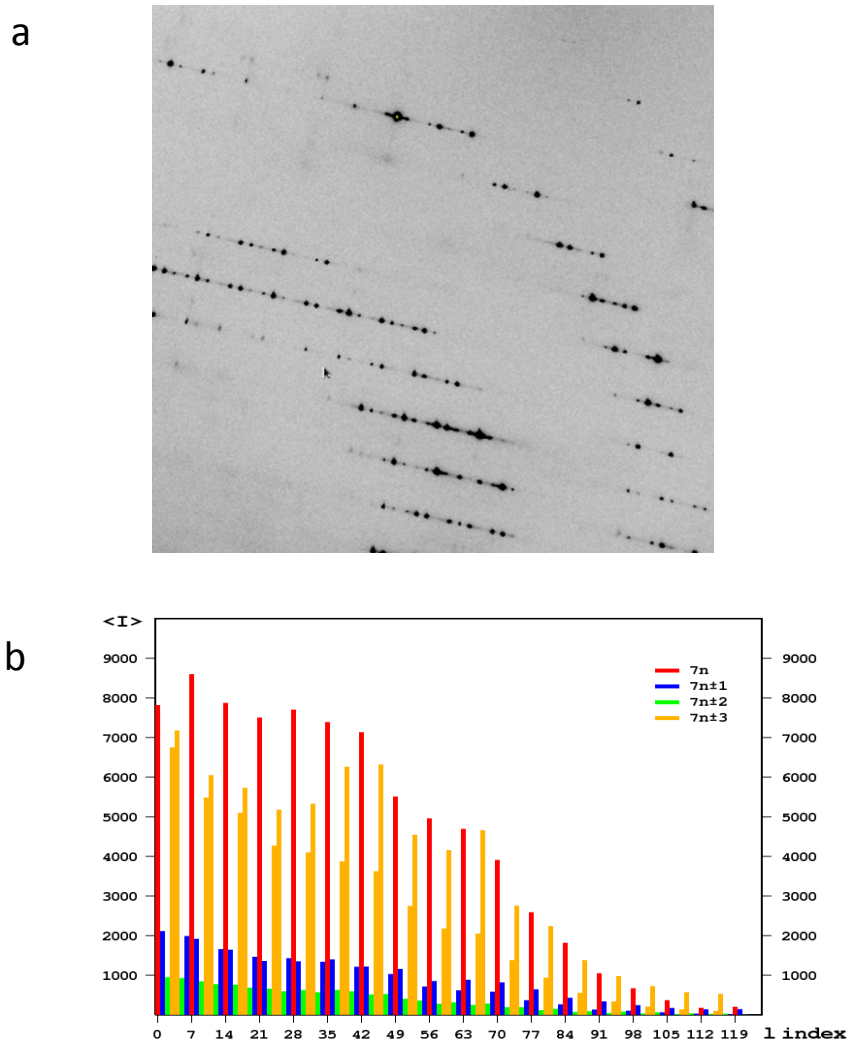


Figure 5. (a) Zoom-in view of an X-ray diffraction image of Hyp-1/ANS crystal; (b) A histogram of intensity distribution in layers of l (from Sliwiak *et al.*, 2015; Publication III).

Structure solution by standard maximum-likelihood (ML) molecular replacement methods was impossible, as they assume uniform structure factor distribution. Therefore, it was necessary to adapt the *Phaser* ML algorithms for the statistical effects of pseudotranslation.

The diffraction data were originally interpreted in (and actually collected with the assumption of) 422 symmetry. However, it turned out during structure analysis that this apparently high symmetry was the result of a nearly ideal pseudomerohedral twinning. A tentative solution in the $P4_122$ space group could not be refined to an R factor better than 48%, suggesting that the true symmetry of the investigated crystal structure was lower.

Since the true symmetry of the crystal structure could not be deduced from the diffraction data, the molecular replacement analysis of the crystal structure was carried out in

P1 space group, after expansion of the diffraction data to triclinic symmetry. In consequence, the MR algorithm with the tNCS-corrected ML targets found 56 copies of the Hyp-1 model in the asymmetric unit. Careful analysis of the symmetry of that solution revealed the true *C2* symmetry of the crystal structure. However, reprocessing of the data in that symmetry resulted in very low completeness, as the data collection protocol had been (incorrectly) adjusted for the 422 symmetry. Fortunately, it was possible to take advantage of the nearly perfect twinning and expand the 422-processed (and thus perfectly twinned) data to monoclinic symmetry. Ultimately, the structural *C2* model containing 28 protein molecules in the asymmetric unit was refined to a very satisfactory *R* factor of 22.3%.

The F_o-F_c electron density map was of excellent quality and allowed the identification of as many as 89 ANS molecules in the asymmetric unit. The Hyp-1 protein molecules within this asymmetric unit are arranged in groups of four and this pseudotetragonal packing is repeated seven times along the *c* axis (Fig. 6). The protein molecules are paired into dimers by $\beta 1-\beta 1$ interactions and the dimers are rotated by $\sim 180^\circ$ and translated by $\sim 1/14$ of the *c* parameter. If this was the end of the structure description, one could conclude that the symmetry of the structure solution should be higher and there was no need for such a big asymmetric unit. However, a detailed analysis of the structure packing reveals peculiar abnormalities, which are generated by the distribution of the ANS molecules and make each of the four columns of seven Hyp-1 copies different.

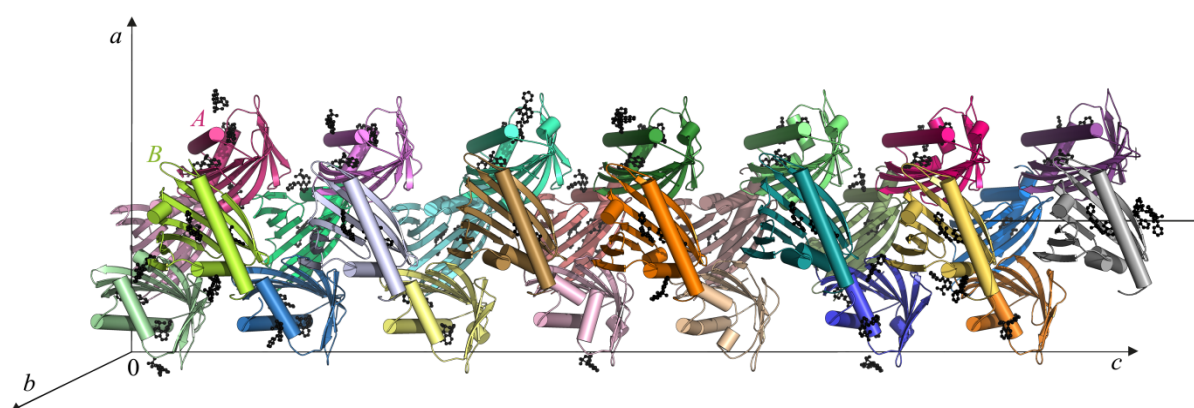


Figure 6. The 28 independent Hyp-1 molecules in the asymmetric unit of the *C2* crystal packing. The protein molecules are arranged in a dimeric pattern with a sevenfold repeat around a noncrystallographic 2_1 screw (indicated) along the crystallographic *c* direction. Dimer AB is labeled (From Sliwak *et al.* 2015; Publication III).

The ANS molecule can be bound both inside the three internal Hyp-1 binding sites as well as at interstitial sites, where they join two or three Hyp-1 molecules together. However, there is a complicated pattern of saturation of the internal and external binding sites by the ANS molecules along the Hyp-1 column, as the Hyp-1 copies can carry 0, 1, 2 or 3 ANS molecules inside the internal cavities without an obvious pattern along the sevenfold column. Another interesting observation is that there is only one column of seven Hyp-1 protein molecules that are fully saturated by the ANS ligand. Moreover, the 29 interstitial ANS molecules are arranged in a way that violates the arrangement of the protein molecules. Whenever a ligand molecule links adjacent, tNCS-related Hyp-1 molecules, those protein molecules are closer to each other. This unusual pattern of molecular packing in the *c* direction together with the unusual pattern of the reflection intensities provide convincing evidence that this structure represents a case of a modulated superstructure. However, since it was possible to refine it using an expanded unit cell, the modulation can be treated as commensurate. The case of the Hyp-1/ANS complex is the

first example of a successful structure determination and refinement of a modulated macromolecular crystal structure.

This study also demonstrates that novel maximum-likelihood algorithms with tNCS corrections accounting for the structure-factor modulations are powerful tools for handling very difficult cases in protein crystallography, which previously have been considered too difficult to solve. The successful refinement of the Hyp-1/ANS complex structure and its in-depth description provide a clear validation of the success of the sophisticated structure solution protocol. Moreover, the structure of the Hyp-1 protein in complex with a fluorescent dye provides an important basis for the interpretation of the ADA displacement assays. It also aids in the identification of potential physiologically relevant novel binding sites, since the aromatic rings of ANS resemble some biologically important ligands.

6.3. Hyp-1 protein from St John's wort as a PR-10 protein with novel type of tripartite cavity (Publications III, IV and V)

The Hyp-1 protein from St John's wort was initially implicated in hypericin biosynthesis from emodin (Bais, *et al.*, 2003), but this proposition was later questioned by Michalska *et al.* (2010). More recently, Hyp-1 has been classified as a PR-10 protein based on its gene structure (Kosuth *et al.*, 2013). In the most recent studies (Karppinen *et al.*, 2016) three genes homologous to *hyp-1* have been identified in *Hypericum*, and all of them together with the *hyp-1* gene, are characterized by constitutive albeit variable expression in roots, stem and leaves. Furthermore, these genes are upregulated by salicylic acid, abscisic acid and wounding, indicating their contribution to the plant's defense mechanisms.

The first crystallographic studies of Hyp-1 by Michalska *et al.* (2010) revealed that it shares the canonical fold with other PR-10 proteins. The binding cavity of the Hyp-1 in that structure (PDB code: 3ie5), which was supposed to illustrate the situation in ligand-free form of the protein, contains in fact serendipitously bound polyethylene glycol (PEG) molecules from the crystallization buffer.

Detailed analysis and comparisons of the cavity of the 3ie5 model with those of the two new structures of Hyp-1 presented in this thesis, in complex with ANS (Chapter 6.2.) and MEL (Chapter 6.4.), provided a unique opportunity for in-depth overview of the cavity shapes of all structures of PR-10 complexes deposited in the Protein Data Bank. These careful analyses revealed that the PR-10 cavities, capable of accommodating different ligands, have different volumes as well as create different types of interactions with the ligands, and that they can be divided into three distinct types:

Type I – the cavity is shallow and accessible only via the E1 entrance. It is characteristic of the MLP/RRP and CSBP proteins and ABA receptors, but it is also found in the IPR-type nodulin MtN13. This type of cavity is usually capable of binding only a single copy of the ligand molecule but in a highly specific manner, which strongly suggests a receptor role.

Type II – the cavity is large, spanning the space between entrances E1 and E2. It is capable of binding more than two ligand molecules, mainly by hydrophobic interactions, which indicates transport/storage function. This type of cavity can be found in all NCS and IPR proteins with the exception of nodulin MtN13.

Type III – represented only by the structures of the Hyp-1 complexes studied in this thesis, which demonstrate that Hyp-1 cannot be classified in any of the above categories, as its binding cavity consists in fact of three separate binding sites, two of which are internal chambers and one is a deep surface pocket.

Examples of the cavity types in different PR-10/hormone complexes are listed in Table 3 and the cavity shapes are illustrated in Figure 7.

Cavity Type	Name of protein	Organism	Ligands	No. of ligand molecules	PDB code
	PYL1	<i>Arabidopsis thaliana</i>	ABA	1	3kdi
	MtN13	<i>Medicago truncatula</i>	ZEA KIN 2iP 6BAP	1	4jhg 4jhh 4gy9 4jhi
I	<u>MtCSBP</u>	<u><i>Medicago truncatula</i></u>	<u>GA3</u>	<u>1</u>	<u>4q0k</u>
	<u>VrCSBP</u>	<u><i>Vigna radiata</i></u>	<u>GA3</u>	<u>1</u>	<u>4psb</u>
	Act d 11*	<i>Actinidia deliciosa</i>	UNL*	1	4igv
	<u>LLPR-10.1A</u>	<u><i>Lupinus luteus</i></u>	<u>ZEA</u>	<u>3</u>	<u>4ryv</u> <u>5c9y</u>
II	<u>LLPR-10.2B</u>	<u><i>Lupinus luteus</i></u>	<u>ZEA</u>	<u>3</u>	<u>2qim</u>
	Bet v 1	<i>Betula verrucosa</i>	DXC	2	4a83
III	<u>Hyp-1</u>	<u><i>Hypericum perforatum</i></u>	<u>MEL/UNL</u>	<u>2+1</u>	<u>5i8f</u>

Table 3. Examples of PDB structures of PR-10/phytohormone complexes arranged according to binding cavity type. The complexes discussed as part of this work are underlined. *In the case of the Act d 11 structure, the ligand bound in the cavity is unknown; however it has been included as the only example of an MLP homolog in complex with a ligand.

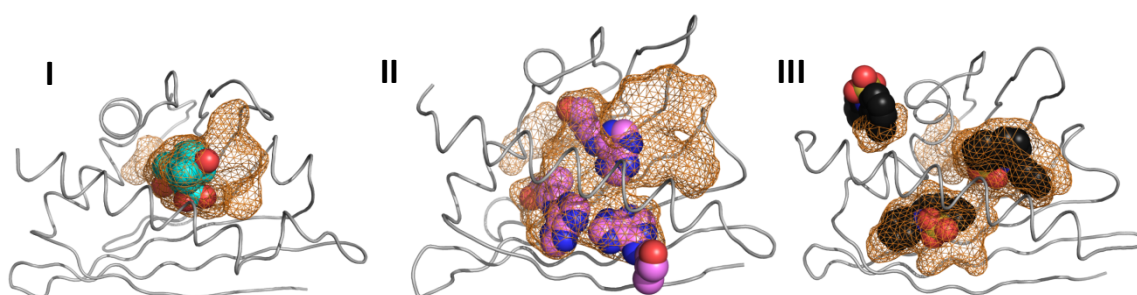


Figure 7. Three types of PR-10 cavities (shown in mesh surface representation) of PR-10 proteins, illustrated by selected PDB structures of PR-10 complexes. (I) MtPhBP in complex with GA3, 4q0k; (II) LLPR-10.2B in complex with ZEA, 2qim; (III) Hyp-1 in complex with ANS, 4n3e, chain K.

Interestingly, Hyp-1 not only reveals a new mode of ligand accommodation within the internal cavity, but also features an unusually deep surface invagination that serves as a new ligand binding pocket. Moreover, Hyp-1 also binds ligands of very different chemical character with amazing positional conservation (Fig. 8). The main hydrophobic cavity of this protein is divided into two chambers separated mainly by the bulky Arg27 side chain, which drags the $\alpha 2$

helix towards the cavity interior. This feature distinguishes Hyp-1 from all other PR-10 proteins, which typically have a residue with a short side chain, such as Gly or Ala, at the position corresponding to Arg27 in Hyp-1. Interestingly, the MLP members also have Arg or Lys at this position but the side chain is directed outside the binding cavity.

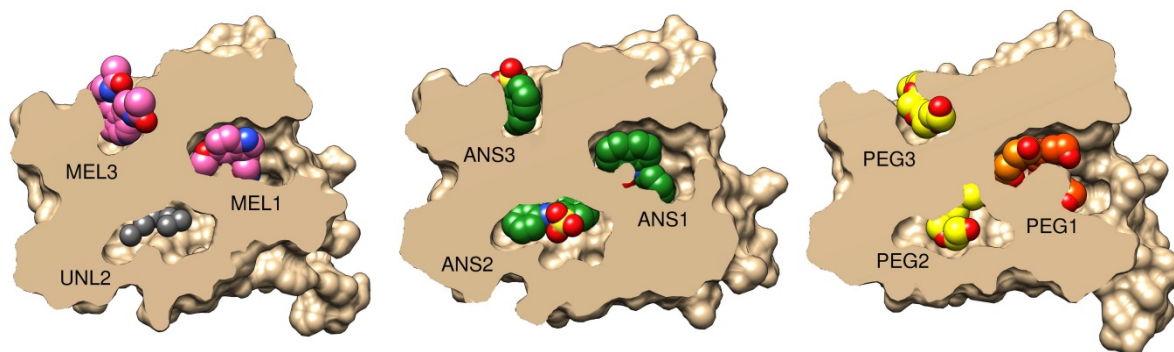


Figure 8. Structures of Hyp-1 complexes with MEL (this work), ANS (this work, chain K) and PEG (3ie5, chain B) in cutaway, surface representation. The ligand molecules are shown as van der Waals models. “PEG” denotes various fragments (oligomers) of polyethylene glycol, a buffer component that was serendipitously bound by the Hyp-1 protein in the experiments conducted by Michalska *et al.* (2010). (From Sliwiak *et al.*, 2016; Publication V)

The Hyp-1 surface invagination leading to the creation of a new ligand binding site is also a novel feature among all PR-10 complexes. The main force that holds the ligand molecule at this site comes from stacking interactions with the residues Lys33 and Tyr150, which act as the jaws of a vice and which are conserved in the PR-10 family. The Hyp-1 structures presented in this thesis are, however, the first to highlight the relevance for ligand binding of these residues. Despite the fact that with other PR-10 proteins studied in our laboratory, similar high ligand:protein ratios were often used in co-crystallization trails, those experiments have never resulted in structures with ligand molecules bound at this site.

The unstructured C-end of Hyp-1 interacts via numerous contacts with helix α_1 , which is unusual for PR-10 members. Such interactions could contribute to the formation the third binding site. Moreover, the ligand molecule bound at this site can act as a lever opening the E1 entrance and facilitating the docking of another ligand molecule at site 1. The high resolution structure 3ie5 of the Hyp-1/PEG complex has two protein chains in the asymmetric unit. In one of them the third site is occupied (by a PEG molecule) and the E1 entrance is open. In the other Hyp-1 molecule the third binding site is empty and the E1 entrance is shut.

I have tried to study the Hyp-1 binding capacity in solution by ITC calorimetry. However, titration of the protein with the studied natural ligands, even if their solubility was adequate, did not produce a measurable heat effect. Therefore, my next goal was to carry out ANS displacement assays. Accordingly, I first crystallized the Hyp-1/ANS complex as its structure would be the basis for the interpretation of the ADA results. However, as explained in Chapter 6.2., the structure of the Hyp-1/ANS complex contains numerous unexpected additional superficial ANS binding sites (Fig. 9), which are absent in the MEL complex despite a high excess of the ligand in the crystallization conditions. These superficial ANS binding sites suggest that the superstructure of the Hyp-1/ANS complex is “glued” together by these additional ANS molecules. It can be concluded that the hydrophobic interaction between the ANS ligand and protein surface patches are strong enough to determine the unusual crystal packing. This structural information can also provide an explanation of the ADA results (Chapter 5.2.). Despite

the high excess of the MEL ligand used for Hyp-1/MEL co-crystallization, the MEL molecules do not occupy any of the superficial binding sites occupied in the crystal of the complex. This could explain why MEL could not compete with ANS in ADA assays.

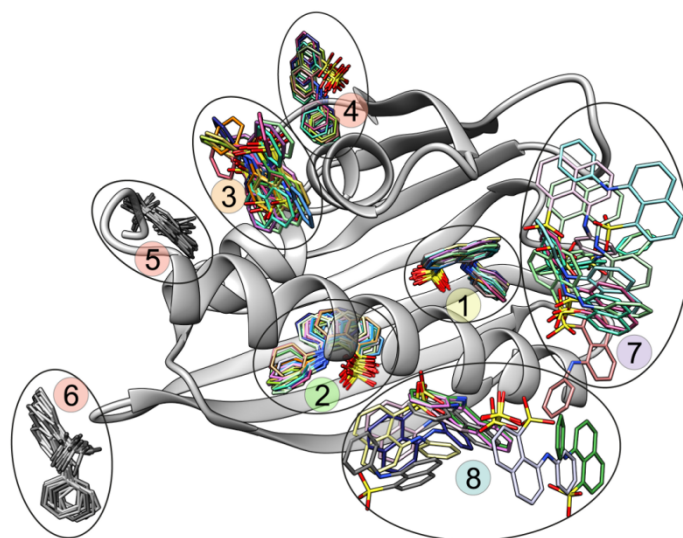


Figure 9. All of the ANS molecules found in the structure of the Hyp-1/ANS complex, superposed using a common frame of C α atoms for the 28 copies of the protein molecule. The ANS molecules are color-coded as their nearest protein molecules in Fig. 6 (From Sliwiak *et al.*, 2015; Publication III).

The crystal structures of the Hyp-1/ANS and Hyp-1/MEL complexes determined within this thesis reveal new binding sites and a novel binding cavity type among all PR-10 members. Moreover, the structure of the Hyp-1/ANS complex shows that the ANS probe can be very strongly bound not only in internal binding sites but also at external binding sites at the protein surface and that this could bias the results of ADA experiments.

6.4. Melatonin, a new ligand for PR-10 proteins (Publication V)

Although it is still under debate whether MEL (Fig. 3) should be included in the list of plant hormones (Hardeland, 2016), recent lines of evidences clearly suggest that it acts in a hormone-like manner in plant organisms. It regulates plant growth in an auxin-like manner, however, by different transduction pathways; it promotes growth of etiolated hypocotyls and induces rhizogenesis; it was also shown to promote growth of lateral roots by inducing a group of genes other than AUX (Murch *et al.*, 2001; Arnao & Hernandez-Ruiz, 2007; Pelagio-Flores *et al.*, 2012). Importantly, MEL acts not only by upregulating particular genes but also directly, being one of the most potent antioxidants. It has the ability to quench up to 10 ROS/RNS radical species thanks to a unique reaction cascade. The products of MEL oxidation, including N1-acetyl-N2-formyl-5-methoxykynuramine (AFMK), N1-acetyl-5-methoxykynuramine (AMK) or cyclic 3-hydroxymelatonin (3OH-MEL), have the same or even stronger antioxidant capacity (Guenther *et al.*, 2005; Thann *et al.*, 2006; Tan *et al.*, 2007; Manchester *et al.*, 2015). The content of MEL in plants varies significantly from nanomolar to micromolar range. The highest level of MEL (200 $\mu\text{g/g}$) was found in the kernels of *Pistacia vera* (Oladi *et al.*, 2014). *Hypericum perforatum* is another species of relatively high MEL levels (4 $\mu\text{g/g}$), which can be linked to the medicinal properties of *Hypericum* preparations (Murch *et al.*, 1997). The concentration of MEL in plants is elevated in response to heat, cold, heavy metal pollution, UV radiation, pathogen attack and other stress conditions, and the stress effects can be alleviated by MEL. MEL plays an important

role also during senescence of leaves and in other vulnerable tissues like seeds, as well as at some special developmental stages, e.g. in flower buds or ripening fruits (Hardeland, 2016). High levels of MEL in some tissues may contradict its potential role as a phytohormone. If at low levels MEL acts as a regulator (in other organs/species/conditions), then functional high-affinity sites for its binding would be required. In situations when MEL concentration is elevated, the high affinity binding sites would be saturated or its receptors internalized, as is observed with other receptors, and signaling through these sites would no longer exist. Hardeland (2016) concludes that an escape of this dilemma would be the assumption of the existence of low-affinity binding sites, which take over the MEL signaling function.

Increased levels of MEL are temporally and spatially correlated with PR-10 expression patterns, e.g. during senescence, upon wounding, or in stress. Moreover, MEL upregulates salicylic acid, abscisic acid and ethylene signaling-related genes, and PR-10 expression is induced by these phytohormones. In view of the above observations, PR-10 members appear to be very good candidates as low-affinity binding sites for MEL in conditions of stress when the levels of MEL are high. Therefore, based on numerous literature reports about the role of MEL in plants, I decided to include this molecule among the ligands used in my crystallization screens.

My crystallization studies indicate that from among numerous ligands tested, the Hyp-1 protein forms crystalline complex only with MEL and ANS. It is important to stress that the indole phytohormone auxin, and the MEL precursor serotonin, did not form crystalline complexes with either the Hyp-1 protein or the LIPR-10.2B isoform, despite numerous crystallization trials. Although both these molecules contain an indole ring, overall they have more hydrophilic substituents than MEL, which could compromise their binding affinity. My observation of PR-10 binding of MEL but not AUX, could be the first molecular evidence of an AUX-independent MEL binder. Moreover, equimolar concentration of ZEA, a common PR-10 binding partner, in co-crystallization experiments did not perturb Hyp-1/MEL crystal formation. Owing to its very high resolution (1.30 Å), the crystal structure of the Hyp-1/MEL complex provides the best available model of the Hyp-1 protein itself.

MEL presence was detected in all three internal binding sites of Hyp-1, in the same position as in the ANS and PEG complexes. The MEL1 ligand has the best defined electron density (Fig. 10).

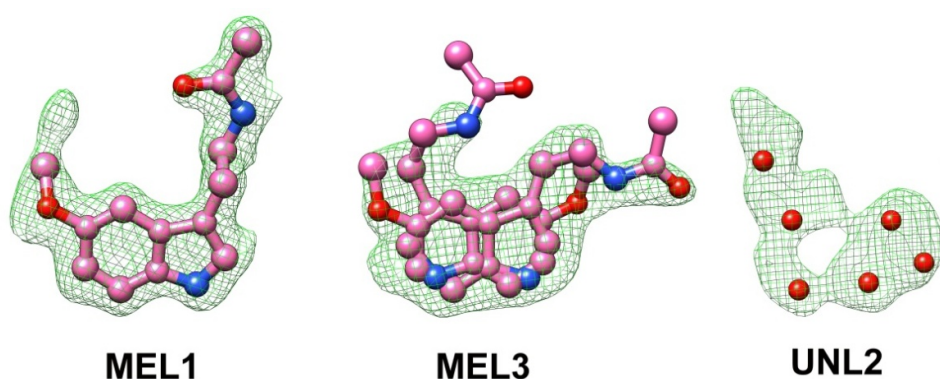


Fig. 10. F_o-F_c OMIT maps contoured at 2.5σ corresponding to MEL bound at Hyp-1 sites 1 and 3, as well as to unknown ligand UNL identified at site 2 (From Sliwiak *et al.*, 2016; Publication V).

It interacts with the protein via numerous hydrophobic interactions and by water-mediated hydrogen bonds. MEL1 has an unusual positive electron density near the methoxyl group. At the 2nd binding site, a density resembling an indole ring was detected. However, the whole MEL molecule could not be modeled there and dummy water molecules were modeled instead. Changed shape of the ligand electron density at the 2nd binding site could indicate a high mobility MEL molecule or some product of its degradation. The ligand at the 2nd binding site creates hydrophobic contacts with the same residues as the corresponding ANS molecule in the Hyp-1/ANS complex. The MEL3 molecule seems to rotate within the vise of the Lys33 and Tyr150 side chains, similarly to ANS3, and it could be modeled in two alternative orientations.

The Hyp-1/MEL structure is highly reproducible, even when it comes to the peculiarities of ligand electron density. The same crystal structure is obtained regardless of the melatonin form used for co-crystallization (methanol stock or powder), or of the presence or absence of *trans*-zeatin in the co-crystallization experiment. Thus, the results of all the crystal screens seem to indicate a physiological role for the Hyp-1-MEL interaction. Moreover, the subsequently determined crystal structure of the yellow lupine LIPR-10.2B isoform in complex with MEL together with the results of competitive crystallization of this isoform with equimolar solution of MEL and ZEA (J. Sliwiak *et al.*, unpublished results), reinforce the hypothesis that PR-10 proteins could be low-affinity MEL binders of biological relevance.

6.5. Calorimetric characterization and determination of binding specificity of Phytohormone Binding Proteins (Publication VI)

CSBP, or cytokinin specific binding proteins were included in the PR-10 family based on a relatively low level (~20%) of amino acid sequence identity. Homologs of these proteins are found only in legume plants and their levels of expression are very low. They were identified by Fujimoto *et al.* (1998) as strong cytokinins binders, although subsequent studies downshifted the binding affinity by five orders of magnitude (Pasternak *et al.*, 2006). Moreover, the crystal structure of the VrCSBP protein in complex with ZEA (2flh) revealed a rather nonspecific binding of *trans*-zeatin, as in four protein molecules in the asymmetric unit there are three distinct ZEA binds modes and two different stoichiometries (1 or 2). Although the studies by Pasternak *et al.* (2006) clearly indicated the inadequate terminology, the term “CSBP” has been well established.

The structural studies conducted by the coauthors of Publication VI resulted in the solution of high resolution structures of CSBP proteins from *Medicago truncatula* (MtCSBP) and *Vigna radiata* (VrCSBP) in complex with gibberellic acid, GA3 (4q0k and 4psb). The structures revealed the presence of one GA3 molecule in the binding cavity of both proteins, with numerous direct and water-mediated hydrogen bonds to the protein. VrCSBP and MtCSBP have different electrostatic surface potential around the E1 entrance of the ligand binding cavity. In MtCSBP, the area surrounding the entrance is only slightly charged, whereas in VrCSBP many charged residues surround the cavity entrance. This could be explain why the two proteins form crystalline complex with GA3 under different pH conditions, MtCSBP at pH 4.0 and VrCSBP at pH 6.5.

My main contribution to the studies presented in Publication VI was to verify the binding specificity of both proteins in solution by the use of ITC microcalorimetry. The calorimetric

titration with GA3 and ZEA were performed for both proteins under two different pH conditions, 5.5 and 7.4 (Table 4).

Protein	VrPhBP		MtPhBP		
	pH	7.4	5.5	7.4	5.5
GA3	<i>N</i>	1	1		1
	K_d	23 ± 1	6.1 ± 0.2		13.4 ± 0.5
	ΔH	-8054 ± 120	-9039 ± 63	¹	-3727 ± 35
	ΔS	-6.3	-7.0		9.6
ZEA	K_{d1}	76 ± 1	181 ± 25		
	ΔH_1	-5186 ± 59	-6645 ± 493		
	ΔS_1	1.16	-5.6		
	K_{d2}	67 ± 1	94 ± 9	²	²
	ΔH_2	1903 ± 71	4348 ± 573		
	ΔS_2	25.6	33.2		

Table 4. Thermodynamic parameters of the interactions of MtCSBP and VrCSBP with GA3 and ZEA. *N* is the stoichiometry; K_d (μM) is the dissociation constant; ΔH (cal mol^{-1}) is the enthalpy change; ΔS ($\text{cal mol}^{-1} \text{K}^{-1}$) is the change of entropy. ¹No heat effect. ²Very small enthalpy change ($\Delta H < 800 \text{ cal/mol}$) and high noise/signal ratio precluded reliable estimation of the derived parameters (From Ruszkowski *et al.*, 2014; Publication VI).

The titrations revealed high affinity and specificity of GA3 binding by MtCSBP and VrCSBP, which appears to be pH-dependent. The heat effect of the interaction of MtCSBP with GA3 was observed only at acidic pH, which is consistent with the structural data and the presence of uncharged surroundings of the cavity entrance. Titration of MtCSBP with GA3 at acidic pH resulted in a relatively high affinity ($K_d = 13 \mu\text{M}$) and in stoichiometry of $N=1$, consistent with the stoichiometry observed in the crystal. Titration of MtCSBP with ZEA under both pH conditions did not result in any detectable heat change. This is consistent with crystal screening results, as it was impossible to obtain crystalline complex of MtCSBP with ZEA.

VrCSBP binds GA3 with high affinity at both pH 7.4 and 5.5. However, binding at lower pH is nearly four times stronger (K_d $23 \mu\text{M}$ and $6 \mu\text{M}$, respectively). Moreover, it appears that GA3 binding by VrCSBP is enthalpy-driven, most likely as a result of ion pairing between the ligand and the charged cavity surroundings. The entropy contribution to MtCSBP/ZEA binding is higher due to hydrophobic interactions with the uncharged surroundings. VrCSBP titration with ZEA produces a considerable heat effect. However, after integration of the heat peaks, it appears that the binding curve is hyperbolic, indicating nonspecific binding, and the affinity constant is rather low. The stronger and more specific GA3 binding by VrPhBP was additionally confirmed by competitive titration of VrPhBP with GA3 in the presence of ZEA. Table 4 reports the thermodynamic parameters obtained in the ITC titration experiments.

The results of the thermodynamic characterization of the “cytokinin specific binding proteins” together with the high resolution structures of their complexes with GA3, suggested a revision of the biological function of these proteins and a more adequate nomenclature. Consequently, in Paper VI the name Phytohormone Binding Proteins, or PHBPs, has been proposed for this class of proteins. Although the studies reported herein were carried out *in vitro* and further experiments are needed to verify the PhBPs’ role as GA3 receptors, the very low levels of their expression together with high affinity for GA3 provide a strong argument in support of such a hypothesis.

7. Summary

The results presented in this thesis provide novel insights and expand our knowledge about the PR-10 protein family. Among the most significant achievements are the extension of the list of PR-10 ligands by melatonin and the demonstration that melatonin is bound by two PR-10 members more strongly than *trans*-zeatin, which has been regarded so far as the best established PR-10 binding partner. The presented studies also help to elucidate the binding mechanism of natural ligands by PR-10 proteins, highlighting the role of the C-terminal helix $\alpha 3$ in ligand recognition. Moreover, new methodological approaches have been developed for studying the PR-10 proteins, such as competitive co-crystallization assays or co-crystallization of the protein with variable molar excess of the ligand. Importantly, the studies conducted within this thesis have clearly demonstrated that, in solution, binding of gibberellic acid by so-called “Cytokinin Specific Binding Proteins” is stronger and more specific than binding of a cytokinin. As an additional achievement, the Hyp-1/ANS complex was crystallized with a modulated superstructure. This structure has been successfully solved and refined with 28 protein molecules in the asymmetric unit, despite the additional complication of a severe crystal twinning, providing the first example of a successful elucidation of a macromolecular modulated structure. The crystal and X-ray diffraction data that I obtained for the Hyp-1/ANS complex, and its structure itself, became a testing ground used by the authors of *Phaser* for the development of new statistical methods for macromolecular structure solution in the presence of translational pseudosymmetry. Last but not least, this work introduces a completely novel classification of PR-10 proteins based on the topology and binding properties of their internal cavities.

8. Acknowledgments

I would like to thank the European Union's European Regional Developmental Fund and the Foundation for Polish Science which supported my MPD program. I also acknowledge support from the Polish Ministry of Science and Higher Education (grant No. NN 301 003739 and the KNOW program) as well as from the National Science Center (2013/10/M/NZ1/00251). The research leading to these results has also received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under BioStruct-X (grant agreement N 283570).

I thank Helmholtz-Zentrum Berlin, Deutsches Elektronen-Synchrotron (DESY, Hamburg), Max-Lab (Lund), and Argonne Photon Source for the allocation of synchrotron radiation beamtime.

I would like to thank Prof. Mariusz Jaskolski for supervising my work and motivation.

I owe thanks to Dr. hab. Zbigniew Dauter for the opportunity of a fruitful internship at the Argonne National Laboratory (USA) and for scientific supervision of my work there. I thank Prof. Michał Sikorski for guidance with protein engineering, Ms. Mirosława Dauter for crystallization tips, and Ms. Alina Kasperska for laboratory support.

My thanks also go to Dr. Katarzyna Glanowska and Ms. Ewa Tomaszewska for proof-reading of the thesis, Dr. Piotr Małecki and Dr. Miłosz Ruszkowski, for brainstorming and support at synchrotrons; Dr. Marta Grzechowiak, for tea, sweats and some polyacrylamide gels; Dr. Joanna Banasiak, for inspiring talks about plants; Mr. Tomasz Manszewski for software advices; Dr. Joanna Raczyńska for help with LINUX and script writing; Dr. Karolina Michalska for introducing me to PR-10 proteins; Dr. Kamil Szpotkowski for circular dichroism measurements, Dr. Magdalena Małgowska for NMR measurements, Dr. Anna Urbanowicz for administrative help with the MPD PhD program, and to all of my colleagues at the CBB and IBCh for help and good atmosphere.

9. References

- Agarwal, P. & Agarwal, P. K. (2014). Pathogenesis related-10 proteins are small, structurally similar but with diverse role in stress signaling. *Mol. Biol. Rep.* 41, 599-611.
- Ahuja, I., Kissen, R. & Bones, A. M. (2012). Phytoalexins in defense against pathogens. *Trends Plants Sci.* 17, 73-90.
- Andrade, L. B., Oliveira, A. S., Ribeiro, J. K., Kiyota, S., Vasconcelos, I. M., de Oliveira, J. T. & de Sales, M. P. (2010) Effects of a novel pathogenesis-related class 10 (PR-10) protein from *Crotalaria pallida* roots with papain inhibitory activity against root-knot nematode *Meloidogyne incognita*. *J. Agric. Food Chem.* 58, 4145-4152.
- Arnao, M. N. & Hernandez-Ruiz, J. (2007). Melatonin promotes adventitious and lateral root regeneration in etiolated hypocotyls of *Lupinus albus* L. *J. Pineal Res.* 42, 147-152.
- 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, 32414-32422.
- Barratt, D. H. P. & Clark, J. A. (1991). Proteins arising during the late stages of embryogenesis in *Pisum sativum* L. *Planta* 184, 14-23.
- Biesiadka, J., Bujacz, G., Sikorski, M. & Jaskolski, M. (2002). Crystal structures of two homologous pathogenesis-related proteins from yellow lupine. *J. Mol. Biol.* 319, 1223-1234.
- Bindschelder, L. V., Dewdney, J., Blee, K. A., Stone, J. M., Asai, T., Plotnikov, J., Denoux, C., Hayes, T., Gerrish, C., Davies, D. R., Ausubel, F. M. & Bolwell, G. P. (2006). Peroxidase-dependent apoplastic oxidative burst in *Arabidopsis* required for pathogen resistance. *Plant J.* 47, 851-863.
- Breiteneder, H. & Ebner, C. (2000). Molecular and biochemical classification of plant-derived food allergens. *J. Allergy Clin. Immunol.* 106, 27-36.
- Chruszcz, M., Ciardiello, M. A., Osinski, T., Majorek, K. A., Giangrieco, I., Font, J., Breiteneder, H., Thalassinou, K. & Minor, W. (2013). Structural and bioinformatic analysis of the kiwifruit allergen Act d 11, a member of the family of ripening-related proteins. *Mol. Immunol.* 56, 794-803.
- Clouse, S. D. & Sasse, J.M. (1998). Brassinosteroids: Essential regulators of plant growth and development". *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 49, 427-51.
- 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., Bujacz, A., Bujacz, G., Jelen, F., Jasinski, M., Kachlicki, P., Otlewski, J., Sikorski, M. & Jaskolski, M. (2009). Cytokinin-induced structural adaptability of a *Lupinus luteus* PR-10 protein. *FEBS J.* 276, 1596-1609.
- Fernandes, H., Michalska, K., Sikorski, M., & Jaskolski, M. (2013). Structural and functional aspects of PR-10 proteins. *FEBS J.* 280, 1169-1199.

- Franco, O. L. (2011). Peptide promiscuity: an evolutionary concept for plant defense. *FEBS Lett.* 6, 995-1000.
- 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.
- Gane, P. J., Dunwell, J. M. & Warwicker, J. (1998). Modeling based on the structure of vicilins predicts a histidine cluster in the active site of oxalate oxidase. *J. Mol. Evol.* 46, 488-493.
- Gasymov, O. K. & Glasgow, B. J. (2007). ANS fluorescence: potential to augment the identification of the external binding sites of proteins. *Biochim. Biophys. Acta* 1774, 403-411.
- Gilchrist, D.G. (1998). Programmed cell death in plant disease: the purpose and promise of cellular suicide. *Annu. Rev. Phytopathol.* 36, 393-414.
- Goulas, E., Richard-Molard, C., Le Dily, F., Le Dantec, C., Ozouf, J. & Ourry, A. (2007) A cytosolic vegetative storage protein (TrVSP) from white clover is encoded by a cold-inducible gene. *Physiol. Plant* 129, 567–577.
- Guenther, A. L., Schmidt, S. I., Laatsch, H., Fotso, S., Ness, H., Ressmeyer, A. R., Poeggler, B. & Hardeland R. (2005). Reactions of the melatonin metabolite AMK (N1-acetyl-5-methoxykynuramine) with reactive nitrogen species: Formation of novel compounds, 3-acetamidomethyl-6-methoxycinnolinone and 3-nitro-AMK. *J. Pineal Res.* 39, 251-260.
- Handschuh, L., Femiak, I., Kasperska, A., Figlerowicz, M. & Sikorski, M. (2007). Structural and functional characteristics of two novel members of pathogenesis-related multigene family of class 10 from yellow lupine. *Acta Biochim. Pol.* 54, 783-796.
- Hardeland, R. (2016). Melatonin: Another Phytohormone? *Res. Rev. J. Bot. Sci.* 5:2.
- Hardeland, R. (2016). Melatonin in plants – diversity of levels and multiplicity of functions. *Front. Plant. Sci.* 7: 198.
- Hewitt, F. R., Hough, T., O'Neill, P., Sasse, J. M., Williams, E. G. & Rowan, K. S. (1985). Effect of brassinolide and other growth regulators on the germination and growth of pollen tubes of *Prunus avium* using a multiple hanging drop assay. *Aust. J. Plant. Physiol.* 12, 201–211.
- Hoffmann-Sommergruber, K., Vanek-Krebitz, M., Radauer, C., Wen, J., Ferreira, F., Scheiner, O. & Breiteneder, H. (1997). Genomic characterization of members of the Bet v 1 family: genes coding for allergens and pathogenesis-related proteins share intron positions. *Gene.* 197, 91–100.
- Holm, L. & Sander, C. (1995). Dali: a network tool for protein structure comparison. *Trends Biochem. Sci.* 20, 478-480.
- Hwang, H. J., Kim, H., Yu, H. J., Oh, M. N., Lee, I. & Kim, S. G. (2003). Gene encoding pathogenesis-related 10 protein of *Lithospermum erythrorhizon* is responsive to exogenous stimuli related to the plant defense system. *Plant Sci.* 165, 1297–1302.
- Hwang, I., Sheen, J. & Müller B. (2012). Cytokinin signaling networks. *Annu. Rev. Plant Biol.* 63, 353-380.

- Iyer, L. M., Koonin, E. V. & Aravind, L. (2001). Adaptations of the helix-grip fold for ligand binding and catalysis in the START domain superfamily. *Proteins* 43, 134-144.
- Jain, S., Kumar, D., Jain, M., Chaudhary, P., Deswal, R. & Sarin, N. B. (2012). Ectopic overexpression of a salt stress-induced pathogenesis-related class 10 protein (PR10) gene from peanut (*Arachis hypogaea* L.) affords broad spectrum abiotic stress tolerance in transgenic tobacco. *Plant Cell Tissue Organ Cult.* 109, 19–31.
- Jain, D., Khandal, H., Khurana, J. P. & Chattopadhyay, D. (2016). A pathogenesis related-10 protein CaARP functions as aldo/keto reductase to scavenge cytotoxic aldehydes. *Plant Mol. Biol.* 90, 171-187.
- Karppinen, K., Derzsó, E., Jaakola, L. & Hohtola, A. (2016). Molecular cloning and expression analysis of *hyp-1* type PR-10 family genes in *Hypericum perforatum*. *Front Plant Sci.* 7:526.
- 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.
- Kofler, S., Asam, C., Eckhard, U., Wallner, M., Ferreira, F. & Brandstetter, H. (2012). Crystallographically mapped ligand binding differs in high and low IgE binding isoforms of birch pollen allergen bet v 1. *J. Mol. Biol.* 422, 109-123.
- Krishnaswamy, S., Baral, P. K., James, M. N. & Kav, N. N. (2011). Site-directed mutagenesis of histidine 69 and glutamic acid 148 alters the ribonuclease activity of pea ABR17 (PR10.4). *Plant Physiol. Biochem.* 49, 958–962.
- Krissinel, E. & Henrick, K. (2004). Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr.* D60, 2256-2268.
- Lebel, S., Schellenbaum, P., Walter, B. & Maillot, P. (2010). Characterisation of the *Vitis vinifera* PR10 multigene family. *BMC Plant Biol.* 10:184.
- Liu, J. J. & Ekramoddoullah, A. K. M. (2006). The family 10 of plant pathogenesis-related proteins: Their structure, regulation, and function in response to biotic and abiotic stresses. *Physiol. Mol. Plant Pathol.* 68, 3-13.
- 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 anti-microbial activity is stress- and pathogen-inducible. *J. Plant Physiol.* 163, 546–556.
- Lovelace, J. J., Simone, P. D., Petricek, V. & Borgstahl, G. E. O. (2008). Protein crystals can be incommensurately modulated. *J. Appl. Cryst.* 41, 600-605.
- Manchester, L. C., Coto-Montes, A., Boga, J. A., Andersen, L. P. H., Zhou, Z., Galano, A., Vriend, J., Tand, D. X. & Reiter, R. J. (2015). Melatonin: an ancient molecule that makes oxygen metabolically tolerable. *J. Pineal Res.* 59, 403-419.
- 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.

- Michalska, K., Fernandes, H., Sikorski, M. & Jaskolski, M. (2010). Crystal structure of Hyp-1, a St. John's wort protein implicated in the biosynthesis of hypericin. *J. Struct. Biol.* 169, 161-171.
- Miyazono, K., Miyakawa, T., Sawano, Y., Kubota, K., Kang, H., Asano, A., Miyauchi, Y., Takahashi, M., Zhi, Y., Fujita, Y., Yoshida, T., Kodaira, K., Yamaguchi-Shinozaki, K. & Tanokura, M. (2009). Structural basis of abscisic acid signalling. *Nature* 462, 609-615.
- Moiseyev, G. P., Beintema, J. J., Fedoreyeva, L. I., & Yakovlev, G. I. (1994). High sequence similarity between a ribonuclease from ginseng callus and fungus-elicited proteins from parsley indicates that intracellular pathogenesis-related proteins are ribonucleases. *Planta* 193, 470-472.
- 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.
- Murch, S. J., Simmons, C. B. & Saxena, P. K. (1997). Melatonin in feverfew and other medicinal plants. *Lancet* 359, 1598-1599.
- Murch, S. J., Campbell, S. S. B. & Saxena, P. K. (2001). The role of serotonin and melatonin in plant morphogenesis: Regulation of auxin-induced root organogenesis in in vitro-cultured explants of St. John's Wort (*Hypericum perforatum* L.). *In Vitro Cell. Dev. Biol.- Plant* 37, 786-793.
- Oladi, E., Mohamadi, M., Shamspur, T. & Mostavi, A. (2014). Spectrofluorimetric determination of melatonin in kernels of four different Pistacia varieties after ultrasound-assisted solid-liquid extraction. *Spectrochim. Acta A. Mol. Biomol. Spectrosc.* 132, 326-329.
- Orengo, C. A., Flores, T. P., Taylor, W. R. & Thornton J. M. (1993). Identification and classification of protein fold families. *Protein Eng.* 6, 485-500.
- 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.
- Park, C. J., Kim, K. J., Shin, R., Park, J., M., Shin, Y. C. & Paek K., H. (2004). Pathogenesis-related protein 10 isolated from hot pepper functions as a ribonuclease in an antiviral pathway. *Plant J.* 37, 186-198.
- 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.
- Pelagio-Flores, R., Munoz-Parra, E., Ortiz-Castro, R. & Lopez-Bucio, J. (2012). Melatonin regulates *Arabidopsis* root system architecture likely acting independently of auxin signaling. *J. Pineal. Res.* 53, 279-288.
- Radauer, C., Lackner, P. & Breiteneder, H. (2008). The Bet v 1 fold: an ancient, versatile scaffold for binding of large, hydrophobic ligands. *BMC Evol. Biol.* 8: 286.
- Ricci, A. & Bertolotti, C. (2008). Urea derivatives on the move: cytokinin-like activity and adventitious rooting enhancement depend on chemical structure. *Plant Biol.* 11, 262-272.

- Samanani, N., Liscombe, D. K. & Facchini, P. J. (2004). Molecular cloning and characterization of norcochlorogenic acid 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.
- Schiebel, J., Radeva, N., Krimmer, S. G., Wang, X., Stieler, M., Ehrmann, F. R., Fu, K., Metz, A., Huschmann, F. U., Weiss, M. S., Mueller, U., Heine, A. & Klebe, G. (2016). Six biophysical screening methods miss a large proportion of crystallographically discovered fragment hits: a case study. *ASC Chem. Biol.* 11, 1693-1701.
- Sikorski, M., Biesiadka, J., Kasperska, A. E, Kopcinska, J., Lotocka, B., Golinowski, W., & Legocki, A. B. (1999). Expression of genes encoding PR10 class pathogenesis-related proteins is inhibited in yellow lupine root nodules. *Plant Sci.* 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.
- Tan, D. X., Manchester, L. C., Terron, M. P., Flores, L. J. & Reiter, R. J. (2007). One molecule, many derivatives: a never-ending interaction of melatonin with reactive oxygen and nitrogen species? *J. Pineal. Res.* 42, 28-42.
- Thann, N. N., Heer, C., Laatsch, H. & Hardeland, R. (2006). Reactions of the melatonin metabolite N1-acetyl-5-methoxykynuramine (AMK) with the ABTS cation radical: identification of new oxidation products. *Redox Rep.* 11, 15-24.
- 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 Environ.* 27, 1112–1121.
- Whitmore, L. & Wallace, B. A. (2008). Protein secondary structure analyses from circular dichroism spectroscopy: methods and reference databases. *Biopolymers* 89, 392–400.
- 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.
- VanEtten, H. D, Mansfield, J. W., Bailey, J. A. & Farmer, E. E. (1994). Two classes of plant antibiotics: phytoalexins versus "phytoanticipins". *Plant Cell* 6, 1191-1192.
- Van Loon, L. C., Rep, M. & Pieterse, C. M. J. (2006). Significance of inducible defense-related proteins in infected plants. *Annu. Rev. Phytopathol.* 44, 135-162.
- Xie, Y. R., Chen, Z. Y., Brown, R. L. & Bhatnagar, D. (2010). Expression and functional characterization of two pathogenesis-related protein 10 genes from *Zea mays*. *J. Plant. Physiol.* 167, 121–130.
- Yamamoto, M., Torikai, S. & Oeda, K. (1997) A major root protein of carrots with high homology to intracellular pathogenesis-related (PR) proteins and pollen allergens. *Plant Cell Physiol.* 38, 1080–1086.

PUBLICATION I



Crystallographic and CD probing of ligand-induced conformational changes in a plant PR-10 protein



Joanna Śliwiak^a, Rafał Dolot^b, Karolina Michalska^{c,1}, Kamil Szpotkowski^a, Grzegorz Bujacz^a, Michał Sikorski^a, Mariusz Jaskolski^{a,c,*}

^a Center for Biocrystallographic Research, Institute of Bioorganic Chemistry, Polish Academy of Sciences, Poznan, Poland

^b Centre of Molecular and Macromolecular Studies, Department of Bioorganic Chemistry, Polish Academy of Sciences, Lodz, Poland

^c Department of Crystallography, Faculty of Chemistry, A. Mickiewicz University, Poznan, Poland

ARTICLE INFO

Article history:

Received 5 November 2015

Received in revised form 26 November 2015

Accepted 27 November 2015

Available online 28 November 2015

PDB references: Crystal structure of yellow lupine LIPR-10.1A protein in complex with *trans*-zeatin, 4RYV; crystal structure of yellow lupine LIPR-10.1A protein in ligand-free form, 4Y31; crystal structure of yellow lupine LIPR-10.1A protein partially saturated with *trans*-zeatin, 5C9Y.

Keywords:

PR-10 fold

Phytohormone

Cytokinin

Trans-zeatin

Ligand-binding

ABSTRACT

Plant pathogenesis-related class 10 (PR-10) proteins are a family of abundant proteins initially identified as elements of the plant defense system. The key structural feature suggesting PR-10 functionality is a huge hydrophobic cavity created in the protein interior by a scaffold composed of an extended β -sheet wrapped around a long and flexible C-terminal α -helix. Several crystallographic and NMR studies have shown that the cavity can accommodate a variety of small molecule ligands, including phytohormones. The article describes ~ 1.3 Å resolution crystal structures of a *Lupinus luteus* PR-10 isoform LIPR-10.1A, in its free form and in complex with *trans*-zeatin, a naturally occurring plant hormone belonging to the cytokinin group. Moreover we present the structure of the same protein where the saturation with zeatin is not complete. This set of three crystal structures allows us to track the structural adaptation of the protein upon *trans*-zeatin docking, as well as the sequence of the ligand-binding events, step-by-step. In addition, titration of LIPR-10.1A with *trans*-zeatin monitored in solution by CD spectra, confirmed the pattern of structural adaptations deduced from the crystallographic studies. The ligand-binding mode shows no similarity to other zeatin complexes of PR-10 proteins. The present work, which describes the first atomic models of the same PR-10 protein with and without a physiological ligand, reveals that the conformation of LIPR-10.1A undergoes a significant structural rearrangement upon *trans*-zeatin binding.

© 2015 Elsevier Inc. All rights reserved.

1. Introduction

Within the huge superfamily of plant pathogenesis-related (PR) proteins, the members of class 10 (PR-10) are defined as small, slightly acidic and mainly cytosolic proteins. In variance with other PR classes, PR-10 proteins do not have a clearly defined biological function and their classification is often made according to

structural similarity, even though it is quite apparent now that not all plant proteins with PR-10 fold and physicochemical properties are linked to pathogenesis mechanisms (van Loon et al., 2006). The key and common feature of the PR-10 fold is the presence of a hydrophobic cavity formed at the interface of the principal secondary structure elements, which are a long and flexible C-terminal α -helix $\alpha 3$, and a seven-stranded, antiparallel β -sheet. This folding scheme, resembling the thumb in a clenched fist, is completed by a number of loops (L1–L9) and by two additional short α -helices, $\alpha 1$ and $\alpha 2$, which form a V-shaped support for the C-terminal end of helix $\alpha 3$. The cavity has two entrances: E1 formed by helix $\alpha 3$ and loops L3, L5 and L7, and E2 situated between helix $\alpha 3$ and strand $\beta 1$ (Fig. 1a).

Numerous structural and biophysical studies have demonstrated that the PR-10 cavity has the ability to bind diversified small molecules, leading to the suggestion that the PR-10 fold could be a generic solution to ligand binding utilized in different physiological processes in plants. Indeed, the PR-10 fold is exhibited by enzymes such as S-norococlaurine synthase (Berkner et al.,

Abbreviations: 2iP, 6-(γ,γ -dimethylallylamino)purine; ABA, abscisic acid; ANS, 8-anilino-1-naphthalene-sulfonate; CD, circular dichroism; CPPU, *N*-(3-chloropyridyl)-*N'*-phenylurea; CSBP, Cytokinin Specific Binding Protein; DPU, diphenylurea; GA3, gibberellic acid; IAA, 3-indoleacetic acid; LIPR, *Lupinus luteus* pathogenesis-related protein; MPD, 2-methyl-2,4-pentanediol; PDB, Protein Data Bank; PEG, polyethylene glycol; PhBP, Phytohormone Binding Protein; PR, pathogenesis-related; R.M.S.D., root mean square deviation; ZEA, *trans*-zeatin.

* Corresponding author at: Department of Crystallography, Faculty of Chemistry, A. Mickiewicz University, Grunwaldzka 6, 60-780 Poznan, Poland.

E-mail address: mariuszj@amu.edu.pl (M. Jaskolski).

¹ Current address: Argonne National Laboratory, Biosciences Division, Argonne, IL, USA.

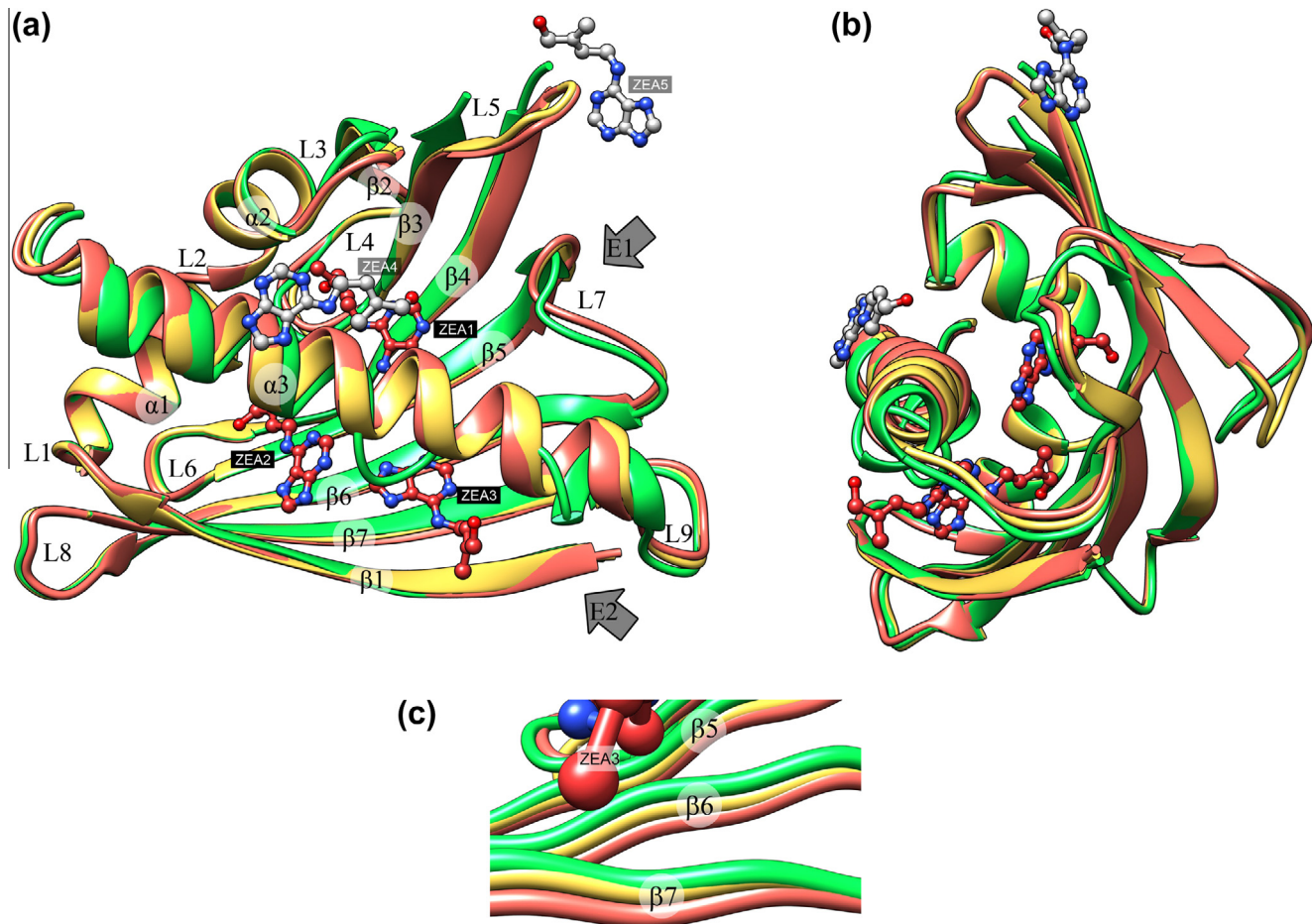


Fig. 1. (a) Superposition of the three LIPR-10.1A structures determined in this work, with annotation of the canonical structural elements of the PR-10 fold. The free form **F** is shown in green, the unsaturated form **U** in yellow, and the saturated form **Z** in red (with the external *trans*-zeatin molecules in gray). (b) A different view of the same superposition. (c) A zoom-in view illustrating the variable degree of caving of the β 5– β 7 fragment of the β -sheet.

2008) or TcmN aromatase/cyclase (Ames et al., 2008), hormone receptors such as Pyl1, which is a component of the abscisic acid (ABA) receptor (Miyazono et al., 2009) or the nodulin MtN13 (Ruszkowski et al., 2013), which takes part in cytokinin signaling in nodulating legumes. Other examples include proteins such as the birch pollen allergen Bet v 1 (Kofler et al., 2012), peanut panallergen Ara h 8 (Hurlburt et al., 2013), or yellow lupine isoform LIPR-10.2B (Fernandes et al., 2008, 2009), with likely storage or transport functions, as they are capable of accommodating numerous copies of chemically diversified ligand molecules in their huge internal cavity. In an attempt to correlate the PR-10 fold with function, Sliwiak et al. (2013) proposed a taxonomy of PR-10 proteins according to their cavities. In this classification, PR-10 members that bind ligands in a highly specific manner and are often involved in signaling, have cavity of type 1, which is small and shallow, has only one entrance (E1) and usually accommodates only one ligand molecule. On the other hand, PR-10 proteins that bind chemically different ligands nonspecifically and thus tend to have storage function, possess a huge cavity of type 2, which spans the entire body of the protein interior from E1 to E2, and is capable of accommodating two or more ligand molecules. Proteins from the latter group bind their ligands in an unusual manner, not only because of the variable number of the cargo molecules, but also because the number of strong, specific and anchoring interactions between the ligand and protein residues is very limited. However, despite this lack of specific interactions, atomic-resolution crystal structures have revealed in a number of cases a perfect order of the different ligand molecules, most often plant hormones.

Notably, plant hormones from the cytokinin group have been reported to be bound by numerous PR-10 proteins. In plant physiology, cytokinins stimulate cell division and additionally control the symbiotic root nodulation in legumes (Hwang et al., 2012). The most common among these adenine N6-derivatives is *trans*-zeatin (Fig. 2). Interestingly, synthetic urea derivatives, e.g. diphenylurea (DPU) or *N*-(3-chloropyridyl)-*N'*-phenylurea (CPPU), show very strong physiological activity almost identical to that of natural cytokinins (Ricci and Bertoletti, 2009) despite no chemical similarity.

PR-10 proteins were first linked to cytokinin binding by Fujimoto et al. (1998), who isolated a cytosolic fraction from *Vigna radiata* with high cytokinin affinity. The protein responsible for this interaction, named Cytokinin Specific Binding Protein (CSBP), was later shown to bind cytokinins with lower affinity (Pasternak et al., 2006) than originally reported and the binding mode (within a type 1 cavity) turned out to have a puzzling diversity. Moreover, a recent study revealed that the CSBP proteins bind gibberellin, which is an entirely different phytohormone, with higher affinity and specificity (Ruszkowski et al., 2014). Accordingly, the term Phytohormone Binding Protein (PhBP) has been proposed to replace CSBP as more appropriate. Nevertheless, another PR-10 protein with type 1 cavity, *Medicago truncatula* nodulation protein MtN13, involved in cytokinin signaling (Ruszkowski et al., 2013), was shown to bind cytokinins in a highly specific and reproducible manner.

The PR-10 members that have a large type 2 cavity and are promiscuous in ligand selection are also capable of cytokinin

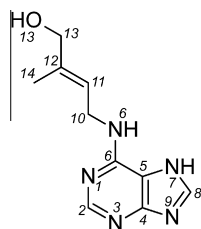


Fig. 2. The chemical structure of *trans*-zeatin, with the correct (Jaskolski, 2013) atom-numbering scheme.

binding, as illustrated by several crystal structures of the birch pollen allergen Bet v 1 in complex with kinetin (PDB IDs 4a85, 4a86) and yellow lupine LIPR-10.2B protein in complex with *trans*-zeatin (2qim). However, no common cytokinin binding pattern can be derived from all these complexes, as the binding modes are characterized by different stoichiometries and protein–ligand interactions. Moreover, the structural studies have emphasized the amazing promiscuity in terms of the chemical nature of the ligands acceptable by these two PR-10 members, as in addition to cytokinins, such molecules as steroids or flavonoids could form complexes as well (Kofler et al., 2012; Śliwiak, unpublished results).

Despite the large number of crystal structures of PR-10 proteins deposited in the PDB, there is no single case where the same protein would be studied in its free and ligand-bound states. The closest situations one can find for such a pair are presented by Bet v 1 (4a88/4a85) and the Hyp-1 protein from St John's wort (3ie5/4n3e), but the “free” states in both these cases correspond in fact to structures in complex with organic molecules from the crystallization buffers, MPD and PEG, respectively, raising doubts if such structures can indeed serve as templates for the ligand-free conformations. On the other hand, there are several crystal structures of PR-10 proteins in their free form, such as the yellow lupine isoforms LIPR-10.1B (1ifv) and LIPR-10.2A (1xdf), for which no matching complexes have been studied.

In this paper, we describe high-resolution crystal structures of the LIPR-10.1A protein from yellow lupine, which represents the subgroup of PR-10 members with a large type 2 cavity, in complex with *trans*-zeatin (1.38 Å) and in free state (1.32 Å). The latter structure significantly extends the resolution of the model already available in the PDB (1icx, 1.95 Å; Biesiadka et al., 2002). Interestingly, unlike the LIPR-10.2B isoform mentioned above, LIPR-10.1A seems to be much less promiscuous in its ligand “choice”, as despite numerous attempts, it was not possible to obtain crystalline complexes with other ligands. Upon co-crystallization with *trans*-zeatin, the protein internalized three hormone molecules in the binding cavity, and two additional ligand molecules were trapped at the protein surface, where they mediate protein–protein interactions in the crystal lattice. In addition, we also present the structure of an LIPR-10.1A complex unsaturated with *trans*-zeatin (1.50 Å), which can be treated as an “intermediate binding state”. This complement of three high-resolution crystal structures allows us to visualize the conformational rearrangements of the protein that are induced by the docking of the ligand molecules in the binding pocket. These structural transformations were additionally monitored and correlated with the situation in solution using circular dichroism (CD) measurements.

2. Materials and methods

2.1. Crystallization and data collection

LIPR-10.1A was expressed and purified as previously described (Biesiadka et al., 2002), with one important exception. Specifically,

the bacteria were grown overnight at 18 °C (instead of 37 °C), which resulted in the expression of the protein exclusively in the soluble fraction (preventing the formation of inclusion bodies and obviating the need for refolding). Crystallization was carried out using the hanging drop vapor diffusion method at 278 K. For the crystallization of the ligand-free form, protein concentration of 11 mg/ml was used. Single crystals of ligand-free LIPR-10.1A were obtained using as precipitant a solution containing 18% PEG 4K, 0.2 M sodium acetate and 0.1 M HEPES pH 7.0. The crystals were cryoprotected in a 1:1 mixture of the mother liquor and 50% PEG400, and vitrified at 100 K in a cold nitrogen gas stream. Diffraction data extending to 1.32 Å resolution (designated **F**) were collected at beamline BL14.1 of the BESSY synchrotron in Berlin.

Prior to crystallization of the zeatin-saturated complex, the LIPR-10.1A protein at 10 mg/ml was incubated with five-fold molar excess of *trans*-zeatin (ZEA). The incubation mixture was then used for setting up the crystallization drops. Crystals of the LIPR-10.1A/ZEA complex grew over a reservoir solution containing 1.8 M $(\text{NH}_4)_2\text{SO}_4$ and 0.1 M MES pH 6.5. Diffraction data extending to 1.38 Å resolution (**Z**) were collected using beamline I911-2 at MAX-lab in Lund. Crystals of a complex unsaturated with *trans*-zeatin were obtained using protein at 20 mg/ml concentration preincubated with a three-fold molar excess of the ligand, and 1.6 M $(\text{NH}_4)_2\text{SO}_4$ as the precipitating agent. Diffraction data extending to 1.50 Å resolution (**U**) were collected at beamline X11 of the DORIS storage ring at EMBL/DESY in Hamburg. Crystals of both complexes were cryoprotected using 25% glycerol in the mother liquor and vitrified at 100 K in a cold nitrogen gas stream.

The diffraction data for **F** and **Z** were indexed, integrated and scaled in *HKL2000* (Otwinowski and Minor, 1997) and for the unsaturated complex **U** in *XDS* (Kabsch, 2010).

2.2. Structure solution and model refinement

The present structure **F** of ligand-free LIPR-10.1A is isomorphous with the orthorhombic structure ($P2_12_12_1$) reported previously at 1.95 Å resolution by Biesiadka et al. (2002) and the coordinates of the protein atoms of the PDB model 1icx could be used directly in the refinement against the new 1.32 Å data. Manual model rebuilding was carried out in *Coot* (Emsley et al., 2010), while for the crystallographic refinement against maximum-likelihood targets the *Refmac5* (Murshudov et al., 1997) program was used.

The crystals of the LIPR-10.1A/ZEA complex **Z** represent a new monoclinic form ($C2$) whereas the crystals of the unsaturated complex **U** have the same space group as the ligand-free structure. The phase problem was solved by molecular replacement in the *Phaser* program (McCoy et al., 2007), using 1icx as the search model in both cases. Improvement of the initial solution was achieved by manual rebuilding in *Coot* and the final crystallographic refinement was carried out in *phenix.refine* (Afonine et al., 2012). A model of *trans*-zeatin for stereochemical restraint targets was taken from the PDB library. Five molecules of ZEA were modeled in the electron density of the ligand-saturated structure **Z**, whereas the unsaturated structure **U** was refined and deposited in the Protein Data Bank (PDB) without ligand atoms because the corresponding electron density, while clearly indicating the presence of the ligand molecules in the binding cavity, was too fragmentary and ambiguous for reliable modeling of the hormone molecules.

All three structures were refined anisotropically with riding H atoms of the protein molecules included in F_c calculations. The final models were validated with *MolProbity* (Chen et al., 2010). The refinement statistics are given in Table 1. Atomic coordinates

Table 1
Data collection and refinement statistics.

Crystal form	F	U	Z
<i>Data collection</i>			
Temperature (K)	100	100	100
Space group	$P2_12_12_1$	$P2_12_12_1$	C2
<i>Cell dimensions</i>			
<i>a</i> , <i>b</i> , <i>c</i> (Å)	35.1, 56.5, 61.9	35.5, 57.9, 62.5	70.1, 63.9, 47.6
α , β , γ (°)	90, 90, 90	90, 90, 90	90, 126.7, 90
Resolution (Å)	30.0–1.32 (1.37–1.32) ^A	50.0–1.50 (1.59–1.50)	99.0–1.38 (1.43–1.38)
R_{merge} (I/σ)	0.059 (0.401) 19.70 (2.00)	0.090 (0.829) 21.96 (2.45)	0.054 (0.209) 24.74 (6.48)
Completeness (%)	99.9 (99.7)	99.5 (98.2)	99.2 (99.6)
Redundancy	3.7 (3.3)	7.2 (2.4)	4.1 (3.5)
<i>Refinement</i>			
Resolution (Å)	28.27–1.32	42.48–1.50	42.49–1.38
Reflections work/test	28,503/1059	21,204/1001	33,856/896
$R_{\text{work}}/R_{\text{free}}$ (%)	15.3/21.2	18.1/24.3	14.3/17.0
<i>No. atoms</i>			
Protein	1248	1243	1239
Ligand/ion	0/4	0/0	80 (ZEA)/ 5 (SO ₄ ²⁻)
Water	178	148	222
<i>(B) factors (Å²)</i>			
Protein	19.98	27.10	14.55
Ligand/ion	0/33.99	–	23.83/36.10
Water	31.03	37.92	30.72
<i>R.M.S. deviations from ideal</i>			
Bond lengths (Å)	0.019	0.018	0.015
Bond angles (°)	1.8	1.8	2.0
<i>Ramachandran statistics (%)</i>			
Favored	99.4	99.4	99.4
Outliers	0	0	0
PDB code	4Y31	5C9Y	4RYV

^A Values in parentheses correspond to the last resolution shell.

and structure factors have been deposited in the PDB with the accession codes 4ryv (**F**), 4y31 (**Z**) and 5c9y (**U**).

2.3. Circular dichroism measurements

Circular dichroism (CD) spectra were recorded on a JASCO J-815 CD spectrometer equipped with a Peltier thermostated cell holder. Protein was dialyzed against 10 mM sodium phosphate buffer pH 7.0, containing 75 mM NaF. Protein samples at the concentration of 0.01 mg/ml were incubated for 1 h with 1-, 2-, ... 10-fold molar excess of *trans*-zeatin before measurements. 600 μ l of protein sample were placed in a 2-mm quartz cuvette for the measurements. Each CD spectrum was the average of three scans at continuous scanning mode, corrected by subtracting the spectrum of the buffer solution at identical conditions. Each scan in the range of 185–340 nm was obtained with a scanning rate of 100 nm/min, 1 nm bandwidth, 0.5 nm data pitch and data integration time of 1 s. The spectra were analyzed using the DICHROWEB server (Whitmore and Wallace, 2004, <http://www.cryst.bbk.ac.uk/cdweb/html/> accessed 01.04.2015).

2.4. Software

The volumes of the protein internal cavities were calculated using the SPACEBALL (Chwastyk et al., 2014) server (<http://www.ifpan.edu.pl/~chwastyk/spaceball/> accessed 20.10.2015). UCSF Chimera (Pettersen et al., 2004) and PyMol (The PyMOL Molecular Graphics System, Version 1.7.4 Schrödinger, LLC) were used for structural alignments and for the preparation of the figures. The ALIGN program (Cohen, 1997) was used for C α superpositions and R.M.S.D. calculations.

3. Results and discussion

3.1. Crystal formation

The final crystallization conditions for the LIPR-10.1A/ZEA complex were established after several trials with the commercial crystal screens (Peg/Ion I and II, Crystal Screen I and II; Hampton Res.). Interestingly, these screens were used for testing co-crystallization of three yellow lupine PR-10 isoforms, LIPR-10.1A, LIPR-10.1B and LIPR-10.2B, with several phytohormones and signal molecules including cytokinins (*trans*-zeatin, kinetin, isopentenyladenine (2iP), DPU, CPPU), gibberellin (GA3), auxin (IAA), abscisic acid (ABA), salicylic acid and flavonoids (naringenin, quercetin). While the LIPR-10.2B protein formed crystalline complexes with *trans*-zeatin (2qim), DPU (3e85), CPPU, 2iP, ANS (8-anilinoanthracene-1-sulfonate, a fluorescent dye) and both flavonoids (Śliwiak, unpublished results), LIPR-10.1A formed only a crystalline complex with *trans*-zeatin, and its co-crystallization with other cytokinins invariably resulted in the ligand-free form only. LIPR-10.1B on the other hand, gave only crystals of the free form in all these tests and it was never possible to detect any complex formation.

3.2. Overall features of the three crystal structures

LIPR-10.1A is monomeric in solution as confirmed by size-exclusion chromatography and native gel electrophoresis (not shown). Moreover, no stable quaternary structure could be predicted by PISA (Krissinel and Henrick, 2007) analysis of both the protein–ligand and ligand-free crystal structures. The crystals of LIPR-10.1A in complex with *trans*-zeatin were created upon crystallization of protein pre-incubated with five- or three-fold molar excess of *trans*-zeatin. The level of ligand saturation also affects the space group of the crystal structure, as the presence of five-fold molar excess of the ligand resulted in the formation of C2 crystals **Z**, whereas three-fold molar excess of the ligand resulted in $P2_12_12_1$ crystals of the same symmetry as the free form but with slightly bigger unit cell (Table 1). The different space group and crystal packing is also reflected in different solvent content, as in the case of **Z** it is 52.5% (with Matthews volume $V_M = 2.59 \text{ \AA}^3/\text{Da}$), whereas in the case of the unsaturated complex **U** it is 35.9% ($V_M = 1.92 \text{ \AA}^3/\text{Da}$), slightly more than for the ligand-free structure **F**, 32.9% ($V_M = 1.83 \text{ \AA}^3/\text{Da}$). In all three forms there is one protein molecule in the asymmetric unit.

All crystals diffracted X-rays to high angles and it was possible to refine the structures at high resolutions of 1.38 Å (**Z**), 1.50 Å (**U**) and 1.32 Å (**F**). Atomic displacement parameters of the protein, ligand and solvent non-H atoms were refined anisotropically for the **Z** and **F** structures, and the same refinement strategy was also used for **U** (with anisotropic ADPs for the protein component only) after confirming that it significantly reduced both R_{work} and R_{free} and as the TLS option was unsatisfactory. Also, anisotropy analysis carried out for **U** using the PARVATI server (Merritt, 2012) (<http://skuld.bmsc.washington.edu/parvati/> accessed 25.11.2015) did not indicate any abnormalities.

Inspection of the electron density maps of the ligand-saturated complex **Z** phased by the protein component only revealed the presence of five ZEA molecules, three of which (1–3) are bound in the protein interior, while the remaining two (4, 5) are located at the protein surface. Additional elongated patches of positive $F_o - F_c$ electron density were observed in the vicinity of the β_3 , L3, β_4 and L5 elements of the protein fold but their poor quality was insufficient to allow reliable modeling of any chemical molecules present in the crystallization buffer. The final **Z** model includes one LIPR-10.1A molecule, five *trans*-zeatin molecules, 222 water molecules and one sulfate anion. The electron density of the pro-

tein molecule is of a very high quality, allowing unambiguous tracing of the entire polypeptide main chain (residues Gly1–Tyr155) and all side chains (Fig. 3a) without any breaks. The absence of the N-terminal Met0 residue is the effect of the *Escherichia coli* methionylaminopeptidase (MAP) activity in the expression system used to produce the recombinant protein. The electron density of the internal ZEA molecules, especially of ZEA2 and ZEA3, is of superb quality, allowing easy modeling and determination of the ZEA conformation (Fig. 3b). The average B_{eq} values for the ZEA molecules after final refinement are 31, 17 and 19 Å² for ZEA1, ZEA2 and ZEA3, respectively, confirming their very good definition in electron density maps. The electron density of the superficial ZEA molecules is less perfect and indicates a degree of rotation of these molecules and a less fixed position. The model refinement converged with R_{work} and R_{free} of 14.3% and 17.0%, respectively. The high stereochemical quality of the model is confirmed by *MolProbity* statistics and the distribution of main-chain torsion angles in the Ramachandran plot (Table 1).

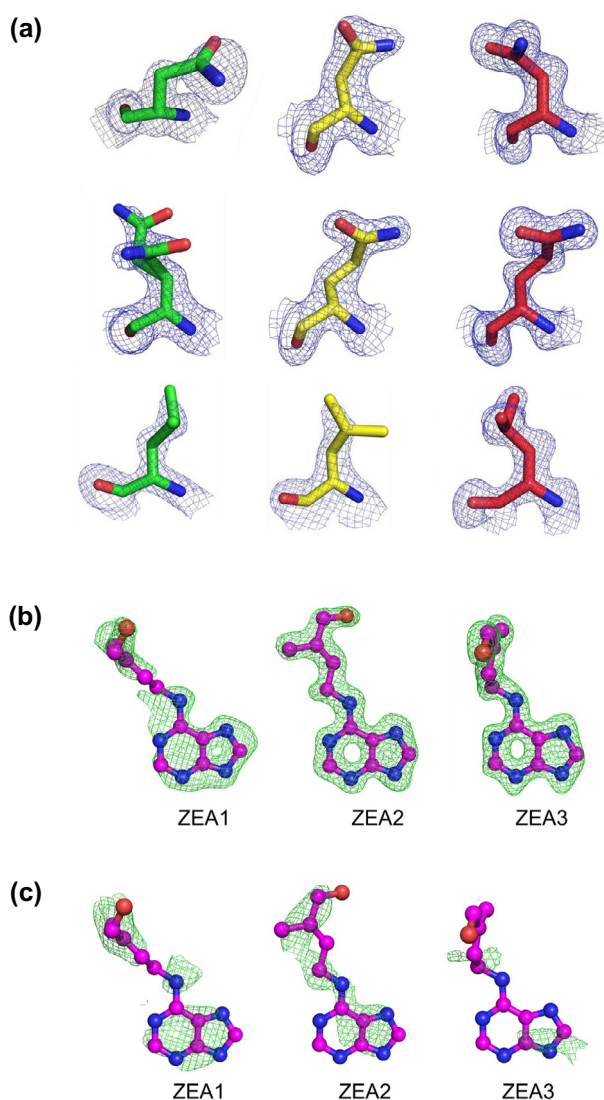


Fig. 3. (a) $2F_o - F_c$ electron density contoured at 1.0σ , illustrating the modeling of Asn7, Gln9 and Leu139 in the three crystal forms, color-coded as in Fig. 1a. (b) Omit $F_o - F_c$ maps contoured at 3.0σ for the internal ZEA molecules in the saturated complex **Z**. (c) $F_o - F_c$ maps contoured at 3.0σ , generated for the unsaturated structure **U** at the locations of the internal ZEA molecules, presented as ball-and-stick models taken from the saturated complex **Z**. *trans*-Zeatin molecules were not modeled in **U** and are shown in (c) only for orientation.

In the structure of the unsaturated complex **U**, flat patches of positive $F_o - F_c$ electron density could be observed in the positions corresponding to ZEA1 and ZEA2, with the general shape indicating the presence of the zeatin purine ring and tail (Fig. 3c). However, the quality of these fragmented maps was too poor to allow responsible modeling of the ligands, which very likely populate the ZEA1 and ZEA2 sites at low occupancy. The $F_o - F_c$ maps also revealed residual electron density at the ZEA3 site (Fig. 3c) and in a new superficial position. However in these cases it was not even possible to figure out the ligand orientation. At the conclusion of the refinement of this unsaturated complex **U** it was decided to exclude the ZEA ligands from the coordinate file in spite of the evidence that the ZEA1 and ZEA2 ligands were present in the crystal structure at low occupancy. The corresponding coordinate file 5c9y was deposited in the PDB with the annotation that the structure might contain unmodeled ZEA ligands. The electron density of the protein model **U** allowed uninterrupted tracing of the entire main chain and most of the side chains (Fig. 3a). There is, however, lack of electron density for the following side chains in the $\alpha 3$ helix: Gln133, Lys135 and Phe136, indicating an increased mobility of this region. The final **U** model consists of one protein molecule and 146 water molecules. The model was refined to R_{work} and R_{free} of 18.1% and 24.3%, respectively.

The ligand-free model of LIPR-10.1A presented here (**F**) has significantly improved resolution (1.32 Å) compared to the previous model (1.95 Å) deposited in the PDB (1icx), thereby allowing better analysis of the ordered and disordered regions of the protein in the absence of stabilizing ligands. Indeed, analyzing the electron density one can observe that almost the entire main chain has excellent electron density with significant breaks corresponding to His59–Gly61 in loop L5 and to Ala134–Lys135 in helix $\alpha 3$, the latter break indicating a higher mobility of the middle part of the helix. Electron density is also missing for side chain atoms close to the above-mentioned mobile regions, namely Gln133, Phe136, Lys137 and His62. The final **F** model was refined to $R_{\text{work}}/R_{\text{free}}$ of 15.3/21.2%. It consists of one protein molecule, 180 water molecules and one acetate anion.

3.3. Overall fold and cavities of the three LIPR-10.1A forms

As all other PR-10 proteins, the LIPR-10.1A molecule consists of a seven-stranded (strands $\beta 1$ – $\beta 7$) antiparallel β -sheet and three α -helices ($\alpha 1$ – $\alpha 3$). The long C-terminal helix $\alpha 3$ (residues Leu126–Ala151), connected to the rest of the protein by loop L9, is a centerpiece of the structure, as it is “gripped” by the β -sheet and supported at its C-terminus by the two short accessory helices $\alpha 1$ and $\alpha 2$. The $\alpha 3$ helix appears to be the key element responsible for the adaptation of the structure to the presence of the *trans*-zeatin ligands. In the ligand-free form **F**, it is unwound at the section of Arg131–Leu139 (Fig. 1a and b). In addition, there is a break in the main-chain electron density in the middle of this segment, at residues Ala134–Lys135, and no electron density for the side chains of Gln133–Lys137. The appearance of the electron density maps indicates that the helix is not only unwound but also very mobile or even disordered in this region in the absence of *trans*-zeatin. In the cases of the two ZEA complexes (**Z** and **U**), the $\alpha 3$ helix is straight and with properly preserved secondary structure.

The secondary structure elements are connected by the nine loops L1–L9. Here again in the case of the ligand-free form **F** the position of loop L5 could not be modeled in the electron density at the segment of His59–Gly61. No such problem was encountered in case of the **Z** and **U** complexes.

As mentioned above, the spatial organization of the secondary structure elements leads to the formation of a huge cavity between the $\alpha 3$ helix and the concave face of the β -sheet. The volume and shape of this cavity seem to be dependent on the cargo content

as the volume of the largest hilum is the smallest in the free form **F** (754 Å³), medium in the unsaturated complex **U** (1080 Å³) and the biggest in the saturated complex **Z** (1979 Å³). The large difference between the calculated volumes of the unsaturated and saturated complexes arises from the way the cavity volume is estimated by the *SPACEBALL* algorithm (Chwastyk et al., 2014), which probes the void spaces with a fixed-radius sphere. In this method, several small, narrowly connected chambers may be assessed very differently than one large, sphere-like cavity of the same geometrical volume. A comparison of the shapes of the three LIPR-10.1A cavities (Fig. 4) indicates that the void of the unsaturated complex **U** is not fully opened and is separated into smaller sub-volumes by a narrowing at the center. Moreover, the pocket for ZEA3 binding, which in the free form **F** is absent altogether, in **U** is not fully formed and separated from the main cavity. Fig. 4 illustrates that the small cavity volume in **F** is the result of not only the absence of a ZEA3 pocket, but also of the doughnut shape of the cavity, which has a significant central constriction.

The side chains pointing to the lumen of the cavity are mostly non-polar, thus creating an environment suitable for hydrophobic ligands. The cavity has two entrances. One of them, E1, is located between the N-terminus of helix α 3 and loops L3, L5 and L7, and is partially gated by the side chain of Lys137 in the saturated complex. The second opening, E2, is formed between the N-terminal fragment of helix α 3 and strand β 1.

3.4. Trans-zeatin binding and conformation

3.4.1. The internal trans-zeatin ligands

The hydrophobic cavity of the saturated complex **Z** is occupied by three *trans*-zeatin molecules (ZEA1, ZEA2, ZEA3) which were modeled with full occupancy. The molecules of ZEA2 and ZEA3 are perfectly defined in the electron density (Fig. 3b) and their stability is confirmed by the low average *B*-factors (17 and 19 Å², respectively). ZEA1 is somewhat less well ordered, with an average *B*-factor of 31 Å². It is possible that this ligand molecule could be bound in several slightly different orientations/conformations, but we were unable to model such a disorder, as introduction of any alternative positions of the ZEA1 molecule invariably led to clashes with the protein atoms. Moreover, refinement with ZEA1 in multiple conformations always resulted in a rise of *R*_{free}. For these reasons, the final **Z** model comprises only one set of coordinates for ZEA1.

ZEA1 is located near helices α 2, α 3 and strands β 3, β 4 (Fig. 1a) and has a vista of the E1 entrance (Fig. 5a). It forms three direct hydrogen bonds with the protein molecule (Fig. 5b). Two of them are formed by the hydroxyl group of the aliphatic tail of the ligand molecule with the side chains of Asp27 and Lys53. The adenine moiety of ZEA1 is also anchored to the protein with a hydrogen bond between the N3 atom and the hydroxyl group of one of the two alternative conformations of Tyr82. Additional stabilization

of ZEA1 is achieved by van der Waals contacts involving Ile55, His68 and Leu141.

ZEA2 sits very deep in the cavity, at the interface between helix α 3 and strands β 6 and β 7 (Fig. 1a) where it is completely buried and has no access to any of the entrances (Fig. 5a). It has only one direct hydrogen bond with the protein molecule but in addition forms four contacts mediated by structural water molecules (Fig. 5c). The adenine N9 atom interacts with the O δ 1 atom of Asn7. This interaction indicates that the ZEA2 molecule has a proton attached to the N9 atom, either as a result of protonation or N7/N9 tautomerism. The remaining hydrogen bonds, involving the N1, N6 and N7 nitrogen atoms and the terminal hydroxyl group of ZEA2, are established via three water molecules one of which is shared by the N6 and N7 atoms and anchors them to the O atom of Lys102. The water molecule interacting with the N1 atom relays this interaction to the O atom of Gly138, whereas the other water molecule, interacting with the ZEA2 hydroxyl group, is ultimately bonded with Leu78 and Ser101. Additionally, Gln9, Leu22, Tyr80, Ser101, Ile115 and Phe142 participate in van der Waals interactions with the ZEA2 ligand.

ZEA3 partially protrudes out of the cavity through the E2 entrance (Fig. 5a). This molecule is directly attached to the protein by only one hydrogen bond, which links the terminal hydroxyl group of ZEA3 and the side chain of Asp132 (Fig. 5d). Additional hydrogen bonds exist between the N3, N6, N7 and N9 atoms and three water molecules. In particular, N3 interacts with a water molecule that relays the interaction to Tyr99. The N6 and N7 atoms again share one hydrogen-bonded water molecule, through which they are connected with the N δ 2 atom of Asn7. The van der Waals contacts of ZEA3 are established with residues Phe5, Asn7, Val117, Phe119, Arg131 and Lys135.

All three *trans*-zeatin molecules bound within the protein cavity have limited numbers of direct hydrogen bonds with the protein molecule. However, the stability of the ZEA2 and ZEA3 ligands seems to be attributed to the dense network of hydrophobic interactions as well as to the water-mediated hydrogen bonds with the protein atoms. In the case of ZEA1, these two types of interactions are less extensive. The weaker docking of the ZEA1 molecule in the binding pocket is related to the fact that the space available for this ligand is much bigger than the molecular volume of a *trans*-zeatin molecule and it is difficult to detect any shape complementarity between the interacting partners. In fact, ZEA1 seems to be accidentally trapped in the cavity and then blocked by the Lys137 gating of the E1 opening. ZEA2 and ZEA3, on the other hand, are found in much better fitting spaces, resulting in stronger binding and improved stabilization.

3.4.2. The superficial trans-zeatin ligands

Apart from the zeatin molecules bound in the protein interior of **Z**, two ligand molecules, ZEA4 and ZEA5, were also identified at the protein surface. ZEA4 is located with 0.5 occupancy at a twofold

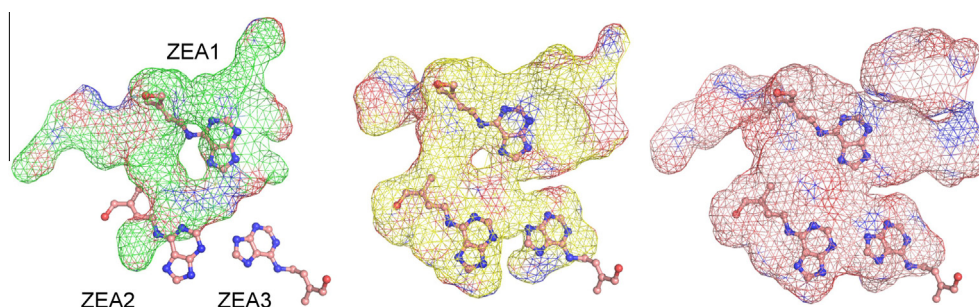


Fig. 4. The shape of the internal cavity in the three forms of LIPR-10.1A, color-coded as in Fig. 1a, with additional coloring of the surfaces at nitrogen (blue) and oxygen (red) atoms. The *trans*-zeatin molecules (ZEA), taken from the saturated structure **Z**, were superposed to illustrate the formation of the binding sites.

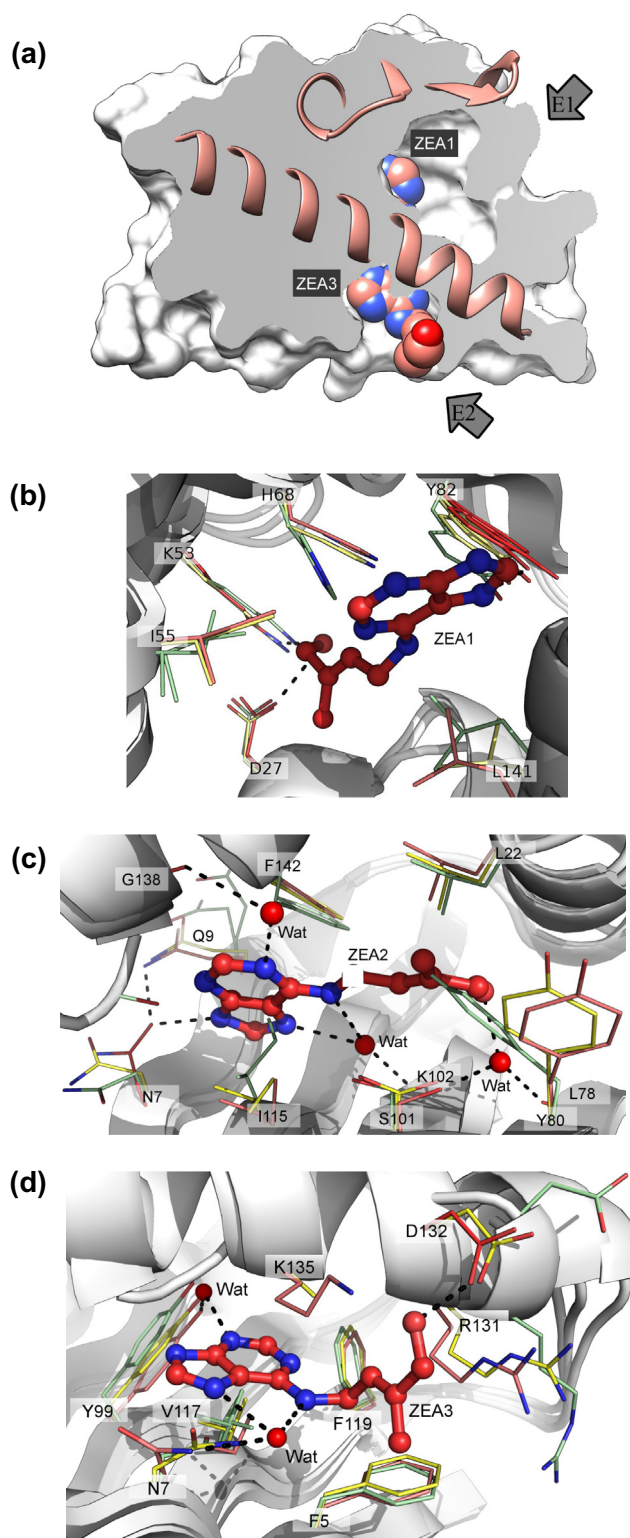


Fig. 5. (a) ZEA molecules in space-filling representation within the surface of the LIPR-10.1A protein in the saturated complex **Z**. The binding sites of ZEA1 (b), ZEA2 (c) and ZEA3 (d), with residues forming direct hydrogen bonds with the ligand molecules shown in red and those forming van der Waals contacts or water-mediated hydrogen bonds shown in salmon. The corresponding residues from the ligand-free (**F**, green) and unsaturated (**U**, yellow) forms were superposed to illustrate their gradual adaptation.

axis and bridges two symmetry-related protein molecules via van der Waals contacts involving residues Gly140, Lys143, Ala144 in the central part of the solvent-exposed face of helix α 3. ZEA5

(0.7 occupancy), found near loop L5 where its adenine ring makes stacking interactions with the imidazole ring of His62, is also located near a twofold axis, but at a distance (~ 1.5 Å) that allows it to form adenine–adenine stacking interactions across the dyad, thus extending the aromatic stack to His62–ZEA5–ZEA5′–His62′. From their protein interactions and location at special sites in the crystal lattice, it is evident that ZEA4 and ZEA5 have an important crystal packing role in **Z**. This observation explains why the unsaturated form **U**, in which the protein molecules are evidently charged (at least to some degree) with the internal *trans*-zeatin molecules, but where there are no interstitial ligands, crystallized isomorphously with the free form **F** and not with the saturated complex **Z**.

In the unsaturated crystal structure **U**, there are patches of electron density that coincide with the positions of ZEA1 and ZEA2 (Fig. 3b). At the location corresponding to ZEA3, however, the patch of residual electron density is too small to unambiguously correlate it with *trans*-zeatin (Fig. 3c). There is also residual electron density near loop L5 but at a different position than ZEA5. In general, there is no electron density to indicate any traces of the ZEA4 and ZEA5 binding sites. The absence of external ZEA molecules in the unsaturated complex **U** may be responsible for retaining the C2 space group symmetry of the free form **F**.

3.4.3. Conformation of the *trans*-zeatin ligands

All the zeatin molecules have evident *trans* configuration, as defined by the C10–C11–C12–C13 torsion angle (Table 2). The orientation of the terminal OH group, defined by the C11–C12–C13–O13 angle is, however, variable, not only *cis* ($\sim 10^\circ$) or *trans* ($\sim 170^\circ$) type, but also nearly perpendicular to the C13–C12–C14 plane. The disposition of the isoprenoid chain (N1–C6–N6–C10) at the exoamino N6 atom is uniformly distal to the imidazole ring, i.e. the substituent is directed towards the N1 atom of the purine system. The remaining two torsion angles (C6–N6–C10–C11, N6–C10–C11–C12) are quite variable, endowing the isoprenoid chain with a range of conformations, of which only one, in ZEA2 (Fig. 5c), is all-*trans* (extended). On comparison with the conformation of the *trans*-zeatin ligands reported in the VrPhBP (Pasternak et al., 2006) and LIPR-10.2B (Fernandes et al., 2008) complexes, one can conclude that this phytohormone is rather flexible, easily adapting its conformation to the requirements of the binding partner, especially using rotations around the N6–C10, C10–C11 and C12–C13 single bonds. A constant feature is the *cis* conformation of the N1–C6–N6–C10 torsion angle.

3.5. Structural adaptation of LIPR-10.1A on binding of *trans*-zeatin

Superposition of the three forms of LIPR-10.1A (Figs. 1, 4 and 5b–d) illustrates their differences, which can be viewed as a gradual adaptation of the protein structure to the *trans*-zeatin ligands.

Table 2

Conformation of the *trans*-zeatin molecules in structure **Z**. The internal ligands (ZEA1, ZEA2, ZEA3) are separated from the superficial ones (ZEA4, ZEA5) with a heavy vertical line. The atom numbering scheme is given in Fig. 2.

Torsion angle ($^\circ$)	ZEA1	ZEA2	ZEA3	ZEA4	ZEA5
N1–C6–N6–C10	–13.1	–0.6	5.9	1.5	–11.4
C6–N6–C10–C11	104.4	–175.8	112.6	–179.8	65.2
N6–C10–C11–C12	167.2	–169.7	101.2	9.0	90.0
C10–C11–C12–C13	178.1	179.2	179.5	–177.7	179.0
C11–C12–C13–O13	93.6	13.8	9.3	–62.8	–167.6

When it comes to the secondary structure elements, the binding events mainly affect the $\alpha 3$ helix, which is unwound in the free form at a segment comprising eight residues (over two turns). It is important to note that the R.M.S.D. value of the structural alignment between the C α atoms of the free form **F** and the unsaturated form **U** is 0.86 Å. This value is somewhat higher (1.03 Å) for the alignment of the free and saturated forms (**F–Z**). However, if only the $\alpha 3$ helices are aligned, this value is ~ 1.3 Å in both cases, indicating that the changes are mostly localized to this structural element. On the other hand, the C α R.M.S.D. value between the two liganded forms (**U–Z**) is only 0.47 Å for all residues, and 0.40 Å for helix $\alpha 3$ alone.

When comparing the protein conformation of the three forms, one can identify several residues that significantly change their position upon interaction with the ligand molecules, leading in general to the increased order of the $\alpha 3$ element. The majority of those residues interact with the ZEA3 molecule but the first stimulus that triggers the helix formation seems to be repulsion between ZEA2 and Leu139, which occupies this binding site in the ligand-free form (Fig. 6). Leu139 caps the E2 entrance in the free form **F** where it forms a van der Waals contact with the C β atom of Asn7, pulling down the $\alpha 3$ helix towards $\beta 1$. In the **Z** complex, the C $\delta 2$ atom of Leu139 in this conformation would be located less than 2 Å from the N9 atom of ZEA2. In the unsaturated form **U** this side chain is slightly pushed away, to be totally moved outside of the E2 entrance in the saturated form **Z**. This change is also mediated by Asn7 and Gln9, both located in strand $\beta 1$, below Leu139. They create two hydrogen bonds in the complex, one between the O $\delta 1$ atom of Asn7 and the N9 atom of ZEA2, and another one between the N $\epsilon 2$ atom of Gln9 and the O $\delta 1$ atom of Asn7, which in effect directs their hydrophilic side chains toward the ZEA2 molecule. At this conformation, these two hydrophilic side chains push the hydrophobic Leu139 outside of the E2 entrance, facilitating the ordering of helix $\alpha 3$. All these conformational changes are very well supported by electron density (Fig. 3a).

The formation (ordering) of helix $\alpha 3$ favors in the end the binding of the ZEA3 molecule as the last internal ligand. In this docking site in the saturated complex **Z**, Arg131 and Lys135 hold the aliphatic tail of ZEA3 in a sort of a “hydrophobic vise”, whereas in the free form **F** Arg131 is directed away from entrance E2 and Lys135 is in the disordered region (Fig. 6). In the unsaturated complex **U**, Arg131 is close to but not quite yet at the position occupied in the saturated complex, and the side chain of Lys135 is disordered, which can be interpreted as indicating that the unsaturated crystal structure **U** represents an intermediate state leading to the

docking of ZEA3 in its binding site. This assumption is supported by the fact that Asp132, which forms a hydrogen bond with the hydroxyl group of ZEA3, is moved farther away from the ZEA3 binding site in the unsaturated complex **U** and is directed totally outside of E2 in the free form **F**. These three side chains (Arg131, Asp132 and Lys135) seem to act like “tentacles” that reach out of the protein surface to draw a ligand molecule to the ZEA3 binding site.

Another residue whose conformational change could stabilize the $\alpha 3$ helix in the complex **Z** is Leu141, which forms van der Waals interactions with ZEA1, while in the free form **F** it penetrates the lumen of the cavity. The interaction with ZEA1 pushes the Leu141 side chain outside of the cavity, straightening at the same time the entire helix $\alpha 3$. Ordering of $\alpha 3$ additionally induces closing up of the E1 entrance by Lys137, sealing off ZEA1 in its binding site.

Another difference at the level of secondary structure is the pronounced change of the conformation (“curvature”) of loops L3, L5 and L7 (the fingertips of the “gripping hand”), which surround the entrance E1, and which upon complex formation change their orientation to tighten the grip of helix $\alpha 3$. Specifically, loop L3 in complex **Z** is moved by ~ 2 Å in the direction of helix $\alpha 3$ relative to the free form **F**. This bending creates van der Waals contacts between Val33-Ile34 and Ala134 from helix $\alpha 3$, clogging the lumen of the E1 entrance.

Although loop L5 in the **F** form is disordered, precluding a detailed conformational comparison, one can observe that the strands $\beta 3$ and $\beta 4$ (connected by L5), which are straightened out in the **F** form, get significantly bent towards the cavity in the **Z** and **U** forms. In this bent conformation, there are van der Waals contacts of residues from the $\beta 3$ and $\beta 4$ strands with ZEA1 in both protein–ligand structures. Moreover, His62 from loop L5 interacts with the external ZEA4 molecule in the saturated structure **Z**, while this interaction is absent in the unsaturated complex **U**, which explains the slightly different conformation of this loop in these two structures.

Loop L7 is also tilted towards helix $\alpha 3$ and the cavity in both complexed forms, but simultaneously it is also stretched towards loop L5 as a consequence of a hydrogen bond between the N $\delta 2$ atom of His62 from L5 and O $\epsilon 2$ of Glu88 (in L7), which is formed in both protein–ligand complexes (**U** and **Z**), but not in **F**.

The remaining loops seem to be insensitive to the presence of the ligand molecules, as illustrated by their close superposition (not shown). The exception is loop L4, which has a different conformation in each of the LIPR-10.1A forms (Fig. 1b). However, this conformational variability seems to be the effect of crystal packing and is not influenced in a direct way by the *trans*-zeatin cargo molecules.

Finally, it is of note that upon *trans*-zeatin binding there is a pronounced bulging of the strands $\beta 5$, $\beta 6$ and $\beta 7$. These structural elements are cambered most significantly in the saturated complex **Z**, whereas in the free form they are almost collapsed into the interior of the protein (Fig. 1c).

These gradual conformational changes are correlated with the increasing volume of the largest chamber of the internal cavity. Interestingly, the volume changes, which reflect the cargo content, are not related to the cavity surface area, which remains constant at ~ 5600 Å². Such an independence can be achieved by means of the above-mentioned caving-in of the $\beta 5$ – $\beta 7$ sheet and by the presence of cavity-exposed side chains of the β -sheet that pad the walls of the cavity in the free form **F** but are displaced upon ligand binding. These residues, especially His68, Tyr80 and Ile115, protrude into the interior of the ZEA1 and ZEA2 binding sites in the free form **F** but are tucked away from these sites in the complex (Fig. 5b and c).

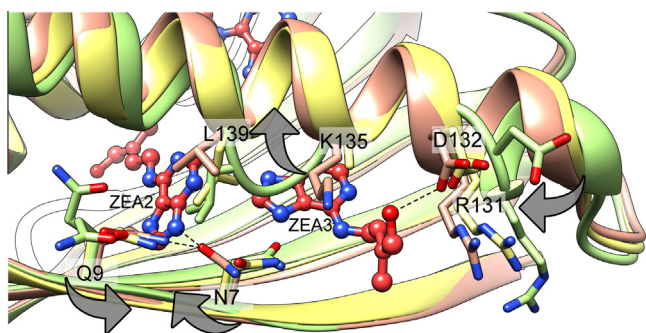


Fig. 6. Cooperativity between the ZEA2 and ZEA3 binding sites, illustrated as a sequence of residue movements that create the ultimate ZEA3 binding site and lead to ordering of helix $\alpha 3$ (light green, free form **F**; yellow, unsaturated form **U**; salmon, saturated form **Z**). The gray arrows illustrate the direction of ligand-induced conformational adaptations.

3.6. Multiple *trans*-zeatin ligands: the sequence and mechanism of binding

Analyzing the orientation of the internal zeatin molecules (Fig. 1), the quality of the $F_o - F_c$ electron density of the ZEA sites in the unsaturated (**U**) and saturated (**Z**) complexes (Fig. 3), as well as the shape of the protein cavity (Fig. 4) and the conformation of the protein side chains in all three forms (Fig. 5b–d), one can try to deduce the order and mechanism of zeatin binding.

The ZEA1 and ZEA2 molecules have a similar orientation inside the cavity, suggesting that they could enter the protein cavity one after another through the same E1 entrance. The main protein cavity has a diverticulum in both complex structures (although in the unsaturated complex **U** it is smaller and partly divided by Asn7, Tyr99, Val115 and Ile117), which extends to the E2 entrance and ultimately becomes the binding pocket for ZEA3. Interestingly, in the free form **F** there is no such a binding pocket at all and in the unsaturated form **U** it is not fully formed, which correlates with the very poor $F_o - F_c$ electron density at the ZEA3 site in the unsaturated complex (Fig. 3c). These observations could suggest that ZEA3 is not present at its binding site before the ZEA1 and ZEA2 molecules have been properly docked at their respective binding sites. With this caveat in mind, the ligand binding scenario might be as follows:

- (1) ZEA 2 and ZEA1 enter the cavity through the large entrance E1.
- (2) ZEA2 is docked via hydrogen bonding with Asn7, leaving room for ZEA1.
- (3) The Gln9 side chain rotates towards Asn7 and a hydrogen bond is created between O δ 1 of Asn7 and N ϵ 2 of Gln9, stabilizing Asn7 for firm docking of ZEA2; ZEA2, Asn7 and Gln9 jointly push the side chain of Leu139 outside of the E2 entrance, initiating the ordering of helix α 3 (Fig. 6).
- (4) A turn of the Asn7 side chain from the N-terminal towards the C-terminal part of β 1 after ZEA2 binding opens the E2 entrance enabling docking of the ZEA3 molecule.
- (5) ZEA3 enters the newly opened binding site, fitting with its aliphatic tail within the vise of Arg131 and Lys135, and forming a hydrogen bond with Asp132 via the hydroxyl group; this finally stabilizes helix α 3; moreover, N7 of ZEA3 interacts with N δ 2 of Asn7 via a water molecule, additionally stabilizing the network of interactions.
- (6) Binding of ZEA3 and the formation of the ZEA3 binding pocket lead to a widening of the internal cavity and disappearance of the separation from the ZEA1 binding site (Fig. 4); this way the ZEA1 pocket becomes large and this ligand molecule gains additional degrees of freedom; this is why the electron density for ZEA1 is the weakest in the saturated complex **Z** (Fig. 3b).

In summary, ZEA3 binds to the protein only after ZEA2 has been docked first. When in place, ZEA3 blocks the E2 entrance, constricting the space around the ZEA2 molecule. At the same time, ZEA3 binding expands the chamber occupied by ZEA1, endowing this ligand with additional freedom within its otherwise sealed-off cavity.

3.7. Crystallographic interpretation of the unsaturated form **U**

While the interpretation of the **F** and **Z** forms is quite clear, because they illustrate, respectively, a ligand-free and 100% *trans*-zeatin-bound protein conformations, the case of **U** is more tricky. We observe here a situation that apparently represents an intermediate protein conformation, and structurally we claim that this reflects partial saturation with the *trans*-zeatin ligand. How-

ever, even if we assume that the ligand binding sites in **U** have fractional occupancy, physically this would mean that some protein molecules in the crystal lattice are empty, while others have the ligands at full occupancy, and that the diffraction experiment gives us a weighted superposition of these two ultimate situations (corresponding to **F** and **Z**). But apparently the protein structure in **U** does not look like a superposition of the states known from **F** and **Z**. If this were the case, we would expect to see a lot more of disorder in the central part of helix α 3, and should even be able to model parts of the backbone in double conformation, where the main chain in **U** gets unwound. This is evidently not the case. The most plausible explanation of this logical puzzle is that the structure **U** represents in fact a partial saturation, where the term “partial” refers to occupation of only *some* of the internal binding sites (which in their own right may also have fractional occupancy in the crystal structure). In this interpretation, the initial binding of the first two ligand molecules at sites ZEA1 and ZEA2, would already induce a major switch in the protein conformation, getting it ready for the docking of the final ligand, ZEA3, whose binding is the most tight, but occurs only after this docking site has been preformed by the “avant-garde” molecules ZEA1 (which ultimately is bound very loosely) and ZEA2.

3.8. Ordering of helix α 3 in the presence of ligands, monitored by CD

The main structural changes of the LIPR-10.1A protein accompanying *trans*-zeatin binding were additionally monitored in solution by circular dichroism measurements in the presence of increasing concentration of the ligand. Circular dichroism is a very sensitive and rapid method for detecting changes in secondary structure content that occur in the presence of denaturing agents or cofactors (Whitmore and Wallace, 2008). Different secondary structures of the protein chain give rise to different CD spectra. The α helix, for example, gives rise to two negative bands at 222 and 208 nm and a positive band at 193 nm, whereas β structures are manifested by one negative band at 218 nm and one positive band at 195 nm (Greenfield, 2006). A change in secondary structure content is typically reflected in a corresponding change of band intensity. Inspection of the CD spectra of LIPR-10.1A recorded in the presence of 0-, 1-, 2-, ... and 10-fold molar excess of *trans*-zeatin revealed a concomitant increase of the intensity of the band at 222 nm, which is associated with the increase of α -helix content in the protein structure. This conclusion was confirmed by a detailed analysis of the CD spectra with the DICHROWEB server (Whitmore and Wallace, 2004) for the detection of the changes in α -helix content (Fig. 7), which additionally showed a saturation (i.e. a plateau of α -helix content) reached at five-fold molar excess of *trans*-zeatin. These experiments in solution confirm the positive effect of *trans*-zeatin on helix α 3 ordering, as deduced from the crystallographic data.

3.9. Comparison of LIPR-10.1A with the previous model and other yellow lupine isoforms in the free state

The model of the free-form LIPR-10.1A protein determined at the resolution of 1.95 Å available thus far in the PDB under the accession code 1icx, is very similar to the **F** structure presented in this work, with an R.M.S.D. of 0.28 Å for the C α superposition. This value is quite low when compared to the R.M.S. deviation of 0.47 Å calculated for the present LIPR-10.1A complexes at different *trans*-zeatin saturation (**Z** vs **U**). The *a/b/c* cell parameters are also very similar: 35.0/56.6/61.8 Å for the 1icx model and 35.1/56.5/61.9 Å in the present crystal. Although the segments His59–Gly61 and Ala134–Lys135, which are disordered in **F** (and thus omitted from the model), were included in the 1icx model without main-chain interruptions, their ambiguous electron

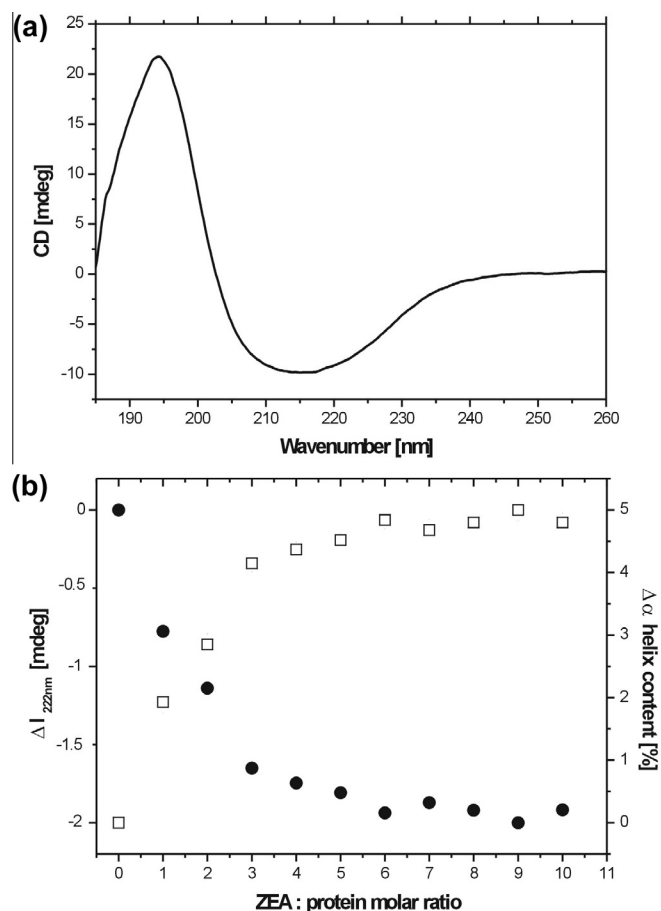


Fig. 7. (a) CD spectrum of the LIPR-10.1A protein without a ligand. (b) Plot of the intensity change at 222 nm (●), or change (%) in α helix content (□) generated in *DICHROWEB* (Whitmore and Wallace, 2004) from the CD spectra collected for protein samples at different ZEA:protein ratio.

density and high average B factors (57 and 76 Å², respectively) indicate poor order of atoms in these regions and make their location doubtful. Apart from this difference, no other significant conformational disparities are observed.

The disorder of helix α 3 in the free form of LIPR-10.1A seems to be a unique feature among the yellow lupine PR-10 isoforms for which ligand-free protein structures are available, namely LIPR-10.1B (1ifv, Biesiadka et al., 2002) and LIPR-10.2A (1xdf, Pasternak et al., 2005) (Fig. 8). Comparison of the corresponding sequences in terms of the residues that take part in *trans*-zeatin binding of LIPR-10.1A, explains the absence of extensive helix α 3 deformation in LIPR-10.1B and LIPR-10.2A. In neither of those proteins is there a triad of residues, Asn7/Gln9/Leu139, which in LIPR-10.1A, after interacting with the ligand molecules, trigger the series of movements that lead to the ordering of helix α 3. Although in LIPR-10.1B there are negatively and positively charged Asp7 and His9 residues at the corresponding positions, Leu139 is replaced by Thr139, whose shorter and polar side chain precludes the structural rearrangements described above.

In LIPR-10.2A (1ivf) the residues at positions 7 and 9 are Asp7 and Ser9. Most importantly, the carbonyl group of Gly139 forms a short hydrogen bond with Tyr82 at the bottom of the cavity, leading to a sharp kink in the middle of helix α 3 and its collapse into the protein interior. Interestingly, Tyr82 is a ligand-interacting residue in most of the PR-10 proteins, while in LIPR-10.2A it is responsible for the helix deformation. Using for comparison the ligand-free model of LIPR-10.2A and a *trans*-zeatin complex of

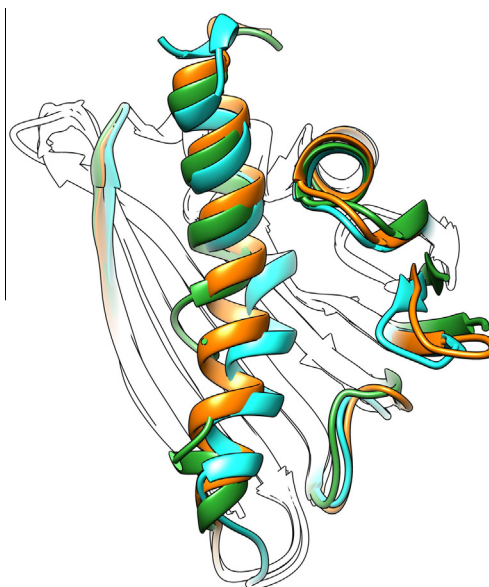


Fig. 8. Superposition of three yellow lupine PR-10 isoforms in their ligand-free state: LIPR-10.1A (F form, green) LIPR-10.1B (2qim, orange) and LIPR-10.2A (1ifv, blue).

LIPR-10.2B, Fernandes et al. (2008) suggested that this collapse of helix α 3 into the protein core could be reverted upon ligand binding. However, in the **Z** and **F** forms of LIPR-10.1A (Fig. 1b) the degree of this collapse is almost the same. The differences between the free and liganded forms of LIPR-10.1A are focused mainly on the degree of order of the α 3 helix and on its distance from the β 1 strand (Fig. 1a). The mode of adaptation of the C-terminal helix, therefore, seems to be specific for each PR-10 case and depends on the particular (generally highly variable) helix α 3 sequence.

3.10. Comparison of the saturated complex with other PR-10 complexes of *trans*-zeatin

α superposition of the available models of PR-10 proteins (VrPhBP, MtN13 and LIPR-10.2B) in complex with *trans*-zeatin (2flh, 4jhg and 2qim, respectively) with the present complex **Z** resulted in R.M.S.D. values of 1.41, 1.63 and 1.72 Å, respectively. Interestingly, the corresponding levels of sequence identity/similarity in pairwise comparisons with LIPR-10.1A (Rice et al., 2000) vary significantly and are equal to 20/43%, 36/55% and 60/77%, respectively, indicating that the primary structure does not affect the conformational (dis)similarity of the complexes.

In terms of conformation, the most conserved structural elements among the available structures of PR-10/*trans*-zeatin complexes are the β -sheet and loops L1 and L6 (Fernandes et al., 2013). On the other hand, the secondary structure elements with highest conformational variability are helix α 3 and loop L8, as well as the loops surrounding the E1 entrance, i.e. L3, L5, L7 and L9 (Fig. 9).

In particular, the C-terminal helix α 3 has a variable distance to the β 1 strand (Fig. 9a) as well as different degree of collapse (inward kinking) (Fig. 9b) into the protein hydrophobic core, in different PR-10/*trans*-zeatin complexes. These structural properties of α 3 seem to depend on the ligand distribution within the protein cavity. The α 3 helix of form **Z** of LIPR-10.1A has the longest distance to the β 1 strand because two of the ZEA ligands (ZEA2 and ZEA3) are accommodated between these structural elements. In the LIPR-10.2B/ZEA complex (2qim, Fernandes et al. (2008)) this distance is somewhat shorter since the (three) *trans*-zeatin molecules are located more centrally in a huge protein cavity

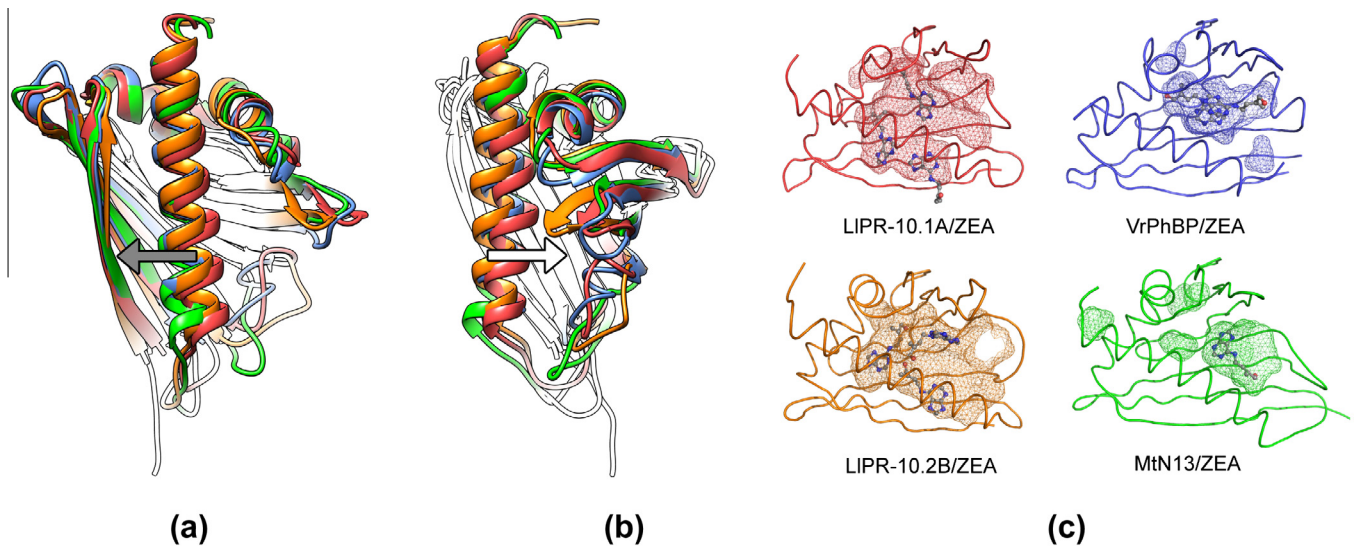


Fig. 9. (a) Superposition of the available PR-10/*trans*-zeatin complexes of MtN13 (4jhg; green), VrPhBP (2flh, chain A; blue), LIPR-10.2B (2qim; orange) and of the present model Z (red), emphasizing the different distances (gray arrow) of helix $\alpha 3$ to strand $\beta 1$. (b) The same superposition shown in a view illustrating the degree of collapse of helix $\alpha 3$ into the hydrophobic core of the protein (white arrow). (c) Ligand positions in the internal cavities of the PR-10/*trans*-zeatin complexes, colored as in (a) and (b).

(Fig. 9c), whereas in the cases of the VrPhBP and MtN13 proteins, which have a rather small cavity and where there is no functional E2 entrance (Sliwiak et al., 2013), helix $\alpha 3$ is even closer to the $\beta 1$ strand. The inward kinking of helix $\alpha 3$ into the hydrophobic core is the most prominent in the case of the LIPR-10.1A/ZEA complex, and the smallest in the LIPR-10.2B/ZEA complex where the big cargo has a central location in the cavity (Fig. 9c). These observations confirm the dominating effect of the $\alpha 3$ helix on the shape of the internal cavity of the PR-10 proteins.

The cavity volumes for the above complexes are 1979 (LIPR-10.1A/ZEA), 2434 (LIPR-10.2B/ZEA), 1080 (VrPhBP/ZEA) and 726 Å³ (MtN13/ZEA), with shapes as presented in Fig. 9c. We note that the two yellow lupine isoforms are examples of PR-10 proteins with type 2 cavity, whereas MtN13 and VrPhBP have type 1 cavity (Sliwiak et al., 2013).

With regard to *trans*-zeatin binding, neither the ligand position and orientation, nor even the stoichiometry is conserved (Fig. 9c). A multiple sequence alignment (Goujon et al., 2010) reveals that there are no conserved residues in the LIPR-10.1A sequence that would be constantly involved in ligand interactions, except Tyr82, which interacts, however, with different *trans*-zeatin atoms and via different interactions. Nevertheless, if only the LIPR-10.1A and 2B lupine isoforms are compared, it is apparent that they share a set of common residues (Phe5, Leu22, Ile115, Val117, Phe142) that form van der Waals interactions with the ligands (even if with their different fragments).

In each of the PR-10/ZEA complex structures available to-date, at least one *trans*-zeatin molecule was bound in a “fork-like” manner at the N6/N7 atoms by a Glu, Gln or Asp residue. In the present complex of LIPR-10.1A (Z), none of the five ligand molecules is bound in this way. Moreover, the N6 and N7 atoms of ZEA2 and ZEA3 are hydrogen bonded with one water molecule in each case, which mediates their hydrogen-bond interactions with the protein atoms.

4. Conclusions

Uncharacteristically, the LIPR-10.1A protein from yellow lupine exhibits no promiscuity in ligand binding, in contrast to other PR-10 proteins with large type 2 cavity, such as the yellow lupine LIPR-10.2B isoform or Bet v 1. The fact that LIPR-10.1A could be only complexed with *trans*-zeatin, suggests that it might be a

specific binder of this phytohormone. Based on the presented models of LIPR-10.1A in the free (F), unsaturated (U) and saturated (Z) forms, one can detect several structural changes that accompany *trans*-zeatin binding. They are visible mainly in (i) the conformation of helix $\alpha 3$ (unwinding-winding of the central two coils), (ii) the curvature of the loops (“fingertips”) surrounding the E1 cavity entrance (gripping of helix $\alpha 3$), (iii) the bulging out of the $\beta 5$ – $\beta 7$ sheet, and (iv) the orientation of the side chains lining the walls of the cavity. The helix formation effect (i) was additionally observed in solution by monitoring the CD spectra at increasing *trans*-zeatin concentration. It was found that at five-fold molar excess the ligand effect reaches a plateau and does not increase the helical content anymore. LIPR-10.1A is capable of accommodating three *trans*-zeatin molecules in its internal cavity. ZEA1 is held rather loosely in the largest hilum by sporadic contacts, whereas ZEA2 and ZEA3 have numerous hydrophobic contacts with the protein as well as hydrogen bonds with residues from the $\alpha 3$ and $\beta 1$ structural elements, which trigger the ordering of the $\alpha 3$ helix. The sequence of the binding events seems to begin with the docking of ZEA1 and ZEA2, which drives the main structural changes and prepares the ultimate binding site of ZEA3, which is bound most tightly, but only in the presence of ZEA1 and ZEA2. The structural adaptation of helix $\alpha 3$ seems to control and orchestrate the whole binding process: the unwound helix facilitates ligand access to the cavity, whereas the fully formed helix blocks the ligand escape routes. The three crystal structures of LIPR-10.1A presented in this work are the first complete set of the same PR-10 protein in different ligand-binding states to enable an in-depth discussion of the structural transformations of plant proteins with PR-10 fold upon binding of physiologically-relevant natural ligands.

Acknowledgments

This publication was supported by the Polish Ministry of Science and Higher Education, under the KNOW program.

References

- Afonine, P.V., Grosse-Kunstleve, R.W., Echols, N., Headd, J.J., Moriarty, N.W., Mustyakimov, M., Terwilliger, T.C., Urzhumtsev, A., Zwart, P.H., Adams, P.D., 2012. Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. D* 68, 352–367.

- Ames, B.D., Korman, T.P., Zhang, W., Smith, P., Vu, T., Tang, Y., Tsai, S.C., 2008. Crystal structure and functional analysis of tetracenomycin ARO/CYC: implications for cyclization specificity of aromatic polyketides. *PNAS* 105, 5349–5354.
- Berkner, H., Schweimer, K., Matecko, I., Rösch, P., 2008. Conformation, catalytic site, and enzymatic mechanism of the PR10 allergen-related enzyme norcochlorine synthase. *Biochem. J.* 413, 281–290.
- 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.
- Chen, V.B., Arendall, W.B., Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J., Murray, L.W., Richardson, J.S., Richardson, D.C., 2010. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D66*, 12–21.
- Chwastyk, M., Jaskolski, M., Cieplak, M., 2014. Structure-based thermodynamic and mechanical stability of plant PR-10 proteins with cavities. *FEBS J.* 281, 416–429.
- Cohen, G.H., 1997. ALIGN: a program to superimpose protein coordinates, accounting for insertions and deletions. *J. Appl. Crystallogr.* 30, 1160–1161.
- Emsley, P., Lohkamp, B., Scott, W., Cowtan, K., 2010. Features and development of coot. *Acta Crystallogr. D66*, 486–501.
- 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., Bujacz, A., Bujacz, G., Jelen, F., Jasinski, M., Kachlicki, P., Otlewski, J., Sikorski, M.M., Jaskolski, M., 2009. Cytokinin-induced structural adaptability of a *Lupinus luteus* PR-10 protein. *FEBS J.* 276, 1596–1609.
- Fernandes, H., Michalska, K., Sikorski, M.M., Jaskolski, M., 2013. Structural and functional aspects of PR-10 proteins. *FEBS J.* 280, 1169–1199.
- 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.
- Goujon, M., McWilliam, H., Li, W., Valentin, F., Squizzato, S., Paern, J., Lopez, R., 2010. A new bioinformatics analysis tools framework at EMBL–EBI. *Nucleic Acids Res.* 38, W695–W699.
- Greenfield, N.J., 2006. Using circular dichroism spectra to estimate protein secondary structure. *Nat. Protoc.* 1, 2876–2890.
- Hurlburt, B.K., Offermann, L.R., McBride, J.K., Majorek, K.A., Maleki, S.J., Chruszcz, M., 2013. Structure and function of the peanut panallergen Ara h 8. *J. Biol. Chem.* 288, 36890–36901.
- Hwang, I., Sheen, J., Müller, B., 2012. Cytokinin signaling networks. *Annu. Rev. Plant Biol.* 63, 353–380.
- Jaskolski, M., 2013. On propagation of errors. *Acta Crystallogr. D69*, 1865–1866.
- Kabsch, W., 2010. XDS. *Acta Crystallogr. D66*, 125–132.
- Kofler, S., Asam, C., Eckhard, U., Wallner, M., Ferreira, F., Brandstetter, H., 2012. Crystallographically mapped ligand binding differs in high and low IgE binding isoforms of birch pollen allergen bet v 1. *J. Mol. Biol.* 422, 109–123.
- Krissinel, E., Henrick, K., 2007. Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* 372, 774–797.
- 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. Crystallogr.* 40, 658–674.
- Merritt, E.A., 2012. To B or not to B: a question of resolution? *Acta Crystallogr. D68*, 468–477.
- Miyazono, K., Miyakawa, T., Sawano, Y., Kubota, K., Kang, H., Asano, A., Miyauchi, Y., Takahashi, M., Zhi, Y., Fujita, Y., Yoshida, T., Kodaira, K., Yamaguchi-Shinozaki, K., Tanokura, M., 2009. Structural basis of abscisic acid signalling. *Nature* 462, 609–615.
- Murshudov, G.N., Vagin, A.A., Dodson, E.J., 1997. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D53*, 240–255.
- Otwinowski, Z., Minor, W., 1997. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* 276, 307–326.
- 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 Crystallogr. D61*, 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.
- Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., Ferrin, T.E., 2004. UCSF Chimera – a visualization system for exploratory research and analysis. *J. Comput. Chem.* 25, 1605–1612.
- Ricci, A., Bertoletti, C., 2009. Urea derivatives on the move: cytokinin-like activity and adventitious rooting enhancement depend on chemical structure. *Plant Biol.* 11, 262–272.
- Rice, P., Longden, I., Bleasby, A., 2000. EMBOSS: the European molecular biology open software suite. *Trends Genet.* 16, 276–277.
- Ruszkowski, M., Szpotkowski, K., Sikorski, M.M., Jaskolski, M., 2013. The landscape of cytokinin binding by a plant nodulin. *Acta Crystallogr. D69*, 2365–2380.
- Ruszkowski, M., Sliwiak, J., Ciesielska, A., Barciszewski, J., Sikorski, M.M., Jaskolski, M., 2014. Specific binding of gibberellic acid by cytokinin-specific binding proteins: a new aspect of plant hormone-binding proteins with the PR-10 fold. *Acta Crystallogr. D70*, 2032–2041.
- Sliwiak, J., Dauter, Z., Jaskolski, M., 2013. Hyp-1 protein from St John's wort as a PR-10 protein. *BioTechnologia* 94, 47–50.
- 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.
- Whitmore, L., Wallace, B.A., 2004. DICHROWEB: an online server for protein secondary structure analyses from circular dichroism spectroscopic data. *Nucleic Acids Res.* 32, 668–673.
- Whitmore, L., Wallace, B.A., 2008. Protein secondary structure analyses from circular dichroism spectroscopy: methods and reference databases. *Biopolymers* 89, 392–400.

PUBLICATION II

Joanna Sliwiak,^a Mariusz Jaskolski,^{a,b*} Zbigniew Dauter,^c Airlie J. McCoy^d and Randy J. Read^{d*}

^aCenter for Biocrystallographic Research, Institute of Bioorganic Chemistry, Polish Academy of Sciences, Noskowskiego 12/14, 61-704 Poznan, Poland, ^bDepartment of Crystallography, Faculty of Chemistry, A. Mickiewicz University, Grunwaldzka 6, 60-780 Poznan, Poland, ^cSynchrotron Radiation Research Section, National Cancer Institute, Argonne National Laboratory, Argonne, IL 60439, USA, and ^dDepartment of Haematology, University of Cambridge, Wellcome Trust/MRC Building, Hills Road, Cambridge CB2 0XY, England

Correspondence e-mail: mariuszj@amu.edu.pl, rjr27@cam.ac.uk

Likelihood-based molecular-replacement solution for a highly pathological crystal with tetartohedral twinning and sevenfold translational noncrystallographic symmetry

Translational noncrystallographic symmetry (tNCS) is a pathology of protein crystals in which multiple copies of a molecule or assembly are found in similar orientations. Structure solution is problematic because this breaks the assumptions used in current likelihood-based methods. To cope with such cases, new likelihood approaches have been developed and implemented in *Phaser* to account for the statistical effects of tNCS in molecular replacement. Using these new approaches, it was possible to solve the crystal structure of a protein exhibiting an extreme form of this pathology with seven tetrameric assemblies arrayed along the *c* axis. To resolve space-group ambiguities caused by tetartohedral twinning, the structure was initially solved by placing 56 copies of the monomer in space group *P*1 and using the symmetry of the solution to define the true space group, *C*2. The resulting structure of Hyp-1, a pathogenesis-related class 10 (PR-10) protein from the medicinal herb St John's wort, reveals the binding modes of the fluorescent probe 8-anilino-1-naphthalene sulfonate (ANS), providing insight into the function of the protein in binding or storing hydrophobic ligands.

Received 1 October 2013

Accepted 6 November 2013

1. Introduction

Hyp-1 is a 165-residue pathogenesis-related class 10 (PR-10) protein from the medicinal herb St John's wort (*Hypericum perforatum*). PR-10 proteins are among the most mysterious plant proteins since no unique biological function can be attributed to them despite their abundance (Fernandes *et al.*, 2013). The mystery shrouding the function of PR-10 proteins is in contrast to their comprehensive structural characterization, which reveals an almost hollow molecular core surrounded by a seven-stranded antiparallel β -sheet gripped around a long α -helix (α 3) supported at the C-terminus by a fork of two shorter helices (Gajhede *et al.*, 1996; Biesiadka *et al.*, 2002). This characteristic fold, termed the PR-10 fold (or the Bet v 1 fold after birch pollen allergen, which was the first PR-10 protein to have its crystal structure solved) strongly suggests the binding/storage of hydrophobic ligands. Such a function would be compatible with signalling and/or regulation, which in plants involve small molecules of diverse structure called phytohormones (Santner & Estelle, 2009).

Fluorescent probes, such as 8-anilino-1-naphthalene sulfonate (ANS), can be used to study the ligand-binding function of PR-10 proteins in ANS displacement assays (ADAs). To facilitate the interpretation of the spectra, accurate structural information is needed and to this end we have crystallized Hyp-1 in complex with ANS. Hyp-1 has been postulated to

Table 1
Diffraction data statistics.

Values in parentheses are for the highest resolution shell.

Beamline	19ID, SER-CAT, APS	
Temperature (K)	100	
Space group	<i>P</i> 422	<i>C</i> 2
Unit-cell parameters		
<i>a</i> (Å)	103.42	146.21
<i>b</i> (Å)	103.42	146.12
<i>c</i> (Å)	298.50	298.35
β (°)	90	90.07
Wavelength (Å)	1.000	1.000
Resolution (Å)	30–2.43 (2.47–2.43)	30–2.43 (2.47–2.43)
Reflections, measured	496579	495931
Reflections, unique	61810	170447
Completeness (%)	99.8 (99.2)	72.7 (65.9)
$\langle I/\sigma(I) \rangle$	26.4 (2.6)	13.4 (1.5)
$R_{\text{merge}}^{\dagger}$ (%)	7.5 (75.8)	6.6 (69.1)
Multiplicity	8.0 (7.1)	2.9 (2.6)

$$\dagger R_{\text{merge}} = \frac{\sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle|}{\sum_{hkl} \sum_i I_i(hkl)}$$

catalyze the oxidative coupling of emodin to hypericin, the main pharmacological ingredient of St John's wort (Bais *et al.*, 2003), although this enzymatic activity has been questioned (Michalska *et al.*, 2010). In this context, the binding of ANS, which contains a large π -electron system similar to that of emodin, is of additional interest.

Structure solution by the method of molecular replacement (MR) turned out to be a daunting problem not only because of tetartohedral twinning, but primarily because the asymmetric unit was found to contain multiple copies of the protein molecule arranged with sevenfold noncrystallographic repetition along *c*. This bizarre structural architecture can be interpreted as a superstructure modulation. In crystals with modulated structures, the short-range translational order from one unit cell to the next is lost, but long-range order is restored by a periodic atomic modulation function (AMF; Lovelace *et al.*, 2013). In general the two periods (of the AMF and of the underlying lattice) can be incommensurate, in which case the superstructure has to be described in a higher-dimensional space (Lovelace *et al.*, 2008). However, if the modulation is commensurate (as found in this work), it is possible to describe the structure in an expanded unit cell. Superstructure modulation in direct space is manifested in the reciprocal lattice by strong main reflections (from the underlying lattice) and much weaker satellite reflections (from the AMF wave). While superstructure modulation is a well studied phenomenon in small-molecule crystallography, it has been less well studied in macromolecular crystallography. In solving this structure, it was sufficient to consider the structure to arise approximately from a sevenfold replication of the underlying unit cell, and not to be concerned about the details of the changes in orientation and translation described by the AMF. A subsequent publication will address the detailed interpretation of this structure in terms of commensurate modulation.

Note that the word 'modulation' is used here in two contexts. In real space, a superstructure modulation causes the atomic positions to vary systematically in different copies in a way that can be represented by a periodic function. In

reciprocal space, the repetition of similarly oriented copies causes a modulation of the diffraction intensities, which vary systematically in a way that can also be represented by a (different) periodic function.

2. The diffraction data set and initial attempts to solve the structure

Large single crystals of a Hyp-1–ANS complex were obtained by co-crystallization with an eightfold molar excess of the ligand. Strong blue fluorescence observed under a UV microscope confirmed the presence of ANS in the crystals. X-ray diffraction data extending to 2.4 Å resolution were collected on the SER-CAT beamline 19ID at the APS synchrotron and were processed with *HKL-2000* (Otwinowski & Minor, 1997). The initial merging of the data appeared to be satisfactory in space group *P*422, with an R_{merge} of 7.5% (Table 1). Solvent-content analysis indicated that between six and 12 protein molecules could be accommodated in the asymmetric unit of *P*422.

The diffraction images revealed a repetitive modulation of reflection intensities along the direction of c^* with a period of 7/2 (Fig. 1*a*), indicating a noncrystallographic translation of a molecular assembly along the longest cell dimension of the crystal, *c*. In the native Patterson (Fig. 1*b*), the peak corresponding to 2/7 of the *c* lattice translation was much stronger (72% of the origin peak height) than the peaks corresponding to 1/7 (18%) or 3/7 (35%) of the *c* axis. In the ultimate crystal structure (Fig. 1*c*), these features were shown to arise from an approximate sevenfold repetition of the unit cell along the *c* axis, where molecules separated by 2/7 of the unit cell are generally more similar in orientation than those separated by 1/7 of the unit cell.

Repeated attempts failed to solve the structure by molecular replacement using existing algorithms, even though an excellent model of the unliganded protein was available (Michalska *et al.*, 2010). We reasoned that the presence of translational noncrystallographic symmetry (tNCS) was violating assumptions in current approaches to molecular replacement, which implicitly assume that the diffraction data vary smoothly over reciprocal space instead of being highly modulated. This structure was therefore used as a test case for new likelihood-based methods taking explicit account of the statistical effects of tNCS.

3. Molecular-replacement likelihood function for tNCS

New likelihood functions that apply corrections for the presence of tNCS were implemented in *Phaser-2.5.4* (McCoy *et al.*, 2007). The tNCS is parameterized by the tNCS vector itself and resolution-dependent Luzzati *D* terms (Luzzati, 1952) that account for deviations in positions between equivalent atoms including the effects of small differences in orientation and small errors in the translation vector. This treatment allows multiple copies of an asymmetric unit substructure to be related by the same tNCS vector, as in this case, in which seven copies are related by approximately the

same translation vector. The parameters are used to generate expected intensity factors for each reflection that model the modulations observed in the data (Read *et al.*, 2013) and are refined against the Wilson distribution (Wilson, 1949) of the data.

3.1. Characterizing tNCS prior to molecular replacement

The structure-factor contributions from molecules related by tNCS are correlated, with similar amplitudes governed by their similar orientations and with relative phase shifts dependent on the translation vector (Read *et al.*, 2013). The relative phase shifts create interference effects that modulate the covariances between structure-factor contributions from tNCS-related copies and, consequently, the variance for the total structure factor, thus altering the expected intensities in different parts of reciprocal space. The strength of the modulation is determined by the degree to which the

structure-factor contributions are correlated, which in turn is determined by how precisely the conformations and orientations of the tNCS-related molecules or molecular assemblies are preserved. When the multiplicity of the tNCS is high and the orientational differences are effectively random, as for our Hyp-1 crystal, small differences in orientation and relative translation between tNCS-related copies are approximated well by Luzzati D parameters (Luzzati, 1952) describing overall random conformational differences among the molecules, ignoring the small directional dependence of the modulation effects introduced by any rotational differences (Read *et al.*, 2013). Although we anticipate that the signal in a molecular-replacement search would be stronger if the deviations in the orientations of the tNCS-related copies and in the exact translation vectors relating successive copies could be modelled in advance, we have not yet developed an algorithm that can model such deviations for more than two copies in advance of structure solution.

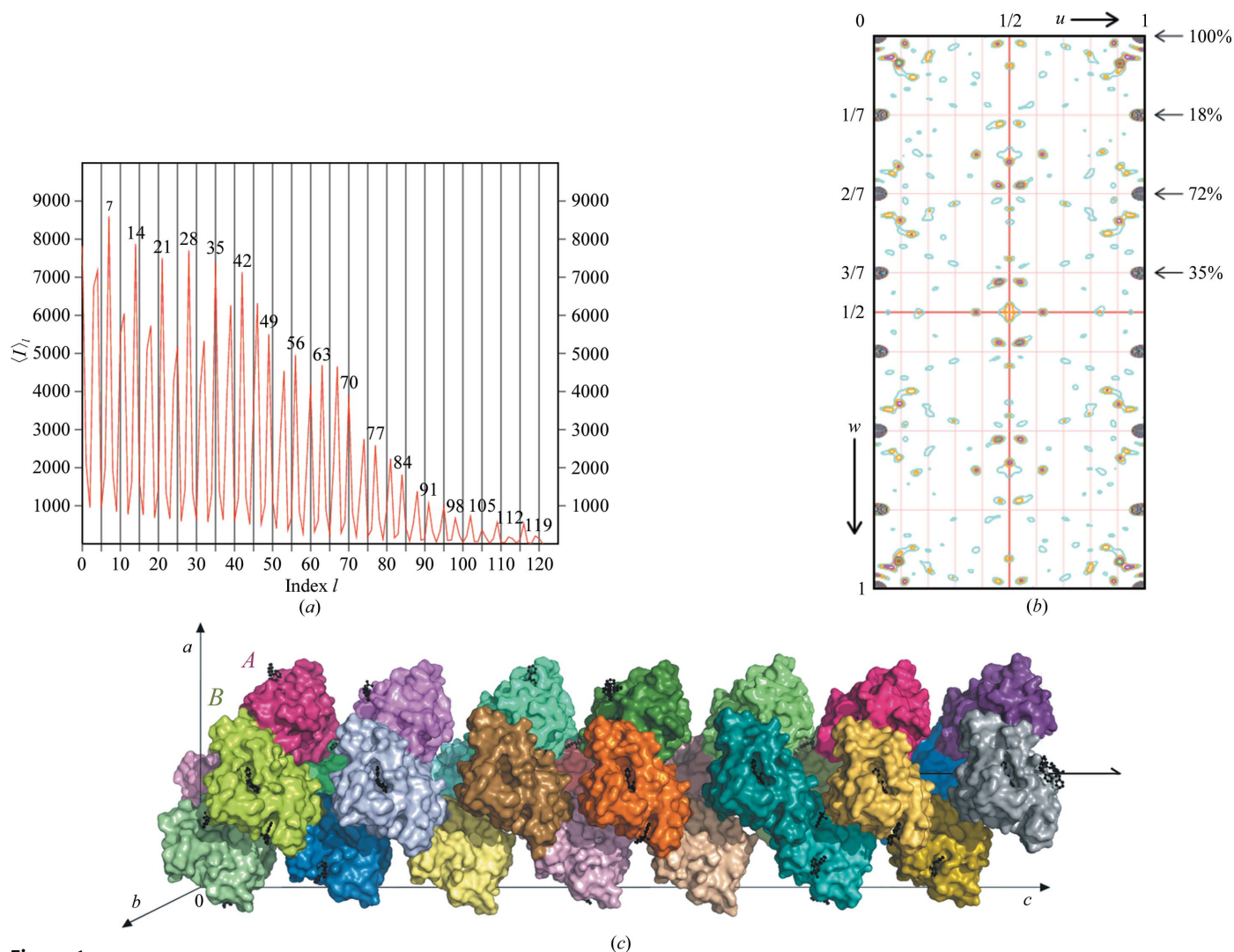


Figure 1

Translational noncrystallographic symmetry in a Hyp-1-ANS crystal. (a) Averaged reflection intensities in layers of constant l index. The pattern of modulation of the intensities, with peaks separated by $7/2$ along c^* , is striking. (b) Patterson map $v = 0$ section, showing the repetitive peaks (with peak height relative to the origin) along $00w$. (c) The 28 independent Hyp-1 molecules forming the asymmetric unit of the $C2$ crystal packing, arranged in a dimeric pattern with a sevenfold repeat around a noncrystallographic 2_1 screw (indicated) along the crystallographic c direction. Dimer AB is labelled.

3.2. tNCS correction in molecular replacement

3.2.1. Covariance elements for true structure factors. To introduce the notation needed for the application to molecular replacement, we start by briefly reviewing the effect of tNCS on intensity distributions (Read *et al.*, 2013). For simplicity, in the following we will ignore the effects of measurement errors, but note that these are introduced into the likelihood targets by incrementing the variances in these targets (McCoy *et al.*, 2007).

The total true structure factor is defined as the sum of contributions from components related by crystallographic (index k below) and noncrystallographic (index m) symmetry (NCS),

$$\mathbf{F} = \sum_{k=1}^{N_{\text{sym}}} \sum_{m=1}^{N_{\text{ncs}}} \mathbf{F}_{km},$$

$$\mathbf{F}_{km} = \sum_{j=1}^N f_{jm} \exp(2\pi i \mathbf{h} \cdot \mathbf{x}_{jkm}), \quad (1)$$

where

$$\begin{aligned} \mathbf{x}_{jkm} &= \mathbf{T}_k[(\mathbf{x}_j + {}_F\delta_{jm}) + {}_F\mathbf{v}_m] + \mathbf{t}_k \\ &= \mathbf{T}_k(\mathbf{x}_j + {}_F\delta_{jm}) + (\mathbf{T}_k {}_F\mathbf{v}_m + \mathbf{t}_k). \end{aligned} \quad (2)$$

This expresses the idea that all of the tNCS-related copies of a component (with coordinates \mathbf{x}_{jkm}) are considered to be derived from a canonical (average) copy centred on the origin (with coordinates \mathbf{x}_j for unique atom j) by a combination of rigid-body translations (translation vector ${}_F\mathbf{v}_m$ for NCS copy m) with perturbations of both coordinates (perturbation vector ${}_F\delta_{jm}$) and B factors (expressed as differences in the scattering factors f_{jm} for different NCS-related copies). The number of atoms in one copy of the component is given by N . In (2), the crystallographic symmetry operator k is expressed as a rotation, \mathbf{T}_k , and a translation, \mathbf{t}_k . The subscripted prefix F indicates a term relating to a component of the true structure factor \mathbf{F} , to distinguish it from terms relating to the calculated structure factor \mathbf{G} introduced below.

The expected intensity for a reflection is obtained by adding up all of the covariance elements relating contributions from different components in the unit cell, which are significant for components related by tNCS. The derivation of the expected intensity expression in (3), given in detail in our earlier publication (Read *et al.*, 2013), is similar to that shown below for the expected values of calculated intensities in (4)–(6),

$$\langle F^2 \rangle = \varepsilon \Sigma_N \left[1 + 2 \sum_{k=1}^{N_{\text{sym}}} \sum_{m=1}^{N_{\text{ncs}}-1} \sum_{n=m+1}^{N_{\text{ncs}}} \frac{{}_{FF}\rho_{mn}(\Sigma_{Fm}\Sigma_{Fn})^{1/2}}{\Sigma_N} \times \cos(2\pi \mathbf{h} \cdot {}_{FF}\mathbf{v}_{kkmn}) \right], \quad (3)$$

where ε is the expected intensity factor arising from crystallographic symmetry, Σ_N is the scattering power of the unit-cell contents, ${}_{FF}\rho_{mn}$ is the correlation between the tNCS-related structure-factor contributions from components m and n of the crystal on the same origin, *i.e.* before tNCS translations have been applied (reduced from unity by any perturbations of coordinates or scattering factors), Σ_{Fm} is the scattering

power of one copy of component m and ${}_{FF}\mathbf{v}_{kkmn}$ is the translation vector relating the k th symmetry copies of components m and n , analogous to ${}_{GG}\mathbf{v}_{kkmn}$ relating components of the model in (5) below. (3) lacks the G -function term (Rossmann & Blow, 1962) of the expression derived earlier [equation (14) in Read *et al.*, 2013] because the tNCS-related copies are treated as being in the same orientation. In the notation used here, the subscripted prefix FF refers to terms relating the contributions of two components of the true structure factor \mathbf{F} ; below, the subscripted prefix GG will be used for terms relating two components of the calculated structure factor \mathbf{G} and the subscripted prefix FG will be used for terms relating one component of \mathbf{F} to a component of \mathbf{G} .

3.2.2. Covariance elements for calculated structure factors. In deriving a likelihood target for tNCS-corrected molecular replacement, the additional covariances relevant to calculated structure factors must also be introduced, including both covariances between tNCS-related contributions to the calculated structure factors and cross-terms between contributions to both the true and calculated structure factors. If it is assumed that the tNCS operations are correctly modelled, then the total calculated structure factors will be governed by modulations similar in size to those of the true structure factors. The same modulations will also apply to terms in the calculation of variances describing the differences between the true and calculated structure factors. Here, we make the approximation that tNCS-related molecules in the model are in an identical orientation and share the same conformation and scattering factors.

As in the case of the true structure factor \mathbf{F} , the calculated structure factor \mathbf{G} can be described as the sum over both crystallographic and noncrystallographic symmetry of the copies of contributions from individual models, shown in (4). Note that, without loss of generality, the model and the true structure can be considered to contain the same N atoms in each copy of the unique structural motif; atoms present in only one of them can be assigned a scattering factor of zero in the other. The positions of these atoms, denoted \mathbf{x} in the true structure and \mathbf{y} in the model, are related by random coordinate errors that will be introduced explicitly later,

$$\begin{aligned} \mathbf{G} &= \sum_{k=1}^{N_{\text{sym}}} \sum_{m=1}^{N_{\text{ncs}}} \mathbf{G}_{km}, \\ \mathbf{G}_{km} &= \sum_{j=1}^N g_j \exp(2\pi i \mathbf{h} \cdot \mathbf{y}_{jkm}), \quad \text{where} \\ \mathbf{y}_{jkm} &= \mathbf{T}_k(\mathbf{y}_j + {}_G\mathbf{v}_m) + \mathbf{t}_k \\ &= \mathbf{T}_k\mathbf{y}_j + (\mathbf{T}_k {}_G\mathbf{v}_m + \mathbf{t}_k). \end{aligned} \quad (4)$$

As for (1) and (2) describing the true structure, the coordinates in the model (coordinates \mathbf{y}_{jkm} for the copy generated by a combination of symmetry operation k and NCS operation m) are represented in terms of those from a canonical copy (coordinates \mathbf{y}_j) of the molecule centred on the origin, translating that copy by a vector ${}_G\mathbf{v}_m$ for NCS copy m ; the major difference from the treatment for the true structure is the lack of the terms describing perturbations of coordinates and

scattering factors between the copies. For convenience, we can take the canonical copy to be in the same orientation as the copy with $k = m = 1$, so that $\mathbf{y}_j = \mathbf{y}_{j1} - {}_G\mathbf{v}_1$. As for the case of the true structure factor, \mathbf{F} , we will only consider the covariances between NCS-related molecules in similar orientations which are assumed to be assigned to the same asymmetric unit. The interesting covariances are those between copies related by tNCS ($m \neq n$ and $k = l$). We can neglect covariances between symmetry-related contributions ($k \neq l$) because these will only be nonzero when the symmetry rotation is parallel to the diffraction vector, and the effect of these will be captured simply by introducing the usual expected intensity factor, ε .

$$\begin{aligned} \langle \mathbf{G}_{km} \mathbf{G}_{kn}^* \rangle &\simeq \sum_{j=1}^N \langle g_j^2 \exp[2\pi i \mathbf{h} \cdot (\mathbf{y}_{jkm} - \mathbf{y}_{jkn})] \rangle \\ &= \sum_{j=1}^N \langle g_j^2 \exp(2\pi i \mathbf{h} \cdot {}_{GG}\mathbf{v}_{kkmn}) \rangle \\ &= \Sigma_G \exp(2\pi i \mathbf{h} \cdot {}_{GG}\mathbf{v}_{kkmn}), \text{ where} \\ {}_{GG}\mathbf{v}_{kkmn} &= \mathbf{T}_k({}_G\mathbf{v}_m - {}_G\mathbf{v}_n). \end{aligned} \quad (5)$$

As discussed previously (Read *et al.*, 2013), terms involving common atoms will dominate, so cross-terms relating different atoms in the NCS copies are ignored in (5). The phase-shift term expressed by the exponential is the same for all atoms, so the sum of squared scattering factors can be factored out as Σ_G , the scattering power of one copy of the tNCS-related component in the asymmetric unit.

The expected calculated intensity is obtained, as for the true intensity, by summing all of the covariance elements,

$$\langle G^2 \rangle = \varepsilon \Sigma_P \left[1 + 2 \sum_{k=1}^{N_{\text{sym}}} \sum_{m=1}^{N_{\text{ncs}}-1} \sum_{n=m+1}^{N_{\text{ncs}}} \frac{\Sigma_G}{\Sigma_P} \cos(2\pi \mathbf{h} \cdot {}_{GG}\mathbf{v}_{kkmn}) \right]. \quad (6)$$

The diagonal elements of the covariance matrix, for which $m = n$, are summed in (6) to give Σ_P , the total scattering power of the model. As noted above, the expected intensity factor ε accounts for correlations between symmetry-related contributions. Off-diagonal elements of the covariance matrix are paired, and their imaginary components cancel to leave only the cosine term from the phase-shift exponential in (5). The term in the square brackets shows how the overall average intensity, $\varepsilon \Sigma_P$, is modulated by the presence of tNCS.

3.2.3. Covariance elements relating contributions to true and calculated structure factors. The covariance elements relating the contributions to the true and calculated structure factors take the following form:

$$\langle \mathbf{F}_{km} \mathbf{G}_{kn}^* \rangle \simeq \sum_{j=1}^N \langle f_{jm} g_j \exp[2\pi i \mathbf{h} \cdot (\mathbf{x}_{jkm} - \mathbf{y}_{jkn})] \rangle. \quad (7)$$

In (7) we assume, as in (5) above, that terms relating common atoms dominate so that there is only a single sum over the unique atoms in a component. We assume that the orientation of the model is correct, on the basis that it will be correct for some orientation in the rotation search, and this orientation should show optimal agreement with the data in the likelihood function. Using the definitions of \mathbf{F}_{km} and

\mathbf{G}_{km} given above, and assuming that the orientations of tNCS-related components in the crystal and the model are identical (with any actual deviations to be modelled by Luzzati D factors), the dot product inside the exponential can be expanded,

$$\begin{aligned} \mathbf{h} \cdot (\mathbf{x}_{jkm} - \mathbf{y}_{jkn}) &= \mathbf{h} \cdot [\mathbf{T}_k(\mathbf{x}_j + {}_F\delta_{jm}) + (\mathbf{T}_k {}_F\mathbf{v}_m + \mathbf{t}_k) \\ &\quad - \mathbf{T}_k \mathbf{y}_j - (\mathbf{T}_k {}_G\mathbf{v}_n + \mathbf{t}_k)]. \end{aligned} \quad (8)$$

We can simplify this by expressing the coordinates of the model in terms of the true positions of the corresponding atoms in the canonical component of the crystal structure,

$$\mathbf{y}_j = \mathbf{x}_j + {}_{FG}\delta_j, \quad (9)$$

where the random error in the position of atom j is given by ${}_{FG}\delta_j$,

$$\begin{aligned} \mathbf{h} \cdot (\mathbf{x}_{jkm} - \mathbf{y}_{jkn}) &= \mathbf{h} \cdot [\mathbf{T}_k({}_F\mathbf{v}_m - {}_G\mathbf{v}_n) + \mathbf{T}_k({}_F\delta_{jm} - {}_{FG}\delta_j)] \\ &= \mathbf{h} \cdot {}_{FG}\mathbf{v}_{kkmn} + \mathbf{h} \cdot {}_{FG}\delta_{jkkmn}, \text{ where} \\ {}_{FG}\mathbf{v}_{kkmn} &= \mathbf{T}_k({}_F\mathbf{v}_m - {}_G\mathbf{v}_n) \\ {}_{FG}\delta_{jkkmn} &= \mathbf{T}_k({}_F\delta_{jm} - {}_{FG}\delta_j). \end{aligned} \quad (10)$$

In (10), ${}_{FG}\mathbf{v}_{kkmn}$ is the translation vector relating the k th symmetry copies of component m in the crystal and component n in the model and ${}_{FG}\delta_{jkkmn}$ is the random coordinate error affecting atom j in these two components. Substituting (10) into (7) gives (11),

$$\begin{aligned} \langle \mathbf{F}_{km} \mathbf{G}_{kn}^* \rangle &= \sum_{j=1}^N \langle f_{jm} g_j \exp(2\pi i \mathbf{h} \cdot {}_{FG}\mathbf{v}_{kkmn}) \exp(2\pi i \mathbf{h} \cdot {}_{FG}\delta_{jkkmn}) \rangle \\ &= {}_{FG}\rho_{mn} (\Sigma_{Fm} \Sigma_G)^{1/2} \exp(2\pi i \mathbf{h} \cdot {}_{FG}\mathbf{v}_{kkmn}), \text{ where} \\ {}_{FG}\rho_{mn} (\Sigma_{Fm} \Sigma_G)^{1/2} &= \left\langle \sum_{j=1}^N f_{jm} g_j \exp(2\pi i \mathbf{h} \cdot {}_{FG}\delta_{jkkmn}) \right\rangle. \end{aligned} \quad (11)$$

In this equation, the phase-shift term arising from the difference in positions of the component copies, ${}_{FG}\mathbf{v}_{kkmn}$, is the same for all atoms, so it has been factored out. ${}_{FG}\rho_{mn}$ is the correlation between the structure-factor contributions of component m in the crystal and component n in the model placed on the same origin (*i.e.* after removing the effect of their relative translation), which is reduced from unity by differences between the coordinates and scattering factors. Note that it can be interpreted as equivalent to a σ_A value, as discussed in the context of molecular-replacement ensemble models [equations (14) and (15) of Read, 2001], so that its value can be estimated in advance of structure solution from the expected r.m.s. error of the model (estimated in turn from the sequence identity and size of the model; Oeffner *et al.*, 2013) and the completeness of the model.

3.2.4. Conditional probability distribution given a model.

The conditional probability of the true structure factor given a model is obtained most easily by starting from the joint distribution of all of the NCS-related contributions to the true and calculated structure factors. This is similar to the strategy used to derive likelihood functions for molecular replacement (Read, 2001) and experimental phasing (Read, 2003). A large covariance matrix, Σ , is partitioned into separate matrices

for the contributions to the true structure factor (Σ_{11}), the contributions to the calculated structure factor (Σ_{22}) and the covariances between them (Σ_{12} and Σ_{21} , related by a Hermitian transpose). The individual submatrices have a block-diagonal structure, with blocks reflecting the correlations among copies related by translational NCS and zeroes for the symmetry-related copies that (after accounting for the crystallographic expected intensity factor ε) can be considered uncorrelated.

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}, \quad (12)$$

$$\Sigma_{11} = \begin{pmatrix} {}_1\Sigma_{11} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & {}_2\Sigma_{11} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & {}_{N_{\text{sym}}}\Sigma_{11} \end{pmatrix}, \text{ where}$$

$${}_k\Sigma_{11} = \begin{pmatrix} \langle \mathbf{F}_{k1} \mathbf{F}_{k1}^* \rangle & \cdots & \langle \mathbf{F}_{k1} \mathbf{F}_{kN_{\text{ncs}}}^* \rangle \\ \vdots & \ddots & \vdots \\ \langle \mathbf{F}_{k1} \mathbf{F}_{kN_{\text{ncs}}}^* \rangle^* & \cdots & \langle \mathbf{F}_{kN_{\text{ncs}}} \mathbf{F}_{kN_{\text{ncs}}}^* \rangle \end{pmatrix}. \quad (13)$$

$$\Sigma_{22} = \begin{pmatrix} {}_1\Sigma_{22} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & {}_2\Sigma_{22} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & {}_{N_{\text{sym}}}\Sigma_{22} \end{pmatrix}, \text{ where}$$

$${}_k\Sigma_{22} = \begin{pmatrix} \langle \mathbf{G}_{k1} \mathbf{G}_{k1}^* \rangle & \cdots & \langle \mathbf{G}_{k1} \mathbf{G}_{kN_{\text{ncs}}}^* \rangle \\ \vdots & \ddots & \vdots \\ \langle \mathbf{G}_{k1} \mathbf{G}_{kN_{\text{ncs}}}^* \rangle^* & \cdots & \langle \mathbf{G}_{kN_{\text{ncs}}} \mathbf{G}_{kN_{\text{ncs}}}^* \rangle \end{pmatrix}. \quad (14)$$

$$\Sigma_{12} = \begin{pmatrix} {}_1\Sigma_{12} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & {}_2\Sigma_{12} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & {}_{N_{\text{sym}}}\Sigma_{12} \end{pmatrix}, \text{ where}$$

$${}_k\Sigma_{12} = \begin{pmatrix} \langle \mathbf{F}_{k1} \mathbf{G}_{k1}^* \rangle & \cdots & \langle \mathbf{F}_{k1} \mathbf{G}_{kN_{\text{ncs}}}^* \rangle \\ \vdots & \ddots & \vdots \\ \langle \mathbf{F}_{kN_{\text{ncs}}} \mathbf{G}_{k1}^* \rangle & \cdots & \langle \mathbf{F}_{kN_{\text{ncs}}} \mathbf{G}_{kN_{\text{ncs}}}^* \rangle \end{pmatrix}. \quad (15)$$

Because the covariance matrix has Hermitian symmetry, $\Sigma_{21} = \Sigma_{12}^H$.

The matrix manipulations used to derive the conditional distribution require inverting the Σ_{22} submatrix and then computing products with the off-diagonal submatrices. Note that the inverse of a block-diagonal matrix is itself a block-diagonal matrix, in which the individual blocks (denoted by a subscripted prefix) are the matrix inverses of the original blocks.

$$\Sigma_{22}^{-1} = \begin{pmatrix} {}_1\Sigma_{22}^{-1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & {}_2\Sigma_{22}^{-1} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & {}_{N_{\text{sym}}}\Sigma_{22}^{-1} \end{pmatrix}. \quad (16)$$

In addition, the product of two block-diagonal matrices is itself a block-diagonal matrix, in which the individual blocks are the products of the corresponding blocks from the original matrices,

$$\Sigma_{12}\Sigma_{22}^{-1} = \begin{pmatrix} {}_1\Sigma_{12}{}_1\Sigma_{22}^{-1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & {}_2\Sigma_{12}{}_2\Sigma_{22}^{-1} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & {}_{N_{\text{sym}}}\Sigma_{12}{}_N\Sigma_{22}^{-1} \end{pmatrix}. \quad (17)$$

Thus, all of the manipulations used to derive the conditional probability distributions involve operations carried out only on the blocks corresponding to the NCS-related contributions to a particular symmetry copy in the crystal and the model.

3.2.5. Conditional probability when the rotational component of the tNCS operator is zero. The terms in the submatrix block ${}_k\Sigma_{12}$, i.e. $\langle \mathbf{F}_{km} \mathbf{G}_{kn}^* \rangle$, can be related to the terms in the submatrix block ${}_k\Sigma_{22}$, i.e. $\langle \mathbf{G}_{km} \mathbf{G}_{kn}^* \rangle$, if we make some reasonable assumptions. The guiding principle is that if we had a clear idea of the systematic differences between the model and the true structure then we would have changed the model accordingly, so any differences that remain should be random. If the NCS translations in the true structure and the model were identical, then the exponential phase-shift terms in (5) and (11) would be identical, giving

$$\langle \mathbf{F}_{km} \mathbf{G}_{kn}^* \rangle = {}_{FG}\rho_{mn} \left(\frac{\Sigma_{Fm}}{\Sigma_G} \right)^{1/2} \langle \mathbf{G}_{km} \mathbf{G}_{kn}^* \rangle. \quad (18)$$

Considering the interpretation of ${}_{FG}\rho_{mn}$ as a σ_A value, as discussed in §3.2.3, and noting the definition of σ_A in terms of model completeness and the Luzzati (1952) D factor (Srinivasan & Ramachandran, 1965), where

$$\sigma_A = D \left(\frac{\Sigma_P}{\Sigma_N} \right)^{1/2}, \quad (19)$$

(in which Σ_P plays the same role as Σ_G , and Σ_N plays the same role as Σ_{Fm}), we obtain a simple relationship between the terms in the submatrix block,

$$\langle \mathbf{F}_{km} \mathbf{G}_{kn}^* \rangle = D \langle \mathbf{G}_{km} \mathbf{G}_{kn}^* \rangle. \quad (20)$$

If we assume that the tNCS translations in the true structure and the model differ instead by a random error that is independent of the model errors, then the correlation between the true and calculated structure-factor contributions will be somewhat lower, which can be modelled by assuming a slightly larger r.m.s. error in computing the values of D as a function of resolution. Note that the effective r.m.s. errors are refined as part of the final step of molecular replacement in *Phaser*.

The same errors should apply to different components, so we can approximate the whole off-diagonal submatrix blocks as

$${}_k \Sigma_{12} = {}_k \Sigma_{21} = D {}_k \Sigma_{22}, \quad (21)$$

so that

$$\Sigma_{12} \Sigma_{22}^{-1} = D \mathbf{I}, \quad (22)$$

where \mathbf{I} is an identity matrix.

With these results in hand, standard manipulations can be applied to obtain the expected values of the symmetry- and NCS-related contributions to the true structure factor, given the corresponding contributions from the model,

$$\left\langle \begin{pmatrix} \mathbf{F}_{11} \\ \mathbf{F}_{12} \\ \vdots \\ \mathbf{F}_{N_{\text{sym}} N_{\text{ncs}}} \end{pmatrix} \right\rangle = \Sigma_{12} \Sigma_{22}^{-1} \begin{pmatrix} \mathbf{G}_{11} \\ \mathbf{G}_{12} \\ \vdots \\ \mathbf{G}_{N_{\text{sym}} N_{\text{ncs}}} \end{pmatrix} = D \begin{pmatrix} \mathbf{G}_{11} \\ \mathbf{G}_{12} \\ \vdots \\ \mathbf{G}_{N_{\text{sym}} N_{\text{ncs}}} \end{pmatrix}. \quad (23)$$

In words, the expected values of the various contributions \mathbf{F}_{km} to the total structure factor are simply the calculated contributions \mathbf{G}_{km} multiplied by D . The covariance matrix expressing the uncertainties in those expected values is

$$\Sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21} = \Sigma_{11} - D^2 \Sigma_{22}. \quad (24)$$

For the probability distribution of the total true structure factor, the variance is given by the sum of the elements of this updated covariance matrix, and the expected value is simply D times the total calculated structure factor. For acentric and centric reflections, the structure-factor probability distributions are thus given by

$$p_a(\mathbf{F}; \mathbf{G}) = \frac{1}{\pi \varepsilon \sigma_{\Delta}^2} \exp\left(-\frac{|\mathbf{F} - D\mathbf{G}|^2}{\varepsilon \sigma_{\Delta}^2}\right) \text{ and} \\ p_c(\mathbf{F}; \mathbf{G}) = \frac{1}{(2\pi \varepsilon \sigma_{\Delta}^2)^{1/2}} \exp\left(-\frac{|\mathbf{F} - D\mathbf{G}|^2}{2\varepsilon \sigma_{\Delta}^2}\right), \text{ where} \\ \sigma_{\Delta}^2 = (\langle F^2 \rangle - D^2 \langle G^2 \rangle) / \varepsilon \\ = \Sigma_N \left[1 + 2 \sum_{k=1}^{N_{\text{sym}}} \sum_{m=1}^{N_{\text{ncs}}-1} \sum_{n=m+1}^{N_{\text{ncs}}} \frac{FF \rho_{mn} (\Sigma_{Fm} \Sigma_{Fn})^{1/2}}{\Sigma_N} \right. \\ \left. \times \cos(2\pi \mathbf{h} \cdot {}_{FF} \mathbf{v}_{kkmn}) \right] \\ - D^2 \Sigma_P \left[1 + 2 \sum_{k=1}^{N_{\text{sym}}} \sum_{m=1}^{N_{\text{ncs}}-1} \sum_{n=m+1}^{N_{\text{ncs}}} \frac{\Sigma_G}{\Sigma_P} \right. \\ \left. \times \cos(2\pi \mathbf{h} \cdot {}_{GG} \mathbf{v}_{kkmn}) \right]. \quad (25)$$

In the general expression for σ_{Δ}^2 , it would be possible for one of the terms to be more highly modulated than the other. If care were not taken with the parameterization or with constraining the relative values of different terms (especially D), then this variance term could become negative. In practice, the modulation factors applied to the true and calculated intensities can often be assumed to be equivalent.

We will consider elsewhere the effects of modelling the rotational differences when there are only two tNCS-related

copies and the approximations inherent in the treatment presented here are poorly satisfied.

4. Hyp-1 tNCS-corrected molecular replacement

4.1. Attempts in *P422*-type symmetry

Molecular-replacement searches were carried out in *Phaser*-2.5.4, which included the likelihood functions able to account for the intensity modulations owing to translational NCS described above. Refinement of the tNCS operators relating pairs of molecules in space group *P422* gave an optimal translation vector of $(-0.004, -0.004, 0.285)$. (Note that the statistical effects of the tNCS operators depend only on the point group, but not on the particular space group.) Searches were carried out in all primitive space groups with 422 point-group symmetry, looking for seven copies related by tNCS. Using Hyp-1 as a model (Michalska *et al.*, 2010), multiple non-equivalent solutions with high signal to noise were found for space group *P4₁22*, showing similar but non-identical packing. However, space group *P4₁22* is ruled out by the presence of strong $00l$ reflections where the index l is not a multiple of 4. This fact, the existence of multiple incompatible solutions and the failure of the model to refine to an R factor better than 48% all suggested that the crystal was pseudo-symmetric, with the true symmetry being lower than *P422*. However, the excellent merging statistics in *P422* suggest that if the crystal is pseudo-symmetric it is also twinned. In agreement with this, the L test (Padilla & Yeates, 2003) suggested the presence of twinning; when reflections offset by multiples of 2 in h and k and multiples of 7 in l were used for the L test, the values $\langle L \rangle = 0.458$ and $\langle L^2 \rangle = 0.288$ were obtained. Pseudo-symmetry and twinning are commonly found in conjunction with one another (Lebedev *et al.*, 2006), and the presence of pseudo-symmetry would explain why the intensity distributions are perturbed less than one would otherwise expect for perfect twinning, where $\langle L \rangle = 3/8$ and $\langle L^2 \rangle = 1/5$, compared with $\langle L \rangle = 1/2$ and $\langle L^2 \rangle = 1/3$ for untwinned data.

4.2. Structure solution in space group *P1*

To identify the true symmetry, the diffraction data were expanded to *P1* and molecular replacement was attempted looking for 56 copies of Hyp-1. It can be difficult to resolve cases of pseudo-symmetry because if a perfectly symmetric solution is generated the symmetry has to be broken in some way, but the symmetric solution is balanced between different ways in which the symmetry can be broken. To avoid this trap, the search in *P1* was carried out in a way designed to avoid perfect symmetry, particularly the sevenfold translational pseudo-symmetry. A search for the first molecules in *P1* was carried out by assuming that the second through seventh molecules would be generated from the first by successive applications of the translation vector $(-0.004, -0.004, 0.285)$, as revealed by refinement of the tNCS operators in the 422 point-group symmetry (see above). After rigid-body refinement of the top solution, seven additional copies of this

assembly of seven molecules were added to yield a solution with 56 copies of Hyp-1 in the unit cell.

4.3. True space group identified as C2

Rigid-body refinement of the solution with 56 copies of the protein molecule in the *P1* unit cell was carried out using *phenix.refine* (Afonine *et al.*, 2012). To determine whether the molecular-replacement solution obeyed higher symmetry than *P1*, the calculated structure factors were examined for evidence of symmetry using *POINTLESS* (Evans, 2006), which looks for agreement between structure factors related

by potential symmetry operators of the lattice. Only one of the diagonal dyads of the initial *P422* space group ([110] direction of the tetragonal lattice) gave good agreement between related structure factors. This twofold operator corresponds to the unique *y* direction of space group *C2*, following the reindexing operation ($h + k, k - h, l$).

Accordingly, the diffraction data were reprocessed in the correct *C2* symmetry, with the results presented in Table 1. Unfortunately, the data-collection strategy had been selected for tetragonal symmetry, and instead of covering the unique 90° of rotation (between directions parallel and perpendicular to the monoclinic twofold axis) necessary for completeness, the same (*i.e.* symmetry-equivalent) 45° region of reciprocal space was covered twice. This led to a completeness of only ~73% in the genuine monoclinic symmetry. Since the R_{merge} value for *P422* (7.5%) was only less than 1% higher than that for *C2* (6.6%), with much higher multiplicity, it was decided to exploit this effect of the crystal twinning and to use in all subsequent calculations a data set expanded from *P422* to *C2* symmetry. This data set is almost fully complete and has the same statistical characteristics as presented in the first column of Table 1. Since the intensities conform to 422 symmetry, they correspond to a pseudo-tetartohedrally twinned crystal. The twinning of the monoclinic data set thus obtained is perfect, although in the real crystal it might have been only nearly perfect.

4.4. Structure solution in space group C2

The *C2* data were used to solve the structure by molecular replacement again, searching for four copies of the set of seven protein molecules found in the first step of the *P1* structure solution. This yielded two clear solutions with identical likelihood scores. Although the two solutions were not crystallographically equivalent, they were related by a fourfold rotation corresponding to one of the tetartohedral twin operators for *C2*. Rigid-body refinement of the 28 copies of the protein molecule in the *C2* solution confirmed that this solution does not obey any higher symmetry, though it is pseudo-symmetric with pseudo-tetragonal symmetry. The fact that the data could be merged well in point group 422 indicates that the additional apparent symmetry arose from twinning (Lebedev *et al.*, 2006).

5. Refinement of the structure

Before the atomic coordinate refinement commenced, data were selected for R_{free} tests using *SHELXPRO* (Sheldrick, 2008) within narrow shells of resolution in order to guarantee the inclusion of NCS-related reflections. The structure was refined in *REFMAC5* (Murshudov *et al.*, 2011) with intensity-based twin detection/refinement and jelly-body refinement. As expected from the molecular replacement and the treatment of the intensity data, four twin domains were found with operators corresponding to the twofold axes of the tetragonal supersymmetry. Upon refinement, all of the twin fractions converged at about 0.25. Application of loose NCS restraints

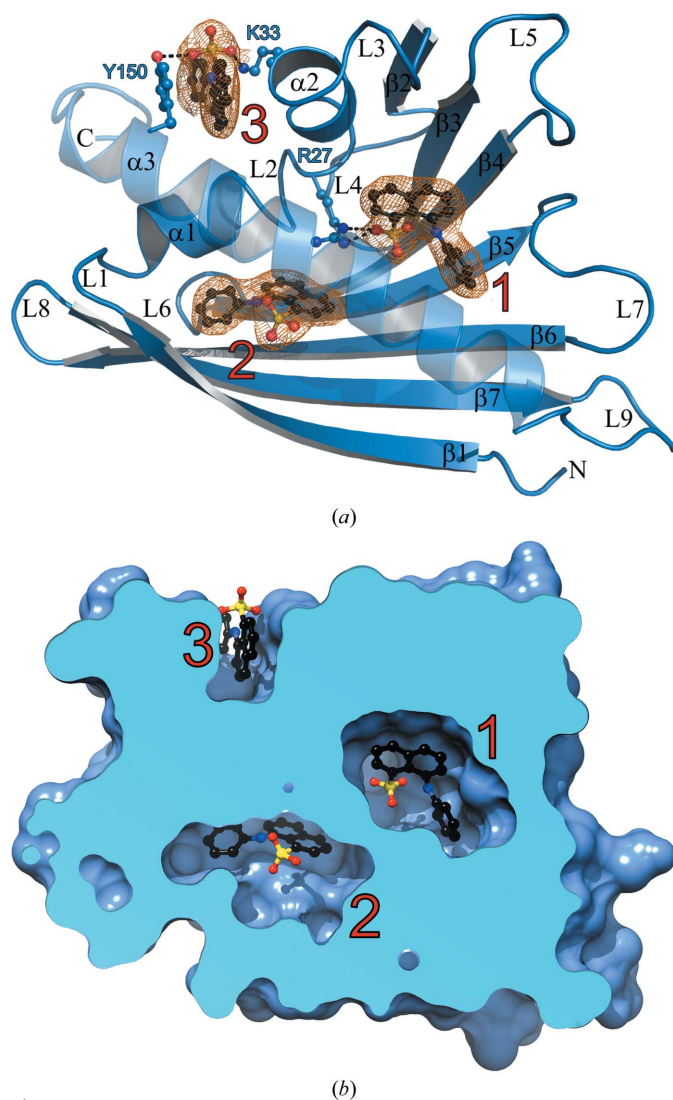


Figure 2
ANS binding to copy *K* of Hyp-1. (a) $2F_o - F_c$ electron density contoured at 1.5σ around the ligands, showing the ANS molecules (red labels). Two ligands are bound in internal chambers (sites 1 and 2) and one in a deep surface pocket (site 3) formed by residues Lys33 and Tyr150. Sites 1, 2 and 3 are occupied in 22, 25 and 13, respectively, of the 28 protein molecules in the asymmetric unit. Dashed lines indicate hydrogen bonds to protein atoms. The ribbon diagram is annotated with numbered secondary-structure elements, with α for helices, β for β -strands and L for loops. (b) A cutaway view of protein molecule *K* generated with *Chimera* (Pettersen *et al.*, 2004), showing ligand positions relative to the protein surface.

to all 28 independent copies of the Hyp-1 molecule resulted in a slight improvement of the refinement statistics. In the final refinement, the NCS restraints were removed without any effect on the refinement statistics. *REFMAC* refinement was alternated with manual rebuilding in *Coot* (Emsley *et al.*, 2010). After modelling 89 ANS molecules and 35 water molecules, the final refinement converged with *R* and *R*_{free} factors of 22.2 and 27.7%, respectively. The r.m.s. deviation from standard bonds was 0.015 Å, with 91.8% of all residues in favoured and 7.0% in allowed Ramachandran regions and just a few Ramachandran outliers in loops L4 and L7, which were partially disordered. The final electron-density maps are of very good quality, showing unambiguously the main-chain trace of all 28 independent protein molecules (*A*, *B*, ... *Z*, *a*, *b*), clear conformations for most side chains and good density for all copies of the C-terminal helix α 3, which is often disordered in PR-10 structures. In addition, the 89 ANS molecules have very good definition in the electron density (Fig. 2*a*).

6. Ligand binding by Hyp-1

The maps show excellent electron density for either one, two or three internal ANS molecules (at sites designated 1, 2 and 3) per Hyp-1 protein (Fig. 2) and 29 interstitial ANS molecules. This structure of the Hyp-1–ANS complex therefore has implications for the ADA method of studying ligand binding to PR-10 proteins using fluorescent probes. The structure shows three clearly defined and separated ligand-binding sites, and the fact that the complex stoichiometry can be 1:1, 1:2 or 1:3 has to be taken into account as a complication when studying the kinetics and stoichiometry of PR-10–ligand complexes using ANS displacement fluorescence. Fortunately, the structure shows that there is no direct interaction between the fluorescing species to further complicate the spectra.

7. Crystal packing and superstructure modulation

The Hyp-1 molecules are arranged into dimers through intermolecular β -sheet formation between β 1– β 1 strands, although the protein is monomeric in solution. Seven of these dimers have the same orientation and nearly equal repetitive spacing along the *c* axis, while the remaining seven are their copies through a noncrystallographic 2_1 axis in the *c* direction. This packing arrangement creates a noncrystallographic screw axis with $\sim 180^\circ$ rotation and 1/14 translation (Fig. 1*c*). The interstitial ANS molecules have a similar but not identical disposition with respect to the sevenfold symmetric packing of the protein molecules. This variation explains why the crystal has a unit cell with a pseudo-sevenfold translation along the *c* axis instead of a smaller cell.

The peculiar pattern of reflection intensities in the *c** direction and the repetitive pattern of molecular packing in the corresponding direction in direct space, leading to a sevenfold expansion of the basic unit cell, are both strong indications that we have a case of a modulated superstructure. Since it was possible to successfully refine the structure using a

sevenfold expanded unit cell, the modulation appears to be commensurate. Modulated structures have been well studied in small-molecule crystallography but are practically unheard of in macromolecular crystallography (Porta *et al.*, 2011). These aspects of the Hyp-1–ANS crystal structure will be treated elsewhere.

8. Conclusion

Our crystal form of the Hyp-1–ANS complex is a case of a modulated superstructure. In protein crystallography such reports are rare (Porta *et al.*, 2011), most likely not because such cases do not exist but because such crystal structures are rejected as too difficult to solve. The present modulation is evidently commensurate, which allows its description in a larger unit cell (here, repeated sevenfold along *c*) without having to resort to description in a higher-dimensional space (Wagner & Schönleber, 2009), which would be very difficult indeed.

In this study, we have demonstrated that novel maximum-likelihood algorithms that account for the structure-factor modulations induced by tNCS are extremely powerful in tackling even the most difficult cases in macromolecular crystallography. In this particular example, the algorithm correctly located 56 copies in space group *P1* of the protein molecule used as a probe, despite near-perfect tetartohedral twinning. The success of our approach is important as it shows that modulated macromolecular superstructures do not have to be discarded but can in fact become sources of structural information on a par with unmodulated structures. Finally, the particular ANS complex of a PR-10 protein shows at atomic detail unexpected protein interactions that have to be taken into account when using ANS as a fluorescent probe in studies of biologically relevant ligand molecules.

The version of *Phaser* that accounts for tNCS using the algorithms described here is available as part of the current releases of both the *CCP4* (Winn *et al.*, 2011) and *PHENIX* (Adams *et al.*, 2010) packages.

Financial support to MJ and JS was provided by the European Union within the European Regional Developmental Fund and by the Polish Ministry of Science and Higher Education (grant No. NN 301 003739). ZD was supported by funds from the Intramural Research Program of the NIH, National Cancer Institute, Center for Cancer Research. RJR was supported by a Principal Research Fellowship from the Wellcome Trust (grant No. 082961/Z/07/Z).

References

- Adams, P. D. *et al.* (2010). *Acta Cryst.* **D66**, 213–221.
- Afonine, P. V., Grosse-Kunstleve, R. W., Echols, N., Headd, J. J., Moriarty, N. W., Mustyakimov, M., Terwilliger, T. C., Urzhumtsev, A., Zwart, P. H. & Adams, P. D. (2012). *Acta Cryst.* **D68**, 352–367.
- Bais, H. P., Vepachedu, R., Lawrence, C. B., Stermitz, F. R. & Vivanco, J. M. (2003). *J. Biol. Chem.* **278**, 32413–32422.
- Biesiadka, J., Bujacz, G., Sikorski, M. M. & Jaskolski, M. (2002). *J. Mol. Biol.* **319**, 1223–1234.

- Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* **D66**, 486–501.
- Evans, P. (2006). *Acta Cryst.* **D62**, 72–82.
- Fernandes, H., Michalska, K., Sikorski, M. & Jaskolski, M. (2013). *FEBS J.* **280**, 1169–1199.
- Gajhede, M., Osmark, P., Poulsen, F. M., Ipsen, H., Larsen, J. N., Joost van Neerven, R. J., Schou, C., Løwenstein, H. & Spangfort, M. D. (1996). *Nature Struct. Biol.* **3**, 1040–1045.
- Lebedev, A. A., Vagin, A. A. & Murshudov, G. N. (2006). *Acta Cryst.* **D62**, 83–95.
- Lovelace, J. J., Murphy, C. R., Daniels, L., Narayan, K., Schutt, C. E., Lindberg, U., Svensson, C. & Borgstahl, G. E. O. (2008). *J. Appl. Cryst.* **41**, 600–605.
- Lovelace, J. J., Simone, P. D., Petříček, V. & Borgstahl, G. E. O. (2013). *Acta Cryst.* **D69**, 1062–1072.
- Luzzati, V. (1952). *Acta Cryst.* **5**, 802–810.
- McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Cryst.* **40**, 658–674.
- Michalska, K., Fernandes, H., Sikorski, M. & Jaskolski, M. (2010). *J. Struct. Biol.* **169**, 161–171.
- Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* **D67**, 355–367.
- Oeffner, R. D., Bunkóczi, G., McCoy, A. J. & Read, R. J. (2013). *Acta Cryst.* **D69**, 2209–2215.
- Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
- Padilla, J. E. & Yeates, T. O. (2003). *Acta Cryst.* **D59**, 1124–1130.
- Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C. & Ferrin, T. E. (2004). *J. Comput. Chem.* **25**, 1605–1612.
- Porta, J., Lovelace, J. J., Schreurs, A. M. M., Kroon-Batenburg, L. M. J. & Borgstahl, G. E. O. (2011). *Acta Cryst.* **D67**, 628–638.
- Read, R. J. (2001). *Acta Cryst.* **D57**, 1373–1382.
- Read, R. J. (2003). *Acta Cryst.* **D59**, 1891–1902.
- Read, R. J., Adams, P. D. & McCoy, A. J. (2013). *Acta Cryst.* **D69**, 176–183.
- Rossmann, M. G. & Blow, D. M. (1962). *Acta Cryst.* **15**, 24–31.
- Santner, A. & Estelle, M. (2009). *Nature (London)*, **459**, 1071–1078.
- Sheldrick, G. M. (2008). *Acta Cryst.* **A64**, 112–122.
- Srinivasan, R. & Ramachandran, G. N. (1965). *Acta Cryst.* **19**, 1008–1014.
- Wagner, T. & Schönleber, A. (2009). *Acta Cryst.* **B65**, 249–268.
- Wilson, A. J. C. (1949). *Acta Cryst.* **2**, 318–321.
- Winn, M. D. *et al.* (2011). *Acta Cryst.* **D67**, 235–242.

PUBLICATION III

ANS complex of St John's wort PR-10 protein with 28 copies in the asymmetric unit: a fiendish combination of pseudosymmetry with tetartohedral twinning

Joanna Sliwiak,^a Zbigniew Dauter,^b Marcin Kowiel,^c Airlie J. McCoy,^d
 Randy J. Read^d and Mariusz Jaskolski^{a,e*}

Received 12 November 2014

Accepted 21 January 2015

Edited by M. Schiltz, Fonds National de la Recherche, Luxembourg

Keywords: pathogenesis-related class 10 protein; St John's wort; *Hypericum perforatum*; 8-anilino-1-naphthalene sulfonate.

PDB reference: Hyp-1, a St John's wort PR-10 protein in complex with the fluorescent probe 8-anilino-1-naphthalene sulfonate, 4n3e

Supporting information: this article has supporting information at journals.iucr.org/d

^aCenter for Biocrystallographic Research, Institute of Bioorganic Chemistry, Polish Academy of Sciences, Poznan, Poland, ^bSynchrotron Radiation Research Section, National Cancer Institute, Argonne National Laboratory, Argonne, IL 60439, USA, ^cDepartment of Organic Chemistry, Poznan University of Medical Sciences, Poznan, Poland, ^dDepartment of Haematology, Cambridge Institute for Medical Research, University of Cambridge, Cambridge CB2 0XY, England, and ^eDepartment of Crystallography, Faculty of Chemistry, A. Mickiewicz University, Poznan, Poland. *Correspondence e-mail: mariuszj@amu.edu.pl

Hyp-1, a pathogenesis-related class 10 (PR-10) protein from St John's wort (*Hypericum perforatum*), was crystallized in complex with the fluorescent probe 8-anilino-1-naphthalene sulfonate (ANS). The highly pseudosymmetric crystal has 28 unique protein molecules arranged in columns with sevenfold translational noncrystallographic symmetry (tNCS) along *c* and modulated X-ray diffraction with intensity crests at $l = 7n$ and $l = 7n \pm 3$. The translational NCS is combined with pseudotetragonal rotational NCS. The crystal was a perfect tetartohedral twin, although detection of twinning was severely hindered by the pseudosymmetry. The structure determined at 2.4 Å resolution reveals that the Hyp-1 molecules (packed as β -sheet dimers) have three novel ligand-binding sites (two internal and one in a surface pocket), which was confirmed by solution studies. In addition to 60 Hyp-1-docked ligands, there are 29 interstitial ANS molecules distributed in a pattern that violates the arrangement of the protein molecules and is likely to be the generator of the structural modulation. In particular, whenever the stacked Hyp-1 molecules are found closer together there is an ANS molecule bridging them.

1. Introduction

The proteins that are expressed by plants under stressful conditions (such as drought, salinity or pathogen invasion), known as pathogenesis-related (PR) proteins, have been divided into 17 classes (Sels *et al.*, 2008). The members of most of these classes have well known biological activity. On this background, PR proteins of class 10 (PR-10) are very unusual because no unique function can be assigned to them despite their abundance, their coexistence as many isoforms in one plant, their differentially regulated expression levels and many years of study (Fernandes *et al.*, 2013). This is particularly surprising since the structure of PR-10 proteins has been very thoroughly studied and even defines a characteristic fold, also known as the Bet v 1 fold after the first protein from this class, a birch pollen allergen, to have its crystal structure determined (Gajhede *et al.*, 1996). The canonical PR-10 fold consists of an extended seven-stranded antiparallel β -sheet with a baseball-glove shape crossed by a long C-terminal helix α_3 , which is the most variable (in terms of both sequence and structural deformations) element of the PR-10 structure (Biesiadka *et al.*, 2002; Pasternak *et al.*, 2006). The seven β -strands form a consecutive progression connected by β -turns and loops,



OPEN ACCESS

except for strands $\beta 1$ and $\beta 2$, which form the edges of the β -sheet and which are connected by a V-shaped motif of two α -helices ($\alpha 1$ and $\alpha 2$) that provides support for the C-terminal end of helix $\alpha 3$. The most intriguing feature of the PR-10 fold is the apparent lack of a proper hydrophobic core, in place of which there is a large hydrophobic cavity formed between the main structural elements, *i.e.* the β -sheet and helix $\alpha 3$. However, the hollow core does not lead to instability, as the PR-10 members are quite robust, resistant to proteases and have a mechanical stability that surpasses that of average globular proteins (Chwastyk *et al.*, 2014). The properties and the size of the internal cavity are modulated by the character of helix $\alpha 3$ in each particular case. The system of conserved β -bulges (which endow the β -sheet with its curvature) and numerous loops (L1–L9), some of which act as gating elements for the cavity, are also important for the PR-10 folding canon. The presence of the internal cavity naturally suggests a biological ligand-binding role. Indeed, several PR-10–ligand complexes have been characterized by crystallography, but their biological significance has only begun to emerge (Ruszkowski *et al.*, 2013, 2014). The persisting concerns are related to the fact that the physiological concentrations of phytohormones, which are the most frequently suggested ligands (Fernandes *et al.*, 2008; Ruszkowski *et al.*, 2014), are low compared with the binding constants, as well as to the observation that while the ligands in the crystal structures usually have excellent definition in electron density, they form diverse protein–ligand binding patterns. For example, similar or identical molecules are bound in multiple ways and even with variable stoichiometry. Additionally, complexes with PR-10 proteins are formed by phytohormones from totally divergent chemical classes, such as cytokinins (Pasternak *et al.*, 2006; Fernandes *et al.*, 2008), brassinosteroids or their analogues (Marković-Housley *et al.*, 2003), gibberellins (Ruszkowski *et al.*, 2014) and abscisic acid (Sheard & Zheng, 2009).

Direct determination of the binding constants, for example by isothermal titration calorimetry (ITC), is often difficult because of the low solubility displayed by most phytohormones. An alternative method, an ANS displacement assay, or ADA, is based on the fact that the fluorescent dye 8-anilino-1-naphthalene sulfonate (ANS) strongly changes its fluorescence in response to the chemical environment (Gasymov & Glasgow, 2007) and therefore can be titrated by another ligand that replaces it in a protein complex. ANS fluorescence is significantly increased after binding to a protein, with a hypsochromic shift of the fluorescence peak. To make full use of this method, the structural properties of ANS complexes with the target proteins should be well understood; as a minimum, the binding stoichiometry should be precisely known. Despite the popularity of the ADA method, it is surprising that there are only two deposited crystal structures (with coordinates) of the ANS anion [entries AMMANS (Weber & Tulinsky, 1980) and ANAPHS (Cody & Hazel, 1977)] in the Cambridge Structural Database (CSD; Allen, 2002) and that structural studies of ANS complexes with PR-10 proteins are scarce and limited to published structures

of Bet v 1 complexes (PDB entries 4a80 and 4a8v; Kofler *et al.*, 2012) and an unpublished structure of a complex with a protein from the Andean crop jicama (PDB entry 1txc; F. Wu, Z. Wei, Z. Zhou & W. Gong, unpublished work). In the former case, the structure helped to explain the anomalous ANS fluorescence data at the molecular level. In the present study (first reported briefly in the context of molecular replacement with translational noncrystallographic symmetry; Sliwiak *et al.*, 2014), we have determined the crystal structure of an ANS complex of Hyp-1, a PR-10 protein from the medicinal herb St John's wort (*Hypericum perforatum*). Hyp-1 was originally implicated (Bais *et al.*, 2003), most likely erroneously (Kořuth *et al.*, 2013), as an enzyme catalyzing the biosynthesis of the pharmacological ingredient of the plant, the dianthrone hypericin, from two molecules of emodin. A subsequent crystallographic study of unliganded Hyp-1 demonstrated that the protein cavity (filled with serendipitous PEG molecules from the crystallization buffer) is indeed compatible with the size of one hypericin or two emodin molecules (PDB entry 3ie5; Michalska *et al.*, 2010). In this context, complex formation between Hyp-1 and ANS is of interest in itself as all of the implicated molecules (hypericin, emodin, ANS) contain large aromatic chromophores.

The Hyp-1–ANS complex studied in this work crystallized in a huge unit cell, with the basic motif of four protein molecules imperfectly repeated along *c*. Such translational noncrystallographic symmetry (tNCS) is sometimes called pseudotranslation. The presence of tNCS causes great difficulties in structure solution for two major reasons. Firstly, it can be difficult to work out how to break the exact lattice translational symmetry correctly. Secondly, most methods assume, at least implicitly, that the structure factors are all drawn from a uniform distribution, whereas in the presence of pseudotranslations there are extreme modulations in the intensity distribution, as seen here. In molecular replacement (MR) this can lead to false solutions because once one copy of a molecule has been placed (correctly or incorrectly), any copy placed in the same orientation but separated by the appropriate translation vector will reproduce the intensity modulation, thus improving the fit to the data without necessarily being correct. The maximum-likelihood methods for MR implemented in *Phaser* (McCoy *et al.*, 2007) depend on an accurate statistical model, so they were found to be highly sensitive to the failure to account for the statistical effects of tNCS. In order to solve the Hyp-1–ANS structure, it was necessary to adapt *Phaser* to account for these effects (Sliwiak *et al.*, 2014). Effectively, the entire set of molecules related by one or more translations is treated as a group, with the molecules rotating in concert during the rotation search and being translated as a group in the translation search. At the same time, the modulation of the error terms in the likelihood target is also accounted for.

To aggravate the problems even further, the crystal was found (belatedly, after the diffraction experiments had been finished) to be tetartohedrally twinned, which not only complicated the structure analysis as such but also resulted in an incomplete data set when indexed in the correct space

group. However, in this case crystal twinning was actually used in a constructive way, *i.e.* to restore data completeness.

2. Materials and methods

2.1. Protein preparation

Hyp-1 was produced in *Escherichia coli* strain DE3 using the pET151/D vector with the *hyp-1* coding sequence and an N-terminal His-tag fusion (Fernandes *et al.*, 2008). 1 l LB medium was inoculated with 10 ml overnight culture grown at 310 K in the presence of 100 $\mu\text{g ml}^{-1}$ ampicillin. At an OD_{600} of ~ 1 , the temperature was lowered to 291 K and isopropyl β -D-1-thiogalactopyranoside (IPTG) was added to a final concentration of 1 mM. After overnight culture, the cells were centrifuged at 6000g for 15 min at 277 K. The pellet was resuspended in lysis buffer [500 mM NaCl, 20 mM Tris-HCl pH 8.0, 20 mM imidazole, 3 mM β -mercaptoethanol, 100 $\mu\text{g ml}^{-1}$ chicken egg-white lysozyme (Sigma-Aldrich)] and sonicated. The lysate was centrifuged at 18 000g for 15 min at 277 K. The supernatant was passed through a HisTrap column equilibrated with wash buffer (20 mM Tris-HCl pH 8.0, 20 mM imidazole, 3 mM β -mercaptoethanol) and eluted with 500 mM imidazole. The His tag was cleaved by His-tagged TEV protease with simultaneous dialysis against

wash buffer at 277 K. After another round of affinity chromatography, the protein was purified on a size-exclusion column in 3 mM citrate buffer pH 6.3 with 150 mM NaCl. After purification, the protein was dialyzed against 3 mM citrate buffer and frozen at 193 K. The purified protein contains an N-terminal hexapeptide extension (GIDPFW-) as a cloning artifact. The final yield of recombinant Hyp-1 was 40 mg per litre of culture.

2.2. Complex formation, characterization and crystallization

For crystallization experiments, the protein solution was concentrated to 15 mg ml^{-1} and pre-incubated at 292 K for 1 h with an eightfold molar excess of ANS added from a 0.1 M stock solution in DMSO. Screening for crystallization conditions using Crystal Screen, PEG/Ion and PEG/Ion 2 (Hampton Research) was performed by the sitting-drop vapour-diffusion method against 120 μl well solution with the use of a Mosquito crystallization robot. The crystallization drops consisted of 0.2 μl protein-ligand solution and 0.2 μl well solution. Small crystals appeared after one week in 0.1 M HEPES pH 7.5 with 1.4 M tribasic sodium citrate as the precipitant. The preliminary crystals were used for seeding in a gradient of PEG 400 or glycerol and tribasic sodium citrate. Large crystals of dimensions 0.1 \times 0.1 \times 0.3 mm (Fig. 1a) appeared in 0.1 M HEPES pH 7.5, 10% glycerol, 1.3 M tribasic sodium citrate. Strong blue fluorescence observed under a UV microscope (Fig. 1b) confirmed the presence of ANS in the crystals.

2.3. X-ray diffraction data collection and processing

Diffraction data collection and processing, including the treatment of data incompleteness resulting from the acceptance of apparent $P422$ crystal symmetry arising from perfect tetartohedral twinning and the eventual choice of $C2$ symmetry following molecular replacement in $P1$, have been described previously (Sliwiak *et al.*, 2014). The diffraction images recorded to 2.43 Å resolution revealed a repetitive sevenfold modulation (Fig. 2) of the reflection intensities along the longest lattice dimension (c), which was interpreted as an indication of a sevenfold noncrystallographic translation of a structural pattern along c .

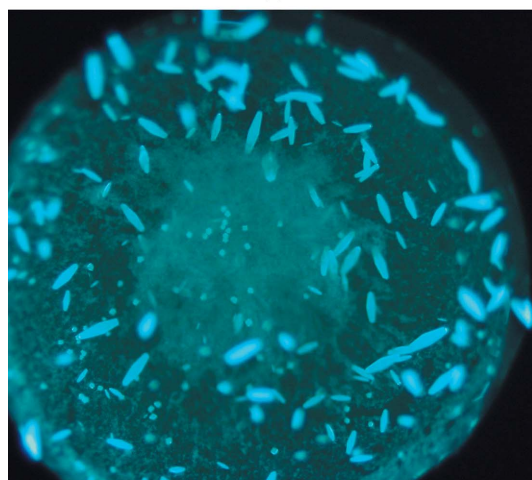
As noted previously, the strategy adopted during data collection, adjusted for tetragonal symmetry, turned out to be inadequate for the $C2$ cell. The 90° of crystal rotation covering the asymmetric unit of the 422 symmetry corresponded to two equivalent 45° ranges instead of the full 90° wide monoclinic asymmetric unit, yielding only $\sim 73\%$ data completeness. However, the presence of perfect tetartohedral twinning suggested an opportunity to expand the data from tetragonal to monoclinic symmetry without introducing significant errors, since in the case of perfect tetartohedry the data agree with the 422 symmetry anyway.

2.4. Structure solution

The procedure that led to the solution of the crystal structure has been outlined before (Sliwiak *et al.*, 2014). Briefly,



(a)



(b)

Figure 1
Single crystals of Hyp-1-ANS under a polarizing microscope (a) and a UV microscope (b).

MR trials in all space groups consistent with a P lattice and point group 422 yielded multiple similar potential solutions in space group $P4_122$, but this symmetry was ruled out by strong $00l \neq 4n$ reflections. Coupled with evidence of twinning, this suggested that the true symmetry was lower, but it was not clear which of the many potential subgroups of 422 point-group symmetry would be correct. Accordingly, structure solution was attempted in space group $P1$, searching for 56

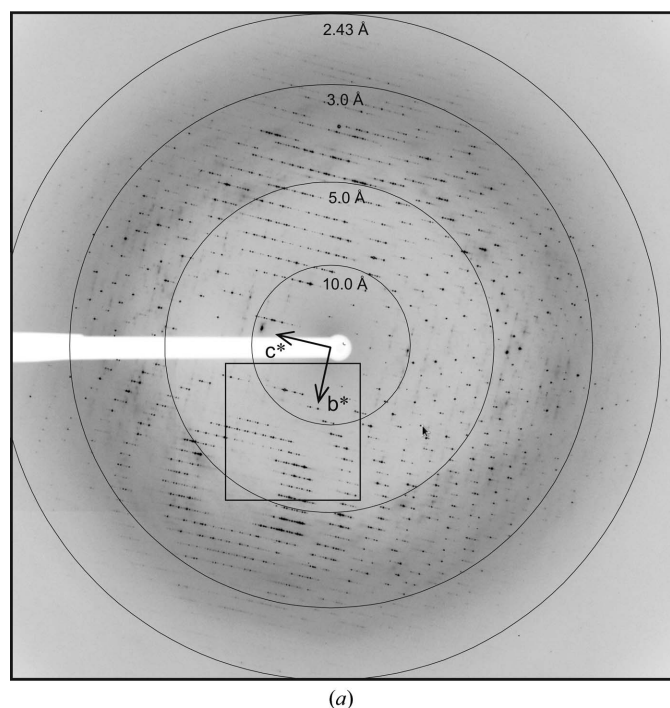
copies of Hyp-1. Although one copy of the model comprises less than 2% of the scattering power, the search accounting for tNCS actually looked for seven copies at a time (in accord with strong native Patterson $0, 0, w$ peaks at $w = n/7$), making the problem tractable. This search succeeded in finding a unique solution, and the correct $C2$ symmetry was deduced by analyzing the symmetry of the calculated structure factors as described below. The MR solution in $C2$ symmetry was obtained by searching for four copies of the first set of seven molecules from the $P1$ solution.

2.5. Structure refinement

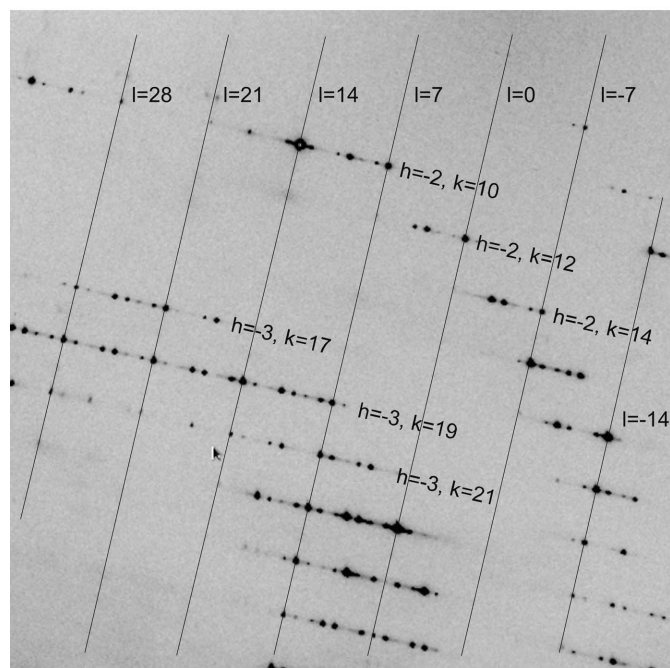
About 3000 (1.3%) R_{free} reflections were selected in *SHELXPRO* (Sheldrick, 2008) in narrow resolution shells to ensure the inclusion of twin-related and NCS-related reflections. The structure was refined in *REFMAC5* (Murshudov *et al.*, 2011) with an intensity-based twin-detection/refinement and jelly-body refinement mode. For the protein molecules, the standard stereochemical restraint library was used (Engh & Huber, 1991). The geometrical restraints for the ANS molecules were created using the coordinates of the magnesium salt of ANS (Cody & Hazel, 1977) found with reference code ANAPHS in the Cambridge Structural Database (Allen, 2002). Briefly, stereochemical targets from this structure were applied to covalent bonds, planar groups and three torsion angles, τ_1 ($O2-S-C9-C10$), τ_2 ($C10-C1-N-C11$) and τ_3 ($C1-N-C11-C16$), with weights adjusted for bonds, planarity and torsions using 0.02 \AA , 0.02 \AA and 20° , respectively, as the standard deviations. Valence sp^2 angles were restrained at $120 (3)^\circ$. The refinement statistics are summarized in Table 1.

2.6. ANS binding assay

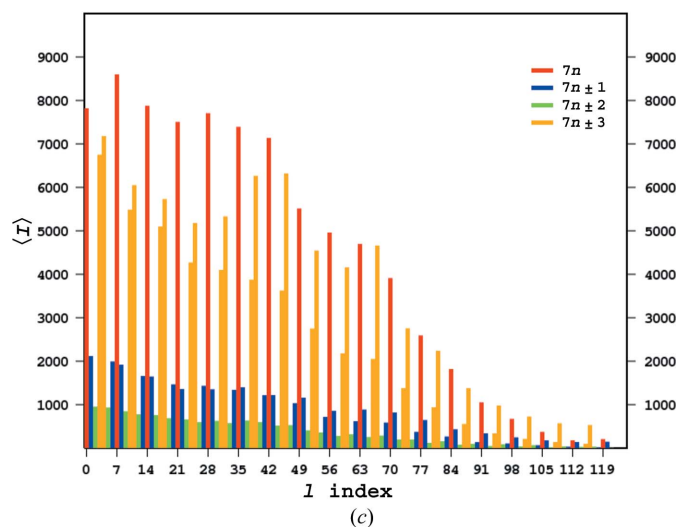
Fluorescence measurements were carried out at room temperature using an RF-5301 Shimadzu spectrofluorimeter



(a)



(b)



(c)

Figure 2 X-ray diffraction images of Hyp-1-ANS crystals. (a) A sample full image, with resolution rings and reciprocal-lattice directions indicated, and (b) an enlarged fragment on which layer lines of constant $l = 7n$ are marked and annotated; the perpendicular layer lines have h, k indices as annotated. (c) A histogram of intensity distribution in layers of l .

Table 1
Refinement statistics.

Values in parentheses are for the last resolution shell.

Space group	<i>C</i> 2
Unit-cell parameters (Å, °)	$a = b = 146.21$, $c = 298.35$, $\beta = 90.07$
Data-collection temperature (K)	100
Resolution (Å)	30–2.43 (2.47–2.43)
Reflections measured	496579
Unique reflections	170447†/61810‡
Completeness (%)	99.8 (99.2)§
Multiplicity	8.0 (7.1)§
Refinement program	REFMAC5
Unique reflections (work + test)	232268§
Test reflections	3077
Molecules in asymmetric unit	
Protein	28
ANS	89
No. of atoms	
Protein	35224
Ligand	1899
Water	35
Metal	15
$\langle B \rangle$ (Å ²)	
Protein	47.3
Ligand	39.9
Water	27.6
Metal	44.0
$R_{\text{work}}/R_{\text{free}}$ (%)	22.3/27.8
R.m.s.d. from ideal geometry	
Bond lengths (Å)	0.015
Bond angles (°)	2.18
Ramachandran statistics¶ (%)	
Favoured	91.80
Allowed	7.04
Outliers	1.16
PDB code	4n3e

† Scaled in *C*2 symmetry. ‡ Scaled in *P*422 symmetry. § After expansion from *P*422 symmetry. ¶ Assessed with *MolProbity* (Chen *et al.*, 2010).

and the following conditions: $\lambda_{\text{exc}} = 378$ nm and $\lambda_{\text{em}} = 470$ nm with 5 nm excitation and emission slits. Concentrated protein (2.6 mM) was titrated in 4–50 μ l aliquots into a cuvette containing 2.5 ml 1 μ M ANS solution in HEPES buffer (25 mM HEPES pH 7.4, 150 mM NaCl, 10 mM MgCl₂, 1 mM β -mercaptoethanol). After each injection, the sample was mixed by pipetting. The fluorescence data (F) plotted against protein concentration were fitted using the equation $F = F_{\text{max}}[\text{protein}]/(K_{\text{d}} + [\text{protein}])$, where K_{d} is the dissociation constant.

2.7. ITC measurements

ITC titrations were carried out at 298 K using a MicroCal iTC200 calorimeter (GE Healthcare). Before the experiment, the protein was dialyzed against a buffer consisting of 150 mM NaCl, 25 mM HEPES pH 7.4, 1 mM β -mercaptoethanol. ANS was dissolved in the dialysis buffer to a concentration of 5 mM. The protein concentration in the sample cell (145 μ M) was determined by the Bradford assay (Bradford, 1976). The ligand solution was injected in 54 aliquots of 1.5 μ l each until saturation was observed. The ITC data were analyzed with the *Origin 7.0* software (OriginLab) to obtain the following parameters: stoichiometry (N), dissociation constant (K_{d}) and

Table 2
Analysis using *POINTLESS* (Evans, 2006).

(a) Analysis of *P*1 model symmetry based on calculated structure factors.

Symmetry element	Symmetry operator	Correlation coefficient	$R_{\text{meas}}^{\dagger}$
Twofold h	$h, -k, -l$	0.93	0.219
Twofold k	$-h, k, -l$	0.95	0.171
Twofold l	$-h, -k, l$	0.92	0.241
Twofold diagonal	$-k, -h, -l$	0.99	0.089
Twofold diagonal	$k, h, -l$	0.92	0.236
Fourfold l	$-k, h, l; k, -h, l; -h, -k, l$	0.94	0.228

(b) Analysis of *C*2 model symmetry based on calculated structure factors in the *P*422 setting.

Symmetry element	Symmetry operator	Correlation coefficient		R_{meas}	
		Rigid body	Final	Rigid body	Final
Twofold h	$h, -k, -l$	0.95	0.67	0.187	0.363
Twofold k	$-h, k, -l$	0.94	0.70	0.196	0.350
Twofold l	$-h, -k, l$	0.94	0.75	0.127	0.177
Twofold diagonal	$k, h, -l$	0.92	0.76	0.235	0.348
Fourfold l	$-k, h, l; k, -h, l; -h, -k, l$	0.95	0.65	0.187	0.424

† $R_{\text{meas}} = \sum_{hkl} \{N(hkl)/[N(hkl) - 1]\}^{1/2} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$ (Diederichs & Karplus, 1997).

the changes in enthalpy (ΔH) and entropy (ΔS) during the complexation reaction. The experimental curves were fitted using one set of binding sites as the model.

3. Results and discussion

3.1. Treatment of diffraction data: transformation from tetragonal to triclinic to monoclinic symmetry

The exploration of possible symmetries has been described previously (Sliwiak *et al.*, 2014), but the details of the statistics on which the decisions were based were not presented.

Because of the initial ambiguity in the true space group of the structure introduced by the physical twinning of the crystal, it was decided to solve the structure by MR in the triclinic *P*1 space group after expansion of the diffraction data to the Ewald hemisphere. The transformation (from *P*422 to *P*1) retains the unit cell but ignores its symmetry, *i.e.* it expands reflections in the same axial system and with the same indices.

After the structure had been solved in the *P*1 unit cell, the 56 copies of Hyp-1 were subjected to rigid-body refinement in *phenix.refine* (Afonine *et al.*, 2009). The symmetry of the MR solution was determined using *POINTLESS* (Evans, 2006) to analyze the relationships among F_{calc} structure amplitudes, which were evaluated in terms of correlation coefficients and merging R factors between reflections related by potential symmetry operations (Table 2a). The agreement was excellent for only one symmetry element, a twofold axis oriented along one of the original tetragonal diagonals, which becomes the unique monoclinic **b** axis after reindexing. The second diagonal becomes the crystallographic **a** direction (without any symmetry), and this choice of axes creates the *C* centring. The original tetragonal **c** direction becomes the **c** axis of the

monoclinic cell and loses its crystallographic symmetry. The asymmetric unit of the monoclinic lattice contains 28 protein molecules, labelled A, B, \dots, Z, a, b .

After the structure was re-solved in the $C2$ unit cell, it was again subjected to rigid-body refinement. This yielded the same R factor (43.2%) as the $P1$ solution, supporting the conclusion that the twofold axis was indeed crystallographic. Analysis with *POINTLESS* (Evans, 2006) showed that there was no further undetected symmetry in the calculated structure factors (Table 2*b*). Although there is significantly better agreement between reflections related by the pseudo-twofold axis parallel to \mathbf{c}^* than for other potential symmetry operators, this operator could only be crystallographic if the other diagonal twofold, which gives much poorer agreement, were also crystallographic.

3.2. Data statistics and detection of twinning in the presence of translational pseudosymmetry

The translational pseudosymmetry causes a modulation of the diffraction pattern in which the strongest intensities tend to have l indices near multiples of 3.5. Accordingly, the native Patterson map has strong peaks at $0, 0, n/7$, with the strongest peak at $w = 2/7$. The modulation broadens the distribution of intensities, thereby masking the statistical effect of twinning. A complete analysis of the statistical effect of tNCS can unmask the effect of twinning on intensities (Read *et al.*, 2013), but the algorithm in *Phaser* is currently only able to model this with sufficient sophistication (including the differences in orientation of tNCS-related copies) in the case of twofold tNCS.

The L -test (Padilla & Yeates, 2003) provides an independent method to unmask the effect of twinning by looking at pairs of reflections separated by vectors in reciprocal space

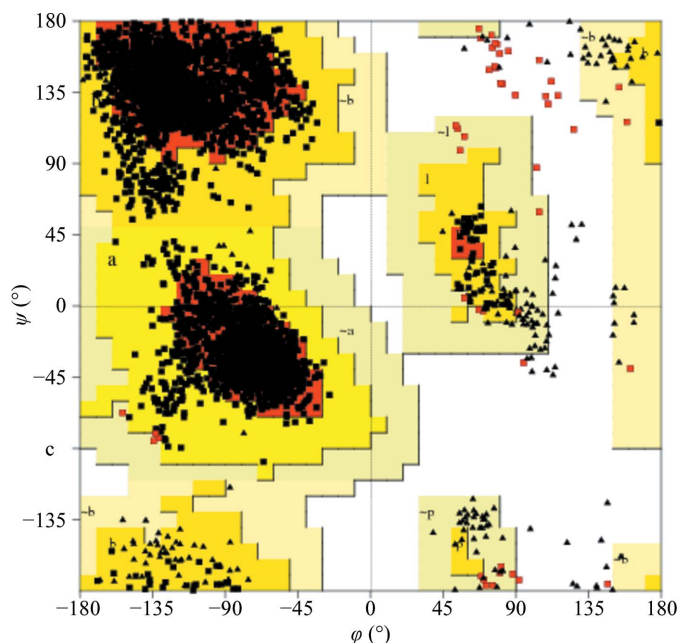


Figure 3 Ramachandran plot for the 28 copies of Hyp-1 in the asymmetric unit, generated by *PROCHECK* (Laskowski *et al.*, 1993). Gly residues are marked by triangles; residues in disallowed regions (2.2%; 0.6%) are in red.

chosen to remove the correlation from tNCS. By default, the L -test uses pairs of reflections separated by multiples of 2 in h, k and l , at least in some implementations. In the present case, reflections separated by 2 in l are actually anticorrelated, because this is approximately half of the distance between peaks in the intensity distribution separated by 3.5 in l . This explains why, in standard applications, the test appears to suggest negative twinning! The *phenix.xtriage* program (Zwart *et al.*, 2005) tries to find a better default separation by taking the inverse of the nonzero coordinates of the top Patterson peak, but unfortunately in the present case the top peak at $w = 2/7$ yields 4 as the nearest integer. This gives a slightly more sensible, but still suboptimal, L -test result. To obtain an optimal L -test for this case, *phenix.xtriage* was run with a separation of multiples of 7 along l , using the expert option `l_test_dhkl = '2,2,7'`. With this separation (and the default of multiples of 2 along h and k), the L -test gives $\langle |L| \rangle = 0.458$ and $\langle L^2 \rangle = 0.288$ for the data merged in $P422$, indicating at least partial twinning. Note that twinning tests based on intensity distributions will underestimate the extent of twinning that parallels the pseudosymmetry, because the intensities of reflections related by pseudosymmetry will be correlated, thus reducing the perturbations in the intensity statistics introduced by twinning. In addition, there are no twin laws for 422 or pseudomerohedral twin laws for this cell in this symmetry, so a crystal cannot both have $P422$ symmetry and suffer from only partial twinning.

3.3. Structure refinement and model quality

The final refinement converged with an R factor of 22.3%, yielding a model with very good stereochemical quality (Table 1). Analysis of the outliers in the Ramachandran plot (presented in Fig. 3 as a consolidated plot generated in *PROCHECK*; Laskowski *et al.*, 1993) carried out in *MolProbity* (Chen *et al.*, 2010) suggests that violations of main-chain conformation are found mainly in four loop areas, L4, L7, L8 and L9, which are usually well defined in other PR-10 structures. Conversely, the C-terminal helix $\alpha 3$, which is often disordered, especially in ligand-free PR-10 structures, is well ordered in Hyp-1. The final electron-density maps allow the tracing of all 28 Hyp-1 chains without gaps. Also, most of the side chains have very clear definition. The high quality of the electron density is illustrated by the fact that 89 copies of the ANS molecule could be confidently modelled in difference electron-density maps phased by the protein component only immediately after MR (Fig. 4*a*). 60 of the ANS ligands are tightly docked (Fig. 4*b*) within three binding sites (denoted 1, 2 and 3) of Hyp-1, but the ligand saturation is not complete (*i.e.* there are fewer than $28 \times 3 = 84$ docked ligands). However, one can easily identify protein chains that are totally empty (T and V) or have one or two binding sites occupied, as well as the 11 copies of Hyp-1 that are fully saturated with three docked ligand molecules. Moreover, an additional 29 ANS molecules with very good electron density were found at selective sites between Hyp-1 molecules. Their locations can be grouped into five superficial sites (denoted 4–8). Surpris-

ingly, despite the huge asymmetric unit cell, only 35 water molecules could be confidently identified in the structure.

3.4. Hyp-1–ANS as a modulated superstructure

The appearance of the diffraction pattern, with an alternation of strong (main) and weak (satellite) reflections in the

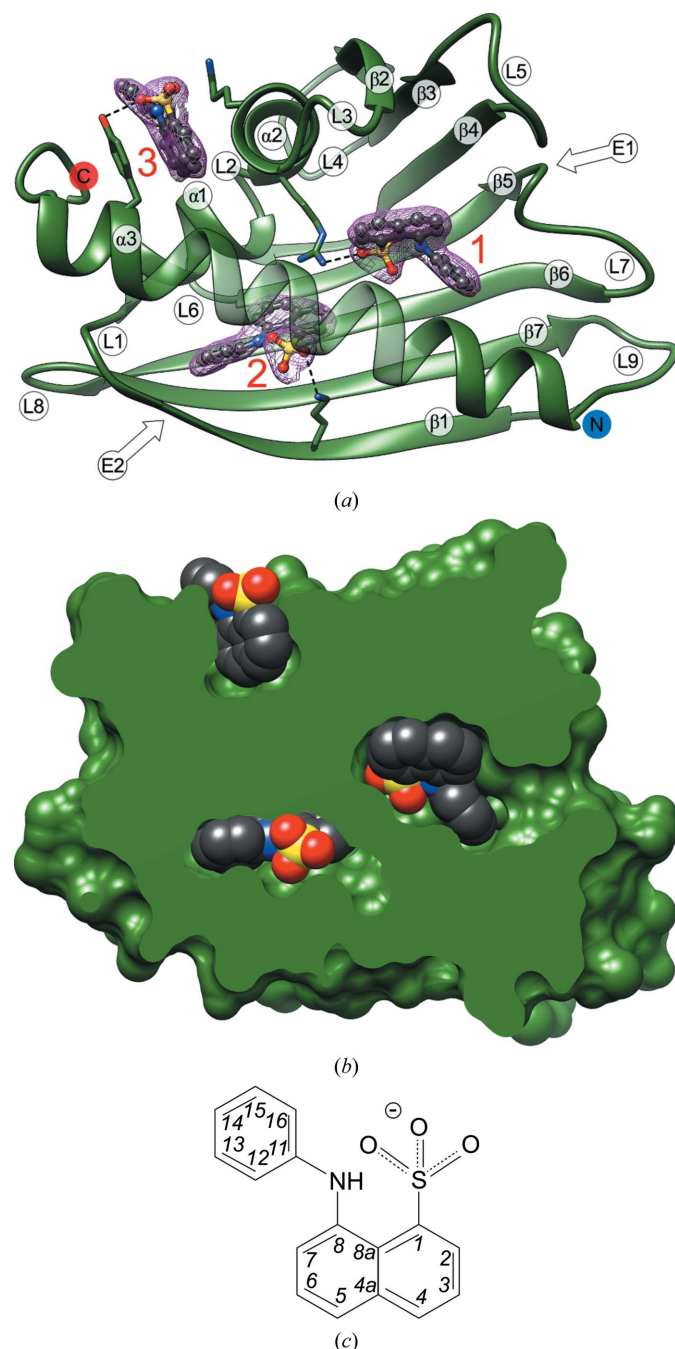


Figure 4
ANS binding sites 1, 2 and 3 in copy *L* of Hyp-1 (*a*) in relation to the (labelled) secondary-structure elements of the Hyp-1 fold, with ANS molecules in ball-and-stick representation, and (*b*) in van der Waals representation to emphasize the excellent fit to the shape of the binding pockets. The ANS molecules in (*a*) are shown in their $2F_o - F_c$ electron density contoured at 1.5σ . Dashed lines indicate hydrogen bonds to protein atoms. Additionally, Lys33 forms hydrogen bonds to ANS at site 3 in some copies of Hyp-1. (*c*) Covalent structure and IUPAC atom numbering of ANS (Jaskolski, 2013).

c^* direction (Fig. 2), and the appearance of the crystal packing in direct space, with a sevenfold translational repetition of the same structural pattern (two Hyp-1 dimers related by $\sim 180^\circ$ rotation and $\sim 1/14$ translation) in the c direction, both suggest that this is a case of a modulated superstructure (Wagner & Schönleber, 2009). However, the successful indexing of the diffraction pattern (of both the main and satellite reflections) with a simple three-dimensional lattice, in which the satellites divide the distances between the main reflections in a rational manner, indicates that the modulation is commensurate. It is thus possible to simplify the description of the structure using an expanded (sevenfold in the c direction) supercell, without resorting to the more rigorous but also much more complicated description in superspace that would be necessary in an incommensurate case.

3.5. Crystal packing of the protein molecules

3.5.1. Dimerization of Hyp-1. In keeping with the majority of other PR-10 proteins, Hyp-1 is monomeric in solution, as confirmed by size-exclusion chromatography and native PAGE electrophoresis (not shown), and is expected to be biologically relevant as a monomer. Nevertheless, we note that there is a precedent of functional dimerization of a PR-10 protein (*Medicago truncatula* MtN13; Ruszkowski *et al.*, 2013) and that in the previously reported unliganded Hyp-1 structure (with PEG molecules found in the protein cavity) the protein molecules were linked into dimers *via* an S–S bond between cysteine residues Cys126 (quite rare in PR-10 sequences). In addition, in another crystal-packing contact two Hyp-1 molecules formed an intermolecular β -sheet through parallel association of their $\beta 1$ strands. It is interesting to note that in the present structure all of the multiple copies of Hyp-1 are also paired into dimers *via* intermolecular $\beta 1$ – $\beta 1$ interactions. At variance with the previous structure, all of the present $\beta 1$ – $\beta 1$ sheets are antiparallel, thus flawlessly extending the intramolecular β -sheet from one molecule to the other (Fig. 5). In the adopted labelling scheme, the Hyp-1 dimers are *AB*, *CD*, ..., *ab*. Seven of these dimers (*AB*, ..., *MN*) have the same orientation and similar, repetitive spacing along the c axis, forming a discernible row (denoted *I*) in this

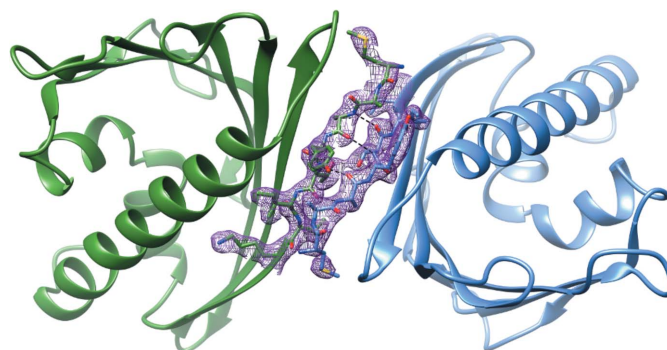


Figure 5
Hyp-1 dimer formation in the crystal structure shown in $2F_o - F_c$ electron density (contoured at 1.5σ) for protein molecules *K* and *L*. The antiparallel intermolecular β -sheet is created by the $\beta 1$ strands from both molecules.

direction. The remaining seven dimers are copies of the former dimers through a noncrystallographic 2_1 screw axis along c and form another row (II) in this direction. In effect, this zigzag packing arrangement follows a noncrystallographic $2_{1/7}$ screw axis with $\sim 180^\circ$ rotation and $\sim 1/14$ translation (Fig. 6). The 29 interstitial ANS molecules have a similar but not identical disposition with respect to the sevenfold symmetric packing of the protein molecules. This deviation from perfection explains why the crystal has a unit cell with pseudo-sevenfold translation along the c axis.

3.5.2. Higher-order association in the crystal lattice. As explained in §3.7, the Hyp-1 dimers form a pillar following a left-handed helical line with a pitch of $c/7$ (black line in Figs. 7*a* and 7*b*). The ANS molecules follow the helical pattern of the protein dimers but can be segregated into three groups. The first group (yellow in Figs. 7*a* and 7*b*), corresponding to binding sites 1 and 2, are located within the protein cavities and are closely associated with unique protein partners and therefore exactly follow the protein helix. The molecules in the second group correspond to binding sites 7 and 8 (green), where they link Hyp-1 molecules, helping to create the helix of dimers. The ANS molecules in the third group (red) lie outside of the protein helix and at sites 4, 5 and 6 glue the neighbouring helices together. This group also includes the surface-pocket site 3. The red molecules follow a (red) helical line that is similar to that of the Hyp-1 helix but has a larger radius. The ANS molecules viewed along the helical axis are shown in Fig. 7(*b*). Even though they follow the respective helical lines, they do not create a regular angular pattern around the helix axis.

3.6. ANS binding

Although the ANS ligand was added to the crystallization buffer as a DMSO solution of the acid form (sulfonic acid), there is no doubt that in view of the pK_a value of -1 the compound is deprotonated to its anionic form (sulfonate) in aqueous solutions and upon interaction with a protein.

3.6.1. Hyp-1-ANS binding assays. ANS binding by Hyp-1 in solution was tested by both calorimetric and fluorometric assays. The calorimetric titration curve (Fig. 8*a*) was fitted using a model of one set of N independent binding sites to yield a stoichiometry of $N = 3$ and a dissociation constant $K_d = 108 \pm$

$3 \mu M$. At the end of the ITC titration, when all three binding sites were saturated, the Hyp-1:ANS molar ratio was 1:12. We note that the eightfold molar excess of the ligand during the crystallization experiments resulted in incomplete occupation (2.14 per protein molecule on average) of the three binding sites, although on the other hand as many as 29 interstitial ANS molecules were still available for docking. It is difficult, however, to directly compare the situation within a crystal lattice with the dynamic equilibrium in solution.

In fluorometric titration, the titration system is inverted and we used a fluorescent ligand at a very low and constant concentration together with a variable concentration of the protein. In such a system, where the ligand concentration is much lower than the expected K_d , we do not achieve full saturation of the protein with the ligand. Moreover, if one of

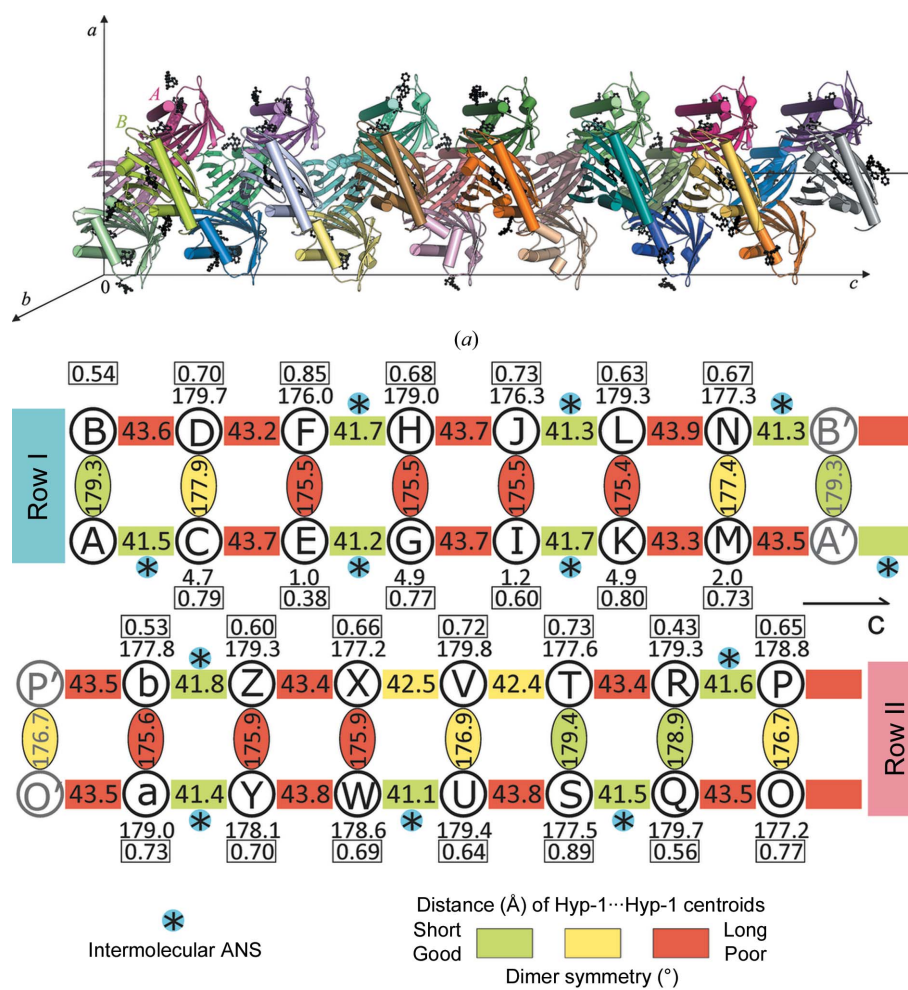


Figure 6 (a) The 28 independent Hyp-1 molecules forming the asymmetric unit of the crystal packing (with molecules *A* and *B* labelled), arranged in a dimeric pattern with a sevenfold repeat around a noncrystallographic 2_1 screw axis along the crystallographic c direction. The ANS ligands are shown as black ball-and-stick models. (b) Arrangement of Hyp-1 dimers (labelled by protein chain identifiers *A, B, ..., b*) in two rows (*AB, ..., MN* and *ab, ..., OP*) along c . Protein copies translated along c are marked with a prime. Centroid distances between consecutive Hyp-1 molecules are marked in Å. Cross-linking of consecutive Hyp-1 pairs through an interstitial ANS molecule is marked by *. NCS symmetry of the dimers is indicated by the degree of rotation between the two chains. The rotation required for the best superposition of molecule *A* onto the remaining Hyp-1 C^α traces is given for each chain, with the corresponding r.m.s.d. (in Å) boxed. All rotations (in $^\circ$) were calculated in *ALIGN* (Cohen, 1997).

the sites has a much higher affinity, the K_d value determined in such an assay could refer to that particular site only. From the fluorometric titration of ANS with Hyp-1 (Fig. 8*b*), a K_d value of $58 \pm 4 \mu\text{M}$ was determined, which is in reasonable agreement with the global value from the ITC experiment. From the analysis of the crystal structure it could be speculated that ANS binding at site 1 is the strongest, as the protein always uses Arg27 to form an ion pair with the ligand with the same binding geometry, in contrast to sites 2 and 3 where mainly hydrophobic interactions are detected supported by sporadic hydrogen bonds. It is therefore likely that the K_d value of $58 \pm 4 \mu\text{M}$ most closely characterizes site 1.

3.6.2. Structural description of the ANS sites. As mentioned above, in addition to the three (internal) ANS docking sites (1, 2 and 3) there are also interstitial sites 4, 5, 6, 7 and 8 occupied by ANS molecules that ‘glue together’ some of the Hyp-1 molecules in the crystal structure. Hereafter, the

ANS sites are denoted using the protein chain label (of the nearest protein molecule for interstitial sites) and the site number, e.g. A1.

3.6.3. Internal Hyp-1 ligand-binding sites. Binding sites 1 and 2 are internal enclosures or chambers within a general PR-10-type cavity that are sealed off and separated from one another. In fact, a typical PR-10 cavity is not present in the Hyp-1 core because the two chambers are nearly completely isolated and binding sites 1 and 2 have their own separate entrances: E1 and E2, respectively. Entrance E1 is surrounded by loops L3, L5 and L7 and the N-terminal part of helix α_3 , whereas entrance E2 is gated by the full length of α_3 and strand β_1 . The main partition between sites 1 and 2 is formed by Arg27 from helix α_2 . Additional residues that form a division between sites 1 and 2 are Ala140 and Phe143 from helix α_3 , Tyr84 from strand β_6 and Tyr101 from strand β_7 . As a consequence, there is no contact between the ANS molecules

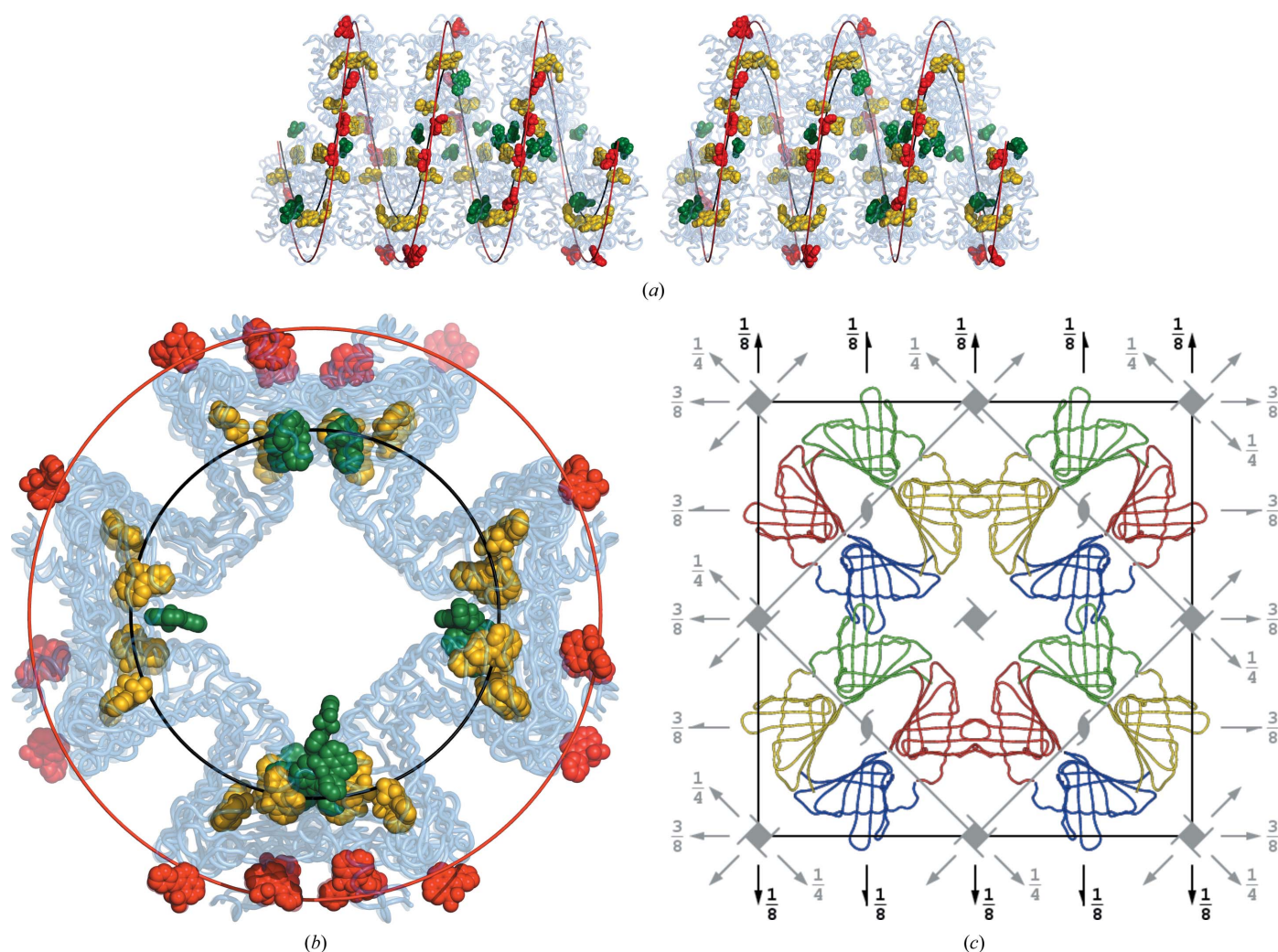


Figure 7
 (a, b) Interpretation of the structure in terms of helical motifs. The black helical line was fitted to the centres of the consecutive *ba-NM-AB-OP-ZY-LK-CD-QR-XW-JI-EF-ST-VU-HG* dimers with $0.0 \leq z < 0.5$. The ANS molecules at binding sites 1 and 2 are in yellow, those at binding sites 3, 4, 5 and 6 are in red and those at binding sites 7 and 8 are in green. The red helical line was fitted to the ANS molecules marked in red. Views down the **b** (stereo) (a) and **c** (b) axes are shown. (c) The $C2$ unit cell and its symmetry elements (black) and the approximate $P4_122$ unit cell and its symmetry elements (grey). The four rows in each Hyp-1 column are shown in different colour (red/green and blue/yellow dimers), in each case representing a set of seven molecules translated along the projection axis **c**.

Table 3

Protein–ANS interactions ($<3.2 \text{ \AA}$) in all 28 copies of Hyp-1 at the internal sites 1, 2 and 3 with the frequency in parentheses.

Residues highlighted in bold form hydrogen bonds to the sulfonate group of ANS (note that Arg27 at site 1 always interacts with ANS *via* hydrogen bonding, while other residues in bold show a variable pattern of hydrogen-bonding/hydrophobic/no interactions at the remaining binding sites).

Binding site (No. of chains occupied by ANS in this binding site)	Residues involved in contact to ANS (No. of chains with this interaction)
1 (23)	Arg27 (22), Gln35 (5), Leu31 (2), Val91 (1), Glu132 (1), Gly136 (1), Lys139 (1), ANS at site 7 (2)
2 (24)	Tyr144 (23), Lys8 (12), Leu19 (6), Ile116 (6), Glu10 (5), Tyr141 (3), Leu23 (2), Arg27 (1), Tyr84 (1)
3 (13)	Tyr150 (11), Lys33 (9), Val30 (3), Val147 (3), Phe158 (3), Val157 (2), Leu151 (1)

at sites 1 and 2. Site 3 is a deep surface-binding pocket formed by a deep invagination of the protein surface between Lys33 and Tyr150.

It is intriguing to note that in the numerous (28) copies of the protein molecule, a given binding site is either fully occupied by an ordered ANS molecule (the most typical situation) or is left completely empty. With just one exception (site R3 with 50% occupancy), there are no intermediate situations observed, for example of partial occupancy of a binding site or of a snapshot of an ANS molecule during its transition to its final binding site.

From the point of view of saturation with the ANS ligand, the two protein rows related in the asymmetric unit by the noncrystallographic 2_1 screw axis along **c** are not equivalent at all (Fig. 9*a*). In row I (dimers *AB/CD/EF/GH/IJ/KL/MN*), the ‘first’ Hyp-1 molecule of each dimer (*A, C, . . . , M*) has the internal docking sites 1, 2 and 3 fully saturated with ANS in all cases and the ‘second’ molecules (*B, D, . . . , N*) are nearly all fully saturated, with the only vacancies left at *D3, F1, F3, J3* and *N3*. The situation in row II (*OP/QR/ST/UV/WX/YZ/ab*) is very different. Here, the first Hyp-1 molecules (*O, Q, . . . , a*) have many vacancies, with site 3 being empty in all of them (with additional vacancies at sites *Q1, S2* and *a1*). The set of the second molecules (*P, R, . . . , b*) of these dimers has nearly the same number of vacancies but with an entirely different pattern, namely with Hyp-1 molecules *T* and *V* having no internal ligands and with additional vacancies at site 3 of *R* (partial), *X* and *b* and at site 2 of *P*.

Considering all of the internal sites of all the Hyp-1 molecules in both rows, it can be summarized that site 1 is empty in five cases, site 2 in four cases and site 3 in 15 cases (15.5 to be exact). Most vacancies (19.5 out of 24.5) are in row II. It appears that this unusual and complicated pattern of docked ANS ligands in the two rows of Hyp-1 molecules repeats itself regularly throughout the crystal lattice because the electron density of the ANS molecules at these sites is very good, clearly indicating well conserved unique orientations and conformations of the ligands.

Table 3 illustrates the interactions between protein residues and the ANS molecules at sites 1, 2 and 3. The ANS molecule at site 1 is mainly anchored by a salt bridge between the

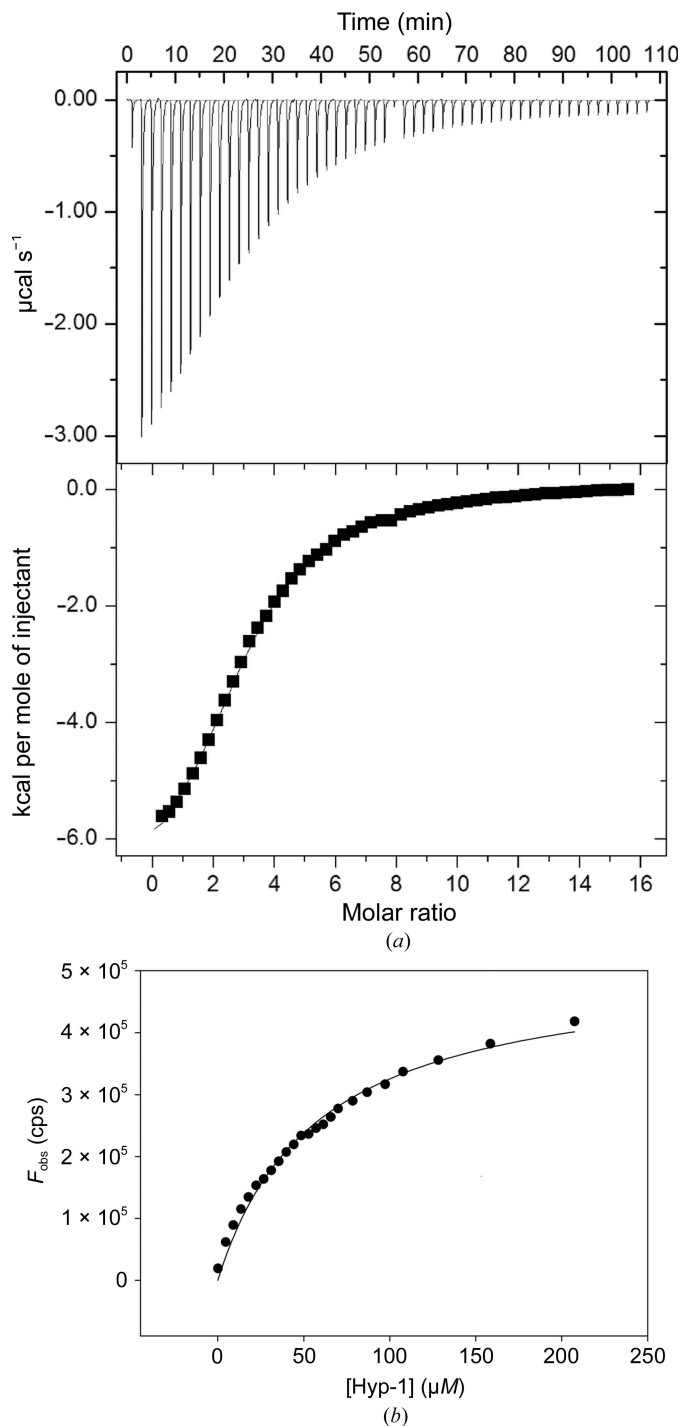


Figure 8

(*a*) ITC titration of Hyp-1 with ANS. The top panel shows raw heat data for 54 consecutive injections of 5 mM ANS into the sample cell (200 μl) containing 0.145 mM Hyp-1 in 25 mM HEPES pH 7.4 at 298 K. The bottom panel shows the binding isotherm created by plotting the heat peak areas against the molar ratio of ANS added to the protein. The line represents the best fit to the model of N independent sites. ANS binding is endothermic with 1:3 stoichiometry ($N = 3.14 \pm 0.02$) and a K_d of $108 \pm 3 \mu\text{M}$. The change in enthalpy ΔH is $-7213 \pm 77 \text{ cal mol}^{-1}$ and that in entropy ΔS is $-6.04 \text{ cal mol}^{-1} \text{ K}^{-1}$. (*b*) ANS binding to Hyp-1 monitored fluorometrically by titration of 1 μM ANS in 25 mM HEPES pH 7.4 with Hyp-1. The line represents the best fit to the equation $F = F_{\text{max}}[\text{protein}] / (K_d + [\text{protein}])$. The calculated K_d value is $58 \pm 4 \mu\text{M}$ (R^2 of fitting 0.9878).

Table 4

Statistics of the conformations (torsion angles τ_1 , τ_2 and τ_3) of the ANS molecules at different Hyp-1 binding sites compared with the CSD structure ANAPHS (Cody & Hazel, 1977).

For each angle at the designated sites, the mean value and its standard deviation are given. Sites 4, 5 and 6 are treated jointly as they correspond to essentially the same position of the ligand at which it glues together three neighbouring Hyp-1 chains. Likewise, sites 7 and 8 are found between two Hyp-1 chains. The statistical analysis takes into account the discontinuity ($+180/-180^\circ$) in torsion-angle definition. The atom numbering of the ANS molecule (Fig. 4c) follows the recommendation of IUPAC, as explained by Jaskolski (2013), regardless of the system adopted by the PDB.

Site	Torsion angle ($^\circ$)					ANAPHS
	1	2	3	4/5/6	7/8	
C2–C1–S–O \ddagger (τ_1)	3 (2)	1 (2)	2 (2)	-5 (2)	-1 (2)	1
C7–C8–N–C11 (τ_2)	38 (2)	12 (1)	-2 (6)	-21 (5)	11 (6)	-54
C8–N–C11–C \ddagger (τ_3)	11 (3)	24 (5)	-3 (11)	-31 (4)	1 (12)	-7

\ddagger The sulfonyl O atom was selected to minimize $|\tau_1|$. \ddagger The aniline C12 atom was selected to minimize $|\tau_3|$.

sulfonate anion and the guanidinium group of Arg27 (Fig. 4a). In two cases, ANS at site 1 is additionally pushed from the outside by hydrophobic contacts with an external ANS molecule at site 7. The main molecular contact at site 2 is based on stacking interactions between the aniline substituent of ANS and the aromatic ring of Tyr144, supported in 11 copies of Hyp-1 by hydrogen bonding to the N $^\delta$ atom of Lys8 from strand β 1, which also delimits this binding pocket. The ligand molecule at site 3 forms vice-type stacking interactions with Lys33 and Tyr150, which additionally form hydrogen bonds to the ANS molecule in one and eight cases, respectively. As ANS binding to proteins is mainly affected by ionic interactions with positively charged residues (Matulis & Lovrien, 1998), one can speculate that in Hyp-1 binding site 1 the dominating interaction is with the positive charge of Arg27. At site 2, this role could be played by Lys8, which in about half of the cases is in hydrogen-bonding contact with the ligand. At site 3, Lys22 is the nearest cationic centre but it forms a hydrogen-bond contact with ANS in only one case.

3.6.4. Interstitial ligands. The 29 interstitial ANS molecules occupy the five superficial sites (4–8) on the surface of the protein molecules much more sparsely and there does not seem to be a discernible pattern of occupancy. The sparsity of the superficial sites is similar around both rows. There is no Hyp-1 molecule that has all of the associated superficial sites occupied. Likewise, none of the superficial sites is occupied in all copies of the protein. Moreover, while the internal sites are always occupied in exactly the same manner, leading to very good superposition of the ligand molecules, particularly at sites 1 and 2 (Fig. 9b), the superficial positions show a higher degree of positional and conformational variability, which at sites 7 and 8 is manifested by a range of locations.

The interstitial ANS molecules in sites 4, 5 and 6 are surrounded by three neighbouring protein chains and are stabilized mainly by hydrogen bonds to the peptide group of Gly47 (in loop L4). This interaction is supported in several cases by contacts (<3.2 Å) with single atoms from loops L6 and L8. The ANS molecules at sites 7 and 8, where they glue

two adjacent protein molecules, interact with protein residues from loops L3 and L5 as well as from helix α 3. Residue Lys138, which in most cases forms a salt bridge to the sulfonate group, seems to play a crucial role in these interactions.

Fig. 9(b) shows all 89 ANS molecules superposed using a common C $^\alpha$ framework of the nearest Hyp-1 molecule. It indicates that the position of the ligand molecule is most stable at sites 1, 2, 4, 5 and 6. At site 3 the ANS molecule appears to be rotating between the jaws of the vice. Sites 7 and 8 are characterized by a large scatter. However, the pattern is not random but is located alongside helix α 3 (8) and the E1 entrance (7) of a neighbouring protein molecule.

3.6.5. ANS conformation. The geometry of the ANS molecules¹ was analyzed using the three rotatable torsion angles τ_1 (C2–C1–S–O; the orientation of the sulfonate group), τ_2 (C7–C8–N–C11; the orientation of the aniline substituent) and τ_3 (C8–N–C11–C; the rotation of the phenyl ring of the aniline substituent). Table 4 illustrates that the conformations at the different binding sites are quite distinct, with the exception of the τ_1 angle, which owing to the threefold symmetry of the substituent is generally close to 0° . The ANS molecules at sites 1 and 2 have well conserved but different conformations. In particular, the aniline substituent at site 1 deviates from the naphthalene plane in a very significant way. The rotational variability of the phenyl substituent is quite large, especially at sites 3 and 7/8, as illustrated by the elevated values of the standard deviations in Table 4. This agrees with the observation that while the vast majority of the ANS molecules have perfect definition in the electron density, in seven cases (five of which are at sites 7 and 8) the electron density of the aniline substituent is blurred.

Although the torsion angles τ_1 and τ_3 of the ANS molecules are similar to those in the ANAPHS structure from the CSD, the τ_2 angle deviates quite significantly (up to 92°).

The 1,8-substituted naphthalene ring in the small-molecule ANAPHS structure (Cody & Hazel, 1977) that served as the source of the ANS restraints is significantly distorted, with the substituents showing particularly large deviations from the naphthalene system. The weight of planarity restraints ($\sigma_{\text{flat}} = 0.02$ Å) applied in *REFMAC* evidently over-restrained the planarity against the experimental evidence, visible for example as a $>10\sigma$ deviation from planarity of the N atom in 33 ANS molecules. An additional round of refinement with $\sigma_{\text{flat}} = 0.2$ Å was able to rectify this and created ANS models with similar deviations from idealized geometry as in ANAPHS. The issue of ANS deformations will be analyzed in depth elsewhere.

3.7. Pseudosymmetric aspects of the crystal structure

The crystal structure of Hyp-1–ANS is highly pseudosymmetric in two aspects: firstly because of the way the protein molecules are arranged in infinite columns along the longest cell dimension and secondly because of the way these columns pack in the unit cell.

¹ The numbering scheme of the ANS molecule (Fig. 4c) follows the recommendation of IUPAC, as explained by Jaskolski (2013), regardless of the system adopted by the PDB.

Fig. 6(a) presents the 28 Hyp-1 molecules grouped into columns built from the pseudo-twofold-symmetric dimers *AB*, *CD*, ..., *ab* that are arranged in a zigzag fashion. One half of this column (row I) is formed by (seven) dimers *AB*, ..., *MN* separated by a shift of $\sim 1/7$ of the cell length. The second half of the column (row II) is formed by a similar series of dimers *OP*, ..., *ab* and can be generated from the first row by a rotation of $\sim 180^\circ$ and a translation of $\sim 1/2$ along the *c* axis, which is equivalent to a translation of $\sim 1/2$ of the interdimer distance. The column composed of these two rows of dimers (red/green and blue/yellow in Fig. 7c) is therefore formed according to a ‘ $2_{1/7}$ ’ screw axis, with a rotation of $\sim 180^\circ$ and a translation of $\sim 1/14$ along the unit-cell *c* axis. In addition, there are two sets of pseudo-twofold axes perpendicular to the column axis. The Hyp-1 dimers are generated by one set and there are 14 such dyads in the unit cell. The dimers across the zigzag pattern of the column are related by the axes from the second set and there are also 14 such dyads; they are perpendicular to the first set and are located halfway between them. The approximate symmetry of the column may be described by the symbol $222_{1/7}$. The distances between the successive Hyp-1 molecules are similar but not equal, and the location of the ANS ligands is also variable. Perfect repetition along the column is achieved only after seven translations.

If the 28 Hyp-1 molecules are collapsed to a sevenfold smaller unit cell, *i.e.* if all dimers are shifted by the appropriate fraction of the *c* cell length ($1/7$, $2/7$ etc.) and overlapped on the *AB* and *ab* dimers, the r.m.s. distance of all 4452 C^α atoms from their mean position in each group of seven molecules is 1.18 Å. The symmetry of such an assembly is approximately 222_1 . If, in addition, all of the molecules are transformed according to that symmetry, the 28 molecules superpose onto one target with an r.m.s.d. of 1.23 Å. The latter value illustrates the difference between the positions of all of the C^α

atoms in the real (pseudosymmetric) and idealized ($222_{1/7}$ symmetric) column.

In the *C2* unit cell there are four columns of Hyp-1 molecules as described above. Owing to their pseudosymmetry, their packing is also pseudosymmetric, as illustrated in Fig. 7(c). After an appropriate shift along the monoclinic *y* axis, the four columns are positioned exactly in each of the four quarters of the unit cell, and at a cursory glance their packing seems tetragonal. Indeed, the *C*-centring moves the $\frac{1}{4}, \frac{1}{4}$ column to $\frac{3}{4}, \frac{3}{4}$, and these two columns are related by a $2_{1/7}$ axis, which also includes a 2_1 operation (as its sevenfold repetition). Similarly, the monoclinic twofold axis transforms the $\frac{1}{4}, \frac{1}{4}$ column to $\frac{3}{4}, \frac{1}{4}$ and the monoclinic 2_1 axis transforms it to $\frac{1}{4}, \frac{3}{4}$. The columns in the latter two pairs are also related by an approximate $4_{-1/7}$ screw axis involving a clockwise 90° rotation and a negative shift by $1/28$ of the *c* axis. This left-handed $4_{-1/7}$ screw axis includes a right-handed 4_1 screw axis (as its sevenfold repetition), a $2_{1/7}$ screw axis (twofold repetition) and a 2_1 screw axis (14-fold repetition). These relations are analogous to the case of the left-handed 6_4 screw axis, which contains the right-handed 3_1 and neutral 2_1 screw axes (Dauter & Jaskolski, 2010).

Taking into account the presence of the (perpendicular) twofold axes, the arrangement of the Hyp-1 molecules in the four columns approximately corresponds to the $P4_122$ and ‘ $P4_{-1/7}22$ ’ space groups. The primitive tetragonal unit cell has one-half of the *C2* cell volume and is rotated by 45° around *c*. In addition, to conform to the location of the twofold axes in the original *C2* symmetry, the origin of the tetragonal cell is shifted along the fourfold axis by $-1/8$. The handedness of the pseudo-tetragonal axis results from the particular shift of the Hyp-1 columns with their local dyads with respect to the crystallographic twofold axes. If the columns were shifted from the current position by an odd multiple of $1/28$ of the cell

Chain	Cavity-docked ANS			Surface-docked ANS				Chain	Cavity-docked ANS			Surface-docked ANS					
	1	2	3	4	5	6	7		8	1	2	3	4	5	6	7	8
A	x	x	x	<i>Y5, C6</i>	<i>Y4, a6</i>	-	<i>M'8</i>	-	B	x	x	x	<i>b5, N'6</i>	<i>b4, Z6</i>	-	-	-
C	x	x	x	-	-	-	-	<i>K'7</i>	D	x	x	-	-	-	-	-	x
E	x	x	x	<i>U5, G6</i>	<i>U4, W6</i>	-	<i>I'8</i>	2x	F	-	x	-	-	-	-	-	-
G	x	x	x	-	-	x	2x	<i>2x / G'7</i>	H	x	x	x	-	-	-	-	x
I	x	x	x	<i>Q5, K6</i>	<i>Q4, S6</i>	-	<i>E'8</i>	2x	J	x	x	-	-	-	x	-	2x
K	x	x	x	-	-	x	x	-	L	x	x	x	<i>R5, J6</i>	<i>R4, P6</i>	-	-	-
M	x	x	x	-	-	-	-	x	N	x	x	-	-	-	-	-	2x
O	x	x	-	-	-	-	-	-	P	x	-	x	-	-	x	-	-
Q	-	x	-	x	-	-	<i>D'8</i>	-	R	x	x	0.5	x	x	-	x	<i>X7</i>
S	x	-	-	-	-	x	-	-	T	-	-	-	-	-	-	x	<i>T'7</i>
U	x	x	-	x	x	-	<i>H'8</i>	-	V	-	-	-	-	-	-	-	<i>T'8</i>
W	x	x	-	-	-	-	<i>J'8</i>	-	X	x	x	-	-	-	-	x	<i>R'7</i>
Y	x	x	-	x	x	-	-	-	Z	x	x	x	-	-	x	-	-
a	-	x	-	-	-	x	<i>N'8</i>	-	b	x	x	-	x	x	-	<i>b'8</i>	x

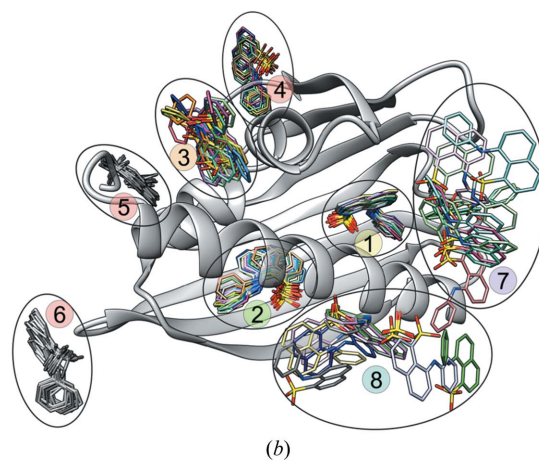


Figure 9

(a) Saturation of the Hyp-1 molecules in the two rows, highlighted in blue (I) and pink (II), with ANS ligands. Full occupancy is marked with an x (or 2x if two ligand molecules are found in the general area of a particular binding site), 0.5 denotes a single case (site R3) of 0.5 occupancy. The docked sites 1, 2 and 3 are highlighted in green. An entry on a grey background with designation of contact sites in adjacent (one or two) Hyp-1 molecules repeats another entry marked x. Symmetry-related protein molecules are indicated with primes. ANS molecules farther than 3.2 Å from a particular protein chain are marked in italics. (b) All of the ANS molecules (sticks) superposed using a common frame of the C^α atoms of the nearest protein molecule (shown in ribbon representation). The intramolecular binding sites 1, 2 and 3 are much more constant than the superficial sites, especially 7 and 8. The ANS molecules are colour-coded by the nearest protein molecules in Fig. 6(a).

length ($1/28, 3/28, \dots, 7/28 = 1/4, \dots$), the pseudo-tetragonal space group would be $P4_322$ or ' $P4_{1/7}22$ '. If all Hyp-1 molecules in the four columns are superposed onto one target according to the idealized $P4_{-1/7}22$ symmetry, the r.m.s.d. value for all C^α atoms is 1.71 Å.

A less intuitive view of the crystal packing, but one that is more amenable to analysis, is obtained by considering four rows of Hyp-1 dimers extending along c as a 'pillar'. In this view, the protein dimers in such a pillar follow a left-handed helical line in the order $ba-NM-AB-OP-ZY-LK-CD-QR-XW-JI-EF-ST-VU-HG$ within $0 \leq z < 0.5$ and then continue smoothly in the unit cell ($0.5 \leq z < 1$) in the order $GH-UV-TS-FE-IJ-WX-RQ-DC-KL-YZ-PO-BA-MN-ab$. The pillar (around the grey 4_1 axis in its centre) viewed along its axis can be seen in Fig. 7(c). The helical line of the protein packing can be traced through the centres (mean coordinate) of the main-chain atoms of each dimer in the pillar. Each dimer is rotated 90° counterclockwise around the helical axis and translated by $1/28$ of the c parameter with respect to the previous point. This helical line (black in Figs. 7a and 7b) has a pitch of $c/7$, *i.e.* it is commensurate with the c axis (has seven periods in one c repeat) and runs as a smooth wave over the Hyp-1 dimers from one unit cell to the next.

The square shape of the unit-cell base and the highly pseudo-tetragonal character of the arrangement of Hyp-1 molecules are conducive to 'erroneous' packing of the Hyp-1 columns in different unit cells without significant distortions or dislocations in the crystal. This explains the occurrence of tetartohedral twinning, in which individual domains of the crystal are related by fourfold rotation around the long cell axis.

The pseudosymmetry of the packing of the Hyp-1 molecules strongly influences the intensity of diffraction. This is visible not only in the sevenfold modulation illustrated in Fig. 2(c), but also in the values of the structure factors related by the pseudo-tetragonal symmetry. Since the crystal of Hyp-1-ANS was perfectly tetartohedrally twinned, the measured intensities I_{obs} conform to 422 symmetry with an R_{merge} of 7.5%. To eliminate the effect of twinning, R_{merge} was also calculated using I_{calc} values obtained after refinement and this value was 26%, significantly less than the value of about 50% usually obtained for merging data in the wrong symmetry.

Normally, the R factors resulting from structure refinement against merohedrally twinned data are lower than expected for nontwinned crystals; whereas a completely wrong model with randomly positioned atoms gives an R factor of 58% for untwinned crystals (Wilson, 1950), for hemihedrally twinned crystals this value is 41% (Murshudov, 2011). From this perspective, the R and R_{free} values of 22.3 and 27.8%, respectively, which would be quite normal for a 'healthy' structure at 2.43 Å resolution, might seem somewhat high for a highly twinned crystal. However, the analysis of Murshudov (2011) corresponds to twinned structures with random distributions of atoms in the unit cell. Contrary to this assumption, the structure of Hyp-1-ANS is highly pseudosymmetric, with atoms distributed in a nearly tetragonal fashion, despite the true monoclinic $C2$ space group. As a result of this pseudo-

tetragonal arrangement, the reflections related by 422 point-group symmetry operations have related intensities, as illustrated by the above R_{merge} of 26% calculated using I_{calc} , *i.e.* corresponding to pseudosymmetric but untwinned data. The F_{calc} statistics are opposite to those expected for twinning, with larger than normal fractions of very weak and very strong data, as is characteristic for tNCS. The twin laws (which also correspond to 422 symmetry) therefore mix reflections that are similar by pseudosymmetry, rather than mixing unrelated contributions from different twin domains. This explains why various twinning criteria, including the L -test, did not clearly indicate the presence of a very high degree of twinning in the experimental set of intensities. For this reason, for twinned but highly pseudosymmetric crystals the refinement R factor will not be expected to be much lower than for ordinary structures, and in this context the value of $\sim 22\%$ for such a huge structure as Hyp-1-ANS should be considered to be quite normal. The correctness of the refined model is further confirmed by the distributions of the scale (close to ~ 1) and R factors (inversely related to average reflection layer intensity) in seven $n = \text{mod}(l, 7)$ groups calculated in different resolution ranges (Supplementary Table S1). Also, the CC_{work} and CC_{free} coefficients, when compared with CC^* , show the expected behaviour, with slight fluctuation in pace with the overall intensity of the subsets considered (Supplementary Table S2).

3.8. Comparison with other PR-10 proteins

3.8.1. Superpositions of the present Hyp-1 models. Structural comparisons of the 28 Hyp-1 models from the present structure show that they are all very similar. In particular, there are no meaningful differences between the C^α traces of the Hyp-1 molecules that are fully occupied by ANS and those without any ligand. For example, the C^α r.m.s.d. for chains K (three ANS ligands) and T (no ligands) is 0.41 Å, *i.e.* it is very similar to the value of 0.46 Å for the A/K pair with both chains fully occupied by ANS. This illustrates that there is no conformational adaptation of the Hyp-1 framework upon ligand binding, at least for ligands such as ANS.

3.8.2. Comparison with the unliganded structure of Hyp-1. The present models of Hyp-1 are also very similar to the previously reported ligand-free form (PDB entry 3ie5; Michalska *et al.*, 2010), with C^α r.m.s.d. values of ~ 0.6 Å. In a structural superposition, one notes that the L5 and $\alpha 2$ elements of chain A of the PDB entry 3ie5 are tilted toward the cavity when compared with chain B from the same structure or with, for example, chain K of the present structure, but in general, in agreement with the above conclusion, there are no clear manifestations of structural adaptability upon ANS binding. It should be noted, however, that the formally ligand-free structure with the PDB code 3ie5 in fact has PEG molecules in the binding cavity. Interestingly, the PEG molecules occupy similar sites as ANS ligands 1, 2 and 3 in the present structure, suggesting conservation of these Hyp-1 binding sites. Also, the residues responsible for ligand interactions (< 3.2 Å) in the PDB entry 3ie5, Lys8 and Lys33 of

Table 5

C^α r.m.s.d. values between chain *K* of the present structure and PR-10 models (identified by their PDB codes) with a large hydrophobic void (I) or a small cavity with one entrance (II).

	Protein	R.m.s.d. (Å)
I	4a80 (Bet v 1)	1.57
	1icx (LIPR-10.1A)	1.83
	1ifv (LIPR-10.1B)	1.94
	1xdf (LIPR-10.2A)	2.08
	2qim (LIPR-10.2B)	1.74
	1txc (SPE16)	1.56
II	2flh (VrPhBP)	2.36
	4q0k (MtPhBP)	1.48
	4jhg (MtN13)	1.93

chain *A* and Arg27 and Gln35 of chain *B*, are the same as those involved in ANS binding (Table 3).

3.8.3. Comparison of Hyp-1 with other PR-10 models. The structure of the Hyp-1–ANS complex reveals an interesting location of ligand-binding sites that is not found in other PR-10 proteins. The structures of PR-10 complexes reported to date have either a huge hydrophobic cavity which spans the entire space between the E1 and E2 entrances or have a small cavity with only one entrance, E1. The former group, represented by proteins such as the birch allergen Bet v 1 (e.g. PDB entry 4a80; Kofler *et al.*, 2012), PR-10 isoforms from yellow lupin (PDB entries 1icx, 1ifv, 1xdf and 2qim; Biesiadka *et al.*, 2002; Pasternak *et al.*, 2005; Fernandes *et al.*, 2008) or SPE16 from jicama (PDB entry 1txc; F. Wu, Z. Wei, Z. Zhou & W. Gong, unpublished work), can accommodate more than two ligand molecules with many hydrophobic contacts, whereas the latter group, represented by phytohormone-binding proteins (PhBP) from *Vigna radiata* (PDB entry 2flh; Pasternak *et al.*, 2006) and *M. truncatula* (PDB entry 4q0k; Ruszkowski *et al.*, 2014) and by *M. truncatula* nodulin 13 (PDB entry 4jhg; Ruszkowski *et al.*, 2013), usually bind only one ligand molecule, typically *via* hydrogen bonding. The two internal binding sites of Hyp-1, each with a separate entrance, are a novelty that is reported for the PR-10 proteins for the first time. Also, the deep surface-invagination binding pocket 3 is a novel feature. The C^α r.m.s.d. values between chain *K* of the present structure and PDB entries belonging to the two PR-10 groups mentioned above are rather high (typically 1.8 Å or more) and are similar for both groups (Table 5).

3.8.4. ANS and other PR-10 ligands. A growing number of crystal structures of small-molecule complexes of PR-10 proteins underscore their ability to bind various physiologically important molecules such as cytokinins, gibberellins, abscisic acid, steroids or flavonoids. These accumulating observations need to be verified in solution to eliminate the possibility of crystallographic artifacts and to characterize the complexes kinetically. ANS as a fluorescence probe, with its aromatic ring and small size, is an excellent mimic of the above natural ligands for such studies.

3.8.5. Comparison of ANS binding in PR-10 complexes. To date, two other PR-10 proteins have been crystallized in complex with ANS, namely isoforms a (PDB entry 4a80) and j (PDB entry 4a8v) of Bet v 1 from birch pollen (Kofler *et al.*, 2012), with one ANS molecule in the same position near the

E2 entrance to the cavity (corresponding roughly to the present site 2), and SPE16 from jicama with two ANS molecules near the E1 entrance (corresponding roughly to the present site 1), which was deposited in the PDB (as entry 1txc) without publication. Superposition of those two structures with Hyp-1–ANS (represented by chain *L*) shows that all three potential binding sites are only occupied in Hyp-1. Moreover, in the case of Bet v 1, additional structural data revealed that natural ligands are bound in a binding site that is not occupied by ANS (Kofler *et al.*, 2012). Mapping of the binding cavities with van der Waals surfaces (Figs. 10*a* and 4*b*) shows that only in Hyp-1 are they structurally well defined and distinct, which is of advantage in the interpretation of ADA results, as no direct interactions can be expected between ligands in different binding sites.

Structural alignment of Hyp-1 (chain *L*) with PDB entries 1txc and 4a80 (Fig. 10*b*), with highlighting of the residues involved in ANS contacts (<3.2 Å), shows that binding site 1 of Hyp-1 has no common residues with PDB entry 1txc. Intriguingly, the conserved residues Lys33 and Tyr150 that form the vice of Hyp-1 site 3 make no ligand interactions in the two other structures.

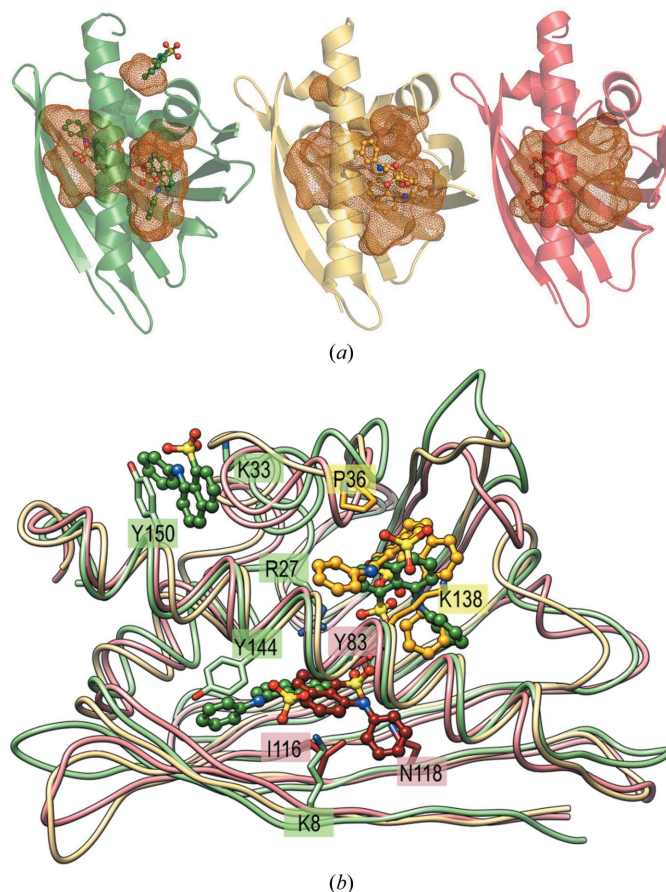


Figure 10
(*a*) ANS molecules within the differently shaped cavities (shown in van der Waals mesh representation) of the present Hyp-1 chain *L* (green), 1txc (yellow) and 4a80 (red). All the available sites are probed by the ligand only in the Hyp-1 complex. (*b*) Structural alignment of these protein chains (colour coded as in *a*), with highlighting of residues involved in ANS contacts (<3.2 Å).

4. Conclusions

A co-crystallization experiment produced tetartohedrally twinned, highly pseudosymmetric Hyp-1–ANS crystals with a modulated superstructure. The modulation is manifested by intensity fluctuations in reciprocal space, with crests at $l = 7n$ and $l = 7n \pm 3$. In direct space, a group of four Hyp-1 molecules (with pseudotetragonal packing) is sevenfold repeated along **c**. Since the modulation appears to be commensurate, the structure could be successfully refined and interpreted in an expanded (sevenfold along **c**) supercell. Because of the severe twinning, the structure was solved by MR using a tNCS-corrected ML algorithm in triclinic symmetry searching for 56 protein molecules, and the correct space group (*C2*) was figured out (in reciprocal space) by analyzing the *P1* solution. The final model is of high quality and reveals an unusual mode of ligand binding consisting of two internal sites and a deep pocket on the surface of the Hyp-1 molecule. The 1:3 complex was characterized in solution by fluorometric and calorimetric measurements. In addition to 60 protein-docked ligands, there are 29 interstitial ANS molecules distributed in a pattern that violates the arrangement of the protein molecules and is likely to be the generator of structural modulation. In particular, the tNCS-related Hyp-1 molecules are found closer together whenever there is an ANS molecule linking them. Twinning detection is very difficult in the presence of tNCS and is further complicated by additional rotational pseudo-symmetry (Lebedev *et al.*, 2006; Zwart *et al.*, 2008). The strength of twinning tests could be analyzed without ambiguity, as the twinning in this case is noncontroversial because of the prohibited symmetry displayed by the diffraction pattern.

Acknowledgements

Financial support for this project was provided by the European Union within the European Regional Developmental Fund and by the Polish Ministry of Science and Higher Education (grant No. NN 301 003739) and National Science Center (2013/10/M/NZ1/00251). RJR was supported by a Principal Research Fellowship from the Wellcome Trust (grant No. 082961/Z/07/Z). ZD was supported in part by the Intramural Research Program of the National Cancer Institute, Center for Cancer Research. The raw images are available from the authors (ZD; zdaute@anl.gov) on request.

References

Afonine, P. V., Grosse-Kunstleve, R. W., Urzhumtsev, A. & Adams, P. D. (2009). *J. Appl. Cryst.* **42**, 607–615.
 Allen, F. H. (2002). *Acta Cryst.* **B58**, 380–388.
 Bais, H. P., Vepachedu, R., Lawrence, C. B., Stermitz, F. R. & Vivanco, J. M. (2003). *J. Biol. Chem.* **278**, 32413–32422.
 Biesiadka, J., Bujacz, G., Sikorski, M. M. & Jaskolski, M. (2002). *J. Mol. Biol.* **319**, 1223–1234.
 Bradford, M. M. (1976). *Anal. Biochem.* **72**, 248–254.
 Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. & Richardson, D. C. (2010). *Acta Cryst.* **D66**, 12–21.

Chwastyk, M., Jaskolski, M. & Cieplak, M. (2014). *FEBS J.* **281**, 416–429.
 Cody, V. & Hazel, J. (1977). *Acta Cryst.* **B33**, 3180–3184.
 Cohen, G. H. (1997). *J. Appl. Cryst.* **30**, 1160–1161.
 Dauter, Z. & Jaskolski, M. (2010). *J. Appl. Cryst.* **43**, 1150–1171.
 Diederichs, K. & Karplus, P. A. (1997). *Nature Struct. Mol. Biol.* **4**, 269–275.
 Engh, R. A. & Huber, R. (1991). *Acta Cryst.* **A47**, 392–400.
 Evans, P. (2006). *Acta Cryst.* **D62**, 72–82.
 Fernandes, H., Michalska, K., Sikorski, M. & Jaskolski, M. (2013). *FEBS J.* **280**, 1169–1199.
 Fernandes, H., Pasternak, O., Bujacz, G., Bujacz, A., Sikorski, M. M. & Jaskolski, M. (2008). *J. Mol. Biol.* **378**, 1040–1051.
 Gajhede, M., Osmark, P., Poulsen, F. M., Ipsen, H., Larsen, J. N., Joost van Neerven, R. J., Schou, C., Løwenstein, H. & Spangfort, M. D. (1996). *Nature Struct. Biol.* **3**, 1040–1045.
 Gasymov, O. K. & Glasgow, B. J. (2007). *Biochim. Biophys. Acta*, **1774**, 403–411.
 Jaskolski, M. (2013). *Acta Cryst.* **D69**, 1865–1866.
 Kofler, S., Asam, C., Eckhard, U., Wallner, M., Ferreira, F. & Brandstetter, H. (2012). *J. Mol. Biol.* **422**, 109–123.
 Košuth, J., Hrehorová, D., Jaskolski, M. & Čellárová, E. (2013). *Plant. Cell. Tiss. Org. Cult.* **114**, 207–216.
 Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
 Lebedev, A. A., Vagin, A. A. & Murshudov, G. N. (2006). *Acta Cryst.* **D62**, 83–95.
 Marković-Housley, Z., Degano, M., Lamba, D., von Roepenack-Lahaye, E., Clemens, S., Susani, M., Ferreira, F., Scheiner, O. & Breiteneder, H. (2003). *J. Mol. Biol.* **325**, 123–133.
 Matulis, D. & Lovrien, N. (1998). *Biophys. J.* **74**, 422–429.
 McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Cryst.* **40**, 658–674.
 Michalska, K., Fernandes, H., Sikorski, M. M. & Jaskolski, M. (2010). *J. Struct. Biol.* **169**, 161–171.
 Murshudov, G. N. (2011). *Appl. Comput. Math.* **10**, 250–261.
 Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* **D67**, 355–367.
 Padilla, J. E. & Yeates, T. O. (2003). *Acta Cryst.* **D59**, 1124–1130.
 Pasternak, O., Biesiadka, J., Dolot, R., Handschuh, L., Bujacz, G., Sikorski, M. M. & Jaskolski, M. (2005). *Acta Cryst.* **D61**, 99–107.
 Pasternak, O., Bujacz, G. D., Fujimoto, Y., Hashimoto, Y., Jelen, F., Otlewski, J., Sikorski, M. M. & Jaskolski, M. (2006). *Plant Cell*, **18**, 2622–2634.
 Read, R. J., Adams, P. D. & McCoy, A. J. (2013). *Acta Cryst.* **D69**, 176–183.
 Ruszkowski, M., Sliwiak, J., Ciesielska, A., Barciszewski, J., Sikorski, M. & Jaskolski, M. (2014). *Acta Cryst.* **D70**, 2032–2041.
 Ruszkowski, M., Szpotkowski, K., Sikorski, M. & Jaskolski, M. (2013). *Acta Cryst.* **D69**, 2365–2380.
 Sels, J., Mathys, J., De Coninck, B. M., Cammue, B. P. & De Bolle, M. F. (2008). *Plant Physiol. Biochem.* **46**, 941–950.
 Sheard, L. B. & Zheng, N. (2009). *Nature (London)*, **462**, 575–576.
 Sheldrick, G. M. (2008). *Acta Cryst.* **A64**, 112–122.
 Sliwiak, J., Jaskolski, M., Dauter, Z., McCoy, A. J. & Read, R. J. (2014). *Acta Cryst.* **D70**, 471–480.
 Wagner, T. & Schönleber, A. (2009). *Acta Cryst.* **B65**, 249–268.
 Weber, L. D. & Tulinsky, A. (1980). *Acta Cryst.* **B36**, 611–614.
 Wilson, A. J. C. (1950). *Acta Cryst.* **3**, 397–398.
 Zwart, P. H., Grosse-Kunstleve, R. W. & Adams, P. D. (2005). *CCP4 Newsl. Protein Crystallogr.* **43**, contribution 7.
 Zwart, P. H., Grosse-Kunstleve, R. W., Lebedev, A. A., Murshudov, G. N. & Adams, P. D. (2008). *Acta Cryst.* **D64**, 99–107.

PUBLICATION IV

Hyp-1 protein from St John's wort as a PR-10 protein

JOANNA ŚLIWIAK¹, ZBIGNIEW DAUTER², MARIUSZ JASKOLSKI^{1,3*}

¹Institute of Bioorganic Chemistry, Polish Academy of Sciences, Poznań, Poland

²Synchrotron Radiation Research Section, National Cancer Institute, Argonne National Laboratory, Argonne, USA

³Faculty of Chemistry, Adam Mickiewicz University, Poznań, Poland

* Corresponding author: mariuszj@amu.edu.pl

Abstract

PR-10 proteins form a large subclass of plant pathogenesis-related proteins that are expressed in response to harmful environmental factors in a wide range of species. Although their function is still not clear, structural data suggest that their characteristic internal hydrophobic cavity can bind relevant plant small-molecule mediators. Hyp-1 from St John's wort (*Hypericum perforatum*), initially proposed as a catalyst for the biosynthesis of hypericin, was eventually shown to share sequence similarity and a folding pattern with PR-10 proteins. The crystal structure of Hyp-1 in complex with fluorescent probe ANS reveals three distinct and separated binding sites that are unique among PR-10 proteins. The structure can provide guidance in our quest for the true physiological ligands of Hyp-1.

Introduction

One of several plant defense mechanisms is based on a number of genes that are expressed in response to harmful abiotic and biotic factors, such as drought, salinity, cold, oxidative stress or pathogens. The pathogenesis-related (PR) proteins encoded by these genes have been divided into 17 classes, based on function, amino acid sequence, and biochemical activity. This classification groups proteins with common function, such as chitinases, defensins, glucanases, peroxidases, proteases, etc. Class 10 (PR-10), however, groups proteins with a well-established canonical fold but without a clearly determined biological function (Fernandes et al., 2013). PR-10 proteins are small (~18 kDa), mostly cytosolic and have acidic isoelectric point (pI). Their accumulation in roots, seeds and senescent leaves indicates their role in signaling and/or development. Indeed, studies on abscisic acid (ABA) receptors in *Arabidopsis thaliana* revealed that the cytosolic ABA-binding moiety has a PR-10 fold (Umezawa et al., 2010). Also, other structural studies demonstrate the ability of PR-10 proteins to bind such plant mediators as cytokinins, flavonoids, gibberellins or steroids.

Common fold of PR-10 proteins

The PR-10 folding canon consists of a seven-stranded antiparallel β -sheet wrapped around a long and elas-

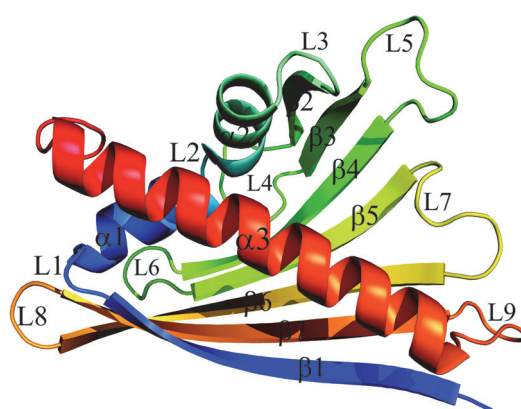


Fig. 1. Secondary structure elements of PR-10 proteins, shown using the example of the Hyp-1 protein (PDB accession code 3IE5; Michalska et al., 2010)

tic C-terminal helix α 3, which shows not only conformational but also sequence variability. This "baseball glove" fold creates a large hydrophobic void in the core which is quite unexpected in a rather small globular protein, and suggests a storage function. The hydrophobic cavity has two clearly discernible main entrances: entrance E1 is created by loops and a helix α 3, while entrance E2 is formed between the same α -helix and the first strand (β 1) of the β -sheet (Fig. 1).

Ligand binding

The number of NMR and X-ray crystal structures of PR-10 proteins is systematically growing. The structures

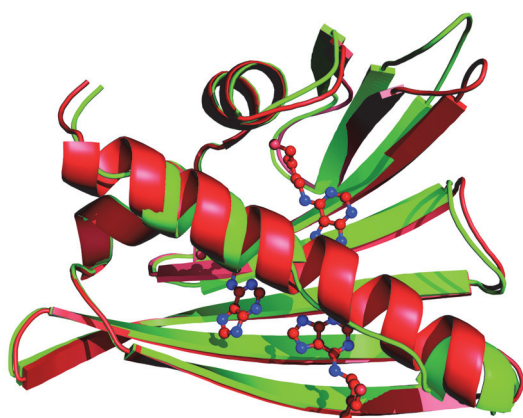


Fig. 2. C α superposition of the yellow lupine LIPR-10.1A protein structure in its free form (green) on the structure from its complex with *trans*-zeatin (red) (unpublished). The *trans*-zeatin molecules are shown as ball-and-stick models

show the ability of PR-10 proteins to bind variety of physiologically important molecules. In addition, the availability of accurate structures, both in the free form as well as in complexes with ligands, offers a possibility to study the adaptability of the PR-10 fold in binding plant hormones. For instance, the example of the LIPR-10.1A protein from *Lupinus luteus* in its free form (Biesiadka et al., 2002) and in complex with *trans*-zeatin (unpublished) can be used to explain the binding mode of *trans*-zeatin (Fig. 2): in the free form the N-terminal part of helix α 3 is disordered and the L7 loop is moved to the outside to open entrance E1. Zeatin binding closes this gate and additionally orders the α 3 helix.

Analysis of the structures of different PR-10 protein complexes with ligands shows that the interactions with the protein interior are mostly hydrophobic with only sporadic hydrogen bonds formed by hydrophilic side chains, which are rare in the cavity. In such complexes as the *Vigna radiata* Cytokinin-Specific Binding Protein (VrCSBP) with *trans*-zeatin (PDB code 2FLH; Pasternak et al., 2006), *Medicago truncatula* Nodulin 13 (MtN13) with *trans*-zeatin (4JHG; unpublished) or Phytohormone Binding Protein (PBP) from the same organism with gibberellic acid (3US7; unpublished data) there is only one E1 entrance to the cavity, which is smaller than in complexes of PR-10 proteins from yellow lupine, such as LIPR-10.1A/zeatin (unpublished) and LIPR-10.2B/zeatin (2QIM; Fernandes et al., 2008) or complexes of the white birch allergen Bet v 1 with deoxycholate, kinetin and naringenin (PDB codes 4A83, 4A85, 4A87; Kofler et al., 2012). In the latter group, the hydrophobic void spans

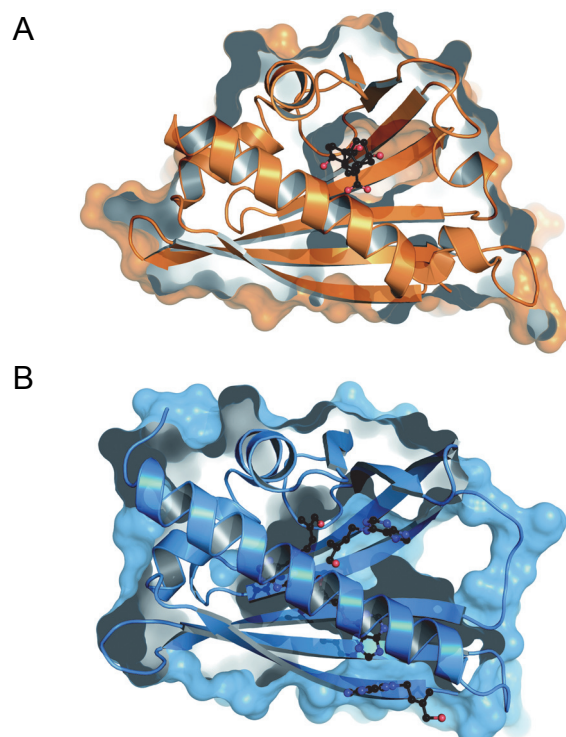


Fig. 3. Two types of shapes of the internal cavity found in PR-10 proteins. A) A cavity with only one entrance (E1), illustrated by the structure of the PBP/gibberellic acid complex (PDB accession code 3US7; unpublished). B) A cavity in the form of an elongated hydrophobic tunnel connecting entrances E1 and E2, illustrated by the structure of the LIPR-10.2B/*trans*-zeatin complex (PDB: 2QIM; Fernandes et al., 2008)

the space from E1 to E2 and can accommodate more than two ligand molecules (Fig. 3).

The Hyp-1 protein story

St John's wort is a millennia-sanctioned medicinal plant, recently rediscovered as a herbal remedy for mild-to-moderate depression thanks to the high levels of its pharmacologically active ingredient, hypericin. Hyp-1 had been proposed to catalyze the biosynthesis of hypericin from emodin *in vivo* (Bais et al., 2003); however, the reaction could not be reproduced (Michalska et al., 2010). Moreover, the highest transcription level of Hyp-1 was found in roots, not leaves (Kosuth et al., 2013). Hyp-1, with a molecular mass of 17.8 kDa and pI of 5.54, shares ~50% sequence similarity with the PR-10 protein family (Fernandes et al., 2008) and its structural classification in the PR-10 class was confirmed by the crystal structure of Hyp-1 (Michalska et al., 2010), although the root-mean-square deviation for C α atom superposition on different PR-10 proteins is about 2 Å. Co-crystallization

of Hyp-1 with the purported substrate (emodin) or product (hypericin), as well as calorimetric assays did not confirm their interaction with the protein either.

ANS as a probe of the ligand binding sites of PR-10 proteins

ANS (8-anilino-1-naphthalene sulfonate) is a fluorescent probe used to investigate hydrophobic ligand-binding sites of macromolecules in ligand-replacement tests coupled with fluorescence measurements. These measurements, which can identify potentially relevant ligands, are based on the ability of ANS to fluoresce (with blue light) in hydrophobic environments (Kundu et al., 2002). The fluorescence is significantly decreased in hydrophilic environments with a bathochromic shift of the fluorescence peak. There are two structures of PR-10 proteins in complex with ANS in the PDB; namely, for proteins from jicama (SPE-16) (PDB: 1TXC; unpublished) and from white birch (Bet v 1) (4A80; Kofler et al., 2012). In the latter case, the structure helped to explain the anomalous ANS displacement data at the molecular level.

Hyp-1 complex with ANS

Our recently determined crystal structure of Hyp-1 in complex with ANS (unpublished), refined at 2.4 Å resolution, confirms the PR-10 fold of the protein and reveals three principal ligand-binding sites (Fig. 4). Noticeable, even at first glance, is the fact that the binding sites are well separated and defined, quite unlike in the previously characterized ligand complexes of PR-10 proteins. The high specificity of the Hyp-1 binding sites is unquestionable, as the ANS molecules are observed with highly conserved conformation and orientation in numerous copies of the 28 Hyp-1 molecules found in the asymmetric unit (!). The binding site designated as A1 has its entrance at E1 and the ANS-1 molecule there is stabilized by a hydrogen bond with R27. The binding site designated as A2 has its entrance at E2 and is based on hydrophobic interactions. The third site, designated as A3, has the shape of a deep pocket formed by invagination of the protein surface, created by tyrosine and lysine residues which bind the ligand mainly through stacking interactions. The superposition of Hyp-1 from the ANS complex with the free form of the protein does not show any significant changes, particularly of helix α 3. This could be explained by the fact that the previously

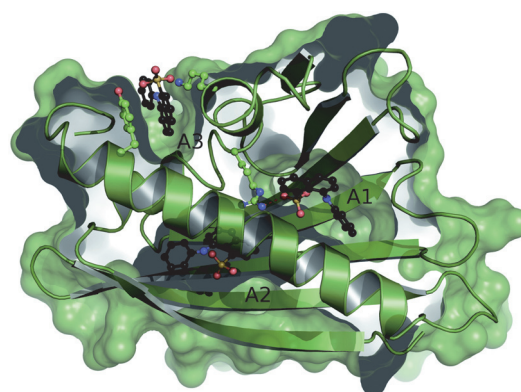


Fig. 4. A crystallographic model of the Hyp-1/ANS complex, illustrating the three separate ligand binding sites of the protein

determined structure of unliganded Hyp-1 (Michalska et al., 2010) has in fact several PEG molecules (from the crystallization buffer) bound non-specifically in the cavity, and this could be sufficient to force the protein to adopt a cargo-bound conformation. Our difficulties with obtaining single crystals of Hyp-1, and of PR-10 proteins in general, in a strictly ligand-free form might be interpreted as indicating that in its unliganded form the PR-10 fold has unstable or disordered structural elements. This speculation is corroborated by the case of the LIPR-10.1A protein mentioned above.

Conclusion

Hyp-1 has been classified as a PR-10 protein both on the basis of a rather high level of sequence similarity and its canonical PR-10 fold. However, the shapes of the ANS binding sites and their evident separation indicate that Hyp-1 is a rather unique protein in the PR-10 category. The crystal structure of the Hyp-1/ANS complex has revealed three well-defined and unique ligand binding sites. Moreover, the structure provides a valuable atomic-level basis for a structural interpretation of the fluorescent displacement assays, which can help in identifying the natural ligands of this protein.

Acknowledgments

Financial support for the project was provided by the European Union within the framework of the European Regional Development Fund and by the Polish Ministry of Science and Higher Education (grant No. NN 301 003739).

References

Bais H., Vepachedu R., Lawrence C., Stermitz F., Vivanco J. (2003) *J. Biol. Chem.* 278: 32413-32422.

- Biesiadka J., Bujacz G., Sikorski M., Jaskolski M. (2002) *J. Mol. Biol.* 319: 1223-1234.
- Fernandes H., Michalska K., Sikorski M., Jaskolski M. (2013) *FEBS J.* 280: 1169-1199.
- Fernandes H., Konieczna M., Kołodziejczyk R., Bujacz G., Sikorski M., Jaskolski M. (2008) *Acta Cryst.* F64: 405-408.
- Fernandes H., Pasternak O., Bujacz G., Bujacz A., Sikorski M., Jaskolski M. (2008) *J. Mol. Biol.* 378: 1040-1051.
- Kofler S., Asam C., Eckhard U., Wallner M., Ferreira F., Brandstetter H. (2012) *J. Mol. Biol.* 422: 109-123.
- Kosuth J., Jaskolski M., Cellarova E. (2013) *Plant Cell Tiss. Org.*, in press.
- Kundu B., Guptasarma P. (2002) *Biochem. Biophys. Res. Commun.* 293: 572-577.
- Michalska K., Fernandes H., Sikorski M., Jaskolski M. (2010) *J. Struct. Biol.* 169: 161-171.
- Pasternak O., Bujacz G., Fujimoto Y., Hashimoto Y., Jelen F., Otlewski J., Sikorski M., Jaskolski M. (2006) *Plant Cell.* 18: 2622-2634.
- Umezawa T., Nakashima K., Miyakawa T., Kuromori T., Tanokura M., Shinozaki K., Yamaguchi-Shinozaki K. (2010) *Plant Cell Physiol.* 51: 1821-1839.

PUBLICATION V



Crystal Structure of Hyp-1, a *Hypericum perforatum* PR-10 Protein, in Complex with Melatonin

Joanna Sliwiak¹, Zbigniew Dauter² and Mariusz Jaskolski^{1,3*}

¹ Center for Biocrystallographic Research, Institute of Bioorganic Chemistry, Polish Academy of Sciences, Poznan, Poland,

² Synchrotron Radiation Research Section, National Cancer Institute, Argonne National Laboratory, Argonne, IL, USA,

³ Department of Crystallography, Faculty of Chemistry, Adam Mickiewicz University in Poznań, Poznań, Poland

OPEN ACCESS

Edited by:

Gregory Franklin,
Polish Academy of Sciences, Poland

Reviewed by:

Kathrin Schrick,
Kansas State University, USA
Doriano Lamba,
Consiglio Nazionale delle Ricerche,
Italy

*Correspondence:

Mariusz Jaskolski
mariuszj@amu.edu.pl

Specialty section:

This article was submitted to
Plant Metabolism
and Chemodiversity,
a section of the journal
Frontiers in Plant Science

Received: 19 February 2016

Accepted: 01 May 2016

Published: 18 May 2016

Citation:

Sliwiak J, Dauter Z and Jaskolski M
(2016) Crystal Structure of Hyp-1,
a *Hypericum perforatum* PR-10
Protein, in Complex with Melatonin.
Front. Plant Sci. 7:668.
doi: 10.3389/fpls.2016.00668

Hyp-1, a PR-10-fold protein from *Hypericum perforatum*, was crystallized in complex with melatonin (MEL). The structure confirms the conserved protein fold and the presence of three unusual ligand binding sites, two of which are internal chambers (1,2), while the third one (3) is formed as an invagination of the protein surface. The MEL ligand in site 1 is well defined while that in site 3 seems to be rotating between the side chains of Lys33 and Tyr150 that act as a molecular vise. The patch of electron density in site 2 does not allow unambiguous modeling of a melatonin molecule but suggests a possible presence of its degradation product. This pattern of ligand occupation is reproducible in repeated crystallization/structure determination experiments. Although the binding of melatonin by Hyp-1 does not appear to be very strong (for example, MEL cannot displace the artificial fluorescence probe ANS), it is strong enough to suggest a physiological role of this interaction. For example, *trans*-zeatin, which is a common ligand of PR-10 proteins, does not overcompete melatonin for binding to Hyp-1 as it does not affect the crystallization process of the Hyp-1/MEL complex, and among a number of potential natural mediators tested, melatonin was the only one to form a crystalline complex with Hyp-1 with the use of standard crystallization screens. Hyp-1 is the second protein in the Protein Data Bank for which melatonin binding has been demonstrated crystallographically, the first one being human quinone reductase.

Keywords: pathogenesis-related protein, PR-10, phytohormone, ligand binding, cytokinin

INTRODUCTION

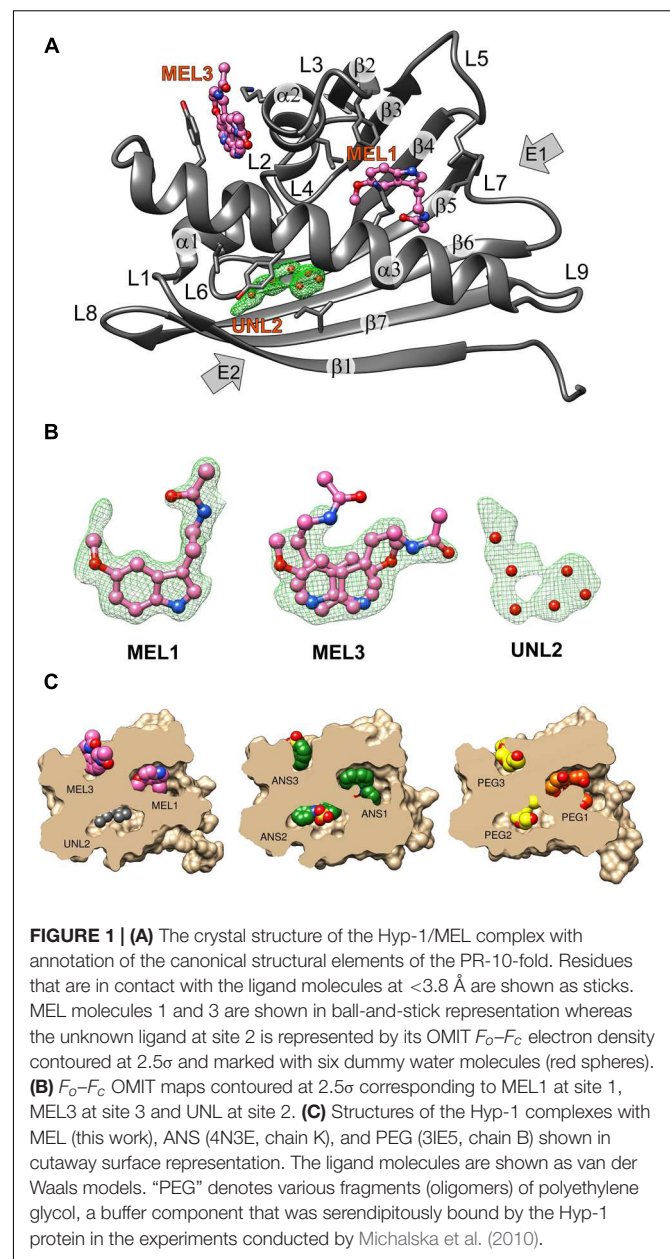
Hyp-1, the protein product (comprised of 159 residues) of the *hyp-1* gene in *Hypericum perforatum* has a picturesque history. It was first described as the enzyme catalyzing the biosynthesis of the pharmacological ingredient of this plant, the dianthrone hypericin, from two molecules of emodin (Bais et al., 2003). That study, however, could never be replicated and instead the crystal structure of the Hyp-1 protein revealed the canonical PR-10-fold (Michalska et al., 2010) strongly suggesting classification in class 10 of the superfamily of plant Pathogenesis-Related (PR-10) proteins. The latter hypothesis was corroborated by genetic data, which showed that *hyp-1* has gene structure analogous to typical *pr-10* genes (Kosuth et al., 2013), but it has to be underlined that Hyp-1 has not been demonstrated so far to be involved in stress response of *H. perforatum*. With regard to the localization in the plant, it was shown that Hyp-1 mRNA expression occurs in all organs of *Hypericum* seedlings with the highest levels in roots (Kosuth et al., 2007), whereas

immunofluorescence assays of plantlets revealed wide distribution of the Hyp-1 protein in different tissues, including roots, stem, and leaves (Qian et al., 2012).

The characteristic PR-10-fold (Fernandes et al., 2013), also known as the Betv1 fold according to the first protein from this class, a birch (*Betula verucosa*) pollen allergen to have its crystal structure determined (Gajhede et al., 1996), consists of a large seven-stranded antiparallel β -sheet forming a baseball-glove grip over a long C-terminal helix α 3, which is the most variable element of the PR-10 structure (Biesiadka et al., 2002; Pasternak et al., 2006). The consecutive β -strands are connected by loops, except for strands β 1 (first) and β 2 (last) at the edges of the β -sheet, which are connected by a V-shaped fork of two α -helices (α 1 and α 2) that provides a support for the C-terminal end of helix α 3. At its N-terminal end, helix α 3 is connected to the protein scaffold by loop L9. A conspicuous feature of the PR-10-fold is a large hydrophobic cavity formed between the main structural elements, i.e., the β -sheet and helix α 3, with the participation of other secondary structures, such as the odd-numbered loops (L3, L5, L7, L9), which are the fingertips of the gripping hand (Figure 1A). The cavity has two entrances connecting it to the outer environment: E1 surrounded by the odd numbered loops (L3, L5, L7) and helix α 3, and entrance E2 located between helix α 3 and strand β 1. Despite the hollow core, the PR-10 proteins are robust, resistant to proteases and have mechanical stability that even surpasses that of average globular proteins (Chwastyk et al., 2014). The properties, size and shape of the internal cavity are mostly modulated by the character of the α 3 helix.

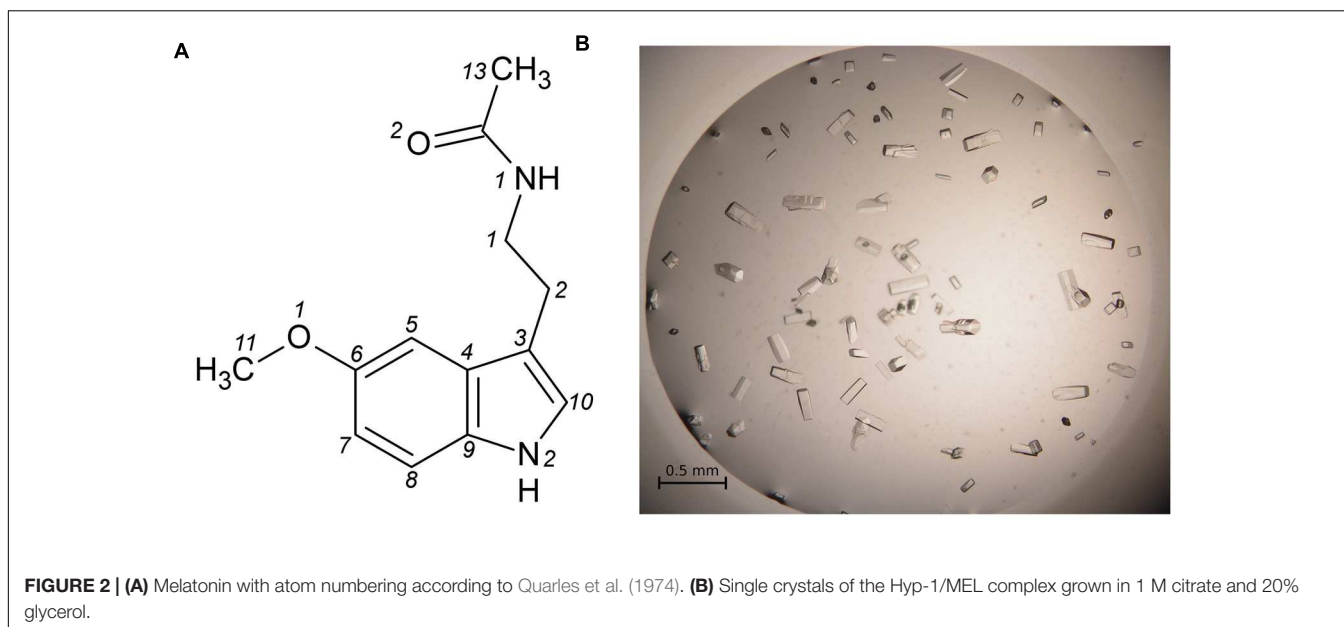
The presence of such an intriguing cavity has led to the hypothesis that PR-10 proteins might have evolved in plants to bind/store/transport important small-molecule mediators, such as plant hormones (Fernandes et al., 2013). Along these lines, a number of PR-10 (or at least PR-10-fold) proteins have been characterized structurally in complex with phytohormones (or their analogs), such as cytokinins (Pasternak et al., 2006; Fernandes et al., 2008, 2009; Kofler et al., 2012; Ruszkowski et al., 2013; Sliwiak et al., 2016), gibberellin (Ruszkowski et al., 2014), brassinosteroids (Markovic-Housley et al., 2003), or abscisic acid (Sheard and Zheng, 2009). Moreover, other plant metabolites, such as flavonoids (Mogensen et al., 2002; Kofler et al., 2012; Casañal et al., 2013) or their glycosylated forms (Seutter von Loetzen et al., 2014, 2015), are also bound by PR-10 proteins.

On the list of recognized plant hormones, melatonin (*N*-acetyl-5-methoxytryptamine, MEL, Figure 2A) is a relatively new addition. Apart from the discovery of the presence of this conservative molecule in plants (Dubbels et al., 1995), relatively little has been learned about phytomelatonin function over the last decade. Melatonin appears to regulate plant growth in an auxin-like manner, regulate the response to photoperiod and increase tolerance to abiotic stress. It is also one of the most efficient antioxidants (Arnao and Hernández-Ruiz, 2015). *H. perforatum*, alongside *Tanacetum parthenium* or the Chinese herb *Scutellaria bicacensis*, appears to contain very high MEL concentrations (Murch et al., 1997) that reach 2 μ g/g of dried weight in leaves and are above 4 μ g/g in flowers. In the case of *H. perforatum*, this could be responsible for the



medicinal effects of St John's wort preparations. Apart from tissue content determinations and studies of the tryptophan-dependent biosynthetic pathway (Murch et al., 2000), studies also focused on the role of melatonin in *H. perforatum*, demonstrating that increased light intensity elevates melatonin synthesis, thus confirming its free radical scavenging function (Murch et al., 2000). It was shown that MEL is able to induce rhizogenesis (Murch et al., 2001), an observation that has been recently confirmed in other plant species (Arnao and Hernández-Ruiz, 2015).

Biophysical and kinetic studies of PR-10/phytohormone complexes are often difficult because of problems with ligand solubility, low heat effect upon binding (in calorimetry) and/or unsuitable spectroscopic properties. A frequently used assay



in such studies is ADA (ANS Displacement Assay), in which the fluorescent dye 8-anilinoanthralene-1 sulfonate (ANS) is displaced by the ligand of interest (Mogensen et al., 2002). Our investigations of ligand-binding properties of the Hyp-1 protein started in fact with the crystallization of a Hyp-1/ANS complex, which turned out to have a complex modulated crystal structure with as many as 28 copies of the protein in the asymmetric unit (Sliwiak et al., 2014, 2015). At the same time, that structure revealed an unprecedented among PR-10-fold proteins ligand binding mode, with two ANS molecules (at sites 1, 2) bound in two tight internal chambers (instead of one large cavity) and another one (3) docked in a deep invagination of the protein surface.

In this work we present high resolution crystal structure of Hyp-1 in complex with melatonin, demonstrating that this physiological ligand utilizes the same internal docking sites as ANS. The Hyp-1/MEL structure is the first example reported in the Protein Data Bank (PDB) of melatonin bound to a plant protein, and the second case with any protein, the first one being human quinone reductase (Calamini et al., 2008). It is also important to stress that among many different phytohormones and hypothetical biologically relevant substrate/product molecules (e.g., emodin, hypericin) tested, melatonin was the only ligand that formed crystalline complex with the Hyp-1 protein.

MATERIALS AND METHODS

Protein Preparation, Complex Formation, and Crystallization

Hyp-1 was produced as described before (Sliwiak et al., 2015). Prior to crystallization, the protein solution was concentrated to 15 mg/ml and pre-incubated at 292 K for 1 h with 10-fold molar

excess of melatonin (Sigma–Aldrich) added from a 0.1 M stock solution in methanol, or with MEL powder. Screening for Crystal Screen, PEG/Ion I and II (Hampton Research) crystallization conditions was performed by the sitting drop vapor diffusion method against 120 μ L well solution with the use of a Mosquito Crystallization Robot. The crystallization drops were mixed from 0.2 μ L protein/ligand solution and 0.2 μ L well solution. Small crystals, which appeared the same day in 1.6 M tribasic sodium citrate, pH 6.5, were used for seeding in a gradient of PEG 400 or glycerol and tribasic sodium citrate in hanging drops. Large, prismatic crystals of dimensions 0.08 mm \times 0.08 mm \times 0.2 mm (**Figure 2B**) grew in 1 M citrate and 20% glycerol.

Competitive Crystallization Assays

Hyp-1 protein was pre-incubated for 1 h with an equimolar solution of MEL (from 0.1 M methanol stock) and *trans*-zeatin (from 0.1 M stock in DMSO), as well as with a solution mixture of MEL and ANS (from 0.1 M stock in DMSO), mixed at the following MEL:ANS ratios: 1:1, 2:1, and 3:1. The Hyp-1:MEL molar ratio was 1:10 in all conditions. Crystallization was performed in all these cases using the final growth conditions established for the crystals of the Hyp-1/MEL complex. In these competition assays, crystals were obtained only in the presence of *trans*-zeatin but they had the prismatic morphology of the Hyp-1/MEL crystals.

Data Collection, Structure Solution, and Refinement

X-Ray diffraction data extending to 1.30 Å resolution were collected at the SER-CAT 22ID beamline of the Advanced Photon Source (APS/ANL) and were processed with HKL-2000 (Otwinowski and Minor, 1997). The data were merged in space group C222₁ with R_{merge} of 5.7% (**Table 1**). For molecular replacement in Phaser (McCoy et al., 2007), the PDB model 3IE5

TABLE 1 | Data collection and refinement statistics.

Data collection	
Space group	C222 ₁
Unitcell parameters <i>a</i> , <i>b</i> , <i>c</i> (Å)	60.86, 89.64, 76.41
Beamline	SER-CAT 22ID (APS)
Wavelength (Å)	1.0000
Data collection temperature (K)	100
Resolution (Å)	30.0–1.30 (1.32–1.30) ^a
<i>R</i> _{merge} (%)	5.7 (51.4)
< <i>I</i> / <i>σ</i> >	28.9 (2.5)
CC _{1/2} /CC* (%) ^b	(86.3)/(96.2)
Completeness (%)	99.9 (99.6)
Redundancy	4.9 (3.9)
Refinement	
Resolution (Å)	25.31–1.30
Reflections work/test	49014/2630
<i>R</i> _{work} / <i>R</i> _{free} (%)	12.8/15.3
Protein/ligand/solvent/water/metal atoms	1405/51(MEL), 12(GOL), 6(UNL)/204/3
< <i>B</i> > protein/ligand/water/metal (Å ²)	22.8/45.3(MEL), 58.8(GOL), 59.8(UNL)/45.2/37.8
R.M.S.D. from ideal geometry	
Bond lengths (Å)/bond angles (°)	0.017/1.6
Ramachandran statistics (%) ^c	
Favored/outliers	98.3/0
PDB code	5I8F

^aValues in parentheses correspond to the last resolution shell. ^bCorrelation coefficients, as defined by Karplus and Diederichs (2012), given for the last resolution shell. ^cAssessed with MolProbity (Chen et al., 2010).

(Michalska et al., 2010) was used. Manual rebuilding was carried out in Coot (Emsley et al., 2010) and anisotropic maximum-likelihood refinement was carried out in phenix.refine (Afonine et al., 2012). Stereochemical restraints for the melatonin molecule were generated from the coordinates found in the CSD deposit MELATN01 of melatonin crystal structure (Quarles et al., 1974). X-Ray diffraction data collected for identical crystals, obtained upon co-crystallization with melatonin added in pulverized form or in the presence of equimolar concentration of *trans*-zeatin, extended to 1.34 Å and 1.40 Å resolution, respectively, and were also merged in the C222₁ space group with *R*_{merge} of 7.8 and 9.6%, respectively.

Other Software

For C α superpositions and R.M.S.D. calculations the ALIGN program (Cohen, 1997) was used. Figures were prepared in UCSF Chimera (Pettersen et al., 2004).

Deposition Note

Atomic coordinates and processed structure factors corresponding to the final model of the Hyp-1/melatonin complex have been deposited with the PDB under the accession code 5I8F. The corresponding raw X-ray diffraction images have been deposited in the RepOD Repository at the Interdisciplinary Centre for Mathematical and Computational Modelling (ICM) of the University of Warsaw, Poland, and are available for

download with the following Digital Object Identifier (DOI): <http://dx.doi.org/10.18150/repod.4711822>.

RESULTS AND DISCUSSION

Crystallization Trials

Hyp-1 co-crystallization experiments were carried out with phytohormones from different classes, including auxin, *trans*-zeatin, gibberellic acid, abscisic acid, and melatonin; and additionally with the flavonoid quercetin, the fluorescence probe ANS, as well as with the hypothetical substrate (emodin) and product (hypericin) molecules. All these trials were performed using the same commercial screens and with similar protein:ligand ratios as for the present complex. The ultimate result of those crystallizations was that crystalline complexes of Hyp-1 could be obtained only with ANS (Sliwiak et al., 2015) or with MEL (added in solution or in powder form). In this context, it is interesting to note that crystallography emerges as a superior approach to the detection of protein-ligand complexes when standard biophysical methods fail (Schiebel et al., 2016).

Competitive crystallization with MEL and *trans*-zeatin resulted in crystals of the same Hyp-1/MEL complex. Thus one can conclude that *trans*-zeatin does not perturb Hyp-1/MEL complex formation under the conditions of Hyp-1/MEL crystal growth.

On the other hand, the presence of ANS in the Hyp-1/MEL crystallization conditions, even at lower concentration than that used for *trans*-zeatin, suppressed the crystal growth entirely. Moreover, addition of melatonin (even at 1:1 ANS:MEL ratio) to Hyp-1/ANS crystallization conditions (Sliwiak et al., 2015) resulted in crystals of the Hyp-1/ANS complex with a new type of modulation (Sliwiak, Unpublished Data). This suggests that ANS blocks the MEL binding sites of Hyp-1 with higher affinity, explaining why it was not possible to detect any signal with MEL titration in ANS Displacement Assays (ADA) performed according to a well-established procedure (Pasternak et al., 2006). On the other hand, it has to be noted that unlike in the Hyp-1/MEL complex, in the crystal structure of the Hyp-1/ANS complex, in addition to the internal binding sites 1,2,3, there are numerous ANS molecules bound at conserved sites on the surface of the Hyp-1 protein (Sliwiak et al., 2015). Since those superficial ANS molecules (which are most likely responsible for the modulation of the crystal structure) are not exchangeable for MEL even at high melatonin concentration, they could additionally mask the ADA signal.

Overall Features of the Crystal Structure

The structure of Hyp-1 described in this work is of the highest resolution (1.30 Å) among all Hyp-1 structures available in the PDB (3IE5, 1.69 Å; 4N3E, 2.43 Å) and therefore provides the most accurate model of this protein. Moreover, as the protein was purified in reducing condition, in variance with the 3IE5 model, in the present structure there are no accidental disulfide bonds bridging the Hyp-1 molecules in the crystals structure. In agreement with this, PISA (Krissinel and Henrick, 2007) analysis did not detect any stable quaternary structure. The

solvent content of the crystal is 56.4% with Matthews volume equal to $2.82 \text{ \AA}^3/\text{Da}$. Thanks to the high resolution of the diffraction data, all atoms in the structure were refined with anisotropic atomic displacement parameters (*B*-factors). Careful examination of difference electron density maps revealed the positions of 204 water molecules as well as of two molecules of MEL (**Figure 1B**), one of which (modeled in two orientations, with average *B*-factor of 54.5 \AA^2) is most likely endowed with rotational degrees of freedom within the surface invagination, and another one (modeled in one orientation with $\langle B \rangle$ of 36.1 \AA^2) is well stabilized in the internal cavity of the protein. Two glycerol molecules with low *B*-factors were modeled at the protein surface. In addition, there are three Na^+ ions included in the model, two of which are octahedrally coordinated by the protein (one by loop L5 and another one by strand $\beta 1$ and the C-end of helix $\alpha 3$), and a third one partially occupied within the protein cavity. Within the cavity, there is also an ambiguous patch of electron density which could not be assigned to any of the components of the crystallization solution. Since there is an indication of an indole ring with a short side chain (**Figure 1B**), it could be a poorly occupied MEL molecule or a product of its degradation. In view of these doubts, we decided to model this density with several water molecules marked as UNL (Unknown Ligand).

The main chain of the protein model could be traced in electron density without any brakes and it was possible to determine the rotamers of all side chains. Only the last two, one and three atoms, respectively, of three lysine side chains, Lys21, Lys40, and Lys113, which are directed toward bulk solvent, were omitted from the model due to their high mobility and lack of electron density. For 13 residues two rotamers of the side chain could be determined. The structure was refined to *R/R*_{free} of 12.8/15.3% and *MolProbity* (Chen et al., 2010) analysis emphasizes the high stereochemical quality of the model (**Table 1**).

Overall Fold and Three New PR-10 Binding Sites

As already established by Michalska et al. (2010), Hyp-1 has the canonical PR-10-fold with all its structural motifs (**Figure 1A**). The residue ranges of each structural motif are given in **Table 2**. Although the protein main chain creates the typical “baseball glove” framework, the peculiarity of the Hyp-1 protein lies in the side chains, which are responsible for shaping a very interesting and unusual internal cavity, quite different from the cavities known from other PR-10 proteins. As discussed before (Sliwiak et al., 2016), the PR-10 proteins of known structure appear to possess two types of cavities; type I, which is small, shallow, opened at the E1 entrance and capable of binding one ligand molecule in a specific manner; and type II, which resembles a spacious bag and spans the entire hydrophobic core between entrances E1 and E2, and is capable to accommodating more than two ligands of different chemical nature at different positions. In the case of Hyp-1, we observe two separated internal chambers (1 and 2), each with its own entrance (E1 and E2, respectively), and a third site (3) which is formed as a deep invagination of the

protein surface. A very unusual feature of the Hyp-1 binding sites as compared to other PR-10 proteins is the amazing conservation of the ligand position; regardless of their chemical character, the ligand molecules always take the same position in the three sites, as clearly illustrated by the cut-away sections of the protein interior in **Figure 1C**.

The binding site 3 is quite mysterious. Although the residues stabilizing the ligand in a vise-type manner (Lys33 and Tyr150) are conserved among almost all PR-10 proteins, Hyp-1 is the only protein where ligand binding is found at this site and is seen there in all available structures of Hyp-1 complexes. An explanation of this observation may lie in the interaction between the unstructured C-terminal end of Hyp-1 and helix $\alpha 1$. Among the aligned sequences (**Figure 3A**) of PR-10 proteins studied structurally in our laboratory, Hyp-1 is the only one to have a long C-terminal peptide forming a C-terminal loop that interacts with helix $\alpha 1$ (**Figure 3B**). This interaction involves hydrogen bonds between the Ne2 atom of His17 and the C-terminal carboxylate group (of Ala159) and between the Ne and N η 2 atoms of Arg18 and the O atom of Val157, as well as hydrophobic interactions of the aromatic ring of Phe158 with Lys21 and the main chain of helix $\alpha 1$ (**Figure 3C**). His17 is unique to the Hyp-1 protein, as in other PR-10 proteins there is a negatively charged Glu or hydrophobic Ala residue at this position (**Figure 3A**). Although the C-terminal sequence of MtN13 is even longer than in Hyp-1 (**Figure 3A**), in all crystal structures of MtN13/cytokinin complexes (4GY9, 4JHG, 4JHH, 4JHI; Ruskowski et al., 2013), this end of the protein is disordered, indicating the absence of such C-terminal stabilization. The interactions mentioned above stabilize the surface invagination of Hyp-1, thereby creating the new ligand binding site 3. The above interactions between the

TABLE 2 | Residue ranges of PR-10 canonical motifs in Hyp-1 protein.

Secondary structure element	No	Residue range	
α -helix	1	Pro16–Leu23	
	2	Arg27–Ala34	
	3	Glu130–Asn154	
	β -sheet	1	Ala2–Ser12
		2	Ser41–Glu46
		3	Val54–Thr58
		4	Tyr67–Asp76
5		Tyr81–Glu88	
6		Lys98–Leu105	
7		Lys113–His121	
Loop	1	Pro13–Ala15	
	2	Val24–Glu26	
	3	Gln35–Lys40	
	4	Gly47–Thr53	
	5	Phe59–Thr66	
	6	Ala77–Phe80	
	7	Gly89–Glu97	
Unstructured	8	Glu106–Ser112	
	9	Pro122–Asn129	
	–	Pro155–Ala159	

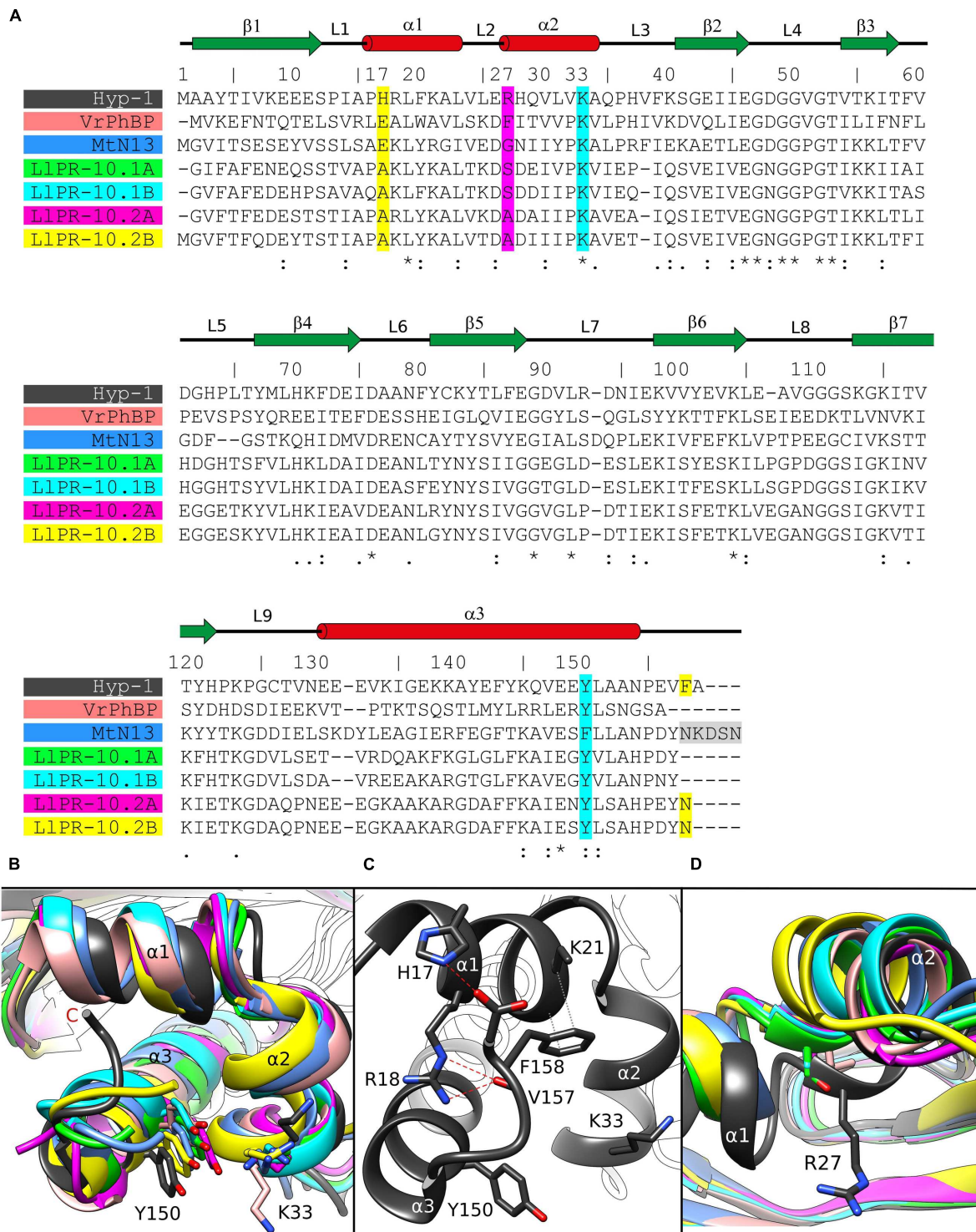


FIGURE 3 | (A) Multiple sequence alignment and **(B)** superposition of PR-10 models (identified by their PDB codes and color) with zoom on the area of the Lys33–Tyr150 invagination of Hyp-1. Green, LIPR-10.1A (4RYV); cyan, LIPR-10.1B (1IFV, chain A); magenta, LIPR-10.2A (1XDF, chain B); yellow, LIPR-10.2B (2QIM); salmon, VrPhBP (2FLH, chain B); blue, MtN13 (4JHG); dark gray, the present model. In the sequence alignment **(A)**, the positions corresponding to Hyp-1 Pro17, Arg27, Lys33, Tyr150 and Phe158 are highlighted as follows: cyan, conservative residues creating the surface invagination; yellow, unique Hyp-1 residues that are involved in C-end stabilization; magenta, the cavity separator Arg27. The disordered C-terminal pentapeptide of MtN13 is highlighted in gray. Identical (*) as well as more (:) and less (.) similar residues are marked at the bottom, while the pictograms above the Hyp-1 sequence numbers illustrate the secondary structure elements (green arrows, β -strands; red cylinders, α -helices) and their annotation. **(C)** Interaction of the C-end of Hyp-1 with $\alpha 1$, stabilizing the novel binding site 3. **(D)** $C\alpha$ superposition of different PR-10 models as in **(B)** with zoom on the $\alpha 2$ structural element, with residues corresponding to Hyp-1 Arg27 shown in stick representation.

C-terminus and helix $\alpha 1$ are present in all experimental models of Hyp-1.

As discussed before (Sliwiak et al., 2015), the main partition between the chambers 1 and 2 in Hyp-1 is the long side chain of Arg27 from helix $\alpha 2$, with further contribution from Tyr84, Tyr101, Ala140, and Phe143. A structural superposition of the $\alpha 2$ helix of different PR-10 proteins (Figure 3D) reveals not only that the Hyp-1-specific Arg27 residue is replaced in other PR-10 sequences by Gly, Ala or Ser, but also that the $\alpha 2$ helix of Hyp-1 penetrates the hydrophobic core in an exceptionally high degree, contributing to this unique partitioning into two separate internal chambers.

Ligand Identification in Electron Density

As mentioned above, MEL1 is the best stabilized ligand in the structure. Its electron density (Figure 1B) clearly indicates the position of each atom. A similar situation was found in the Hyp-1/ANS complex, where the superposition of the 28 Hyp-1 molecules in the asymmetric unit produced an exceptionally consistent overlay of the ligand molecules at site 1 (Sliwiak et al., 2015). We note, however, that there is a strange positive electron density peak less than 2 Å from the MEL1 methoxy group (Figure 1B), for which we do not have a plausible explanation.

Although we did not model melatonin at site 2, it is quite obvious that the electron density there is consistent with the shape of the indole ring. However, we were unable to build a satisfactory model of MEL or 5-methoxyindole there. One possibility is that the ligand at site 2 is very mobile. Alternatively, a melatonin degradation product could be bound there. As a free radical scavenger, melatonin is rather unstable and could undergo X-ray-induced degradation. To test this possibility, we irradiated an NMR probe containing 0.6 M solution of melatonin in deuterated methanol with a synchrotron X-ray dose ~ 10 times higher than that used in the diffraction experiment. However, the NMR spectrum after irradiation was unchanged, suggesting that photodegradation was not a likely mechanism of the observed effect. Notwithstanding this result, we were also unable to model N1-acetyl-N2-formyl-5-methoxykynuramine (AFMK) or 6-hydroxymelatonin at this site, the two known photodegradation products of melatonin (Maharaj et al., 2002).

MEL3 has flat electron density, indicating in-plane rotation of the ligand. To account for this effect, we modeled MEL3 in two orientations. It is interesting to note that rotation of the flat ANS molecule at site 3 site was also observed in the Hyp-1/ANS complex, with the caveat that the rotation could be deduced from the superposition of the 28 copies of the Hyp-1 molecule, whereas in each individual case the ANS3 ligand could be modeled in a unique orientation. Nevertheless, despite the rotation of MEL3, it is safe to conclude that the ligand molecule is firmly docked between the jaws of the Lys33–Tyr150 vise.

It is very important to stress that the ligand electron densities described above were perfectly reproducible in several independent structure determinations utilizing differently produced Hyp-1/MEL crystals, namely either in the presence of solid (pulverized) melatonin or in the presence of equimolar concentration of *trans*-zeatin. The reproducibility includes

even the inexplicable electron density peak near the MEL1 molecule.

Ligand Binding

The MEL molecule at site 1 that has the best definition in electron density, makes direct contacts with the protein only via weak (3.6–3.8 Å) hydrophobic interactions with Phe39, Leu31, Leu65, Val91, Gly136, and Lys139, as well as via water-mediated hydrogen bonds of its O2¹ atom with Gly136 (O), Ala140 (N), and Met68 (S8), and of its N1 atom with Arg93 (N η 2) and Glu132 (O).

Interestingly, the MEL1 ligand is additionally stabilized and pushed to its binding site by a direct hydrogen bond of its N2 atom with the carboxylate group of Asp48 from loop L4 of an adjacent Hyp-1 molecule. This interaction is additionally stabilized by a hydrogen bond between the “intruding” Asp48 carboxylate and the N ϵ atom of His63 from loop L5 of the MEL1-binding protein molecule (Figure 4). Such a situation is not new among PR-10 proteins. In the structures of *Medicago truncatula* Nodulin 13 (MtN13) in complex with cytokinins (Ruszkowski et al., 2013), there is a similar interaction with Asp62 from loop L5 of another copy of MtN13, which forms a fork of hydrogen bonds with the N6 and N7 atoms of the cytokinin molecule. In variance with the Hyp-1 situation, however, the cytokinin...Asp62 interaction in MtN13 is mutual, leading to (quite exceptional) dimer formation of that PR-10 protein.

As mentioned above, the MEL3 ligand is evidently rotating between the jaws of the vise formed by the side chains of Lys33 and Tyr150. Its stacking interactions with these residues have van der Waals character.

¹The atom numbering scheme of melatonin (Figure 2A) follows that of Quarles et al. (1974) because the system used by the PDB is quite irrational and inconsistent with established rules, as also noted for other ligands (Jaskolski, 2013).

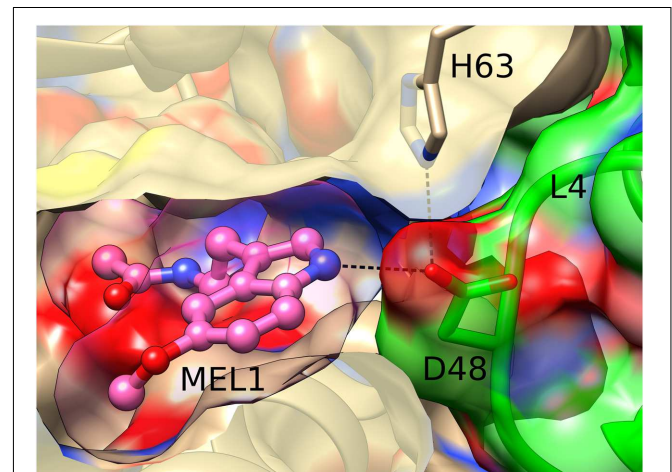


FIGURE 4 | Hydrogen bonding between MEL1 and Asp48 from loop L4 of an adjacent Hyp-1 molecule (green). MEL1 is shown as a ball-and-stick model surrounded by its (semitransparent) van der Waals surface, and the two interacting Hyp-1 molecules (sand and green) are presented with their semitransparent van der Waals surfaces.

The unidentified ligand marked by dummy (UNL) water molecules at site 2 is in van der Waals distance to residues Leu19, Ile116, and Tyr144. The same residues stabilize the ANS2 ligand in the binding site 2 of the Hyp-1/ANS complex (Sliwiak et al., 2015).

Conformational Differences between the Available Hyp-1 Complexes

We note at the outset that there is no truly ligand-free structure of Hyp-1 in the PDB. The closest case of a protein crystallized without any intentional ligand is 3IE5 (Michalska et al., 2010) but even in that structure there are serendipitous PEG molecules found in the binding sites of the two protein chains A and B. A superposition of the C α atoms revealed that the structures of Hyp-1 complexed with melatonin, ANS (4N3E, represented by chain K) and PEG (3IE5, chains A and B) are quite similar, with R.M.S.D. values within ~ 1 Å (Table 3). That means that the chemical character of the ligand does not influence the Hyp-1-fold to a significant degree and that the three conserved binding sites are capable of accommodating different hydrophobic and amphiphilic ligands from the aqueous environment. Interestingly, chain A of the 3IE5 structure seems to differ most significantly from all the remaining Hyp-1 models, even when compared with chain B from the same structure (Figure 5). This difference can be correlated with the fact that in chain A of 3IE5 the binding site 3 is empty, allowing Tyr150 and Lys33 to form a direct stacking contact. This interaction brings the end of helix $\alpha 3$ closer to helix $\alpha 2$ and, in consequence, pulls the loops L3 and L5 toward $\alpha 3$. One can speculate that binding of a ligand molecule at site 3 widens the E1 entrance, facilitating access of another ligand molecule to site 1. We can therefore hypothesize that the ligand binding mechanism of Hyp-1 has a cooperative character. Moreover the PEG molecules in 3IE5 (Figure 5) seem to pull the main cavity separator (Arg27) away from the hydrophobic core, resulting in a less solid separation between chambers 1 and 2.

CONCLUSION

Hyp-1, a protein from *H. perforatum*, has the characteristic PR-10-fold. However, despite the overall similarity, it has three highly unique and characteristic ligand binding sites, which may suggest a unique ligand-binding mechanism among the PR-10 proteins. Although the interaction of Hyp-1 with melatonin

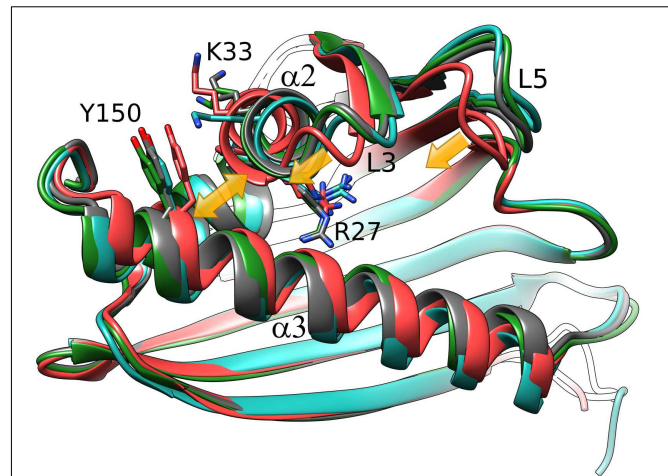


FIGURE 5 | C α superposition of the available models of Hyp-1. Color code: Hyp-1/MEL (this work), dark gray; Hyp-1/ANS (4N3E, chain K), green; Hyp-1/PEG (3IE5), chain A – red, chain B – blue. Yellow arrows indicate the conformational rearrangements in chain A of 3IE5, which has an empty site 3, in particular the approach of helices $\alpha 2$ and $\alpha 3$ (highlighted by Tyr150) leading to a tighter grip of the fingertip loops L3 and L5.

does not appear to be particularly strong, the structure of the Hyp-1/MEL complex is quite robust and reproducible in a number of crystal structure determination experiments. The reproducibility regards also the unidentified electron density at site 2 and an unattributed peak near the methoxy group of MEL1. Moreover, co-crystallization trials with other phytohormones and natural ligands using the same crystallization screens as in the Hyp-1/MEL experiments, produced no results. The three binding sites identified in the Hyp-1/MEL complex are exactly the same as in the Hyp-1/ANS complex. They comprise a well ordered MEL (site 1), a rotationally disordered one (3) and possibly an unidentified melatonin degradation product (2). Considering all the facts together, one can conclude that Hyp-1 may be capable of melatonin storage/transport under stress conditions in *H. perforatum*. A shortlist of the supporting observations is as follows: (i) Hyp-1, as a probable pathogenesis-related protein (a superfamily, whose members are expressed *inter alia* during abiotic and biotic stress) was detected in the roots and other parts of *Hypericum* plantlets and its mRNA expression has the highest level in the roots. (ii) Melatonin concentration is very high in vulnerable parts, like seedlings, as well as in the leaves and flowers of mature *Hypericum* plants, and it is further elevated during, e.g., radiation stress, and thus it could be bound by Hyp-1 despite of a relatively low affinity. (iii) Melatonin and its precursor tryptophan have been reported to be absorbed by plant roots from soil and media (Tan et al., 2007). (iv) *trans*-Zeatin, which is frequently reported as a natural ligand of PR-10 proteins, did not affected Hyp-1/MEL crystallization, while the artificial fluorescent probe ANS - did. However, taking into account that some PR-10 proteins show pleiotropic binding capacity (Sliwiak et al., 2016), it should not be ruled out that Hyp-1 may also play other role(s) in *H. perforatum*.

TABLE 3 | R.M.S.D. (Å) values of C α superpositions of the following Hyp-1 models: Hyp-1/MEL complex (this work), Hyp-1/ANS complex (4N3E, chain K), and chains A/B from the “ligand free” (i.e., Hyp-1/PEG) form (3IE5).

	3IE5, chain A	3IE5, chain B	4N3E, chain K
Hyp-1/MEL	1.01	0.72	0.61
4N3E, chain K	1.04	0.75	
3IE5, chain B	1.07		

AUTHOR CONTRIBUTIONS

JS designed the experiments, conducted the experiments and calculations, analyzed the results, drafted the manuscript. ZD supervised initial experiments, conducted X-ray diffraction experiments and data processing, participated in ms preparation. MJ supervised the project, analyzed the results, wrote the paper.

FUNDING

X-Ray diffraction data were collected at the Southeast Regional Collaborative Access Team (SER-CAT) beamline of the APS/ANL. Use of the Advanced Photon Source was supported by the US Department of Energy, Office of Science, Office of

Basic Energy Sciences, under Contract No. W-31-109-Eng-38. Financial support of the project was provided by the European Union within the European Regional Developmental Fund and by the Polish National Science Centre (grant No. 2013/10/M/NZ1/00251). This publication was supported by the Polish Ministry of Science and Higher Education under the KNOW program.

ACKNOWLEDGMENTS

We would like to thank Dr. Magdalena Małgowska for NMR measurements and advice, Mirosława Dauter for crystal seeding tips and Prof. Michał Sikorski for Hyp-1 clone.

REFERENCES

- Afonine, P. V., Grosse-Kunstleve, R. W., Echols, N., Headd, J. J., Moriarty, N. W., Mustyakimov, M., et al. (2012). Towards automated crystallographic structure refinement with phenix refine. *Acta Crystallogr. D* 68, 352–367. doi: 10.1107/S0907444912001308
- Arnao, M., and Hernández-Ruiz, J. (2015). Functions of melatonin in plants: a review. *J. Pineal Res.* 59, 133–150. doi: 10.1111/jpi.12253
- Bais, H. P., Vepachedu, R., Lawrence, C. B., Stermitz, F. R., and 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. doi: 10.1074/jbc.M301681200
- Biesiadka, J., Bujacz, G., Sikorski, M. M., and Jaskolski, M. (2002). Crystal structures of two homologous pathogenesis-related proteins from yellow lupine. *J. Mol. Biol.* 319, 1223–1234. doi: 10.1016/S0022-2836(02)00385-6
- Calamini, B., Santarsiero, B. D., Boutin, J. A., and Mesecar, A. D. (2008). Kinetic, thermodynamic and X-ray structural insights into the interaction of melatonin and analogues with quinone reductase 2. *Biochem. J.* 413, 81–91. doi: 10.1042/BJ20071373
- Casañal, A., Zander, U., Muñoz, C., Dupeux, F., Luque, I., Botella, M. A., et al. (2013). The strawberry pathogenesis-related 10 (PR-10) Fra proteins control flavonoid biosynthesis metabolic intermediates. *J. Biol. Chem.* 288, 35322–35332. doi: 10.1074/jbc.M113.501528
- Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., et al. (2010). MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* 66, 12–21. doi: 10.1107/s0907444909042073
- Chwastak, M., Jaskolski, M., and Cieplak, M. (2014). Structure-based thermodynamic and mechanical stability of plant PR-10 proteins. *FEBS J.* 281, 416–429. doi: 10.1111/febs.12611
- Cohen, G. H. (1997). ALIGN: a program to superimpose protein coordinates, accounting for insertions and deletions. *J. Appl. Crystallogr.* 30, 1160–1161. doi: 10.1107/S0021889897006729
- Dubbels, R., Reiter, R. J., Klenke, E., Goebel, A., Schnakenberg, E., Ehlers, C., et al. (1995). Melatonin in edible plants identified by radioimmunoassay and by high performance liquid chromatography-mass spectrometry. *J. Pineal Res.* 18, 28–31. doi: 10.1111/j.1600-079X.1995.tb00136.x
- Emsley, P., Lohkamp, B., Scott, W., and Cowtan, K. (2010). Features and development of Coot. *Acta Crystallogr. D* 66, 486–501. doi: 10.1107/S0907444910007493
- Fernandes, H., Bujacz, G., Bujacz, A., Jelen, F., Jasinski, M., Kachlicki, P., et al. (2009). Cytokinin-induced structural adaptability of a *Lupinus luteus* PR-10 protein. *FEBS J.* 276, 1596–1609. doi: 10.1111/j.1742-4658.2009.06892.x
- Fernandes, H., Michalska, K., Sikorski, M., and Jaskolski, M. (2013). Structural and functional aspects of PR-10 proteins. *FEBS J.* 280, 1169–1199. doi: 10.1111/febs.12114
- Fernandes, H., Pasternak, O., Bujacz, G., Bujacz, A., Sikorski, M. M., and Jaskolski, M. (2008). L. luteus pathogenesis-related protein as a reservoir for cytokinins. *J. Mol. Biol.* 378, 1040–1051. doi: 10.1016/j.jmb.2008.03.027
- Gajhede, M., Osmark, P., Poulsen, F. M., Ipsen, H., Larsen, J. N., Joost van Neerven, R. J., et al. (1996). X-ray and NMR structure of Bet v 1, the origin of birch pollen allergy. *Nat. Struct. Biol.* 3, 1040–1045. doi: 10.1038/nsb1296-1040
- Jaskolski, M. (2013). On the propagation of errors. *Acta Crystallogr. D* 69, 1865–1866. doi: 10.1107/s090744491301528x
- Karplus, P. A., and Diederichs, K. (2012). Linking crystallographic model and data quality. *Science* 336, 1030–1033. doi: 10.1126/science.1218231
- Kofler, S., Asam, C., Eckhard, U., Wallner, M., Ferreira, F., and Brandstetter H. (2012). Crystallographically mapped ligand binding differs in high and low IgE binding isoforms of birch pollen allergen bet v 1. *J. Mol. Biol.* 422, 109–123. doi: 10.1016/j.jmb.2012.05.016
- Kosuth, J., Hrehorova, D., Jaskolski, M., and Cellarova, E. (2013). Stress-induced expression and structure of the putative gene hyp-1 for hypericin biosynthesis. *Plant Cell Tiss. Organ. Cult.* 114, 207–216. doi: 10.1007/s11240-013-0316-0
- Kosuth, J., Katkovicinová, Z., Olexová, P., and Cellárová, E. (2007). Expression of the hyp-1 gene in early stages of development of *Hypericum perforatum* L. *Plant Cell Rep.* 26, 211–217. doi: 10.1007/s00299-006-0240-4
- Krissinel, E., and Henrick, K. (2007). Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* 372, 774–797. doi: 10.1016/j.jmb.2007.05.022
- Maharaj, D. S., Anoopkumar-Duki, S., Glass, B. D., Antunes, E. M., Lack, B., Walker, R. N., et al. (2002). The identification of the UV degradants of melatonin and their ability to scavenge free radicals. *J. Pineal Res.* 32, 257–261. doi: 10.1034/j.1600-079X.2002.01866.x
- Markovic-Housley, Z., Degano, M., Lamba, D., von Roepenack-Lahaye, E., Clemens, S., Susani, M., et al. (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. doi: 10.1016/S0022-2836(02)01197-X
- McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007). Phaser crystallographic software. *J. Appl. Crystallogr.* 40, 658–674. doi: 10.1107/S0021889807021206
- Michalska, K., Fernandes, H., Sikorski, M. M., and Jaskolski, M. (2010). Crystal structure of Hyp-1, a St John's wort protein with implication in the biosynthesis of hypericin. *J. Struct. Biol.* 169, 161–171. doi: 10.1016/j.jsb.2009.10.008
- Mogensen, J. E., Wimmer, R., Larsen, J. N., Spangfort, M. D., and 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. doi: 10.1074/jbc.M202065200
- Murch, S. J., Campbell, S. S. B., and Saxena, P. K. (2001). The role of serotonin and melatonin in plant morphogenesis: regulation of auxin-induced root organogenesis in in vitro-cultured explants of St. John's Wort (*Hypericum perforatum* L.). *In Vitro Cell. Dev. Biol. Plant* 37, 786–793. doi: 10.1007/s11627-001-0130-y
- Murch, S. J., KrishnaRaj, S., and Saxena, P. K. (2000). Tryptophan is a precursor for melatonin and serotonin biosynthesis in in vitro regenerated St. John's wort (*Hypericum perforatum* L. cv. Anthos) plants. *Plant Cell Rep.* 19, 698–704. doi: 10.1007/s002990000206

- Murch, S. J., Simmons, C. B., and Saxena, P. K. (1997). Melatonin in feverfew and other medicinal plants. *Lancet* 359, 1598–1599. doi: 10.1016/S0140-6736(05)64014-7
- Otwinowski, Z., and Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* 276, 307–326. doi: 10.1016/S0076-6879(97)76066-X
- Pasternak, O., Bujacz, G. D., Fujimoto, Y., Hashimoto, Y., Jelen, F., Otlewski, J., et al. (2006). Crystal structure of *Vigna radiata* cytokinin-specific binding protein in complex with zeatin. *Plant Cell* 18, 2622–2634. doi: 10.1105/tpc.105.037119
- Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., et al. (2004). UCSF Chimera - a visualization system for exploratory research and analysis. *J. Comput. Chem.* 25, 1605–1612. doi: 10.1002/jcc.20084
- Qian, J., Wu, J., Yao, B., and Lu, Y. (2012). Preparation of a polyclonal antibody against hypericin synthase and localization of the enzyme in red-pigmented *Hypericum perforatum* L. plantlets. *Acta Biochim. Polon.* 59, 639–645.
- Quarles, W. G., Templeton, D. H., and Zalkin, A. (1974). The crystal and molecular structure of melatonin. *Acta Crystallogr.* B30, 99–103. doi: 10.1107/S0567740874002287
- Ruszkowski, M., Sliwiak, J., Ciesielska, A., Barciszewski, J., Sikorski, M. M., and Jaskolski, M. (2014). Specific binding of gibberellic acid by Cytokinin-Specific Binding Proteins: a new aspect of plant hormone-binding proteins with PR-10 fold. *Acta Crystallogr.* D70, 2032–2041. doi: 10.1107/s1399004714010578
- Ruszkowski, M., Szpotkowski, K., Sikorski, M. M., and Jaskolski, M. (2013). The landscape of cytokinin binding by a plant nodulin. *Acta Crystallogr.* D69, 2365–2380. doi: 10.1107/s0907444913021975
- Schiebel, J., Radeva, N., Krimmer, S. G., Wang, X., Stieler, M., Ehrmann, F. R., et al. (2016). Six biophysical screening methods miss a large proportion of crystallographically discovered fragment hits: a case study. *Chem. Biol.* doi: 10.1021/acscmbio.5b01034 [Epub ahead of print].
- Seutter von Loetzen, C., Hoffmann, T., Hartl, M. J., Schweimer, K., Schwab, W., Rosch, P., et al. (2014). Secret of the major birch pollen allergen Bet v 1: identification of the physiological ligand. *Biochem. J.* 457, 379–390. doi: 10.1042/BJ20130413
- Seutter von Loetzen, C., Jacob, T., Hartl-Spiegelhauer, O., Vogel, L., Schiller, D., Spörlein-Güttler, C., et al. (2015). Ligand recognition of the major birch pollen allergen bet v 1 is isoform dependent. *PLoS ONE* 10:e0128677. doi: 10.1371/journal.pone.0128677
- Sheard, L. B., and Zheng, N. (2009). Plant biology: Signal advance for abscisic acid. *Nature* 462, 575–576. doi: 10.1038/462575a
- Sliwiak, J., Dauter, Z., Kowiel, M., McCoy, A. J., Read, R. J., and Jaskolski, M. (2015). ANS complex of St John's wort PR-10 protein with 28 copies in the asymmetric unit: a fiendish combination of pseudosymmetry with tetrahedral twinning. *Acta Crystallogr.* D71, 829–843. doi: 10.1107/s139900471501388
- Sliwiak, J., Dolot, R., Michalska, K., Szpotkowski, K., Bujacz, G., Sikorski, M. M., et al. (2016). Crystallographic and CD probing of ligand-induced conformational changes in a plant PR-10 protein. *J. Struct. Biol.* 193, 55–66. doi: 10.1016/j.jsb.2015.11.008
- Sliwiak, J., Jaskolski, M., Dauter, Z., McCoy, A. J., and Read, R. J. (2014). Likelihood-based molecular-replacement solution for a highly pathological crystal with tetrahedral twinning and sevenfold translational noncrystallographic symmetry. *Acta Crystallogr.* D70, 471–480. doi: 10.1107/s1399004713030319
- Tan, D., Manchester, L. C., Helton, P., and Reiter, R. J. (2007). Phytoremediative capacity of plants enriched with melatonin. *Plant Signal. Behav.* 2, 514–516. doi: 10.4161/psb.2.6.4639

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Sliwiak, Dauter and Jaskolski. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

PUBLICATION VI

Specific binding of gibberellic acid by Cytokinin-Specific Binding Proteins: a new aspect of plant hormone-binding proteins with the PR-10 fold

Milosz Ruskowski,^a Joanna Sliwiak,^a Agnieszka Ciesielska,^a Jakub Barciszewski,^a Michal Sikorski^a and Mariusz Jaskolski^{a,b*}

^aCenter for Biocrystallographic Research, Institute of Bioorganic Chemistry, Polish Academy of Sciences, Poznan, Poland, and ^bDepartment of Crystallography, Faculty of Chemistry, A. Mickiewicz University, Poznan, Poland

Correspondence e-mail: mariuszj@amu.edu.pl

Pathogenesis-related proteins of class 10 (PR-10) are a family of plant proteins with the same fold characterized by a large hydrophobic cavity that allows them to bind various ligands, such as phytohormones. A subfamily with only ~20% sequence identity but with a conserved canonical PR-10 fold have previously been recognized as Cytokinin-Specific Binding Proteins (CSBPs), although structurally the binding mode of *trans*-zeatin (a cytokinin phytohormone) was found to be quite diversified. Here, it is shown that two CSBP orthologues from *Medicago truncatula* and *Vigna radiata* bind gibberellic acid (GA3), which is an entirely different phytohormone, in a conserved and highly specific manner. In both cases a single GA3 molecule is found in the internal cavity of the protein. The structural data derived from high-resolution crystal structures are corroborated by isothermal titration calorimetry (ITC), which reveals a much stronger interaction with GA3 than with *trans*-zeatin and pH dependence of the binding profile. As a conclusion, it is postulated that the CSBP subfamily of plant PR-10 proteins should be more properly linked with general phytohormone-binding properties and termed phytohormone-binding proteins (PhBP).

Received 6 April 2014

Accepted 8 May 2014

PDB references:

VrPhBP–GA3 complex, 4psb;

MtPhBP–GA3 complex, 4q0k

1. Introduction

Plant hormones (phytohormones) are chemical messengers that coordinate numerous cellular functions. This large group of molecules includes (but is not restricted to) ten main, chemically very diverse, classes: auxins, cytokinins, gibberellins, abscisic acid, brassinosteroids, ethylene, jasmonates, polypeptide hormones, salicylic acid and strigolactones (Santner *et al.*, 2009). Gibberellins, such as gibberellic acid (GA3; Fig. 1), are diterpenoid tetracyclic or pentacyclic growth regulators. They induce *inter alia* seed development and germination, organ elongation and flowering (Yamaguchi, 2008). Gibberellins were first discovered in *Gibberella fujikuroi*, a fungal pathogen of rice that causes extreme stem elongation, finally leading to plant collapse and death (Yabuta & Sumitki, 1938). Plants produce endogenous gibberellins and their cellular level is regulated not only *via* a negative-feedback loop but also by the concentration of auxins and ethylene (Fleet & Sun, 2005; Yamaguchi, 2008). The gibberellin receptor is known as gibberellin-insensitive dwarf1 protein (GID1), as loss-of-function mutations in the *gid1* gene cause dwarfism (Peng *et al.*, 1999), a feature desired in the cultivation of rice. Complexes of *Arabidopsis thaliana* GID1 with gibberellins have been investigated structurally by Murase *et al.* (2008) (PDB entries 2zsh and 2zsi). The GID1 receptor can bind DELLA proteins possessing a conserved Asp-Glu-Leu-Leu-Ala N-terminal sequence, which are

negative regulators of gibberellin response (Schwechheimer, 2008; Schwechheimer & Willige, 2009). In this mechanism, gibberellin binding by *GID1* initiates *GID1*–*DELLA* complex formation. As a result, the *DELLA* proteins can no longer function as transcription repressors of gibberellin-dependent genes and are instead ubiquitinated and targeted for degradation.

On the other hand, cytokinins, such as *trans*-zeatin (*ZEA*), which are adenine derivatives, stimulate cell division (cytokinesis) and differentiation in various developmental processes. Cytokinins take part, for example, in apical dominance, axillary bud growth, leaf senescence, flowering and response to pathogens. In legume plants such as *Medicago truncatula* and *Vigna radiata* cytokinins also control root nodulation during symbiosis with nitrogen-fixing soil bacteria.

Pathogenesis-related proteins of class 10 (*PR*-10) are small (up to 19 kDa), usually monomeric, slightly acidic, cytosolic plant-specific proteins (Fernandes *et al.*, 2013). The name pathogenesis-related is quite misleading as it is now well established that *PR*-10 proteins are expressed not only during pathogenesis (van Loon *et al.*, 2006). Despite many years of study, however, the exact biological function of *PR*-10 proteins remains unknown. In contrast, the structural conservation of *PR*-10 proteins is well established and forms the basis of the canonical *PR*-10 fold. The fold consists of a curved seven-stranded antiparallel β -sheet crossed by a long C-terminal helix (α 3), supported at its carboxy end by a V-shaped motif of two shorter helices (α 1 and α 2). The most prominent feature of this fold is a large hydrophobic cavity formed between helix α 3 and the β -sheet (Chwastyk *et al.*, 2014) that is evidently the binding site for *PR*-10 ligands. However, even the nature of the physiological binding partners of the *PR*-10 proteins remains obscure. A group of promising candidates are cytokinins, but while several crystallographic studies did confirm the potential of *PR*-10 proteins to bind cytokinins, the complexes also revealed a perplexing diversity of ligand interactions and binding modes, and even a highly variable stoichiometry (Pasternak *et al.*, 2006; Fernandes *et al.*, 2008, 2009; Kofler *et al.*, 2012). Only recently, a *PR*-10 protein involved in root nodulation, nodulin MtN13, was found to form highly specific and structurally conserved complexes with different cytokinins, although the complexes turned out to have an unusual dimeric structure (Ruszkowski *et al.*, 2013).

Historically, a subclass of proteins with a conserved *PR*-10 fold, despite very low sequence identity (\sim 20%) to other *PR*-10 proteins, was identified by Fujimoto *et al.* (1998) as strong cytokinin binders. The reported exceedingly high (nanomolar) cytokinin affinity was later corrected (Pasternak *et al.*, 2006) by five orders of magnitude (high micromolar), but the term Cytokinin-Specific Binding Proteins (*CSBPs*) has been well established in the literature. *CSBP* proteins are only found in legume plants and are expressed at such low levels that Fujimoto *et al.* (1998) had to use 95 kg of etiolated mung bean (*V. radiata*) seedlings for the detection and N-terminal sequencing of Vr*CSBP*. The crystal structure of Vr*CSBP* in complex with *ZEA* (Pasternak *et al.*, 2006) revealed that neither the binding mode nor the stoichiometry were conserved in the four protein molecules found in the asymmetric unit. This led to the assumption that cytokinins might be in fact not the preferred physiological ligands of the *CSBP* proteins. In an independent study, Zawadzki *et al.* (2010) reported that Vr*CSBP* also interacted with gibberellins. The latter results were obtained in indirect experiments using fluorescence correlation spectroscopy to monitor the displacement of *trans*-zeatin (labelled with a large chromophore moiety) by gibberellic acid.

In the present work, we have focused on the question of whether two *CSBP* proteins from *M. truncatula*, which is a model legume plant, and from *V. radiata* could be demonstrated to interact with gibberellins in direct experiments using crystallography and isothermal titration calorimetry (*ITC*). The question has been answered positively, as our crystallographic results unambiguously show that both proteins bind GA3 with 1:1 stoichiometry and that the binding mode is strictly conserved in the two complexes. The crystallographic results were supported by *ITC* measurements, which showed that the proteins bind GA3 much more strongly than *trans*-zeatin and that the binding is pH-dependent. From these observations, we conclude that the proteins classified so far as *CSBPs* should be more properly recognized as more general phytohormone binders. Consequently, we propose to replace the term *CSBP* with PhBP (phytohormone-binding proteins), and we use the latter acronym throughout the remaining part of this article.

2. Materials and methods

2.1. Cloning, overexpression and purification of PhBP proteins

The MtPhBP DNA coding sequence was amplified by polymerase chain reaction (*PCR*) using *M. truncatula* (ecotype J5) cDNA as template. The reaction product was cloned into the pET-TOPO-151D vector (Invitrogen) and the correctness of the insert was confirmed by sequencing. The vector introduces an N-terminal His₆ tag followed by the cleavage site for TEV (*Tobacco etch virus*) protease and a hexapeptide linker (*GIDPFT*) that precede the genuine protein sequence. Overexpression was carried out in *Escherichia coli* BL21 (DE3) cells. The cells were disrupted by sonication using bursts of

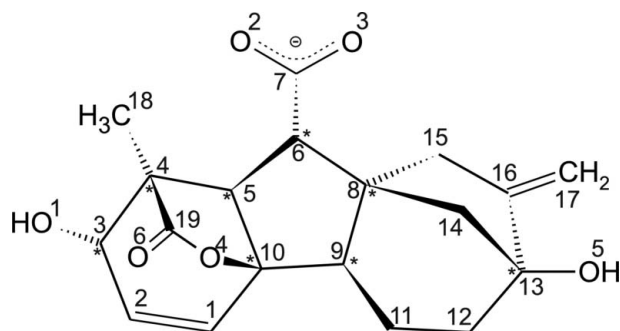


Figure 1
Chemical structure of gibberellic acid (GA3) with atom numbering. Asterisks indicate chiral C atoms.

total duration 4 min with appropriate intervals for cooling. Cell debris was removed by centrifugation at 15 000 rev min⁻¹ for 30 min at 4°C. The supernatant was applied onto a column packed with 6 ml HisTrap HP resin (GE Healthcare). After binding, the column was washed four times with 30 ml binding buffer (50 mM Tris–HCl pH 8.0, 500 mM NaCl, 20 mM imidazole) and the purified protein was eluted with 15 ml elution buffer (50 mM Tris–HCl pH 8.0, 500 mM NaCl, 200 mM imidazole). The His₆ tag was cleaved with TEV protease and the excess imidazole was removed by dialysis (overnight at 4°C) simultaneously. The solution was mixed in a column with HisTrap HP resin to remove the His₆-tag debris and the His₆-tagged TEV protease. The flowthrough was collected, concentrated to 4 ml and applied onto a HiLoad Superdex 200 16/60 column (GE Healthcare) equilibrated with 5 mM sodium citrate pH 6.3. The sample was concentrated to 10 mg ml⁻¹ as determined by the method of Bradford (1976) and used for crystallization experiments. VrPhBP was produced and purified as described previously (Bujacz *et al.*, 2003).

2.2. Crystallization and data collection

Protein solutions at 10 mg ml⁻¹ (MtPhBP) and 13 mg ml⁻¹ (VrPhBP) concentration were incubated overnight with GA3 (Sigma–Aldrich, catalogue No. 63492) added as a 10 mg ml⁻¹ solution in 10% aqueous ethanol. A threefold molar excess of GA3 was used in both crystallization experiments. Following overnight incubation, the protein–ligand mixtures were centrifuged at 14 000 rev min⁻¹ for 5 min at room temperature to remove the precipitated protein. The crystallizations were carried out in hanging drops using the vapour-diffusion method. The crystallization reservoir for MtPhBP was composed of 0.1 M ADA buffer pH 6.5, 1.0 M ammonium sulfate and the drops were composed of 1 µl protein–ligand solution and 1 µl reservoir solution. In the case of VrPhBP, the crystals were grown using a reservoir solution consisting of 0.1 M MMT buffer pH 4.0, 25% PEG 1500 and the crystallization drops were composed of 4 µl protein–ligand solution and 2 µl reservoir solution. The crystals of the complexes appeared after ten months (MtPhBP) or one week (VrPhBP) at 19°C. The reservoir solutions supplemented with 30 or 20% glycerol were used for cryoprotection of the crystals of the MtPhBP or VrPhBP complexes, respectively. The crystals were vitrified in liquid nitrogen and stored for synchrotron-radiation data collection. The diffraction data were processed and scaled using XDS (Kabsch, 2010) with data statistics as summarized in Table 1.

2.3. Determination and refinement of the crystal structures

The crystal structures of both complexes were solved by molecular replacement using Phaser (McCoy *et al.*, 2007). Protein chain A of VrPhBP retrieved from its complex with ZEA (PDB entry 2flh; Pasternak *et al.*, 2006) served as the search probe. In the case of MtPhBP, automatic model building was carried out with the online version of ARP/wARP (Langer *et al.*, 2008). Coot (Emsley *et al.*, 2010) was

Table 1

Data-collection and refinement statistics.

Values in parentheses are for the highest resolution shell.

GA3 complex with	MtPhBP	VrPhBP
Data collection		
Radiation source	BESSY, Berlin	PETRA III, DESY Hamburg
Beamline	14.2	P14
Wavelength (Å)	0.918000	0.975507
Temperature (K)	100	100
Space group	<i>P</i> 6 ₅	<i>C</i> 2
Unit-cell parameters (Å, °)	<i>a</i> = <i>b</i> = 55.8, <i>c</i> = 100.0	<i>a</i> = 33.4, <i>b</i> = 54.3, <i>c</i> = 71.1, β = 98.7
Resolution (Å)	34.74–1.34 (1.44–1.34)	35.06–1.42 (1.46–1.42)
Reflections (total/unique)	415195/38199	119545/24350
Completeness (%)	97.8 (88.3)	99.5 (99.3)
Multiplicity	10.9 (5.6)	4.9 (4.7)
<i>R</i> _{merge} † (%)	4.3 (87.2)	7.9 (52.0)
<i>I</i> / <i>σ</i> (<i>I</i>)	30.64 (1.91)	9.20 (1.98)
Refinement		
Unique reflections (work + test)	38196	24350
Test reflections	1000	1023
Matthews coefficient (Å ³ Da ⁻¹)	2.58	1.87
Solvent volume (%)	52.4	34.4
No. of atoms (non-H)		
Protein	1265	1268
GA3/glycerol	25/30	25/0
Solvent	197	143
<i>R</i> _{work} / <i>R</i> _{free} (%)	12.4/15.7	15.4/20.4
R.m.s.d. from ideal geometry		
Bond lengths (Å)	0.019	0.019
Bond angles (°)	1.6	1.9
Ramachandran statistics (%)		
Favoured	98.7	98.7
Allowed	1.3	1.3
PDB code	4q0k	4psb

† $R_{\text{merge}} = \frac{\sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle|}{\sum_{hkl} \sum_i I_i(hkl)}$, where $I_i(hkl)$ is the intensity of observation i of reflection hkl .

used for manual fitting in electron-density maps between rounds of model refinement in *phenix.refine* (Adams *et al.*, 2010; Afonine *et al.*, 2012). Anisotropic atomic displacement parameters were refined for all (non-H) atoms. Riding H atoms for the protein molecules were included in F_c calculations for both complexes. Geometrical restraints for the GA3 ligand were generated in *phenix.elbow* (Moriarty *et al.*, 2009) using target values from entry BUWZAU (Kutschabsky & Gunter, 1983) in the Cambridge Structural Database (CSD; Allen, 2002). The final models were validated with *MolProbity* (Chen *et al.*, 2010). The refinement statistics are listed in Table 1.

2.4. Isothermal titration calorimetry

All ITC experiments were carried out at 20°C using a MicroCal iTC200 calorimeter (GE Healthcare). Both proteins, MtCSBP and VrCSBP, were dialyzed against either 150 mM NaCl, 25 mM MES pH 5.5 or 150 mM NaCl, 25 mM HEPES pH 7.4 buffer before titration. GA3 and ZEA were dissolved in the dialysis buffers to concentrations of 0.9 and 1.5 mM, respectively. The protein concentration in the sample cell was determined by the Bradford assay (Bradford, 1976) and was adjusted each time to within the range 80–100 µM.

Phytohormone concentrations were determined gravimetrically by weighing a sample at least 100-fold heavier than the balance sensitivity of 0.1 mg. The ligands (GA3 or ZEA) were injected in 1.5 μl aliquots until saturation was observed. Raw ITC data were analyzed with the *Origin 7.0* software (OriginLab) to obtain the following parameters: stoichiometry (N), dissociation constant (K_d) and the changes in the enthalpy (ΔH) and entropy (ΔS) during the complexation reaction. For the hyperbolic curves (for which determination of N is impossible) which were obtained from titrations of VrPhBP with ZEA, a sequential binding sites model for two binding sites was imposed on the basis of the crystal structure of the VrCSBP–ZEA complex (Pasternak *et al.*, 2006). For

sigmoidal curves obtained from titrations with GA3, a one set of sites model was fitted and N (stoichiometry) was determined from the titration experiment. A competitive binding assay (GA3 *versus* ZEA) performed at pH 5.5 was designed in the same way as the simple titration with GA3, with the additional presence of ZEA at 433 μM concentration in the sample cell. All ITC experiments were performed in triplicate.

2.5. Other software used

Assignment of secondary-structure elements was based on the *DSSP* algorithm (Kabsch & Sander, 1983). Potential quaternary structures were analyzed with *PISA* (Krissinel & Henrick, 2007). Surfaces of protein internal cavities were calculated with *SURFNET* (Laskowski, 1995). *UCSF Chimera* (Pettersen *et al.*, 2004) was used for structural alignments and for the preparation of molecular figures. *ClustalW* (Larkin *et al.*, 2007) was used for sequence alignment.

3. Results and discussion

3.1. Overall structural properties of the PhBP–GA3 complexes

Both MtPhBP and VrPhBP are monomeric in solution, as they elute as monomers in size-exclusion chromatography (not shown). Likewise, no stable quaternary structure could be predicted from crystal packing by *PISA* (Krissinel & Henrick, 2007). The two complexes crystallize in different space groups, namely $P6_5$ (MtPhBP) and $C2$ (VrPhBP). In both crystals there is one copy of a 1:1 PhBP complex with GA3 in the asymmetric unit. The different crystal packing is reflected in a different Matthews volume (Matthews, 1968) and in different solvent contents: 52.4 and 34.4% for MtPhBP and VrPhBP, respectively. Despite the different levels of hydration, crystals of both complexes diffracted X-rays to very high angles, which allowed refinement of the crystal structures at the very high resolutions of 1.34 and 1.42 \AA , respectively. Owing to the high data resolution, the atomic displacement parameters could be refined anisotropically. The final electron-density maps are of excellent quality in both cases. The entire protein chains starting with Met1 could be modelled with confidence, except for two residues (–GA) at the C-terminus in the case of MtPhBP and three (–GSA) in the case of VrPhBP, which were disordered and were thus omitted from the models. The GA3 ligand in both complexes had superb definition in $F_o - F_c$ electron-density maps phased using the protein atoms only,

and could be modelled without any ambiguity (Fig. 2). Five glycerol molecules from the cryoprotectant solution could be traced in the electron-density maps of the MtPhBP complex. Two of those glycerol molecules are found near the entrance to the internal cavity. One glycerol molecule is hydrogen-bonded to Ser23 and one to Asp47. Another

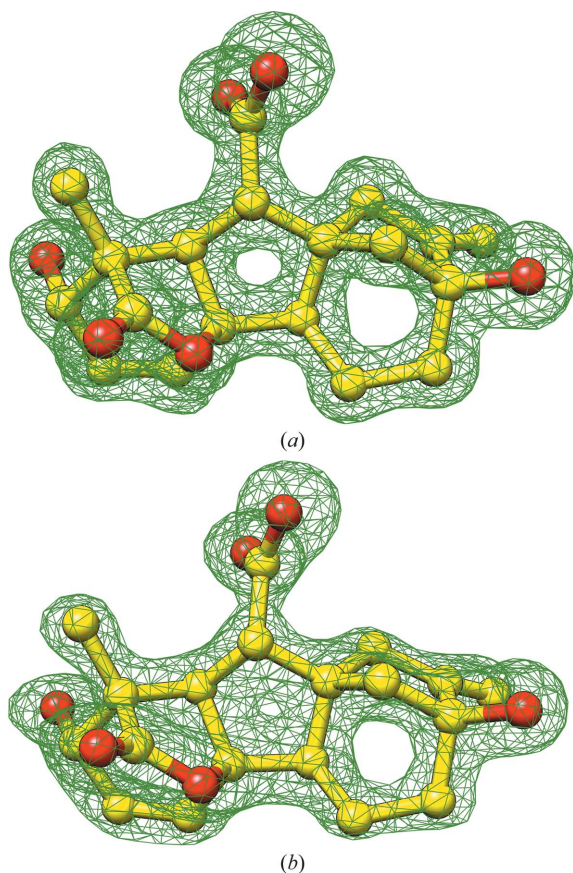


Figure 2
Electron-density maps of gibberellic acid (ball-and-stick representation) bound in the presented complexes with MtPhBP (a) and VrPhBP (b). OMIT $F_o - F_c$ electron-density maps (green mesh) are contoured at the 4σ level.

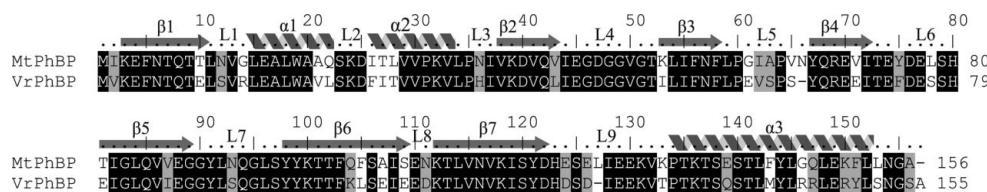


Figure 3
Sequence alignment of MtPhBP and VrPhBP, with annotation of secondary-structure elements as assigned by *DSSP* (Kabsch & Sander, 1983). Loops are labelled L1–L9. Black and grey shadings indicate identical and similar residues, respectively.

glycerol molecule is close to the N-terminus, interacting *via* a water-mediated hydrogen bond with Glu4.

The two protein molecules in this study, MtPhBP and VrPhBP, share 74% sequence identity and 86% similarity. Their sequence alignment with secondary-structure assignment is shown in Fig. 3. Both proteins have the canonical PR-10 fold (Fig. 4*a*), with a seven-stranded antiparallel β -sheet wrapped around the C-terminal helix α 3. The consecutive β -strands are connected by β -hairpins and loops, except for the β 1– β 2 crossover, which is formed by helices α 1 and α 2 that link the edges of the β -sheet. The β -sheet has a highly curved shape induced by eight β -bulges. Overall, the fold of the proteins resembles a right-handed baseball glove, where the odd-numbered loops (L3, L5, L7 and L9) form the ‘fingers’. The two short helices α 1 and α 2 create a V-shaped support for the C-terminal part of the long α 3 helix, which forms the ‘thumb’ of the glove. A single α -helical turn within loop L7, formed by four residues (Gly90–Asn93 in MtPhBP and Gly89–Ser92 in VrPhBP), is disregarded in the following discussion to maintain consistency of secondary-structure

numbering with other PR-10 proteins. In the topology of PR-10 proteins, the internal cavity, which is often a ligand-docking site, is formed between the β -sheet and the α 3 helix. This is also the case for the present PhBP complexes, where the gibberellic acid molecule is located inside this internal cavity in both structures (Figs. 4*b* and 4*c*). The structural details of GA3 docking are discussed in the next section.

The backbones of MtPhBP and VrPhBP are quite similar, as illustrated by the r.m.s. (root-mean-square) deviation between their C $^{\alpha}$ positions of 0.69 Å. Most of the few differences of significance are observed within loops, in particular in loop L9, which is the point of entry for helix α 3 (Fig. 4*a*). Loop L9 is one of the most variable structural elements of the PhBP subfamily of PR-10 proteins. In particular, in the VrPhBP–ZEA complex (Pasternak *et al.*, 2006) loop L9 of chain *A* could not be traced in the electron-density maps (residues 123–129) or was involved in Na $^{+}$ coordination in chains *B* and *C*, while in chain *D* it was visible but metal-free. In contrast, the so-called glycine-rich loop L4 with the sequence motif ${}_{44}(\text{IV})\text{EG}(\text{ND})\text{GG}(\text{PV})\text{GT}_{52}$ is sequentially conserved and structurally rigid as in all other PR-10 structures.

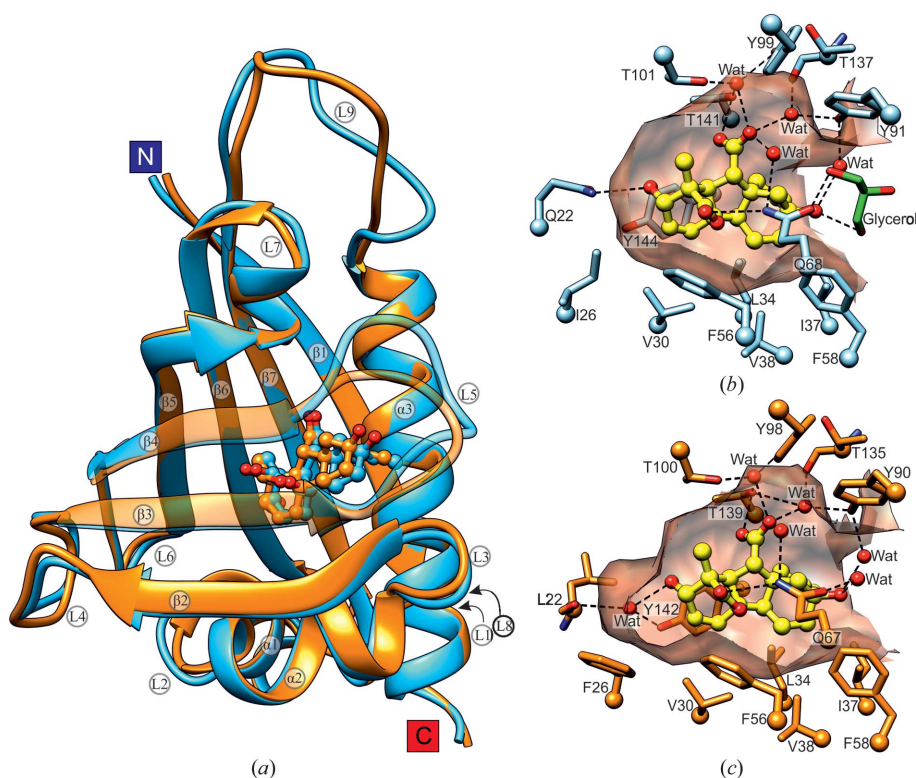


Figure 4

Gibberellic acid binding by PhBP proteins. (*a*) Overall fold of the PhBP proteins (C $^{\alpha}$ superposition) in their GA3 complexes, shown as a cartoon model with MtPhBP in blue and VrPhBP in orange. Secondary-structure elements are numbered according to the PR-10 canon. The N- and C-termini are also marked. Strands β 3 and β 4 and loop L5 are semitransparent to visualize the phytohormone molecule in the internal binding cavity. Note that the GA3 molecules (ball-and-stick representation) are in the same position and orientation in both complexes. Close-up view of the GA3 binding site of MtPhBP (*b*) and VrPhBP (*c*). Hydrogen bonds are shown as dashed lines. The entrance to the internal cavity is on the right side of each panel. Protein surfaces that form the walls of the cavity have been clipped to show a maximum vista and are presented in semitransparent red. C $^{\alpha}$ atoms are highlighted as balls. A glycerol molecule (*b*, green) originates from the cryoprotectant buffer. The backbone C, O and N atoms of the protein chains have been omitted for clarity unless they take part in binding interactions, as is the case for Thr137 in MtPhBP and Leu22 and Thr135 in the VrPhBP complex.

3.2. Structural details of gibberellic acid binding

The GA3 molecules bound to MtPhBP and VrPhBP are found deep in the internal cavities of the proteins. The cavity is the most fascinating structural element of all PR-10 proteins, as it gives rise to an essentially hollow protein core without degrading the mechanical properties of the molecule (Chwastyk *et al.*, 2014). In the present structures several hydrophobic residues shape the walls of the cavities (Figs. 4*b* and 4*c*). These nonpolar residues are perfect partners for interactions with the hydrophobic fragments of the GA3 molecule. The hydrophobic interaction surface of the protein cavity is formed by the side chains of Ile26/Phe26 (MtPhBP/VrPhBP), Val30, Leu34, Ile37, Val38, Phe56 and Phe58. There is also a stabilizing stacking interaction between the π -electrons of the double C1=C2 bond of GA3 and a parallel aromatic ring of Tyr144 (MtPhBP) or Tyr142 (VrPhBP), located within a distance of ~ 4 Å. The conservation of these hydrophobic and Tyr residues strongly suggests that they are required for specific interactions with a ligand, and hence that the complexes observed in our crystals are in all likelihood of biological significance.

Table 2

Hydrogen bonds between GA3 and MtPhBP/VrPhBP, with donor–acceptor distances (Å) in parentheses.

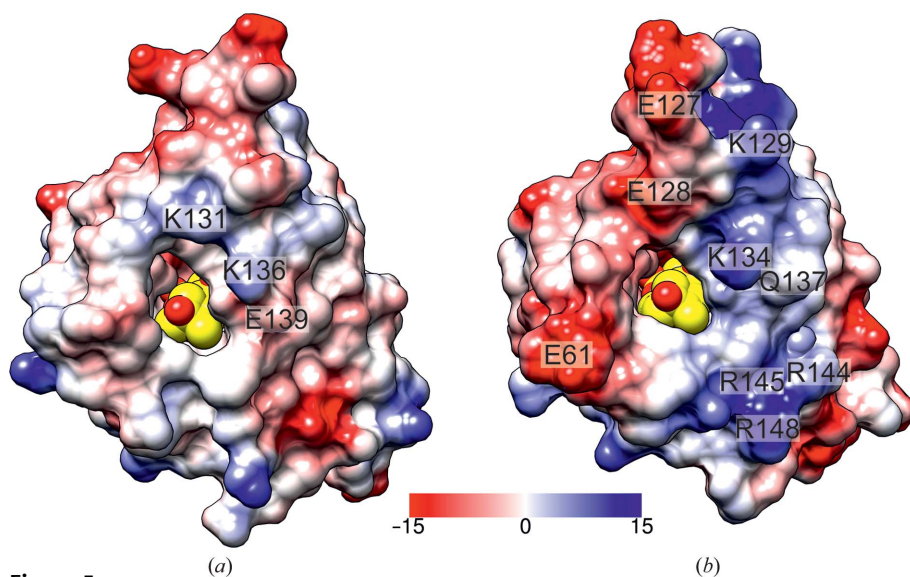
In the case of water-mediated or glycerol-mediated hydrogen bonds only the immediate interactions with the solvent molecules are listed and the protein residues serving as the ultimate docking sites are discussed in the text.

Atom	MtPhBP	VrPhBP
O1	Gln22 N ^ε (3.2)	Wat1† (2.7)
O2	Thr141 O ^γ (2.7)	Thr139 O ^γ (2.7)
O3	Wat1 (2.7)	Wat2 (2.6)
	Wat2 (2.7)	Wat3 (2.7)
	Wat3 (2.7)	Wat4 (2.8)
O5	Glycerol (3.1)	Wat5 (2.6)
	Wat4 (2.7)	Wat6 (3.1)
O6	Gln68 N ^ε (3.1)	Gln67 N ^ε (3.2)

† Labels of water molecules in the table are assigned in each complex sequentially and do not correspond to the numbers in the PDB entries.

In addition to the hydrophobic interactions, several hydrogen bonds are formed between the gibberellic acid and the protein. The GA3 molecule has six O atoms of different chemical character. Two of them belong to the carboxyl group, two are in hydroxyl groups and two other form the lactone moiety. In the present complexes, all of the O atoms of GA3 except for the endocyclic O4 atom interact with the protein *via* hydrogen bonds (Figs. 4*b* and 4*c*, Table 2). Three (Gln22, Gln68 and Thr141) or two (Gln67 and Thr139) residues form direct hydrogen bonds to GA3 in the MtPhBP or VrPhBP complexes, respectively. In addition, there are several solvent-mediated hydrogen bonds that ultimately link the GA3 molecules to Gln68, Tyr91, Tyr99, Thr101 and Thr137 in MtPhBP or to Leu22, Gln67, Tyr90, Tyr98, Thr100, Thr135 and Tyr142 in VrPhBP.

In contrast to the extensive similarities of the interior of the GA3 binding sites, there are significant differences in the


Figure 5

Electrostatic potential surface around the entrance to the internal cavity. MtPhBP (*a*) and VrPhBP (*b*) are viewed in the same orientation to show the GA3 molecule (yellow C-atom spheres) bound inside the cavity. The electrostatic potential is colour-coded according to the scale bar, which is calibrated in $kT e^{-1}$ units.

Table 3

Thermodynamic characterization of the interactions of MtPhBP and VrPhBP with GA3 (gibberellic acid) and ZEA (*trans*-zeatin).

The stoichiometry N , the dissociation constant K_d (μM), the change in enthalpy ΔH (cal mol^{-1}) and the change in entropy ΔS ($\text{cal mol}^{-1} \text{K}^{-1}$) were determined by ITC titrations at pH 5.5 and 7.4.

Protein	MtPhBP		VrPhBP	
	pH 7.4	pH 5.5	pH 7.4	pH 5.5
GA3				
N	†	1.14 ± 0.01	1.17 ± 0.01	1.09 ± 0.01
K_d		13.4 ± 0.5	23 ± 1	6.1 ± 0.2
ΔH		-3727 ± 35	-8054 ± 120	-9039 ± 63
ΔS		9.6	-6.3	-7.0
ZEA				
K_{d1}	‡	‡	76 ± 1	181 ± 25
ΔH_1			-5186 ± 59	-6645 ± 493
ΔS_1			1.16	-5.6
K_{d2}			67 ± 1	94 ± 9
ΔH_2			1903 ± 71	4348 ± 573
ΔS_2			25.6	33.2

† No heat effect. ‡ A very small enthalpy change ($\Delta H < 800 \text{ cal mol}^{-1}$) and a high noise-to-signal ratio precluded reliable estimation of the derived parameters.

surface areas adjacent to the entrance leading to the cavity (Fig. 5). In MtPhBP this portion of the protein surface is only slightly charged, with only three residues, Lys131, Lys136 and Glu139, contributing to the electrostatic potential of this side of the protein surface. Moreover, the positive charge of Lys136 is compensated by the negative charge of the Glu139 side chain. The situation is very different in the VrPhBP protein, where nine charged residues (Glu61, Glu127, Glu128, Lys129, Lys134, Gln137, Arg144, Arg145 and Arg148) surround the entrance to the internal cavity. Moreover, these highly polar residues are segregated sidewise, meaning that the negative charge is concentrated on one side of the cavity (left in the view presented in Fig. 5*b*), while the positive charge is concentrated on the opposite (right) side. The difference in the surface-charge distribution in the vicinity of the entrance to the cavity explains why, according to the ITC experiments (see below), VrPhBP is capable of binding GA3 (as an anion) at pH 7.4 whereas MtPhBP is not.

The binding mode of gibberellic acid in PhBP complexes is quite different from that reported for the gibberellin receptor GID1 (Murae *et al.*, 2008). There are, however, three similar aspects between the PhBP and GID1 complexes that need to be addressed. Firstly, in both cases the GA3 molecule is oriented in the ligand-binding site in such a way that the lactone moiety is buried deeply in the cavity, whereas the C13 OH group points towards solvent/cytoplasm. Secondly, the docking of GA3 occurs *via* only three direct hydrogen bonds, whereas the remaining

heteroatoms of GA3 interact with the protein through water-mediated contacts. Thirdly, the endocyclic O4 atom of the lactone ring is the only GA3 O atom that does not form any hydrogen bonds.

It is noted that a GA3 restraint library based on the CSD structure BUWZAU corresponds to the acidic (protonated) form of the carboxylic group of the ligand. Because there is no

crystal structure of GA3 in the anionic form, this library was used with both carboxylate C—O bond lengths set to 1.254 Å. In the crystal structure of MtPhBP, which was formed at pH 6.5, the ligand is expected to be in the anionic state because the pK_a value of gibberellic acid is 4.0 (Tomlin, 1997). In the structure of the VrPhBP complex formed at pH 4.0, one could theoretically expect a half-protonated carboxylic group of

GA3. However, since refinement at even 1.42 Å resolution does not allow carboxylic/carboxylate groups to be unambiguously distinguished and since the pattern of hydrogen bonds at the GA3 molecule is not univocal, the ligand has been treated as an anion in this paper.

3.3. Thermodynamic characterization of PhBP–GA3 and PhBP–ZEA interactions

Table 3 presents the results of ITC titrations of VrPhBP and MtPhBP with GA3 and ZEA conducted at two pH conditions: 7.4 and 5.5. The assays revealed pH dependence of the binding of these two phytohormones, as well as confirming the specificity of the proteins towards GA3 binding and the nonspecific character of ZEA binding. Binding stoichiometry (N ligand molecules per one protein molecule) can be determined from titration curves that have sigmoidal shape, and this is the case for titrations of MtPhBP and VrPhBP with GA3 at pH 5.5 (Fig. 6) and for VrPhBP also at pH 7.4. For the PhBP–GA3 complexes, N can be (slightly) rounded down to 1, which agrees with the stoichiometry observed in the crystal structures. The small discrepancies (1.14 for MtPhBP and 1.09 for VrPhBP) can be attributed to the presence of a small amount of inactive protein molecules that for one reason or another (e.g. misfolding, denaturation or precipitation) lost their binding capability.

MtPhBP binds GA3 strongly at pH 5.5, with a dissociation constant K_d of $13.4 \pm 0.5 \mu\text{M}$. No significant heat change was

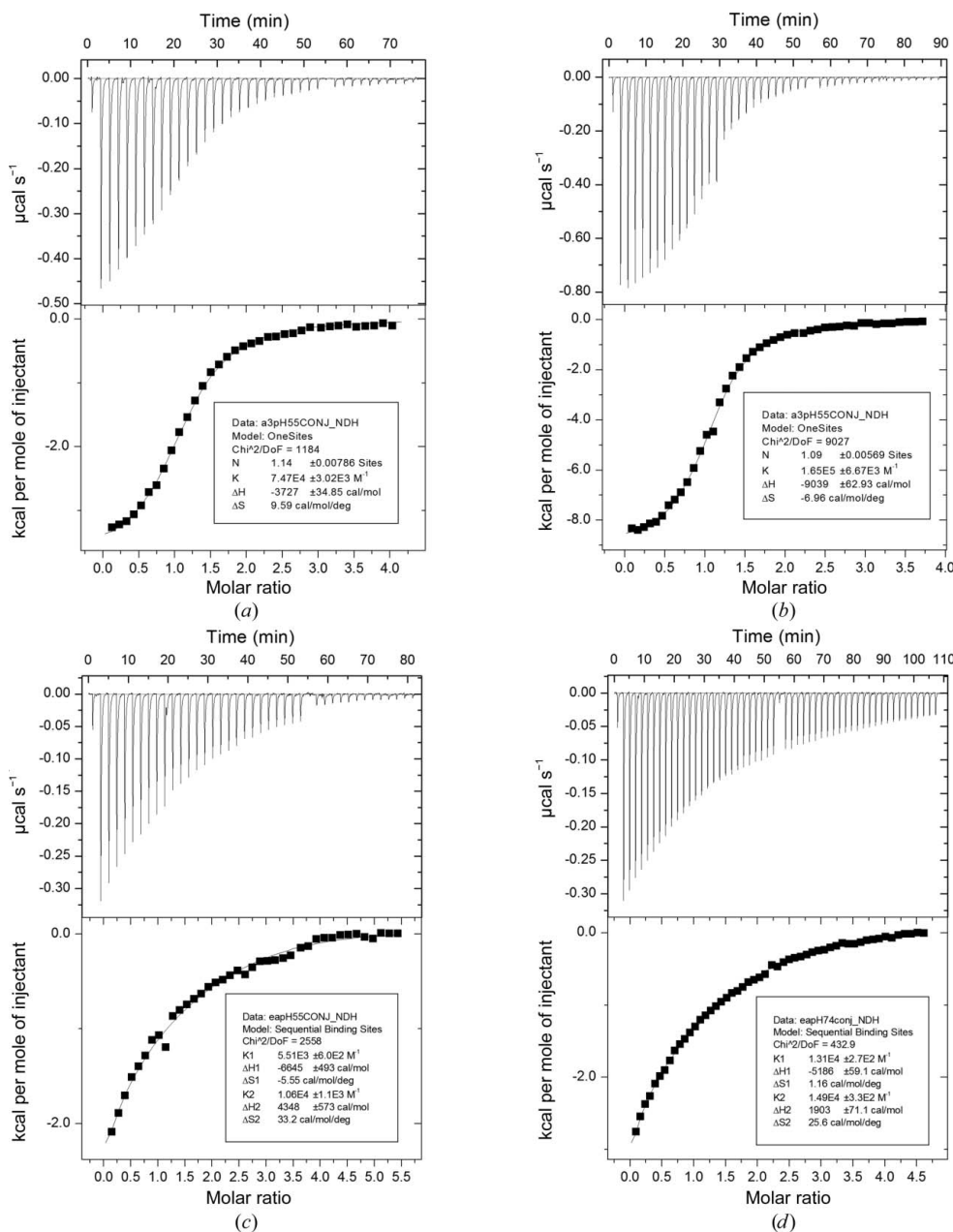


Figure 6 Calorimetric titrations of PhBP proteins with GA3 (*a, b*) and ZEA (*c, d*). The top plot of each panel shows the raw heat data obtained from ~40 consecutive injections of GA3 or ZEA into the sample cell (200 µl) containing MtPhBP (*a*) or VrPhBP (*b, c, d*). The experimental conditions were as follows: (*a*) 105 µM MtPhBP, 1.5 mM GA3, pH 5.5, (*b*) 80 µM VrPhBP, 0.9 mM GA3, pH 5.5, (*c*) 93 µM VrPhBP, 1.5 mM ZEA, pH 5.5 and (*d*) 100 µM VrPhBP, 1 mM ZEA, pH 7.4. The titrations were performed at 290.15 K in a buffer composed of (*a, b, c*) 25 mM MES pH 5.5 or (*d*) 25 mM HEPES pH 7.4 supplemented with 150 mM NaCl. At the bottom of each panel, the binding isotherm has been created by plotting the heat peak areas against the molar ratio of GA3 or ZEA added to the protein present in the cell. The line represents the best fit of a model with one independent binding site (*a, b*) or two sequential binding sites (*c, d*). One of three experiments is shown in each panel.

observed during the titration of MtPhBP with GA3 at pH 7.4. The very small change in enthalpy ($\Delta H < 800 \text{ cal mol}^{-1}$) during titration of MtPhBP with ZEA at pH 5.5 and 7.4 makes these measurements very unreliable and suggests a nonspecific interaction. All of these observations strongly suggest that MtPhBP is a very specific GA3 binder but only under slightly acidic conditions.

In the case of the VrPhBP protein, as previously reported by Pasternak *et al.* (2006), titration with ZEA yields hyperbolic curves despite the use of a high protein concentration ($\sim 100 \mu\text{M}$) in the sample cell (Figs. 6*c* and 6*d*). Our results (Table 3) are in very good agreement with the values reported for the VrPhBP–ZEA interaction ($K_d = 106 \pm 12 \mu\text{M}$ at pH 6.5) by Pasternak *et al.* (2006). Both experiments are suggestive of a rather nonspecific character of ZEA binding. Because of the hyperbolic shape of the titration curves, the stoichiometry had to be fixed before fitting other model parameters. The stoichiometry of the VrPhBP–ZEA complex was assigned on the basis of the crystal structure (PDB entry 2flh), in which the protein:ligand ratio is 1:2 in three complex molecules and 1:1 in one complex molecule in the asymmetric unit. The binding model was selected as sequential binding sites as the second ligand molecule, bound near the entrance to the binding cavity, can only be docked after the deeper binding site has been filled. From this model, we can conclude that the binding of the first ZEA molecule is enthalpy-driven, whereas the binding of the second molecule leads to an entropy increase (Table 3). Fitting the data with a one set of binding sites model and a 1:1 stoichiometry yields a K_d value for ZEA binding that is only slightly lower than K_{d2} of the sequential binding sites model ($66 \pm 5 \mu\text{M}$ at pH 5.5 and $60 \pm 3 \mu\text{M}$ at pH 7.4).

On the other hand, our results disagree with those reported by Zawadzki *et al.* (2010) at pH 7.2, where the K_d values for VrPhBP were estimated at 409 ± 32 and $383 \pm 15 \mu\text{M}$ for the interactions with ZEA and GA3, respectively. These authors, however, used a very large chromophore to label the ligands in their assays, which makes their results controversial, especially when the limited volume of the internal cavity is taken into account. Based on our results, VrPhBP shows a much higher affinity for GA3. Our K_d values of 23 ± 1 and $6.1 \pm 0.2 \mu\text{M}$ at pH 7.4 and 5.5, respectively, indicate strong interactions.

The ITC data clearly demonstrate that proteins from two different plants previously classified as Cytokinin-Specific Binding Proteins (CSBPs) are in fact relatively weak *trans*-zeatin binders and show a much higher affinity towards gibberellic acid. At acidic pH their affinity for GA3 is either (i) additionally increased, as is the case for VrPhBP, where the K_d decreases from ~ 23 to $6 \mu\text{M}$ between pH 7.4 and 5.5, or (ii) switches from an absence of binding at pH 7.4 to enhanced binding at pH 5.5 ($K_d = 13.4 \mu\text{M}$), as is the case for MtPhBP. Additionally, while GA3 binding by VrPhBP is exclusively enthalpy-driven, in the case of MtPhBP enthalpy and entropy drive the association process almost equally. This can be explained by the different chemical character of the residues surrounding the entrance to the internal cavity (Fig. 5). Specifically, charged residues in VrPhBP form hydrogen

bonds to the ligand, contributing to an enthalpic effect, while hydrophobic interactions in the case of MtPhBP contribute to an entropy change.

The specificity of VrPhBPs towards gibberellic acid was additionally confirmed by a competitive binding (displacement) assay at pH 5.5, in which the protein was titrated with GA3 in the presence of *trans*-zeatin, added in advance at a concentration assuring saturation. The presence of a competing ligand (ZEA) changed the K_d of GA3 binding from 6 to $23 \mu\text{M}$ and ΔH from -9039 to $-4683 \text{ cal mol}^{-1}$. The titration curve remained sigmoidal. In addition, the binding parameters for ZEA in the first binding site of VrPhBP ($K_{d\text{ZEA}}$ and ΔH_{ZEA}) were tested using this displacement assay. From the apparent K_d ($K_{d\text{app}}$) and ΔH (ΔH_{app}) one can determine the binding parameters of the low-affinity ligand (ZEA; *i.e.* the ligand being displaced) using (1) and (2) derived from Zhang & Zhang (1998),

$$K_{d\text{ZEA}} = \frac{K_{d\text{GA3}}}{K_{d\text{app}}} \times [\text{ZEA}], \quad (1)$$

$$\Delta H_{\text{ZEA}} = (\Delta H_{\text{GA3}} - \Delta H_{\text{app}}) \left(1 + \frac{K_{d\text{ZEA}}}{[\text{ZEA}]} \right), \quad (2)$$

where [ZEA] is the concentration of the low-affinity ligand (ZEA) and $K_{d\text{GA3}}$ and ΔH_{GA3} are the dissociation constant and enthalpy change, respectively, obtained from titration with the high-affinity ligand only (GA3; Table 3). The calculated values of $113 \mu\text{M}$ for $K_{d\text{ZEA}}$ and of $-5489 \text{ cal mol}^{-1}$ for ΔH_{ZEA} are in very good agreement with the values obtained in the direct titration experiment (Table 3).

It is very interesting to note that the dissociation constants for PhBP–GA3 interactions, 13.4 and $6.1 \mu\text{M}$ at pH 5.5, are close to the value of $4 \mu\text{M}$ reported for the gibberellin receptor *GID1* (Ueguchi-Tanaka *et al.*, 2005). However, the above value of K_d for the *GID1* receptor was determined at pH 7.6, which suggests that the PhBPs may be physiologically relevant gibberellin binders that are switched on by local pH decreases. While it is tempting to suggest that GA3 binding becomes relevant under acidic conditions, we note that there is no experimental evidence to suggest vacuolar or endosomal localization of the PhBP proteins.

3.4. PhBP proteins are adapted to bind gibberellic acid more potently than *trans*-zeatin

In the crystal structure of VrPhBP in complex with *trans*-zeatin (PDB entry 2flh), the phytohormone was bound in three different modes in the four copies of the protein molecule in the asymmetric unit (Pasternak *et al.*, 2006). Two instances of a head-to-head orientation of a tandem of ligand molecules (with adenine-ring stacking), in chains *A* and *D*, are almost identical. In chain *B* two ligands are bound in a head-to-tail fashion, whereas in chain *C* there is only one ZEA molecule inside the cavity. This binding diversity even within one crystal structure strongly suggests that *trans*-zeatin is not an optimal, and perhaps also not a biologically relevant, ligand for the protein. The picture with gibberellic acid binding,

where a single GA3 molecule is bound in exactly the same manner by two ‘CSBP’ orthologues, is certainly more convincing as biologically significant. Of course, this does not preclude the possibility that some other as yet unknown ligands (not only phytohormones) could also be binding partners of the PhBP proteins, especially in view of the large diversity of plant metabolites and signalling molecules and the documented ability of PR-10 proteins to bind different ligands (Fernandes *et al.*, 2013). However, at the present moment GA3 appears to be the best ligand for the PhBP subfamily.

The conformation of the VrPhBP protein is practically insensitive to the binding of either of the phytohormones, GA3 or ZEA. This is best illustrated by the small r.m.s.d. values on comparing the C α atoms of the VrPhBP–GA3 complex with those of the VrPhBP–ZEA complexes, which were calculated in UCSF Chimera (Pettersen *et al.*, 2004) as 0.55 Å (for chains C and D of PDB entry 2flh) or 0.64 Å (chains A and B). The volume of the internal cavity is also unchanged and is calculated using SURFNET (Laskowski, 1995) as 915–950 Å³. Since there is no crystal structure of a PhBP protein without a ligand, it is not possible to tell how much conformational change is needed for the binding of a given phytohormone. However, from the analysis of similar situations with other PR-10 proteins (Fernandes *et al.*, 2008; Sliwiak *et al.*, 2014) one can quite safely assume that the structural adjustment of the PhBP proteins is also minimal.

Our results have another important implication. For a number of years it has been postulated that PR-10 proteins show higher intraspecific than interspecific conservation (Wen *et al.*, 1997; Finkler *et al.*, 2005; Schenk *et al.*, 2009; Lebel *et al.*, 2010), and this assumption has made PR-10 proteins very good phylogenetic markers. Our work shows that the PhBP subfamily may be an exception in this context and that two proteins from different organisms can actually have conserved function.

4. Conclusions and outlook

This paper describes the crystal structures of two proteins from a PR-10 subfamily, originally classified as Cytokinin-Specific Binding Proteins (CSBP), in complex with a completely different phytohormone, gibberellic acid (GA3). These proteins bind GA3 strongly and specifically with a 1:1 stoichiometry, and the binding mode of this phytohormone is conserved. The crystallographic observations are corroborated by calorimetric experiments showing that the dissociation constants of the GA3 complexes are in the low micromolar range at pH 5.5. From this experimental evidence, supplemented with the observation that binding of a cytokinin ligand (*trans*-zeatin) is nonspecific, much weaker or absent altogether, we propose a revision of the annotation of these proteins as phytohormone-binding proteins (PhBP) to reflect their more likely (and more general) biological function.

The complexes with gibberellic acid described in this paper do not explain the universal role of PR-10 proteins. Nevertheless, the presented results show the PhBP subfamily in the context of gibberellic acid binding, which has not been

considered before. The PhBP proteins share a low level of sequence identity (~20%) with classic PR-10 proteins and they appear to have evolved to bind gibberellic acid and perhaps other gibberellins as well. The PhBPs could have evolved from an unknown common ancestor from which the abscisic acid (ABA) receptors have also originated. The ABA receptors have the PR-10 fold (Nishimura *et al.*, 2009; Santiago *et al.*, 2009) despite only marginal sequence identity. The assumption of a common ancestor could explain why these evolutionarily very distant proteins have retained the same overall fold. Apparently, it is a perfect fold for binding small-molecule, largely hydrophobic ligands such as phytohormones.

Based on a BLAST search (Altschul *et al.*, 1997), the PhBP homologues are only present in legume plants. This observation, however, has not been explained from a functional point of view. The strikingly low expression levels of PhBPs might suggest their biological role in plant hormone signalling pathways, as phytohormones are also present at very low concentrations.

Financial support for this project was provided by the European Union within the European Regional Developmental Fund and by the Polish Ministry of Science and Higher Education (grant Nos. NN 301 003739 and NN 301 204233). The research leading to these results received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) BioStruct-X under grant agreement No. 283570. We thank Alina Kasperska for excellent technical assistance with recombinant protein production and purification.

References

- Adams, P. D. *et al.* (2010). *Acta Cryst.* **D66**, 213–221.
- Afonine, P. V., Grosse-Kunstleve, R. W., Echols, N., Headd, J. J., Moriarty, N. W., Mustyakimov, M., Terwilliger, T. C., Urzhumtsev, A., Zwart, P. H. & Adams, P. D. (2012). *Acta Cryst.* **D68**, 352–367.
- Allen, F. H. (2002). *Acta Cryst.* **B58**, 380–388.
- Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997). *Nucleic Acids Res.* **25**, 3389–3402.
- Bradford, M. M. (1976). *Anal. Biochem.* **72**, 248–254.
- Bujacz, G., Pasternak, O., Fujimoto, Y., Hashimoto, Y., Sikorski, M. M. & Jaskolski, M. (2003). *Acta Cryst.* **D59**, 522–525.
- Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. & Richardson, D. C. (2010). *Acta Cryst.* **D66**, 12–21.
- Chwastyk, M., Jaskolski, M. & Cieplak, M. (2014). *FEBS J.* **281**, 416–429.
- Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* **D66**, 486–501.
- Fernandes, H., Bujacz, A., Bujacz, G., Jelen, F., Jasinski, M., Kachlicki, P., Otlewski, J., Sikorski, M. M. & Jaskolski, M. (2009). *FEBS J.* **276**, 1596–1609.
- Fernandes, H., Michalska, K., Sikorski, M. & Jaskolski, M. (2013). *FEBS J.* **280**, 1169–1199.
- Fernandes, H., Pasternak, O., Bujacz, G., Bujacz, A., Sikorski, M. M. & Jaskolski, M. (2008). *J. Mol. Biol.* **378**, 1040–1051.
- Finkler, C., Giacomet, C., Muschner, V. C., Salzano, F. M. & Freitas, L. B. (2005). *Genetica*, **124**, 117–125.
- Fleet, C. M. & Sun, T.-P. (2005). *Curr. Opin. Plant Biol.* **8**, 77–85.
- Fujimoto, Y., Nagata, R., Fukasawa, H., Yano, K., Azuma, M., Iida, A., Sugimoto, S., Shudo, K. & Hashimoto, Y. (1998). *Eur. J. Biochem.* **258**, 794–802.

- Kabsch, W. (2010). *Acta Cryst.* **D66**, 125–132.
- Kabsch, W. & Sander, C. (1983). *Biopolymers*, **22**, 2577–2637.
- Kofler, S., Asam, C., Eckhard, U., Wallner, M., Ferreira, F. & Brandstetter, H. (2012). *J. Mol. Biol.* **422**, 109–123.
- Krissinel, E. & Henrick, K. (2007). *J. Mol. Biol.* **372**, 774–797.
- Kutschabsky, L. & Gunter, A. (1983). *J. Chem. Soc. Perkin Trans. 1*, pp. 1653–1655.
- Langer, G., Cohen, S. X., Lamzin, V. S. & Perrakis, A. (2008). *Nature Protoc.* **3**, 1171–1179.
- Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J. & Higgins, D. G. (2007). *Bioinformatics*, **23**, 2947–2948.
- Laskowski, R. A. (1995). *J. Mol. Graph.* **13**, 323–330.
- Lebel, S., Schellenbaum, P., Walter, B. & Maillot, P. (2010). *BMC Plant Biol.* **10**, 184.
- Loon, L. C. van, Rep, M. & Pieterse, C. M. (2006). *Annu. Rev. Phytopathol.* **44**, 135–162.
- Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
- McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Cryst.* **40**, 658–674.
- Moriarty, N. W., Grosse-Kunstleve, R. W. & Adams, P. D. (2009). *Acta Cryst.* **D65**, 1074–1080.
- Murase, K., Hirano, Y., Sun, T.-P. & Hakoshima, T. (2008). *Nature (London)*, **456**, 459–463.
- Nishimura, N., Hitomi, K., Arvai, A. S., Rambo, R. P., Hitomi, C., Cutler, S. R., Schroeder, J. I. & Getzoff, E. D. (2009). *Science*, **326**, 1373–1379.
- Pasternak, O., Bujacz, G. D., Fujimoto, Y., Hashimoto, Y., Jelen, F., Otlewski, J., Sikorski, M. M. & Jaskolski, M. (2006). *Plant Cell*, **18**, 2622–2634.
- Peng, J., Richards, D. E., Hartley, N. M., Murphy, G. P., Devos, K. M., Flintham, J. E., Beales, J., Fish, L. J., Worland, A. J., Pelica, F., Sudhakar, D., Christou, P., Snape, J. W., Gale, M. D. & Harberd, N. P. (1999). *Nature (London)*, **400**, 256–261.
- Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C. & Ferrin, T. E. (2004). *J. Comput. Chem.* **25**, 1605–1612.
- Ruszkowski, M., Szpotkowski, K., Sikorski, M. & Jaskolski, M. (2013). *Acta Cryst.* **D69**, 2365–2380.
- Santiago, J., Dupeux, F., Round, A., Antoni, R., Park, S.-Y., Jamin, M., Cutler, S. R., Rodriguez, P. L. & Márquez, J. A. (2009). *Nature (London)*, **462**, 665–668.
- Santner, A., Calderon-Villalobos, L. I. & Estelle, M. (2009). *Nature Chem. Biol.* **5**, 301–307.
- Schenk, M. F., Cordewener, J. H., America, A. H., Van't Westende, W. P., Smulders, M. J. & Gilissen, L. J. (2009). *BMC Plant Biol.* **9**, 24.
- Schwechheimer, C. (2008). *Curr. Opin. Plant Biol.* **11**, 9–15.
- Schwechheimer, C. & Willige, B. C. (2009). *Curr. Opin. Plant Biol.* **12**, 57–62.
- Sliwiak, J., Jaskolski, M., Dauter, Z., McCoy, A. J. & Read, R. J. (2014). *Acta Cryst.* **D70**, 471–480.
- Tomlin, C. D. S. (1997). Editor. *The Pesticide Manual: A World Compendium*, 11th ed., p. 639. Alton: BCPC.
- Ueguchi-Tanaka, M., Ashikari, M., Nakajima, M., Itoh, H., Katoh, E., Kobayashi, M., Chow, T.-Y., Hsing, Y.-I., Kitano, H., Yamaguchi, I. & Matsuoka, M. (2005). *Nature (London)*, **437**, 693–698.
- Wen, J., Vanek-Krebitz, M., Hoffmann-Sommergruber, K., Scheiner, O. & Breiteneder, H. (1997). *Mol. Phylog. Evol.* **8**, 317–333.
- Yabuta, T. & Sumitaki, Y. (1938). *J. Agric. Chem. Soc. Jpn*, **14**, 1526.
- Yamaguchi, S. (2008). *Annu. Rev. Plant Biol.* **59**, 225–251.
- Zawadzki, P., Slósarek, G., Boryski, J. & Wojtaszek, P. (2010). *Biol. Chem.* **391**, 43–53.
- Zhang, Y.-L. & Zhang, Z.-Y. (1998). *Anal. Biochem.* **261**, 139–148.

CO-AUTHORS' STATEMENTS

Joanna Śliwiak
Instytut Chemii Bioorganicznej PAN
Noskowskiego 12/14
61-704 Poznań

Poznań, 28 lutego 2017

Oświadczenie o współautorstwie:

Oświadczam, iż mój udział w publikacji:

Sliwiak, J., Dolot, R., Michalska, K., Szpotkowski, K., Bujacz, G., Sikorski, M. & Jaskolski, M. (2016). Crystallographic and CD probing of ligand-induced conformational changes in a plant PR-10 protein. *J. Struct. Biol.* 193, 55-66.

polegał na:

- oczyszczeniu białka LIPR-10.1A
- przeprowadzeniu współkryształizacyjnych testów przesiewowych tego białka z różnymi ligandami
- otrzymaniu kryształu, rozwiązaniu i udokładnieniu struktury białka LIPR-10.1A w kompleksie niewysyconym *trans*-zeatyną
- udokładnieniu struktury białka LIPR-10.1A w formie wolnej, otrzymanej i rozwiązanej przez dr. Karolinę Michalską
- analizie porównawczej powyższych struktur ze strukturą kompleksu białka LIPR-10.1A nasyconego *trans*-zeatyną otrzymaną przez dr. Rafała Dolota oraz udokładnioną przez dr. Karolinę Michalską
- zaprojektowaniu eksperymentu miareczkowania białka LIPR-10.1A z wykorzystaniem dichroizmu kołowego oraz przeprowadzenie eksperymentu wraz z dr. Kamilem Szpotkowskim, który dodatkowo przygotował podrozdział publikacji dotyczący tej metody
- przygotowaniu figur oraz szkicu manuskryptu

Z poważaniem



Karolina Michalska
Argonne National Laboratory
9700 South Cass Ave, Argonne IL 60439, USA

Argonne, 27-02-2017

Oświadczenie o współautorstwie:

Oświadczam, iż mój udział w publikacji:

Sliwiak, J., Dolot, R., Michalska, K., Szpotkowski, K., Bujacz, G., Sikorski, M. & Jaskolski, M. (2016).
Crystallographic and CD probing of ligand-induced conformational changes in a plant PR-10
protein. *J. Struct. Biol.* 193, 55-66.

polegał na procesowaniu danych dyfrakcyjnych i udokładnieniu struktury białka LLPR-
10.1A w kompleksie z zeatyną.

Z poważaniem



Poznań, 27.02.2017

dr Kamil Szpotkowski
Zakład Krystalografii
Instytut Chemii Bioorganicznej PAN
ul. Z. Noskowskiego 12/14
61-704 Poznań

Oświadczenie o współautorstwie:

Oświadczam, iż mój udział w publikacji:

Sliwiak, J., Dolot, R., Michalska, K., Szpotkowski, K., Bujacz, G., Sikorski, M. & Jaskolski, M. (2016).
Crystallographic and CD probing of ligand-induced conformational changes in a plant PR-10
protein. *J. Struct. Biol.* 193, 55-66.

polegał na rejestracji danych dichroizmu kołowego oraz ich analizie i opracowaniu
wspólnie z doktorantem. Przygotowałem również fragment tekstu publikacji dotyczący
badań dichroizmu kołowego wraz z doktorantem.

Z poważaniem



dr hab. Michał M. Sikorski prof. IChB PAN
Pracownia Inżynierii Białek
Zakład Krystalografii – CBB
Instytut Chemii Bioorganicznej PAN
Noskowskiego 12/14
61-704 Poznań

Poznań, 28 lutego 2017

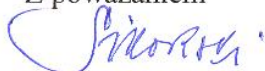
Oświadczenie o współautorstwie:

Oświadczam, iż mój udział w publikacji:

Sliwiak, J., Dolot, R., Michalska, K., Szpotkowski, K., Bujacz, G., Sikorski, M. & Jaskolski, M. (2016).
Crystallographic and CD probing of ligand-induced conformational changes in a plant PR-10
protein. *J. Struct. Biol.* 193, 55-66.

polegał na opracowaniu procedury oczyszczania preparatu białka i udziale w przygotowaniu
manuskryptu.

Z poważaniem



Michał M. Sikorski

Prof. Mariusz Jaskólski
Zakład Krystalografii - CBB
Instytut Chemii Bioorganicznej PAN
w Poznaniu

26 lutego 2017r.

Oświadczenie o współautorstwie:

Oświadczam, że mój udział w publikacji:

Sliwiak, J., Dolot, R., Michalska, K., Szpotkowski, K., Bujacz, G., Sikorski, M. & Jaskolski, M. (2016).
Crystallographic and CD probing of ligand-induced conformational changes in a plant PR-10
protein. *J. Struct. Biol.* 193, 55-66.

polegał na:

- koordynacji projektu
- wkładzie do analizy wyników krystalograficznych
- koordynacji przygotowania publikacji.



prof. dr hab. Mariusz Jaskólski

Prof. Mariusz Jaskólski
Department of Crystallography - Center for Biocrystallographic Research
Institute of Bioorganic Chemistry, Polish Academy of Sciences
Poznań, Poland

February 26, 2017

To Whom It May Concern:

Regarding the PhD thesis of Joanna Śliwiak.

This is to certify my contribution to the following publication of which I am a co-author:

Sliwiak, J., Jaskolski, M., Dauter, Z., McCoy, A. J. & Read, R. J. (2014). Likelihood-based molecular-replacement solution for a highly pathological crystal with tetartohedral twinning and sevenfold translational noncrystallographic symmetry. *Acta Crystallogr. D70*, 471-480.

My contribution consisted in:

- coordination of the project
- participation in the analysis of the results
- participation in manuscript preparation



Prof. Mariusz Jaskólski, PhD

Zbigniew Dauter
Argonne National Laboratory
Argonne, IL 60439
USA

2.03.2017

To Whom It May Concern:

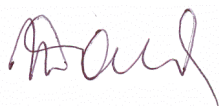
Regarding PhD thesis of Joanna Śliwiak.

Hereby I certify that my contribution to the following publication of which I was a co-author:

Sliwiak, J., Jaskolski, M., Dauter, Z., McCoy, A. J. & Read, R. J. (2014). Likelihood-based molecular-replacement solution for a highly pathological crystal with tetartohedral twinning and sevenfold translational noncrystallographic symmetry. *Acta Crystallogr. D70*, 471-480,

was related to the provision to Mrs. Sliwiak of the possibility of collecting diffraction data from crystals of Hyp-1 protein complex with ANS at a synchrotron beam line of the Argonne National Laboratory.

Sincerely



Zbigniew Dauter, Ph.D.

Chief, Synchrotron Radiation Research Section, MCL, NCI
Section Editor, Acta Crystallographica F
Argonne National Laboratory, Bldg. 202, Rm. Q142
9700 S. Cass Ave., Argonne, IL 60439
Tel: (630)-252-3960
Fax: (630)-252-3622
Email: dauter@anl.gov

To Whom It May Concern:

Regarding PhD thesis of Joanna Śliwiak.

This is letter to certify my contribution to the following publication of which I was a co-author.

Sliwiak, J., Jaskolski, M., Dauter, Z., McCoy, A. J. & Read, R. J. (2014). Likelihood-based molecular-replacement solution for a highly pathological crystal with tetartohedral twinning and sevenfold translational noncrystallographic symmetry. *Acta Crystallogr. D70*, 471-480.

Phaser software

(Description of contribution)

Airlie McCoy

(Signature)

Randy J Read

Department of Haematology
University of Cambridge
Cambridge Institute for Medical Research
Hills Road
Cambridge CB2 0XY
UK

27 February 2017

To Whom It May Concern:

Regarding PhD thesis of Joanna Śliwiak.

This is letter to certify my contribution to the following publication of which I was a co-author.

Sliwiak, J., Jaskolski, M., Dauter, Z., McCoy, A. J. & Read, R. J. (2014). Likelihood-based molecular-replacement solution for a highly pathological crystal with tetartohedral twinning and sevenfold translational noncrystallographic symmetry. *Acta Crystallogr. D70*, 471-480.

My contributions were to solve the molecular replacement problem for this crystal structure (using the data provided by other coauthors), to contribute (with Airlie McCoy) to developing the required underlying theory for the structure solution approach, and to take part in all aspects of writing the paper.



Randy J Read

Zbigniew Dauter
Argonne National Laboratory
Argonne, IL 60439
USA

2.03.2017

To Whom It May Concern:

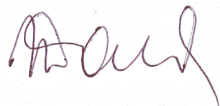
Regarding PhD thesis of Joanna Śliwiak.

Hereby I certify that my contribution to the following publication of which I was a co-author:

Sliwiak, J., Dauter, Z., Kowiel, M., McCoy, A. J., Read, R. J. & Jaskolski, M. (2015). ANS complex of St. John's wort PR-10 protein with 28 copies in the asymmetric unit: a fiendish combination of pseudosymmetry with tetartohedral twinning. *Acta Crystallogr.* D71, 829-843.

was related to the provision to Mrs. Sliwiak of the possibility of collecting diffraction data from crystals of Hyp-1 protein complex with ANS at a synchrotron beam line of the Argonne National Laboratory.

Sincerely



Zbigniew Dauter, Ph.D.

Chief, Synchrotron Radiation Research Section, MCL, NCI
Section Editor, *Acta Crystallographica F*
Argonne National Laboratory, Bldg. 202, Rm. Q142
9700 S. Cass Ave., Argonne, IL 60439
Tel: (630)-252-3960
Fax: (630)-252-3622
Email: dauter@anl.gov

Dr. Marcin Kowiel
Instytut Chemii Bioorganicznej PAN
Noskowskiego 12/14
61-704 Poznań

Poznań, 28 lutego 2017

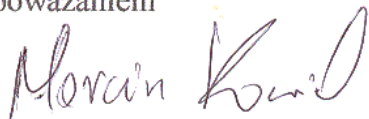
Oświadczenie o współautorstwie:

Oświadczam, iż mój udział w publikacji:

Sliwiak, J., Dauter, Z., Kowiel, M., McCoy, A. J., Read, R. J. & Jaskolski, M. (2015). ANS complex of St John's wort PR-10 protein with 28 copies in the asymmetric unit: a fiendish combination of pseudosymmetry with tetartohedral twinning. *Acta Crystallogr. D* 71, 829-843.

polegał na analizie obsadzenia cząsteczek ANS wokół cząsteczek białka w modelu struktury krystalicznej i związku położenia cząsteczek związku ANS z zaobserwowanym efektem pseudosymetrii oraz translacji niekrystalograficznej w modelu.

Z poważaniem



Airlie McCoy

16-03-2017

Cambridge Institute for Medical Research

To Whom It May Concern:

Regarding PhD thesis of Joanna Śliwiak.

This is letter to certify my contribution to the following publication of which I was a co-author.

Sliwiak, J., Dauter, Z., Kowiel, M., McCoy, A. J., Read, R. J. & Jaskolski, M. (2015). ANS complex of St. John's wort PR-10 protein with 28 copies in the asymmetric unit: a fiendish combination of pseudosymmetry with tetartohedral twinning. *Acta Crystallogr.* D71, 829-843.

Phase software

(Description of contribution)

Airlie McCoy

(Signature)

Randy J Read

Department of Haematology
University of Cambridge
Cambridge Institute for Medical Research
Hills Road
Cambridge CB2 0XY
UK

27 February 2017

To Whom It May Concern:

Regarding PhD thesis of Joanna Śliwiak.

This is letter to certify my contribution to the following publication of which I was a co-author.

Sliwiak, J., Dauter, Z., Kowiel, M., McCoy, A. J., Read, R. J. & Jaskolski, M. (2015). ANS complex of St. John's wort PR-10 protein with 28 copies in the asymmetric unit: a fiendish combination of pseudosymmetry with tetartohedral twinning. *Acta Crystallogr. D* 71, 829-843.

I contributed to the data analysis, particularly regarding symmetry and twinning, and to writing parts of the paper concerning that analysis.

A handwritten signature in black ink that reads "Randy J Read". The signature is written in a cursive style with a large, prominent 'R' and 'J'.

Randy J Read

Prof. Mariusz Jaskólski
Department of Crystallography - Center for Biocrystallographic Research
Institute of Bioorganic Chemistry, Polish Academy of Sciences
Poznań, Poland

February 26, 2017

To Whom It May Concern:

Regarding the PhD thesis of Joanna Śliwiak.

This is to certify my contribution to the following publication, of which I am a co-author:

Sliwiak, J., Dauter, Z., Kowiel, M., McCoy, A. J., Read, R. J. & Jaskolski, M. (2015). ANS complex of St. John's wort PR-10 protein with 28 copies in the asymmetric unit: a fiendish combination of pseudosymmetry with tetartohedral twinning. *Acta Crystallogr. D* 71, 829-843.

My contribution consisted in:

- coordination of the project
- participation in the analysis of the results
- coordination of manuscript preparation



Prof. Mariusz Jaskólski, PhD

Zbigniew Dauter
Argonne National Laboratory
Argonne, IL 60439
USA

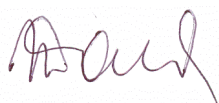
2.03.2017

Oświadczenie o współautorstwie:

Oświadczam, iż mój udział w publikacji:
Sliwiak, J., Dauter, Z. & Jaskolski, M. (2013). Hyp-1 protein from St John's wort as a PR-10 protein. *BioTechnologia* 94, 47-50.

polegał na udostępnieniu Pani Śliwiak możliwości zebrania danych dyfrakcyjnych z kryształu Hyp-1 na synchrotronowej linii eksperymentalnej w Argonne.

Z poważaniem



Zbigniew Dauter, Ph.D.

Chief, Synchrotron Radiation Research Section, MCL, NCI
Section Editor, Acta Crystallographica F
Argonne National Laboratory, Bldg. 202, Rm. Q142
9700 S. Cass Ave., Argonne, IL 60439
Tel: (630)-252-3960
Fax: (630)-252-3622
Email: dauter@anl.gov

Prof. Mariusz Jaskólski
Zakład Krystalografii - CBB
Instytut Chemii Bioorganicznej PAN
w Poznaniu

26 lutego 2017r.

Oświadczenie o współautorstwie:

Oświadczam, że mój udział w publikacji:
Sliwiak, J., Dauter, Z. & Jaskolski, M. (2013). Hyp-1 protein from St John's wort as a PR-10
protein. *BioTechnologia* 94, 47-50.

polegał na współredagowaniu publikacji.



prof. dr hab. Mariusz Jaskólski

Zbigniew Dauter
Argonne National Laboratory
Argonne, IL 60439
USA

2.03.2017

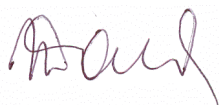
Oświadczenie o współautorstwie:

Oświadczam, iż mój udział w publikacji:

Sliwiak, J., Dauter, Z. & Jaskolski, M. (2016). Crystal structure of Hyp-1, a *Hypericum perforatum* PR-10 protein, in complex with melatonin. *Front. Plant Sci.* 7, a.668.

polegał na udostępnieniu Pani Śliwiak możliwości zebrania danych dyfrakcyjnych z kryształu kompleksu proteiny Hyp-1 z melatoniną na synchrotronowej linii eksperymentalnej w Argonne.

Z poważaniem



Zbigniew Dauter, Ph.D.

Chief, Synchrotron Radiation Research Section, MCL, NCI
Section Editor, Acta Crystallographica F
Argonne National Laboratory, Bldg. 202, Rm. Q142
9700 S. Cass Ave., Argonne, IL 60439
Tel: (630)-252-3960
Fax: (630)-252-3622
Email: dauter@anl.gov

Prof. Mariusz Jaskólski
Zakład Krystalografii - CBB
Instytut Chemii Bioorganicznej PAN
w Poznaniu

26 lutego 2017r.

Oświadczenie o współautorstwie:

Oświadczam, że mój udział w publikacji:

Sliwiak, J., Dauter, Z. & Jaskolski, M. (2016). Crystal structure of Hyp-1, a *Hypericum perforatum* PR-10 protein, in complex with melatonin. *Front. Plant Sci.* 7, a.668.

polegał na:

- koordynacji projektu
- nadzorowaniu analizy wyników
- koordynacji przygotowania manuskryptu



prof. dr hab. Mariusz Jaskólski

Miłosz Ruszkowski, National Cancer Institute, Valparaiso, IN, USA, 26 lutego 2017

Oświadczenie o współautorstwie:

Oświadczam, iż mój udział w publikacji:

Ruszkowski, M., Sliwiak, J., Ciesielska, A., Barciszewski, J., Sikorski, M. & Jaskolski, M. (2014).

Specific binding of gibberellic acid by Cytokinin-Specific Binding Proteins: a new aspect of plant hormone-binding proteins with the PR-10 fold. *Acta Crystallogr. D70*, 2032-2041.

polegał na rozwiązaniu i udokładnieniu struktury krystalicznej białka MtPhBP oraz VrPhBP oraz przygotowaniu większości manuskryptu.

Z poważaniem

Miłosz Ruszkowski

Joanna Śliwiak
Instytut Chemii Bioorganicznej PAN
Noskowskiego 12/14
61-704 Poznań

Poznań, 28 lutego 2017

Oświadczenie o współautorstwie:

Oświadczam, iż mój udział w publikacji:

Ruszkowski, M., Sliwiak, J., Ciesielska, A., Barciszewski, J., Sikorski, M. & Jaskolski, M. (2014).
Specific binding of gibberellic acid by Cytokinin-Specific Binding Proteins: a new aspect of plant
hormone-binding proteins with the PR-10 fold. *Acta Crystallogr. D70*, 2032-2041.

polegał na:

- przeprowadzeniu pomiarów kalorymetrycznych oddziaływań białek MtPhBP oraz VrPhBP z *trans*-zeatyną i gibereliną i wyznaczeniu parametrów termodynamicznych tych oddziaływań
- analizie otrzymanych wyników i przygotowaniu części manuskryptu poświęconej w/w pomiarom kalorymetrycznym

Z poważaniem



Jakub Barciszewski
Instytut Chemii Bioorganicznej PAN
Noskowskiego 12/14

Poznań, 28 lutego 2017

Oświadczenie o współautorstwie:

Oświadczam, iż mój udział w publikacji:

Ruszkowski, M., Sliwiak, J., Ciesielska, A., Barciszewski, J., Sikorski, M. & Jaskolski, M. (2014).
Specific binding of gibberellic acid by Cytokinin-Specific Binding Proteins: a new aspect of plant
hormone-binding proteins with the PR-10 fold. *Acta Crystallogr. D70*, 2032-2041.

polegał na otrzymaniu kryształów białka MtPhBP w kompleksie z GA3.

Z poważaniem



dr hab. Michał M. Sikorski prof. IChB PAN
Pracownia Inżynierii Białek
Zakład Krystalografii – CBB
Instytut Chemii Bioorganicznej PAN
Noskowskiego 12/14
61-704 Poznań

Poznań, 28 lutego 2017

Oświadczenie o współautorstwie:

Oświadczam, iż mój udział w publikacji:

Ruszkowski, M., Sliwiak, J., Ciesielska, A., Barciszewski, J., Sikorski, M. & Jaskolski, M. (2014).

Specific binding of gibberellic acid by Cytokinin-Specific Binding Proteins: a new aspect of plant hormone-binding proteins with the PR-10 fold. *Acta Crystallogr. D70*, 2032-2041.

polegał na opracowaniu procedury oczyszczania preparatu białka i udziale w przygotowaniu manuskryptu.

Z poważaniem



Michał M. Sikorski

Prof. Mariusz Jaskólski
Zakład Krystalografii - CBB
Instytut Chemii Bioorganicznej PAN
w Poznaniu

26 lutego 2017r.

Oświadczenie o współautorstwie:

Oświadczam, że mój udział w publikacji:

Ruszkowski, M., Sliwiak, J., Ciesielska, A., Barciszewski, J., Sikorski, M. & Jaskolski, M. (2014).
Specific binding of gibberellic acid by Cytokinin-Specific Binding Proteins: a new aspect of plant
hormone-binding proteins with the PR-10 fold. *Acta Crystallogr. D70*, 2032-2041.

polegał na:

- wkładzie do analizy wyników krystalograficznych
- współredagowaniu publikacji.



prof. dr hab. Mariusz Jaskólski