International Journal of Biological, Life and Agricultural Sciences
ISSN: 2415-6612
Vol:3, No:6, 2009

Computational Analysis of the Membrane Targeting Domains of Plant-specific PRAF Proteins

Ewa Wywial and Shaneen M. Singh

***Abstract*—**The PRAF family of proteins is a plant specific family of proteins with distinct domain architecture and various unique sequence/structure traits. We have carried out an extensive search of the *Arabidopsis* genome using an automated pipeline and manual methods to verify previously known and identify unknown instances of PRAF proteins, characterize their sequence and build 3D structures of their individual domains. Integrating the sequence, structure and whatever little known experimental details for each of these proteins and their domains, we present a comprehensive characterization of the different domains in these proteins and their variant properties.

***Keywords*—**PRAF proteins, homology modeling, *Arabidopsis thaliana*.

I. INTRODUCTION

THE PRAF family of proteins consist of a Pleckstrin Homology (PH) domain, one or more regulator of chromosome condensation 1 (RCC1) repeats, a variant Fab1p, YOTB, Vac1 and EEA1 (FYVE) domain and a Brevis radix (BRX , also referred to as the DZC domain) [1, 2].

The first PRAF protein, named PRAF-1 (i.e. AT1G65920), was isolated from *Arabidopsis thaliana* in 2002 [3]. PRAF-1 has been shown to be present at high levels only in flowers and not any other reported tissue [3]. The PH domain of PRAF-1 has been shown to preferentially bind phosphatidylinositol 4,5-bisphosphate (PtdIns(4,5)P_2) [4]. Sequence analysis shows that PRAF-1 PH domain is very similar to the mammalian PLCδ1 PH domain [3], which is known to play a role in Ca^{2+} uptake in liver mitochondria [5] and bind PtdIns(4,5)P_2 [6-8]. Not much is known about other members of this protein family. In general, PRAF proteins have been found only in plants where they are thought to be localized within or near the nucleus [9-12]. Their overall function remains largely unknown. The signature domains of this family are traditionally implicated in membrane-targeting (PH and FYVE; [13-19]), protein-protein interactions (BRX; [1]) and mitotic spindle assembly (RCC1; [20]).

We have carried out an extensive search of the *Arabidopsis* genome using an automated pipeline and manual methods to verify previously known and identify unknown instances of PRAF proteins, characterize their sequence and build 3D structures of their individual domains. Integrating the sequence, structure and whatever little known experimental details for each of these proteins and their domains, we present a domain-based comprehensive characterization of *Arabidopsis* PRAF proteins. Our study provides the first glimpse into the role(s) that PRAF proteins play in *Arabidopsis* and how it compares to their role in other plants.

Authors are with the Brooklyn College CUNY, USA.
e-mail: ssingh@brooklyn.cuny.edu

II. METHODS

A. Arabidopsis *PRAF proteins*

The *Arabidopsis* PRAF proteins were identified by database searches and via their constituent domains using a computational pipeline for automated high-throughput modeling [21], which was run against the *Arabidopsis* protein sequence database (TAIR6_pep_20051108) using the coordinates of known structures from the protein data bank (PDB) [22, 23]. The *Arabidopsis* PRAF protein sequences corresponding to the identified accession numbers were retrieved from KEGG GENES [24, 25] and verified for presence of PH, RCC1 and FYVE domains with SMART [26-28].

B. *Sequence verification*

To verify the total number and individual accession numbers of *Arabidopsis* PRAF proteins obtained with pipeline, three additional methods were employed: 1) search of publicly available sequence databases: Swiss-Prot/TrEMBL [29, 30], NCBI [31, 32], and UniProt [33-35]; 2) search of the TAIR database, which maintains a contains genetic and molecular biology data for *Arabidopsis* only [36]; and 3) MOTIF search [25] using a manually derived pattern specific to *Arabidopsis* PRAF proteins in PROSITE format offered by GenomeNet service [25].

C. *Domain architecture analyses and classification*

Each *Arabidopsis* PRAF protein sequence was analyzed for its constituent domains and how they are organized by searching against Pfam [37], SMART [26-28], Conserved Domain Database v2.10 using conserved domain (CD)-Search [38-42] and Clusters of Orthologous Groups [43, 44], and then subgrouped according to the derived consensus domain architecture.

D. Verification of domain boundaries

The boundaries of the *Arabidopsis* PH, RCC1 and FYVE domain sequences were ascertained based on the consensus output from a number of secondary structure prediction programs: Jpred [45-47], PSIPRED [48-50], PHDsec [51], Prof [52], SAM-T06 [53, 54], SAM-T99 [55], Hierarchical Neural Network [56] and SSPro [57, 58]. Protein secondary structure prediction is an important step towards understanding how proteins fold in 3D. Recently, the combination of multiple sequence family-based data with sophisticated algorithms and computing techniques has significantly improved accuracy of secondary structure predictions, for example prediction accuracies of up to 81.5% have been achieved with Jpred3 [59] and 79.5% with SPINE [60]. To validate our secondary structure prediction approach we have predicted secondary structure elements of solved PH and FYVE domains deposited in the PDB [22, 23]. Following the initial validation, secondary structure prediction was performed for all PH, RCC1 and FYVE domain sequences. This step was performed to ensure that the sequences encompass all of secondary structure elements characteristic of PH, RCC1 and FYVE domains completely as domain classification programs can often be inaccurate in defining the boundaries of the domains. The consensus secondary structure prediction for each sequence was also used to verify the accuracy of template–target alignments used in modeling their 3D structure.

E. Molecular modeling of PH and FYVE domains

There is no single modeling program that produces reliable models for all sequences at all times [61]. Therefore to generate high-quality models for the domains present in the *Arabidopsis* PRAF proteins, we have used a number of programs to create many different alternative alignments and models followed by a quality assessment and selection process. We used two separate approaches: automated and manual. The automated approach is based on the previously mentioned computational pipeline with its own built-in alignment, modeling and evaluation methods [21] and the web server Pudge [62]. The automated pipeline takes in a coordinate file in PDB format, extracts the protein's amino acid sequence, identifies homologous sequences, builds 3D models for each sequence and assesses their quality (see Mirkovic *et al.,* [21] for a detailed description). The manual approach was based on a scheme previously applied by Singh and Murray [63]. The scheme involves the use of multiple programs at each step: 1) choice of a suitable structural template, 2) alignment of the template and target sequences, 3) model building, and 4) model evaluation and refinement. Loop refinement and side chain packing optimization was performed using individual modeling programs whenever available and additionally, loop refinement was done with Loopy [64] and the prediction of side-chain conformations with SCWRL3.0 [65] and SCAP [66-68].

1) Structural templates

Currently, there are over twenty non-redundant structures of PH domains [69] and over five non-redundant structures of FYVE domains available in the PDB database. All templates were identified as suitable structural templates for the respective sequences based on: 1) fold recognition as implemented by 123D+ [70], FUGUE [71], LOOPP [72-74], and PHYRE [75]; 2) sequence and structure homology via BLAST and PSI-BLAST [36, 76-78], 3D-JIGSAW [79-81], automatic and manual Homology Modeling with HOMER [82, 83], and CPH [84]; 3) manual multiple sequence alignment using the alignment editor, GeneDoc [85], and 4) Rosetta *ab initio* modeling [86-92].

All results were scrutinized for normalized rank scores or statistical parameters such as Z-scores or E-values and also for percentage coverage of target sequence and accepted only if they were significant and of sufficient length to model the entire domain.

2) Sequence alignment

Alignments were generated using the programs 123D+ [70], FUGUE [71], LOOPP [72-74], PHYRE [75], 3D-JIGSAW [79-81], HOMER [82, 83], CPH [84], and manually edited in certain cases using GeneDoc [85].

Once generated, the alignments were assessed and manually edited to ensure the correspondence of the positions of the secondary structure elements of the template with the predicted consensus secondary structure assignments for the target sequence. This alignment refinement was performed iteratively in conjunction with model evaluation to judge the effect of changes made to the alignments.

3) Model building

The quality of comparative models depends on the closeness of the evolutionary relationship on which they are based [93]. It has been established that the pairwise sequence identity of the protein sequence with its known template should be greater than 30% and 80 or over residues long for the straightforward homology approach to perform with good or high accuracy. This cutoff is five percentage points above the threshold for structural homology [94], in an attempt to provide high-quality multiple homology models. In cases with sufficient pairwise sequence similarity valuable 3D models of the protein sequence can be constructed by routine homology modeling methods. In cases with low sequence identity to known proteins, fold recognition and *ab initio* methods have to be used and more user input is required compared to routine homology modeling. Fold recognition methods identify the likely protein fold even in cases where there is no clear sequence homology and *ab initio* methods build 3D protein models based on first principles rather than structural information available from known structures. It has been shown that predictions derived from a consensus of different methods can reach accuracy as high as 80% [95]. To maximize the reliability of the generated models, many different model-building programs were used in the proposed study. The following is a list of the programs: 3D-JIGSAW [79-81], Modeller 8v1 [96-99], NEST [100], LOOPP [72-74], HOMER [82, 83], CPH [84], and PHYRE [75]. 3D-JIGSAW and NEST are based on rigid-body assembly method while Modeller uses modeling by satisfaction of spatial restraints. LOOPP is based on various structural signals, which merge

into a single score. HOMER and CPH employ homology searching while PHYRE uses an advanced fold-recognition system designed to model the entropy of a folding protein. Detailed comparison and performance scores of some of the programs mentioned are reviewed in Wallner and Elofsson [61].

4) Model refinement and evaluation

Loop refinement and side chain packing iterations were performed using individual modeling programs whenever available. In addition, loop refinement was done using the stand alone module, Loopy [64], and the prediction of side-chain conformations with SCWRL3.0 [65] and SCAP [66-68].

The quality of the models was assessed using Verify3D [101-103] and Prosa [104].

F. Electrostatic and hydrophobic interactions

The analysis of biophysical properties including the electrostatics, hydrophobicity, and shape of each model was conducted using the surface property analysis tools in the program GRASP [105].

The pKa values of ionizable amino acid side chains in *Arabidopsis* PH and FYVE domains as well as total charges were computed using the automated system H++ [106-108], which is based on solutions to the Poisson-Boltzmann equation for the given modeled structure. The calculations were performed using default settings. The reported total charges were calculated at pH 6.5 for the PH and FYVE domains because most PH domains, for example, the pleckstrin PH domain [109], were crystallized at pH of 6.5 and the EEA1-FYVE was estimated to exist in bound state at low pH of 6.0 – 6.6 and only half of the protein was estimated to remain active at pH of 7.3 [110].

G. Lipid binding and specificity via docking

Ins(1,*3)P*$_2$ and Ins(1,*5)P*$_2$ (headgroups of PtdIns(3)*P* and PtdIns(5)*P*, respectively) were assembled from the Protein Data Bank (PDB) database by extracting coordinates from various PDB files of determined structures and energy minimizing them. If not available in the PDB database, as in the case of Ins(1,*5)P*$_2$, liganda were created from other ligands using the UCSF Chimera package from the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco (supported by NIH P41 RR-01081) [111] and energy minimized. Our docking study focused only on the interaction of Ins(1,*3)P*$_2$ and Ins(1,*5)P*$_2$ with the FYVE domains and did not consider the interaction of phosphoinositide lipid fragments with the membrane. Consequently, there are no lipid chains involved in the docked ligands.

H. Phosphoinositide docking and analysis of resulting interactions

Rigid and flexible docking was performed using DOCK 6.1 [112] and DOCK 6.1 suite of programs. A molecular surface of the receptor was created with DMS [113, 114]. Spheres were generated with Sphgen_cpp v1.2, which was modified by Andrew Magis from its original version called Sphgen [112]. The resulting file was edited to include only spheres grouped within the first cluster. Grids were generated with GRID [115, 116]. Contact scores and energy scores were calculated using an energy cutoff distance of 5.0 A. Our docking technique was validated by docking Ins(1,*3)P*$_2$ of known FYVE domains into their corresponding solved structures. Following the initial validation we used our approach to dock Ins(1,*3)P*$_2$ and Ins(1,*5)P*$_2$ using rigid and flexible docking scenarios into the modeled sequences of *Arabidopsis* FYVE domains. In the end, each model was subjected to six docking runs for each headgroup. A given residue is reported to interact with the headgroup only if it does so 50% or more of the time (i.e. three or more times) as evaluated by the Ligand-Protein Contacts server [117].

İ. Multiple sequence alignments

The analysis of biophysical properties including the electrostatics conservation of residues in evolutionarily related proteins suggest that they are likely to be functionally and/or structurally important. These conservation patterns can be delineated using multiple sequence alignment programs. We have utilized multiple sequence alignments of *Arabidopsis* PRAF domains created by the program ClustalW [118] to reveal any significant conservation patterns with emphasis on the presence or absence of consensus sequences known to bind specific phsophoinositides.

III. Results

A. Domain Architecture of Arabidopsis *PRAF proteins*

The *Arabidopsis* PRAF family includes nine proteins, which share similar domain architecture, i.e. a PH domain, followed by RCC1 regions/blades (overlapping with Alpha Tubulin Suppressor 1 (ATS1)) and a FYVE domain.

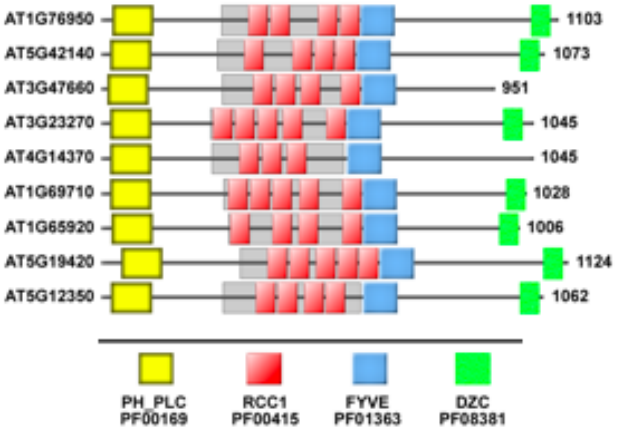

Fig. 1. *Arabidopsis* PRAF proteins. The Pfam (PF), SMART (SM), Conserved Domains (CD) and Clusters of Orthologous Groups (COG) accession numbers for the different domains are given under the figure. The lengths of each protein are indicated on the right.

In addition, seven out of nine PRAF proteins are characterized by the presence of a BRX motif found near the C-terminus. UniProtKB/TrEMBL annotates function for PRAF proteins as either disease resistance protein-like, e.g. AT5G42140 and AT4G14370, Ran GTPase binding / chromatin binding / zinc ion binding, e.g. AT1G65920, AT1G69710, AT3G23270 and AT5G12350, or hypothetical / unknown, e.g. AT3G47660, AT1G76950, and AT5G19420. AT4G14370 is a misannotated *in silico* fusion of nucleotide-binding site-leucine-rich repeat (NBS-LRR) and an RCC1-encoding gene [10].

B. The PH of Arabidopsis PRAF proteins

The modeled PRAF PH domains are typical PH domains with distinct electrostatic polarization and strongly basic canonical binding surfaces (Fig. 2) but show unique secondary structure elements in addition to the ones that form the core PH domain fold (Fig. 3). The secondary structure predictions identify an N-terminal α-helix, five consecutive β-strands followed by an additional α-helix, two more β-strands and a C-terminal α-helix.

Sequence analysis shows that PRAF PH domains are very similar to the mammalian PLCδ1 PH domain (Fig. 3). Moreover, both PRAF-1 and PLCδ1 have been shown to preferentially bind PtdIns(4,*5*)P_2 with high specificity and affinity [4, 6-8]. Given that plant PLCs lack the N-terminal PH domain seen in PLCs of other organisms, including PLCδ1, it is intriguing to find that there is a highly related PH domain sequence present in this non-related family of PRAF proteins. We speculate that there may be a missing link that has not yet been established, which ties the two families of proteins evolutionarily.

We predict that most PRAFs will have a lowered PtdIns(4,*5*)P_2 – specific binding affinity compared to PLCδ1 PH domain due to a partial PtdIns(4,*5*)P_2 binding signature (Fig. 3) but a stronger non-specific electrostatic contribution to membrane binding from their large basic patches for most members of the family (Fig. 2). The existence of the unique secondary structure elements i.e. the additional N-terminal α-helix and the α-helix within the β5/β6 loop is also seen in the PLC family of PH domains, reinforcing the idea that these two protein families share some evolutionary link.

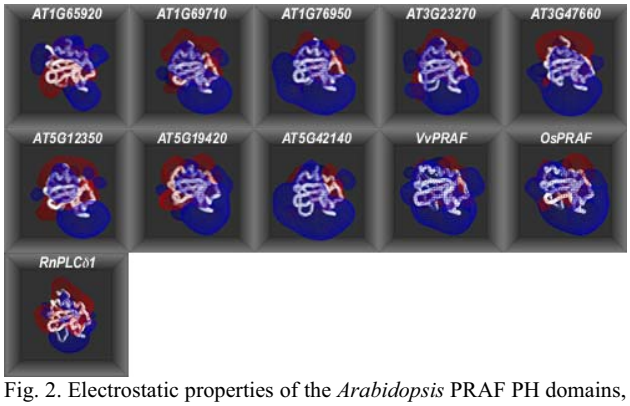

Fig. 2. Electrostatic properties of the *Arabidopsis* PRAF PH domains, their *V. vinifera* (A5B4Z5) and *O. sativa* (Q0JFZ5) putative homologs and *R. norvegicus* PLCδ1 PH domain (PDB ID: 1MAI [119]). All PH domain models are in the same orientation. In all panels, the electrostatic potentials were calculated at +25 mV and -25 mV equipotential contours in 0.1 M KCl. All images of the electrostatic potential contours were calculated with GRASP [105]. The models are represented as Cα backbone traces and their accession numbers/abbreviated names are displayed in the upper part of each profile.

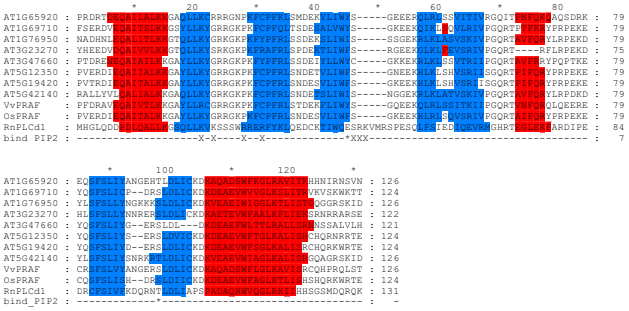

Fig. 3. The alignment of consensus secondary structure predictions for the *Arabidopsis* PRAF PH domains, their *V. vinifera* (A5B4Z5) and *O. sativa* (Q0JFZ5) putative homologs and *R. norvegicus* PLCδ1 PH domain (PDB ID: 1MAI [119]). Rows depict 1-letter code of PH domain sequences, α-helices are highlighted in red, β-strands in blue, X identifies residues which hydrogen bond directly and * identifies residues, which hydrogen bond via a water molecule.

C. The RCC1 domains of Arabidopsis PRAF proteins

The SMART database recognizes between three and five RCC1 regions within the *Arabidopsis* PRAF proteins, whereas the CD-search identifies additionally yeast domain with similarity to human RCC1 domain, ATS1 domain, overlapping the RCC1 blades (Fig. 4). In some cases, only the ATS1 domain is detected by the CD-search or the number of RCC1 blades does not correspond to the number obtained from SMART database (data not shown). These inconsistencies prompted further inquiry into the number and nature of the putative RCC1 repeats identified in the *Arabidopsis* PRAF proteins. Up to now, RCC1 and RCC1-like domains that have been described are within cytoplasmic proteins associated with membrane structures, e.g. endosomes (Alsin) [120] and Golgi apparatus (HERC1) [121]. Figure 4 shows an internal sevenfold sequence repeat of 51–68 residues present in the solved structure of human RCC1 [122] aligned with putative RCC1 regions of the *Arabidopsis* PRAF proteins. In human RCC1, one half of the first sequence repeat, the C and D repeats, is made from the N-terminal end of the protein, and the other half, the A and B repeats, is made from the C-terminal end [122]. It has been suggested that this arrangement stabilize the circular arrangement of secondary structural elements through a molecular clasp mechanism similar to a belt closure [122]. Our data show that putative RCC1 blades of *Arabidopsis* PRAF proteins align well with six of human seven RCC1 blades. In fact, the seven highly conserved residues, i.e. four glycines, a tyrosine, a leucine and a *cis*-proline, identified in human RCC1 repeats are also mostly conserved among putative *Arabidopsis* RCC1 blades (boxed residues in Fig. 4). However, it appears that the first blade of human RCC1 shares little or no primary and/or

secondary sequence similarity with most putative *Arabidopsis* RCC1 blades. Consequently, it is possible that the first blade of seven putative *Arabidopsis* RCC1 proteins is not an actual repeat (Fig. 4 and Fig. 5). It is likely however, that the first blade of AT5G19420 and AT5G12350 is a genuine repeat given the slightly higher residue similarity (Fig. 4 and Fig. 5).

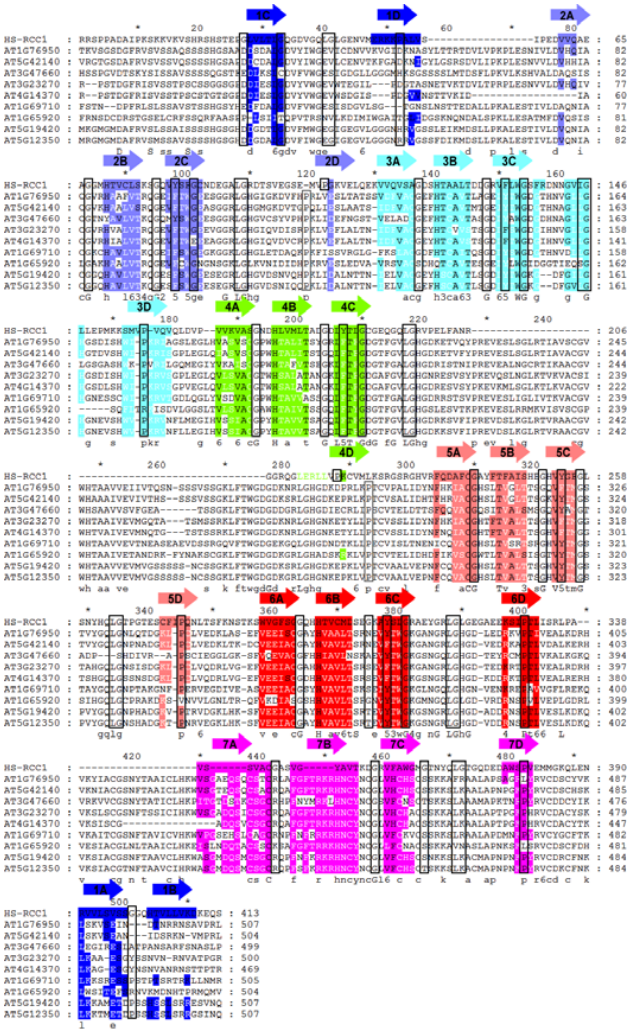

Fig. 4. Sequence alignment of human RCC1 and putative *Arabidopsis* homologues. The secondary structures are adapted from solved structure of human RCC1 with minor modifications [122]. Residues absolutely conserved within the secondary structures are black on a colored background. Residues moderately conserved within the secondary structures and/or among *Arabidopsis* repeats and not human RCC1 are white on a colored background. Boxed residues correspond to amino acids, which are highly conserved among each blade of the seven propeller structure [122].

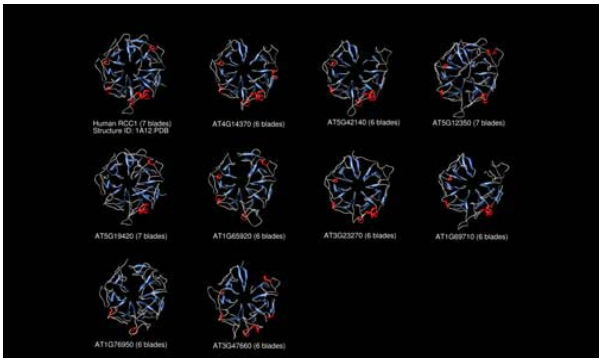

Fig. 5. Structural representation of *Arabidopsis* RCC1 proteins.

D. The FYVE of Arabidopsis PRAF proteins

The PRAF FYVE domains do not share the classical FYVE N-terminal WxxD motif. Instead they have either the N-terminal WxxG motif, only the G residue or residues that share no similarity to others (Fig. 6). Moreover, the classical FYVE R(R/K)HHCR motif is also replaced by the (K/R)(R/K)HNCY motif, which is atypical [4].

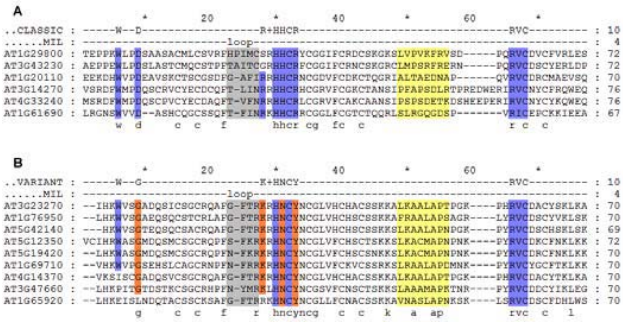

Fig. 6. A. Alignment of six FYVE domains representing classic *Arabidopsis* FYVE proteins. The conserved sequence motif of classic *Arabidopsis* FYVE domains is highlighted in blue. B. Alignment of nine FYVE domains representing the variant *Arabidopsis PRAF* FYVE proteins. The conserved sequence motif of *Arabidopsis* PRAF FYVE domains is highlighted in blue (if the same as the classic FYVE motif) and orange (if different from the classic FYVE motif). The variable turret loop is highlighted in grey and the putative dimer interface as corresponding to EEA1-FYVE dimer region is highlighted in yellow.

Preliminary docking studies show that the PRAF FYVE domains use the variant signature of residues to potentially bind headgroups of PtdIns(3)*P* and PtdIns(5)*P* i.e. xRKxHNxY motif and (L/F/P)YR motif overlapping the RVC motif (Fig. 6). In addition to the variant residues, our data indicate that (H/K/N)xx(S/T)(S/N)(K/R)K motif located immediately prior to the dimerization region (i.e. HxCSSKK) is used by the PRAF FYVE domains to recognize either of the headgroups (Fig. 6).

In addition, the putative dimerization interface region of the *Arabidopsis* PRAF FYVE domains is highly hydrophobic and very conserved with at least three absolutely conserved residues, i.e. AxxAP.

All models of the PRAF FYVE domains have a highly positive overall net charge, which varies from +9 to +16 (Fig. 7).

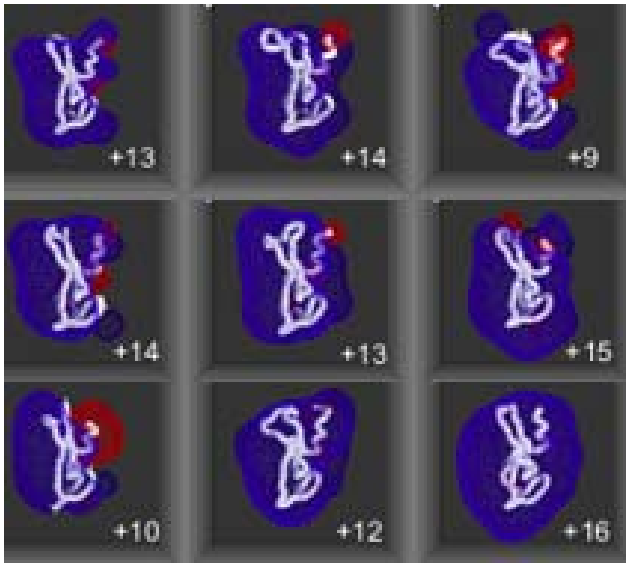

Fig. 7. The electrostatic profiles of *Arabidopsis* FYVE domains. All profiles are shown in the same orientation with the membrane binding regions facing down. The red and blue meshes represent the -1 kT/e and +1kT/e equipotential contours of the FYVE domains. From left to right, AT1G76950, AT5G42140, AT3G47660, AT3G23270, AT4G14370, AT1G69710, AT1G65920, AT5G19420 and AT5G12350. The numbers in the lower right corner correspond to total charges on each *Arabidopsis* FYVE domain model.

E. The BRX of Arabidopsis PRAF proteins

The BRX domain mediates homotypic and heterotypic interactions within and between the BRX and PRAF protein families in *S. cerevisiae*. Consequently, it has been suggested that the BRX domain represents a novel protein-protein interaction domain. Structurally, the BRX domain is yet to be solved. Our preliminary modeling analyses reveal that the BRX domain is most likely comprised of a single helix (Fig. 8).

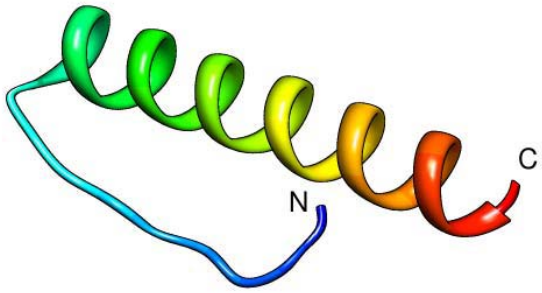

Fig. 8. A schematic representation of Arabidopsis AT5G12350 BRX model.1.

IV. Discussion

The *Arabidopsis thaliana* genome contains nine plant-specific PRAF proteins, which share moderate to high sequence homology and possibly redundant membrane binding behaviors to elicit their functional role. Our studies show that not only is this family of proteins unique to plants but also that their membrane targeting domains are unlike their counterparts in other organisms such as mammals, worm and yeast, suggesting variant functionality. For example, the distinctive highly positive electrostatic potentials of FYVE domains, unique substrate specificity to PtdIns(5)*P* and plant exclusive signature motif are likely to contribute to their unique functions.

The characterization of *Arabidopsis* lipid-binding PH and FYVE domains presented in this study represents the beginning of our understanding of PH and FYVE domains in plants. Moreover, the identified difference in β-propeller organization of RCC1 repeats could provide direction for experimental study of protein-binding characteristics either via mutational studies or molecular dynamics simulations. Additionally, our results also provide groundwork for domain-based swapping studies or extensive comparative studies among various species.

References

[1] G.C. Briggs, C.F. Mouchel, C.S. Hardtke, Plant Physiol, 140 (2006) 1306.
[2] C.F. Mouchel, G.C. Briggs, C.S. Hardtke, Genes Dev, 18 (2004) 700.
[3] B. Heras, B.K. Drobak, J Exp Bot, 53 (2002) 565.
[4] R.B. Jensen, T. La Cour, J. Albrethsen, M. Nielsen, K. Skriver, Biochem J, 359 (2001) 165.
[5] C.D. Knox, A.E. Belous, J.M. Pierce, A. Wakata, I.B. Nicoud, C.D. Anderson, C.W. Pinson, R.S. Chari, Am J Physiol Gastrointest Liver Physiol, 287 (2004) G533.
[6] P. Garcia, R. Gupta, S. Shah, A.J. Morris, S.A. Rudge, S. Scarlata, V. Petrova, S. McLaughlin, M.J. Rebecchi, Biochemistry, 34 (1995) 16228.
[7] F.M. Flesch, J.W. Yu, M.A. Lemmon, K.N. Burger, Biochem J, 389 (2005) 435.
[8] M.A. Lemmon, K.M. Ferguson, R. O'Brien, P.B. Sigler, J. Schlessinger, Proc Natl Acad Sci U S A, 92 (1995) 10472.
[9] A. Hayakawa, S.J. Hayes, D.C. Lawe, E. Sudharshan, R. Tuft, K. Fogarty, D. Lambright, S. Corvera, J Biol Chem, 279 (2004) 5958.

[10] W. van Leeuwen, L. Okresz, L. Bogre, T. Munnik, Trends Plant Sci, 9 (2004) 378.
[11] B.K. Drobak, B. Heras, Trends Plant Sci, 7 (2002) 132.
[12] M. Lemmon, Biochem Soc Trans. , 32 (2004) 707.
[13] T. Gibson, M. Hyvonen, A. Musacchio, M. Saraste, E. Birney, Trends Biochem Sci. , 19 (1994) 349.
[14] J.M. Kavran, D.E. Klein, A. Lee, M. Falasca, S.J. Isakoff, E.Y. Skolnik, M.A. Lemmon, J Biol Chem, 273 (1998) 30497.
[15] M.A. Lemmon, K.M. Ferguson, Biochem J, 350 (2000) 1.
[16] A. Musacchio, T. Gibson, P. Rice, J. Thompson, M. Saraste, Trends Biochem Sci. , 18 (1993) 343.
[17] A. Petiot, J. Faure, H. Stenmark, J. Gruenberg, J. Cell Biol., 162 (2003) 971.
[18] A.E. Wurmser, J.D. Gary, S.D. Emr, J. Biol. Chem., 274 (1999) 9129.
[19] A. Simonsen, R. Lippe, S. Christoforidis, J.M. Gaullier, A. Brech, J. Callaghan, B.H. Toh, C. Murphy, M. Zerial, H. Stenmark, Nature, 394 (1998) 494.
[20] W. Moore, C. Zhang, P.R. Clarke, Curr Biol, 12 (2002) 1442.
[21] N. Mirkovic, Z. Li, A. Parnassa, D. Murray, Proteins, 66 (2006) 766.
[22] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne, Nucleic Acids Res, 28 (2000) 235.
[23] H. Berman, K. Henrick, H. Nakamura, Nat Struct Biol, 10 (2003) 980.
[24] M. Kanehisa, Novartis Found Symp, 247 (2002) 91.
[25] M. Kanehisa, S. Goto, M. Hattori, K.F. Aoki-Kinoshita, M. Itoh, S. Kawashima, T. Katayama, M. Araki, M. Hirakawa, Nucl. Acids Res., 34 (2006) D354.
[26] J. Schultz, R.R. Copley, T. Doerks, C.P. Ponting, P. Bork, Nucleic Acids Res, 28 (2000) 231.
[27] J. Schultz, F. Milpetz, P. Bork, C.P. Ponting, PNAS, 95 (1998) 5857.
[28] I. Letunic, R.R. Copley, B. Pils, S. Pinkert, J. Schultz, P. Bork, Nucleic Acids Res, 34 (2006) D257.
[29] A. Bairoch, B. Boeckmann, Nucleic Acids Res, 22 (1994) 3578.
[30] B. Boeckmann, M.C. Blatter, L. Famiglietti, U. Hinz, L. Lane, B. Roechert, A. Bairoch, C R Biol, 328 (2005) 882.
[31] D.A. Benson, I. Karsch-Mizrachi, D.J. Lipman, J. Ostell, D.L. Wheeler, Nucleic Acids Res, 34 (2006) D16.
[32] D.L. Wheeler, T. Barrett, D.A. Benson, S.H. Bryant, K. Canese, V. Chetvernin, D.M. Church, M. DiCuccio, R. Edgar, S. Federhen, L.Y. Geer, W. Helmberg, Y. Kapustin, D.L. Kenton, O. Khovayko, D.J. Lipman, T.L. Madden, D.R. Maglott, J. Ostell, K.D. Pruitt, G.D. Schuler, L.M. Schriml, E. Sequeira, S.T. Sherry, K. Sirotkin, A. Souvorov, G. Starchenko, T.O. Suzek, R. Tatusov, T.A. Tatusova, L. Wagner, E. Yaschenko, Nucleic Acids Res, 34 (2006) D173.
[33] R. Apweiler, A. Bairoch, C.H. Wu, W.C. Barker, B. Boeckmann, S. Ferro, E. Gasteiger, H. Huang, R. Lopez, M. Magrane, M.J. Martin, D.A. Natale, C. O'Donovan, N. Redaschi, L.S. Yeh, Nucleic Acids Res, 32 (2004) D115.
[34] A. Bairoch, R. Apweiler, C.H. Wu, W.C. Barker, B. Boeckmann, S. Ferro, E. Gasteiger, H. Huang, R. Lopez, M. Magrane, M.J. Martin, D.A. Natale, C. O'Donovan, N. Redaschi, L.S. Yeh, Nucleic Acids Res, 33 (2005) D154.
[35] C.H. Wu, R. Apweiler, A. Bairoch, D.A. Natale, W.C. Barker, B. Boeckmann, S. Ferro, E. Gasteiger, H. Huang, R. Lopez, M. Magrane, M.J. Martin, R. Mazumder, C. O'Donovan, N. Redaschi, B. Suzek, Nucleic Acids Res, 34 (2006) D187.
[36] S.F. Altschul, T.L. Madden, A.A. Schaffer, J. Zhang, Z. Zhang, W. Miller, D.J. Lipman, Nucleic Acids Res, 25 (1997) 3389.
[37] A. Bateman, L. Coin, R. Durbin, R.D. Finn, V. Hollich, S. Griffiths-Jones, A. Khanna, M. Marshall, S. Moxon, E.L. Sonnhammer, D.J. Studholme, C. Yeats, S.R. Eddy, Nucleic Acids Res, 32 (2004) D138.
[38] A. Marchler-Bauer, J.B. Anderson, P.F. Cherukuri, C. DeWeese-Scott, L.Y. Geer, M. Gwadz, S. He, D.I. Hurwitz, J.D. Jackson, Z. Ke, C.J. Lanczycki, C.A. Liebert, C. Liu, F. Lu, G.H. Marchler, M. Mullokandov, B.A. Shoemaker, V. Simonyan, J.S. Song, P.A. Thiessen, R.A. Yamashita, J.J. Yin, D. Zhang, S.H. Bryant, Nucleic Acids Res, 33 (2005) D192.
[39] A. Marchler-Bauer, J.B. Anderson, M.K. Derbyshire, C. DeWeese-Scott, N.R. Gonzales, M. Gwadz, L. Hao, S. He, D.I. Hurwitz, J.D. Jackson, Z. Ke, D. Krylov, C.J. Lanczycki, C.A. Liebert, C. Liu, F. Lu, S. Lu, G.H. Marchler, M. Mullokandov, J.S. Song, N. Thanki, R.A. Yamashita, J.J. Yin, D. Zhang, S.H. Bryant, Nucleic Acids Res, 35 (2007) D237.
[40] A. Marchler-Bauer, J.B. Anderson, C. DeWeese-Scott, N.D. Fedorova, L.Y. Geer, S. He, D.I. Hurwitz, J.D. Jackson, A.R. Jacobs, C.J. Lanczycki, C.A. Liebert, C. Liu, T. Madej, G.H. Marchler, R. Mazumder, A.N. Nikolskaya, A.R. Panchenko, B.S. Rao, B.A. Shoemaker, V. Simonyan, J.S. Song, P.A. Thiessen, S. Vasudevan, Y. Wang, R.A. Yamashita, J.J. Yin, S.H. Bryant, Nucleic Acids Res, 31 (2003) 383.
[41] A. Marchler-Bauer, S.H. Bryant, Nucleic Acids Res, 32 (2004) W327.
[42] A. Marchler-Bauer, A.R. Panchenko, B.A. Shoemaker, P.A. Thiessen, L.Y. Geer, S.H. Bryant, Nucleic Acids Res, 30 (2002) 281.
[43] R.L. Tatusov, N.D. Fedorova, J.D. Jackson, A.R. Jacobs, B. Kiryutin, E.V. Koonin, D.M. Krylov, R. Mazumder, S.L. Mekhedov, A.N. Nikolskaya, B.S. Rao, S. Smirnov, A.V. Sverdlov, S. Vasudevan, Y.I. Wolf, J.J. Yin, D.A. Natale, BMC Bioinformatics, 4 (2003) 41.
[44] R.L. Tatusov, E.V. Koonin, D.J. Lipman, Science, 278 (1997) 631.
[45] J.A. Cuff, G.J. Barton, Proteins, 40 (2000) 502.
[46] J.A. Cuff, G.J. Barton, Proteins, 34 (1999) 508.
[47] J.A. Cuff, M.E. Clamp, A.S. Siddiqui, M. Finlay, G.J. Barton, Bioinformatics, 14 (1998) 892.
[48] K. Bryson, L.J. McGuffin, R.L. Marsden, J.J. Ward, J.S. Sodhi, D.T. Jones, Nucleic Acids Res, 33 (2005) W36.
[49] D.T. Jones, J Mol Biol, 292 (1999) 195.
[50] L.J. McGuffin, K. Bryson, D.T. Jones, Bioinformatics, 16 (2000) 404.
[51] B. Rost, Methods Enzymol, 266 (1996) 525.
[52] M. Ouali, R.D. King, Protein Sci, 9 (2000) 1162.
[53] K. Karplus, C. Barrett, R. Hughey, Bioinformatics, 14 (1998) 846.
[54] K. Karplus, S. Katzman, G. Shackleford, M. Koeva, J. Draper, B. Barnes, M. Soriano, R. Hughey, Proteins, 61 Suppl 7 (2005) 135.
[55] K. Karplus, B. Hu, Bioinformatics, 17 (2001) 713.
[56] Y. Guermeur, PhD thesis (1997).
[57] P. Baldi, S. Brunak, P. Frasconi, G. Soda, G. Pollastri, Bioinformatics, 15 (1999) 937.
[58] J. Cheng, A.Z. Randall, M.J. Sweredoski, P. Baldi, Nucleic Acids Res, 33 (2005) W72.
[59] C. Cole, J.D. Barber, G.J. Barton, Nucleic Acids Res, 36 (2008) W197.
[60] O. Dor, Y. Zhou, Proteins, 66 (2007) 838.
[61] B. Wallner, A. Elofsson, Protein Sci, 14 (2005) 1315.
[62] D. Petrey, B. Honig, Mol Cell, 20 (2005) 811.
[63] S.M. Singh, D. Murray, Protein Sci, 12 (2003) 1934.
[64] Z. Xiang, C.S. Soto, B. Honig, Proc Natl Acad Sci U S A, 99 (2002) 7432.
[65] A.A. Canutescu, A.A. Shelenkov, R.L. Dunbrack, Jr., Protein Sci, 12 (2003) 2001.
[66] Z. Xiang, P.J. Steinbach, M.P. Jacobson, R.A. Friesner, B. Honig, Proteins (2007).
[67] Z. Xiang, B. Honig, J Mol Biol, 311 (2001) 421.
[68] M.P. Jacobson, R.A. Friesner, Z. Xiang, B. Honig, J Mol Biol, 320 (2002) 597.
[69] J.L. Sussman, D. Lin, J. Jiang, N.O. Manning, J. Prilusky, O. Ritter, E.E. Abola, Acta Crystallogr D Biol Crystallogr, 54 (1998) 1078.
[70] N.N. Alexandrov, R. Nussinov, R.M. Zimmer, Pac Symp Biocomput (1996) 53.
[71] J. Shi, T.L. Blundell, K. Mizuguchi, J Mol Biol, 310 (2001) 243.
[72] J. Meller, R. Elber, Proteins, 45 (2001) 241.
[73] O. Teodorescu, T. Galor, J. Pillardy, R. Elber, Proteins, 54 (2004) 41.
[74] D. Tobi, R. Elber, Proteins, 41 (2000) 40.
[75] L.A. Kelley, R.M. MacCallum, M.J. Sternberg, J Mol Biol, 299 (2000) 499.
[76] S.F. Altschul, E.V. Koonin, Trends Biochem Sci, 23 (1998) 444.
[77] A.A. Schaffer, L. Aravind, T.L. Madden, S. Shavirin, J.L. Spouge, Y.I. Wolf, E.V. Koonin, S.F. Altschul, Nucleic Acids Res, 29 (2001) 2994.
[78] A.A. Schaffer, Y.I. Wolf, C.P. Ponting, E.V. Koonin, L. Aravind, S.F. Altschul, Bioinformatics, 15 (1999) 1000.
[79] P.A. Bates, L.A. Kelley, R.M. MacCallum, M.J. Sternberg, Proteins, Suppl 5 (2001) 39.
[80] P.A. Bates, M.J. Sternberg, Proteins, Suppl 3 (1999) 47.
[81] B. Contreras-Moreira, P.A. Bates, Bioinformatics, 18 (2002) 1141.
[82] S.C.E. Tosatto, (Submitted).
[83] S.C.E. Tosatto, Journal of Computational Biology, 12 (2005) 1316.
[84] O. Lund, M. Nielsen, C. Lundegaard, P. Worning, Abstract at the CASP5 conference A102 (2002).
[85] K. Nicholas, H. Nicholas, D. Deerfield, EMBNEW.NEWS, 14 (1997).

[86] R. Bonneau, C.E. Strauss, C.A. Rohl, D. Chivian, P. Bradley, L. Malmstrom, T. Robertson, D. Baker, J Mol Biol, 322 (2002) 65.
[87] R. Bonneau, J. Tsai, I. Ruczinski, D. Chivian, C. Rohl, C.E. Strauss, D. Baker, Proteins, Suppl 5 (2001) 119.
[88] D. Chivian, D.E. Kim, L. Malmstrom, P. Bradley, T. Robertson, P. Murphy, C.E. Strauss, R. Bonneau, C.A. Rohl, D. Baker, Proteins, 53 Suppl 6 (2003) 524.
[89] D. Chivian, D.E. Kim, L. Malmstrom, J. Schonbrun, C.A. Rohl, D. Baker, Proteins, 61 Suppl 7 (2005) 157.
[90] D.E. Kim, D. Chivian, D. Baker, Nucl. Acids Res., 32 (2004) W526.
[91] K.T. Simons, C. Kooperberg, E. Huang, D. Baker, J Mol Biol, 268 (1997) 209.
[92] K.T. Simons, I. Ruczinski, C. Kooperberg, B.A. Fox, C. Bystroff, D. Baker, Proteins, 34 (1999) 82.
[93] J. Moult, Curr Opin Struct Biol, 15 (2005) 285.
[94] C. Sander, R. Schneider, Proteins, 9 (1991) 56.
[95] M. Pellegrini-Calace, S. Soro, A. Tramontano, Febs J, 273 (2006) 2977.
[96] A. Fiser, R.K. Do, A. Sali, Protein Sci, 9 (2000) 1753.
[97] M.A. Marti-Renom, A.C. Stuart, A. Fiser, R. Sanchez, F. Melo, A. Sali, Annu Rev Biophys Biomol Struct, 29 (2000) 291.
[98] U. Pieper, N. Eswar, H. Braberg, M.S. Madhusudhan, F.P. Davis, A.C. Stuart, N. Mirkovic, A. Rossi, M.A. Marti-Renom, A. Fiser, B. Webb, D. Greenblatt, C.C. Huang, T.E. Ferrin, A. Sali, Nucleic Acids Res, 32 (2004) D217.
[99] A. Sali, T.L. Blundell, J Mol Biol, 234 (1993) 779.
[100] D. Petrey, Z. Xiang, C.L. Tang, L. Xie, M. Gimpelev, T. Mitros, C.S. Soto, S. Goldsmith-Fischman, A. Kernytsky, A. Schlessinger, I.Y. Koh, E. Alexov, B. Honig, Proteins, 53 Suppl 6 (2003) 430.
[101] J.U. Bowie, R. Luthy, D. Eisenberg, Science, 253 (1991) 164.
[102] D. Eisenberg, R. Luthy, J.U. Bowie, Methods Enzymol, 277 (1997) 396.
[103] R. Luethy, J.U. Bowie, D. Eisenberg, Nature, 356 (1992) 83.
[104] M.J. Sippl, J Comput Aided Mol Des, 7 (1993) 473.
[105] A. Nicholls, K.A. Sharp, B. Honig, Proteins, 11 (1991) 281.
[106] J.C. Gordon, J.B. Myers, T. Folta, V. Shoja, L.S. Heath, A. Onufriev, Nucleic Acids Res, 33 (2005) W368.
[107] J. Myers, G. Grothaus, S. Narayanan, A. Onufriev, Proteins, 63 (2006) 928.
[108] D. Bashford, M. Karplus, Biochemistry, 29 (1990) 10219.
[109] S.G. Jackson, Y. Zhang, R.J. Haslam, M.S. Junop, BMC Struct Biol, 7 (2007) 80.
[110] T.G. Kutateladze, Biochim Biophys Acta, 1761 (2006) 868.
[111] E.F. Pettersen, T.D. Goddard, C.C. Huang, G.S. Couch, D.M. Greenblatt, E.C. Meng, T.E. Ferrin, J Comput Chem, 25 (2004) 1605.
[112] I.D. Kuntz, J.M. Blaney, S.J. Oatley, R. Langridge, T.E. Ferrin, J Mol Biol, 161 (1982) 269.
[113] F.M. Richards, Annu Rev Biophys Bioeng, 6 (1977) 151.
[114] T.E. Ferrin, C.C. Huang, L.E. Jarvis, R. Langridge, J. Mol. Graph., 6 (1988) 13.
[115] B.K. Shoichet, D.L. Bodian, I.D. Kuntz, J. Comp. Chem., 13 (1992) 380.
[116] E.C. Meng, B.K. Shoichet, I.D. Kuntz, J. Comp. Chem., 13 (1992) 505.
[117] V. Sobolev, A. Sorokine, J. Prilusky, E.E. Abola, M. Edelman, Bioinformatics, 15 (1999) 327.
[118] R. Chenna, H. Sugawara, T. Koike, R. Lopez, T.J. Gibson, D.G. Higgins, J.D. Thompson, Nucleic Acids Res, 31 (2003) 3497.
[119] K.M. Ferguson, M.A. Lemmon, J. Schlessinger, P.B. Sigler, Cell, 83 (1995) 1037.
[120] A. Otomo, S. Hadano, T. Okada, H. Mizumura, R. Kunita, H. Nishijima, J. Showguchi-Miyata, Y. Yanagisawa, E. Kohiki, E. Suga, M. Yasuda, H. Osuga, T. Nishimoto, S. Narumiya, J.E. Ikeda, Hum Mol Genet, 12 (2003) 1671.
[121] J.L. Rosa, R.P. Casaroli-Marano, A.J. Buckler, S. Vilaro, M. Barbacid, Embo J, 15 (1996) 4262.
[122] L. Renault, N. Nassar, I. Vetter, J. Becker, C. Klebe, M. Roth, A. Wittinghofer, Nature, 392 (1998) 97.
[123] A. Rodrigues, J. Santiago, S. Rubio, A. Saez, K.S. Osmont, J. Gadea, C.S. Hardtke, P.L. Rodriguez, Plant Physiol., 149 (2009) 1917.