

**GENOME-WIDE ANALYSIS OF PH AND FYVE DOMAINS IN
ARABIDOPSIS THALIANA BY COMPUTATIONAL METHODS**

by

EWA WYWIAL

A dissertation submitted to the Graduate Faculty in Biology in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The City University of New York

2009

© 2009

EWA WYWIAL

All Rights Reserved

This manuscript has been read and accepted for the
Graduate Faculty in Biology in satisfaction of the
dissertation requirement for the degree of Doctor of Philosophy.

_____	_____
Date	Shaneen M. Singh, Ph.D. Chair of Examining Committee

_____	_____
Date	Laurel Eckhardt, Ph.D. Executive Officer

Shaneen M. Singh, Ph.D. _____

Theodore R. Muth, Ph.D. _____

Juergen Polle, Ph.D. _____
Supervision Committee

THE CITY UNIVERSITY OF NEW YORK

Abstract

GENOME-WIDE ANALYSIS OF PH AND FYVE DOMAINS IN *ARABIDOPSIS THALIANA* BY COMPUTATIONAL METHODS

by

Ewa Wywiał

Mentor: Dr. Shaneen M. Singh

Phosphoinositide-binding domains have emerged as molecules responsible for trafficking and anchoring of membrane associated proteins in various organisms ranging from yeast to mammalian cells. Two such domains, i.e. the Pleckstrin Homology (PH) and Fab1p, YOTB, Vac1 and EEA1 (FYVE), are of special interest because of their unique membrane binding characteristics and genome wide prevalence. Structurally, the PH domains share a core fold of seven β -strands followed by a C-terminal α -helix whereas the FYVE domains comprise two small double-stranded β -sheets, a C-terminal α -helix and eight Zn^{2+} ion-binding cysteines. The PH and FYVE domains are found in functionally diverse families in different organisms; some of these are well studied and others remain unexplored. Plants stand out from the rest of the organisms where most of protein families containing these domains remain largely uncharacterized. 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 PH and FYVE proteins, characterize their sequence and model/analyze their 3D structure. Integrating the sequence, structure and known experimental details for each of these domains, we propose a comprehensive genome-wide domain-based classification of both the *Arabidopsis* PH and FYVE proteins: we classify forty-nine *Arabidopsis* PH and fifteen *Arabidopsis* FYVE proteins into twelve and five classes, respectively. Our study provides valuable details on the biophysical traits of the individual modeled domains, family specific characteristics as well as genomic trends for these domains in *Arabidopsis* and predictions of their membrane binding behaviors. We report the exclusive presence of plant specific domain architectures, variant binding signatures and biophysical profiles in individual members or complete families of *Arabidopsis* PH and FYVE

domain-containing proteins such as the PRAF proteins. Our study provides the first glimpse into the putative roles of these lipid-binding domains based on which we can begin to understand the role(s) they play in plants and how it compares to their role in other organisms.

To Rodrigo, Nicole and Noah

Acknowledgments

I am indebted first and foremost to Dr. Singh for her support, guidance and patience as well as her careful reading of my original manuscript and her detailed, perceptive comments, which greatly contributed to the shape and clarity of the finished thesis.

To the members of my doctoral committee, Dr. Muth, Dr. Polle, Dr. Reddy, Dr. Stuart and Dr. Qiu, for their valuable time, guidance and careful supervision.

To my mother for her love and support throughout my life.

To my husband and children for being incredibly understanding and allowing me to study at the expense of household chores, holidays and countless other little things, all of which only his unconditional love could endure.

To Lois Katz, Dr. Rochelle Buffenstein and Dr. Timothy O'Connor for being a source of great inspiration.

To my colleagues at the Computational Biology Group at Brooklyn College and the Murray Laboratory at Columbia University for their help and constructive criticism all throughout these years. Special thanks to Saodat Shamuratova, Nikita Brasse and Ramy Abdel-Naby for being my friends and for listening to both my joys and woes.

To the faculty and staff of the Biology Department at Brooklyn College and the Graduate Center for their invaluable assistance and support.

Table of Contents

LIST OF ABBREVIATIONS	X
LIST OF TABLES	XII
LIST OF FIGURES	XIII
CHAPTER 1 INTRODUCTION	1
<i>ARABIDOPSIS THALIANA</i>	1
PHOSPHOINOSITIDES AND PHOSPHOINOSITIDE-BINDING DOMAINS	3
PHOSPHOINOSITIDES AND PHOSPHOINOSITIDE-BINDING DOMAINS IN PLANT PROTEINS	9
A. PH DOMAINS	9
B. FYVE DOMAINS	13
PROJECT JUSTIFICATION	17
CHAPTER 2 METHODS	18
<i>ARABIDOPSIS</i> PH AND FYVE PROTEINS	18
SEQUENCE VERIFICATION	18
DOMAIN ARCHITECTURE ANALYSES AND CLASSIFICATION	18
VERIFICATION OF PH AND FYVE DOMAIN BOUNDARIES	19
MOLECULAR MODELING	19
A. STRUCTURAL TEMPLATES	23
B. SEQUENCE ALIGNMENT	25
C. MODEL BUILDING	27
D. MODEL REFINEMENT AND EVALUATION	31
ANALYSIS OF THE MODELS	33
E. ELECTROSTATIC AND HYDROPHOBIC INTERACTIONS	33
F. LIPID BINDING AND SPECIFICITY VIA DOCKING	33
G. CONSENSUS MOTIFS	36
H. FUNCTIONAL ANNOTATION	37
PHYLOGENETIC TREE RECONSTRUCTION	38
ELECTRONIC SUPPLEMENT	38
CHAPTER 3 <i>ARABIDOPSIS</i> PH DOMAINS	39
SEQUENCE ANALYSES	39
A. SEQUENCE VERIFICATION	39
B. SECONDARY STRUCTURE PREDICTION	39
C. DOMAIN-BASED CLASSIFICATION	40
STRUCTURE PREDICTION AND ANALYSES	42
A. MOLECULAR MODELING	42
B. ELECTROSTATIC PROFILES	42
C. PI DOCKING	45
FUNCTIONAL ANNOTATION	47
DISCUSSION	48
A. THE OSBP CLASS	48
B. THE START CLASS	53

C. THE PLD ζ CLASS	57
D. THE PRAF CLASS	61
E. THE PDK CLASS	65
F. THE ARFGAP CLASS	69
G. THE RHO GAP CLASS	73
H. THE DYN CLASS	76
I. THE ATPH1 CLASS	80
J. THE DUF828, UNK1 AND UNK2 CLASSES	83
SUMMARY	88
CHAPTER 4 <i>ARABIDOPSIS</i> FYVE DOMAINS	89
IDENTIFICATION, CHARACTERIZATION AND CHROMOSOMAL LOCALIZATION OF FYVE PROTEINS ENCODED IN THE <i>ARABIDOPSIS</i> GENOME	89
DOMAIN ARCHITECTURE OF <i>ARABIDOPSIS</i> FYVE PROTEINS	91
MOLECULAR MODELS OF THE <i>ARABIDOPSIS</i> FYVE DOMAINS AND THEIR BIOPHYSICAL PROPERTIES	97
SEQUENCE MOTIFS OF THE <i>ARABIDOPSIS</i> FYVE DOMAINS.	100
FYVE DOMAINS HAVE THE POTENTIAL TO BIND PTDIns(3)P AND PTDIns(5)P.	102
DISCUSSION	105
CHAPTER 5 CONCLUSIONS AND FUTURE DIRECTIONS	111
APPENDIX A	112
REFERENCES	119

List of Abbreviations

ADL1	<i>Arabidopsis</i> dynamin-like 1
ADL2	<i>Arabidopsis</i> dynamin-like 2
ADL3	<i>Arabidopsis</i> dynamin-like 3
ADL6	<i>Arabidopsis</i> dynamin-like 6
AGI	<i>Arabidopsis</i> Genome Initiative
Akt	protein kinase B
ANK	ankyrin
APPL1	adaptor protein containing a PH domain, a PTB domain, and a leucine zipper motif
ArfGAP	adenosine diphosphate-ribosylation factor-guanosine triphosphatase activating proteins
ARF1/Arf1p	ADP-ribosylation factor 1
AtPDK1	<i>Arabidopsis thaliana</i> pyruvate dehydrogenase kinase 1
AtPH1	<i>Arabidopsis thaliana</i> pleckstrin homologue 1
AtSeh1	<i>Arabidopsis</i> Sec13 homolog 1
ATS1	alpha tubulin suppressor 1
BAR	BIN/amphiphysin/RVS
cDNA	complementary DNA
CD-search	conserved domain-search
C-DART	Conserved Domain Architecture Retrieval Tool
DAPP1	dual adaptor for phosphotyrosine and 3-phosphoinositides
DUF828	domain of unknown function 828
DUF500	domain of unknown function 500
DUF1336	domain of unknown function 1336
DYN	dynamin
DYNc	dynamin GTPase
EDR2	enhanced disease resistant 2
EST	expressed sequence tag
FAPP1	phosphatidylinositol-four-phosphate adaptor protein-1
Fgd1	faciogenital dysplasia 1
FYVE	domain name derived from the first letter of Fab1p, YOTB, Vac1 and EEA1 proteins
GED	GTPase effector
GDP	guanosine diphosphate
GEF	guanine nucleotide exchange factor
GmOSBP	<i>Glycine max</i> oxysterol binding protein
Grp1	general receptor for phosphoinositides isoform 1
GTP	guanosine triphosphate
Hrs	hepatocyte growth factor-regulated tyrosine kinase substrate
InsPtd(s)	Inositol Phosphate(s)
MIPS	Munich Information Center for Protein Sequences
MIL	membrane insertion loop
NBS-LRR	nucleotide-binding site-leucine-rich repeat
NCBI	National Center for Biotechnology Information
OSBP	oxysterol-binding protein
Osh1p	<i>Saccharomyces cerevisiae</i> OSBP 1
Osh2p	<i>Saccharomyces cerevisiae</i> OSBP 2
PA	phosphatidic acid
PDB	Protein Data Bank
PEPP1	phosphatidylinositol-three-phosphate-binding PH-domain protein-1
PH	pleckstrin homology
PI	phosphoinositide
PIKfyve	phosphatidylinositol 3-phosphate 5-kinase

PIP ₂	phosphatidylinositol bisphosphate
PIP ₃	phosphatidylinositol trisphosphate
PIPKc	phosphatidylinositol phosphate kinases
PI3-kinase	phosphatidylinositol 3-kinase
PI4-kinase	phosphatidylinositol 4-kinase
PLC δ	phospholipase C δ
PLD	phospholipase D
PLD ζ	phospholipase D ζ
PDK	3-phosphoinositide-dependent kinase 1
PKB	protein kinase B (Akt)
PLC δ 1	phospholipase C delta 1
PRAF	PH, RCC1 and FYVE domains-containing protein
PtdIns(3) <i>P</i>	phosphatidylinositol (3) monophosphate
PtdIns(4) <i>P</i>	phosphatidylinositol (4) monophosphate
PtdIns(5) <i>P</i>	phosphatidylinositol (5) monophosphate
PtdIns(3,4) <i>P</i> ₂	phosphatidylinositol (3,4) bisphosphate
PtdIns(4,5) <i>P</i> ₂	phosphatidylinositol (4,5) bisphosphate
PtdIns(3,5) <i>P</i> ₂	phosphatidylinositol (3,5) bisphosphate
PtdIns(3,4,5) <i>P</i> ₃	phosphatidylinositol (3,4,5) trisphosphate
PX	phox
RCC1	regulator of chromosome condensation 1
RhoGAP	Rho GTPase-activating proteins
ROS	reactive oxygen species
SARA	smad anchor for receptor activation
Sbf1	set binding factor 1
SMART	simple modular architecture research tool
STARD11	star-related lipid transfer protein 11
START	steroidogenic acute regulatory protein-related lipid-transfer
SVM	support vector machine
TAIR	The <i>Arabidopsis</i> Information Resource
TAPP1	tandem PH-domain-containing protein-1
TIGR	The Institute for Genomic Research
UniProt	Universal Protein Resource
UNK1	unknown 1
UNK2	unknown 2
VAN3	vascular network defective 3
Vps1p	vacuolar protein sorting protein 1
Vps27p	vacuolar protein sorting protein 27
Vps34	vacuolar protein sorting protein 34 (phosphatidylinositol 3-kinase)

List of Tables

Table 1.1	Localization and abundance of PIs in unstimulated eukaryotic cells.	5
Table 1.2	The specificity of PH domains.	8
Table 2.1	PDB accession numbers of structural templates used in the study.	24
Table 2.2	Verify3D plots and Prosa scores for the AT1G65920 PH domain models depicted in Fig. 2.3.	32
Table 2.3	Ligands used in the study, their parent coordinate files, resolutions, and molecular weights.	35
Table 0.1	Individual members of the domain-architecture based twelve classes of <i>Arabidopsis</i> PH proteins.	112
Table 0.2	<i>Arabidopsis</i> PH domains: accession numbers, techniques, and templates used to build their structural models.	113
Table 0.3	Functional annotation of <i>Arabidopsis</i> PH domains in TAIR database.	114

List of Figures

Figure 1.1	PI metabolism.	4
Figure 1.2	Differential distribution of PIs in membrane of various organelles present in eukaryotic cells.	6
Figure 1.3	Similarities and differences between the PI-binding PH domains.	11
Figure 1.4	The FYVE domain of 1JOC, EEA1 as a monomer (A) and a dimer (B).	15
Figure 2.1	A generic flow-chart for molecular (comparative) modeling.	21
Figure 2.2	Flowchart of the automated comparative modeling pipeline.	22
Figure 2.3	A graphical representation of the PH domain from PLC δ protein and four models of the AT1G65920 PH domain built using PLC δ as a template.	30
Figure 3.1	Cartoon of representative examples of the different classes of <i>Arabidopsis</i> PH proteins.	41
Figure 3.2	Electrostatic properties of the best A. OSBP, B. START, C. PLD ζ , D. PRAF, E. PDK, F. ArfGAP, G. RhoGAP, H. DYN, I. AtPH1, J. DUF828; K. UNK1, L. UNK2 PH domain models.	44
Figure 3.3	<i>Arabidopsis</i> PH domains and their PI specificity as identified via docking.	46
Figure 3.4	The alignment of consensus secondary structure predictions for the <i>Arabidopsis</i> OSBP PH domains and their <i>H. sapiens</i> (OSBP8, OSBP11, and ORP3), <i>S. cerevisiae</i> (Osh1p and Osh2p) and <i>G. max</i>) homologs.	51
Figure 3.5	Electrostatic properties of <i>Arabidopsis</i> OSBP PH domains and their <i>H. sapiens</i> (OSBP8, OSBP11, and ORP3), <i>S. cerevisiae</i> (Osh1p and Osh2p) and <i>G. max</i>) homologs.	52
Figure 3.6	The alignment of consensus secondary structure predictions for the <i>Arabidopsis</i> START domains, their <i>O. sativa</i> , <i>M. truncatula</i> , <i>P. pratensis</i> and <i>H. sapiens</i> STARD11 homologs.	55
Figure 3.7	Electrostatic properties of the <i>Arabidopsis</i> START PH domains, their <i>O. sativa</i> , <i>M. truncatula</i> , <i>P. pratensis</i> and <i>H. sapiens</i> STARD11 homologs.	56

Figure 3.8	The alignment of consensus secondary structure predictions for the <i>Arabidopsis</i> PLD ζ PH domains and their <i>H. sapiens</i> (PLD1, PLD2), <i>S. cerevisiae</i> , <i>C. elegans</i> and <i>D. discoideum</i> homologs.	59
Figure 3.9	Electrostatic properties of the <i>Arabidopsis</i> PLD ζ PH domains and their <i>H. sapiens</i> (PLD1, PLD2), <i>S. cerevisiae</i> , <i>C. elegans</i> and <i>D. discoideum</i> homologs.	60
Figure 3.10	The alignment of consensus secondary structure predictions for the <i>Arabidopsis</i> PRAF PH domains, their <i>V. vinifera</i> and <i>O. sativa</i> putative homologs and <i>R. norvegicus</i> PLC δ 1 PH domain.	63
Figure 3.11	Electrostatic properties of the <i>Arabidopsis</i> PRAF PH domains, their <i>V. vinifera</i> and <i>O. sativa</i> putative homologs and <i>R. norvegicus</i> PLC δ 1 PH domain.	64
Figure 3.12	The alignment of consensus secondary structure predictions for the <i>Arabidopsis</i> PDK PH domains and their <i>H. sapiens</i> and <i>C. elegans</i> PDPK1 homologs.	67
Figure 3.13	Electrostatic properties of the <i>Arabidopsis</i> PDK PH domains and their <i>H. sapiens</i> and <i>C. elegans</i> PDPK1 homologs.	68
Figure 3.14	The alignment of consensus secondary structure predictions for the <i>Arabidopsis</i> ArfGAP PH domains, <i>H. sapiens</i> Arf1 GAP PH domain, <i>H. sapiens</i> APPL1 PH domain, <i>H. sapiens</i> centaurin β 1 PH domain, <i>H. sapiens</i> centaurin β 2 PH domain and putative <i>O. sativa</i> ArfGAP PH domain homolog.	71
Figure 3.15	Electrostatic properties of the <i>Arabidopsis</i> ArfGAP PH domains, <i>H. sapiens</i> Arf1 GAP PH domain, <i>H. sapiens</i> APPL1 PH domain, <i>H. sapiens</i> centaurin β 1 PH domain, <i>H. sapiens</i> centaurin β 2 PH domain and putative <i>O. sativa</i> ArfGAP PH domain homolog.	72
Figure 3.16	The alignment of consensus secondary structure predictions for the <i>Arabidopsis</i> RhoGAP PH domains and <i>H. sapiens</i> RhoGAP21, <i>H.</i>	

	<i>sapiens</i> Oligophrenin 1, <i>C. elegans</i> GAP and putative <i>O. sativa</i> RhoGAP PH domain homologs.	74
Figure 3.17	Electrostatic properties of the <i>Arabidopsis</i> RhoGAP PH domains and <i>H. sapiens</i> RhoGAP21, <i>H. sapiens</i> Oligophrenin 1, <i>C. elegans</i> GAP and putative <i>O. sativa</i> RhoGAP PH domain homologs.	75
Figure 3.18	The alignment of consensus secondary structure predictions for the <i>Arabidopsis</i> DYN PH domains, <i>H. sapiens</i> dynamin 1, <i>H. sapiens</i> dynamin 2 and putative <i>O. sativa</i> dynamin.	78
Figure 3.19	Electrostatic properties of the <i>Arabidopsis</i> DYN PH domains, <i>H. sapiens</i> dynamin 1, <i>H. sapiens</i> dynamin 2 and putative <i>O. sativa</i> dynamin.	79
Figure 3.20	The alignment of consensus secondary structure predictions for the <i>Arabidopsis</i> AtPH PH domains, <i>H. sapiens</i> N-terminal pleckstrin, <i>H. sapiens</i> C-terminal pleckstrin, putative <i>O. sativa</i> homolog and <i>M. musculus</i> cytohesin-3.	81
Figure 3.21	Electrostatic properties of the <i>Arabidopsis</i> AtPH PH domains, <i>H. sapiens</i> N-terminal pleckstrin, <i>H. sapiens</i> C-terminal pleckstrin, putative <i>O. sativa</i> homolog and <i>M. musculus</i> cytohesin-3.	82
Figure 3.22	The alignment of consensus secondary structure predictions for the <i>Arabidopsis</i> DUF828 PH domains and their putative <i>V. vinifera</i> and <i>B. oleracea</i> homologs.	84
Figure 3.23	Electrostatic properties of the <i>Arabidopsis</i> DUF828 PH domains and their putative <i>V. vinifera</i> and <i>B. oleracea</i> homologs.	85
Figure 3.24	The alignment of consensus secondary structure predictions for the <i>Arabidopsis</i> UNK 1-2 PH domains and <i>C. elegans</i> muscle protein UNC-89.	86
Figure 3.25	Electrostatic properties of the <i>Arabidopsis</i> UNK 1-2 PH domains and <i>C. elegans</i> muscle protein UNC-89.	87

Figure 4.1	A. Multiple sequence alignment of the FYVE-finger domain sequences of fifteen <i>Arabidopsis</i> proteins and EEA1 C-terminal FYVE domain. B. A schematic localization of <i>Arabidopsis</i> FYVE proteins within the <i>Arabidopsis</i> genome.	90
Figure 4.2	Cartoon of representative examples of the five different classes of <i>Arabidopsis</i> FYVE proteins.	93
Figure 4.3	Sequence alignment of human RCC1 and putative <i>Arabidopsis</i> homologues (class V).	95
Figure 4.4	The electrostatic profiles of five known FYVE domain structures and fifteen <i>Arabidopsis</i> FYVE homology models i.e. classes I – V.	98
Figure 4.5	The electrostatic profiles and sequence alignment of <i>Arabidopsis</i> class I FYVE homologues from <i>S. cerevisiae</i> (Fab1p) and <i>C. elegans</i> (hypothetical protein ppk-3).	99
Figure 4.6	A. Alignment of six FYVE domains representing class I – IV of <i>Arabidopsis</i> FYVE proteins. B. Alignment of nine FYVE domains representing class V of <i>Arabidopsis</i> FYVE proteins.	101
Figure 4.7	Alignment of classic and variant FYVE domains.	103
Figure 4.8	Schematic representation of the coordination of PtdIns(3)P headgroup, Ins(1,3)P ₂ , and PtdIns(5)P headgroup, Ins(1,5)P ₂ , by the conserved sequence motif of the class IIA <i>Arabidopsis</i> AT3G43230-FYVE (classic) and the class V <i>Arabidopsis</i> AT1G65920-FYVE (variant).	104
Figure 0.1.	Protein tree of <i>Arabidopsis</i> PH domains.	117
Figure 0.2.	Protein tree of <i>Arabidopsis</i> FYVE domains.	118

Chapter 1 Introduction

Arabidopsis thaliana

Arabidopsis thaliana is a small flowering annual dicotyledonous plant of the *Brassicaceae* family. It has a broad natural distribution throughout Europe, Asia, and North America [1]. There are many different ecotypes, with the Columbia and Landsberg ecotypes accepted as the standards for biological studies [1]. It has no agronomic significance but it is of immense scientific utility as a model organism. It has emerged as a highly suitable model for plant biology as first suggested by Laibach [2]. Some of the advantages offered by *Arabidopsis* include large number of progeny, rapid development (six-week life cycle), easy cultivation in limited space, abundant natural variation, fertile hybrid production and a relatively low chromosome number [2, 3].

The concept of *Arabidopsis* as a model plant organism did not hold until the mid/late 1980s [3] when transformation became as simple as spraying the bacteria, *Agrobacterium tumefaciens*, which contained a gene of interest in a plasmid onto the *Arabidopsis*' flowers [3, 4] and when it was demonstrated that *Arabidopsis* has a small genome amenable to detailed molecular analysis [1]. Consequently, a large collection of mutants was created and many fundamental biological discoveries were made using *Arabidopsis*. Some areas where *Arabidopsis* has been used to address questions pertaining to plant biology include cell morphogenesis [5], seed coat development [6], root development [7], floral induction [8, 9], flower and fruit development [10, 11], plant light perception [12, 13], plant disease resistance [14], vitamin synthesis in plants [15], fatty acid signaling in plants [16], plant-pathogen interactions [17], plant communication [18], and plant hormone action [19, 20].

Arabidopsis thaliana was the first higher plant to have its genome fully sequenced [21]. The genome is contained within five chromosomes. Its total size is approximately 135 megabases (TAIR Genome Assembly webpage, 2006). It is interrupted by five centromeric regions, five 5S rDNA regions and twenty other gaps and potential gaps (TAIR Genome Assembly webpage, 2006). The *Arabidopsis* Genome Initiative (AGI) sequencing effort predicted approximately 25,500 protein-coding genes [21], of which 11,794 have been structurally verified by full-length

complementary DNA (cDNA) sequencing as of 2003 [22].

About 58-60% of the *Arabidopsis* genome is present in 24 duplicated segments, each of more than 100 kb in size [21, 23]. As a result many genes have at least one duplicate, which may or may not have undergone further rearrangements, such as local inversions, after duplication events. For example, all members of the SHATTERPROOF gene family of MADS-box transcription factors share the same redundant function. Since there are multiple copies of the same genes, there are also multiple copies of the same protein present in the *Arabidopsis*. In fact, 65% of the predicted proteins are encoded more than once in the genome [23], and duplicate or multiple copies of the same protein are anticipated in the *Arabidopsis* proteome. About half of all *Arabidopsis* genes appear to be closely related in sequence to genes found in other organisms ranging from bacteria to humans [1].

After the completion of the *Arabidopsis* genome [21], the community of plant biologists proposed a new initiative known as the AT2010 or “The 2010 Project” [24]. The goal of the AT2010 was to use *Arabidopsis* genome to gain understanding of the molecular function of all plant genes by the year 2010 [24]. The process of annotation can be defined as the attachment of information to a sequence [25]. The information may include the sequencing center, experimental source, name, description, function, classification, etc [25]. The annotation can be performed automatically or manually. The sequenced *Arabidopsis thaliana* genome was initially annotated by the AGI sequencers [21]. A total of 25,498 predicted gene functions were analyzed [21, 23]. The initial annotation of the *Arabidopsis* genome relied on the manual synthesis of the output from several prediction programs, including Genscan and GenemarkHMM with extrinsic data such as protein similarities and expressed sequence tag (EST) matches [21]. Recently, the genome was re-annotated using full-length cDNA sequence data available from many projects [26]. The functional reannotation was based on either protein sequence similarity information to annotate functional categories as in the case of The Institute for Genomic Research (TIGR), or various sources of data such as the literature, sequence analysis, other database annotations, and community input as in the case of The *Arabidopsis* Information Resource (TAIR) [26]. The re-annotation was made compatible via vocabulary terms developed by the Gene Ontology

Consortium. As of 2002, 10,302 unique genes have been functionally annotated [26].

Once genes are identified, the protein sequence that they encode form the basis for a second layer of annotation [25]. The automatic proteome annotation for the *Arabidopsis* predicted proteins was based on the yeast proteome and the assumption that sequences conserved between the yeast and *Arabidopsis* genomes reflect common functional relationship [23]. In this way, it was possible to assign a likely function to approximately 69% of *Arabidopsis* protein-coding genes [23, 27]. These are mostly hypothetical functions, since only 9% of the genes have been experimentally characterized [27]. The remaining 31% of the predicted protein-coding genes could not be assigned to functional categories [23, 27]. Functional annotation was performed with Gene Ontology [28, 29] and the Munich Information Center for Protein Sequences (MIPS) functional classifications [30-32].

Phosphoinositides and phosphoinositide-binding domains

Phosphoinositides (PIs) are a family of inositol-containing phospholipids found in the membrane of all eukaryotic cells. Eukaryotic cell membranes are composed of phosphatidylserine, phosphatidic acid (PA), phosphatidylinositol and much less abundant (approximately 10% of total phospholipids) but functionally critical PIs (Table 1.1). PIs include $\text{PtdIns}(3)P$, $\text{PtdIns}(4)P$, $\text{PtdIns}(5)P$, $\text{PtdIns}(3,4)P_2$, $\text{PtdIns}(4,5)P_2$, $\text{PtdIns}(3,5)P_2$ and $\text{PtdIns}(3,4,5)P_3$ (Fig. 1.1; see [33-35] for a recent review). In unstimulated eukaryotic cells, by far the most abundant PIs are $\text{PtdIns}(4)P$ and $\text{PtdIns}(4,5)P_2$ followed by $\text{PtdIns}(3)P$ and $\text{PtdIns}(5)P$, which are found in lower levels (Table 1.1). The levels of $\text{PtdIns}(3,4)P_2$, $\text{PtdIns}(3,5)P_2$ and $\text{PtdIns}(3,4,5)P_3$ are very low in unstimulated cells and as such these PIs may be undetectable by standard assay procedures. However, it is known that the levels of $\text{PtdIns}(3,4)P_2$, $\text{PtdIns}(3,5)P_2$ and $\text{PtdIns}(3,4,5)P_3$ can rise dramatically upon stimulation [36], for example in response to hyperosmotic stress as it has been shown for $\text{PtdIns}(3,5)P_2$ in yeast, plants and animal cells [37]. Plants differ from other eukaryotic organisms in that two PIs, $\text{PtdIns}(3,4)P_2$ and $\text{PtdIns}(3,4,5)P_3$, have not been detected in any of their tissues [38, 39].

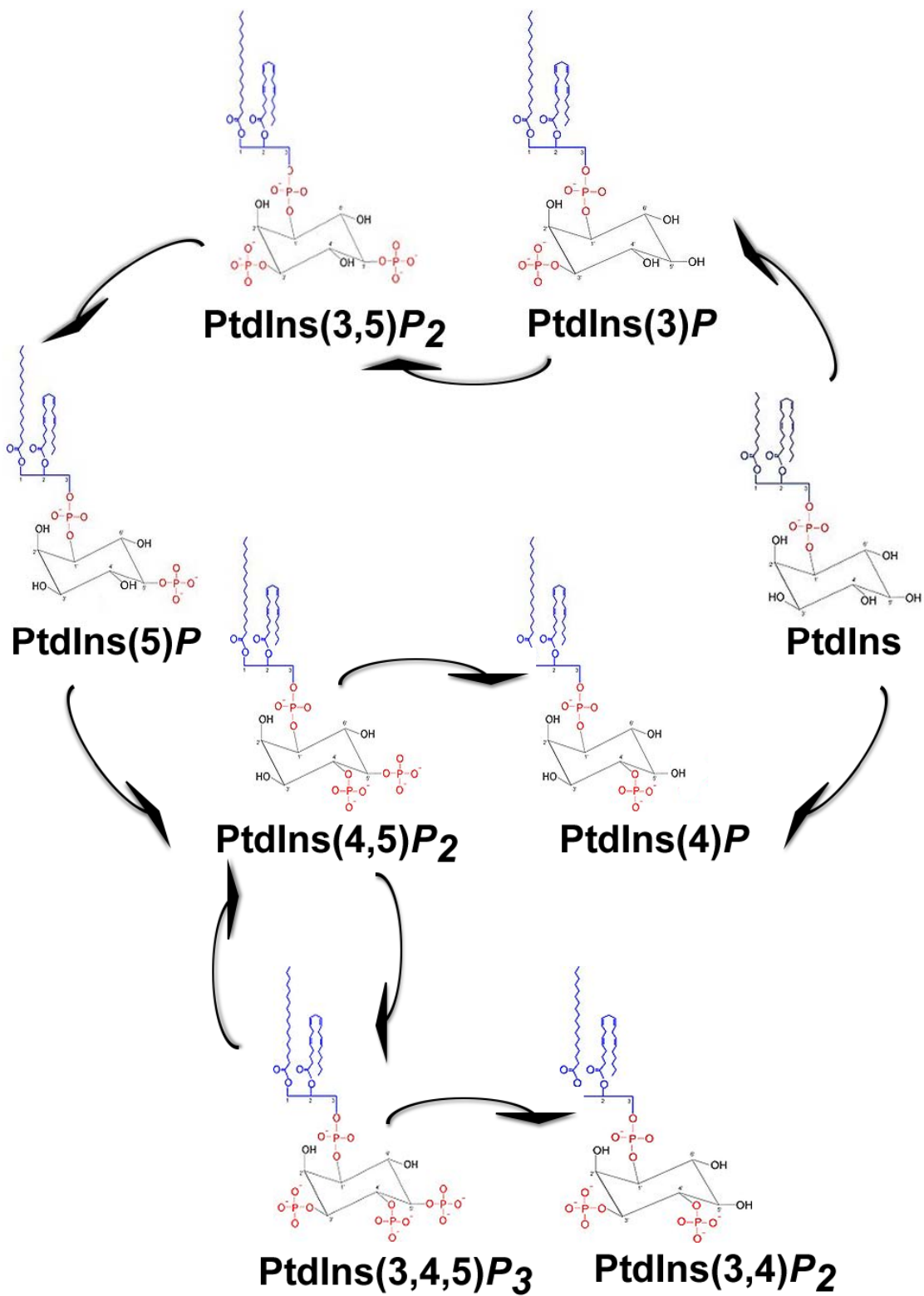


Figure 1.1. PI metabolism. Figure adapted from (Sasaki *et al.*,) [34].

Table 1.1. Localization and abundance of PIs in unstimulated eukaryotic cells.

PIs	Localization				Abundance			
	Cytostolic	Ref	Nuclear	Ref	Mammalian	Ref	Plant	Ref
PtdIns(3) <i>P</i>	Early endosome	[33]	Matrix	[35]	0.2%	[36]	12%	[40]
PtdIns(4) <i>P</i>	Golgi apparatus		Matrix		5%		80%	
PtdIns(5) <i>P</i> *	Not yet clear		Matrix, chromatin		0.2%		8%	
PtdIns(4,5) <i>P</i> ₂	Plasma membrane		Matrix, chromatin		5%			
PtdIns(3,4) <i>P</i> ₂	Plasma membrane		Matrix		0.01		***	[39]
PtdIns(3,5) <i>P</i> ₂ **	Late endosome				0.01			
PtdIns(3,4,5) <i>P</i> ₃	Plasma membrane	[33]	Matrix	[35]	0.01	[36]	***	[38]

*PtdIns(5)*P* has not been detected in yeast [41, 42].

**PtdIns(3,5)*P*₂ levels increase by up to 30-fold in yeast and 2-6-fold in plants and animal cells in response to hyperosmotic stress [37]

***not yet reported

Note: precise levels vary considerably depending on the cell type and growth conditions [42].

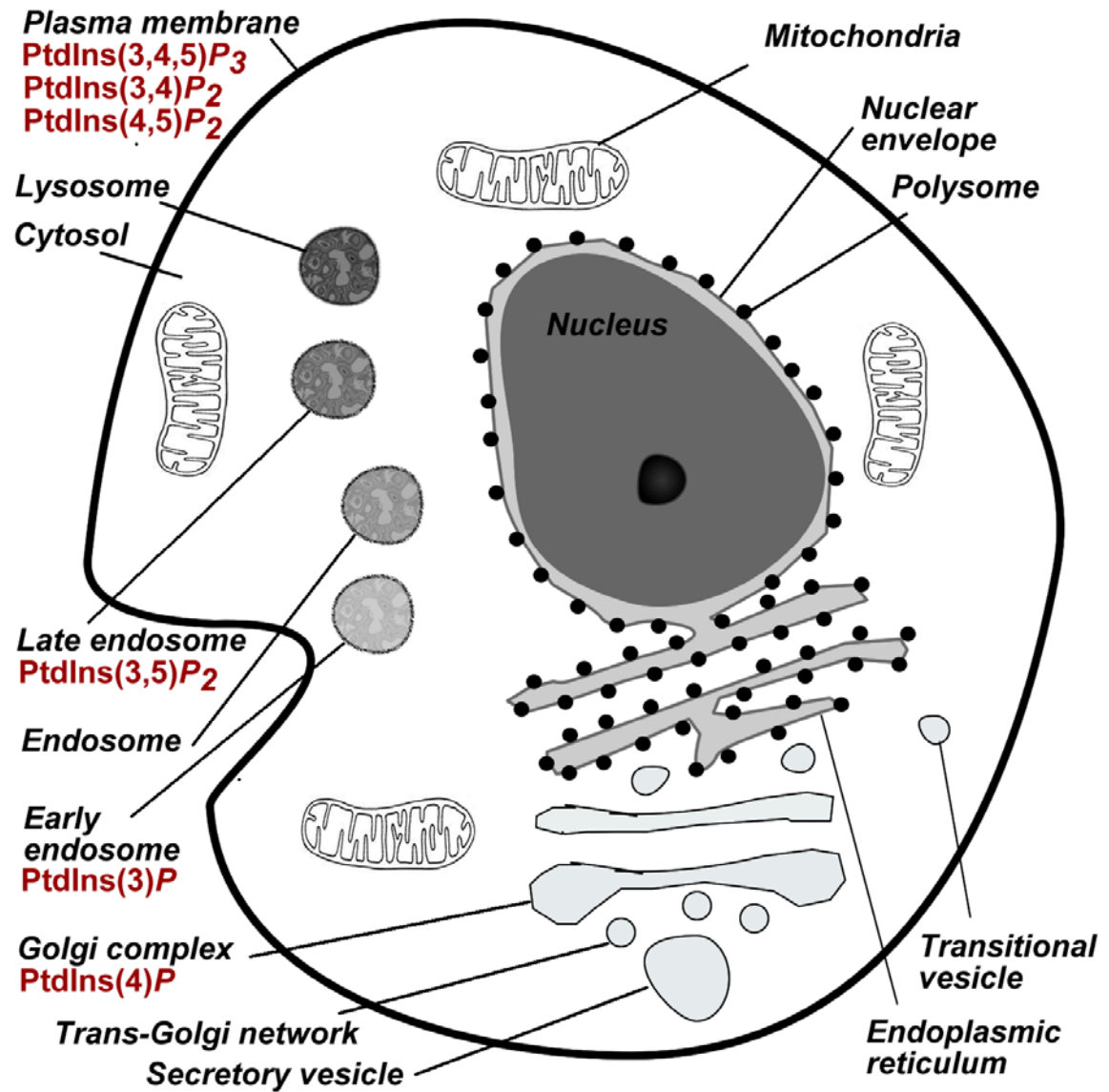


Figure 1.2. Differential distribution of PIs in membrane of various organelles present in eukaryotic cells. Only the most predominant PIs are indicated for each organelle.

PIs have been shown to exhibit distinct membrane-based sub-cellular and nuclear distribution patterns (Table 1.1; Fig. 1.2). For example, PtdIns(4)*P* is predominantly found in the Golgi membranes and nuclear matrix [33, 35, 43]. Based on their differential distribution it has been proposed that PIs contribute to the organelle's identity by serving as ligands, which allow for PI-binding proteins to associate with a membrane of a particular organelle [43-47]. In most cases, the protein associated with the membrane has a specific lipid-binding domain involved in the PI-binding, such as Pleckstrin Homology (PH) or Fab1p, YOTB, Vac1 and EEA1 (FYVE) domain [48]. The interactions of these domains with the PI as well as membrane surfaces are either highly specific or non-specific, driven by charge, amphiphilicity and curvature or a combination of these (Table 1.2). Moreover, penetration of hydrophobic residues in the membrane interface can further enhance the protein-membrane interaction. For example, molecular dynamics simulations have revealed that PtdIns(4,5)*P*₂ binding to the PH domain of phospholipase C δ 1 facilitates tighter intra-complex interaction and increases membrane penetration [49]. Similarly, PtdIns(3,4,5)*P*₃ binding to the General receptor for phosphoinositides isoform 1 (Grp1) PH domain triggers insertion of the Grp1 PH domain into the membrane as demonstrated by NMR, monolayer surface tension and surface plasmon resonance experiments [50].

Table 1.2. The specificity of PH domains.

Protein	Specificity	Reference
PLC δ 1-PH	PtdIns(4,5) P_2	[51, 52]
β -ARK-PH	PtdIns(3,4,5) P_3 , PtdIns(4,5) P_2	[53-55]
Spectrin-PH	PtdIns(3,4,5) P_3 , PtdIns(4,5) P_2	[54, 55]
Pleckstrin (N)-PH	PtdIns(3,4,5) P_3 , PtdIns(4,5) P_2	[53, 54]
Grp1-PH	PtdIns(3,4,5) P_3	[54, 56]
Btk-PH	PtdIns(3,4,5) P_3	[55, 57]
mSos-PH	PtdIns(3,4,5) P_3 , PtdIns(4,5) P_2	[50, 55]
Atk/PKB-PH	PtdIns(3,4,5) P_3 , PtdIns(3,4) P_2	[58, 59]
DAPP1-PH	PtdIns(3,4,5) P_3 , PtdIns(3,4) P_2	[60]
TAPP1-PH	PtdIns(3,4) P_2	[61]
TAPP2-PH	PtdIns(3,4) P_2	[61]
FAPP1-PH	PtdIns(4) P	[61]
AtPH1-PH	PtdIns(3) P	[61]
PEPP1-PH	PtdIns(3) P	[61]
centaurin- β 2-PH	PtdIns(3,5) P_2	[61]

Phosphoinositides and phosphoinositide-binding domains in plant proteins

Plants appear to lack several of the most widely adapted signaling pathways and basic cellular regulators found in vertebrates, flies and worms, such as the Wingless and Hedgehog pathways, or cyclin-dependent kinases [27]. Instead, they make use of different signal transduction pathways and cellular regulators to perform similar functions [27]. Some of the cellular regulators used by plants include PIs. Thus far, no homologues of the class I PI3-kinase, which can phosphorylate $\text{PtdIns}(4)P$ and/or $\text{PtdIns}(4,5)P_2$ on their 3-position, have been found in plants, no $\text{PtdIns}(3,4,5)P_3$ has been detected in any plant tissue [38] and formation of $\text{PtdIns}(3,4)P_2$ in *Arabidopsis* has not been reported [39] suggesting that even though plants utilize PIs in cellular pathways, they do so with marked differences from other eukaryotic organisms. The remaining PIs have all been detected in plants but their exact roles in signaling pathways are mostly unknown [38]. Similarly, it is not known whether plant PI-binding proteins use similar molecular mechanisms and play similar roles in plants as seen in other well-studied organisms.

A. PH domains

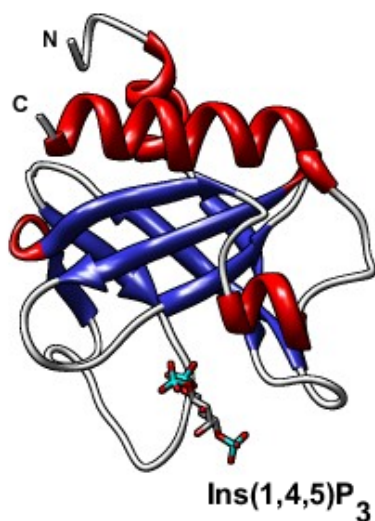
The PH domain is the eleventh most abundant domain class in the human genome [62]. It was originally identified in 1993 [63-65] as a stretch of 100-120 amino acids in the hematopoietic protein Pleckstrin [66]. PH domains are now known to occur in more than 100 different proteins involved in intracellular signaling, protein trafficking, cytoskeletal organization, and modification of membrane phospholipids, the majority of which require membrane association to elicit their function [54, 65, 67, 68]. Reported ligands for PH domains include PIs [53, 55, 58, 69-71], InsPtds [52, 58, 69, 72-75], $\beta\gamma$ of heterotrimeric G proteins [76, 77], protein kinase C [78], filamentous actin [79], and WD motifs (also known as the Trp-Asp or WD40 motifs) [80].

1. Structure

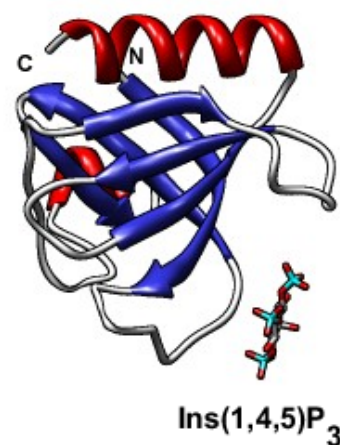
PH domains fold up into essentially the same structure (Fig. 1.3) as first observed by Yoon et al. 1994 [81], even though, on average, the pairwise sequence comparison of unrelated

PH domains suggests a low sequence identity of between 7% and 23% [82]. PH domains have seven β -strands separated into two nearly orthogonal β -sheets. One sheet consists of four β -strands (β 1- β 4) and the other of just three (β 5- β 7). The two nearly orthogonal β -sheets are closed off at one corner by a C-terminal amphipathic α -helix. With the exception of the recently determined structure of Unc89 PH domain [83], the experimentally solved structures of all other PH domains are electrostatically polarized [84] with the three most variable loops, i.e. β 1/ β 2, β 3/ β 4, and β 6/ β 7, falling in the positively charged half of the domain [63-65, 84-86]. Based on experimental evidence, from both nuclear magnetic resonance and X-ray crystallography studies, the positively charged surface surrounds the ligand-binding site for PIs [53, 57, 74, 82, 87-89], which is also known as the 'canonical membrane binding site'.

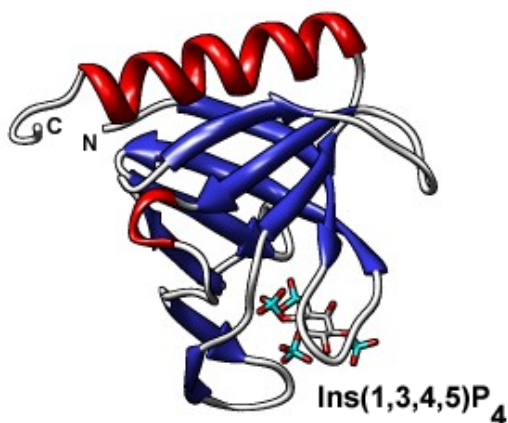
A. PLC δ 1 PH



B. Spectrin PH



C. Grp1 PH



D. Akt PH

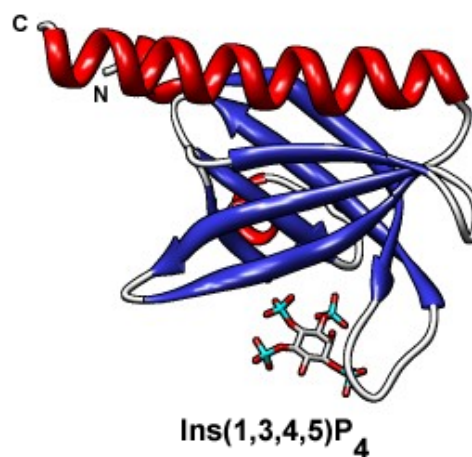


Figure 1.3. Similarities and differences between the PI-binding PH domains. PI binding of PH domains mostly occurs in a pocket formed by the β 1/ β 2 and β 3/ β 4 sheets and connecting loops (A, D). The spectrin PH domain binds the PI ligand using the β 1/ β 2 and β 5/ β 6 sheets and connecting loops (B) whereas the Grp1 PH domain uses, additionally, a long insertion within the β 5/ β 6 loop (C). The PDB accession numbers of these PH domain structures are 1MAI, PLC δ 1 PH [90], 1BTN, Spectrin PH [74], 1FHX, Grp1 PH [91], and 1H10, Akt PH [92]. Molecular graphics images were produced 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) [93].

2. Phosphoinositide binding mechanism

Most, if not all, PH domains are thought or shown to bind PIs [53, 68, 94, 95]. The binding specificity and affinity depends on both the individual PH domain and the respective PI ligand [54, 55, 96]. In approximately 10% of cases, the PI binding is strong and highly specific [48, 53, 54, 68, 90]. The majority of PH domains, however, have been reported to bind PIs or their head groups, the InsPtds with weak or no specificity [54, 55, 96, 97]. Although no specific signatures have been identified that link PH domain specificity for particular PIs, Isakoff *et al.* derived a consensus sequence, which successfully predicts high affinity binding to PtdIns(3,4) P_2 and/or PtdIns(3,4,5) P_3 [98]. The signature is centered on the loop that connects strands $\beta 1$ and $\beta 2$ of the PH domain, and may be written as suggested by Isakoff *et al.* [98] and adapted from Lemmon and Ferguson [48] as $\Phi xKx[G/A/S/P]x^*[K/R]^*x[R/K]xRx[F/L]$, where Φ denotes an amino acid with a hydrophobic side chain, x represents any amino acid, and * represents any number of amino acids. The mutation of conserved phospholipid-binding residues tryptophan 26 and arginine 29 generates proteins with decreased PtdIns(3,4,5) P_3 binding, and subsequently decreased protein retention near the receptor and less sustained signaling [99, 100].

The variable affinity of PH domain for different PIs may be one of the mechanisms underlying targeting of PH proteins to many different cellular compartments such as the plasma, cytosol, cytoskeleton, various organelles, and the nucleus. Therefore, characterizing PH domains and their binding affinities for PIs can greatly enhance the understanding of protein localization, function and regulation in the organism being investigated [101, 102].

3. Function in plants

Despite the functional importance of PH domains in signal transduction pathways as membrane targeting modules or by interaction with other cellular components [103], the PH domains are fairly under-investigated in plants. There are only a handful of experimental studies dealing with PH domains in plants. Specifically, there are three studies that have looked at PH domains and their ability to bind different phospholipids in *Arabidopsis*: (a) PI4-kinase was found

to selectively bind PA, PtdIns(4)*P*, PtdIns(4,5)*P*₂ [102]; (b) *Arabidopsis thaliana* pyruvate dehydrogenase kinase 1 (PDK1) was found to interact with PA, PtdIns(3)*P*, PtdIns(3,4)*P*₂ and PtdIns(3,4,5)*P*₃ and to a lesser extent with PtdIns(4)*P* and PtdIns(4,5)*P*₂ [104]. The human PDK1 acts to inactivate the enzyme pyruvate dehydrogenase by phosphorylating it using ATP. PDK1 and the pyruvate dehydrogenase complex are located in the mitochondrial matrix of eukaryotic cells (Fig. 1.2); and (c) *Arabidopsis* dynamin-like 6 (ADL6) was identified to bind PtdIns(3)*P* with high affinity *in vivo* and PtdIns(3)*P* or PtdIns(4)*P* *in vitro* [101]. There is only one study that has examined the functional role of PI4-kinase PH domain and concludes that it acts as an allosteric modulator for the protein's enzyme activity [105]. To date there is no study that has examined in detail, the role of PH domains in plants either at the level of the individual protein or from a genomic perspective using experimental or computational approaches.

B. FYVE domains

The FYVE lipid-binding domains were named after the first letter of the four proteins in which they were originally discovered: Fab1, YOTB, Vac1, and EEA1 [106]. FYVE proteins have primarily been associated with functions related to endosomal trafficking e.g. a FYVE hepatocyte growth factor-regulated tyrosine kinase substrate (Hrs) is involved in sorting of down-regulated receptor molecules in early endosomes [107], a FYVE domain-containing vacuolar protein sorting protein 27 (Vps27p) has been implicated in endosome maturation [108], whereas a FYVE EEA1 has been shown experimentally to function in endocytic membrane fusion [109] and regulation of endosome-to-TGN retrograde transport via phosphatidylinositol 3-phosphate 5-kinase (PIKfyve) [110]. However, they may play other important roles in cell signaling as exemplified by a FYVE Faciogenital dysplasia 1 (Fgd1), which functions in cytoskeletal regulation [111], a FYVE Fab1p involved in regulation of membrane homeostasis [112-114], whereas a FYVE Smad Anchor for Receptor Activation (SARA) and a FYVE endofin in growth factor signaling [115-118].

1. Structure

Structurally, FYVE domains share a fold comprising of two small double-stranded β-sheets and a C-terminal α-helix (Fig. 1.4) as deduced from experimentally solved structures such

as the crystal structure of the FYVE domain from yeast Vps27p [119]. The fold is stabilized by eight Zn^{2+} coordinating cysteines residues, which bind Zn^{2+} in pairs such that the first and third pairs bind one zinc atom, while the second and fourth pairs bind the other zinc atom [119]. The FYVE domains have been characterized as PI-binding domains that are highly specific for $\text{PtdIns}(3)P$ [120-123]. This ligand recognition is Zn^{2+} -dependent [124] and stems primarily from a conserved ligand-binding motif, i.e. (R/K)(R/K)HHCR surrounding the third and fourth cysteine residues [119]. Mutagenesis of either the cysteines involved in Zn^{2+} coordination or the ligand binding conserved residues results in decreased affinity for $\text{PtdIns}(3)P$ [120, 124-126].

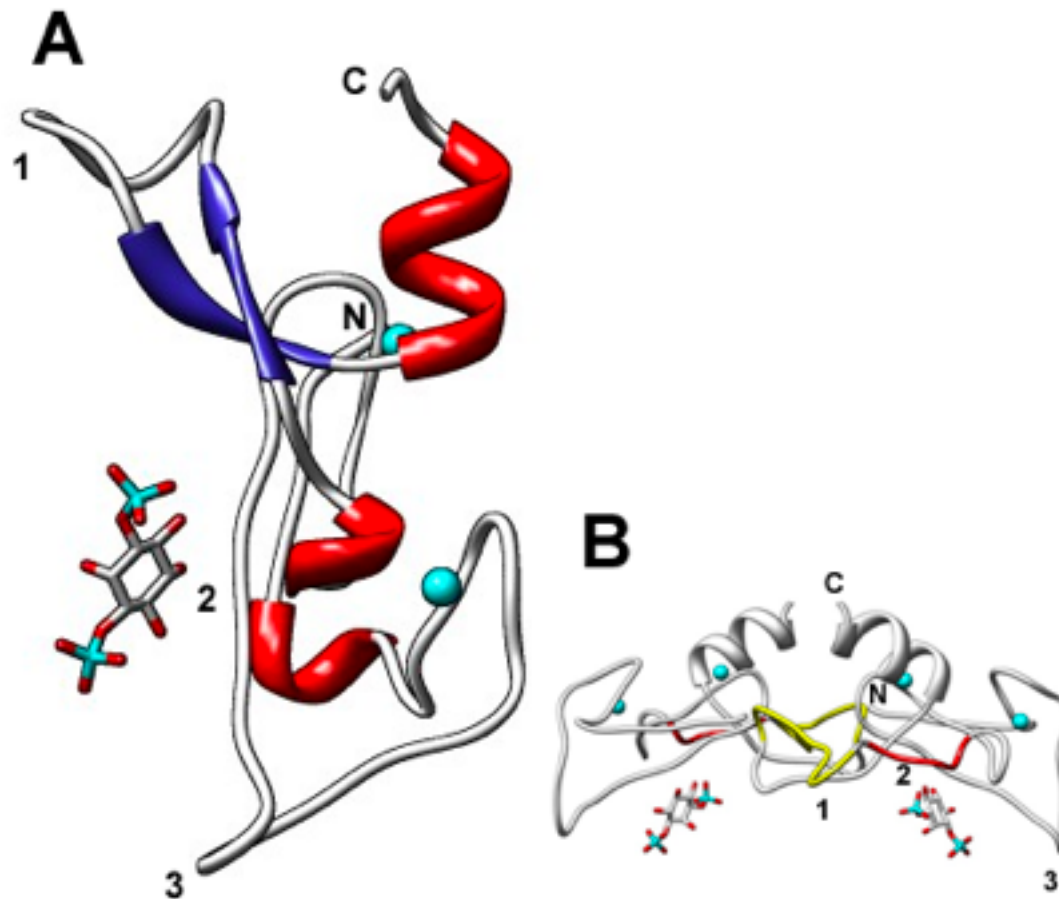


Figure 1.4. The FYVE domain of 1JOC, EEA1 [127] as a monomer (A) and a dimer (B). α -helices are colored red and β -strands are colored blue. The membrane insertion loop (MIL), Ins(1,3) P_2 binding pocket and dimerization region are labeled 1, 2 and 3, respectively. Molecular graphics images were produced 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) [93].

2. Phosphoinositide binding mechanism

The PtdIns(3)*P* binding signature contains three classic conserved regions: the N-terminal WxxD, the central R(R/K)HHCR and the C-terminal R(V/I)C motifs [128], which drives the PtdIns(3)*P* specific membrane recruitment of FYVE domains. However, there are several factors in addition to PtdIns(3)*P* binding that are thought to contribute to the membrane affinity of FYVE domains: nonspecific electrostatic interactions between basic face of the domain and the anionic membrane surface [129-131], hydrophobic interactions between the residues located in the “turret loop” near the PtdIns(3)*P* binding pocket and the membrane bilayer [119, 129, 131-133] and dimerization [124, 134]. In addition to working out the structural and functional role of various amino acids comprising the binding motifs, it has also been shown that the binding of PtdIns(3)*P* to the ligand-binding pocket of FYVE domains neutralizes nearby basic residues to reduce the local positive potential and allow conserved hydrophobic residues to penetrate the membrane interface enhancing membrane attachment [129, 130, 133]. It has been well documented that the PtdIns(3)*P*-binding FYVE domains are well conserved from yeast to mammalian cells and PtdIns(3)*P* is also present in various organelles of plant cells [135]. It is possible to envision that plant cells utilize the same or highly similar lipid binding and membrane targeting mechanism [135] for FYVE domains given that both the FYVE domains and type III PI3-kinase, which makes PtdIns(3)*P*, are present in plant cells [136]. However some recent reports suggest that PtdIns(3)*P* may not be the only lipid-ligand of plant FYVE domains and that PtdIns(5)*P* could also bind to selected plant FYVE domains [137].

3. Function in plants

Similar to PH domains, FYVE domains are fairly unknown and have not been investigated much in plants. There are some experimental studies dealing with FYVE domains in plants, which are restricted to a few individual proteins [128, 137, 138], while most FYVE domains remain uncharacterized. Specifically, there are two studies that have looked at *Arabidopsis* FYVE domains and their PI-binding preferences. In the first study, the authors examined *Arabidopsis*

FYVE domains for their binding specificity and identified classic vs. variant FYVE domains [128]. In the second study, the authors characterized variant *Arabidopsis* FYVE domain present in PRAF-1 proteins and identified its novel PtdIns(5)*P* specificity [137]. As in the case of PH domains, to date there is no study that has examined in detail, the role of FYVE domains in plants either at the level of the individual protein or from a genomic perspective using experimental or computational approaches.

Project justification

Technical difficulties in tracing lipids in plant tissues have hampered the progress of PI signaling research in plants. Consequently we are just beginning to understand it. However, many studies have shown that phospholipids and their binding partners are critical for various aspects of plant growth and development [139-143]. The previous scant studies on various aspects of *Arabidopsis* lipid-binding proteins, such as the PH and FYVE proteins, provide an extremely abbreviated view of the role of these domains and domain containing proteins in plant PI signaling. All studies so far lack a complete plant genome/proteome approach and there have been no attempts to compare their functional properties/roles in plants with other organisms. This study was designed to carry out a systematic examination of the lipid binding properties of the entire family of PH and FYVE proteins encoded in the *Arabidopsis* genome to address the current lack of knowledge about these two important classes of lipid binding domains in plants. The results from this study significantly increases the current understanding about PH and FYVE protein families in *Arabidopsis* and how they relate to their domain families in other organisms, with respect to prevalence, domain architecture, properties such as ligand specificity, and overall function. Our study provides the initial starting points in understanding the molecular basis of plant PI signaling and trafficking through the integrated computational approach of combining sequence and structural analyses with experimental data to give a comprehensive functional annotation of the family of *Arabidopsis* PH and FYVE proteins.

Chapter 2 Methods

***Arabidopsis* PH and FYVE proteins**

The accession numbers of the *Arabidopsis* PH and FYVE proteins were identified using a computational pipeline for automated high-throughput modeling [144], which was run against the *Arabidopsis* protein sequence database (TAIR6_pep_20051108) using the coordinates of known structures from the protein data bank (PDB) [145, 146]. The *Arabidopsis* PH and FYVE protein sequences corresponding to the identified accession numbers were retrieved from KEGG GENES [147, 148] and verified for presence of PH and FYVE domain with SMART [149-151]

Sequence verification

To verify the total number and individual accession numbers of *Arabidopsis* PH and FYVE proteins obtained with pipeline, three additional methods were employed: 1) search of publicly available sequence databases: Swiss-Prot/TrEMBL [152, 153], NCBI [154, 155], and UniProt [156-158]; 2) search of the TAIR database, which maintains a contains genetic and molecular biology data for *Arabidopsis* only [159]; and 3) MOTIF search [148] using a manually derived pattern specific to *Arabidopsis* PH and FYVE domains in PROSITE format offered by GenomeNet service [148].

Domain architecture analyses and classification

Each *Arabidopsis* PH and FYVE protein sequence was analyzed for its constituent domains and how they are organized by searching against Pfam [160], SMART [149-151], Conserved Domain Database v2.10 using conserved domain (CD)-Search [161-165] and Clusters of Orthologous Groups [166, 167], and then subgrouped according to the derived consensus domain architecture.

Verification of PH and FYVE domain boundaries

The boundaries of the *Arabidopsis* PH and FYVE domain sequences were ascertained based on the consensus output from a number of secondary structure prediction programs: Jpred [168-170], PSIPRED [171-173], PHDsec [174], Prof [175], SAM-T06 [176, 177], SAM-T99 [178], Hierarchical Neural Network [179] and SSPro [180, 181]. Protein secondary structure prediction is an important step towards understanding how proteins fold in three dimensions. 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 [182] and 79.5% with SPINE [183]. To validate our secondary structure prediction approach we have predicted secondary structure elements of solved PH and FYVE domains deposited in the PDB [145, 146]. Following the initial validation, secondary structure prediction was performed for all PH and FYVE domain sequences. This step was performed to ensure that the sequences encompass all of secondary structure elements characteristic of PH 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 three-dimensional structure.

Molecular modeling

Molecular modeling (schematically depicted in Fig. 2.1) also known as comparative modeling, is a class of methods used for constructing a three-dimensional model of a protein from its amino acid sequence, often referred to as a query. Molecular modeling relies on the identification of a solved protein structure, called a template, which is closely related to the query sequence, and on alignment with the query, maps residues its residues to the corresponding residues in the query sequence. The sequence alignment and template structure are then used to produce a structural model of the query using a modeling program. However, there is no single modeling program that produces reliable models for all sequences at all times [184]. Therefore to generate high-quality models for the *Arabidopsis* FYVE and PH domains, 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 (schematically depicted in Fig. 2.2) [144] and the web server Pudge [185]. 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.*, [144] for a detailed description). The manual approach was based on a scheme previously applied by Singh and Murray [186]. 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 [187] and the prediction of side-chain conformations with SCWRL3.0 [188] and SCAP [189-191].

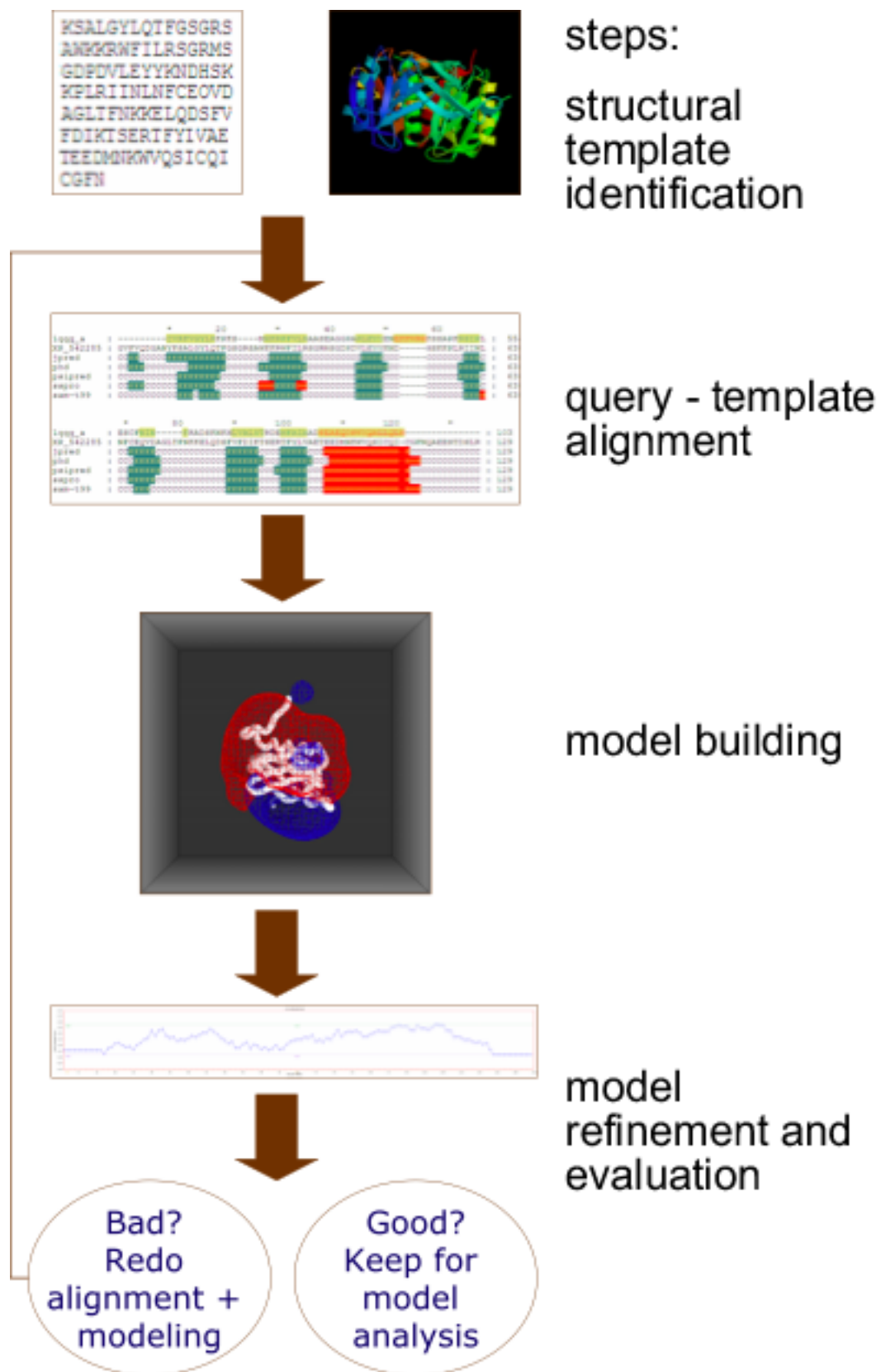


Figure 2.1. A generic flow-chart for molecular (comparative) modeling.

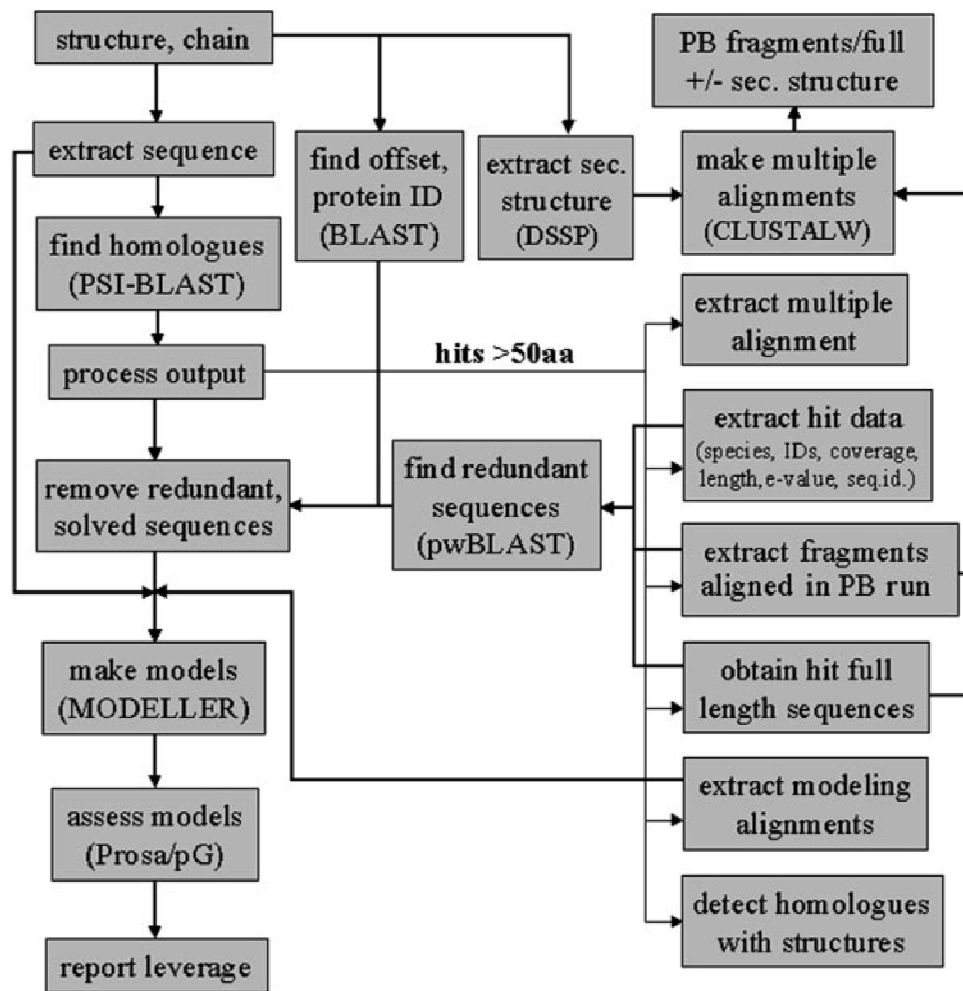


Figure 2.2. Flowchart of the automated comparative modeling pipeline as extracted from Mirkovic *et al.*, [144].

A. Structural templates

Currently, there are over twenty non-redundant structures of PH domains [192] and over five non-redundant structures of FYVE domains available in the PDB database (Table 2.1). All templates were identified as suitable structural templates for the respective sequences based on: 1) fold recognition as implemented by 123D+ [193], FUGUE [194], LOOPP [195-197], and PHYRE [198]; 2) sequence and structure homology via BLAST and PSI-BLAST [159, 199-201], 3D-JIGSAW [202-204], automatic and manual Homology Modeling with HOMER [205, 206], and CPH [207]; 3) manual multiple sequence alignment using the alignment editor, GeneDoc [208], and 4) Rosetta *ab initio* modeling [209-215].

For example, the AT1G65920 PH domain sequence was submitted, among others, to the CPH [207], LOOPP [195-197] and FUGUE [194] programs for template identification. The CPH [207] program reported the best match to have the PDB ID code: 1MAI [90] (E-score 1e-05), which is a crystallized structure of the PH domain from phospholipase C δ (PLC δ [90]). The LOOPP [195-197] program reported PDB ID codes: 1MAI [90] (score 4.820) and 1PLS [81] (score 1.732) as the two best templates. Similarly, the FUGUE [194] program reported PDB ID codes: 1MAI [90] (Z-score 21.89) and 1PLS [81] (Z-score 9.22).

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.

In the case of the AT1G65920 PH domain, described earlier, the CPH [207], LOOPP [195-197] and FUGUE [194] programs agreed that the 1MAI [90] is the best template based on their scoring algorithms. Consequently, the 1MAI [90] template was then selected as the structural template for alignments performed in the next step (B).

Table 2.1. PDB accession numbers of structural templates used in the study.

	PH domain	FYVE domain
Templates	1awe, 1b55, 1bak, 1btk, 1btn, 1bwn, 1dbh, 1dro, 1dyn, 1eaz, 1fao, 1fb8, 1fgy, 1fgz, 1fho, 1fhw, 1fhx, 1foe, 1h10, 1ki1, 1kz7, 1kzg, 1lb1, 1mai, 1mph, 1nty, 1omw, 1p6s, 1pls, 1pms, 1qqg, 1rj2, 1txd, 1u27, 1u29, 1u2b, 1u5d, 1u5e, 1u5f, 1u5g, 1unp, 1unq, 1unr, 1upq, 1upr, 1v5m, 1v5p, 1v5u, 1v61, 1v88, 1v89, 1wg7, 1wgq, 1wi1, 1wjm, 1x05, 1x1f, 1x1g, 1x86, 1xcg, 1xd4, 1xdv, 1xx0, 1ym7, 1zc3, 1zc4, 1zm0, 2adz, 2bcj, 2coa, 2coc, 2cod, 2cof, 2d9v, 2d9w, 2d9x, 2d9y, 2d9z, 2da0, 2dfk, 2dhi, 2dhj, 2dhk, 2dkp, 2dn6, 2dtc, 2dx1, 2dyn, 2elb, 2fjl, 2i5c, 2i5f, 2j59, 2nz8, 2otx, 2p0d, 2p0f, 2p0h, 2pz1, 2q13, 2r09, 2r0d, 2rgn, 2uvm, 2uzr, 2uzs, 2yry, 2ys3	1dvp, 1hyi, 1hyj, 1joc, 1vfy, 1wfk, 1x4u, 1y02, 1z2q, 2yqm

B. Sequence alignment

Alignments were generated using the programs 123D+ [193], FUGUE [194], LOOPP [195-197], PHYRE [198], 3D-JIGSAW [202-204], HOMER [205, 206], CPH [207], and manually edited in certain cases using GeneDoc [208].

The following are examples of the alignments created between the amino acid of the AT1G65920 PH domain and the PLC δ (PDB ID code: 1MAI [90]) PH domain. The 1MAI [90] was identified as the best template for the *Arabidopsis* PH domain from the AT1G65920 protein in the template identification step (A).

The CPH [207] alignment:

```

Query: 1      DEQAILALKKGAQLLKCRRRGNPKFCPFKLSMDEKYLIWYSGEE-----ERQLRLSSVIT 60
Templ: 1      DDPDLQALLKGSQLLKVKSSSWRRERFYKLQEDCK-TIWQESRKVMRSPESQLFSIEDIQ 60

Query: 61     IVR-GQITPNFQKQAQSDRKEQSFSLIYANGEHTLDLICKDKAQADSWFKGLRAVI 116
Templ: 61     EVRMGHRTEGLEKFARDIPEDRCFSIVFKDQRNTLDLIAPSPADAQHWWQGLRKII 116

```

The LOOPP [195-197] alignment:

```

-----HHHHHHHHHH--EEEE-----EEEEEE-----EEEE-----EEE-EEEEEE-----HHHHH-----
Query: 2 DRTDEQAILALKKGAQLLKCRRRGNPKFCPFKLSMDEKYLIWYSGEEERQ-----LRLSSVITIVRGQITPNFQKQAQSD
Sbjct:12 GLQDDPDLQALLKGSQLLKVKSSSWRRERFYKLQEDCKTIWQESRKVMRSPESQLFSIEDIQEVRMGHRTEGLEKFARDI
-----HHHHHHHHHH--EEEEEE-----EEEEEE-----EEEE-----EEE-EEEEEE-----HHHHHH-----

-----EEEEEE-----EEEE--HHHHHHHHHHHHHHHH
Query:77 RKEQSFSLIYANGEHTLDLICKDKAQADSWFKGLRAVI 114
Sbjct:92 PEDRCFSIVFKDQRNTLDLIAPSPADAQHWWQGLRKII 129
-----EEEEEE-----EEEE--HHHHHHHHHHHHHHHH

```

The FUGUE [194] alignment:

```

          10      20      30      40      50
hs1mai    ( 12 )  glqddpdLgaLlkGsqLlKVkssswrreRfYkLgedcktIwqesrkvmr
QUERY      RDRTEQAILALKKGAQLLKCRRRGNPKFCPFKLSMDEKYLIWYSGE---
          aaaaaaa bbbbbb      bbbbbb      bbbb

          60      70      80      90      100
hs1mai    ( 61 )  spesqlFsiedIgeVrmGhrteglekfArdipedrCFSivFkdqrntLdL
QUERY      --EERQLRLSSVITIVRGQITPNFQKQAQSDRKEQSFSLIYANGEHTLDL
          bbb333bbbbbb      aaaaaa      333bbbbbb      bbb

          110      120      130
hs1mai    ( 111 )  IApspadAqhWVqGLrkiih
QUERY      ICKDKAQADSWFKGLRAVITKHHNIRNSVNHR
          b      aaaaaaaaaaaaaaa

```

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.

C. Model building

The quality of comparative models depends on the closeness of the evolutionary relationship on which they are based [216]. Table 2.1 lists PDB accession numbers of structural templates used in the study. 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 [217], 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% [218]. 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 [202-204], Modeller 8v1 [219-222], NEST [223], LOOPP [195-197], HOMER [205, 206], CPH [207], and PHYRE [198]. 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 [184].

For example, the models of the AT1G65920 PH domain were built using the CPH [207], LOOPP [195-197], Modeller 8v1 [219-222] and NEST [223] programs (Fig. 2.3). The CPH [207] and LOOPP [195-197] programs used an alignment generated by each program individually (described in sequence alignment section (B)). The Modeller 8v1 [219-222] and NEST [223] used any one of the CPH [207], LOOPP [195-197] or FUGUE [194] alignments reformatted to match the format of the Modeller 8v1 [219-222] and NEST [223] programs, respectively. For example:

The FUGUE [194] alignment reformatted for Modeller 8v1 [219-222] run:

```
>P1;1MAI
structure:1MAI: 12: : 130: : : :
-GLQDDPDLQALLKGSQLLKVKSSSWRRERFYKLQEDCKTIWQESRKVMRSPESQLFSIEDIQEVRMGHR
TEGLEKFARDIPEDRCFSIVFKDQRNTLDLIAPSPADAQHWVQGLRKIIH-----*

>P1; AT1G65920
sequence: AT1G65920: 1: :127: : : : :
RDRTDEQAILALKKGAQLLKRRRGNPKFCPFKLSMDEKYLIWYSGE-----EERQLRLSSVITIVRGQI
TPNFQKQAQSDRKEQSFSLIYANGEHTLDLICKDKAQADSWFKGLRAVITKHHNIRNSVNHR*
```

The FUGUE [194] alignment reformatted for NEST [223] run:

```
>P1;1MAI
structure:1MAI:
-GLQDDPDLQALLKGSQLLKVKSSSWRRERFYKLQEDCKTIWQESRKVMRSPESQLFSIEDIQEVRMGHR
TEGLEKFARDIPEDRCFSIVFKDQRNTLDLIAPSPADAQHWVQGLRKIIH-----

>P1; AT1G65920
sequence: AT1G65920:
RDRTDEQAILALKKGAQLLKRRRGNPKFCPFKLSMDEKYLIWYSGE-----EERQLRLSSVITIVRGQI
TPNFQKQAQSDRKEQSFSLIYANGEHTLDLICKDKAQADSWFKGLRAVITKHHNIRNSVNHR
```

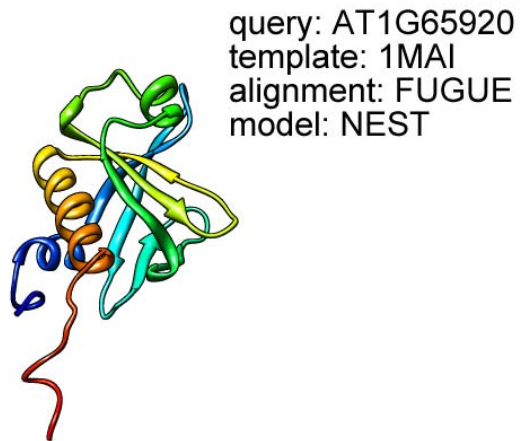
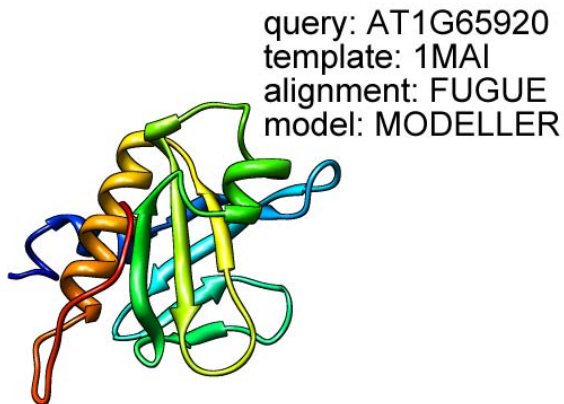
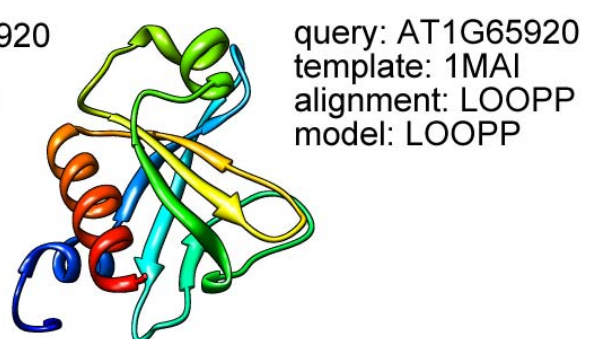
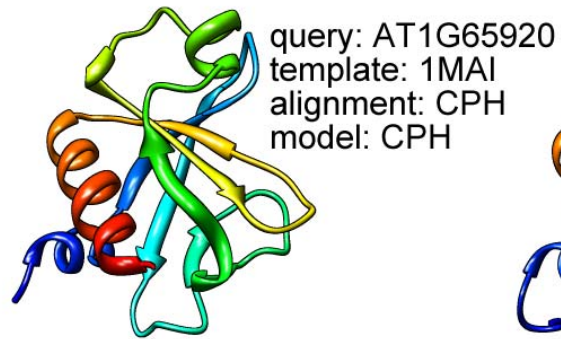
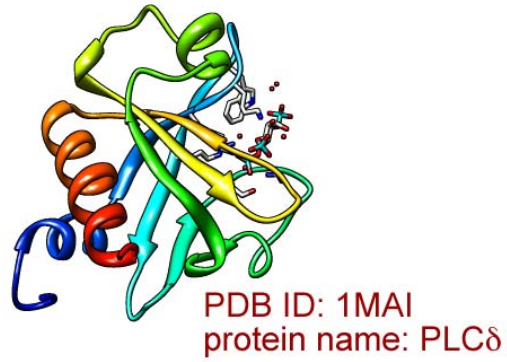


Figure 2.3. A graphical representation of the PH domain from PLC δ protein (PDB ID: 1MAI) and four models of the AT1G65920 PH domain built using PLC δ as a template.

D. Model refinement and evaluation

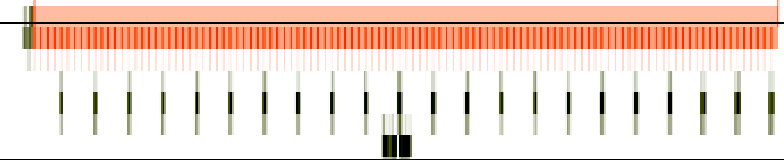
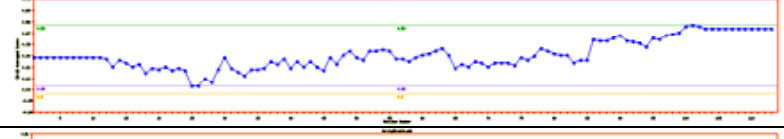
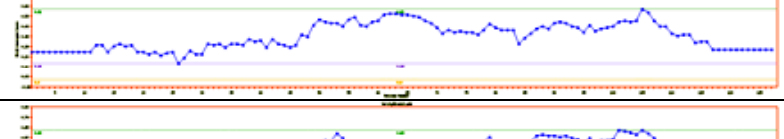
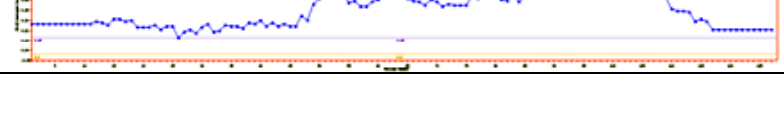
One of the current major challenges in model building is the refining of comparative models so that they can be considered reliable and match any known experimental details [216]. Following the generation of the C-alpha trace for the protein and side-chains at identical positions within the alignment, loops for which no structural information is available in the template structures remain undefined and must be constructed. The number of side chains that need to be built is dictated by the degree of sequence identity between target and template sequences. The most probable rotamers for each amino acid side chain depending on their backbone conformation are used in the final model, which is energy minimized for a good fit. However, loops being the most variable region in the sequence and packing of the side-chains often remain modeled sub-optimally.

Additional loop refinement and side chain packing iterations after creating the initial models were performed using individual modeling programs whenever available. In addition, loop refinement was done using the stand alone module, Loopy [187], and the prediction of side-chain conformations with SCWRL3.0 [188] and SCAP [189-191].

The quality of the models was assessed using Verify3D [224-226] and Prosa [227]. Verify3D tests the compatibility of a protein structure with its amino acid sequence. Prosa calculates energy profiles for a structural model by using molecular mechanics force fields. An acceptable model had a consistent Verify3D score higher than 0.2 using the default window of amino acid residues. Prosa plots are considered acceptable if they have low energy values across the entire plot. In addition, the predicted models were checked for correlation of their biophysical and structural properties with experimental observations.

Table 2.2 shows the plots and scores for the four models of the AT1G65920 PH domain depicted in Fig. 2.3. Based on these assessments, the best AT1G65920 PH domain model, Verify 3D-1D averaged score > 22 and Prosa = -4.5, was built using the FUGUE [194] alignment (described in sections B and C) and Modeller 8v1 [219-222] modeling program.

Table 2.2. Verify3D [22]
models depicted in Fig.

		Prosa
template: 1MAI alignment: CPH model: CPH		-4.37
template: 1MAI alignment: LOOPP model: LOOPP		-4.23
template: 1MAI alignment: FUGUE model: MODELLER		-4.5
template: 1MAI alignment: FUGUE model: NEST		-4.35

Analysis of the models

The models were analyzed for their sequence, structural, and biophysical properties, as detailed below.

E. 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 [228].

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++ [229-231], 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 [232], 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 [126].

F. Lipid binding and specificity via docking

1. Ligand preparation

A dataset of seven InsPtds (head groups of PIs) was assembled from the Protein Data Bank (PDB) database by extracting coordinates of various InsPtds from PDB files of determined structures and energy minimizing them. If not available in the PDB database, ligands, e.g. Ins(1,5) P_2 , 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) [93] and energy minimized. Our docking study focused only on the interaction of the InsPtds, PI head groups, with the PH and FYVE domains and did not consider the interaction of PIs lipid fragments with the membrane. Consequently, there are no lipid chains involved in the docked ligands. The generated set includes the following: one phosphatidic acid (PA), and three of each Ins(1,3) P_2 , Ins(1,4) P_2 , Ins(1,5) P_2 , Ins(1,4,5) P_3 ,

Ins(1,3,4) P_3 and Ins(1,3,4,5) P_3 . To consider all possible theoretical interactions Ins(1,3,4) P_3 and Ins(1,3,4,5) P_3 were tested even though they have not been reported in plants [38, 39]. The PA group contains only one representative ligand due to the lack of PDB files containing PA ligand and/or PDB files that could be manually manipulated to generate a PA ligand.

Table 2.3. Ligands used in the study, their parent coordinate files, resolutions, and molecular weights.

	PDB ID	Resolution[Å]	MV[g/mol]
Phosphatidic Acid	1SIW	2.20	703.990
Ins(1,3) P_2	1HYI		340.116
	1JOC	2.20	
Ins(1,4) P_2	1I9Z	1.80	
	1MAI*	1.90	
	1FGY*	1.50	
Ins(1,5) P_2	1MAI*	1.90	
	1FGY*	1.50	
	2P0D*	1.81	
Ins(1,3,4) P_3	1Z2P	1.22	420.096
	1FAO*	1.80	
	1FGY*	1.50	
Ins(1,4,5) P_3	1MAI	1.90	
	1H0A	1.70	
	2P0D	1.81	
Ins(1,3,4,5) P_3	1FAO	1.80	500.075
	1FGY	1.50	
	1H10	1.40	

* InsPtds files generated by manual deletion of phosphate groups from their corresponding PDB files.

2. Phosphoinositide docking and analysis of resulting interactions

Rigid and flexible docking was performed using DOCK 6.1 [233] and DOCK 6.1 suite of programs. A molecular surface of the receptor was created with DMS [234, 235]. Spheres were generated with Sphgen_cpp v1.2, which was modified by Andrew Magis from its original version called Sphgen [233]. The resulting file was edited to include only spheres grouped within the first cluster. Grids were generated with GRID [236, 237]. Contact scores and energy scores were calculated using an energy cutoff distance of 5.0 Å. Our docking technique was validated by docking InsPtds of known PH and FYVE domains into their corresponding solved structures. Following the initial validation we used our approach to dock InsPtd ligands using rigid and flexible docking scenarios into the modeled sequences of *Arabidopsis* PH and 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 [238].

G. Consensus motifs

1. Ligand binding regions

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* PH and FYVE domains created by the program ClustalW [239] to reveal any significant conservation patterns with emphasis on the presence or absence of consensus sequences known to bind specific PIs.

2. Subcellular localization analysis

The eukaryotic cell has many compartments, each of which host very different biochemical and biological processes, carried out by different proteins. Identifying the subcellular localization of a protein sequence is a crucial step in finding the biological process in which the

protein may partake and the molecular function it may play. Consequently, predictive subcellular localization analyses were performed with ChloroP 1.1 [240], MitoProt II 1.0a4 [241], WoLF PSORT [242-244] for all *Arabidopsis* PH and FYVE proteins. All three programs predict the subcellular localization sites of proteins based on their amino acid sequences. They take the amino acid sequence of a protein and, in the case of WoLF PSORT, also the type of organism from which it was obtained. Each program then searches for several features that may reflect the subcellular localization of the protein. For instance, it has been found that the trafficking of proteins to some subcellular compartments is dictated by short signal peptides at the N- or C-terminal of the protein. The ChloroP 1.1 [240] server searches for the presence of chloroplast transit peptides, MitoProt II 1.0a4 [241] for mitochondrial targeting sequences, and WoLF PSORT [242-244] for a library of the known signals peptides, e.g. the presence of an N-terminal mitochondrial targeting signal, which is calculated from the amino acid composition of the N-terminal twenty residues. WoLF PSORT also checks predicted structural features such as topology that may indicate that protein is transmembrane, reflect its amino acid composition, and PROSITE motifs, usually short consensus signatures associated with functional sites in proteins [242-244].

H. Functional annotation

In the context of protein data, annotation generally refers to all information about proteins other than protein sequence. In collection of protein data, each protein sequence is usually complemented by annotations such as name, authors, submission date of protein data, etc. For our purpose, three main sources of protein annotations were used: (1) those taken from various protein data sources submitted by authors of protein data themselves from their published experimental results, (2) those that are obtained by analysis of protein sequence, 3D structure, and biophysical properties with various tools extracting biological information from other protein data collections, and (3) those that are based on support vector machine (SVM) method [245], which uses physical and chemical properties of the sequence of amino acids irrespective of sequence similarity. In particular, a web-based tool called SVMProt was used for the analysis [246].

Whenever necessary, the original functional annotation of *Arabidopsis* PH and FYVE domains was re-annotated using current resources that may not have been available at the time that the original annotation was performed [247]. It is possible that the published database annotation may not have been updated since the database entry was originally submitted and thus it was outdated or incomplete [247]. After all known and computed annotation was associated with a particular sequence, we used the sequence details and biophysical properties of our modeled sequence to predict function and compare it with the compiled annotation information.

Phylogenetic Tree Reconstruction

A multiple *Arabidopsis* PH and FYVE domain sequence alignments were generated using ClustalW, version 1.83 [248] to serve as inputs for the bayesian phylogeny software, MrBayes, version 3.12 [249]. MrBayes run using the Jones amino acid substitution matrix at fixed rate among sites [249]. The program proceeded for one million runs, sampling every 100th generation, and the final 2,500 trees were summarized.

Electronic supplement

The sequences, NCBI identifiers and coordinate files representing our models for all *Arabidopsis* PH and FYVE domains as well as other supplementary information, for example, GRASP images and structure verification plots, and alignment files are available at the following websites: <http://userhome.brooklyn.cuny.edu/ssingh/Arabidopsis/PH/ph.html> and <http://userhome.brooklyn.cuny.edu/ssingh/Arabidopsis/FYVE/fyve.html>.

Chapter 3 *Arabidopsis* PH domains

Sequence analyses

A. Sequence verification

Forty-nine *Arabidopsis* PH proteins were identified in the *Arabidopsis* proteome (Appendix Table 0.1). The final number of *Arabidopsis* PH proteins is close to the reported number of forty seven PH proteins by the SMART domain database [150, 151] and previously published number of fifty three PH proteins [250]. In the first case, the discrepancy is most likely due to incomplete annotation and in the second case, due to missannotations in the published reports, which are discussed below.

B. Secondary structure prediction

The *Arabidopsis* RhoGAP, AtPH1, UNK2, and three of the START (i.e. AT2G28320, AT3G54800, and AT5G35180) PH domains share the same structure of secondary elements as other known PH domains (Fig. 3.6, 3.16, 3.20 and 3.24). The common core fold of the PH domain includes seven β -strands followed by a C-terminal α -helix. In addition, the loop lengths are also of equal or of comparative length to the loop lengths found in most known PH domains. The *Arabidopsis* OSBP and ArfGAP PH domains have unusual β 3/ β 4 loop lengths of approximately 40 – 50 residues (Fig. 3.4 and 3.14). The AT4G19040 and AT5G45560 START PH domains have an additional α -helix predicted within the β 3/ β 4 loop (Fig. 3.6). The β 1/ β 2 loop of *Arabidopsis* PLD ζ PH domains is approximately 25 residues long, which is two times as long as seen in other PH domains (Fig. 3.8). The *Arabidopsis* PDK, DUF828, UNK2 and PRAF PH domains share an extra N-terminal α -helix prior to the typical seven β -strands, followed by a C-terminal α -helix (Fig. 3.10, 3.12, 3.22 and 3.24). The AT4G14370 PH domain has an unusual secondary structure prediction, which does not agree with predictions for the remaining PRAF proteins because the AT4G14370 PH protein is missannotated [250] and its PH domain appears to be incomplete. Two UNK2 PH domains, AT1G17820 and AT1G73200, have an additional α -helix within the β 1/ β 2 loop (Fig. 3.24) whereas the PRAF PH domains have one within the β 5/ β 6 loop (Fig. 3.10). Unlike

the $\beta 1/\beta 2$ loop of the two UNK2 PH domains however, the $\beta 5/\beta 6$ loop of the PRAF PH domains is about 20 residues long (Fig. 3.10 and 3.24). The lengthy $\beta 5/\beta 6$ loop is also predicted for DYN PH domains but with no additional secondary structure elements localized within it (Fig. 3.18). Finally, the *Arabidopsis* UNK1 PH domain has an additional α -helix in the $\beta 6/\beta 7$ loop (Fig. 3.24).

C. Domain-based classification

On the basis of domain architecture twelve classes of *Arabidopsis* PH proteins can be identified (Fig. 3.1, Appendix A Table 0.1 and Appendix A Fig. 0.1) based on the output of SMART [149-151] and NCBI Conserved Domain Architecture Retrieval Tool (C-DART) [251]. Other than the PH domain, there is little functional or structural similarity among the different classes. For eight of these classes, i.e. the OSBP, START, PLD ζ , PDK, ArfGAP, RhoGAP, DYN and AtPH1, we find putative homologues in other organisms. For two classes, i.e. the PRAF and DUF828, we find putative homologues among plants but not among other eukaryotes such as yeast, worm or vertebrate proteins. For the remaining two classes we find no known homologues.

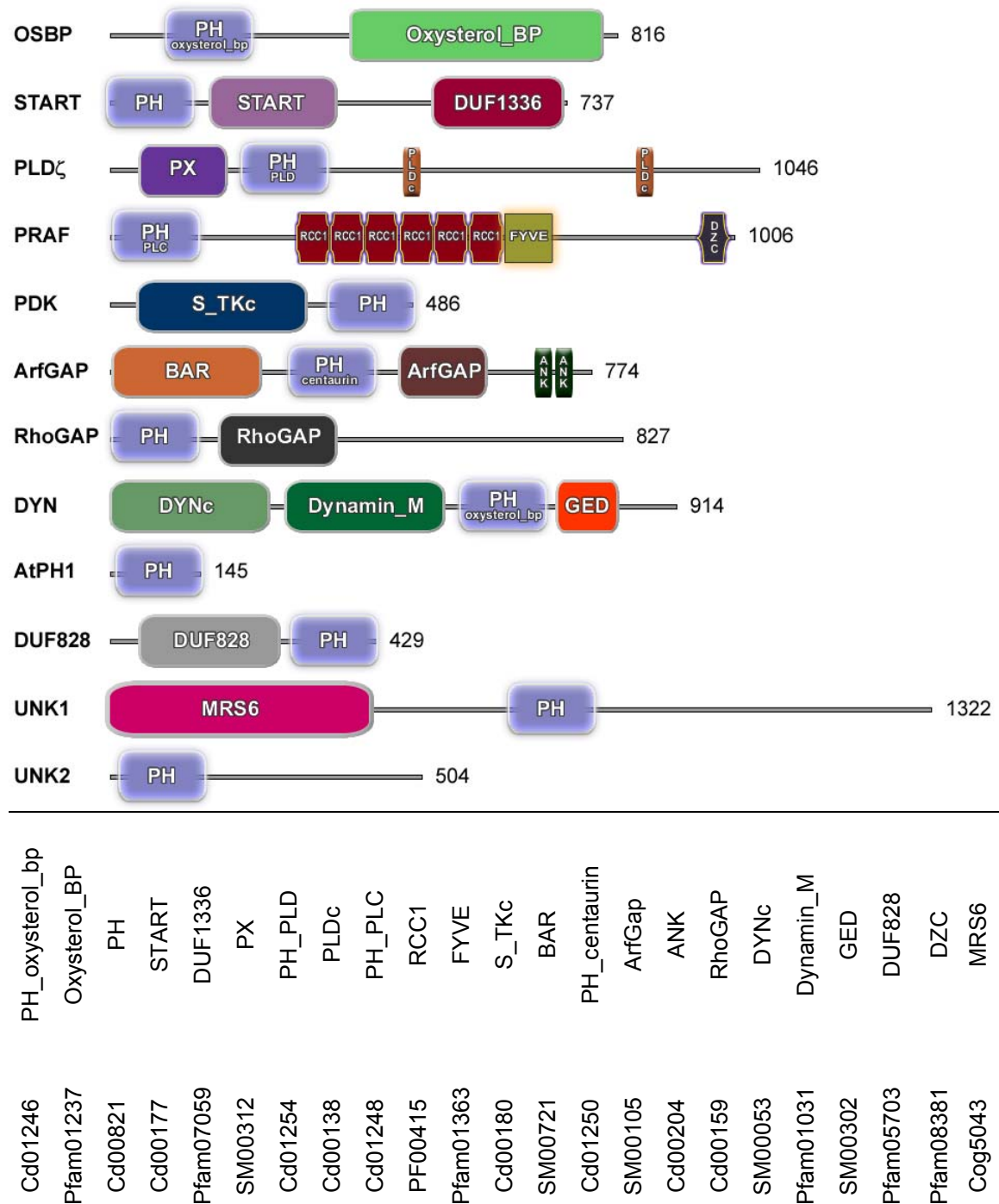


Figure 3.1. Cartoon of representative examples of the different classes of *Arabidopsis* PH proteins. The Pfam (PF), SMART (SM), Conserved Domains (CD) and Clusters of Orthologous Groups (COG) accession numbers for the different domains are given under the figure. The length of each protein is indicated on the right.

Structure prediction and analyses

A. Molecular modeling

High quality models were built for all *Arabidopsis* PH domains using multiple approaches thus giving rise to many different models for each sequence. Appendix Table 0.2 lists data corresponding to the top model for each *Arabidopsis* PH domain sequence. The table includes *Arabidopsis* PH domain class as assigned earlier, accession number, technique used to build the model, name of the protein used as a modeling template, and sequence identity/similarity between the modeled sequence and its template. The majority of models, twenty-nine out of forty-nine, were built using fold recognition techniques (Appendix Table 0.2), which was anticipated given the low sequence similarity that is characteristic of the *Arabidopsis* PH domains and other PH domains in general. Five models were built using homology searches, twelve models with the automated pipeline and three using the *de novo* approach (Appendix Table 0.2). All *Arabidopsis* PRAF PH domains were built based on the same PLC δ 1 PH domain as a template, both *Arabidopsis* PDK PH domains using PDK1 PH domain, and both DYN PH domains using Pleckstrin PH domain. Members of the remaining classes were modeled using more than one template per class. For example, the OSBP class models were built based on five different template PH proteins: Protein kinase B (PKB), tandem PH protein 1 (TAPP1), SET binding factor 1 (Sbf1), and dual adaptor of phosphotyrosine and 3- phosphoinositides (DAPP1) (Appendix Table 0.2).

B. Electrostatic profiles

Fig. 3.2 shows GRASP images of the electrostatic potential contours as calculated for high quality models of all *Arabidopsis* PH domain sequences. The modeled structures of AT3G16785, AT4G14740, AT5G5770 and AT4G23895 PH domains display a different overall pattern (Fig. 3.2C, J and L). In particular, the electrostatic potential in the membrane-binding region of the three most variable loops, i.e. β 1/ β 2, β 3/ β 4, and β 6/ β 7, appear to be predominantly negative for these domains. Consequently, AT3G16785, AT4G14740, AT5G5770 and

AT4G23895 PH domains may not be able to bind PIs due to the lack of a basic binding pocket, which typically houses the PI-binding site [53, 57, 74, 82, 87-89].

The electrostatic profiles of AT4G12460, AT4G22540, AT5G42140, AT3G22810, AT3G63300, AT4G17350, AT4G32785, AT5G43870, AT5G47440 and AT2G30880 PH domains appear to have a strong overall positive electrostatic potential (Fig. 3.2A, D, J and L) and seem to lack electrostatic polarization.

The remaining PH domains have a typical electrostatically polarized structure such as that reported first for β -spectrin [84] with a large positive potential opposite the negatively charged half containing the C-terminal α -helix.

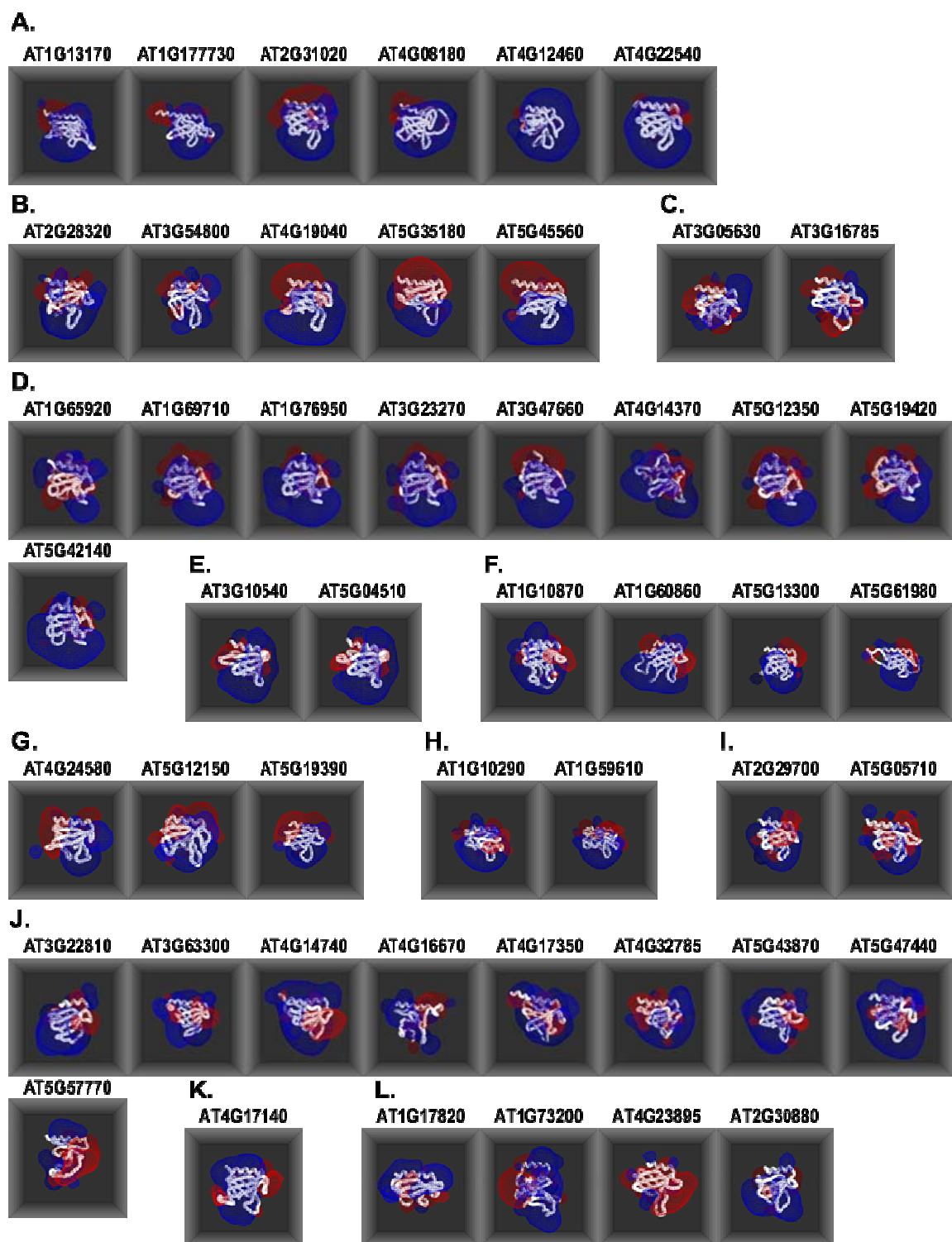


Figure 3.2. Electrostatic properties of the best A. OSBP, B. START, C. PLD ζ , D. PRAF, E. PDK, F. ArfGAP, G. RhoGAP, H. DYN, I. AtPH1, J. DUF828; K. UNK1, L. UNK2 PH domain models. In all panels, the electrostatic potentials were calculated at +25 mV and -25 mV equipotential contours in 0.1 M KCl. All images of the electrostatic potential contours were calculated with GRASP [228]. The models are represented as C α backbone traces and their accession numbers are displayed in the upper part of each profile.

C. PI docking

Docking calculations were performed with the top 3D modeled structure of each *Arabidopsis* PH domain (Appendix A Table 0.2) and various InsPtds, to identify potential InsPtds binding partners. The obtained results were evaluated in terms of binding energy and docking position in the binding site. Fig. 3.3 shows the preferred InsPtds for each *Arabidopsis* PH domain, which vary both within each class as well as between classes.

A. OSBP					
AT1G13170	AT1G77730	AT2G31020	AT4G08180	AT4G12460	AT4G22540
Ins(1,3)P2	PA	Ins(1,3)P2	Ins(1,5)P2		
Ins(1,4)P2	Ins(1,3)P2				
	Ins(1,4)P2				
	Ins(1,5)P2				
	Ins(1,4,5)P3				

B. START				C. PLD ζ		
AT2G28320	AT3G54800 PA	AT4G19040 PA	AT5G35180	AT5G45560 Ins(1,3,4,5)P4	AT3G05630	AT3G16785
		Ins(1,4)P2			Ins(1,5)P2	

D. PRAF								
AT1G65920	AT1G69710	AT1G76950	AT3G23270	AT3G47660	AT4G14370	AT5G12350	AT5G19420	AT5G42140
	Ins(1,5)P2		Ins(1,3)P2	Ins(1,3)P2	Ins(1,3)P2			Ins(1,3)P2
	Ins(1,3,4,5)P4		Ins(1,4)P2	Ins(1,3,4)P3	Ins(1,5)P2			Ins(1,4)P2
			Ins(1,3,4)P3	Ins(1,3,4,5)P4	Ins(1,4,5)P3			Ins(1,4,5)P3
					Ins(1,3,4)P3			Ins(1,3,4)P3

E. PDK		F. ArfGAP				G. RhoGAP		
AT3G10540	AT5G04510	AT1G10870	AT1G60860	AT5G13300	AT5G61980	AT4G24580	AT5G12150	AT5G19390
Ins(1,3)P2	PA		Ins(1,3)P2	Ins(1,5)P2	Ins(1,4)P2	Ins(1,3)P2	Ins(1,4)P2	PA
Ins(1,4)P2	Ins(1,3,4)P3					Ins(1,4)P2	Ins(1,5)P2	
Ins(1,5)P2						Ins(1,5)P2		
Ins(1,4,5)P3						Ins(1,3,4)P3		
Ins(1,3,4)P3								

H. DYN		I. AtPH1		K. UNK1		L. UNK2		
AT1G10290	AT1G59610	AT2G29700	AT5G05710	AT4G17140	AT1G17820	AT1G73200	AT4G23895	AT2G30880
	Ins(1,3)P2			Ins(1,3)P2	PA	Ins(1,4)P2		
	Ins(1,4)P2			Ins(1,4)P2				
	Ins(1,5)P2			Ins(1,5)P2				
	Ins(1,3,4)P3							

J. DUF828						
AT3G22810	AT3G63300	AT4G14740	AT4G16670	AT4G17350	AT4G32785	AT5G43870
					PA	Ins(1,3)P2
						Ins(1,3,4)P3

Figure 3.3. *Arabidopsis* PH domains and their InsPtd specificity as identified via docking. Highlighted in red are two InsPtd, Ins(3,4)P₂ and Ins(3,4,5)P₃, which have not yet been detected in plants [38, 39].

Functional annotation

Although the functions of many *Arabidopsis* PH proteins are currently unknown (Appendix A Table 0.3), their roles and possible ligand specificities can be inferred to some extent by the existence of multiple homologues in other eukaryotes. Our classification updates an earlier one [250] in light of the new sequence information and analysis. In particular, the PI 4-kinase group of *Arabidopsis* PH domains (i.e. AT1G60490, AT1G51040, AT1G49340, AT5G09350 and AT5G64070), which was found to act as an allosteric modulator [105] and selectively bind PA, PtdIns(4)P, PtdIns(4,5)P₂ [102] was not included in the current classification because we are unable to confirm the presence of PH domains within the PI 4-kinases of *Arabidopsis thaliana*. The AT1G48090 sequence was excluded for the same reason. Two additional *Arabidopsis* sequences, not reported previously [250], were included in our classification, i.e. AT4G23895 and AT5G57770 because the automated pipeline identified PH domain sequences present within them. These PH domain sequences were later confirmed using secondary structure predictions, fold recognition and SMART [149-151] analyses.

The functional annotation of proteins based on sequence characteristics alone is a work in progress and needs continual updating as more data becomes available. To date, the suggestions for functions associated with the various PH domain sequences are detailed in Appendix A Table 0.3. As pointed out earlier, most PH proteins have an unknown function and are targeted to an unknown location(s) within the cell. Some of the exceptions include the AT5G13300 and AT1G10290 proteins, which are involved in experimentally established vesicle transport [252] and Golgi to vacuole transport [253]. The broad role predicted for most PH proteins is that they would function either as transferases or hydrolases. Most of the meaningful information regarding the function of these domains comes from the analysis of their modeled structure and is discussed in detail below.

Discussion

A. The OSBP class

The oxysterol-binding proteins (OSBPs) bind oxidized forms of cholesterol to move them from the cytoplasm to the Golgi [254, 255]. OSBPs have been studied in many organisms, including *H. sapiens* [256, 257], *M. musculus* [258], *D. melanogaster* [259], *S. cerevisiae* [260] and *A. thaliana* [261]. They have been implicated in membrane trafficking [262], tumor metastasis [263], cell cycle progression [259] and sterol homeostasis [264]. OSBPs are often found to be part of a multi-gene family where some of OSBPs contain N-terminal PH domains together with the oxysterol-binding domain whereas others do not [256, 261]. Recently, the *Arabidopsis* OSBPs/OSBP-related proteins family has been reported to include twelve members with five proteins containing the N-terminal PH domains [250, 261]. We report an additional OSBP PH domain present in the *Arabidopsis* genome, i.e. AT1G77730 PH domain. Together these six proteins make up the OSBP class of *Arabidopsis* PH proteins depicted in Fig. 3.1. In addition to the PH and oxysterol-binding domain, all *Arabidopsis* OSBPs are predicted to house a coiled-coiled region, which is potentially involved in dimerization of the protein [261]. Gene expression analysis points to variable expression profiles for the various *Arabidopsis* OSBP proteins in the plant tissues, with AT2G31020 predominantly expressed in roots, AT4G08180 expressed in pollen, and the remaining OSBPs expressed in stem, leaf, seedling, root and flower [261].

The various OSBPs studied so far exhibit different patterns in their sub-cellular targeting, e.g. mammalian OSBP PH domains localize predominantly to the cytosol and the membrane of the Golgi apparatus [254, 255], yeast OSBP (Osh1p) is targeted to both the late Golgi via its PH domain and the nucleus-vacuole junction via the ankyrin (ANK) repeats, full-length yeast Osh2p localizes to the plasma membrane while its PH domain when expressed alone is weakly targeted to the Golgi [260]. The difference in strong vs. weak Golgi targeting of yeast OSBPs has been attributed to the third residue of strand $\beta 7$, which in strongly targeted Osh1p PH domain is histidine (RWHLKG) instead of arginine in the weakly targeted Osh2p (RWRLKG) [265]. This residue acts as a second interaction site in Golgi targeting in addition to the PI-binding site. It lies adjacent to the PI-binding site and it is believed to interact with a protein, ADP-ribosylation factor

1 (ARF1)/Arf1p, rather than a membrane lipid [265]. Secondary structure predictions identify the residues of strand $\beta 7$ in *Arabidopsis* OSBP PH domain sequences to be: KTLHLR in AT4G12460 and AT4G22540 PH domains, KRLHLR in AT2G31020 and AT4G08180 PH domains, KSLHLR in AT1G77730 PH domain and TLHLRC in AT1G13170 PH domain (Fig. 3.4). By analogy to the yeast Osh1p PH domains, the presence of histidine residue in strand $\beta 7$ of the *Arabidopsis* OSBP PH domains may serve as an additional interaction site and be a determinant of strong Golgi targeting.

The secondary structure predictions reveal unusually long $\beta 3/\beta 4$ loop in *Arabidopsis* OSBP PH domains not present in other OSBP/OSBP related proteins (Fig. 3.4). This $\beta 3/\beta 4$ loop is approximately 30 residues long (Fig. 3.4). The human OSBP related protein 8 also has a long loop, which is located between the $\beta 6$ and $\beta 7$ strands instead of the $\beta 3$ and $\beta 4$ strands (PDB ID: 1V88 [266]). The exact significance of the long loops remains to be determined, however, in the *Arabidopsis* OSBP PH domains their putative location in the canonical membrane binding site and abundance in basic residues seem to point to their involvement in phospholipid-binding, contribution to predominantly basic electrostatic profiles (Fig. 3.5) and non-specific membrane interactions.

Although the PH domain of a mammalian OSBP has been shown to bind both PtdIns(4)*P* and PtdIns(4,5)*P*₂ *in vitro* [254], the *in vivo* studies of yeast Osh1p PH domains suggest that only PtdIns(4)*P* is involved in Golgi targeting [267]. The docking results for the *Arabidopsis* OSBP PH domains suggest variable PI preferences among the different OSBP PH domains with the AT4G22540 and AT4G12460 PH domains showing no apparent affinity for any tested InsPtd ligand and AT1G77730 PH domain showing the most promiscuous behavior in its InsPtd-binding specificity (Fig. 3.3). Only At1G13170 OSBP PH domain showed a marked preference for Ins(1,4)*P*₂ based on the docking results.

We suggest that the *Arabidopsis* OSBP PH domains target the Golgi membrane by a combination of non-specific electrostatic interactions as well as the Golgi-specific interactions with the ARF1 protein mediated by the histidine residue present in the $\beta 7$ strand. It is also possible that *Arabidopsis* OSBP PH domains are targeted to the nuclear membrane based on the

sequence/structural similarity of the *Arabidopsis* OSBP PH domains with the *Glycine max* oxysterol binding protein (GmOSBP) PH domain, which has been shown to localize in/around the nucleus of onion cells [268].

```

      *           20           *           40           *           60           *           80
AT4G12460 : RSVAGILYKWTNYG---KGWRSRWFLLR--DGILSYSKIRRPENVNLLSPSDD-LRLIGDISTDRLLRMKS-----CSGRSR : 71
AT4G22540 : ETVAGILYKWTNFG---KGWRSRWFLLR--NGILSYSKIRRPENLNLLSPDDD-VRLIGDISGERLSRMDSD-----CSG--R : 69
AT2G31020 : NSIAGILYKVVNYG---KGWRPRWFVLQ--DGVLSSYYKIKGPDKIVVHHETEKGSRVIGEESTRMISRNRK---HAATNNTNH : 75
AT4G08180 : NGIAGILYKVVNYG---RGWRPRWFVLQ--DGVLSSYYKIHGPDKIFVSPETEKSSKVIIGDESARMISRNR---RGGSSSS-C : 74
AT1G77730 : NGVAGILYKVVNYG---QGWRPRWFVLQ--DGVLSSYYKIHGPDKISLVEMDRRSKLIGGESLRFICRHS-----KRGDV : 70
AT1G13170 : AGVAGIMYKVVNYG---KGWRARWFLE--DGVLSSYYKIHGPDKIVMNPSPREKGVVIGEEVSVYIRKASCGSSNRLGASAVAA : 79
GmOSBP : NGISGILYKVVNYG---KGWRPRWFVLQ--DGVLSSYYKIHGPDKIVNPETEKGSMTVIGEEVSVYIRKASCGSSNRLGASAVAA : 70
Osh1p : PTYKGYLKKWTNFA---QGYKLRWFILSS-DGKLSYYIDQ-----S : 37
Osh2p : PTYKGYLKKWTNFA---HGYKLRWFILSS-DGKLSYYIDQ-----S : 37
OSBP11 : ENVYGYLMKYTNLV---TQGYRFFVNNNEAGLLEYVNEOS-----S : 39
OSBP8 : GSSGIVMADWLKRGTLKSWTKLWCVLK--PGVLLIYKTQKNQGVGTVLNACRIIERP-----S : 59
ORP3 : PVQKGFLLKRRKWP--LKGWTKRFFYLD--KGILKYASQST-----D : 38

      *           100           *           120           *           140
AT4G12460 : RKHHKNIQIVHLKQVSSYRESKSDHRKFYIFTA---TKTLHLRTDSRSDRAAWLQALASTRGIV : 132
AT4G22540 : RKQKQTVQIVHLK-VSSFRESKSDDRKFYIFTA---TKTLHLRTDSITDRAAWLQALASTKCIF : 129
AT2G31020 : QLRRKPFGEVHLK-VSSIRSRSDDKRFSIFTG---TKRLHLRAETREDREAWIEALQAVKDMF : 135
AT4G08180 : QLRRKPFGEVHLK-VSSVRESRSDDKRFSIFTG---TKRLHLRAETREDRTTWVEALQAVKDMF : 134
AT1G77730 : HSPGKPLGQIHLK-VSSIGQSISDGKRFTVFTG---TKSLHLRAATSEDRAAWIEALQAVKETF : 130
AT1G13170 : SRPCKPFGEIHLK-VSSIRASKSDDKRLAIFTG---TKTLHLRCVSKENRAAWVEAFQVAKDLF : 139
GmOSBP : QHIRKPFGEIHLK-VSTIRSRSDDKRFSIFTG---TKRLHLRAETREDRVAMMEALQAVKDMF : 130
Osh1p : DTKNACRGSNMS-SCSLHLDSEKLLKFEIIGCNNGVIRWHLKGNHPIETNRWVWAIQGAIRYA : 100
Osh2p : HVDPR-RGTLKVS-TCRLHLDSEKLLNFELGGITGTIRWRLKGNHPIETTRWVWAIQSAIRFA : 99
OSBP11 : -RNQKPRGTLQLA-GAVLSPSDEDSHTFTVNAAS--GEQYKLRTDAKERQHWVSRQICTQHE : 99
OSBP8 : KKGDFCFKLFHPLEQSIWAVKGPKEAVGSITQPLPSSYLIRATSESDGRCWMDALELAKSG : 123
ORP3 : IEREKLHGCIDVG-LVSMVSKSS-KCIDLTE---EIIYHLKVSSEVFDWVSKLRHHRYR : 97

```

Figure 3.4. The alignment of consensus secondary structure predictions for the *Arabidopsis* OSBP PH domains and their *H. sapiens* (OSBP8 (PDB ID: 1V88 [266], OSBP11 (PDB ID: 2D9X [269], and ORP3 (NP_056365)), *S. cerevisiae* (Osh1p (P35845) and Osh2p (Q12451)) and *G. max* (BAF96543)) homologs. Rows depict 1-letter code of PH domain sequences, α -helices are highlighted in red, β -strands in blue, β_6/β_7 OSBP8 loop in yellow and critical histidine residue in orange.

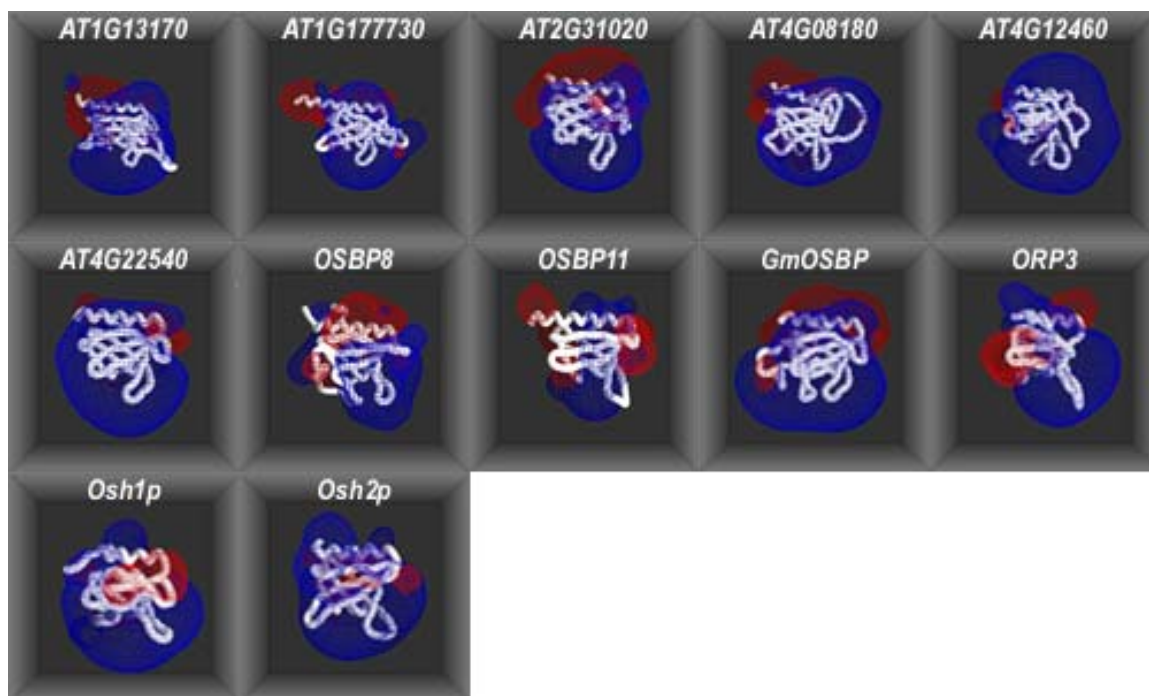


Figure 3.5. Electrostatic properties of *Arabidopsis* OSBP PH domains and their *H. sapiens* (OSBP8 (PDB ID: 1V88 [266], OSBP11 (PDB ID: 2D9X [269], and ORP3 (NP_056365)), *S. cerevisiae* (Osh1p (P35845) and Osh2p (Q12451)) and *G. max* (BAF96543)) homologs. PH domain models are in the same orientation. 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 [228]. The models are represented as C α backbone traces and their accession numbers are displayed in the upper part of each profile.

B. The START class

There are fifteen START proteins identified in human and mouse genomes [270] while thirty five START proteins are known to be present in *Arabidopsis* genome where they are often found in homeodomain transcription factors [271, 272]. Of the thirty five *Arabidopsis* START proteins only five contain an N-terminal PH domain, which is accompanied by a lipid-transfer steroidogenic acute regulatory protein-related lipid-transfer (START) domain and a C-terminal domain of unknown function 1336, DUF1336 (Fig. 3.1). The START domains are lipid binding domains that could bind cholesterol [273], phosphatidylcholine [274-276] or ceramide [277] depending on the nature of the binding tunnel of their host protein. However, the exact determinants for binding specificity remain unknown [278].

Members of the *Arabidopsis* START class lack the well characterized serine-rich domains responsible for serine/threonine kinase activity and Goodpasture antigen binding found in the mammalian START proteins [272]. Instead, the plant START proteins contain a 250 amino acid-long plant-specific C-terminal domain of unknown function, DUF1336 [272, 279]. One member of the *Arabidopsis* START class is the enhanced disease resistant 2 (EDR2) protein (AT4G19040), which has been shown to regulate plant defense responses and cell death in *Arabidopsis* possibly functioning inside mitochondria [279]. Loss-of-function mutations in the EDR2 gene confer enhanced disease resistance to powdery mildew [279]. The remaining members of *Arabidopsis* START class have an unknown function.

The human star-related lipid transfer protein 11 (STARD11) has been implicated to utilize its PH domain to bind $\text{PtdIns}(4)P$ and target the protein to the Golgi apparatus [280-282]. The data from our docking experiments suggests that only the EDR2 PH domain binds preferentially to $\text{Ins}(1,4)P_2$, the headgroup of $\text{PtdIns}(4)P$ (Fig. 3.3), implying a similar binding specificity as the human STARD11 PH domain despite having very different sequences (Fig. 3.6). In fact, a conserved glycine residue important for $\text{PtdIns}(4)P$ -recognition in START and other subfamilies of PH domains [282] is not conserved in any of the *Arabidopsis* START PH domains (Fig. 3.6). Our docking results instead implicate ARG15, ARG16, LYS84, HIS86, ARG87 and THR89 to be involved in the EDR2 PH domain recognition of $\text{PtdIns}(4)P$. Sub-cellular localization predictions

using the iPSORT and TargetP algorithms also suggest that EDR2 localizes to the mitochondria rather than the Golgi. Two other members of the *Arabidopsis* START class, i.e. AT3G54800 and AT5G45560 are also predicted to localize to the mitochondria [272, 279].

The *Arabidopsis* START PH domain models are typical in their electrostatic polarization (Fig. 3.7). However, all of them share large basic patches that are created by surface exposed basic residues not just from $\beta 1/\beta 2$, and $\beta 6/\beta 7$ variable loops but also by the $\beta 3/\beta 4$ variable loop, which has not been typically implicated in contributing to membrane binding. Models of the *Arabidopsis* START PH domains and their plant homologs are in stark contrast to the human STARD11 PH domain model, which has a much weaker basic surface patch in the membrane-binding site (Fig. 3.7). Consequently, we suggest that non-specific electrostatics play a major role in the membrane targeting of *Arabidopsis* START PH domains. The electrostatic profiles, together with the unique domain architecture and sequence variations suggest a novel role for the PH domains of the *Arabidopsis* START class.

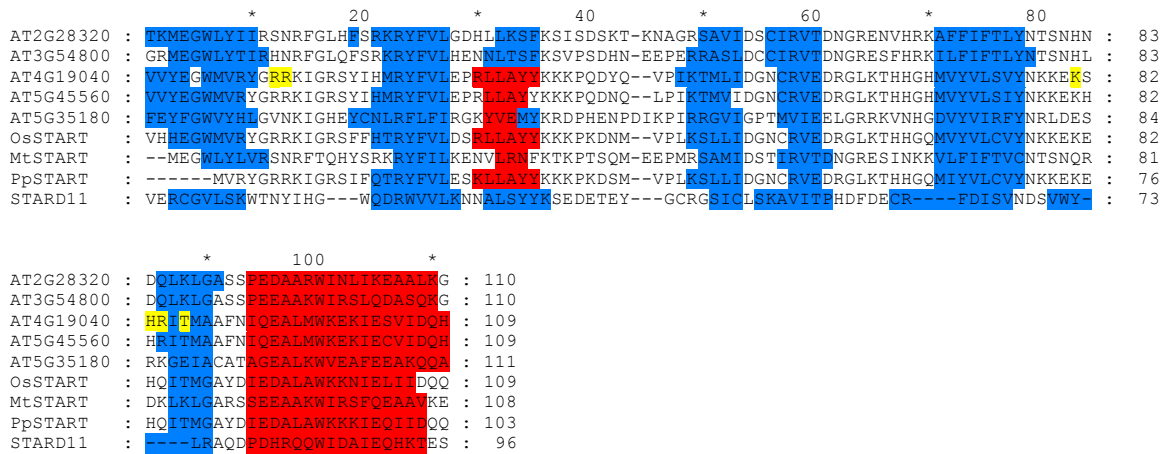


Figure 3.6. The alignment of consensus secondary structure predictions for the *Arabidopsis* START domains, their *O. sativa* (Q6YU77), *M. truncatula* (A4Q010), *P. pratensis* (Q5K2N0) and *H. sapiens* STARD11 (Q9Y5P4) homologs. Rows depict 1-letter code of PH domain sequences, α -helices are highlighted in red, β -strands in blue, residues implicated in PtdIns(4)P recognition are highlighted in yellow.

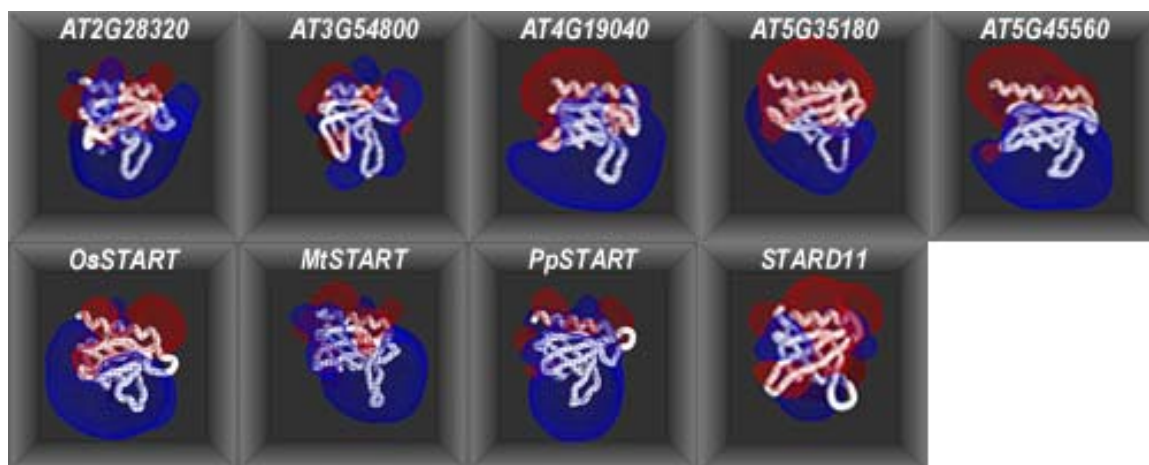


Figure 3.7. Electrostatic properties of the *Arabidopsis* START PH domains, their *O. sativa* (Q6YU77), *M. truncatula* (A4Q010), *P. pratensis* (Q5K2N0) and *H. sapiens* STARD11 (Q9Y5P4) homologs. PH domain models are in the same orientation. In all panels, the electrostatic potentials were calculated at +25 mV and -25 mV equipotential contours in 0.1 M KCl. All images of the electrostatic potential contours were calculated with GRASP [228]. The models are represented as C α backbone traces and their accession numbers are displayed in the upper part of each profile.

C. The PLD ζ class

The Phospholipase D ζ (PLD ζ) class of *Arabidopsis* PH domain proteins includes PLD ζ 1 (i.e. AT3G16785) and PLD ζ 2 (i.e. AT3G05630). Both sequences share a Phox (PX) domain, a PH domain (Fig. 3.1) and a C-terminal catalytic site comprised of a dimer of two H-x-K-x(4)-D (i.e. HKD) motifs characteristic to other PLD enzymes [283-285]. The two PLD ζ genes are located on chromosome number three. They share about 74% sequence similarity, which is far greater than the 46% sequence similarity observed for the other *Arabidopsis* PLDs [103]. Both *Arabidopsis* PLD ζ sequences seem to be Ca²⁺-independent PLDs because they lack the Ca²⁺/PI-binding C2 domain characteristic of the remaining ten PLD proteins encoded in *Arabidopsis* genome [103, 286]. PLD ζ 1 is unique in that it exhibits narrow substrate specificity as compared to other *Arabidopsis* PLDs hydrolyzing mainly phosphatidylcholine and not phosphatidylserine or phosphatidylethanolamine [103]. Functionally, *Arabidopsis* PLD ζ 1 and PLD ζ 2 were identified to regulate root development in response to nutrient limitation [287] and supply inorganic phosphorus during phosphorus starvation via hydrolysis of phosphatidylcholine [288, 289]. Additionally, *Arabidopsis* PLD ζ 2 was found to regulate vesicle trafficking and auxin responses, e.g. transport and distribution [290] and suggested to possibly be involved in lipid transfer between chloroplast and mitochondria for possible lipid modification that occurs in mitochondria [289].

In most mammalian cells, PLD ζ 1 shows dynamic localization between the plasma membrane and the intracellular membranes of endocytic and secretory compartments, including Golgi apparatus, endoplasmic reticulum, early and late endosomes, and multivesicular bodies [291-295] whereas PLD ζ 2 is mainly found at the plasma membrane [291, 296]. Unlike the homologous PLD PH domains of *H. sapiens*, *C. elegans* and *D. discoideum*, the *Arabidopsis* PLD ζ 1 PH domain lacks the strong positive potential surrounding the PI-binding site (Fig. 3.9). In fact, the electrostatic profiles of *Arabidopsis* PLD ζ PH domains appear most similar to the profile of the *S. cerevisiae* PLD PH domain (Fig. 3.9).

The secondary structure predictions of the PH domain sequences suggest extended $\beta 1/\beta 2$ and $\beta 5/\beta 6$ loops present within the *Arabidopsis* PLD ζ PH domains when compared to *H. sapiens* PLD1, *H. sapiens* PLD2 and *C. elegans* PLD (Fig. 3.8). The *D. discoideum* PLD PH domain shares an extended $\beta 1/\beta 2$ loop with the *Arabidopsis* PLD ζ whereas the *S. cerevisiae* PLD PH domain has an extended $\beta 5/\beta 6$ loop instead of the extended $\beta 1/\beta 2$ loop compared to the *Arabidopsis* PLD ζ s PH domains. The human PLD1 PH domain is selective for PtdIns(4,5) P_2 and PtdIns(3,4) P_2 [297] whereas the PH domain of *Arabidopsis* PLD $\zeta 2$ binds, possibly, Ins(1,5) P_2 and the PH domain of *Arabidopsis* PLD $\zeta 1$ does not bind any PI ligand as indicated by our docking analyses (Fig. 3.3). These results are in agreement with the atypical electrostatic profiles observed for the *Arabidopsis* PLD ζ PH domains (Fig. 3.9).

A complete removal of the PH domain ablates activity and prevents association of human PLD with membrane [297] but it does not affect enzymatic activity or phosphatidylinositol bisphosphate (PIP $_2$) dependence [298, 299]. There is evidence that PIP $_2$ binds to PLD at an additional motif outside of the PH domain sequence [297, 300]. The additional motif involved in the PIP $_2$ binding is characterized by five (i.e. R-x(3)-R-x(4)-R-x(4)-K-x(7)-R/K) conserved basic amino acids [103, 300]. The *Arabidopsis* PLD ζ s have four (i.e. R-x(3)-R-x(4)-R-x(7)-K-x(3)-E) of the five residues [103, 300], so they may retain this PH domain independent PIP $_2$ binding albeit with reduced affinity.

```

      *           20           *           40           *           60           *           80
AT3G16785 : KLKEDYIMVHLPKFSKSDDDSNRCG-CWFCCQNDNWQKVGVLKPGFLALLEDPFPAKLDDIIVFDVLPVS-NG----- : 75
AT3G05630 : KMKEGYVTVKHLRDVP--GSDGVRCLPETHCLGFGTSWTKVWAVLKPGFLALLEDPFSGKLLDIMVFDTLGLQGTK----- : 75
PLD1      : KGIKGMIMKRSRSG-----GHRIPG--LNCCGQGRACYRWSKRWLIVKDSFLLYMK-PDSGALAFVLLVDK----- : 61
PLD2      : KGLEGMIRKRSRSG-----GHRVPG--LTCCGRDQVCYRWSKRWLIVKDSFLLYMK-LETGAISFVQLADP----- : 61
CePLD     : KHTEGFFVKRSPG-----GSRVFLGWKQCCVR--YLLPWSKRWLMVRDSFVAYMD-HRTEQIRMVLLMDR----- : 61
ScPLD     : QGKQGYLVIRSTAK----AQGWVSHFGKHAHKDMIRHTTKWFLVRNSYLTYSVSDLSSTPLDVFLIDWKFKVRFSGNKNNIL : 80
DdPLD     : INMNGQLTKIKPKTIFNFENISHRRTLRLRDGFLIVYKKKKFKHDEDKPSKVIFLP---GSTVKVVVGSQ----- : 66

      *           100          *           120          *           140          *           160
AT3G16785 : -----NDGVDISLAEFLKDHNPRLHAFKVTSGNRSIRIRAKNSAKVKDWVASINQ : 125
AT3G05630 : -----ESSEQPKLABGVKEHNPLRFQFKVTSGLDRTVRLRTTSRKVKENWKAVDE : 125
PLD1      : -----EFKIKVGGKETETK-YGIRIDNLSRTLILKCNSTRHARWVGGAIEE : 106
PLD2      : -----GFEVQVGGKSTEAR-HGVRIDTSHRSLILKCSSYRQARWVAQEITE : 106
CePLD     : -----DFKVAAGGKETEGIPTGLIITNSQHELHLKCRRLQDTATWKYIIEQ : 107
ScPLD     : DNENEINWIIHDPNLEINDELEEFGIENDANNILDKNKSKTHQKKSNISSKLLLTLENSEKKLKIKCKSESSLKQWMSSTIR : 164
DdPLD     : -----PSKKKKFLSKKTHNYGFRVTKGNYNKFFLMENRDEALNWVNARF : 112

      *
AT3G16785 : AALRP : 130
AT3G05630 : AGCYS : 130
PLD1      : FIKKH : 111
PLD2      : LAQGP : 111
CePLD     : AMGGI : 112
ScPLD     : MSTST : 169
DdPLD     : HSRQG : 117

```

Figure 3.8. The alignment of consensus secondary structure predictions for the *Arabidopsis* PLD ζ PH domains and their *H. sapiens* (PLD1 (Q13393), PLD2 (O14939)), *S. cerevisiae* (P36126), *C. elegans* (Q17637) and *D. discoideum* (Q54WR3). Rows depict 1-letter code of PH domain sequences, α -helices are highlighted in red, β -strands in blue.

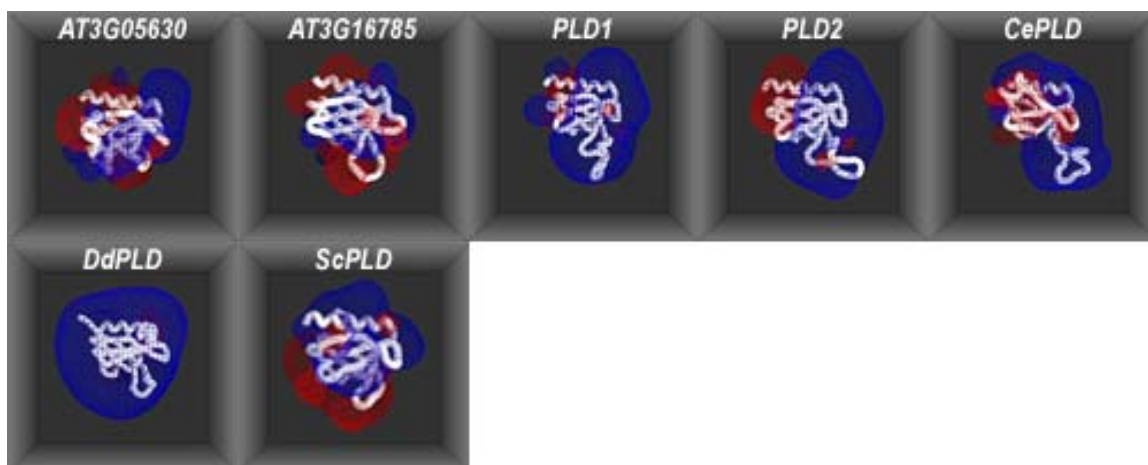


Figure 3.9. Electrostatic properties of the *Arabidopsis* PLD ζ PH domains and their *H. sapiens* (PLD1 (Q13393), PLD2 (O14939)), *S. cerevisiae* (P36126), *C. elegans* (Q17637) and *D. discoideum* (Q54WR3). PH domain models are in the same orientation. In all panels, the electrostatic potentials were 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 [228]. The models are represented as C α backbone traces and their accession numbers are displayed in the upper part of each profile.

D. The PRAF class

The PRAF class includes AT1G65920, AT1G69710, AT1G76950 AT3G23270, AT3G47660, AT4G14370, AT5G12350, AT5G19420 and AT5G42140 *Arabidopsis* PH proteins. The PRAF proteins consist of a PH domain, one or more regulator of chromosome condensation 1 (RCC1) repeats, a variant FYVE domain and a DZC domain [301, 302] (Fig. 3.1). So far, PRAF proteins have been found only in plants where they are thought to be localized within or near the nucleus [134, 138, 250, 303]. Their function remains largely unknown. The signature domains of this family are traditionally implicated in membrane-targeting (PH and FYVE; [54, 65, 67, 68, 107-109]), protein-protein interactions (DZC; [301]) and mitotic spindle assembly (RCC1; [304]). AT4G14370, a potential member of the PRAF family of proteins, is a misannotated *in silico* fusion of nucleotide-binding site-leucine-rich repeat (NBS-LRR) and an RCC1-encoding gene [250]. As a result AT4G14370 is omitted from analysis.

The modeled *Arabidopsis* PRAF PH domains are typical PH domains with distinct electrostatic polarization and strongly basic canonical binding surfaces (Fig. 3.11) but show unique secondary structure elements in addition to the ones that form the core PH domain fold (Fig. 3.10). 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 PLC δ 1 PH domain (Fig. 3.10). Moreover, both PRAF-1 and PLC δ 1 have been shown to preferentially bind PtdIns(4,5) P_2 with high specificity and affinity [51, 52, 128, 305]. Given that plant PLCs lack the N-terminal PH domain seen in PLCs of other organisms, including PLC δ 1, 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 PLC δ 1 PH domain due to a partial PtdIns(4,5) P_2 binding signature (Fig. 3.10) but a stronger non-specific electrostatic contribution to membrane binding from their large basic

patches for most members of the family (Fig. 3.11). The existence of the unique secondary structure elements i.e. the additional N-terminal α -helix and the α -helix within the $\beta 5/\beta 6$ loop is also seen in the PLC family of PH domains, reinforcing the idea that these two protein families share some evolutionary link.

```

      *           20           *           40           *           60           *           80
AT1G65920 : PRDRTDEQAILALKKGGAQLLKRRRGNPKFCPPFLSMDEKYLIIWYS-----GEEERQLRLSSVITIVRGQITENFQKQAQSDRK : 79
AT1G69710 : FSERDVEQAITSLLKGSYLLKYGRRGKPKFCPPFLTSDESALVWYS-----GKEEKQIKLQVLRIVPGQRTPTFKRYPRPEKE : 79
AT1G76950 : NADHNLQALITLKKGTQLLKYGRKGKPKFYPPFLSSDEKSLIWI-----SSGEKRLKLASVSKIIVPGQRTAVFQRYLRPEKD : 79
AT3G23270 : YHEEDVQQAIVVLKKGTTQLLKYSRKGKPKKFRFAFLSPDEKTLIIWFS-----RGEKGLKLEVSRIIVPGQRT----RFLRPEKD : 75
AT3G47660 : PTDRENEQAIAILKKGAYLLKYGRRGKPKFYPPFLSSDEIYLLWYC-----GKKEKRLKLSSVTRIIPGQRTAVFRYPQPTKE : 79
AT5G12350 : PVERDIQQAITALKKGAYLLKYGRRGKPKFCPPFLSNDETSLIWF-----GNEEKHLKLSHVSRITSGQRTPIFQRYPRPEKE : 79
AT5G19420 : PVTRDIQQAITALKKGAYLLKYGRRGKPKFCPPFLSNDESVLIIWF-----GKEEKHLKLSHVSRITSGQRTPIFQRYPRPEKE : 79
AT5G42140 : RALLYVLQALIALKKGGAQLLKYSRKGKPKFCPPFLSNDETSLIWI-----NGGEKRLKLATVSKIIVPGQRTAVFQRYLRPEKD : 79
VvPRAF : PFDRAVEQAIVTLKKGAYLLRCGRRGKPKFCPPFLSTDEKFLIWI-----GQEEKQLRLSSITKIIPGQRTVNFQRLQEERE : 79
OsPRAF : PVERDIQQAITALKKGAYLLKYGRRGKPKFCPPFLSNDESVLIIWF-----GKEEKHLRLSQVSRIVPGQRTAIFQRYPRPEKE : 79
RnPLCd1 : MHGLQDDFDLQALKGSQLLKVKSSSWRREFFYKLQEDCKTIWQESRKVMRSPESQLFSIEDIQEVRMGHRTGLEKKAARDIPE : 84
bind_PIP2 : -----X-X-----X-X-----*XXX----- : 7

      *           100           *           120           *
AT1G65920 : EQSFSLIYANGEHTLDLICKDKAQADSWFKGLRAVITRHHNIRNSVN : 126
AT1G69710 : YQSFSLICP--DRSLDLICKDKDEAEVWVVGKSLITRVKVSQWKTT : 124
AT1G76950 : YLSFSLLYNGKKKSLDLICKDKVEAEIWIGLKLITSGQGRSKID : 126
AT3G23270 : HLSFSLLYNNRERSLDLICKDKAETEVNFAALKFLIEKSRNRRARSE : 122
AT3G47660 : YQSFSLIYG--ERSLDL--DKDEAEFWLTTLRALLSRNNSSALVLH : 121
AT5G12350 : YQSFSLIYS--ERSLDVICKDKDEAEVWFTGLKALISHCHQRNRTE : 124
AT5G19420 : YQSFSLIYD--ERSLDLICKDKDEAEVWFSGLKALISHRCHQRKWRTE : 124
AT5G42140 : YLSFSLIYSNRKRTLDLICKDKVEAEVWIAGLKALISHGQAGRSKID : 126
VvPRAF : CRSFSLVYANGERSLDLICKDKAQADSWFGLKAVISRCHQHPRLST : 126
OsPRAF : CQSFSLISH--DRSLDLICKDKDEAEVWFAGLKLITRSHQRKWRTE : 124
RnPLCd1 : DRCSFIVFKDQRNTLDLITAPSADAAQHVVQGLRKLIHHSGSMDQRQK : 131
bind_PIP2 : -----*----- : -

```

Figure 3.10. 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 [90]). 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.



Figure 3.11. Electrostatic properties of the *Arabidopsis* PRAF PH domains, their *V. vinifera* (A5B4Z5) and *O. sativa* (Q0JFZ5) putative homologs and *R. norvegicus* PLC δ 1 PH domain (PDB ID: 1MAI [90]). 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 [228]. The models are represented as C α backbone traces and their accession numbers/abbreviated names are displayed in the upper part of each profile.

E. The PDK class

The 3-phosphoinositide-dependent kinase 1 (PDK) class includes AT3G10540 and AT5G04510 *Arabidopsis* PH proteins, which possess a kinase domain at the N-terminus and a PH domain at the C-terminus (Fig. 3.1). Like mammalian PDK proteins, *Arabidopsis* PDKs require phosphorylation at a number of residues, i.e. D167, T176, S177, T211 and S276 for their downstream phosphorylation activity [306]. *H. sapiens* PDK1 negatively regulates TGF-beta-mediated signaling via a direct physical interaction with Smad proteins [307, 308]. The *in vitro* target of *Arabidopsis* PDK proteins is thought to be the S6 kinase 2 [306] but the molecular function of *Arabidopsis* PDK proteins remains to be determined.

PDK proteins are known to phosphorylate and activate several protein kinases, e.g. human protein kinase B (PKB) [309]. In fact, the interaction of the PDK with PKB takes place via the PH domain of PDK, which is known to bind $\text{PtdIns}(3,4,5)\text{P}_3$ and/or $\text{PtdIns}(3,4)\text{P}_2$ [310-312] and with reduced affinity to $\text{Ins}(1,3,4,5,6)\text{P}_5$ and/or InsP_6 [313]. The *Arabidopsis* PDK counterpart also phosphorylates PKB but at a much lower rate than observed for human PDK [104]. It is possible that the observed difference is a consequence of diverse ligand binding specificities, since *Arabidopsis* PDK PH domains were identified to interact with phosphatidic acid (PA), $\text{PtdIns}(3)\text{P}$, $\text{PtdIns}(3,4)\text{P}_2$, $\text{PtdIns}(3,4,5)\text{P}_3$ and to a lesser extent with $\text{PtdIns}(4)\text{P}$ and $\text{PtdIns}(4,5)\text{P}_2$ using the protein lipid overlay assays [104]. This promiscuity with regard to PI recognition for the *Arabidopsis* PDK PH domain is not intriguing given the fact that *Arabidopsis* PDK PH domains lack many residues critical to PI3-kinase associated binding [104]. The docking results confirm the reported binding promiscuity for the AT3G10540 PH domain whereas AT5G04510 PH domain is suggested to bind only PA and $\text{PtdIns}(3,4)\text{P}_2$ (Fig. 3.3). If the ligand binding properties of these two PDK PH domains do indeed vary *in vitro*, these can be explained primarily by differences in loop orientations given their high pairwise sequence identity.

Both *Arabidopsis* PDK PH models, however, exhibit highly similar biophysical profiles, which suggest a large contribution of non-specific electrostatics in membrane targeting (Fig. 3.13) that stems from the positively charged surface localized within the three most variable loops i.e. $\beta 1/\beta 2$, $\beta 3/\beta 4$, and $\beta 6/\beta 7$ (Fig. 3.13). The modeled *Arabidopsis* PDK PH domains carry significantly

more positive charge on their surface than their *H. sapiens* homolog and consequently, share greater resemblance with the *C. elegans* homolog. We predict that *Arabidopsis* PDK PH domains are targeted to membranes primarily by non-specific electrostatics.

```

      *           20           *           40           *           60           *           80
AT3G10540 : WQQLEPGE SVLMISAVKKI QKITSKKVQLILTNKPLRIYVDPSKLVVKGNIISDNSNDLNVQVSSPSHFKICTPKKVLSFED : 84
AT5G04510 : WQQLEPGE SVLMISAVKKI QKITSKKVQLILTNKPLRIYVDPSKLVVKGNIISDNSNDLNVVTSPSHFKICTPKKVLSFED : 84
PDK1      : WHQFVEN-NLILKMGVPVDRKGLFARRRQLLLTEGPHLYYVDPVNKVLKGEIPWS---QELRPEAKNFKTFFVHTPNRTYYLMD : 80
PDPK1     : -----ILKQGYLKKRGLFARRRMFLLTEGPHLLYIDVPNLVLKGEVPWT---PCMQVELKNSGTEFFIHTPNRVYYLFD : 71

      *           100
AT3G10540 : AKQRAIQWKKAIETLQ : 100
AT5G04510 : AKQRASVWKKAIETLQ : 100
PDK1      : PSGNAHKWCRKIQEVW : 96
PDPK1     : LEKKADEWCKRAINDVR : 87

```

Figure 3.12. The alignment of consensus secondary structure predictions for the *Arabidopsis* PDK PH domains and their *H. sapiens* (PDB ID: 1W1H [313]) and *C. elegans* PDPK1 (Q9Y1J3) homologs. Rows depict 1-letter code of PH domain sequences, α -helices are highlighted in red, β -strands in blue.

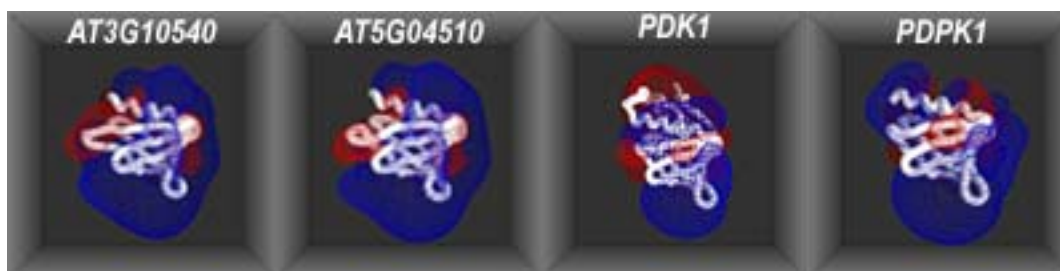


Figure 3.13. Electrostatic properties of the *Arabidopsis* PDK PH domains and their *H. sapiens* (PDB ID: 1W1H [313]) and *C. elegans* PDPK1 (Q9Y1J3) homologs. PH domain models are in the same orientation. In all panels, the electrostatic potentials were calculated at +25 mV and -25 mV equipotential contours in 0.1 M KCl. All images of the electrostatic potential contours were calculated with GRASP [227]. The models are represented as Ca backbone traces and their accession numbers are displayed in the upper part of each profile.

F. The ArfGAP class

The adenosine diphosphate-ribosylation factor-guanosine triphosphatase activating proteins (ArfGAPs) class includes five PH *Arabidopsis* proteins. The *Arabidopsis* ArfGAPs contain four domains: a BIN/amphiphysin/RVS (BAR) domain, a PH domain, an ArfGAP domain and two or three ankyrin repeats (Fig. 3.1). The BAR domain and ANK repeats are involved in protein-protein interactions [314]. The BAR domain is a membrane binding and a curvature-sensing module [315, 316] also known to be responsible for ArfGAP dimerization. It plays an important role in fission of synaptic vesicles, endocytosis, regulation of the actin cytoskeleton, transcriptional repression, cell-cell fusion, apoptosis, secretory vesicle fusion and tissue differentiation [317, 318]. The ArfGAP domain stimulates guanosine triphosphate (GTP) hydrolysis [319]. The ArfGAPs are involved in membrane-trafficking and vesicle formation by inactivating the small G-protein Arf [320]. An example of *Arabidopsis* ArfGAPs is a vascular network defective 3 (VAN3; i.e. AT5G13300) protein, which has been localized to the trans-Golgi network where it is involved in the vesicle transport responsible for the auxin signaling required for vascular differentiation [252]. VAN3 has been shown to share the same domain structure as human centaurin $\beta 1$ and centaurin $\beta 2$ together with high sequence identities of 32% and 33%, respectively [252, 321].

Recently, a BAR-PH structure of adaptor protein containing a PH domain, a PTB domain, and a leucine zipper motif (APPL1) has been solved [322]. It is similar to the *Arabidopsis* ArfGAPs, in that the BAR domain of APPL1 is located at the N-terminal of the protein and it is followed by the PH domain [323, 324]. It has been shown that the BAR and PH domains function together in a union mediating Rab5 recruitment to the membrane [325] and ArfGAP's binding to curved membranes as demonstrated for centaurin $\beta 2$ [316]. It is possible that the BAR and PH domains of *Arabidopsis* ArfGAPs also function as a unit. More specifically, since the membrane interaction of APPL1 is thought to be mediated by the $\alpha 2/\alpha 3$ loop of the BAR domain and the $\beta 5/\beta 6$ loop of the PH domain [322], it is possible that the $\beta 5/\beta 6$ loop of the ArfGAP PH domains is also involved in membrane interactions. This interaction might be further reinforced by the

involvement of the $\beta 3/\beta 4$ loop, which has been predicted to be unusually long, 40 – 50 residues, in the *Arabidopsis* ArfGAP PH domains (Fig. 3.14).

The *Arabidopsis* ArfGAP PH domains have predominantly positive electrostatic profiles (Fig. 3.15), suggesting that non-specific interactions play a major role in targeting the domain to the membrane. The ArfGAP PH domain binds $\text{PtdIns}(4,5)\text{P}_2$ in mammalian cells [326] and induces a conformational change, which stimulates the activity of the protein up to 30-fold [327]. *Arabidopsis* VAN3 (i.e. AT5G13300) has been shown to bind with high affinity $\text{PtdIns}(4)\text{P}$ and with reduced affinity $\text{PtdIns}(4,5)\text{P}_2$ [252]. Our docking results suggest that three ArfGAP PH domains bind preferentially to headgroups of monophosphorylated PIs, e.g. AT1G60860 to $\text{Ins}(1,3)\text{P}_2$, AT5G61980 to $\text{Ins}(1,4)\text{P}_2$ and AT5G13300 to $\text{Ins}(1,5)\text{P}_2$ (Fig. 3.3). We predict that the different *Arabidopsis* ArfGAP PH domains mediate membrane recruitment using a combination of non-specific electrostatics and ligand specificity for one or more monophosphorylated PIs, which may vary from member to member within this class of proteins.

```

      *           20           *           40           *           60           *           80
AT1G10870 : VIKQGYLLKR--SSSLRTDWKRRKFFVLD SHGSMYYYYFTNGNKSMSGSHHHYS---GSSDHNTGVFGRFRAR--HNRSGSLTEGS : 76
AT1G60860 : VTKQGYLLKR--SASLRADWKRRRFFVLDNHGSLYYYYNTGNKSAKSHHHYS---GLGEHSSGVFGRFRTR--HNRSSAS--QGS : 74
AT5G13300 : TIIRQGYLSKR--SSNLRGDWKRRRFFVLD SRGMLYYYYRKQCSKPSGSGSQLSGQRN--SSELGSGLSRWLSSNNHGHGGVHDEKS : 81
AT5G61980 : TIIRQGYLSKR--SSNLRGDWKRRRFFIIDS RGMYYYYKPNWWSGNGSRSVVHRNMASENSPGLLSRWLSS--HYHGGVHDEKP : 80
CENB1 : LVMEGHLEKRR--ASNAFKTWRRRWFTIQSNQLVYQKKYK-----DP : 39
CENB2 : IVMEGYLFKRR--ASNAFKTNRRWFSTQNNQLVYQKKFK-----DN : 39
APPL1 : TRKAGYLNARNKTGLVSSTWD RQFYFTQGGNLMSCARG-----DV : 40
Arf1GAP : SEKKGYLLKK--SDGIRKVVQRRKCSVKNGILTISHATS-----N : 38
OsArfGAP : IIKQGYLFKR--SENSR-EWRRRFFVLD SHGTLFYTGNGKQSQGAASQQTATA---GEG---TGVFSRFRFL--NQKAPTQGDNA : 72

      *           100           *           120           *           140
AT1G10870 : LGYNTIILRTSLIKLDAEDMDLRLCFRIISP--QKTYTLQAENGADRMDWVNKITKAIGTL : 135
AT1G60860 : LDCNMIDLETSLIKLDAEDTDLRLCFRIISP--QKTYTLQAENGADRMDWVNKITAAITIR : 133
AT5G13300 : VARHTVNLLTSTIKVDADQSDLRFCFRIISP--TKNYTLQAESALDQMDWIEKITGVIASL : 140
AT5G61980 : VARHTVNLLTSTIKVDADQTDLRFCFRIISP--TKVYTLQAENAQDQMDWIEKITGVIASL : 139
CENB1 : VTVVVDLRLRLCTVKLCPD-SERRFCFEVVT--SKSCLLQADSERLLQLWVSQVSSIASA : 97
CENB2 : PTVVVELLRLCTVKHCED-IERRFCFEVVT--TKSCMLQADSEKLRQAWIKAVQTSIATA : 97
APPL1 : AGGLAMDIDNCSV-MAVDCEDRRYCFQITSFDGKKSSILQAESKKDHEEWICTINNISKQ- : 99
Arf1GAP : RQPAKLNLLTCQVKPNAE---DKKSFDLISH--NRTYHQAEDEQDYVAWISVLTNSKEEA : 94
OsArfGAP : LSCHKIILRLSTIKIDAEENDLRFCFRVISP--MKAYTLQAESGADQKDWIEKVTGVIASL : 131

```

Figure 3.14. The alignment of consensus secondary structure predictions for the *Arabidopsis* ArfGAP PH domains, *H. sapiens* Arf1 GAP PH domain (PDB ID: 2DA0 [328]), *H. sapiens* APPL1 PH domain (PDB ID: 2ELB [322]), *H. sapiens* centaurin β 1 PH domain (Q15027), *H. sapiens* centaurin β 2 PH domain (Q15057) and putative *O. sativa* ArfGAP PH domain (Q6H462) homolog. Rows depict 1-letter code of PH domain sequences, α -helices are highlighted in red, β -strands in blue.

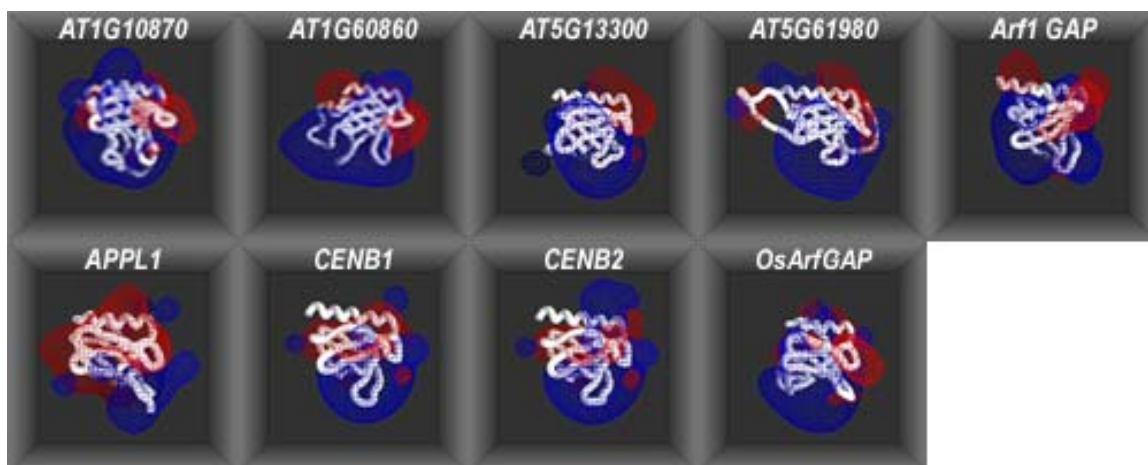


Figure 3.15. Electrostatic properties of the *Arabidopsis* ArfGAP PH domains, *H. sapiens* Arf1 GAP PH domain (PDB ID: 2DA0 [328]), *H. sapiens* APPL1 PH domain (PDB ID: 2ELB [322]), *H. sapiens* centaurin β 1 PH domain (Q15027), *H. sapiens* centaurin β 2 PH domain (Q15057) and putative *O. sativa* ArfGAP PH domain (Q6H462) homolog. 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 [227]. The models are represented as C α backbone traces and their accession numbers are displayed in the upper part of each profile.

G. The RhoGAP class

Rho GTPases are ubiquitously expressed across the eukaryotes where they act as molecular switches that cycle between an active guanosine triphosphate (GTP)-bound and an inactive guanosine diphosphate (GDP)-bound state participating in the regulation of many cellular processes [329-331]. The Rho GTPases can be regulated by positive regulators called guanine nucleotide exchange factors (GEFs) and negative regulators called Rho GTPase-activating proteins (RhoGAPs) encoded by ARHGAP family genes, such as Cdc42, Rac, Arhu and Arhv. Many of the RhoGAPs are implicated in various types of cancer, i.e. RhoGAP4, RhoGAP13, RhoGAP14 and RhoGAP20 in breast cancer [332, 333] and RhoGAP7 in liver cancer [334, 335]. RhoGAPs have been identified in various eukaryotes, ranging from yeast to human [336, 337]. For example, there are twenty three genes containing the RhoGAP domain in *C. elegans* genome [338], forty six in the *D. discoideum* genome, two times as many in the *H. sapiens* genome [339-341]. However, there seem to be only three RhoGAP genes in *Arabidopsis* based on our studies and as published previously [250] that we classify in the *Arabidopsis* RhoGAP class of *Arabidopsis* PH proteins. The N-terminal region of the *Arabidopsis* RhoGAPs contains a PH domain followed by a RhoGAP domain (Fig. 3.1), which is a known fold comprised of nine α -helices [342].

The *Arabidopsis* RhoGAP proteins share their domain architecture with various proteins, such as oligophrenin 1 (implicated in chromosome X-linked cognitive impairment [343]), RhoGAP9 (involved in regulating adhesion of hematopoietic cells to extracellular matrix [344]) and RhoGAP15 (implicated in cell contraction and localizes to the cell surface via its PH domain [345]). The modeled RhoGAP PH domains have biophysical characteristics of typical PH domains with polarized electrostatic profiles (Fig. 3.17) and resemble most homologs except oligophrenin 1, which seems to carry a prominent negative patch in the canonical membrane-binding site. Our docking data suggests that the AT4G24580 and AT5G12150 RhoGAP PH domains exhibit promiscuous PI binding and AT5G19390 RhoGAP PH domain may have affinity only for PA (Fig. 3.3). Although the PH domains share high sequence identity, the binding preferences may be dependant on the orientation of their loops.

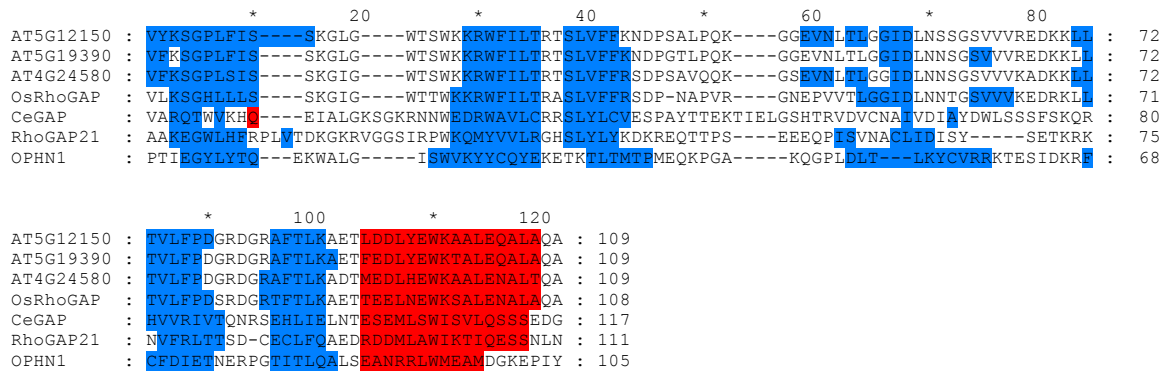


Figure 3.16. The alignment of consensus secondary structure predictions for the *Arabidopsis* RhoGAP PH domains and *H. sapiens* RhoGAP21 (PDB ID: 2DHJ [346]), *H. sapiens* Oligophrenin 1 (O60890), *C. elegans* GAP (P34288) and putative *O. sativa* RhoGAP (Q10Q51) PH domain homologs. Rows depict 1-letter code of PH domain sequences, α -helices are highlighted in red, β -strands in blue.

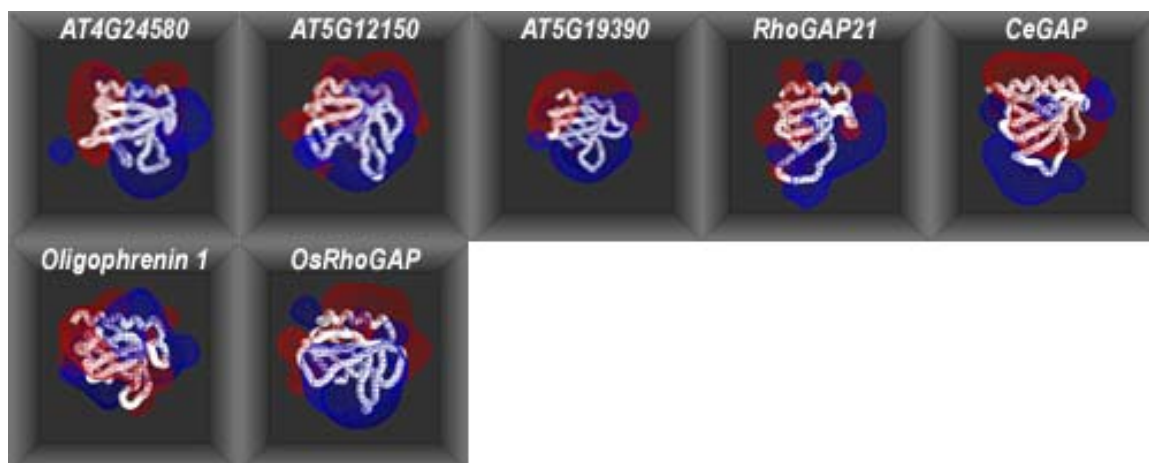


Figure 3.17. Electrostatic properties of the *Arabidopsis* RhoGAP PH domains and *H. sapiens* RhoGAP21 (PDB ID: 2DHJ [346]), *H. sapiens* Oligophrenin 1 (O60890), *C. elegans* GAP (P34288) and putative *O. sativa* RhoGAP (Q10Q51) PH domain homologs. PH domain models are in the same orientation. In all 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 [227]. The models are represented as C α backbone traces and their accession numbers are displayed in the upper part of each profile.

H. The DYN class

Dynamins and dynamin-like proteins play important roles in various biological processes. In animal cells, dynamin proteins function in the scission event during endocytosis at the plasma membrane [347, 348]. In yeast, the dynamin-like protein called vacuolar protein sorting protein 1 (Vps1p) functions in protein transport from Golgi to an endosomal compartment [349, 350], participates in the organization of the actin cytoskeleton [351] and is involved in peroxisome fission via interaction with peroxins [352]. The dynamin-like protein Mgm1p acts in mitochondria maintenance [353]. The *Arabidopsis thaliana* genome contains four dynamin-like proteins, e.g. *Arabidopsis* dynamin-like 1 (ADL1), *Arabidopsis* dynamin-like 2 (ADL2), *Arabidopsis* dynamin-like 3 (ADL3) and ADL6 but ADL1 and ADL2 lacks a PH domain and are unique to plants [347]. AT1G59610 (i.e. ADL3 also known as AtDRP2B) and AT1G10290 (i.e. ADL6 also known as AtDRP2A) are dynamin-like sequences that comprise the dynamin (DYN) class of *Arabidopsis* PH proteins. They are both located on chromosome number one and they share a 96% sequence similarity. AT1G59610 and AT1G10290 are small GTPases that are characterized by the presence of an N-terminal dynamin GTPase (DYNc), a dynamin M, a PH and a C-terminal GTPase effector (GED) domain (Fig. 3.1). The dynamin M region interacts with DYNc and myosin head, but its biological function remains unknown [160]. The GED functions as an assembly domain [354]. Occasionally, some authors mention that the PRD motif characteristic to mammalian dynamin I is also present in ADL6 [355], but not in AT1G59610 [356]. However, we are unable to confirm the presence of the PRD motif in AT1G10290 as well.

The PH domain of dynamin I was shown to mediate membrane association and bind specifically to $\text{PtdIns}(4,5)P_2$ [57, 357]. Recently, a two-step model for dynamin function was put forward, in which initially dynamin functions as a scaffolding molecule and later, after self-assembly, as an active participant in the process of membrane fission in vesicle formation [358].

AT1G10290 is highly related to mammalian dynamin I and has been shown to be required for trafficking from the trans-Golgi network to the central vacuole [253]. It has been shown that AT1G10290 primarily associates with membrane *in vivo* and binds with high affinity and specificity to $\text{PtdIns}(3)P$. It has also been shown to bind to $\text{PtdIns}(4)P$ *in vitro*, although to a

lesser extent [101]. The PH domain of AT1G10290 shows equal binding affinity to both $\text{PtdIns}(3)P$ and $\text{PtdIns}(4)P$ [101]. It has been suggested that the AT1G10290 PH domain facilitates the PI-binding, but the binding specificity is determined by a combination of the AT1G10290 PH domain and the C-terminal domain of AT1G10290 [101]. The *Arabidopsis* Sec13 homolog 1 (AtSeh1) was suggested to regulate bound vs. soluble cycling of AT1G10290 through inhibition of its PH domain binding to $\text{PtdIns}(3)P$ via the competition between the AtSeh1 and the C-terminal domain of AT1G10290 for the AT1G10290 PH domain [359]. However, we do not pick up these binding specificities in our docking results for AT1G10290 (Fig. 3.3); it is likely that the differences may arise from the suboptimal orientation of loops in the models.

Despite the overall 96% sequence similarity between AT1G10290 and AT1G59610, AT1G59610 PH domain is thought not to be homologous to human dynamin 1 [356]. The N-terminal extension of AT1G59610 is unique amongst other dynamin-like proteins and it has been suggested that AT1G59610 exhibits unique subcellular localization as yet not observed for other dynamin-like proteins [356]. In fact, AT1G59610 is predicted to translocate to chloroplast stroma, mitochondrial matrix space, microbody, nucleus and/or associate with membrane via phospholipid-binding through its PH domain [356]. The docking results suggest the modeled AT1G59610 PH domain can bind to $\text{Ins}(1,3)P_2$, $\text{Ins}(1,4)P_2$, and $\text{Ins}(1,5)P_2$ showing a preference for the headgroups of monophosphorylated PIs (Fig. 3.3). However, both *Arabidopsis* DYN models exhibit highly similar profiles with typical electrostatic polarization and most likely interact with the membrane non-specifically with similar affinity (Fig. 3.19).

```

      *           20           *           40           *           60           *           80
AT1G10290 : EITAGYLMKKS AKTNGWSRRWFVINEKTGKLGYT KKQERNFRGVTITLEE CTIEETPEDEVEKSKSSKD-KKANGPDSKGPGLV : 83
AT1G59610 : EITAGYLMKKS AKTNGWSRRWFVINEKTGKLGYT KKQERNFRGVTITLEE CSIEETSDDEGEKSKSSKD-KKSNGPDSKGPGLV : 83
DYN1      : MKTSG----NQDEILVIRKGWLTIN---NIGIMKGGSKKEYWFLTAENLSWYKDDEEKEKKYMLSVDNLLKLRDVEKGFMSKH : 76
DYN2      : -GSSG----SSG---VIRGWLTIN---NISLMKGGSKKEYWFLTAESLSWYKDDEEKEKKYMLPLDNLKLRDVEKGFMSKH : 72
OsDyn     : EITAGFLFKKRAKANDWSNRWFVLNERSGKLGYT KKQERHFRGVINLEECNLEEVLDDE-DPTRSLKDPKKPNVPDIGTPGLI : 83

      *           100           *           120           *
AT1G10290 : FKITCKVPYKTVLKAHNALVLKAESVVDKNEWINKLQKVIQAR----- : 126
AT1G59610 : FKITCRVPYKTVLKAHNALVLKAESVVDKNEWINKLQKVIQAR----- : 126
DYN1      : IFALFNTEQRNVYKDYRQLELAETQEEVDSWKASFLRAGVYTERVGDK : 125
DYN2      : VFALENTEQRNVYKDLRQIELACDSQEDVDSWKASFLRAGV----- : 113
OsDyn     : FKITHKVAYKNVLKAHKAIIILKAESMAEKTEWVTKIRSIVQK----- : 126

```

Figure 3.18. The alignment of consensus secondary structure predictions for the *Arabidopsis* DYN PH domains, *H. sapiens* dynamin 1 (PDB ID: 1DYN [86]), *H. sapiens* dynamin 2 (PDB ID: 2YS1 [86]) and putative *O. sativa* dynamin (Q7EZV4). Rows depict 1-letter code of PH domain sequences, α -helices are highlighted in red, β -strands in blue.



Figure 3.19. Electrostatic properties of the *Arabidopsis* DYN PH domains, *H. sapiens* dynamin 1 (PDB ID: 1DYN [86]), *H. sapiens* dynamin 2 (PDB ID: 2YS1 [86]) and putative *O. sativa* dynamin (Q7EZV4). 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 [227]. The models are represented as C α backbone traces and their accession numbers are displayed in the upper part of each profile.

I. The AtPH1 class

The *Arabidopsis thaliana* pleckstrin homologue 1 (AtPH1) class includes AT2G29700 and AT5G05710 *Arabidopsis* PH proteins. The AtPH1 class consists of a PH domain only (Fig. 3.1). Not much is known about either proteins but expression analysis studies for AT2G29700 suggests that it is expressed in almost all plant parts, flower, root, leaf, silique, stem and seed tissues but at varying levels, e.g. strong expression in root, moderate expression in leaf, silique, stem and seed and weak expression in flower [360]. Although, these proteins are annotated as homologues of Pleckstrin, AT5G5710 and AT2G29700 PH domains share high sequence identities of 34% and 32%, respectively, and biophysical profiles with cytohesin-3.

AtPH1 PH domains exhibit weaker positive potentials than the most of the other classes of *Arabidopsis* PH domains (Fig. 3.2) and show no specific recognition of InsPtd ligands (Fig. 3.3) based on our docking analyses. One study however suggests that the PH domain of AT2G29700 binds with high affinity to PtdIns(3)*P* [61]. We predict that this class of PH domains bind membranes non-specifically with low to moderate affinity and probably need another determinant such as interaction with a membrane associated protein or oligomerization to effectively carry out their membrane targeting function.

```

      *           20           *           40           *           60           *           80
AT5G05710 : P E R T G W L T K Q G E Y I K T W R R R W F V L K Q G K L F W F F D S D V T -- R V S R P R G V V F V E S C L T A K G A E D V L N --- K Q N A F E L S T --- : 72
AT2G29700 : P E R S G W L T K Q G D Y I K T W R R R W F V L K R G K L L W F F D Q A A A G I R G S T P R G V I S V G D C L T V K G A E D V V N --- K P F A F E L S S --- : 74
N-Pleck   : R I R E G Y L V K K G S V F N T W K P M W V V L L E D G I E F Y K K K S D N --- S P K G M I F L K G S T L T S P C Q D F G K --- R M F V F K I T T --- : 69
C-Pleck   : I I K Q G C L L K Q G H R R K N K V R K F I L R E D P A V L H Y Y F P A G --- A E D P L G A I H L R G C V V T S V E S N S N G R K S E E E N L F E I I T --- : 75
OsPleck   : G E R T G W L N K Q G E Y I K T W R R R W F V L K Q G R L F W F F D A A V T -- R G S V P R G V I F V A T L T V K G A E D V I N --- R Q F A F E L S T --- : 72
Cytohesin3 : P D R E G W L L K I G G R V K T W K R R W F I L T D N C L Y Y F E Y T T D K --- E P R G I I F L E N - L S I R E V E D P R --- K P N C F E L Y N P S H K G : 72

      *           100          *           120
AT5G05710 : -----R N E T M Y F I A D - S E K E K E D W I N S I G R S I V Q N : 101
AT2G29700 : -----G S Y T M F F I A D - N E K E K E E W I N S I G R S I V Q H : 103
N-Pleck   : -----T K Q Q D H F F Q A A F L E E R D A W V R D I N K A I K C I : 99
C-Pleck   : -----A D E V H Y F L Q A A T P K E R T E W I K A I Q M A S R T G : 105
OsPleck   : -----P T D T M Y F I A D - S E K E K E E W I N S I G R S I V Q H : 101
Cytohesin3 : Q V I K A C K T E A D G R V V H G N H V V Y R I S A P S P E E K E E W M K S I K A S I S R D : 118

```

Figure 3.20. The alignment of consensus secondary structure predictions for the *Arabidopsis* AtPH PH domains, *H. sapiens* N-terminal pleckstrin (PDB ID: 1PLS [81]), *H. sapiens* C-terminal pleckstrin (PDB ID: 1XX0 [361]), putative *O. sativa* (NP_001055692) homolog and *M. musculus* cytohesin-3 (PDB ID: 2R0D [362]). Rows depict 1-letter code of PH domain sequences, α -helices are highlighted in red, β -strands in blue.

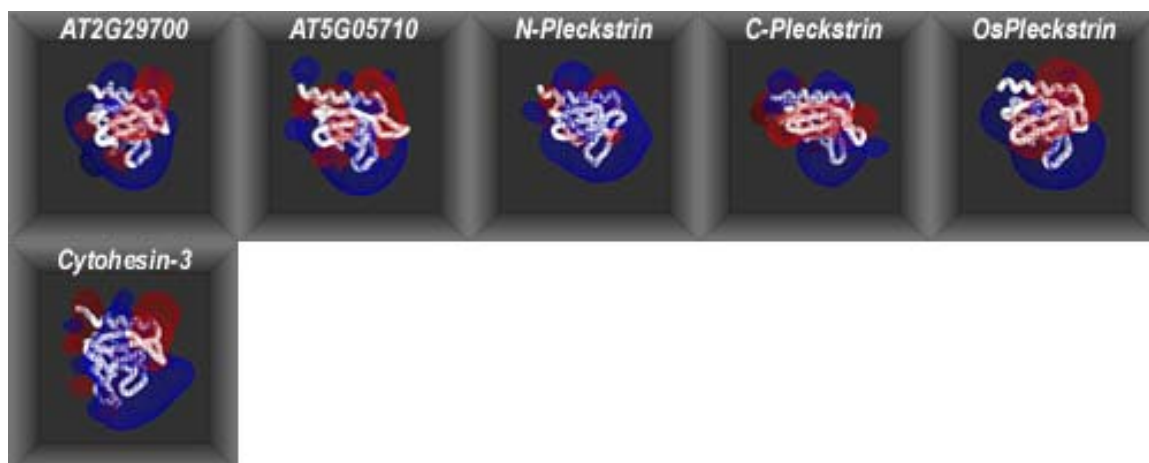


Figure 3.21. Electrostatic properties of the *Arabidopsis* AtPH PH domains, *H. sapiens* N-terminal pleckstrin (PDB ID: 1PLS [81]), *H. sapiens* C-terminal pleckstrin (PDB ID: 1XX0 [361]), putative *O. sativa* (NP_001055692) homolog and *M. musculus* cytohesin-3 (PDB ID: 2R0D [362]). PH domain models 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 [227]. The models are represented as α backbone traces and their accession numbers are displayed in the upper part of each profile.

J. The DUF828, UNK1 and UNK2 classes

The Pfam database entry describes domain of unknown function 828 (DUF828) family as actually consisting of several plant proteins of unknown function [160]. Currently, there are at least eighteen DUF828 proteins in *Arabidopsis* of which nine make up one of the largest classes of *Arabidopsis* PH proteins. The *Arabidopsis* DUF828 proteins consist of an N-terminal DUF828 domain and a C-terminal PH domain (Fig. 3.1). The unknown 1 (UNK1) protein has a PH domain in combination with an N-terminal MRS6 domain that is also found in vacuolar protein sorting-associated proteins known to be involved in intracellular trafficking and secretion [251] (Fig. 3.1). The unknown 2 (UNK2) proteins contain only a PH domain (Fig. 3.1).

Most of the PH domains from DUF828, UNK1 and UNK2 classes have additional secondary structure elements that may possibly contribute to their function. For example, the DUF828 PH domains have an extra N-terminal α -helix prior to the typical six/seven β -strands, followed by a C-terminal α -helix (Fig. 3.22) whereas the UNK1 PH domain has an additional α -helix in the $\beta 6/\beta 7$ loop (Fig. 3.24).

The function and ligand-specificity of the *Arabidopsis* members of DUF828, UNK1 and UNK2 classes and their PH domains have not been established. Most of the PH domain models show mostly positive electrostatic profiles (Fig. 3.23 and 3.25). Others, for example DUF828 AT5G57770 and UNK2 AT4G23895 PH domains have a strongly negative overall electrostatic potential (Fig. 3.23 and 3.25), which resembles the electrostatic profile reported for the structure of the PH domain from *C. elegans* muscle protein UNC-89 [83]. These domains seem to adopt the classic PH domain fold, but have binding pockets lined with negative charges, which may be indicative of non-canonical functions or membrane targeting via novel mechanisms.

Our docking results suggest that UNK1 may bind $\text{Ins}(1,3)P_2$, $\text{Ins}(1,4)P_2$ and $\text{Ins}(1,5)P_2$, UNK2 AT1G17820 PH domain may be specific for PA and UNK2 AT1G73200 PH domain for $\text{Ins}(1,4)P_2$ whereas the remaining UNK2 PH domains do not show a clear-cut specificity for any particular InsPtds (Fig. 3.3).

```

      *           20           *           40           *           60           *           80
AT4G17350 : LPFEGELMQCARNGVQRTKRVCVYINKKSVMIKLSKSHVGGAFSKKIKCVVYGV CDEKSAWPY-RKEREN-SEEVYFGLKGTGQ : 82
AT5G47440 : LPLEGELVQCARNGLQRNKRVCVYINKKSVMIKLSKSHVGGAFSKKIKCVVYGV CDEISAWPC-RKEREN-SEEVYFGLKGTGQ : 82
AT4G16670 : EESTELMQCTRNGVLRWKHVYVYINKKSVVIVEIKSKHVVGAFSMKSGIVNDVCETVSGLQN-GKDTENTEEELYFGIGTGK : 83
AT3G22810 : LARGGQLLKRTRKGDHLHWKIVSVYINRLNQVILKMKSRHVGGTFTKKNKNVVIDVIKNVQAWPGR-HLLEGGEDLRYFGLKTVP : 83
AT4G14740 : LARGCELKRTRKGDHLHWKIVSVYINKMNVMLKMKSRHVGGTFTKKKKNIVLDVIKNVPAWPGR-HLLEGGDDLRYFGLKTVM : 83
AT5G43870 : LAKGCELKRTRKGDHLHWKVSVYINRTKQVILKTSKHVAGTTTKKKKNVVVGIVKGLPAWPGR-EMLEGGENLRYFGLKTVE : 83
AT3G63300 : LAKGTELLKRTRGGELHWKIVSVYINKAGQAVLKMKSKHVGGTFTKKKKHMVLEV RKDIPAWAGR-DLFNG-DKHHYFGLKTET : 82
AT4G32785 : VSRGEELLKRTRKGDHLHWQVSFNINSNWQVVLKMKSKHVGGTFTKTKKCVVNGVCRDIPEWAHRGRADKMVERRAYFGVKTVE : 84
AT5G57770 : LAKGTDLFVETPDGDFKARTVSMVLNKDGVILKMKKHNL---LRTKKE SIVTNVHVELYKDSDS--EDNNIEDTCYLIVLRNTN : 79
VvDUF828 : LARGSELKRTRNGDLHWKMVFVYIHRTGQVMLKMKSRHVAGTTTKKKKIVVLEVVKDMPAWPGR-HLLEGGEQRHYFGLKTTP : 83
BoDUF828 : LPLEGELMQCARNGVQRTKRVCVYINKKSVVILKLSKSHVGGAFSKKIKCVVYGV CDEKSAWPY-RKEREN-SEEVYFGLKGTGQ : 82

      *           100           *
AT4G17350 : G-LLEFKCKSKT-HKQRWVDGVQSLLRQVNCFF : 113
AT5G47440 : G-LLEFKCKSKT-HKQRWVAGIQSNLRLVSCLE : 113
AT4G16670 : G-LTKFKCKSKA-DKQTWVDSIRNLLHRVTAVE : 114
AT3G22810 : RGIVEFCCKSKR-EYEMWTQGVSRLLIAVAERN : 115
AT4G14740 : RGDVEFEVKSQR-EYEMWTQGVSRLLVLAERK : 115
AT5G43870 : KRVEFECKSKR-EYDLWTQGVSMLLSIASDRK : 115
AT3G63300 : KRVEFECKRNQR-EYEIWTQGVSRLLAIAAEKX : 114
AT4G32785 : R-VIEFECKNKR-EKQMWIEGIQQLNLSLEIGS : 115
AT5G57770 : RGAIKLDMADDYNRYKTWVTTIQHMLTLSSSSL : 112
VvDUF828 : RGVVEFECKRNQR-EYDIWTQGVVERLLTIV---- : 111
BoDUF828 : G-LLEFKCKSKV-HKQRWVDGVESLLHQV---- : 109

```

Figure 3.22. The alignment of consensus secondary structure predictions for the *Arabidopsis* DUF828 PH domains and their putative *V. vinifera* (A5BJ67) and *B. oleracea* (Q2A9D6) homologs. Rows depict 1-letter code of PH domain sequences, α -helices are highlighted in red, β -strands in blue.



Figure 3.23. Electrostatic properties of the *Arabidopsis* DUF828 PH domains and their putative *V. vinifera* (A5BJ67) and *B. oleracea* (Q2A9D6) homologs. 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 [227]. The models are represented as C α backbone traces and their accession numbers are displayed in the upper part of each profile.

```

      *          20          *          40          *          60          *          80
AT1G17820 : -----KQKKRGLLETHPLRKF--ARIK--DHKLILSDADSTQSETTVSLIG----CSIEA : 49
AT1G73200 : LNKQGVIWYLELDENVKDWMEKLPKEOKKHVRDLELVHPVRR--ARIK--DHKLILSDS-LDGPQTFITLKG----CFVDA : 74
AT2G30880 : -----NLLQGFLFKRS--ETLRKWNERRVILLPTTGKMEYKTRRNE-----PTIKG : 44
AT4G23895 : -----LEHVDVLRSAISGYLFIRS--RVALPGEEPPIDDTEVDDWLPFRFVVLQGPCLFFYYLLS : 59
AT4G17140 : -----ADFEGLWLSILSWKG--REASNRRYLCLVGPFIYVLESPGSKSYKK--YTSIRG : 50
UNC-89 : -----MGDTGKLGRIIRHDAFCVWEGDEPPKLRVVFLEFRNKIMFTECD-----ASTS : 47

      *          100          *          120          *          140          *          160
AT1G17820 : VSGSDLPTAKKWARFPFK-VESKISP-----ALYKGN---QVFYIYLE---TSWEKESWCKALRLASCEN----- : 107
AT1G73200 : VSGSG-PTAKKWARFPFQ-VESKTS-----VLYKGN---RVFYIFLE---TSWEKESWCKALRLAACEN----- : 130
AT2G30880 : TILFDENSTISISPVNFQGLPKYNGC-----CIYIGTPQKKDYFLCAE---TPGAAKAWVTTLHATQLVLK----- : 107
AT4G23895 : TDLSPQDSTLADIVEIGSVPSYTREFDETNYCFYILTQGLRFECST---SKTQVDSWLSVLRLDCKSE----- : 127
AT4G17140 : KHIYKVPVELAGGVEHVLSTNASR-----ISEKVMEDVNALILMD---SEDSRKTWHSRLQGANVYRA----- : 111
UNC-89 : PPSYTHYSIRLDKYNIRQHTTDEDT-----IVLQPQEPGLPSIRIKPKDFETSEYVRKANLRDIAEEQEKYAAERD : 119

```

Figure 3.24. The alignment of consensus secondary structure predictions for the *Arabidopsis* UNK 1-2 PH domains and *C. elegans* muscle protein UNC-89 (PDB ID: 1FHO [363]). Rows depict 1-letter code of PH domain sequences, α -helices are highlighted in red, β -strands in blue.

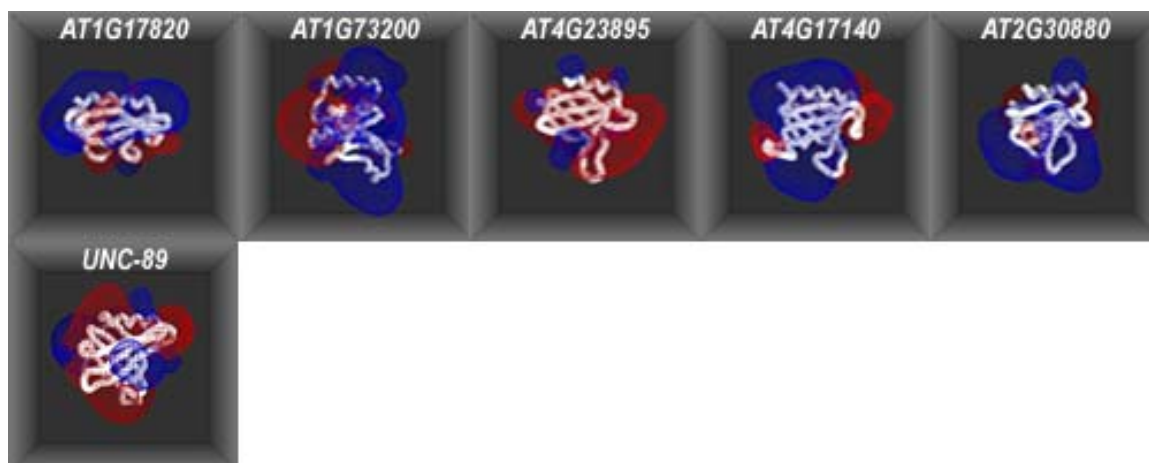


Figure 3.25. Electrostatic properties of the *Arabidopsis* UNK 1-2 PH domains and *C. elegans* muscle protein UNC-89 (PDB ID: 1FHO [363]). PH domain models are in the same orientation. In all panels, the electrostatic potentials were calculated at +25 mV and -25 mV equipotential contours in 0.1 M KCl. All images of the electrostatic potential contours were calculated with GRASP [227]. The models are represented as Cα backbone traces and their accession numbers are displayed in the upper part of each profile.

Summary

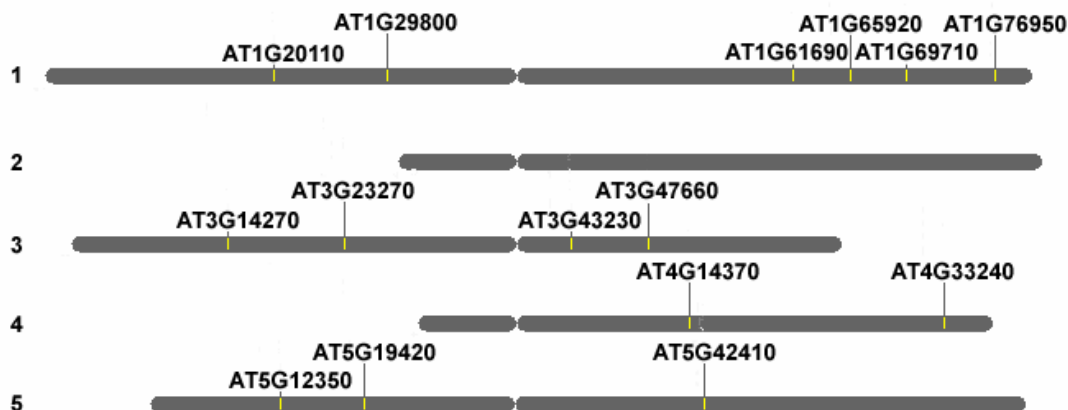
The *Arabidopsis thaliana* genome contains forty-nine PH proteins, which can be grouped into twelve classes. Each *Arabidopsis* PH protein class has a characteristic domain architecture. Within individual classes, members often share moderate to high sequence homology and possibly redundant membrane binding behaviors to elicit their functional role. Outside the classes, there is little functional or structural similarity; the only thing in common is the presence of a PH domain. The examination of PH proteins encoded in the entire proteome of *Arabidopsis thaliana* leads us to the conclusion that there is a considerable heterogeneity in their sequence based characteristics and function. Twenty-seven of plant PH proteins have counterparts in worm, fly and vertebrate proteomes but the remaining twenty-two are unique to plants, e.g. the PRAF, DUF828, UNK1 and UNK2 proteins.

On average the pairwise sequence identity of *Arabidopsis* PH domains ranges from 4% to 71% for different members of the different classes. As mentioned earlier, PH domains are associated with many functions including the PI-binding and membrane targeting. Of the twelve PH classes identified in the *Arabidopsis* proteome, experimental studies suggest high affinity binding for members of PRAF PH domains to $\text{PtdIns}(4,5)P_2$ [128], PDK PH domains to PA, $\text{PtdIns}(3)P$, $\text{PtdIns}(3,4)P_2$, $\text{PtdIns}(3,4,5)P_3$ [104], ArfGAP PH domains to $\text{PtdIns}(4)P$ [252], DYN and AtPH1 PH domains to $\text{PtdIns}(3)P$ [61, 359]. Similarly, the function has been demonstrated for only a few of the *Arabidopsis* PH proteins, e.g. START protein called EDR2 [279], PLD ζ 1 and PLD ζ 2 [287], PLD ζ 2 [290], PDK [364, 365], ArfGAP protein referred to as VAN3 [252] and DYN protein termed ADL6 [253]. Currently, most *Arabidopsis* PH proteins, in particular members of the RhoGAP, AtPH1, DUF828, UNK1 and UNK2 classes remain uncharacterized and our computational analyses provides new insight to the properties and possible mechanisms of membrane targeting and function.

Chapter 4 *Arabidopsis* FYVE domains

Identification, characterization and chromosomal localization of FYVE proteins encoded in the *Arabidopsis* genome

The total number of FYVE proteins seems to be directly correlated with the total estimated number of genes for a given organism, e.g. twenty-seven FYVE encoding genes in a total of 42,000 in *H. sapiens*, thirteen in a total of 18,000 in *C. elegans* and five in a total of 6,000 in *S. cerevisiae* [366]. We identified fifteen *Arabidopsis* FYVE proteins in the *Arabidopsis* protein sequence database i.e. TAIR6_pep_20051108. Our finding of fifteen FYVE proteins encoded within predicted 25,500 genes [21] of the *Arabidopsis* genome (Fig. 4.1A) falls in line with the above observation. At the very least, the observed correlation is based on a coincidence, and at the very best, it is a novel phenomenon suggesting that some proteins might maintain a specific ratio to other proteins to function optimally. The initial identification was done using an automated pipeline [144]. Later, the total number of *Arabidopsis* FYVE proteins and their individual accession numbers were verified through manual searches performed in various databases. The fifteen *Arabidopsis* FYVE proteins represent the entire family of *Arabidopsis* proteins (Fig. 4.1A). Fig. 4.1B displays the schematic localization of the fifteen *Arabidopsis* FYVE proteins within the *Arabidopsis* genome. The fifteen identified encoding sequences of genuine *Arabidopsis* FYVE proteins are dispersed throughout the *Arabidopsis* genome, being located on all chromosomes except chromosome two (Fig. 4.1B). The disagreement of our total with previously reported totals of nine [128], over ten [138] and most recently, sixteen [250] stems from misannotations, incomplete and incorrect identification of the FYVE domains. For example, AT1G61620, AT1G66040, AT1G66050 and AT5G39550 proteins are all misannotated as FYVE proteins but do not actually possess FYVE domains.



Domain Architecture of *Arabidopsis* FYVE proteins

On the basis of domain architecture we propose five classes of *Arabidopsis* FYVE proteins (Fig. 4.2). Class I comprises two out of four documented *Arabidopsis* Fab1p homologues expressed in plants, i.e. AT3G14270 and AT4G33240 [367, 368]. The other two Fab1 homologues do not contain a FYVE domain [367, 368]. Class I members, i.e. AT3G14270 and AT4G33240, contain a FYVE domain, followed by Fab1_TCP(chaparonin-like) and PIPKc domains. AT3G14270 is annotated in UniProtKB/TrEMBL database as a member of “phosphatidylinositol-4-phosphate 5-kinase family” while AT4G33240 as a “hypothetical protein”. Conversely, the Munich Information Center for Protein Sequences (MIPS) annotation for AT4G33240 is “similar to FAB1 protein” and for AT3G14270 is “unknown protein” [250]. Our Blast analysis reveals similarity of both class I members to hypothetical proteins ppk-3 (*C. elegans*), Fab1p (*S. cerevisiae*) and FYV1 (*H. sapiens*) (Fig. 4.5). Ppk-3 and Fab1p proteins share domain architecture identical to class I members and FYV1 protein has an additional DEP domain (data not shown). The sub-classes IIA and IIB are represented by one sequence each. According to SMART database, AT3G43230 of IIA and AT1G29800 of IIB possess a set of two domains: a FYVE domain and a domain of unknown function 500 (DUF500). However, we also found a contradicting indication that AT3G43230 of IIA has a CD04366 domain while AT1G29800 of IIB contains a COG2930 domain instead of the DUF500 domain, which prompted the sub-classification. Class III comprises the AT1G61690 protein and class IV comprises the AT1G20110 protein. Both classes are unique in that they contain only a FYVE domain but they differ in the placement of the FYVE domain (N-terminus versus C-terminus) and its biophysical properties. UniProtKB/TrEMBL annotates function for class II – IV as either hypothetical or unknown. The representation of class II – IV members in the literature is full of contradictions. They are not mentioned in the classification by Drobak and Heras [138] and class IIB and III are omitted from the classification by Jensen et al [128]. Intriguingly, the class IV protein are identified as AtAAF79901 and shown to contain a FYVE domain followed by a plant specific SGNH-plant-lipase-like domain [128]. Our data suggests, however, that class IV protein is over 300 amino acids shorter than AtAAF79901, and it does not contain a SGNH-plant-lipase-like domain. Class

II sequences, on the other hand seem to contain additional domains not represented by van Leeuwen, e.g. class IIB AT1G29800 protein has an extra DUF500 (or COG2930) domain [250]. Class V is the largest class and it includes nine *Arabidopsis* FYVE 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. In addition, seven out of nine class V proteins are characterized by the presence of a DZC motif found near the C-terminus. UniProtKB/TrEMBL annotates function for class V members 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.

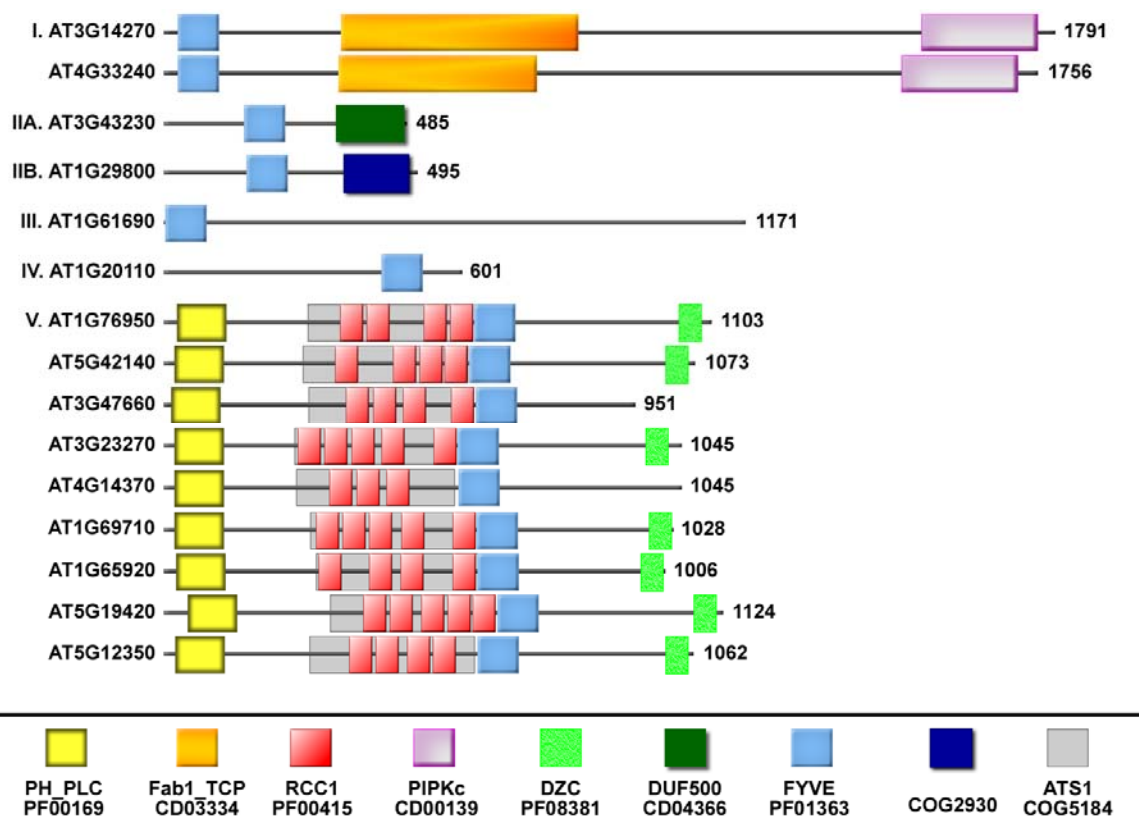


Figure 4.2. Cartoon of representative examples of the five different classes of *Arabidopsis* FYVE proteins. The Pfam (PF), SMART (SM), Conserved Domains (CD) and Clusters of Orthologous Groups (COG) accession numbers for the different domains are given under the figure. The lengths of each protein are indicated on the right.

The SMART database recognizes between three and five RCC1 regions within class V of *Arabidopsis* FYVE proteins, whereas the CD-search identifies additionally yeast domain with similarity to human RCC1 domain, ATS1 domain, overlapping the RCC1 blades (Fig. 4.2). 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 class V of *Arabidopsis* FYVE 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) [369] and Golgi apparatus (HERC1) [370]. Figure 4.3 shows an internal sevenfold sequence repeat of 51–68 residues present in the solved structure of human RCC1 [371] aligned with putative RCC1 regions of class V *Arabidopsis* FYVE 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 [371]. 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 [371]. Our data show that putative RCC1 blades of *Arabidopsis* FYVE 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. 4.3). 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. In reality, the first blade of putative *Arabidopsis* RCC1 may not even be a potential repeat for at least seven out of nine class V *Arabidopsis* FYVE proteins because they share a low sequence similarity in the corresponding region as compared to other regions. It is likely however, that the first blade of AT5G19420 and AT5G12350 is a genuine repeat given the slightly higher residue similarity (Fig. 4.3).

Diagram illustrating the domain architecture of the HS-RCC1 protein, showing various domains (1A, 1B, 1C, 1D, 2A, 2B, 2C, 2D, 3A, 3B, 3C, 3D, 4A, 4B, 4C, 4D, 5A, 5B, 5C, 6A, 6B, 6C, 6D, 7A, 7B, 7C, 7D, 7E, 7F, 7G, 7H, 7I, 7J, 7K, 7L, 7M, 7N, 7O, 7P, 7Q, 7R, 7S, 7T, 7U, 7V, 7W, 7X, 7Y, 7Z) and their corresponding amino acid sequences across different species (HS-RCC1, AT1G76950, AT5G42140, AT3G47660, AT3G23270, AT4G14370, AT1G69710, AT1G65920, AT5G19420, AT5G12350).

The diagram shows the alignment of the HS-RCC1 protein sequence with other species (AT1G76950, AT5G42140, AT3G47660, AT3G23270, AT4G14370, AT1G69710, AT1G65920, AT5G19420, AT5G12350) across various domains (1A, 1B, 1C, 1D, 2A, 2B, 2C, 2D, 3A, 3B, 3C, 3D, 4A, 4B, 4C, 4D, 5A, 5B, 5C, 6A, 6B, 6C, 6D, 7A, 7B, 7C, 7D, 7E, 7F, 7G, 7H, 7I, 7J, 7K, 7L, 7M, 7N, 7O, 7P, 7Q, 7R, 7S, 7T, 7U, 7V, 7W, 7X, 7Y, 7Z).

Key domains and motifs identified include:

- 1A**: HS-RCC1: RRSPPADAIPKSKKVKVSHRSHSTET... (AT1G76950: TKVSGSDGFRVSVSSAQSSSSHGSA...)
- 1B**: HS-RCC1: ... (AT1G76950: ...)
- 1C**: HS-RCC1: ... (AT1G76950: ...)
- 1D**: HS-RCC1: ... (AT1G76950: ...)
- 2A**: HS-RCC1: ... (AT1G76950: ...)
- 2B**: HS-RCC1: ... (AT1G76950: ...)
- 2C**: HS-RCC1: ... (AT1G76950: ...)
- 2D**: HS-RCC1: ... (AT1G76950: ...)
- 3A**: HS-RCC1: ... (AT1G76950: ...)
- 3B**: HS-RCC1: ... (AT1G76950: ...)
- 3C**: HS-RCC1: ... (AT1G76950: ...)
- 3D**: HS-RCC1: ... (AT1G76950: ...)
- 4A**: HS-RCC1: ... (AT1G76950: ...)
- 4B**: HS-RCC1: ... (AT1G76950: ...)
- 4C**: HS-RCC1: ... (AT1G76950: ...)
- 4D**: HS-RCC1: ... (AT1G76950: ...)
- 5A**: HS-RCC1: ... (AT1G76950: ...)
- 5B**: HS-RCC1: ... (AT1G76950: ...)
- 5C**: HS-RCC1: ... (AT1G76950: ...)
- 6A**: HS-RCC1: ... (AT1G76950: ...)
- 6B**: HS-RCC1: ... (AT1G76950: ...)
- 6C**: HS-RCC1: ... (AT1G76950: ...)
- 6D**: HS-RCC1: ... (AT1G76950: ...)
- 7A**: HS-RCC1: ... (AT1G76950: ...)
- 7B**: HS-RCC1: ... (AT1G76950: ...)
- 7C**: HS-RCC1: ... (AT1G76950: ...)
- 7D**: HS-RCC1: ... (AT1G76950: ...)
- 7E**: HS-RCC1: ... (AT1G76950: ...)
- 7F**: HS-RCC1: ... (AT1G76950: ...)
- 7G**: HS-RCC1: ... (AT1G76950: ...)
- 7H**: HS-RCC1: ... (AT1G76950: ...)
- 7I**: HS-RCC1: ... (AT1G76950: ...)
- 7J**: HS-RCC1: ... (AT1G76950: ...)
- 7K**: HS-RCC1: ... (AT1G76950: ...)
- 7L**: HS-RCC1: ... (AT1G76950: ...)
- 7M**: HS-RCC1: ... (AT1G76950: ...)
- 7N**: HS-RCC1: ... (AT1G76950: ...)
- 7O**: HS-RCC1: ... (AT1G76950: ...)
- 7P**: HS-RCC1: ... (AT1G76950: ...)
- 7Q**: HS-RCC1: ... (AT1G76950: ...)
- 7R**: HS-RCC1: ... (AT1G76950: ...)
- 7S**: HS-RCC1: ... (AT1G76950: ...)
- 7T**: HS-RCC1: ... (AT1G76950: ...)
- 7U**: HS-RCC1: ... (AT1G76950: ...)
- 7V**: HS-RCC1: ... (AT1G76950: ...)
- 7W**: HS-RCC1: ... (AT1G76950: ...)
- 7X**: HS-RCC1: ... (AT1G76950: ...)
- 7Y**: HS-RCC1: ... (AT1G76950: ...)
- 7Z**: HS-RCC1: ... (AT1G76950: ...)

Figure 4.3. Sequence alignment of human RCC1 and putative *Arabidopsis* homologues (class V). The secondary structures are adapted from solved structure of human RCC1 with minor modifications [371]. 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 [371].

Molecular models of the *Arabidopsis* FYVE domains and their biophysical properties

We built models of all FYVE domains present in fifteen *Arabidopsis* FYVE proteins listed in Fig. 4.1. Electrostatic profiles of class I *Arabidopsis* AT3G14270 FYVE model, which has a net charge of +7 (including zinc ions), bears a resemblance to *D. melanogaster* Hrs FYVE domain (PDB ID: 1DVP [132]) (Fig. 4.4). The model of the other class I member, AT4G33240 FYVE domain, exhibits weaker positive potential. Both members of class II *Arabidopsis* FYVE domains share a potential profile that is similar to the profile observed for *H. sapiens* EEA1 FYVE domain (PDB ID: 1HYI [133]). Class IIA AT3G43230 FYVE domain has a net charge of +6 and class IIB AT1G29880 FYVE domain has a net charge of +8. Intriguingly, the model of class III AT1G61690 FYVE domain has a very strong positive potential observed also for all models of class V *Arabidopsis* FYVE domains, *M. musculus* 19 protein FYVE domain (PDB ID: 1WFK [372]) and to *H. sapiens* 27 isoform b FYVE domain (PDB ID: 1X4U [373]). Their overall net charge is highly positive, but varies from +9 to +16. The electrostatic profile of class IV *Arabidopsis* FYVE model is distinctive in that its positive potential is the weakest of all potentials observed for *Arabidopsis* FYVE domain models. Additional electrostatic profiles for the alternative models, their PDB coordinate files and verification profiles are available at our website.

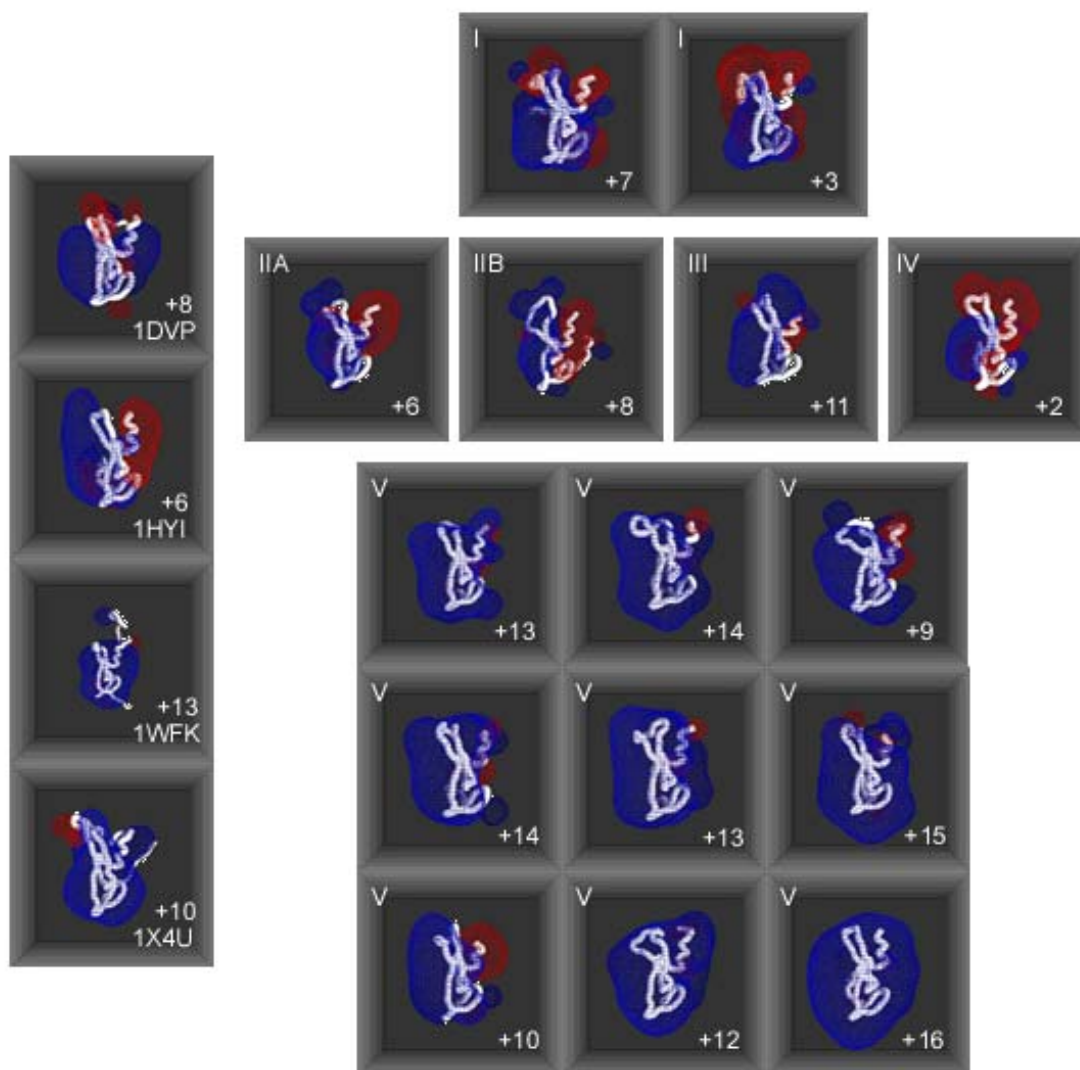


Figure 4.4. The electrostatic profiles of five known FYVE domain structures, i.e. 1DVP [132], 1HYI [133], 1WFK [372] and 1X4U [373] and fifteen *Arabidopsis* FYVE homology models i.e. groups I – V. All FYVE domains are shown in the same orientation with the membrane binding regions facing down. The red and blue meshes represent the -1 kT/e and $+1$ kT/e equipotential contours of the FYVE domains. The numbers in the left corner correspond to the *Arabidopsis* FYVE classification described earlier (Fig. 4.2). From left to right, class I AT3G14270 and AT4G33240; class IIA AT3G43230; class IIB AT1G29880; class III AT1G61690; class IV AT1G20110; class V 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.

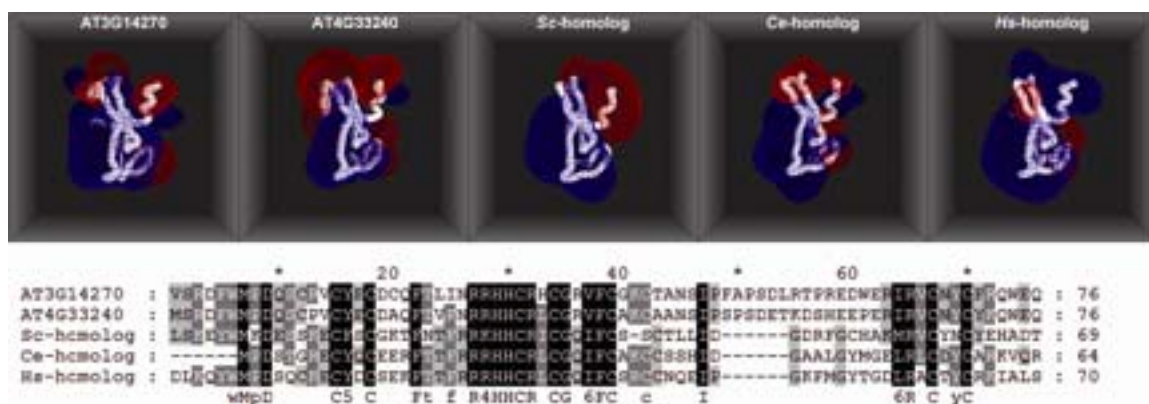


Figure 4.5. The electrostatic profiles and sequence alignment of *Arabidopsis* class I FYVE homologues from *S. cerevisiae* (Fab1p) and *C. elegans* (hypothetical protein ppk-3). Both FYVE domain models are shown in the same orientation with the membrane binding regions facing down. As previously, the red and blue meshes represent the -1 kT/e and +1kT/e equipotential contours of the FYVE domains.

Sequence motifs of the *Arabidopsis* FYVE domains.

Arabidopsis FYVE domains can be divided into two distinctive groups based on different consensus sequences identified via ClustalW multiple sequence alignment (Fig. 4.6). Fig. 4.6A depicts *Arabidopsis* FYVE domains, which belong to class I – IV. These *Arabidopsis* domains were previously referred to as classic FYVE domains because they contain three classic conserved regions: the N-terminal WxxD, the central R(R/K)HHCR and the C-terminal R(V/I)C motifs [128] implicated in binding the PI ligand PtdIns(3)*P*. Class I – IV *Arabidopsis* FYVE proteins have a classic FYVE domain (Fig. 4.6A) with a conserved motif for PtdIns(3)*P* binding (Fig. 4.7A) that is found in *H. sapiens* [366], *S. cerevisiae*, *C. elegans* [374], and various other organisms, e.g. *P. troglodytes*, *M. musculus*, *R. norvegicus*, *C. familiaris*, *B. taurus*, *G. Gallus*. Class V *Arabidopsis* FYVE domains do not share the 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. 6B). Moreover, the central R(R/K)HHCR motif is (K/R)(R/K)HNCY, which is atypical and hence the name variant FYVE domains [128].

We observe that the variable turret loop prior to the R(R/K)HHCR motif associated with membrane penetration of the FYVE domain and the putative dimerization interface region are made up of residues, which are quite diverse in identity in the various FYVE domains. Despite the observed differences in residue identities, however, all classic *Arabidopsis* FYVE domains share at least one hydrophobic residue within the turret loop and highly hydrophobic dimerization interface region. AT1G29800 and AT3G43230 FYVE domains have an insertion of an additional hydrophobic residue within the turret loop and with one exception of AT3G47660 FYVE domain, contain a conserved phenylalanine residue in a second position and conserved arginine in the last position. Similar to class I – IV *Arabidopsis* FYVE domains, the putative dimerization interface region of class V *Arabidopsis* FYVE domains is highly hydrophobic. Unlike the class I – IV *Arabidopsis* FYVE domains, however, class V *Arabidopsis* FYVE domains dimerization interface regions seem quite conserved with at least three absolutely conserved residues, i.e. AxxAP.

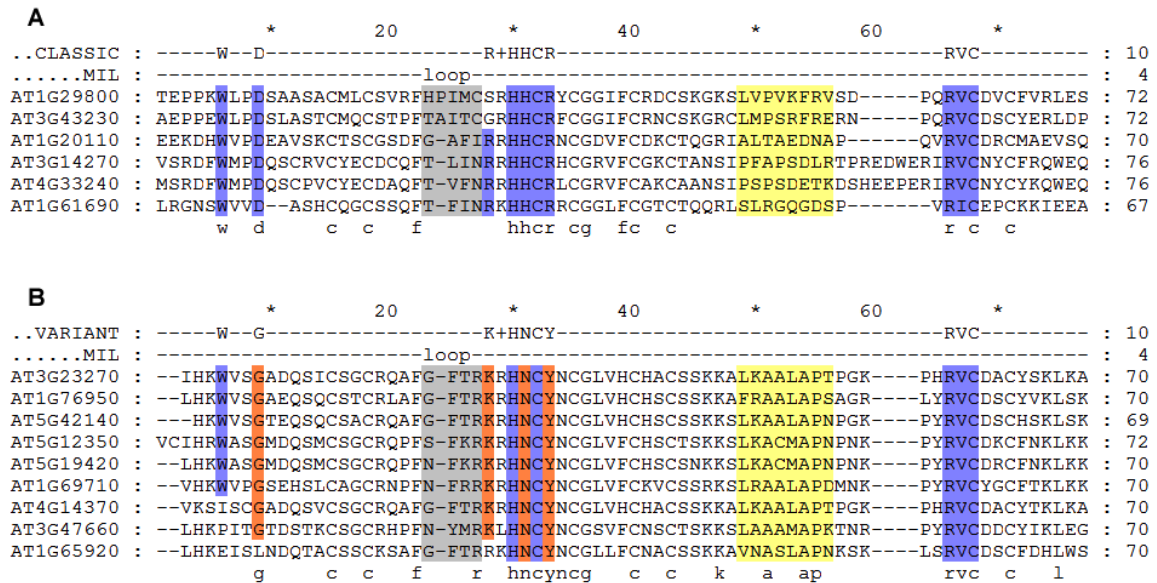


Figure 4.6. A. Alignment of six FYVE domains representing class I – IV of *Arabidopsis* FYVE proteins. The conserved sequence motif of classic *Arabidopsis* FYVE domains is highlighted in blue. B. Alignment of nine FYVE domains representing class V of *Arabidopsis* FYVE proteins. The conserved sequence motif of variant *Arabidopsis* 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.

FYVE domains have the potential to bind PtdIns(3)P and PtdIns(5)P.

Preliminary docking studies represented in Fig. 4.7 and Fig. 4.8 shows that class I – IV *Arabidopsis* FYVE domains have a potential to bind headgroups of PtdIns(3)P and PtdIns(5)P using the same set of residues previously identified to bind Ins(1,3)P₂, i.e. the RHHxR motif and arginine residue of RVC motif (Fig. 4.7). Class V *Arabidopsis* 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. 4.7). 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 is used by the class V *Arabidopsis* FYVE domains to recognize either of the headgroups (Fig. 4.7).

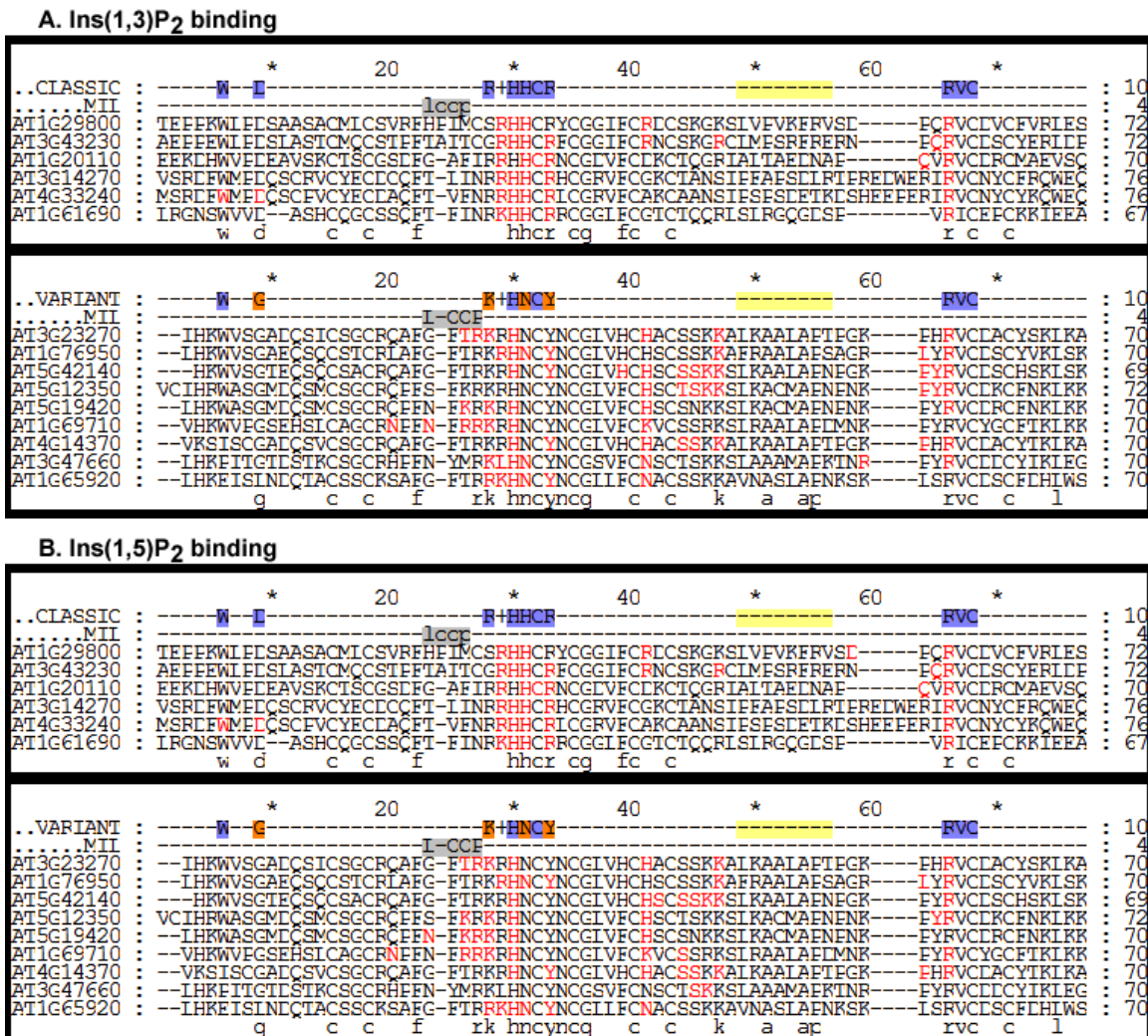


Figure 4.7. Alignment of classic and variant *Arabidopsis* FYVE domains. The conserved sequence motifs of classic and variant *Arabidopsis* FYVE domains are depicted as the top two sequences on each figure. The highlighted motifs in the two top rows correspond to the description found in Figure 4.6. Residues highlighted in red represent the A. hydrogen bound residues to Ins(1,3)P₂ B. hydrogen bound residues to Ins(1,5)P₂.

ORTEP diagram of the molecular structure of 2,3-bis(4-oxo-4H-pyridin-2-yl)-2,3-dihydro-1,4-benzodioxine. The structure shows a central benzodioxine ring system with two pyridine rings attached at the 2 and 3 positions. The pyridine rings are in a trans configuration. The structure is labeled with atom names (C, N, O) and bond lengths (3.00, 3.22, 2.89, 2.93, 2.84 Å). The structure is also labeled with 'Arg 47(A)', 'Arg 61(A)', 'His 30(A)', and 'His 31(A)'.

104

Discussion

Proteins that contain FYVE zinc finger domains have so far been known as effectors of PtdIns(3)*P*, playing a major role in endocytic-trafficking [376]. PtdIns(3)*P* is a PI that is present at very low levels in plant cells (Brearley and Hanke, 1992). It is synthesized by phosphatidylinositol 3-kinase (PI3-kinase), which in turn is encoded by the only PI3-kinase gene reported in plants, vacuolar protein sorting protein 34 (Vps34) [377]. Despite its low levels in plant cells, PtdIns(3)*P* and PI3-kinase are essential for normal plant growth [136] and have been implicated in diverse physiological functions, including root nodule formation [378], auxin-induced production of reactive oxygen species (ROS) and root gravitropism [379], root hair curling and *Rhizobium* infection in *M. truncatula* [380], maintenance of the processes essential for root hair cell elongation [381], increased plasma membrane endocytosis and the intracellular production of ROS in the salt tolerance response [382], and stomatal closing movement [383, 384]. So if we envision plant FYVE domains as being potential effectors of PtdIns(3)*P*, they could be playing important roles in various physiological processes and not just restricted to endocytosis-related pathways. In this study we have modeled the structure of all *Arabidopsis* FYVE domains and predict their membrane targeting behavior based on the biophysical profiles of the modeled structures.

Based on the domain architecture and derived function transferred from homologous proteins of known function, we have classified *Arabidopsis* FYVE proteins into five distinct classes (Fig. 4.2). Similar domain based classifications previously performed for FYVE protein families encoded in *H. sapiens*, *C. elegans*, and *S. cerevisiae* genomes [366, 374] previously suggested a remarkable degree of correspondence among the different FYVE containing protein families in the various organisms [366]. In fact, out of the twenty seven human FYVE proteins, grouped into twelve classes, there are only eight groups, which can be matched with putative homologues in worm proteome and three in yeast proteome [366]. However, *Arabidopsis* is striking in showing no obvious similarities or correspondence with these domain architecture based classifications. More specifically, only one class of *Arabidopsis* FYVE proteins corresponds to what was reported in other organisms, i.e. class I in our classification and the corresponding

PIKfyve, MmPIKfyve and ScFab1p groups [366, 367, 374] in the other classifications. Even that correspondence, however, is partial since the *Arabidopsis* counterparts lack the disheveled, Egl-10 and pleckstrin (DEP) domain observed in mammals and worms [366, 367, 374, 385]. The remaining members of *Arabidopsis* FYVE proteins class II - V are unique and exhibit completely different domains suggesting that FYVE domains in plants play a unique role in plant signaling pathways and/or play similar/parallel roles in signaling as other organisms but use different protein players/signaling mechanisms.

The two class I sequences are potential homologues of the PIKfyve/Fab1 family of phosphatidylinositol phosphate 5-kinases that phosphorylate the D-5 position in phosphatidylinositol (PtdIns) and PtdIns(3)*P* to make PtdIns(5)*P* and PtdIns(3,5)*P*₂, respectively [386]. PIKfyve/Fab1 proteins bind PtdIns(3)*P* with high specificity through their FYVE domains [120, 386] and are known to participate in several aspects of endosomal trafficking functions [385], transduction of osmotic shock signals [387] and other cellular functions [367] in mammals and yeast. The plant members of this family of proteins have never been investigated experimentally and our molecular models provide the initial insight into their mode of function. Both class I members possess the complete classic signature for binding PtdIns(3)*P* and the conserved hydrophobic motif suggesting that they likely bind membranes using the general mechanism of non-specific electrostatic interactions, membrane penetration of hydrophobic residues close to the PtdIns(3)*P* binding pocket facilitated by an electrostatic switch coupled with specific interactions with PtdIns(3)*P* as proposed by previous computational modeling studies [130]. These studies also show that essentially all human FYVE domains have electrostatic equipotential profiles similar to those of the Hrs and EEA1 FYVE domains. This electrostatic polarity seems to be characteristic for the *Arabidopsis* class I FYVE domains and their *S. cerevisiae* and *C. elegans* homologues (Fig. 4.5 and Fig. 4.6). Despite the overall electrostatic profile similarity AT3G14270 FYVE domain has a higher net charge (+7 at pH 6.5; Zn ions included) than AT4G33240 FYVE domain (+3 at pH 6.5; Zn ions included) (Fig 4.3). Based on the net charge difference, we predict that AT4G33240 FYVE domain will have a reduced non-specific electrostatic contribution to membrane targeting. Moreover, we predict that its hydrophobic

contribution will also be reduced because AT3G14270 FYVE domain conserved hydrophobic motif possesses valine versus leucine in AT3G14270 FYVE domain (Fig. 4.5). It is possible that a distinctive contribution such as through dimerization may be important for a functional membrane association of AT3G14270 FYVE domain.

The proteins of class II – IV are untouched sequences in terms of functional assignment and remain annotated as “hypothetical” or “unknown function” in the databases. All of them share the complete/nearly complete conserved PtdIns(3)*P* binding motif and large basic binding pocket except AT1G20110 FYVE domain of class IV which has a significantly reduced basic surface patch in the potential ligand binding pocket (Figure 4.5; class II – IV domains have net charges of +6, +8, +11, and +2, respectively). The class II FYVE domains possess a classic FYVE domain electrostatic profile but their signature is missing the first of the arginines in the R(R/K)HHCR motif, which recognizes the 1-phosphate of the headgroup of PtdIns(3)*P* (Kutateladze, 2006). Even though this residue does not participate in the direct recognition of the 3-phosphate, mutational studies suggest that substitution of this arginine substantially reduces the FYVE domain's affinity for PtdIns(3)*P* containing membranes and potential for membrane localization. On the other hand, the altered signature may slightly reduce the local basic charge in the vicinity of the hydrophobic motif and a lowered barrier to their membrane penetration. Further theoretical calculations and mutational studies could easily verify the subtle differences that we predict in the classic mode of PtdIns(3)*P* specific membrane localization for the class II FYVE domains. The class III FYVE domain on the other hand has the full binding signature and an electrostatic equipotential profile similar to those of the Hrs and EEA1 FYVE domains. We predict that this domain will localize to PtdIns(3)*P* containing membranes using the classic mechanism of action of previous studies FYVE domains with a strong contribution from non-specific electrostatic interactions. Subcellular localization predictions ($p= 0.1422$) suggest that the class III FYVE protein may be targeted to the mitochondria.

The class IV *Arabidopsis* FYVE domain has the most reduced basic surface patch and the lowest net charge of +2. It is possible that class IV member depends mostly on hydrophobic interactions for membrane insertion. A similar suggestion has been made for the FENS FYVE

domain [130], which localizes to endosomal membranes [388] even though it has a weaker positive potential than other FYVE domains [130]. Subcellular localization analyses suggest that class IV *Arabidopsis* FYVE proteins may be targeted to the chloroplast.

The class V FYVE domain proteins are the most intriguing although much remains to be understood about their function. Previously, it has been suggested that association with membrane(s) may be crucial for their functioning given the presence of two PI-binding domains, i.e. PH and FYVE domains [128]. Experimental data suggest that class V AT1G65920 PH domain binds to $\text{PtdIns}(4,5)\text{P}_2$ while its FYVE domain binds to $\text{PtdIns}(3)\text{P}$ as well as $\text{PtdIns}(5)\text{P}$ [128]. The various members of the class V *Arabidopsis* FYVE domains show a high degree of sequence conservation within an enrichment of basic residues throughout the length of the FYVE domain (Fig. 4.6). The most striking feature of these FYVE domains is the presence of a variant PI binding motif (Fig. 4.6B) which seems to be unique to plants as is the overall domain architecture of these proteins [128]; Fig. 4.6B). When a manually derived class V FYVE domains conserved residue motif, (K/R)(R/K)HNCY, is used to search for other FYVE domains, only sequences from plants are retrieved, e.g. Q1SA17 (*M. truncatula*), Q1SIN6 (*M. truncatula*), Q84RS2 (*M. sativa*), Q5JL00 (*O. sativa*) Q5N8I7 (*O. sativa*) Q6AV10 (*O. sativa*) Q6L5B2 (*O. sativa*) Q259N3 (*O. sativa*), Q5XWP1 (*S. tuberosum*), Q60CZ5 (*S. tuberosum*), Q60CZ5 (*S. demissum*) or Q5EWZ4 (*T. turgidum*).

The obvious question that comes to mind is whether this variant signature is responsible for an altered binding specificity in this class of FYVE proteins and therefore associated with a novel pattern of membrane/sub-cellular targeting. Within mammalian cells, FYVE domains are highly conserved and seem to select $\text{PtdIns}(3)\text{P}$ over other PIs [121]. Possibly, there might still be a significant difference in the specificity and affinity of individual FYVE domains towards PIs. In fact, EEA1 has affinity for $\text{PtdIns}(5)\text{P}$ as well, perhaps because $\text{PtdIns}(3)\text{P}$ and $\text{PtdIns}(5)\text{P}$ have the phosphomonoester in the *meta* position [138]. Consequently, $\text{PtdIns}(5)\text{P}$ induces small but important chemical shift changes similar to those induced by $\text{PtdIns}(3)\text{P}$ [138]. The only exception is arginine in the proximal cluster of conserved residues, the resonances of which remain practically unaltered by $\text{PtdIns}(5)\text{P}$ [138]. $\text{PtdIns}(3)\text{P}$ – specific recognition by the FYVE domain

seems to involve indirect recognition of this specific ligand by exclusion of alternatively phosphorylated PIs: the two residues implicated in this are the aspartic acid of the N-terminal WxxD motif and the second histidine of the central HHCR motif [126]. Both of these are substituted in the class V variant *Arabidopsis* FYVE domain (Fig. 4.6) and possess conserved WxxG and HNCY motifs instead. This opens up the possibility that the class V of FYVE domains may have the potential to interact equally or better with PI ligands other than PtdIns(3)*P*. Our preliminary docking analysis of classic as well as variant motif-containing *Arabidopsis* FYVE domains seem to suggest that both have the potential to interact with PtdIns(3)*P* and PtdIns(5)*P* headgroups using practically the same set of residues (Fig. 4.6 and 4.7). However, our analyses reveal an additional ligand-recognizing, highly conserved motif located immediately prior to the dimerization region present only within the class V proteins (Fig. 4.7). Class V *Arabidopsis* FYVE domains also exhibit the greatest net charge observed among FYVE domains classified to date [130, 389] and possess prominent hydrophobic motifs. We predict that class V *Arabidopsis* FYVE domains target to the membrane with significant contributions of non-specific electrostatics and hydrophobic interactions, coupled with specific interactions with Ins(1,3)*P*₂ and/or Ins(1,5)*P*₂ ligands using the classic binding residue and an additional conserved motif specific to this class of FYVE domains.

Based on experimental studies, it has been suggested that the strength of the positive potential and the identity of the hydrophobic residues near the binding site may be two key factors, which are critical in determining which FYVE domains act alone, undergo dimerization or require additional partners before anchoring to the membrane [130]. For example, SARA FYVE domain was predicted and verified experimentally to membrane associate with significant contributions of non-specific electrostatic and hydrophobic interactions given its net charge of +12 (zinc ions included) and the presence of phenylalanine at the conserved hydrophobic position [130, 390, 391]. Our data suggests that *Arabidopsis* FYVE domains engage in both non-specific electrostatic and PtdIns(3)*P*-induced hydrophobic interactions for membrane localization with various degrees of contribution as described earlier. Dimerization also plays an important role in the membrane recruitment of FYVE domains [125, 392] and it appears that the electrostatic free

energy contributions to the membrane associations are additive for each part of the EEA1 FYVE dimer [130]. The dimer interface regions of all *Arabidopsis* FYVE domains are longer and more hydrophobic (Fig. 4.6) than the equivalent region of EEA1-FYVE and predicted region of SARA FYVE domain [389, 393] suggesting that all of the *Arabidopsis* FYVE domains could potentially interact with the membrane as dimers. Further theoretical calculations will provide better quantitative picture of the energetics of such an association relative to a monomeric interaction with the membrane.

Overall, *Arabidopsis* FYVE proteins seem to be quite distinct from FYVE proteins of other organisms exhibiting unique domain architectures. The biophysical profiles of the modeled FYVE domains in *Arabidopsis* suggest various membrane-targeting mechanisms. Our predictions provide a foundation for designing directed mutational studies to confirm these behaviors; this is crucial to understand the role of these domains in important plant signaling pathways, something that has so far not been explored.

Chapter 5 Conclusions and future directions

Given data presented here, it is obvious that *Arabidopsis* PH and FYVE domains are hardly related to their animal, worm and yeast counterparts. 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. Moreover, various combinations of domain architectures observed among *Arabidopsis* PH and FYVE proteins as well as FYVE domains' blends of conserved and variable regions are all expected to be key structural membrane association determinants. Consequently, the *Arabidopsis* PH and FYVE domains are highly likely to function as regulators of lipid signaling through a variety of ways including head-on membrane association, aggregation, dimerization and/or other protein involvement.

Clearly, the physiological role of each *Arabidopsis* PH and FYVE domain must be experimentally determined within its protein to leave no doubts as to their function in plant cells. However, 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. For example, sequence analysis of the *Arabidopsis* FYVE domains of the class V identified differences in β -propeller organization that could provide direction for experimental study of protein-binding characteristics either via mutational studies or molecular dynamics simulations. In addition, based on our domain-based classification experimental researchers can chose two PH or FYVE proteins of the same class for swapping experiments to study protein localization and targeting. Our results also provide groundwork for extensive comparative studies among various species.

Appendix A

Table 0.1 Individual members of the domain-architecture based twelve classes of *Arabidopsis* PH proteins. The numbers in parentheses represent length of each protein in amino acids.

Class	Members
OSBP	AT1G13170(816), AT2G31020(760), AT4G08180(814), AT4G12460(694), AT4G22540(721), AT1G77730(265);
START	AT2G28320(737), AT3G54800(733), AT4G19040(718), AT5G35180(778), AT5G45560(719);
PLD□	AT3G05630(1046), AT3G16785(1096);
PRAF	AT1G65920(1006), AT1G69710(1028), AT1G76950(1103), AT3G23270(1045), AT3G47660(951), AT4G14370(1045), AT5G12350(1069), AT5G19420(1124), AT5G42140(1073);
PDK	AT3G10540(486), AT5G04510(491);
ArfGAP	AT1G10870(775), AT1G60860(776), AT5G13300(768); AT5G61980(850),
RhoGAP	AT4G24580(902), AT5G12150(827), AT5G19390(870);
DYN	AT1G10290(914), AT1G59610(920);
AtPH1	AT2G29700(145), AT5G05710(144);
DUF828	AT3G22810(472), AT3G63300(498), AT4G14740(475), AT4G16670(429), AT4G17350(405), AT4G32785(124), AT5G43870(453), AT5G47440(406), AT5G57770(409);
UNK1	AT4G17140(1322);
UNK2	AT1G17820(803), AT1G73200(779); AT2G30880(504); AT4G23895(247);

Table 0.2. *Arabidopsis* PH domains: accession numbers, techniques and templates used to build their structural models.

<i>Arabidopsis</i> PH domain class	Accession number	Technique	PH domain template	Seq. id (%) / seq. similarity (%)
OSBP	AT1G13170	Pipeline	PKB	15% / 30%
	AT1G77730	Pipeline	PKB	15% / 34%
	AT2G31020	Fold recognition	TAPP1	12% / 29%
	AT4G08180	Pipeline	PKB	18% / 34%
	AT4G12460	Homology	Sbf1	18% / 28%
	AT4G22540	Fold recognition	DAPP1/PHISH	13% / 24%
START	AT2G28320	Fold recognition	DAPP1/PHISH	16% / 29%
	AT3G54800	Fold recognition	Sbf1	17% / 30%
	AT4G19040	Fold recognition	PKB	19% / 38%
	AT5G35180	Fold recognition	Pleckstrin	10% / 21%
	AT5G45560	Fold recognition	PKB	19% / 36%
PLD□	AT3G05630	Fold recognition	GRP1	12% / 27%
	AT3G16785	Fold recognition	SKAP-Hom	10% / 15%
PRAF	AT1G65920	Fold recognition	PLCd1	27% / 49%
	AT1G69710	Pipeline	PLCd1	25% / 46%
	AT1G76950	Pipeline	PLCd1	24% / 50%
	AT3G23270	Fold recognition	PLCd1	27% / 48%
	AT3G47660	Homology	PLCd1	18% / 44%
	AT4G14370	Homology	PLCd1	15% / 26%
	AT5G12350	Pipeline	PLCd1	22% / 46%
	AT5G19420	Pipeline	PLCd1	24% / 46%
	AT5G42140	Fold recognition	PLCd1	24% / 50%
PDK	AT3G10540	Homology	PDK1	16% / 30%
	AT5G04510	Homology	PDK1	17% / 30%
ArfGAP	AT1G10870	Pipeline	Pleckstrin	21% / 32%
	AT1G60860	Pipeline	SKAP55	23% / 42%
	AT5G13300	Pipeline	PKB	13% / 31%
	AT5G61980	Fold recognition	GRP1	13% / 26%
RhoGAP	AT4G24580	Fold recognition	PKB	15% / 35%
	AT5G12150	Fold recognition	Sbf1	17% / 39%
	AT5G19390	Fold recognition	SKAP-Hom	21% / 41%
DYN	AT1G10290	Pipeline	Pleckstrin	24% / 37%
	AT1G59610	Pipeline	Pleckstrin	24% / 37%
AtPH1	AT2G29700	Fold recognition	DAPP1/PHISH	30% / 46%
	AT5G05710	Fold recognition	PKB	26% / 40%
DUF828	AT3G22810	Fold recognition	□-SPECTRIN	13% / 20%
	AT3G63300	Fold recognition	SCOP b.55.1.1	10% / 27%
	AT4G14740	Fold recognition	Pleckstrin	10% / 29%
	AT4G16670	De novo	De novo	n/a
	AT4G17350	Fold recognition	Exo84	9% / 31%
	AT4G32785	Fold recognition	Pleckstrin	14% / 29%
	AT5G43870	Fold recognition	SKAP55	19% / 34%
	AT5G47440	Fold recognition	DAPP1/PHISH	9% / 29%
	AT5G57770	De novo	De novo	n/a
UNK1	AT4G17140	Fold recognition	PEPP1	13% / 24%
UNK2	AT1G17820	Fold recognition	KIAA1914	12 % 27%
	AT1G73200	De novo	De novo	n/a
	AT4G23895	Fold recognition	SKAP55	12% / 27%
	AT2G30880	Fold recognition	β-SPECTRIN	7% / 25%

Table 0.3. Functional annotation of *Arabidopsis* PH domains in TAIR database. EC 2.1.-.-: Transferases - Transferring One-Carbon Groups; EC 2.3.-.-: Transferases - Acyltransferases; EC 2.4.-.-: Transferases - Glycosyltransferases; EC 2.7.-.-: Transferases - Transferring Phosphorus-Containing Groups; EC 3.1.-.-: Hydrolases - Acting on Ester Bonds; EC 3.2.-.-: Hydrolases - Glycosylases; EC 3.4.-.-: Hydrolases - Acting on peptide bonds (Peptidases); EC 4.1.-.-: Lyases - Carbon-Carbon Lyases; EC 5.3.-.-: Isomerases - Intramolecular Oxidoreductases; TC 1.C. Channels/Pores - Pore-forming toxins (proteins and peptides); TC 3.A.1 ATP-binding cassette (ABC) family; TC 3.A.3 P-type ATPase (P-ATPase) family. The numerical values represent the probability of correct classification (P-value (%)). Only the highest probabilities, if any, are reported.

Class	Accession	GO			SVMProt Predicted functional class	Probability (%)
		Biological Process	Cellular Component	Molecular Function		
ArfGAP	AT1G10870	regulation of ARF GTPase activity	cytoplasm	protein binding, zinc ion binding		
ArfGAP	AT1G60860	regulation of ARF GTPase activity	cytoplasm	protein binding, zinc ion binding	EC 2.7	99
ArfGAP	AT5G13300	vesicle transport [252]				
ArfGAP	AT5G61980	cytoskeleton organization and biogenesis	endocytic vesicle	protein binding, zinc ion binding		
AtPH1	AT2G29700	not annotated	unknown	unknown		
AtPH1	AT5G05710	not annotated	unknown	unknown	EC 2.7	98
DUF828	AT3G22810	not annotated	unknown	not annotated	EC 3.4	84
DUF828	AT3G63300	unknown	not annotated	unknown	EC 3.1	84
DUF828	AT4G14740	not annotated	not annotated	not annotated		
DUF828	AT4G16670	unknown	unknown	unknown	EC 2.7	95
DUF828	AT4G17350	unknown	unknown	unknown	TC 1.C	59
DUF828	AT4G32785	unknown	unknown	unknown	EC 2.3	82
DUF828	AT5G43870	unknown	unknown	not annotated		
DUF828	AT5G47440	unknown	unknown	unknown		
DUF828	AT5G57770	unknown	unknown	unknown	EC 3.1	99
DYN	AT1G10290	Golgi to vacuole transport [253]				
DYN	AT1G59610	not annotated	unknown	GTPase activity	EC 3.4	80
OSBP	AT1G13170	steroid metabolic process	not annotated	oxysterol binding	EC 3.1/ EC 2.4/ EC 2.7	99 / 98 / 98
OSBP	AT1G77730	steroid metabolic process	chloroplast	unknown	EC 3.1	99
OSBP	AT2G31020	steroid metabolic process	not annotated	oxysterol binding		
OSBP	AT4G08180	steroid metabolic process	unknown	oxysterol binding	EC 3.4	94

Table 0.3. continued

OSBP	AT4G12460	steroid metabolic process	unknown	oxysterol binding	EC 2.3	97
OSBP	AT4G22540	steroid metabolic process	unknown	oxysterol binding	EC 3.1	99
PDK	AT3G10540	protein amino acid phosphorylation	unknown	kinase activity	EC 2.7	69
PDK	AT5G04510	positive regulation of protein kinase activity	unknown	protein binding, PI binding		
PLD ζ	AT3G05630	root development in response to nutrient limitation [287], supply inorganic phosphorus during phosphorus starvation via hydrolysis of phosphatidylcholine [288] and vesicle trafficking and auxin responses [290].				
PLD ζ	AT3G16785	root development in response to nutrient limitation [287], supply inorganic phosphorus during phosphorus starvation via hydrolysis of phosphatidylcholine [288]				
PRAF	AT1G65920	not annotated	not annotated	zinc ion binding		
PRAF	AT1G69710	not annotated	unknown	Ran GTPase, chromatin and zinc ion binding	EC 2.7	98
PRAF	AT1G76950	unknown	unknown	Ran GTPase, chromatin and zinc ion binding	EC 4.1	74
PRAF	AT3G23270	unknown	unknown	Ran GTPase and chromatin binding	EC 3.1	74
PRAF	AT3G47660	not annotated	not annotated	Ran GTPase and chromatin binding	EC 3.1	80
PRAF	AT4G14370*	defense response	intrinsic to membrane	ATP, nucleotide, protein and zinc ion binding	EC 5.3/ TC 1.C	59 / 59
PRAF	AT5G12350	not annotated	not annotated	Ran GTPase, chromatin and zinc ion binding	EC 2.1	59
PRAF	AT5G19420	not annotated	not annotated	Ran GTPase, chromatin and zinc ion binding		

Table 0.3. continued

PRAF	AT5G42140	not annotated	not annotated	Ran GTPase, chromatin and zinc ion binding	EC 3.4	69
RhoGAP	AT4G24580	signal transduction	intracellular	Rho GTPase activator activity	EC 3.1	69
RhoGAP	AT5G12150	signal transduction	intracellular	unknown	EC 2.7	76
RhoGAP	AT5G19390	signal transduction	chloroplast thylakoid membrane	not annotated	EC 3.4	89
START	AT2G28320	unknown	unknown	unknown		
START	AT3G54800	unknown	unknown	unknown		
START	AT4G19040	regulator of plant defense responses and cell death functioning inside mitochondria [279]				
START	AT5G35180	unknown	unknown	unknown	EC 2.4	80
START	AT5G45560	not annotated	mitochondrion	lipid binding	EC 2.7	69
UNK1	AT4G17140	unknown	unknown	unknown	EC 2.7	96
UNK2	AT1G17820	not annotated	endomembrane system	not annotated	EC 3.1	84
UNK2	AT1G73200	unknown	endomembrane system	unknown	EC 2.4	90
UNK2	AT4G23895	unknown	unknown	unknown		
UNK2	AT2G30880	unknown	chloroplast	not annotated	EC 3.1	76

*missannotated [250]

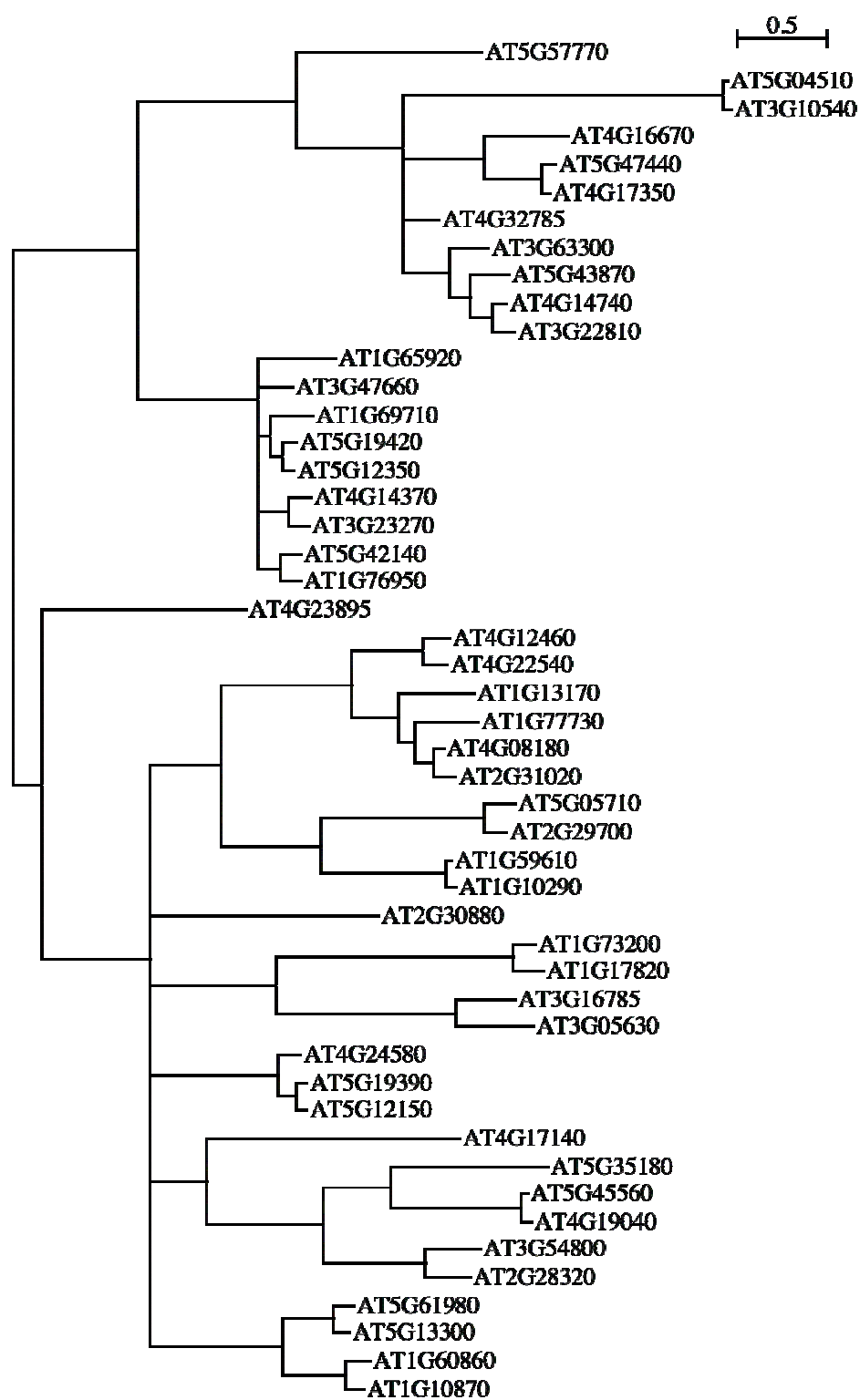


Figure 0.1. Protein tree of *Arabidopsis* PH domains (unresolved).

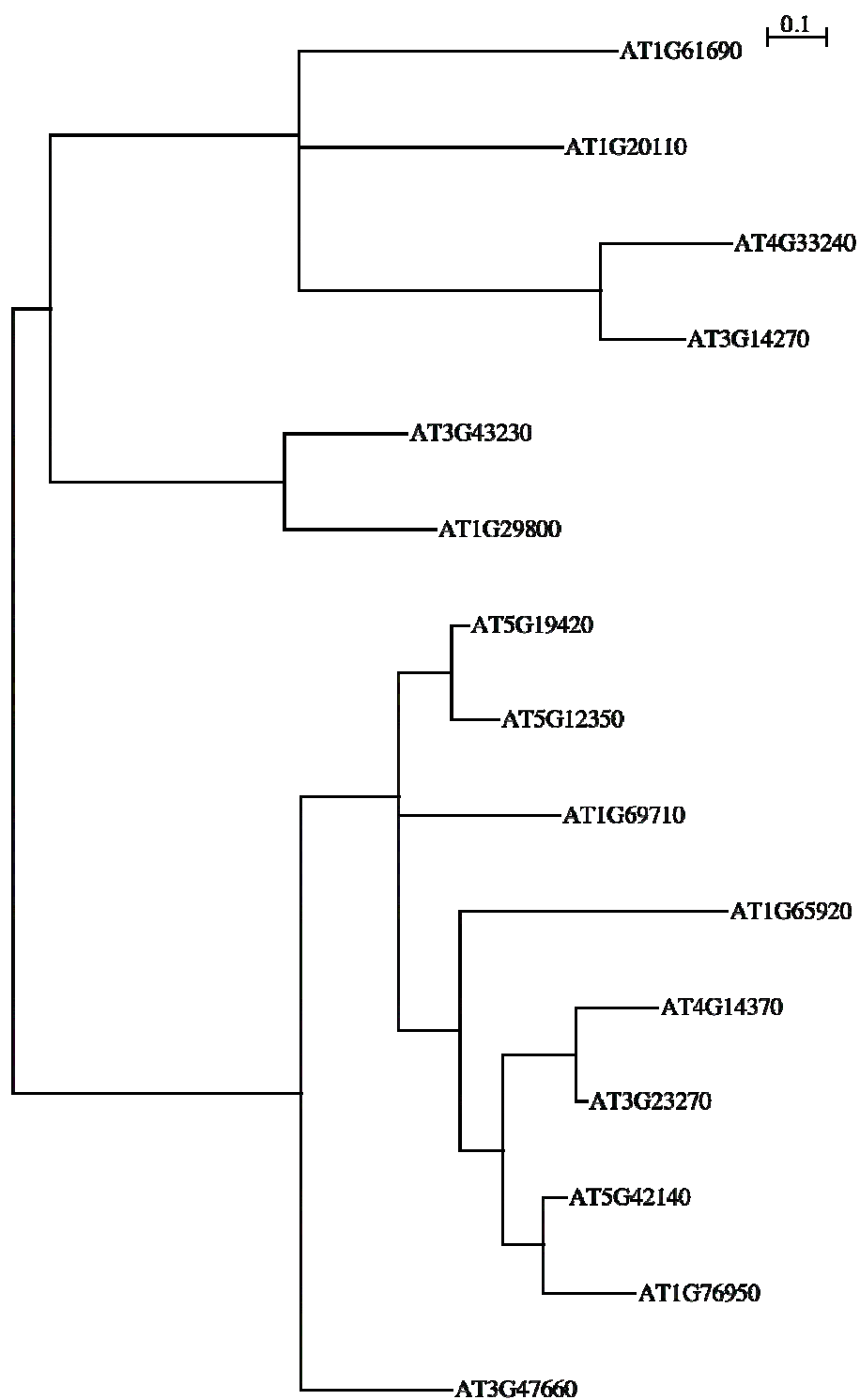


Figure 0.2. Protein tree of *Arabidopsis* FYVE domains (unresolved).

References

1. Meinke DW, Cherry JM, Dean C, Rounsley SD, Koornneef M: **Arabidopsis thaliana: a model plant for genome analysis.** *Science* 1998, **282**:662, 679-682.
2. Laibach F: **Arabidopsis thaliana (L.) Heynh. Als Object Fur Genetische Und Entwicklungsphysiologische Untersuchungen.** *Bot Archiv* 1943:439-455.
3. Somerville C, Koornneef M: **A fortunate choice: the history of Arabidopsis as a model plant.** *Nat Rev Genet* 2002, **3**:883-889.
4. Bechtold N, Ellis J, Pelletier G: **In-planta Agrobacterium-mediated gene transfer by infiltration of adult Arabidopsis thaliana plants.** *C R Acad Sci Sér* 1993, **III** 1194-1199.
5. Hülskamp M, Folkers U, Grini P: **Cell morphogenesis in Arabidopsis.** *BioEssays* 1998, **20**:20-29.
6. Haughn G, Chaudhury A: **Genetic analysis of seed coat development in Arabidopsis.** *Trends in Plant Science* 2005, **10**:472.
7. Hardtke CS: **Root development -- branching into novel spheres.** *Current Opinion in Plant Biology* 2006, **9**:66.
8. Komeda Y: **Genetic regulation of time to flower in Arabidopsis thaliana.** *Annu Rev Plant Biol* 2004, **55**:521-535.
9. Koornneef M, Alonso-Blanco C, Peeters AJ, Soppe W: **Genetic Control of Flowering Time in Arabidopsis.** *Annu Rev Plant Physiol Plant Mol Biol* 1998, **49**:345-370.
10. Robles P, Pelaz S: **Flower and fruit development in Arabidopsis thaliana.** *Int J Dev Biol* 2005, **49**:633-643.
11. van Doorn WG, van Meeteren U: **Flower opening and closure: a review.** *J Exp Bot* 2003, **54**:1801-1812.
12. Fankhauser C: **Light perception in plants: cytokinins and red light join forces to keep phytochrome B active.** *Trends in Plant Science* 2002, **7**:143.
13. Sullivan JA, Deng XW: **From seed to seed: the role of photoreceptors in Arabidopsis development.** *Developmental Biology* 2003, **260**:289.
14. Parker JE, Aarts N, Austin MA, Feys BJ, Moisan LJ, Muskett P, Rusterucci C: **Genetic analysis of plant disease resistance pathways.** *Novartis Found Symp* 2001, **236**:153-161; discussion 161-154.
15. DellaPenna D, Pogson BJ: **Vitamin synthesis in plants: tocopherols and carotenoids.** *Annu Rev Plant Biol* 2006, **57**:711-738.
16. Farmer EE, Weber H, Vollenweider S: **Fatty acid signaling in Arabidopsis.** *Planta* 1998, **206**:167.
17. Davis KR: **Arabidopsis thaliana. A robust model system for studying plant-pathogen interactions.** *Subcell Biochem* 1998, **29**:253-285.

18. Gough NR: **Focus Issue: Plant Communication**. *Sci STKE* 2006, **2006**:eg3-.
19. De Paepe A, Van Der Straeten D, Gerald L: **Ethylene Biosynthesis and Signaling: An Overview**. In *Vitamins & Hormones*. Volume Volume 72: Academic Press; 2005: 399
20. Vert G, Nemhauser JL, Geldner N, Hong F, Chory J: **Molecular mechanisms of steroid hormone signaling in plants**. *Annu Rev Cell Dev Biol* 2005, **21**:177-201.
21. AGI: **Analysis of the genome sequence of the flowering plant Arabidopsis thaliana**. *Nature* 2000, **408**:796-815.
22. Yamada K, Lim J, Dale JM, Chen H, Shinn P, Palm CJ, Southwick AM, Wu HC, Kim C, Nguyen M, et al: **Empirical analysis of transcriptional activity in the Arabidopsis genome**. *Science* 2003, **302**:842-846.
23. Bevan M, Mayer K, White O, Eisen JA, Preuss D, Bureau T, Salzberg SL, Mewes HW: **Sequence and analysis of the Arabidopsis genome**. *Curr Opin Plant Biol* 2001, **4**:105-110.
24. Chory J, Ecker JR, Briggs S, Caboche M, Coruzzi GM, Cook D, Dangl J, Grant S, Guerinot ML, Henikoff S, et al: **National Science Foundation-Sponsored Workshop Report: "The 2010 Project" functional genomics and the virtual plant. A blueprint for understanding how plants are built and how to improve them**. *Plant Physiol* 2000, **123**:423-426.
25. Schoof H, Karlowski WM: **Comparison of rice and Arabidopsis annotation**. *Curr Opin Plant Biol* 2003, **6**:106-112.
26. Garcia-Hernandez M, Berardini TZ, Chen G, Crist D, Doyle A, Huala E, Knee E, Lambrecht M, Miller N, Mueller LA, et al: **TAIR: a resource for integrated Arabidopsis data**. *Funct Integr Genomics* 2002, **2**:239-253.
27. Wigge PA, Weigel D: **Arabidopsis genome: life without notch**. *Curr Biol* 2001, **11**:R112-114.
28. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, et al: **Gene ontology: tool for the unification of biology. The Gene Ontology Consortium**. *Nat Genet* 2000, **25**:25-29.
29. GO: **Creating the gene ontology resource: design and implementation**. *Genome Res* 2001, **11**:1425-1433.
30. Frishman D, Albermann K, Hani J, Heumann K, Metanowski A, Zollner A, Mewes HW: **Functional and structural genomics using PEDANT**. *Bioinformatics* 2001, **17**:44-57.
31. Frishman D, Mokrejs M, Kosykh D, Kastenmuller G, Kolesov G, Zubrzycki I, Gruber C, Geier B, Kaps A, Albermann K, et al: **The PEDANT genome database**. *Nucleic Acids Res* 2003, **31**:207-211.
32. Riley ML, Schmidt T, Wagner C, Mewes HW, Frishman D: **The PEDANT genome database in 2005**. *Nucleic Acids Res* 2005, **33**:D308-310.
33. Takenawa T, Itoh T: **Membrane targeting and remodeling through phosphoinositide-binding domains**. *IUBMB Life* 2006, **58**:296-303.

34. Sasaki T, Sasaki J, Sakai T, Takasuga S, Suzuki A: **The physiology of phosphoinositides.** *Biol Pharm Bull* 2007, **30**:1599-1604.
35. Gonzales ML, Anderson RA: **Nuclear phosphoinositide kinases and inositol phospholipids.** *J Cell Biochem* 2006, **97**:252-260.
36. Stephens L, McGregor A, Hawkins P: **Phosphoinositide 3-kinases: regulation by cell-surface receptors and function of 3-phosphorylated lipids.** in *Biology of Phosphoinositides* (ed Cockcroft, S) 2000.
37. Michell RH, Heath VL, Lemmon MA, Dove SK: **Phosphatidylinositol 3,5-bisphosphate: metabolism and cellular functions.** *Trends Biochem Sci* 2006, **31**:52-63.
38. Munnik T, Irvine RF, Musgrave A: **Phospholipid signalling in plants.** *Biochim Biophys Acta* 1998, **1389**:222-272.
39. Meijer HJ, Munnik T: **Phospholipid-based signaling in plants.** *Annu Rev Plant Biol* 2003, **54**:265-306.
40. Meijer HJ, Berrie CP, Iurisci C, Divecha N, Musgrave A, Munnik T: **Identification of a new polyphosphoinositide in plants, phosphatidylinositol 5-monophosphate (PtdIns5P), and its accumulation upon osmotic stress.** *Biochem J* 2001, **360**:491-498.
41. Pettitt TR, Dove SK, Lubben A, Calaminus SD, Wakelam MJ: **Analysis of intact phosphoinositides in biological samples.** *J Lipid Res* 2006, **47**:1588-1596.
42. Lemmon MA: **Membrane recognition by phospholipid-binding domains.** *Nat Rev Mol Cell Biol* 2008, **9**:99-111.
43. Krauss M, Haucke V: **Phosphoinositides: regulators of membrane traffic and protein function.** *FEBS Lett* 2007, **581**:2105-2111.
44. Behnia R, Munro S: **Organelle identity and the signposts for membrane traffic.** *Nature* 2005, **438**:597-604.
45. De Matteis MA, Godi A: **PI-loting membrane traffic.** *Nat Cell Biol* 2004, **6**:487-492.
46. De Matteis MA, Godi A: **Protein-lipid interactions in membrane trafficking at the Golgi complex.** *Biochim Biophys Acta* 2004, **1666**:264-274.
47. Di Paolo G, De Camilli P: **Phosphoinositides in cell regulation and membrane dynamics.** *Nature* 2006, **443**:651-657.
48. Lemmon MA, Ferguson KM: **Molecular determinants in pleckstrin homology domains that allow specific recognition of phosphoinositides.** *Biochem Soc Trans* 2001, **29**:377-384.
49. Psachoulia E, Sansom MS: **Interactions of the pleckstrin homology domain with phosphatidylinositol phosphate and membranes: characterization via molecular dynamics simulations.** *Biochemistry* 2008, **47**:4211-4220.
50. He J, Haney RM, Vora M, Verkhusha VV, Stahelin RV, Kutateladze TG: **Molecular mechanism of membrane targeting by the GRP1 PH domain boxes.** *J Lipid Res* 2008, **49**:1807-1815.

51. Garcia P, Gupta R, Shah S, Morris AJ, Rudge SA, Scarlata S, Petrova V, McLaughlin S, Rebecchi MJ: **The pleckstrin homology domain of phospholipase C-delta 1 binds with high affinity to phosphatidylinositol 4,5-bisphosphate in bilayer membranes.** *Biochemistry* 1995, **34**:16228-16234.
52. Lemmon MA, Ferguson KM, O'Brien R, Sigler PB, Schlessinger J: **Specific and high-affinity binding of inositol phosphates to an isolated pleckstrin homology domain.** *Proc Natl Acad Sci U S A* 1995, **92**:10472-10476.
53. Harlan J, Hajduk P, Yoon H, Fesik S: **Pleckstrin homology domains bind to phosphatidylinositol-4,5-bisphosphate.** *Nature* 1994, **371**:169-170.
54. Kavran JM, Klein DE, Lee A, Falasca M, Isakoff SJ, Skolnik EY, Lemmon MA: **Specificity and promiscuity in phosphoinositide binding by pleckstrin homology domains.** *J Biol Chem* 1998, **273**:30497-30508.
55. Rameh LE, Arvidsson A-k, Carraway Iii KL, Couvillon AD, Rathbun G, Crompton A, VanRenterghem B, Czech MP, Ravichandran KS, Burakoff SJ, et al: **A Comparative Analysis of the Phosphoinositide Binding Specificity of Pleckstrin Homology Domains.** *J Biol Chem* 1997, **272**:22059-22066.
56. Klarlund JK, Guilherme A, Holik JJ, Virbasius JV, Chawla A, Czech MP: **Signaling by phosphoinositide-3,4,5-trisphosphate through proteins containing pleckstrin and Sec7 homology domains.** *Science* 1997, **275**:1927-1930.
57. Salim K, Bottomley MJ, Querfurth E, Zvelebil MJ, Gout I, Scaife R, Margolis RL, Gigg R, Smith CI, Driscoll PC, et al: **Distinct specificity in the recognition of phosphoinositides by the pleckstrin homology domains of dynamin and Bruton's tyrosine kinase.** *Embo J* 1996, **15**:6241-6250.
58. Frech M, Andjelkovic M, Ingley E, Reddy KK, Falck JR, Hemmings BA: **High affinity binding of inositol phosphates and phosphoinositides to the pleckstrin homology domain of RAC/protein kinase B and their influence on kinase activity.** *J Biol Chem* 1997, **272**:8474-8481.
59. James SR, Downes CP, Gigg R, Grove SJ, Holmes AB, Alessi DR: **Specific binding of the Akt-1 protein kinase to phosphatidylinositol 3,4,5-trisphosphate without subsequent activation.** *Biochem J* 1996, **315** (Pt 3):709-713.
60. Dowler S, Currie RA, Downes CP, Alessi DR: **DAPP1: a dual adaptor for phosphotyrosine and 3-phosphoinositides.** *Biochem J* 1999, **342** (Pt 1):7-12.
61. Dowler S, Currie RA, Campbell DG, Deak M, Kular G, Downes CP, Alessi DR: **Identification of pleckstrin-homology-domain-containing proteins with novel phosphoinositide-binding specificities.** *Biochem J* 2000, **351**:19-31.
62. **International Human Genome Sequencing Consortium.** *Nature (London)* 2001, **409**:860-921.
63. Haslam RJ, Koide HB, Hemmings BA: **Pleckstrin domain homology.** *Nature* 1993, **363**:309.
64. Mayer BJ, Ren R, Clark KL, Baltimore D: **A putative modular domain present in diverse signaling proteins.** *Cell* 1993, **73**:629.

65. Musacchio A, Gibson T, Rice P, Thompson J, Saraste M: **The PH domain: a common piece in the structural patchwork of signalling proteins.** *Trends Biochem Sci* 1993, **18**:343-348.
66. Tyers M, Rachubinski RA, Stewart MI, Varrichio AM, Shorr RG, Haslam RJ, Harley CB: **Molecular cloning and expression of the major protein kinase C substrate of platelets.** *Nature* 1988, **333**:470-473.
67. Gibson T, Hyvonen M, Musacchio A, Saraste M, Birney E: **PH domain: the first anniversary.** *Trends Biochem Sci* 1994, **19**:349-353.
68. Lemmon MA, Ferguson KM: **Signal-dependent membrane targeting by pleckstrin homology (PH) domains.** *Biochem J* 2000, **350**:1-18.
69. Kojima T, Fukuda M, Watanabe Y, Hamazato F, Mikoshiba K: **Characterization of the pleckstrin homology domain of Btk as an inositol polyphosphate and phosphoinositide binding domain.** *Biochem Biophys Res Commun* 1997, **236**:333-339.
70. Milburn CC, Deak M, Kelly SM, Price NC, Alessi DR, Van Aalten DM: **Binding of phosphatidylinositol 3,4,5-trisphosphate to the pleckstrin homology domain of protein kinase B induces a conformational change.** *Biochem J* 2003, **375**:531-538.
71. Thomas CC, Dowler S, Deak M, Alessi DR, van Aalten DM: **Crystal structure of the phosphatidylinositol 3,4-bisphosphate-binding pleckstrin homology (PH) domain of tandem PH-domain-containing protein 1 (TAPP1): molecular basis of lipid specificity.** *Biochem J* 2001, **358**:287-294.
72. Baraldi E, Carugo KD, Hyvonen M, Surdo PL, Riley AM, Potter BV, O'Brien R, Ladbury JE, Saraste M: **Structure of the PH domain from Bruton's tyrosine kinase in complex with inositol 1,3,4,5-tetrakisphosphate.** *Structure* 1999, **7**:449-460.
73. Hirata M, Kanematsu T, Takeuchi H, Yagisawa H: **Pleckstrin homology domain as an inositol compound binding module.** *Jpn J Pharmacol* 1998, **76**:255-263.
74. Hyvonen M, Macias MJ, Nilges M, Oschkinat H, Saraste M, Wilmanns M: **Structure of the binding site for inositol phosphates in a PH domain.** *Embo J* 1995, **14**:4676-4685.
75. Takeuchi H, Kanematsu T, Misumi Y, Yaakob HB, Yagisawa H, Ikehara Y, Watanabe Y, Tan Z, Shears SB, Hirata M: **Localization of a high-affinity inositol 1,4,5-trisphosphate/inositol 1,4,5,6-tetrakisphosphate binding domain to the pleckstrin homology module of a new 130 kDa protein: characterization of the determinants of structural specificity.** *Biochem J* 1996, **318** (Pt 2):561-568.
76. Inglese J, Koch WJ, Touhara K, Lefkowitz RJ: **G beta gamma interactions with PH domains and Ras-MAPK signaling pathways.** *Trends Biochem Sci* 1995, **20**:151-156.
77. Touhara K, Inglese J, Pitcher JA, Shaw G, Lefkowitz RJ: **Binding of G protein beta gamma-subunits to pleckstrin homology domains.** *J Biol Chem* 1994, **269**:10217-10220.
78. Yao L, Kawakami Y, Kawakami T: **The pleckstrin homology domain of Bruton tyrosine kinase interacts with protein kinase C.** *Proc Natl Acad Sci U S A* 1994, **91**:9175-9179.

79. Yao L, Janmey P, Frigeri LG, Han W, Fujita J, Kawakami Y, Apgar JR, Kawakami T: **Pleckstrin homology domains interact with filamentous actin.** *J Biol Chem* 1999, **274**:19752-19761.
80. Wang DS, Shaw R, Winkelmann JC, Shaw G: **Binding of PH domains of beta-adrenergic receptor kinase and beta-spectrin to WD40/beta-transducin repeat containing regions of the beta-subunit of trimeric G-proteins.** *Biochem Biophys Res Commun* 1994, **203**:29-35.
81. Yoon HS, Hajduk PJ, Petros AM, Olejniczak ET, Meadows RP, Fesik SW: **Solution structure of a pleckstrin-homology domain.** *Nature* 1994, **369**:672-675.
82. Fushman D, Najmabadi-Haske T, Cahill S, Zheng J, LeVine H, 3rd, Cowburn D: **The solution structure and dynamics of the pleckstrin homology domain of G protein-coupled receptor kinase 2 (beta-adrenergic receptor kinase 1). A binding partner of Gbetagamma subunits.** *J Biol Chem* 1998, **273**:2835-2843.
83. Blomberg N, Baraldi E, Sattler M, Saraste M, Nilges M: **Structure of a PH domain from the C. elegans muscle protein UNC-89 suggests a novel function.** *Structure Fold Des* 2000 **8**:1079-1087.
84. Macias M, Musacchio A, Ponstingl H, Nilges M, Saraste M, H. O: **Structure of the pleckstrin homology domain from beta-spectrin.** *Nature* 1994, **369**:675-677.
85. Blomberg N, Baraldi E, Nilges M, Saraste M: **The PH superfold: a structural scaffold for multiple functions.** *Trends Biochem Sci* 1999, **24**:441-445.
86. Ferguson KM, Lemmon MA, Schlessinger J, Sigler PB: **Crystal structure at 2.2 Å resolution of the pleckstrin homology domain from human dynamin.** *Cell* 1994, **79**:199-209.
87. Koshiba S, Kigawa T, Kim J, Shirouzu M, Bowtell D, Yokoyama S: **The solution structure of the pleckstrin homology domain of mouse Son-of-sevenless 1 (mSos1).** *J Mol Biol* 1997, **269**:579-591.
88. Zheng J, Cahill SM, Lemmon MA, Fushman D, Schlessinger J, Cowburn D: **Identification of the binding site for acidic phospholipids on the pH domain of dynamin: implications for stimulation of GTPase activity.** *J Mol Biol* 1996, **255**:14-21.
89. Zheng J, Chen RH, Corblan-Garcia S, Cahill SM, Bar-Sagi D, Cowburn D: **The solution structure of the pleckstrin homology domain of human SOS1. A possible structural role for the sequential association of diffuse B cell lymphoma and pleckstrin homology domains.** *J Biol Chem* 1997, **272**:30340-30344.
90. Ferguson KM, Lemmon MA, Schlessinger J, Sigler PB: **Structure of the high affinity complex of inositol trisphosphate with a phospholipase C pleckstrin homology domain.** *Cell* 1995, **83**:1037-1046.
91. Ferguson KM, Kavran JM, Sankaran VG, Fournier E, Isakoff SJ, Skolnik EY, Lemmon MA: **Structural basis for discrimination of 3-phosphoinositides by pleckstrin homology domains.** *Mol Cell* 2000, **6**:373-384.
92. Thomas CC, Deak M, Alessi DR, van Aalten DM: **High-resolution structure of the pleckstrin homology domain of protein kinase b/akt bound to phosphatidylinositol (3,4,5)-trisphosphate.** *Curr Biol* 2002, **12**:1256-1262.

93. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE: **UCSF Chimera--a visualization system for exploratory research and analysis.** *J Comput Chem* 2004, **25**:1605-1612.
94. Bottomley MJ, Salim K, Panayotou G: **Phospholipid-binding protein domains.** *Biochim Biophys Acta* 1998, **1436**:165-183.
95. Rebecchi M, Scarlata S: **Pleckstrin homology domains: a common fold with diverse functions.** *Annu Rev Biophys Biomol Struct* 1998, **27**:503-528.
96. Takeuchi H, Kanematsu T, Misumi Y, Sakane F, Konishi H, Kikkawa U, Watanabe Y, Katan M, Hirata M: **Distinct specificity in the binding of inositol phosphates by pleckstrin homology domains of pleckstrin, RAC-protein kinase, diacylglycerol kinase and a new 130 kDa protein.** *Biochim Biophys Acta* 1997, **3**:275-285.
97. Yu JW, Mendrola JM, Audhya A, Singh S, Keleti D, DeWald DB, Murray D, Emr SD, Lemmon MA: **Genome-wide analysis of membrane targeting by *S. cerevisiae* pleckstrin homology domains.** *Mol Cell* 2004, **13**:677-688.
98. Isakoff S, Cardozo T, Andreev J, Li Z, Ferguson K, Abagyan R, Lemmon M, Aronheim A, Skolnik E: **Identification and analysis of PH domain-containing targets of phosphatidylinositol 3-kinase using a novel in vivo assay in yeast.** *EMBO J* 1998, **17**:5374-5387.
99. Maroun C, Holgado-Madruga M, Royal I, Naujokas M, Fournier T, Wong A, Park M: **The Gab1 PH domain is required for localization of Gab1 at sites of cell-cell contact and epithelial morphogenesis downstream from the met receptor tyrosine kinase.** *Mol Cell Biol* 1999, **19**:1784-1799.
100. Maroun CR, Moscatello DK, Naujokas MA, Holgado-Madruga M, Wong AJ, Park M: **A conserved inositol phospholipid binding site within the pleckstrin homology domain of the Gab1 docking protein is required for epithelial morphogenesis.** *J Biol Chem* 1999, **274**:31719-31726.
101. Lee SH, Jin JB, Song J, Min MK, Park DS, Kim YW, Hwang I: **The intermolecular interaction between the PH domain and the C-terminal domain of Arabidopsis dynamin-like 6 determines lipid binding specificity.** *J Biol Chem* 2002, **277**:31842-31849.
102. Stevenson JM, Perera IY, Boss WF: **A phosphatidylinositol 4-kinase pleckstrin homology domain that binds phosphatidylinositol 4-monophosphate.** *J Biol Chem* 1998, **273**:22761-22767.
103. Qin C, Wang X: **The Arabidopsis phospholipase D family. Characterization of a calcium-independent and phosphatidylcholine-selective PLD zeta 1 with distinct regulatory domains.** *Plant Physiol* 2002, **128**:1057-1068.
104. Deak M, Casamayor A, Currie RA, Downes CP, Alessi DR: **Characterisation of a plant 3-phosphoinositide-dependent protein kinase-1 homologue which contains a pleckstrin homology domain.** *FEBS Lett* 1999, **451**:220-226.
105. Stevenson-Paulik J, Love J, Boss WF: **Differential regulation of two Arabidopsis type III phosphatidylinositol 4-kinase isoforms. A regulatory role for the pleckstrin homology domain.** *Plant Physiol* 2003, **132**:1053-1064.

106. Stenmark H, Aasland R, Toh BH, D'Arrigo A: **Endosomal localization of the autoantigen EEA1 is mediated by a zinc-binding FYVE finger.** *J Biol Chem* 1996, **271**:24048-24054.
107. Petiot A, Faure J, Stenmark H, Gruenberg J: **PI3P signaling regulates receptor sorting but not transport in the endosomal pathway.** *J Cell Biol* 2003, **162**:971-979.
108. Wurmser AE, Gary JD, Emr SD: **Phosphoinositide 3-Kinases and Their FYVE Domain-containing Effectors as Regulators of Vacuolar/Lysosomal Membrane Trafficking Pathways.** *J Biol Chem* 1999, **274**:9129-9132.
109. Simonsen A, Lippe R, Christoforidis S, Gaullier JM, Brech A, Callaghan J, Toh BH, Murphy C, Zerial M, Stenmark H: **EEA1 links PI(3)K function to Rab5 regulation of endosome fusion.** *Nature* 1998, **394**:494-498.
110. Rutherford AC, Traer C, Wassmer T, Pattni K, Bujny MV, Carlton JG, Stenmark H, Cullen PJ: **The mammalian phosphatidylinositol 3-phosphate 5-kinase (PIKfyve) regulates endosome-to-TGN retrograde transport.** *J Cell Sci* 2006, **119**:3944-3957.
111. Estrada L, Caron E, Gorski JL: **Fgd1, the Cdc42 guanine nucleotide exchange factor responsible for faciogenital dysplasia, is localized to the subcortical actin cytoskeleton and Golgi membrane.** *Hum Mol Genet* 2001, **10**:485-495.
112. Cooke FT, Dove SK, McEwen RK, Painter G, Holmes AB, Hall MN, Michell RH, Parker PJ: **The stress-activated phosphatidylinositol 3-phosphate 5-kinase Fab1p is essential for vacuole function in *S. cerevisiae*.** *Curr Biol* 1998, **8**:1219-1222.
113. Gary JD, Wurmser AE, Bonangelino CJ, Weisman LS, Emr SD: **Fab1p is essential for PtdIns(3)P 5-kinase activity and the maintenance of vacuolar size and membrane homeostasis.** *J Cell Biol* 1998, **143**:65-79.
114. Odorizzi G, Babst M, Emr SD: **Fab1p PtdIns(3)P 5-kinase function essential for protein sorting in the multivesicular body.** *Cell* 1998, **95**:847-858.
115. Tsukazaki T, Chiang TA, Davison AF, Attisano L, Wrana JL: **SARA, a FYVE domain protein that recruits Smad2 to the TGFbeta receptor.** *Cell* 1998, **95**:779-791.
116. Seet LF, Hong W: **Endofin, an endosomal FYVE domain protein.** *J Biol Chem* 2001, **276**:42445-42454.
117. Seet LF, Hong W: **Endofin recruits clathrin to early endosomes via TOM1.** *J Cell Sci* 2005, **118**:575-587.
118. Seet LF, Liu N, Hanson BJ, Hong W: **Endofin recruits TOM1 to endosomes.** *J Biol Chem* 2004, **279**:4670-4679.
119. Misra S, Hurley JH: **Crystal structure of a phosphatidylinositol 3-phosphate-specific membrane-targeting motif, the FYVE domain of Vps27p.** *Cell* 1999, **97**:657-666.
120. Burd CG, Emr SD: **Phosphatidylinositol(3)-phosphate signaling mediated by specific binding to RING FYVE domains.** *Mol Cell* 1998, **2**:157-162.
121. Gaullier JM, Simonsen A, D'Arrigo A, Bremnes B, Stenmark H, Aasland R: **FYVE fingers bind PtdIns(3)P.** *Nature* 1998, **394**:432-433.

122. Patki V, Virbasius J, Lane WS, Toh BH, Shpetner HS, Corvera S: **Identification of an early endosomal protein regulated by phosphatidylinositol 3-kinase.** *Proc Natl Acad Sci U S A* 1997, **94**:7326-7330.
123. Patki V, Lawe DC, Corvera S, Virbasius JV, Chawla A: **A functional PtdIns(3)P-binding motif.** *Nature* 1998, **394**:433-434.
124. Gaullier JM, Ronning E, Gillooly DJ, Stenmark H: **Interaction of the EEA1 FYVE finger with phosphatidylinositol 3-phosphate and early endosomes. Role of conserved residues.** *J Biol Chem* 2000, **275**:24595-24600.
125. Kutateladze TG, Ogburn KD, Watson WT, de Beer T, Emr SD, Burd CG, Overduin M: **Phosphatidylinositol 3-phosphate recognition by the FYVE domain.** *Mol Cell* 1999, **3**:805-811.
126. Kutateladze TG: **Phosphatidylinositol 3-phosphate recognition and membrane docking by the FYVE domain.** *Biochim Biophys Acta* 2006, **1761**:868-877.
127. Dumas JJ, Merithew E, Sudharshan E, Rajamani D, Hayes S, Lawe D, Corvera S, Lambright DG: **Multivalent endosome targeting by homodimeric EEA1.** *Mol Cell* 2001, **8**:947-958.
128. Jensen RB, La Cour T, Albrethsen J, Nielsen M, Skriver K: **FYVE zinc-finger proteins in the plant model *Arabidopsis thaliana*: identification of PtdIns3P-binding residues by comparison of classic and variant FYVE domains.** *Biochem J* 2001, **359**:165-173.
129. Stahelin RV, Long F, Diraviyam K, Bruzik KS, Murray D, Cho W: **Phosphatidylinositol 3-phosphate induces the membrane penetration of the FYVE domains of Vps27p and Hrs.** *J Biol Chem* 2002, **277**:26379-26388.
130. Diraviyam K, Stahelin RV, Cho W, Murray D: **Computer modeling of the membrane interaction of FYVE domains.** *J Mol Biol* 2003, **328**:721-736.
131. Kutateladze TG, Capelluto DG, Ferguson CG, Cheever ML, Kutateladze AG, Prestwich GD, Overduin M: **Multivalent mechanism of membrane insertion by the FYVE domain.** *J Biol Chem* 2004, **279**:3050-3057.
132. Mao Y, Nickitenko A, Duan X, Lloyd TE, Wu MN, Bellen H, Quirocho FA: **Crystal structure of the VHS and FYVE tandem domains of Hrs, a protein involved in membrane trafficking and signal transduction.** *Cell* 2000, **100**:447-456.
133. Kutateladze T, Overduin M: **Structural mechanism of endosome docking by the FYVE domain.** *Science* 2001, **291**:1793-1796.
134. Hayakawa A, Hayes SJ, Lawe DC, Sudharshan E, Tuft R, Fogarty K, Lambright D, Corvera S: **Structural basis for endosomal targeting by FYVE domains.** *J Biol Chem* 2004, **279**:5958-5966.
135. Kim DH, Eu YJ, Yoo CM, Kim YW, Pih KT, Jin JB, Kim SJ, Stenmark H, Hwang I: **Trafficking of phosphatidylinositol 3-phosphate from the trans-Golgi network to the lumen of the central vacuole in plant cells.** *Plant Cell* 2001, **13**:287-301.
136. Welters P, Takegawa K, Emr SD, Chrispeels MJ: **AtVPS34, a Phosphatidylinositol 3-Kinase of *Arabidopsis thaliana*, is an Essential Protein with Homology to a Calcium-Dependent Lipid Binding Domain.** *PNAS* 1994, **91**:11398-11402.

137. Heras B, Drobak BK: **PARF-1: an Arabidopsis thaliana FYVE-domain protein displaying a novel eukaryotic domain structure and phosphoinositide affinity.** *J Exp Bot* 2002, **53**:565-567.
138. Drobak BK, Heras B: **Nuclear phosphoinositides could bring FYVE alive.** *Trends Plant Sci* 2002, **7**:132-138.
139. Kost B, Lemichez E, Spielhofer P, Hong Y, Tolias K, Carpenter C, Chua NH: **Rac homologues and compartmentalized phosphatidylinositol 4, 5-bisphosphate act in a common pathway to regulate polar pollen tube growth.** *J Cell Biol* 1999, **145**:317-330.
140. Loewus MW, Loewus F: **myo-Inositol 1-Phosphate Synthase Inhibition and Control of Uridine Diphosphate-d-glucuronic Acid Biosynthesis in Plants.** *Plant Physiol* 1974, **54**:368-371.
141. Preuss ML, Schmitz AJ, Thole JM, Bonner HK, Otegui MS, Nielsen E: **A role for the RabA4b effector protein PI-4Kbeta1 in polarized expansion of root hair cells in Arabidopsis thaliana.** *J Cell Biol* 2006, **172**:991-998.
142. Vincent P, Chua M, Nogue F, Fairbrother A, Mekeel H, Xu Y, Allen N, Bibikova TN, Gilroy S, Bankaitis VA: **A Sec14p-nodulin domain phosphatidylinositol transfer protein polarizes membrane growth of Arabidopsis thaliana root hairs.** *J Cell Biol* 2005, **168**:801-812.
143. Williams ME, Torabinejad J, Cohick E, Parker K, Drake EJ, Thompson JE, Hortter M, Dewald DB: **Mutations in the Arabidopsis phosphoinositide phosphatase gene SAC9 lead to overaccumulation of PtdIns(4,5)P2 and constitutive expression of the stress-response pathway.** *Plant Physiol* 2005, **138**:686-700.
144. Mirkovic N, Li Z, Parnassa A, Murray D: **Strategies for high-throughput comparative modeling: Applications to leverage analysis in structural genomics and protein family organization.** *Proteins* 2006, **66**:766-777.
145. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE: **The Protein Data Bank.** *Nucleic Acids Res* 2000, **28**:235-242.
146. Berman H, Henrick K, Nakamura H: **Announcing the worldwide Protein Data Bank.** *Nat Struct Biol* 2003, **10**:980.
147. Kanehisa M: **The KEGG database.** *Novartis Found Symp* 2002, **247**:91-101; discussion 101-103, 119-128, 244-152.
148. Kanehisa M, Goto S, Hattori M, Aoki-Kinoshita KF, Itoh M, Kawashima S, Katayama T, Araki M, Hirakawa M: **From genomics to chemical genomics: new developments in KEGG.** *Nucl Acids Res* 2006, **34**:D354-357.
149. Schultz J, Copley RR, Doerks T, Ponting CP, Bork P: **SMART: a web-based tool for the study of genetically mobile domains.** *Nucleic Acids Res* 2000, **28**:231-234.
150. Schultz J, Milpetz F, Bork P, Ponting CP: **SMART, a simple modular architecture research tool: Identification of signaling domains.** *PNAS* 1998, **95**:5857-5864.
151. Letunic I, Copley RR, Pils B, Pinkert S, Schultz J, Bork P: **SMART 5: domains in the context of genomes and networks.** *Nucleic Acids Res* 2006, **34**:D257-260.

152. Bairoch A, Boeckmann B: **The SWISS-PROT protein sequence data bank: current status.** *Nucleic Acids Res* 1994, **22**:3578-3580.
153. Boeckmann B, Blatter MC, Famiglietti L, Hinz U, Lane L, Roechert B, Bairoch A: **Protein variety and functional diversity: Swiss-Prot annotation in its biological context.** *C R Biol* 2005, **328**:882-899.
154. Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Wheeler DL: **GenBank.** *Nucleic Acids Res* 2006, **34**:D16-20.
155. Wheeler DL, Barrett T, Benson DA, Bryant SH, Canese K, Chetvernin V, Church DM, DiCuccio M, Edgar R, Federhen S, et al: **Database resources of the National Center for Biotechnology Information.** *Nucleic Acids Res* 2006, **34**:D173-180.
156. Apweiler R, Bairoch A, Wu CH, Barker WC, Boeckmann B, Ferro S, Gasteiger E, Huang H, Lopez R, Magrane M, et al: **UniProt: the Universal Protein knowledgebase.** *Nucleic Acids Res* 2004, **32**:D115-119.
157. Bairoch A, Apweiler R, Wu CH, Barker WC, Boeckmann B, Ferro S, Gasteiger E, Huang H, Lopez R, Magrane M, et al: **The Universal Protein Resource (UniProt).** *Nucleic Acids Res* 2005, **33**:D154-159.
158. Wu CH, Apweiler R, Bairoch A, Natale DA, Barker WC, Boeckmann B, Ferro S, Gasteiger E, Huang H, Lopez R, et al: **The Universal Protein Resource (UniProt): an expanding universe of protein information.** *Nucleic Acids Res* 2006, **34**:D187-191.
159. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ: **Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.** *Nucleic Acids Res* 1997, **25**:3389-3402.
160. Bateman A, Coin L, Durbin R, Finn RD, Hollich V, Griffiths-Jones S, Khanna A, Marshall M, Moxon S, Sonnhammer EL, et al: **The Pfam protein families database.** *Nucleic Acids Res* 2004, **32**:D138-141.
161. Marchler-Bauer A, Anderson JB, Cherukuri PF, DeWeese-Scott C, Geer LY, Gwadz M, He S, Hurwitz DI, Jackson JD, Ke Z, et al: **CDD: a Conserved Domain Database for protein classification.** *Nucleic Acids Res* 2005, **33**:D192-196.
162. Marchler-Bauer A, Anderson JB, Derbyshire MK, DeWeese-Scott C, Gonzales NR, Gwadz M, Hao L, He S, Hurwitz DI, Jackson JD, et al: **CDD: a conserved domain database for interactive domain family analysis.** *Nucleic Acids Res* 2007, **35**:D237-240.
163. Marchler-Bauer A, Anderson JB, DeWeese-Scott C, Fedorova ND, Geer LY, He S, Hurwitz DI, Jackson JD, Jacobs AR, Lanczycki CJ, et al: **CDD: a curated Entrez database of conserved domain alignments.** *Nucleic Acids Res* 2003, **31**:383-387.
164. Marchler-Bauer A, Bryant SH: **CD-Search: protein domain annotations on the fly.** *Nucleic Acids Res* 2004, **32**:W327-331.
165. Marchler-Bauer A, Panchenko AR, Shoemaker BA, Thiessen PA, Geer LY, Bryant SH: **CDD: a database of conserved domain alignments with links to domain three-dimensional structure.** *Nucleic Acids Res* 2002, **30**:281-283.

166. Tatusov RL, Fedorova ND, Jackson JD, Jacobs AR, Kiryutin B, Koonin EV, Krylov DM, Mazumder R, Mekhedov SL, Nikolskaya AN, et al: **The COG database: an updated version includes eukaryotes.** *BMC Bioinformatics* 2003, **4**:41.
167. Tatusov RL, Koonin EV, Lipman DJ: **A genomic perspective on protein families.** *Science* 1997, **278**:631-637.
168. Cuff JA, Barton GJ: **Application of multiple sequence alignment profiles to improve protein secondary structure prediction.** *Proteins* 2000, **40**:502-511.
169. Cuff JA, Barton GJ: **Evaluation and improvement of multiple sequence methods for protein secondary structure prediction.** *Proteins* 1999, **34**:508-519.
170. Cuff JA, Clamp ME, Siddiqui AS, Finlay M, Barton GJ: **JPred: a consensus secondary structure prediction server.** *Bioinformatics* 1998, **14**:892-893.
171. Bryson K, McGuffin LJ, Marsden RL, Ward JJ, Sodhi JS, Jones DT: **Protein structure prediction servers at University College London.** *Nucleic Acids Res* 2005, **33**:W36-38.
172. Jones DT: **Protein secondary structure prediction based on position-specific scoring matrices.** *J Mol Biol* 1999, **292**:195-202.
173. McGuffin LJ, Bryson K, Jones DT: **The PSIPRED protein structure prediction server.** *Bioinformatics* 2000, **16**:404-405.
174. Rost B: **PHD: predicting one-dimensional protein structure by profile-based neural networks.** *Methods Enzymol* 1996, **266**:525-539.
175. Ouali M, King RD: **Cascaded multiple classifiers for secondary structure prediction.** *Protein Sci* 2000, **9**:1162-1176.
176. Karplus K, Barrett C, Hughey R: **Hidden Markov models for detecting remote protein homologies.** *Bioinformatics* 1998, **14**:846-856.
177. Karplus K, Katzman S, Shackleford G, Koeva M, Draper J, Barnes B, Soriano M, Hughey R: **SAM-T04: what is new in protein-structure prediction for CASP6.** *Proteins* 2005, **61 Suppl 7**:135-142.
178. Karplus K, Hu B: **Evaluation of protein multiple alignments by SAM-T99 using the BAliBASE multiple alignment test set.** *Bioinformatics* 2001, **17**:713-720.
179. Guermeur Y: **Combinaison de classifieurs statistiques, application à la prédiction de la structure secondaire des protéines.** . *PhD thesis* 1997.
180. Baldi P, Brunak S, Frasconi P, Soda G, Pollastri G: **Exploiting the past and the future in protein secondary structure prediction.** *Bioinformatics* 1999, **15**:937-946.
181. Cheng J, Randall AZ, Sweredoski MJ, Baldi P: **SCRATCH: a protein structure and structural feature prediction server.** *Nucleic Acids Res* 2005, **33**:W72-76.
182. Cole C, Barber JD, Barton GJ: **The Jpred 3 secondary structure prediction server.** *Nucleic Acids Res* 2008, **36**:W197-201.

183. Dor O, Zhou Y: **Achieving 80% ten-fold cross-validated accuracy for secondary structure prediction by large-scale training.** *Proteins* 2007, **66**:838-845.
184. Wallner B, Elofsson A: **All are not equal: A benchmark of different homology modeling programs.** *Protein Sci* 2005, **14**:1315-1327.
185. Petrey D, Honig B: **Protein structure prediction: inroads to biology.** *Mol Cell* 2005, **20**:811-819.
186. Singh SM, Murray D: **Molecular modeling of the membrane targeting of phospholipase C pleckstrin homology domains.** *Protein Sci* 2003, **12**:1934-1953.
187. Xiang Z, Soto CS, Honig B: **Evaluating conformational free energies: the colony energy and its application to the problem of loop prediction.** *Proc Natl Acad Sci U S A* 2002, **99**:7432-7437.
188. Canutescu AA, Shelenkov AA, Dunbrack RL, Jr.: **A graph-theory algorithm for rapid protein side-chain prediction.** *Protein Sci* 2003, **12**:2001-2014.
189. Xiang Z, Steinbach PJ, Jacobson MP, Friesner RA, Honig B: **Prediction of side-chain conformations on protein surfaces.** *Proteins* 2007.
190. Xiang Z, Honig B: **Extending the accuracy limits of prediction for side-chain conformations.** *J Mol Biol* 2001, **311**:421-430.
191. Jacobson MP, Friesner RA, Xiang Z, Honig B: **On the role of the crystal environment in determining protein side-chain conformations.** *J Mol Biol* 2002, **320**:597-608.
192. Sussman JL, Lin D, Jiang J, Manning NO, Prilusky J, Ritter O, Abola EE: **Protein Data Bank (PDB): database of three-dimensional structural information of biological macromolecules.** *Acta Crystallogr D Biol Crystallogr* 1998, **54**:1078-1084.
193. Alexandrov NN, Nussinov R, Zimmer RM: **Fast protein fold recognition via sequence to structure alignment and contact capacity potentials.** *Pac Symp Biocomput* 1996:53-72.
194. Shi J, Blundell TL, Mizuguchi K: **FUGUE: sequence-structure homology recognition using environment-specific substitution tables and structure-dependent gap penalties.** *J Mol Biol* 2001, **310**:243-257.
195. Meller J, Elber R: **Linear programming optimization and a double statistical filter for protein threading protocols.** *Proteins* 2001, **45**:241-261.
196. Teodorescu O, Galor T, Pillardy J, Elber R: **Enriching the sequence substitution matrix by structural information.** *Proteins* 2004, **54**:41-48.
197. Tobi D, Elber R: **Distance-dependent, pair potential for protein folding: results from linear optimization.** *Proteins* 2000, **41**:40-46.
198. Kelley LA, MacCallum RM, Sternberg MJ: **Enhanced genome annotation using structural profiles in the program 3D-PSSM.** *J Mol Biol* 2000, **299**:499-520.
199. Altschul SF, Koonin EV: **Iterated profile searches with PSI-BLAST--a tool for discovery in protein databases.** *Trends Biochem Sci* 1998, **23**:444-447.

200. Schaffer AA, Aravind L, Madden TL, Shavirin S, Spouge JL, Wolf YI, Koonin EV, Altschul SF: **Improving the accuracy of PSI-BLAST protein database searches with composition-based statistics and other refinements.** *Nucleic Acids Res* 2001, **29**:2994-3005.
201. Schaffer AA, Wolf YI, Ponting CP, Koonin EV, Aravind L, Altschul SF: **IMPALA: matching a protein sequence against a collection of PSI-BLAST-constructed position-specific score matrices.** *Bioinformatics* 1999, **15**:1000-1011.
202. Bates PA, Kelley LA, MacCallum RM, Sternberg MJ: **Enhancement of protein modeling by human intervention in applying the automatic programs 3D-JIGSAW and 3D-PSSM.** *Proteins* 2001, **Suppl 5**:39-46.
203. Bates PA, Sternberg MJ: **Model building by comparison at CASP3: using expert knowledge and computer automation.** *Proteins* 1999, **Suppl 3**:47-54.
204. Contreras-Moreira B, Bates PA: **Domain fishing: a first step in protein comparative modelling.** *Bioinformatics* 2002, **18**:1141-1142.
205. Tosatto SCE: **The Victor package for 3D protein structure modeling.** Submitted.
206. Tosatto SCE: **The Victor/FRST Function for Model Quality Estimation.** *Journal of Computational Biology* 2005, **12**:1316-1327.
207. Lund O, Nielsen M, Lundegaard C, Worning P: **CPHmodels 2.0: X3M a Computer Program to Extract 3D Models.** *Abstract at the CASP5 conference A102* 2002.
208. Nicholas K, Nicholas H, Deerfield D: **GeneDoc: Analysis and Visualization of Genetic Variation.** *EMBNEWNEWS* 1997, **14**.
209. Bonneau R, Strauss CE, Rohl CA, Chivian D, Bradley P, Malmstrom L, Robertson T, Baker D: **De novo prediction of three-dimensional structures for major protein families.** *J Mol Biol* 2002, **322**:65-78.
210. Bonneau R, Tsai J, Ruczinski I, Chivian D, Rohl C, Strauss CE, Baker D: **Rosetta in CASP4: progress in ab initio protein structure prediction.** *Proteins* 2001, **Suppl 5**:119-126.
211. Chivian D, Kim DE, Malmstrom L, Bradley P, Robertson T, Murphy P, Strauss CE, Bonneau R, Rohl CA, Baker D: **Automated prediction of CASP-5 structures using the Robetta server.** *Proteins* 2003, **53 Suppl 6**:524-533.
212. Chivian D, Kim DE, Malmstrom L, Schonbrun J, Rohl CA, Baker D: **Prediction of CASP6 structures using automated Robetta protocols.** *Proteins* 2005, **61 Suppl 7**:157-166.
213. Kim DE, Chivian D, Baker D: **Protein structure prediction and analysis using the Robetta server.** *Nucl Acids Res* 2004, **32**:W526-531.
214. Simons KT, Kooperberg C, Huang E, Baker D: **Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and Bayesian scoring functions.** *J Mol Biol* 1997, **268**:209-225.
215. Simons KT, Ruczinski I, Kooperberg C, Fox BA, Bystroff C, Baker D: **Improved recognition of native-like protein structures using a combination of sequence-dependent and sequence-independent features of proteins.** *Proteins* 1999, **34**:82-95.

216. Moult J: **A decade of CASP: progress, bottlenecks and prognosis in protein structure prediction.** *Curr Opin Struct Biol* 2005, **15**:285-289.
217. Sander C, Schneider R: **Database of homology-derived protein structures and the structural meaning of sequence alignment.** *Proteins* 1991, **9**:56-68.
218. Pellegrini-Calace M, Soro S, Tramontano A: **Revisiting the prediction of protein function at CASP6.** *Febs J* 2006, **273**:2977-2983.
219. Fiser A, Do RK, Sali A: **Modeling of loops in protein structures.** *Protein Sci* 2000, **9**:1753-1773.
220. Marti-Renom MA, Stuart AC, Fiser A, Sanchez R, Melo F, Sali A: **Comparative protein structure modeling of genes and genomes.** *Annu Rev Biophys Biomol Struct* 2000, **29**:291-325.
221. Pieper U, Eswar N, Braberg H, Madhusudhan MS, Davis FP, Stuart AC, Mirkovic N, Rossi A, Marti-Renom MA, Fiser A, et al: **MODBASE, a database of annotated comparative protein structure models, and associated resources.** *Nucleic Acids Res* 2004, **32**:D217-222.
222. Sali A, Blundell TL: **Comparative protein modelling by satisfaction of spatial restraints.** *J Mol Biol* 1993, **234**:779-815.
223. Petrey D, Xiang Z, Tang CL, Xie L, Gimpelev M, Mitros T, Soto CS, Goldsmith-Fischman S, Kernytsky A, Schlessinger A, et al: **Using multiple structure alignments, fast model building, and energetic analysis in fold recognition and homology modeling.** *Proteins* 2003, **53 Suppl 6**:430-435.
224. Bowie JU, Luthy R, Eisenberg D: **A method to identify protein sequences that fold into a known three-dimensional structure.** *Science* 1991, **253**:164-170.
225. Eisenberg D, Luthy R, Bowie JU: **VERIFY3D: assessment of protein models with three-dimensional profiles.** *Methods Enzymol* 1997, **277**:396-404.
226. Luthy R, Bowie JU, Eisenberg D: **Assessment of protein models with three-dimensional profiles.** *Nature* 1992, **356**:83-85.
227. Sippl MJ: **Boltzmann's principle, knowledge-based mean fields and protein folding. An approach to the computational determination of protein structures.** *J Comput Aided Mol Des* 1993, **7**:473-501.
228. Nicholls A, Sharp KA, Honig B: **Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons.** *Proteins* 1991, **11**:281-296.
229. Gordon JC, Myers JB, Foltz T, Shoja V, Heath LS, Onufriev A: **H++: a server for estimating pKas and adding missing hydrogens to macromolecules.** *Nucleic Acids Res* 2005, **33**:W368-371.
230. Myers J, Grothaus G, Narayanan S, Onufriev A: **A simple clustering algorithm can be accurate enough for use in calculations of pKs in macromolecules.** *Proteins* 2006, **63**:928-938.

231. Bashford D, Karplus M: **pKa's of ionizable groups in proteins: atomic detail from a continuum electrostatic model.** *Biochemistry* 1990, **29**:10219-10225.
232. Jackson SG, Zhang Y, Haslam RJ, Junop MS: **Structural analysis of the carboxy terminal PH domain of pleckstrin bound to D-myo-inositol 1,2,3,5,6-pentakisphosphate.** *BMC Struct Biol* 2007, **7**:80.
233. Kuntz ID, Blaney JM, Oatley SJ, Langridge R, Ferrin TE: **A geometric approach to macromolecule-ligand interactions.** *J Mol Biol* 1982, **161**:269-288.
234. Richards FM: **Areas, volumes, packing and protein structure.** *Annu Rev Biophys Bioeng* 1977, **6**:151-176.
235. Ferrin TE, Huang CC, Jarvis LE, Langridge R: **The MIDAS display system.** *J Mol Graph* 1988, **6**:13-27.
236. Shoichet BK, Bodian DL, Kuntz ID: **Molecular docking using shape descriptors.** *J Comp Chem* 1992, **13**:380-397.
237. Meng EC, Shoichet BK, Kuntz ID: **Automated docking with grid-based energy evaluation.** *J Comp Chem* 1992, **13**:505-524.
238. Sobolev V, Sorokine A, Prilusky J, Abola EE, Edelman M: **Automated analysis of interatomic contacts in proteins.** *Bioinformatics* 1999, **15**:327-332.
239. Chenna R, Sugawara H, Koike T, Lopez R, Gibson TJ, Higgins DG, Thompson JD: **Multiple sequence alignment with the Clustal series of programs.** *Nucleic Acids Res* 2003, **31**:3497-3500.
240. Emanuelsson O, Nielsen H, von Heijne G: **ChloroP, a neural network-based method for predicting chloroplast transit peptides and their cleavage sites.** *Protein Sci* 1999, **8**:978-984.
241. Claros MG, Vincens P: **Computational method to predict mitochondrially imported proteins and their targeting sequences.** *Eur J Biochem* 1996, **241**:779-786.
242. Nakai K, Horton P: **Computational prediction of subcellular localization.** *Methods Mol Biol* 2007, **390**:429-466.
243. Horton P, Park KJ, Obayashi T, Fujita N, Harada H, Adams-Collier CJ, Nakai K: **WoLF PSORT: protein localization predictor.** *Nucleic Acids Res* 2007, **35**:W585-587.
244. Horton P, Park K, Obayashi T, Nakai K: **Protein Subcellular Localization Prediction with WoLF PSORT.** *Proceedings of the 4th Annual Asia Pacific Bioinformatics Conference APBC06* 2006:39-48.
245. Burges CJC: **A Tutorial on Support Vector Machines for Pattern Recognition.** *Data Mining and Knowledge Discovery* 1998, **2**:121 - 167.
246. Cai CZ, Han LY, Ji ZL, Chen X, Chen YZ: **SVM-Prot: Web-based support vector machine software for functional classification of a protein from its primary sequence.** *Nucleic Acids Res* 2003, **31**:3692-3697.
247. Brendel V, Zhu W: **Computational modeling of gene structure in Arabidopsis thaliana.** *Plant Mol Biol* 2002, **48**:49-58.

248. Higgins DG: **CLUSTAL V: multiple alignment of DNA and protein sequences.** *Methods Mol Biol* 1994, **25**:307-318.
249. Huelsenbeck JP, Ronquist F: **MRBAYES: Bayesian inference of phylogenetic trees.** *Bioinformatics* 2001, **17**:754-755.
250. van Leeuwen W, Okresz L, Bogre L, Munnik T: **Learning the lipid language of plant signalling.** *Trends Plant Sci* 2004, **9**:378-384.
251. Geer LY, Domrachev M, Lipman DJ, Bryant SH: **CDART: protein homology by domain architecture.** *Genome Res* 2002, **12**:1619-1623.
252. Koizumi K, Naramoto S, Sawa S, Yahara N, Ueda T, Nakano A, Sugiyama M, Fukuda H: **VAN3 ARF-GAP-mediated vesicle transport is involved in leaf vascular network formation.** *Development* 2005, **132**:1699-1711.
253. Jin JB, Kim YA, Kim SJ, Lee SH, Kim DH, Cheong GW, Hwang I: **A new dynamin-like protein, ADL6, is involved in trafficking from the trans-Golgi network to the central vacuole in Arabidopsis.** *Plant Cell* 2001, **13**:1511-1526.
254. Levine TP, Munro S: **The pleckstrin homology domain of oxysterol-binding protein recognises a determinant specific to Golgi membranes.** *Curr Biol* 1998, **8**:729-739.
255. Ridgway ND, Dawson PA, Ho YK, Brown MS, Goldstein JL: **Translocation of oxysterol binding protein to Golgi apparatus triggered by ligand binding.** *J Cell Biol* 1992, **116**:307-319.
256. Jaworski CJ, Moreira E, Li A, Lee R, Rodriguez IR: **A family of 12 human genes containing oxysterol-binding domains.** *Genomics* 2001, **78**:185-196.
257. Levanon D, Hsieh CL, Francke U, Dawson PA, Ridgway ND, Brown MS, Goldstein JL: **cDNA cloning of human oxysterol-binding protein and localization of the gene to human chromosome 11 and mouse chromosome 19.** *Genomics* 1990, **7**:65-74.
258. Anniss AM, Apostolopoulos J, Dworkin S, Purton LE, Sparrow RL: **An oxysterol-binding protein family identified in the mouse.** *DNA Cell Biol* 2002, **21**:571-580.
259. Alphey L, Jimenez J, Glover D: **A Drosophila homologue of oxysterol binding protein (OSBP)--implications for the role of OSBP.** *Biochim Biophys Acta* 1998, **1395**:159-164.
260. Levine TP, Munro S: **Dual targeting of Osh1p, a yeast homologue of oxysterol-binding protein, to both the Golgi and the nucleus-vacuole junction.** *Mol Biol Cell* 2001, **12**:1633-1644.
261. Skirpan AL, Dowd PE, Sijacic P, Jaworski CJ, Gilroy S, Kao TH: **Identification and characterization of PiORP1, a Petunia oxysterol-binding-protein related protein involved in receptor-kinase mediated signaling in pollen, and analysis of the ORP gene family in Arabidopsis.** *Plant Mol Biol* 2006, **61**:553-565.
262. Fang M, Kearns BG, Gedvilaite A, Kagiwada S, Kearns M, Fung MK, Bankaitis VA: **Kes1p shares homology with human oxysterol binding protein and participates in a novel regulatory pathway for yeast Golgi-derived transport vesicle biogenesis.** *Embo J* 1996, **15**:6447-6459.

263. Fournier MV, Guimaraes da Costa F, Paschoal ME, Ronco LV, Carvalho MG, Pardee AB: **Identification of a gene encoding a human oxysterol-binding protein-homologue: a potential general molecular marker for blood dissemination of solid tumors.** *Cancer Res* 1999, **59**:3748-3753.
264. Brown MS, Goldstein JL: **A proteolytic pathway that controls the cholesterol content of membranes, cells, and blood.** *Proc Natl Acad Sci U S A* 1999, **96**:11041-11048.
265. Roy A, Levine TP: **Multiple pools of phosphatidylinositol 4-phosphate detected using the pleckstrin homology domain of Osh2p.** *J Biol Chem* 2004, **279**:44683-44689.
266. Li H, Tomizawa T, Koshiba S, Inoue M, Kigawa T, Yokoyama S: **Solution Structure of the Pleckstrin Homology Domain of Oxysterol-Binding Protein-Related Protein 8 (KIAA1451 Protein).** To be Published.
267. Levine TP, Munro S: **Targeting of Golgi-specific pleckstrin homology domains involves both PtdIns 4-kinase-dependent and -independent components.** *Curr Biol* 2002, **12**:695-704.
268. Li DY, Inoue H, Takahashi M, Kojima T, Shiraiwa M, Takahara H: **Molecular characterization of a novel salt-inducible gene for an OSBP (oxysterol-binding protein)-homologue from soybean.** *Gene* 2007.
269. Li H, Sato M, Koshiba S, Inoue M, Kigawa T, Yokoyama S: **Solution structure of the PH domain of Oxysterol binding protein-related protein 11** To be Published.
270. Soccio RE, Breslow JL: **StAR-related lipid transfer (START) proteins: mediators of intracellular lipid metabolism.** *J Biol Chem* 2003, **278**:22183-22186.
271. Ponting CP, Aravind L: **START: a lipid-binding domain in StAR, HD-ZIP and signalling proteins.** *Trends Biochem Sci* 1999, **24**:130-132.
272. Schrick K, Nguyen D, Karlowski WM, Mayer KF: **START lipid/sterol-binding domains are amplified in plants and are predominantly associated with homeodomain transcription factors.** *Genome Biol* 2004, **5**:R41.
273. Tsujishita Y, Hurley JH: **Structure and lipid transport mechanism of a StAR-related domain.** *Nat Struct Biol* 2000, **7**:408-414.
274. Westerman J, Wirtz KW, Berkhout T, van Deenen LL, Radhakrishnan R, Khorana HG: **Identification of the lipid-binding site of phosphatidylcholine-transfer protein with phosphatidylcholine analogs containing photoactivable carbene precursors.** *Eur J Biochem* 1983, **132**:441-449.
275. Roderick SL, Chan WW, Agate DS, Olsen LR, Vetting MW, Rajashankar KR, Cohen DE: **Structure of human phosphatidylcholine transfer protein in complex with its ligand.** *Nat Struct Biol* 2002, **9**:507-511.
276. de Brouwer AP, Bouma B, van Tiel CM, Heerma W, Brouwers JF, Bevers LE, Westerman J, Roelofsen B, Wirtz KW: **The binding of phosphatidylcholine to the phosphatidylcholine transfer protein: affinity and role in folding.** *Chem Phys Lipids* 2001, **112**:109-119.

277. Romanowski MJ, Soccio RE, Breslow JL, Burley SK: **Crystal structure of the *Mus musculus* cholesterol-regulated START protein 4 (StarD4) containing a StAR-related lipid transfer domain.** *Proc Natl Acad Sci U S A* 2002, **99**:6949-6954.
278. Perry RJ, Ridgway ND: **Molecular mechanisms and regulation of ceramide transport.** *Biochim Biophys Acta* 2005, **1734**:220-234.
279. Tang D, Ade J, Frye CA, Innes RW: **Regulation of plant defense responses in *Arabidopsis* by EDR2, a PH and START domain-containing protein.** *Plant J* 2005, **44**:245-257.
280. Hanada K: **Discovery of the molecular machinery CERT for endoplasmic reticulum-to-Golgi trafficking of ceramide.** *Mol Cell Biochem* 2006, **286**:23-31.
281. Hanada K, Kumagai K, Tomishige N, Kawano M: **CERT and intracellular trafficking of ceramide.** *Biochim Biophys Acta* 2007.
282. Hanada K, Kumagai K, Yasuda S, Miura Y, Kawano M, Fukasawa M, Nishijima M: **Molecular machinery for non-vesicular trafficking of ceramide.** *Nature* 2003, **426**:803-809.
283. Ponting CP, Kerr ID: **A novel family of phospholipase D homologues that includes phospholipid synthases and putative endonucleases: identification of duplicated repeats and potential active site residues.** *Protein Sci* 1996, **5**:914-922.
284. Hammond SM, Altshuller YM, Sung TC, Rudge SA, Rose K, Engebrecht J, Morris AJ, Frohman MA: **Human ADP-ribosylation factor-activated phosphatidylcholine-specific phospholipase D defines a new and highly conserved gene family.** *J Biol Chem* 1995, **270**:29640-29643.
285. Koonin EV: **A duplicated catalytic motif in a new superfamily of phosphohydrolases and phospholipid synthases that includes poxvirus envelope proteins.** *Trends Biochem Sci* 1996, **21**:242-243.
286. Pappan K, Zheng L, Krishnamoorthi R, Wang X: **Evidence for and characterization of Ca^{2+} binding to the catalytic region of *Arabidopsis thaliana* phospholipase Dbeta.** *J Biol Chem* 2004, **279**:47833-47839.
287. Han LY, Lin HH, Li ZR, Zheng CJ, Cao ZW, Xie B, Chen YZ: **PEARLS: Program for Energetic Analysis of Receptor-Ligand System.** *J Chem Inf Model* 2006, **46**:445-450.
288. Li M, Welti R, Wang X: **Quantitative profiling of *Arabidopsis* polar glycerolipids in response to phosphorus starvation. Roles of phospholipases D zeta1 and D zeta2 in phosphatidylcholine hydrolysis and digalactosyldiacylglycerol accumulation in phosphorus-starved plants.** *Plant Physiol* 2006, **142**:750-761.
289. Yamaryo Y, Dubots E, Albrieux C, Baldan B, Block MA: **Phosphate availability affects the tonoplast localization of PLDzeta2, an *Arabidopsis thaliana* phospholipase D.** *FEBS Lett* 2008, **582**:685-690.
290. Li G, Xue HW: ***Arabidopsis* PLDzeta2 regulates vesicle trafficking and is required for auxin response.** *Plant Cell* 2007, **19**:281-295.
291. Colley WC, Sung TC, Roll R, Jenco J, Hammond SM, Altshuller Y, Bar-Sagi D, Morris AJ, Frohman MA: **Phospholipase D2, a distinct phospholipase D isoform with novel**

- regulatory properties that provokes cytoskeletal reorganization.** *Curr Biol* 1997, **7**:191-201.
292. Du G, Altshuller YM, Vitale N, Huang P, Chasserot-Golaz S, Morris AJ, Bader MF, Frohman MA: **Regulation of phospholipase D1 subcellular cycling through coordination of multiple membrane association motifs.** *J Cell Biol* 2003, **162**:305-315.
 293. Lucocq J, Manifava M, Bi K, Roth MG, Ktistakis NT: **Immunolocalisation of phospholipase D1 on tubular vesicular membranes of endocytic and secretory origin.** *Eur J Cell Biol* 2001, **80**:508-520.
 294. Morris AJ, Hammond SM, Colley C, Sung TC, Jenco JM, Sciorra VA, Rudge SA, Frohman MA: **Regulation and functions of phospholipase D.** *Biochem Soc Trans* 1997, **25**:1151-1157.
 295. Vitale N, Caumont AS, Chasserot-Golaz S, Du G, Wu S, Sciorra VA, Morris AJ, Frohman MA, Bader MF: **Phospholipase D1: a key factor for the exocytotic machinery in neuroendocrine cells.** *Embo J* 2001, **20**:2424-2434.
 296. Du G, Huang P, Liang BT, Frohman MA: **Phospholipase D2 localizes to the plasma membrane and regulates angiotensin II receptor endocytosis.** *Mol Biol Cell* 2004, **15**:1024-1030.
 297. Hodgkin MN, Masson MR, Powner D, Saqib KM, Ponting CP, Wakelam MJ: **Phospholipase D regulation and localisation is dependent upon a phosphatidylinositol 4,5-bisphosphate-specific PH domain.** *Curr Biol* 2000, **10**:43-46.
 298. Sung TC, Altshuller YM, Morris AJ, Frohman MA: **Molecular analysis of mammalian phospholipase D2.** *J Biol Chem* 1999, **274**:494-502.
 299. Sung TC, Zhang Y, Morris AJ, Frohman MA: **Structural analysis of human phospholipase D1.** *J Biol Chem* 1999, **274**:3659-3666.
 300. Sciorra VA, Rudge SA, Prestwich GD, Frohman MA, Engebrecht J, Morris AJ: **Identification of a phosphoinositide binding motif that mediates activation of mammalian and yeast phospholipase D isoenzymes.** *Embo J* 1999, **18**:5911-5921.
 301. Briggs GC, Mouchel CF, Hardtke CS: **Characterization of the plant-specific BREVIS RADIX gene family reveals limited genetic redundancy despite high sequence conservation.** *Plant Physiol* 2006, **140**:1306-1316.
 302. Mouchel CF, Briggs GC, Hardtke CS: **Natural genetic variation in Arabidopsis identifies BREVIS RADIX, a novel regulator of cell proliferation and elongation in the root.** *Genes Dev* 2004, **18**:700-714.
 303. Lemmon M: **Pleckstrin homology domains: not just for phosphoinositides.** *Biochem Soc Trans* 2004, **32**:707-711.
 304. Moore W, Zhang C, Clarke PR: **Targeting of RCC1 to chromosomes is required for proper mitotic spindle assembly in human cells.** *Curr Biol* 2002, **12**:1442-1447.
 305. Flesch FM, Yu JW, Lemmon MA, Burger KN: **Membrane activity of the phospholipase C-delta1 pleckstrin homology (PH) domain.** *Biochem J* 2005, **389**:435-441.

306. Otterhag L, Gustavsson N, Alsterfjord M, Pical C, Lehrach H, Gobom J, Sommarin M: **Arabidopsis PDK1: identification of sites important for activity and downstream phosphorylation of S6 kinase.** *Biochimie* 2006, **88**:11-21.
307. Seong HA, Jung H, Choi HS, Kim KT, Ha H: **Regulation of transforming growth factor-beta signaling and PDK1 kinase activity by physical interaction between PDK1 and serine-threonine kinase receptor-associated protein.** *J Biol Chem* 2005, **280**:42897-42908.
308. Seong HA, Jung H, Kim KT, Ha H: **3-Phosphoinositide-dependent PDK1 negatively regulates transforming growth factor-beta-induced signaling in a kinase-dependent manner through physical interaction with Smad proteins.** *J Biol Chem* 2007, **282**:12272-12289.
309. Alessi DR, James SR, Downes CP, Holmes AB, Gaffney PR, Reese CB, Cohen P: **Characterization of a 3-phosphoinositide-dependent protein kinase which phosphorylates and activates protein kinase Balpha.** *Curr Biol* 1997, **7**:261-269.
310. Stephens L, Anderson K, Stokoe D, Erdjument-Bromage H, Painter GF, Holmes AB, Gaffney PR, Reese CB, McCormick F, Tempst P, et al: **Protein kinase B kinases that mediate phosphatidylinositol 3,4,5-trisphosphate-dependent activation of protein kinase B.** *Science* 1998, **279**:710-714.
311. Currie RA, Walker KS, Gray A, Deak M, Casamayor A, Downes CP, Cohen P, Alessi DR, Lucocq J: **Role of phosphatidylinositol 3,4,5-trisphosphate in regulating the activity and localization of 3-phosphoinositide-dependent protein kinase-1.** *Biochem J* 1999, **337 (Pt 3)**:575-583.
312. Walker KS, Deak M, Paterson A, Hudson K, Cohen P, Alessi DR: **Activation of protein kinase B beta and gamma isoforms by insulin in vivo and by 3-phosphoinositide-dependent protein kinase-1 in vitro: comparison with protein kinase B alpha.** *Biochem J* 1998, **331 (Pt 1)**:299-308.
313. Komander D, Fairservice A, Deak M, Kular GS, Prescott AR, Peter Downes C, Safrany ST, Alessi DR, van Aalten DM: **Structural insights into the regulation of PDK1 by phosphoinositides and inositol phosphates.** *Embo J* 2004, **23**:3918-3928.
314. Navarro P, Durrens P, Aigle M: **Protein-protein interaction between the RVS161 and RVS167 gene products of Saccharomyces cerevisiae.** *Biochim Biophys Acta* 1997, **1343**:187-192.
315. Lee MC, Schekman R: **Cell biology. BAR domains go on a bender.** *Science* 2004, **303**:479-480.
316. Peter BJ, Kent HM, Mills IG, Vallis Y, Butler PJ, Evans PR, McMahon HT: **BAR domains as sensors of membrane curvature: the amphiphysin BAR structure.** *Science* 2004, **303**:495-499.
317. Dawson JC, Legg JA, Machesky LM: **Bar domain proteins: a role in tubulation, scission and actin assembly in clathrin-mediated endocytosis.** *Trends Cell Biol* 2006, **16**:493-498.
318. Ren G, Vajjhala P, Lee JS, Winsor B, Munn AL: **The BAR domain proteins: molding membranes in fission, fusion, and phagy.** *Microbiol Mol Biol Rev* 2006, **70**:37-120.

319. Cukierman E, Huber I, Rotman M, Cassel D: **The ARF1 GTPase-activating protein: zinc finger motif and Golgi complex localization.** *Science* 1995, **270**:1999-2002.
320. Donaldson JG, Klausner RD: **ARF: a key regulatory switch in membrane traffic and organelle structure.** *Curr Opin Cell Biol* 1994, **6**:527-532.
321. Jackson TR, Brown FD, Nie Z, Miura K, Foroni L, Sun J, Hsu VW, Donaldson JG, Randazzo PA: **ACAPs are arf6 GTPase-activating proteins that function in the cell periphery.** *J Cell Biol* 2000, **151**:627-638.
322. Li J, Mao X, Dong LQ, Liu F, Tong L: **Crystal structures of the BAR-PH and PTB domains of human APPL1.** *Structure* 2007, **15**:525-533.
323. Bradshaw JM, Waksman G: **Molecular recognition by SH2 domains.** *Adv Protein Chem* 2002, **61**:161-210.
324. Machida K, Mayer BJ: **The SH2 domain: versatile signaling module and pharmaceutical target.** *Biochim Biophys Acta* 2005, **1747**:1-25.
325. Miaczynska M, Christoforidis S, Giner A, Shevchenko A, Uttenweiler-Joseph S, Habermann B, Wilm M, Parton RG, Zerial M: **APPL proteins link Rab5 to nuclear signal transduction via an endosomal compartment.** *Cell* 2004, **116**:445-456.
326. Kam JL, Miura K, Jackson TR, Gruschus J, Roller P, Stauffer S, Clark J, Aneja R, Randazzo PA: **Phosphoinositide-dependent activation of the ADP-ribosylation factor GTPase-activating protein ASAP1. Evidence for the pleckstrin homology domain functioning as an allosteric site.** *J Biol Chem* 2000, **275**:9653-9663.
327. Randazzo PA, Kahn RA: **GTP hydrolysis by ADP-ribosylation factor is dependent on both an ADP-ribosylation factor GTPase-activating protein and acid phospholipids.** *J Biol Chem* 1994, **269**:10758-10763.
328. Li H, Tomizawa T, Koshiba S, Inoue M, Kigawa T, Yokoyama S: **Solution structure of the PH domain of PIP2-dependent ARF1 GTPase-activating protein from human.** To be Published.
329. Burridge K, Wennerberg K: **Rho and Rac take center stage.** *Cell* 2004, **116**:167-179.
330. Wennerberg K, Der CJ: **Rho-family GTPases: it's not only Rac and Rho (and I like it).** *J Cell Sci* 2004, **117**:1301-1312.
331. Wennerberg K, Rossman KL, Der CJ: **The Ras superfamily at a glance.** *J Cell Sci* 2005, **118**:843-846.
332. Katoh M, Katoh M: **FBNP2 gene on human chromosome 1q32.1 encodes ARHGAP family protein with FCH, FBH, RhoGAP and SH3 domains.** *Int J Mol Med* 2003, **11**:791-797.
333. Katoh M, Katoh M: **Identification and characterization of human KIAA1391 and mouse Kiaa1391 genes encoding novel RhoGAP family proteins with RA domain and ANXL repeats.** *Int J Oncol* 2003, **23**:1471-1476.
334. Yuan BZ, Miller MJ, Keck CL, Zimonjic DB, Thorgeirsson SS, Popescu NC: **Cloning, characterization, and chromosomal localization of a gene frequently deleted in**

- human liver cancer (DLC-1) homologous to rat RhoGAP.** *Cancer Res* 1998, **58**:2196-2199.
335. Ching YP, Wong CM, Chan SF, Leung TH, Ng DC, Jin DY, Ng IO: **Deleted in liver cancer (DLC) 2 encodes a RhoGAP protein with growth suppressor function and is underexpressed in hepatocellular carcinoma.** *J Biol Chem* 2003, **278**:10824-10830.
 336. Tcherkezian J, Lamarche-Vane N: **Current knowledge of the large RhoGAP family of proteins.** *Biol Cell* 2007, **99**:67-86.
 337. Moon SY, Zheng Y: **Rho GTPase-activating proteins in cell regulation.** *Trends Cell Biol* 2003, **13**:13-22.
 338. Delawary M, Nakazawa T, Tezuka T, Sawa M, Iino Y, Takenawa T, Yamamoto T: **Molecular characterization of a novel RhoGAP, RRC-1 of the nematode *Caenorhabditis elegans*.** *Biochem Biophys Res Commun* 2007, **357**:377-382.
 339. Rivero F, Dislich H, Glockner G, Noegel AA: **The Dictyostelium discoideum family of Rho-related proteins.** *Nucleic Acids Res* 2001, **29**:1068-1079.
 340. Rivero F, Somesh BP: **Signal transduction pathways regulated by Rho GTPases in Dictyostelium.** *J Muscle Res Cell Motil* 2002, **23**:737-749.
 341. Wilkins A, Szafranski K, Fraser DJ, Bakthavatsalam D, Muller R, Fisher PR, Glockner G, Eichinger L, Noegel AA, Insall RH: **The Dictyostelium genome encodes numerous RasGEFs with multiple biological roles.** *Genome Biol* 2005, **6**:R68.
 342. Vlahou G, Rivero F: **Rho GTPase signaling in Dictyostelium discoideum: insights from the genome.** *Eur J Cell Biol* 2006, **85**:947-959.
 343. Billuart P, Bienvenu T, Ronce N, des Portes V, Vinet MC, Zemni R, Roest Crolius H, Carrie A, Fauchereau F, Cherry M, et al: **Oligophrenin-1 encodes a rhoGAP protein involved in X-linked mental retardation.** *Nature* 1998, **392**:923-926.
 344. Furukawa Y, Kawasoe T, Daigo Y, Nishiwaki T, Ishiguro H, Takahashi M, Kitayama J, Nakamura Y: **Isolation of a novel human gene, ARHGAP9, encoding a rho-GTPase activating protein.** *Biochem Biophys Res Commun* 2001, **284**:643-649.
 345. Seoh ML, Ng CH, Yong J, Lim L, Leung T: **ArhGAP15, a novel human RacGAP protein with GTPase binding property.** *FEBS Lett* 2003, **539**:131-137.
 346. Li H, Tochio N, Koshiba S, Inoue M, Kigawa T, Yokoyama S: **Solution structure of the PH domain of Rho GTPase activating protein 21 from human.** To be Published.
 347. van der Blik AM: **Functional diversity in the dynamin family.** *Trends Cell Biol* 1999, **9**:96-102.
 348. Vallee RB, Herskovits JS, Aghajanian JG, Burgess CC, Shpetner HS: **Dynamin, a GTPase involved in the initial stages of endocytosis.** *Ciba Found Symp* 1993, **176**:185-193; discussion 193-187.
 349. Wilsbach K, Payne GS: **Dynamic retention of TGN membrane proteins in *Saccharomyces cerevisiae*.** *Trends Cell Biol* 1993, **3**:426-432.

350. Wilsbach K, Payne GS: **Vps1p, a member of the dynamin GTPase family, is necessary for Golgi membrane protein retention in *Saccharomyces cerevisiae*.** *Embo J* 1993, **12**:3049-3059.
351. Yu X, Cai M: **The yeast dynamin-related GTPase Vps1p functions in the organization of the actin cytoskeleton via interaction with Sla1p.** *J Cell Sci* 2004, **117**:3839-3853.
352. Vizeacoumar FJ, Vreden WN, Fagarasanu M, Eitzen GA, Aitchison JD, Rachubinski RA: **The dynamin-like protein Vps1p of the yeast *Saccharomyces cerevisiae* associates with peroxisomes in a Pex19p-dependent manner.** *J Biol Chem* 2006, **281**:12817-12823.
353. Jones BA, Fangman WL: **Mitochondrial DNA maintenance in yeast requires a protein containing a region related to the GTP-binding domain of dynamin.** *Genes Dev* 1992, **6**:380-389.
354. Marks B, Stowell MH, Vallis Y, Mills IG, Gibson A, Hopkins CR, McMahon HT: **GTPase activity of dynamin and resulting conformation change are essential for endocytosis.** *Nature* 2001, **410**:231-235.
355. Hong Z, Geisler-Lee CJ, Zhang Z, Verma DP: **Phragmoplastin dynamics: multiple forms, microtubule association and their roles in cell plate formation in plants.** *Plant Mol Biol* 2003, **53**:297-312.
356. Mikami K, Iuchi S, Yamaguchi-Shinozaki K, Shinozaki K: **A novel *Arabidopsis thaliana* dynamin-like protein containing the pleckstrin homology domain.** *J Exp Bot* 2000, **51**:317-318.
357. Achiriloaie M, Barylko B, Albanesi JP: **Essential role of the dynamin pleckstrin homology domain in receptor-mediated endocytosis.** *Mol Cell Biol* 1999, **19**:1410-1415.
358. Narayanan R, Leonard M, Song BD, Schmid SL, Ramaswami M: **An internal GAP domain negatively regulates presynaptic dynamin in vivo: a two-step model for dynamin function.** *J Cell Biol* 2005, **169**:117-126.
359. Lee MH, Lee SH, Kim H, Jin JB, Kim DH, Hwang I: **A WD40 repeat protein, *Arabidopsis* Sec13 homolog 1, may play a role in vacuolar trafficking by controlling the membrane association of AtDRP2A.** *Mol Cells* 2006, **22**:210-219.
360. Mikami K, Takahashi S, Katagiri T, Yamaguchi-Shinozaki K, Shinozaki K: **Gene note. Isolation of an *Arabidopsis thaliana* cDNA encoding a pleckstrin homology domain protein, a putative homologue of human pleckstrin.** *J Exp Bot* 1999, **50**:729-730.
361. Edlich C, Stier G, Simon B, Sattler M, Muhle-Goll C: **Structure and phosphatidylinositol-(3,4)-bisphosphate binding of the C-terminal PH domain of human pleckstrin.** *Structure* 2005, **13**:277-286.
362. DiNitto JP, Delprato A, Gabe Lee MT, Cronin TC, Huang S, Guilherme A, Czech MP, Lambright DG: **Structural basis and mechanism of autoregulation in 3-phosphoinositide-dependent Grp1 family Arf GTPase exchange factors.** *Mol Cell* 2007, **28**:569-583.

363. Blomberg N, Baraldi E, Sattler M, Saraste M, Nilges M: **Structure of a PH domain from the *C. elegans* muscle protein UNC-89 suggests a novel function.** *Structure* 2000, **8**:1079-1087.
364. Anthony RG, Khan S, Costa J, Pais MS, Bogre L: **The Arabidopsis protein kinase PTI1-2 is activated by convergent phosphatidic acid and oxidative stress signaling pathways downstream of PDK1 and OXI1.** *J Biol Chem* 2006, **281**:37536-37546.
365. Anthony RG, Henriques R, Helfer A, Meszaros T, Rios G, Testerink C, Munnik T, Deak M, Koncz C, Bogre L: **A protein kinase target of a PDK1 signalling pathway is involved in root hair growth in Arabidopsis.** *Embo J* 2004, **23**:572-581.
366. Stenmark H, Aasland R, Driscoll PC: **The phosphatidylinositol 3-phosphate-binding FYVE finger.** *FEBS Lett* 2002, **513**:77-84.
367. Cooke FT: **Phosphatidylinositol 3,5-bisphosphate: metabolism and function.** *Arch Biochem Biophys* 2002, **407**:143-151.
368. Mueller-Roeber B, Pical C: **Inositol phospholipid metabolism in Arabidopsis. Characterized and putative isoforms of inositol phospholipid kinase and phosphoinositide-specific phospholipase C.** *Plant Physiol* 2002, **130**:22-46.
369. Otomo A, Hadano S, Okada T, Mizumura H, Kunita R, Nishijima H, Showguchi-Miyata J, Yanagisawa Y, Kohiki E, Suga E, et al: **ALS2, a novel guanine nucleotide exchange factor for the small GTPase Rab5, is implicated in endosomal dynamics.** *Hum Mol Genet* 2003, **12**:1671-1687.
370. Rosa JL, Casaroli-Marano RP, Buckler AJ, Vilaro S, Barbacid M: **p619, a giant protein related to the chromosome condensation regulator RCC1, stimulates guanine nucleotide exchange on ARF1 and Rab proteins.** *Embo J* 1996, **15**:4262-4273.
371. Renault L, Nassar N, Vetter I, Becker J, Klebe C, Roth M, Wittinghofer A: **The 1.7 Å crystal structure of the regulator of chromosome condensation (RCC1) reveals a seven-bladed propeller.** *Nature* 1998, **392**:97-101.
372. Nagashima T, Hayashi F, Yokoyama S: **FYVE domain of FYVE domain containing 19 protein from *Mus musculus*.** To be Published.
373. Tomizawa T, Kigawa T, Koshiba S, Inoue M, Yokoyama S: **Solution structure of the FYVE domain from human FYVE domain containing 27 isoform b protein.** To be Published.
374. Stenmark H, Aasland R: **FYVE-finger proteins--effectors of an inositol lipid.** *J Cell Sci* 1999, **112** (Pt 23):4175-4183.
375. Wallace AC, Laskowski RA, Thornton JM: **LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions.** *Protein Eng* 1995, **8**:127-134.
376. Hu Y, Chuang JZ, Xu K, McGraw TG, Sung CH: **SARA, a FYVE domain protein, affects Rab5-mediated endocytosis.** *J Cell Sci* 2002, **115**:4755-4763.
377. Panaretou C, Domin J, Cockcroft S, Waterfield MD: **Characterization of p150, an adaptor protein for the human phosphatidylinositol (PtdIns) 3-kinase. Substrate presentation by phosphatidylinositol transfer protein to the p150.PtdIns 3-kinase complex.** *J Biol Chem* 1997, **272**:2477-2485.

378. Hong Z, Verma DP: **A phosphatidylinositol 3-kinase is induced during soybean nodule organogenesis and is associated with membrane proliferation.** *Proc Natl Acad Sci U S A* 1994, **91**:9617-9621.
379. Joo JH, Yoo HJ, Hwang I, Lee JS, Nam KH, Bae YS: **Auxin-induced reactive oxygen species production requires the activation of phosphatidylinositol 3-kinase.** *FEBS Lett* 2005, **579**:1243-1248.
380. Peleg-Grossman S, Volpin H, Levine A: **Root hair curling and Rhizobium infection in *Medicago truncatula* are mediated by phosphatidylinositide-regulated endocytosis and reactive oxygen species.** *J Exp Bot* 2007, **58**:1637-1649.
381. Lee Y, Bak G, Choi Y, Chuang WI, Cho HT, Lee Y: **Roles of phosphatidylinositol 3-kinase in root hair growth.** *Plant Physiol* 2008, **147**:624-635.
382. Leshem Y, Seri L, Levine A: **Induction of phosphatidylinositol 3-kinase-mediated endocytosis by salt stress leads to intracellular production of reactive oxygen species and salt tolerance.** *Plant J* 2007, **51**:185-197.
383. Jung JY, Kim YW, Kwak JM, Hwang JU, Young J, Schroeder JI, Hwang I, Lee Y: **Phosphatidylinositol 3- and 4-phosphate are required for normal stomatal movements.** *Plant Cell* 2002, **14**:2399-2412.
384. Park KY, Jung JY, Park J, Hwang JU, Kim YW, Hwang I, Lee Y: **A role for phosphatidylinositol 3-phosphate in abscisic acid-induced reactive oxygen species generation in guard cells.** *Plant Physiol* 2003, **132**:92-98.
385. Shisheva A: **PIKfyve: the road to PtdIns 5-P and PtdIns 3,5-P(2).** *Cell Biol Int* 2001, **25**:1201-1206.
386. Sbrissa D, Ikononov OC, Shisheva A: **Phosphatidylinositol 3-phosphate-interacting domains in PIKfyve. Binding specificity and role in PIKfyve. Endomembrane localization.** *J Biol Chem* 2002, **277**:6073-6079.
387. Dove SK, Cooke FT, Douglas MR, Sayers LG, Parker PJ, Michell RH: **Osmotic stress activates phosphatidylinositol-3,5-bisphosphate synthesis.** *Nature* 1997, **390**:187-192.
388. Ridley SH, Ktistakis N, Davidson K, Anderson KE, Manifava M, Ellson CD, Lipp P, Bootman M, Coadwell J, Nazarian A, et al: **FENS-1 and DFCP1 are FYVE domain-containing proteins with distinct functions in the endosomal and Golgi compartments.** *J Cell Sci* 2001, **114**:3991-4000.
389. Blatner NR, Stahelin RV, Diraviyam K, Hawkins PT, Hong W, Murray D, Cho W: **The molecular basis of the differential subcellular localization of FYVE domains.** *J Biol Chem* 2004, **279**:53818-53827.
390. Panopoulou E, Gillooly DJ, Wrana JL, Zerial M, Stenmark H, Murphy C, Fotsis T: **Early endosomal regulation of Smad-dependent signaling in endothelial cells.** *J Biol Chem* 2002, **277**:18046-18052.
391. Itoh F, Divecha N, Brocks L, Oomen L, Janssen H, Calafat J, Itoh S, Dijke Pt P: **The FYVE domain in Smad anchor for receptor activation (SARA) is sufficient for localization of SARA in early endosomes and regulates TGF-beta/Smad signalling.** *Genes Cells* 2002, **7**:321-331.

392. Callaghan J, Simonsen A, Gaullier JM, Toh BH, Stenmark H: **The endosome fusion regulator early-endosomal autoantigen 1 (EEA1) is a dimer.** *Biochem J* 1999, **338** (Pt 2):539-543.
393. Qin BY, Lam SS, Correia JJ, Lin K: **Smad3 allostery links TGF-beta receptor kinase activation to transcriptional control.** *Genes Dev* 2002, **16**:1950-1963.