# Structure prediction, evolution and ligand interaction of CHASE domain

Jakub Pas[a,c,*], Marcin von Grotthuss[a], Lucjan S. Wyrwicz[b], Leszek Rychlewski[a], Jan Barciszewski[c]

[a] *BioInfoBank Institute, ul. Limanowskiego 24A, 60-744 Poznan, Poland*
[b] *Bioinformatics Unit, Department of Physics, Adam Mickiewicz University, ul. Umultowska 85, 61-614 Poznan, Poland*
[c] *Institute of Bioorganic Chemistry, Polish Academy of Sciences Noskowskiego 12, 61-704 Poznan, Poland*

Received 22 June 2004; revised 7 September 2004; accepted 7 September 2004

Available online 25 September 2004

Edited by Takashi Gojobori

**Abstract Cytokinins are plant hormones involved in the essential processes of plant growth and development. They bind with receptors known as CRE1/WOL/AHK4, AHK2, and AHK3, which possess histidine kinase activity. Recently, the sensor domain cyclases/histidine kinases associated sensory extracellular (CHASE) was identified in those proteins but little is known about its structure and interaction with ligands. Distant homology detection methods developed in our laboratory and molecular phylogeny enabled the prediction of the structure of the CHASE domain as similar to the photoactive yellow protein-like sensor domain. We have identified the active site pocket and amino acids that are involved in receptor–ligand interactions. We also show that fold evolution of cytokinin receptors is very important for a full understanding of the signal transduction mechanism in plants.**


*Keywords:* Cytokinin; Structure prediction; Ligand interaction; CHASE

## 1. Introduction

Early studies of cytokinin receptor suggested that cytokinins are recognized by proteins of the two component system [1]. First results indicated that overexpression of histidine kinase CKI1 produces cytokinin-like effect in plant tissues [1]. Subsequent experiments have shown that CKI1 is rather constitutively active and normally expressed only in female gametophytes [2]. The recent identification of CRE1 [3] – a histidine kinase identical to AHK4 [4] and WOL [5] as the cytokinin receptor of *Arabidopsis thaliana* – represents a milestone in cytokinin research. The mutations that cause phenotype of loss-of-function identified in CRE1 shed new light on the role of cytokinins in plant development. Low cytokinin concentrations were sufficient for complementation in the fission yeast and *Escherichia coli* assays [6,7], indicating that the receptor is responsive to the cytokinins at physiological level.

* Corresponding author. Fax: +48-618520532.
*E-mail address:* kuba@bioinfo.pl (J. Pas).

*Abbreviations:* ACT, aspartate kinase–chorismate mutase–TyrA; CACHE, $Ca^{2+}$ channels and CHEmotaxis receptors; CHASE, cyclases/histidine kinases associated sensory extracellular; GAF, cGMP phosphodiesterase, adenyl cyclase, FhlA domain; GRDB, gene related data base; PAS, periodic clock protein, aryl hydrocarbon receptor and single-minded protein; PDB, protein data bank; PYP, photoactive yellow protein

CRE1 is mapped to chromosome 2 and codes for a histidine kinase gene. One predicted form of CRE1 is a 1057 amino acid protein and the second form carries 23 additional amino acids at the N-termini. An analysis of the predicted polypeptide sequence reveals the histidine kinase domain on C terminus followed by two receiver domains responsible for signal transduction. N terminal 277 amino acids long domain is suspected to bind ligands. It contains two predicted hydrophobic membrane spanning regions. Mutation of the threonine 278 in this domain is responsible for the loss of function [5].

Protein domains organization indicates that the cytokinin receptor protein initiates intracellular phosphotransfer and it is a part of two-component regulatory systems, which are common in prokaryotes, lower eukaryotes and plants [8]. The name of the system is derived from the standard configuration in bacteria, where a sensor kinase directly activates the second component, a response regulator, which regulates transcription of the target genes.

Recently, two independent groups of researchers identified the sensor domain of the cytokinin receptor [9,10]. It was named the cyclases/histidine kinases associated sensory extracellular (CHASE) domain because of its presence in diverse receptor-like proteins with histidine kinase and nucleotide cyclase domains. The CHASE domain is an extracellular domain of 200–230 amino acids, which is found in transmembrane receptors from bacteria, lower eukaryotes and plants. The domain always occurs N-terminally in extracellular or periplasmic locations, followed by an intracellular tail housing diverse enzymatic signaling domains such as histidine kinase, adenyl cyclase, GGDEF-type nucleotide cyclase and EAL-type phosphodiesterase domains, as well as non-enzymatic domains such periodic clock protein, aryl hydrocarbon receptor and single-minded protein (PAS), GAP, phosphohistidine and response 1regulatory domains. The CHASE domain is predicted to bind diverse low molecular weight ligands, such as the cytokinin-like adenine derivatives or peptides, and mediate signal transduction through the respective receptors.

The CHASE has a predicted a + b fold, with two extended helices on both boundaries and two central helices separated by sheets. The termini are less conserved compared with the central part of the domain. The central part contains strongly conserved motifs.

Although the CHASE domain was characterized by the mechanism of ligand binding, evolutionary relationships between other sensing domains remains unrevealed. Using sensitive distant homology detection methods and molecular modeling show structural properties of CHASE with their ligands.

0014-5793/$22.00 
doi:10.1016/j.febslet.2004.09.020

### 1.1. Structure prediction of the CHASE domain

Using standard methods such as BLAST, it is currently impossible to find homologs of the CHASE domain with known structure. As shown before, sensing domains may become very diverse during evolution [11,12] but still possess the same fold, which sometimes can be detected using more sensitive tools. Latest evaluation of fold recognition algorithms [13,14] shows that the recently developed fully automated methods [15] greatly increase the prediction sensitivity and accuracy.

We applied the gene relate sequence database (GRDB) [16] to identify distant homologs of the CHASE domain. The GRDB system contains the characteristic profiles computed for many protein families collected from Pfam, cluster of ortholog groups, the protein data bank 7 (PDB7) and from other genomic sources. The system facilitates the comparison of the target family with about 100 000 other families, using the Meta-BASIC [17] program. Meta-BASIC, unlike many other methods used in fold recognition, does not require any information about any native structure to draw the comparison.

We searched PDB database to find distant homologs with known tertiary structure. The sequence of CRE1a protein without extra amino acids at the N-terminus was used as a query. Two related structures were found: the structure of the sensory domain of fumarate sensor Dcus from *E. coli* (PDB code: IojgA) and the structure of the periplasmic ligand-binding domain of the kinase Cita sensor from *Klebsiella pneumoniae* (PDB code: Ip0zA). Both proteins possess sensor kinase activity and are parts of a two-component system [18,19] as is the CHASE domain. They are structurally classified as the PAS-like fold. These results confirm the assumptions that the CHASE, the $Ca^{2+}$ channels and CHEmotaxis receptors (CACHE) and the PAS domains are phylogenetically related [11]. Multiple sequence alignment between the consensus sequence of the CHASE domain and the two families shows conservation of sequence patterns and secondary structure despite the low amino acid identity (Fig. 1). Molecular modeling and docking were performed to examine the possibility of ligand binding. Information about ligands bound to related proteins was used as initial guide. The structure of one of the homologs identified by GRDB contained a ligand (Ip0zA). We performed additional database search using the 3D-Hit program [20] to find structurally similar domains in complex with ligands.

Using 1jogA structure as a query, we discovered two other structures: regulatory segment of mouse 3′, 5′-cyclic nucleotide phosphodiesterase (PDB code: 1mc0a) complexed with cyclic guanosine monophosphate and a more diverse photoactive yellow protein (PYP) bound with hydroxycinnamic (PDB code: 1.f98). Both of them bind ligand in a pocket in the same manner that is shared among many PAS domains with some differences according to the chemical properties of the bound compound. To discover the nature of such a variation, we performed molecular modeling with the MODELER program [21] of CRE1 on the basis of the alignment generated by GRDB and performed molecular docking of cytokinin compound to receptor.

To evaluate the local environment and inter-residue contacts in model (Fig. 4), we used the Verify3D program [22]. The Verify3D program provides assessment of structures on the residue level, which enables the user to locate parts of the protein that are likely to have the correct conformation or to look for misfolded regions. The program uses information about the local secondary structure, solvent accessibility and the fraction of side-chain area that is covered by polar atoms. We decided to take advantage of this feature and apply Verify3D to improve incorrectly folded model sections. The alignment in regions poorly scored by Verify3D was corrected. This idea of verification and correction of the model fragments was successfully used by our group in critical assessment of techniques for protein structure prediction 5 experiment [16].

We superimposed the model and related structures with their ligand. As it is known, crel binds *trans*-zeatin [3,23]. We manually docked the *trans*-zeatin molecule in the pocket in the same manner as other ligands in their structures. The docked ligand was entirely buried. None of its atoms is available outside the protein. Molecular modeling confirms the importance of threonine 278. It is located at the entry of the sensing pocket and is at a contact distance from ligand (Fig. 2). The role of this amino acid in the function loss of the protein may

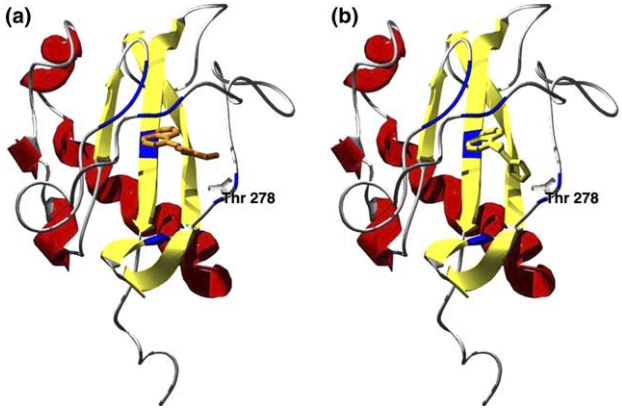


Fig. 2. Molecular model of CRE1a receptor from *A. thaliana* docked with (a) *trans*-zeatin, (b) kinetin. The yellow and red colors indicate strands and helices, respectively. Ligand is colored in orange. The part of the chain whose residues are at contact distance with ligand is colored in blue. The visible side chain belongs to threonine 278 whose mutation is responsible for loss of function.

```
           1    .   10    .   20    .   30    .   40    .   50    .   60    .   70    .   80    .   90    .  100    .  110
CHASE score SPSAIDQETFAEYTARTSFERPLLSGVAYAQKVTHSEREQFERQAGWVIKKMYQIENIPAPAQDEYAPVIFAQPSSNSHVIGLDMMSGEERRENILRARASGKAVLTSPFR
1jogA 13.71 --SDMTRDGLANKALAVARTLADSPEIRQGLQKKPQESGIQAIAEAVRKRNDLLFIVVTDMQSLRYS--------HPEAQRIGQPFKGDDILK----ALNGEENVAINRGF
1p0zA 08.56 ---------------PHDAWLITRGLSTLNIRMKAESENAMKVAEYLKSHPAVEKVYYPGFEDHEGHDIAKKQMRMYGSMITFILKSGFEGAKKLLDNLKLITLAVS----
1e4eA 09.36 EERLHYQVGQRALIQAMQISAMPELVEAVQKRDLARIKALIDPMRSFSDATYITVGDASGQRLY--------HVNPDEIGKSMEGGDSDE-------ALINAKSYVS---V
1ehiA 08.68  -------EDGSIQGLFELSGIPFVGCDIQSSAICMDKSLTYIVAKNAGIATPAFWVINPVAATFTYPV-FVKPARSGSSFGVKKVNSADEDYAIESARQYDSKI------

      100 .  120    .  130    .  140    .  150    .  160    .  170    .  180    .  190    .  200    .  210    .  220    .
CHASE score LLQTNHRGVILTFPVYKSSLPSNPTVEERQEATIGFLGGAFDVESLVESLLQQLASNQSIDVEVYDVTNTSSPIVFYGNQVSAEDLSQPHZLFFGDPFRKHEMA-------
1jogA 13.71 L----AQALRVFTPIYDEN----------HKQIGVVAIGLELSRVTQQIND-----------------------------------------------------------
1p0zA 08.56 -----LGGCESLIQHPASMTHAVVPKEEREAAGITDGMIRLSVDADELIADFKQGL-----------------------------------------------------
1e4eA 09.36 RKGSLGSSLRGKSPIQDAT----------GKVIGIVSVGYTIEQL-----------------------------------------------------------------
1ehiA 08.68 LIEQAVSGCEVGCAVLGNSAAL---------------------------------------------------------------------------------------
```

Fig. 1. Alignment performed by GRDB. Predicted secondary structure of CHASE domain agrees with secondary structure of proteins of known three dimensional structure. Helices are colored red, sheets in blue. 1jogA – sensory domain of the membraneous two-component fumarate sensor Dcus of *E. coli*, 1pOzA – sensor kinase cita. 1e4eA and 1ehiA – structures of D-alanine-D-lactate ligase.

Table 1
Selected structures of sensing domains with their ligands detected by 3D-Hit as related to loga sorted by score

| Domain | 3D-Hit score | PDB code | Function | Organism | Ligand |
|---|---|---|---|---|---|
| CACHE | 84.3 | Ip0z | $Ca^{2+}$ channels and chemotaxis receptors | *K. pneumoniae* | Citrate anion |
| GAF | 55.6 | Imc0 | cGMP phosphodiesterase, adenyl cyclase, FhlA domain | *M. musculus* | Cyclic guanosine monophosphate |
| PAS/PYP | 47.2 | If98 | Periodic clock protein, aryl hydrocarbon receptor and single-minded protein/photoactive yellow protein | *E. halophila* | 4′-Hydroxycinnamic acid |
| Profiln | 46.2 | Ig5u | monomeric actin binding | *H. brasiliensis* | Actin |
| ACT | – | IpsdAl | Aspartate kinase–chorismate mutase–Tyr A | *E. coli* | Nicotinamide-adenine-dinucleotide |

CACHE, GAP, and PAS/PYP belong to PAS-like fold. Profilin is more diverged although related to PAS. ACT sensing domain in contrast is diverged beyond recognition.

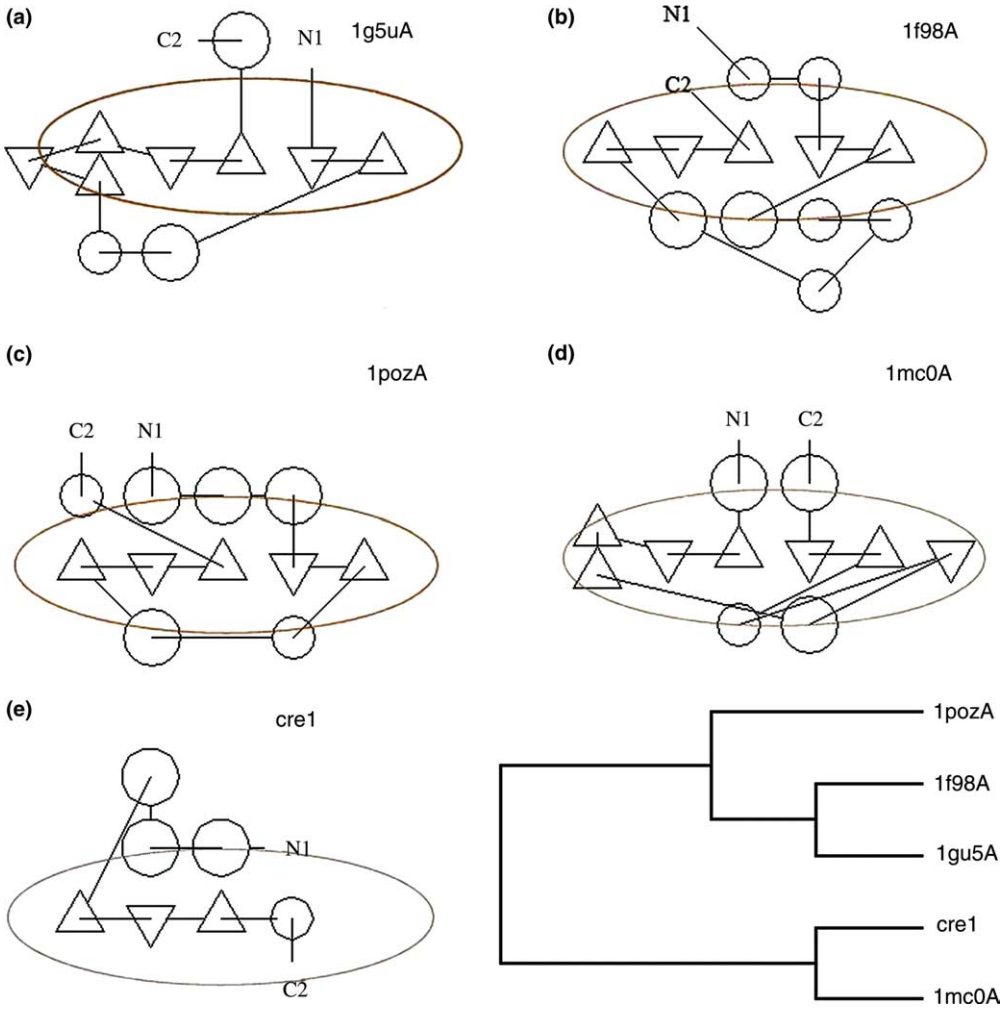


Fig. 3. Topological representation and fold evolution of PYP-like family. The cartoon representation of topology of selected related sensing domains of: (a) Profilin; (b) PAS; (c) CACHE; (d) GAP; and (e) CHASE. The triangular symbols represent beta strand, circular symbols represent helices. The direction information for strands is indicated by up and down pointing triangles. Common structural motif in all structures is circled red. Cartoons were performed using the TOPS [28] program. Dendrogram is calculated using PRIDE [24] based on the measure of the degree of similarity between proteins and shows distances between structurally diverged 3D structures. The evolutionary distant sensing domain ACT was used as an outgroup.

be due to the ligand interaction or/and formation of entry to the binding pocket.

Additional amino acids Arg 235, Asp 236, Glu 237,Leu 254, Asp 255, Leu 277, Ser 276, His 288, and Gly 160 are in contact with ligand and are highly conserved among all plant cytokinin receptors (data not shown).

Because of the low sequence similarity between the proteins, it is difficult to perform phylogenetic analysis. The evolutionary distance between domains may be approximated using structural similarity. We analyzed the domains discovered by the 3D-Hit program (Table 1) and the model of the CHASE domain using the PRIDE server [24]. The server calculates

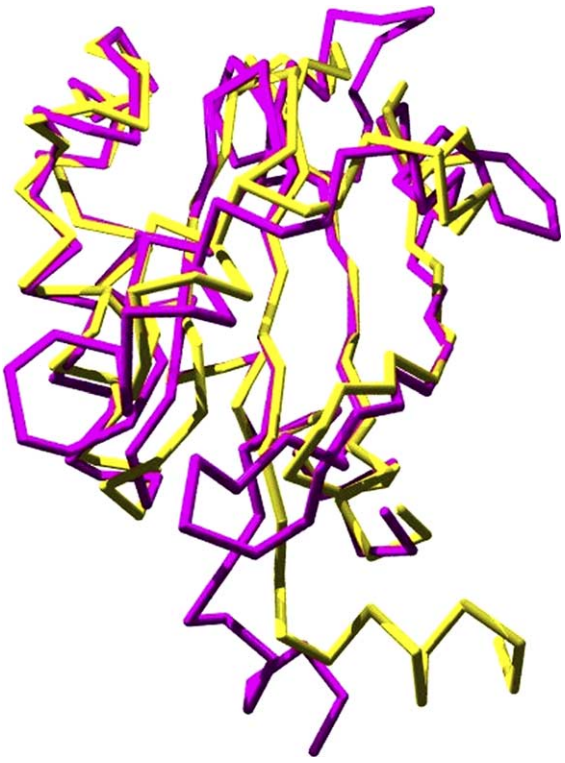

Fig. 4. Superimposition of model (pink) of crel receptor and template structure 1jogA (yellow). The root mean square deviation between structures is 2.4. The main differences between model and template are in the loop regions.

dendrogram, which shows how the structures are grouped by their structure similarity. Calculations confirm that PAS-like fold is the framework for CHASE as well as for PAS [25,26] and GAP [27] domains. On the structural basis, CRE1 and Imc0 are located on the same clade. This may indicate that CACHE and GAP domains have common origin. Despite the fact that crel is lacking in a strand (Fig. 3e), they are structurally much more similar to one another than to other members of PAS/PYP superfamily. The ligands (Cyclic guanosine monophosphate and *trans*-zeatin) bound by both domains are also structurally similar. CACHE, PAS/PYP and Profilin form a separate branch, which indicates that these two groups separated very early in the course of evolution.

It seems that most of the protein–ligand interaction is carried out by a small number of closely related receptors which are diverse in sequence due to the nature of ligands they bind, but share a common structure. Moreover, relationships between evolutionary distant ligand-protein sets may be detected using sensitive tools. The description of receptor-associated ligand binding domain helps in better understanding of cytokinin signaling in plants, and also to understand the evolution of signaling components.

*Acknowledgements:* M.V.G. is a fellow of the Foundation for Polish Science.

## References

[1] Kakimoto, T. (1996) CKI1, a histidine kinase homolog implicated in cytokinin signal transduction. Science 274 (5289), 982–985.
[2] Pischke, M.S. et al. (2002) An arabidopsis histidine kinase is essential for megagametogenesis. Proc. Natl. Acad. Sci. USA 99 (24), 15800–15805.
[3] Inoue, T. et al. (2001) Identification of CRE1 as a cytokinin receptor from Arabidopsis. Nature 409 (6823), 1060–1063.
[4] Taniguchi, M. et al. (1998) Expression of Arabidopsis response regulator homologs is induced by cytokinins and nitrate. FEBS Lett. 429 (3), 259–262.
[5] Mahonen, A.P. et al. (2000) A novel two-component hybrid molecule regulates vascular morphogenesis of the Arabidopsis root. Genes Dev. 14 (23), 2938–2943.
[6] Mizuno, T. (1998) His-Asp phosphotransfer signal transduction. J. Biochem. (Tokyo) 123 (4), 555–563.
[7] Suzuki, T. et al. (2001) The Arabidopsis sensor His-kinase, AHk4, can respond to cytokinins. Plant Cell. Physiol. 42 (2), 107–113.
[8] Parkinson, J.S. and Kofoid, E.C. (1992) Communication modules in bacterial signaling proteins. Annu. Rev. Genet. 26, 71–112.
[9] Mougel, C. and Zhulin, I.B. (2001) CHASE: an extracellular sensing domain common to transmembrane receptors from prokaryotes, lower eukaryotes and plants. Trends Biochem. Sci. 26 (10), 582–584.
[10] Anantharaman, V. and Aravind, L. (2001) The CHASE domain: a predicted ligand-binding module in plant cytokinin receptors and other eukaryotic and bacterial receptors. Trends Biochem. Sci. 26 (10), 579–582.
[11] Anantharaman, V. et al. (2001) Regulatory potential, phyletic distribution and evolution of ancient, intracellular small-molecule-binding domains. J. Mol. Biol. 307 (5), 1271–1292.
[12] Aravind, L. et al. (2002) Trends in protein evolution inferred from sequence and structure analysis. Curr. Opin. Struct. Biol. 12 (3), 392–399.
[13] Kinch, L.N. et al. (2003) CASP5 assessment of fold recognition target predictions. Proteins 53 (Suppl 6), 395–409.
[14] Rychlewski, L. et al. (2003) LiveBench-6: large-scale automated evaluation of protein structure prediction servers. Proteins 53 (Suppl 6), 542–547.
[15] Bujnicki, J.M. et al. (2001) Structure prediction meta server. Bioinformatics 17 (8), 750–751.
[16] von Grotthuss, M. et al. (2003) Application of 3D-Jury, GRDB, and Verify3D in fold recognition. Proteins 53 (Suppl 6), 418–423.
[17] Ginalski, K. et al. (2004) Detecting distant homology with Meta-BASIC. Nucleic Acids Res. 32 (Web Server Issue), W576–W581.
[18] Pappalardo, L. et al. (2003) The NMR structure of the sensory domain of the membranous two-component fumarate sensor (histidine protein kinase) DcuS of *Escherichia coli*. J. Biol. Chem. 278 (40), 39185–39188.
[19] Reinelt, S. et al. (2003) The structure of the periplasmic ligand-binding domain of the sensor kinase cita reveals the first extracellular pas domain. J. Biol. Chem. 278 (40), 39189–39196.
[20] Plewczynski, D., et al. (2002) 3D-Hit, fast structural comparison of proteins. in: Poster on 10th International Conference on Intelligent Systems For Molecular Biology.
[21] Sali, A. and Blundell, T.L. (1993) Comparative protein modelling by satisfaction of spatial restraints. J. Mol. Biol. 234 (3), 779–815.
[22] Luthy, R. et al. (1992) Assessment of protein models with three-dimensional profiles. Nature 356 (6364), 83–85.
[23] Yamada, H. et al. (2001) The Arabidopsis AHK4 histidine kinase is a cytokinin-binding receptor that transduces cytokinin signals across the membrane. Plant Cell Physiol. 42 (9), 1017–1023.
[24] Carugo, O. and Pongor, S. (2002) Protein fold similarity estimated by a probabilistic approach based on C(alpha)-C(alpha) distance comparison. J. Mol. Biol. 315 (4), 887–898.
[25] Ponting, C.P. and Aravind, L. (1997) PAS: a multifunctional domain family comes to light. Curr. Biol. 7 (11), R674–677.
[26] Zhulin, I.B. et al. (1997) PAS domain S-boxes in Archaea, Bacteria and sensors for oxygen and redox. Trends Biochem. Sci. 22 (9), 331–333.
[27] Aravind, L. and Ponting, C.P. (1997) The GAP domain: an evolutionary link between diverse phototransducing proteins. Trends Biochem. Sci. 22 (12), 458–459.
[28] Michalopoulos, I. et al. (2004) TOPS: an enhanced database of protein structural topology. Nucleic Acids Res. 32, D251–D254, Database issue.