Miłosz Ruszkowski

Structural and biochemical studies on proteins involved in hormonal regulation in plants

TABLE OF CONTENTS

Abbreviations

2iP	N^6-isopentenyladenine
ADP	atomic displacement darameter
BAP	N^6-benzyladenine
CHASE	cyclases/histidine kinases-associated sensory extracellular domain
DLS	dynamic light scattering
FPLC	fast protein liquid chromatography
HK	histidine kinase
HPt	histidine-containing phosphotransfer protein
IMAC	immobilized metal affinity chromatography
IPTG	isopropyl-D-thiogalactopyranoside
KIN	kinetin
LB	Luria Bertani
LIC	ligation-independent cloning
MBP	maltose binding protein
MPD	2-methyl-2,4-pentanediol
MST	micro-scale thermophoresis
MtCRE1	*Medicago truncatula* cytokinin receptor 1
MtCRE1’	intracellular part of MtCRE1
NF	Nod factor
PEG	polyethylene glycol (MME, PEG monomethyl ether)
PCR	polymerase chain reaction
PDB	Protein Data Bank
PIPE	polymerase incomplete primer extension
REC	receiver domain
RMSD	root-mean-square deviation
RR	response regulator
SAXS	small-angle X-ray scattering
SEC	size-exclusion chromatography
TEV	Tobacco Etch Virus
TCEP	tris(2-carboxyethyl)phosphine
TLS	translation/libration/screw
ZEA	*trans*-zeatin

Preface

This dissertation describes several novel findings concerning structural biology of plants. More precisely, it is focused on proteins that can bind cytokinins or are able to transduce signal initiated by these phytohormones. Studies on this proteins were performed mainly with the use of biomolecular crystallography but are also supported by various *in vitro* assays.

The thesis itself is divided into four parts. The first, *Introduction*, provides biological background on the hormonal regulation in plants. It also introduces the reader with nodulation of legumes and pathogenesis-related proteins of class 10 that are known to interact with some plant hormones. The second part, *Methods*, presents the techniques used within this thesis. These include protein production, purification, crystallization as well as several additional methods developed and used to provide both structural and functional information. The very basic concepts of X-ray crystallography and, in particular, biomolecular crystallography are also treated in this part as briefly as possible. Part *Experimental procedures* describes all the experiments relevant to this thesis in details, in particular the routes to solving the crystal structures of the proteins of interest. The last, nonetheless everything but least part, *Results and Discussion*, is largest and focuses on a very comprehensive structural analysis of the ultimate results, the solved protein structures. The supporting experiments are described only if they are relevant to the structures. It is also worth to emphasize here that in order to keep this thesis concise and not to mislead the reader, only the solved structures are regarded as results.

List of publications related to this thesis

Ruszkowski M, Szpotkowski K, Sikorski M & Jaskolski M (2013) The landscape of cytokinin binding by a plant nodulin. *Acta Cryst.* D**69**,

Ruszkowski M, Brzezinski K, Jedrzejczak R, Dauter M, Dauter Z, Sikorski M & Jaskolski M (2013) Medicago truncatula histidine-containing phosphotransfer protein: Structural and biochemical insights into the cytokinin transduction pathway in plants. *FEBS J* **280**, 3709-3720.

Ruszkowski M, Sikorski M, Jaskolski M (2013) Crystal Structure of a PR-10 nodulin in complex with *trans*-zeatin. *BioTechnologia* **94**, 42-46.

Ciesielska A, Ruszkowski M, Kasperska A, Femiak I, Michalski Z, Sikorski M (2012) New insights into the signaling and function of cytokinins in higher plants. *BioTechnologia* **93**, 400-413.

Introduction

Plant hormones

Throughout its lifecycle, each living organism undergoes substantial changes, that either contribute to growth, development or are adaptation to environmental stimuli. For this reason, plants and animals developed hormones, chemical messengers that coordinate cellular functions. However, hormonal regulation in both kingdoms is completely different. Animals synthesize a large number of hormones that all have specific functions and trigger a particular response. These hormones are secreted in one, explicit organ (gland), transferred through the blood and influence a physiological process in a distant tissue. In addition, hormonal regulation in animals cooperates with their nervous system. Plants lack a nervous system. Although, recent studies have shown that plants can remember or respond to information using their intelligence (Karpinski & Szechynska-Hebda, 2010), however, hormones are considered the factors that regulate every aspect of plant life. Plant hormones (phytohormones) can be synthesized in most plant cells and usually affect the cell of origin itself, as well as, more remote cells. Plant hormones are also fewer in number, when compared to animals. The set of plant hormones have expanded over the years and now includes (but is not restricted to) ten main classes: auxins, cytokinins, gibberellins, abscisic acid, brassinosteroids, ethylene, jasmonates, polypeptide hormones, salicylic acid and strigolactones (Santner *et al.*, 2009). In contrast to animals, in plants, it is usually not a single hormone, but the cross-talk between hormones from different classes, that controls a particular physiological process.

Auxins are usually represented by indole-3-acetic acid (IAA, Fig. 1). They are involved in plant growth and development at almost every step. In general, auxins induce cell growth and allow a plant to move in response to a stimuli, such as light or gravity (Went, 1938, 1942). It is expected, that auxins control cell-wall-related genes that influence growth (Peng *et al.*, 2009), but the details have not been elucidated yet. IAA can be synthesized from tryptophan *via* tryptamine (Quittenden *et al.*, 2009) or indole-3-pyruvic acid (Strader & Bartel, 2008) or from indole, in a pathway that do not involve tryptophan (Woodward & Bartel, 2005). Local maxima and minima of auxins are of a key relevance in context of the ultimate effect, thus these hormones are transported in the plant organism via an elaborated cell-to-cell transport system (Vieten *et al.*, 2007).

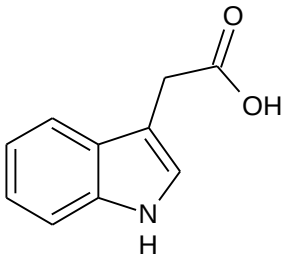


Figure 1
Indole-3-acetic acid.

In *Arabidopsis thaliana* auxins are perceived by a transport inhibitor response 1 protein (TIR1) that is an F-box protein (Dharmasiri *et al.*, 2005). The crystal structures of TIR1 protein are known (PDB ID: 2p1m, 2p1n, 2p1o, 2p1p, 2p1q) and explain the structural features of auxin recognition (Tan *et al.*, 2007). Auxins can trigger either nongenomic or genomic response and the latter has been much better described (Badescu & Napier, 2006). Auxin-regulated transcription can occur through Auxin Response Factors (ARFs) that directly bind DNA and regulate transcription of ARF-depending genes. ARFs can be repressed by auxin/IAA proteins of short life-expectancy, which degradation involves ubiquitin-proteasome pathway (Mockaitis & Estelle, 2008). The knowledge about auxin response network beyond the proteins mentioned above is very limited. It is known, however, that these phytohormones regulate transcription of very broad range of genes with various functions (Peng *et al.,* 2009). Some of these genes stimulate the expression of auxin/IAA proteins, meaning that the auxin response is regulated *via* negative feedback loop (Mockaitis & Estelle, 2008).

Cytokinins, which are the main subject of this thesis are extensively described in a separate chapter below.

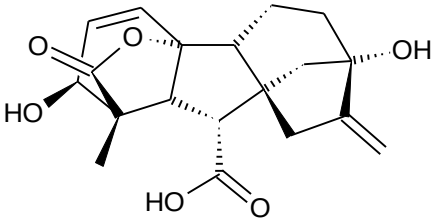


Figure 2
Gibberellic acid.

Gibberellins, such as gibberellic acid (Fig. 2**Błąd! Nie można odnaleźć źródła odwołania.**) are diterpenoid, tetracyclic growth regulators. They induce *inter alia* seed development and germination, organ elongation and flowering (Yamaguchi, 2008). Gibberellins were first discovered in *Gibberella fujikuroi* (thus the name), a fungal pathogen of rice that causes extreme elongation of the stem, finally leading to plant death (Yamaguchi, 2008). In plant cells, endogenous gibberellins are produced from geranylgeranyl diphosphate in a complex pathway that can be adjusted at several steps. The level of gibberellins in the plant cell is regulated not only *via* negative feedback loop, but also by concentration of auxins and ethylene (Fleet & Sun, 2005; Yamaguchi, 2008). The receptor of gibberellins, named Gibberellin Insensitive Dwarf1 protein (GID1), have been identified in rice (Ueguchi-Tanaka *et al.*, 2005). *A. thaliana* GID1 complexes with various gibberellins were investigated from structural point of view by Murase *et al.* (2008), (PDB ID: 2zsh, 2zsi). The name "dwarf" indicates, that loss-of-function mutations in *gid1* gene cause dwarfism (Peng *et al.*, 1999), a feature desired in

agriculture. GID1 proteins can bind DELLA (proteins with conserved Asp-Glu-Leu-Leu-Ala sequence at N-terminus) that are negative regulators of gibberellin response (Schwechheimer, 2008; Schwechheimer & Willige, 2009). GID1-DELLA interaction depends on the level of gibberellins. More precisely, binding of gibberellin by GID1 enhances GID1 affinity to DELLAs, causes GID1-DELLA complex formation. As a result, DELLA no longer functions as a transcription repressor of gibberellin dependent proteins, is ubiquitinated and degraded.

Abscisic acid (ABA) has an isoprenoid skeleton (Fig. 3) and controls drought responses, seed dormancy and other responses to abiotic stresses (Nambara & Marion-Poll, 2005). More accurately, abscisic acid induces leaves abscission in plants during preparation for winter (thus the name), causes the closure of stomata in the shortage of water and inhibits seed germination, meaning that it is an antagonist of gibberellins in the latter process. Pathways leading to ABA biosynthesis have been nearly completely identified (Nambara & Marion-Poll, 2005; Israelsson *et al.*, 2006). However, the molecular basis of ABA recognition and the response pathway have been clarified only recently (Soon *et al.*, 2012; Zhou *et al.*, 2012). Shortly, ABA controls the activity of type 2C protein phosphatases (PP2Cs) and Snf1-related kinases, SnRK (Fujii *et al.*, 2009; Fujita *et al.*, 2009; Umezawa *et al.*, 2009). Unless ABA is present, PP2C inactivate SnRKs by dephosphorylating the active Ser residue which phosphorylation is necessary for kinase activity. However, if ABA is bound to PYR/PYL/RCAR family of ABA receptors, the latter form complex with PP2C and inhibit their dephosphorylating activity. In the consequence, kinase activity of SnRKs is restored and the signal can be passed downstream and trigger the ultimate response.

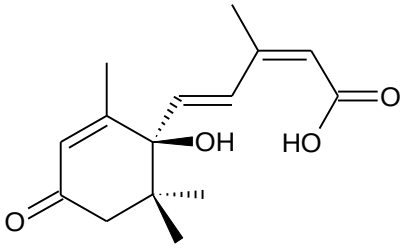


Figure 3
Abscisic acid.

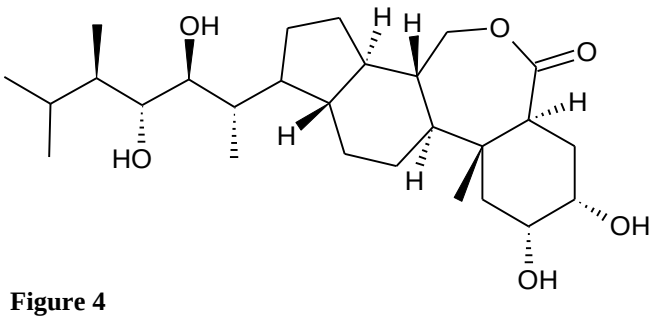


Figure 4
Brassinolide.

Brassinosteroids are represented by brassinolide (Fig. 4). Hormones from this class induce cell expansion and elongation (Clouse & Sasse, 1998) in cooperation with auxins (Nemhauser *et al.*, 2004).

Brassinosteroids promote vascular differentiation (Cano-Delgado *et al.*, 2004). They also endorse pollen development, accelerate senescence (Clouse & Sasse, 1998). Biosynthesis pathway of brassinosteroids starts from campesterol (Fujioka & Sakurai, 1997; Sakurai & Fujioka, 1997). Brassinosteroids are recognized by membrane-bound receptor-like kinases with Leu-rich repeat (She *et al.*, 2011). Binding of the hormone initiates a phosphorylation cascade finally influencing expression of brassinosteroid-dependent genes.

Ethylene is the only gaseous plant hormone. Its role in fruit ripening is generally known and used by retailers that ship raw fruits overseas. Ethylene influences more aspects of a plant life (Kendrick & Chang, 2008), but they are outside the scope of this thesis. This gaseous hormone is biosynthesized from methionine in the Yang cycle (Adams & Yang, 1979). Ethylene biosynthesis pathway can occur in most of the plant cells but is dependent on environmental stimuli, fruit ripening itself (a ripe fruit secretes more ethylene) and auxin (De Paepe & Van der Straeten, 2005). The ethylene response pathway is similar to those of cytokinins. The first element is the HK which is anchored in the plasma membrane of the endoplasmatic reticulum (Kendrick & Chang, 2008). This HK is activated upon ethylene binding and starts the phosphorylation cascade that leads to the ultimate response.

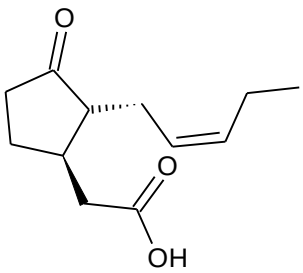


Figure 5
Jasmonic acid.

Jasmonic acid (JA, Fig. 5) and its derivatives, commonly named jasmonates, mediate response to biotic and abiotic stress. JA level increases rapidly (within 30 min) upon wounding such as damage caused by a herbivore (Wasternack, 2007). Accumulation of JA (or its more active form, isoleucine conjugated, JA-Ile) releases the expression of genes that have antiherbivory capabilities, as for instance insect digestive or antimicrobial proteins. Methyl jasmonate is a volatile derivative of JA that has been found not only to attract carnivorous arthropods, but also functions as alarming messenger for other branches of the same plant or plants in the neighborhood (Wasternack, 2007). The jasmonate signaling pathway resembles the one of auxin. Coronatine-Insensitive1 (COI1), an F-box protein, is the receptor for jasmonates (Xie *et al.*, 1998; Melotto *et al.*, 2008). Binding of jasmonates to COI1 results in ubiquitination of repressors called jasmonate ZIM-domain proteins (JAZ) (Chini *et al.*, 2007; Thines *et*

al., 2007). In a consequence, JAZs can no longer suppress the expression of response genes.

Polypeptide plant hormones is a large family that gathers various factors of several amino acid residues long that are recognized by a membrane-anchored receptor kinases (Matsubayashi & Sakagami, 2006). The peptide-receptor interaction is always specific and triggers a particular and precise response. For instance, systemin (Ryan & Pearce, 2003) is a long distance messenger that initiates chemical defense against herbivores. *ENOD40* gene encodes two peptides of 12 and 18 amino acid residues (Rohrig *et al.*, 2002) that have been shown to influence nodulation by interacting with sucrose synthase (Rohrig *et al.*, 2004). Peptide hormones are usually biosynthesized as longer inactive precursors that undergo proteolytic cleavage (McGurl *et al.*, 1992; Yang *et al.*, 1999).

Salicylic acid (SA, Fig. 6) is an important element in plant defense system (Vlot *et al.*, 2008). SA is induced by a pathogen infection and promotes the transcription of *PATHOGENESIS RELATED* genes (Van Huijsduijnen *et al.*, 1986). Systemic acquired resistance (SAR) is a very interesting aspect of SA signaling (Durrant & Dong, 2004). During SAR, SA is methylated to its volatile derivative, methyl salicylate, which, similarly to methyl jasmonate, passes the information about infection to other plant organs and plants in the neighborhood.

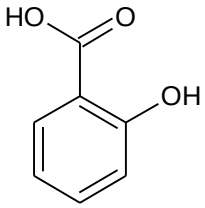


Figure 6
Salicylic acid.

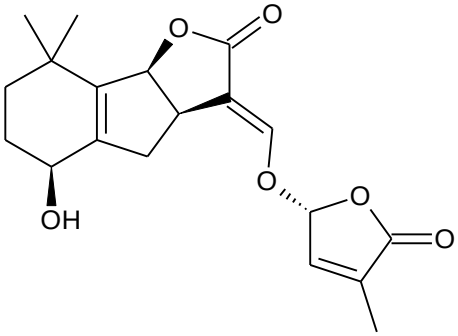


Figure 7
(+)-Strigol.

Strigolactones, carotenoid-derived phytohormones (e.g. (+)-strigol, Fig. 7) for years have been considered as signaling molecules in the rhizosphere. Recently it was found that they are also essential growth regulators (Ruyter-Spira *et al.*, 2013). Strigolactones are connected with pathways of other phytohormones (Cheng *et al.*, 2013). Mainly, auxins induce biosynthesis of strigolactones in the roots. Strigolactones are subsequently transferred to the shoot where they reduce bud outgrowth (Lin *et al.*, 2009).

Cytokinins

Cytokinins are a family of phytohormones that stimulates cell division (cytokinesis) and differentiation. They take part in regulation of e.g. apical dominance, axillary bud growth, leaf senescence, flowering and immunity. Cytokinins are also important factors during nodulation, as described in a separate chapter. Naturally occurring cytokinins are N6-substituted adenine derivatives (Fig. 8). The remaining part of this paragraph is focused on a brief description of cytokinins used in experiments for this thesis. In 1961 (Miller) *trans*-zeatin was isolated from a natural source, namely corn endosperm. Authors termed *trans*-zeatin a kinetin-like compound because kinetin, the first cytokinin, had been identified six years before *trans*-zeatin (Miller *et al.*, 1955). Kinetin, however, isolated from autoclaved herring sperm in the latter report was not believed to be a natural phytohormone for many years. Eventually, it has been found also in plant cell extracts (Barciszewski *et al.*, 1996). N6-isopentenyl adenine is an intermediate during cytokinin biosynthesis (Mok & Mok, 2001) but has been proven to be active *in vivo* (Huneeus *et al.*, 1980; Seyedi *et al.*, 2001). N6-benzyl adenine is one of the first synthetic cytokinins, nevertheless, it has been also found in plant extracts (Ernst *et al.*, 1983; Nandi *et al.*, 1989).

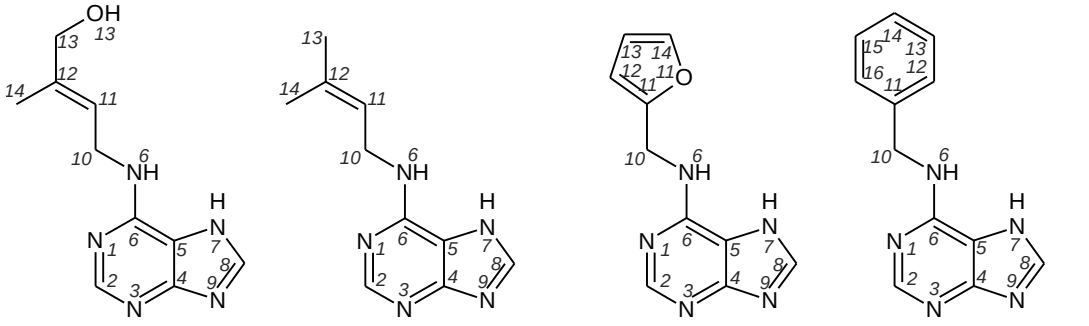


Figure 8
Chemical structures of four naturally occurring cytokinins. The atom numbering and abbreviations are shown for further reference.

Currently, two independent pathways of cytokinin biosynthesis are postulated: (i) oxidatative degradation of RNA and (ii) *de novo* synthesis (Schmulling, 2004). Cytokinins (in riboside forms) have been found in tRNA (Skoog *et al.*, 1966; Armstrong *et al.*, 1969; Barciszewska *et al.*, 1988), rRNA (Murai *et al.*, 1978) and mRNA (de la Serve *et al.*, 1984). The tRNA recycling results in *cis*-zeatin type cytokinins (Miyawaki *et al.*, 2006; Hirose *et al.*, 2008) that can be converted into biologically active form,

trans-zeatin, by *cis*-/*trans*-isomerases. The turnover rate of tRNA is, however, relatively slow and unable to provide a sufficient amount of cytokinins (Barciszewski *et al.*, 2000; Haberer & Kieber, 2002; Czerpak & Piotrowska, 2003). In *de novo* biosynthetic pathway adenosine phosphate-isopentenyltransferases (IPT) use ATP, ADP or AMP as substrates. Cytokinin ribosides and glycosides, that are inactive *in planta* (Kurakawa *et al.*, 2007; Kudo *et al.*, 2010), are used as a storage reservoir and can be easily converted to active forms. Cytokinin degradation is important in context their biological activity. For instance, cytokinin oxidases/dehydrogenases (CKXs) degrade and in this manner regulate the amount of cytokinins (Ashikari *et al.*, 2005). Expression profiles of *CKX*s genes show variability not only within different plant tissues, but are also influenced by changing environmental conditions or physiological state (Perilli *et al.*, 2010).

Phenylurea derivatives are synthetic compounds that *in planta* cause similar to cytokinins reactions. For instance, 1,3-diphenylurea (DPU) as well as thidiazuron are known to have cytokinin activity. However, none of the phenylurea-type cytokinins were found in natural plant tissues so far.

Cytokinin signal transduction

Phosphorelay systems are used by all living organisms as regulation mechanism. In prokaryotes, the two-component system consists of a hybrid sensor/histidine kinase and a response regulator protein with conserved His and Asp residues, respectively. The two latter amino acid residues are consecutive phosphorylation sites in the cascade of signal transduction. The phosphorelay itself can be triggered by various stimuli, such as osmotic stress (West & Stock, 2001). Eukaryotic phosphorylation pathways usually involve either tyrosine or serine/threonine kinases, however, the two-component-like signaling operates also in fungi and plants (Schaller G, 2002). Beside the obvious differences, resulting from evolutionary divergence, bacteria, fungi and plants have to face the same problem of adaptation to environmental conditions. This can explain why the two-component systems is absent only in animals which can either move towards beneficial or escape from adverse circumstances.

The eukaryotic cascade, however, is more complex and comprises of three proteins: (i) a hybrid sensor/histidine kinase that is anchored in the plasma membrane, (ii) a histidine-containing phosphotransfer protein (HPt), and (iii) a response regulator (RR) (Fig. 9). The sensory kinase is a three-domain protein, with the following arrangement

(N- to C-terminal): an extracellular cyclases/histidine kinases-associated sensory (CHASE) domain, and of two intracellular moieties, namely histidine kinase (HK) and receiver (REC) domains. In higher plants, either environmental stress or a phytohormone can trigger the cascade. Ethylene (Chang & Stadler, 2001) and cytokinins (Schmulling, 2001; Haberer & Kieber, 2002; Ferreira & Kieber, 2005) are phytohormones known to operate *via* this type of response pathways.

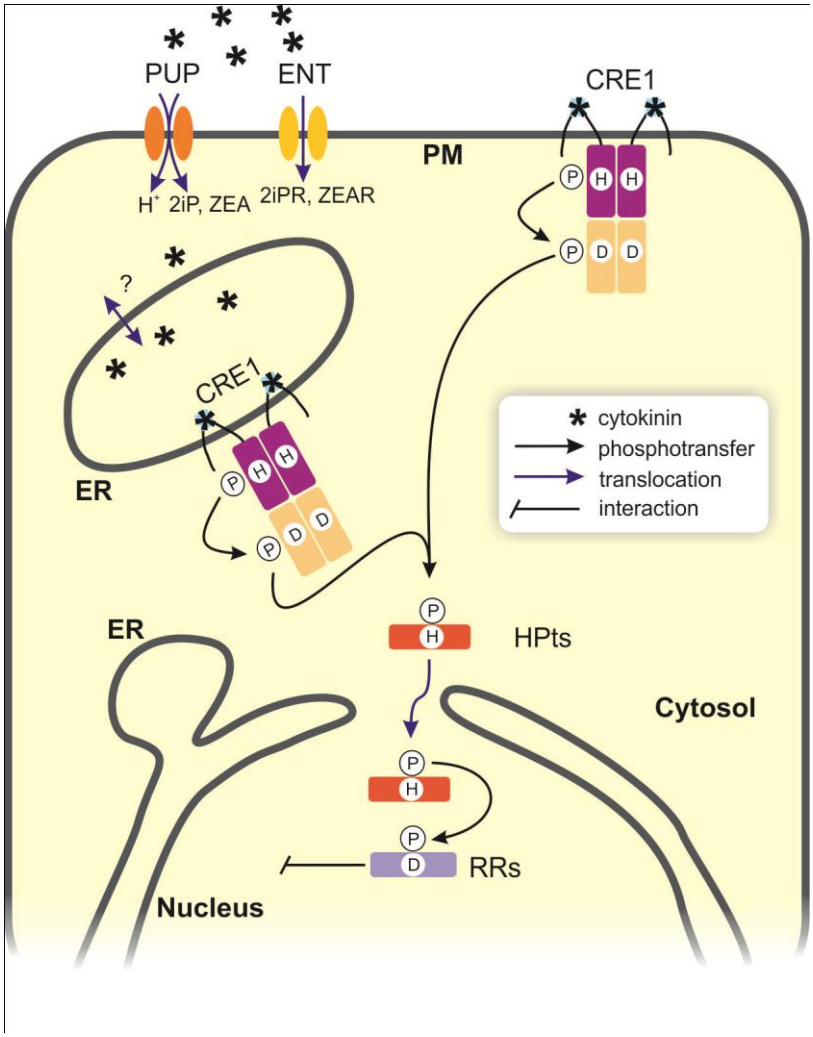


Figure 9
The cytokinin transduction pathway and active transport. Only one example of cytokinin receptors (CRE1) is shown for clarity. A detailed description is provided in text. Active histidine (H) and aspartate (D) residues transfer phosphate (P). Purine permeases (PUP) and equilibrative nucleoside transporters (ENT) transport cytokinins in either native forms (ZEA, 2iP) or as ribosides (ZEAR, 2iPR), respectively. ER, endoplasmatic reticulum; RM, plasma membrane. Adapted with modifications from (Ciesielska *et al.*, 2012; Hwang *et al.*, 2012; Nongpiur *et al.*, 2012)

The cytokinin transduction pathway is initiated by ligand binding to the CHASE domain. As a result, the sensory protein gains its kinase activity and a His residue in the HK domain is autophosphorylated. The His residue in turn phosphorylates an Asp

residue in the REC domain. Subsequently, an HPt protein is bound to the REC domain and phosphoryl group is transferred to the active His residue of the HPt protein. Then, the HPt protein phosphorylates an Asp residue of the RR protein. The latter protein initiates the final response by binding to a specific DNA element or to a target protein.

For years, cytokinin perception was considered an extracellular process, meaning that the phytohormones do not cross the cell membrane to activate the pathway. Recently, cytokinin receptors have been identified in the endoplasmatic reticulum (ER) (Caesar *et al.*, 2011; Lomin *et al.*, 2011; Wulfetange *et al.*, 2011). This discovery has shed new light on cytokinin signaling because the diffusion of cytokinins within the plant cell is still obscure and there may possibly be factors that mediate their translocation.

Legumes and nodulation

Nitrogen is one of the key macroelements. All living organisms require nitrogen to synthesize nucleotides, amino acids and many other vital compounds. Nitrogen is also very abundant since it is the main ingredient of the atmosphere. However, the atmospheric, molecular nitrogen is very difficult to assimilate. Only a very limited number of living organisms are capable of using N_2 directly. Namely, certain prokaryotes elaborated nitrogenases, enzymes which allow them to convert atmospheric nitrogen. However, nitrogenases are easily and irreversibly blocked by molecular oxygen, thus organisms that are dependent on O_2 do not have these enzymes. A family of plants, the legumes (*Fabaceae*), have developed a very extraordinary solution to this drawback. They encapsulate nitrogen-fixing bacteria from the *Rhizobia* genus in special root organs called nodules (Fig. 10) in a process termed nodulation. Nodulation is species-specific for both symbionts, the plant host and the bacteria, meaning that the partners have to precisely recognize each other. The bacteria assimilate atmospheric nitrogen, convert it chemically and supply the plant host with ammonia or glutamine. In exchange, the plant provides the bacteria with organic nutrients and an ideal place to live (low in oxygen). Thanks to nodulation, legumes do not require at all or need only very little nitrogen fertilization. This symbiotic interaction made legumes the perfect crops, so important to humans. Pea, bean, soybean, lens, peanut, and chickpea are only some examples of legumes that have been cultivated for millennia and whose nutritious values are appreciated also today. The legumes provide a roughly equal amount of fixed nitrogen to chemical fertilizer industry, reducing the ecological footprint and cost of cropping. A huge effort is made to implement a nodulation-like symbiotic interactions

with non-legume species. Companies, such as Azotic Technologies Ltd (http://www.azotictechnologies.com; Chorley, United Kingdom) patented a technology, which allows to infect a non-legume species with the nitrogen-fixing bacteria from sugarcane roots. These bacteria, however, do not live in nodules but grow inside the cytoplasm of the plant cells instead.

Figure 10
Nodules on *Medicago truncatula* root.
Source: commons.wikimedia.org

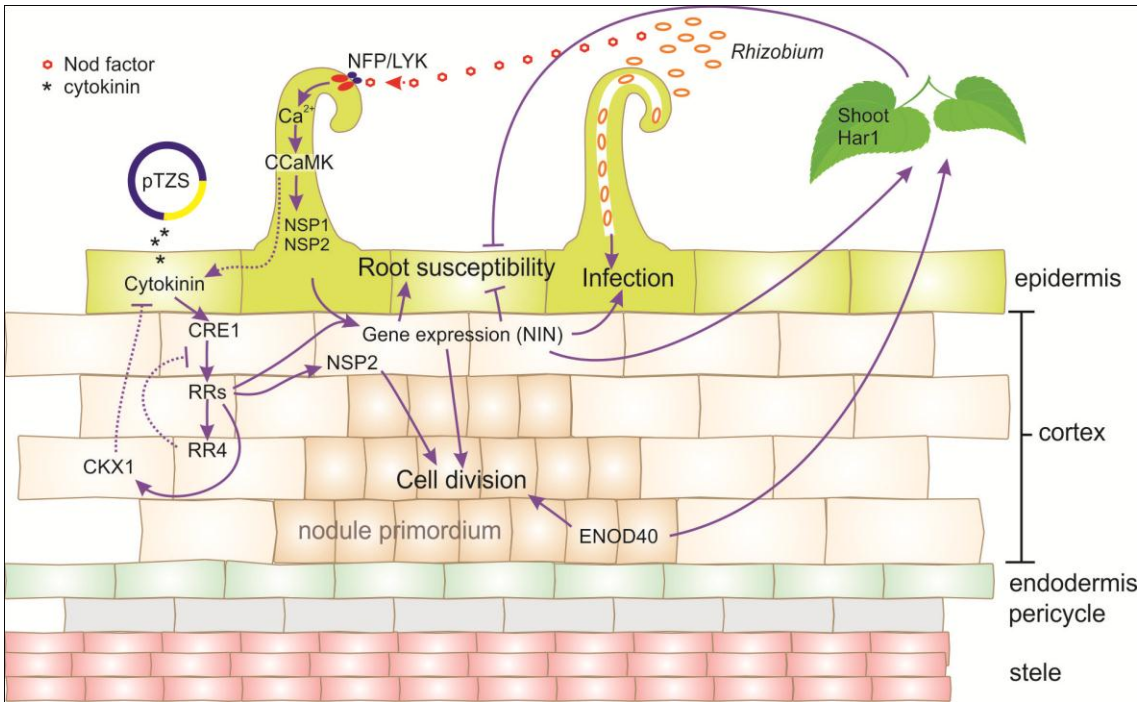


Figure 11
Nodulation pathways. Solid and dotted lines represent proven and putative interactions, respectively. Nod factor (NF) is perceived by NF perception/Lys-M kinase receptor (NFP/LYK). Calcium and calmodulin dependent kinase activates NF signaling pathways 1 and 2 (NSP1-2). Subsequently, the signal is transduced to nodule inception transcription factor (NIN). The NF signaling cascade can be bypassed via cytokinin transduction pathway initiated e.g. by a bacteria carrying pTZS plasmid. Cytokinin oxidase/dehydrogenase (CKX1) acts as a regulator of cytokinin concentration. Adapted with modifications from (Frugier *et al.*, 2008; Ariel *et al.*, 2012; Ciesielska *et al.*, 2012).

The nodulation is a complex process that involves multiple proteins clustered in either of two pathways, the typical one and the one leading *via* cytokinin signaling (Fig. 11). Unless a legume plant grows in a nitrogen-rich environment, it releases to the soil certain signaling flavonoids (Wasson *et al.*, 2006). These flavonoids act as chemoattractants for the soil-dwelling *Rhizobia* in the neighborhood of the plant roots. When the bacteria perceive the signal, they synthesize and secrete a variously modified lipochitin oligosaccharide, termed the Nod Factor (NF). NF is perceived in the root epidermis by the Nod Factor Perception/Lys-M kinase receptor (NFP/LYK), which triggers a signaling pathway that involves Ca^{2+} spiking and activation of calmodulin/calcium-dependent kinase, CCaMK (Gleason *et al.*, 2006). CCaMK activates Nod factor signaling pathways 1 and 2 (NSP1/2). The response involves also ethylene response transcription factor (ERN) and nodule inception transcription factor (NIN) (Roche *et al.*, 1991a; Roche *et al.*, 1991b). Multiple mutation-based studies have shown that this typical pathway can be bypassed after the CCaMK activation step (Tirichine *et al.*, 2006a; Tirichine *et al.*, 2006b; Murray *et al.*, 2007; Tirichine *et al.*,

2007; Frugier *et al.*, 2008; Ariel *et al.*, 2012). In this second pathway (that bypasses the typical pathway), the nodulation events involve (i) increased cytokinin production, (ii) activation of a cytokinin receptor, and (iii) a phosphorylation relay, known as the regular mechanism of response to cytokinins (Kakimoto, 1996; Hwang & Sheen, 2001). Both pathways eventually lead to increased rate of cell division in the nodule primordium and in this way cause nodule formation.

Cytokinins in nodulation

Results of several experiments that proved the involvement of cytokinins in nodulation have been reported. For instance, studies on two gain-of-function mutants of proteins involved in NF signaling or cytokinin transduction pathway, namely CCaMK (Gleason *et al.*, 2006; Tirichine *et al.*, 2006a) or *L. japonicus* cytokinin receptor, histidine kinase 1 from (LHK1) (Murray *et al.*, 2007; Tirichine *et al.*, 2007) demonstrated that the pathways involving either the CCaMK or LHK1 are sufficient for nodule organogenesis. In both cases, the nodule formation did not involve actual infection by *Rhizobia* and, of course, such barren organs were unable to fix nitrogen. Noteworthy, these mutations have additive effect, meaning that a double mutant generates more nodules than a single one. Other studies have also indicated that cytokinins are the key signaling molecules in the nodulation process. To recall two with opposite effects: (i) overexpression of genes responsible for catabolism of cytokinins diminished organogenesis of nodules in *Lotus japonicus* (Lohar *et al.*, 2004) and (ii) non-symbiotic bacteria that carried a plasmid coding for isopentyl transferase (*IPT*), a gene responsible for cytokinin biosynthesis, were able to initiate the morphogenesis of nodules (Cooper & Long, 1994). One experiment has been also performed on *Medicago truncatula*, the model legume and the source organism of all proteins from this thesis. It has revealed that when the cytokinin receptor (MtCRE1) is suppressed, the formation of nodules is defective (Gonzalez-Rizzo *et al.*, 2006).

PR-10 Proteins

PR-10 class gathers small (up to 19 kDa) monomeric, slightly acidic, cytosolic plant-specific proteins that are expressed in response to various biotic and abiotic stress factors. PR-10 proteins are divided into three subclasses, based on sequence homology. Sequence alignment of PR-10 proteins from each subclass is shown in figure 12. Proteins from the first subclass, classic PR-10, are expressed upon induction by

pathogens. Members of the second subclass, major latex proteins (MLP), are found in the latex, hence their name. Cytokinin-Specific Binding Proteins (CSBP), make the third subclass, notwithstanding they share a very low level of sequence identity (<20%) when compared to the other groups.

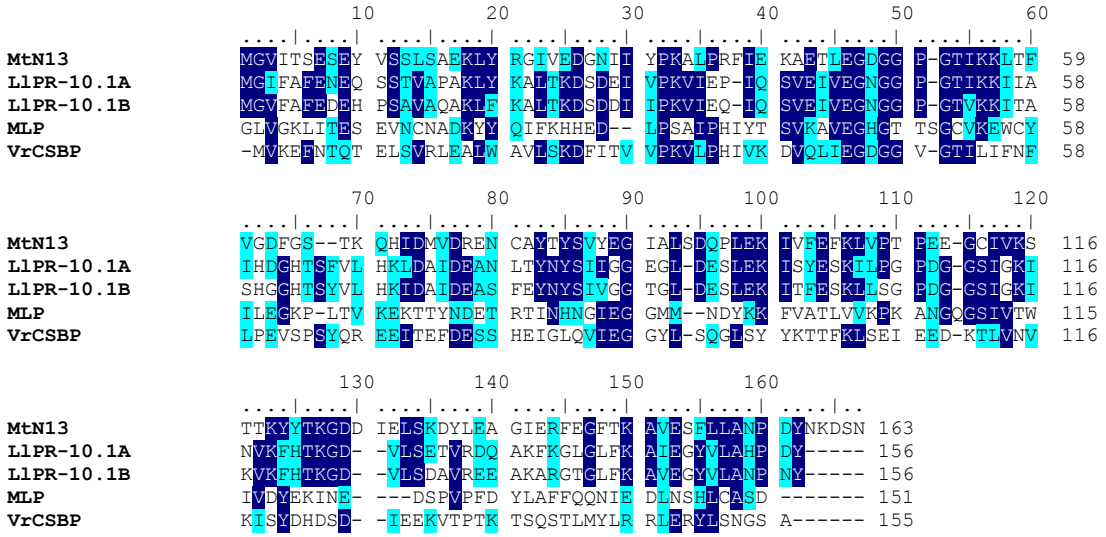


Figure 12
Sequence alignment of PR-10 proteins from three classes of PR-10. IPR (intracellular pathogenesis-related) with three examples given - MtN13, LlPR10.1A, LlPR10.2. Cytokinin Specific Binding Proteins (CSBPs) with one example from *Vigna radiata* shares only low identity to IPR. Major latex proteins (MLP) are represented here by MLP15 protein from *Papaver somniferum*. Dark and light blue represent residues highly and moderately conserved, respectively.

Despite only marginal sequence identity, shared between some PR-10 proteins, they have the same overall fold (Fig. 13). This fold consists of a seven-stranded antiparallel β-sheet that is wrapped around the C-terminal helix α3. The β-strands are connected by β-hairpins with the exception of the β1-β2 crossover, formed by helices α1 and α2, joining the edges of the β-sheet. The overall shape remains a baseball glove. The β-hairpins and loops, in particular, the odd-numbered loops L3, L5, L7, L9 resemble the "fingers" of the glove. The two short helices (α1 and α2) create a V-shaped support for the C-terminal part of the long α3 helix. In many examples of PR-10 proteins, this helix α3 is slightly kinked in its middle toward the protein core.

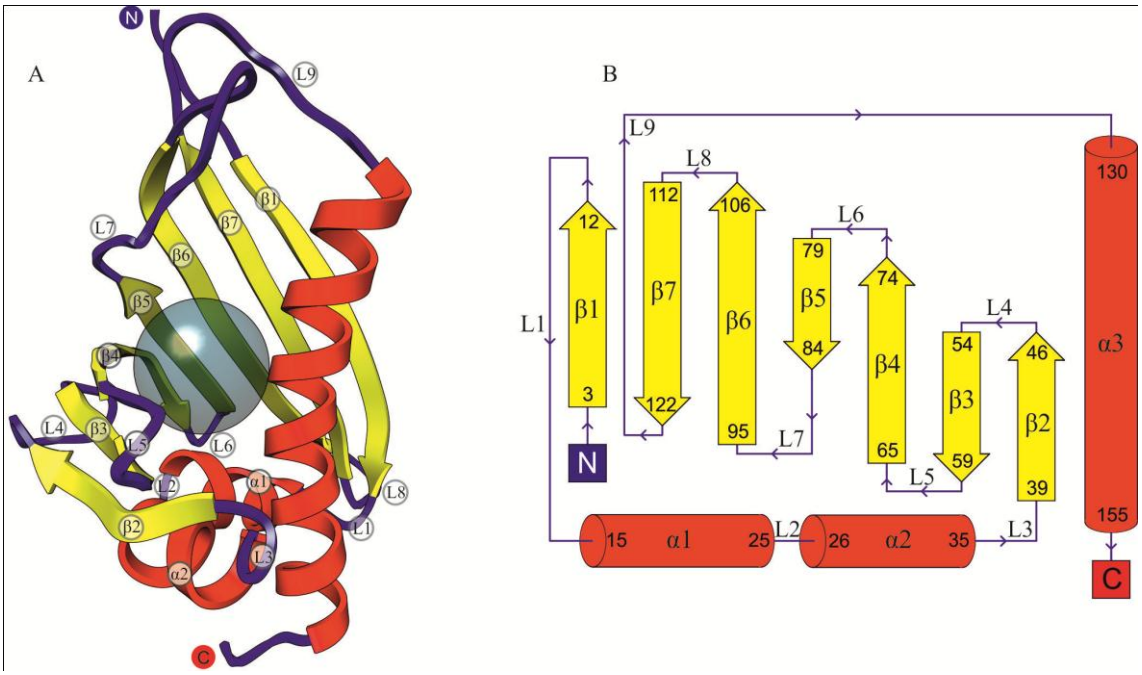


Figure 13
Overall structure of PR-10 proteins. **A** The fold consists of seven-stranded antiparallel β-sheet (yellow) and three α helices (red). Loops are colored in blue. A semitransparent sphere represents the location of the internal cavity. **B** Topology diagram with residues numbering corresponding to MtN13.

A large cavity is formed between the β-sheet and the longest, C-terminal α-helix α3. These PR-10 cavities are created primarily by hydrophobic residues in the core of the protein. However, a few polar side chains point into the lumen of the cavity, allowing to create selective hydrogen bonds with suitable heteroatom partners of the ligand molecules. The cavities of PR-10 proteins have been shown to be able to bind diverse ligands, such as phytohormones (Markovic-Housley *et al.*, 2003; Pasternak *et al.*, 2006; Fernandes *et al.*, 2008; Pasternak *et al.*, 2008; Fernandes *et al.*, 2009; Kofler *et al.*, 2012). This cavity is believed to be the key structural element that determines the function of each PR-10 protein (Fernandes *et al.*, 2013). Some examples of PR-10 proteins have been reported to bind cytokinin molecules (Pasternak *et al.*, 2006; Fernandes *et al.*, 2008; Fernandes *et al.*, 2009; Kofler *et al.*, 2012). However, the conclusion from these studies is rather perplexing, since there is no unique mode of ligand binding (or constant stoichiometry) and even the same protein can bind the same cytokinin in several different but well defined manners (Pasternak *et al.*, 2006).

Methods

Recombinant protein production

Only a small fraction of proteins for crystallography is isolated directly from their natural sources. In most cases overexpression in *E. coli* cells is a method of choice. This applies also for eukaryotic proteins despite the different biosynthetic machineries of prokaryotes and eukaryotes. More importantly, in many examples it was shown that recombinant proteins maintain biochemical properties of their native equivalents [REF]. Recent advances in DNA manipulation techniques, PCR, cutting-edge polymerases, ligases and restriction enzymes have led to development of protocols that allow to clone target genes into expression vectors in a variety of cellular hosts. This has proven most useful not only when the protein of interest is difficult to obtain but also when in the natural source the expression level is very low or is dependent on some other factors.

Before any actual experiment, it is best to start with *in silico* data mining. Bioinformatic tools pay a great contribution and provide enormous amount of cross-linked information in on-line meta-servers. Basic Local Alignment Search Tool or BLAST (Altschul *et al.*, 1990) as well as its subprograms are most helpful during the search of a particular gene or protein. BLAST also indicates matching motifs and recognizes domains, it is, therefore useful when a parallel approach is planned with a separate expression of such truncated domains. Having the sequence coding for a particular protein it is worth to perform some extra *in silico* analysis that could save both time and money in the laboratory. For example, XtalPred (Slabinski *et al.*, 2007) predicts secondary structure elements, disordered regions, isoelectric point and potential signal peptides or transmembrane domains. The latter would require a different approach and unless appropriate adjustments are introduced, attempting to produce, purify and crystallize a protein that has such fragments is *a priori* an experiment destined to fail.

After *in silico* steps, one must obtain the appropriate generic material of the source organism. It must be remembered, that eukaryotic DNA contains introns, fragments of genes that are removed by RNA splicing machinery. Hence, if an eukaryotic gene is to be cloned, it is best to start with isolation of RNA followed by reverse transcription. In this procedure cDNA (complementary DNA) is obtained that covers coding fragments of DNA (exons). cDNA is, therefore, a perfect template for amplification of a particular gene using a PCR. The PCR has to be preceded by the choice of a vector and appropriate primer design so the resulting insert has ends compatible with the vector.

In this work, LIC (Dieckman *et al.*, 2002; Kim *et al.*, 2011) method was applied for obtaining most of the constructs. LIC allows to clone the target gene into a vector of choice without restriction endonucleases or DNA ligase. In LIC method T4 DNA polymerase is used instead. This T4 polymerase has a 3'-5' exonuclease activity and cleaves 3' ends of both, vector and primer DNA. The 3'-5' exonuclease activity of T4 DNA polymerase stops when the enzyme encounters the deoxynucleotide that is present in the reaction mixture. The PCR primers for DNA amplification are designed in a way so both, the insert and the vector after treatment with T4 polymerase have sticky 5' ends of fifteen bases length. A resulting plasmid can be used directly for transformation of chemicompetent host *E. coli* cells. The transformed cells are spread over a medium supplemented with selective antibiotics that allow to grow only colonies that have incorporated a proper resistance gene (that is also a part of the transformation vector). The colonies are subsequently tested for the protein expression and plasmids isolated from these cells are sequenced to ensure there was no coincidental mutation during the entire procedure. Having the correct clones, a large scale overexpression can be performed using optimal conditions for each particular protein.

Sometimes it is necessary to incorporate changes in the amino acid sequence of a protein. Generation of inactive mutants to evaluate the active residues, surface entropy reduction (*vide infra*) are only two examples, when site-directed mutagenesis is needed. Thanks to novel, very efficient and proof-reading DNA polymerases these mutations can be introduced using relatively simple procedures, such as PIPE cloning (Klock & Lesley, 2009). In this method the mutation is included in primer sequences that overlap each other. Plasmid carrying the original sequence can be used as template for PCR reaction and product used directly for transformation of the competent *E. coli* cells.

Protein purification

Crystallization requires a protein sample of a very high quality, therefore protein purification is one of the most important steps when preparing a sample. If the protein was overexpressed in *E. coli* cells, and is expected to be present in the cytosol, the cell walls must be disrupted. To this end, enzymatic lysis with lysozyme, French press or sonication is usually a method of choice. If the protein of interest is soluble, the supernatant after lysis can be used directly for purification. If this is not the case and it is found in inclusion bodies, the protein may be dissolved e.g. in urea or guanidine hydrochloride and refolded *in vitro*. This approach, however is only sometimes helpful and solely for one-domain proteins. Another approach is to return to cloning and express the protein with a fusion tag that improves the solubility.

MBP or NusA are the ones that have the highest success rate (Kim *et al.*, 2011). NusA itself is a ~60 kDa protein that takes part in termination of transcription in *E. coli.*

Having the protein of interest in solution, the next task is to choose optimal purification strategy. One may attempt to use affinity chromatography to get rid of most of the other impurities, such as other proteins, DNA fragments, lipids and carbohydrates. For this reason affinity tags are usually incorporated within the construct and expressed at either N- or C-terminus of the protein. His_6-tag is the most commonly used because it provides a great selection rate and is relatively cheap and effective in IMAC with Ni^{2+}. A protein tagged with His_6 can be bound to the resins that bound previously Ni^{2+}. The sample can be washed to obtain pure protein which is eluted with a buffer containing high concentration of imidazole. Other popular tags used for affinity chromatography are MBP, glutathione-S-transferase (GST) (Kaplan *et al.*, 1997), Strep tag (Schmidt & Skerra, 2007) and Halo tag (Los & Wood, 2007).

Ion exchange chromatography allows to separate proteins based on their net charge. The net charge of a protein is a resultant of both, the protein itself and a buffer composition, thus adjustment of buffer pH or salt concentration affects binding properties. Protein is usually bound to the column in a buffer with low salt concentration, and is washed or eluted as the salt concentration is raised slightly or radically, respectively.

Size-exclusion chromatography (SEC) is used for separation of proteins based on their size using columns with porous gels. The principle of this method is that the smaller molecule the larger is the volume of the matrix it has to migrate through. As a consequence, large proteins migrate faster through the column and are collected prior to the small ones. SEC can also provide information on the oligomeric state of the protein of interest. This purification method can be used either for small-scale sample analysis or for protein preparation in preparative scale.

Protein crystallization

Protein crystals are assemblies of periodically packed macromolecules that are detained together by a network of specific but weak intermolecular interactions and, therefore, are very fragile. On average, half of the volume of a protein crystal is filled with solvent that occupies voids between the protein molecules. The network of the voids is connected with the environment, meaning that small molecules can diffuse *via* these solvent channels. Protein crystal formation is driven by entropy, although one may find it surprising, because crystal

itself is a piece of well-ordered matter of low entropy at a first glance. However, the release of water molecules that in solubilized protein form a hydration shell contributes to a high entropy gain that overcompensates the entropy loss caused by the decrease of motional degrees of freedom for the protein molecules.

Protein crystallization is a phase separation of a supersaturated and thermodynamically metastable but homogenous protein solution (Fig. 14). Supersaturated solution can be achieved by adjusting several variables, such as protein concentration, pH, precipitant concentration (or a precipitating agent itself), temperature. This metastable solution equilibrates into a protein-rich phase and a saturated protein solution. The protein-rich phase can form crystals if the nucleation conditions and growth kinetics are favorable. In a less fortuitous case, the protein-rich phase can form either solid precipitates or "protein oils".

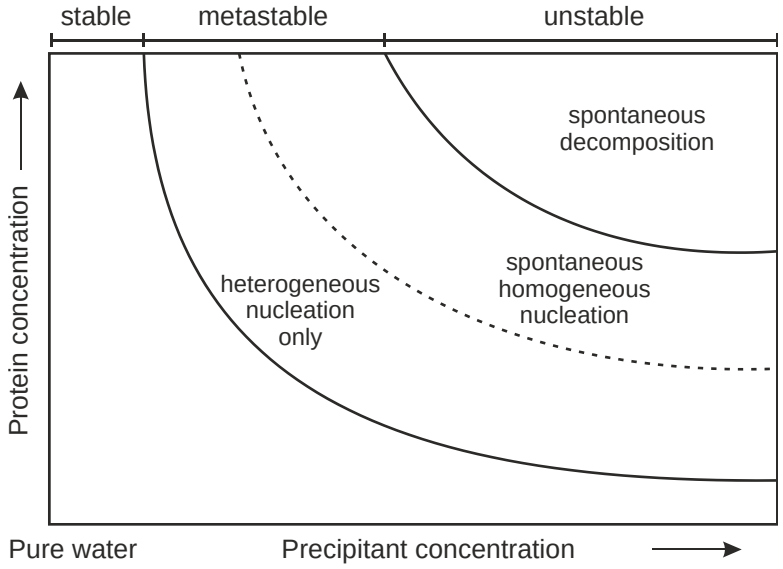


Figure 14
Crystallization diagram. Higher supersaturation is required for formation of stable crystallization nuclei. At low supersaturation crystallization can occur only upon initiation by external seeding. Adapted with modifications from (Rupp, 2010).

Despite decades of research, crystallization is still considered a serendipitous process. In most cases one must screen through tens or hundreds of conditions and still feel lucky if one or a few of them yield crystals. Commercially available sets of crystallization cocktails together with recent advances in robotic set up techniques come with a great help while searching for a perfect condition. In general, each crystallization cocktail contains reagents that can be gathered into three groups: (i) buffers that define the pH, (ii) precipitants and (iii) additive compounds, such as ions, detergents etc. Both organic and inorganic salts are common

precipitants, widely used in crystallization experiments. In addition, polyethylene glycol (PEG) and its monomethyl ether of varying chain lengths are often used.

A few crystallization techniques have been developed particularly for proteins. These include (but are not restricted to) batch crystallization, free-interface diffusion, micro-batch under oil, micro-dialysis or vapor-diffusion. The latter is nowadays most commonly applied due to simple and easy to automate set-up. This method relies on the presence of a reservoir that absorbs water from the crystallization drop and causes the supersaturated state within the drop. The vapor-diffusion technique can be used at two approaches, namely sitting- and hanging-drop. Sitting-drop is easiest for automation, it is therefore used while screening through a sparse matrix of crystallization conditions. Hanging-drop vapor diffusion, however, is usually applied during optimization of initial hits in a larger scale, because individual wells can be easily open and the crystals handles more precisely.

Once the crystals have grown, they need to be harvested and mounted for data collection. Since the data collection at cryogenic temperatures significantly reduces radiation damage, the crystals are usually vitrified in liquid nitrogen and stored in this condition for diffraction experiment. A proper cryoprotection that prevents formation of ice, that would strongly affect the diffraction images, is required. Some crystals, however, are too fragile to withheld the shock caused by either introducing cryoprotection or quenching in liquid nitrogen. Growing crystals in capillaries may be the method of choice in this case because the crystals do not need to be harvested and data collection can be performed at room temperature.

Protein modifications that improve crystallization

If crystallization trials of a protein have failed, it is worth to introduce modifications as the salvage strategies that can help obtaining crystals. One of the methods is reductive methylation of solvent-exposed lysine residues that are usually highly disordered and cause increased entropy at the protein surface. This modification at a protein level can be used to lower the conformational freedom of lysine residues (Rypniewski *et al.*, 1993; Rayment, 1997; Walter *et al.*, 2006) by changing solvent-exposed, primary amines (Nζ atoms of lysines and N-terminus) to tertiary dimethylamines.

Surface entropy reduction is a method of a protein modification at a construct level (Cooper *et al.*, 2007). Mutating long, hydrophilic amino acid residues, that can undertake various conformations to e.g. alanines may rescue the crystallization of the target protein. This approach is most useful whenever there are several of the amino acids with multiple degrees

of freedom in a row. With the use of SER server (Goldschmidt *et al.*, 2007) the protein sequence can be extensively analyzed for putative mutation sites. The biggest advantage of using the latter is that it also performs secondary structure matching with PSIPRED (Bryson *et al.*, 2005) and suggests only regions that are not expected to be involved in protein tertiary structure.

Another method that has proven to be useful while rescuing the proteins that would not form crystals otherwise is limited proteolysis. Using a small concentration of a selected protease one can get rid of the regions that do not possess a compact structure (flexible loops, linkers or termini) because only these fragments can bind to the active site of the protease. Prior to actual crystallization experiments with a partially digested protein of interest, one should perform screening with different proteases to establish optimal enzyme, its concentration and conditions. After wards, either the crystallization may be set up, or the resulting protein can be analyzed using e.g. mass spectrometry to identify most stable fragments of the sequence. If it is possible to distinguish a new, truncated but consecutive amino acid sequence, a good approach is to generate a new plasmid that would cause expression of the stable fragments solely.

Basics of X-ray diffraction on crystals

X-ray diffraction along with any other scattering experiments, including DLS and SAXS, is based on the fact that electromagnetic radiation is scattered by inducing polarization of the electrons that build matter. More precisely, electric field vectors of electromagnetic radiation interacts with electrons of matter. Only a small fraction of X-ray beam is scattered, most of it travels through the matter unchanged. Scattering from a single molecule is too weak to be measured. For this reason, in a diffraction experiment, crystals must be used. A crystal is an assembly of symmetrically related molecules that compose unit cells. Those unit cells are stacked in a three-dimensional space to form a crystal, and therefore are the easiest and complete description of the crystal structure itself. In a diffraction experiment, molecules organized periodically in a crystal lattice scatter X-rays that recombine into a resulting photons. The waves that are in phase (shifted by 2π) interfere with each other in a maximally constructive manner. On the other hand, phase shift of π, causes maximum destructive interference. The scattering by crystal is proportional to the number of electrons of the scattering atoms. It is also negatively influenced by positional displacements, such as disorder in the crystal lattice or thermal vibration around resting position of an atom. The wavelength

of emitted light remains unchanged if there is no electronic transition between energy levels (the scattering process is elastic).

Interpretation of X-ray diffraction can be simplified by looking at a crystal as a set of planes that reflect X-rays. The concept of Bragg planes was introduced by Sir William Lawrence Bragg, as per analogy to optical reflection, where the angle of incidence is equal to the angle of reflection. Bragg planes that are equidistant, equivalent and parallel are grouped into sets with the same Miller indices (*hkl*). With this model, Bragg equation can be introduced, that explains the condition when an reflected X-ray can be observed:

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

where n is an integer, λ the wavelength, d_{hkl} the spacing between the planes and θ the diffraction angle.

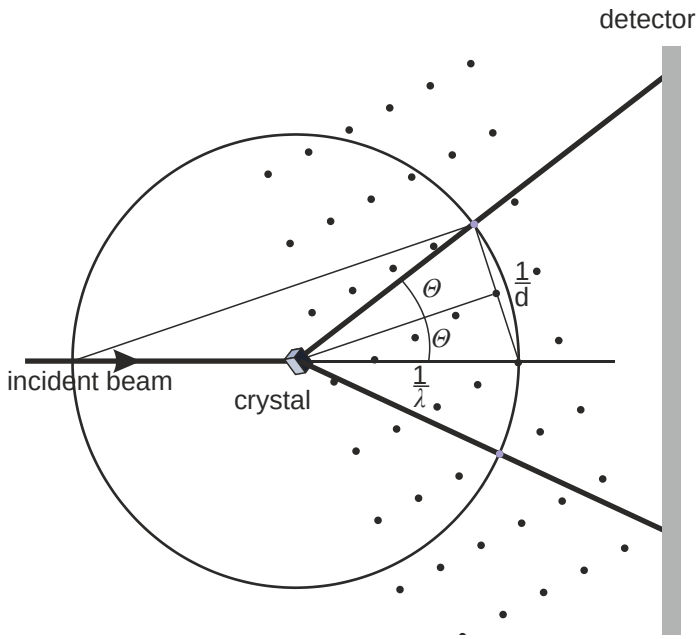


Figure 15
Ewald construction. The trigonometric condition provided by the Bragg equation is fulfilled when a reciprocal lattice point intersects (depicted as blue dot) with the surface of the Ewald sphere of a radius 1/ λ, where λ is the wavelength. The resulting reflection can be recorded on a detector. d Is the interplanar spacing. Adapted with modifications from (Dauter, 1999; Rupp, 2010).

Bragg equation, however, demonstrates the reflection condition in a two-dimensional space. The Ewald construction (Fig. 15) extends this concept to the third dimension in reciprocal lattice. The reciprocal lattice is a Fourier transform of the spacial wavefunctions of the real lattice. In the Ewald construction, the diffracting crystal is represented by its reciprocal lattice and wavelength λ by Ewald sphere of radius $1/\lambda$. The origin of the reciprocal lattice is at the

intersection of the incident beam and the Ewald sphere. In this construction any reciprocal lattice point that lies on the Ewald sphere fulfills the diffraction condition derived from the Bragg equation:

$$d^*_{hkl} = \frac{2\sin\theta}{n\lambda}$$

where $d^*_{hkl} = 1/d_{hkl}$ is the reciprocal lattice vector, normal to the set of lattice planes (*hkl*), that extends from the reciprocal lattice origin to a reciprocal lattice point (*h, k, l*). As a result, a reflection emerges in a direction from the center of the Ewald sphere towards the reciprocal lattice point. Only a small fraction of reciprocal lattice, and resulting diffraction spots, can be observed in a single and static orientation of a crystal. To bring the other reciprocal points to intersect with the Ewald sphere one must rotate the crystal. Other approach is to use non-monochromatic X-ray radiation, but in case of protein crystallography this method is hardly ever a method of choice.

In a diffraction experiment the intensity of reflected X-rays emanating from the crystal is measured. The intensity of a measured reflection *hkl* is proportional to squared structure factor amplitude, $|F_{hkl}|^2$, given by equation:

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

where *N* is the number of atoms within the unit cell and f_j is the atomic scattering factor of the *j*-th atom with coordinates (x_j, y_j, z_j) expressed as the fractions of the unit cell lengths a, b, c. The equation indicates evidently that each and every atom in the unit cell contributes to each and every reflection.

X-ray diffraction data collection

Diffraction data collection is the last experimental step of crystal structure determination. All the next stages are performed on a computer, thus can be easily repeated in case of failure. This means, that one should pay great attention while collecting data, as even best diffracting crystal is useless if wrongfully measured. Protein crystals usually diffract weakly, hence extremely bright sources of X-ray radiation are required. Synchrotrons, rings where electrons are accelerated to nearly relativistic speed, provide this ultra-intensive radiation. Nowadays, due to mushrooming synchrotron facilities all over the world, access to the beamlines is not as

limited as before and the granted beamtime is rarely a limiting factor in a diffraction experiment.

The first and probably most important aspect during diffraction data collection is the completeness of the data. In a rotation method, one must make sure to provide a sufficient rotation angle so almost all reciprocal lattice points intersect at least once the Ewald sphere. If the native data are collected the required rotation range is reduced by Friedel's law, stating that the intensities of reflections *h k l* and *-h -k -l* are equal. In addition, these reciprocal lattice points are related by symmetry of the crystal, meaning that even smaller range has to be covered. For this reason it is crucial to elucidate at least the point group of the crystal *a priori* to collection of the entire data set.

Optimal exposure time is another factor that greatly influences data quality. In general, longer exposure results in enhanced signal-to-noise ratio. However, two aspects have to be considered while estimating optimal radiation dose, namely overloads and radiation damage. Overloads result from a limited dynamic range of detector and occur usually in low-resolution zones of data. Reflections which are recorded outside the dynamic range of the detector are neglected, and the resulting data incomplete. This limitation was especially important in case of 16-bit charge-coupled device (CCD) detectors. Modern pixel array detectors, with 20-bit counters, have much wider dynamic range. However, even with this cutting-edge technology, too long exposure causes radiation damage of the crystal. Radiation damage is strongly reduced if the data are collected at cryogenic temperatures, but at the most powerful synchrotrons, the lifetime of a crystal is as short as several minutes.

Subsequent issues concerning the diffraction data collection depend on either properties of a particular crystal. In particular, crystals that diffract to very high resolution (above ~1.3 Å) usually diffract so strongly at low resolution that it is impossible to collect a full data set in a single pass without overloads. Therefore, low-resolution data are collected first, followed by medium or high resolution data with longer exposure time. Crystals with long cell axis also require a different approach as the diffraction spots are too close to each other. Decreasing the rotation angle per single frame is usually sufficient to separate the reflections. Pixel array detectors that work in a shutter-less mode allow to collect data with "fine slicing" of 0.1° or less. However, from my experience, fine slicing, which is always advised at the most powerful synchrotrons, should not be the strategy for weakly diffracting crystals. These crystals are usually small and vulnerable to radiation damage. It is therefore better to collect

data with 0.5° oscillation range per frame because in this case it is possible to record more high resolution spots. In other words, higher resolution can be achieved if a crystal is exposed e.g. for 5 s per 0.5° than if 5 subsequent images are recorded with 1 s exposure per 0.1°. This is probably owed to the fact that shorter exposure is insufficient to cause emission of a single photon corresponding to a high-resolution reflection.

Various experiments and applications also dictate the strategy of diffraction data collection. For instance, anomalous phasing requires very accurately measured reflection intensities so data are collected with high multiplicity (redundancy), not necessarily high resolution. Low-resolution completeness is important in both phasing methods, molecular replacement and anomalous phasing, but less (but still) important in high-resolution structure refinement. For ligand screening or searching, however, one may pay less attention to completeness, accuracy or resolution and automate the data collection to test more crystals.

Diffraction data processing

The raw image frames resulting from diffraction data collection must be converted to a list of measured reflections with their intensities and standard errors. Nowadays, this is performed by elaborated computer programs, such as XDS (Kabsch, 2010) or HKL2000 (Otwinowski & Minor, 1997), and a modern crystallographer can only imagine and salute the pioneers of macromolecular crystallography for their tremendous amount of work.

Data processing starts with corrections of raw frames. These adjustments include, but are not restricted to: detector-specific corrections, beam geometry and crystal orientation. The next step is integration of the reflection intensities. Reflections recorded partially on a few frames are combined and two- or three-dimensional spot profiles are created. Integration results in a list of reflections (indices, intensity and standard error) that are all treated independently. Then the data are merged and scaled accordingly to the initial indexing, and, unless the Bijvoet differences are needed for anomalous phasing, Friedel pairs are combined. Scaling during this process involves corrections of the unit cell parameters, fluctuating beam intensity, radiation damage etc. A reduced list of reflections is obtained as an output.

At the end of data processing the final space group is assigned with careful inspection of systematic absences indicating screw axes. Also, the high resolution cut-off needs to be determined. Common practice is to include the data until the intensity to standard error ratio is about 2. However, if for instance R_{merge} is too high, the data should be cleaved at lower resolution. R_{merge} is a quality indicator of diffraction data, given by equation:

$$R_{merge} = \sum_{hkl} \sum_{i} |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_{i} I_i(hkl)$$

where $I_i(hkl)$ is the intensity of observation *i* of reflection *hkl*.

Structure solution: coping with the phase problem

Reconstruction of the electron density

Crystal structure solution is a reconstruction of the electron density map from the complex structure factors, it is therefore, by some means, the inversion of the diffraction process. The electron density $\rho(x, y, z)$ is given by equation:

$$\rho(x, y, z) = \frac{1}{V} \sum_{hkl} |F_{hkl}| \exp(i\alpha_{hkl}) \exp[-2\pi i(hx + ky + lz)]$$

where V is the cell volume and α_{hkl} the phase angle associated with the structure factor amplitude $|F_{hkl}|$. The electron density in a grid point (x, y, z) is therefore the sum of the contributions of all waves scattered from planes (hkl) added with the appropriate phases. The amplitude of these waves is related to the number of electrons in the plane.

Fourier transform allows transformations between the reciprocal space (structure factors) and the real space (electron density). These mathematical operations are reversible without any loss of information. Fourier transform, however, requires two components for each reflection, namely the amplitude and the phase. The structure factor amplitudes can be easily calculated from the measured intensities. Unfortunately, information about the phase is lost during diffraction experiment and several approaches have been developed to retrieve it (see next sections).

Patterson maps

Patterson function (Patterson, 1934) that represents autocorrelation of the electron density is used to calculate the Patterson map. The Patterson maps, in contrary to electron density maps, are calculated from reflection intensities alone (without the phases). The Patterson maps have $N(N-1)$ peaks at the tips of interatomic distance vectors, where N is the number of atoms. Centrosymmetry of Patterson maps is caused by the fact that any vector $\overrightarrow{AB}$ has a complimentary vector $\overrightarrow{BA}$ with the opposite direction. Applications of Patterson function (and Patterson maps) in macromolecular crystallography include: (i) determination of heavy atom

positions or anomalous scatterers, (ii) determination of the correct orientation of the search model in molecular replacement and (iii) detecting Non-crystallographic symmetries (NCS).

Direct methods

Direct methods, based on phase relations between some sets of structure factors, are mostly used in small-molecule crystallography. In macromolecular crystallography they are applied for *ab initio* structure determination of relatively small proteins that diffract to atomic resolution (above 1.2 Å) and for finding positions of heavy atoms in experimental phasing (*vide supra*).

Isomorphous replacement

Isomorphous replacement is the oldest method of protein structure determination. The pioneers, Max Perutz and John Kendrew used isomorphous replacement for solving the first structure of hemoglobin. In this method, the protein crystal is soaked with a solution containing heavy atoms. These heavy atoms contribute to measurable rises of the reflection intensities. Moreover, contributions from lighter atoms (e.g. carbon) will be decreased because they scatter with a different phase angle. A contribution of different atoms to the scattered intensity is proportional to the square of the number of electrons the atoms contain. The differences of measured intensities can be used to calculate the Paterson map and deduce the positions of heavy atoms. That is, of course, if the derivative crystal is isomorphous to the native one. In practice multiple data sets of different derivatives and the native crystal need to be collected, because a single derivative leads to a phase ambiguity. The phase ambiguity can be also resolved with the use of anomalous scattering (*vide supra*).

Anomalous scattering

Anomalous scattering is nowadays the most common method of *de novo* determination of protein crystal structures. *De novo* in this context means that there is no similar structure known that can be used as a search model in molecular replacement (see the next section). Anomalous scattering occurs at the element-specific absorption edge, when an X-ray photon has a correct energy to excite an electron from one of the inner shells (usually K). As a result, real and imaginary (*f'* and *f''*, respectively), λ-dependent corrections to the atomic scattering factor f_j are introduced or, mathematically:

$$f_j = f_j^0(\theta) + f_j'(\lambda) + if_j''(\lambda)$$

where f_j^0 is the normal scattering factor, dependent on diffraction angle θ. This leads to the breakdown of Friedel's law and reciprocal lattice is no longer centrosymmetric. The Bijvoet differences between the Friedel pairs of reflections *h k l* and *-h -k -l* are proportional to the atomic number of the anomalous scatterer. For instance, anomalous scattering from sulphur atom is very subtle when compared to e.g. mercury or uranium. In practice, either one of two approaches during diffraction data collection is used. Multi-wavelength anomalous diffraction (MAD) (Hendrickson, 1991) requires collection of multiple (usually three or four) data sets at different wavelengths. Single-wavelength anomalous diffraction (SAD, on the other hand, relies on one data set and, when combined with phase extension techniques is a very efficient method. Unfortunately, it is impossible to predict *a priori* whether SAD will yield a structure solution. For this reason, presented by Z. Dauter (Dauter, 2002) One-and-a-half wavelength approach is optimal. It involves collecting the first data set at the absorption edge, trying to process the data and solve the structure while collecting subsequent data sets at different wavelengths.

Molecular Replacement

When a structure similar to the target protein is available, molecular replacement (Rossmann, 2001) is usually the first phasing method chosen by a macromolecular crystallographer. Molecular replacement is based on a search for the correct orientation and translation of the model (search probe) in the unit cell of the unknown crystal structure. Many trials of the search are required but thanks to fast computers and elaborated software, such as Phaser (McCoy *et al.*, 2007) the solution will be ultimately found, provided it exists. The orientation of the search model is determined by matching the intramolecular Patterson vectors. Intermolecular Patterson vectors, on the other hand, are used for finding the proper location (translation) of the oriented molecule.

Of course, one may only predict whether these, model and target, structures are similar based on their sequence. In general, it is assumed that proteins which share more that 30% identity in their amino acid sequence (Rupp, 2010) can serve as models for molecular replacement and the higher the identity, the most probable the successful structure solution is. Growing number of protein structures in PDB (Bergman *et al.*, 2000) makes this method very popular. It is a good practice to appropriately prepare the model before using it as a search probe. This involves trimming the disordered parts, loops and domains (or fragments) that are not

expected to have a similar structure in the target. The solvent and other ions are automatically removed by modern software.

Model building and refinement

A successful structure solution provides only a rough view over the model and electron density maps. The obtained model always needs to be improved in many consecutive steps of model building and refinement in real and reciprocal spaces. If the data resolution is sufficiently good (usually above 2 Å) automatic model builders, such as ARP/wARP (Cohen *et al.*, 2008; Langer *et al.*, 2008) can be used for the first round of real-space modeling. Subsequently, programs for manual real-space refinement, e.g. COOT (Emsley *et al.*, 2010), are used for corrections of the model that is built into electron density maps. The model is refined with respect to different types of electron density maps, mainly $2F_o$-F_c or difference map F_o-F_c, where F_o and F_c are observed and calculated structure factors, respectively. The proper values for bond lengths, angles, planar groups, chirality, torsion angles and forbidden non-bonding interactions must be also considered during the refinement. Most of these values are provided as geometrical restraints in refinement programs and can be treated more freely with increasing data resolution which in macromolecular crystallography is rarely high enough to disregard the restraints completely. Manual structure refinement must also include modeling of metal cations according to their coordination sphere, determination of proper orientation (flip) of glutamine, asparagine and histidine residues with respect to hydrogen bonding network and ADPs. Ligands are often of particular interest and should be modeled with most careful inspection of the difference electron density maps and geometry.

Each round of manual real-space model building and refinement is followed by automated refinement of both: real- and reciprocal-space refinement in programs such as Refmac (Murshudov *et al.*, 2011), phenix.refine (Adams *et al.*, 2010) or SHELXL (Sheldrick, 2008). The first two programs use maximum likelihood minimization function, which is advantageous at lower resolutions or early stages of refinement because it accounts for errors in the model and incompleteness. SHELXL, on the other hand, uses more accurate, least-squares target functions which are superior to maximum likelihood during refinement against atomic resolution data (1.2 Å or higher). The agreement of the refined model and data is usually expressed by R_{work} value, given by equation:

$$R_{work} = \frac{\sum_{hkl \notin free} |F_o - F_c|}{\sum_{hkl \notin free} F_o}$$

where F_o and F_c are observed and calculated structure factors, respectively. R_{work}, however is very susceptible to overparametrization of the model, thus R_{free}, calculated with the use of approx. 1000 reflections excluded from refinement is a valuable tool for cross-validation. If the R_{free} - R_{work} difference is too high (more than 0.05) the number of parameters included in the refinement should be decreased. The data-to-parameter ration is a very important issue in macromolecular structure refinement and needs to be considered in context of extensive solvent modeling and number of parameters refined for each atom. During isotropic refinement (all atoms expressed as spheres) there are four parameters for each individual atom to be refined, namely the coordinates (x, y, z) and ADP. If one may want to refine the structure anisotropically (atoms are ellipsoids), five more parameters need to be refined. Sufficient amount of data (number of unique reflections) for anisotropic refinement is provided only by crystals diffracting to very high resolutions (usually better than 1.4 Å). Anisotropic movements of atoms have of course reasonable explanation as, for example, side-chain atoms will move more freely parallel to the rigid protein backbone than in perpendicular direction. The account of this sort of movement is taken into account in translation/libration/screw (TLS) (Winn *et al.*, 2001) approach that allows for treating groups of atoms as moving groups with common anisotropic properties.

The refinement steps are also combined with model validation with the use of e.g. MolProbity (Chen *et al.*, 2010) that inspects whether the proposed model does not break the laws of nature. Once there are no interpretable peaks in the difference density maps, no unexplained deviations from geometric target values, the model makes biological and chemical sense and R_{work} and R_{free} factors are low and do not differ significantly from each other, the refinement can be abandoned. Abandoned, because, quoting George Sheldrick (2008): *One should bear in mind that a macromolecular refinement against high resolution data is never finished, only abandoned*.

Micro-Scale Thermophoresis

Micro-scale thermophoresis (MST) is a method developed for quantitative analysis of interactions between biomolecules (Duhr & Braun, 2006). It interprets the directed movement of particles in a temperature gradient. Changes in macromolecular shape, size or hydration shell can be observed using MST. One of the biomolecule used in the experiment needs to be labeled with a fluorescent marker. The movement of the biomolecule is monitored on a concentration scale, where the buffer is kept constant. The experiment is performed in capillaries with different ratios of the interacting partners. An IR laser is pointed towards the

sample (usually for 30 s) causing local temperature increase (by 2-8K) and the resulting directed movement of particles in the temperature gradient (Fig. 16). Relaxation to the initial state is also measured after the laser is switched off. With the use of this method, binding affinities and kinetics can be determined. MST is most useful when protein-protein interactions are evaluated, since the changes in molecular movement are substantial after binding of a large partner.

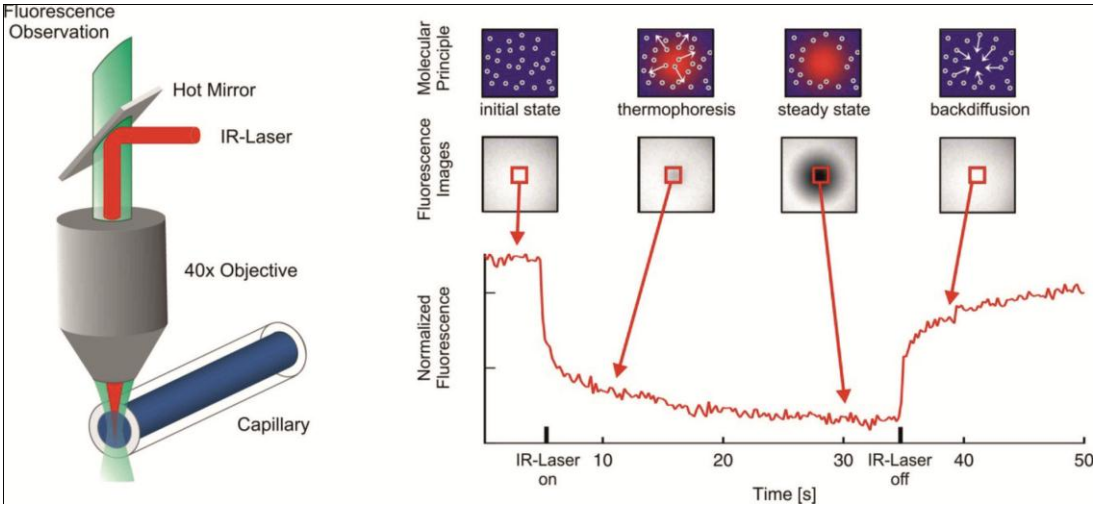


Figure 16
Micro-scale thermophoresis. Left panel, the solution inside the capillary is locally heated with a focussed IR-laser, which is coupled into the path of light using a hot mirror. Right panel, the fluorescence inside the capillary is imaged with a photodiode and the normalized fluorescence in the heated spot is plotted against time. The IR-laser is switched on at t=5s and the fluorescence decreases as the temperature increases and the labelled molecules/complexes move away from the heated spot due to thermophoresis. Courtesy of NanoTemper Technologies.

Autoradiography

Autoradiography is a method used to study the distribution of radioactive substances in a given object, based on emission of radiation from the sample itself (thus the prefix auto-). In molecular biology it is very commonly applied to measure the distribution of a radioactively labeled sample of a protein or nucleic acid. Best results are obtained if the sample is separated by electrophoresis prior to actual autoradiography because in this case radiation from other sources can be removed. The radiation from the spots on the gel can be subsequently analyzed using either photographic films that are sensitive to radiation emission or enhancing screens and scanned.

Dynamic light scattering

The size of a typical protein molecule (hydrodynamic radius in the 10-100 Å range) is much smaller than the wavelength of visible light (e.g. 5000 Å for green light), that is used in DLS experiment. The nature of scattering is determined by the Mie ratio ($x_M = 2\pi r / \lambda$, where r is the particle radius, λ is the wavelength) which, in case of DLS experiments is much smaller than one. In this case, the Rayleigh scattering occurs, meaning that the visible light is isotropically scattered on entire protein molecules. As a contrary, X-rays, with wavelengths in ~1 Å range, scatter on electrons of individual atoms (Thomson scattering). For the $x_M << 1$ (the limit for Rayleigh scattering), the total scattering intensity is given by an equation:

$$I = I_0 \frac{8\pi^4 \alpha^2}{\lambda^4 R^2} (1 + cos^2 2\theta)$$

where I_0 in the intensity of primary beam, λ the wavelength, R the detector distance, θ the scattering angle and α the polarization volume (expressing interaction between the particle and the electric field vector). The scattering intensity, therefore, highly depends on the wavelength, but even higher on the particle radius (α is proportional to r^3, so α^2 to r^6). This means, that any contamination with e.g. protein aggregates or a dust particle will dominate the intensity. In solution, Brownian motions of particles influence their diffusion coefficient (and hydrodynamic radius, R_H). Hence, by measuring the fluctuations of the scattered intensity in the DLS experiment the hydrodynamic radius can be estimated. The DLS method is useful not only for the measurement of the protein size in solution, but also for analysis of homogeneity of a protein sample. The latter property can be used to predict the chance of success in crystallization trials, because heterogeneous (polydisperse) samples are less likely to crystallize (D'Arcy, 1994).

Small-angle X-ray scattering

Compared to visible light, scattering of X-rays provides us with more information because the wavelength is four orders of magnitude shorter. In contrast to diffraction experiments on periodic, crystalline materials, in SAXS the scattering molecules are randomly oriented. As a result, instead of a three dimensional scattering function (diffraction image), SAXS measurements result in a one dimensional scattering curve. The small-angle scattering curve is a description of an overall shape of a particle with spherical approximation. The intensity of the scattered X-rays from a solution sample decays rapidly with an increasing scattering angle, thus can only be measured in the forward direction, close to the primary beam (at very

low angles). During the analysis, the scattered intensity, I_s is interpreted as a function of the scattering vector ($s = 4\pi\sin\theta / \lambda$; θ is the scattering angle and λ the wavelength) and is given by Guinier approximation:

$$I_s = I_0 \exp(-s^2 R_g^2/3)$$

where I_0 is the extrapolated scattering intensity at zero scattering angle and R_g the radius of gyration. R_g can be determined from the linear part of the log (I_s / I_0) curve. Noteworthy, the approximate relation between R_H (from DLS experiment) and R_g for elongated particles (Berry, 2010) is given by the equation $R_g = 0.8 \cdot R_H$. Additional information about the shape of the molecular envelope can be derived using sophisticated software, such as *PRIMUS* (Petoukhov *et al.*, 2012) and *CRYSOL* (Svergun *et al.*, 1995). The resulting $P(r)$ (pair-distance distribution) function (Svergun & Koch, 2003) describes the distribution of all possible distances within the particle of interest, therefore is a valuable tool for detecting conformational changes within a macromolecule. To provide graphical information about the shape of the molecule one needs to use a model building program, such as *DAMMIN* (Svergun, 1999) or *GASBOR* (Svergun *et al.*, 2001).

Art work

Molecular figures were created with UCSF Chimera (Pettersen *et al.*, 2004)

Experimental procedures

General remarks

1. The touchdown PCR protocol (Don *et al.*, 1991) with KOD Hot Start DNA Polymerase (Merck) was used to amplify coding sequences of proteins. The original PCR reaction buffer was supplemented with 1 M betaine to suppress the negative effect of GC-rich regions. Cycling parameters were as follows: initial denaturation 95°C for 3 min, primer annealing for 30 s and elongation at 70°C for appropriate time (25 s per thousand of base pairs). The annealing temperature was decreased by 0.5°C every cycle from the initial 62°C. Once the annealing temperature reached 55°C, the PCR was continued for 18 more cycles. Final extension step was not included in the protocols and the samples were cooled down to 4°C.
2. PCR products were purified using PureLink PCR Purification Kit (Invitrogen).
3. Plasmids were obtained using GeneMATRIX Plasmid Miniprep DNA Purification Kit (Eurx).
4. For large-scale overexpression, *E. coli* cells were cultured in 2.5 l plastic bottles in 1 l of media for proper aeration.
5. Concentration of the protein samples was performed using Amicon centrifugal filters at 6 000 G with 10 kDa molecular weight cut-off (Millipore).
6. SEC was performed using ÄKTA prime FPLC system.
7. SDS-PAGE electrophoreses were performed using Mini Protean (Bio-Rad) aparatus and gradient 4–15% precast polyacrylamide gels (Bio-Rad).

Medicago truncatula Histidine-containing Phosphotransfer protein 1

Cloning

The homologous sequence of *A. thaliana ahp*1 (GenBank: **AB015141**) was used in BLAST (Altschul *et al.*, 1997) as a probe to search for the gene coding for MtHPt1 in *M. truncatula* genome transcript (Moskal *et al.*, 2008). The MtHPt1 protein has "Evidence at protein level" annotation in the UniProt Knowledgebase and is accessible under the code **B7FGU6**.

The coding sequence of MtHPt1 was amplified by PCR using *M. truncatula* (ecotype J5) cDNA from roots as template (Forward CACCATGGAGGTGGGTCAG and Reverse TCAGTTCAGTTCCATCATAGGGA PCR primers were used). The reaction product was cloned into the pET-TOPO-151D (Invitrogen) vector. This vector introduces the His_6 tag at N-terminus. The correctness of the insert was confirmed by DNA sequencing. Small-scale overexpression tests in *E. coli* BL21 cells showed poor solubility of the protein. The pMCSG9 vector (from the Midwest Center for Structural Genomics), which adds an N-terminal His_6 followed by Maltose Binding Protein (MBP) fusion tag was used to improve the solubility. The pMCSG9-Mthpt1 construct was obtained by LIC (Kim *et al.*, 2011) after PCR gene amplification with the pET-TOPO-151D plasmid as template. With the use of the pMCSG9-Mthpt1 construct, the solubility of the expressed protein was significantly higher.

Overexpression

Overexpression of MtHPt1 protein was carried out in BL21 Magic *E. coli* cells. The bacteria were cultured with shaking in LB media supplemented with 120 µg/ml ampicillin and 25 µg/ml kanamycin at 37°C until the OD_{600} reached 1.0. After that, the temperature was set to 18°C and IPTG was added to a final concentration of 0.5 mM. The culture was grown for 18 h and then centrifuged at 4500 rpm for 10 min at 4°C. Cell pellet from 1l culture was resuspended in 30 ml of binding buffer (50 mM Tris-HCl, pH 8.0, 500 mM NaCl, 20 mM imidazole, 1 mM TCEP), flash-frozen in liquid nitrogen and stored at -80°C prior to purification.

Purification

The samples were thawed and the cells were disrupted by sonication using 3 s bursts with appropriate intervals for cooling. The sonication was carried out for 4 min of total working time of the probe. Cell debris was pelleted by centrifugation at 15 000 rpm for 30 min at 4°C. The supernatant was applied on a column packed with 6 ml of Ni Sepharose HP resin (GE Healthcare), connected to VacMan (Promega). The chromatographic process was accelerated

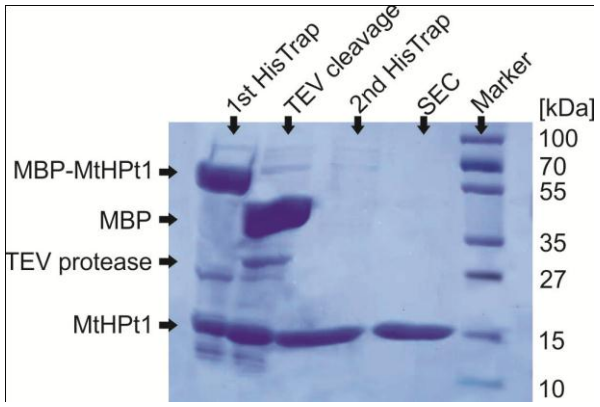


Figure 17
SDS-PAGE electrophoresis of MtHPt1 purification steps.

by a vacuum pump. There is no need to incubate the sample with Ni Sepharose HP resin for binding, hence, the column was washed 4 times with 30 ml of binding buffer and the purified protein was eluted with 15 ml of elution buffer (50 mM Tris-HCl, pH 8.0, 500 mM NaCl, 300 mM imidazole, 1 mM TCEP). The His_6-MBP-tag was cleaved with TEV protease and the excess of imidazole was removed by dialysis (overnight, 4°C) simultaneously. The solution was mixed with Ni Sepharose HP resin to get rid of the His_6-MBP-tag and the His_6-tagged TEV protease. The flow-through was collected, concentrated to 4 ml and applied on a HiLoad Superdex 200 16/60 column (GE Healthcare) equilibrated with a buffer composed of 50 mM Tris-HCl, pH 8.0, 200 mM NaCl and 1 mM TCEP. The size exclusion chromatography step yielded a homogenous, monomeric protein fraction of about 6 ml. The sample was concentrated to 6 mg/ml, as determined by the method of Bradford (Bradford, 1976), and used for crystallization.

Crystallization

Robotic Sitting Drop Vapor Diffusion setup (Mosquito) was applied for initial screening of crystallization conditions. Screening was performed using Crystal Screen HT, Index HT and Salt Rx reagents from Hampton Research. Manual optimization of the initial crystallization hits in hanging drops gave the following crystallization conditions: 100 mM Bis-Tris, pH 5.5, 25% PEG3350 and 200 mM ammonium acetate. The drop was mixed from 1 µl of the reservoir solution and 3 µl of protein solution. Crystals grew within 4 days at 18°C to dimensions of approx. 0.5 x 0.2 x 0.2 mm (Fig. 18). After 2 months they were harvested with 0.1 mm nylon loops (Hampton Research), washed with cryo-protectant solution containing 20% (v/v) glycerol in the reservoir cocktail, and vitrified in liquid nitrogen for synchrotron-radiation data collection.

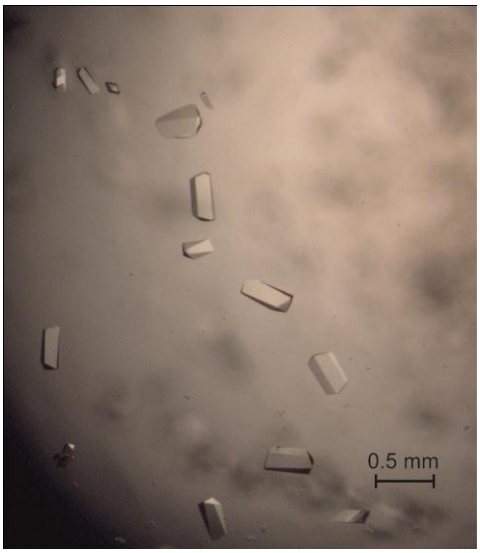


Figure 18
Crystals of MtHPt1 protein

Diffraction data collection and processing

X-Ray diffraction data were collected at the Advanced Photon Source (ANL, Argonne, IL, USA) 22-BM beamline. A total number of 200 diffraction images with 0.5^{o} oscillation were indexed, integrated and scaled using XDS (Kabsch, 2010). Details of data collection and processing are presented in Table 1.

Table 1. MtHPt1: Data collection

Wavelength (Å)	1.0000
Crystal-to-detector distance (mm)	153
Oscillation (°)	0.5
Temperature (K)	100
Space group	$P2_12_12_1$
Unit cell parameters	
a, *b*, *c* (Å)	38.6, 44.9, 85.9
Resolution (Å)	31.0-1.45 (1.53-1.45)[a]
Total reflections	98484
Unique reflections	25816
Completeness (%)	94.3 (73.7)
Multiplicity	3.8 (2.5)
R_{merge} (%)	6.1 (52.4)
$\langle I/\sigma(I)\rangle$	14.0 (2.0)

[a]Values in parentheses correspond to the highest resolution shell.

Structure solution and refinement

The crystal structure of MtHPt1 was solved by molecular replacement using Phaser (McCoy *et al.*, 2007). The crystal structure of a homologous protein from rice (*Oryza sativa*; PDB code 1yvi; Center for Eukaryotic Structural Genomics) was used as a search probe. ARP/wARP (Langer *et al.*, 2008) was applied for automatic model building. Coot (Emsley *et al.*, 2010) was used for manual fitting in electron density maps between model refinement and validation cycles carried out in phenix.refine (Adams *et al.*, 2010). Hydrogen atoms were added at riding positions. An attempt to refine anisotropic ADPs resulted in a significant increase of R_{free}. Therefore, the isotropic model, with seven TLS groups (Winn *et al.*, 2001), defined as suggested by the TLSMD server (Painter & Merritt, 2006), was accepted as final. The final refinement statistics are listed in Table 2.

Table 2. MtHPt1: Refinement

No. of reflections, working set	24814
No. of reflections, test set	1002
No. of atoms (non H)	
protein	1250
solvent	159
R_{work}/R_{free} (%)	16.7/19.0
RMSD from ideal geometry	
bond lengths (Å)	0.019
bond angles (°)	1.6
Ramachandran statistics (%)	
most favored	97.7
allowed	2.3

Generation of H79N mutant of MtHPt1

Site-directed mutagenesis of MtHPt1 which changed the active His79 residue to Asn was performed using the PIPE technique (Klock & Lesley, 2009) with the use of KOD Hot Start DNA Polymerase (Merck). The pMCSG9-Mthpt1 construct was used as a template. The mutation (bold in underlined codon) was introduced using two primers: 5'-GCTCATGTT**A**AACCAGTTTAAGGGTAG-3' (coding strand) and 5'-CTGGT**T**AACATGAGCATCAACTTTTTTG-3' (complementary strand). The template plasmin was digested with DpnI nuclease (New England Biolabs) at 37°C for 3 h followed by denaturation of the enzyme (80°C, 20 min) . The reaction product was used for transformation of BL21 Magic *E. coli* cells. The correctness of the resulting vector was confirmed by DNA

sequencing. The mutated MtHPt1 protein was overexpressed and purified as described for the wild-type protein above.

In vitro phosphorylation

The procedure of obtaining the recombinant MtCRE1' is described in a separate chapter later in this part. To perform *in vitro* phosphorylation, MtCRE1' was mixed with wild-type MtHPt1 or its H79N mutant at different molar ratios in a buffer containing 50 mM Tris-HCl, pH 8.0, 200 mM NaCl, 10 mM $MgCl_2$ and 1 mM TCEP in 20 µl aliquots. *In vitro* phosphorylation reactions (at room temperature) were initiated by the addition of 1 µl of [γ-^{32}P]ATP solution (1 µCi/µl and 5 mM final ATP). After 20 min, the reactions were quenched by the addition of SDS-PAGE loading buffer, and immediately subjected to electrophoretic analysis using a Mini-Protean TGX system (Bio-Rad). The autoradiography was performed using an enhancing screen followed by radioisotope imaging with an FLA 5100 image analyzer (Fuji). The gel was stained with Coomassie Brilliant Blue R-250 and scanned using a regular desktop scanner.

Micro-scale thermophoresis

MtCRE1' was labeled with NT-647 dye using NanoTemper's Protein Labeling Kit RED (L001, NanoTemper Technologies) to perform the MST experiment. The labeled MtCRE1' was kept at a constant concentration of 20 nM, while the MtHPt1 solution (initial concentration 500 µM) was prepared as a dilution series of 15 concentrations. The analysis was carried out in a buffer containing 50 mM Tris-HCl pH 7.4, 150 mM NaCl, 10 mM $MgCl_2$ and 0.05% Tween. Each reaction mixture was incubated in 20 µl for 10 min prior to thermophoresis in hydrophilic MST-grade glass capillaries. The experiment was performed with Monolith NT.115 apparatus (NanoTemper Technologies).

Medicago truncatula Histidine-containing Phosphotransfer protein 2

Cloning

The coding sequence of MtHPt2 gene was used as a search probe for analogous sequences in M. truncatula transcript (Moskal *et al.*, 2008) using BLAST (Altschul *et al.*, 1997). There are several HPt sequences in the *M. truncatula* genome, as revealed by the recently completed genome sequencing project (Branca *et al.*, 2011; Young *et al.*, 2011). The coding sequence of MtHPt2 was amplified by PCR using *M. truncatula* (ecotype J5) cDNA from roots as template. Following PCR primers were used: Forward TACTTCCAATCCAATGCCATGGATCACTTACATAGGAAGCTACGC and Reverse TTATCCACTTCCAATGTTATTAACCCCTTGTGTGAATAGATCTACCG (fragments of primers pairing the MtHPt2 coding sequence are underlined, the remaining nucleotides are complementary to the cloning site of the pMCSG vectors). The reaction product was cloned into the pMCSG48 vector (from the Midwest Center for Structural Genomics), which adds an N-terminal His_6 followed by NusA fusion tag. The pMCSG48-Mthpt2 construct was obtained by LIC method (Kim *et al.*, 2011). The correctness of the insert was confirmed by DNA sequencing.

Overexpression

MtHPt2 protein was overexpressed in BL21 Magic *E. coli* cells. The bacteria were cultured with shaking in LB media supplemented with 120 μg/ml ampicillin and 25 μg/ml kanamycin at 37°C until the OD_{600} reached 1.0. Subsequently, the temperature was set to 18°C and IPTG was added to a final concentration of 0.5 mM. The culture was grown for 18 h and then centrifuged at 6000 rpm for 6 min at 4°C. Cell pellet from 1l culture was resuspended in 30 ml of binding buffer (50 mM Tris-HCl, pH 8.0, 500 mM NaCl, 20 mM imidazole, 1 mM TCEP) and stored at -80°C.

Purification

The samples were thawed and the cells were disrupted by sonication using 3 s bursts with appropriate intervals for cooling. The sonication was carried out for 4 min of total working time of the probe. Cell debris was removed by centrifugation at 15 000 rpm for 30 min at 4°C. The supernatant was applied on a column packed with 6 ml of Ni Sepharose HP resin (GE Healthcare), connected to VacMan (Promega). The chromatographic process was accelerated by a vacuum pump. The column was washed 4 times with 30 ml of binding buffer and the purified protein was eluted with 15 ml of elution buffer (50 mM Tris-HCl, pH 8.0, 500 mM

NaCl, 300 mM imidazole, 1 mM TCEP). The His_6-NusA-tag was cleaved with TEV protease and the excess of imidazole was removed by dialysis (overnight, 4°C) simultaneously. The sample was mixed with Ni Sepharose HP resin to get rid of the His_6-NusA-tag and the His_6-tagged TEV protease. Only the flow-through was collected, concentrated to 5 ml and applied on a HiLoad Superdex 200 16/60 column (GE Healthcare) equilibrated with a buffer composed of 50 mM Tris-HCl, pH 8.0, 200 mM NaCl and 1 mM TCEP. The size exclusion chromatography step yielded a homogenous, monomeric protein fraction of about 8 ml. The sample was concentrated to 8 mg/ml, as determined by the method of Bradford (Bradford, 1976), and used for crystallization.

Crystallization

Robotic Sitting Drop Vapor Diffusion setup (Gryphoon, Art Robins Inc.) was applied for initial screening of crystallization conditions. Screening was performed using a set of three screens: JCSG plus, PACT Premier, Morpheus (Molecular Dimensions). The Morpheus screen proved to be most useful while searching for best crystallization conditions because of two reasons. Firstly, crystals appeared in all eight rows within the first column (Buffer system 1, pH 6.5, and a mixture of PEG550MME and PEG 20K). Secondly, there is no need to search for cryprotectant, since each condition provides such protection. Manual optimization of the initial crystallization hits in hanging drops gave the following crystallization conditions: Buffer system 1, pH 6.5, 30% PEG 550 MME and PEG 20K, 0.09M Halogens. The mixture was obtained from genuine stocks from Molecular Dimensions. The drop was mixed from 1 µl of the reservoir solution and 3 µl of protein solution. On one of the cover slides only a single crystal grew within 7 days at 19°C to enormous dimensions (for macromolecular crystallography) of approx. 1 x 0.5 x 0.5 mm (Fig. 19). After 2 months the crystal was harvested with 0.3 mm nylon loops (Hampton Research) and vitrified in liquid nitrogen for synchrotron-radiation data collection.

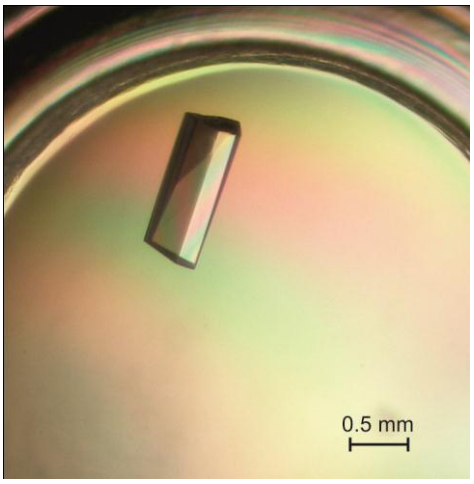


Figure 19
Single crystal of MtHPt2 protein.

Diffraction data collection and processing

X-Ray diffraction data were collected at the BESSY (Berlin, Germany) 14.2 beamline. The crystal diffracted X-ray radiation to very high angles, hence, the data were collected in three passes (Fig. 20), each containing 240 images of 1° oscillation range. The first pass, that

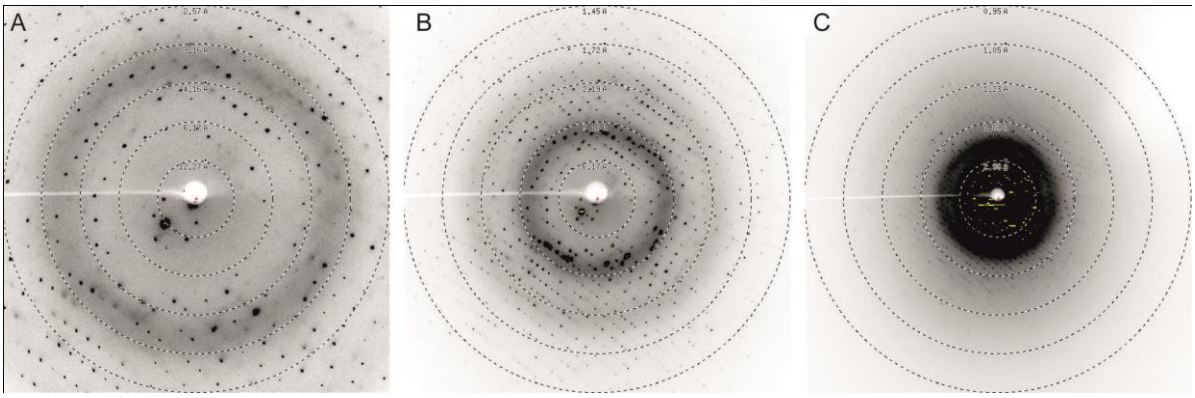


Figure 20
Three passes of diffraction data collection on MtHPt2 crystal. The data resolution rings extend to 2.6 Å (A), 1.47 Å(B) and 0.95 Å (C). Crystal is in the same orientation, with respect to the incident beam, in all three frames. Data from the edges of the detector was also included, therefore the final resolution is 0.92 Å.

included low resolution reflections up to 2.6 Å, was collected with 1.2 s exposure time and with the beam attenuated by a 0.12 mm Al filter. The second, medium resolution pass to 1.47 Å was collected with the same attenuation and 3.2 s exposure time. The third pass was collected without attenuation and 10 s exposure time. Each set has been processed separately in XDS (Kabsch, 2010) and the data were merged using XSCALE (Kabsch, 2010). Details of data collection and processing are listed in Table 3.

Table 3. MtHPt2: Final statistics of data collection

Wavelength (Å)	0.91841
Temperature (K)	100
Space group	$P2_1$
Unit cell parameters	
a, *b*, *c* (Å)	25.2, 50.2, 50.4
β (°)	97
Resolution (Å)	50-0.02 (0.94-0.92)[a]
Unique reflections	84634
Completeness (%)	97.8 (75.1)
Multiplicity	6.3
R_{merge} (%)	3.6 (73.8)
$\langle I/\sigma(I)\rangle$	21.8 (1.9)

[a]Values in parentheses correspond to the highest resolution shell.

Structure solution and refinement

The crystal structure of MtHPt1 (Ruszkowski *et al.*, 2013) was used as a search probe for molecular replacement using Phaser (McCoy *et al.*, 2007). ARP/wARP (Langer *et al.*, 2008) was applied for automatic model building. Coot (Emsley *et al.*, 2010) was used for manual fitting in electron density maps between model refinement SHELXL (Sheldrick, 2008).

Hydrogen atoms were added at riding positions. ADPs were refined anisotropically for all protein atoms. Validation of the final model was carried out in MolProbity (Chen *et al.*, 2010). The final refinement statistics are listed in Table 4.

Table 4. MtHPt2: Refinement

No. of reflections, working set	83626
No. of reflections, test set	1008
No. of atoms (non H)	
protein	1182
solvent	207
R_{work}/R_{free} (%)	12.1/15.5
RMSD from ideal geometry	
bond lengths (Å)	0.023
angle distances (Å)	0.040
Ramachandran statistics (%)	
most favored	98.3
allowed	1.1
outliers	0.6

Medicago truncatula Cytokinin Receptor

Cloning

Sequence analysis of the *M. truncatula* cytokinin receptor MtCRE1 (GenBank: **357521450**) revealed that it contains an N-terminal part consisting of an extracellular cytokinin-binding domain and a transmembrane domain. The intracellular C-terminal part (MtCRE1') comprised of residues Ile352 -- Ser1003 was amplified using PCR reaction with *M. truncatula* (ecotype J5) cDNA from roots as template. The primers used for amplification of the gene fragment allowed to use LIC method and clone the reaction product into the pMCSG9 vector (Forward TACTTCCAATCCAATGCCATGATTTTATATGGTGCTGGGAATCAT and Reverse TTATCCACTTCCAATGTTATCATGAATCTACTGAAGTAGGTTTTGT, fragments of primers pairing the MtCRE1' coding sequence are underlined, the remaining nucleotides are complementary to the cloning site of the pMCSG vectors). The reaction product was cloned into the pMCSG9 vector.

Overexpression

Overexpression of MtCRE1' protein was carried out in BL21 Magic *E. coli* cells. The bacteria were cultured with shaking in LB media supplemented with 120 µg/ml ampicillin and 25 µg/ml kanamycin at 37°C until the OD_{600} reached 1.0. After that, the temperature was set to 18°C and IPTG was added to a final concentration of 0.5 mM. The culture was grown for 18 h and then centrifuged at 6000 rpm for 6 min at 4°C. Cell pellet from 2 l culture was resuspended in 60 ml of binding buffer (50 mM Tris-HCl, pH 8.0, 500 mM NaCl, 20 mM imidazole, 1 mM TCEP), flash-frozen in liquid nitrogen and stored at -80°C prior to purification.

Purification

The bacterial pellet from 2 l culture, suspended in binding buffer, was thawed and the cells were disrupted by sonication using 3 s bursts with appropriate intervals for cooling. The sonication was carried out for 4 min of total working time of the probe. Cell debris was removed by centrifugation at 20 000 rpm for 30 min at 4°C. The purification steps are presented on an SDS-PAGE gel in fig. 21.

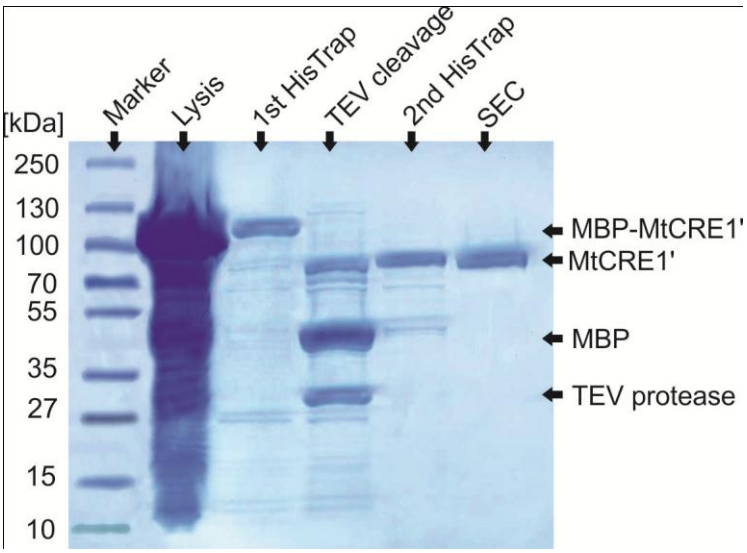


Figure 21
SDS-PAGE electrophoresis of MtCRE1' purification steps.

The supernatant was applied on a column packed with 6 ml of Ni Sepharose HP resin (GE Healthcare), connected to VacMan (Promega). The chromatographic process was accelerated by a vacuum pump. The column was washed 4 times with 30 ml of binding buffer and the purified protein was eluted with 15 ml of elution buffer (50 mM Tris-HCl, pH 8.0, 500 mM NaCl, 300 mM imidazole, 1 mM TCEP). The His_6-MBP-tag was cleaved with TEV protease and the excess of imidazole was removed by dialysis (overnight, 4°C) simultaneously. The solution was mixed with Ni Sepharose HP resin to get rid of the His_6-MBP-tag and the His_6-tagged TEV protease. The flow-through was collected, concentrated to 4 ml and applied on a HiLoad Superdex 200 16/60 column (GE Healthcare) equilibrated with a buffer composed of 50 mM Tris-HCl, pH 8.0, 200 mM NaCl and 1 mM TCEP. The final purification step, size exclusion chromatography, yielded a homogenous fraction of dimeric MtCRE1' (6 mg of protein from 2 l culture). The dimerization was also confirmed by DLS experiments (not shown).

Reductive lysine methylation

The procedure was derived from that reported previously (Rypniewski *et al.*, 1993; Rayment, 1997). A sample of MtCRE1' after the second affinity chromatography step, at 1 mg/ml concentration in a volume 20 ml of was used for the experiment. Tris-based buffer was removed by three-fold dialysis against buffer composed of 50mM HEPES, pH 7.5, 250 mM NaCl, 1 mM TCEP at 277K. Entire reductive lysine methylation was carried out at 277K. Reaction was initiated by addition of 400 µl of freshly prepared 1 M dimethylamine-borane

complex (Sigma) and 800 µl of 1 M formaldehyde (made from 37% stock, Merck). After 2 h, the same amounts of both reagents were added to the protein solution and the reaction was continued for another 2 h. Following the final addition of 200 µl of 1 M dimethylamine-borane complex and 400 µl of 1 M formaldehyde, the mixture was incubated overnight at 277K. The reaction was quenched by the addition of 10 ml of 50 mM Tris-HCl, pH 8.0, 200 mM NaCl and 1 mM TCEP. Small amount of precipitated protein was removed by centrifugation at 16 000 G and 277K. Soluble fraction of the methylated MtCRE1' was concentrated to 3 ml and subjected to SEC with HiLoad Superdex 200 16/60 column (GE Healthcare) equilibrated with a buffer composed of 50 mM Tris-HCl, pH 8.0, 200 mM NaCl and 1 mM TCEP. The fractions were analyzed using SDS-PAGE electrophoresis and those of pure protein were concentrated to 4.5 mg/ml and subjected to crystallization.

Limited proteolysis

Limited proteolysis was performed with the use of two sets of proteases from Hampton Research, namely Proti-Ace (α-Chymotrypsin, Trypsin, Elastase, Papain, Subtilisin and Endoproteinase Glu-C) and Proti-Ace 2 (Proteinase K, Clostripain, Pepsin, Thermolysin, Bromelain and Actinase). Small-scale screening experiments were performed on a 20 µl samples of MtCRE1' at 7.5 mg/ml concentration. One µl of each protease (at 10 µg/ml concentration) was added to a separate MtCRE1' aliquot. The proteolysis was performed at 292K and samples were aspirated and quenched after 2.5 h and after an overnight incubation, giving two measuring points. The samples were subjected for SDS-PAGE electrophoresis (Fig. 22). Large scale procedure was performed with the use of the same protocol on a thirty-fold larger volumes of the samples. After an overnight treatment the samples were diluted to 3 ml and dialyzed in Slide-a-lyzer (Thermo Scientific) against 50 mM Tris-HCl, pH 8.0, 200 mM NaCl and 1 mM TCEP buffer to get rid of the small peptides. The resulting protein solutions were concentrated to 5 mg/ml and subjected to crystallization.

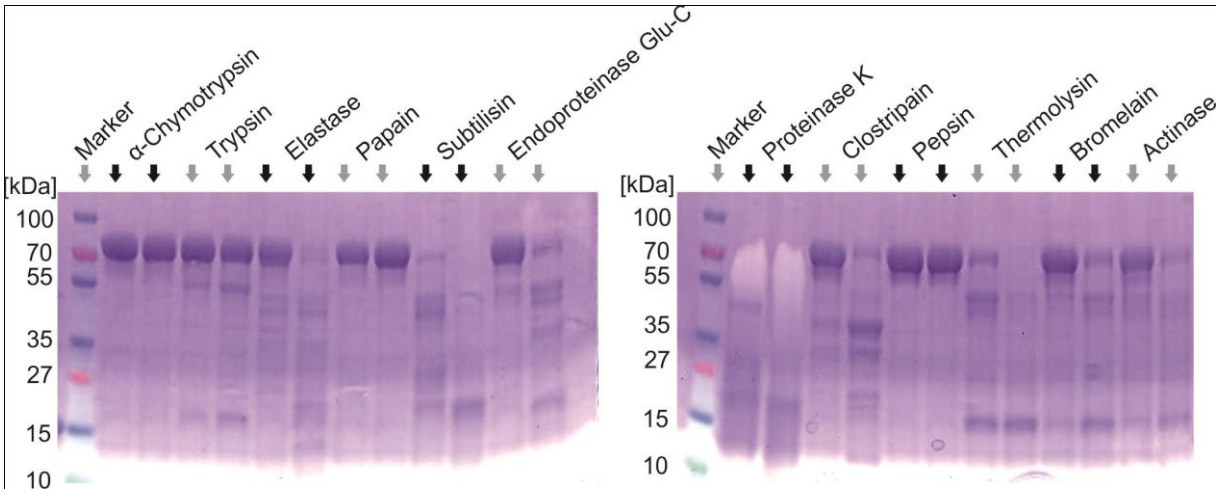


Figure 22
SDS-PAGE electrophoresis of limited proteolysis of MtCRE1'. Left and right panels correspond to Proti-Ace 1 and 2, respectively. Two measuring points were examined for each protease. Namely, left arrow represents the sample after 2.5 h and right arrow after overnight incubation. The gels were not completely destained to show that the sample after treatment with Proteinase K is not stained with Coomassie Brilliant Blue R-250.

Crystallization

Crystallization trials of an unmodified MtCRE1' at 3, 6, 10 mg/ml concentrations were performed using Robotic Sitting Drop Vapor Diffusion setup (Gryphoon, Art Robins Inc.). Multiple screens were used to this end, namely JCSG plus, PACT Premier, Morpheus, PGA, Midas (Molecular Dimensions). All trials failed, meaning no protein crystals were obtained and precipitate was observed in a vast majority of crystallization drops. MtCRE1' sample (4.5 mg/ml) with methylated lysine residues was tested in crystallization experiments with the use of the same set of screens. Similarly to the case of the unmodified sample, in nearly all conditions the protein molecules formed heavy, brown precipitate and no protein crystal appeared. The three promising samples of MtCRE1' after treatment with Subtilisin, Thermolysin and Proteinase K were used in crystallization experiments with the previous set of screens. Crystals appeared in the sample treated with thermolysin after one week in three conditions from Morpheus screen that contained Buffer Systems 1, 2 and 3 (pH 6.5, 7.5 and 8.5, respectively); 0.06 M mixture of divalent cations and 37.5% of mixture of MPD, PEG 1k and PEG3350. Manual optimization using hanging drops showed that the ultimate conditions are Buffer System 3 (pH 8.5) and the latter crystallization cocktail. The crystals diffracted to approx. 4 Å resolution and further optimization is required.

Medicago truncatula Nodulin 13

Cloning

The coding sequence for MtN13 protein was found in *M. truncatula* transcriptome (Moskal *et al.*, 2008) by Protein BLAST (Altschul *et al.*, 1997). The MtN13 sequence has "Evidence at protein level" annotation in the UniProt Knowledgebase and is accessible under the code P93330. The MtN13 DNA coding sequence was amplified by polymerase chain reaction (PCR) using *M. truncatula* (ecotype J5) cDNA from root nodules as template (Forward CACCATGGGTGTTATCACTTCAGAAAGC and Reverse TTAGTTGCTGTCTTTGTTGT-AATCAG PCR primers were used). The reaction product was cloned into the pET-TOPO-151D vector (Invitrogen) and the correctness of the insert was confirmed by DNA sequencing.

Overexpression

Overexpression was carried out in BL21 Magic *E. coli* cells. The bacteria were cultured with shaking at 210 rpm in LB media supplemented with 120 μg/ml ampicillin and 25 μg/ml kanamycin at 37°C until the OD_{600} reached 1.0. The temperature was lowered to 18°C and IPTG was added to a final concentration of 0.5 mM. The culture was grown for 18 h and then centrifuged at 4500 rpm for 10 min at 4°C. Cell pellet from 1 l culture was resuspended in 30 ml of binding buffer (50 mM Tris-HCl pH 8.0; 500 mM NaCl; 20 mM imidazole; 1 mM TCEP), flash-frozen in liquid nitrogen and stored at -80°C.

Purification

The samples were thawed and the cells were disrupted by sonication using 3 s bursts of total duration of 4 min, with appropriate intervals for cooling. Cell debris was removed by centrifugation at 15 000 rpm for 30 min at 4°C. The supernatant was applied to a column packed with 6 ml of HisTrap HP resin (GE Healthcare), connected to VacMan (Promega) and the chromatographic process was accelerated with a vacuum pump. After binding, the column was washed four times with 30 ml of the binding buffer and the purified protein was eluted with 15 ml of elution buffer (50 mM Tris-HCl pH 8.0; 500 mM NaCl; 300 mM imidazole; 1 mM TCEP). The His_6-tag was cleaved with TEV protease and the excess of imidazole was removed by dialysis (4 h at 4°C) simultaneously. The solution was mixed with HisTrap HP resin to get rid of the His_6-tag and the His_6-tagged TEV protease. The flow-through was collected, concentrated to 4 ml and applied on a HiLoad Superdex 200 16/60 column (GE Healthcare) equilibrated with a buffer composed of 50 mM Tris-HCl pH 8.0, 200 mM NaCl

and 1 mM TCEP. The size exclusion chromatography step yielded a homogenous protein fraction of about 6 ml. The sample was concentrated to 7 mg/ml as determined by the method of Bradford (1976) and used immediately for crystallization experiments.

Crystallization

Because of differences in solubility, the ligands were added to the protein solution using two different protocols. In both cases the protein-ligand mixture was incubated overnight prior to crystallization. *Trans*-zeatin hydrochloride (20 mg/ml, Sigma-Aldrich Z2753) was added as aqueous solution at three-fold molar excess. The three remaining hormones (2iP, KIN and BAP) were added in pulverized form in submilligram quantities to 300 µl of protein solution. It is of note that addition of these ligands as solutions in organic solvents or hydrochlorides invariably led to either precipitation of the protein or to a failure of the subsequent crystallization experiments. Also, the ligands had to be added immediately after the final size exclusion chromatography step, or the protein would not crystalize. Following an overnight incubation, the protein-ligand solutions were centrifuged at 14 000 rpm for 5 min at room temperature to get rid of undissolved ligands. Initial screening for crystallization conditions was performed using a Robotic Sitting Drop Vapor Diffusion setup (Mosquito) with Crystal Screen HT, Index HT and Salt Rx reagents from Hampton Research. All four complexes crystallized from sodium malonate conditions at pH 7.0. Manual optimization of the initial crystallization hits in hanging drops gave the following malonate concentrations: 1.9 (complex with ZEA), 1.5 (2iP), 1.85 (KIN) and 1.7 M (BAP). The hanging drops were mixed from 2 µl of reservoir solution and 2 µl of protein-ligand solution. Hexagonal bipyramidal crystals (Fig. 23) grew to typical dimensions (in mm) of approx. 0.50 x 0.40 x 0.40 (ZEA complex), 0.40 x 0.30 x 0.30 (2iP), 0.25 x 0.20 x 0.20 (KIN) and 0.15 x 0.10 x 0.10 (BAP) within a week at 19°C. The crystals were transferred to reservoir solutions with elevated concentration of sodium malonate for controlled dehydration. Dehydration was carried out stepwise (0.1 M per day) and the final sodium malonate concentration was 0.3 M higher in each case, compared to the crystallization conditions. After one month, the crystals were harvested with 0.1 mm nylon loops (Hampton Research) and vitrified in liquid nitrogen for synchrotron-radiation data collection. There was no need for additional cryo-solutions since sodium malonate at concentrations above 1.8 M is expected to provide sufficient cryoprotection (Bujacz *et al.*, 2010).

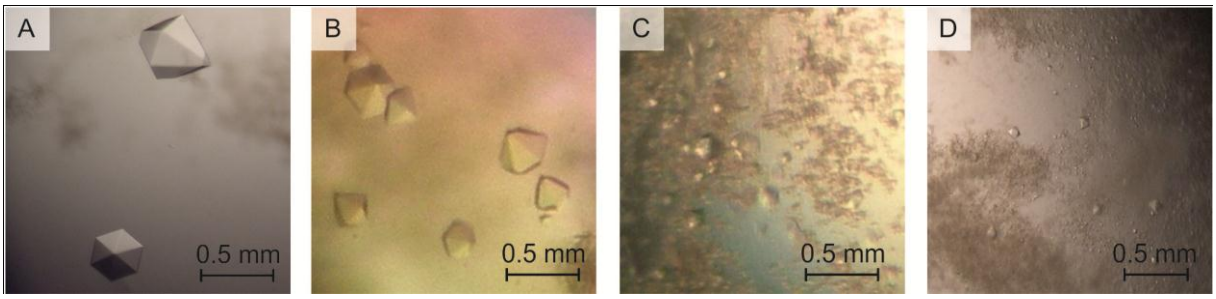


Figure 23
Crystals of MtN13 protein with different cytokinin ligands, namely **A**, ZEA; **B**, 2iP; **C**, KIN; **D**, BAP.

Diffraction data collection and processing

X-Ray diffraction data were collected at 100 K at four different synchrotron facilities worldwide. The diffraction data for the MtN13/ZEA complex were processed using the *HKL2000* package (Otwinowski & Minor, 1997) while *XDS* (Kabsch, 2010) was used for the three remaining cases (2iP, KIN and BAP complexes). Details and statistics of data collection and processing are summarized in Table 4.

Solution and refinement of the crystal structures

The crystal structure of MtN13 in complex with ZEA was solved by molecular replacement using *PHASER* (McCoy *et al.*, 2007). To this end, the crystal structure of a PR-10 protein from *Lupinus luteus* (Fernandes *et al.*, 2008) was used as a search probe (PDB code 2qim). Automatic model building was carried out with the on-line version of *ARP/wARP* (Langer *et al.*, 2008). The three remaining crystal structures (complexes with 2iP, KIN and BAP) are isomorphous, thus, the protein chain from the MtN13/ZEA complex was used as the initial model for refinement of these complexes. *COOT* (Emsley *et al.*, 2010) was used for manual fitting in electron density maps between rounds of model refinement in *phenix.refine* (Adams *et al.*, 2010). Hydrogen atoms at riding positions for the protein chain were included in the complexes with ZEA, 2iP and KIN for refinement. For these three models, TLS parameters (Winn *et al.*, 2001) were refined for six, five and six groups, respectively, as suggested by the refinement program. The models were validated with *MolProbity* (Chen *et al.*, 2010). The refinement statistics are listed in Table 4.

Experiments in solution

Dynamic light scattering (DLS) experiments were performed with a Zetasizer μV (Malvern Instruments) on an MtN13 sample at 1 mg/ml concentration. Small Angle X-ray Scattering (SAXS) data from protein solutions at 5 mg/ml concentration were collected at the Bio-SAXS

beamline BM29 of the storage ring at the European Synchrotron Radiation Facility (ESRF, Grenoble, France). All SAXS measurements were performed at 288 K for the scattering vectors: $0.05<s<5.0$ nm^{-1} ($s=4\pi\sin\theta/\lambda$; θ is the scattering angle and λ is the X-ray wavelength). Integration, scaling, and buffer subtraction were done using *PRIMUS* (Petoukhov *et al.*, 2012). *CRYSOL* (Svergun *et al.*, 1995) was applied for the evaluation of the solution scattering patterns using the present MtN13 crystal structure in monomeric and dimeric form. *Ab initio* modeling from the scattering curves was performed in *GASBOR* (Svergun *et al.*, 2001) with two-fold symmetry restraints.

Table 4. Data collection and refinement of MtN13 complexes

MtN13 complex with:	ZEA	2iP	KIN	BAP
Data collection				
Radiation source	APS, ANL Argonne	PETRA III, DESY Hamburg	BESSY, Berlin	MAX-lab, Lund
Beamline	22-BM	P14	BL14.1	I911-2
Wavelength (Å)	1.0000	1.22343	0.91801	1.04172
Temperature (K)	100	100	100	100
Space group	$P6_222$	$P6_222$	$P6_222$	$P6_222$
Unit cell parameters *a*, *c* (Å)	96.1, 113.2	96.6,112.0	96.4, 113.4	96.0, 113.4
Oscillation range (°)	0.5	1	0.5	1
No. of images	180	120	100	90
Resolution (Å)	30.0-1.85 (1.92-1.85)[a]	48.3-2.04 (2.16-2.04)	48.2-2.20 (2.33-2.2)	46.9-2.60 (2.76-2.60)
Reflections collected/unique	281527 / 26831	287179 / 20324	94999 / 16292	105014 / 10018
Completeness (%)	99.8 (99.4)	99.7 (98.2)	99.0 (94.4)	99.8 (99.6)
Multiplicity	10.5 (8.9)	14.1 (14.0)	5.8 (4.9)	10.5 (10.7)
R_{merge} [a](%)	7.5 (98.1)	9.9 (70.6)	9.4 (64.7)	12.8 (68.4)
$\langle I/\sigma(I)\rangle$	9.5 (2.1)	24.3 (4.0)	14.5 (2.0)	17.0 (4.3)
Refinement				
Unique reflections (work+test)	26803	20321	16286	10017
Test reflections	1349	1004	1002	802
No. of atoms (non-H)				
protein	1305	1303	1290	1274
cytokinin / malonate	16 / 7	15 / 7	16 / 7	17
metal (Na)	3	6	1	1
solvent	121	146	94	25
R_{work}/R_{free} (%)	18.2/21.6	16.5/19.8	18.6/21.5	17.0/22.5
RMSD from ideal geometry				
bond lengths (Å)	0.019	0.019	0.020	0.015
bond angles (°)	1.6	1.4	1.7	1.6
Ramachandran statistics (%)				
favored	96.4	96.4	96.9	98.1
outliers	0.6	0.0	0.6	0.0
PDB code	4jhg	4gy9	4jhh	4jhi

[a]Values in parentheses correspond to the highest resolution shell.

Results and discussion

Medicago truncatula Histidine-containing Phosphotransfer protein 1

Overall features of MtHPt1 crystal structure

MtHPt1 protein crystallizes in the orthorhombic $P2_12_12_1$ space group and the asymmetric unit contains one protein molecule. It is a monomer in solution, as confirmed by size exclusion chromatography. PDBe PISA web server (Krissinel & Henrick, 2007) also indicates no quaternary assemblies of MtHPt1. The complete coding sequence of MtHPt1 (including the residues missing in the coordinates) is 153 amino acid residues long and its molecular weight is 17.8 kDa. The Matthews coefficient (V_m) is 2.1 $Å^3$/Da, meaning that the solvent content is 41.3%.

The SNA- cloning artifact from pMCSG9 vector, N-terminal methionine and four residues at the C terminus (-MELN) are missing in the electron density maps and are not present in the final model. The rest of the protein chain has excellent definition in electron density maps. The entire Glu2-Met149 main chain of the native protein and the majority of side chains were traced unambiguously. Eleven side chains were modeled in double conformation. The refined model includes also 159 water molecules and no metal cation.

The overall fold of MtHPt1 consists of six α-helices (Fig. 24). The lengths of the helices vary from 8 (α2) to 34 residues (α6). Four of the helices (α3, α4, α5 and α6) form an antiparallel coiled-coil four-helix bundle at the C terminus (Fig. 24C). In general, 78% (116 out of 148) of amino acid residues that are present in the coordinate set, have α-helical conformation. Helix α3 is slightly kinked and leans away from the protein core which distorts the characteristic 3.6_{13} α-helical hydrogen-bonding pattern.

There are five n+3 (β) turns between residues 21-24, 35-38, 36-39, 38-41 and 66-69. The three overlapping turns map to the longest disordered region, within loop 33-41. Additionally, two n+2 (γ) turns in the structure occur between residues 69-71 and 89-91.

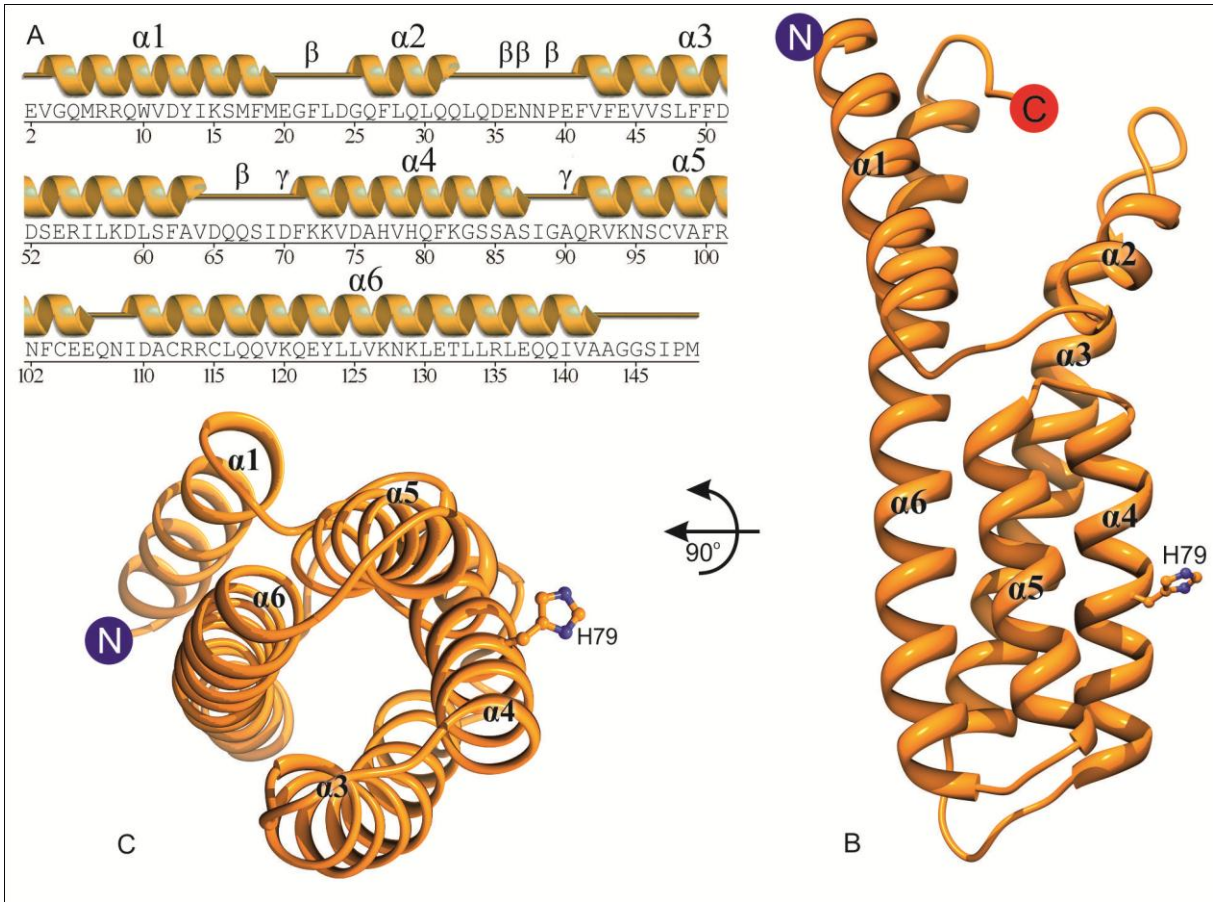


Figure 24
Overall structure of MtHPt1. **A** Secondary structure elements (helices α1- α6) corresponding to the amino acid sequence with denotation of five β and two γ turns. **B** A ribbon diagram illustrating the six α-helices and the active residue (phosphorylation site) His79. Helices α3-α6 form a four-helix bundle. **C** View down the axis of the four-helix bundle showing the His79 side chain exposed to the solvent.

Sulfur-aromatic interactions

MtHPt1 has a very high content (6.5%) of sulfur-containing residues (six Met and four Cys). However, none of the cysteines create a disulfide bond. Despite the side chains of Cys104 and Cys112 point toward each other, the distance between their S atoms (>4 Å) disqualifies from forming a disulfide bond. Many of the S atoms, including the former two, interact with aromatic residues (Reid *et al.*, 1985). Sulphur-aromatic interactions are strongest when the distance between the S atom and the centroid of the aromatic ring is less than 4 Å with the sulfur atom lying on the normal to the aromatic ring plane (Ringer *et al.*, 2007). It has been reported that these interactions can stabilize ligand binding and in this way participate in the enzymatic mechanism, as observed e.g. in NodS N-methyltransferase from *B. japonicum* in complex with S-adenosyl-L-homocysteine (SAH) (Cakici *et al.*, 2010). The situation is different in the core region of proteins, where S-centroid line is usually inclined to the aromatic ring at an angle of less than 60° and a distance of up to 6 Å (Reid *et al.*, 1985; Ringer *et al.*, 2007). This is also the case of MtHPt1, where an extended network stabilizes

the protein structure. There is an interaction between Met17 and Tyr13 within helix α1 (Fig. 25A). Six S-aromatic interactions are found in the four-helix bundle (Fig. 25B), clearly contributing to the compactness of the bundle. Residues Phe71, Phe100, Phe103, Cys104, Cys112 and Cys115 create a network of S-aromatic interactions that stabilize one end of the four-helix bundle. Phe100, Cys104 and Cys115 in this network are modeled with double conformations but for both an S-aromatic interaction is formed (Fig. 25B). Additionally, the interaction between Phe81 and Cys97 links the helices α4 and α5 together.

Some of the residues that create the S-aromatic interactions in MtHPt1 are highly conserved among homologous sequences from other plants (Fig. 26). More precisely, the amino acids at positions corresponding to Phe71, Cys97, Phe100, Cys104, Cys112 and Cys115 in MtHPt1 are nearly universally conserved in the plant orthologs. The fact that these residues are highly conserved in plant kingdom suggests their essential role in the protein structure and function.

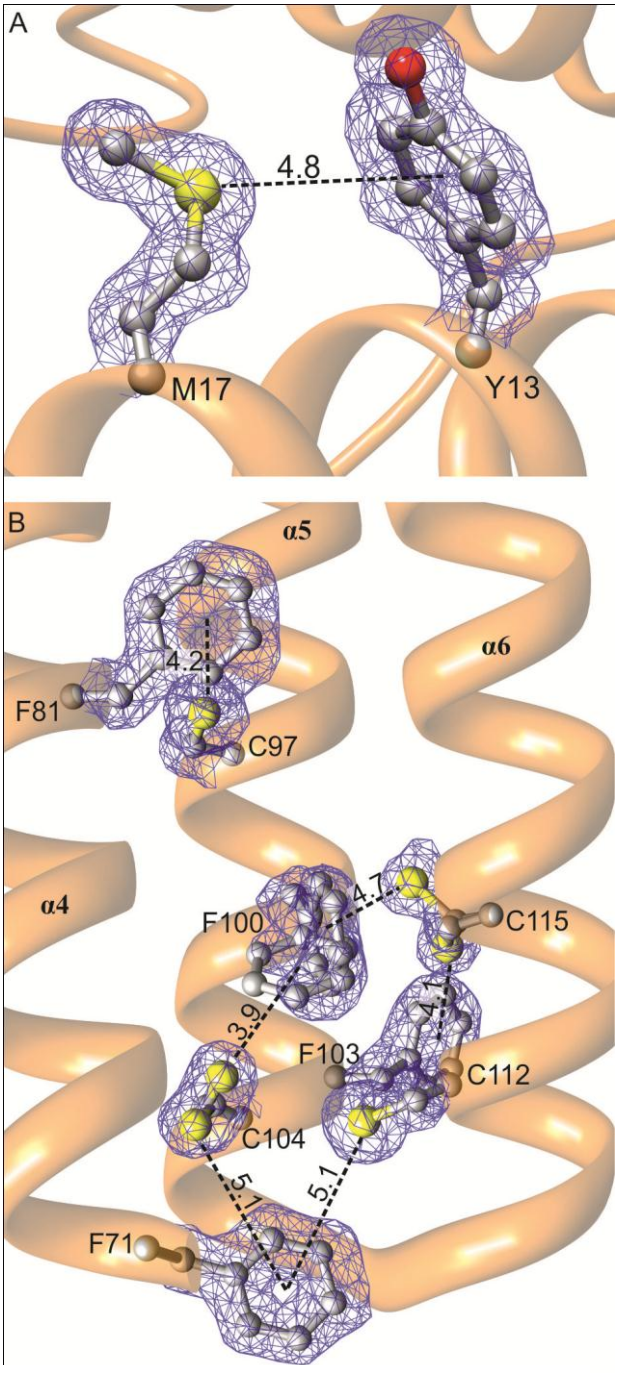
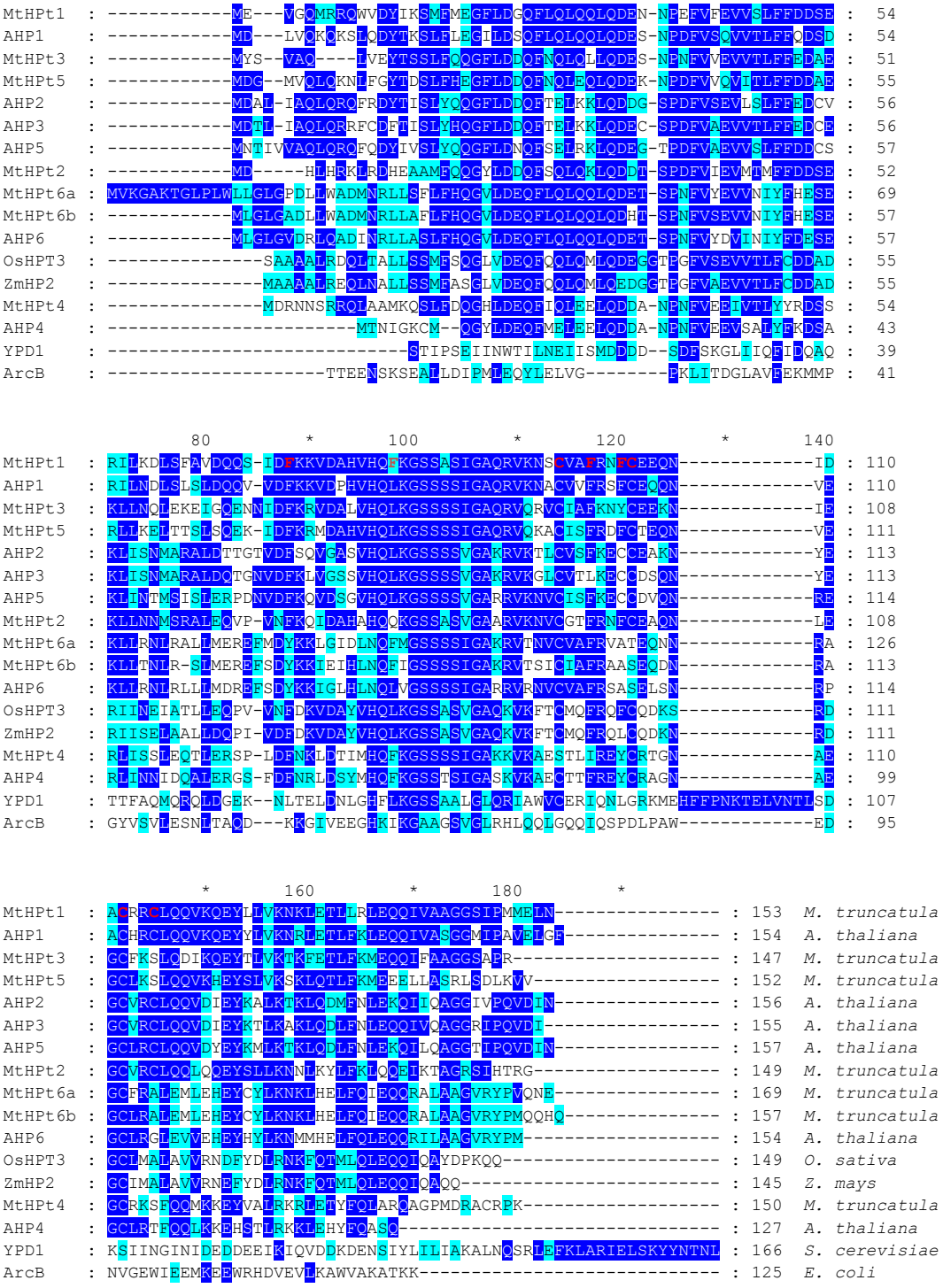


Figure 25
Sulfur-aromatic interactions in MtHPt1. The interactions are depicted as dashed lines and the S-centroid distances given in Å. The $2F_o$-F_c electron density map is contoured at 1σ.



1

Figure 26
Alignment of known amino acid sequences of HPt proteins from plants (*M. truncatula*, *A. thaliana*, *O. sativa*, *Z. mays*), yeast (*S. cerevisiae*) and bacteria (*E. coli*). The level of conservation is expressed by the darkness of the lettering background. Residues forming sulfur-aromatic interactions in MtHPt1 are highlighted in bold red font. Source organisms are appended to the sequences. The following UniProt sequences have been used: MtHPt1 [B7FGU6], MtHPt2 [G7I2T8], MtHPt3 [A2Q3H8], MtHPt4 [G7IHZ7], MtHPt5 [G7I7G5], MtHPt6a [G7JGY0], MtHPt6b [G7II71], AHP1 [Q9ZNV9], AHP2 [Q9ZNV8], AHP3 [Q9SAZ5], AHP4 [Q9LU15], AHP5 [Q8L9T7], AHP6 [Q9SSC9], OsHPT3 [Q6VAK4], ZmHPT2 [Q9SLX1], YPD1 [Q07688], ArcB [C6EH17].

Comparison of HPt proteins

Phosphorylation cascades with the use of histidine kinases and HPt proteins themself are found in bacteria, fungi and plants. A comparison of amino acid sequences from these kingdoms is shown in a full sequence alignment (Fig. 26) and phylogram (Fig. 27). In this work, the numbering of helices and residues corresponds to MtHPt1. It is also worth to divide the HPt proteins into enzymatically active and inactive orthologs. The active orthologs are able to carry out the phosphotransfer due to a conserved histidine residue (the active site) in a position corresponding to His79 in MtHPt1. The active residue of HPt proteins has been initially identified in maize (Sakakibara *et al.*, 1999). In the present studies, His79 in MtHPt1 have been additionally verified as the phosphorylation site by site-directed mutagenesis (H79N, *vide supra*). The inactive HPt orthologs have Asn instead of the active His residue and act as inhibitors of the cytokinin receptor (Mahonen *et al.*, 2006). Since the inactive orthologs are unable to transfer the phosphate group, they block the cytokinin signaling pathway. In *M. truncatula* there are two proteins that potentially function in this inhibitory manner. In the phylogram and the alignment they are termed MtHPt6a and MtHPt6b.

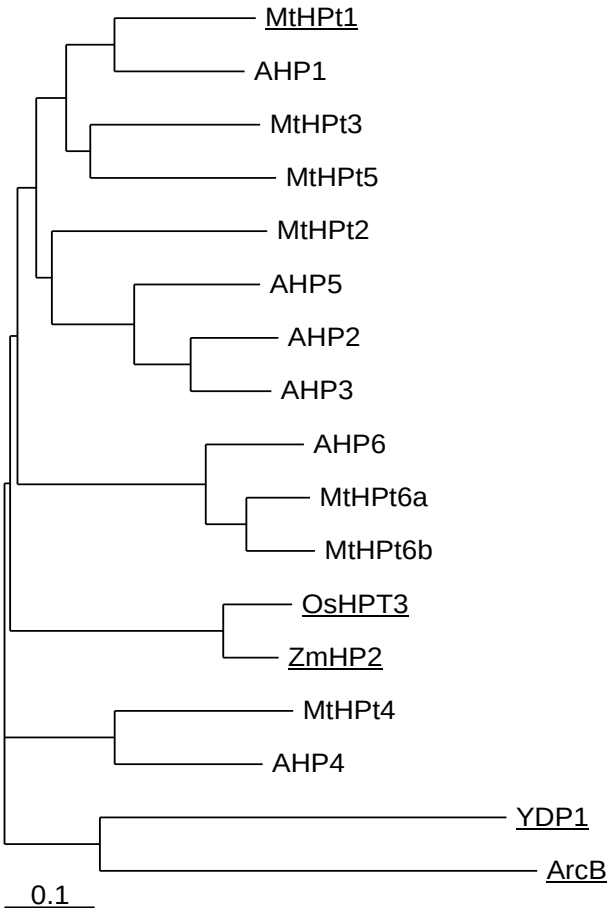


Figure 27
A phylogram of amino acid sequences of HPt proteins. The same sequences have been used as in Fig. 26. Proteins with determined crystal structures are underlined. The scale bar at the bottom indicates the number of amino acid substitution per site.

Homologues from different kingdoms of both the active and inactive HPt proteins share low sequence similarity in the neighborhood of the active site. However, when only the active orthologs from plants are concerned, the conservation level is much higher. The sequence motif is in this case is $_{79}$**H**Q[LFQ]KGSS[SAT]S[IV]GAx[RK]V$_{93}$, with the active histidine highlighted in bold.

There are many structural differences between prokaryotic, fungal and plant HPt proteins. Fig. 28 depicts a superposition of: (i) the HPt domain of the *E. coli* hybrid sensor ArcB, (ii) *S. cerevisiae* phosphorelay intermediate protein 1 (YPD1), and (iii) the plant MtHPt1 structure from this work. The shown HPt proteins from bacteria and fungi lack the α1 helix at the N terminus. The α2 helix in ArcB lies farther away from the protein core when compared to eukaryotic homologues. The helices α3-α5, on the other hand, superpose very well, and the active histidine in α4 occupies approximately the same position. Other differences are observed as approaching the C terminus. In fungal YPD1, there is a long and mostly disordered insert between helices α5 and α6. Helix α6 is the longest helix in the MtHPt1 structure and follows the same general direction as in YPD1. This C-terminal helix is in ArcB, kinked towards the long axis of the molecule and occupies part of the space used by α1 in MtHPt1 structure.

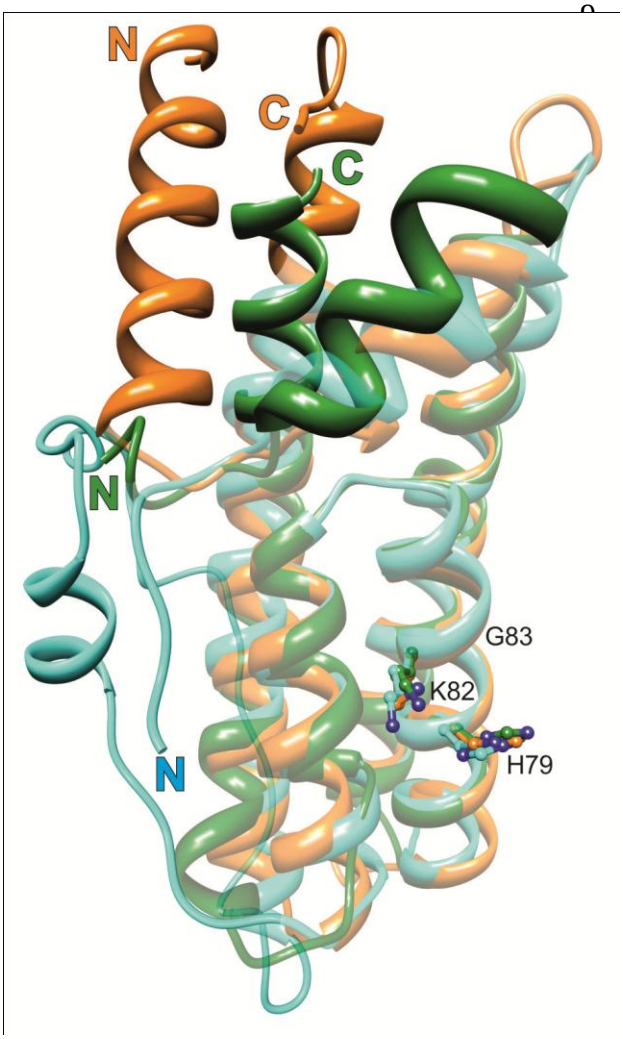


Table 5 shows RMSDs and Q-scores for the structures of eukaryotic HPt proteins with known three-dimensional structure. In plant HPt sequences, the conserved residues are located predominantly in the helices forming the four-helix bundle and, particularly, surrounding the active His79. Conservation of these amino acids suggests that they might take part in binding an important partner protein, either the REC

Figure 28
Structural alignment of HPt orthologs from three different kingdoms. MtHPt1 (plants), orange (this work); YDP1 (fungi), cyan (S. cerevisiae), and ArcB (bacteria), green (E. coli). The latter two proteins have been extracted from their complexes (PDB IDs 1oxk and 1bdj, respectively). The well-superposed elements are semi-transparent while the divergent elements are in solid color. Note the absence of helix α1 in the lower organisms and a long insert between helices α5 and α6 in YDP1. The active histidine (ball-and-stick model) occupies nearly the same position in all proteins. The lysine and glycine residues from the KGSS sequence motif that is conserved in all kingdoms are also shown to indicate the structural location of this motif in HPt proteins. In general, the structural divergence increases with the distance from the active site. Numbering of the residues corresponds to MtHPt1.

domain of the sensory kinase or one of the RRs. That has been actually shown in by Zhao and coworkers (2008). Authors studied a complex of *S. cerevisiae* osmosensing histidine protein kinase RR domain (SLN1-R1) and YPD1. They observed the phosphate mimic, BeF_3^-, next to this His residue and many of the residues conserved in the sequence motif (*vide supra*) involved in complex formation. The active His79, together with Lys82 and Gly83, are conserved in all active HPt orthologs. This histidine and lysine form the positive charge patch that surrounds the latter residue (see next section). Two consequent serines from the HXXKGSS motif of unknown function are present only in eukaryotic HPt proteins. The most important inference is that the structural similarities between HPt proteins are highest in the vicinity of the active histidine.

Table 5. Structural comparison of MtHPt1 with other eukaryotic HPt proteins.

PDB[a]	RMSD of Cα atoms (Å)	Number of aligned residues	Sequence identity (%)	Number of secondary structure elements	Q-score[b]
1oxk	1.28	108	19	5	0.43
1yvi	1.28	136	48	6	0.78
1wn0	1.29	136	50	6	0.77

[a]The PDB IDs correspond to the following HPt structures: 1oxk, *Saccharomyces cerevisiae*; 1yvi, *Oryza sativa*; 1wn0, *Zea mays*

[b]Q-score represents the quality function of Cα alignment. It reduces the effect of RMSD - N_{algn} (number of aligned residues) balance on the estimation of alignments (N_{res1} and N_{res2} stand for the number of residues in the aligned proteins, and the empirical parameter R_0 is set to 3 Å):
$Q = (N_{algn} \cdot N_{algn})/[(1+(RMSD/R_0)^2) \cdot N_{res1} \cdot N_{res2}]$

His79: the active site

The proper orientation of the imidazole ring of the active His79 in MtHPt1 can be determined due to two hydrogen bonds with water molecules (Fig. 29A). The protein crystallized at pH 5.5, at which both nitrogen atoms of the His79 imidazole ring are protonated and the positive charge is delocalized over both N atoms. Therefore, these water molecules can only act as hydrogen-bond acceptors. There is a strong positive patch in the vicinity of the active His79, created by His79 itself and by Lys82 and Lys94 (Fig. 29B). These three residues are part of the conserved HPt sequence motif and are expected to take part in binding of a phosphorelay partner.

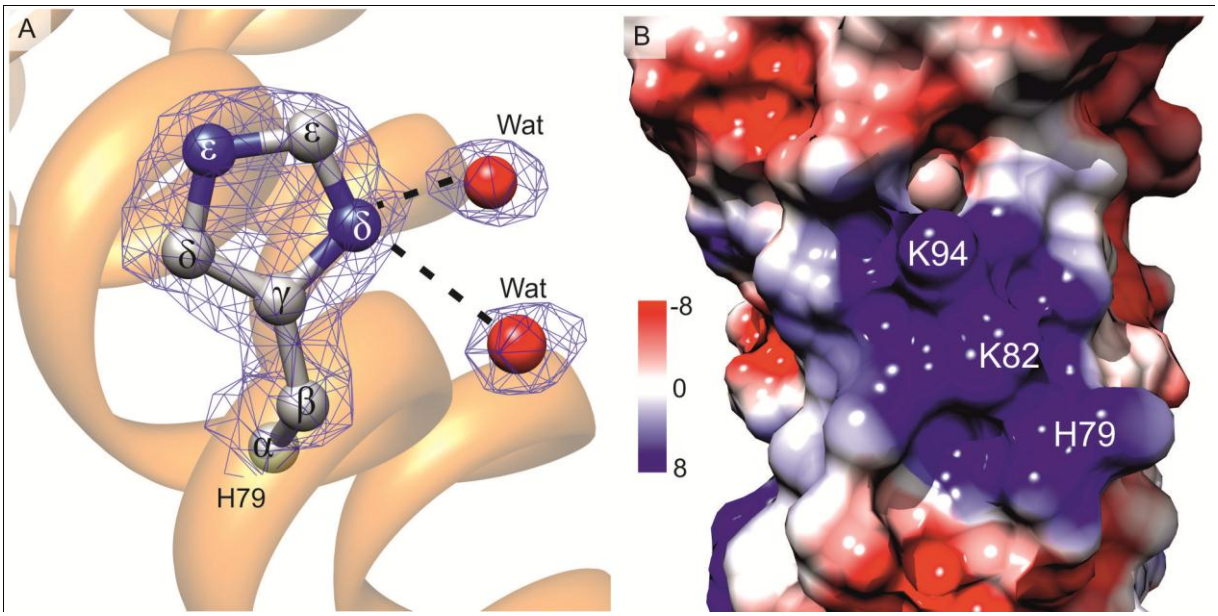


Figure 29
The active site His79. **A** Ball-and-stick representation of His79 with $2F_o$-F_c electron density map contoured at the 1.2σ level. Two water molecules form the only hydrogen bonds with the His79 side chain. **B** Electrostatic potential of the molecular surface of MtHPt1 in the vicinity of the active His79. The electrostatic potential is color-coded according to the scale in *k*T units shown on the left. Note the large patch of positive electrostatic charge created by Lys82, Lys94 and the active His79.

My attempts to crystallize MtHPt1 in the phosphorylated state have been unsuccessful, even though the phosphorylation itself was proven to take place. At acidic pH, at which MtHPt1 normally crystallizes, N-phosphorylated histidine is very labile (Quezada *et al.*, 2005; Zu *et al.*, 2007; Puttick *et al.*, 2008). It was possible to carefully increase the pH to 7.0, obtain diffracting crystals and solve structures, the electron density for those crystals exhibited absence of a phosphate group (not shown).

There are three crystal structures of proteins containing a phosphorylated histidine residue in the PDB (HIP, PDB symbol for a modified His residue) but all are unrelated to HPt proteins. Nonetheless, it is useful to examine the modification scheme of the histidine side chain. In all three cases, the N atom proximal to the Cα atom (Nδ) is the phosphorylation site. Two of the structures (PDB ID: 1nsq, 1nsp) were obtained from crystals that grown at pH 8.0 (Morera *et al.*, 1995). Unfortunately, no structure factors were deposited with those structures, therefore it is impossible to inspect the corresponding electron density maps. The third structure, Discoidin II from *D. discoideum* (PDB ID: 2vme) was determined at pH 7.5 (Aragao *et al.*, 2008). Although the difference (F_o-F_c) electron density show a negative peak in map at –4.6σ, implying only partial phosphorylation, there is no doubt about the phosphate group location.

Biochemical properties

The activity of MtHPt1 as a phosphate receptor was confirmed by *in vitro* phosphorylation assay. In this experiment, recombinant MtHPt1 was phosphorylated only when MtCRE1' was present (Fig. 30A,B). The intensity of the band on the autoradiography gel was the highest when the MtCRE1':MtHPt1 molar ratio was 1:50. To verify the specificity of the phosphorylation reaction, exclude signal from impurities and confirm His79 as the phosphorylation site, an H79N site-directed mutant of MtHPt1 was utilized in an analogous set of assays. H79N site-directed mutant was used per analogy to biologically inactive HPt variants (Mahonen *et al.*, 2006). No phosphorylation of the mutated protein was observed.

MST allowed to determine the dissociation constant (K_d) of the MtCRE1':MtHPt1 interaction to be 14 μM (Fig. 30C). This means that binding between MtHPt1 and the intracellular portion of the cytokinin receptor is quite strong even in the absence of phosphorylation, as ATP was not present in the reaction mixture. Substantial changes of MtCRE1' mobility upon binding of MtHPt1 make the thermophoresis method very reliable in this case.

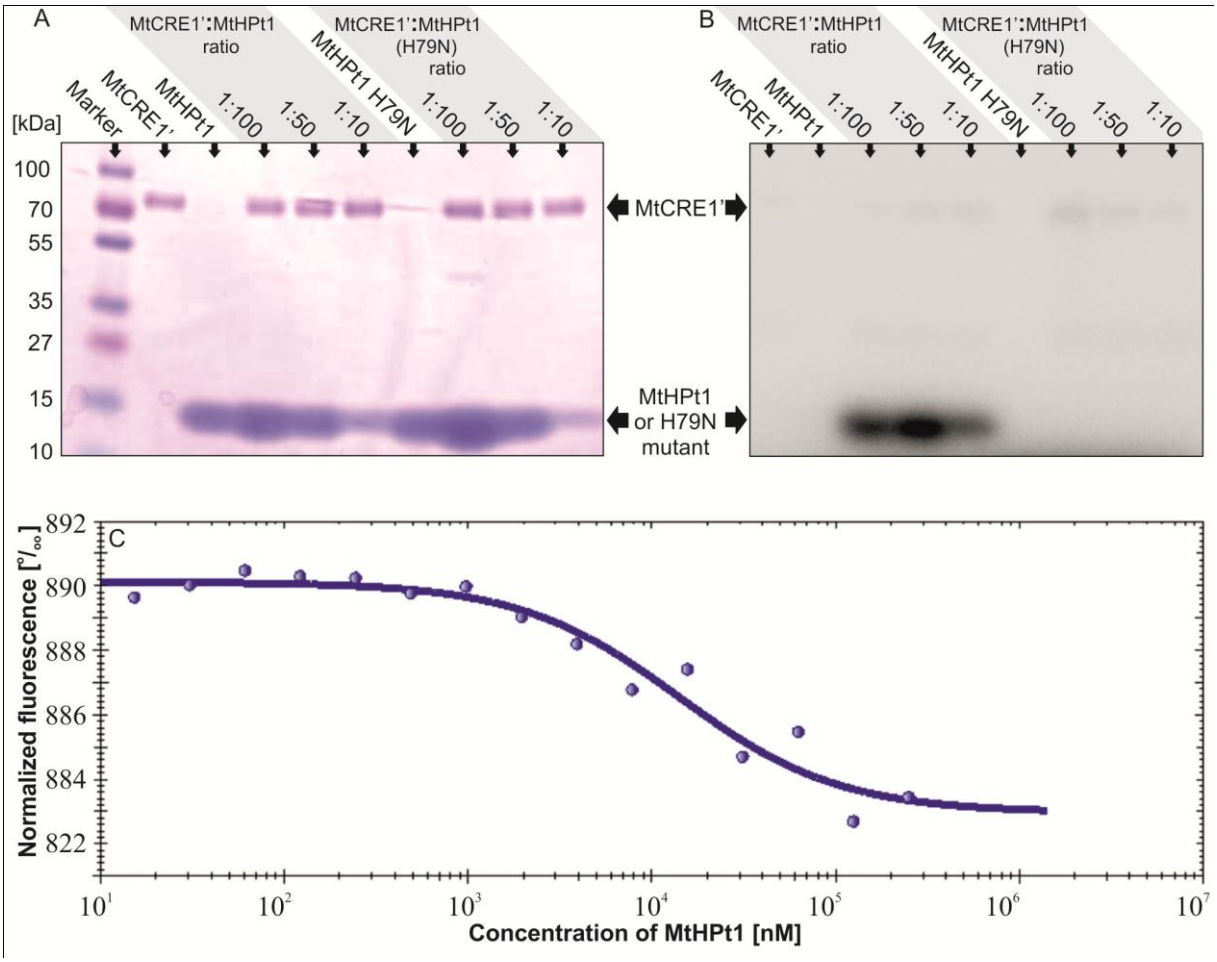


Figure 30
Biochemical properties of MtHPt1. **A, B** Autoradiography of enzymatic phosphorylation of MtHPt1 by MtCRE1' with γ^{32}P- labeled ATP as a phosphate donor. The H79N mutant of MtHPt1 was utilized as a negative control. Different ratios of the interacting proteins were used as indicated above the gels. The same gel was analyzed using autoradiography (B) and stained with Coomassie Brilliant Blue R-250 (A). **C** Results of Micro-Scale Thermophoresis. The experimental points and the fitted curve are shown. Fifteen measurements were taken at different MtCRE1':MtHPt1 ratios. The change in fluorescence is the result of complex formation. The migration rate of the labeled protein (MtCRE1') in a thermal gradient is slower upon binding to MtHPt1.

Medicago truncatula Histidine-containing Phosphotransfer protein 2

Overall structure of MtHPt2

According to UniProt, the sequence of MtHPt2 comprises of 149 amino acids (UniProtKB: G7I2T8) that correspond to molecular weight 17.2 kDa. In the crystal structure, three additional residues (SNA-), that precede the genuine Met1 at N-terminus, are visible in the electron density maps. This is a cloning artifact from the pMCSG48 vector. On the other hand, five residues at C-terminus (-IHTRG) were impossible to trace and are not present in the coordinate set. The coordinate set contains, therefore, 147 amino acid residues, 36 of which were modeled with alternate conformations. Most of these residues, due to subatomic data resolution, have excellent definition in the very high quality electron density maps. Except for the protein atoms, also 207 water molecules were modeled. No additional ion was observed in the MtHPt2 structure, analogously to MtHPt1.

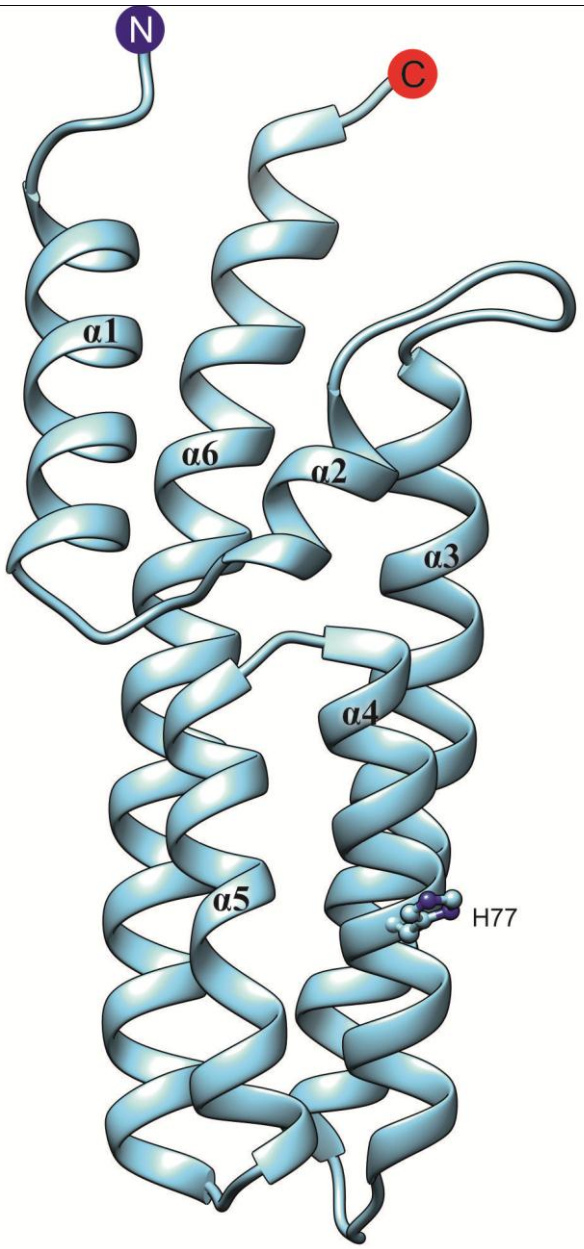


Figure 31
Overall structure of MtHPt2. The four C-terminal helices form a four-helix bundle. The active residue, His77 is shown in ball-and-stick representation. N and C termini are also denoted.

The coordinates of the refined crystal structure of MtHPt2 protein together with the structure factors were deposited in PDB (PDB ID: 4g78). It is definitely of note, that as of April 2013, MtHPt2 is the highest resolution structure of a plant protein deposited in PDB, with only one exception of crambin refined to 0.48 Å resolution (Schmidt *et al.*, 2011) (PDB ID: 3nir). Crambin itself is a very small storage protein of only 48 amino acids, meaning it is one third the size of MtHPt2.

The overall structure of MtHPt2 protein consists of six α-helices (Fig. 31) which lengths vary from 7 (α2) to 36 residues (α6). One the helices (α5) is has a distorted pattern of hydrogen bonds. More precisely, a water molecule mediates the framework of hydrogen bonds between the carbonyl O atom of Gly96 and backbone N atom of Asn100 (Fig. 32). As a result, the backbone N atoms of Phe98 and Arg99 do not form hydrogen bonds expected for α helices.

Four of the helices (α3, α4, α5 and α6) form an antiparallel coiled-coil structure, the four-helix bundle. There are four n+3 (β) turns in the MtHPt2 structure. They occur between residues (-1)-2, 19-22, 32-35, 64-67. One of the turns involves a residue from the cloning artifact (Asn-1). In addition, one inverse n+2 (γ) turn is formed between the residues 87-89. Interestingly, the turns map to the similar positions in the structures of both HPt proteins from *M. truncatula*.

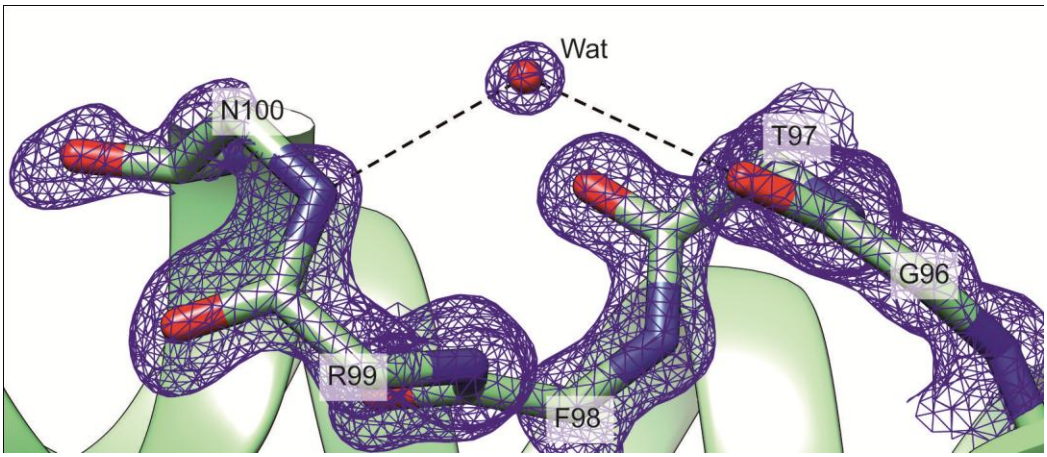


Figure 32
Distortion of the hydrogen-bonding-network canonical for α helices. A water molecule mediates the interaction between the carbonyl O of Thr97 and backbone N of Asn100. Side-chain atoms were omitted for clarity. $2F_o$-F_c electron density maps contour the protein model at 1.5 σ level.

Crystal packing

MtHPt2 protein crystallizes in the monoclinic $P2_1$ space group and the asymmetric unit contains one protein molecule. The molecules in the crystal are packed very tightly, as expressed by low Matthews coefficient (V_m) of 1.8 $Å^3$/Da. This means that the solvent content in the crystal is only 32%, which is nearly ten percent less than in the crystal of MtHPt1. A detailed insight into crystal packing of these two proteins is provided in Fig. 33. The most important conclusion from this analysis is that the void volumes, occupied by bulk solvent in the crystal lattice, are much smaller in case of MtHPt2. This means also, that the surface areas buried upon the MtHPt2 crystal formation are very large and intermolecular contacts are vast. On the other hand, very little of the surface is left that is exposed to the solvent channels in

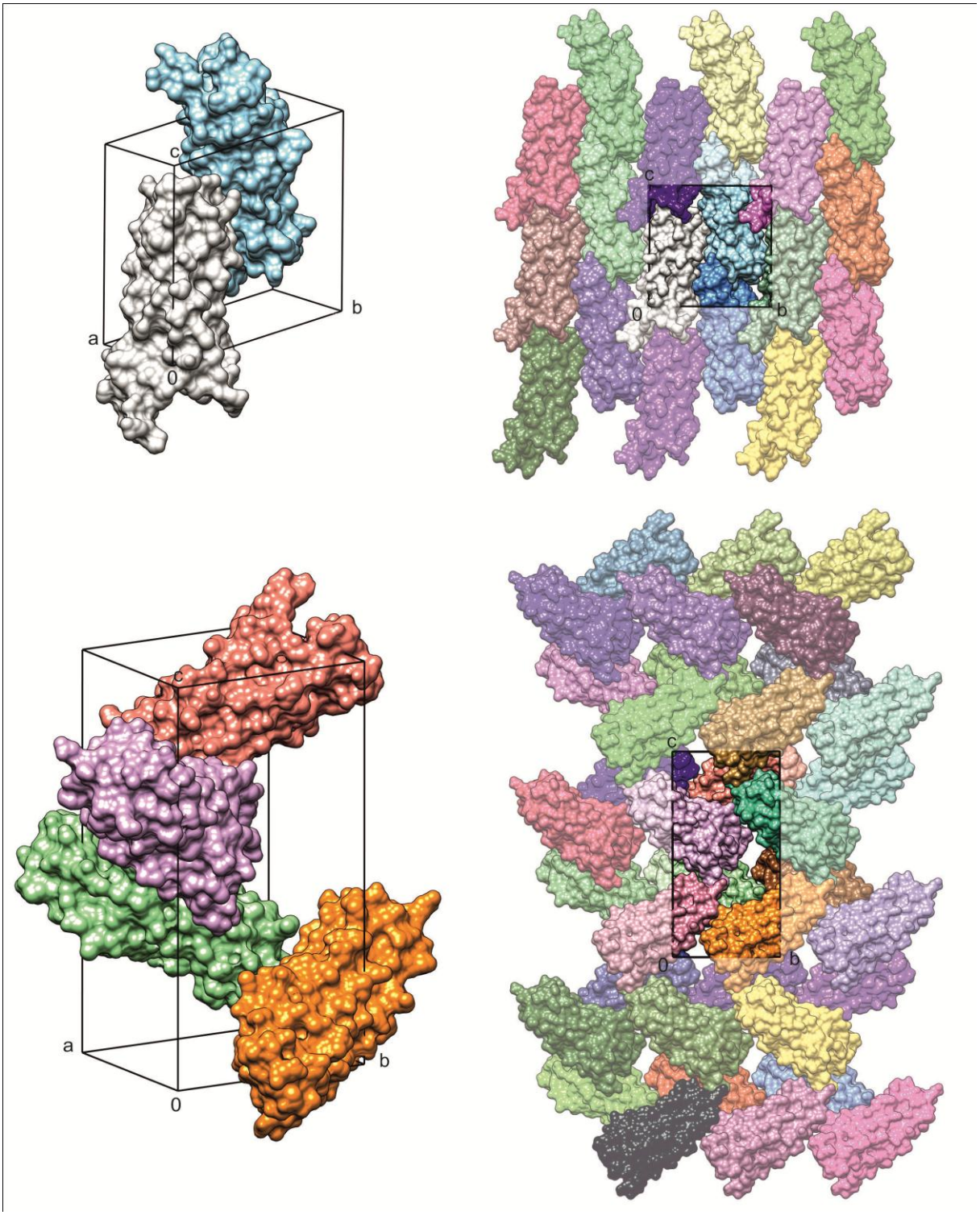


Figure 33
Comparison of crystal packing in MtHPt2 (top panel) and MtHPt1 (bottom panel). Single unit cells are shown on the left, while the arrangement of the molecules and view along the a axis is presented on the right. Protein molecules are represented by their surfaces.

the crystal lattice. This tight packing of protein molecules undoubtedly improves the rigidity of the crystal lattice. It is probably owed mainly (but not only) to this rigidity, that crystals of MtHPt2 diffract X-rays to higher angles and the obtained data resolution is by 0.5 Å better than in case of MtHPt1.

The active site of MtHPt2

By analogy to all the other HPt proteins, His77 is the active site of MtHPt2 protein as it occupies the same structural position than in all homologs. MtHPt1 is actually the best reference here, because for this particular protein I have validated the active site myself by site-directed mutagenesis.

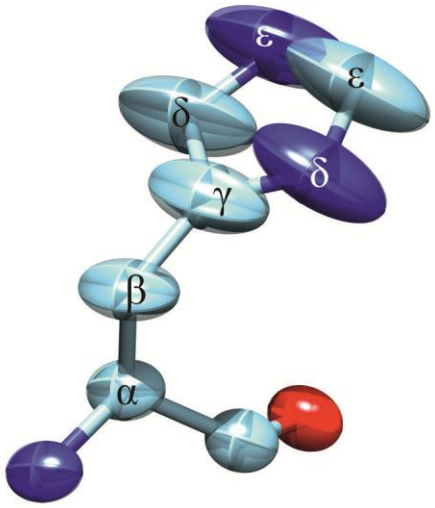


Figure 34
Thermal ellipsoids of atoms of His77 (the active residue). Note substantial differences in the volumes of the ellipsoids between the main- and side-chain atoms.

In the topology of MtHPt2 protein His77 belongs to the helix α4 within the four-helix bundle. His77 is exposed to the solvent when a single protein molecule is concerned. Also in the crystal lattice, His77 points towards the void occupied by bulk solvent. The side chain of His77 does not form any hydrogen bonds and its orientation (the flip) was determined solely by the comparison of the ADPs after refinement of both alternate conformations. Because, apart from the obvious covalent bonds formed by its main-chain atoms, His77 has no binding partners in the crystal structure, it is also very flexible. This flexibility is best visualized by thermal ellipsoids (Fig. 34) that can be shown thanks to the subatomic resolution of the crystal structure and anisotropic refinement. The thermal ellipsoids depict the large differences of atomic displacement between the main- and side-chain atoms of His77.

Future outlook

The subatomic resolution structure of MtHPt2 protein provided very detailed structural information about the molecule. It is of my next extreme interest to examine whether MtHPt2 can be also phosphorylated by the MtCRE1 as is MtHPt1. If so, the remaining question is why there would be two proteins of the same and specific function. Perhaps they have different affinities to the receptor kinase or/and can activate different response regulators. Unless

MtHPt2 is the phosphate carrier in a different transduction pathway e. g. the one initiated by ethylene. As mentioned in the introduction, ethylene receptor is also a histidine kinase (Kendrick & Chang, 2008) and needs a phosphate carrier to transduce the signal. Further studies are required to find and examine the binding partners of MtHPt2 that, at the very least, should include autoradiography with the use of histidine kinases of *M. truncatula*.

Medicago truncatula Nodulin 13

Remarks concerning the protein sample

It is of note that the entire purification procedure (from thawing the cell pellet to addition of a ligand) had to be finished within 10 h. Even with a 24-hr delay the protein sample exhibits a significant level of dimerization (Fig. 35). As explained below, the dimerized protein probably loses its ability to bind cytokinins.

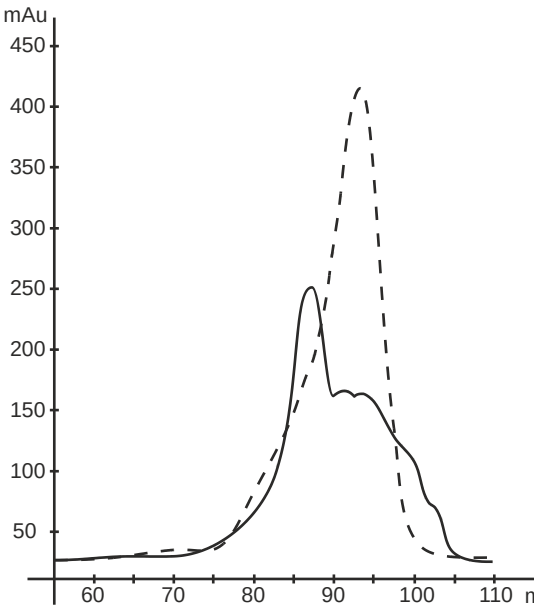


Figure 35
SEC of the MtN13 protein. The curves show the elution profiles of MtN13 samples applied on a column immediately after the second HisTrap chromatography (dashed line, monomers) or with a 24 h delay (continuous line, mostly dimers).

Attempts to crystallize MtN13 without a ligand or in complex with other plant hormones (such as synthetic cytokinins, auxins, gibberellic acid, abscisic acid, salicylic acid, jasmonic acid and methyl jasmonate) have failed (no crystals obtained).

Overall features of MtN13 complexes with cytokinins

The crystals of MtN13 complexes with cytokinins are isomorphous. All four structures were solved in the $P6_222$ space group with one protein molecule in the asymmetric unit, despite the complexes are actually dimeric (*vide supra*). The Matthews coefficient for all four crystal structures is about 4 $Å^3$/Da and corresponds to a high solvent content of ~70%. Obviously, the quality of electron density maps varies due to different data resolution (from 1.85 Å to 2.6 Å). Nevertheless, they were very clear and allowed the tracing of almost all the residues of the MtN13 protein sequence. The only exception is five residues at the C-terminus (NKDSN). On the other hand, five (IDPFT) or four (DPFT) additional residues, preceding the genuine MtN13 protein sequence, introduced as cloning artifacts (a linker between the TEV cleavage site and the MtN13 sequence), were modeled at the N terminus in the complexes with ZEA, 2iP and KIN, or in the complex with BAP, respectively. The cytokinin ligands were precisely

defined in difference electron density maps and could be modeled without any ambiguity (Fig. 36). Moreover, the refined ADPs for each cytokinin molecule are similar or lower than those for the protein side-chains of the surrounding residues.

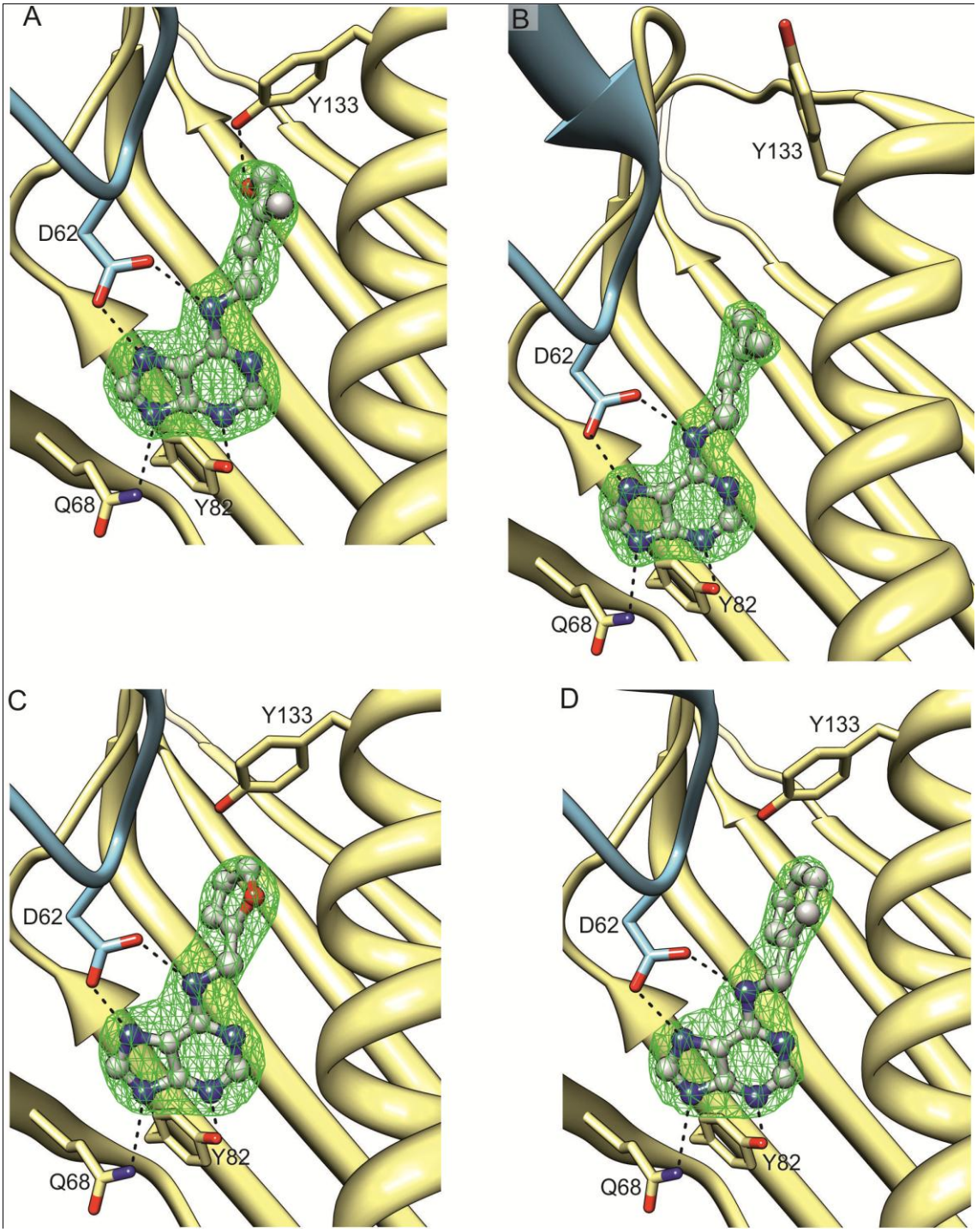


Figure 36
Cytokinin ligands bound in the MtN13 cavity. (A) ZEA; (B) 2iP; (C) KIN or (D) BAP, in ball-and-stick representation, are shown with OMIT F_o-F_c electron density maps contoured at 4.5σ level

MtN13 protein has the canonical PR-10 fold (Fig. 13), consisting of a seven-stranded antiparallel β-sheet wrapped around the C-terminal helix α3. β-Hairpins connect the β-strands, except for the β1-β2 crossover, which is formed by helices α1 and α2 and links the edges of the β-sheet. The β-sheet has highly curved baseball-glove shape owed to six β-bulges, where the odd-numbered loops (L3, L5, L7, L9) resemble the "fingers" of the glove. The two short helices (α1 and α2) create a V-shaped support for the C-terminal part of the long α3 helix. The α3 helix is connected to the β7 strand of the β-sheet by loop L9 and is slightly kinked toward the protein core.

All the cytokinin molecules studied in this project are found inside the hydrophobic cavity which is created between the β-sheet and the α-helices. In the MtN13 topology, the only prominent entrance to the cavity leads through a narrowing tunnel. The entrance is bordered by the loops L5 and L7 and the N-terminal part of helix α3. If unplugged, meaning that the second protein molecule from the dimer is not considered, the entrance has a diameter of about 8 Å.

Additional ions

The protein was crystalized using sodium malonate as the precipitant and the malonate ion was found and modeled next to Gly87 in complexes with ZEA, 2iP and KIN. Three, six, one and one sodium cations were modeled in the ZEA, 2iP, KIN and BAP complexes, respectively. The identification of metal cations was based on the octahedral coordination and metal...O distances that were shorter (~2.2 - 2.5 Å) than for typical hydrogen bonds. In addition, the presence of Na^+ cations was verified by the calcium bond-valence sum (CBVS) method (Muller *et al.*, 2003). In each structure, one Na^+ cation stabilizes the L5 loop and is in the same position. Two additional Na^+ cations, located in the complexes with ZEA and 2iP, occupy the same sites, next to Thr0 (preceding the genuine Met1) and Ser65. Subsequent three Na^+ cations were found in the 2iP complex only.

Dimerization of MtN13

PR-10 proteins are monomeric in solution with only a few exceptions that have been reported (Ma *et al.*, 2006; Scholl *et al.*, 2005). These exceptions, however, were neither examined from the structural point of view nor the functional importance of the oligomerization was specified. MtN13, on the other hand forms very tight dimers in all the presented crystal structures which is one of the two most exciting features of MtN13. The dimers have crystallographic two-fold symmetry generated by the x, x-y, 1/3-z operation. The molecular

surface area buried upon dimerization is very large, ~4000-5000 Å^2 per dimer. It is particularly impressive in comparison with the surface area of the entire dimer itself (~16000 Å^2).

Two experiments, namely DLS and SAXS, were performed to verify the dimerization observed in MtN13 crystal structures. The DLS measurements for both, the unliganded and liganded forms of MtN13 provided a particle mass of 37.6±11.8 kDa, which corresponds to a homodimer. It is of note that the exact agreement of the calculated formula weight ($2 \cdot 18.8 = 37.6$) is actually an idiosyncrasy in DLS experiments. The hydrodynamic radius R_H obtained from the DLS experiment is 25.4 Å was used in farther comparison with SAXS experiments.

The radius of gyration, R_g obtained from SAXS experimental curves of unliganded MtN13 (Fig. 37A) and extrapolated to zero concentration is 24.3 Å. Theoretical calculations based on the crystal structure of MtN13 (in complex with ZEA) give for the R_g values of 16.5 Å or 21.8 Å for the monomer and the homodimer, respectively. The theoretical R_g value estimated for the homodimer is in a very good agreement with the experimental data. The minor discrepancy (2.5 Å) can be owed to the hydration shell and differences in molecular compactness caused by crystal packing and/or by the presence of a bound ligand. The DLS and SAXS results were compared with the use of the relation $R_g = 0.8 \cdot R_H$ (Berry, 2010). The equation is obeyed in the case of the MtN13 protein in solution.

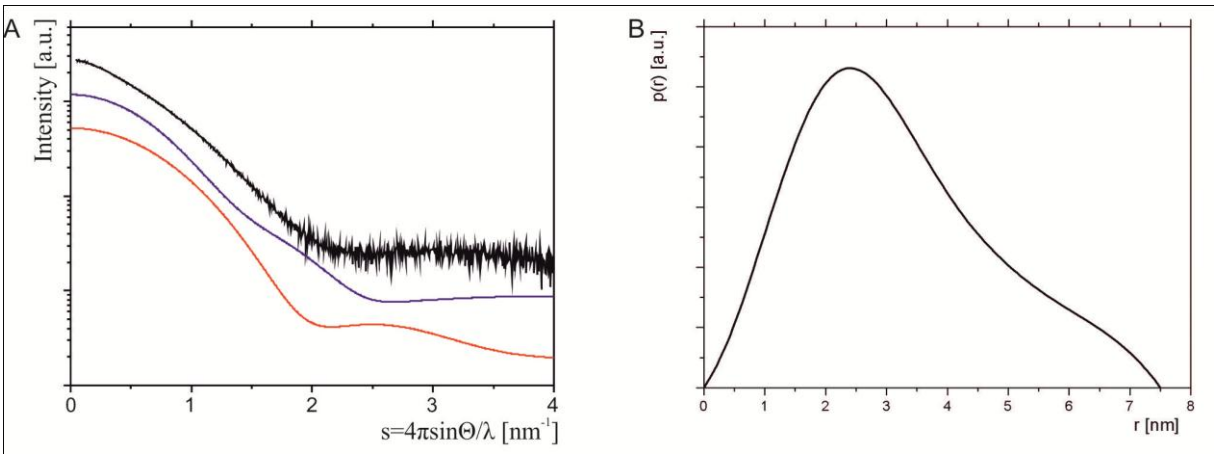


Figure 37
SAXS data. **A** Comparison of the experimental curve for MtN13 solution (black) with theoretical SAXS scattering curved calculated for the monomer (red) and dimer (blue). Crystallographic coordinates from the present study were used for the calculations. B The $p(r)$ function that suggests the overall shape of MtN13 protein is an elongated ellipsoid with the maximum dimension of 75 Å.

The *p*(*r*) function (Svergun & Koch, 2003) (Fig. 37B) suggests that in solution the protein is an elongated ellipsoid with a maximum diameter of 75 Å. The SAXS experiments served also for simulation of the shape of the MtN13 protein in solution, the so-called molecular envelope. A comparison between the homodimeric crystal structure and the molecular envelope, calculated *ab initio* from the SAXS scattering curves, is presented in Fig. 38A,B. The overall agreement between the crystallographic homodimer and the molecular envelope is very satisfactory. The more obtuse angle between the two halves of the SAXS model (in comparison to the crystallographic model) can be attributed to crystal packing forces and/or to the effect of ligand binding. A detailed insight into the MtN13 homodimers with annotation of the secondary structure elements and location of the ligand molecules is provided in fig. 38C.

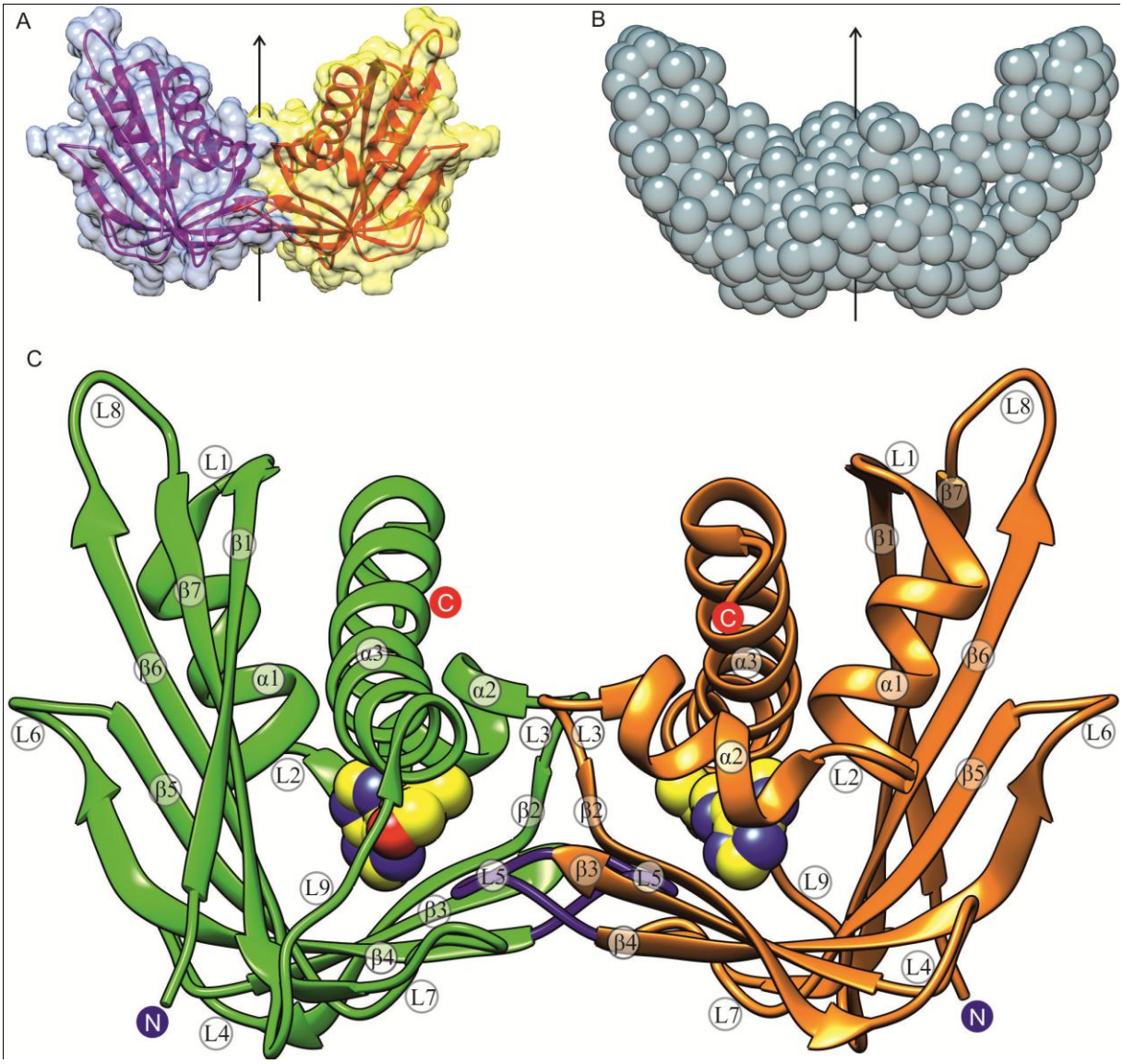


Figure 38
Dimerization of MtN13. **A** The crystallographic twofold-symmetric dimer of MtN13 with semitransparent surface. **B** A model of MtN13 in solution generated by ab initio calculations using SAXS data. Twofold axes are indicated on A and B panels that are both in the same scale. **C** An overall structure of the crystallographic twofold-symmetric dimer of MtN13. The two MtN13 subunits are colored green and orange. ZEA molecule is shown in van der Waals representation in both ninding cavities. The L5 loops that are mutually exchanged are depicted in blue.

As a conclusion of this key section, it is worth to point major features concerning MtN13 dimerization. Firstly, MtN13 forms dimer in solution, as confirmed by DLS and SAXS experiments. Secondly, in the presence of cytokinin molecules, the dimer is formed by two protein molecules and two cytokinin ligands and thus the complex stoichiometry is best described as 2:2. Thirdly, the dimer formation seems to be independent of cytokinin binding since the SAXS experiments were performed using unliganded protein sample. It would be of a great interest to investigate the oligomeric state of MtN13 *in planta*. Aiming this, at the very least, the physiological concentrations of both MtN13 and of cytokinins during different phases of nodulation in *M. truncatula* would have to be determined.

The mutually exchanged L5 loop

Formation of the crystallographic dimer involves a mutual insertion of the loop L5 into the internal cavity of the dimer mate (Fig. 39). Loop L5 connects strands β3 and β4 and is five residues long ($_{60}$VGDFG$_{64}$). It is formed as an inverse γ-turn between Gly61 and Phe63, meaning that Asp62 is the central residue of the loop apex. The conformation of loop L5 in MtN13 crystal structure is stabilized by a sodium cation with almost all of the bonds between the metal cation and the electron donors shorter than 2.5 Å. The coordination sphere of the Na^+ cation is formed by the main-chain O atoms of Phe59 and Gly61 and water molecules. Only in the two highest-resolution structures, with ZEA (1.85 Å resolution) and 2iP (2.0 Å) the coordination octahedron is complete (six O ligands). In the complexes with KIN (2.2 Å) and BAP (2.6 Å), one and three, respectively, water molecules are defined by only weak peaks of positive electron density and, therefore, were not modeled. It is worth to note that in the 2.6 Å resolution structure of the BAP complex, a total of only 25 water molecules have been modeled. However, even in the BAP complex the sodium cation has been modeled by analogy to the structures of the other complexes resolved with higher resolution. Despite the coordination sphere in the BAP complex is incomplete (three ligands missing), the visible part has a geometry that is consistent with octahedral coordination of a sodium cation by oxygen ligands. Additionally, the angles $O_{equator}$…Na…O_{apex} are close to 90° and Na…O distances are approximately 2.5 Å.

In MtN13 topology, loop L5 penetrates the binding cavity of the second protein molecule through the only prominent entrance, located in its own neighborhood. The two L5 loops that are exchanged by the subunits pass, therefore, each other in an antiparallel fashion, as required by the twofold symmetry of the dimer. The L5 loops from both subunits interact with each other *via* water-mediated hydrogen bonds between the carboxylic O atom of Asp62 and the backbone N atom of Ser65 and carbonyl O atom of Phe63 from the dimer mate. The loop penetrates the second protein subunit so deeply that the Asp62 side chain at its apex creates a fork of two hydrogen bonds with the N^6 and N7 atoms of a cytokinin molecule docked inside the cavity. Therefore, the loop L5 can be considered an intermolecular "plug" sealing the only entrance to the cavity and pinning the cytokinin ligand in its binding site.

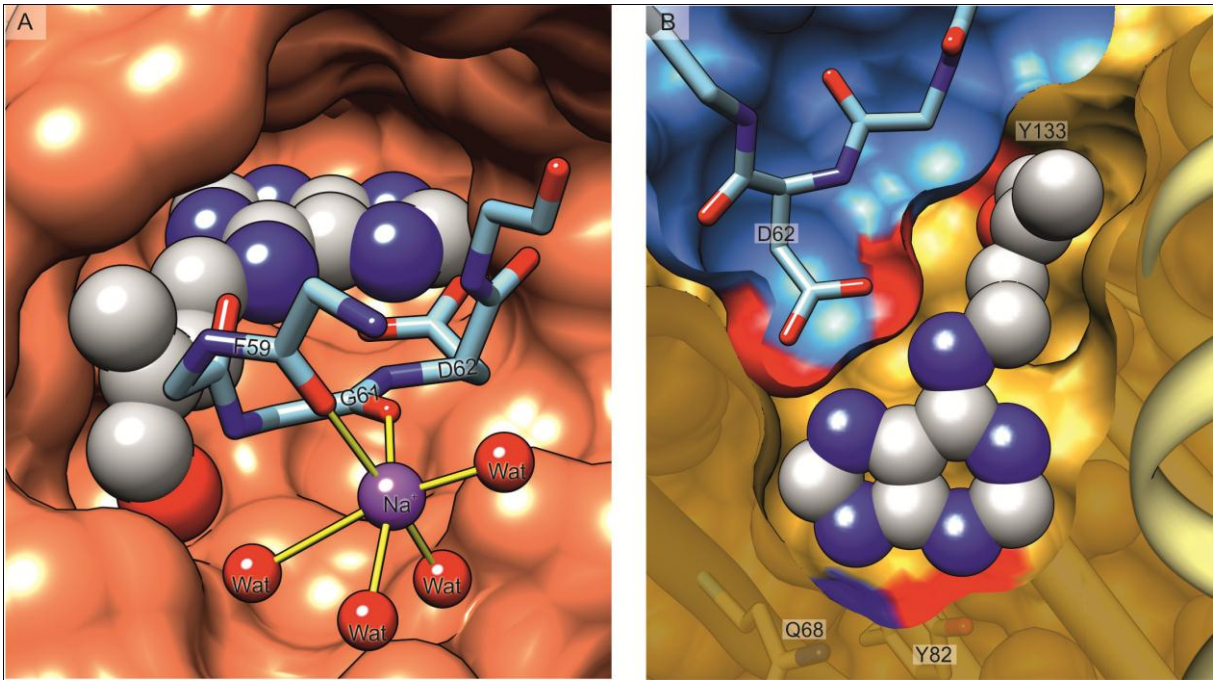


Figure 39
Loop L5, the "plug". MtN13 molecules are represented as their surfaces and the loop is shown as sticks with only main-chain atoms shown except for Asp62 for which the side-chain is also presented. ZEA molecule is depicted in van der Waals representation. **A** The internal cavity presented from the perspective of the L5 loop of the other protein subunit. Coordination bonds of a Na^+ cation are shown as yellow sticks. **B** The view perpendicular to the plane of ZEA adenine ring. The surface corresponding to the dimer mate is colored in blue. Note that the entrance to the cavity is entirely secured by the "plug".

Numerous bonding interactions secure the L5 "plug" in its "socket". Asp62 is the only amino acid with a polar side chain in the L5 loop. Apart from binding the cytokinin ligand, Asp62 also forms a hydrogen bond with Thr66 from the other protein subunit. In addition, a few water-mediated hydrogen bonds are formed between the L5 loop and the other protein molecule. It is of note that the data resolution limits the total number of water molecules

visible in the electron density maps. Therefore the water(or sodium)-mediated interactions at the interface between the "plug" and the "socket" are described below for the highest resolution complexes of MtN13 with ZEA and 2iP. One water-mediated hydrogen bond is formed by carbonyl O atom of Val60 and both Nη atoms of Arg140. Subsequent interactions of this type are created between two water molecules from the coordination sphere of the Na^+ cation (coordinated by backbone O atoms of Phe59 and Gly61) and the carbonyl O atoms of Gly87 and Ala89. Another docking interaction is mediated by additional Na^+ cation (present only in ZEA and 2iP complexes) adjacent to the L5 loop. This second Na^+ cation is coordinated by the carbonyl O atom of Asp62 from the "plug" and two atoms from the "socket", namely the carbonyl O atom of Ser65 and hydroxyl O atom of Thr66. The L5 loop is stabilized in its "socket additionally by hydrophobic contacts. More precisely, the side chain of Val60 interacts with Ala36 and the aromatic ring of Phe63 interacts with the aromatic rings of Phe38 and Phe59.

The internal cavity of MtN13: a cytokinin binding site

The ligand-binding cavities of PR-10 proteins are formed mainly by hydrophobic residues in the core of the protein. In addition, a few polar side chains point into the lumen of the cavity, and allow to create selective hydrogen bonds with suitable heteroatom partners of the ligand molecules. These internal cavities are most fascinating structural elements of PR-10 proteins. In MtN13 complexes several hydrophobic residues, namely Val84, Ile98, Phe100, Phe141 and Phe144 shape the walls of the cavity (Fig. 40). Hydrogen bonds with the cytokinin molecules are formed by Gln68, Tyr82 and (in the complexes with ZEA, KIN and BAP) Tyr133.

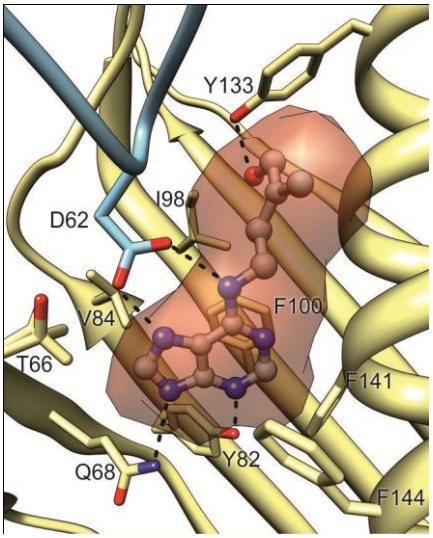


Figure 40
The internal cavity of MtN13. The cavity (with ZEA bound) is represented as a semitransparent surface. Hydrophobic residues create the walls of the cavity and polar amino acids that allow to bind to a ligand *via* hydrogen bonds.

The volume of the cavity of the MtN13 dimer is additionally reduced by the presence of the L5 loop from the second protein molecule that creates two additional hydrogen bonds with Asp62 from the apex of L5 loop. As a consequence, the available volume of the cavity in the MtN13/ZEA complex is only 270 $Å^3$, which is close to the volume of 186 $Å^3$ calculated for an isolated *trans*-zeatin molecule (contributions of H-atoms is not included). The volumes of the cavities of the other MtN13/cytokinin complexes are (values in parentheses refer to the volume of an isolated ligand molecule): 2iP, 300 $Å^3$ (179 $Å^3$); KIN, 250 (174 $Å^3$) and BAP 280 (188 $Å^3$). It has been noted that the cavity is the largest in the complex with 2iP (which itself is the smallest ligand). This is attributed to a different orientation of the Tyr133 side chain (see next paragraph). Another small cavity of about 80 $Å^3$ is found deep in the protein core. This small chamber is neither connected with the main cavity nor accessible from the outside of the protein and is occupied by two water molecules.

Two of the cytokinins used in this study, ZEA and 2iP, have substituents at the adenine N^6 atom (referred to as tails) that are very different in their chemical character. Specifically, the 2iP tail lacks the terminal hydroxyl group and, hence, is highly hydrophobic. Structural comparison of the MtN13 complexes with ZEA and 2iP reveals important differences in the arrangement of the upper part of the cavity (Fig. 41). If ZEA is bound in the cavity the side chain of Tyr133 is flipped towards the tail of the cytokinin ligand and forms a hydrogen bond with terminal oxygen atom of ZEA. In contrast, in the complex with 2iP Tyr133 points away from the hormone molecule, whereas Leu90 from the adjacent L7 loop is much closer (Cα shift of ~2 Å) to the tail of 2iP than in the ZEA complex. This means that the cavity can rearrange its binding elements and adjust their conformation for optimal binding of cytokinin ligands with chemically very different tails.

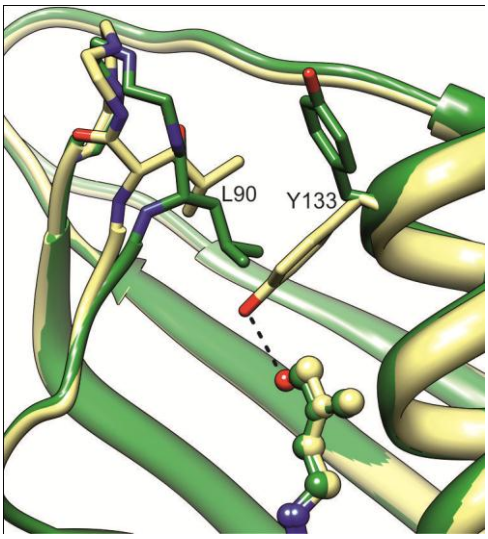


Figure 41
Structural adaptation of the upper-part of the internal cavity. The rearrangement upon binding two cytokinin ligands polar (ZEA, khaki) or non-polar (2iP, green). Note the flip of the Tyr133 and the shift of the L7 loop (marked by Leu90).

Except for the rearrangement of Tyr133 and loop L7 on 2iP binding, only minor tuning of the binding cavity is observed when the other cytokinins are bound. The four MtN13/cytokinin complexes are shown in Fig. 36. In all four cases, Gln68 and Tyr82 create hydrogen bonds with N9 and N3 atoms of the adenine ring, respectively. Also, the hydrogen bonds with the side chain of Asp62 from the other protein molecule of the homodimer are strictly conserved. Due to these interactions, the adenine framework of all four cytokinin molecules studied in this project is fixed in the same orientation. The Tyr133 side chain points towards the cytokinins that have an aromatic group in the tail (KIN and BAP) as it was described for ZEA complex. In the complex with KIN a weak hydrogen bond is formed between a C-H donor from the furanyl ring (KIN) and hydroxyl O atom from Tyr133. However, in the BAP complex a C-H…π interaction is present which involves the H-atom of the BAP benzyl ring that points towards the centroid of the aromatic ring of Tyr133. The fact that MtN13 can bind different cytokinin ligands in a conserved mode is the second (after dimerization) most exciting and novel feature, never described for PR-10 proteins.

The tautomeric form of the cytokinin ligands bound in the cavity of MtN13 can be resolved by investigation of the hydrogen bond network in which the cytokinins participate (Fig. 42). Typically, adenine derivatives are protonated at the N9 atom of the adenine ring. This is, however, not the case in the MtN13 complexes. The fork of hydrogen bonds to the N^6 and N7

atoms of cytokinin formed by the side chain of Asp62 (from the complementary protein molecule) could, theoretically, use either a carboxylate (with N7 protonated) or carboxylic (with N7 unprotonated) form. The latter circumstance is, however, highly unlikely when the pH (7.0) of the crystallization buffers is considered. Furthermore, the N9 atom cannot be protonated because it is an acceptor of the Nε–H…N9 hydrogen bond from the side-chain amide of Gln68 in all complexes. Noteworthy, the orientation of the Gln68 side chain was confirmed by the test calculations with the amide group flipped and careful analysis of the ADPs. Hence, the system of H-bond interactions makes sense only if N7 is protonated and N9 is not.

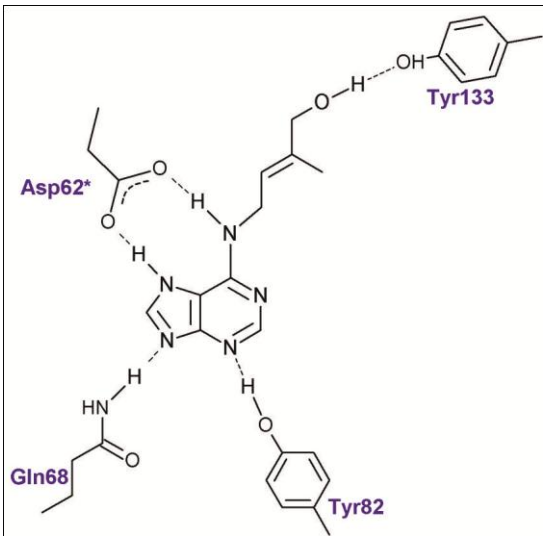


Figure 42
Tautomery of the bound cytokinins. Only mode of hydrogen-bonding network is logically consistent. ZEA complex is presented.

Another hydrogen bond is formed by the adenine N3 atom and the hydroxyl group of Tyr82 in all four complexes studied in this project. N1 is the only heteroatom of the adenine ring that is not involved in any hydrogen-bond interactions. This may be understood when the orientation of the cytokinin tails is concerned, as all cases these tails are *cis* oriented with respect to N1. Therefore, N^6 substituents cause steric hindrance and N1 is not able to form any hydrogen bond with the protein. It is noted that *trans*-zeatin was added to the incubation solution as a salt, in a protonated form of hydrochloride. There is however no hint indicating that the cytokinin in the MtN13/ZEA complex is protonated.

Considering both, the dimerization and the shape of the internal cavity of MtN13, one can postulate a mechanism of cytokinin binding by this PR-10 protein. As demonstrated by the

light and X-ray scattering studies of solutions of ligand-free MtN13, the presence of cytokinins is not a necessary condition for the homodimer to form. However, once it is formed, there is no entrance to nor exit from the internal cavity. This means that the MtN13 molecules that have dimerized without a ligand, most probably lose their binding properties. On the other hand, the MtN13 monomers that have bound a cytokinin prior to dimerization hinder the ligand permanently. It seems unlikely that the dimers can dissociate back to the monomers, at least when the protein concentration is relatively high, as in the scattering experiments on MtN13 solutions. The only possible scenario is thus that cytokinin binding must either precede or occur simultaneously with the dimerization event. In the first option, cytokinins would be bound in the cavity of two monomeric protein molecules and, subsequently, two monomeric complexes would dimerize sealing each other's cavity with their L5 loops. In the second option, cytokinin molecules would be fetched from the solvent by the N^6,N7-forks of hydrogen bonds with Asp62, and this assemblies would be mutually inserted in the ligand binding cavities of dimer partners. At this stage of studies it is impossible to distinguish between the two possibilities as it would require additional experiments such as site-directed mutagenesis of the Asp62 or residues at the dimer interface. Binding of cytokinins can and probably does change the dissociation constant (K_D) of the $(MtN13)_2$ homodimers. More precisely, K_D can be rather high in the absence of cytokinins and becomes reduced with the presence of cytokinin ligands inside the binding cavities. In this case, especially at low, physiological concentrations MtN13 would exist mostly as monomeric species and the homodimers would be formed upon binding of cytokinins. Unfortunately this is only a speculation and its further consideration, at the very least would require knowledge of profiles of physiological concentrations of both the MtN13 protein and its cytokinin ligands during nodulation in *Medicago truncatula*.

Structural comparison of MtN13 with other PR-10 homologs

The protein chains in the four studied here MtN13 complexes are very similar, as illustrated by the RMSDs of their Cα superpositions (~0.2-0.3 Å) shown in Table 5. Unless specified otherwise, in the subsequent structural comparisons the four MtN13 models are represented by the MtN13/ZEA structure, which has the highest resolution.

Table 5. Comparison of the four models of MtN13 from the cytokinin complexes studied in this work, characterized by RMSD (Å) values of pairwise Cα superpositions.

MtN13 complex with	2iP	KIN	BAP
ZEA	0.31	0.28	0.21
2iP	-	0.29	0.24
KIN	-	-	0.19

ZEA, *trans*-zeatin; 2iP, N^6-isopentenyladenine; KIN, kinetin; BAP, N^6-benzyladenine. The superpositions were calculated with UCSF Chimera (Pettersen et al., 2004).

Structural alignment of representative PR-10 proteins with the present MtN13/ZEA structure are characterized by RMSD values of Cα superpositions in Table 6. The RMSD values are similar (1.45-1.88 Å) even when distant homologs are compared. The following part of this section provides a detailed comparison of selected structural elements. The β-sheet of PR-10 proteins is a very conserved structural element as it is the case of MtN13 when compared to the other PR-10 proteins. For example, in the crystal structures of the yellow lupine LlPR-10.2B protein in complexes with ZEA (Fernandes *et al.*, 2008) and N,N'-diphenylurea (DPU) (Fernandes *et al.*, 2009), a water molecule disrupts the canonical hydrogen-bonding pattern between strands β5-β6 close to loop L7. This water molecule mediates the backbone interactions between the O of Val85 and N atom of Lys96. In the MtN13 complexes a similar motif is present with a water molecule that is a donor in hydrogen bonds with main-chain O atoms, of Tyr85 and Glu96 from the strands β5 and β6, respectively. This rupture of the β-sheet certainly influences the conformation of loop L7 that itself must make room for the L5 loop that is mutually exchanged with the other protein subunit of the homodimer.

Table 6. Structural superpositions of selected PR-10 proteins with the present MtN13 structure, represented by the protein chain from the MtN13/ZEA complex. The superpositions were calculated with PDBe *Fold v2.55* (Krissinel & Henrick, 2004).

Protein structure	PDB code (chain ID)[a]	RMSD of Cα atoms (Å)	Number of aligned residues	Sequence identity (%)	Q-score[b]
VrCSBP/ZEA	2flh (A)	1.45	135	22	0.63
LlPR-10.1B	1ifv (A)	1.59	155	39	0.69
LlPR-10.2B/ZEA	2qim	1.77	149	44	0.64
LlPR-10.2B/DPU	3e85	1.80	146	44	0.62
Bet v 1a	1bv1	1.80	152	44	0.65
LlPR-10.1A	1icx	1.88	155	38	0.64

[a] Only the chain with the lowest RMSD was used for the analysis, if more than one protein chain was present in crystal structure,
[b] Q-score represents the quality function of Cα alignment. It reduces the effect of RMSD - N_{algn} (number of aligned residues) balance on the estimation of alignments (N_{res1} and N_{res2} stand for the number of residues in the aligned proteins, and the empirical parameter R_0 is set to 3 Å): $Q = (N_{algn} \cdot N_{algn})/[(1+(RMSD/R_0)^2) \cdot N_{res1} \cdot N_{res2}]$.

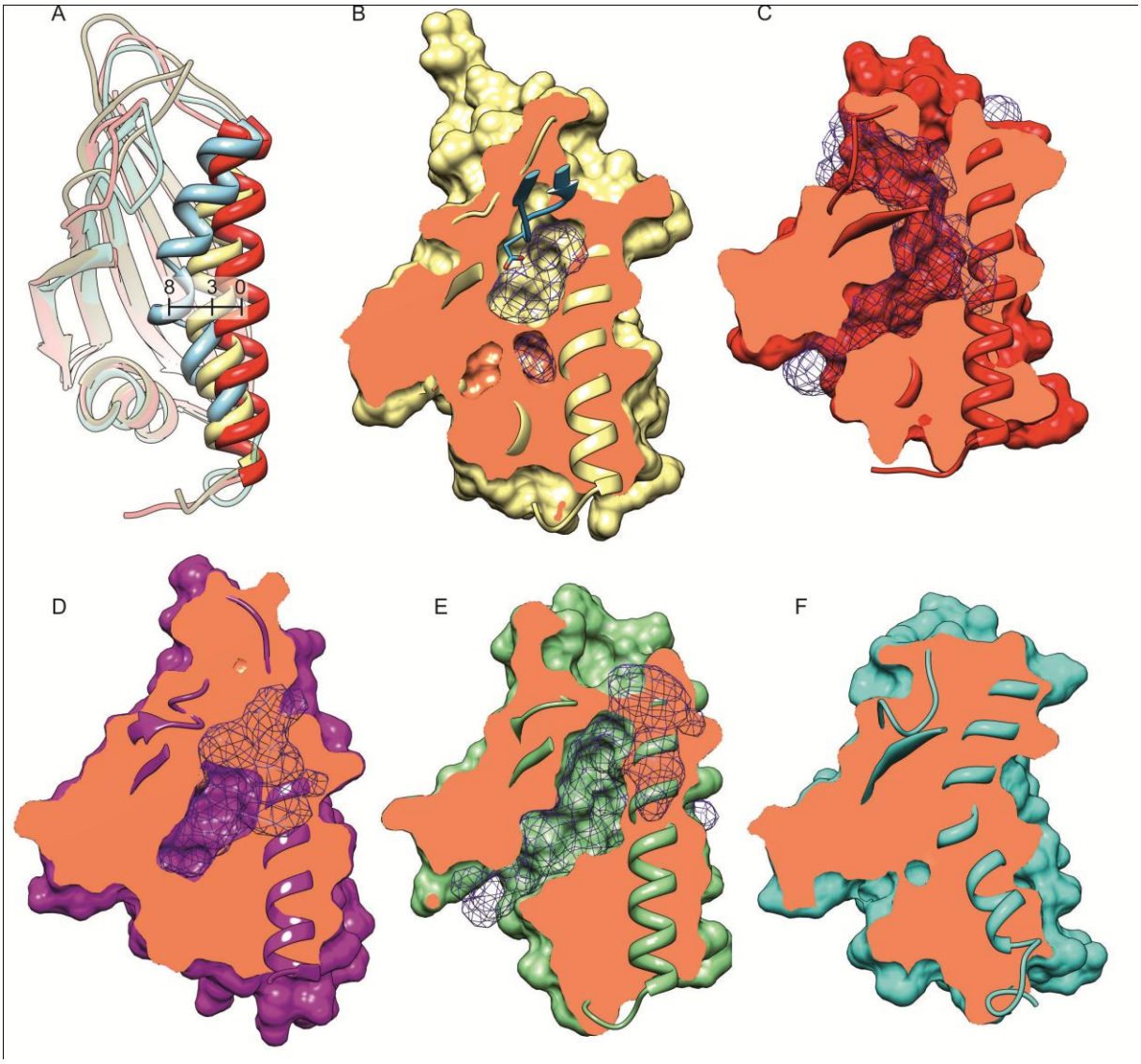
The glycine-rich loop L4 has been most comprehensively analyzed among the numerous loops in the PR-10 structures (Pasternak *et al.*, 2006; Fernandes *et al.*, 2008; Fernandes *et al.*, 2009; Fernandes *et al.*, 2013). The sequence signature of loop L4, EGxGGxGT, is also conserved in MtN13. This signature is similar to the phosphate-binding loop of nucleotide-binding proteins (Saraste *et al.*, 1990). Nucleotides, however, have never been confirmed to be binding partners of PR-10 proteins (Koistinen *et al.*, 2005). Interestingly, the L4 loop is highly conserved in the PR-10 proteins not only in sequence but also from the structural point of view. The latter observation is surprising when the high content of Gly residues in the L4 loop is considered. Typically, the rigid L4 conformation is stabilized by hydrogen bonds between the side chain of a Thr in its sequence and the main-chain atoms which is also preserved in the MtN13 structures. The L4 loop of MtN13 protein is very well defined in the electron density maps and its atoms have low ADP parameters. Despite the numerous studies, no exact function for the L4 loop of the PR-10 proteins has been suggested.

The odd-number loops L3-L9 of PR-10 proteins, compared above to the "fingers" of the baseball glove, surround the main entrance to the internal cavity. Among these loops only L3 and L9 have been reported to be involved in metal coordination in previously solved structures. In most examples, a single sodium cation interacts with either one of these loops. The only two exceptions were: (i) two Na^+ cations coordinated by both loops in the LlPR-10.2B/DPU complex (Fernandes *et al.*, 2009) and (ii) a Ca^{2+} cation coordinated by the L3 loop in the crystal structure of the LlPR-10.2B/ZEA complex (Fernandes *et al.*, 2008). MtN13 is so far the only example of a PR-10 protein with loop L5 involved in coordination of a metal cation. This Na^+ cation causes deformation of loop L5 that forces it to bend away from the protein core. The bending of the L5 loop itself is associated with the dimer-formation process, in which loop L5 plays a crucial role.

Because of two reasons, namely very low sequence conservation and a very high degree of structural variability, the α3 helix at the C-terminus is the most variable structural element of PR-10 proteins. On the other hand, the α3 helix is an essential component of the canonical PR-10 fold. The α3 helix is responsible for the shape of the internal cavity, as it can be either

straight or kinked towards the protein core, it has impact on the volume of the cavity. The structural and sequential variability of this helix allow for modulation of the volume and chemical character of the internal cavity. Fernandes *et al.* (2008) suggested that the α3 helix is quite elastic and that ligand binding causes straightening of the helix. The authors compared two PR-10 proteins from yellow lupine, namely LlPR-10.2B in complex with ZEA and its homolog LlPR-10.2A (91% sequence identity), crystallized without a ligand (Pasternak *et al.*, 2005). The strong kink towards the protein core in the middle of the α3 helix that actually fulfills the cavity was observed in the "empty" LlPR-10.2A structure. In LlPR-10.2B/ZEA complex the cavity is significantly larger as compared to the unliganded structure of LlPR-10.2A. Unfortunately, there has not been a crystallographic study of the same PR-10 protein in both forms, with and without a ligand. The recently reported crystal structures of the most comprehensively studied PR-10 protein, the birch pollen allergen Bet v 1 (Kofler *et al.*, 2012) do not solve the problem either. Although the authors determined the crystal structures without the ligand of their interest, ingredients of the crystallization cocktail are still present inside the cavity making the analysis of helix α3 conformation unreliable. The structural alignment in context of the helix α3 deformation for MtN13/ZEA complex with two extreme examples, namely LlPR-10.2A and LlPR-10.2B/ZEA is presented in Figure 43A. The α3 helix of the MtN13 protein has a relatively moderate kink. It is ~3 Å more profound as compared to the almost straight line of LlPR-10.2B/ZEA and ~5 Å less bent than the helix of LlPR-10.2A. As a result of the bending of helix α3, the internal cavity of the MtN13 protein is very small in contrast to other structures of PR-10 proteins with ligands inside the cavity. One has to remember that the volume of the cavity of the MtN13 homodimer is additionally reduced by the L5 loop from the other protein subunit.

Of course the inward kink of the α3 helix is not the only factor influencing the shape and the volume of the internal cavity as they are also regulated by the relative location of all structural elements that surround the chamber. Figure 43B-F presents the internal cavities of five PR-10 members, including MtN13. In MtN13, there is only one solvent-accessible entrance to the internal cavity. Moreover, as mentioned above, it is sealed upon MtN13 homodimer formation. On the contrary, PR-10 proteins with largest cavities can have up to three entrances, as observed for instance in the LlPR-10.2B/ZEA complex (Fernandes *et al.*, 2008).



1

Figure 43
Structural analysis of MtN13 and other PR-10 proteins in the context of their internal cavities. Helix α2 as well as strands β2 and β3 have been omitted for clarity. **A** Comparison of the inward kink of helix α3. The three cases are illustrated by the crystal structures of (i) MtN13 in complex with ZEA (this work, yellow), (ii) LlPR-10.2B in complex with ZEA (PDB ID: 2qim, red), and (iii) LlPR-10.2A without any ligand in the protein cavity (1xdf, blue). All secondary structure elements, except helix α3, are semitransparent. Note that in the LlPR-10.2A structure the α3 helix is bent so strongly that the hydrogen-bond pattern is distorted and in fact it should be regarded as two separate helices. The approximate indentations of the helices toward the lumen of the cavity are given in Å. **B-F** The internal cavities of five representative PR-10 proteins: (**B**) MtN13/ZEA complex (this work) with the cavity surface including the elements from the second protein subunit (shown in blue), cavity volume 270 + 80 $Å^3$; (**C**) LlPR-10.2B in complex with ZEA (PDB 2qim), 2440 $Å^3$; (**D**) VrCSBP in complex with ZEA (2flh, chain A), 950 $Å^3$; (**E**) Bet v 1 in complex with KIN (4a85), 1900 $Å^3$; (**F**) LlPR-10.2A without any ligand (1xdf) with no solvent accessible cavity. The cavity volumes were calculated with *Surfnet* (Laskowski, 1995) and displayed as a 3D blue mesh. A 1.5-Å gap spheres radius was used for the calculations to represent solvent accessible areas. Solvent atoms, ions, ligands and hydrogen atoms were disregarded in the calculations. The protein surfaces have been clipped in a cut-away mode to show a maximum outlook of each cavity.

2

Cytokinin binding by PR-10 proteins

Three proteins from PR-10 family have been shown to bind cytokinins. Interestingly, in each example the cytokinin molecules are bound in a completely different way and even the stoichiometry of the complex for the same PR-10/cytokinin dyad is not conserved (Fernandes *et al.*, 2013). This diversity is combined with a surprisingly excellent definition of the ligand molecules in electron density maps. To sum up the previous results, LlPR-10.2B can bind either three ZEA or four DPU molecules inside the cavity (Fernandes *et al.*, 2008; Fernandes *et al.*, 2009). Bet v 1 a was shown to bind a single KIN molecule in two orientations, or both KIN and 8-anilino-1-naphthalene sulfonate (ANS) in the same cavity (Kofler *et al.*, 2012). In ZEA complexes of *Vigna radiata* CSBP (VrCSBP), one or two cytokinin molecules are present in the internal cavity, and the binding can occur in two ways (Pasternak *et al.*, 2006). Only one ligand in the VrCSBP/ZEA complexes has a conserved binding mode. In general, however, no common pattern of cytokinin binding can be pointed out as they can occupy different positions and have various orientations in the internal cavity.

On this background, the unique and conserved cytokinin binding site of the MtN13 protein is an exciting novelty, never observed before for PR-10 proteins. Analysis of the amino acid residues forming hydrogen bonds with the cytokinin molecules in other crystal structures of PR-10 complexes reveals that the binding mode is highly variable (Fig. 44). There is no residue that would be universally involved in cytokinin binding and conserved throughout the entire family of PR-10 proteins. Amino acid residues at positions corresponding to Gln68 and Tyr82 of MtN13 interact with cytokinins in some PR-10 complexes, however, in each case a different atom of the ligand is recognized by these residues. More precisely, in MtN13, Gln68 forms a hydrogen bond with the N9 atom of each cytokinin molecule, while in the VrCSBP/ZEA complex Glu69, which occupies the same structural position, binds the N7 atom of the ligand. Tyr82 forms binds the N3 atom of the adenine ring in MtN13/cytokinin complexes while in the LlPR10.2B/ZEA the same residue forms hydrogen bond with the terminal hydroxyl group from the cytokinin tail. Hence, neither the orientation nor the position of the cytokinin ligands is preserved in the internal cavity of the PR-10 proteins studied before.

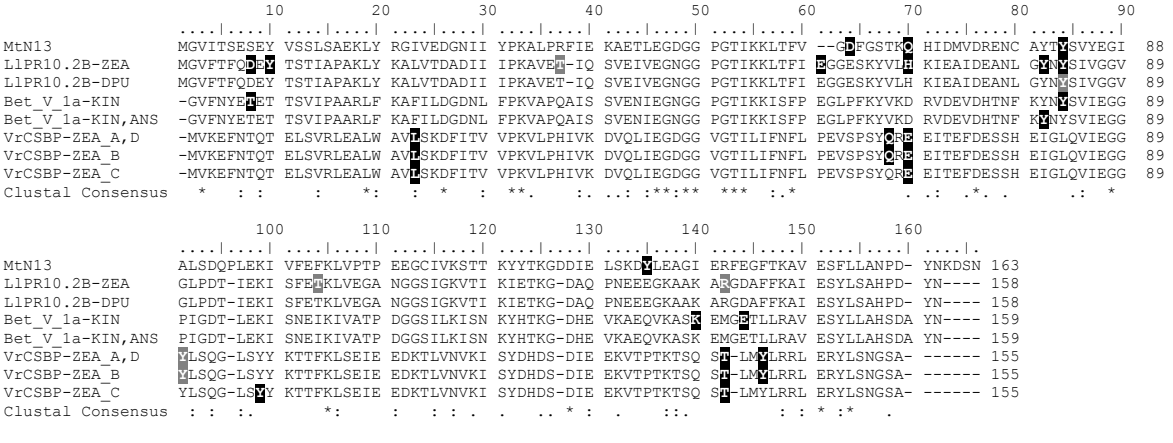


Figure 44
Sequence alignment of PR-10 proteins with reported structures of cytokinin complexes. Identical residues are marked with an asterisk. Colon indicates conservative, and a period - semiconservative substitutions. Residues marked in black form hydrogen bonds with a cytokinin molecule directly and those in gray interact *via* a water-mediated hydrogen bond. In the case of the VrCSBP/ZEA complex with four protein molecules in the asymmetric unit (A-D), the two molecules with identical ligand stoichiometry (1:2) and binding (A and D) are shown together. ANS stands for 8-anilino-1-naphthalene sulfonate.

Despite all the differences and variability, one common feature of cytokinin recognition by the PR-10 proteins can be deduced. Namely, the fork-like interaction between the N^6 and N7 atoms of the cytokinin ligand and an acid or amide side chain of binding amino acid. This type of interaction has been observed in the LlPR-10.2B/ZEA complex with the involvement of Asp7 and in the VrCSBP/ZEA complex where Glu69 (in all protein chains A-D) and Gln67 (chains A, D) participate in binding two separate ZEA molecules. In the VrCSBP complexes where Gln67 is involved, the binding scheme indicates that the ligand molecule must have the alternative tautomeric form, with N9 atom of ZEA protonated. One must, however, keep in mind that the fork-like hydrogen-bonding can be realized in the internal cavity by different residues and from various locations within the cavity. MtN13 is the most spectacular example, because these interactions are formed by Asp62 from loop L5 loop coming from the complementary protein molecule in the homodimeric assembly.

Comparison of MtN13 and ligand binding domain of cytokinin receptor

Crystal structure of the extracellular cytokinin binding domain of the cytokinin receptor has been solved in 2011 (Hothorn *et al.*). Authors reported a set of crystal structures of the sensor domain of *Arabidopsis thaliana* histidine kinase 4, AHK4, in complexes with a large set of various natural and synthetic cytokinins. AHK4 had been known before as a cytokinin receptor (Kakimoto, 2003). The receptor is a dimer, but the dimerization interface is different from that of MtN13. The overall fold of the hormone binding domain of AHK4 is different

from MtN13, however, the hormone-binding cavity is created in a similar manner between an antiparallel β-sheet and an α-helix. The binding mode of different natural cytokinins is conserved in all the studied complexes with the AHK4 sensor domain. The location and orientation of the bound cytokinins is, of course, do not resemble the case of MtN13 due to different fold of the proteins itself. Therefore, MtN13 is the second protein, which crystal structure has been solved, that is able to bind various cytokinins in the same fashion. Despite the differences in the location of the cytokinin ligands in the internal cavity, important similarities between MtN13 and AHK4 can be inferred (Fig. 45). In both protein clusters, MtN13 and AHK4 complexes, the N^6 and N7 atoms of the cytokinin framework are recognized by a fork-like hydrogen-bonding interactions with the participation of Asp residues. More precisely, in AHK4 Asp262 binds cytokinins while in MtN13 the interacting Asp62 is provided by the second protein subunit of the homodimeric assembly. Another similar feature is the tautomeric form of the bound cytokinins, as it is the same in both clusters of complexes. In all cases, the N7 atom of the cytokinin adenine ring must be protonated because this is the only state that makes logically consistent hydrogen-bond-network.

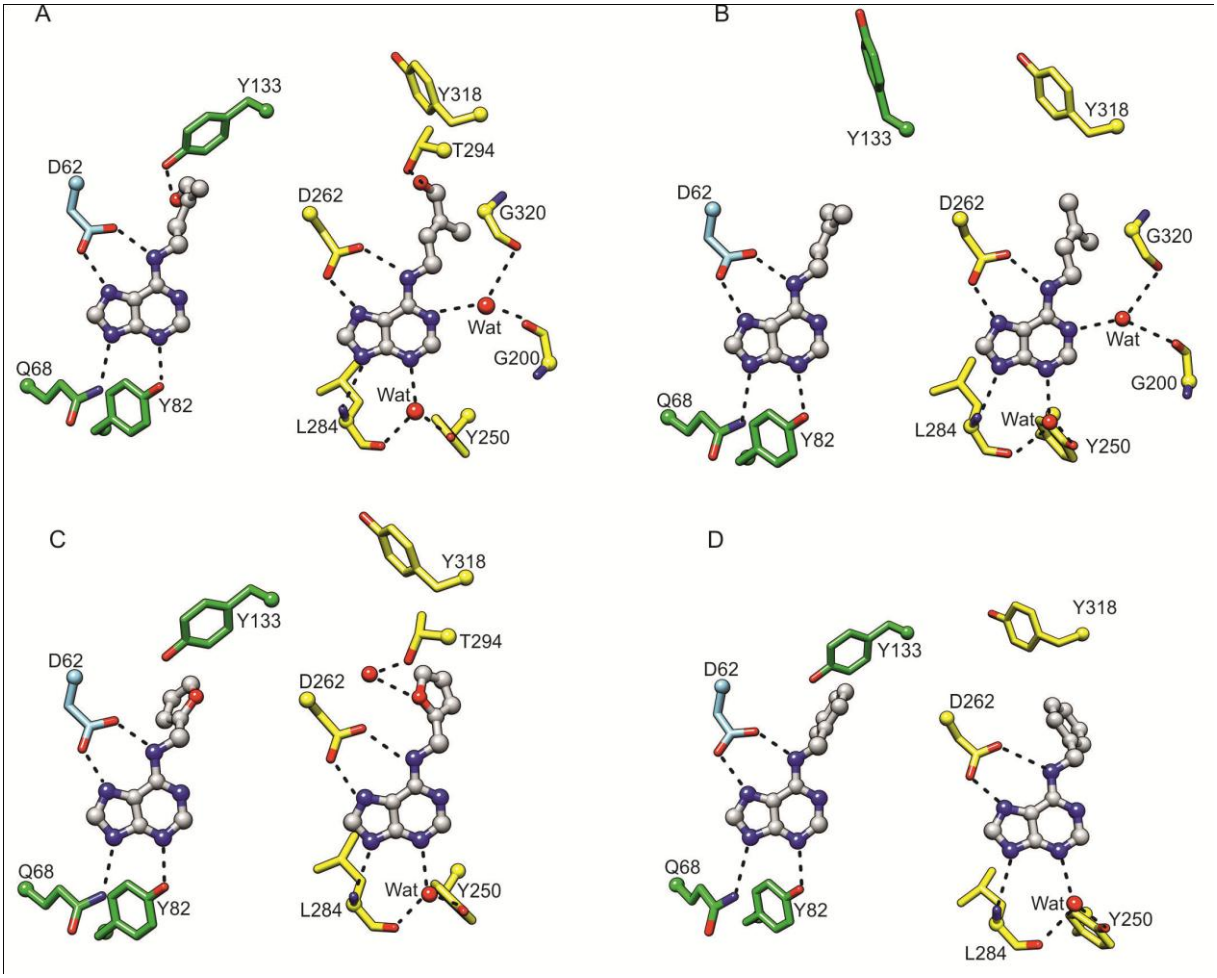


Figure 45
Comparison of cytokinin binding mode of MtN13 (the present complexes) and of the cytokinin-binding domain of cytokinin receptor, AHK4, studied in complexes with ZEA (PDB ID: 3t4l), 2iP (3t4j), KIN (3t4s), and BAP (3t4k) (Hothorn *et al.*, 2011). Each panel (**A**-**D**) compares the structures of MtN13 (left, green) and AHK4 (right, yellow) with the same ligand, as follows ZEA (**A**), 2iP (**B**), KIN (**C**) and BAP (**D**). The second molecule of the MtN13 homodimer is colored in blue. Cytokinin ligands are grey (C atoms) and the heteroatoms colored in a usual mode. The views are perpendicular to the purine ring of the cytokinins, disregarding any protein alignment. Cα atoms are highlighted as balls and water molecules are shown as red spheres. Unless binding cytokinins (Gly200, Leu284 and Gly320 of AHK4), the main-chain atoms were omitted for clarity.

Additional feature that needs to be discussed concerns the portion of the internal cavity that is responsible for binding the cytokinin tail. In the MtN13/cytokinin complexes, as mentioned before, Tyr133 is flipped towards the cytokinin tail whenever a binding interaction is possible (complexes with ZEA, KIN and BAP). On the contrary, in the AHK4 complexes Tyr318 (that is at a similar location to Tyr133 in MtN13) is always flipped away from the ligand molecule. In other words, the upper part of the cavity in AHK4 complexes resembles the case of the MtN13/2iP complex. In the AHK4 complexes, hydrogen bonds (if possible) with the cytokinin tails are formed by Thr294. More exactly, the hydroxyl group of Thr294 forms a direct hydrogen bond with the terminal hydroxyl oxygen atom of the ZEA tail in the AHK4/ZEA complex. In AHK4/KIN complex, the hydrogen bond between Thr294 and with the oxygen atom from the furanyl ring of KIN is mediated by a water molecule.

Conformation of the ligand molecules

Conclusions

References

Adams DO & Yang SF (1979) Ethylene biosynthesis: Identification of 1-aminocyclopropane-1-carboxylic acid as an intermediate in the conversion of methionine to ethylene. *Proc Natl Acad Sci U S A* **76**, 170-174.

Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung LW, Kapral GJ, Grosse-Kunstleve RW, McCoy AJ, Moriarty NW, Oeffner R, Read RJ, Richardson DC, Richardson JS, Terwilliger TC & Zwart PH (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Cryst. D* **66**, 213-221.

Altschul SF, Gish W, Miller W, Myers EW & Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* **215**, 403-410.

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W & Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**, 3389-3402.

Aragao KS, Satre M, Imberty A & Varrot A (2008) Structure determination of Discoidin II from Dictyostelium discoideum and carbohydrate binding properties of the lectin domain. *Proteins* **73**, 43-52.

Ariel F, Brault-Hernandez M, Laffont C, Huault E, Brault M, Plet J, Moison M, Blanchet S, Ichante JL, Chabaud M, Carrere S, Crespi M, Chan RL & Frugier F (2012) Two direct targets of cytokinin signaling regulate symbiotic nodulation in Medicago truncatula. *Plant Cell* **24**, 3838-3852.

Armstrong DJ, Burrows WJ, Skoog F, Roy KL & Soll D (1969) Cytokinins: distribution in transfer RNA species of Escherichia coli. *Proc. Natl. Acad. Sci. USA* **63**, 834-841.

Ashikari M, Sakakibara H, Lin S, Yamamoto T, Takashi T, Nishimura A, Angeles ER, Qian Q, Kitano H & Matsuoka M (2005) Cytokinin oxidase regulates rice grain production. *Science* **309**, 741-745.

Badescu GO & Napier RM (2006) Receptors for auxin: will it all end in TIRs? *Trends Plant Sci* **11**, 217-223.

Barciszewska M, Keith G, Dirheimer G & Barciszewski J (1988) The nucleotide sequence of tRNA(Ser) and tRNA(Ile) of yellow lupin seeds. *Nucleic Acids Res.* **16**, 8175.

Barciszewski J, Mielcarek M, Stobiecki M, Siboska G & Clark BF (2000) Identification of 6-furfuryladenine (kinetin) in human urine. *Biochem. Biophys. Res. Commun.* **279**, 69-73.

Barciszewski J, Siboska GE, Pedersen BO, Clark BF & Rattan SI (1996) Evidence for the presence of kinetin in DNA and cell extracts. *FEBS Lett* **393**, 197-200.

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

Berry GC (2010) The hydrodynamic and conformational properties of denatured proteins in dilute solutions. *Protein Sci* **19**, 94-98.

Bradford MM (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem* **72**, 248-254.

Branca A, Paape TD, Zhou P, Briskine R, Farmer AD, Mudge J, Bharti AK, Woodward JE, May GD, Gentzbittel L, Ben C, Denny R, Sadowsky MJ, Ronfort J, Bataillon T, Young ND & Tiffin P (2011) Whole-genome nucleotide diversity, recombination, and linkage disequilibrium in the model legume Medicago truncatula. *Proc Natl Acad Sci U S A* **108**, E864-870.

Bryson K, McGuffin LJ, Marsden RL, Ward JJ, Sodhi JS & Jones DT (2005) Protein structure prediction servers at University College London. *Nucleic Acids Res* **33**, W36-38.

Bujacz G, Wrzesniewska B & Bujacz A (2010) Cryoprotection properties of salts of organic acids: a case study for a tetragonal crystal of HEW lysozyme. *Acta Cryst. D* **66**, 789-796.

Caesar K, Thamm AM, Witthoft J, Elgass K, Huppenberger P, Grefen C, Horak J & Harter K (2011) Evidence for the localization of the Arabidopsis cytokinin receptors AHK3 and AHK4 in the endoplasmic reticulum. *J Exp Bot* **62**, 5571-5580.

Cakici O, Sikorski M, Stepkowski T, Bujacz G & Jaskolski M (2010) Crystal structures of NodS N-methyltransferase from Bradyrhizobium japonicum in ligand-free form and as SAH complex. *J Mol Biol* **404**, 874-889.

Cano-Delgado A, Yin Y, Yu C, Vafeados D, Mora-Garcia S, Cheng JC, Nam KH, Li J & Chory J (2004) BRL1 and BRL3 are novel brassinosteroid receptors that function in vascular differentiation in Arabidopsis. *Development* **131**, 5341-5351.

Chang C & Stadler R (2001) Ethylene hormone receptor action in Arabidopsis. *Bioessays* **23**, 619-627.

Chen VB, Arendall WB, 3rd, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS & Richardson DC (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Cryst. D* **66**, 12-21.

Cheng X, Ruyter-Spira C & Bouwmeester H (2013) The interaction between strigolactones and other plant hormones in the regulation of plant development. *Front Plant Sci* **4**, 199.

Chini A, Fonseca S, Fernandez G, Adie B, Chico JM, Lorenzo O, Garcia-Casado G, Lopez-Vidriero I, Lozano FM, Ponce MR, Micol JL & Solano R (2007) The JAZ family of repressors is the missing link in jasmonate signalling. *Nature* **448**, 666-671.

Ciesielska A, Ruszkowski M, Kasperska A, Femiak I, Michalski Z & Sikorski M (2012) New insights into the signaling and function of cytokinins in higher plants. *BioTechnologia* **93**, 400-413.

Clouse SD & Sasse JM (1998) BRASSINOSTEROIDS: Essential Regulators of Plant Growth and Development. *Annu Rev Plant Physiol Plant Mol Biol* **49**, 427-451.

Cohen SX, Ben Jelloul M, Long F, Vagin A, Knipscheer P, Lebbink J, Sixma TK, Lamzin VS, Murshudov GN & Perrakis A (2008) ARP/wARP and molecular replacement: the next generation. *Acta Crystallogr D Biol Crystallogr* **64**, 49-60.

Cooper DR, Boczek T, Grelewska K, Pinkowska M, Sikorska M, Zawadzki M & Derewenda Z (2007) Protein crystallization by surface entropy reduction: optimization of the SER strategy. *Acta Crystallogr D Biol Crystallogr* **63**, 636-645.

Cooper JB & Long SR (1994) Morphogenetic Rescue of Rhizobium meliloti Nodulation Mutants by trans-Zeatin Secretion. *Plant Cell* **6**, 215-225.

Czerpak R & Piotrowska A (2003) Cytokininy, ich struktura, metabolizm i aktywność biologiczna. *Kosmos* **52**, 203-215.

D'Arcy A (1994) Crystallizing proteins - a rational approach? *Acta Crystallogr D Biol Crystallogr* **50**, 469-471.

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

Dauter Z (2002) One-and-a-half wavelength approach. *Acta Crystallogr D Biol Crystallogr* **58**, 1958-1967.

de la Serve BT, Jouanneau JP & Peaud-Lenoel C (1984) Incorporation of N-Benzyladenine into Messenger Poly(A)-RNA of Tobacco Cells Incubated at Stimulatory or Cytostatic Cytokinin Concentration. *Plant Physiol.* **74**, 669-674.

De Paepe A & Van der Straeten D (2005) Ethylene biosynthesis and signaling: an overview. *Vitam Horm* **72**, 399-430.

Dharmasiri N, Dharmasiri S & Estelle M (2005) The F-box protein TIR1 is an auxin receptor. *Nature* **435**, 441-445.

Dieckman L, Gu M, Stols L, Donnelly MI & Collart FR (2002) High throughput methods for gene cloning and expression. *Protein Expr Purif* **25**, 1-7.

Don RH, Cox PT, Wainwright BJ, Baker K & Mattick JS (1991) 'Touchdown' PCR to circumvent spurious priming during gene amplification. *Nucleic Acids Res* **19**, 4008.

Duhr S & Braun D (2006) Why molecules move along a temperature gradient. *Proc Natl Acad Sci U S A* **103**, 19678-19682.

Durrant WE & Dong X (2004) Systemic acquired resistance. *Annu Rev Phytopathol* **42**, 185-209.

Emsley P, Lohkamp B, Scott WG & Cowtan K (2010) Features and development of Coot. *Acta Cryst. D* **66**, 486-501.

Ernst D, Schafer W & Oesterhelt D (1983) Isolation and identification of a new, naturally occurring cytokinin (6-benzylaminopurine riboside) from an old anise cell culture (Pimpinella anisum L.). *Planta* **159**, 222-225.

Fernandes H, Bujacz A, Bujacz G, Jelen F, Jasinski M, Kachlicki P, Otlewski J, Sikorski MM & 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.

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

Ferreira FJ & Kieber JJ (2005) Cytokinin signaling. *Curr Opin Plant Biol* **8**, 518-525.

Fleet CM & Sun TP (2005) A DELLAcate balance: the role of gibberellin in plant morphogenesis. *Curr Opin Plant Biol* **8**, 77-85.

Frugier F, Kosuta S, Murray JD, Crespi M & Szczyglowski K (2008) Cytokinin: secret agent of symbiosis. *Trends Plant Sci* **13**, 115-120.

Fujii H, Chinnusamy V, Rodrigues A, Rubio S, Antoni R, Park SY, Cutler SR, Sheen J, Rodriguez PL & Zhu JK (2009) In vitro reconstitution of an abscisic acid signalling pathway. *Nature* **462**, 660-664.

Fujioka S & Sakurai A (1997) Brassinosteroids. *Nat Prod Rep* **14**, 1-10.

Fujita Y, Nakashima K, Yoshida T, Katagiri T, Kidokoro S, Kanamori N, Umezawa T, Fujita M, Maruyama K, Ishiyama K, Kobayashi M, Nakasone S, Yamada K, Ito T, Shinozaki K & Yamaguchi-Shinozaki K (2009) Three SnRK2 protein kinases are the main positive regulators of abscisic acid signaling in response to water stress in Arabidopsis. *Plant Cell Physiol* **50**, 2123-2132.

Gleason C, Chaudhuri S, Yang T, Munoz A, Poovaiah BW & Oldroyd GE (2006) Nodulation independent of rhizobia induced by a calcium-activated kinase lacking autoinhibition. *Nature* **441**, 1149-1152.

Goldschmidt L, Cooper DR, Derewenda ZS & Eisenberg D (2007) Toward rational protein crystallization: A Web server for the design of crystallizable protein variants. *Protein Sci* **16**, 1569-1576.

Gonzalez-Rizzo S, Crespi M & Frugier F (2006) The Medicago truncatula CRE1 cytokinin receptor regulates lateral root development and early symbiotic interaction with Sinorhizobium meliloti. *Plant Cell* **18**, 2680-2693.

Haberer G & Kieber JJ (2002) Cytokinins. New insights into a classic phytohormone. *Plant Physiol* **128**, 354-362.

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

Hirose N, Takei K, Kuroha T, Kamada-nobusada T & Hayashi H (2008) Regulation of cytokinin biosynthesis, compartmentalization and translocation. *J. Exp. Bot.* **59**, 75-83.

Hothorn M, Dabi T & Chory J (2011) Structural basis for cytokinin recognition by Arabidopsis thaliana histidine kinase 4. *Nat Chem Biol* **7**, 766-768.

Huneeus VQ, Wiley MH & Siperstein MD (1980) Isopentenyladenine as a mediator of mevalonate-regulated DNA replication. *Proc Natl Acad Sci U S A* **77**, 5842-5846.

Hwang I & Sheen J (2001) Two-component circuitry in Arabidopsis cytokinin signal transduction. *Nature* **413**, 383-389.

Hwang I, Sheen J & Muller B (2012) Cytokinin signaling networks. *Annu Rev Plant Biol* **63**, 353-380.

Israelsson M, Siegel RS, Young J, Hashimoto M, Iba K & Schroeder JI (2006) Guard cell ABA and CO2 signaling network updates and Ca2+ sensor priming hypothesis. *Curr Opin Plant Biol* **9**, 654-663.

Kabsch W (2010) Xds. *Acta Cryst. D* **66**, 125-132.

Kakimoto T (1996) CKI1, a histidine kinase homolog implicated in cytokinin signal transduction. *Science* **274**, 982-985.

Kakimoto T (2003) Perception and signal transduction of cytokinins. *Annual review of plant biology* **54**, 605-627.

Kaplan W, Husler P, Klump H, Erhardt J, Sluis-Cremer N & Dirr H (1997) Conformational stability of pGEX-expressed Schistosoma japonicum glutathione S-transferase: a detoxification enzyme and fusion-protein affinity tag. *Protein Sci* **6**, 399-406.

Karpinski S & Szechynska-Hebda M (2010) Secret life of plants: from memory to intelligence. *Plant Signal Behav* **5**, 1391-1394.

Kendrick MD & Chang C (2008) Ethylene signaling: new levels of complexity and regulation. *Curr Opin Plant Biol* **11**, 479-485.

Kim Y, Babnigg G, Jedrzejczak R, Eschenfeldt WH, Li H, Maltseva N, Hatzos-Skintges C, Gu M, Makowska-Grzyska M, Wu R, An H, Chhor G & Joachimiak A (2011) High-throughput protein purification and quality assessment for crystallization. *Methods* **55**, 12-28.

Klock HE & Lesley SA (2009) The Polymerase Incomplete Primer Extension (PIPE) method applied to high-throughput cloning and site-directed mutagenesis. *Methods Mol Biol* **498**, 91-103.

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.

Koistinen KM, Soininen P, Venalainen TA, Hayrinen J, Laatikainen R, Perakyla M, Tervahauta AI & Karenlampi SO (2005) Birch PR-10c interacts with several biologically important ligands. *Phytochemistry* **66**, 2524-2533.

Krissinel E & Henrick K (2007) Inference of macromolecular assemblies from crystalline state. *J Mol Biol* **372**, 774-797.

Kudo T, Kiba T & Sakakibara H (2010) Metabolism and Long-distance Translocation of Cytokinins. *J. Integr. Plant Biol.* **52**, 53-60.

Kurakawa T, Ueda N, Maekawa M, Kobayashi K, Kojima M, Nagato Y, Sakakibara H & Kyozuka J (2007) Direct control of shoot meristem activity by a cytokinin-activating enzyme. *Nature* **445**, 652-655.

Langer G, Cohen SX, Lamzin VS & Perrakis A (2008) Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. *Nat Protoc* **3**, 1171-1179.

Lin H, Wang R, Qian Q, Yan M, Meng X, Fu Z, Yan C, Jiang B, Su Z, Li J & Wang Y (2009) DWARF27, an iron-containing protein required for the biosynthesis of strigolactones, regulates rice tiller bud outgrowth. *Plant Cell* **21**, 1512-1525.

Lohar DP, Schaff JE, Laskey JG, Kieber JJ, Bilyeu KD & Bird DM (2004) Cytokinins play opposite roles in lateral root formation, and nematode and Rhizobial symbioses. *Plant J* **38**, 203-214.

Lomin SN, Yonekura-Sakakibara K, Romanov GA & Sakakibara H (2011) Ligand-binding properties and subcellular localization of maize cytokinin receptors. *J Exp Bot* **62**, 5149-5159.

Los GV & Wood K (2007) The HaloTag: a novel technology for cell imaging and protein analysis. *Methods Mol Biol* **356**, 195-208.

Mahonen AP, Bishopp A, Higuchi M, Nieminen KM, Kinoshita K, Tormakangas K, Ikeda Y, Oka A, Kakimoto T & Helariutta Y (2006) Cytokinin signaling and its inhibitor AHP6 regulate cell fate during vascular development. *Science* **311**, 94-98.

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

Matsubayashi Y & Sakagami Y (2006) Peptide hormones in plants. *Annu Rev Plant Biol* **57**, 649-674.

McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC & Read RJ (2007) Phaser crystallographic software. *J Appl Crystallogr* **40**, 658-674.

McGurl B, Pearce G, Orozco-Cardenas M & Ryan CA (1992) Structure, expression, and antisense inhibition of the systemin precursor gene. *Science* **255**, 1570-1573.

Melotto M, Mecey C, Niu Y, Chung HS, Katsir L, Yao J, Zeng W, Thines B, Staswick P, Browse J, Howe GA & He SY (2008) A critical role of two positively charged amino acids in the Jas motif of Arabidopsis JAZ proteins in mediating coronatine- and jasmonoyl isoleucine-dependent interactions with the COI1 F-box protein. *Plant J* **55**, 979-988.

Miller C, Skoog F, Von Saltza M & Strong F (1955) Kinetin, a cell division factor from deoxyribonucleic acid. *J. Am. Chem. Soc.* **77**.

Miller CO (1961) A Kinetin-Like Compound in Maize. *Proc Natl Acad Sci U S A* **47**, 170-174.

Miyawaki K, Tarkowski P, Matsumoto-Kitano M, Kato T, Sato S, Tarkowska D, Tabata S, Sandberg G & Kakimoto T (2006) Roles of Arabidopsis ATP/ADP isopentenyltransferases and tRNA isopentenyltransferases in cytokinin biosynthesis. *Proc. Natl. Acad. Sci. USA* **103**, 16598-16603.

Mockaitis K & Estelle M (2008) Auxin receptors and plant development: a new signaling paradigm. *Annu Rev Cell Dev Biol* **24**, 55-80.

Mok DW & Mok MC (2001) Cytokinin Metabolism and Action. *Annu Rev Plant Physiol Plant Mol Biol* **52**, 89-118.

Morera S, Chiadmi M, LeBras G, Lascu I & Janin J (1995) Mechanism of phosphate transfer by nucleoside diphosphate kinase: X-ray structures of the phosphohistidine intermediate of the enzymes from Drosophila and Dictyostelium. *Biochemistry* **34**, 11062-11070.

Moskal W, Chan A, Cheung F, Xiao Y & Town CD (2008) Direct Submission. *Plant Genomics, J. Craig Venter Institute, 9704 Medical Center Dr, Rockville, MD 20850, USA*.

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

Murai N, Armstrong DJ, Taller BJ & Skoog F (1978) Distribution of incorporated, synthetic cytokinins in ribosomal RNA preparations from tobacco callus. *Plant Physiol.* **61**, 318-322.

Murase K, Hirano Y, Sun TP & Hakoshima T (2008) Gibberellin-induced DELLA recognition by the gibberellin receptor GID1. *Nature* **456**, 459-463.

Murray JD, Karas BJ, Sato S, Tabata S, Amyot L & Szczyglowski K (2007) A cytokinin perception mutant colonized by Rhizobium in the absence of nodule organogenesis. *Science* **315**, 101-104.

Murshudov GN, Skubak P, Lebedev AA, Pannu NS, Steiner RA, Nicholls RA, Winn MD, Long F & Vagin AA (2011) REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr D Biol Crystallogr* **67**, 355-367.

Nambara E & Marion-Poll A (2005) Abscisic acid biosynthesis and catabolism. *Annu Rev Plant Biol* **56**, 165-185.

Nandi S, Letham DS, Palni L, Wong O & Summons R (1989) 6-Benzylaminopurine and its glycosides as naturally occurring cytokinins. *Plant Science* **61**, 189-196.

Nemhauser JL, Mockler TC & Chory J (2004) Interdependency of brassinosteroid and auxin signaling in Arabidopsis. *PLoS Biol* **2**, E258.

Nongpiur R, Soni P, Karan R, Singla-Pareek SL & Pareek A (2012) Histidine kinases in plants: cross talk between hormone and stress responses. *Plant Signal Behav* **7**, 1230-1237.

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

Painter J & Merritt EA (2006) Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. *Acta Crystallogr D Biol Crystallogr* **62**, 439-450.

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

Pasternak O, Bujacz A, Biesiadka J, Bujacz G, Sikorski M & Jaskolski M (2008) MAD phasing using the (Ta6Br12)2+ cluster: a retrospective study. *Acta Cryst. D* **64**, 595-606.

Pasternak O, Bujacz GD, Fujimoto Y, Hashimoto Y, Jelen F, Otlewski J, Sikorski MM & Jaskolski M (2006) Crystal structure of Vigna radiata cytokinin-specific binding protein in complex with zeatin. *Plant Cell* **18**, 2622-2634.

Patterson AL (1934) A Fourier Series Method for the Determination of the Components of Interatomic Distances in Crystals. *Phys. Rev.* **46**, 372-376.

Peng J, Richards DE, Hartley NM, Murphy GP, Devos KM, Flintham JE, Beales J, Fish LJ, Worland AJ, Pelica F, Sudhakar D, Christou P, Snape JW, Gale MD & Harberd NP (1999) 'Green revolution' genes encode mutant gibberellin response modulators. *Nature* **400**, 256-261.

Peng ZY, Zhou X, Li L, Yu X, Li H, Jiang Z, Cao G, Bai M, Wang X, Jiang C, Lu H, Hou X, Qu L, Wang Z, Zuo J, Fu X, Su Z, Li S & Guo H (2009) Arabidopsis Hormone Database: a comprehensive genetic and phenotypic information database for plant hormone research in Arabidopsis. *Nucleic Acids Res* **37**, D975-982.

Perilli S, Moubayidin L & Sabatini S (2010) The molecular basis of cytokinin function. *Curr. Opin. Plant Biol.* **13**, 21-26.

Petoukhov MV, Franke D, Shkumatov AV, Tria G, Kikhney AG, Gajda M, Gorba C, Mertens HDT, Konarev PV & Svergun DI (2012) New developments in the ATSAS program package for small-angle scattering data analysis. *J. Appl. Crystallogr* **45**, 342-350.

Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC & Ferrin TE (2004) UCSF Chimera - a visualization system for exploratory research and analysis. *J Comput Chem* **25**, 1605-1612.

Puttick J, Baker EN & Delbaere LT (2008) Histidine phosphorylation in biological systems. *Biochim Biophys Acta* **1784**, 100-105.

Quezada CM, Hamel DJ, Gradinaru C, Bilwes AM, Dahlquist FW, Crane BR & Simon MI (2005) Structural and chemical requirements for histidine phosphorylation by the chemotaxis kinase CheA. *J Biol Chem* **280**, 30581-30585.

Quittenden LJ, Davies NW, Smith JA, Molesworth PP, Tivendale ND & Ross JJ (2009) Auxin biosynthesis in pea: characterization of the tryptamine pathway. *Plant Physiol* **151**, 1130-1138.

Rayment I (1997) Reductive alkylation of lysine residues to alter crystallization properties of proteins. *Methods Enzymol* **276**, 171-179.

Reid KSC, Lindley PF & Thornton JM (1985) Sulphur-aromatic interactions in proteins. *FEBS* **190**, 209-213.

Ringer AL, Senenko A & Sherrill CD (2007) Models of S/pi interactions in protein structures: comparison of the H2S benzene complex with PDB data. *Protein Sci* **16**, 2216-2223.

Roche P, Debelle F, Maillet F, Lerouge P, Faucher C, Truchet G, Denarie J & Prome JC (1991a) Molecular basis of symbiotic host specificity in Rhizobium meliloti: nodH and nodPQ genes encode the sulfation of lipo-oligosaccharide signals. *Cell* **67**, 1131-1143.

Roche P, Lerouge P, Ponthus C & Prome JC (1991b) Structural determination of bacterial nodulation factors involved in the Rhizobium meliloti-alfalfa symbiosis. *J Biol Chem* **266**, 10933-10940.

Rohrig H, John M & Schmidt J (2004) Modification of soybean sucrose synthase by S-thiolation with ENOD40 peptide A. *Biochem Biophys Res Commun* **325**, 864-870.

Rohrig H, Schmidt J, Miklashevichs E, Schell J & John M (2002) Soybean ENOD40 encodes two peptides that bind to sucrose synthase. *Proc Natl Acad Sci U S A* **99**, 1915-1920.

Rossmann MG (2001) Molecular replacement--historical background. *Acta Crystallogr D Biol Crystallogr* **57**, 1360-1366.

Rupp B (2010). *Biomolecular crystallography : principles, practice, and application to structural biology*. New York: Garland Science.

Ruszkowski M, Brzezinski K, Jedrzejczak R, Dauter M, Dauter Z, Sikorski M & Jaskolski M (2013) Medicago truncatula histidine-containing phosphotransfer protein: Structural and biochemical insights into the cytokinin transduction pathway in plants. *FEBS J* **280**, 3709-3720.

Ruyter-Spira C, Al-Babili S, van der Krol S & Bouwmeester H (2013) The biology of strigolactones. *Trends Plant Sci* **18**, 72-83.

Ryan CA & Pearce G (2003) Systemins: a functionally defined family of peptide signals that regulate defensive genes in Solanaceae species. *Proc Natl Acad Sci U S A* **100 Suppl 2**, 14577-14580.

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

Sakakibara H, Hayakawa A, Deji A, Gawronski SW & Sugiyama T (1999) His-Asp phosphotransfer possibly involved in the nitrogen signal transduction mediated by cytokinin in maize: molecular cloning of cDNAs for two-component regulatory factors and demonstration of phosphotransfer activity in vitro. *Plant Mol Biol* **41**, 563-573.

Sakurai A & Fujioka S (1997) Studies on biosynthesis of brassinosteroids. *Biosci Biotechnol Biochem* **61**, 757-762.

Santner A, Calderon-Villalobos LI & Estelle M (2009) Plant hormones are versatile chemical regulators of plant growth. *Nat Chem Biol* **5**, 301-307.

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

Schaller G MD, Gribskov M, Walker JC (2002) Two-component signaling elements and histidyl-toaspartyl phosphorelays. *C Somerville, E Meyerowitz, The Arabidopsis Book, American Society of Plant Biologists*.

Schmidt A, Teeter M, Weckert E & Lamzin VS (2011) Crystal structure of small protein crambin at 0.48 A resolution. *Acta Crystallogr Sect F Struct Biol Cryst Commun* **67**, 424-428.

Schmidt TG & Skerra A (2007) The Strep-tag system for one-step purification and high-affinity detection or capturing of proteins. *Nat Protoc* **2**, 1528-1535.

Schmulling T (2001) CREam of cytokinin signalling: receptor identified. *Trends Plant Sci* **6**, 281-284.

Schmulling T (2004) Cytokinin. *Encyclopedia of Biological Chemistry (Eds. Lennarz, W., Lane, M.D.)* **Academic Press/Elsevier Science**.

Schwechheimer C (2008) Understanding gibberellic acid signaling--are we there yet? *Curr Opin Plant Biol* **11**, 9-15.

Schwechheimer C & Willige BC (2009) Shedding light on gibberellic acid signalling. *Curr Opin Plant Biol* **12**, 57-62.

Seyedi M, Selstam E, Timko MP & Sundqvist C (2001) The cytokinin 2-isopentenyladenine causes partial reversion to skotomorphogenesis and induces formation of prolamellar bodies and protochlorophyllide657 in the lip1 mutant of pea. *Physiol Plant* **112**, 261-272.

She J, Han Z, Kim TW, Wang J, Cheng W, Chang J, Shi S, Yang M, Wang ZY & Chai J (2011) Structural insight into brassinosteroid perception by BRI1. *Nature* **474**, 472-476.

Sheldrick GM (2008) A short history of SHELX. *Acta Crystallogr A* **64**, 112-122.

Skoog F, Armstrong DJ, Cherayil JD, Hampel AE & Bock RM (1966) Cytokinin activity: localization in transfer RNA preparations. *Science* **154**, 1354-1356.

Slabinski L, Jaroszewski L, Rychlewski L, Wilson IA, Lesley SA & Godzik A (2007) XtalPred: a web server for prediction of protein crystallizability. *Bioinformatics* **23**, 3403-3405.

Soon FF, Ng LM, Zhou XE, West GM, Kovach A, Tan MH, Suino-Powell KM, He Y, Xu Y, Chalmers MJ, Brunzelle JS, Zhang H, Yang H, Jiang H, Li J, Yong EL, Cutler S, Zhu JK, Griffin PR, Melcher K & Xu HE (2012) Molecular mimicry regulates ABA signaling by SnRK2 kinases and PP2C phosphatases. *Science* **335**, 85-88.

Strader LC & Bartel B (2008) A new path to auxin. *Nat Chem Biol* **4**, 337-339.

Svergun D, Barberato C & Koch MHJ (1995) CRYSOL - A program to evaluate x-ray solution scattering of biological macromolecules from atomic coordinates. *Journal of Applied Crystallography* **28**, 768-773.

Svergun D & Koch MH (2003) Small-angle scattering studies of biological macromolecules in solution. *Rep. Prog. Phys.* **66**, 1735-1782.

Svergun DI (1999) Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophys J* **76**, 2879-2886.

Svergun DI, Petoukhov MV & Koch MH (2001) Determination of domain structure of proteins from X-ray solution scattering. *Biophys J* **80**, 2946-2953.

Tan X, Calderon-Villalobos LI, Sharon M, Zheng C, Robinson CV, Estelle M & Zheng N (2007) Mechanism of auxin perception by the TIR1 ubiquitin ligase. *Nature* **446**, 640-645.

Thines B, Katsir L, Melotto M, Niu Y, Mandaokar A, Liu G, Nomura K, He SY, Howe GA & Browse J (2007) JAZ repressor proteins are targets of the SCF(COI1) complex during jasmonate signalling. *Nature* **448**, 661-665.

Tirichine L, Imaizumi-Anraku H, Yoshida S, Murakami Y, Madsen LH, Miwa H, Nakagawa T, Sandal N, Albrektsen AS, Kawaguchi M, Downie A, Sato S, Tabata S, Kouchi H, Parniske M, Kawasaki S & Stougaard J (2006a) Deregulation of a Ca2+/calmodulin-dependent kinase leads to spontaneous nodule development. *Nature* **441**, 1153-1156.

Tirichine L, James EK, Sandal N & Stougaard J (2006b) Spontaneous root-nodule formation in the model legume Lotus japonicus: a novel class of mutants nodulates in the absence of rhizobia. *Mol Plant Microbe Interact* **19**, 373-382.

Tirichine L, Sandal N, Madsen LH, Radutoiu S, Albrektsen AS, Sato S, Asamizu E, Tabata S & Stougaard J (2007) A gain-of-function mutation in a cytokinin receptor triggers spontaneous root nodule organogenesis. *Science* **315**, 104-107.

Ueguchi-Tanaka M, Ashikari M, Nakajima M, Itoh H, Katoh E, Kobayashi M, Chow TY, Hsing YI, Kitano H, Yamaguchi I & Matsuoka M (2005) GIBBERELLIN INSENSITIVE DWARF1 encodes a soluble receptor for gibberellin. *Nature* **437**, 693-698.

Umezawa T, Sugiyama N, Mizoguchi M, Hayashi S, Myouga F, Yamaguchi-Shinozaki K, Ishihama Y, Hirayama T & Shinozaki K (2009) Type 2C protein phosphatases directly regulate abscisic acid-activated protein kinases in Arabidopsis. *Proc Natl Acad Sci U S A* **106**, 17588-17593.

Van Huijsduijnen RAMH, Alblas SW, De Rijk RH & Bol JF (1986) Induction by Salicylic Acid of Pathogenesis-related Proteins and Resistance to Alfalfa Mosaic Virus Infection in Various Plant Species. *Journal of General Virology* **67**, 2135.

Vieten A, Sauer M, Brewer PB & Friml J (2007) Molecular and cellular aspects of auxin-transport-mediated development. *Trends Plant Sci* **12**, 160-168.

Vlot AC, Klessig DF & Park SW (2008) Systemic acquired resistance: the elusive signal(s). *Curr Opin Plant Biol* **11**, 436-442.

Walter TS, Meier C, Assenberg R, Au KF, Ren J, Verma A, Nettleship JE, Owens RJ, Stuart DI & Grimes JM (2006) Lysine methylation as a routine rescue strategy for protein crystallization. *Structure* **14**, 1617-1622.

Wasson AP, Pellerone FI & Mathesius U (2006) Silencing the flavonoid pathway in Medicago truncatula inhibits root nodule formation and prevents auxin transport regulation by rhizobia. *Plant Cell* **18**, 1617-1629.

Wasternack C (2007) Jasmonates: an update on biosynthesis, signal transduction and action in plant stress response, growth and development. *Ann Bot* **100**, 681-697.

Went FW (1938) Specific Factors Other Than Auxin Affecting Growth and Root Formation. *Plant Physiol* **13**, 55-80.

Went FW (1942) Growth, Auxin, and Tropisms in Decapitated Avena Coleoptiles. *Plant Physiol* **17**, 236-249.

West AH & Stock AM (2001) Histidine kinases and response regulator proteins in two-component signaling systems. *Trends Biochem Sci* **26**, 369-376.

Winn MD, Isupov MN & Murshudov GN (2001) Use of TLS parameters to model anisotropic displacements in macromolecular refinement. *Acta Cryst. D* **57**, 122-133.

Woodward AW & Bartel B (2005) Auxin: regulation, action, and interaction. *Ann Bot* **95**, 707-735.

Wulfetange K, Lomin SN, Romanov GA, Stolz A, Heyl A & Schmulling T (2011) The cytokinin receptors of Arabidopsis are located mainly to the endoplasmic reticulum. *Plant Physiol* **156**, 1808-1818.

Xie DX, Feys BF, James S, Nieto-Rostro M & Turner JG (1998) COI1: an Arabidopsis gene required for jasmonate-regulated defense and fertility. *Science* **280**, 1091-1094.

Yamaguchi S (2008) Gibberellin metabolism and its regulation. *Annu Rev Plant Biol* **59**, 225-251.

Yang H, Matsubayashi Y, Nakamura K & Sakagami Y (1999) Oryza sativa PSK gene encodes a precursor of phytosulfokine-alpha, a sulfated peptide growth factor found in plants. *Proc Natl Acad Sci U S A* **96**, 13560-13565.

Young ND, Debelle F, Oldroyd GE, Geurts R, Cannon SB, Udvardi MK, Benedito VA, Mayer KF, Gouzy J, Schoof H, Van de Peer Y, Proost S, Cook DR, Meyers BC, Spannagl M, Cheung F, De Mita S, Krishnakumar V, Gundlach H, Zhou S, Mudge J, Bharti AK, Murray JD, Naoumkina MA, Rosen B, Silverstein KA, Tang H, Rombauts S, Zhao PX, Zhou P, Barbe V, Bardou P, Bechner M, Bellec A, Berger A, Berges H, Bidwell S, Bisseling T, Choisne N, Couloux A, Denny R, Deshpande S, Dai X, Doyle JJ, Dudez AM, Farmer AD, Fouteau S, Franken C, Gibelin C, Gish J, Goldstein S, Gonzalez AJ, Green PJ, Hallab A, Hartog M, Hua A, Humphray SJ, Jeong DH, Jing Y, Jocker A, Kenton SM, Kim DJ, Klee K, Lai H, Lang C, Lin S, Macmil SL, Magdelenat G, Matthews L, McCorrison J, Monaghan EL, Mun JH, Najar FZ, Nicholson C, Noirot C, O'Bleness M, Paule CR, Poulain J, Prion F, Qin B, Qu C, Retzel EF, Riddle C, Sallet E, Samain S, Samson N, Sanders I, Saurat O, Scarpelli C, Schiex T, Segurens B, Severin AJ, Sherrier DJ, Shi R, Sims S, Singer SR, Sinharoy S, Sterck L, Viollet A, Wang BB, et al. (2011) The Medicago genome provides insight into the evolution of rhizobial symbioses. *Nature* **480**, 520-524.

Zhao X, Copeland DM, Soares AS & West AH (2008) Crystal structure of a complex between the phosphorelay protein YPD1 and the response regulator domain of SLN1 bound to a phosphoryl analog. *J Mol Biol* **375**, 1141-1151.

Zhou XE, Soon FF, Ng LM, Kovach A, Suino-Powell KM, Li J, Yong EL, Zhu JK, Xu HE & Melcher K (2012) Catalytic mechanism and kinase interactions of ABA-signaling PP2C phosphatases. *Plant Signal Behav* **7**, 581-588.

Zu XL, Besant PG, Imhof A & Attwood PV (2007) Mass spectrometric analysis of protein histidine phosphorylation. *Amino Acids* **32**, 347-357.