Structural basis of *Weak-similarity Guanine nucleotide Exchange Factor* (WGEF) activation in Wnt-planar cell polarity pathway

by

Omble Aishwarya Dilip
10BB18A26034

A thesis submitted to the
Academy of Scientific & Innovative Research
for the award of the degree of
DOCTOR OF PHILOSOPHY

in

SCIENCE

Under the supervision of

Dr. Kiran Kulkarni

CSIR-National Chemical Laboratory, Pune

Academy of Scientific and Innovative Research
AcSIR Headquarters, CSIR-HRDC campus
Sector 19, Kamla Nehru Nagar,
Ghaziabad, U.P. – 201 002, India

October, 2024

Certificate

This is to certify that the work incorporated in this Ph.D. thesis entitled, "Structural basis of *Weak-similarity Guanine nucleotide Exchange Factor* (WGEF) activation in Wnt-planar cell polarity pathway*",* submitted by Omble Aishwarya Dilip to the Academy of Scientific and Innovative Research (AcSIR) in fulfillment of the requirements for the award of the Degree of Doctor of Philosophy in Science, embodies original research work carried-out by the student. We, further certify that this work has not been submitted to any other University or Institution in part or full for the award of any degree or diploma. Research material(s) obtained from other source(s) and used in this research work has/have been duly acknowledged in the thesis. Image(s), illustration(s), figure(s), table(s) etc., used in the thesis from other source(s), have also been duly cited and acknowledged.

Omble Aishwarya Dilip
(Student)
Date: 30/10/2024

Dr. Kiran Kulkarni
(Supervisor)
Date: 30/10/2024

STATEMENTS OF ACADEMIC INTEGRITY

I Omble Aishwarya Dilip, a Ph.D. student of the Academy of Scientific and Innovative Research (AcSIR) with Registration No. 10BB18A26034 hereby undertake that, the thesis entitled "Structural basis of *Weak-similarity Guanine nucleotide Exchange Factor* (WGEF) activation in Wnt-planar cell polarity pathway" has been prepared by me and that the document reports original work carried out by me and is free of any plagiarism in compliance with the UGC Regulations on *"Promotion of Academic Integrity and Prevention of Plagiarism in Higher Educational Institutions (2018)"* and the CSIR Guidelines for "*Ethics in Research and in Governance (2020)".*

Signature of the Student
Date : 30/10/2024
Place : Pune

It is hereby certified that the work done by the student, under my supervision, is plagiarism-free in accordance with the UGC Regulations on *"Promotion of Academic Integrity and Prevention of Plagiarism in Higher Educational Institutions (2018)"* and the CSIR Guidelines for "*Ethics in Research and in Governance (2020)".*

Signature of the Supervisor
Name : Dr. Kiran Kulkarni
Date : 30/10/2024
Place : Pune

Acknowledgement

I want to take this opportunity to express my deepest gratitude to God and to all those who have supported and guided me throughout my PhD journey.

First and foremost, I sincerely thank my supervisor, Dr. Kiran Kulkarni, for his unwavering guidance and constant encouragement. His dedication, patience and expertise helped me grow as a researcher. He has taught me structural biology by simplifying various aspects and concepts, making them easier to understand. His enthusiasm for research and tackling challenging scientific problems has been a continuous source of inspiration throughout my journey. I am really thankful to him for believing in me and giving me the opportunity to pursue PhD under his guidance.

I am also deeply grateful to my Doctoral Advisory Committee members; Dr. Durba Sengupta, Dr. Mahesh Kulkarni and Dr. Kumar Vanka, for their helpful suggestions, discussions and guidance, which greatly enriched this thesis.

I am also thankful to the Council of Scientific & Industrial Research, India, for the fellowship, Director (NCL) and Head of Department (Biochemical Sciences Division) for allowing me to conduct my research and providing me with the necessary infrastructure and facilities required for that. I thank the Academy of Scientific & Innovative Research (AcSIR) for my PhD registration. I also acknowledge the administrative and academic staff of NCL for their timely help with my documentation process and other formalities.

I want to extend my sincere thanks to Prof. Dawid Igor for generously sharing various constructs of Dvl, Frizzled and WGEF. I also acknowledge the European Synchrotron Radiation Facility (ESRF), ID23-2 beamline, for the provision of synchrotron radiation facilities and express my gratitude to Dr. Christoph Mueller-Dieckmann for his assistance and support in using the beamline. Additionally, I appreciate the help of BL-21 PX beamline scientists Dr. Ravindra D. Makde, Dr. Ashwani Kumar and Dr. Biplab Ghosh at RRCAT, Indore, India, for helping with data collection. I thank the Central Instrumentation Facility (CIF), Savitribai Phule Pune University (SPPU), for the MST facility. My gratitude also goes to Dr. Durba Sengupta for her valuable inputs on MD simulations, Dr. Radha Chauhan from NCCS, Pune, for granting me access to the SEC-MALS facility and Jyotsna for helping me with the SEC-MALS experiments. I am also grateful to the NCCS-FACS facility for allowing me to perform FACS experiments and to Dr. Mahesh Kulkarni, for providing access to the

Mass spectrometry facility and reagents. I extend my thanks to Monika for her help with Mass spectrometry sample preparations and analysis.

Furthermore, I am thankful to Dr. Gayathri Pananghat and Dr. Saikrishnan Kayarat from IISER-Pune for allowing me to use their plate reader facility. I would like to acknowledge Dr. Ashok Giri, Dr. Narendra kadoo, Dr. Rakesh Joshi, Dr. Subashchandrabose Chinnathambi, Dr. Dhanasekaran Shanmugam, Dr. V.Koteswara Rao, Dr. Rahul Bhambure and Dr. Bhushan Chaudhari for allowing me to use their lab facilities.

I am deeply thankful to all my seniors Anand, Sneha, Ashwini, Zenia and Debopriya, as they have played a major role in teaching me various techniques and sharing their experience and knowledge in addressing the scientific challenges that I came across during my PhD. I am also grateful to my present labmates, Ravi, Neeraj, Pratiksha, Ashwini, Prajakta and Shrutika, for their scientific inputs, assistance with the experiments, and for making this journey more pleasant. I would like to take this opportunity to thank my former labmates, Disha, Simerdeep, Raveena, Shreegauri, Meghana and my trainees for creating a cheerful and exciting lab environment that made conducting research truly enjoyable.

Now, I would like to acknowledge my incredible friends who have made NCL feel like a second home to me. A special thanks to Smita, Kritika, Kranti, Rajnigandha, Hari, Ashish and Soumya for all the support and care they have provided. I am also particularly grateful to Pratiksha and Ravi for being around when needed. Further, extending my gratitude to Pranav, Vinay, Aashima, Vibhavari, Aakash, Amit, Kush, Suryadev, Kailash, Jayram, Mahendra, Pavan, Ravi Ranjan and Ravi Kumar for all the timely breaks of fun. I truly cherish all the memories created with these wonderful people.

A heartfelt thank you to my parents and my brother for believing in me and for providing constant support throughout the ups and downs of my journey. I truly appreciate all the sacrifices you have made to ensure that my education went smoothly. I would also like to express my deep gratitude to my husband, Rakesh, not only for his support during this journey but also for constantly motivating me to stay strong and become a better version of myself. I am equally thankful to my in-laws for supporting me, just as my parents in completing my degree. Each of you has believed in me more than I have believed in myself, and that has always been a source of motivation for me to keep moving forward.

- ***Aishwarya Omble***

Table of Contents

List of Figures

List of Tables

Abbreviations

GDP	Guanosine diphosphate
GTP	Guanosine triphosphate
TE buffer	Tris-EDTA buffer
NaOH	Sodium hydroxide
SDS	Sodium Dodecyl Sulfate
DNA	Deoxyribonucleic acid
PCR	Polymerase chain reaction
EDTA	Ethylenediaminetetraacetic acid
Mant-GDP	N-Methylanthraniloyl- Guanosine diphosphate
PMSF	Phenylmethylsulfonyl fluoride
TEV	Tobacco etch virus
CMV	Cytomegalovirus
CHAPS	3-[(3-cholamidopropyl)dimethylammonio]propane-1-sulfonate
FC-12	Fos-Choline-12
LDAO	Lauryldimethylamine oxide
DM	Decyl β-D-maltopyranoside
RPM	Revolutions per minute
PEG	Polyethylene glycol
Ni^{2+}-NTA	Nickel-Nitriloacetic acid
GST	Glutathione-S-transferase
$MgCl_2$	Magnesium chloride
DTT	Dithiothreitol
NaCl	Sodium chloride
kDa	kilodalton
CHES	N-cyclohexyl-2-aminoethanesulfonic acid
β-met	β-mercaptoethanol
K	Kelvin
h	Hours
min	Minutes
sec	Seconds
RT	Room temperature
SEC	Size exclusion chromatography
IEC	Ion-exchange chromatography

PDB	Protein Data Bank
HEK	Human embryonic kidney
SD	Standard deviation
YFP	Yellow fluorescent protein
DMEM	Dulbecco's Modified Eagle Medium
FBS	Fetal bovine serum
PEI	Polyethylenimine
ATCC	American Type Culture Collection
PBS	Phosphate-buffered saline
DAPI	4′,6-diamidino-2-phenylindole
IPTG	Isopropyl ß-D-1-thiogalactopyranoside
dNTP	deoxynucleotide triphosphate
C-terminal	Carboxy-terminal
N-terminal	Amino-terminal
DCCM	Dynamical cross-correlation matrix
Å	Angstrom
His tag	Histidine tag
nL	Nanoliter
℃	Degree celsius
µM	Micromolar
µg and mg	Microgram and milligram
µL and mL	Microliter and Milliliter
RMSD	Root mean square deviation
IC	Independent component
kb	kilobase
µm	Micrometer

Chapter 1

Introduction

Cellular signalling involves the generation, propagation and processing of information critical for cellular functions. Processing of biological information by employing signalling events is ubiquitous across the genre of life. However, the complexity of signalling events in multicellular organisms is significantly greater than those operating in unicellular organisms. Cells use a myriad spectrum of molecules, such as proteins, sugars, lipids and ions, to exert these signalling events. In higher organisms, intricate interlacing of the signalling pathways adds to their complexity. These signalling cascades control a wide array of cellular activities, including gene expression, metabolism, cell growth and differentiation. In multicellular organisms, signalling plays critical roles in maintaining homeostasis, regulating immune responses and for developmental processes. In higher eukaryotes, Wnt signalling is known for its critical role in regulating of numerous cellular functions such as cell polarization, differentiation, proliferation and migration during the developmental stages of the embryo as well as in tissue homeostasis in adult organisms. This pathway is mediated by Wnts ligands, which are lipidated, secreted glycoproteins. In humans, the Wnt family is shown to be composed of 19 members that activate various intracellular signalling cascades by binding to their membrane bound receptors, Frizzled. Frizzled are the seven transmembrane helical receptors that belongs to the family of G protein-coupled receptors (GPCRs)[1,2]. Apart from Frizzled receptors, Wnts are also shown to signal via heterodimeric complexes consisting of Frizzled receptors and other co-receptors including *Lipoprotein-receptor related protein 5 and 6* (LRP5 and LRP6), *Tyrosine kinase-related* (RYK) receptors and *Receptor tyrosine kinase-like orphan receptors 1/2* (ROR1/2) to transmit the signalling cues to the downstream effectors[3]. The mechanism of selectivity between Frizzled receptors and Wnt ligands and the subsequent modes of signalling cascade they exert remains unclear. Notably, this selectivity is also regulated by the co-receptor involved in the signalling event[4,5]. The Wnt signalling pathway is primarily divided into two categories: the β-catenin dependent canonical pathway and the β-catenin independent non-canonical pathway.

1.1 Canonical Wnt signalling

Canonical Wnt signalling is majorly involved in cellular differentiation and proliferation. In the absence of Wnt signalling, a β-catenin destruction complex consisting of cytoplasmic proteins such as *Adenomatous polyposis coli* (APC) protein, Casein kinase 1 (CK1), *Ser/Thr kinases glycogen synthase kinase-3 beta* (GSK-3β), Axin, *E3-ubiquitin ligase β-TrCP*, and *Protein phosphatase 2A* (PP2A) is formed. It has been shown that this destruction complex phosphorylates β-catenin, leading to its ubiquitination and breakdown by proteasome mediated degradation (Fig 1.1A). Once the Wnt ligand interacts with the corresponding Frizzled (FZD) receptors and co-receptors (LRP5/LRP6), the former recruits a scaffold protein called Dishevelled (Dvl) at its cytosolic domains. Dvl further interacts with the destruction complex proteins such as Axin and GSK-3β, thereby dissociating the destruction complex[3,6,7]. This event stabilizes the β-catenin in the cytosol, which further translocate to the nucleus, where it interacts with the *T-cell factor/ Lymphoid enhancer factor* (TCF/LEF) transcription factors and regulates expression of gene under their control (Fig 1.1B).

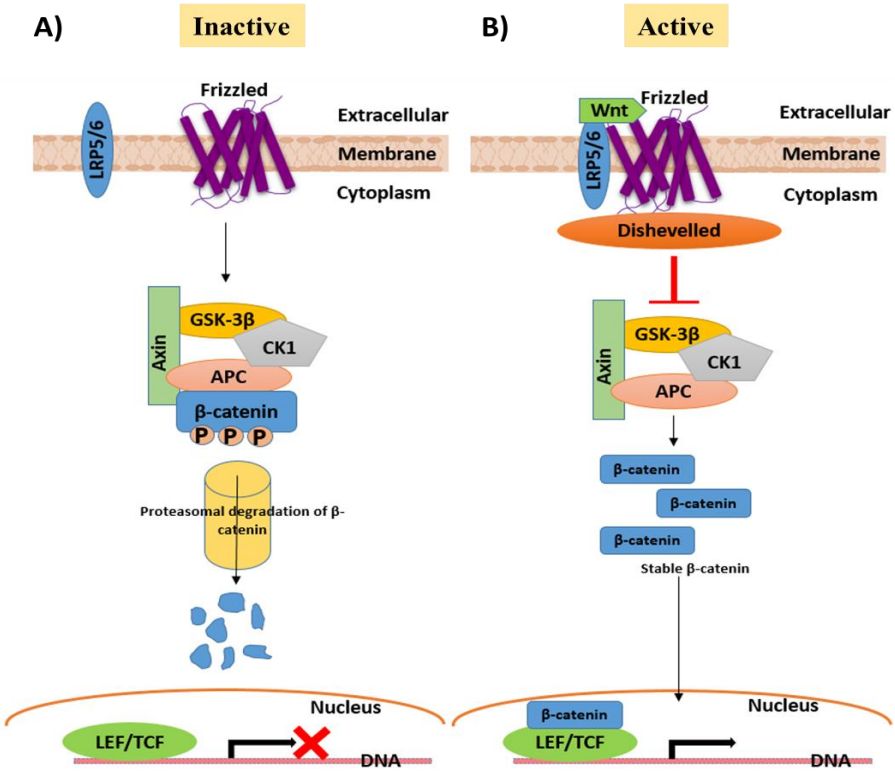


Figure 1.1: Overview of β-catenin dependent canonical Wnt signalling. A) *Wnt signalling in the absence of Wnt ligand.* ***B)*** *Wnt signalling in the presence of Wnt ligand.*

This signalling plays a vital role in various biological processes involving different stages of development and tissue maintenance. During mouse embryogenesis, it is shown to be involved in locally triggering the epithelial to mesenchymal cell transition and mesendodermal differentiation[8]. In adult tissues, particularly in organs requiring constant renewal like the intestine, stomach and skin, Wnt signalling is important for stem cell proliferation and tissue regeneration. For example, Wnt signalling in the intestinal epithelium brings about continuous regeneration of intestinal tissue by promoting the proliferation of the local stem cells[9,10]. Similarly, Wnts, in their canonical mode of operation, are also shown to promote the regeneration of interfollicular epidermis[11,12] and proliferation of pancreatic beta cells[13]. Furthermore, Wnt signalling also controls hematopoiesis, emphasizing its importance in the maintenance of the hematopoietic system[14].

1.2. Non-canonical Wnt signalling

On the contrary to its canonical counterpart, non-canonical Wnt signalling operates independently of β-catenin stabilization and it is subsequently divided into the Wnt-calcium pathway and the Wnt-planar cell polarity (Wnt-PCP) pathway. This mode of Wnt signalling is shown to be involved in cell migration and polarization. Along with the Frizzled receptors (FZD), these Wnt ligands are also reported to interact with co-receptors such as ROR1/2, RYK receptors, Syndecan, Glypican and *Protein tyrosine kinase 7* (PTK7) receptor to transmit signals intracellularly[15].

The Wnt-calcium pathway involves recruitment of Dishevelled (Dvl) as well as heterotrimeric G-protein activation by the Wnt receptors, Frizzled. This pathway activates Phospholipase C (PLC), which converts phosphatidylinositol 4,5-bisphosphate (PIP2) into inositol 1,4,5-triphosphate (IP3) and 1,2 diacylglycerol (DAG) which acts as signalling molecules. IP3 releases Ca^{2+} from the endoplasmic reticulum (ER), further activating calcium regulated proteins like *Calmodulin-dependent protein kinase II* (CaMKII)[16], Protein Kinase C (PKC)[17] and Calcineurin[18] in *Xenopus*. Further, CaMKII is reported to activate regulatory proteins such as NF-κB[19] and CREB[20]. Similarly, Calcineurin in *Xenopus* is shown to activate the *nuclear factor associated with T cells* (NFAT)[18], another transcription factor. These transcription factors, NF-κB, CREB and NFAT, transcribe the regulatory genes by their nuclear translocation (Fig 1.2A). Activation of another calcium sensitive protein, PKC is reported to activate Cdc42 and to control tissue separation during gastrulation in *Xenopus*[21,22].

Non-canonical Wnt-PCP signalling is stimulated when the appropriate Wnt ligand interacts with its cognate FZD receptor. This further recruits Dishevelled (Dvl) to the receptor and activates it to form a complex with *Dishevelled-associated activator of morphogenesis 1* (Daam1)[23]. Dvl-Daam1 complex is reported to subsequently interact with *Weak-similarity Guanine Nucleotide Exchange Factor* (WGEF), which activates RhoA, leading to cell movement during *Xenopus* gastrulation[24]. Apart from RhoA, Dvl is also shown to activate another Rho GTPase, Rac, which is independent of Daam1 interaction. Dvl activated Rac is reported to stimulate *c-Jun-N-terminal Kinase* (JNK), which eventually leads to the regulation of cell polarity and migration[25] (Fig 1.2B). During embryonic development, Wnt-PCP signalling is required for cellular movement during neural crest closure and gastrulation[26]. In higher eukaryotes, the Wnt-PCP pathway is also known to regulate hair follicle orientation in mammalian skin[27] and is crucial for neural tube closure. Disruption of Wnt-PCP proteins has been linked to convergent extension defects in the mouse kidney, cochlea and heart[28–30].

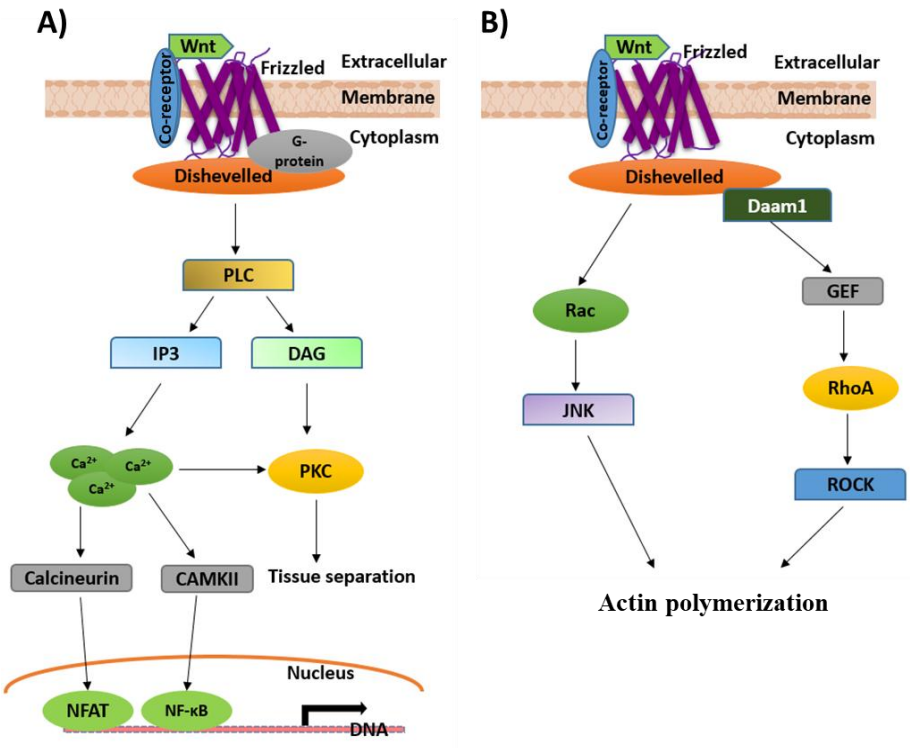


Figure 1.2: Overview of β-catenin independent non-canonical Wnt signalling. A) *Wnt-calcium pathway.* ***B)*** *Wnt-planar cell polarity (Wnt-PCP) pathway.*

In the Wnt-PCP pathway, *Prickle* protein is known to inhibit this pathway in the absence of Wnt ligand. The binding of Wnt ligand to the receptor recruits proteins such as *Inversion*, Dvl2, Par-6 polarity protein and *smad ubiquitination regulatory factor* (smurf),

which ubiquitinates *Prickle* leading to its proteasomal degradation and activation of Wnt-PCP signalling. This indicates that the Wnt-PCP pathway is also regulated, like canonical Wnt signalling. Celsr1 and Vangl2 are the other known receptors that also interact with Wnt ligands.

1.3 Wnt ligands, Receptors and Co-receptors in Wnt signalling

1.3.1 Wnt ligands

The word 'Wnt' is derived from the *wingless* (wg) gene, essential for the correct development of most of the appendages and tissue in Drosophila[31,32]. In humans there exist nineteen different genes that encode Wnt ligands. Wnts are cysteine-rich proteins consisting of approximately 22 cysteine residues that form disulphide bonds required for the proper folding of the proteins. After translation, Wnt proteins are targeted to the ER and Golgi apparatus for post-translational modifications which include acylation and glycosylation. The amount of glycosylation on Wnts varies significantly across the family. On post-translational modifications, these proteins are secreted out of the cells through a secretory pathway[33]. Previously, they were classified as canonical and non-canonical Wnts based on the pathways they majorly activate. For instance, Wnt1, Wnt2, Wnt3, and Wnt8A are some of the Wnts involved in canonical Wnt signalling[34] whereas Wnt4, Wnt5A/B, Wnt6, Wnt7A, and Wnt11, are known to activate the non-canonical Wnt signalling[35]. It is to be noted that there is no one-to-one correspondence between the Wnt ligands and the type of pathway they exert.

1.3.2 Frizzled, the non-canonical GPCRs

Frizzled (FZD) are seven transmembrane helical receptors of Wnt ligands and belong to the class F of the GPCR family. They are also referred to as atypical or non-canonical GPCRs, as their primary signalling does not involve canonical G-proteins[36]. GPCRs encompass a wide range of membrane receptors that regulate crucial physiological functions, making 50% of these proteins important drug targets. Although they vary in sequence and functions, the seven transmembrane helical architecture of GPCRs is highly conserved across the family[37]. The class F of GPCRs consists of ten FZD and one Smoothened (SMO) receptor. They are known to possess an extracellular cysteine-rich domain (CRD), which mediates binding with the Wnt ligands[38]. Based on their sequence FZD have been grouped into four clusters: FZD1, 2 and 7 share approximately 75% sequence identity; FZD5 and 8 share about 70% sequence identity; FZD4, 9 and 10 exhibit about 65% sequence identity; whereas FZD3 and 6 share 50% sequence

identity[38]. The specific downstream Wnt signalling is activated depending on the interaction between the corresponding Wnt ligand, FZD receptor and co-receptor[39]. Apart from these interactions, the linker domain between the CRD and the transmembrane domain of FZD receptors also plays a role in determining the specificity of Wnt-FZD interactions[40].

1.3.3 Wnt ligand co-receptors

In addition to FZDs, Wnt ligands interact with co-receptors such as LRP5 and LRP6[41]. These single-span non-FZD transmembrane proteins heterodimerize with FZD receptors upon Wnt binding. Phosphorylation of LRP receptors by various protein kinases operating in the cytoplasmic region recruits Axin, facilitating the activation of Wnt signalling[42]. Other receptors of Wnts include RYK receptors, which consist an extracellular *Wnt inhibitory factor* (WIF) domain that interacts with the Wnts to form a Frizzled-Wnt-RYK ternary complex[43]. Additionally, the ROR1/2 also serve as co-receptors of Wnts. These receptors have an extracellular CRD domain similar to that of FZD receptors, a single-pass transmembrane domain, intracellular kinase domain and a proline-rich domain[44,45]. These proteins also form heterodimers with FZD receptors upon Wnt binding to initiate the signal transduction[46].

1.4 Recruitment of Dishevelled and Daam1 in non-canonical Wnt-PCP pathway

1.4.1 Recruitment of adapter protein Dishevelled

The recruitment of Dishevelled (Dvl) at the cytosolic end of the FZDs is a key event in Wnt signalling as it plays a crucial role in canonical as well as in non-canonical Wnt signalling. Based on the specific Wnt-FZD interaction, the Wnt signalling diverges from the point of Dishevelled recruitment. In canonical Wnt pathway it is known to prevent β-catenin from proteasomal degradation, whereas in non-canonical Wnt-PCP pathway it forms complex with Daam1 to activate RhoA. It can also activate Rac, independent of Daam1 interaction[47]. In vertebrates, Dvl is known to possess three isoforms (Dvl1, Dvl2 and Dvl3) and all of these comprise of three conserved domains, namely, an N-terminal DIX (Dishevelled-Axin) domain, a central PDZ (PSD-95, DLG, ZO1) domain, and a C-terminal DEP (Dishevelled, EGL-10, Pleckstrin) domain (Fig 1.3). These domains play a vital role in signal transmission during Wnt signalling. The N-terminal DIX domain brings about dynamic polymerization to form Wnt signalosomes and also promotes protein-protein interaction. In canonical Wnt signalling, the

DIX domain of Dvl interacts with Axin, which leads to the disruption of the β-catenin destruction complex and stabilizes the β-catenin to activate Wnt signalling[48,49]. The central PDZ domain plays an crucial role in establishing protein-protein interactions during canonical and non-canonical Wnt signalling[50,51]. For instance, Dvl^{PDZ} domain is reported to bind with the KTXXXW internal peptide motif of Frizzled receptors and also interacts with Daple, which further mediates Dvl phosphorylation to regulate both canonical and non-canonical Wnt signalling (involving Rac GTPase)[52,53]. On the other hand, the DEP domain present at the C-terminal contributes to the assembly of signalosomes[54,55] (Fig 1.3). The C-terminal DEP domain is responsible for interaction with Frizzled receptors for Dvl recruitment and also dimerizes to cross-link Dvl polymers, contributing to the assembly of signalosome during Wnt signalling[55]. Additionally, Dvl possesses other regions, such as a basic region of positive amino acids between the DIX-PDZ domain, a nuclear import signal and a proline-rich region between the PDZ-DEP domains and a C-terminal nuclear export signal. These regions are shown to assist during interactions with other proteins, signal transduction and cellular localization of Dvl[47] (Fig 1.3). For instance, phosphorylation of the basic region is linked to interaction of proteins such as *Naked Cuticle* which act as antagonist of canonical Wnt signalling. This region is also important for interaction of Dvl with membrane lipids. The proline-rich region of Dvl1 interacts with NEDL1 ubiquitin ligase, which further leads to the degradation of Dvl1[56]. Thus, these interactions enable Dvl to coordinate complex signalling events involving regulation of both canonical and non-canonical signal transduction.

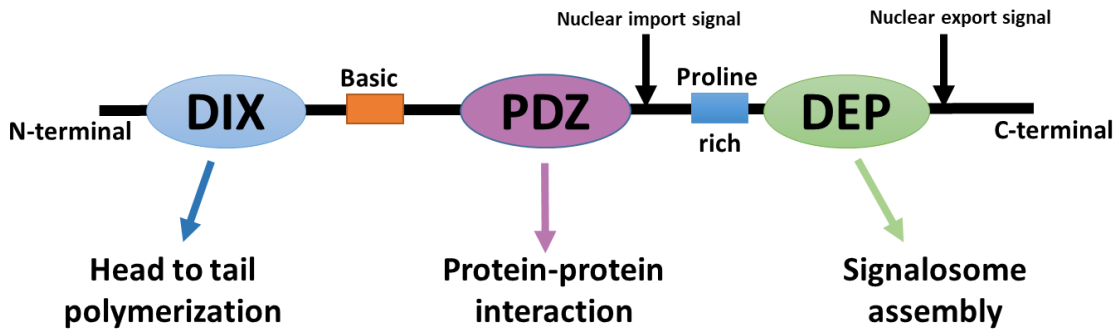


Figure 1.3: Different domains of Dishevelled (Dvl) proteins and their respective functions during signalling.

1.4.2 Daam1, another critical player in non-canonical Wnt signalling

Daam1 belongs to the group of formin proteins and consists of four domains: GTPase binding domain (GBD), the formin homology 1 and 2 (FH1 and FH2) domains, and the diaphanous autoinhibitory domain (DAD). Daam1 is autoinhibited through the intradomain interactions between its N-terminal GBD and the C-terminal DAD domains. Dvl binds with the DAD domain of Daam1 to release its autoinhibition. Rho GTPase is also known to bind the GBD domain to partially release Daam1 from an autoinhibitory state[57]. Thus, recruitment of the Dvl-Daam1 complex is an important step in the activation of RhoA dependent cytoskeletal rearrangement in the Wnt-PCP pathway[58,59].

1.5 Rho GTPases in non-canonical Wnt signalling

1.5.1 Rho family of GTPases

The Rho (RAS homolog) family of GTPases belongs to the RAS superfamily of Guanine nucleotide binding small GTPases (~21 to 25 kDa). GTPases such as Rho (A-D, G and H), Rac (1-3), Cdc42 and RND (1-3) belong to the Rho family of GTPases[60]. These GTPases are crucial for several cellular processes, such as cell migration, cell polarity establishment, cytoskeleton rearrangement, etc. Like other GTPases, these proteins act as biomolecular switches that cycle between the GTP-bound "ON" (active) and the GDP-bound "OFF" (inactive) states. The most conserved structural motifs of these proteins include G1 (phosphate-binding loop, P-loop), G2 (Switch I region/SWI), G3 (Switch II region/SWII), G4 and G5 motifs, important for guanine nucleotide binding. The G1 motif consists of a conserved sequence GxxxxGK[S/T], where x denotes any amino acid. This motif binds to the β-phosphate of the guanine nucleotide as well as to the Mg^{2+} ion (Fig 1.4).

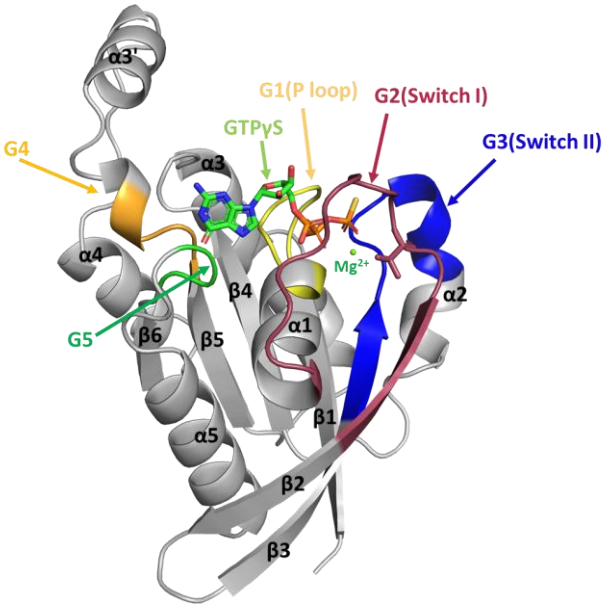


Figure 1.4: Crystal structure of RhoA bound with GTP analogue (GTPγS) (PDB ID: 1A2B). *Conserved motifs of RhoA are represented with different colours; G1 (P loop): Yellow, G2 (SWI): Bordeaux, G3 (SWII): Blue, G4: Dark yellow, G5: Green.*

The threonine (Thr37 in RhoA) residue of the SWI region forms a hydrogen bond with the GTP γ-phosphate as well as with the Mg^{2+} ion through its side chain (Fig 1.5). Similarly, the glycine (Gly62 in RhoA) residue of the SWII region, which is part of the conserved DxxG (x is any amino acid) motif, also establishes a hydrogen bond with the GTP γ-phosphate. Thus, conformational reorganization of SWI and SWII regions on account of these hydrogen bonds with the γ-phosphate was asserted to be a signature of activation of Ras superfamily of proteins, known as the "loaded spring mechanism"[61,62]. After GTPase-driven hydrolysis of the GTP γ-phosphate, GTP is converted to GDP and the two hydrogen bonds are lost, thereby relaxing the GTPase to GDP specific state (Fig 1.6A and B). Thus, this transition between the GTP to GDP-bound state is accompanied by a structural change in the SWI and SWII regions of the GTPases (Fig 1.6C and D). The G4 (NKxD) motif and G5 (TSAK) motifs are shown to make interactions with the guanine nucleotide base. The aspartate side chain of the DxxG motif is responsible for binding specificity towards guanine nucleotide base[61,62].

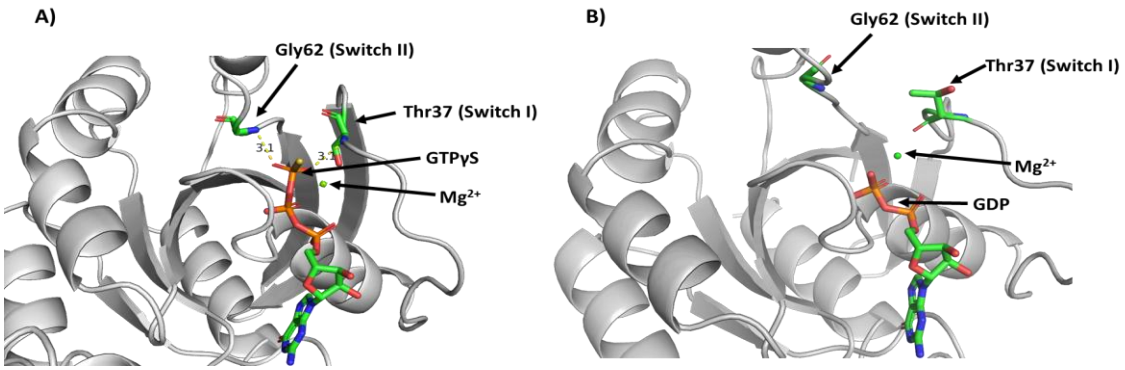


Figure 1.5: Crystal structure of RhoA in active (PDB ID: 1A2B) and inactive state (PDB ID: 1FTN). A) *RhoA bound to GTP analogue. Thr37 and Gly62 form bonds with the γ-phosphate of the bound GTPγS.* ***B)*** *RhoA bound to GDP. The bonds involving Thr37 and Gly62 are disrupted after GTPase-driven hydrolysis of the γ-phosphate.*

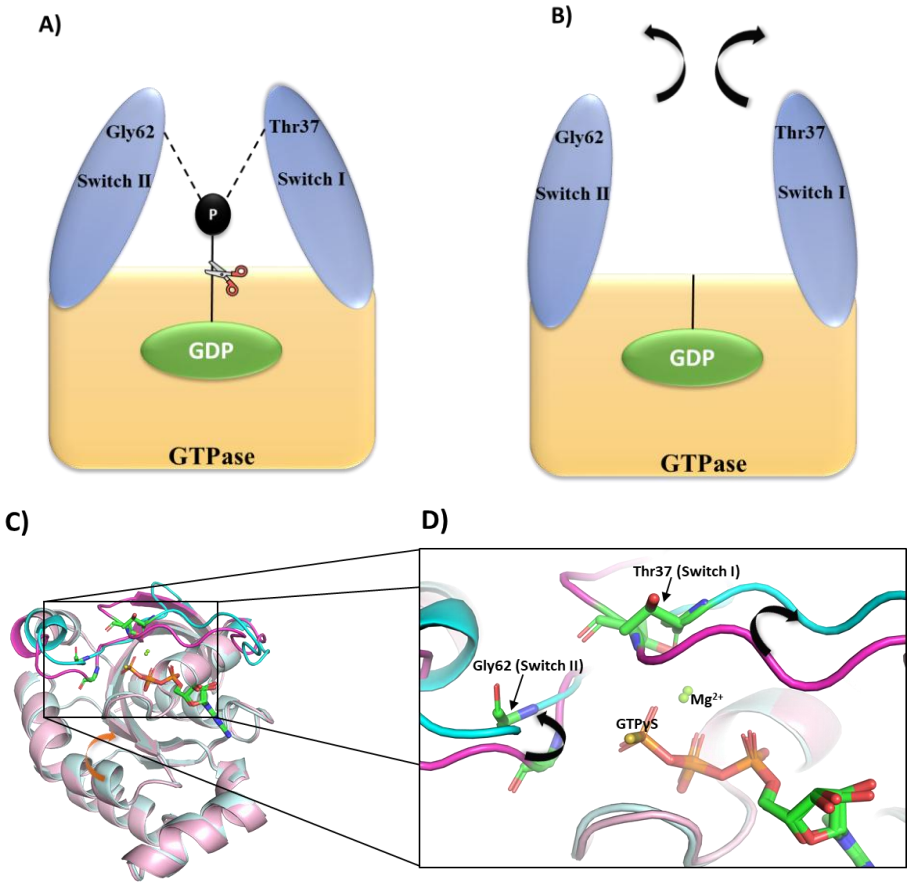


Figure 1.6: Loaded spring mechanism and switching mechanism of GTPases. A) *In the active state, the conserved Thr in the SWI region and Gly in the SWII region form hydrogen bonds with the γ-phosphate, a conformation referred to as "loaded spring mechanism".* ***B)*** *After the γ-phosphate is cleaved, the structure is relaxed.* ***C and D)*** *Superimposition of the active (1A2B-Pink colour) and inactive (1FTN- Teal colour) structures of RhoA shows a shift in the SWI and SWII regions after GTP hydrolysis.*

1.5.2 Rho pathway meeting the Wnts

Rho GTPases are recognized as key downstream regulators in non-canonical Wnt signalling. Operating under this pathway, they control actions that drive cytoskeletal dynamics for cell movements during gastrulation[63]. Amongst the Rho family, RhoA and Rac have been shown to be activated by Wnt11 and Frizzled7 (FZD7) interaction in the Wnt-PCP signalling. As observed from the majority of Wnt-FZD pathways, Wnt11-FZD7 interaction recruits the adapter protein, Dvl, which relays the signal to the effector proteins. The Wnt11-FZD7 signalling cascade that activates RhoA, depends on the interaction between Dvl, Daam1 and a Rho specific GEF known as WGEF[23,24]. However, activation of Rac under the same pathway is independent of Dvl interaction with Daam1 and WGEF[25]. Thus, RhoA and Rac activations are proposed to be mutually exclusive during the convergent extension movements. Furthermore, Wnt11-xFZD7 interaction is also reported to activate another Rho family member, Cdc42, by the Gβγ subunit of the heterotrimeric G-protein, which acts through Protein kinase C (PKC) during *Xenopus* gastrulation[21]. It is also suggested that Cdc42 regulates convergent extension movements through the activation of PKC, brought about by the Wnt/Ca^{2+} pathway[64].

1.6 Regulators of Rho GTPases

The switch between GTP-bound active and GDP-bound inactive conformation of the GTPases is controlled by two groups of regulatory proteins known as *Guanine nucleotide exchange factors* (GEFs) and *GTPase activating proteins* (GAPs). For Rho GTPases, there is an additional level of regulation, which is exerted by *Guanine nucleotide dissociation inhibitors* (GDIs) (Fig 1.7).

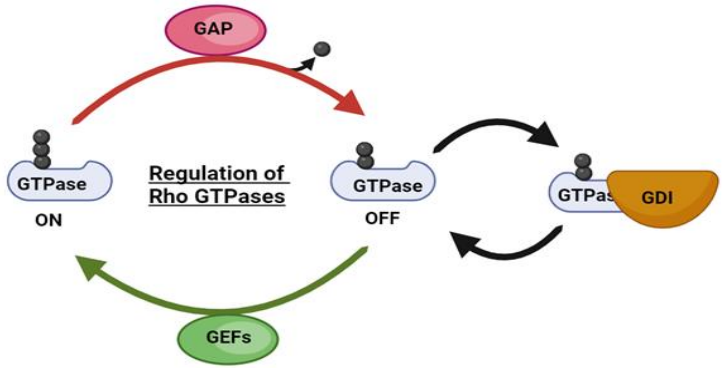


Figure 1.7: Rho GTPase regulation cycle. *Made with biorender.*

1.6.1 Guanine nucleotide exchange factors (GEFs)

GEFs are regulatory proteins which activates Rho GTPases by facilitating the exchange of their bound GDP with GTP (Fig 1.7). The intrinsic dissociation of GDP after GTP hydrolysis is slow as GTPases have a high affinity for the guanine nucleotides (GDP/GTP). GEFs bring about conformational change in the SWI and SWII region of the Rho GTPase, leading to the reduced affinity of GTPase for GDP/GTP, eventually leading to the dissociation of GDP[65]. Subsequently, GEF forms an intermediate complex with nucleotide-free GTPase (Fig 1.8), which further binds immediately to GTP, as the cytoplasmic concentration of GTP is tenfold excess than GDP. The final step in the GTPase activation involves dissociation of the GTP bound GTPase from the GEF. Thus, GEFs activate GTPases by exchanging their bound GDP with the cytosolic GTP[62,66].

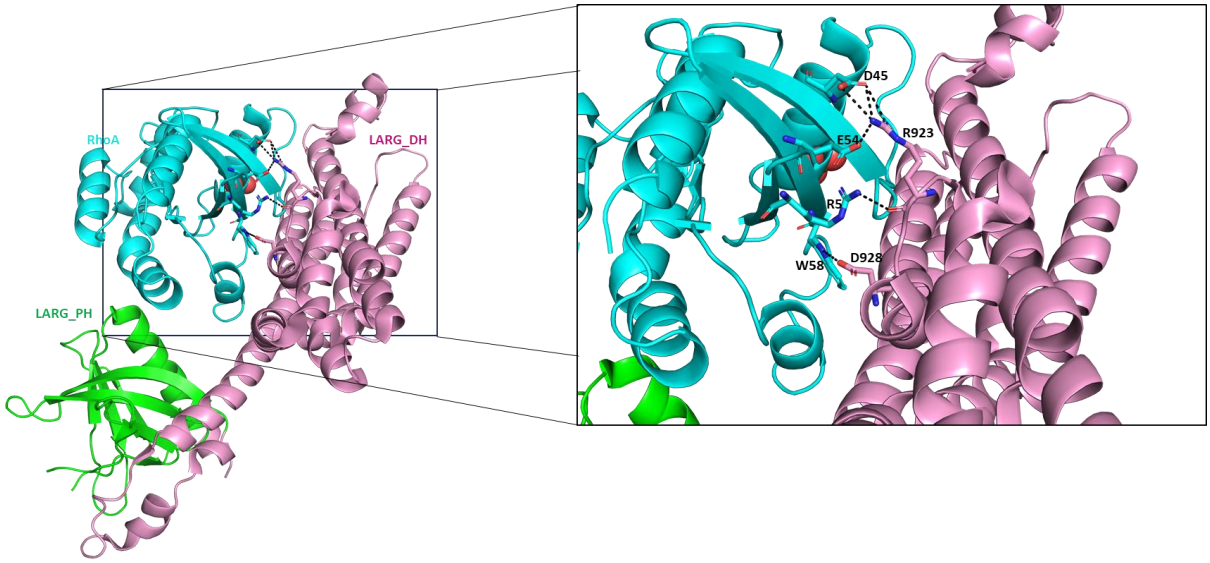


Figure 1.8: Crystal structure of RhoA in complex with Leukaemia-associated Rho GEF (LARG) (PDB ID: 1X86). *Diffuse B-cell lymphoma (Dbl) homology/DH (pink) and pleckstrin homology/PH (green) domain containing LARG is shown to form complex with RhoA (blue). Different residues of the catalytic DH domain, which form salt-bridges and hydrogen bonds to interact with RhoA are shown as black dotted lines*[67].

1.6.2 GTPase activating proteins (GAPs)

The intrinsic hydrolysis of the GTP by the GTPase is slow but can be significantly accelerated from their interaction with Rho GAPs (Fig 1.7). Apart from the catalytic (GAP) domain, GAPs comprise other functional domains such as PH and SH3 domains[60]. Despite more than three decades of structural and biochemical studies, the GAP catalyzed GTPase hydrolysis

mechanism is still not clear. From the structures of Rho GAP in complex with RhoA-GDP·AlF4$^-$ (tetrafluoroaluminate) and other GAP bound GTPase structures (Fig 1.9), the most plausible mechanism for the hydrolysis of GTP has been proposed[60,68]. The catalytic mechanism involves the reorientation of glutamine residue at the 63rd position of the GTPase (RhoA) and the conserved arginine of the GAP domain (also known as arginine finger) into the nucleotide binding pocket of the former[68] (Fig 1.9). This new favourable position of Q63 makes it a potent nucleophile, which subsequently activates the nearby bound water molecule to initiate the cleavage of γ-phosphate. The arginine finger stabilizes the transition state. Thus, a coordinated action of these two residues triggers the hydrolysis of GTP to GDP. It is interesting to note that till date, there are only 66 GAPs identified in humans, whereas there are approximately 85 canonical GEFs known[60]. Therefore these numbers show that GAPs may have higher promiscuity towards their ligands (GTPases) compared to GEFs.

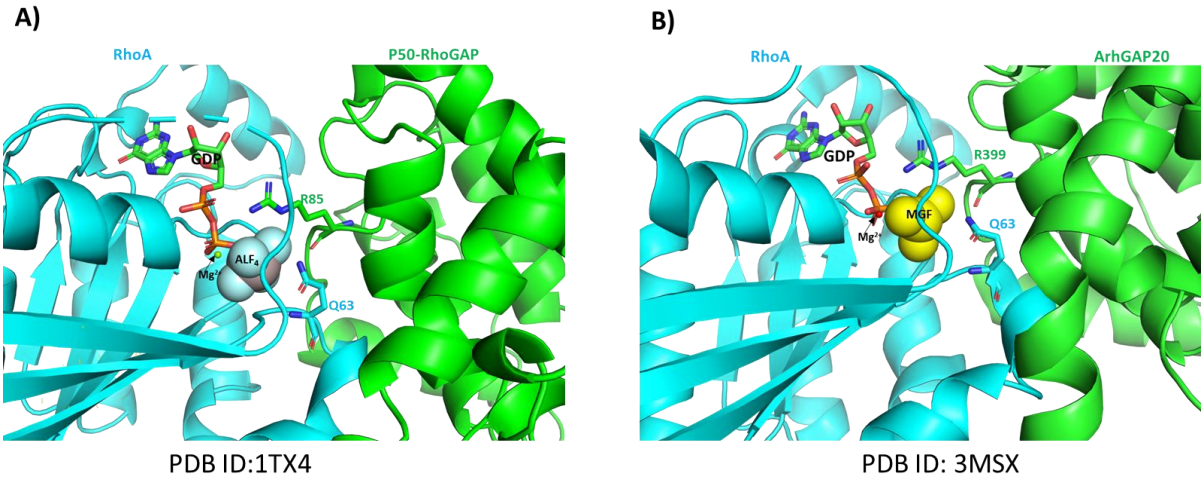


Figure 1.9: Conserved mechanism of GAP mediated GTP hydrolysis by RhoA GTPase. A) *Crystal structure of P50-RhoGAP bound to RhoA-GDP·AlF4$^-$ (PDB IB: 1TX4).* ***B)*** *Crystal structure of ArhGAP20 bound to RhoA-GDP.MgF3 (PDB IB: 3MSX). From these crystal structures, it is observed that the conserved arginine of the GAP domain (R399 in ArhGAP20 and R85 in P50-RhoGAP) facilitates the stabilization of GTPase transition state and promotes the GTP hydrolysis activity of the GTPase.*

1.6.3 Guanine nucleotide dissociation inhibitors (GDIs)

As explained previously, GDIs provide an additional level of regulation that is seen exclusively for Rho GTPases. GDIs sequester the GDP-bound Rho GTPases from the cytosol, thereby preventing their degradation and/or unwanted activation[60]. Firstly, GDIs interact with Rho GTPases through their N-terminal arm, which binds to the switch regions of Rho GTPase. This

step prevents the hydrolysis of GTP as well as dissociation of GDP from the GTPase. Next, the N-terminal of GDIs interacts with the hypervariable region of GTPase, which is known to interact with the phospholipid in the membrane. After this interaction, the hydrophobic moiety of Rho GTPase is inserted into the hydrophobic pocket of the GDI protein, thus releasing the GTPase from the membrane[69] (Fig 1.10).

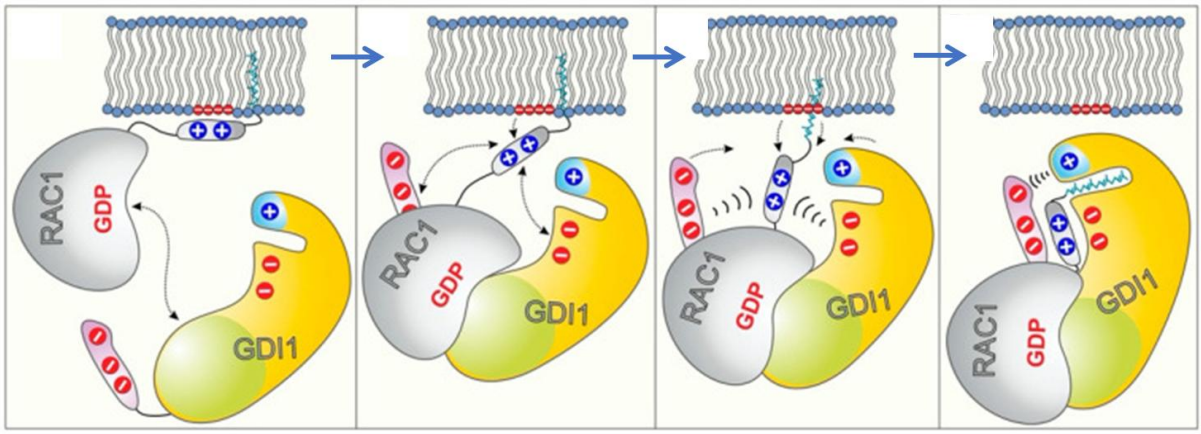


Figure 1.10: GDI1 mediated dissociation of Rac1 GTPase from membrane. *Image adapted from Mosaddeghzadeh et al.*[69].

1.7 Autoinhibition of GEFs

In terms of hierarchy of regulation, GEFs, too, require activation so that they can act upon the downstream GTPases to further propagate the signal. This is achieved by releasing GEFs from their autoinhibitory state on account of their inter-protein and other interactions. In higher eukaryotes, such as humans, two clades of GEFs are known: the canonical (DH-PH)/ Diffuse B-cell lymphoma (Dbl) GEFs and the non-canonical or Docker GEFs (Table 1.1). The canonical GEFs essentially consists of the catalytic Diffuse B-cell lymphoma (Dbl) homology (DH) domain and the membrane anchoring pleckstrin homology (PH) domain[70,71]. On the contrary Docker GEFs have DOCK homology regions-2 and 1 (DHR-2 and DHR-1) domains known to exert similar functions when compared to that of DH and PH domains, respectively[72–74]. Apart from the DH and PH domains, Dbl family of GEFs may also contain additional domains such as Src homology 2 (SH2), Regulator of G protein signalling (RGS), Src homology 3 (SH3), Ras binding domains (RBD) and PSD-95, DLG, ZO1 domains (PDZ) which assists in protein-protein binding and regulating the activity[60,75]. The interdomain interaction induced autoinhibitory state renders the GEFs inactive by precluding their active sites for GTPases binding[76] (Fig 1.11).

Guanine nucleotide exchange factors (GEFs)	GTPase specificity	Tissue specificity	
		High expression	**Moderate/ low expression**
Canonical GEFs			
LARG	RhoA	Peripheral blood leukocytes, ovary, spleen, testis, small intestine, prostate and colon	Thymus
PDZ-RhoGEF	RhoA	Brain, testis, heart, ovary, central nervous system (CNS) and placenta	Kidney, spleen, colon, skeletal muscle, pancreas, lung, prostate and liver
p115-RhoGEF	RhoA	Hematopoietic cells	
AKAP-Lbc	RhoA	Cardiac tissues	
p63RhoGEF	RhoA, RhoB, RhoC	Heart and brain	
p114-RhoGEF	RhoA, Rac1	Kidney and pancreas	Liver, skeletal muscle and testis
P-Rex1	Rac, RhoG	Neutrophils, macrophages, platelets, endothelial cells, neurons and brain tissues	Bone marrow, thymus, spleen, lymph nodes and lungs
αPIX	Rac1 , Cdc42	Heart, skeletal muscle and immune cells	
βPIX	Cdc42 , Rac	Ubiquitous	
Vav1	RhoA, Rac1, Rac2, RhoG, Cdc42	Hematopoietic cells	Pancreas, the tooth enamel, and the trophoblast layer
Vav2	RhoA, RhoG, Rac, Cdc42	Broadly expressed	
Vav3	RhoA, RhoG	Broadly expressed, highest in brain and hematopoietic cells	
WGEF	RhoA	Intestine, liver, heart, and kidney	
TIAM1	Rac1	Brain, testis and epidermis	
TIAM2	Rac1	Only restricted to neuronal cells	
Non-canonical GEFs			
DOCK180/ DOCK1	Rac	Broadly expressed	
DOCK2	Rac	Hematopoietic cells	

Table 1.1: List of few canonical and non-canonical GEFs, along with the GTPases they activate and their tissue-specific expression.

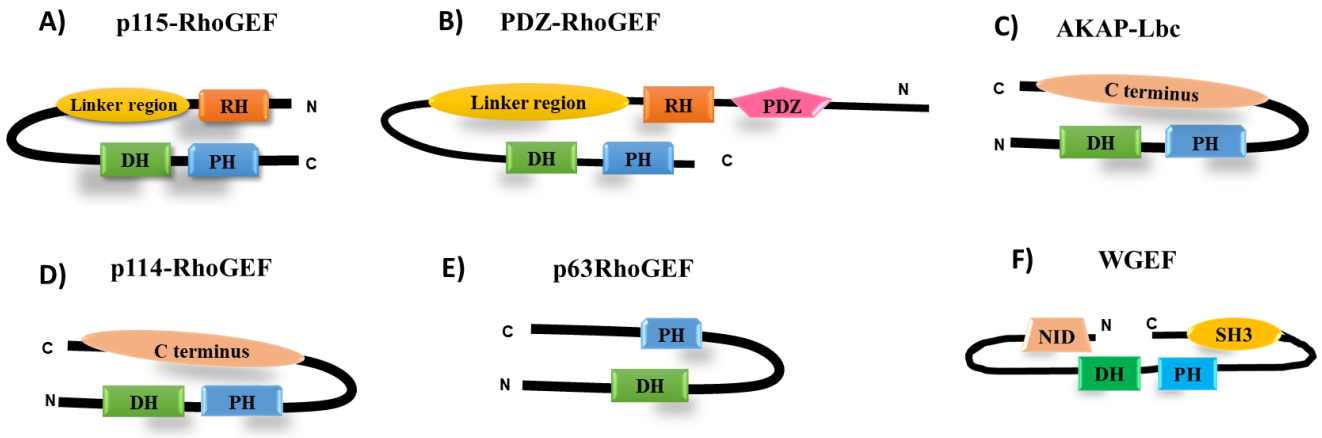


Figure 1.11: Illustrations showing autoinhibition of few Dbl family GEFs.

The activation of GEFs, achieved by releasing them from their autoinhibited state, involves numerous mechanisms such as phosphorylation of GEFs, lipid-mediated interactions and protein-protein interactions. These modifications lead to the structural rearrangement in GEFs, thereby activating them[76]. Kinases are shown to release autoinhibition of GEFs by phosphorylating them at specific positions leading to their activation. For instance, *p21-activated kinase interacting exchange proteins α and β* (α-PIX and β-PIX) are phosphorylated by kinases such as Src kinase, p21-activated kinase (PAK) and protein kinase A (PKA). Another similar example would be the Janus kinase (JAK) mediated phosphorylation of Vav GEFs. In contrast, *T-cell lymphoma invasion and metastasis inducing factor 1* (TIAM1) are reported to be activated by both phosphorylation and lipid-mediated interactions. Furthermore, a few GEFs are activated through protein-protein interaction. For instance, PDZ-RhoGEF and *Leukaemia-Associated Rho GEF* (LARG) are activated through their direct binding with the GPCR, whereas WGEF is reported to be activated by interaction with Dvl2 adapter protein[76]. This highlights the regulation of GEF activation, which occurs via different mechanisms depending on the signalling context.

1.8 Aberration in non-canonical Wnt signalling

Upregulation or overexpression of Wnt ligands, such as Wnt5A, Wnt7A and Wnt11, which predominantly operate under the non-canonical Wnt-PCP pathway, have been linked with various cancers, including those of the lungs, liver, pancreas, ovaries and breasts[77]. Furthermore, non-canonical Wnt signalling is also implicated in cardiovascular diseases[78].

Amongst the non-canonical Wnts, Wnt5A is studied extensively for its role in cancer and other diseases. For instance, Wnt5A has been reported to promote cell migration as well as cell invasion in gastric and colon cancer[79,80]. Dysregulation of Wnt5A is also associated with other pathological conditions including, fibrosis and inflammation[81]. Along with Wnt5A, the FZD7 receptor, which predominantly activates non-canonical Wnt signalling, but also plays a role in canonical Wnt signalling, is gaining attention as a potential therapeutic target[81–83]. Upregulation of FZD7 and other Wnt-PCP pathway proteins has also been detected in ovarian cancer, leading to cell proliferation, progression of cell cycle and migration[84]. Similarly, FZD7 is involved in tumour cell migration and invasion in colorectal cancer and hepatocellular carcinoma[85]. Wnt11, another bonafide member of the Wnt family, has been shown to control tumour cell migration and invasion in colon cancer[86]. Overexpression of another Wnt-PCP component, Daam1, is linked with poor prognosis and metastasis in breast cancer[87]. Other Wnt-PCP proteins, such as *Vangl* and *Prickle*, are shown to be involved in cancer cell metastasis[85]. While hyper activation of Wnt-PCP components is linked with cancer, their downregulation or loss of function has also been linked to various disorders, including neural tube closure defects, ciliopathies and kidney disorders[88–90]. Several reviews highlight the significance of Wnt-PCP in cancer and developmental disorders, emphasizing its potential as a target for therapeutic interventions[77,91].

1.9 Statement of problem

Cytoskeletal remodelling through the activation of RhoA has been shown to be one of the primary manifestations of the Wnt-PCP pathway. This signalling is critical for establishing cell polarity during the developmental stages and also for maintaining tissue homeostasis. Furthermore, dysregulation of the Wnt-PCP pathway has been implicated in various pathological conditions, including cancer and autoimmune disorders. Therefore, it is imperative to understand how this signalling propagates at the molecular level for therapeutic interventions targeting this pathway. Additionally, as one of the critical pathways that regulate life processes, it is essential to unravel how Wnt signalling exhibits diverse signalling modes by utilising a combination of Wnts and Frizzleds.

With decades of research on Rho GTPases, it is now well established that RhoA, a bonafide member of the Rho family, regulates a multitude of cellular processes by acting as a molecular switch. Activation of these switches requires GEFs that facilitate the exchange of

bound GDP for GTP in GTPases. Thus, the pathways that utilize Rho GTPases, employ GEFs that are activated by the upstream signalling events, which essentially involve propagation of signalling cascade emanating on account of the interaction between membrane receptors and their extracellular ligands. In the current context, RhoA is activated by the Wnt-PCP signalling, which is triggered upon the binding of secreted glycoprotein, Wnts and their membrane bound receptors, Frizzled. The GEF that operates under this pathway is shown to be WGEF. Although the signalling mechanism for the WGEF mediated activation of RhoA under the Wnt-PCP pathway has been shown, the molecular basis of this signalling event at the atomic level has remained unaddressed. The present study aims to unravel these aspects of WGEF mediated activation of RhoA, which is relayed under the Wnt-PCP pathway. From earlier studies, it is clear that cellular processes have evolved multiplayer regulation for triggering GTPases, which involve activation of GEFs that are generally known to exist in an autoinhibited state. In the present case, the autoinhibition of WGEF is released by an adapter protein, Dvl2, which is recruited to the C-terminal domain of the activated Frizzled receptors. The PDZ domain of Dvl2 engages with WGEF to release the autoinhibition of GEF. However, the precise molecular basis of Dvl2-WGEF interaction and its subsequent effects on RhoA activation is not clear. Thus, in pursuit of unravelling the Wnt-PCP controlled RhoA activation mechanism, the current thesis addresses key questions like a) which regions of WGEF interact with the $Dvl2^{PDZ}$ domain? b) What is the structural basis of their interaction? and c) How does the WGEF-$Dvl2^{PDZ}$ interaction influences the catalytic activity of the GEF?

1.10 Objectives of the thesis

As detailed in the previous section, this thesis work aims to investigate the molecular mechanism of Wnt-PCP signal that propagates from the Frizzled7 receptor to activate RhoA, leading to cytoskeletal rearrangements. By applying biochemical, structural and computational methods, we seek to achieve the following three objectives:

Objective 1: To investigate the mechanism of RhoA activation through WGEF

WGEF possesses the catalytic canonical DH domain, which shares a modest 42.9%, 45.1%, 38.9% and 46% sequence similarity with the well-characterized RhoA GEFs, LARG, PDZ-RhoGEF, Vav1 and AKAP_Lbc, respectively. Sequence alignment of DH domains of WGEF and other GEFs is given in Figure 1.12. However, it has been shown that subtle differences in the sequence and structure between the GEFs could result in a divergent catalytic mechanism[92–94]. Furthermore, Dbl family GEFs harbour a spectrum of functional domains that largely influence their mode of action[75]. Thus, unravelling the WGEF mediated RhoA activation requires an understanding of the structural basis of their interaction. Such a study would also provide insights into the role of other domains present in the GEF on the catalytic mechanism of the protein. Therefore, the first objective of the study is to elucidate the structural basis of WGEF-RhoA interaction using X-ray crystallography, which involves the formation of stable *in vitro* complex of both the proteins.

```
WGEF_DH           KLQEAKFELITSEASYIHSLSVAVGHFLGSAELSECLGAQDKQWLFSKLPEVKSTSERFL 60
Vav1_DH           KRCCCLREIQQTEEKYTDTLGSIQQHFLKPL--QRFLKPQDIEIIFINIEDLLRVHTHFL 58
AKAP_Lbc_DH       KRQEVIYELMQTEFHHVRTLKIMSGVYSQGMMADLLFEQQMVEKLFPCLDELISIHSQFF 60
LARG_DH           KRQEVINELFYTERAHVRTLKVLDQVFYQRVSREGILSPSELRKIFSNLEDILQLHIGLN 60
PDZ_RhoGEF_DH     DRQEVINELFVTEASHLRTLRVLDLIFYQRMKKENLMPREELARLFPNLPELIEIHNSWC 60
                  .      *:  :*  :  :*      :      .  :  .    :*  : ::

WGEF_DH           QDLEQRLEADVLRFSVC--------DVVLD-----HCPAFRRVYLPYVTNQAYQERTYQR 107
Vav1_DH           KEMKEALGTPGAA-------------NLYQVFIK--YKERFLVYGRYCSQVESASKHLDR 103
AKAP_Lbc_DH       QRILERKKESLVDKSEKNFLIKRIGDVLVNQFSGENAERLKKTYGKFCGQHNQSVNYFKD 120
LARG_DH           EQMKA------VRKRNETSVIDQIGEDLLTWFSGPGEEKLKHAAATFCSNQPFALEMIKS 114
PDZ_RhoGEF_DH     EAMKK------LR--EEGPIIKEISDLMLARFDGPAREELQQVAAQFCSYQSIALELIKT 112
                  : :                        :             .   :        .  .

WGEF_DH           LLLENPRFPGILARLEESPVCQRLPLTSFLILPFQRITRLKMLVENILKRTAQGSEDEDM 167
Vav1_DH           VAAAREDVQMKLEECSQRANNGRFTLRDLLMVPMQRVLKYHLLLQELVKHTQEA-MEKEN 162
AKAP_Lbc_DH       LYAKDKRFQAFVKKKMSSSVVRRLGIPECILLVTQRITKYPVLFQRILQCTKDNEVEQED 180
LARG_DH           RQKKDSRFQTFVQDAESNPLCRRLQLKDIIPTQMQRLTKYPLLLDNIAKYTEWP-TEREK 173
PDZ_RhoGEF_DH     KQRKESRFQLFMQEAESHPQCRRLQLRDLIISEMQRLTKYPLLLESIIKHTEGGTSEHEK 172
                         .   :    .     *: : . :    **: :  :*.: : : *     :.:

WGEF_DH           ATKAFNALKELVQECNAS   185
Vav1_DH           LRLALDAMRDLAQCVNEV   180
AKAP_Lbc_DH       LAQSLSLVKDVIGAVDSK   198
LARG_DH           VKKAADHCRQILNYVNQA   191
PDZ_RhoGEF_DH     LCRARDQCREILKYVNEA   190
                     : .  :::    :
```

Figure 1.12: Sequence alignment of DH domains of WGEF and other RhoA GEFs.

Objective 2: To unravel the mechanism of WGEF activation through the adapter protein Dishevelled2

Next, we aim to understand the activation of WGEF by the Wnt adapter protein Dvl2. Members of this GEF family are intrinsically autoinhibited in the absence of signalling events and are known to be activated by divergent mechanisms[76]. The autoinhibition of WGEF is due to the intra-protein interactions involving its N-terminal autoinhibitory domain (NID) and the C-terminal Src homology 3 (SH3) domain. Autoinhibition of the GEF is reported to be released by protein-protein binding involving WGEF and the PDZ domain of Dishevelled2 ($Dvl2^{PDZ}$)[24]. However, the specific regions of WGEF that interact with $Dvl2^{PDZ}$ are not known. Under the current objective, we employ biochemical, structural and computational methods to map the $Dvl2^{PDZ}$ interacting regions of WGEF and further investigate how this interaction influences the activation of RhoA by the GEF.

Objective 3: To investigate the molecular interaction between WGEF and Dvl2 in the context of Wnt-PCP pathway

On elucidating the $Dvl2^{PDZ}$ interacting regions of WGEF, we aim to understand how the interaction between these two proteins manifests in the Wnt-PCP pathway. Here, we utilize fluorescence microscopy to establish a cellular system to study the mode of $Dvl2^{PDZ}$- WGEF interaction operating under the FZD7-Wnt5A (Wnt-PCP ligand) pathway[58]. The formation of stress fiber upon activation of the Wnt-PCP pathway shall be used as a readout to characterize the $Dvl2^{PDZ}$-WGEF interaction and the consequent activation of RhoA. To achieve this objective, we shall establish breast cancer cell lines that conditionally express the Wnt-PCP components.

Chapter 2

Materials and Methods

This chapter discusses the materials and methods used to achieve the thesis objectives. The first section provides a list of reagents and materials used in the experiments, while the second section summarizes the key methodologies employed for biochemical characterization and protein structure determination using X-ray crystallography. Methods like molecular, analytical, biochemical, cell culture and computational techniques used to meet the thesis objectives are also briefly discussed.

2.1 Materials

List of reagents and materials used for the thesis work is given in Table 2.1

Name	**Company**
Taq DNA Polymerase	New England Biolabs (NEB), USA
Oligonucleotide primers	Eurofins, India
Restriction endonucleases	New England Biolabs (NEB), USA
20x SYBR Green	Invitrogen, USA
LA taq	Takara Bio, Japan
T4 DNA ligase	Promega, USA
Ultra pure agarose	Invitrogen, USA
Gel extraction and PCR purification kit	Qiagen, Germany and Promega, USA
Luria Bertani (LB) media	HiMedia, India
Luria Bertani (LB) agar	HiMedia, India
Isopropyl-β-D-thiogalactopyranoside	MP Biomedicals, USA
Ampicillin	MP Biomedicals, USA
Kanamycin	Gibco, ThermoFisher, USA
Chloramphenicol	Sigma-Aldrich, USA
Gentamicin	Invitrogen, USA
Tetracycline	HiMedia, India
Tris-HCl	MP Biomedicals, USA
Sodium chloride	MP Biomedicals, USA
Dithiothreitol (DTT)	MP Biomedicals
2β-mercaptoethanol	HiMedia, India
Phenylmethylsulfonyl fluoride (PMSF)	Sigma-Aldrich, USA
Glycerol	SRL, India
Lysozyme	MP Biomedicals, USA
cOmplete EDTA-free protease	Roche, Switzerland
Glutathione reduced	MP Biomedicals, USA

Coomassie Brilliant Blue R-250	SRL, India
Glutathione sepharose 4B slurry	Cytiva, USA
$Ni2^{+}$-nitrilotriacetate (Ni^{2+}-NTA) slurry	Roche, USA
Strep-Tactin XT Sepharose	Cytiva, USA
Sodium dodecyl sulfate (SDS)	Sigma-Aldrich, USA
d-Desthiobiotin	Sigma-Aldrich, USA
Calcium chloride	MP Biomedicals
TOPO-TA cloning kit	Invitrogen, USA
Tris-acetate	Sigma-Aldrich, USA
1kb DNA ladder	New England Biolabs (NEB), USA
Acrylamide	Sigma-Aldrich, USA
Ethylenediaminetetraacetic acid (EDTA)	ANJ Biomedicals, USA
Heparin trap	Cytiva, USA
Blue sepharose trap	Cytiva, USA
Q sepharose trap	Cytiva, USA
Sucrose	Sigma-Aldrich, USA
Size exclusion columns	Cytiva, USA
Precision plus protein standards	Bio-Rad, USA
Imidazole	MP Biomedicals, USA
Mant-GDP	Sigma-Aldrich, USA
Guanosine triphosphate (GTP)	Sigma-Aldrich, USA
Magnesium chloride	Sigma-Aldrich, USA
Commercial Crystallization suits	Qiagen, Germany & Molecular Dimensions, UK
Polyethylene glycol (PEG)	Sigma-Aldrich, USA
Isopropanol	MP Biomedicals, USA
Tri-sodium citrate	Sigma-Aldrich, USA
sodium-acetate	Sigma-Aldrich, USA
Peptides	Sai biotech, India
CHAPS	SRL,India
Triton X-100	HiMedia, India
Dimethylsulfoxide (DMSO)	MP Biomedicals, USA
Diluent for nucleic acid extraction	ANJ Biomedicals, USA
Fetal Bovine Serum, HI	Gibco, ThermoFisher, USA
GeneJuice Transfection Reagent	Sigma-Aldrich, USA
SF900-III Insect Cell Media	Gibco, ThermoFisher, USA
Polyethylenimine (PEI)	Sigma-Aldrich, USA
Lipofectamine 3000	Invitrogen, USA
OPTI-MEM media	Gibco, ThermoFisher, USA
DMEM media	Gibco, ThermoFisher, USA
Poly L-lysine	Sigma-Aldrich, USA
GlutaMAX supplement	Gibco, ThermoFisher, USA
TrypLE	Gibco, ThermoFisher, USA
Sodium pyruvate solution	Gibco, ThermoFisher, USA
Geneticin powder	Gibco, ThermoFisher, USA
Phosphate Buffer Saline (PBS)	Gibco, ThermoFisher, USA
Penstrep	Gibco, ThermoFisher, USA

Zeocin	Invitrogen, USA
Trypan blue	Gibco, ThermoFisher, USA
Anti-6xHis tag primary antibody	ThermoFisher, USA
HRP-conjugated secondary antibody	ThermoFisher, USA

Table 2.1: List of chemicals and reagents used.

2.2 Methods

2.2.1 Molecular biology methods

2.2.1.1 Primer designing

NCBI database was used to obtain a complete nucleotide sequence of genes of interest. Nucleotides consisting of the start and end of the gene sequence that need to be cloned serve as a template to design primers. Although a nucleotide length of 20-30 base pairs is considered ideal for primers, this may vary depending upon the GC content (GC content of 50% is considered ideal) and the melting temperature (Tm) of the primer[95]. To calculate the GC content and Tm for primers, the free online tool *Oligo Calc: Oligonucleotide Properties Calculator* was used[96]. Primers were synthesized and procured from Eurofins, Bangalore.

2.2.1.2 Polymerase chain reaction (PCR)

PCR is a technique in molecular biology used to produce multiple copies of a specific DNA fragment. PCR based gene amplification utilizes Taq polymerase (thermostable DNA polymerase) and gene specific forward and reverse primers. PCR reactions were carried out in a 30 µL reaction volume containing 30-60 ng of DNA template, 0.1-0.5 µM of forward and reverse primers each, 0.2 mM final concentration of each dNTPs, 1U of Taq DNA polymerase enzyme, 1X DNA polymerase buffer and RNase free water to makeup the reaction volume. Thermal cycler PCR instrument (Agilent Technologies Sure cycler 8800) was used for DNA amplification and the specific cycling program is given below (Table 2.2).

Step	Stage	Temperature ℃	Time	No of cycles
1	Denaturation	95	1 minute	1
2	Denaturation	95	10 seconds	5
	Annealing	61	10 seconds	
	Extension	68	As per length of gene (1 kb/minute)	
	Extension	72	10 seconds	
3	Denaturation	95	10 seconds	5
	Annealing	58	10 seconds	
	Extension	68	As per length of gene (1 kb/minute)	
	Extension	72	10 seconds	
4	Denaturation	95	10 seconds	30
	Annealing	55	10 seconds	
	Extension	68	As per length of gene (1 kb/minute)	
	Extension	72	10 seconds	
5	Extension	72	10 minutes	1
6	Hold	4	Infinite	

Table 2.2: PCR program for gene amplification.

This method of decreasing the annealing temperature after a few cycles is known as touch-down PCR, this helps to get specific PCR products[97]. The amplified PCR product was subsequently analyzed using agarose gel electrophoresis to get a pure and isolated product, as there may be non-specific and primer dimer bands in the reaction. Upon visualizing the gel, a band obtained at the correct size was excised from the gel and extracted using gel extraction and PCR clean-up system (Promega).

2.2.1.3 Restriction digestion and Ligation

The cloning vector of interest and the PCR amplified product were digested using the same set of restriction endonuclease enzymes. 0.5-1 µg plasmid DNA, 1X of buffer and 1U of restriction enzymes were added to the reaction followed by incubation at 37°C/3 hours (h). This process generates complementary sticky ends in the PCR product and the vector facilitating their ligation. For ligation reaction, the digested PCR product and vector were mixed in a ratio of 3:1 or 5:1 in a reaction mixture of 10-12 µL consisting of T4 ligase enzyme (Promega) and 1X manufacturer's ligase buffer. This ligation reaction was incubated at 4°C/16 h or room temperature (RT)/2 h, followed by transformation in chemically competent *E.coli* strain DH5α. Plates were kept at 37°C/16-18 h and the colonies obtained were further processed for screening of positive clones.

2.2.1.4 Transformation

Chemically competent cells of *E.coli* strains Rosetta (DE3), B834 rare (DE3), BL21* (DE3) and DH5α were generated by the Inoue method[98]. For transformation, the competent cells from -80^{o}C were allowed to thaw for 5 min on ice. After thawing, the ligation reaction was added to cells, followed by gentle tapping to ensure proper mixing of DNA. The cells-DNA mixture was then incubated for 35 min on ice to promote DNA binding to the cell surface. After incubation, cells were given heat shock in a 42°C water bath for 60 seconds (sec), temporarily creating pores in the cell membrane and allowing DNA entry. Following heat shock, 1 mL of Luria-Bertani (LB) broth was added to the cells, which were then incubated under shaking condition 37°C/1 h. After incubation, the cells were centrifuged and the media was discarded, such that around 100 µL media was left for cell resuspension before plating on the desired antibiotic-containing agar plate. Similar transformation steps were performed in *E.coli* cells during protein expression and plasmid amplification, wherein 100-150 ng of desired plasmid was added to the competent cells. Plates were incubated 37°C/overnight.

2.2.1.5 Purification of plasmid DNA

A single isolated colony obtained on the antibiotic-containing plate was cultured in 5-10 mL of antibiotics supplemented LB broth at 37°C/12-14 h, followed by cell pelleting. Purification of plasmid was done as per the manufacturer's instructions (Qiagen). The cell pellet was first resuspended in a resuspension buffer (Tris buffer, pH 8.5 or TE buffer, pH 8) followed by

alkaline lysis with NaOH and SDS containing buffer for 5 minutes (min) in presence of RNase (final concentration 100 µg/mL). After cell lysis, a neutralization step was carried out, during which the chromosomal DNA, proteins and other insoluble components were precipitated. This mixture was centrifuged at high speed to separate the precipitate from the plasmid DNA present in the supernatant. The supernatant was then added to a pre-equilibrated column, enabling the binding of plasmid DNA to the column and eliminating other components, such as salts and metabolites, in the flow through. This binding of the plasmid to the column is based on the principle of ion-exchange, where the plasmid adheres to the anion-exchange column. Column washing was done with a low salt, low alcohol wash buffer to remove any contaminants without affecting the bound plasmid. Finally, the plasmid DNA was eluted with pre-warmed (60°C) MilliQ purified water or a high salt buffer.

2.2.1.6 Quantification of DNA

Plasmid DNA concentration was measured using a Nanodrop spectrophotometer (NanoVue GE Healthcare) based on the optimal absorption at 260 nm, as double-stranded DNA absorbs at this wavelength. Purity of the DNA was evaluated using the 260/280 ratio. A ratio between 1.8-2.0 indicates pure DNA, a ratio of 2.0 and above suggests presence of RNA and a ratio below 1.8 suggests protein or phenol contamination.

2.2.1.7 Site-directed mutagenesis

Site-directed mutagenesis was employed to incorporate specific nucleotide changes into a DNA sequence, typically by using PCR. Primers for mutation were designed such that the intended mutation should be present in the center of the primers with, 10-15 base pairs (bp) matching on each side, this ensures proper primer annealing to the DNA template. Eurofins, Bangalore, synthesized the primers. The parent DNA plasmid was used as a template for PCR amplification. LA Taq DNA polymerase (Takara bio) was used for PCR by using the program given in Table 2.3. Amplified DNA was then treated with Dpn1 enzyme at 37°C/3 h, to digest the methylated parent plasmid. After this, DH5α cells were transformed with the plasmid and plated on the antibiotic-containing agar plate. The colonies obtained were further screened for positive clones.

Step No	Stage	Temperature (°C)	Time	Cycles
Step1	Denaturation	95	1 minute	1
Step2	Denaturation	95	10 seconds	30
	Annealing	54	10 seconds	
	Extension	68	As per plasmid length 1kb/minute	
Step3	Extension	72	10 minutes	1
Step4	Hold	4	Infinite	

Table 2.3: PCR program for site-directed mutagenesis.

2.2.1.8 TOPO-TA subcloning

TOPO subcloning was used to transfer PCR products from one vector to another. This cloning method relies on the 3´ A-overhangs introduced by Taq polymerase during PCR amplification. The TOPO vector is designed with T-overhangs to facilitate efficient ligation with the PCR-amplified product. Once the PCR product was obtained, it was incubated at RT with the TOPO-TA vector which is covalently bound with Topoisomerase I (Invitrogen™ TOPO-TA Cloning Kit). This reaction after incubation, was used for transformation of DH5α cells. This cloning method does not require a separate ligation step as Topoisomerase I acts as restriction enzyme and ligase.

2.2.1.9 Clone conformation and sequencing of plasmid

After cloning or mutagenesis, the colonies from the plates were screened to identify positive clones. Firstly, the colonies were inoculated for plasmid extraction. The plasmid was extracted using the protocol in section 2.2.1.5. The extracted plasmid was further subjected to restriction digestion using specific restriction enzymes. The digested products were then run on 0.7% agarose gel, and the resulting band patterns were compared with the predicted band pattern profile obtained using NEBcutter v2.0[99] for the digested plasmid. Once the digested bands are confirmed, positive clones are further sequenced using Sanger sequencing (Eurofins, Bangalore) and the sequence is analyzed using Snapgene viewer and Clustal Omega (Multiple Sequence Alignment) softwares.

2.2.2 Analytical Methods

2.2.2.1 Electrophoresis

Electrophoresis is used to separate molecules, such as proteins, DNA and RNA according to their size and charge under the influence of an electrical field. This separation was done by using agarose gel for DNA and polyacrylamide gel for proteins. The separated DNA and protein bands were visualized using stains, such as SYBR-safe and Coomassie blue, respectively.

2.2.2.1.1 Agarose gel electrophoresis

Agarose gels were made in TAE buffer (40 mM Tris-acetate pH 8, 1 mM EDTA). The gels were casted and run using a Mini electrophoresis system (Bio-Rad, USA). Percentage of agarose gel to be casted was determined based on the length of the DNA fragments to be separated. Solidified gel was placed in the electrophoresis unit containing the TAE buffer. DNA samples were mixed with SYBR-safe gel loading dye (contains 0.25% Xylene cyanol FF, 40% sucrose, 0.25% Bromophenol blue) and loaded into the wells of the gel. 1 kb DNA ladder (200 ng) (New England Biolabs, UK) served as a marker for molecular weight. Electrophoresis was carried at 5 V/cm for 40 min or until the ladder was resolved, and the DNA was observed under UV light using the Bio-Rad Gel Doc XR system.

2.2.2.1.2 SDS-Polyacrylamide gel electrophoresis (SDS-PAGE)

SDS-PAGE was used to resolve protein bands depending on their molecular weight. Protein samples were prepared with 1X SDS-PAGE sample loading buffer and heated at 95℃/5 min. All SDS-PAGE gels used in this study were prepared in the lab and casted using Bio-Rad 1.0 mm mini gel cassettes. Depending on the size of the protein, either 12% or 15% resolving gels were used with a 5% stacking gel. SDS-PAGE was carried out in 1X SDS running buffer in an electrophoresis unit containing both the cathode and anode chambers at 150 V/2 h. Further, the gels were removed and stained with a lab-made instantBlue coomassie blue protein stain to observe the protein bands.

2.2.2.2 Affinity Chromatography

Protein purification using affinity chromatography is based on their specific interaction with a ligand immobilized on the chromatographic matrix. To selectively isolate proteins of interest, these proteins were tagged with 6xHis, Glutathione S-transferase (GST) and Strep-tag II affinity tags. GST-tagged proteins attach to glutathione-coated resin and are eluted by using free glutathione[100]. Besides facilitating purification, the GST tag can also enhance the solubility of the fusion protein. His tag typically consists of six histidine residues (6xHis) and binds to metal ions like nickel or cobalt charged resins and elution was typically achieved by increasing concentrations of imidazole, which competes with histidine tag for binding the charged slurry[101]. Strep-tag II peptide tag comprises 8 amino acids (WSHPQFEK) that attach to the streptavidin slurry[102]. Elution for strep II-tagged proteins was achieved by using biotin or biotin analogues, which have a high affinity for streptavidin.

2.2.2.3 Ion-exchange Chromatography (IEC)

IEC is a technique used for separating molecules based on their net charge. The stationary phase consists of an ion-exchange resin that contains charged groups, the negatively charged cation-exchange resins interact with positively charged molecules, whereas the positively charged anion-exchange resins interact with negatively charged molecules. Elution of the attached protein was achieved by raising the ionic strength of the mobile phase, which is achieved by having a gradient of increasing salt concentration. Q sepharose anion-exchange column (GE Healthcare) was used to perform ion-exchange chromatography.

2.2.2.4 Size exclusion chromatography (SEC)

SEC is used to resolve molecules based on their molecular weight. The SEC columns are packed with porous beads made from dextran, agarose or polyacrylamide materials. The choice of the bead depends on the molecule to be separated, as each of these beads has a defined pore size range. Large molecules that cannot enter the pores travel around them, whereas small molecules enter and elute through the pores, delaying their elution. Thus, the molecules are separated based on size, with the large molecules eluting first followed by the small molecules[103].

2.2.2.5 Size exclusion chromatography- Multi angle light scattering (SEC-MALS)

SEC combined with Multi-angle light scattering is a highly sensitive and commonly used technique for determining the absolute molecular weight of proteins[104]. SEC separates the molecules based on their hydrodynamic volumes. At the same time, MALS measures the intensity and angular dependence of scattered light to determine the size (radius of gyration) and absolute molar mass in the solution. These techniques together provide more accurate mass determinations than either methods individually. SEC-MALS analysis of the purified proteins was done using a Superdex 200 10/300 GL column (GE Healthcare) in line with an Agilent HPLC system consisting of an 18-angle light scattering detector (Wyatt Dawn HELIOS II) and a refractive index detector (Wyatt Optilab TrEX). ASTRA software (Wyatt Technologies) was used to determine the molecular weight.

2.2.3 Biochemical Methods

2.2.3.1 Protein expression and purification

2.2.3.1.1 Expression protocol

After confirming the recombinant constructs, transformation of *E.coli* expression strains was done. A single transformed colony was used to inoculate 5 mL of LB broth supplemented with antibiotics (depending on the resistance encoded by the plasmid used) and incubated at 37°C/overnight in a shaker at 180 RPM. 1 mL of this primary culture was used to inoculate 100 mL of LB broth with antibiotics for secondary culture. This culture was grown at 37°C in a shaker. When the optical density (OD 600nm) of 0.6 is achieved, expression of protein was induced with 0.5 mM isopropyl β-d-thiogalactopyranoside (IPTG) and cells were incubated at 18°C-20°C/14-16 h. Similarly, an equal volume of uninduced culture was also processed as a control. Following incubation, cells were centrifuged at 4000 RPM/20 min. The harvested cell pellet was stored at -80°C for subsequent use.

2.2.3.1.2 Expression analysis

The cell pellet was resuspended in Tris buffer followed by sonication for 30 seconds (sec) (45% amplitude, 5 sec ON/8 sec OFF cycle). Cell extract was further centrifuged at 20,000 RPM/30 min at 4°C, and the resulting clear supernatant and pellet fractions were loaded onto 12% or

15% (depending on the size of protein to be studied) SDS-PAGE to assess protein expression. Similar steps were repeated for the uninduced cell pellet. A comparison of both the induced and uninduced bands can be analyzed for protein expression.

2.2.3.1.3 Solubility analysis

The solubility analysis of the overexpressed protein was performed by resuspending the harvested cells in 5-10 mL of buffer with 0.70 mg/mL of lysozyme. The lysate was then subjected to sonication and centrifugation at 20,000 RPM/30 min at 4°C. The clear supernatant and the pellet were analyzed for the presence of overexpressed protein by SDS-PAGE. Presence of protein in the clear supernatant fraction indicates soluble protein.

2.2.3.1.4 Large scale protein expression and purification

Large scale expression was performed using 4 liters to 8 liters of secondary culture. 1% of primary culture was inoculated in antibiotics-containing LB broth. Protein expression was induced as discussed in 2.2.3.1.1. Cell pellet was resuspended in a 10X chilled lysis buffer containing 20 mM Sucrose, 2 mM PMSF, lysozyme and a cOmplete EDTA-free protease inhibitor (PI) tablet. Lysis was done either by sonication or homogenization: for sonication, the sample was subjected to 3-4 rounds of 5 min cycle (45% amplitude, 5 sec ON/8 sec OFF) at 4°C; for homogenization, pellet was lysed at 4°C/20 min under 350-400 bar pressure. After lysis, the lysate was spun at 20,000 RPM/1 h to obtain a clear supernatant, and filtered using a 0.4 µm syringe filter to remove any residual debris and then loaded onto the respective pre-equilibrated affinity column. After loading the protein on column, buffer wash was given to remove non-specifically bound proteins. Elution was carried out using 20 mM reduced glutathione for GST-tagged protein, 200 mM imidazole gradient elution for His-tagged proteins and 5 mM desthiobiotin for elution of strep II-tagged proteins. To remove cleavable affinity tags, the elute fractions were dialysed at 10°C/overnight in presence of PreScission protease (for GST-tagged proteins in p3E vector) or TEV (for His-tagged proteins in pET vector) proteases, both purified in our lab. After dialysis, the tags, proteases and uncleaved protein were removed by performing desalting and a reverse affinity purification. Target protein present in the flow-through of reverse affinity chromatography was further purified by IEC, followed by SEC or directly used for SEC. SEC was done using Superdex or Sephacryl

columns. Peak fractions containing the proteins were concentrated for biochemical and crystallization studies.

2.2.3.2 Guanine nucleotide exchange factor (GEF) assay

GEF assay is used to measure the activity of GEFs, which catalyze the exchange of GDP with GTP, switching the GTPases from inactive (GDP-bound) to active (GTP-bound) state. In this assay, the GTPase is pre-loaded with fluorescently labelled Mant-GDP and incubated with and without GEF. After the fluorescence stabilizes, an excess GTP is added, which replaces the fluorescent Mant-GDP, leading to a decrease in the fluorescence[105] (Fig 2.1A). The intrinsic rate of nucleotide exchange in GTPase is slow, however in the presence of GEF, this exchange is enhanced, indicated by an exponential decrease in the fluorescence. This exchange is monitored for 20-30 min to determine the rate of exchange reaction in the presence and absence of GEF (Fig 2.1B).

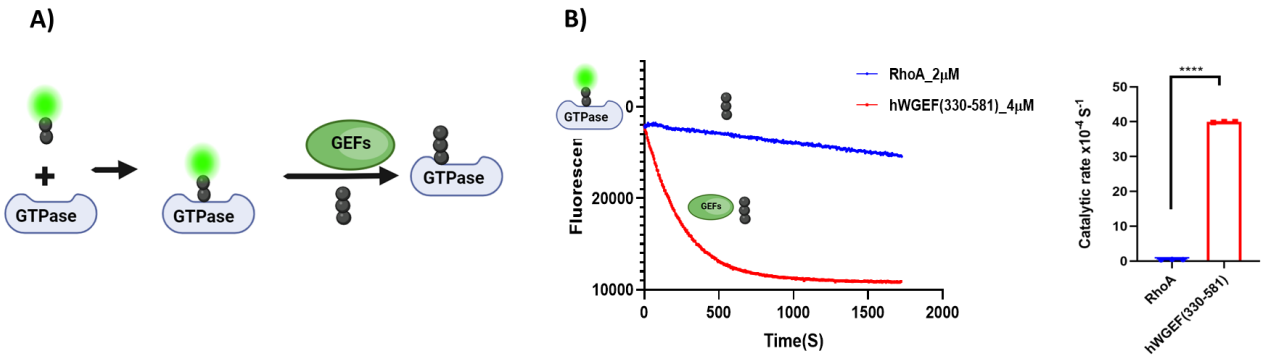


Figure 2.1: Guanine nucleotide exchange assay. A) *Addition of GEF, enhances the GDP to GTP exchange of GTPase.* ***B)*** *The exchange activity can be plotted as exponential decrease in the fluorescence and the catalytic activity is determined. The graph shows exponential decay in the presence of human WGEF$^{330-581}$ construct (hWGEF330). The catalytic rate of nucleotide exchange was calculated using the one-phase exponential decay mode in GraphPad Prism.*

2.2.3.3 Microscale Thermophoresis (MST)

MST is a technique used to study binding affinities between proteins, nucleic acids and small molecules. Binding affinities are measured based on the detection of changes in the movement of molecules in the presence of temperature gradients. This technique is sensitive to any binding-dependent changes in molecular properties such as size, hydration shell, conformation or charge of the molecule. In MST, typically, one of the interaction partners is labelled with a fluorescent dye which is tracked using fluorescence spectroscopy[106,107]. The labelled molecule

was added in a fixed quantity, whereas the unlabelled molecule is serially diluted across these tubes and incubated for 30 min. Once the incubation is done, each of these mixtures were loaded into glass capillaries, which were then loaded on the MST instrument. When a microscale temperature gradient is applied using an infrared laser, molecules in the capillary move away. Changes in the movement of the molecules which are dependent on the binding event or conformational changes are detected by the changes in the fluorescence. The relative fluorescence between the initial fluorescence and fluorescence after thermal diffusion is used to measure the binding affinity through MST[106,107] (Fig 2.2).

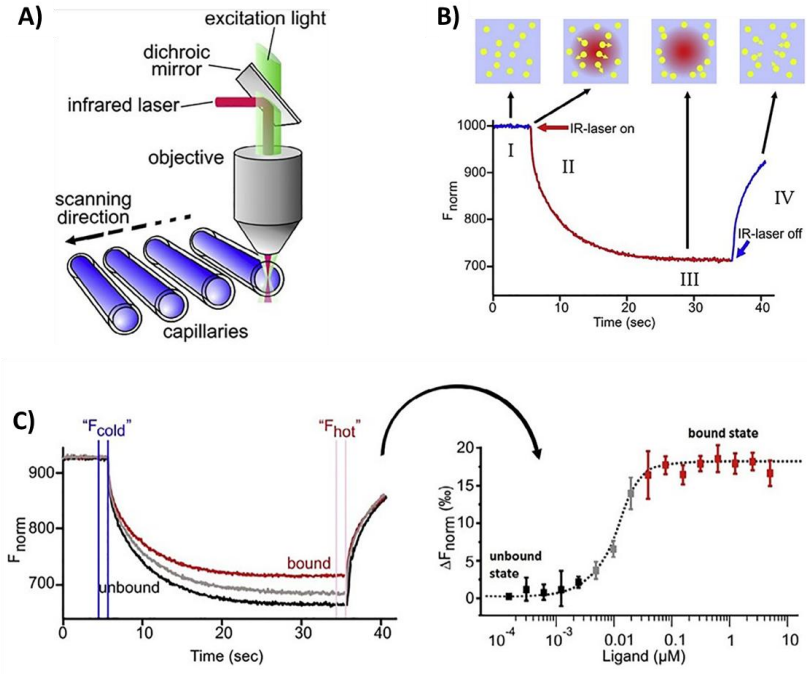


Figure 2.2: Microscale Thermophoresis (MST) assay. A) *MST instrument.* ***B)*** *Movement of molecules with respect to temperature change.* ***C)*** *Signals for bound and unbound molecules (left) that are further used to determine the binding curves (right). Image adapted from Asmari et al.* [107,108].

2.2.4 Macromolecular X-ray crystallography

Macromolecular X-ray crystallography is used to determine the atomic and macromolecular structure of large biological macromolecules, such as nucleic acids and proteins. Using this, we can visualize the three-dimensional arrangement of atoms inside these macromolecules with high resolution, revealing details on their function, mechanism and interactions. For this, the target macromolecule should be first crystallized, followed by X-ray diffraction of the crystal. Exposing the crystals to X-rays, leads to scattering of these rays by the atoms within the crystal

lattice, producing sharp, intense spots on an X-ray detector. The intensities and positions of these diffraction spots can be used to obtain the structure factors are necessary to calculate the electron density map for the molecule[109]. This structure determination process has significant limitations, as achieving crystals of diffractable quality and size can be difficult. Furthermore, even after collecting high-quality diffraction data, obtaining accurate phase information remains a significant hurdle in X-ray crystallography. Consequently, crystallization and phase determination steps are often considered the rate-limiting steps in macromolecular crystallographic studies. Despite these challenges, X-ray crystallography remains one of the important methods for determining protein and other macromolecule structures. Moreover, recent advancements such as the development of tunable synchrotron X-ray sources, automated robotic systems for sample handling, improved cryoprotection techniques, and the use of CCD detectors have significantly accelerated and simplified the procedure. Experimental steps involved in structure determination are outlined in a flow diagram (Fig 2.3).

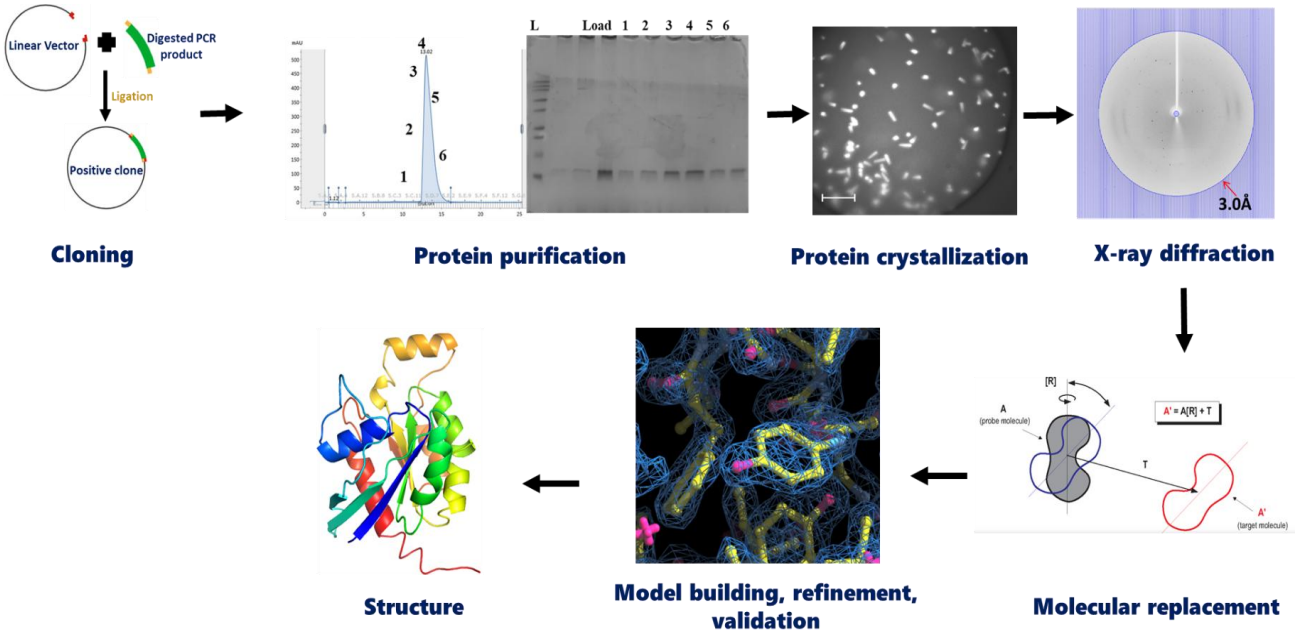


Figure 2.3: Major steps involved in protein structure determination.

2.2.4.1 Protein crystallization

Protein crystallization is a fundamental step in X-ray crystallography, allowing the study of protein structures at the molecular level. However, it is often a challenging process as the success of obtaining high-quality crystals majorly depends on the purity and homogeneity of the protein. Impurities such as aggregates or misfolded proteins can interfere with the crystal

formation. Other parameters that influence the crystal formation may include the concentration of protein, disordered regions in the proteins and crystallization conditions. The basic principle of crystallization is to bring the protein into a supersaturated state, where the protein concentration is more than its solubility. A protein crystallization phase diagram helps to explain the relationship between protein concentration and precipitation concentration (Fig 2.4A). In undersaturated regions, no crystallization occurs. In the moderately supersaturated region, nuclei begin to form. If the supersaturation is too high, the protein will form precipitation instead of nucleation. Therefore, controlled saturation of the protein is critical for the formation of crystals. After nucleation, the crystal growth phase begins, where more protein molecules add up to the nucleation, forming a stable crystal lattice. At this stage, the protein molecules arrange in a regular, repeating pattern, forming a crystal.

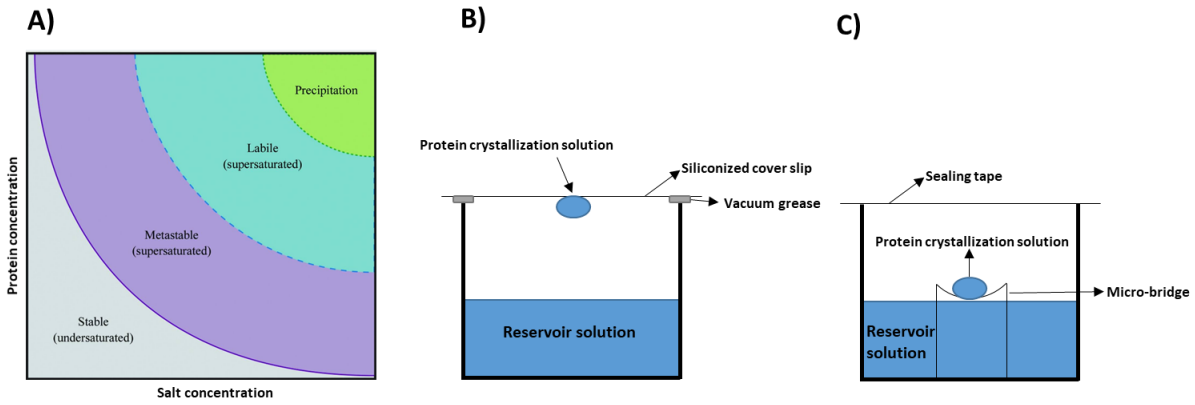


Figure 2.4: Protein crystallization phase diagram and methods: A) *Phase diagram of crystallization. Image adapted from McPherson et al.* [110]. ***B)*** *Hanging drop crystallization method.* ***C)*** *Sitting drop crystallization method.*

Several crystallization methods have been developed over time, with vapour diffusion being the most commonly used method. The protein solution and precipitant were mixed and equilibrated against a reservoir of precipitant solution. The protein solution becomes concentrated as the water vapour diffuses, leading to supersaturation. There are two variations of vapour diffusion method, hanging drop: in this method, a small droplet containing the protein and precipitant mixture is suspended on a coverslip over a reservoir of precipitant (Fig 2.4B), sitting drop: in this method, the droplet sits on a platform in a well which also consists of a reservoir well for the precipitant (Fig 2.4C). In the present study, these two methods were majorly employed to obtain protein crystals. Microbatch method is another method in which the protein and precipitant are mixed in small droplets and covered with a layer of silicon or

paraffin oil, preventing evaporation, allowing the protein and precipitant to interact over time and promotes supersaturation. Other methods would include the dialysis method in which a semi-permanent membrane separates the protein and precipitant. Crystal formation is achieved by diffusion of precipitant into the protein chamber over a period of time. Another method is free interface diffusion, in which the protein and precipitant solutions are layered on top of each other in a capillary, leading to the diffusion and subsequent increase in protein concentration. Commercially available screening kits provide a wide range of conditions as these kits contain a combination of precipitants, salts and buffers at different concentrations and pH levels. Once initial crystals are obtained, these parameters can be fine-tuned to improve the quality of the crystals.

2.2.4.2 Single crystal X-ray diffraction and data collection

X-ray diffraction is based on constructive interference. Constructive interference is formed when the path difference between X-rays diffracted by two adjacent crystal lattice planes is an integral multiple of the wavelength of the collimated monochromatic X-ray used[111]. This is also known as Bragg's Law [112] and is given by the following equation 1:

$$n\lambda = 2d\sin(\theta) \qquad (1)$$

Where,

n = integer

λ= X-rays wavelength

d= distance between two adjacent crystal lattice planes generating diffraction

θ= angle of diffraction/incident

For protein crystallography, there are two radiation sources: Laboratory/Home source and synchrotron source. These consist of three main components which include, X-ray radiation source, optical elements and detectors. In laboratory/home sources, the X-ray radiations can be obtained with a rotating anode. X-rays are produced by bombarding metals like copper and molybdenum with high-speed electrons[113]. Another way for obtaining X-rays in laboratory/home sources is by using vacuum-sealed tube X-ray generators[114]. At synchrotron

facilities, highly collimated X-ray beams are generated by using high-energy electrons circulating in the storage ring. Paths of these electrons are maintained via bending magnets or undulators. Some advanced synchrotron facilities use free-electron lazers as a source to produce more collimated and intense X-ray beams[115]. To generate high-quality data, optical elements also play an important role. These optical elements include collimators, monochromators, mirrors, beam shaping optics and beamstop[116]. Amongst the detectors, Hybrid Photon Counting detectors facilitate high resolution data collection and better signal-to-noise ratio[117]. Modern synchrotrons are better than conventional laboratory/home sources, mainly due to their elevated electron flux and ability to fine-tune X-ray wavelength.

Data collection for protein crystals is conducted at cryogenic temperatures (100 K) to prevent radiation damage caused due to interaction of X-ray beams with the crystal molecules. This interaction generates heat, leading to bond cleavage in protein atoms depending on X-ray exposure dose. Therefore, once the protein crystals were obtained, they were soaked in a cryoprotectant buffer before being frozen in liquid nitrogen for storage[118,119]. Successful data collection depends on the quality of the crystal, its symmetry and mosaicity, amongst other factors. To design an optimal data collection strategy, a few diffraction frames were collected and indexed to determine crystal lattice, unit cell parameters and also the optimum distance between crystal and detector. Data obtained from X-ray diffraction can be processed further using following steps: Indexing of data, integration of spot intensities, scaling and merging. Indexing involves analysis of diffraction patterns and assigning Miller indices (*hkl*) values for each plane in the crystal lattice, which contribute to the diffraction spots. It helps in defining the lattice plane orientation and also provides a preliminary approximation regarding the dimensions of the unit cell and crystal symmetry[120,121]. Integration involves quantifying the intensity of each diffraction spot. It can be achieved by the summation integration method, which requires addition of each pixel value that forms the spot. Alternatively, the profile fitting method is employed in which a 3D integration profile is created using intensities from the spots in different images, followed by calculating the refined intensity of the diffraction spot. This refined intensity is more accurate than the intensity obtained from summation integration[121]. Various factors such as crystal size, X-ray beam intensity fluctuations and radiation damage lead to differences in the scale of recorded intensities. Therefore, scaling was done to normalize these intensities on a common scale followed by merging which merges the scaled data sets into a single, merged dataset. This was done by averaging the intensities of equivalents from

different data sets. The dataset was further used for structure determination. This can be done in crystallographic software packages like CCP4 and Phenix[121].

After scaling, parameters such as resolution, completeness, multiplicity, signal-to-noise ratio and Rmerge can be used for assessing the quality of diffraction data. Data completeness describes how complete the data is for the reflections that are recorded. Multiplicity helps to determine the average number of measurements for equivalent reflections. Rmerge corresponds to different measurements of intensities of a particular reflection and its agreement with the average[122]. It is defined as (Equation 2)

$$\text{Rmerge} = \frac{\sum_{hkl}\sum_{i=1}^{n}|I_i(hkl)-\bar{I}\,(hkl)|}{\sum_{hkl}\sum_{i=1}^{n} I_i(hkl)} \qquad (2)$$

Where *n* stands for independent intensity measurements, *I(hkl)* denotes the reflection intensity and $\bar{I}(hkl)$ denotes average of the intensities. Rmerge value for a good data set is ~10%[122]. In this study, diffraction data were indexed and integrated using XDS and scaling was done using AIMLESS.

XDS was used for the reduction of data obtained from single crystal diffraction collected using the rotation method in X-ray crystallography. This software identifies and signs the crystal's unit cell parameters. The software also calculates the intensities of individual reflections by adding the diffraction spots over different images, accounting for the orientation of the crystal and other parameters. This software is composed of three programs, which include XDS, XSCALE and XDSCONV. XDS predicts the crystal symmetry and generates a list of integrated reflection intensities. This program is organized into 8 steps, namely, XYCORR, INIT, COLSPOT, IDXREF, DEFPIX, XPLAN, INTEGRATE and CORRECT, which are used in succession by the program. XSCALE aligns the data sets processed by XDS on a common scale and optionally combines them into one or more sets of unique reflections. A report on the accuracy of the integrated intensities and the completeness of the data is also provided by this program. XDSCONV is used to convert reflection data files produced by XDS and XSCALE into different formats needed by various structure determination softwares. It can also provide test reflections to calculate the Rfree factor required for monitoring the structure refinement progress[123]. After XDS, AIMLESS is used for scaling and merging the data. It applies scaling corrections to the diffraction data to take care of the experimental

variations, such as changes in crystal orientation, decay of crystal over time and detector response. It combines multiple datasets of the same reflection to produce a set of unique reflections with average intensities. It also provides various parameters of the scaled and merged data, which includes completeness, redundancy, Rmerge and signal-to-noise ratio. These parameters help to assess the data quality[124].

2.2.4.3 Structure determination and Phase problem

After scaling the diffraction data, the next step is to calculate the structure factors which can be used to calculate the electron density map for the molecule. The structure factor (F_{hkl}) is expressed in terms of amplitude and phase of the scattered wave for a *hkl* reflection[125] (Equation 3).

$$F_{hkl} = |F_{hkl}| \, e^{\phi i_{hkl}} \quad (3)$$

The amplitude of the structure factor is directly proportional to the square root of the measured intensity from the diffraction (Equation 4).

$$|F_{hkl}| \propto \sqrt{I_{hkl}} \quad (4)$$

For a unit cell, the structure factor F_{hkl} is the summation of all scattered waves by the individual atoms present in the unit cell (Equation 5).

$$F_{hkl} = \sum_{j=1}^{n} f_j \, e^{2\pi i(hx+ky+lz)} \quad (5)$$

Where *n* is the number of diffracting atoms; f_j is the atomic scattering factor (depends on the atom type and diffraction direction); *x, y* and *z* are atomic coordinates; $e^{2\pi i(hx+ky+lz)}$ is the phase factor. Structure factors can be used to calculate the electron density at a particular point (*xyz*) in a unit cell through an inverse Fourier transform[126] (Equation 6).

$$\rho(xyz) = 1/V \sum |F_{hkl}| \, e^{\phi i_{hkl}} e^{-2\pi i(hx+ky+lz)} \quad (6)$$

Thus, it is essential to know both the amplitude and the phase of the diffracted waves to calculate electron density. The amplitude can be obtained from the intensity of the diffraction spots, but the phase information is lost during the process of measurement, which is known as

'Phase problem'. For small molecules, the phases of reflections can be determined directly from their diffraction data by analyzing strong reflections in specific regions of the unit cell. This approach is known as a direct method of phase determination. However, for larger molecules such as proteins, the phase problem is typically solved using methods such as experimental phasing or molecular replacement.

2.2.4.3.1 Molecular replacement

Molecular replacement (MR) is a computational method used in X-ray crystallography to solve the phase problem while determining the structure of proteins. In this method, the structure of a homologous protein is used to determine the phases of the target protein. The starting homologous model should have at least 40% sequence identity with the target protein[126,127]. The known structure is oriented and positioned within the unit cell of the target molecule so that the calculated diffraction pattern matches the experimental diffraction data. The search model can be oriented and positioned in three-dimensions each, making it a six dimension problem. This problem can be solved by the equation given below (Equation 7):

$$X' = [R]\, X + T \quad (7)$$

where X represents a set of vectors corresponding to the original atomic positions of the molecule and X′ denotes the new position of the atoms after transformation, R and T represent the rotation matrix and translation vector, respectively[128]. The equation signifies that, to transform the original molecule from its starting position, X to a new position X′, the molecule is first rotated by the rotation matrix [R] and then translated by the vector T. In the conventional MR approach, the Patterson function of a known model can be compared with the Patterson function of experimental data to determine the model orientation and position in the unit cell. The Patterson function gives a map of interatomic vectors in which each peak corresponds to a vector between two atoms in the unit cell, showing the distances and directions between atoms. The maps generated by the two Patterson functions will agree well only if the model is accurately oriented and positioned correctly within the unit cell. The Patterson function, *P (u, v, w)* is defined as (Equation 8):

$$P(u,v,w) = \frac{1}{V}\sum_{hkl} |F_{hkl}|^2 \cos 2\pi\,(hu + kv + lw) \quad (8)$$

Where $P(u,v,w)$ is the Patterson function; $|F_{hkl}|^2$ is the squared structure factor amplitudes for the reflection (*hkl*) and V is the unit cell volume. *u, v, w* are distance between the atoms at positions (x_1, y_1 and z_1) and (x_2, y_2 and z_2), respectively.

The Patterson function enables us to divide the six variables problem (three for each position and orientation of the model) into two problems, the rotation matrix and the translation vector, as solving a six variable problem becomes computationally expensive. The rotation function relies on the principle that self-vectors depend only on the molecule's orientation, not its position. The rotation matrix is calculated by rotating the model Patterson function (P_{self}) (from self-convolution of electron density) to align with the observed Patterson function (P_{cryst}) (from experimental data). The rotation function F(R) can be denoted as the sum of the product of these Patterson functions at individual points where r is the integration radius (Equation 9).

$$F(R) = \int_r P_{cryst}\ (u)\ P_{self}\ (Ru)du \quad (9)$$

After determining the orientation of the molecule in the unit cell, the next step is to find the translation vector, which locates the definite molecule position. As the molecule is translated, the intermolecular vectors shift. The correct position is achieved when the cross vectors from the observed and model Patterson functions align well[128]. The translation function T (*t*) can be given as (Equation 10):

$$T(t) = \int_v P_{cryst}\ (u)\ P_{cross}\ (ut)du \quad (10)$$

The experimental Patterson function is abbreviated as P_{cryst} and the Patterson function P_{cross} is determined from the probe oriented in the crystal, translation vector is denoted as *t* whereas intermolecular vector present between two symmetry related molecules is given as *u.*

Several softwares are available for MR solutions, this includes Phaser[129], AMoRe[130] and MOLREP[131]. This study used the CCP4 Phaser module[132] to solve the structures. Phaser is a widely utilized program that automates the phase problem by employing maximum likelihood probability theory this approach evaluates how well the model aligns with the data through a likelihood function, which represents the probability of the observed data given the model. Phaser enables rapid and sensitive searches for rotation and translation likelihood and includes likelihood-based corrections for anisotropy, along with a pruned tree search feature that

accommodates multiple molecules, thus making it effective for complex molecular replacement challenges.

2.2.4.4 Model building, refinement and density improvement

Molecular Replacement gives the initial electron density map, based on which a three-dimensional model of the molecule is built from the X-ray diffraction data. Model building requires interpretation of the electron density to assign the correct positions of atoms, residues and ligands. $2F_O$-F_C and F_O-F_C are the two Fourier-derived electron density maps used to identify the positioning of atoms. The $2F_O$-F_C map shows the overall electron density, while the F_O-F_C map highlights the differences between observed F_O and the calculated F_C densities, revealing the presence of unmodelled atoms or residues[133]. Software tools such as COOT[134] were used to view and build the model. The starting model should be validated as it may have geometric errors such as incorrect backbone torsion angles, bond angles and bond lengths. Therefore, the initial model needs to be refined iteratively against the experimental data, which minimizes the differences between the observed and calculated structure factors. Programs like Phenix[135] and Refmac[136] were used for model refinement which use different algorithms to minimize discrepancies between the model and the observed data while preventing the geometry of the structure. After the computational refinement cycle, the model is further improved by visually examining the map and adjusting the fit accordingly. This process was repeated iteratively, leading to the optimization of both the map and the model. During refinement, other factors such as atomic displacement parameters (B factors), R-factors, bulk solvent, constraints and restraints were also accounted for to ensure a good fit quality.

2.2.4.4.1 R-factor

R-factor was used to assess the agreement between the observed (F_{obs}) and the calculated (F_{cal}) structure factors during the refinement process. It is calculated as follows (Equation 11):

$$R_{factor} = \frac{\sum ||F_{obs}| - |F_{cal}||}{\sum |F_{obs}|} \quad (11)$$

The R-factor was monitored to ensure model improvement during the refinement process. As refinement progresses, the R-factor is expected to decrease as the model fits better with the experimental data. However, the R-factor alone does not guarantee the correctness of a model, as overfitting can occur. This is why additional parameters like R-free are also used to assess

model quality[137,138]. During refinement, a subset of 5%-10% of the reflections is randomly excluded from the refinement process. These reflections are referred to as the test set and the R-free was calculated using only this set. The rest of the reflections (containing 90% data) are known as the working set which were used to optimize the model. The R-free value is an indicator of how well the refined model predicts the test set data and prevents overfitting[139]. The R-free value is slightly higher than the R-work and the difference between them depends on the resolution of the data. Large difference between R-free and R-work indicates overfitting. A good resolution data can give a difference of approximately 0.02, whereas data with low resolution can give a difference of 0.05-0.08[140].

2.2.4.4.2 B-factors or atomic displacement parameters (ADP)

B-factors or atomic displacement parameters (ADP) or temperature factors, and they are used to describe the thermal motion or static disorder of atoms in a crystal structure. The B-factor (B) is related to the mean square displacement of an atom (u^2) through the given equation 12[141]:

$$B = 8\pi^2 (u^2) \quad (12)$$

ADP can be isotropic or anisotropic. The isotropic ADP assumes that atomic motion is the same in all directions (spherical) and is used when the data resolution is low[142]. The anisotropic model is used for high resolution and assumes atomic displacements in different directions to accurately represent the atomic vibrations. It is described by six parameters that account for the movement in three-dimensional space. At lower resolution, the data-to-parameter ratio may be insufficient for detailed macromolecular refinement. As a result, isotropic ADP refinement, often combined with TLS (Translation, Librarian and Screw) parameters, is commonly applied[142].

2.2.4.4.3 Geometric constraints and restraints in refinement

Restraints and constraints help to maintain the chemically relevant fitting of the experimental data. This requires prior information regarding the bond angles, length, torsions, hydrogen bonds and van der Waals interactions. Thus, 'restrained refinement' is majorly used in macromolecule structure determination as it combines the use of both experimental data and allowed geometry. The primary restraints used in macromolecule structure refinement are stereochemical restraints, which define the geometry of ligands and monomer residues[142].

Constraints are more stringent than restraints as they use mathematical relations and enforce exact relationships between atoms to reduce independent parameters.

2.2.4.4.4 Bulk solvent correction

Protein crystals contain about 30%-70% solvent, most of which are disordered in the crystal lattice between protein molecules[143]. These disordered solvent molecules contribute to the low-resolution diffraction, making it important to perform bulk solvent correction to obtain a good-quality electron density map. The electron density of the continuous bulk solvent that surrounds the protein is about 0.33 e/Å^3 and for the protein, it is about 0.43 e/Å^3 [144]. The total structure factor (F) is calculated as Equation 13:

$$F = k\{F_{calc}\, exp[-\Delta B(sin\theta/\lambda)^2] + d_{solv}F_{solv}\, exp[-B_{solv}(sin\theta/\lambda)^2]\} \quad (13)$$

where scale factor is denoted by k, F_{calc} is the structure factor of protein, d_{solv} denotes electron density of bulk solvent, F_{solv} denotes the solvent mask scattering, which is a step function having value 1 in the solvent and 0 outside[145].

2.2.4.5 Validation

Structure validation is a critical process in ensuring the accuracy and reliability of macromolecular models. It is also an essential step during structure refinement to evaluate the improvement in the model as the refinement progresses. Refinement should continue until all structural parameters reach their optimal values, aligning with the quality of the experimental data. The validation process assesses the geometric and stereochemical quality of the protein model, as well as its fit to the experimental data. Ramachandran plots are one of the tools that help to verify the allowed and disallowed conformations of backbone dihedral angles and also checks the bond lengths, bond angles and torsional angles to make sure that the structure complies with the known chemical constraints. Other parameters, such as R-free and R-factor, are used to evaluate how well the model fits the observed diffraction data. Additionally, tools such as MolProbity, provide an analysis of steric clashes, side-chain rotamers and hydrogen bonding patterns[146].

2.2.5 Computational methods

2.2.5.1 Molecular Dynamics simulation

Molecular Dynamics (MD) simulation is a computational method used to model the physical movements of atoms and molecules over a period of time. It can provide insights into the structural dynamics, thermodynamics and mechanical properties of biomolecules like proteins, nucleic acids, ligands, etc. In MD simulations, the behaviour of a group of interacting particles is tracked overtime by solving Newton's equations of motion (Equation 14):

$$F_i(t) = m_i \frac{d^2 r_i(t)}{dt^2} \quad (14)$$

where r_i is the position vector of the ith particle at time *t*, F_i is the force which acts on the ith particle and m_i is the particle mass[147]. The force on the ith atom is derived from the potential energy function *U(r)* of the system where *r* is the atom position (Equation 15):

$$F_i = -\frac{\partial U(r)}{\partial r_i} \quad (15)$$

The force allows for the calculation of each atom's velocity, position and acceleration by solving Newton's equations of motion. By repeating this process at distinct time intervals, the movement of individual particle is determined over time. As a result, MD simulations generate a series of sequential conformations that help to map the energy landscape of the system. In this work, simulations were performed using the Desmond module from the Schrödinger suite[148]. Following steps were performed for simulation runs:

a. **System preparation:** The protein molecule is placed in a defined simulation box filled with solvent, typically water. After adding solvent, counterions are added to neutralize the system and mimic the physiological condition. Non-bonded and bonded parameters are further assigned using the force field. The choice of force field is also essential at this stage, as it dictates the potential energy functions and parameters that describe the interactions within the system.
b. **Energy minimization and equilibration:** Energy minimization is performed to remove unfavourable steric clashes and optimize the system's initial geometry. This step adjusts atomic positions to reach a local energy minimum. Further temperature (NVT) and

pressure (NPT) equilibration is performed using the respective thermostat and barostat. This is done to stabilize the system under the desired thermodynamic conditions.

c. **Production run:** Following system set up, the simulation run involves integrating Newton's equations of motion over a specified period of time. At each timestep, the positions, velocities and forces acting on the atoms are determined, generating a trajectory that captures the dynamics of the system. These trajectories can be further analyzed to extract meaningful information regarding the stability, interactions, conformational changes, etc. within the system.

2.2.5.2 DCCM, dynamically coupled sectors and independent components

Molecular dynamics simulation trajectories provide information regarding the dynamic behaviour of the proteins. These trajectories can be analyzed to gain information regarding regions of the protein which have highly correlated or anti-correlated motions indicating that these regions are dynamically coupled and can be represented as the dynamical cross-correlation matrix (DCCM), dC_{ij}[149] (Equation 16),

$$dC_{ij} = \langle r_i r_j \rangle - \langle r_i \rangle \langle r_j \rangle \qquad (16)$$

Where, r_i and r_j are the position vectors of the i^{th} and j^{th} Cα atoms, respectively of the protein, at time t. In DCCM, a positive correlation value of C_{ij} indicates that the two residues move in the same direction, while a negative C_{ij} suggests anti-correlated motion, where the residue moves in the opposite direction. The value C_{ij}=1, indicates the two residues are perfectly correlated, meaning their movement has the same period and phase. In contrast, C_{ij}= -1 signifies complete anti-correlation, where residues have the same period and phase, but move in opposite directions. If the residues have the same period and phase but their displacements are oriented at 90^o angle, then the C_{ij}= 0. In the DCCM, strong correlations between residue motions appear as large off-diagonal peaks. This highlights the regions with significant interaction or coordination in their movements[149]. This information is crucial for understanding processes like allosteric transitions and ligand binding.

The DCCM obtained from MD studies can be further analyzed by using spectral/eigen decomposition algorithms to obtain the sectors of dynamically coupled amino acids. In spectral decomposition the matrix is represented as a product of three matrices (Equation 17):

$$\tilde{C}^{P} = \tilde{V}\tilde{\Delta}\tilde{V}^{T} \qquad (17)$$

Where $\tilde{\Delta}$ represents a diagonal matrix containing the eigenvalues and $\tilde{V}$represents a matrix with corresponding eigenvectors. Spectral decomposition of DCCM gives eigenmodes, which correspond to the specific patterns or modes of correlated or anti-correlated motions. Each eigenvector represents a distinct eigenmode of correlated or anti-correlated motion, while the eigenvalue indicates the significance of that mode. Further *k** significant eigenmodes are determined to capture the most meaningful correlated motions in the system. Typically, the top few eigenmodes are selected[150].

Dynamically couple sectors are defined based on the residues that contribute strongly to each of the significant eigenmodes. They share similar motion patterns and are therefore dynamically coupled. This clustering of residues based on their contribution to the eigenmodes leads to the identification of dynamical sector groups of residues that are functionally or structurally linked by their correlated dynamics. The top *k** eigenmodes of a correlation matrix can be transformed into independent components by using independent component analysis (ICA)[150,151] (Equation 18).

$$\tilde{V}^{p}{}_{1....k} = W^{P}\,\tilde{V}_{1....k} \qquad (18)$$

Where, $\tilde{V}_{1....k}$ represents the top *k** eigenvectors, $\tilde{V}^{p}{}_{1....k}$represents the *k** maximally independent components and W is the matrix deduced from the ICA. Independent components often give information regarding the domain movement, conformational changes or collective motions that contribute to the function of protein.

2.2.6 Cell based assays

2.2.6.1 Cell culture transfection

Cell culture transfection is a technique used to introduce foreign DNA or RNA into eukaryotic cells to study the function of that gene, for transient protein expression, to create a stable cell

line which involves integration of foreign DNA into the host genome. There are several methods which can be used to facilitate the uptake of these nucleic acids by the cells[152]. These methods are:

Chemical methods: This uses lipofectin, where cationic polymers or liposomes are used to encapsulate the negatively charged nucleic acids. This complex then integrates the cell membrane enabling the entry of nucleic acids into the cell cytoplasm.

Physical methods: In physical methods, techniques like electroporation (involves use of electric field to create pores in the cell membrane) or microinjection are used.

Viral vectors: Viral vectors which can infect host cells and introduce genetic material in the cells are also used for transfection.

2.2.6.2 Wound healing assay

Wound healing assay (scratch assay) is used to study cell migration and processes involved in wound healing. This assay helps to understand how cells migrate to close a wound and the factors that influence this process. In order to perform wound healing assay, cells are first cultured in a monolayer. Once a uniform layer is formed, a wound or a scratch is created using a pipette tip or specialised wound making tools. After the wound is created, the cells are incubated in a medium containing specific growth factors, cytokines or ligands which promotes healing. The area of the wound is observed and imaged at regular intervals to monitor cell migration into the scratched area. The extent of closure is then quantified by measuring the area of wound overtime and analyzed using imaging software.

2.2.6.3 Fluorescence-Activated Cell Sorting (FACS)

FACS is a flow cytometry technique used to sort cells from a heterogeneous cell population. Cells fluorescently labelled for specific markers or express fluorescent proteins after transfection can be sorted using this technique. Cells can also be sorted depending on their physical properties, like granularity and size. In this process, a suspension of labelled cells passes through the flow cytometer one by one. A laser beam is focused on the cells, exciting the fluorescent tags or proteins.

As the lazer interacts with each cell, light is scattered. Scattered light is captured by two detectors. The side scatter (SSC) detector measures the light scattered at 90-degree angle to the laser. The SSC reflects cellular complexity and granularity. The forward scatter (FSC) detector measures the light scattered in the direction of the laser. This measure provides information about the cell size. Together, SSC and FSC data offer insights into the physical characteristics of different cell types within the population[153]. For instance, Figure 2.5 shows SSC vs FSC flow cytometry plot for whole blood cell population. Different cell types such as lymphocytes, monocytes and granulocytes show different scattering patterns based on their granularity and size[154]. In the presence of a fluorescence tag, the cells which emit fluorescence can be identified by their fluorescence emission and can be subsequently sorted for further experiments.

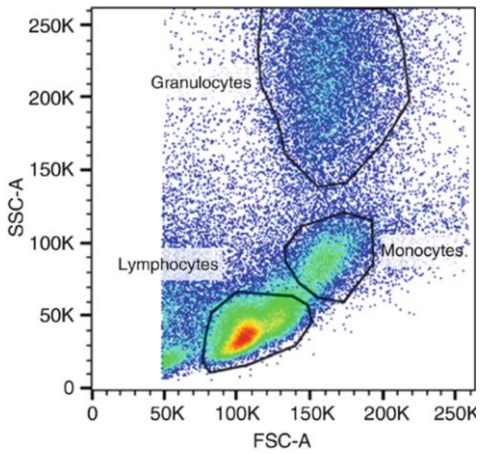


Figure 2.5: Side scatter (SSC) vs Forward scatter (FSC) plot for whole blood cells obtained from flow cytometry. *Image adapted from Proserpio et al.*[154].

Chapter 3

To understand the mechanism of RhoA activation by WGEF

3.1 Background

Non-canonical Wnt-PCP signalling regulates convergent extension (CE) movements during embryo development and adult tissue homeostasis through actin cytoskeleton remodelling, by activating small GTPases such as RhoA and Rac[24,25,155,156]. This signalling pathway gets initiated upon binding of the secreted glycoprotein ligands, Wnts, to their membrane receptors, Frizzleds. This event recruits adapter protein Dishevelled2 (Dvl2) to the cytosolic domain of Frizzleds. Subsequently, other scaffold protein known as *Dishevelled-associated activator of morphogenesis 1* (Daam1) is activated and it further recruits Rho Guanine nucleotide exchange factor, eventually leading to the activation of effector RhoA GTPase[23,24]. Activation of Rac GTPase takes place independent of Daam1 recruitment[25] (Fig 3.1A). As detailed in the Introduction chapter, Rho GTPases exerts signal relay by shuttling between GTP-bound active (ON) and GDP-bound inactive states (OFF). GEFs activate Rho GTPases by exchanging their bound GDP with GTP (Fig 1.7). Aberrant activation of the GTPases is often driven by the dysregulation of these regulatory proteins, which are involved in many pathological conditions, including cancer. For example, dysregulation of various GEFs, such as Vav 1/2/3 and Tiam 1/2, has been implicated in cancer and developmental diseases[157]. Therefore, understanding the mechanism by which RhoA is activated during non-canonical Wnt signalling is imperative for designing various drugs, as disruption in this pathway is shown to contribute to various cancers and developmental disorders[77,158].

Weak-similarity Guanine nucleotide Exchange Factor (WGEF) is a Dbl family GEF known to operate in the non-canonical Wnt11/Frizzled7 pathway, which is shown to activate the downstream GTPase, RhoA (Fig 3.1B). WGEF shares weak similarity with another member of the Dbl family, TIM (RhoGEF 5)[159]. WGEF is suggested to consist of four major domains, including an N-terminal inhibitory domain (NID) responsible for the autoinhibition of GEF activity, a catalytic Dbl-homology (DH) domain that facilitates the GDP to GTP exchange on RhoA, a pleckstrin homology (PH) domain that plays a regulatory role and contributes to the membrane localization of WGEF and an SRC Homology 3 (SH3) domain

that is involved in establishing protein-protein interactions with its binding partners (Fig 3.1C). Like many of the canonical Dbl GEFs, WGEF is known to exist in an autoinhibited state due to the interdomain interactions of the protein involving the NID and SH3 domains. Earlier studies have shown that WGEF interacts with Dvl2 as well as Daam1 during Wnt-PCP signalling[24]. However, the precise role of all these domains in the activation of RhoA is not clear. In this chapter, we report our attempts to elucidate the structural basis of WGEF-RhoA interaction, using X-ray crystallography to further our understanding of the RhoA activation mechanism as well as the consequent signal transduction events. This chapter details optimization of WGEF and RhoA constructs for forming their *in vitro* complex and subsequent crystallization studies.

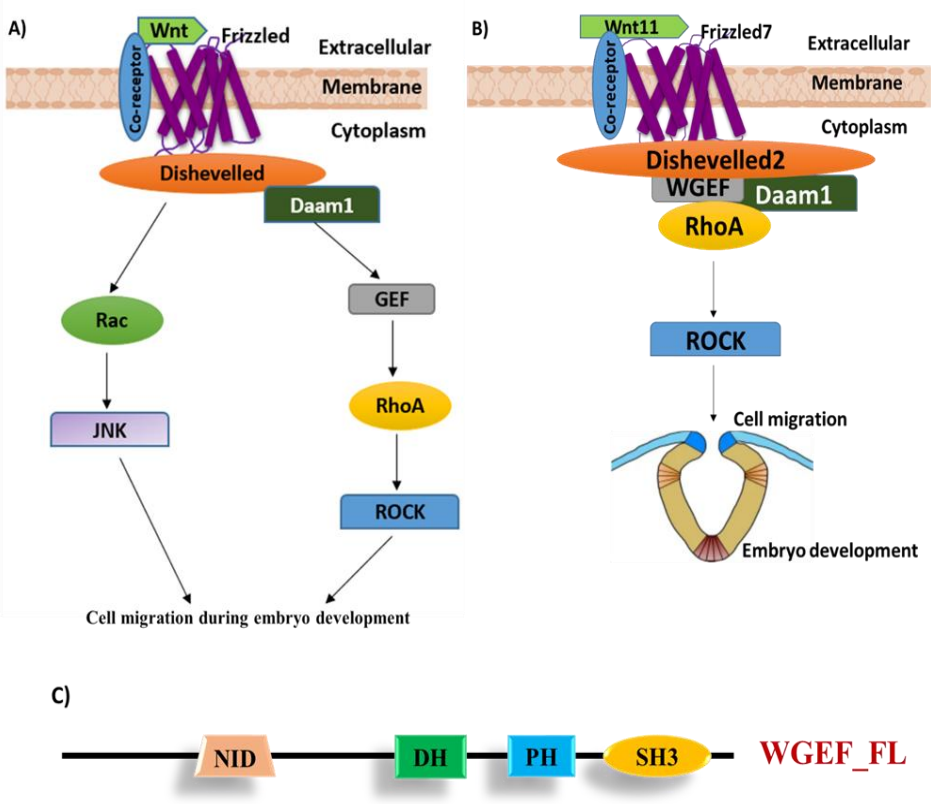


Figure 3.1: Schematic representation of non-canonical Wnt-PCP signalling. ***A)*** *Overview of RhoA and Rac GTPase activation during non-canonical Wnt signalling leading to cell migration during embryo development.* ***B)*** *Detailed pathway for activation of RhoA axis during Wnt11/Frizzled7 activated non-canonical Wnt signalling.* ***C)*** *Different domains of WGEF protein; N-terminal inhibitory domain (NID), Dbl-homology (DH) domain, pleckstrin homology (PH) domain and SRC Homology 3 (SH3) domain.*

3.2 Methodology

3.2.1 Cloning

To express the target protein in the *E.coli* system, the gene encoding the protein of interest is cloned into a suitable *E.coli* expression vector. These vectors are optimized to achieve higher yields of desired protein. In this study, different constructs of *Xenopus laevis* WGEF (xWGEF/ Arhgef19), spanning various ranges of amino acids, were generated by cloning the respective gene segments into a modified pGEX-6P-1 (p3E) vector. To enhance the solubility of the protein, PreScission protease cleavable GST tag was fused at the N-terminal of the protein (Fig 3.2). Thus, the GEF constructs were overexpressed as N-terminal GST tag fusion proteins, which also facilitated the purification of proteins using glutathione sepharose beads based affinity purification method. After affinity chromatography, the fused GST tag was removed by using PreScission protease.

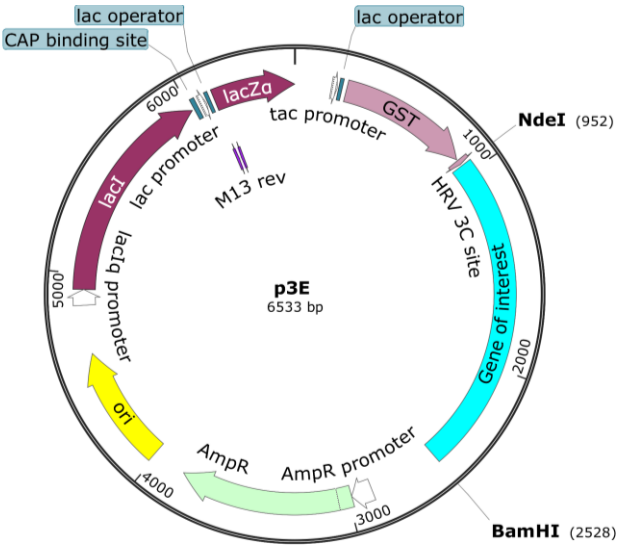


Figure 3.2: Sequence map of p3E vector highlighting the gene of interest and other plasmid features.

The cloning process begins with the amplification of the xWGEF genes via PCR. The amplified gene products were digested with restriction enzymes (Nde1 and BamH1) to create sticky ends compatible with the vector. The vector was similarly linearized using the same restriction enzymes to facilitate the ligation of the digested gene. After ligation, *E.coli* DH5α cells were transformed using the reaction mixture, and the resulting colonies obtained on the

selection media plate were cultured for plasmid amplification and purification. Restriction digestion was used to screen plasmids for ligated target gene, followed by plasmid sequencing based confirmation. Confirmed positive clones were subsequently used for expression and purification of the target proteins.

3.2.2 Protein expression and purification

Recombinant xWGEF constructs were expressed in *E.coli* C41 (DE3) strain (Table 3.1). Secondary cultures were grown at 37°C in LB broth supplemented with 100 µg/mL ampicillin. After IPTG induction, cells were incubated at 18°C/16 h. Following incubation, cells were centrifuged at 4000 RPM/20 min. The resulting cell pellet was resuspended and lysed in respective Buffer A (50 mM Tris pH 8-8.5, 500 mM NaCl, 2 mM β-met, 2 mM EDTA and 5-10% glycerol) supplemented with 2 mM PMSF, PI tablet, 20 mM sucrose and approximately 1 mg lysozyme. Lysis was performed using a homogenizer at 4°C/20 min at 350-400 bar pressure. The lysate was spun at 20,000 RPM/1 h to separate the cell debris from the supernatant. The clear supernatant was passed through a 0.4 µm syringe filter to eliminate residual debris and then used for protein purification. For purification, the filtered supernatant was loaded onto the Glutathione Sepharose-4B (GSH) column pre-equilibrated with Buffer A, allowing the GST-tagged xWGEF proteins to bind. The flow-through was collected, and the column was washed with Buffer A to remove non-specifically bound proteins. After washing, the target protein was eluted by using Buffer A containing reduced glutathione (10 mM concentration). Fractions from pellet, supernatant, flow-through, wash and elutes were analyzed on SDS-PAGE to monitor the expression, presence of contaminant protein bands, solubility and degradation of the target protein. The eluted protein was subjected to overnight dialysis for buffer exchange (10 mM Tris pH 8-8.5, 350-250 mM NaCl, 2 mM DTT, 5-10% glycerol). GST tag of the protein was cleaved by adding 10 fold molar excess of PreScission protease in the dialysis bag. Next day, the dialyzed and cleaved protein was desalted to remove the residual glutathione, followed by reverse affinity purification to eliminate the cleaved GST tag and uncleaved protein. The target protein present in flow-through after the reverse affinity column was concentrated and further purified by SEC using a buffer consisting of 10 mM Tris pH 8-8.5, 350-150 mM NaCl, 2 mM DTT, 5-10% glycerol. The purified protein was then concentrated, snap-frozen in liquid nitrogen and stored at -80°C.

Similar to xWGEF constructs, recombinant RhoA with an N-terminal non-cleavable Strep-tag II, cloned in modified pET33b vector with kanamycin resistance, was expressed in B834 (rare) cells. For purification, the clear supernatant obtained after cell lysis was loaded onto Strep-Tactin XT Sepharose slurry pre-equilibrated with Buffer A containing 50 mM Tris pH 8, 200 mM NaCl, 2 mM β-met, 5 mM EDTA and 5% glycerol. Following loading and washing steps, RhoA was eluted with a 5 mM desthiobiotin containing Buffer A. Eluted RhoA was further purified using SEC in a buffer consisting of 10 mM Tris pH 8, 150 mM NaCl, 2 mM DTT, 2 mM EDTA and 5% glycerol. The protein from the stable fraction was then concentrated and used for subsequent studies.

xWGEF constructs	Stability
xWGEF330 	Stable when co-expressed with RhoA
xWGEF404 	Stable when co-expressed with RhoA
xWGEF386 	Not stable

Table 3.1: Different constructs of xWGEF made for crystallization with RhoA

3.2.3 Co-expression and purification of xWGEF and RhoA

For co-expression, BL21*(DE3) cells were co-transformed with plasmids carrying RhoA (with kanamycin selection gene) and xWGEF (with ampicillin selection gene) genes. Cells were plated on LB agar supplemented with 100 µg/mL ampicillin and 50 µg/mL kanamycin antibiotics to selectively culture the cells co-transformed with both the plasmids. Once the co-transformed colonies were obtained, they were grown for co-expression in the presence of both antibiotics, following the previously discussed protocol. After cell lysis and clarification of the supernatant, the co-expressed proteins were purified using GST affinity chromatography. This allowed purification of the GST-tagged xWGEF and co-expressed RhoA, which was bound to xWGEF. Further purification steps involved dialysis, GST tag cleavage, desalting and reverse

affinity purification, similar to the steps discussed earlier involving xWGEF purification. To separate *apo* xWGEF from the RhoA-bound xWGEF, IEC was employed. For this, the salt concentration of the protein mixture was reduced to 50 mM NaCl and then loaded onto a Q sepharose column, pre-equilibrated with 10 mM Tris pH 8, 50 mM NaCl, 2 mM EDTA, 2 mM DTT and 5% glycerol. Next, the column was washed with the same low salt buffer, following which the bound protein was eluted using a gradient of 1M NaCl containing buffer (10 mM Tris pH 8, 1 M NaCl, 2 mM DTT, 2 mM EDTA and 5% glycerol). The purified protein was further subjected to the SEC in a buffer containing 10 mM Tris pH 8, 250 mM NaCl, 2 mM EDTA, 2 mM DTT and 5% glycerol. Fractions corresponding to the xWGEF-RhoA complex were concentrated and used for the protein crystallization. Both xWGEF404 and xWGEF404-RhoA proteins were analyzed by SEC-MALS to determine their molecular weight. The SEC column connected in line with the MALS setup was equilibrated with a buffer comprising 10 mM Tris pH 8, 250 mM NaCl, 2 mM DTT, 2 mM EDTA and 5% glycerol.

3.2.4 Crystallization

Protein samples obtained from the size exclusion step were concentrated and spun at 14,000 RPM/20 min to eliminate soluble aggregates or precipitation, if any. These samples were subjected to crystallization trials using commercial crystallization screening suites, including JCSG+, PACT, proplex and PEGs. In a 96-well, 3-drop sitting drop plate, 30 µL of precipitant was added to each reservoir well using a multichannel pipette. Different concentrations of the protein samples (xWGEF or xWGEF-RhoA) were mixed with the precipitant in a 1:1 ratio, by keeping the drop volume 200 nL, using the Mosquito (TTP labtech) crystallization robot (liquid handling system). The sitting drop plates were then sealed and incubated at 20°C, and the drops were periodically observed for protein crystal formation.

3.3 Results

3.3.1 Co-expression of xWGEF330 and RhoA stabilize xWGEF330

For structural studies, a stable *in vitro* WGEF-RhoA complex is essential. To obtain such complexes, optimization of the constructs, spanning different domain boundaries, is imperative. Our initial construct design was guided by *in silico* tools, such as disorder prediction and AlphaFold predicted models. However, the AlphaFold predicted structure of WGEF indicates a significant portion of the protein to be disordered and even many of the ordered regions were predicted to have low confidence scores (Fig 3.3). Secondary structure and disorder predictions using *Phyre2* (secondary structure prediction software) also projected a similar trend, suggesting these regions to be disordered. Therefore, we designed WGEF constructs with varying construct boundaries by carefully excluding these disordered regions.

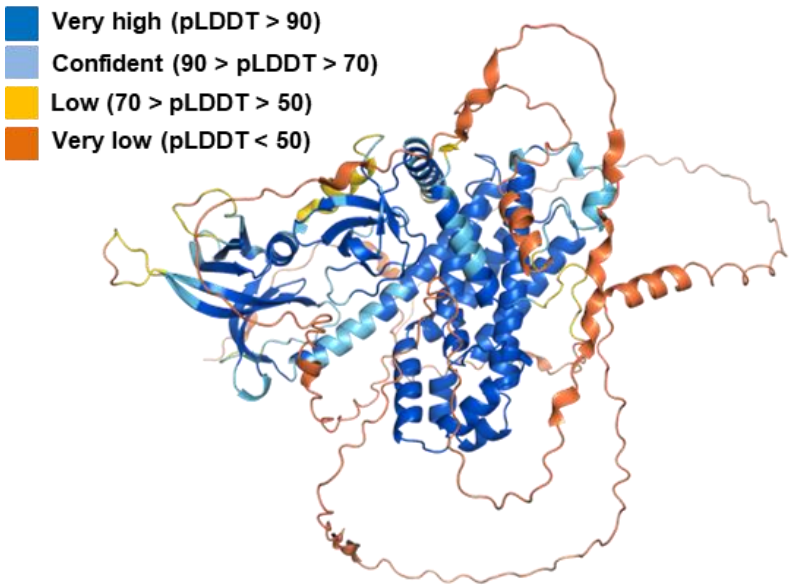


Figure 3.3: Overall AlphaFold predicted structure of hWGEF (Uniprot ID: Q8IW93). *Regions of the structure are coloured based on per-residue confidence score (pLDDT).*

Earlier studies conducted in our laboratory have shown that some human WGEF (hWGEF) constructs were either insoluble and/or poorly expressed in *E.coli*. Therefore, we opted for WGEF from *Xenopus laevis* (xWGEF) as a surrogate system for further studies. To ensure proper folding, stable expression and purification of WGEF, we excluded the major N-terminal disordered region of xWGEF, as predicted by *Phyre2* (Fig 3.4). The resulting truncated construct of xWGEF (330-856), hereafter referred to as xWGEF330, retained all the

domains and it was used for complexation with purified RhoA (Fig 3.5). Unfortunately, the SEC purification profile of xWGEF330 revealed that most protein eluted under the void peak, indicating the protein to be aggregated (Fig 3.6). However, a small amount of the protein was found to be stable and non-aggregated, and it was further used for complexation with RhoA (Fig 3.7). xWGEF330 obtained from this step was mixed with RhoA in a molar ratio 1:3 and subjected to SEC. The SEC peak corresponding to the protein complex was identified from SDS-PAGE and it was concentrated for further studies. However, the xWGEF330-RhoA complex started precipitating when concentrated above 2 mg/mL, indicating instability of the complex at higher concentrations (Fig 3.8). To circumvent this problem, various buffers, including CHES, Trizma and Bis-tris propane, were used to identify optimal conditions that might stabilize the protein complex at higher concentrations. However, all buffers produced similar results, with no significant improvement in the stability.

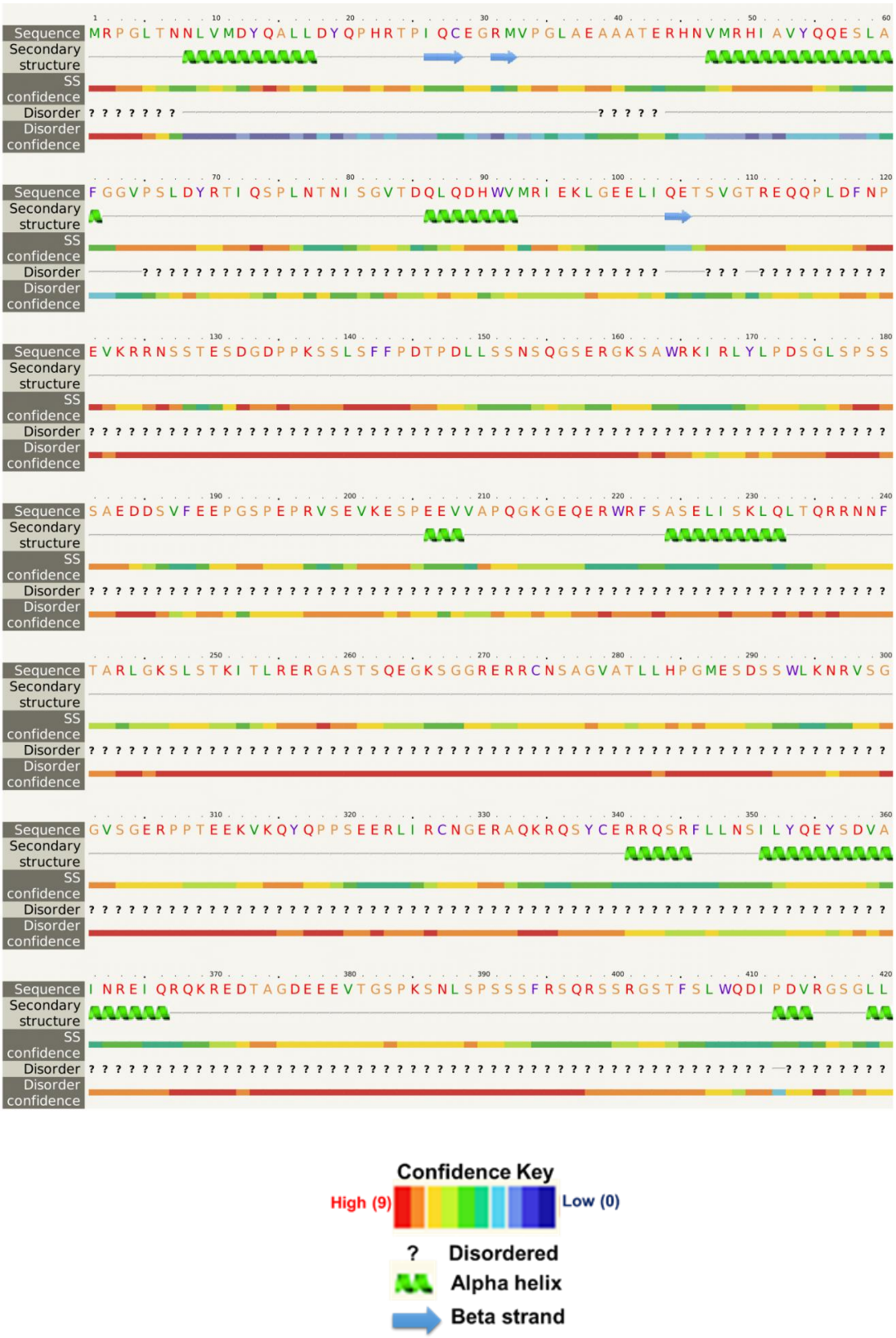


Figure 3.4: Secondary structure and disorder prediction of xWGEF protein by Phyre2. *Residues from 1-420 are shown, where the majority of the protein is shown to be disordered. The confidence key for the structure and disorder prediction is given below.*

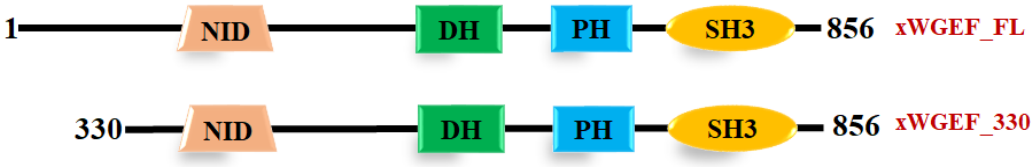


Figure 3.5: Schematic representation of xWGEF_full length and xWGEF330 protein domains.

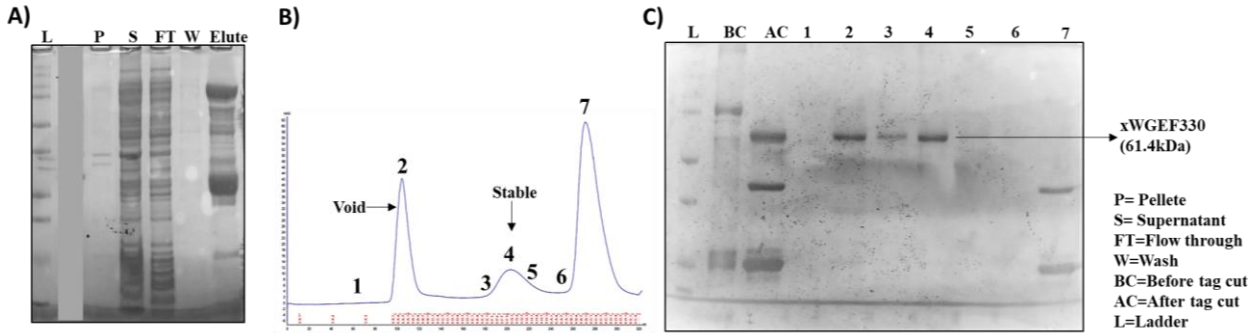


Figure 3.6: xWGEF330 purification. A) *SDS-PAGE for GST affinity chromatography* ***B)*** *Size exclusion chromatogram.* ***C)*** *SDS-PAGE for size exclusion chromatography.*

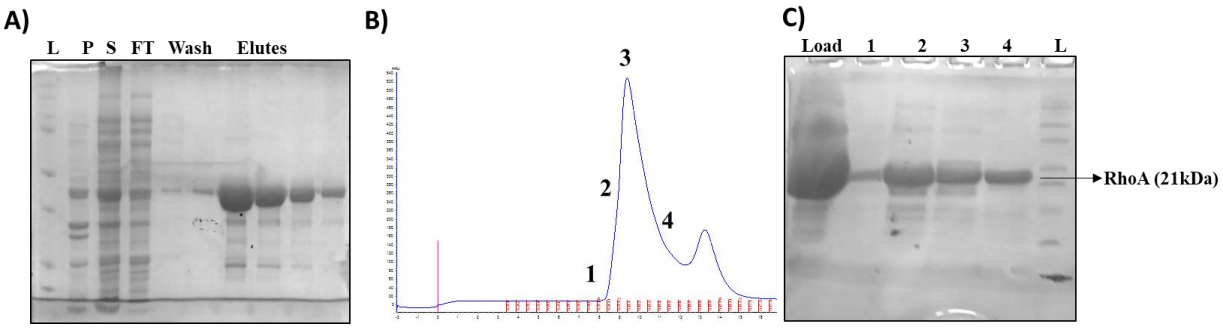


Figure 3.7: RhoA purification. A) *SDS-PAGE for GST affinity chromatography* ***B)*** *Size exclusion chromatogram* ***C)*** *SDS-PAGE for size exclusion chromatography.*

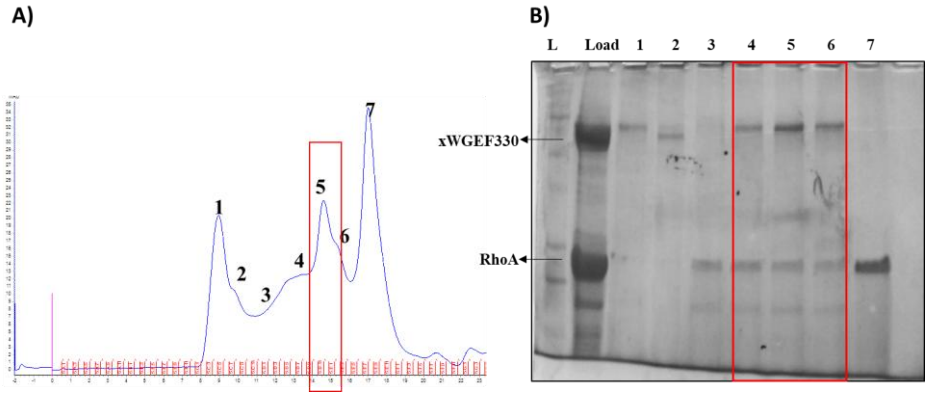


Figure 3.8: xWGEF330 and RhoA complexation. A) *Size exclusion chromatogram for complex* ***B)*** *SDS-PAGE for size exclusion chromatography. Red box indicates the peak containing both proteins.*

Francis & Page report that the co-expression of unstable proteins with their cognate binding partner might improve stability of the unstable proteins[160]. Based on this, we co-expressed GST-tagged xWGEF330 with streptactin-tagged RhoA and purified the xWGEF330-RhoA complex using the GST affinity purification method. Formation of the xWGEF330-RhoA complex on the GST affinity column was detected using SDS-PAGE (Fig 3.9). The peak corresponding to the protein complex obtained from SEC was concentrated and set up for crystallization screening by using protein concentrations up to 15 mg/mL. Thus, it is clear that co-expression of xWGEF330 with RhoA stabilizes the protein complex, which is contrary to what we had observed previously. Forming the *in vitro* complex of xWGEF330 & RhoA on the SEC column, after individually purifying them, resulted in precipitation of the complex when the concentration of protein was close to 2 mg/mL. Despite numerous crystallization trials using various screens and protein concentrations, it did not produce any crystal hits.

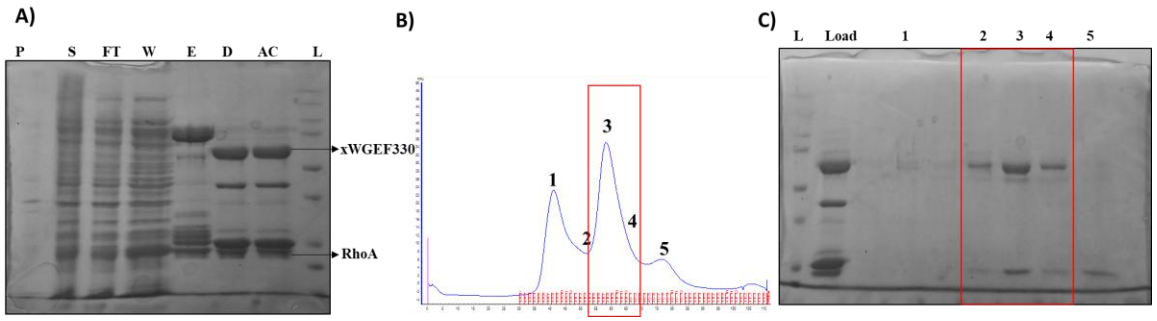


Figure 3.9: xWGEF330 and RhoA co-purification. A) *SDS-PAGE for GST-pulldown affinity chromatography* ***B)*** *Size exclusion chromatogram.* ***C)*** *SDS-PAGE for size exclusion chromatogram. Red box indicates the peak containing both proteins.*

3.3.2 Crystallization of xWGEF404 and RhoA complex

Since our initial crystallization trials of xWGEF330 and RhoA complex met with no success, we further optimized the xWGEF construct by truncating the construct to eliminate the NID region. This new construct will be referred to as xWGEF404 (Table 3.1). The idea was to obtain the GEF that is not completely autoinhibited (Fig 3.10). The modified xWGEF construct was cloned into the p3E vector, encoding an N-terminal GST tag (Fig 3.11).

Figure 3.10: Schematic representation of xWGEF404 protein domains.

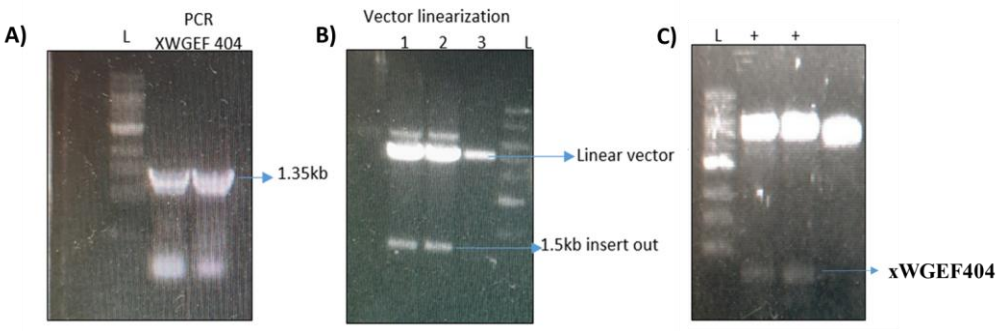


Figure 3.11: Cloning of xWGEF404 in p3E vector. A) *xWGEF404 PCR product* ***B)*** *Linearization of the p3E vector.* ***C)*** *Restriction digestion confirmation of cloned xWGEF404. Positive clones are denoted with '+' symbol.*

Following successful cloning and expression, solubility tests for xWGEF404 were done by expressing the protein in different *E.coli* expression cell lines, such as BL21*, C41, Rosetta and B834. This exercise was meant to determine the optimal cell line that yields the most stable protein. Results indicated that in all tested cells, the majority of expressed xWGEF404 protein was found to be insoluble and appeared in the pellet fraction (Fig 3.12), again suggesting that xWGEF404, too, was unstable when expressed by itself. We approached this problem by co-expressing xWGEF404 with RhoA (contained Strep-tag II) and the complex was purified using the GST pulldown affinity method. Co-expression significantly improved the solubility and stability of the complex (Fig 3.13).

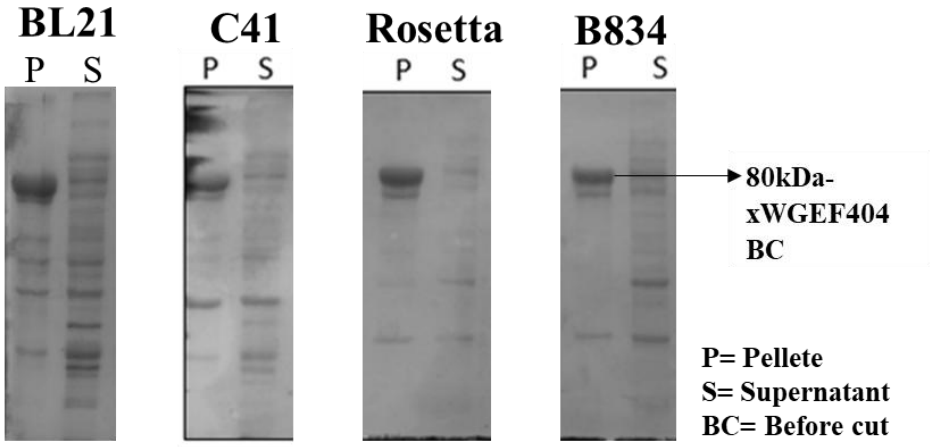


Figure 3.12: Test expression of xWGEF404 in BL21*, C41, Rosetta and B834 E.coli expression cell lines.

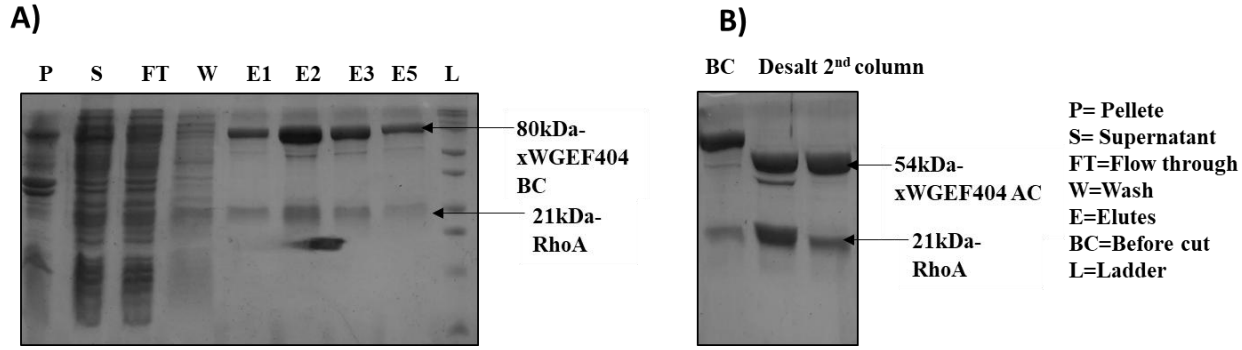


Figure 3.13: xWGEF404 and RhoA co-expression and affinity purification. A) *SDS-PAGE for GST-pulldown affinity chromatography.* ***B)*** *SDS-PAGE for protein after desalt and reverse affinity purification.*

After affinity purification and tag cleavage, ion-exchange chromatography was performed to separate the unbound xWGEF404 from the xWGEF404-RhoA complex (Fig 3.14A-C). The unbound xWGEF404 was found in the wash fraction, while the protein complex

was obtained from the elute fractions of the ion-exchange chromatography. Both the *apo* and protein complex were subjected to size exclusion chromatography, after which the respective protein fractions were concentrated and used for setting up crystallization trials (Fig 3.14 D-F). Despite extensive crystallization trials, no crystals of the protein complex were obtained. Although we successfully separated the *apo* protein from the protein complex using ion-exchange chromatography, achieved a symmetrical elution peak for the complex on the SEC column and concentrated the complex to a relatively higher concentration, crystallization attempts remained unsuccessful. This observation suggests that the complex may not be stable, which could be hindering the crystallization.

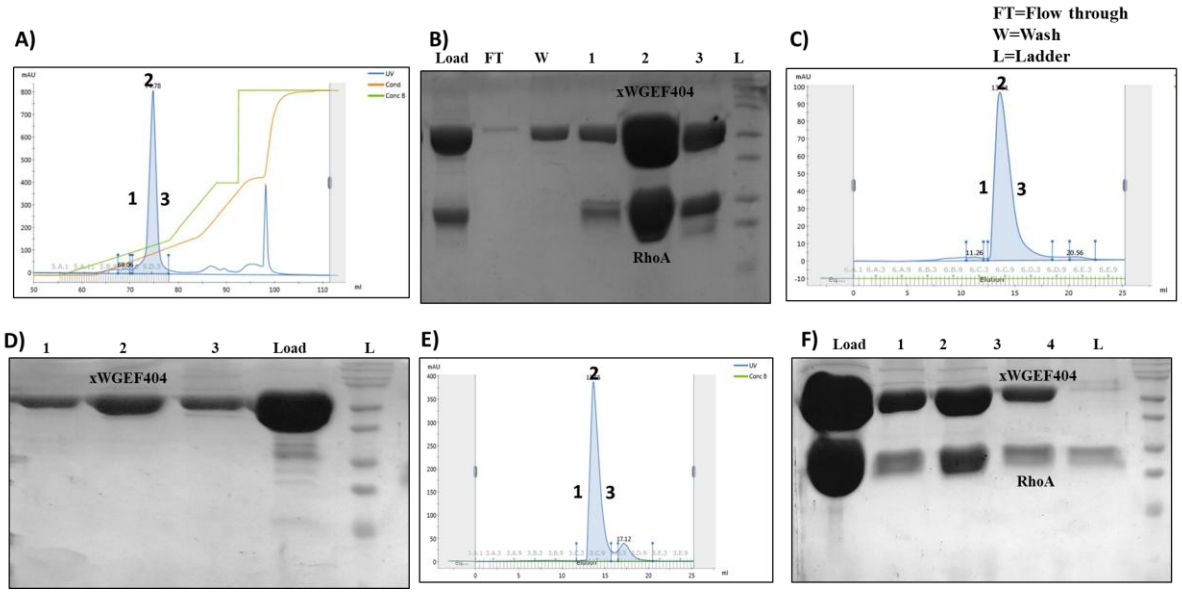


Figure 3.14: xWGEF404 and RhoA purification. A) *Chromatogram for Ion-exchange chromatography (IEC).* ***B)*** *SDS-PAGE for peaks corresponding to IEC protein* ***C)*** *Size exclusion chromatogram for xWGEF404* ***D)*** *SDS-PAGE for size exclusion chromatography of xWGEF404* ***E)*** *Size exclusion chromatogram for xWGEF404 and RhoA complex* ***F)*** *SDS-PAGE for size exclusion chromatography of xWGEF404 and RhoA complex.*

To further investigate whether the proteins eluted from size exclusion were forming a stoichiometric complex or merely co-eluting as higher oligomers of individual proteins, we subjected the protein complex to SEC-MALS. This technique combines SEC with MALS, which enables the determination of molecular weights of different fractions, as seen under the SEC. As a control, we used the *apo* xWGEF404. This analysis revealed that the proteins were not forming a stable complex, but rather they were co-eluting as individual oligomers (Fig 3.15). The lack of complexation could be due to the weak or transient interactions between them. To circumvent this problem, we adopted a strategy of reducing the purification steps and

setting up crystallization after IEC, with the hope to prevent dissociation and degradation of the complex due to the storage. This approach led to a crystal hit in a condition with , 0.1 M Tris pH 8.5 , 0.2 M TMAO (Trimethylamine N-oxide) and 20% w/v PEG2000 MME. This crystal was plate-like and needed further optimization to obtain better quality crystals (Fig 3.16). However, attempts to optimize the crystallization conditions and/or reproduce the crystal did not yield any success.

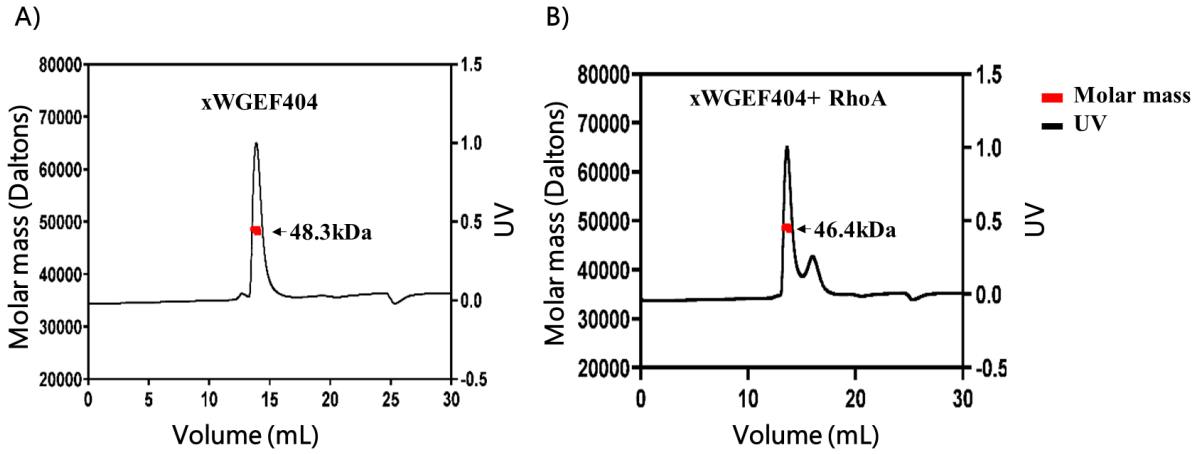


Figure 3.15: SEC-MALS analysis for xWGEF404 and xWGEF404+RhoA. A) *xWGEF404* ***B)*** *xWGEF404+RhoA.*

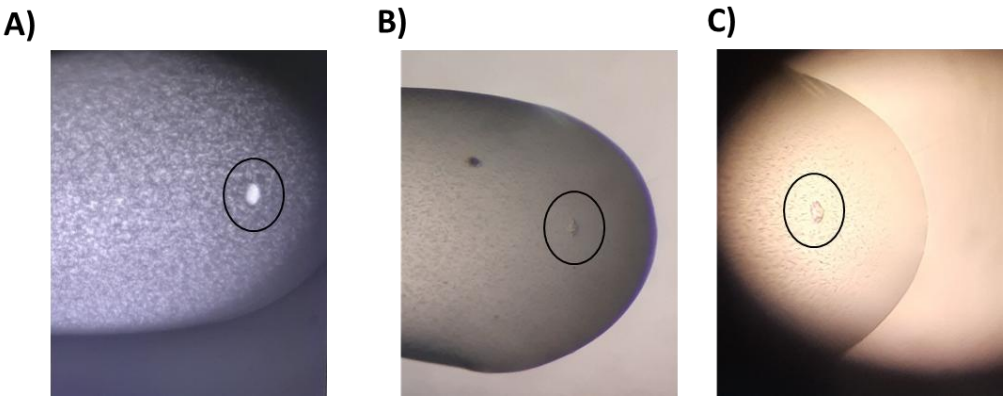


Figure 3.16: Protein crystal obtained at a 6.8 mg/mL concentration of xWGEF404-RhoA complex. A) *UV image* ***B)*** *Brightfield image* ***C)*** *Light microscope image.*

3.3.3 Cloning and expression of xWGEF386

Given that the crystallization of RhoA in complex with xWGEF330 and xWGEF404 met with limited success, we cloned another shorter construct of xWGEF, excluding both the NID and SH3 domains. This construct has just the DH-PH domains of the GEF, which were found to be

crystallizing for other Dbl family GEFs. xWGEF gene sequence encoding amino acids 386-760 (xWGEF386) was cloned in the p3E vector (Fig 3.17A). Although this particular construct was expressed and purified, the resulting protein appeared to be aggregated as it was eluted in the void fraction during size exclusion chromatography (Fig 3.17B-D). Consequently, no further experiments were conducted with this construct.

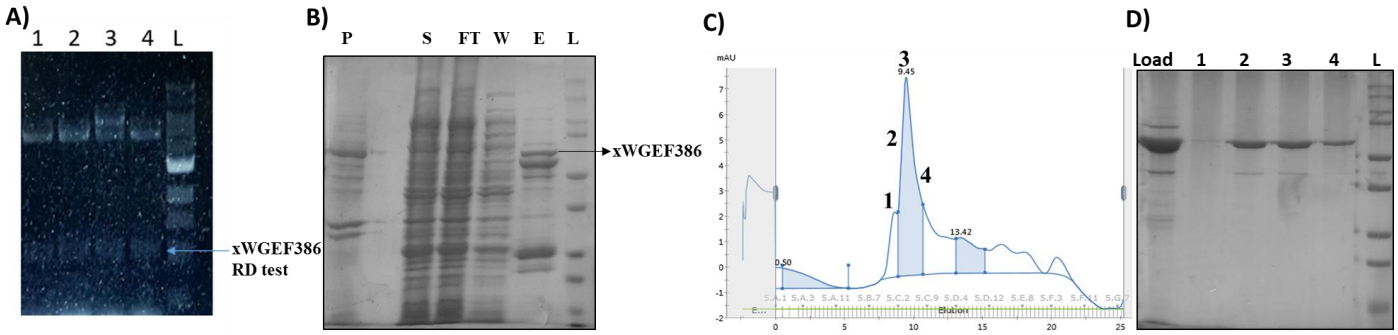


Figure 3.17: Cloning and purification of xWGEF386. A) *Agarose gel electrophoresis of positive clones after Restriction Digestion (RD) test for the presence of xWGEF386 gene insert.* ***B)*** *SDS-PAGE for GST affinity chromatography of xWGEF386.* ***C)*** *Size exclusion chromatogram for xWGEF386.* ***D)*** *SDS-PAGE for size exclusion chromatography of xWGEF386.*

3.4 Discussion

The structural characterization of WGEF-RhoA complex is essential for understanding the molecular basis of RhoA activation by WGEF. We designed different constructs of xWGEF and tried to form an *in vitro* complex with RhoA. However, due to the intrinsic properties of WGEF, we repeatedly observed that all of the xWGEF constructs produced aggregated protein when individually expressed in the *E. coli*. This phenomenon was observed across a spectrum of *E. coli* strains. Thus, the primary challenge in this study was the propensity of WGEF constructs to aggregate or remain unstable during purification. This observation aligns with the structural features predicted by computational tools, which indicated a high degree of disordered region in WGEF. However, the same constructs, except the xWGEF386, were found to be stable and soluble when expressed along with RhoA. The xWGEF386 construct was not subjected to the co-expression studies, as the protein showed no signs of stability when expressed in *E.coli*. These findings underscore the complexity of studying WGEF-RhoA interactions, particularly due to structural characteristics of WGEF.

Furthermore, all the constructs of xWGEF were observed to form sub-stoichiometric or unstable complexes on the SEC column. Our SEC-MALS analysis shows that the SEC peak identified as the one corresponding to the protein complex could be of the individual proteins, eluting as higher oligomers. Although a crystal hit for the xWGEF404-RhoA complex was obtained, when we reduced the number of purification steps and set up crystallization plates immediately after the ion-exchange chromatography step, we could not reproduce the crystal. From our extensive xWGEF-RhoA *in vitro* complexation and crystallization studies, we can infer that the protein complex is highly unstable or feeble. Given the multi-domain induced autoinhibitory nature of WGEF, it is really challenging to design a construct of the protein that retains all the functionally interesting domains and yet forms a stable complex with RhoA. Perhaps we could have designed a construct with just the DH domain of the GEF and formed its complex with RhoA. However, the structure of such a complex shall be an insignificant advancement in our understanding of RhoA activation by WGEF, as there are many structures of RhoA bound to canonical DH-PH domains available, and they can be used to model the complex of our interest. In fact, we have dwelled on such a strategy in the subsequent chapter of this thesis. However, the information obtained here on the nature of the recombinant xWGEF-RhoA complex shall act as a stepping stone in further optimization of the constructs and/or expression systems to elucidate its structure.

Chapter 4

Elucidation of Dishevelled2 induced WGEF activation mechanism in Wnt/Frizzled PCP pathway

4.1 Background

It has been shown that RhoA mediated actin cytoskeleton rearrangement is one of the manifestations of the non-canonical Wnt-PCP signalling pathway, which essentially requires RhoA activation. In this particular pathway, the GEF that activates RhoA is shown to be *Weak-similarity Guanine nucleotide Exchange Factor* (WGEF). Like many canonical GEFs, WGEF, too, exists in an autoinhibitory state and releasing it from the inhibitory state acts as another level of regulation. WGEF autoinhibition is governed by the intra-protein interactions formed by the NID and the SH3 domains of the protein with the residues present in the vicinity of DH and PH domains[161,162] (Fig 4.1). The N-terminal inhibitory region comprises of a helix with an inhibitory tyrosine residue (Y295 in hWGEF and Y353 in xWGEF), which interacts with the serine 394 (as per hWGEF numbering) of DH domain and blocks the access of RhoA. Substitution of Ser 394 to Ala is observed to increase the GEF activity by four folds, thus underlining the role of this interaction in rendering WGEF in an autoinhibited state[161]. Similarly, the interdomain interactions made by the SH3 domain with the polyproline-rich region of the GEF bring WGEF to an autoinhibited state[161,162] (Fig 4.1). However, in this case, the residues that are involved in establishing these interdomain interactions are not clear. Thus, regulation of WGEF involves releasing the protein from its autoinhibitory state, which could involve different activation mechanisms, such as protein-protein interaction with the domains involved in exerting the autoinhibition. It is worth noting that several other GEF activation mechanisms include protein-lipid binding, phosphorylation of the GEF, etc[66,76]. However, for many GEFs, their activation mechanism is not clear. Previous reports suggest that the interaction of WGEF with the adapter protein Dishevelled2 (Dvl2), operating under the Wnt-PCP pathway, is responsible for releasing the GEF from its autoinhibitory state[24].

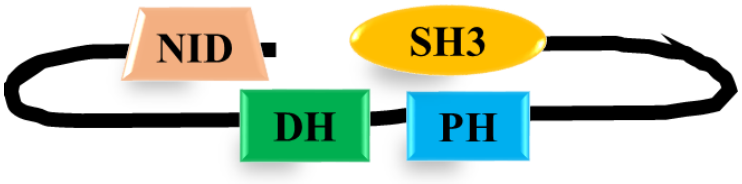


Figure 4.1: Schematic representation illustrating how the NID and SH3 domain blocks the DH domain for RhoA access.

As per the reports, the crucial step in canonical as well as non-canonical Wnt signalling involves the recruitment of a scaffold protein Dishevelled (Dvl) to the cytosolic domain of Wnt-bound membrane receptors, Frizzled[56] (Fig 1.1 and 1.2). Depending on the type of Wnt and Frizzled interaction, Wnt signalling modes diverge from the point of Dvl recruitment. As mentioned in the Introduction chapter, Dvl proteins have three major domains, namely, DIX, PDZ and DEP domains (Fig 1.3). With respect to WGEF activation, the PDZ domain of Dvl2 ($Dvl2^{PDZ}$) is shown to be responsible for releasing WGEF from its autoinhibitory state[24]. However, the exact region of WGEF that interacts with $Dvl2^{PDZ}$ and its mode of interaction with Dvl2 is not known.

The elucidation of WGEF-$Dvl2^{PDZ}$ interaction and its consequent role in the activation of GEF is very crucial in understanding the Wnt-PCP regulated cell migration. In this chapter, we have set out to identify the regions of WGEF and $Dvl2^{PDZ}$ that are involved in their interaction. Further, we systematically probe this interaction to determine the WGEF activation mechanism. This chapter is divided into two parts: Part A details about biochemical studies used to identify an evolutionary conserved internal motif in WGEF ($WGEF^{pep}$), which interacts with $Dvl2^{PDZ}$ and releases GEF from its autoinhibitory state. The next one, Part B, deals with molecular dynamics simulation studies to elucidate the mode of $Dvl2^{PDZ}$-WGEF interaction.

4.2 Part A: To identify the site of interaction between WGEF and the PDZ domain of human Dishevelled2

4.2.1 Background

PDZ domains are ubiquitous signalling modules found in almost all eukaryotes. They play a crucial role in myriad signalling events, serving as signalling nodes by assembling proteins involved in the pathway through their ability to establish protein-protein interactions. Structurally, these domains are predominantly composed of 5-6 β strands and 2-3 α-helices. They consist of a binding site, which is present in the groove formed by the α2-β2 region of the protein[51]. The carboxylate binding loop, which is an integral part of this groove, is shown to possess a conserved X-Φ-G-Φ motif (where X corresponds to any amino acid and Φ corresponds to a hydrophobic residue) (Fig 4.2A). The PDZ domains can recognize specific amino acid motifs that could be present either at their C-terminal or within the polypeptide chain of the target proteins. Thus, based on the presence of the location of the binding motif, the PDZ-target protein binding has been categorised into two modes: the C-terminal and the internal binding (Fig 4.2B). Further, the PDZ domains also exhibit variations in binding with the C-terminal motifs and on the basis of their interaction with these motifs, they are classified into three different classes depending on the motif consensus sequence, class I (motif recognized, S/T-X-Φ), class II (motif recognized, Φ-X-Φ) and class III (motif recognized, D/E-X-Φ) where Φ is a hydrophobic residue. In a few instances, the free carboxylate group of the C-terminal residue of the target drives the protein-protein interaction by interacting with the X-Φ-G-Φ motif, while the amino acids present at other positions of the motif also contribute to stabilizing the interaction. Depending on the importance of these residues lying upstream of the $P0^{th}$ residue (C-terminal residue which interacts with the X-Φ-G-Φ motif), this classification is further extended to 16 classes[163]. Some PDZ domains are also known to interact with internal sequence motifs of binding partners (Fig 4.2B). For example, the internal motif of the Pals1 protein is shown to be involved in its interaction with the $par\text{-}6^{PDZ}$ domain[164]. Similarly, Dvl^{PDZ} is known to interact with the internal motif (KTxxxW/I) of the Frizzled receptor and Idax protein[52,165]. In the present context, however, the sequence motif of WGEF, which interacts with $Dvl2^{PDZ}$, is not known. Therefore, we carried out a systematic sequence

and biochemical analysis to discover the Dvl2^{PDZ} binding motif of WGEF and further we probed the role of this interaction on its GEF activity.

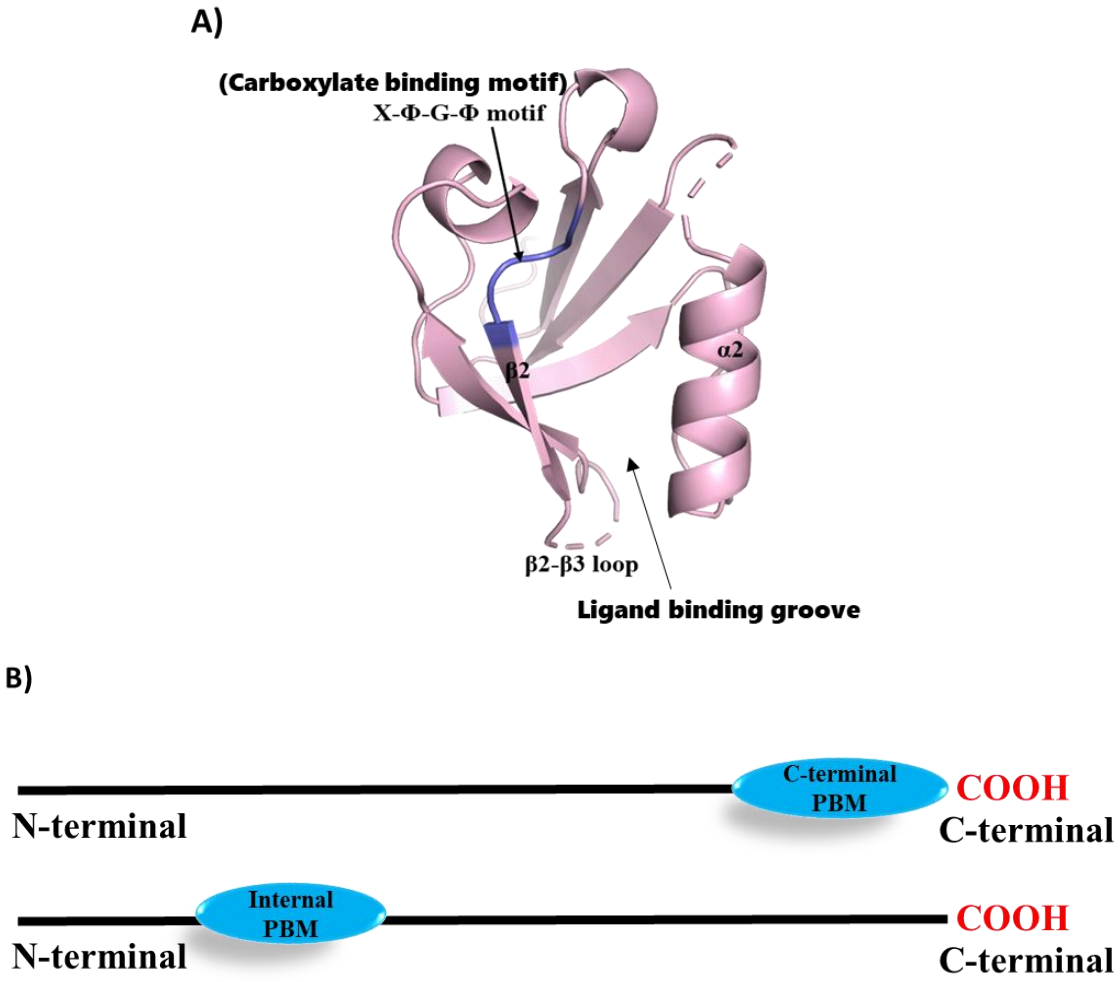


Figure 4.2: PDZ domains and their binding motifs in target proteins. A) *Canonical PDZ domains comprise 6 β strands and 2 α helices. The ligand binding groove is formed by α2 and β2 regions. The carboxylate binding loop is present at the beginning of β2 strand.* ***B)*** *PDZ domains can bind target proteins having C-terminal PDZ binding motif (PBM) or internal PBM.*

4.2.2 Methodology

4.2.2.1 Constructs, cloning, and mutagenesis

Human Arhgef19 (encoding hWGEF 330-581) gene was amplified using PCR and cloned into a modified p3E vector (with N-terminal cleavable GST tag) using the same methods described in Chapter 3. Human Dishevelled2 PDZ (hDvl2 PDZ) was a gift from Nicola Burgess-Brown (RRID: Addgene_38876); this vector was modified by introducing a stop codon after C354

residue of $Dvl2^{PDZ}$ to eliminate the fused peptide, cloned at the C-terminal of the PDZ. Fused constructs of $hDvl2^{PDZ}$-$WGEF^{pep}$ were commercially synthesized from Twist Biosciences. $Dvl2^{PDZ}$ constructs have an N-terminal cleavable 6xHis tag. Mutations were introduced into the xWGEF (330-856) construct by using the site-directed mutagenesis method and confirmed by DNA sequencing. Synthetic peptides such as $WGEF^{pep}$, $WGEF^{pep}$ variants and $N3^{pep}$ were procured from Sai Biotech.

4.2.2.2 Protein Expression and purification

Recombinant hWGEF (330-581) was expressed into *Escherichia coli* (*E. coli*) BL21*(DE3) strain, using the same method described previously for GST-tagged proteins. SEC was performed with 10 mM MOPS pH 6.5, 150 mM NaCl, 2 mM DTT, and 10% glycerol. 6xHis tagged $hDvl2^{PDZ}$ and fused $hDvl2^{PDZ}$-$WGEF^{pep}$ proteins were expressed in Rosetta (DE3) cells and purification was done using Ni^{2+}-nitrilotriacetate (Ni^{2+}-NTA) affinity chromatography. 6xHis tag was further cleaved by using TEV protease during overnight dialysis. TEV protease and His tag were eliminated after desalting and reverse Ni^{2+}-NTA affinity purification. Pure protein was obtained from SEC using a buffer containing 10 mM Tris pH 8, 2 mM DTT, 150 mM NaCl and 5% glycerol for $hDvl2^{PDZ}$ and 10 mM Tris pH 8, 2 mM DTT, 350 mM NaCl and 5% glycerol for fused $hDvl2^{PDZ}$-$WGEF^{pep}$. Site-directed mutagenesis method was used to introduce mutations into xWGEF (330-856); expression, and purification were carried out as described above for the GST-tagged proteins. RhoA expression and purification were done as per the previous chapter. These proteins were concentrated and stored at -80°C.

4.2.2.3 Protein crystallization, data collection, and structure determination

Protein purified from SEC was concentrated and used for setting up plates using a crystallization robot, Mosquito® (TTP Labtech, Royston UK). Crystals of human $Dvl2^{PDZ}$ ($Dvl2^{PDZ}$) were grown using sitting drop vapour diffusion method at 293 K by mixing 1:1 ratio of $Dvl2^{PDZ}$ (10 mg/mL) with 6 M excess of $WGEF^{pep}$ and reservoir buffer consisting of 0.1 M tri-sodium citrate (pH 5.6), 20% (w/v) PEG4000 and 20% (v/v) isopropanol. Similarly, crystals of human $Dvl2^{PDZ}$ ($Dvl2^{PDZ}$) fused with $WGEF^{pep}$ (10 mg/mL) were grown with a reservoir buffer consisting of 0.1 M sodium acetate (pH 4.5) and 3 M sodium chloride (NaCl). Crystals were frozen in cryoprotectants containing crystallization conditions supplemented with 28% glycerol. X-ray diffraction data for $Dvl2^{PDZ}$ co-crystallized with $WGEF^{pep}$ was collected at 100

K on microfocus beamline ID23-2 of the European Synchrotron Radiation Facility (ESRF), Grenoble, France[166]. Crystals of the fused hDvl2^{PDZ}-WGEF^{pep} protein were diffracted at Indus-2 synchrotron on BL-21 PX beamline at RRCAT, Indore. Data integration and scaling was done by using XDS[123] and AIMLESS[124,167], implemented on CCP4 software[168], respectively. Structure solving was done with molecular replacement using PHASER[129]. Human Dvl2-PDZ structure (PDB entry: 2REY) served as a search model. The structure was further iteratively built using COOT[169] and refined using PHENIX[135,170]. The structure validation was performed using Molprobity[171]. Coordinates and structure factors for hDvl2^{PDZ} and WGEF^{pep} fused hDvl2^{PDZ} (no electron density for the peptide) are deposited in the Protein Data Bank with PDB IDs 8WWR and 8YR7, respectively.

4.2.2.4 *In vitro* Guanine nucleotide exchange assay

GEF assays were carried out with fluorescently labelled RhoA[93]. Briefly, 150 µM of RhoA was incubated with 5 Molar excess of fluorescent label, i.e., Mant-GDP in a buffer consisting of 10 mM Tris pH 8, 2 mM DTT, 150 mM Nacl, 5 mM EDTA, 0.5 mM $MgCl_2$, 5% glycerol. The reaction was incubated on ice for 30 min, after which 10 mM $MgCl_2$ was added to terminate the reaction. Excess Mant-GDP was removed through buffer exchange and Mant-GDP loaded RhoA was concentrated and stored at -80°C. For assays, Mant-GDP loaded RhoA (2 µM) was mixed and incubated with xWGEF (5 µM) in the presence or absence of hDvl2^{PDZ} (0 µM, 25 µM, and 50 µM) at 25°C/30 min. To observe the effect of synthetic peptides WGEF^{pep} (GSTFSLWQDIP), and WGEF^{pep}*^{D9A}* (GSTFSLWQAIP) of WGEF origin on Dvl2^{PDZ}-WGEF interaction, these peptides were added in the respective reactions such that their final concentration is 200 µM. Once the fluorescence was stable, excess GTP (10x of labelled GTPase) was added to start the reaction. All GEF assays were carried out in a buffer containing 10 mM Tris, pH 8, 2 mM DTT, 150 mM NaCl, 10 mM $MgCl_2$, and 5% glycerol. A decrease in fluorescence was recorded using a BioTek cytation 5 plate reader with excitation and emission of 360 nm and 440 nm, respectively. Fluorescence was normalized and decrease was fitted using a single exponential decay mode in GraphPad Prism. Assays were performed in triplicates.

4.2.2.5 Microscale Thermophoresis (MST)

MST experiments were performed using the Monolith NT.115 instrument from Nanotemper technologies. Dvl2^{PDZ} domain was fluorescently labelled as per instructions, using Monolith Redmaleimide 2nd generation-cysteine reactive label (Nanotemper technologies). The experimental condition comprise of 10 mM Tris pH 8, 2 mM DTT, 150 mM NaCl, 5% glycerol, and 0.05% tween20. For Dvl2^{PDZ}-peptide affinity determination, synthetic WGEF peptides were serially diluted (1:2) 16 times starting with an approximate concentration of 3.4 mM. 50 nM of fluorescently labelled Dvl2^{PDZ} domain was added to the serially diluted peptides and incubated at 25°C/30 min. Thermophoresis signals were recorded at 20% of excitation power and 40% MST at 25°C. Data was fitted using the K_d fit model in the MO. Affinity Analysis v2.3 software[107]. Like peptides, hWGEF (330-581) protein was serially diluted to obtain a protein concentration range of 138 μM to 0.0042 μM. Experimental conditions were the same as mentioned above for peptides. All MST experiments were done in triplicates.

4.2.3 Results

4.2.3.1 WGEF contains a conserved internal peptide motif which shares sequence homology with Dvl2^{PDZ} binding peptides

Previous studies have shown that Dvl2^{PDZ} binds to WGEF and releases it from the autoinhibitory state. However, the precise motif through which WGEF interacts with Dvl2^{PDZ} is not clear, with studies suggesting the involvement of different regions of the protein. For instance, Igor *et al.* have shown that the Dvl2^{PDZ} binding site in WGEF is present towards the N-terminus, as deletion of the first 213 residues of hWGEF (hWGEF$^{\Delta 213}$) reduces its affinity for Dvl2^{PDZ} while significantly enhancing the GEF activity. This finding indicates that the binding of Dvl2^{PDZ} within this domain may be responsible for releasing WGEF from autoinhibition[24]. A different study by Yohe *et al*. showed that deletion of the first 302 residues of hWGEF (hWGEF$^{\Delta 302N\text{-term}}$) also releases the GEF from its autoinhibition[161]. Both these studies indicate that the Dvl2^{PDZ} binding region precedes the autoinhibitory domain of the GEF (spanning residues 291-NSVLYQEY-298) (Fig 4.3). But the exact motif of interaction remained elusive. Like other PDZ domains, the Dvl2^{PDZ} domain shows promiscuity in recognizing different peptide motifs. PDZ domains canonically recognize *C-terminal peptide motifs* composed of 7-10 amino acids present at the C-terminus of their interacting partners.

Some PDZ domains can recognize *internal peptide motifs* having similar lengths as the C-terminal peptide motifs. In contrast to the C-terminal peptide motifs, the internal peptide motifs can be located anywhere within the binding partner[51,172,173] (Fig 4.2B). The interaction between internal motif and PDZ domains mimics that of the C-terminal peptide binding, with slight variations in the peptide residues involved in interaction[173,174].

Figure 4.3: Schematic representation of WGEF domains and the Dvl2^{PDZ} binding region suggested from previous studies.

As mentioned in section 4.2.1, PDZ binding *C-terminal peptide motifs* shows some consensus in sequence. Thus, PDZ domains can be divided into three classes (class I, class II and class III) based on the consensus sequence of the C-terminal peptide they bind to[163]. However, there is no statistically significant data to define a consensus sequence for internal peptides recognized by PDZ domain. Therefore, to identify the Dvl2^{PDZ} interacting motif of WGEF, we used the Dvl2^{PDZ} binding motifs identified from the peptide-phage display assay to search the sequences of WGEF family members[175]. We analyzed WGEF protein sequence for the presence of Dvl2^{PDZ} binding motifs at both the C-terminal as well within the protein (internal peptide motifs). However, no conserved sequence corresponding to the C-terminal motif was identified, consistent with the findings reported by Zhang *et al*., which showed that WGEF (Ephexin-2) lacks a C-terminal PDZ binding motif[162] . For internal motifs, we used three prominent peptide motifs, X-Y-G-W-Φ^a-D/G, X-W-Φ^a-D-G-P, and W-Φ^s-D-X-P (where X, Φ^a, and Φ^s are any, aliphatic and short side chain hydrophilic amino acids [S/T], respectively), as templates for the search. From this motif search, the only hit that we obtained was 349-GSTFSLWQDIP-359 (hereafter referred to as WGEF^{pep}) (Fig 4.4). This motif is located between the DH and the N-terminal inhibitory domain of hWGEF and is evolutionarily conserved, suggesting it could be the putative site for Dvl2^{PDZ} interaction. This finding differs from the previous observations where the Dvl2^{PDZ} binding region was mapped towards the N-terminal of hWGEF (between amino acid residues 1-213)[24] (Fig 4.3), whereas the motif identified by us is present between the DH domain and the NID domain.

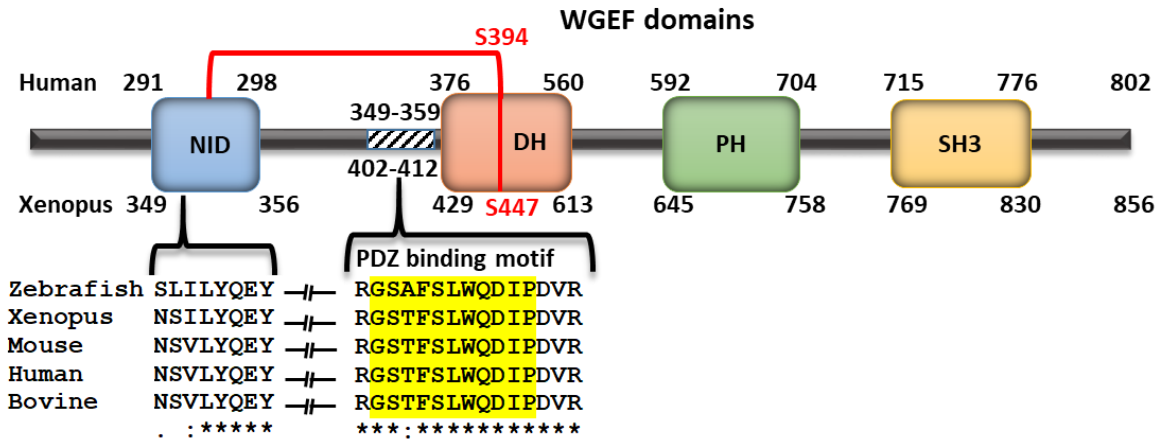


Figure 4.4: Schematic representation of WGEF domains and sequence alignment of the conserved PDZ binding motif of different WGEF species. *Zebrafish (Uniprot ID: E7EY470), Xenopus (Uniprot ID: A5X5J0), Mouse (Uniprot ID: Q8BWA8), Human (Uniprot ID: Q8IW93-1) and Bovine (Uniprot ID: E1BQ24) ahead of DH domain in WGEF. The amino acid residue numbers in the upper and lower parts of the diagram correspond to hWGEF and xWGEF, respectively.*

4.2.3.2 The identified N-terminal conserved 'internal peptide' motif of WGEF facilitates its interaction with Dvl2^{PDZ}

Since the peptide motif identified by us lies away from the earlier reported Dvl2^{PDZ} binding region. We needed to validate that this motif is indeed involved in Dvl2^{PDZ} binding. With this aim, we designed GEF constructs comprising both the autoinhibitory and the identified motif of interest, yet excluding major N-terminal disordered regions of human WGEF (hWGEF). Due to insolubility and/or poor expression of these hWGEF constructs in *E.coli*, we used WGEF from *Xenopus laevis* (xWGEF330) as a surrogate system for further experiments. To study the effect of Dvl2^{PDZ} interaction on the catalytic activity of xWGEF330, we performed guanine nucleotide exchange (GEF) assay on xWGEF330 (5 µM concentration) in presence and absence of the Dvl2^{PDZ} domain. From the assay, it could be observed that addition of 50 µM of Dvl2^{PDZ} (10-fold excess compared to xWGEF330) increased the activity of xWGEF330 by fourfold, whereas the addition of a lower concentration of Dvl2^{PDZ} (25 µM concentration) did not exhibit any notable change in the xWGEF330 activity (Fig 4.5). This indicates that even though the binding of Dvl2^{PDZ} enhances the activity of xWGEF330, the binding affinity between Dvl2^{PDZ} and xWGEF330 is low. Thus, higher concentrations of Dvl2^{PDZ} are required for xWGEF330 activation. Further, to investigate that the WGEF internal peptide motif identified by us 402(349)-GSTFSLWQDIP-412(359)/ WGEF^{pep} [residue number shown in the

parentheses corresponds to hWGEF] is responsible for Dvl2^{PDZ}-xWGEF330 interaction, we performed GEF assay in the presence of both Dvl2^{PDZ} and WGEF^{pep} peptide. Since the internal peptides are shown to have relatively low affinity for PDZ domains, we used a fourfold molar excess of WGEF^{pep} (200 µM concentration) compared to Dvl2^{PDZ} to disrupt the WGEF-Dvl2^{PDZ} interaction. This experiment clearly showed that WGEF^{pep} competitively binds to the PDZ domain, displacing xWGEF330 from Dvl2^{PDZ}-xWGEF330 interaction, thereby reducing the catalytic activity of WGEF (Fig 4.5). Our findings suggest that the 402-412 region of xWGEF represents the 'internal peptide motif' that mediated its interaction with Dvl2^{PDZ}.

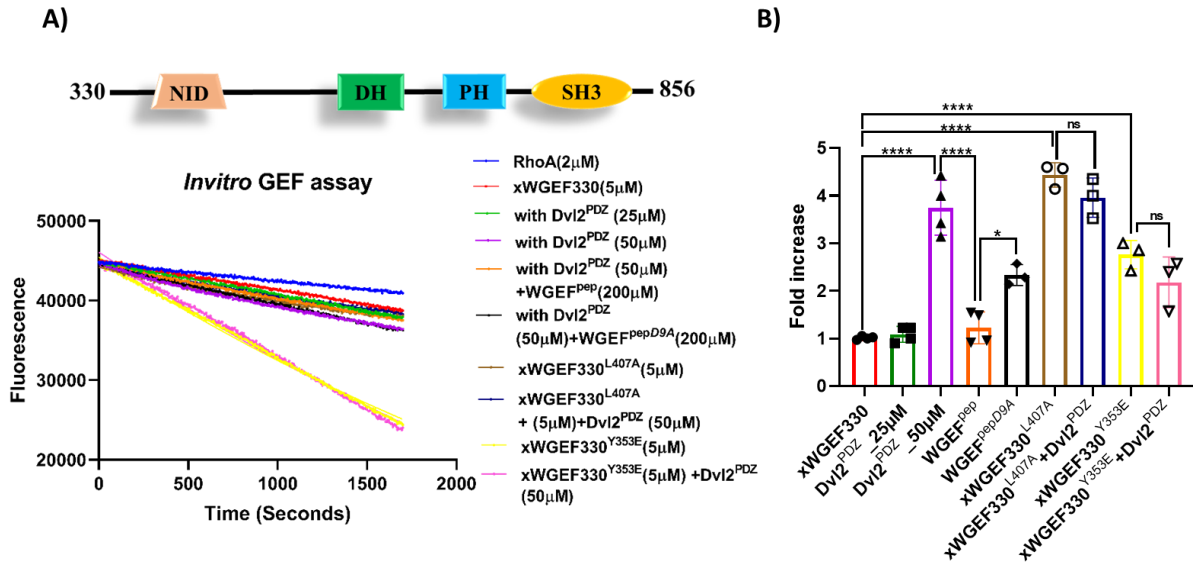


Figure 4.5: Guanine nucleotide exchange assay for xWGEF330 and mutants. A) *GEF assay plots of xWGEF330, xWGEF330^{L407A}, and xWGEF330^{Y353E}, in the presence and absence of Dvl2^{PDZ} and peptide variants.* ***B)*** *Bar graph showing relative GEF activities of xWGEF variants normalized against that of wild-type (autoinhibited xWGEF330) protein. The error bar and * represent standard deviation (± SD) and significance as ****P <0.0001, ***P <0.001, **P <0.01, and *P ≤0.01, respectively.*

4.2.3.3 WGEF^{pep} motif-Dvl2^{PDZ} interaction is sufficient to partially release the autoinhibition of WGEF

So far, we have observed that binding of Dvl2^{PDZ} and WGEF takes place through the identified internal peptide motif leading to an increase in activity of xWGEF330. Next, to understand the mechanism behind this activation, we examined other reported Dvl2^{PDZ} binding internal motifs. The internal peptide motif identified by us shares sequence homology with the previously reported Dvl2^{PDZ} binding internal peptides N1^{pep}, N2^{pep} and N3^{pep}, with N3^{pep} showing the

highest sequence homology[175] (Fig 4.6). In these peptides, the aspartate side chain mimics the free C-terminal carboxylate group, allowing it to bind to the carboxylate binding loop of $Dvl2^{PDZ}$, thereby stabilizing the $Dvl2^{PDZ}$-effector interaction[175] (Fig 4.7). Here, we wanted to determine if the aspartate in our peptide motif also plays a similar role in peptide-PDZ interaction. To further explain our study, we defined labels to different residues of the motifs for better understanding their positions. PDZ domains typically interacts with the C-terminal peptides, the free carboxyl group of the terminal residue of the peptide interacts with the X-Φ-G-Φ motif of the carboxylate binding loop (X represents any amino acid and Φ represents hydrophobic residues) referred to as the 0th or P_0 residue. Residues lying towards the N-terminus of P_0 residue are numbered in reverse order as '-1', '-2', '-3' or P_{-1}, P_{-2}, P_{-3} and so forth[164,176,177]. $Dvl2^{PDZ}$ binds to both C-terminal motifs as well as non-canonical internal motifs[175,178]. Therefore, we labelled the residue which interacts with the X-Φ-G-Φ motif of PDZ as P_0, followed by P_{-1}, P_{-2}, P_{-3} for residues preceding P_0 and P_1, P_2, P_3 for the residues that follow P_0[51] (Fig 4.6). To determine if the $WGEF^{pep}$ follows the typical internal peptide-$Dvl2^{PDZ}$ interaction, we substituted the conserved aspartate of the peptide with alanine ($WGEF^{pepD9A}$). Interestingly, while the $WGEF^{pepD9A}$ peptide reduced GEF activity of xWGEF330, the effect was not as significant as with the wild type $WGEF^{pep}$ (Fig 4.5). We also created a corresponding mutation in the xWGEF330 construct ($xWGEF330^{D410A}$), but this mutant was not stable upon expression (Fig 4.8), precluding further analysis.

Position	P_{-8} P_{-7} P_{-6} P_{-5} P_{-4} P_{-3} P_{-2} P_{-1} P_0 P_1 P_2
N1	-WKDYGWIDGK
N2	-SGNEVWIDGP
N3	--EIVLWSDIP
WGEF	GSTFSLWQDIP
	* *

Figure 4.6: Conserved internal motif identified in WGEF. *Shows sequence alignment of WGEF internal peptide motif with $N1^{pep}$, $N2^{pep}$ and $N3^{pep}$ internal peptides, which are reported to form complex with $Dvl2^{PDZ}$.*

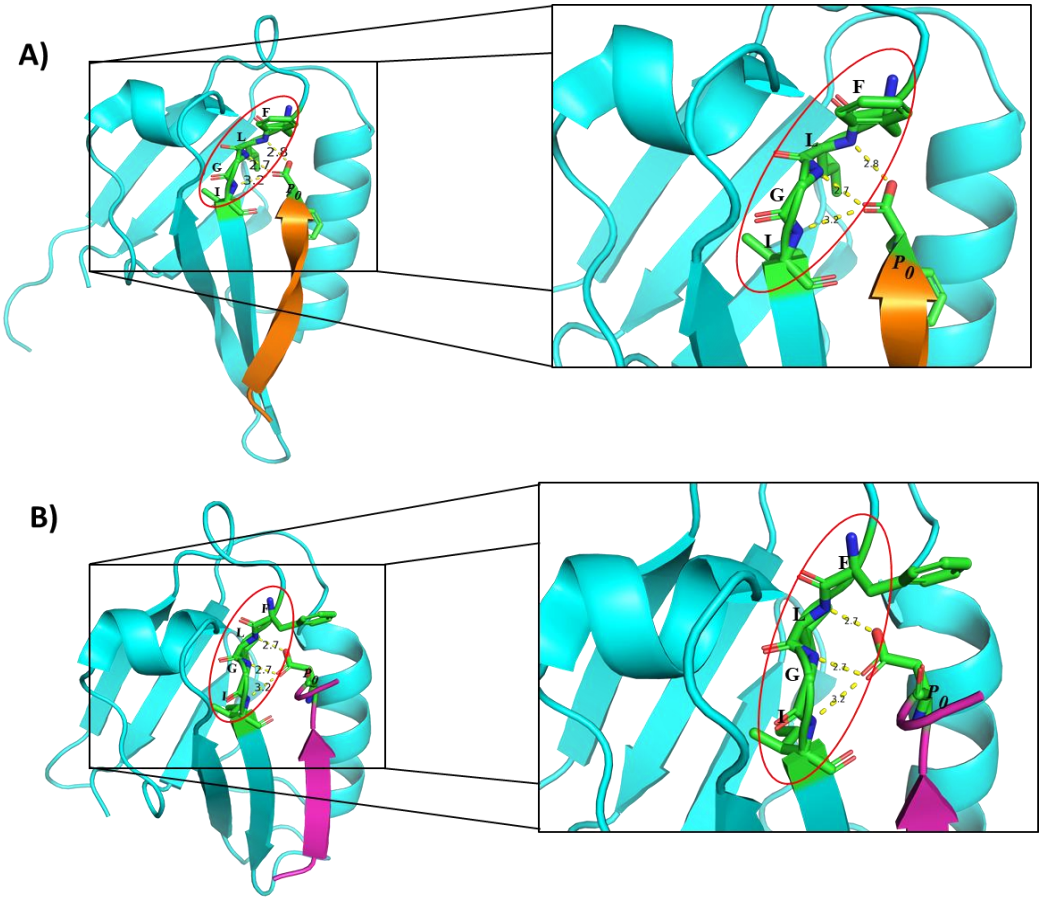


Figure 4.7: Interaction of C-terminal ($C1^{pep}$) and internal ($N3^{pep}$) peptides with the carboxylate binding motif of $Dvl2^{PDZ}$. A) *Crystal structure of $Dvl2^{PDZ}$ in complex with C-terminal peptide (PDB ID: 3CBX). The terminal carboxyl group of the P_0 residue of the C-terminal peptide makes contact with the carboxylate binding motif of PDZ (marked in red circle).* ***B)*** *Crystal structure of $Dvl2^{PDZ}$ bound to internal peptide (PDB ID: 3CC0) which shows highest homology with the $WGEF^{pep}$. The carboxyl side chain of the conserved aspartate residue (P_0) in the peptide interacts with the carboxylate binding motif of PDZ (marked in red circle).*

Next we wanted to study the role of other residues, spanning the P_0 position in the interaction. Previous studies indicate that proline, tryptophan and leucine at positions P_2, P_{-2} and P_{-3}, respectively, also play key roles in stabilizing $Dvl2^{PDZ}$-peptide interactions. Also, these residues are conserved between $N3^{pep}$ and $WGEF^{pep}$. Therefore, to explore the significance of these residues in GEF-PDZ interaction, we generated $xWGEF330^{P412A}$, $xWGEF330^{W408A}$ and $xWGEF330^{L407A}$ mutants (Fig 4.8-4.11). Amongst these, only $xWGEF330^{L407A}$ was soluble (Fig 4.9). Surprisingly, this mutant showed higher GEF activity which appeared to be independent of $Dvl2^{PDZ}$, as addition of 50 µM $Dvl2^{PDZ}$ to the assay mixture did not alter its GEF activity (Fig 4.5). The higher GEF activity observed for this mutant might be due to the

structural rearrangement caused by the mutation; however, the exact mechanism by which L407A substitution affects GEF autoinhibition is not known and requires further investigation.

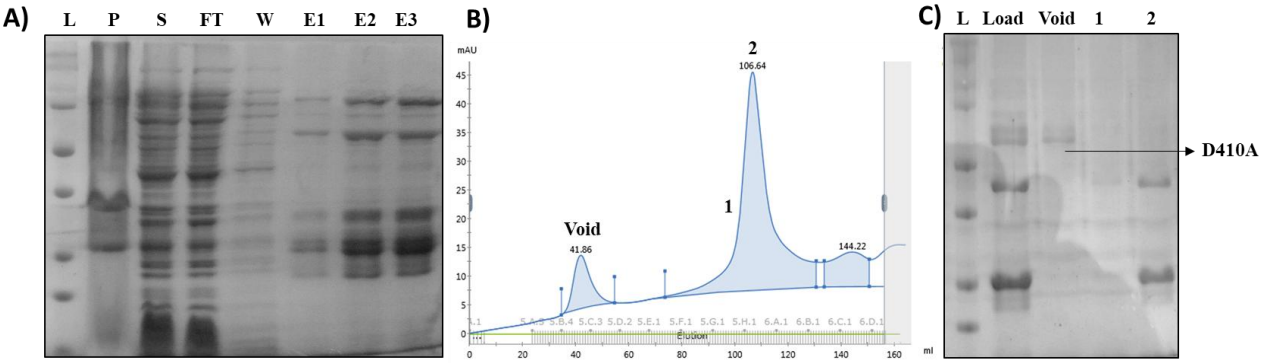


Figure 4.8: Purification of xWGEF330^{D410A} mutant. A) *SDS-PAGE for GST affinity chromatography* ***B)*** *Size exclusion chromatogram* ***C)*** *SDS-PAGE for size exclusion chromatography.*

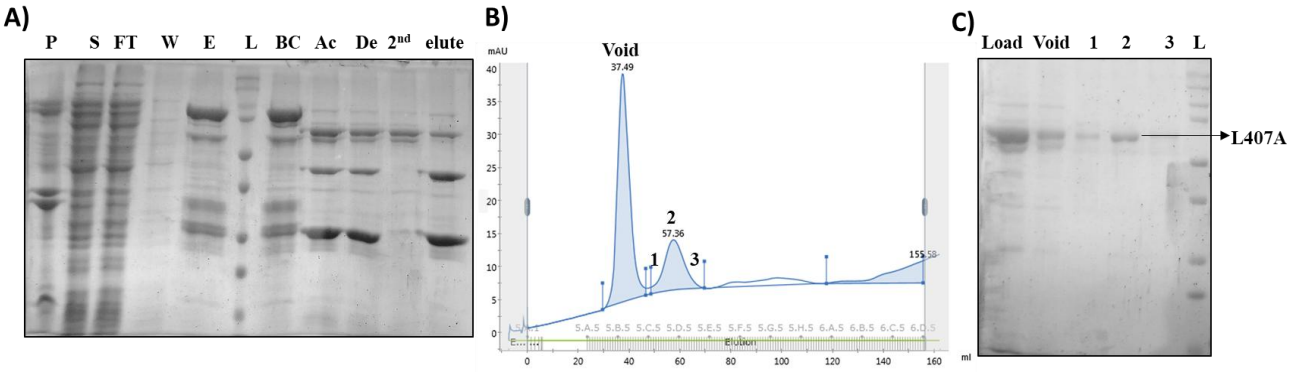


Figure 4.9: Purification of xWGEF330^{L407A} mutant. A) *SDS-PAGE for GST affinity chromatography* ***B)*** *Size exclusion chromatogram* ***C)*** *SDS-PAGE for size exclusion chromatography.*

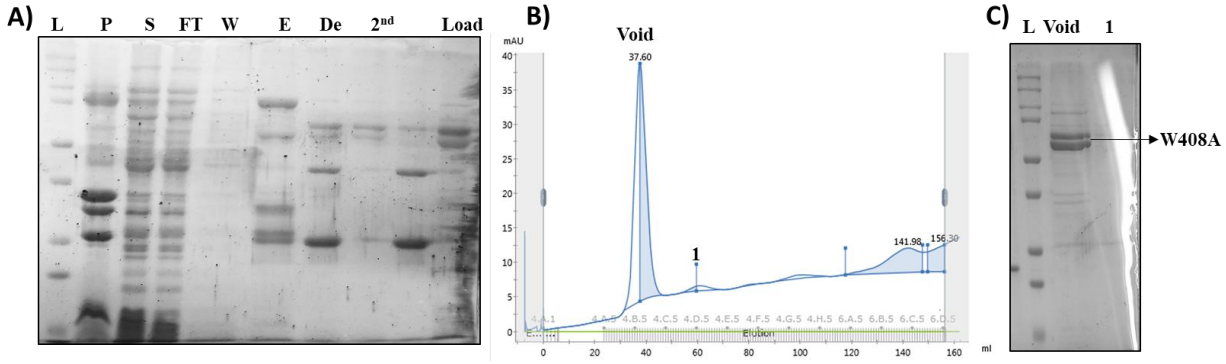


Figure 4.10: Purification of xWGEF330^{W408A} mutant. A) *SDS-PAGE for GST affinity chromatography* ***B)*** *Size exclusion chromatogram* ***C)*** *SDS-PAGE for size exclusion chromatography.*

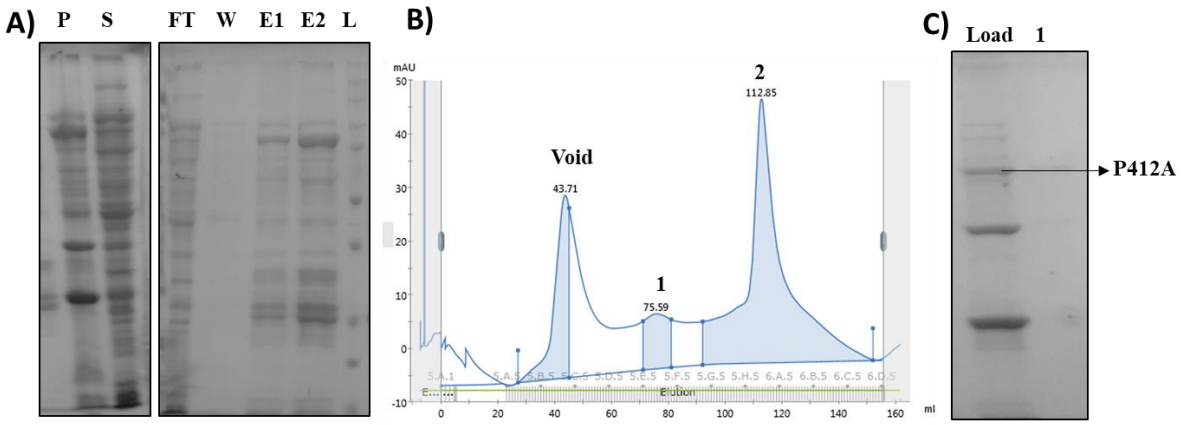


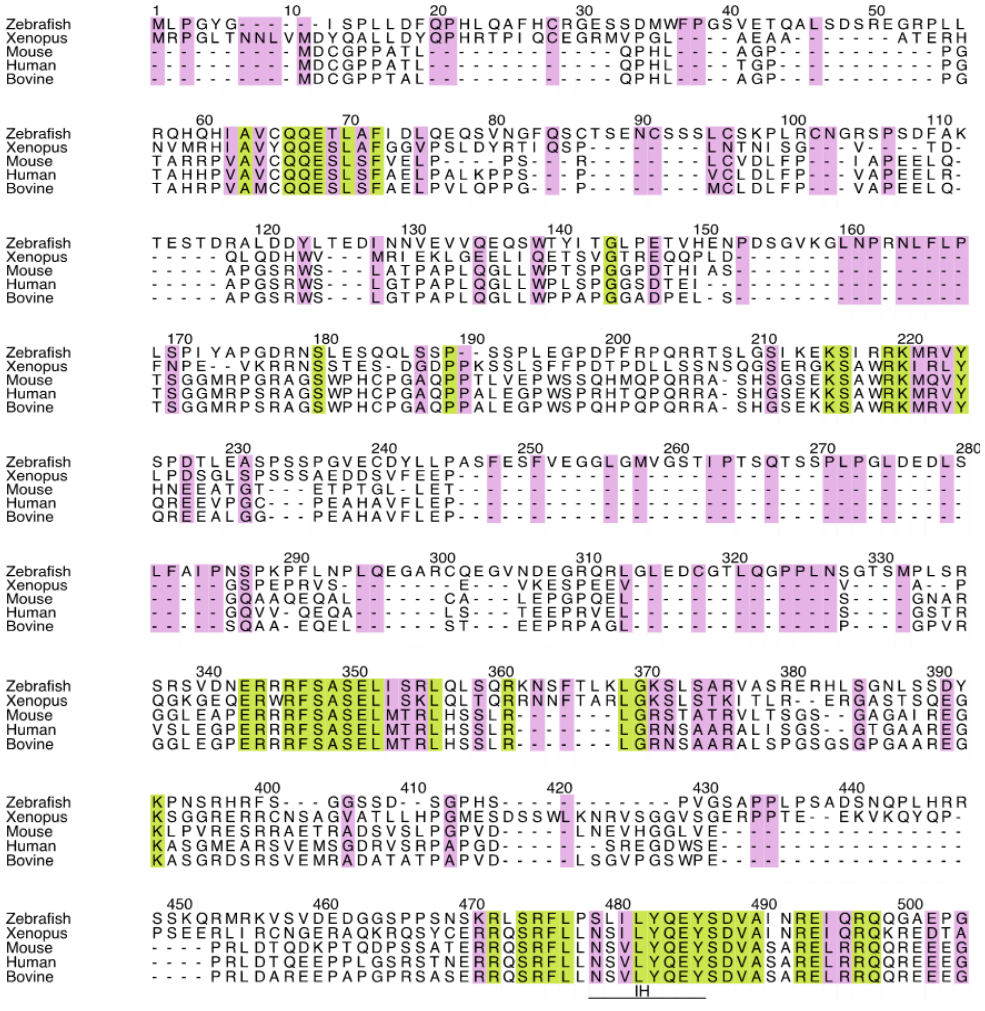
Figure 4.11: Purification of xWGEF330^{P412A} mutant. A) *SDS-PAGE for GST affinity chromatography* ***B)*** *Size exclusion chromatogram* ***C)*** *SDS-PAGE for size exclusion chromatography.*

Further, to determine if WGEF-Dvl2^{PDZ} enhances GEF activity by releasing autoinhibition or involves an unknown activation mechanism, we generated an Y353E mutant of xWGEF330. This tyrosine residue (Y353 in xWGEF/ Y295 in hWGEF) is part of the NID region and is known to inhibit the GEF activity by interacting with the DH domain of the GEF. Previous studies have shown that substituting this tyrosine to glutamate (Y295E) in hWGEF releases the GEF from its autoinhibitory state[161]. Remarkably, the xWGEF330^{Y353E} mutant demonstrated nearly a threefold enhanced GEF activity compared to the wild type xWGEF330, and addition of Dvl2^{PDZ} (50 µM concentration) did not affect its GEF activity (Fig 4.5). These findings suggest that Dvl2^{PDZ} interacts with the WGEF^{pep} motif, present between the NID and the DH domains, partially relieving WGEF from its autoinhibitory state regulated by the NID. However, the mechanism involving the release of autoinhibition, which is governed by the SH3 domain, is not clear.

4.2.3.4 Residues at P_{-2}, P_0, and P_2 positions of the binding motif mediates the binding of WGEF^{pep} with Dvl2^{PDZ}

From the GEF assay it is observed that while the WGEF^{pep} region is substantially conserved among the WGEF family (Fig 4.12), the interaction between WGEF and Dvl2^{PDZ} is feeble. This is consistent with earlier reported PDZ-internal peptide interactions[173,175]. Perhaps differential affinity of the Dvl with its effectors is one of the strategies adopted by Wnt-Frizzled pathways to exert diverse signalling modes[47,179]. While our GEF assay-based studies confirmed that the residue at P_0 position of WGEF^{pep} is crucial for WGEF-Dvl2^{PDZ} binding, however, the

influence of residues at other positions on this interaction was not clear from this study, as the xWGEF mutants used to probe this aspect were found to be unstable (Fig 4.8, 4.10 and 4.11). Therefore, to gain structural insights into the Dvl2^{PDZ}-WGEF interaction, we employed a multipronged approach in which we tried to crystallize both the protein complex (hWGEF330-Dvl2^{PDZ}) and the protein-peptide (Dvl2^{PDZ}-WGEF^{pep}) complex. The hWGEF$^{330-581}$ is a minimal construct of human WGEF, which consists of the WGEF^{pep} motif and the DH domain. Before conducting further studies, we used RhoA activation assay and confirmed that the protein was active (Fig 2.1). Although both strategies produced crystals, the crystals obtained from Dvl2^{PDZ}-WGEF^{pep} in a condition consisting of 0.1 M tri-sodium citrate (pH 5.6), 20% (w/v) PEG4000 and 20% (v/v) isopropanol diffracted to a highest resolution of 1.75 Å. However, on elucidation of the structure, Dvl2^{PDZ} was found to be in its *apo* form as the electron density for the bound peptide was not observed (PDB ID: 8WWR) (Fig 4.13, 4.14 and Table 4.1).



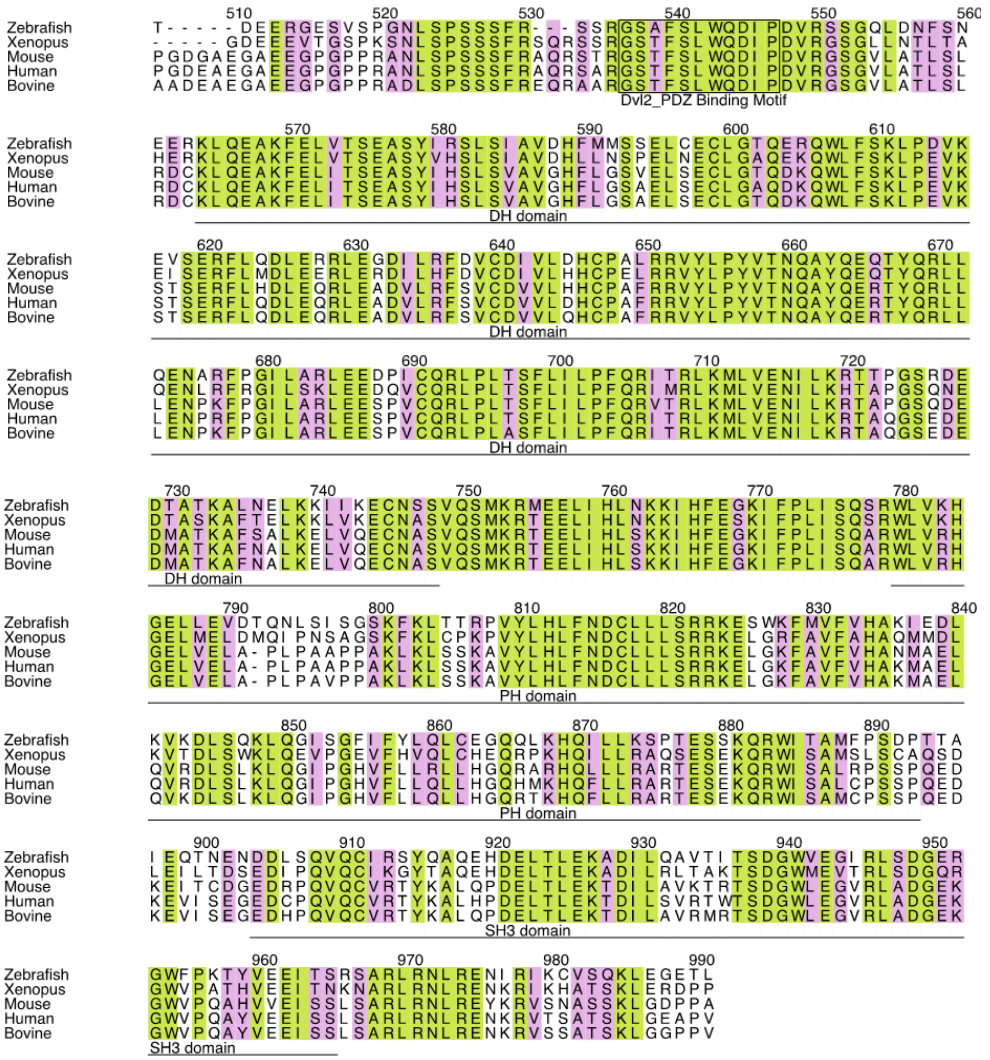


Figure 4.12: Sequence alignment of WGEF from different organisms (Uniprot IDs: Zebrafish_E7EY47, Xenopus_A5X5J0, Mouse_Q8BWA8, Human_Q8IW93_1 and Bovine_E1BQ24). *Sequences with 100% conservation along different organisms are highlighted in green, whereas sequences with 70% conserved sequence identity are highlighted in purple. The conserved Dvl2^{PDZ} binding motif is marked with a black box. Similarly, other regions like the N-terminal Inhibitory (IH), DH, PH and SH3 domains present in WGEF are labelled and underlined with a black line below the sequence.*

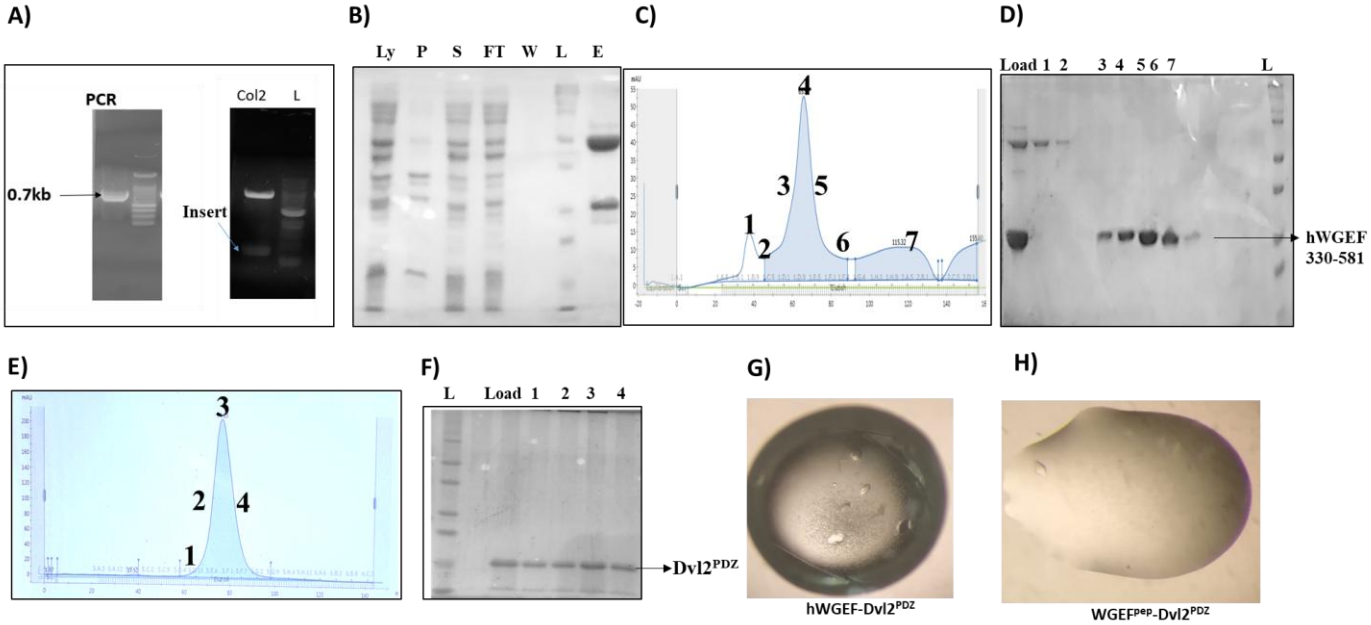


Figure 4.13: Purification of hWGEF330 and Dvl2^{PDZ} for crystallization. A) *Agarose gel of positive clone of hWGEF$^{330-581}$ after restriction digestion.* ***B)*** *SDS-PAGE of GST affinity chromatography for hWGEF$^{330-581}$.* ***C)*** *Size exclusion chromatogram for hWGEF$^{330-581}$.* ***D)*** *SDS-PAGE of size exclusion chromatography for hWGEF$^{330-581}$.* ***E)*** *Size exclusion chromatogram for Dvl2^{PDZ}.* ***F)*** *SDS-PAGE of size exclusion chromatography for Dvl2^{PDZ} G) Crystals of hWGEF-Dvl2^{PDZ}.* ***H)*** *Crystals of WGEF^{pep}-Dvl2^{PDZ}. Ly=Lysate, P= Pellet, S=Supernatant, FT=Flow through, W=Wash, E=Elute.*

Figure 4.14: Crystal structure of apo Dvl2^{PDZ} obtained from the crystals of WGEF^{pep}-Dvl2^{PDZ} complex.

Another approach involves fusing of WGEF^{pep} peptide with the C-terminal of the PDZ domain. This strategy has been described for studying the interaction involving the PDZ domain and N1, N2 and N3 peptides[175]. WGEF^{pep} fused Dvl2^{PDZ} construct was purified and crystallization trials were set up (Fig 4.15A and B). Even for this construct, we obtained crystals

with the crystallization condition containing 0.1 M sodium acetate pH 4.5 and 3 M NaCl. However, as observed with our co-crystallization approach, even in this condition, there was no density for the fused peptide (PDB ID: 8YR7) (Fig 4.15C- E and Table 4.1). We surmise that the presence of very high salt concentration (3M NaCl) in the crystallization condition might have impeded the protein-protein binding, resulting in the abrogation of interaction between the C-terminal peptide and the PDZ domain. Different constructs of $Dvl2^{PDZ}$-peptide with varying linkers and configurations did not yield any crystals. The only construct that crystallized has a PDZ binding peptide (GSTFSLWQDIP) separated by four amino acid linker (SGGG). Thus, despite extensive efforts, the structure of $WGEF^{pep}$ bound to the $Dvl2^{PDZ}$ domain could not be elucidated. However, to fill the gaps in our understanding due to the non-availability of $WGEF^{pep}$-$Dvl2^{PDZ}$ complex structure, we systematically designed variants of $WGEF^{pep}$ and biochemically analyzed their interaction with $Dvl2^{PDZ}$. For this, we used $Dvl2^{PDZ}$-$N3^{pep}$ complex structure (PDB ID: 3CC0) as template to create $WGEF^{pep}$ variants by substituting residues from P_{-3} to P_2 with alanine, as majority of these residues are conserved between $N3^{pep}$ and $WGEF^{pep}$ and in the structure of $Dvl2^{PDZ}$-$N3^{pep}$ complex they are seen to be within 5Å from $Dvl2^{PDZ}$ residues.

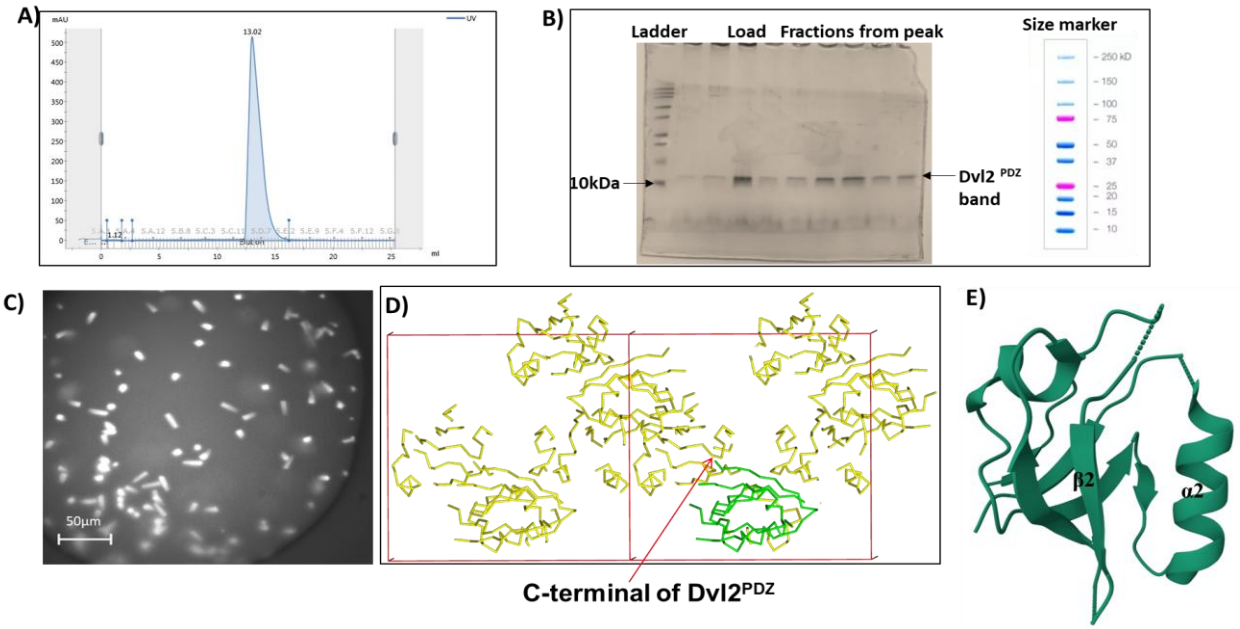


Figure 4.15: Purification and crystallization of WGEF^{pep} fused construct of Dvl2^{PDZ}. ***A)*** *Size exclusion chromatogram for Dvl2^{PDZ} fused with WGEF^{pep}.* ***B)*** *SDS-PAGE of size exclusion chromatography for Dvl2^{PDZ} fused with WGEF^{pep}.* ***C)*** *UV image of Crystals obtained for Dvl2^{PDZ} fused with WGEF^{pep}.* ***D)*** *The fused Dvl2^{PDZ} crystallized in the $I4_1$ space group with ~0.6 solvent content. The crystal packing was found to be markedly different from the other reported peptide fused Dvl2^{PDZ} structures (PDB ID: 2REY, 3CC0, 3CBX, 3CBY, and 3CBZ). The electron density for the fused binding motif at the C-terminal of Dvl2^{PDZ} is not observed (PDB ID: 8YR7), as this particular segment of the structure is fully exposed to the bulk solvent.* ***E)*** *Crystal structure of Dvl2^{PDZ} obtained from the crystals of fused Dvl2^{PDZ} with WGEF^{pep}.*

	***hDvl2*^{PDZ} crystallized with WGEF^{pep}**	***hDvl2*^{PDZ} fused with WGEF^{pep}**
Data collection		
Source	ID23-2 beamline of ESRF	BL21-PX beamline of Indus-2 synchrotron
Space Group	$P4_32_12$	$I\,4_1$
a (Å)	45.98	59.49
b (Å)	45.98	59.49
c (Å)	78.42	58.25
α(°)	90	90
β(°)	90	90
γ(°)	90	90
Resolution limits (Å)	32.51-1.75	42.07-3.00
R_{merge}	0.067(0.731)	0.058(1.091)
I/s (I)	23.3(3.8)	27.8(2.7)
Number of reflections	157526(7686)	30769(5090)
Unique reflections	9023(482)	2071(336)
Completeness (%)	100(100)	99.9(100)
Multiplicity	17.5(15.9)	14.9(15.1)
CC(1/2)	0.999(0.915)	1(0.731)
Refinement		
Resolution limits (Å)	32.51-1.75	29.75-3.00
Number of reflections	8984	2067
Working set	8594	1951
Test set	390	116
R_{work}/R_{free}	0.227/0.252	0.299/0.335
Number of atoms		
Protein	1245	512
Water	20	
B factors		
Protein atoms (Å^2)	39.26	114.32
Water	37.25	
RMSD from ideal values		
Bond length (Å)	0.007	0.009
Bond angles (°)	1.089	1.361
Ramachandran plot		
Preferred (%)	97.53	87.32
Allowed (%)	2.47	12.68
PDB code	8WWR	8YR7

Table 4.1: Data collection and refinement statistics for the structures of human Dishevelled2 PDZ domain, co-crystallized and fused with WGEF^{pep}, respectively*. Single crystal was used for data collection. Values in parenthesis are for the highest-resolution shell.*

To determine the differential contribution of these residues to the WGEF^{pep}-Dvl2^{PDZ} interaction, we determined the dissociation constants of the WGEF^{pep} variants for Dvl2^{PDZ}, employing Microscale Thermophoresis (MST). As a protein control, we used the hWGEF$^{330-581}$ construct in the MST study. The affinities of WGEF^{pep} and N3^{pep} for Dvl2^{PDZ} were found to be nearly identical, with their dissociation constant (K_d) being 45±5 μM and 54±7 μM, respectively. This observation validates our structural model (N3^{pep}-Dvl2^{PDZ}, PDB ID: 3CC0) adopted for studying WGEF-Dvl2^{PDZ} interaction. Further, we made interesting observations on the contribution of peptide residues to the WGEF^{pep}-Dvl2^{PDZ} interaction. Amongst the variant peptides, substitution of Gln at the *P$_{-1}$* (Fig 4.16A, C and Table 4.2) with Ala (WGEF^{pep}*^{Q8A}*) enhances the affinity of the peptide by almost 6-fold, whereas substitution of Ile at *P$_1$* to Ala (WGEF^{pep}*^{I10A}*) did not affect the interaction (Fig 4.16A, C and Table 4.2). Notably, amongst these two residues, only the side chain of Ser at *P$_{-1}$* makes direct interaction with Dvl2^{PDZ}, in the Dvl2^{PDZ}-N3^{pep} complex crystal structure. On the contrary, the substitution of Leu at the *P$_{-3}$* position (WGEF^{pep}*^{L6A}*) significantly reduces the affinity of the peptide for Dvl2^{PDZ} (Fig 4.16B, C and Table 4.2). Furthermore, substitution of tryptophan (WGEF^{pep}*^{W7A}*), aspartate (WGEF^{pep}*^{D9A}*), and proline (WGEF^{pep}*^{P11A}*) residues at the *P$_{-2}$, P$_0$*, and *P$_2$* positions, respectively, with alanine completely abolishes the WGEF-Dvl2^{PDZ} interactions (Table 4.2). Thus, from MST studies it is clear that residues at *P$_{-2}$, P$_0$*, and *P$_2$* positions (W, D & P) make crucial contributions to the WGEF-Dvl2^{PDZ} interaction, whereas leucine at the *P$_{-3}$* position might influence the peptide-Dvl2^{PDZ} interaction either by stabilizing the conformation or by enhancing the solubility of the peptide.

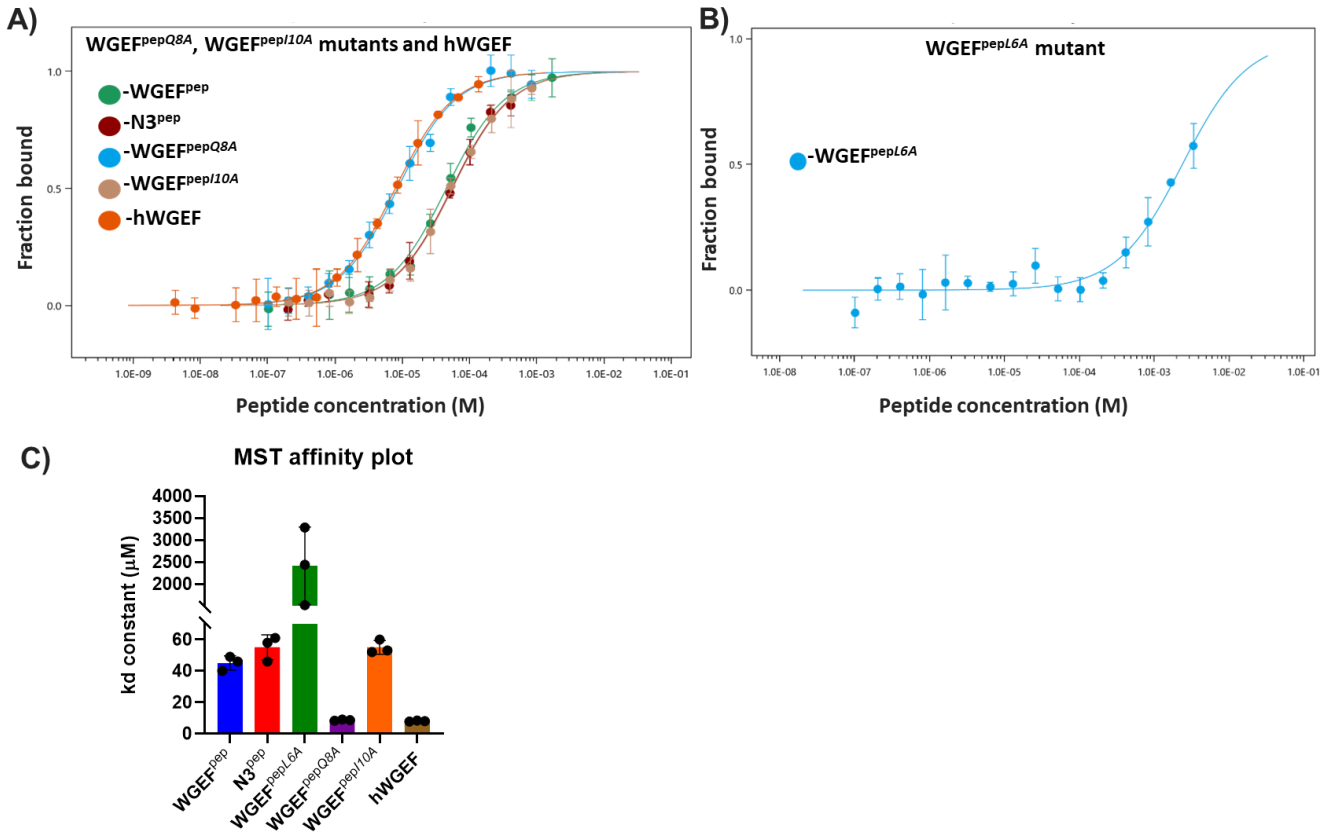


Figure 4.16: MST binding assays of $Dvl2^{PDZ}$ and $WGEF^{pep}$ variants. A and B) *Binding curve corresponding to the interaction of $Dvl2^{PDZ}$ with $WGEF^{pep}$, $N3^{pep}$, $WGEF^{pepQ8A}$, $WGEF^{pepI10A}$ peptides, $hWGEF^{330-581}$, and $WGEF^{pepL6A}$ peptide, respectively.* ***C)*** *Bar plots showing dissociation constants between $Dvl2^{PDZ}$ and different peptides.*

Peptides	Sequence	K_d (μM)
$WGEF^{pep}$	GSTFSLWQDIP	45±5μM
$N3^{pep}$	EIVLWSDIP	54±7μM
$WGEF^{pepL6A}$	GSTFSAWQDIP	2400±900μM
$WGEF^{pepW7A}$	GSTFSLAQDIP	N.D
$WGEF^{pepQ8A}$	GSTFSLWADIP	8.5±0.8μM
$WGEF^{pepD9A}$	GSTFSLWQAIP	N.D
$WGEF^{pepI10A}$	GSTFSLWQDAP	56±4.3μM
$WGEF^{pepP11A}$	GSTFSLWQDIA	N.D
hWGEF	330-581	7.9±0.3μM

Table 4.2: Consolidated table showing dissociation constant (K_d) values for binding studies between different peptides, hWGEF protein and $Dvl2^{PDZ}$.

4.3 Part B: To determine the binding mode of WGEF^{pep}-Dvl2^{PDZ} interaction

4.3.1 Background

From our biochemical studies it is observed that WGEF^{pep} interacts with Dvl2^{PDZ} through a conserved internal motif and we also identified the residues of the *internal peptide motif* that drive the WGEF^{pep}-Dvl2^{PDZ} interaction. However, due to lack of structural information, the mode of WGEF^{pep}-Dvl2^{PDZ} binding was not clear. The problem gets compounded as PDZ domains exhibit a spectrum of peptide binding modes. For example, in both canonical and non-canonical (internal peptide), the ligand adopts antiparallel β strand conformation and forms a β sheet by associating with the β2 strand of the PDZ domain[180]. In the canonical binding mode, the carboxyl group of the P_0 residue (C-terminal residue) engages with the X-Φ-G-Φ motif of the carboxylate binding loop (CBL) to stabilize the complex. However, in the non-canonical PDZ binding mode, the CBL may not adopt a closed conformation and instead engages with the internal peptide through a network of hydrogen bonds involving either the main chain or the side chain atoms of the peptide[164,175,181]. Earlier studies have shown that certain PDZ domains like Erbin^{PDZ} and Syntrophin^{PDZ} exhibit limited flexibility during peptide interaction. Perhaps that is why Erbin^{PDZ} specifically accommodates C-terminal peptides with higher sequence specificity. However, Synrophin^{PDZ} can bind with both, the C-terminal and the internal peptides, provided the peptide undergoes a change in conformation to form a β finger, in order to be incorporated within the peptide binding loop, as seen in nNOS-Syntrophin^{PDZ} complex[182,183]. Another PDZ domain containing protein, Par-6, displays allostery mediated binding for C-terminal peptides, which exhibits an increase in affinity for the ligand by 10-folds when Cdc42 interacts with its adjacent *Cdc42/Rac interactive binding* (CRIB) domain[184]. Par-6^{PDZ} domain also binds to the internal peptide motif of Pals1, via a change in conformation within the PDZ domain, which is independent of allosteric alterations[164]. Thus, it is clear from these studies that PDZ domains exhibit huge conformational dynamics to accommodate their binding motifs, leading to different modes of binding. Due to the limited availability structural data on the WGEF^{pep}-Dvl2^{PDZ} complex, their binding mode has remained unclear. Hence, to address this aspect we performed a total of 9 μs molecular dynamics simulations of the Dvl2^{PDZ} structure, complexed with WGEF^{pep}, WGEF^{pep}^{Q8A} and N3^{pep} peptides. Since for two of the former peptides, there was no crystal structure available, we modelled the respective Dvl2^{PDZ}-peptide complexes using the coordinates of the *apo* (PDB ID: 8WWR) (Table 4.1) and N3^{pep}

bound structures of Dvl2^{PDZ}. The trajectories obtained were then analyzed to study the mode of interaction between Dvl2^{PDZ} and WGEF peptides.

4.3.2 Methodology

4.3.2.1 Molecular Dynamics (MD) simulations of hDvl2^{PDZ} and interacting peptides

Molecular dynamics (MD) simulations were performed by using the Desmond module from the Schrödinger suite[148]. The crystal structure of *apo* human Dvl2^{PDZ} (PDB code: 8WWR) was used as a template for MODELLER based homology modelling in order to build the missing loops[185]. WGEF^{pep}, WGEF^{pep}*Q8A* and N3^{pep} (as positive control) peptides were modelled and docked on Dvl2^{PDZ} by using the previously available internal (N3^{pep}) peptide structure (PDB code: 3CC0). These peptides, along with the modelled Dvl2^{PDZ} structure, were further used for 1 µs MD simulation. *Apo* form of Dvl2^{PDZ} was also simulated by adopting a similar methodology. All the structures were placed in an orthogonal simulation box with a buffer distance of 10 Å and further solvated using TIP3P water molecules and Na^+ to neutralize the system[186]. Next, temperature equilibration was done at a constant temperature of 300 K using the Nosé-Hoover thermostat, and the pressure was equilibrated at 1.01 bar under NPT ensemble using Martyna-Tobias-Klein barostat[187–190]. Followed by temperature and pressure equilibration, MD runs were performed using OPLS_2005 force field[191,192]. MD runs were performed in triplicates.

4.3.2.2 Interaction propensity

Residue-wise peptide-PDZ domain interaction propensity was determined by plotting the probability of the residue side-chain lying within 4 Å from the PDZ residue throughout the simulation time. These interaction propensities were plotted as a heat map where the intensity of colour is directly proportional to the interaction propensity.

4.3.2.3 Dynamical coupling Analysis

The dynamical cross-correlation matrix (DCCM), dC_{ij}[193], was calculated as:

$$dC_{ij} = \langle r_i r_j \rangle - \langle r_i \rangle \langle r_j \rangle$$

Where, r_i and r_j are the position vectors of the i^{th} and j^{th} Cα atoms, respectively of $Dvl2^{PDZ}$, at time t. DCCM calculations were done using the R implementation of the Bio3D program (version 2.4-1)[194]. The spectral decomposition algorithm[151] was used to identify the dynamical sectors using the methodology (Statistical coupling analysis) reported earlier[93,195]. MD trajectories from triplicate runs were merged for dynamical coupling analysis.

4.3.3 Results

4.3.3.1 Residue-wise interaction propensity of $Dvl2^{PDZ}$ with $WGEF^{pep}$, $WGEF^{pepQ8A}$ and $N3^{pep}$ internal peptides

The MD simulation trajectories of the modelled $Dvl2^{PDZ}$-$WGEF^{pep}$, $Dvl2^{PDZ}$-$WGEF^{pepQ8A}$ and $Dvl2^{PDZ}$-$N3^{pep}$ structures were used to calculate the residue-wise interaction propensity. Consistent with earlier reports and our biochemical studies, the aspartate at the P_0 position of the internal peptides was found to be important, showing the highest propensity of interaction with the X-Φ-G-Φ motif of the CBL and with the α2 helix residues of the PDZ domain (Fig 4.17A and D). Apart from Asp at the P_0 position, it was observed that residues at the positions P_{-1}, P_{-2}, and P_{-3} also interact with the β2-α2 region of $Dvl2^{PDZ}$ (Fig 4.17A and D). Interestingly, most of the interactions between $WGEF^{pep}$-$Dvl2^{PDZ}$ and $WGEF^{pepQ8A}$-$Dvl2^{PDZ}$ are similar to those observed in the $N3^{pep}$-$Dvl2^{PDZ}$ (Fig 4.17). However, subtle differences in the interactions made by the P_{-4} and P_{-5} residues of $WGEF^{pep}$ residues make the mode of $WGEF^{pep}$ binding with the PDZ domain divergent when compared with $N3^{pep}$-$Dvl2^{PDZ}$ interactions. In $N3^{pep}$-$Dvl2^{PDZ}$ complex, residues at P_{-4} and P_{-5} positions interact with β2 strand of the $Dvl2^{PDZ}$ domain, which was absent in the $WGEF^{pep}$-$Dvl2^{PDZ}$ model (Fig 4.17). Additionally, the residues towards the N-terminus of $WGEF^{pep}$ did not seem to contribute significantly to the

interaction. Perhaps that is why in WGEF^{pep}-Dvl2^{PDZ} simulations, residues from *P-4* onwards (towards N-terminus) exhibit higher RMSF values as compared to N3^{pep} (Fig 4.18).

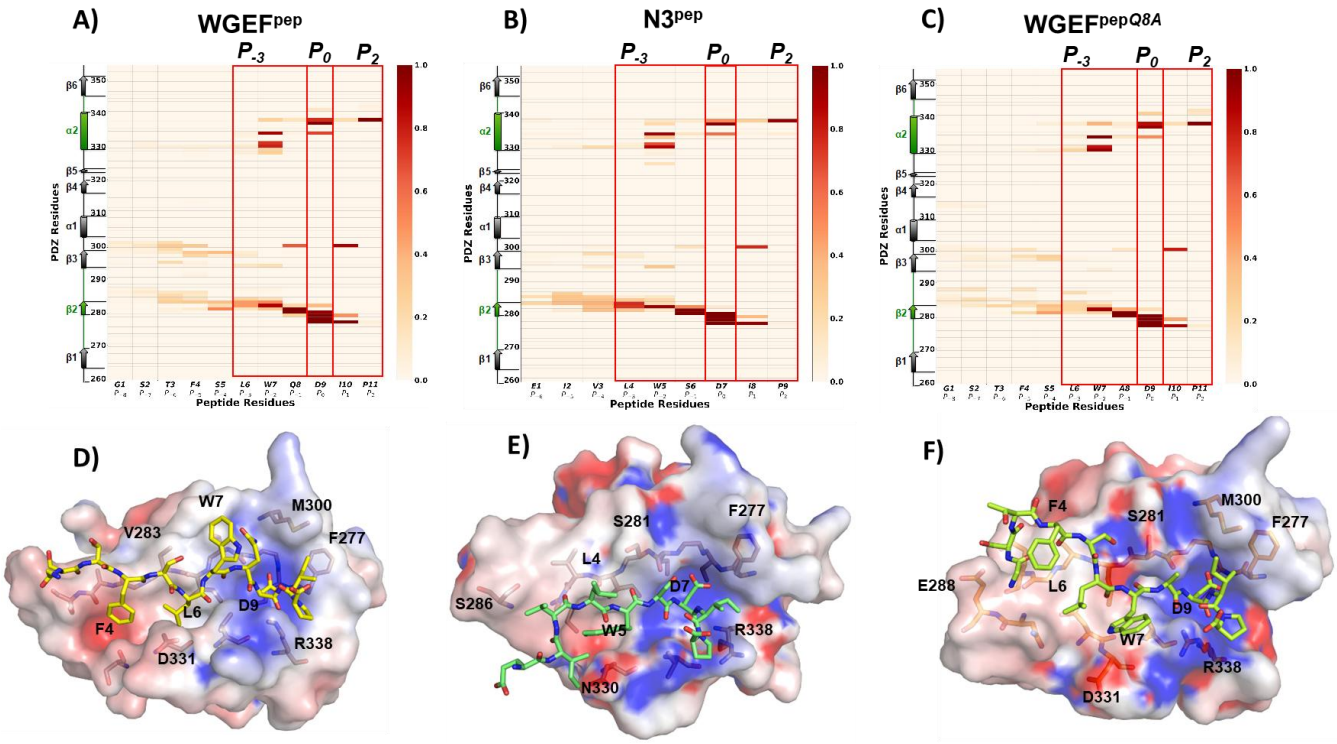


Figure 4.17: Interaction propensity of Dvl2^{PDZ} with peptide ligands. ***A)*** *Interaction propensity with WGEF^{pep}* ***B)*** *N3^{pep} and* ***C)*** *WGEF^{pepQ8A}* ***D-F)*** *show the structure of Dvl2^{PDZ} (represented as surface) bound with peptide variants, WGEF^{pep}, N3^{pep} and WGEF^{pepQ8A} (represented in ball-and–sticks), belonging to one of the stable trajectories, respectively.*

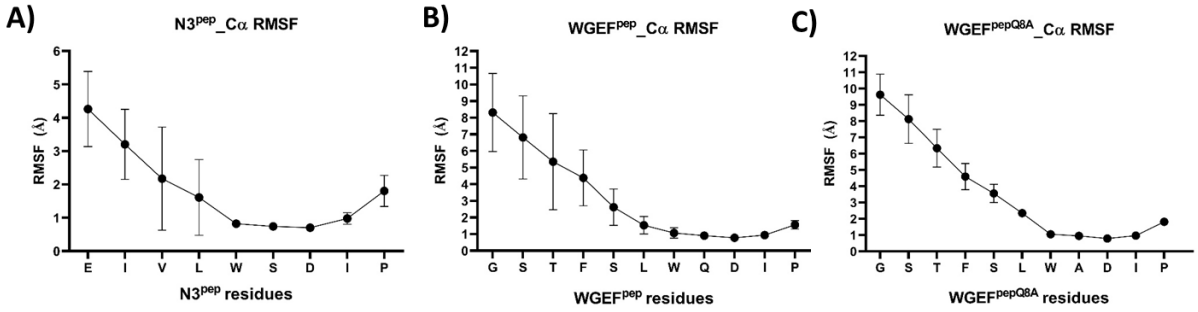


Figure 4.18: Root-mean-square fluctuation (RMSF) plots of the Cα atoms of peptides from MD simulation studies with Dvl2^{PDZ}. ***A)*** *RMSF of N3^{pep},* ***B)*** *WGEF^{pep} and* ***C)*** *WGEF^{pepQ8A} used for studying different internal peptide–Dvl2^{PDZ} interaction employing MD simulations.*

Our MST studies reveal that the substitution of Gln at *P-1* with Ala (WGEF^{pepQ8A}) in WGEF^{pep} considerably increases affinity of this peptide variant towards Dvl2^{PDZ} by 6-folds (Fig 4.16A and C). To understand this increase in the affinity, we first plotted the difference in

interaction propensity of WGEF^{pepQ8A} and WGEF^{pep} to observe if there is any significant change in the binding. However, the difference propensity plot showed no significant difference in the interaction propensity (Fig 4.19A). Therefore, we closely examined the MD trajectories pertaining to both of these systems, sampled at the known time intervals. The structure analysis of the WGEF^{pepQ8A}-Dvl2^{PDZ} complex from MD simulation indicates a higher propensity for the Trp at the *P-2* position to stay in the hydrophobic groove formed by the Ile 280, Leu 278, Leu 337 and Val 341 residues of the β2-α2 region of the PDZ domain (Fig 4.19B). However, the backbone conformation of this residue that favours the hydrophobic interaction with the PDZ domain is governed by the residue at the *P-1* position. The presence of a residue with a longer side chain at *P-1* could restrict the conformational space of Trp at *P-2* and, hence, may hinder the additional hydrophobic interactions between this residue and the PDZ domain (Fig 4.19B). Thus, the substitution of Gln with a shorter-side chain amino acid, like Ala, at *P-1* position promotes WGEF^{pep}-Dvl2^{PDZ} interaction, as seen from MST studies. Thus, our MD studies provide insights into the Dvl2^{PDZ}-WGEF interaction and corroborate our biochemical studies on the Dvl2^{PDZ}-peptide complexes.

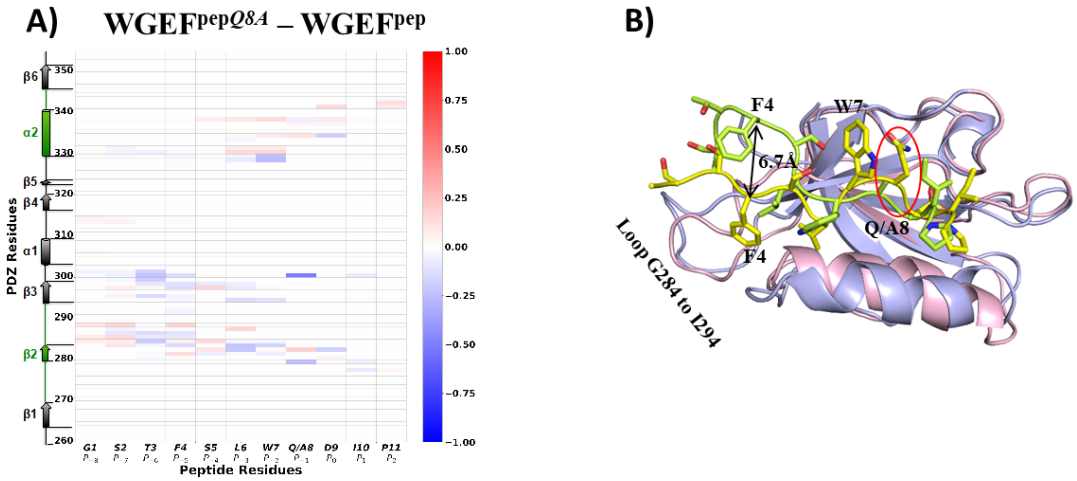


Figure 4.19: Interaction propensity difference for WGEF^{pepQ8A} and WGEF^{pep}. ***A)*** *Difference in the interaction propensity of WGEF^{pepQ8A} and WGEF^{pep}.* ***B)*** *Superimposition of WGEF^{pep} (yellow peptide) and WGEF^{pepQ8A} (green peptide) structures from MD simulation. The residues marked in the red circle correspond to yellow-glutamine and green-alanine.*

4.3.3.2 Internal peptide ligands of Dvl2^{PDZ} exhibit divergent modes of interaction

Furthermore, our MD studies and previous reports on Dvl2-internal peptide complexes show no conserved mode of interaction between the internal peptides and Dvl2^{PDZ} [52,183,196]. In

the $Dvl2^{PDZ}$-$N3^{pep}$ crystal structure, it is observed that $N3^{pep}$ forms antiparallel β strand interaction with the β2 strand of the $Dvl2^{PDZ}$ domain (Fig 4.20).

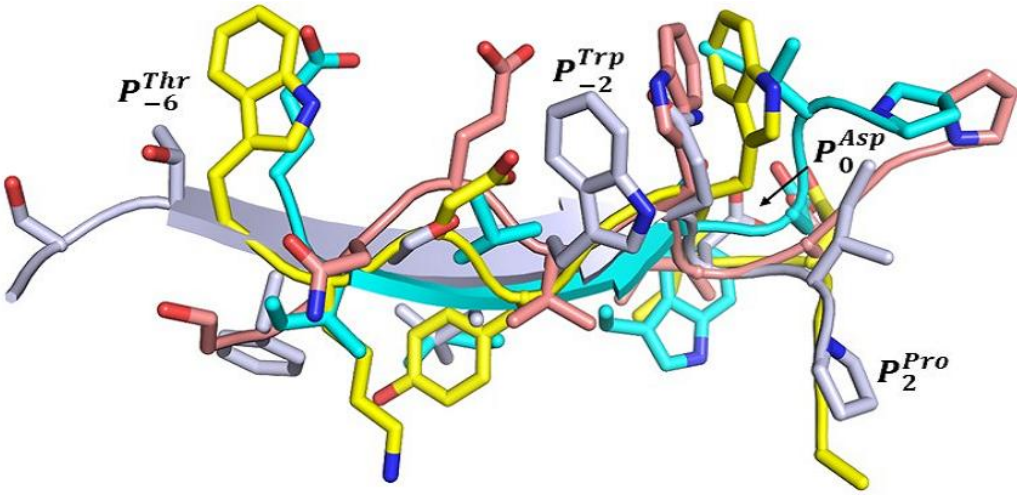


Figure 4.20: Superimposition of internal peptides present in binding the pocket of $Dvl2^{PDZ}$ and Secondary structural propensity of residues corresponding to the peptide*. Yellow colour peptide represents N1 inhibitory peptide (PDB Code: 3CBY), pink colour peptide represents N2 inhibitory peptide (PDB Code: 3CBZ), blue colour peptide represents N3 inhibitory peptide (PDB Code: 3CC0) and grey colour peptide represents $WGEF^{pep}$ peptide from one of the stable trajectories. Residue positions of $WGEF^{pep}$ are labelled.*

Thus, to assess whether there is a conservation of β sheet formation of the internal peptides with the β2 strand of the $Dvl2^{PDZ}$ domain, we calculated the secondary structural propensity of the peptide residues using their respective MD trajectories. In the case of $Dvl2^{PDZ}$ domain and internal peptide complexes ($WGEF^{pep}$, $WGEF^{pepQ8A}$ and $N3^{pep}$ peptides), it was observed that these ligands did not show consistency in forming β sheet with the β2 strand of the $Dvl2^{PDZ}$ domain, as the ligand may not essentially adopt β strand conformation. Perhaps this loss of secondary structural conformation might aid the flexibility of the peptide and, hence, enhance its interaction with $Dvl2^{PDZ}$ (Fig 4.21). Thus, it is observed that the internal peptide does not adopt one secondary structure but undergoes a conformational change to be recognized by the $Dvl2^{PDZ}$ domain.

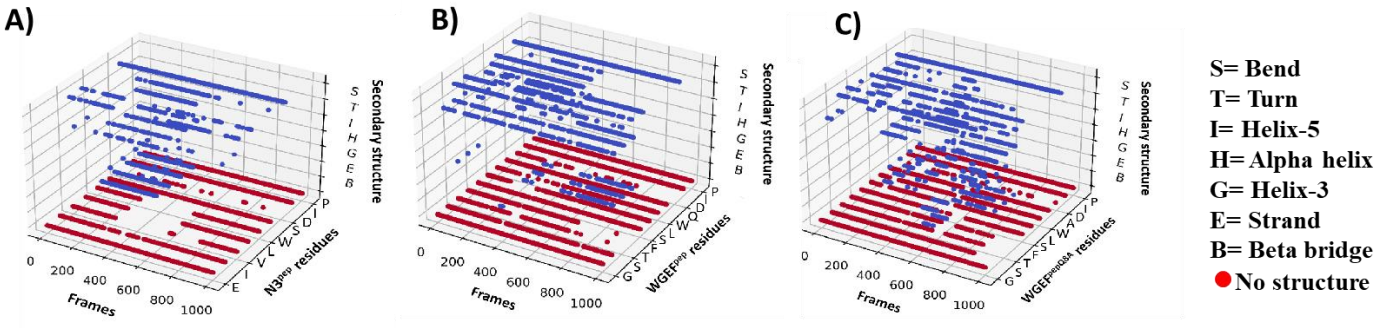


Figure 4.21: Secondary structure probability of the peptides, obtained from the MD trajectories. A) *$N3^{pep}$,* ***B)*** *$WGEF^{pep}$* ***and C)*** *$WGEF^{pepQ8A}$ residues during 1μs simulation of peptide–$Dvl2^{PDZ}$ interaction. The occurrence of a specific secondary structure is denoted by a blue dot within its corresponding axis. Red colour spots indicate the absence of secondary structure.*

4.3.3.3 $Dvl2^{PDZ}$ exhibits ligand-dependent conformational changes

The secondary structure probability plot shows that the peptide undergoes conformational changes, which perhaps promotes its recognition by the $Dvl2^{PDZ}$ domain. Next, we wanted to explore whether $Dvl2^{PDZ}$, too, exhibits signatural conformational dynamics in the presence and absence of the peptide. From the earlier reported MD simulation studies on PDZ domains, we can underline that the dynamic features of the PDZ domains presented by different MD studies vary extensively[183,197]. This could be due to variations in the type of PDZ structures (*apo*, or bound with C-terminal and/or internal peptide) employed in the MD simulations and the time scale of the simulation[197,198]. Since we have elucidated the *apo* structure of human $Dvl2^{PDZ}$ (PDB ID: 8WWR), we looked into the dynamics of this structure and other modelled structures bound to different peptides. Furthermore, superposition of available $Dvl2^{PDZ}$ structures shows deviations with RMSD values in the range of 1.7-2Å. Importantly, significant conformational variations are observed in the α2 helix and β2-β3 loop, which are a part of the peptide binding pocket. This conformational variation is more prominent in the structures of human $Dvl2^{PDZ}$, bound with different peptides (Fig 4.22). This opens up the question of whether the peptide binding pocket of $Dvl2^{PDZ}$ is preformed or it exhibits peptide induced conformational fit.

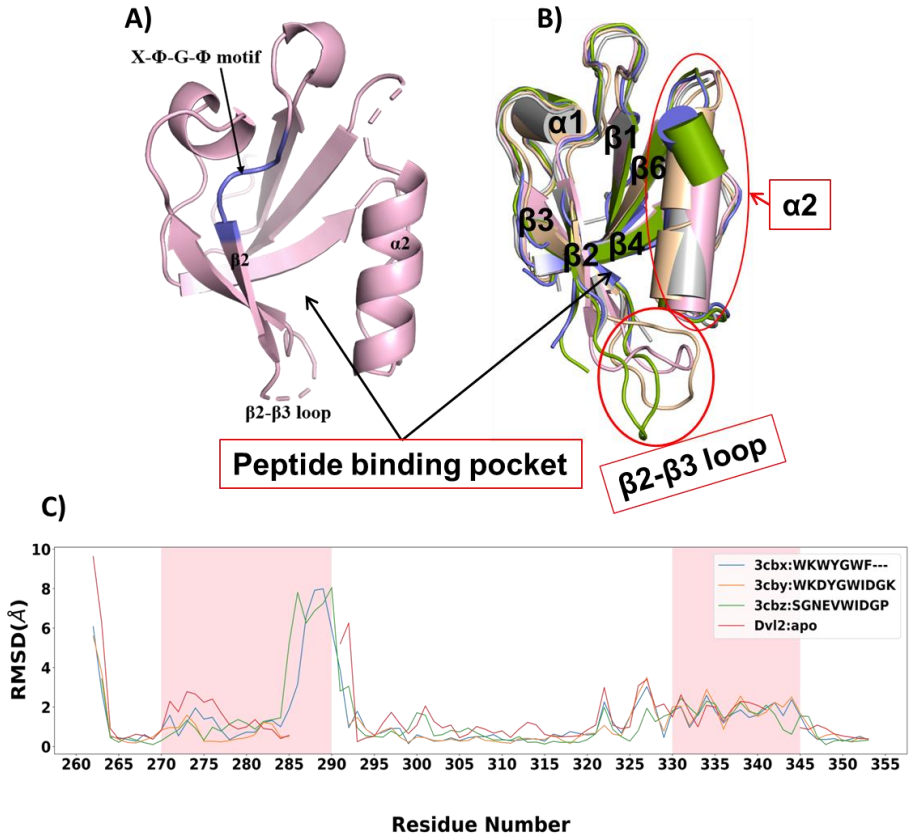


Figure 4.22: Conformational variations between apo and ligand bound Dvl2^{PDZ} structures. ***A)*** *Apo structure of human Dvl2^{PDZ}* ***B)*** *Superimposition of apo Dvl2^{PDZ} and peptide ligand bound Dvl2^{PDZ} structures colour coded as Dvl2^{PDZ} (PDB ID: 8WWR)-grey, C1^{pep} (PDB ID: 3CBX)-green, N1^{pep} (PDB ID: 3CBY)-blue, N2^{pep} (PDB ID: 3CBZ)-peach, N3^{pep} (PDB ID: 3CC0)-pink, major conformational deviations in the α2 helix and β2-β3 loop regions of Dvl2^{PDZ} marked with a red circle.* ***C)*** *Residue-wise RMSD plot between the apo and internal peptide ligand bound Dvl2^{PDZ} structures.*

To address this question, we analyzed the MD run trajectories to identify which amino acids in the Dvl2^{PDZ} are dynamically coupled during the simulation, both in presence and absence of peptide ligands. For this analysis, we used our hDvl2^{PDZ} *apo* structure, as the other available *apo* Dvl2^{PDZ} structure from *Xenopus* (PDB ID's: 3FY5 and 2F0A) differs from our structure with a substitution of an amino acid (Human M^{323}/*Xenopus* I^{310}).

The rationale behind this analysis is to figure out whether there are any dynamical changes in the PDZ domain during peptide recognition. Therefore, to identify the amino acids that are dynamically coupled, we performed spectral decomposition of the *dynamic cross-coupling matrix (cij)* (Fig 4.23A) obtained from all the MD simulations trajectories. This

identifies dynamically coupled clusters of residues, referred to as sectors. These sectors can be divided into five independent components (ICs), depending on their hierarchy of coupling, corresponding to the five highest eigenvalues of the matrix. The Statistical coupling analysis (SCA) provides two main clusters of dynamically coupled amino acids, referred to as Sector 1 and Sector 2.

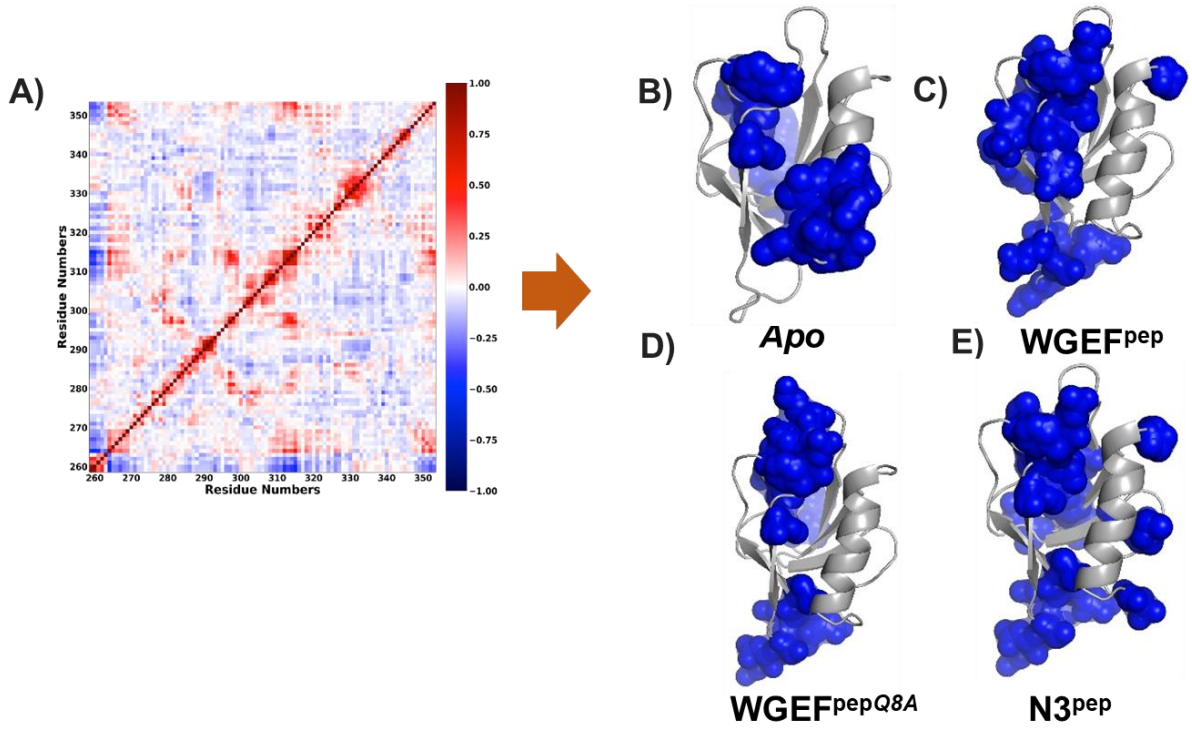


Figure 4.23: Dynamical coupling Analysis. ***A)*** *Representative plot of the dynamic cross-coupling matrix (cij) used to calculate dynamically coupled amino acids belonging to Sector1 which are shown as blue spheres on the* ***B)*** *apo,* ***C)*** *WGEF^{pep}* ***D)*** *WGEF^{pepQ8A}, and* ***E)*** *N3^{pep} bound Dvl2^{PDZ} structures, respectively.*

It is interesting to note that Sector 1, corresponding to all the simulations, which primarily comprises residues belonging to the peptide binding pocket, shows significant variations in terms of their residue composition and location of residues in the protein structure (Fig 4.23B-E). Interestingly, Sector 1 of *apo* Dvl2^{PDZ} was found to consist of residues residing in β2 and α2 regions of the domain (Fig 4.23B), whereas Sector 1 of Dvl2^{PDZ} internal peptide (WGEF^{pep}, WGEF^{pepQ8A} and N3^{pep}) complexes had residues from the β2 strand, β1-β2 and β2-β3 loops (Fig 4.23C-E). These sectors can be divided into five ICs (Fig 4.24). These findings further validate the internal allostery of PDZ domains as elucidated by other studies[199]. Additionally, variations in Sector 1 composition highlight peptide-dependent conformational

adaptation of the Dvl2^{PDZ} domain. In contrast to the previous study by Munz *et al.*, we observe that the α2 helix of the *apo* Dvl2^{PDZ} structure has limited conformational flexibility rather than having the ability to explore all possible conformational space specific to the ligand-bound forms. Perhaps, divergent observations of our studies from those published before could be due to the absence of a "true" *apo* structure of Dvl2^{PDZ}. Furthermore, the shorter duration of earlier MD simulation (200 ns)[182] might have exaggerated the conformational sampling of the PDZ domain. Thus, our structural and MD simulation studies suggest that Dvl2^{PDZ} adopts peptide-dependent conformational change due to its intrinsic flexibility. In short, we can infer that Dvl2^{PDZ} may not have a preformed binding pocket, rather it modulates the binding pocket based on the substrate available for binding. This adaptation is perhaps responsible for the peptide (ligand) promiscuity of Dvl2^{PDZ}.

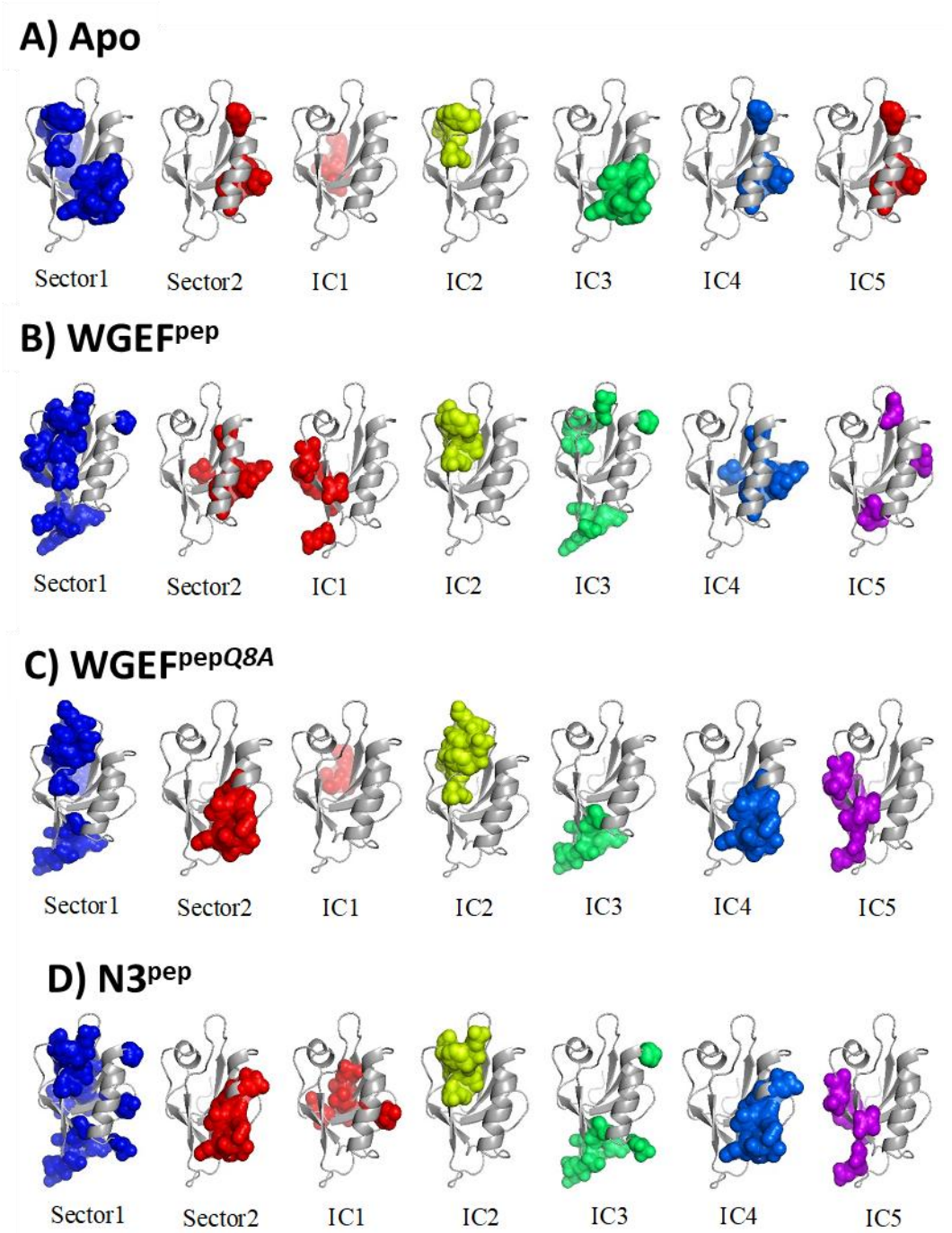


Figure 4.24: Sectors and Independent components (ICs) mapped on Dvl2^{PDZ}. ***A)*** *Sectors and ICs of apo Dvl2^{PDZ},* ***B)*** *Dvl2^{PDZ} simulated with WGEF^{pep},* ***C)*** *Dvl2^{PDZ} simulated with WGEF^{pepQ8A} and* ***D)*** *Dvl2^{PDZ} simulated with N3^{pep}. Dvl2^{PDZ} structure is shown in cartoon, and residues corresponding to sectors and ICs are shown as spheres with different colours.*

4.4 Proposed mechanism of WGEF activation based on the AlphaFold predicted structure of hWGEF

To understand how $Dvl2^{PDZ}$ binds to the identified binding motif within the structure of hWGEF, we looked at the AlphaFold predicted structure of hWGEF (Fig 4.25A and B). Notably, a significant portion of the structure, including the helix that connects PDZ binding motif to the NID domain, is predicted with low confidence levels (Fig 4.25A and B). Furthermore, the spatial disposition of different domains of the structure appears to be less reliable as the loops that connect these domains are predicted with low confidence levels. However, to propose a hypothesis using this model and our experimental results, we superimposed the part of the hWGEF structure that is predicted with higher confidence on its closest structure homolog, *Leukemia-associated RhoGEF* (PDB ID: 1X86). This comparison indicates that in the absence of $Dvl2^{PDZ}$ interaction, GTPase (RhoA) interacting DH domain of hWGEF is obstructed by the NID and the SH3 domain (Fig 4.25C and D). Thus, the intra-protein interaction of the protein could render the GEF into its inactive state (Fig 4.25E). When the $Dvl2^{PDZ}$ domain engages with the "binding motif", it could introduce conformational changes in the NID, resulting in the shifting away of this domain from the GTPase binding pocket of the GEF. Thus, hWGEF-$Dvl2^{PDZ}$ association could partially release the autoinhibitory state of the former (Fig 4.25F). Perhaps, some other unknown mechanism might also disengage the SH3 domain from DH domain, which would result in the complete activation of GEF (Fig 4.25G). Since the putative mechanism of hWGEF activation proposed here is based on the AlphaFold predicted structure, substantive GEF activation mechanism warrants further structural studies.

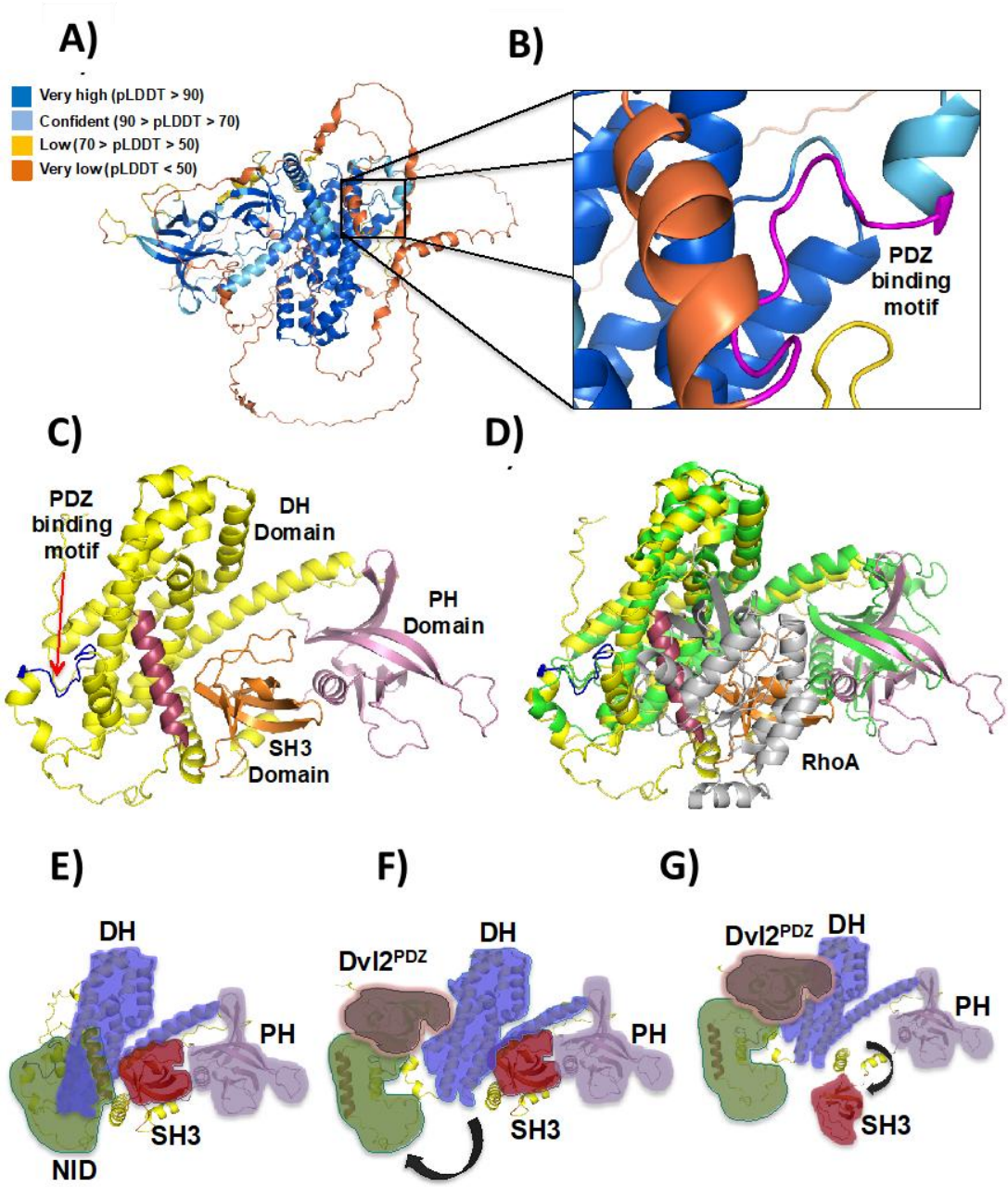


Figure 4.25: hWGEF activation mechanism based on its AlphaFold structural model. ***A)*** *Overall AlphaFold predicted structure of hWGEF. Regions of the structure are coloured based on per-residue confidence score (pLDDT). The level of increase in prediction confidence is scaled from red (low) to blue (high).* ***B)*** *The Dvl2^{PDZ} binding motif of hWGEF (shown in magenta) is zoomed.* ***C)*** *Spatial disposition of DH (yellow), PH (pink), SH3 (orange) and NID (brick) domains of hWGEF. The Dvl2^{PDZ} binding motif is shown in blue.* ***D)*** *Superimposition of hWGEF and Leukemia-associated RhoGEF (PDB ID: 1X86) structures. The latter is complexed with RhoA (shown in grey).* ***E)-G)*** *Putative mechanism of Dvl2^{PDZ} induced activation of hWGEF. Binding of Dvl2^{PDZ} to hWGEF could dismantle the intra-protein interaction, thereby unblocking the GTPase binding pocket of the GEF.*

4.5 Discussion

Despite nearly three decades of studies on PDZ domains, their modes of interaction with their binding partners have remained elusive. It is quite intriguing to learn that despite having highly conserved domain structure, PDZ domains exhibit divergent modes of interaction with their binding partners[51,182,200]. Dishevelled, being a signalling hub, recognizes diverse binding partners through its different domains, especially involving its PDZ domain[201]. Therefore, as seen in other PDZ domains, $Dvl2^{PDZ}$ exhibits significant promiscuity by recognizing binding partners having both C-terminal as well as internal binding peptide motifs[52,163,178]. Thus, it is expected that $Dvl2^{PDZ}$ displays divergent binding modes with its partners[175]. In the context of Wnt-PCP signalling, Dvl2 assembles at the cytosolic domain of the activated Frizzled receptors and triggers RhoA by activating the downstream GEF, WGEF. The function of Dvl2 in this signalling event is to activate WGEF by releasing it from its autoinhibited state. This step involves binding of the PDZ domain of Dvl2 with the WGEF. However, the region of WGEF that facilitates its interaction with $Dvl2^{PDZ}$ was not known. Here, we have successfully identified an internal peptide motif of WGEF, lying between the N-terminal inhibitory and the DH domains, that facilitates the interaction of the protein with Dvl2 to activate itself. Indeed, this is one of the crucial mechanisms involved in releasing the autoinhibition of WGEF. Our biochemical assays show that residues at positions P_2, P_0 and P_{-2} of the binding motif are vital for the $Dvl2^{PDZ}$ - peptide interaction. Additionally, substitution of the leucine at the P_{-3} position significantly reduced the interaction indicating the role of this residue in the solubility of the peptide. Despite extensive efforts, to elucidate a detailed molecular mechanism of WGEF-$Dvl2^{PDZ}$ interaction, which included co-crystallization and fusion strategies, we could not resolve the structure of $Dvl2^{PDZ}$-$WGEF^{pep}$ complex. The inability to observe electron density for $WGEF^{pep}$ in crystal structures suggest dynamic or weak interactions, consistent with the low binding affinity observed biochemically. Consequently, we used computational modelling, MD simulations and comparative structural analysis, to show that $Dvl2^{PDZ}$ exhibits diverse modes of ligand recognition, which is perhaps true of $WGEF^{pep}$, as well. Thus, this study advances our understanding of WGEF regulation through its interaction with $Dvl2^{PDZ}$. The identification of $WGEF^{pep}$ as an internal motif crucial for this interaction provides a new perspective on the molecular mechanism underlying Wnt signalling and GEF activation. Furthermore, in terms of consensus sequence, the substitution of P11A at the P_2 position of $WGEF^{pep}$ makes the peptide homologous to the class III PDZ binding peptide. However, in

terms of the mode of binding, they differ significantly. Class III peptide, bound to nNOS^{PDZ} (PDB ID: 1B8Q)[202], appears to be relatively shifted towards the N-terminal of the α2 helix of the Dvl2^{PDZ} domain (Fig 4.26). Therefore, just based on the sequence of ligand peptides, mode of binding cannot be inferred.

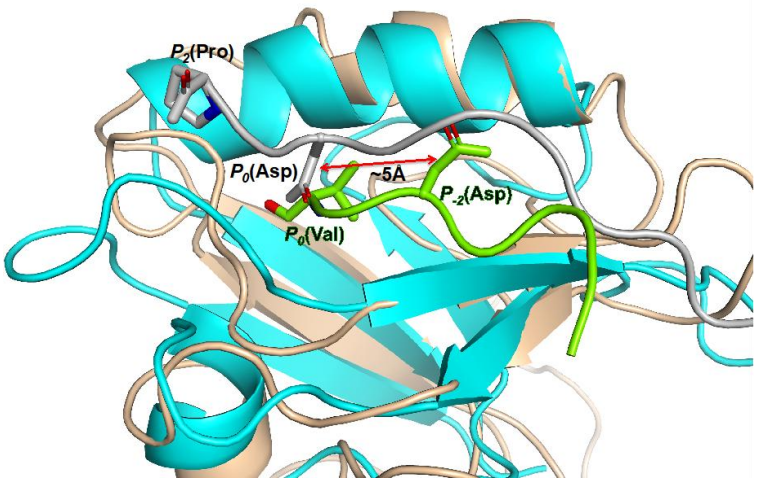


Figure 4.26: Superimposition of Dvl2^{PDZ} (light brown) - internal peptide (grey) & nNOS^{PDZ} (cyan) - Class III peptide (green) (PDB ID: 1B8Q) complex structures. *The binding pocket of conserved Asp (P_0 in internal peptide and P_{-2} in Class III peptide) is shifted by ~5Å, thus indicating that the PDZ-peptide interactions in these two classes are highly divergent.*

Additionally, from MD studies, we observed that WGEF^{pep} and N3^{pep} peptides do not seem to adopt the stringent β strand conformation, as seen in case of the crystal structure of other PDZ-*internal peptide* complexes[175]. In short, the ligand (peptide) and the Dvl2^{PDZ} undergo conformational adaptation for mutual recognition. A genome wide screen for internal PDZ binding motifs indicates that the tendency of PDZ domains to interact with internal peptides is more comm on than earlier understood[181].

Chapter 5

To determine how the WGEF-Dvl2 molecular interaction manifests during Wnt-PCP pathway

5.1 Background

From biochemical and MD studies, we have identified that $Dvl2^{PDZ}$ interacts with WGEF through an evolutionary conserved internal peptide motif. However, to scale down the complexity of the problem, we have probed their binding by studying the *in vitro* interaction between $WGEF^{pep}$ and the PDZ domain of Dvl2. But a physiologically more relevant scenario would be to probe the interaction formed by the internal peptide motif when it is actually part of the GEF with Dvl2, particularly in the light of Wnt-Frizzled7 (FZD7) signal transduction events. In this study, we have developed a cell-based assay to probe Dvl2-WGEF interaction operating under the non-canonical Wnt5A-FZD7 pathway.

Wnt ligands such as Wnt5A and Wnt11 have been studied extensively for their function in the Wnt-PCP pathway. These ligands interact with FZD receptors and other co-receptors, such as RYK, ROR1/2, etc., to transmit intracellular signals. Wnt11 is responsible for cell proliferation, migration and transformation of mouse intestinal epithelial cells[203]. Furthermore, Wnt11/Frizzled7 interaction is also reported to bring about convergent extension movement during *Xenopus* gastrulation through activation of the WGEF-RhoA axis[24]. Similarly, Wnt5A is one of the most highly studied Wnt ligands for its role in the non-canonical Wnt pathway[58]. This protein is reported to regulate epithelial-to-mesenchymal cell transition in mammary epithelial cells, thus regulating their migration and invasion[204]. It is also shown to regulate collective vascular cell migration during angiogenesis[205]. Also, Wnt5A is reported to bring about migration of breast cancer cells through the Dvl2/Daam1/RhoA axis[58].

Therefore, we engineered Human embryonic kidney (HEK) 293 cell lines by instilling the non-canonical Wnt5A/FZD7 pathway by transfecting the cell lines with plasmids carrying all the essential genes that operate under this pathway and required for us to study Dvl2-WGEF interaction. This chapter summarizes our efforts to standardize various phenotypic readouts to

study Dvl2-WGEF interaction operating under the Wnt-PCP mode of the non-canonical Wnt signalling.

5.2 Methodology

5.2.1 Plasmids and transfection

Modified *pEF4/myc-His* vectors (two vectors corresponding to construct 1 and construct 2) containing gene sequences encoding Human Frizzled7, Dvl2, WGEF (Construct 1) and RhoA (Construct 2) were procured from M/s Geneart (Fig 5.1). WGEF in construct 1 was tagged with Yellow fluorescent protein (YFP) and RhoA in construct 2 had Green fluorescent protein (GFP) tag to track the expression of proteins. For fluorescence microscopy experiments, HEK293 cells were cultured in 6-well plates with poly L-lysine coated coverslips, one day before transfection. They were grown in 10% FBS added DMEM medium at 37℃ with 5% CO_2 until 80-90% confluency was achieved. Co-transfection of the cell lines with both construct 1 and construct 2, was done by using Lipofectamine 3000 reagent as per the manufacturer's manual (Invitrogen). Briefly, the transfection reagent and the constructs were added in OPTI-MEM medium without FBS and incubated at RT/30 min. Further, this mixture was added to a PBS-washed cell monolayer, followed by incubation at 37℃/5 h with 5% CO_2 in a humidified incubator. After 5 h, the media was replaced with 10% FBS added DMEM and incubated at 37℃/48 h before fixing and staining the cells for microscopy. For generating conditional cell lines expressing genes from construct 1, (MDA-MB-231) human breast adenocarcinoma cells were transfected with this construct in a T75 flask (Tarson) using a Polyethylenimine (PEI) transfection reagent. For this, PEI and DNA were mixed in a ratio of 3:1 in serum-free OPTI-MEM medium and incubated at RT/30 min[206]. This mixture was then added to a PBS-washed monolayer, followed by incubation at 37℃/5 h with 5% CO_2 in a humidified incubator. After 5 h, 10% FBS was added and cells were incubated at 37℃/48 h before harvesting.

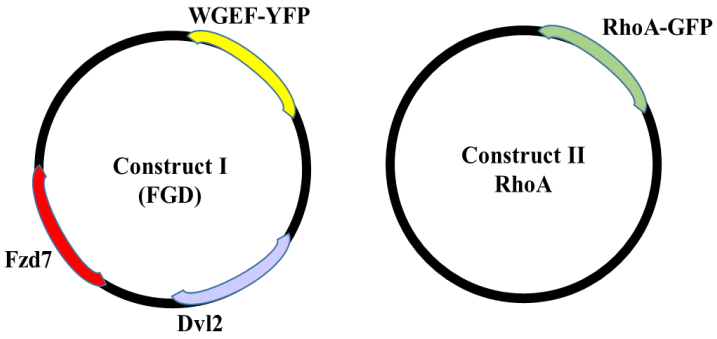


Figure 5.1: Constructs containing genes of Wnt-PCP pathway proteins used for transfection. *WGEF is tagged with Yellow fluorescent protein and RhoA is tagged with Green fluorescent protein.*

5.2.2 Preparation of Wnt5A conditioned medium

Stably transfected Wnt5A-producing mouse fibroblast-like cell line ATCC# CRL-2814 (L-Wnt5A) was purchased from ATCC and grown in DMEM containing 0.4 mg/mL geneticin, 4 mM L-glutamine, 10% FBS and 1 mM sodium pyruvate. After two passages, cells were split (1:10) at 90% confluency and grown for 4 days, following which the media containing extracellularly secreted Wnt5A was collected (batch 1). These cells were replenished with new media and further grown for an additional 3 days before harvesting the media (batch 2). Both the conditioned media (CM) batches were pooled and passed through a 0.2 µm filter and stored at 4°C. Control CM was also made using the same protocol from HEK293 cells.

5.2.3 Purification steps for Wnt5A

Wnt5A purification steps were done according to the Wnt purification protocol[207]. 1% Triton X-100 was added to the filtered conditioned media (CM). This Wnt5A-containing CM was loaded on a blue sepharose column pre-equilibrated with 20 mM Tris, pH 7.5, 150 mM KCl and 1% CHAPS followed by washing. Gradient elution was done with a buffer containing 1.5 M KCl. Elute fractions were pooled and concentrated for size exclusion chromatography (SEC) using 1% CHAPS containing PBS. After SEC, elute peak was loaded on a Heparin cation-exchange column equilibrated with 1% CHAPS containing PBS and elution was done using a gradient of 1 M NaCl containing buffer. Elute fractions were run on SDS-PAGE to analyze the presence of Wnt5A.

5.2.4 Wound healing assay

MDA-MB-231 cells were grown in a 24-well plate in serum-free medium and incubated for 24 h to form a cell monolayer. A wound was made in the monolayer using a 200 µL pipette tip, PBS wash was done to remove the detached cells. Cells were allowed to migrate for 24 h in the presence of Wnt5A CM and Control CM. Images were taken at 0, 8, 16 and 24 h using a BioTek cytation 5 plate reader, and the wound closure area was quantified with the use of ImageJ software. The data was normalized and plotted as the percentage of wound closure using GraphPad Prism. The assay was performed in triplicate.

5.2.5 Fluorescence microscopy sample preparation

Coverslips, containing cell monolayers, were rinsed twice with PBS, after which cells were fixed using 4% paraformaldehyde at RT/20 min. Cells were again washed with PBS and then permeabilized with 0.1% Triton X-100 at RT/20 min. Repeated washing of the cells with PBS was done to remove the detergent (Triton X-100). CF® 594 phalloidin stain (Biotium) was used to stain actin filaments and Fluoromount-G^{TM} mounting medium, with DAPI (Invitrogen) was used to mount the coverslip and stain the nucleus, respectively. STELLARIS confocal microscope (Leica) was used for the imaging.

5.2.6 Fluorescence-Activated Cell Sorting (FACS) and antibiotic based selection of transfected cells

FACS was performed to selectively isolate the transfected MDA-MB-231 cells which express the YFP-tagged WGEF. After 48 h of transfection, cells were detached using TrypLE express reagent (Invitrogen) and centrifuged to eliminate the media. Cells were resuspended in the FBS-free medium such that the final count was $1x10^6$ cells in 1 mL. Cells were then passed through a cell strainer to remove clumps and transferred to sterile FACS tubes for further use. After FACS, to generate cell lines conditionally expressing Wnt-PCP proteins, sorted cells were seeded in a 24-well plate such that each well had approximately 100 cells in it. These cells were allowed to adhere for 24 h before growing them with various concentrations of zeocin antibiotic (0-500 µg/mL). For antibiotics based selection of transfected cells, firstly, lowest concentration of zeocin to kill non-transfected cells was determined (250 µg/mL), then the transfected cells were grown in presence of 250 µg/mL zeocin to select cells conditionally expressing Wnt-PCP proteins.

5.3 Results

5.3.1 Attempts to purification of Wnt5A

Purification of Wnt proteins in their active form has proven to be challenging due to their hydrophobic nature and post-translational modifications. These proteins are post-translationally modified with covalently attached lipids and glycans, which tend to reduce their solubility. Despite these difficulties, Wnts, such as Wnt3A and Wnt5A have been successfully purified from larger conditioned medium (CM) volumes (2-6 liters)[208]. For our experiments, we expressed Wnt5A in 300 mL-500 mL of cell culture media using the reported protocols. However, in the first round of purification, involving blue sepharose column, Wnt5A bands were not seen in the eluted fractions. This lack of visibility of the protein band on SDS-PAGE could be likely due to the very low protein concentration in the expression medium (Fig 5.2). Therefore, to achieve higher protein yields, one would require larger volumes of cell culture. Some studies have used western blotting to track the presence of Wnt proteins throughout the purification steps. Therefore, instead of purifying the secreted Wnt5A from the expression media, we decided to use the conditioned medium that contains Wnt5A for our subsequent studies[209,210].

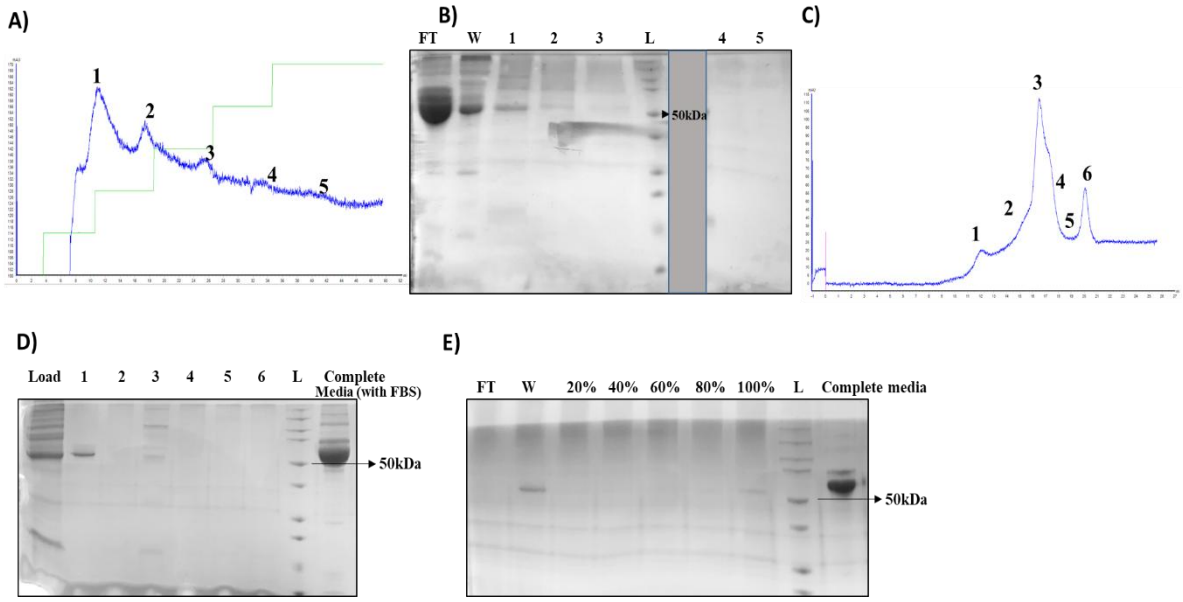


Figure 5.2: Purification of Wnt5A from conditioned medium (CM). A) *Blue sepharose affinity purification chromatogram. Green line shows gradient elution* ***B)*** *SDS-PAGE for Blue sepharose affinity chromatography* ***C)*** *Size exclusion chromatogram for Wnt5A purification* ***D)*** *SDS-PAGE corresponding to size exclusion chromatography* ***E)*** *SDS-PAGE for Heparin cation-exchange chromatography.*

5.3.2 Wnt5A conditioned medium promotes cell migration in MDA-MB-231 cells

To verify the presence of active Wnt5A in the CM, we performed wound healing assay using MDA-MB-231 cell lines. A scratch was made in the cell monolayer using a pipette tip. When the cells were treated with Wnt5A CM, showed a markedly faster rate of wound closure compared to those treated with the control CM, which is devoid of Wnt5A (Fig 5.3). Thus, our observation indicates that Wnt5A present in the CM is functionally active, and it promotes enhanced migratory behaviour in the MDA-MB-231 cells.

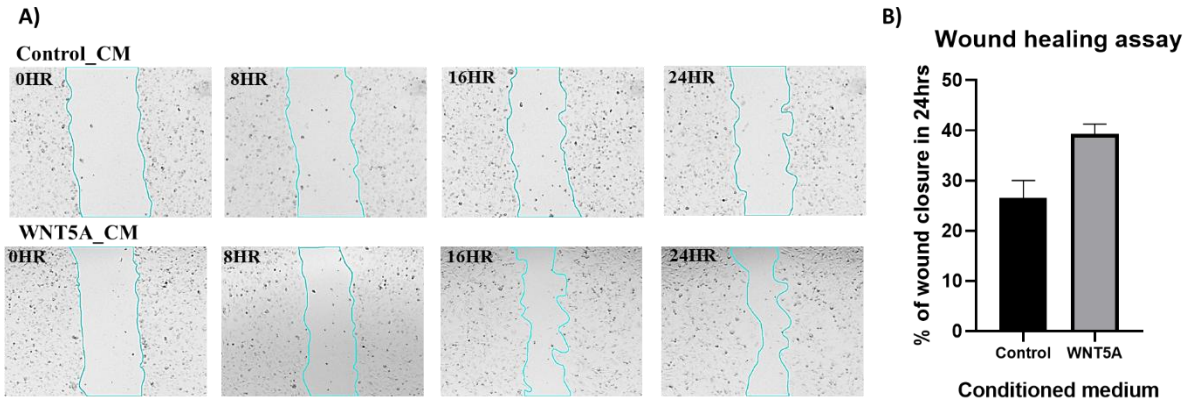


Figure 5.3: Wound healing assay of MDA-MB-231 cells in the presence and absence of Wnt5A CM. A) *The upper panel shows images taken at 0, 8, 16 and 24 h, for cells treated with control CM and the lower panel shows images taken at 0, 8, 16 and 24 h, for cells treated with Wnt5A CM* ***B)*** *Bar plot showing the percent of wound closure.*

5.3.3 Wnt5A condition medium induces stress fiber formation in HEK293 cells

After confirming the presence of active Wnt5A in the conditioned medium (CM), we probed the role of the Wnt5A-Frizzled7 pathway on the Dvl2-WGEF interaction by monitoring its effect on the cytoskeletal arrangement in the transfected HEK293 cells. The expression of WGEF in the model cell line system was visualized via the YFP tag (pseudo-colored red), while RhoA expression was monitored through GFP fluorescence. For imaging, the actin cytoskeleton was stained with CF® 594 phalloidin stain (Biotium) and the nucleus was stained with DAPI. Fluorescent microscopy revealed that the transfected cells (marked with red arrows) (Fig 5.4) when treated with Wnt5A CM, exhibited increased stress fiber formation, a hallmark of RhoA activation. RhoA effectors, such as *Rho-associated coiled-coil containing kinase* (ROCK) and mDia[211,212], are shown to be involved in the assembly of actin filaments to form stress fibers. On the contrary, the Wnt5A untreated cells showed basal or no stress fiber formation. These findings suggest that, indeed, Wnt5A induces RhoA-dependent stress fiber

formation by acting through the Dvl2/WGEF/RhoA axis. Furthermore, stress fibers formed by these engineered cells could be used as a readout to assess the interactions between various proteins operating under this particular pathway.

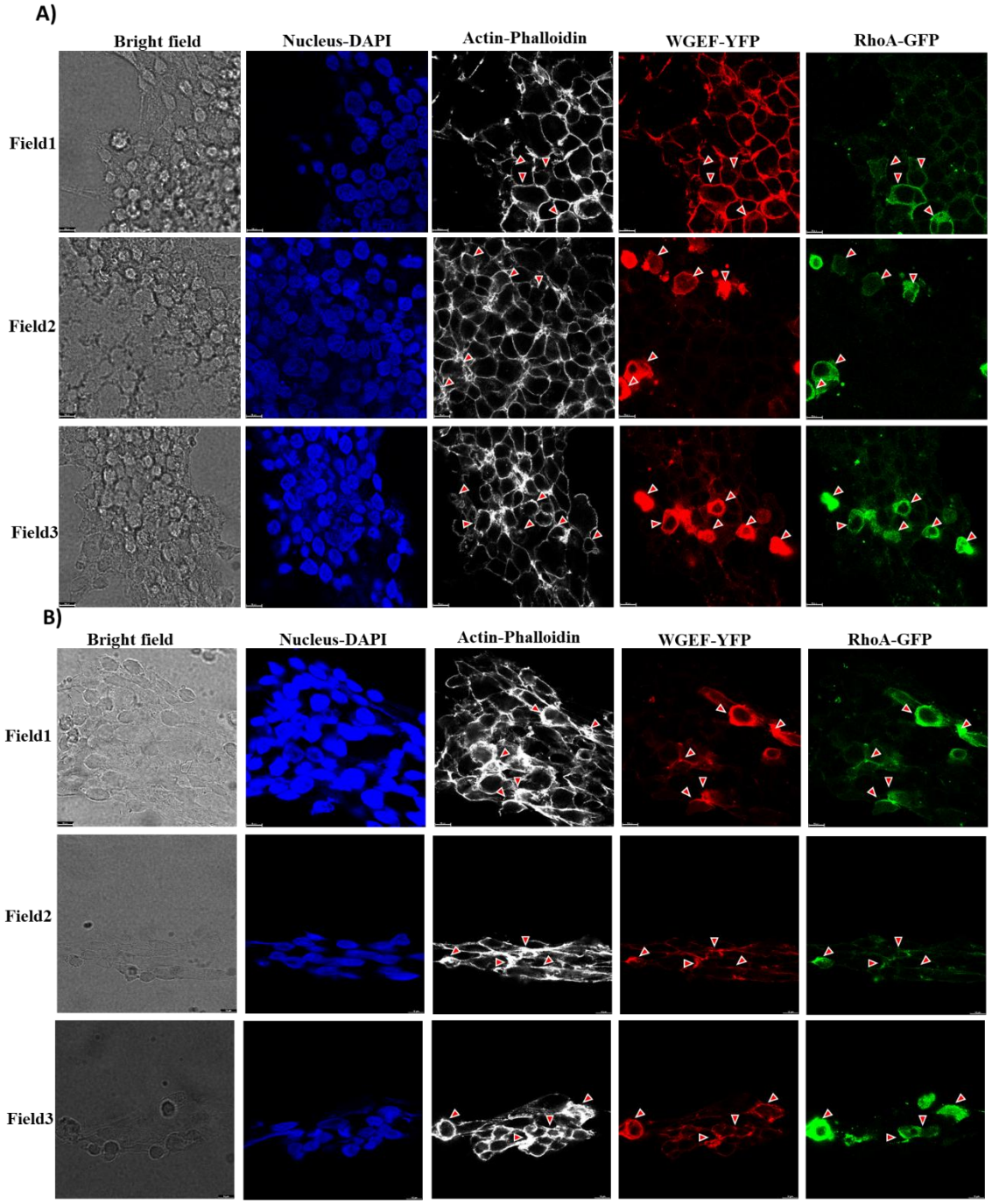


Figure 5.4: Stress fiber formation assay in HEK293 cells in the presence and absence of Wnt5A conditioned medium. ***A)*** *Transfected HEK293 cells treated with control conditioned medium (CM)* ***B)*** *Transfected HEK293 cells treated with Wnt5A CM. The nucleus is stained with DAPI and actin filaments are stained with phalloidin. WGEF is tagged with YFP, shown as pseudo colour red and RhoA is tagged with GFP. The red arrows mark cells showing the presence of both GFP and YFP fluorescence.*

5.3.4 Attempts to generate MDA-MB-231 cell lines with conditional expression of Wnt-PCP proteins

Although our stress fiber formation experiment showed that Wnt5A activates RhoA through the pathway involving Dvl2 and WGEF, we could not capture the specific interactions between Dvl2-WGEF using fluorescent imaging. Furthermore, the aforementioned experiment provided a qualitative nature of RhoA activation. However, to make the readout more quantitative in nature, we need to have cell lines that conditionally express all the proteins of interest. To achieve this, instead of HEK293 cell lines, we transfected construct 1 into the MDA-MB-231 breast cancer cell lines, as Wnt5A mediated activation RhoA, through Dvl2, is shown to promote cell migration in these cells[58]. To generate conditionally expressing cell lines, we used two popular selection methods: 1. Fluorescence-Activated Cell Sorting (FACS) and 2. Antibiotic based selection[213]. FACS was performed 48 h post-transfection, which sorted 3088 cells based on the YFP fluorescent intensity. These cells were diluted and seeded into a 24-well plate with 100 cells per well (Fig 5.5). Cells were then treated with zeocin (0-500 µg/mL) after 24 h of seeding to select for cells conditionally expressing proteins of interest.

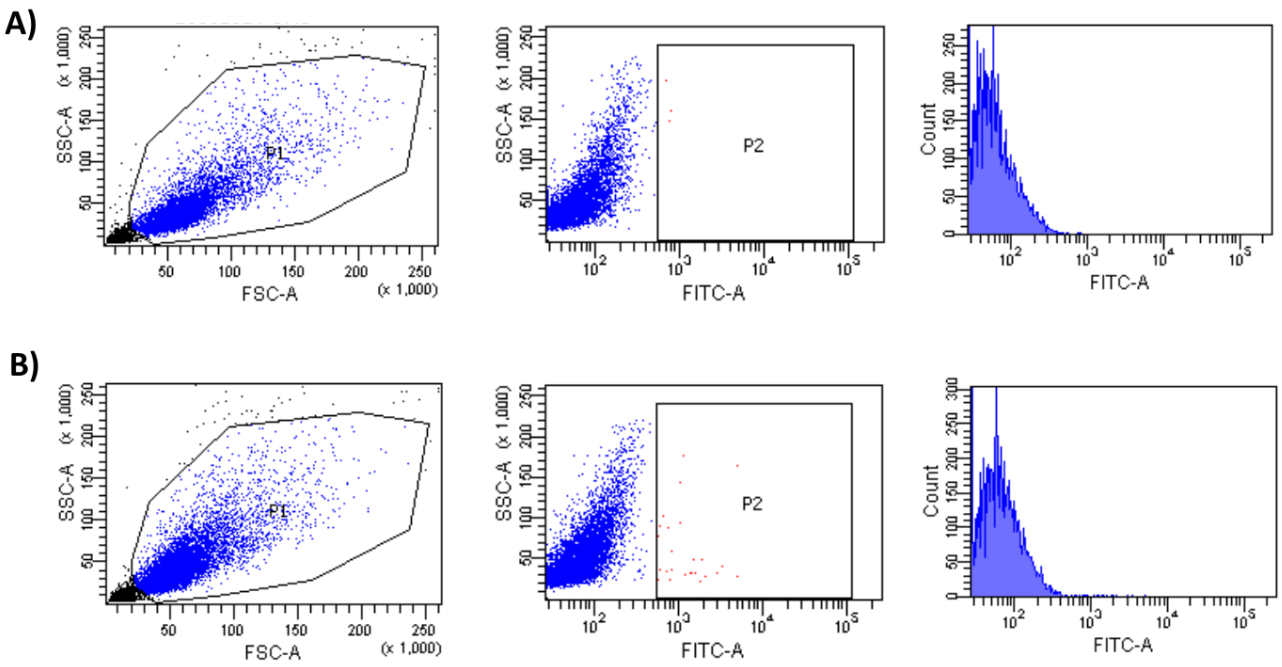


Figure 5.5: Fluorescence-Activated Cell Sorting (FACS) of transfected and non-transfected MDA-MB-231 cells. A) *FACS done for non-transfected cells (control)* ***B)*** *FACS done for transfected cells.*

In parallel, cells were treated with zeocin without subjecting them to FACS. The idea here was to avoid the mechanical stress on the cells introduced by the FACS. During this process, transfected cells were treated with a minimal zeocin concentration (250 µg/mL), the lowest concentration of zeocin required to kill non-transfected cells in 15 days. We observed that the antibiotic-treated transfected cells maintained normal morphology for 20 days, compared to the non-transfected cells, which deleted in less than 14 days (Fig 5.6). However, transfected cells treated with the antibiotic could not be passaged. This could be perhaps due to the cytotoxicity induced by the proteins that were being overexpressed on account of transfection. Perhaps transfection at higher cell densities might produce a greater yield of transfected cells, which could be propagated for further studies.

On the other hand, cells sorted from FACS and treated with the antibiotic did not survive. This could be due to the stress imparted on the cells as a result of continuous FACS sorting and antibiotic selection. Better results could be obtained by introducing sequential FACS, as described by Kaufman *et al*.[214].

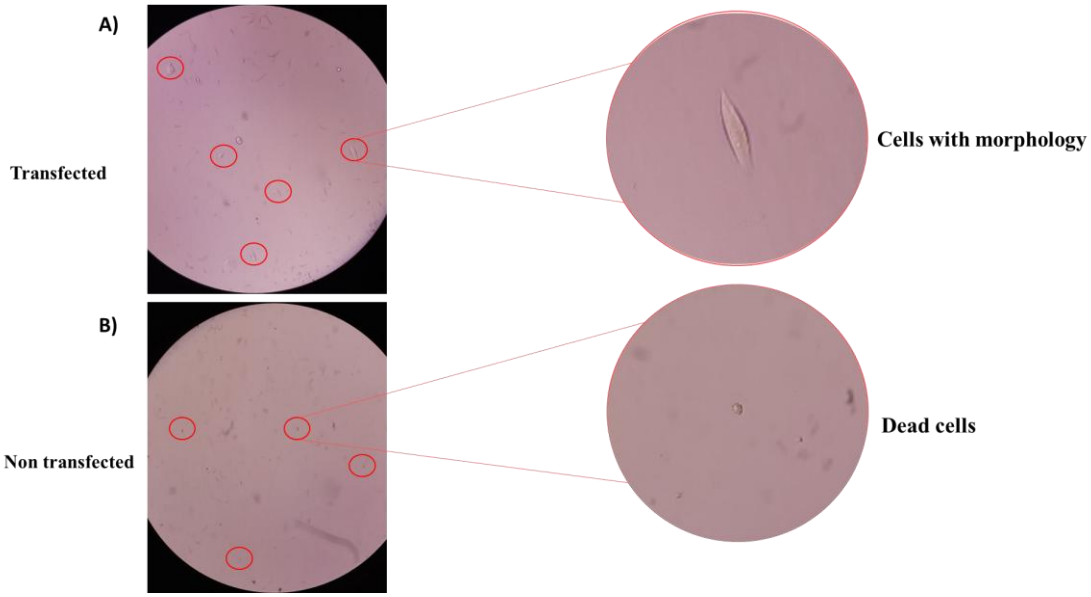


Figure 5.6: Images showing zeocin based antibiotic selection in transfected and non-transfected MDA-MB-231 cells after 14 days of treatment. A) *Transfected cells* ***B)*** *Non-transfected cells.*

5.4 Discussion

In Chapter 4, we discovered the peptide of WGEF that interacts with Dvl2, to activate the GEF and, hence, its downstream effector RhoA. This study provided insights into WGEF-Dvl2 interaction when the proteins are in isolation from the cellular environment. Physiologically a more realistic situation would be to evaluate this particular interaction when these two proteins are operating under the non-canonical Wnt-PCP pathway. Here, we established a cellular system that mimics the physiological situation and provides an opportunity to probe *in vivo* WGEF-Dvl2 interaction. The challenges associated with the purification of hydrophobic and post-translationally modified Wnt5A protein made us to adopt alternative strategies, such as utilizing conditioned media (CM) containing Wnt5A, which has facilitated functional analysis and downstream experimental designs. From cellular imaging experiments, we showed that, indeed, Wnt5A activates FZD7, which further triggers the downstream GTPase, RhoA, and leads to the formation of stress fibers. The outcome of this experiment qualitatively proved the activation of the WGEF-RhoA pathway. However, quantitative measurement of the strength of this interaction, operating under this specific pathway, requires conditionally expressed Wnt pathway proteins. This particular requirement is also due to the complex activation mechanism of WGEF. For instance, both the N-terminal and C-terminal domains of WGEF interact with the catalytic domain or the domains present in the vicinity of the catalytic domain, rendering the protein to be inactive. The Dvl2 interacting motif was found to be lying at the N-terminal of the protein. This may indicate that Dvl2-WGEF interaction, mediated by the peptide motif, might take care of the autoinhibitory aspects introduced by the N-terminal domain. However, the release of WGEF autoinhibition due to the interaction of the SH3 domain is not clear. Hence, a holistic approach that involves all the proteins of the pathway in play would shed light on the strength of Dvl2-WGEF interaction and also its role in the activation of the GEF. Thus, some of these aforementioned criteria demand cell lines that conditionally express the non-canonical Wnt-FZD7-Dvl2-WGEF-RhoA pathway proteins. In our pursuit to generate such a cell line, we employed FACS, antibiotic selection and a combination of these two methods. We observed that, although FACS could sort the cells expressing proteins of interest, imposing antibiotic stress on these did not produce the desired effect. On the other hand, simple antibiotic based selection also met with limited success. These outcomes are may be due to the limited transfection efficiency and/or toxicity of the recombinantly expressed proteins. Perhaps these

problems could be circumvented by using other genome editing techniques, such as CRISPR-Cas.

Thus, our study, for the first time, shows the activation of RhoA operating under the Wnt5A/Dvl2/WGEF/RhoA pathway in a holistic manner. Further fine-tuning of cellular engineering methods shall provide deeper insights into this pathway. This warrants further investigations, which is beyond the scope of this work.

Conclusion and Outlook

Wnt-PCP signalling is essential for cellular movement processes, including migration of neural crest cells and gastrulation during embryonic development. In higher eukaryotes, it regulates the orientation of hair follicles and is essential for proper neural tube closure. Aberrations in the Wnt-PCP signalling pathway may result in various pathological conditions, which includes cancer and disorders of the kidney and heart, where convergent extension movements are affected.

Cytoskeletal remodelling driven by the activation of RhoA GTPases is a critical step in the Wnt-PCP signalling. For instance, activated FZD7 receptor in this pathway recruits the adapter protein Dishevelled2 (Dvl2) and Daam1. Dvl2 further interacts and activates WGEF by relieving it from its autoinhibited state. WGEF then triggers RhoA, leading to the cytoskeletal rearrangement and subsequent cell migration. The signalling mechanisms controlled by the canonical Wnt signalling are studied in much more detail compared to the non-canonical Wnt signalling. Thus, the molecular mechanism underlying Wnt-Frizzled mediated signal propagation, which leads to the activation of RhoA in the non-canonical Wnt-PCP pathway, remains poorly understood. Given the dysregulation of proteins in the Wnt-PCP pathway across various cancers, this pathway is attracting increasing attention as a target for cancer therapeutics. This thesis, aims to understand the molecular mechanism underlying the signal relaying from the Frizzled 7 receptor, leading to the activation of RhoA.

Like other GTPases, RhoA relies on GEFs for its activation, which facilitates the exchange of GDP for GTP in the GTPases, transitioning them from an inactive to an active state. In the Wnt-PCP pathway, WGEF activates RhoA to bring about convergent extension movements. However, the molecular mechanism of WGEF-RhoA interaction during cell migration is not known. Therefore, to determine the structural basis of RhoA activation by WGEF, we designed, cloned, expressed and purified different constructs of xWGEF to form a stable complex with RhoA. Our experiments revealed that xWGEF constructs were highly unstable when expressed and purified in the bacterial system (*E. coli*). This could be attributed to the disordered regions present in the protein. However, these constructs were found to be stable when co-expressed and purified with RhoA. Therefore, we formed the *in vitro* xWGEF-RhoA complex by co-expressing them in BL21* strain of *E. coli*. However, extensive

crystallization attempts did not yield any crystals, suggesting poor stability of the complex and/or feeble interaction between xWGEF and RhoA. In order to address this issue we tinkered with the purification steps and also reduced the purification time. Although these strategies yielded one crystal hit, reproducing the crystal proved challenging. From all our *in vitro* xWGEF-RhoA complexation attempts, we can infer that trimming the protein length to eliminate the disordered regions could help stabilize the complex for crystallization. Further optimization of WGEF constructs and expression systems is beyond the scope of this thesis. However, from our observations we can suggest that variants of WGEF, consisting of the DH domain of WGEF, such as the hWGEF$^{330\text{-}581}$, could be used for crystallization trials.

In the absence of structural information on the WGEF-RhoA complex, we resorted to biochemical and computational methods to elucidate the mechanisms of GEF activation by Dvl2^{PDZ}. Due to the presence of interdomain interactions, WGEF intrinsically exists in an autoinhibited state. Previous studies have shown that Dvl2^{PDZ} domain releases WGEF from its autoinhibitory state, but the exact region of WGEF, which interacts with Dvl2^{PDZ}, was not known. With an aim to map this region and also to study the effect of WGEF-Dvl2^{PDZ} interaction on the activation of GEF, we systematically examined the sequences of the WGEF subfamily by screening them for putative PDZ binding motifs. This exercise resulted in the identification of an evolutionarily conserved motif present in WGEF that shares a significant sequence homology with the known internal peptide motifs, which are shown to interact with Dvl2^{PDZ}. GEF assay in the presence and absence of Dvl2^{PDZ} revealed that WGEF-Dvl2^{PDZ} interaction partially releases the GEF autoinhibition. However, the interaction was feeble as the amount of Dvl2^{PDZ} required to significantly activate WGEF was found to be very high (50 μM). Presence of WGEF^{pep} peptide and its variant WGEF$^{pep\textit{D9A}}$ in the GEF assay mixture containing Dvl2^{PDZ} revealed that both WGEF and WGEF^{pep} competitively bind to Dvl2^{PDZ} and the conserved aspartate residue at P_0 position of the motif plays a significant role in stabilizing this interaction. Furthermore, the requirement of higher WGEF^{pep} quantity to observe the competitive inhibition suggests that the binding affinity of WGEF for Dvl2^{PDZ} is quite feeble.

Further, we made attempts to elucidate the structural basis of WGEF-Dvl2^{PDZ} interaction, by crystallizing Dvl2^{PDZ} in presence of hWGEF330 and WGEF^{pep}. Although these efforts produced crystals, electron density for the ligand (hWGEF330 or WGEF^{pep}) was not observed in the crystal structure. We also tried to get the structure of the WGEF^{pep}-Dvl2^{PDZ}

complex by forming their chimera, wherein WGEF^{pep} was fused at C-terminal of Dvl2^{PDZ} domain. Even in this instance, the density for the WGEF^{pep} was not observed, as due to the crystal packing, the peptide segment was seen to be lying in the solvent region of the crystal. In short, our extensive attempts to elucidate the structural basis of WGEF-Dvl2^{PDZ} complex met with limited success and provided structures of Dvl2^{PDZ} in its *apo* form. Therefore, we designed different peptide variants of WGEF^{pep} and through binding studies, we determined the mode of interaction between WGEF and Dvl2^{PDZ}. This exercise showed that the residues at P_{-3}, P_{-2}, P_0 and P_2 positions are crucial for Dvl2^{PDZ}-WGEF interaction. This observation was further substantiated by employing WGEF mutants, carrying substitutions in their Dvl2^{PDZ} binding motif. Notably, some of the mutants were observed to be relatively less stable compared to the wild-type protein, indicating the role of this particular motif in protein stabilization.

We augmented the information obtained from biochemical studies with the MD simulation studies to determine the structural basis of WGEF-Dvl2^{PDZ} interaction. MD simulation studies showed that the internal peptide ligands do not form a consistent secondary structure but rather undergo conformational rearrangement to interact with the Dvl2^{PDZ} domain. This observation was in contrast with known Dvl2^{PDZ}-internal peptide motif interaction. We employed the Dynamical coupling analysis to delineate the cluster of residues from the Dvl2^{PDZ} domain that recognize the peptide motif and stabilize its interaction with the bound peptide. The ligand-dependent clustering of dynamically coupled residues shows that the Dvl2^{PDZ} domain exhibits intrinsic flexibility to accommodate a spectrum of ligands.

Together, these findings suggest that WGEF contains a unique “internal- PDZ binding motif” that interacts with Dvl2^{PDZ} and partially activates the WGEF from its autoinhibitory state. Through this information we also propose a putative mechanism of WGEF activation by Dvl2^{PDZ}.

To understand WGEF-Dvl2^{PDZ} interaction in the context of the Wnt-PCP signalling, we had set out to establish a cell based assay. Recombinant Wnt5A, essential for the activation of the Wnt-PCP pathway, was produced from L-Wnt5A fibroblast-like cells. The formation of stress fibers and induced cell migration were used as readouts to assay RhoA activation operating under this pathway. For the former, transient expression of the Wnt-PCP components in HEK293 cells was achieved by transfecting the cells with the plasmids encoding genes for

FZD7, Dvl2, WGEF and RhoA. Expression of a few proteins was confirmed through fluorescence microscopy, utilizing the YFP and GFP tags attached to WGEF and RhoA, respectively. Stress fibers were observed when these cells were treated with the Wnt5A conditioned medium. Similarly, MDA-MB-231 cells were prepared for conditional expression of Wnt-PCP proteins by means of plasmid enabled antibiotic selection. Since the transfection efficiency for MDA-MB-231 cells was observed to be low, we sorted the successfully transfected cells using FACS and further subjected these cells to antibiotic selection. Although the outcome of this exercise was not very encouraging, the method developed here could essentially serve as a starting point for future optimization of these cell lines. Thus, we have successfully installed the Wnt-PCP pathway, particularly the branch that activates RhoA, in HEK293 cell lines and developed protocols for the stable expression of Wnt-PCP proteins in MDA-MB-231 cells. This work would serve as a stepping stone in further characterization of Dvl2-WGEF interaction and the consequent RhoA activation controlled by the Wnt-PCP pathway.

To sum up, this thesis reports an extensive amount of work carried out to investigate the structural and molecular aspects of RhoA activation under the Wnt-PCP pathway. The important step in this particular pathway involves activation of the RhoA effector WGEF by the Wnt adapter protein Dvl2. We show that an evolutionarily conserved peptide motif of WGEF facilitates its interaction with Dvl2, resulting in the activation of the GEF. Furthermore, we showed how the dynamical nature of the $Dvl2^{PDZ}$ domain recognizes diverse ligands, including WGEF that vary with respect to sequence and structure. We delineated the impact of WGEF-$Dvl2^{PDZ}$ interaction on the catalytic activity of the GEF and its consequent role in RhoA activation. Furthermore, we established cell based assays to further characterize Wnt-PCP mediated RhoA activation. The current work has provided invaluable insights into the lesser studied Wnt-PCP pathway and its signalling cascade leading to the activation of RhoA. The findings of this study could have a considerable impact on designing therapeutic agents targeting the Wnt-PCP pathway.

The work reported here can be extended by optimizing WGEF constructs and elucidate the structural basis of RhoA activation by this GEF. The more desired structure would be that of Dvl2-WGEF-RhoA ternary complex. As reported in the thesis, the mutual interaction between these proteins is highly transient, which poses serious challenges in forming a stable

in vitro complex for structural studies. However, the efforts reported here would serve as a beacon for optimizing constructs to produce stable complexes. The cell based assays reported here could be used to understand Wnt-PCP signalling events by introducing structure based mutations and characterize the crosstalk between all the protein players operating under this pathway.

The next milestone in the RhoA-Wnt-PCP pathway would be the elucidation of the FZD7-Dvl2-WGEF ternary complex structure. This structure shall help us to understand the assembling of signalosomes at the cytosolic end of the FZD receptor by Dvl. This structure would also provide insights into how the same set of Wnt-FZD pairs exert different modes of Wnt pathways. Given the extreme complexities of the Wnt signalling, rigorous efforts employing structural, biochemical, computational and cell based techniques shall unravel this most perplexing cellular signalling pathway.

References

1. Vinson, C. R., Conover, S. & Adler, P. N. A Drosophila tissue polarity locus encodes a protein containing seven potential transmembrane domains. *Nature* (1989). doi:10.1038/338263a0.

2. Schulte, G. & Bryja, V. The Frizzled family of unconventional G-protein-coupled receptors. *Trends in Pharmacological Sciences* (2007). doi.org/10.1016/j.tips.2007.09.001.

3. Komiya, Y. & Habas, R. Wnt signal transduction pathways. *Organogenesis* (2008). doi.org/10.4161/org.4.2.5851.

4. Eubelen, M. *et al.* A molecular mechanism for Wnt ligand-specific signaling. *Science (80-.).* (2018). doi:10.1126/science.aat1178

5. Verkaar, F. & Zaman, G. J. R. A model for signaling specificity of Wnt/Frizzled combinations through co-receptor recruitment. *FEBS Lett.* (2010). doi:10.1016/j.febslet.2010.08.030.

6. Schwarz-Romond, T., Metcalfe, C. & Bienz, M. Dynamic recruitment of axin by Dishevelled protein assemblies. *J. Cell Sci.* (2007). doi:10.1242/jcs.002956.

7. Kikuchi, A., Yamamoto, H., Sato, A. & Matsumoto, S. New Insights into the Mechanism of Wnt Signaling Pathway Activation. in *International Review of Cell and Molecular Biology* (2011). doi:10.1016/B978-0-12-386035-4.00002-1.

8. ten Berge, D. *et al.* Wnt Signaling Mediates Self-Organization and Axis Formation in Embryoid Bodies. *Cell Stem Cell* (2008). doi:10.1016/j.stem.2008.09.013.

9. Fevr, T., Robine, S., Louvard, D. & Huelsken, J. Wnt/β-Catenin Is Essential for Intestinal Homeostasis and Maintenance of Intestinal Stem Cells. *Mol. Cell. Biol.* (2007). doi:10.1128/mcb.01034-07.

10. van Es, J. H. *et al.* A Critical Role for the Wnt Effector Tcf4 in Adult Intestinal Homeostatic Self-Renewal. *Mol. Cell. Biol.* (2012). doi:10.1128/mcb.06288-11.

11. Choi, Y. S. *et al.* Distinct functions for Wnt/β-Catenin in hair follicle stem cell proliferation and survival and interfollicular epidermal homeostasis. *Cell Stem Cell* (2013). doi:10.1016/j.stem.2013.10.003.

12. Lim, X. *et al.* Interfollicular epidermal stem cells self-renew via autocrine Wnt signaling. *Science (80-.).* (2013). doi:10.1126/science.1239730.

13. Rulifson, I. C. *et al.* Wnt signaling regulates pancreatic β cell proliferation. *Proc. Natl. Acad. Sci. U. S. A.* (2007). doi:10.1073/pnas.0701509104.

14. Luis, T. C. *et al.* Canonical wnt signaling regulates hematopoiesis in a dosage-dependent fashion. *Cell Stem Cell* (2011). doi:10.1016/j.stem.2011.07.017.

15. Qin, K. *et al.* Canonical and noncanonical Wnt signaling: Multilayered mediators, signaling mechanisms and major signaling crosstalk. *Genes and Diseases* (2024). doi.org/10.1016/j.gendis.2023.01.030.

16. Kühl, M., Sheldahl, L. C., Malbon, C. C. & Moon, R. T. Ca2+/calmodulin-dependent protein kinase II is stimulated by Wnt and Frizzled homologs and promotes ventral cell fates in Xenopus. *J. Biol. Chem.* (2000). doi:10.1074/jbc.275.17.12701.

17. Sheldahl, L. C. *et al.* Dishevelled activates Ca2+ flux, PKC, and CamKII in vertebrate embryos. *J. Cell Biol.* (2003). doi:10.1083/jcb.200211094.

18. Saneyoshi, T., Kume, S., Amsaki, Y. & Mikoshiba, K. The wnt/calcium pathway activates nf-at and promotes ventral cell fate in xenopus embryos. *Nature* (2002). doi:10.1038/417295a.

19. Oruganti, S. R., Edin, S., Grundström, C. & Grundström, T. CaMKII targets Bcl10 in T-cell receptor induced activation of NF-κB. *Mol. Immunol.* (2011). doi:10.1016/j.molimm.2011.03.020.

20. Yan, X. *et al.* CaMKII-Mediated CREB phosphorylation is involved in Ca^{2+}-Induced BDNF mRNA transcription and neurite outgrowth promoted by electrical stimulation. *PLoS One* (2016). doi:10.1371/journal.pone.0162784.

21. Penzo-Mendèz, A., Umbhauer, M., Djiane, A., Boucaut, J. C. & Riou, J. F. Activation of Gβγ signaling downstream of Wnt-11/Xfz7 regulates Cdc42 activity during Xenopus gastrulation. *Dev. Biol.* (2003). doi:10.1016/S0012-1606(03)00067-8.

22. Winklbauer, R., Medina, A., Swain, R. K. & Steinbeisser, H. Frizzled-7 signalling controls tissue separation during Xenopus gastrulation. *Nature* (2001). doi:10.1038/35101621.

23. Habas, R., Kato, Y. & He, X. Wnt/Frizzled activation of Rho regulates vertebrate gastrulation and requires a novel formin homology protein Daam1. *Cell* (2001). doi:10.1016/S0092-8674(01)00614-6.

24. Tanegashima, K., Zhao, H. & Dawid, I. B. WGEF activates Rho in the Wnt-PCP pathway and controls convergent extension in Xenopus gastrulation. *EMBO J.* (2008). doi:10.1038/emboj.2008.9.

25. Habas, R., Dawid, I. B. & He, X. Coactivation of Rac and Rho by Wnt/Frizzled signaling is required for vertebrate gastrulation. *Genes Dev.* (2003). doi:10.1101/gad.1022203.

26. Shi, D. L. Wnt/planar cell polarity signaling controls morphogenetic movements of gastrulation and neural tube closure. *Cellular and Molecular Life Sciences* (2022). doi.org/10.1007/s00018-022-04620-8.

27. Guo, N., Hawkins, C. & Nathans, J. Frizzled6 controls hair patterning in mice. *Proc. Natl. Acad. Sci. U. S. A.* (2004). doi:10.1073/pnas.0402802101.

28. Henderson, D. J. *et al.* Cardiovascular defects associated with abnormalities in midline development in the Loop-tail mouse mutant. *Circ. Res.* (2001). doi:10.1161/hh1301.092497.

29. Wang, J. *et al.* Regulation of polarized extension and planar cell polarity in the cochlea by the vertebrate PCP pathway. *Nat. Genet.* (2005). doi:10.1038/ng1622.

30. Fischer, E. *et al.* Defective planar cell polarity in polycystic kidney disease. *Nat. Genet.* (2006). doi:10.1038/ng1701.

31. Klingensmith, J. & Nusse, R. Signaling by wingless in drosophila. *Developmental Biology* (1994). doi.org/10.1006/dbio.1994.1325.

32. Gonsalves, F. C. & DasGupta, R. Function of the Wingless Signaling Pathway in Drosophila. *Methods in Molecular Biology* (2008). doi.org/10.1007/978-1-60327-469-2_10.

33. Willert, K. & Nusse, R. Wnt proteins. *Cold Spring Harb. Perspect. Biol.* (2012). doi:10.1101/cshperspect.a007864.

34. Chae, W. J. & Bothwell, A. L. M. Canonical and Non-Canonical Wnt Signaling in Immune Cells.

Trends in Immunology (2018). doi.org/10.1016/j.it.2018.08.006.

35. Ackers, I. & Malgor, R. Interrelationship of canonical and non-canonical Wnt signalling pathways in chronic metabolic diseases. *Diabetes and Vascular Disease Research* (2018). doi.org/10.1177/1479164117738442.

36. Schulte, G. & Wright, S. C. Frizzleds as GPCRs – More Conventional Than We Thought! *Trends in Pharmacological Sciences* (2018). doi.org/10.1016/j.tips.2018.07.001.

37. Davies, M. N. *et al.* On the hierarchical classification of G protein-coupled receptors. *Bioinformatics* (2007). doi:10.1093/bioinformatics/btm506.

38. Huang, H. C. & Klein, P. S. The frizzled family: Receptor for multiple signal transduction pathways. *Genome Biology* (2004). doi.org/10.1186/gb-2004-5-7-234.

39. Agostino, M. & Pohl, S. Ö. G. Wnt binding affinity prediction for putative frizzled-type cysteine-rich domains. *Int. J. Mol. Sci.* (2019). doi:10.3390/ijms20174168.

40. Ko, S. B. *et al.* Functional role of the Frizzled linker domain in the Wnt signaling pathway. *Commun. Biol.* (2022). doi:10.1038/s42003-022-03370-4.

41. He, X., Semenov, M., Tamai, K. & Zeng, X. LDL receptor-related proteins 5 and 6 in Wnt/β-catenin signaling: Arrows point the way. *Development* (2004). doi.org/10.1242/dev.01117.

42. Nusse, R. & Clevers, H. Wnt/β-Catenin Signaling, Disease, and Emerging Therapeutic Modalities. *Cell* (2017). doi.org/10.1016/j.cell.2017.05.016.

43. Lu, W., Yamamoto, V., Ortega, B. & Baltimore, D. Mammalian Ryk is a Wnt coreceptor required for stimulation of neurite outgrowth. *Cell* (2004). doi:10.1016/j.cell.2004.09.019.

44. Masiakowski, P. & Carroll, R. D. A novel family of cell surface receptors with tyrosine kinase-like domain. *J. Biol. Chem.* (1992). doi:10.1016/s0021-9258(18)35733-8.

45. Saldanha, J., Singh, J. & Mahadevan, D. Identification of a Frizzled-like cysteine rich domain in the extracellular region of developmental receptor tyrosine kinases. *Protein Sci.* (1998). doi:10.1002/pro.5560070718.

46. Petrova, I. M., Malessy, M. J., Verhaagen, J., Fradkin, L. G. & Noordermeer, J. N. Wnt Signaling through the Ror Receptor in the Nervous System. *Molecular Neurobiology* (2014). doi.org/10.1007/s12035-013-8520-9.

47. Sharma, M., Castro-Piedras, I., Simmons, G. E. & Pruitt, K. Dishevelled: A masterful conductor of complex Wnt signals. *Cellular Signalling* (2018). doi.org/10.1016/j.cellsig.2018.03.004.

48. Schwarz-Romond, T. *et al.* The DIX domain of Dishevelled confers Wnt signaling by dynamic polymerization. *Nat. Struct. Mol. Biol.* (2007). doi:10.1038/nsmb1247.

49. Kishida, S. *et al.* DIX Domains of Dvl and Axin Are Necessary for Protein Interactions and Their Ability To Regulate β-Catenin Stability. *Mol. Cell. Biol.* (1999). doi:10.1128/mcb.19.6.4414.

50. Mieszczanek, J. *et al.* Selective function of the PDZ domain of Dishevelled in noncanonical Wnt signalling. *J. Cell Sci.* (2022). doi:10.1242/jcs.259547.

51. Lee, H. J. & Zheng, J. J. PDZ domains and their binding partners: Structure, specificity, and modification. *Cell Communication and Signaling* vol. 8 (2010). doi.org/10.1186/1478-811X-8-8.

52. Wong, H. C. *et al.* Direct binding of the PDZ domain of Dishevelled to a conserved internal sequence in the C-terminal region of Frizzled. *Mol. Cell* (2003). doi:10.1016/S1097-2765(03)00427-1.

53. Shi, Q. & Chen, Y. G. Regulation of Dishevelled protein activity and stability by post-translational modifications and autophagy. *Trends in Biochemical Sciences* (2021). doi.org/10.1016/j.tibs.2021.07.008.

54. Gammons, M. V., Rutherford, T. J., Steinhart, Z., Angers, S. & Bienz, M. Essential role of the Dishevelled DEP domain in a Wnt-dependent human-cell-based complementation assay. *J. Cell Sci.* (2016). doi:10.1242/jcs.195685.

55. Gammons, M. V., Renko, M., Johnson, C. M., Rutherford, T. J. & Bienz, M. Wnt Signalosome Assembly by DEP Domain Swapping of Dishevelled. *Mol. Cell* (2016). doi:10.1016/j.molcel.2016.08.026.

56. Gao, C. & Chen, Y. G. Dishevelled: The hub of Wnt signaling. *Cellular Signalling* (2010). doi.org/10.1016/j.cellsig.2009.11.021.

57. Liu, W. *et al.* Mechanism of activation of the Formin protein Daam1. *Proc. Natl. Acad. Sci. U. S. A.* (2008). doi:10.1073/pnas.0707277105.

58. Zhu, Y. *et al.* Dvl2-dependent activation of Daam1 and RhoA regulates Wnt5a-induced breast cancer cell migration. *PLoS One* (2012). doi:10.1371/journal.pone.0037823.

59. Liu, G. *et al.* Daam1 activates RhoA to regulate Wnt5a-induced glioblastoma cell invasion. *Oncol. Rep.* (2018). doi:10.3892/or.2017.6124.

60. Mosaddeghzadeh, N. & Ahmadian, M. R. The rho family gtpases: Mechanisms of regulation and signaling. *Cells* (2021). doi.org/10.3390/cells10071831.

61. Vetter, I. R. & Wittinghofer, A. The guanine nucleotide-binding switch in three dimensions. *Science* (2001). doi.org/10.1126/science.1062023.

62. Toma-Fukai, S. & Shimizu, T. Structural insights into the regulation mechanism of small GTPases by GEFs. *Molecules* (2019). doi.org/10.3390/molecules24183308.

63. Tahinci, E. & Symes, K. Distinct functions of Rho and Rac are required for convergent extension during Xenopus gastrulation. *Dev. Biol.* (2003). doi:10.1016/S0012-1606(03)00206-9.

64. Choi, S. C. & Han, J. K. Xenopus Cdc42 regulates convergent extension movements during gastrulation through Wnt/Ca2+ signaling pathway. *Dev. Biol.* (2002). doi:10.1006/dbio.2002.0602.

65. Shimizu, H., Toma-Fukai, S., Kontani, K., Katada, T. & Shimizu, T. GEF mechanism revealed by the structure of SmgGDS-558 and farnesylated RhoA complex and its implication for a chaperone mechanism. *Proc. Natl. Acad. Sci. U. S. A.* (2018). doi:10.1073/pnas.1804740115.

66. Bos, J. L., Rehmann, H. & Wittinghofer, A. GEFs and GAPs: Critical Elements in the Control of Small G Proteins. *Cell* (2007). doi.org/10.1016/j.cell.2007.05.018.

67. Kristelly, R., Gao, G. & Tesmer, J. J. G. Structural determinants of RhoA binding and nucleotide exchange in leukemia-associated Rho guanine-nucleotide exchange factor. *J. Biol. Chem.* (2004). doi:10.1074/jbc.M406056200.

68. Rittinger, K., Walker, P. A., Eccleston, J. F., Smerdon, S. J. & Gamblin, S. J. Structure at 1.65 Å of RhoA and its GTPase-activating protein in complex with a transition-state analogue. *Nature* (1997). doi:10.1038/39651.

69. Mosaddeghzadeh, N. *et al.* Electrostatic forces mediate the specificity of RHO GTPase-GDI interactions. *Int. J. Mol. Sci.* (2021). doi:10.3390/ijms222212493.

70. Zheng, Y. Dbl family guanine nucleotide exchange factors. *Trends in Biochemical Sciences* (2001). doi.org/10.1016/S0968-0004(01)01973-9.

71. Jaiswal, M., Dvorsky, R. & Ahmadian, M. R. Deciphering the Molecular and Functional Basis of Dbl Family Proteins. *J. Biol. Chem.* (2013). doi:10.1074/jbc.m112.429746.

72. Boland, A., Côté, J. F. & Barford, D. Structural biology of DOCK-family guanine nucleotide exchange factors. *FEBS Lett.* (2023). doi:10.1002/1873-3468.14523.

73. Premkumar, L. *et al.* Structural basis of membrane targeting by the Dock180 family of Rho family guanine exchange factors (Rho-GEFs). *J. Biol. Chem.* (2010). doi:10.1074/jbc.M110.102517.

74. Kulkarni, K., Yang, J., Zhang, Z. & Barford, D. Multiple factors confer specific Cdc42 and Rac protein activation by dedicator of cytokinesis (DOCK) nucleotide exchange factors. *J. Biol. Chem.* (2011). doi:10.1074/jbc.M111.236455.

75. Rossman, K. L., Der, C. J. & Sondek, J. GEF means go: Turning on Rho GTPases with guanine nucleotide-exchange factors. *Nature Reviews Molecular Cell Biology* (2005). doi.org/10.1038/nrm1587.

76. Omble, A. & Kulkarni, K. GPCRs that Rhoar the Guanine nucleotide exchange factors. *Small GTPases* (2022). doi.org/10.1080/21541248.2021.1896963.

77. Chen, Y., Chen, Z., Tang, Y. & Xiao, Q. The involvement of noncanonical Wnt signaling in cancers. *Biomedicine and Pharmacotherapy* (2021). doi.org/10.1016/j.biopha.2020.110946.

78. Akoumianakis, I., Polkinghorne, M. & Antoniades, C. Non-canonical WNT signalling in cardiovascular disease: mechanisms and therapeutic implications. *Nature Reviews Cardiology* (2022). doi.org/10.1038/s41569-022-00718-5.

79. Bakker, E. R. M. *et al.* Wnt5a promotes human colon cancer cell migration and invasion but does not augment intestinal tumorigenesis in apc1638N mice. *Carcinogenesis* (2013). doi:10.1093/carcin/bgt215.

80. Kurayoshi, M. *et al.* Expression of Wnt-5a is correlated with aggressiveness of gastric cancer by stimulating cell migration and invasion. *Cancer Res.* (2006). doi:10.1158/0008-5472.CAN-06-2359.

81. Kumawat, K. & Gosens, R. WNT-5A: Signaling and functions in health and disease. *Cellular and Molecular Life Sciences* (2016). doi.org/10.1007/s00018-015-2076-y.

82. Phesse, T., Flanagan, D. & Vincan, E. Frizzled7: A promising achilles' heel for targeting the Wnt receptor complex to treat cancer. *Cancers* (2016). doi.org/10.3390/cancers8050050.

83. Larasati, Y., Boudou, C., Koval, A. & Katanaev, V. L. Unlocking the Wnt pathway: Therapeutic potential of selective targeting FZD7 in cancer. *Drug Discovery Today* (2022). doi.org/10.1016/j.drudis.2021.12.008.

84. Asad, M. *et al.* FZD7 drives in vitro aggressiveness in stem-A subtype of ovarian cancer via regulation of non-canonical wnt/PCP pathway. *Cell Death Dis.* (2014). doi:10.1038/cddis.2014.302.

85. VanderVorst, K., Hatakeyama, J., Berg, A., Lee, H. & Carraway, K. L. Cellular and molecular mechanisms underlying planar cell polarity pathway contributions to cancer malignancy. *Seminars in Cell and Developmental Biology* (2018). doi.org/10.1016/j.semcdb.2017.09.026.

86. Nishioka, M. *et al.* Possible involvement of Wnt11 in colorectal cancer progression. *Mol. Carcinog.* (2013). doi:10.1002/mc.21845.

87. Mei, J. *et al.* Overexpressed DAAM1 correlates with metastasis and predicts poor prognosis in breast cancer. *Pathol. Res. Pract.* (2020). doi:10.1016/j.prp.2019.152736.

88. Zallen, J. A. Planar Polarity and Tissue Morphogenesis. *Cell* (2007). doi.org/10.1016/j.cell.2007.05.050.

89. Yang, Y. & Mlodzik, M. Wnt-Frizzled/Planar Cell Polarity Signaling: Cellular Orientation by Facing the Wind (Wnt). *Annu. Rev. Cell Dev. Biol.* (2015). doi:10.1146/annurev-cellbio-100814-125315.

90. Sebbagh, M. & Borg, J. P. Insight into planar cell polarity. *Experimental Cell Research* (2014). doi.org/10.1016/j.yexcr.2014.09.005.

91. Katoh, M. WNT/PCP signaling pathway and human cancer (Review). *Oncology Reports* (2005). doi.org/10.3892/or.14.6.1583.

92. Cherfils, J. & Chardin, P. GEFs: Structural basis for their activation of small GTP-binding proteins. *Trends in Biochemical Sciences* (1999). doi.org/10.1016/S0968-0004(99)01429-2.

93. Roy, D., Sengupta, D. & Kulkarni, K. Substrate induced dynamical remodeling of the binding pocket generates GTPase specificity in DOCK family of guanine nucleotide exchange factors. *Biochem. Biophys. Res. Commun.* (2022). doi:10.1016/j.bbrc.2022.09.059.

94. Snyder, J. T. *et al.* Structural basis for the selective activation of rho gtpases by dbl exchange factors. *Nat. Struct. Biol.* (2002). doi:10.1038/nsb796.

95. Goller, S. P., Gorfer, M., Mach, R. L. & Kubicek, C. P. Gene Cloning Using PCR 2. *Appl. PCR Mycol.* (1998).

96. Kibbe, W. A. OligoCalc: An online oligonucleotide properties calculator. *Nucleic Acids Res.* **35**, (2007).

97. Korbie, D. J. & Mattick, J. S. Touchdown PCR for increased specificity and sensitivity in PCR amplification. *Nat. Protoc.* **3**, (2008).

98. Sambrook, J. & Russell, D. W. The Inoue Method for Preparation and Transformation of Competent E. Coli : "Ultra-Competent" Cells . *Cold Spring Harb. Protoc.* (2006). doi:10.1101/pdb.prot3944.

99. Vincze, T., Posfai, J. & Roberts, R. J. NEBcutter: A program to cleave DNA with restriction enzymes. *Nucleic Acids Res.* (2003). doi:10.1093/nar/gkg526.

100. Smith, D. B. & Johnson, K. S. Single-step purification of polypeptides expressed in Escherichia coli as fusions with glutathione S-transferase. *Gene* (1988). doi:10.1016/0378-1119(88)90005-4.

101. Bornhorst, J. A. & Falke, J. J. Purification of proteins using polyhistidine affinity tags. *Methods in Enzymology* (2000). doi.org/10.1016/s0076-6879(00)26058-8.

102. Schmidt, T. G. M. & Skerra, A. The Strep-tag system for one-step purification and high-affinity detection or capturing of proteins. *Nat. Protoc.* (2007). doi:10.1038/nprot.2007.209.

103. Andrews, P. The gel-filtration behaviour of proteins related to their molecular weights over a wide range. *Biochem. J.* (1965). doi:10.1042/bj0960595.

104. Sahin, E. & Roberts, C. J. Size-exclusion chromatography with multi-angle light scattering for

elucidating protein aggregation mechanisms. *Methods Mol. Biol.* (2012). doi:10.1007/978-1-61779-921-1_25.

105. Kanie, T. & Jackson, P. Guanine Nucleotide Exchange Assay Using Fluorescent MANT-GDP. *BIO-PROTOCOL* **8**, (2018). doi: 10.21769/BioProtoc.2795

106. Seidel, S. A. I. *et al.* Microscale thermophoresis quantifies biomolecular interactions under previously challenging conditions. *Methods* (2013). doi.org/10.1016/j.ymeth.2012.12.005.

107. Asmari, M., Ratih, R., Alhazmi, H. A. & El Deeb, S. Thermophoresis for characterizing biomolecular interaction. *Methods* (2018). doi.org/10.1016/j.ymeth.2018.02.003.

108. Jerabek-Willemsen, M. *et al.* MicroScale Thermophoresis: Interaction analysis and beyond. *J. Mol. Struct.* (2014). doi:10.1016/j.molstruc.2014.03.009.

109. Smyth, M. S. & Martin, J. H. J. x Ray crystallography. *Journal of Clinical Pathology - Molecular Pathology* (2000). doi.org/10.1136/mp.53.1.8.

110. McPherson, A. & Gavira, J. A. Introduction to protein crystallization. *Acta Crystallographica Section F:Structural Biology Communications* (2014). doi.org/10.1107/S2053230X13033141.

111. Bunaciu, A. A., Udriştioiu, E. gabriela & Aboul-Enein, H. Y. X-Ray Diffraction: Instrumentation and Applications. *Critical Reviews in Analytical Chemistry* (2015). doi.org/10.1080/10408347.2014.949616.

112. The reflection of X-rays by crystals. *Proc. R. Soc. London. Ser. A, Contain. Pap. a Math. Phys. Character* (1913). doi:10.1098/rspa.1913.0040.

113. Jaeger, J. Macromolecular Structure Determination by X-ray Crystallography. in *eLS* (2005). doi:10.1038/npg.els.0002723.

114. Skarzynski, T. Collecting data in the home laboratory: Evolution of X-ray sources, detectors and working practices. *Acta Crystallogr. Sect. D Biol. Crystallogr.* (2013). doi:10.1107/S0907444913013619.

115. Willmott, P. R. X-Ray Sources at Large-Scale Facilities. in *Springer Proceedings in Physics* (2021). doi:10.1007/978-3-030-64623-3_1.

116. Hendrickson, W. A. Determination of macromolecular structures from anomalous diffraction of synchrotron radiation. *Science* (1991). doi.org/10.1126/science.1925561.

117. Broennimann, C. *et al.* The PILATUS 1M detector. *J. Synchrotron Radiat.* (2006). doi:10.1107/S0909049505038665.

118. Garman, E. F. & Schneider, T. R. Macromolecular Cryocrystallography. *J. Appl. Crystallogr.* (1997). doi:10.1107/S0021889897002677.

119. Pflugrath, J. W. Practical macromolecular cryocrystallography. *Acta Crystallogr. Sect. F Struct. Biol. Commun.* (2015). doi:10.1107/S2053230X15008304.

120. Morawiec, A. Indexing of diffraction patterns for determination of crystal orientations. *Acta Crystallogr. Sect. A Found. Adv.* (2020). doi:10.1107/S2053273320012802.

121. Powell, H. R. X-ray data processing. *Bioscience Reports* (2017). doi.org/10.1042/BSR20170227.

122. Karplus, P. A. & Diederichs, K. Linking crystallographic model and data quality. *Science (80-.).*

(2012). doi:10.1126/science.1218231.

123. Kabsch, W. *et al. XDS*. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, (2010). doi: 10.1107/S0907444909047337

124. Evans, P. R. & Murshudov, G. N. How good are my data and what is the resolution? *Acta Crystallogr. Sect. D Biol. Crystallogr.* **69**, (2013). doi: 10.1107/S0907444913000061.

125. Drenth, J. & Mesters, J. *Principles of protein X-ray crystallography: Third edition*. *Principles of Protein X-Ray Crystallography: Third Edition* (2007). doi:10.1007/0-387-33746-6.

126. Taylor, G. The phase problem. in *Acta Crystallographica - Section D Biological Crystallography* (2003). doi:10.1107/S0907444903017815.

127. Schwarzenbacher, R., Godzik, A., Grzechnik, S. K. & Jaroszewski, L. The importance of alignment accuracy for molecular replacement. *Acta Crystallogr. Sect. D Biol. Crystallogr.* (2004). doi:10.1107/S0907444904010145.

128. Ilari, A. & Savino, C. Protein structure determination by X-ray crystallography. *Methods Mol. Biol.* (2008). doi:10.1007/978-1-60327-159-2_3.

129. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, (2007). doi: 10.1107/S0021889807021206.

130. Navaza, J. & Saludjian, P. AMoRe: An automated molecular replacement program package. *Methods Enzymol.* (1997). doi:10.1016/S0076-6879(97)76079-8.

131. Vagin, A. & Teplyakov, A. MOLREP: An Automated Program for Molecular Replacement. *J. Appl. Crystallogr.* (1997). doi:10.1107/S0021889897006766.

132. The CCP4 suite: Programs for protein crystallography. *Acta Crystallogr. Sect. D Biol. Crystallogr.* (1994). doi:10.1107/S0907444994003112.

133. Rupp, B. Biomolecular Crystallography. *Biomolecular Crystallography* (2009). doi:10.1201/9780429258756.

134. Emsley, P. & Cowtan, K. Coot: Model-building tools for molecular graphics. *Acta Crystallogr. Sect. D Biol. Crystallogr.* (2004). doi:10.1107/S0907444904019158.

135. Adams, P. D. *et al.* PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, (2010). doi: 10.1107/S0907444909052925.

136. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallographica Section D: Biological Crystallography* (1997). doi.org/10.1107/S0907444996012255.

137. Brünger, A. T. Free R value: A novel statistical quantity for assessing the accuracy of crystal structures. *Nature* (1992). doi:10.1038/355472a0.

138. Tickle, I. J., Laskowski, R. A. & Moss, D. S. Rfree and the Rfree ratio. I. Derivation of expected values of cross-validation residuals used in macromolecular least-squares refinement. *Acta Crystallogr. Sect. D Biol. Crystallogr.* (1998). doi:10.1107/s0907444997013875.

139. Kleywegt, G. J. & Brünger, A. T. Checking your imagination: Applications of the free R value. *Structure* (1996). doi:10.1016/S0969-2126(96)00097-4.

140. Kleywegt, G. J. & Jones, T. A. Model building and refinement practice. *Methods Enzymol.* (1997). doi:10.1016/S0076-6879(97)77013-7.

141. Sun, Z., Liu, Q., Qu, G., Feng, Y. & Reetz, M. T. Utility of B-Factors in Protein Science: Interpreting Rigidity, Flexibility, and Internal Motion and Engineering Thermostability. *Chemical Reviews* (2019). doi.org/10.1021/acs.chemrev.8b00290.

142. Shabalin, I. G., Porebski, P. J. & Minor, W. Refining the macromolecular model–achieving the best agreement with the data from X-ray diffraction experiment. *Crystallography Reviews* (2018). doi.org/10.1080/0889311X.2018.1521805.

143. Kantardjieff, K. A. & Rupp, B. Matthews coefficient probabilities: Improved estimates for unit cell contents of proteins, DNA, and protein–nucleic acid complex crystals. *Protein Sci.* (2003). doi:10.1110/ps.0350503.

144. Fokine, A. & Urzhumtsev, A. Flat bulk-solvent model: Obtaining optimal parameters. *Acta Crystallogr. Sect. D Biol. Crystallogr.* (2002). doi:10.1107/S0907444902010284.

145. Rees, B., Jenner, L. & Yusupov, M. Bulk-solvent correction in large macromolecular structures. *Acta Crystallogr. Sect. D Biol. Crystallogr.* (2005). doi:10.1107/S0907444905019591.

146. Chen, V. B. *et al.* MolProbity: All-atom structure validation for macromolecular crystallography. *Acta Crystallogr. Sect. D Biol. Crystallogr.* (2010). doi:10.1107/S0907444909042073.

147. Meller, J. Molecular Dynamics. in *Encyclopedia of Life Sciences* (2001). doi:10.1038/npg.els.0003048.

148. Bowers, K. J. *et al.* Scalable algorithms for molecular dynamics simulations on commodity clusters. in *Proceedings of the 2006 ACM/IEEE Conference on Supercomputing, SC'06* (2006). doi:10.1145/1188455.1188544.

149. Karplus, M. & Ichiye, T. Comment on a "Fluctuation and cross-correlation analysis of protein motions observed in nanosecond molecular dynamics simulations". *J. Mol. Biol.* (1995). doi:10.1006/jmbi.1995.0514.

150. Ranganathan, R. & Rivoire, O. Note 109 : A summary of SCA calculations. *Order A J. Theory Ordered Sets Its Appl.* (2011).

151. Rivoire, O., Reynolds, K. A. & Ranganathan, R. Evolution-Based Functional Decomposition of Proteins. *PLOS Comput. Biol.* **12**, e1004817 (2016). doi: 10.1371/journal.pcbi.1004817.

152. Fus-Kujawa, A. *et al.* An Overview of Methods and Tools for Transfection of Eukaryotic Cells in vitro. *Frontiers in Bioengineering and Biotechnology* (2021). doi.org/10.3389/fbioe.2021.701031.

153. Parks, D. R. & Herzenberg, L. A. [19] Fluorescence-Activated Cell Sorting: Theory, Experimental Optimization, and Applications in Lymphoid Cell Biology. *Methods Enzymol.* (1984). doi:10.1016/S0076-6879(84)08086-1.

154. Proserpio, V., Conti, L. & Oliviero, S. Flow Cytometry for Beginners: Hints and Tips for Approaching the Very First Single-Cell Technique. in *Methods in Molecular Biology* (2022). doi:10.1007/978-1-0716-1771-7_3.

155. Lojk, J. & Marc, J. Roles of non-canonical wnt signalling pathways in bone biology. *International Journal of Molecular Sciences* (2021). doi.org/10.3390/ijms221910840.

156. Sarabia-Sánchez, M. A. & Robles-Flores, M. WNT Signaling in Stem Cells: A Look into the Non-

Canonical Pathway. *Stem Cell Reviews and Reports* (2024). doi.org/10.1007/s12015-023-10610-5.

157. Cook, D. R., Rossman, K. L. & Der, C. J. Rho guanine nucleotide exchange factors: Regulators of Rho GTPase activity in development and disease. *Oncogene* (2014). doi.org/10.1038/onc.2013.362.

158. Xiao, Q., Chen, Z., Jin, X., Mao, R. & Chen, Z. The many postures of noncanonical Wnt signaling in development and diseases. *Biomedicine and Pharmacotherapy* (2017). doi.org/10.1016/j.biopha.2017.06.061.

159. Wang, Y. *et al.* WGEF is a novel RhoGEF expressed in intestine, liver, heart, and kidney. *Biochem. Biophys. Res. Commun.* **324**, (2004). doi: 10.1016/j.bbrc.2004.09.153.

160. Francis, D. M. & Page, R. Strategies to optimize protein expression in E. coli. *Current Protocols in Protein Science* (2010). doi.org/10.1002/0471140864.ps0524s61.

161. Yohe, M. E., Rossman, K. & Sondek, J. Role of the C-terminal SH3 domain and N-terminal tyrosine phosphorylation in regulation of Tim and related Dbl-family proteins. *Biochemistry* (2008). doi:10.1021/bi702543p.

162. Zhang, M., Lin, L., Wang, C. & Zhu, J. Double inhibition and activation mechanisms of Ephexin family RhoGEFs. *Proc. Natl. Acad. Sci. U. S. A.* **118**, (2021). doi: 10.1073/pnas.2024465118.

163. Tonikian, R. et al. A specificity map for the PDZ domain family. PLoS Biol. (2008). doi:10.1371/journal.pbio.0060239.

164. Penkert, R. R., DiVittorio, H. M. & Prehoda, K. E. Internal recognition through PDZ domain plasticity in the Par-6-Pals1 complex. Nat. Struct. Mol. Biol. 11, (2004). doi: 10.1038/nsmb839.

165. London, T. B. C., Lee, H. J., Shao, Y. & Zheng, J. Interaction between the internal motif KTXXXI of Idax and mDvl PDZ domain. *Biochem. Biophys. Res. Commun.* (2004). doi:10.1016/j.bbrc.2004.07.113.

166. Flot, D. *et al.* The ID23-2 structural biology microfocus beamline at the ESRF. *J. Synchrotron Radiat.* **17**, (2010). doi: 10.1107/S0909049509041168.

167. Evans, P. R. An introduction to data reduction: Space-group determination, scaling and intensity statistics. *Acta Crystallogr. Sect. D Biol. Crystallogr.* (2011). doi:10.1107/S090744491003982X.

168. Winn, M. D. *et al.* Overview of the CCP4 suite and current developments. *Acta Crystallographica Section D: Biological Crystallography* vol. 67 (2011). doi.org/10.1107/S0907444910045749.

169. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, (2010). doi: 10.1107/S0907444910007493.

170. Liebschner, D. *et al.* Macromolecular structure determination using X-rays, neutrons and electrons: Recent developments in Phenix. *Acta Crystallogr. Sect. D Struct. Biol.* **75**, (2019). doi: 10.1107/S2059798319011471.

171. Williams, C. J. *et al.* MolProbity: More and better reference data for improved all-atom structure validation. *Protein Sci.* (2018). doi:10.1002/pro.3330.

172. Songyang, Z. *et al.* Recognition of unique carboxyl-terminal motifs by distinct PDZ domains. *Science (80-.).* **275**, (1997). doi: 10.1126/science.275.5296.73.

173. Ali, M. *et al.* Integrated analysis of Shank1 PDZ interactions with C-terminal and internal binding motifs. *Curr. Res. Struct. Biol.* (2021) doi:10.1016/j.crstbi.2021.01.001.

174. Zhu, Y. *et al.* Deciphering the Unexpected Binding Capacity of the Third PDZ Domain of Whirlin to Various Cochlear Hair Cell Partners. *J. Mol. Biol.* (2020). doi:10.1016/j.jmb.2020.09.012.

175. Zhang, Y. *et al.* Inhibition of Wnt signaling by Dishevelled PDZ peptides. *Nat. Chem. Biol.* (2009). doi:10.1038/nchembio.152.

176. Aasland, R. *et al.* Normalization of nomenclature for peptide motifs as ligands of modular protein domains. *FEBS Letters* vol. 513 (2002). doi.org/10.1016/S0014-5793(01)03295-1.

177. Skelton, N. J. *et al.* Origins of PDZ domain ligand specificity. Structure determination and mutagenesis of the erbin PDZ domain. *J. Biol. Chem.* (2003). doi:10.1074/jbc.M209751200.

178. Lee, H. J., Shi, D. L. & Zheng, J. J. Conformational change of dishevelled plays a key regulatory role in the wnt signaling pathways. *Elife* (2015). doi:10.7554/eLife.08142.

179. Nielsen, C. P., Jernigan, K. K., Diggins, N. L., Webb, D. J. & MacGurn, J. A. USP9X Deubiquitylates DVL2 to Regulate WNT Pathway Specification. *Cell Rep.* (2019). doi:10.1016/j.celrep.2019.06.083.

180. Ivarsson, Y. Plasticity of PDZ domains in ligand recognition and signaling. *FEBS Letters* (2012). doi.org/10.1016/j.febslet.2012.04.015.

181. Mu, Y., Cai, P., Hu, S., Ma, S. & Gao, Y. Characterization of diverse internal binding specificities of PDZ domains by yeast two-hybrid screening of a special peptide library. *PLoS One* **9**, (2014). doi: 10.1371/journal.pone.0088286.

182. Münz, M., Hein, J. & Biggin, P. C. The Role of Flexibility and Conformational Selection in the Binding Promiscuity of PDZ Domains. *PLoS Comput. Biol.* (2012). doi:10.1371/journal.pcbi.1002749.

183. Hillier, B. J., Christopherson, K. S., Prehoda, K. E., Bredt, D. S. & Lim, W. A. Unexpected modes of PDZ domain scaffolding revealed by structure of nNOS-syntrophin complex. *Science (80-.).* (1999). doi:10.1126/science.284.5415.812.

184. Whitney, D. S., Peterson, F. C. & Volkman, B. F. A conformational switch in the CRIB-PDZ module of Par-6. *Structure* **19**, (2011). doi: 10.1016/j.str.2011.07.018.

185. Šali, A. & Blundell, T. L. Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* (1993). doi:10.1006/jmbi.1993.1626.

186. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, (1983). doi.org/10.1063/1.445869

187. Evans, D. J. & Holian, B. L. The Nose-Hoover thermostat. *J. Chem. Phys.* **83**, (1985). doi.org/10.1063/1.449071

188. Nosé, S. A unified formulation of the constant temperature molecular dynamics methods. *J. Chem. Phys.* **81**, (1984). doi.org/10.1063/1.447334

189. Martyna, G. J., Tobias, D. J. & Klein, M. L. Constant pressure molecular dynamics algorithms. *J. Chem. Phys.* (1994). doi:10.1063/1.467468.

190. Martyna, G. J., Tuckerman, M. E., Tobias, D. J. & Klein, M. L. Explicit reversible integrators for extended systems dynamics. *Mol. Phys.* (1996). doi:10.1080/00268979600100761.

191. Jorgensen, W. L., Maxwell, D. S. & Tirado-Rives, J. Development and testing of the OPLS all-atom

force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.* (1996). doi:10.1021/ja9621760.

192. Shivakumar, D. *et al.* Prediction of absolute solvation free energies using molecular dynamics free energy perturbation and the opls force field. *J. Chem. Theory Comput.* (2010). doi:10.1021/ct900587b.

193. Karplus, M. & Ichiye, T. Comment on a "Fluctuation and cross-correlation analysis of protein motions observed in nanosecond molecular dynamics simulations". *J. Mol. Biol.* **252**, 492–503 (1995). doi: 10.1006/jmbi.1995.0514.

194. Grant, B. J., Rodrigues, A. P. C., ElSawy, K. M., McCammon, J. A. & Caves, L. S. D. Bio3d: An R package for the comparative analysis of protein structures. *Bioinformatics* (2006). doi:10.1093/bioinformatics/btl461.

195. Singh, S., Thulasiram, H. V., Sengupta, D. & Kulkarni, K. Dynamic coupling analysis on plant sesquiterpene synthases provides leads for the identification of product specificity determinants. *Biochem. Biophys. Res. Commun.* (2021). doi:10.1016/j.bbrc.2020.12.041.

196. Tochio, H. *et al.* Formation of nNOS/PSD-95 PDZ dimer requires a preformed β-finger structure from the nNOS PDZ domain. *J. Mol. Biol.* (2000). doi:10.1006/jmbi.2000.4148.

197. Lu, C., Knecht, V. & Stock, G. Long-Range Conformational Response of a PDZ Domain to Ligand Binding and Release: A Molecular Dynamics Study. *J. Chem. Theory Comput.* (2016). doi:10.1021/acs.jctc.5b01009.

198. Amacher, J. F., Brooks, L., Hampton, T. H. & Madden, D. R. Specificity in PDZ-peptide interaction networks: Computational analysis and review. *Journal of Structural Biology: X* (2020). doi.org/10.1016/j.yjsbx.2020.100022.

199. Stevens, A. O. & He, Y. Allosterism in the PDZ Family. *International Journal of Molecular Sciences* (2022). doi.org/10.3390/ijms23031454.

200. Harris, B. Z. & Lim, W. A. Mechanism and role of PDZ domains in signaling complex assembly. *Journal of Cell Science* (2001). doi: 10.1242/jcs.114.18.3219.

201. Wharton, K. A. Runnin' with the Dvl: Proteins that associate with Dsh/Dvl and their significance to Wnt signal transduction. *Developmental Biology* (2003). doi.org/10.1006/dbio.2002.0869 (2003).

202. Tochio, H., Zhang, Q., Mandal, P., Li, M. & Zhang, M. Solution structure of the extended neuronal nitric oxide synthase PDZ domain complexed with an associated peptide. *Nat. Struct. Biol.* (1999). doi:10.1038/8216.

203. Ouko, L., Ziegler, T. R., Gu, L. H., Eisenberg, L. M. & Yang, V. W. Wnt11 signaling promotes proliferation, transformation, and migration of IEC6 intestinal epithelial cells. *J. Biol. Chem.* (2004). doi:10.1074/jbc.M402877200.

204. Prasad, C. P., Chaurasiya, S. K., Guilmain, W. & Andersson, T. WNT5A signaling impairs breast cancer cell migration and invasion via mechanisms independent of the epithelial-mesenchymal transition. *J. Exp. Clin. Cancer Res.* (2016). doi:10.1186/s13046-016-0421-0.

205. Carvalho, J. R. *et al.* Non-canonical Wnt signaling regulates junctional mechanocoupling during angiogenic collective cell migration. *Elife* (2019). doi:10.7554/elife.45853.

206. Longo, P. A., Kavran, J. M., Kim, M. S. & Leahy, D. J. Transient mammalian cell transfection with polyethylenimine (PEI). in *Methods in Enzymology* (2013). doi:10.1016/B978-0-12-418687-

3.00018-5.

207. Willert, K. H. Isolation and application of bioactive Wnt proteins. *Methods Mol. Biol.* (2008). doi:10.1007/978-1-59745-249-6_2.

208. Mikels, A. J. & Nusse, R. Purified Wnt5a protein activates or inhibits β-catenin-TCF signaling depending on receptor context. *PLoS Biol.* (2006). doi:10.1371/journal.pbio.0040115.

209. Yang, J. *et al.* Wnt5a increases properties of lung cancer stem cells and resistance to cisplatin through activation of Wnt5a/PKC signaling pathway. *Stem Cells Int.* (2016). doi:10.1155/2016/1690896.

210. Li, S., Wang, W., Zhang, N., Ma, T. & Zhao, C. IL-1β mediates MCP-1 induction by Wnt5a in gastric cancer cells. *BMC Cancer* (2014). doi:10.1186/1471-2407-14-480.

211. Narumiya, S. & Thumkeo, D. Rho signaling research: history, current status and future directions. *FEBS Letters* (2018). doi.org/10.1002/1873-3468.13087.

212. Narumiya, S., Tanji, M. & Ishizaki, T. Rho signaling, ROCK and mDia1, in transformation, metastasis and invasion. *Cancer and Metastasis Reviews* (2009). doi.org/10.1007/s10555-008-9170-7.

213. Adams, P. D., Lopez, P., Sellers, W. R. & Kaelin, W. G. Fluorescence-activated cell sorting of transfected cells. *Methods Enzymol.* (1997). doi:10.1016/S0076-6879(97)83007-8.

214. Kaufman, W. L. *et al.* Homogeneity and persistence of transgene expression by omitting antibiotic selection in cell line isolation. *Nucleic Acids Res.* (2008). doi:10.1093/nar/gkn508.

Appendix A

Structural studies on Frizzled 7 receptor

A.1 Background

Frizzled (FZD) receptors are categorized within the class F of the GPCRs. They are also referred to as atypical GPCRs, as their primary signalling does not involve canonical G-proteins[1,2]. There are ten paralogs of FZD receptors found in mammals which can interact with the 19 secreted glycoproteins, called Wnt ligands, to activate either canonical (β-catenine dependent) or non-canonical (β-catenine independent) Wnt signalling. These receptors are shown to play an important role in cell proliferation, cell and tissue polarity, neural synapse formation, etc. However, their exact role in the physiology of adult organisms is not clear. These receptors have conserved structural features such as N-terminal cysteine-rich domain (CRD), which interacts with the Wnt ligands, a hydrophilic linker region, a hepta helical transmembrane domain and a cytosolic C-terminal domain that binds the downstream effectors such as Dvl proteins during signalling[3] (Fig A.1) In pathological conditions, FZD receptors have been linked to cancer progression[4].

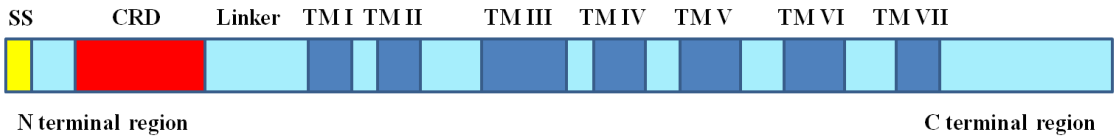


Figure A.1: Domain arrangement of Frizzled receptors: *SS, signal sequence region., CRD, Cysteine-rich domain for Wnt and binding of other ligands., TM, transmembrane domains. The cytosolic C-terminal region interacts with downstream effectors during signalling.*

Among these receptors FZD7 is the most extensively studied receptor due to its ability to activate both canonical and non-canonical Wnt signalling[5]. Perhaps because of this promiscuous nature of FZD7, it is implicated in several processes like cell proliferation, cell differentiation and tissue homeostasis. It is also majorly involved in mesendodermal cell migration and stem cell differentiation[6,7]. In *Xenopus laevis,* the homolog of FZD7 has been shown to be involved in convergent extension movements during gastrulation[8] and is vital for intestinal stem cell maintenance[9]. Apart from these roles, upregulation and mutations of this

receptor have been implicated in various cancers, where it has been shown to promote tumour cell metastasis, proliferation, chemoresistance and cancer stem cell maintenance[10,11]. Therefore this receptor has been recognized as a potential target for cancer therapeutics.

Therefore, understanding the structural basis of FZD7 activation is crucial to decipher how it selectively engages different signalling pathways, and also this knowledge could pave the way for designing drugs that specifically target FZD7 signalling, specifically for treating cancer. With the objective of determining the crystal structure of FZD7, primarily to unravel the signalling mechanism at the molecular level, we set out to produce recombinant protein and crystallize the same. However, during the course of our work, Lu Xu *et al*. reported the crystal structure of FZD7 from humans (Hs_FZD7)[12]. This chapter reports our work on establishing recombinant mouse FZD7 production, which involves optimization of the gene construct and expression system.

A.2 Methodology

A.2.1 Cloning and transformation of DH10 cells

To achieve stable mouse Frizzled 7 (FZD7) protein, the gene encoding this protein was PCR amplified and cloned into the BacMam vector with GFP tag (Fig A.2A). This is a baculovirus based protein expression vector system primarily used for gene expression in insect cells. However, they are engineered to allow transduction and gene expression in mammalian cells without replicating them. It combines the advantages of baculovirus vectors for gene delivery with the ability to express proteins in mammalian cells. BacMam vector has the p10 and polyhedrin promoters from the *Autographa californica* multiple nuclear polyhedrosis virus (AcMNPV) for expression in insect cells and cytomegalovirus (CMV) promoter for expression in mammalian cells (HEK293F). The polyhedrin promoter is in frame with the CMV promoter and the target gene is inserted in the BacMam vector under the CMV promoter (Fig A.2B). The cloned BacMam vector is transfected into *Spodoptera frugiperda (Sf9)* insect cells, where it replicates and produces high-titre recombinant baculoviruses. These baculovirus particles are then harvested to infect mammalian cells for protein expression. Additionally, FZD7 constructs were also commercially synthesized using pFastBac vector with a GFP tag, another baculovirus expression vector system to express recombinant proteins in *Spodoptera frugiperda (Sf9)* insect

cells (Fig A.2C). In the pFastBac vector, the gene expression is under the control of p10 or polyhedrin promoters, which enables high level expression in insect cells.

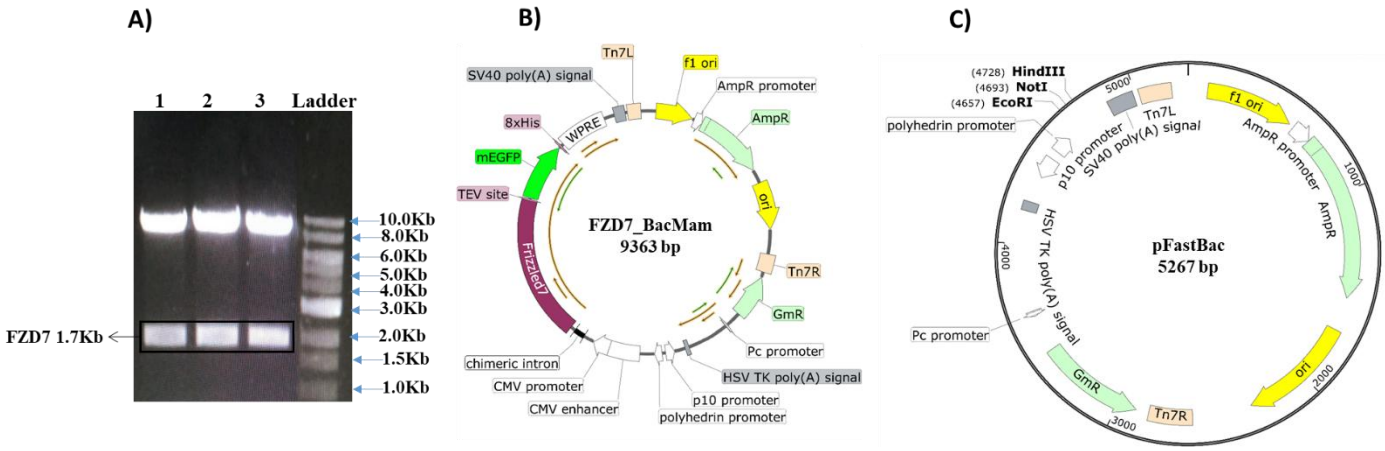


Figure A.2: FZD7 constructs and vector maps. A) *Agarose gel for positive clones of FZD7 in BacMam vector* ***B)*** *BacMam vector map* ***C)*** *pFastBac vector map.*

The FZD7-pFastBac or FZD7-BacMam plasmids, referred to as donor plasmids (carry gentamicin and ampicillin resistance), were transformed into DH10*Bac E. coli* cells. These cells contain a bacmid (with kanamycin resistance) and a helper (with tetracycline resistance) plasmid. Following the transformation of cells with the donor plasmid, a transposition event occurs between the donor plasmid and the bacmid, which contains a mini-*att*Tn7 site that facilitates transposition, producing a recombinant bacmid. The bacmid also contains a lacZ gene that encodes β-galactosidase, which complements the absence of the lacZ gene on the bacterial chromosome. In the presence of Bluo-gal substrate and IPTG, it gives blue colonies. IPTG binds to and inhibits the lac repressor protein (prevents the transcription of lacZ gene). This enables the expression of β-galactosidase to hydrolyze Bluo-gal into blue colour precipitate. The lacZ gene also includes a sequence for mini-*att*Tn7, a bacterial transposition site. Once the donor plasmid gets inserted at this transposition site, the lacZ gene is disrupted, preventing the breakdown of Bluo-gal giving white colonies, which are considered as positive clones and further given for sequence confirmation[13]. The antibiotics are added to the plates at final concentrations of 50 µg/mL Ampicillin, 40 µg/mL IPTG, 7 µg/mL Gentamicin, 50 µg/mL Kanamycin, 100 µg/mL Bluo-gal and 10 µg/mL Tetracycline to selective isolate the transformed colonies.

Other FZD7 constructs that were commercially synthesized for stable expression in *Sf9* insect cells are given in Figure A.3.

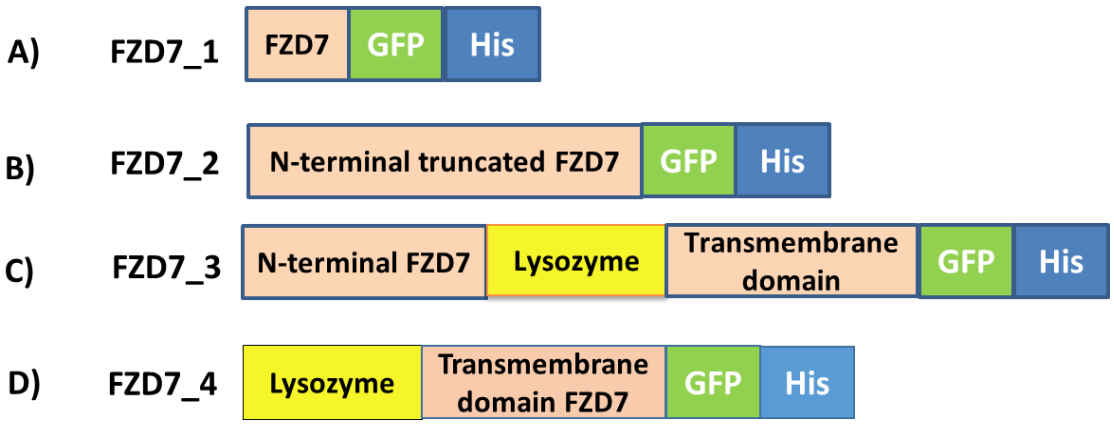


Figure A.3: FZD7 constructs in pFastBac vector. ***A)*** *Full-length FZD7 construct* ***B)*** *N-terminal cysteine-rich domain (CRD) of FZD7 is truncated* ***C)*** *Lysozyme gene sequence is inserted between the CRD and the transmembrane domains of FZD7* ***D)*** *CRD domain is replaced with the lysozyme encoding gene sequence. All these constructs have a GFP tag for visualization and a His tag for affinity purification.*

A.2.2 Bacmid isolation

The positive clones obtained after blue-white screening and sequence confirmation were grown with antibiotics containing media. Bacmid DNA was purified by the alkaline lysis method. The bacterial cell pellet was first reconstituted in a resuspension buffer (Tris buffer pH 8.5 or TE buffer pH 8) followed by alkaline lysis with NaOH and SDS for 5 min in the presence of 100 µg/mL RNase. Following cell lysis, the chromosomal DNA, proteins and other insoluble components were precipitated during the neutralization step. This mixture was then centrifuged at high speed to separate the supernatant containing bacmid DNA from the precipitate. The supernatant was removed in a fresh tube, and approximately 800 µL of isopropanol was added to precipitate the bacmid DNA. The mixture was gently mixed and incubated at RT/20 min. After incubation, the mixture was centrifuged at 14,000 RPM/30 min. The supernatant was carefully removed under sterile conditions and the bacmid pellet was left to air dry in the laminar for 30 min. After drying, the pellet was resuspended in sterile MilliQ water (60 µL).

A.2.3 Transfection and baculovirus amplification

After purifying the bacmids, transfection of *Sf9* insect cells with the bacmids was carried out using the GeneJuice (Sigma-Aldrich), as per the manufacturer's protocol. Briefly, a day prior to transfection, *Sf9* cells were split to ensure they were in the log phase at the time of transfection. On the day of transfection, the bacmid and Genejuice were mixed at different concentrations (optimal recommended ratio is 3 µL Genejuice: 1µg DNA) in Sf-900™ III SFM medium, followed by incubation at RT/15-20 min. *Sf9* cells were diluted to 1 x10^6 cells/ mL and one mL of this diluted culture was added to the wells of a 6-well plate. Cells were allowed to adhere at 27°C/30 min. Media was then replaced with 800 µL of fresh media, and the transfection reagent-bacmid mixture was added to the cells dropwise and incubated 27°C/5 h. After this, the media was again replaced with fresh media and incubated for 72 h. After 72 h, cells were scraped into the media using a cell scraper. The media containing the cells was then collected and stored with 2% FBS at 4°C. This was referred to as the P1 virus, which is the first generation of recombinant baculovirus produced after transfection. This P1 viral stock was subsequently used for virus amplification to generate P2 virus, which was further amplified to produce P3. Finally, P3 was used for protein expression.

For viral amplification and protein expression, the multiplicity of infection (MOI) is a crucial factor. This is known as the number of virus particles per cell. In order to amplify the viral stock, an MOI in the range of 0.05 to 0.1 is recommended, whereas for protein expression, an MOI of 1 to 5 is recommended[14]. This optimal MOI may vary depending on the specific cell lines. Following formula can be used to determine the inoculum volume of the viral stock to be added for a specific MOI:

$$Inoculum\ volume\ (mL) = \frac{Desired\ MOI\ X\ Number\ of\ cells}{Viral\ stock\ titer\ (pfu/\ mL)}$$

For viral amplification, the *Sf9* cells were infected during the mid-log phase of growth, at a cell density of approximately 2 x10^6 cells/ mL and cell viability of over 95%. Cells were incubated at 27°C/72 h, with viability monitored at 48 and 72 h. As viral production progresses, it leads to cell death, resulting in a decrease in the cell viability after 72 h, which is an indicator of viral production. After incubation, cells were centrifuged and the supernatant that contains the baculovirus was supplemented with 2% FBS and stored at 4°C for subsequent use.

A.2.4 Protein expression and affinity purification

For expression of protein, *Sf9* cells were cultured to a cell density of approximately 2 x10^6 cells/ mL with a cell viability of >95%. Cells were infected according to the MOI for protein expression and incubated at 27℃/72 h. Compared to the viral production, the cell viability during the protein expression does not decrease much. At 48 h, cells were monitored for GFP fluorescence, indicating protein production. After 72 h, the cells were centrifuged and the media was removed. The cell pellet containing the protein was washed twice with 1X PBS and stored at -80℃ for further use. For His tag based affinity purification, cells were resuspended in a 10X volume of lysis buffer (50 mM Tris pH 8.5, 300 mM NaCl, 2 mM β-met, 5% glycerol, 3 mM imidazole and PI tablet). Cells were subjected to sonication of a 1.30 min cycle with a 2 sec ON/8 sec OFF pulse at a 35% amplitude. After sonication the lysate was ultracentrifuged to pellet the membrane fraction containing the protein. The supernatant was removed and the membrane fraction was resuspended using a hand homogenizer in solubilization buffer (50 mM Tris pH 8.5, 300 mM NaCl, 3 mM imidazole, 2 mM β-met, 5% glycerol and PI tablet) containing 1% of detergent. This was incubated at 4ºC/2 h followed by ultracentrifugation. Now, the solubilized protein is present in the supernatant which is further used for the Ni^{2+}-NTA affinity purification using a buffer containing one critical micelle concentration (CMC) of respective detergent.

A.3 Results

A.3.1 Expression of FZD7 constructs in HEK293F and *Sf9* cell lines

Initially, we planned to express FZD7 in HEK293F cells using the viral transfection method. To achieve this, we cloned the full-length FZD7 receptor into the BacMam vector, as this vector contains both p10 and CMV promoters required for protein expression in insect and mammalian cell lines, respectively. Furthermore, this vector also carries the polyhedrin promoter required for the generation of viral bacmid in the DH10 bacterial strain. The baculovirus was produced by transfecting *Sf9* cells with the bacmid. The viral titer was enhanced by successive amplification of the virus using *Sf9* cells. For protein expression in mammalian cells, the virus containing media was added to HEK293F cells. However, production of FZD7 was not detected, after 72 h of post-infection growth of HEK293F cells.

Next, we tried expression of FZD7 in *Sf9* cells for which we designed the constructs with the gene cloned in pFastBac vector. Although, BacMam vector carries all essential parts required for the production of recombinant protein in *Sf9* (insect) cell lines, due to the positioning of polyhedrin promoter upstream of CMV promoter, genes cloned in this vector are known to express poorly in insect cell lines. We observed very feeble expression of FZD7 in *Sf9* cell lines. We tried to improve the expression levels by designing the full-length (FZD7_1) and a CRD truncated version of FZD7 in pFastBac (FZD7_2) (Fig A.3).

Furthermore, it has been shown that membrane proteins, when expressed as fusion proteins with bulkier tags like GFP, MBP and lysozyme, exhibit enhanced expression levels and improved solubility[15]. Also, there is an added advantage of enhancing or increasing crystallization likelihood of membrane proteins if lysozyme is inserted in their loops or if they are expressed as lysozyme fusion proteins[16,17]. Therefore, we adopted this strategy and designed constructs with lysozyme inserted in the loop connecting the CRD and transmembrane domains (FZD7_3). We also designed another construct that has its CRD replaced with the lysozyme (FZD7_4) (Fig A.3), which connects with the N-terminal segment of the helical membrane domain through a 10 amino acid linker. To track the expression levels, we had introduced a C-terminal GFP tag to all these constructs. From cell imaging it was clear that FZD7_1-3 constructs were producing decent levels of protein in *Sf9* cells (Fig A.4).

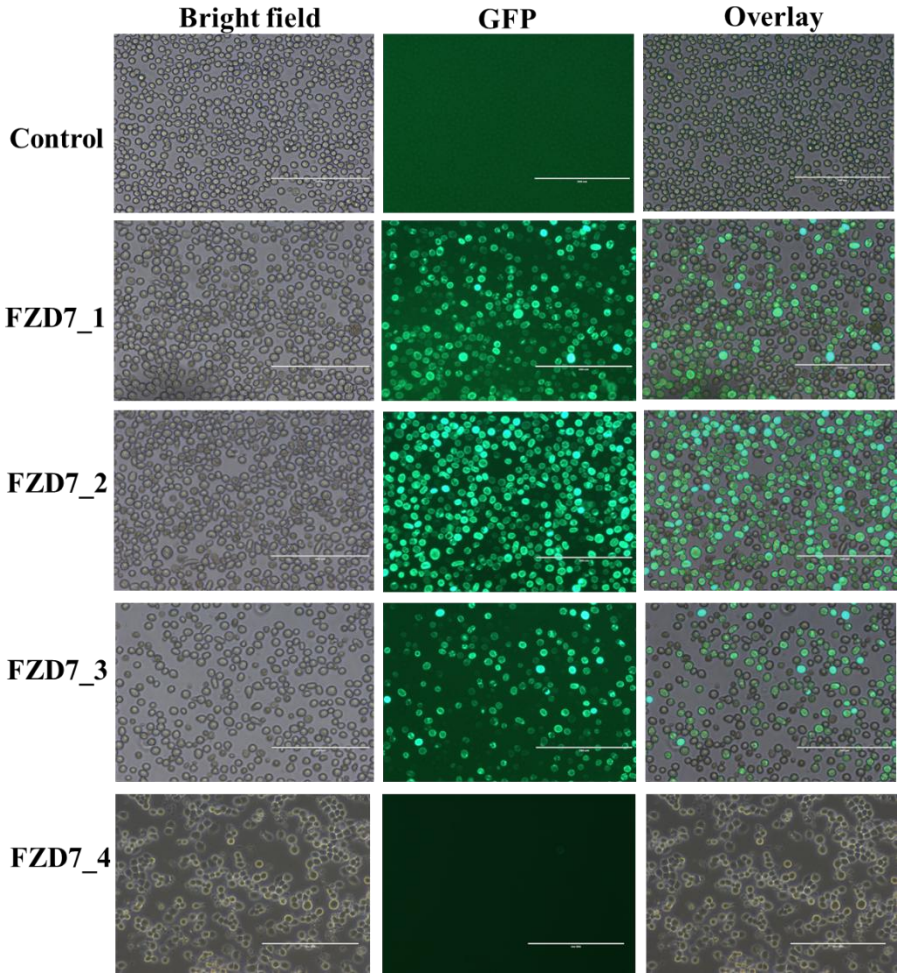


Figure A.4: Protein expression of different FZD7 constructs in Sf9 cells detected by GFP tag expression.

A.3.2 Screening and optimization of FZD7 protein stability with different detergents

The next logical step in recombinant membrane protein purification will be to extract the protein from the membrane using detergents. Since there is no well-defined rule for selecting detergents for particular proteins, the extraction process involves screening and optimizing detergents. To check the stability of FZD7 proteins in different detergents and to further optimize their purification for crystallization studies, we used some common detergents used for the purification of membrane proteins[18]. It has been observed that, for the FZD7_1 construct, Lauryldimethylamine oxide (LDAO) and Decyl β-D-maltopyranoside (DM) detergents were found to be optimal for the extraction. We tested the solubility of FZD7 in the presence of these detergents by subjecting the extract to the Ni^{2+}-NTA affinity purification

step. The solubility levels were assessed by exposing the SDS-PAGE gels carrying the affinity column elution fractions to the UV light. Since the constructs were expressed as GFP-tagged protein bands, their size, expression level and solubility could be gauged from the aforementioned step. For FZD7_1, the protein band was found in the pellet fraction, indicating poor protein solubility (Fig A.5). For the shorter construct of FZD7, that had CRD deletion (FZD7_2), the protein band was seen in the supernatant, suggesting decent protein solubility. However, the observed protein band appeared to be of much lower molecular weight (below 37 kDa) compared to the expected size (69.7 kDa), indicating possible protein degradation (Fig A.6). The lysozyme fused FZD7 was found to be detected with a lower size of approx. 75 kDa, instead of the expected 112 kDa size of the fused protein (Fig A.7). Thus, even this construct was also found to be unstable and degrading. In many instances, the extracellular loops of GPCR proteins are not stable with large insertions or deletions, therefore, N-terminal extracellular CRD domain of FZD7 was replaced with the lysozyme (FZD7_4)[17]. This construct, however, showed poor expression in *Sf9* cells, as indicated by the limited GFP fluorescence observed in the cells during expression (Fig A.4). From these different constructs of FZD7, we tried to analyze the stability of protein in different detergents and further wanted to optimize their purification. Meanwhile, the Cryo-EM structure (3.2 Å resolution) of constitutively active FZD7 in complex with heterotrimeric Gs was published by Lu Xu *et al*[12] (Fig A.8A). Therefore, we stopped our pursuit with the mouse FZD7 construct. Additionally, a recent study has also reported Cryo-EM structure of inactive FZD7 at a resolution of 1.9 Å[19] (Fig A.8B).

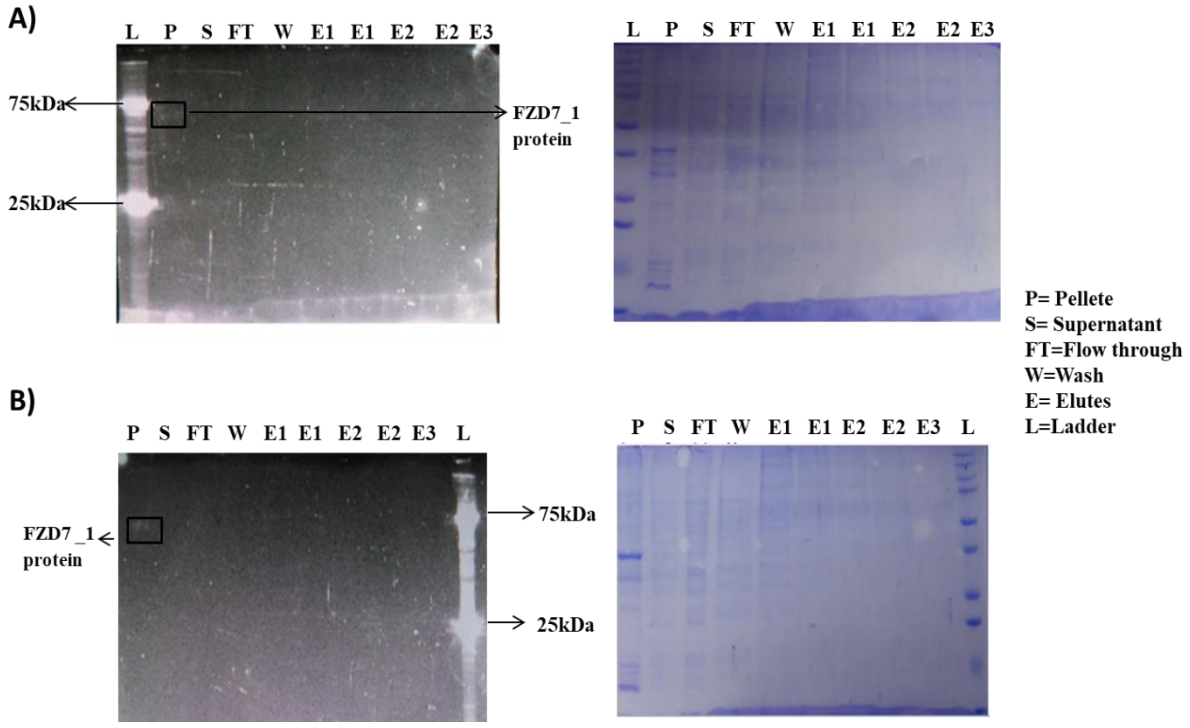


Figure A.5: Affinity purification of FZD7_1 protein. ***A)*** *Purification done using DM detergent.* ***B)*** *Purification done using LDAO detergent. Pictures represent UV image and stained image of the SDS-PAGE gel.*

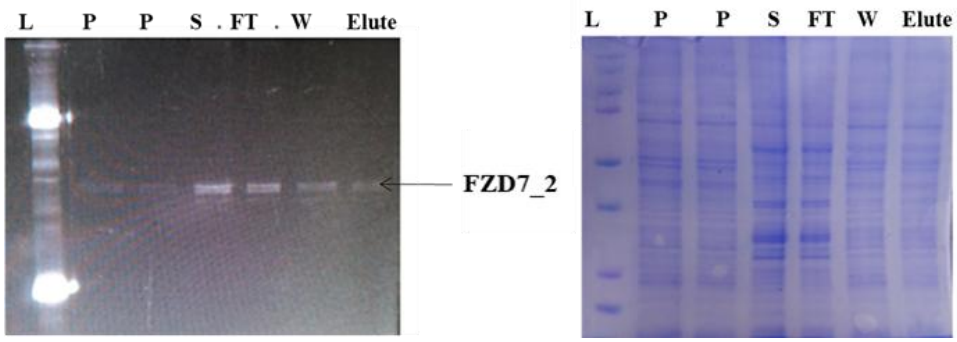


Figure A.6: Affinity purification of FZD7_2 protein using LDAO detergent. *Gel images represent UV image and stained image of the SDS-PAGE gel.*

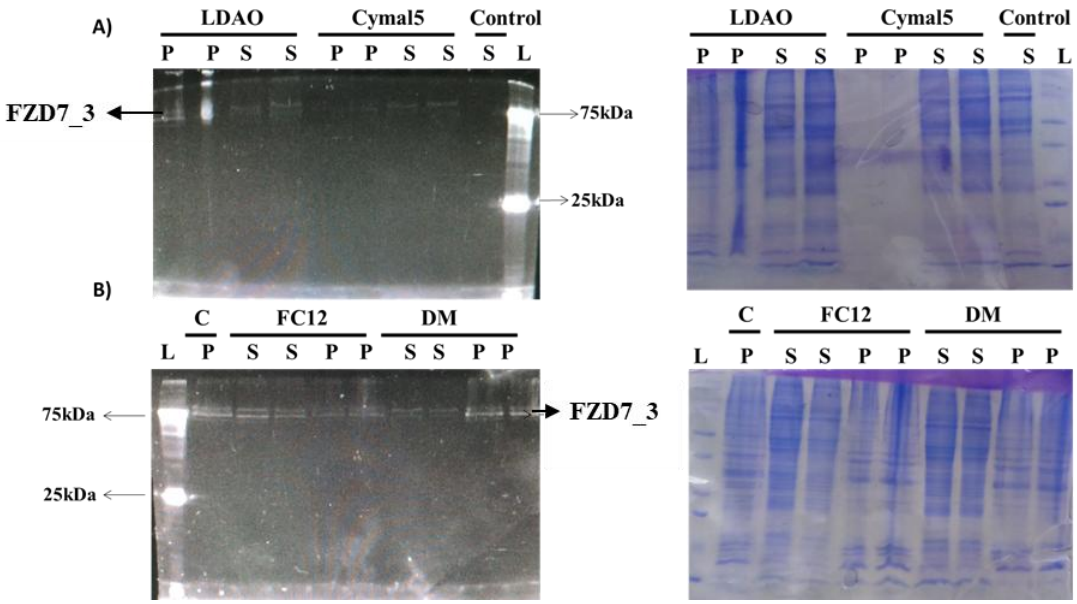


Figure A.7: Detergent screening for FZD7_3 protein using A) *LDAO and Cymal5 detergents.* ***B)*** *FC12 and DM detergents.* ***Gel*** *images represent UV image and stained image of the SDS-PAGE gel.*

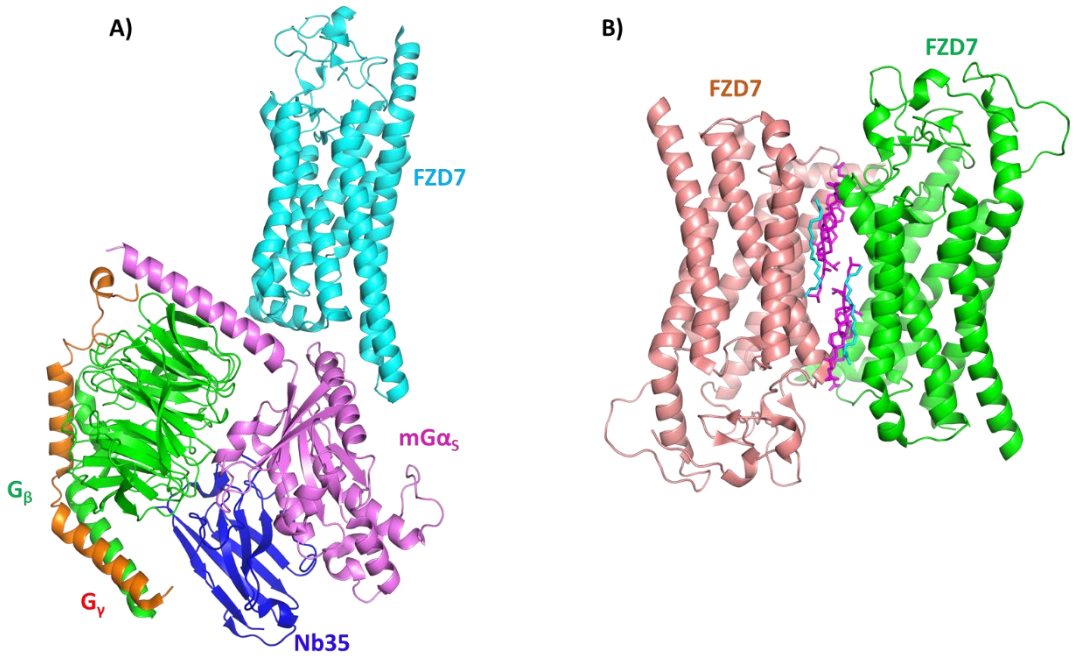


Figure A.8: Cryo-EM structures of FZD7 in constitutively active and inactive states. A) *Constitutively active FZD7 (cyan) in complex with heterotrimeric mini Gαs (pink),Gβ (green), Gγ (orange) subunits and Nanobody35 (Nb35) (blue) which stabilizes these subunits in nucleotide-free form.* ***B)*** *Structure of inactive FZD7 dimer (peach and green) formed by lipid interface consisting of palmitic acid (cyan) and cholesterol hemisuccinate (pink).*

A.4 Discussion

As a part of our pursuit to elucidate the structural basis of non-canonical Wnt-PCP signalling, we aimed to determine the crystal structure of the FZD7. In this direction, various constructs of FZD7 were designed and tested for their expression, stability and solubility in presence of various detergents. During our expression trials, we encountered many challenges such as instability of the protein and protein being prone to aggregation or degradation. Before we could further optimize the recombinant protein production of FZD7, a cryo-EM structure of constitutively active FZD7 (PDB ID: 8YY8), in complex with heterotrimeric Gs, was published by another group[12]. This structure reveals the mechanism of signal relay from FZD7 to the G_S subunit of the heterotrimeric G proteins. Additionally, this study also reveals that FZD7 has developed a distinct mechanism to preserve homologous movements characteristics of class A and class B GPCR activation, while also adapting its unique structural features to facilitate G protein activation. Although not completely, but partially, this report had essentially made our efforts redundant. Hence, we stopped pursuing the structure determination of *apo* FZD7. However, we could advance our understanding if we determine the structure of FZD7 in complex with its cytosolic ligand Dvl. Given the feeble and unstable expression of our constructs, production of recombinant FZD7-Dvl complex was too far fetching for this thesis work. But the work reported here shall provide invaluable insights on the expression of mouse FZD7 constructs in heterologous systems. The knowledge thus generated here could be used in optimizing the constructs for further structural studies on Frizzleds and their complexes with other signalling molecules.

References

1. Nichols, A. S., Floyd, D. H., Bruinsma, S. P., Narzinski, K. & Baranski, T. J. Frizzled receptors signal through G proteins. *Cell. Signal.* (2013). doi:10.1016/j.cellsig.2013.03.009.

2. Arthofer, E. *et al.* Class Frizzled GPCRs in GtoPdb v.2023.1. *IUPHAR/BPS Guid. to Pharmacol. CITE* (2023). doi:10.2218/gtopdb/f25/2023.1.

3. Huang, H. C. & Klein, P. S. The frizzled family: Receptor for multiple signal transduction pathways. *Genome Biology* (2004). doi.org/10.1186/gb-2004-5-7-234.

4. Zeng, C. M., Chen, Z. & Fu, L. Frizzled receptors as potential therapeutic targets in human cancers. *International Journal of Molecular Sciences* (2018). doi.org/10.3390/ijms19051543.

5. Phesse, T., Flanagan, D. & Vincan, E. Frizzled7: A promising achilles' heel for targeting the Wnt receptor complex to treat cancer. *Cancers* (2016). doi.org/10.3390/cancers8050050.

6. Gumber, D. *et al.* Selective activation of FZD7 promotes mesendodermal differentiation of human pluripotent stem cells. *Elife* (2020). doi:10.7554/ELIFE.63060.

7. Čapek, D. *et al.* Light-activated Frizzled7 reveals a permissive role of non-canonical wnt signaling in mesendoderm cell migration. *Elife* (2019). doi:10.7554/eLife.42093.

8. Kim, G. H., Her, J. H. & Han, J. K. Ryk cooperates with Frizzled 7 to promote Wnt11-mediated endocytosis and is essential for Xenopus laevis convergent extension movements. *J. Cell Biol.* (2008). doi:10.1083/jcb.200710188.

9. Flanagan, D. J. *et al.* Frizzled7 functions as a Wnt receptor in intestinal epithelial Lgr5+ stem cells. *Stem Cell Reports* (2015). doi:10.1016/j.stemcr.2015.03.003.

10. Ueno, K. *et al.* Frizzled-7 as a potential therapeutic target in colorectal cancer. *Neoplasia* (2008). doi:10.1593/neo.08320.

11. Yang, L. *et al.* FZD7 has a critical role in cell proliferation in triple negative breast cancer. *Oncogene* (2011). doi:10.1038/onc.2011.145.

12. Xu, L. *et al.* Cryo-EM structure of constitutively active human Frizzled 7 in complex with heterotrimeric Gs. *Cell Research* (2021). doi.org/10.1038/s41422-021-00525-6.

13. Luckow, V. A., Lee, S. C., Barry, G. F. & Olins, P. O. Efficient generation of infectious recombinant baculoviruses by site-specific transposon-mediated insertion of foreign genes into a baculovirus genome propagated in Escherichia coli. *J. Virol.* (1993). doi:10.1128/jvi.67.8.4566-4579.1993.

14. Sanchez, A. Baculovirus expression protocols. *Virus Res.* (1995). doi:10.1016/0168-1702(95)90027-6.

15. Liu, S. & Li, W. Protein Fusion Strategies for Membrane Protein Stabilization and Crystal Structure Determination. *Crystals* (2022). doi.org/10.3390/cryst12081041.

16. Thorsen, T. S., Matt, R., Weis, W. I. & Kobilka, B. K. Modified T4 Lysozyme Fusion Proteins

Facilitate G Protein-Coupled Receptor Crystallogenesis. *Structure* (2014). doi:10.1016/j.str.2014.08.022.

17. Zou, Y., Weis, W. I. & Kobilka, B. K. N-Terminal T4 Lysozyme Fusion Facilitates Crystallization of a G Protein Coupled Receptor. *PLoS One* (2012). doi:10.1371/journal.pone.0046039.

18. Stetsenko, A. & Guskov, A. An overview of the top ten detergents used for membrane protein crystallization. *Crystals* (2017). doi:10.3390/cryst7070197.

19. Bous, J. *et al.* Structural basis of frizzled 7 activation and allosteric regulation. *Nat. Commun.* **15**, 1–15 (2024). doi: 10.1038/s41467-024-51664-4.

Appendix B

List of primers

Primer name	5'-3' sequence
xWGEF404_P3E_FP	GGT CAT ATG ACC TTC TCC CTG TGG CAG GAT ATC
xWGEF404_P3E_RP	GGT GGA TCC CTC TCC AGT TTG CTT GTT GCA TGT TTT ATG
xWGEF386_P3E_FP	GGT CAT ATG ACG GGC AGT CCC AAG AGT AAC CTG
xWGEF386_P3E_RP	GGT GGA TCC TTA TGA TTG GGC GCA GGA GAG GCT CAT
hWGEF330_P3E_FP	GGT CAT ATG CCG CGG GCC AAC CTC TC
hWGEF330_P3E_RP	CGT GGA TCC TTA GCC CTC AAA GTG GAT CTT CTT G
xWGEF_L407A_FP	CAC CTT CTC CGC GTG GCA GGA TAT
xWGEF_L407A_RP	ATA TCC TGC CAC GCG GAG AAG GTG
xWGEF_W408A_FP	CCT TCT CCC TGG CGC AGG ATA TC
xWGEF_W408A_RP	GAT ATC CTG CGC CAG GGA GAA GG
xWGEF_Q409A_FP	CTC CCT GTG GGC GGA TAT CCC AGA T
xWGEF_Q409A_RP	ATC TGG GAT ATC CGC CCA CAG GGA G
xWGEF_D410A_FP	CCT GTG GCA GGC TAT CCC AGA TG
xWGEF_D410A_FP	CAT CTG GGA TAG CCT GCC ACA GG
xWGEF_I411A_FP	GTG GCA GGA TGC CCC AGA TGT GC
xWGEF_I411A_RP	GCA CAT CTG GGG CAT CCT GCC AC
xWGEF_P412A_FP	GGC AGG ATA TCG CAG ATG TGC GG
xWGEF_P412A_RP	CCG CAC ATC TGC GAT ATC CTG CC
xWGEF_Y353E_FP	TTG AAC TCC ATT CTC GAG CAG GAA TAC AGT GA
xWGEF_Y353E_RP	TCA CTG TAT TCC TGC TCG AGA ATG GAG TTC AA
Dvl^{PDZ}_stop_FP	CTG TGG CCA AGT GCT GAG AAA CTT CTG TT
Dvl^{PDZ}_stop_RP	AAC AGA AGT TTC TCA GCA CTT GGC CAC AG

Abstract

Name of the Student: Omble Aishwarya Dilip **Registration No. :** 10BB18A26034
Faculty of Study: Biological Sciences **Year of Submission:** 2024
AcSIR academic centre/CSIR Lab: CSIR-National Chemical Laboratory, Pune
Name of the Supervisor(s): Dr. Kiran Kulkarni
Title of the thesis: Structural basis of *Weak-similarity Guanine nucleotide Exchange Factor* (WGEF) activation in Wnt-planar cell polarity pathway

The Wnt-planar cell polarity (Wnt-PCP) pathway regulates cell polarity & migration during embryo development and adult tissue homeostasis. This pathway is dysregulated in many pathological conditions, including cancer and autoimmune disorders. The signalling events in this pathway gets triggered upon binding of secreted glycoproteins, Wnts, to their membrane receptors Frizzleds, which leads to the recruitment of a scaffold protein, *Dishevelled2* (Dvl2). This signal cascade further propagates to activate Rho GTPase, RhoA, through its effector *Weak-similarity Guanine nucleotide Exchange Factor* (WGEF). Despite extensive studies, the molecular basis of WGEF mediated RhoA activation, triggered by the interaction of the former with Dvl2, is not clear. To elucidate the mechanism of Wnt-PCP mediated RhoA activation, an *in vitro* complex of xWGEF and RhoA was formed and it was subjected to extensive crystallization trials. Due to the intrinsic instability of the complex, the structural studies on this met with limited success. Furthermore, WGEF is known to exist in an autoinhibitory state. The PDZ domain of Dishevelled2 ($Dvl2^{PDZ}$) binds and activates WGEF by releasing it from its autoinhibitory state. However, the actual $Dvl2^{PDZ}$ binding site of WGEF and the consequent activation mechanism of the GEF have remained elusive. Using biochemical and MD studies, we show that a unique "internal-PDZ binding motif" (IPM) of WGEF mediates the WGEF-$Dvl2^{PDZ}$ interaction to activate the GEF. The residues at P_2, P_0, P_{-2} and P_{-3} positions of IPM play an important role in stabilizing the $WGEF^{pep}$-$Dvl2^{PDZ}$ interaction. Furthermore, MD simulations of modelled $Dvl2^{PDZ}$-$WGEF^{IPM}$ peptide complexes suggest that WGEF-$Dvl2^{PDZ}$ interaction may differ from the reported $Dvl2^{PDZ\text{-}IPM}$ interactions. Additionally, the *apo* structure of human $Dvl2^{PDZ}$ shows conformational dynamics different from its IPM peptide bound state, suggesting an induced fit mechanism for the $Dvl2^{PDZ}$-peptide interaction. The current study provides a model for Dvl2 induced activation of WGEF. To study the WGEF-$Dvl2^{PDZ}$ interaction in the context of Wnt signalling, we established a cellular system. The Wnt-PCP pathway was established in HEK293 cells by transfecting these cells with plasmids carrying genes encoding the proteins that operate in this pathway. The installation of the pathway was confirmed by observing the formation of stress fibers on activation of the pathway using Wnt5A conditioned medium. Thus, the work reported here provides novel insights into the molecular basis of WGEF-RhoA activation and also provides a system to explore the intricacies of the pathway using cell based assay. The outcome of the study has significant implications for designing the therapeutic agents that target the Wnt-PCP pathway.

List of publications emanating from the thesis work

1. Omble, A., Mahajan, S., Bhoite, A., & Kulkarni, K. (2024). Dishevelled2 activates WGEF via its interaction with a unique internal peptide motif of the GEF. Communications Biology, 7(1), 543.
2. Omble, A., & Kulkarni, K. (2022). GPCRs that *Rh* oar the Guanine nucleotide exchange factors. Small GTPases, 13(1), 84-99.

List of poster presentations with abstract

Propagation of Wnt- Frizzled signalling leading to the activation of RhoA by Weak similarity guanine nucleotide exchange factor (WGEF)

Aishwarya Omble and Kiran Kulkarni*

50th National Seminar on Crystallography, 2023

CSIR-IMTECH, Chandigarh

Wnt-Frizzled signalling controls diverse cellular processes including embryogenesis, morphogenesis and homeostasis. This signalling operates in two modes: the β-catenin dependent Wnt canonical pathway and the β-catenin independent Wnt non-canonical pathway. The latter pathway is involved in regulating the planar cell polarity based cell migration. Abrasions in this pathway are implicated in many pathological conditions like cancer and autoimmune disorders. Cellular dynamics, including cell polarity and migration require activation of Rho family GTPases; the master regulator of many cellular processes, including cell migration. In case of Wnt11 controlled planar cell polarity pathway, the frizzled7 receptor (Fz7) activates a Rho GTPase RhoA, through *Weak similarity guanine nucleotide exchange factor* (WGEF). The Wnt-Fz7 specific adaptor protein dishevelled2 (Dvl2) transduces the signal from the Fz7 receptor to WGEF. Due to the lack of structural knowledge, the mechanism of these signalling events is not clear. Here, using biochemical, crystallographic and computational studies we probe how Dvl2 interacts with WGEF to release it from its autoinhibitory state. This action further propagates the migratory signal to activate RhoA to bring about cell motility.

Pull down assay to study interaction between Dishevelled (Dvl) protein and Weak-similarity guanine nucleotide exchange factor (WGEF)

Aishwarya Omble and Kiran Kulkarni*

International Symposium on Cell Surface Macromolecules 2020

IISER, Pune

Dishevelled (Dvl) protein plays an important role in wnt signalling pathway. This protein acts as a signalosome which relays signals from the Frizzled receptor in the plasma membrane to different downstream cytoplasmic effectors. The activation of Dvl protein is mainly regulated by phosphorylation, ubiquitination and degradation. This protein consists of 3 main domains namely; DIX (Dishevelled, Axin) domain containing 80 amino acids, the middle domain is PDZ (Postsynaptic density 95, Discs large, Zona occludens-1 domain) containing 90 amino acids and a DEP (DvL, Egl-10, Pleckstrin) domain of 80 amino acids.

Dvl protein interacts with multiple proteins to carry out different functions. In the canonical wnt signalling pathway when Dvl is activated and recruited by Frizzled receptor, it prevents the destruction of cytosolic beta catenin. In non-canonical Wnt Planar cell polarity (PCP) signalling it relays signals to various effector proteins leading to the activation of GTPases like Rac, RhoA which leads to actin cytoskeleton rearrangement finally leading to cell migration. It has been reported that Weak similarity Guanine nucleotide exchange factor (WGEF) protein, which activates RhoA interacts with Dvl through the PDZ domain of Dvl. But, no evidence is reported showing the molecular level interaction of WGEF with DIX, PDZ or DEP domains of Dvl.

Our aim is to study whether Dvl directly interacts with WGEF and if so, to map the WGEF interacting domain of Dvl at molecular level. We prepared different constructs of DIX, PDZ and DEP domains. Standardization for purification of the DEP domain is done and is under study for DIX and PDZ domains. Attempts are being done to pull down studies of these domains with heterologously expressed WGEF in order to find its direct interaction with Dvl protein. Using X- ray crystallography we aim to delineate the mechanism by which Dvl protein contributes in activation of WGEF. This finding will help to gain knowledge regarding how dishevelled relays signal via the Daam1-RhoA axis and bring about cell migration.

communications biology

Article

https://doi.org/10.1038/s42003-024-06194-6

Dishevelled2 activates WGEF via its interaction with a unique internal peptide motif of the GEF

Aishwarya Omble[1,2], Shrutika Mahajan[1], Ashwini Bhoite[1,2] & Kiran Kulkarni[1,2]

The Wnt-planar cell polarity (Wnt-PCP) pathway is crucial in establishing cell polarity during development and tissue homoeostasis. This pathway is found to be dysregulated in many pathological conditions, including cancer and autoimmune disorders. The central event in Wnt-PCP pathway is the activation of *Weak-similarity guanine nucleotide exchange factor* (WGEF) by the adapter protein Dishevelled (Dvl). The PDZ domain of Dishevelled2 ($Dvl2^{PDZ}$) binds and activates WGEF by releasing it from its autoinhibitory state. However, the actual $Dvl2^{PDZ}$ binding site of WGEF and the consequent activation mechanism of the GEF have remained elusive. Using biochemical and molecular dynamics studies, we show that a unique "internal-PDZ binding motif" (IPM) of WGEF mediates the WGEF-$Dvl2^{PDZ}$ interaction to activate the GEF. The residues at P_2, P_0, P_{-2} and P_{-3} positions of IPM play an important role in stabilizing the $WGEF^{pep}$-$Dvl2^{PDZ}$ interaction. Furthermore, MD simulations of modelled $Dvl2^{PDZ}$-$WGEF^{IPM\ peptide}$ complexes suggest that WGEF-$Dvl2^{PDZ}$ interaction may differ from the reported $Dvl2^{PDZ}$-IPM interactions. Additionally, the *apo* structure of human $Dvl2^{PDZ}$ shows conformational dynamics different from its IPM peptide bound state, suggesting an induced fit mechanism for the $Dvl2^{PDZ}$-peptide interaction. The current study provides a model for Dvl2 induced activation of WGEF.

Wnt signalling pathways regulate diverse cellular functions during embryonic development and tissue homoeostasis[1]. The central event in Wnt pathways involves activation of hepta helical membrane receptors, Frizzleds, by the secreted glycoproteins, Wnts. The signalling further propagates on the subsequent recruitment of a scaffold protein, Dishevelled (Dvl), by the C-terminal target binding domain of the Frizzled receptors. This signal further diverges into two modes; canonical β-catenin-dependent signalling and non-canonical β-catenin-independent signalling. The canonical β-catenin-dependent signalling promotes the expression of target genes by facilitating the nuclear translocation of β-catenin and subsequent activation of TCF/LEF transcription factors associated with those target genes. This mode of Wnt signalling is majorly involved in cellular processes like cell proliferation[2]. On the other hand, the non-canonical β-catenin-independent signalling is further subdivided into Wnt/Calcium and Wnt-planar cell polarity (Wnt-PCP) pathways[3,4]. These non-canonical pathways are shown to control cell polarity and migration[2]. Aberrations in the Wnt signalling pathways are implicated in many pathological conditions, including cancer, developmental and autoimmune disorders[5,6].

One of the key features of Wnt-PCP pathway is the regulation of actin cytoskeleton reorganization through the activation of Rho GTPases, namely RhoA, Rac and Cdc42[7]. Rho GTPases function as molecular switches that cycle between GTP bound ON (active) and the GDP bound OFF (inactive) states[8]. This switching between ON and OFF states is brought about by the regulatory proteins known as Guanine nucleotide exchange factors (GEFs) and GTPase activating proteins (GAPs), respectively[9]. Appropriate signalling cues, transducing from the membrane receptors such as G-protein coupled receptors (GPCRs)[10], receptor tyrosine kinase[11] and integrin receptors[12], activate their downstream GEFs, which in turn trigger Rho GTPases by exchanging their bound GDP with GTP. However, the majority of GEFs exist in an autoinhibitory state and require other signalling proteins for their activation[10]. In the context of Wnt-PCP pathway WGEF (Weakly similar to RhoGEF 5/TIM), also known as Arhgef19/ Ephexin2), activates RhoA[13,14]. In *Xenopus*, WGEF-RhoA exerts convergent extension during gastrulation. Thus, marking the Wnt-PCP/RhoA pathway as one of the critical pathways involved in developmental and tissue regeneration processes in the higher eukaryotes[15–18]. Similar to canonical DH-PH GEFs, like TIM and Ephexin, WGEF consists of the conserved catalytic Dbl homology

[1]Division of Biochemical Sciences, CSIR-National Chemical Laboratory, Dr. Homi Bhabha Road, Pune 411008, India. [2]Academy of Scientific and Innovative Research (AcSIR), Ghaziabad 201002, India. e-mail: ka.kulkarni@ncl.res.in

(DH) domain responsible for its GEF activity and the Pleckstrin homology (PH) domain, which is involved in the membrane localization of the GEF[13]. Furthermore, WGEF consists of a C-terminal Src homology 3 (SH3)[13] and an N-terminal inhibitory (NID) domain. It has been shown that the phosphorylation of a conserved tyrosine in the NID region is critical in the activation of the GEF[19–21]. Besides NID, the SH3 domain also contributes to the autoinhibition of the GEF through its intra- and intermolecular interactions[21–23].

In the context of the Wnt-PCP pathway, the adapter protein Dishevelled is shown to be involved in releasing the autoinhibitory state of WGEF[17]. Dishevelled (Dvl), which has three paralogs in humans, viz., Dvl1, Dvl2 and Dvl3, operate as a signalling hub in exerting both canonical and non-canonical Wnt signalling pathways[24,25]. This protein consists of three conserved domains, namely, an N-terminal DIX (Dishevelled-Axin) domain, a central PDZ (PSD-95, DLG, ZO1) domain and a C-terminal DEP (Dishevelled, EGL-10, Pleckstrin) domain[26]. These domains play diverse roles during the Wnt signalling pathways[27–30]. Regarding the activation of WGEF, the PDZ domain of Dishevelled2 (Dvl2) is shown to be the key player. Dvl2 binds to WGEF through its PDZ domain to release the GEF from its autoinhibitory state[17]. However, the exact region of WGEF that interacts with $Dvl2^{PDZ}$ and its mode of interaction with Dvl2 have remained elusive. In the present study, we have identified an evolutionary conserved $Dvl2^{PDZ}$ binding motif in WGEF (hereafter referred to as $WGEF^{pep}$), which facilitates the activation of the GEF by disrupting its autoinhibitory state. Using biochemical assays and molecular dynamics simulation studies, we have elucidated the mode of $Dvl2^{PDZ}$–WGEF interaction and its consequence on the activation of the GEF.

Results

A novel N-terminal conserved 'internal peptide' motif of WGEF mediates its interaction with $Dvl2^{PDZ}$

Previously, Igor et al. have shown that the PDZ domain of Dvl2 binds to the N-terminal domain of WGEF, resulting in activation of the GEF. Further, deletion of the first 213 residues from the N-terminal of human WGEF ($hWGEF^{\Delta 213}$) has been shown to reduce its affinity for Dvl2 considerably but at the same time to significantly enhance its GEF activity. Thus, it was suggested that the N^{213} region of the GEF could be involved in WGEF–$Dvl2^{PDZ}$ interaction[17]. In another study, $hWGEF^{\Delta 302N\text{-}term}$ mutant is shown to be free from autoinhibition, suggesting that its PDZ binding motif precedes the autoinhibitory domain of the GEF (spanning residues 291-NSVLYQEY-298)[17,21]. However, the exact motif of WGEF that binds to $Dvl2^{PDZ}$ is unclear from these studies. Like other PDZ domains, the $Dvl2^{PDZ}$ also exhibits promiscuity in terms of the peptide motifs that it binds to. The peptide motifs that PDZ domains recognize are primarily *C-terminal peptide motifs*, which are composed of a stretch of 7–10 residues situated at the C-terminus of the binding partners. Apart from these, some PDZ domains can also recognize the *internal motifs* that possess a similar number of residues but can lie anywhere in their binding partners[31–33]. Binding of PDZ with internal motifs mimic that of the C-terminal peptide binding with slight variations in the residues of the peptide involved in the binding[33,34].

Although not very stringent, there is some consensus in the sequences of *C-terminal peptide motifs* that bind to PDZ domains[31,32]. However, there is no statistically significant data to arrive at a consensus sequence for the PDZ domain binding *internal peptide motifs*. Therefore, to identify the $Dvl2^{PDZ}$ interacting *internal motif* of WGEF, we searched the sequences of this family of GEFs with the putative $Dvl2^{PDZ}$ binding motifs identified from peptide-phage display studies[35]. We employed three prominent peptide motifs, X-Y-G-W-Φ^a-D/G, X-W-Φ^a-D-G-P and W-Φ^s-D-X-P (where X, Φ^a and Φ^s are any, aliphatic and short side chain hydrophilic amino acids [S/T], respectively), as templates for the search. The only hit, 349-GSTFSLWQDIP-359 (hereafter referred to as $WGEF^{pep}$), obtained from the sequence search, lies between the DH and the inhibitory domain of hWGEF (Fig. 1a). This particular motif is highly conserved amongst WGEFs family, indicating it to be the most putative $Dvl2^{PDZ}$ binding site present in WGEFs (Fig. 1a and Supplementary Fig. 1). This observation is contrary to the previous study in which the $Dvl2^{PDZ}$ binding was mapped towards the N-terminus of hWGEF (between amino acid residues 1–213)[17], whereas the motif identified by us is present between amino acid residues 349–359 of hWGEF. Hence, to validate our observations, we designed constructs of the GEF comprising both the autoinhibitory and the $Dvl2^{PDZ}$ binding motif and yet excluding the major portion of its N-terminus. Since some of these hWGEF constructs were found to be insoluble and/or poorly expressed in *E. coli*, we used WGEF from *Xenopus laevis* (xWGEF) as a surrogate system for further studies. To test the influence of $Dvl2^{PDZ}$ binding on the catalytic activity of WGEFs, we performed nucleotide exchange assays on these constructs, both in the presence and absence of the PDZ domain. Clearly, the xWGEF330 (equivalent to hWGEF272) construct, which is composed of both the inhibitory and the PDZ binding domains, showed insignificant GEF activity. However, in the presence of 10-fold excess (50 μM) of $Dvl2^{PDZ}$, there is a fourfold increase in its GEF activity (Fig. 1b, c). It is worth noting that the lower concentration of $Dvl2^{PDZ}$ (25 μM) did not show any observable change in the GEF activity. These observations suggest that the binding of $Dvl2^{PDZ}$ indeed promotes the activity of GEF; however, the affinity of the xWGEF330 for $Dvl2^{PDZ}$ is considerably low (Fig. 1b, c). Next, to show that the identified internal peptide motif of WGEF, 402(349)-GSTFSLWQDIP-412(359)/ $WGEF^{pep}$ [residue numbers shown in the parentheses corresponds to hWGEF], binds to $Dvl2^{PDZ}$, we performed the GEF assays in the presence of both $Dvl2^{PDZ}$ and $WGEF^{pep}$. Since the internal peptides are shown to have relatively low affinity for PDZ domains, we used a fourfold molar excess of $WGEF^{pep}$ (200 μM) to disrupt the WGEF–$Dvl2^{PDZ}$ interaction (Fig. 1b, c). From this exercise, it is clear that $WGEF^{pep}$ competitively binds to the PDZ domain and replaces the WGEF in $Dvl2^{PDZ}$–WGEF interaction, causing a decrease in the catalytic activity of WGEF. Thus, our observations unambiguously suggest that the 402–412 region of xWGEF is the 'internal peptide motif' that mediates its interactions with $Dvl2^{PDZ}$.

$WGEF^{pep}$ motif–$Dvl2^{PDZ}$ interaction is sufficient to partially release the autoinhibition of WGEF

Canonically, PDZ domains bind to *C-terminal* peptides, and therefore, the residue at the extreme C-terminal of the peptide is referred to as the 0th or P_0 residue. The remaining residues towards the N-terminus of the peptide are numbered in reverse order as '−1', '−2', '−3' or P_{-1}, P_{-2}, P_{-3}, and so on[36–38]. $Dvl2^{PDZ}$ binds to both *C-terminal* as well as the non-canonical *internal peptides*[35,39]. Therefore, we will refer to the residue that interacts with the X-Φ-G-Φ motif of the carboxylate binding loop (X represents any amino acid and Φ represents hydrophobic residues)[31] as P_0, followed by P_{-1}, P_{-2}, P_{-3} for residues preceding P_0 and P_1, P_2, P_3 for the residues that lie towards the C-terminus.

Zhang et al. have reported earlier that the side chain of aspartate residue present in the internal peptide mimics the free C-terminal carboxylate group and binds to the carboxylate binding loop of $Dvl2^{PDZ}$. Thus, this particular interaction is asserted to stabilize the $Dvl2^{PDZ}$-effector binding[35]. Therefore, to test whether the $WGEF^{pep}$ follows this "usual" internal peptide–$Dvl2^{PDZ}$ interaction, we substituted the aspartate present at the P_0^{th} (Fig. 1d) position of the peptide with alanine ($WGEF^{pepD9A}$). Interestingly, with an excess amount of $WGEF^{pepD9A}$ peptide, there was a decrease in the exchange activity of WGEF, but not to an extent brought out by the wild-type $WGEF^{pep}$ (Fig. 1b, c). We performed a comparable exercise on GEF by replacing this particular aspartate ($xWGEF330^{D410A}$) in xWGEF construct. Since this xWGEF mutant was found to be insoluble, no further studies could be conducted. Other studies also suggest that proline, tryptophan and leucine at P_2, P_{-2} and P_{-3}, positions, respectively, also play a key role in stabilizing $Dvl2^{PDZ}$–effectors interactions[35]. To test the role of these residues when they are part of the GEF, we produced $xWGEF330^{P412A}$, $xWGEF330^{W408A}$ and $xWGEF330^{L407A}$ mutants; however, only the $xWGEF330^{L407A}$ was found to be soluble. Surprisingly, the enhanced GEF activity of $xWGEF330^{L407A}$ is independent of its interaction with $Dvl2^{PDZ}$ (50 μM) (Fig. 1b, c), as the addition of $Dvl2^{PDZ}$ (50 μM) does not

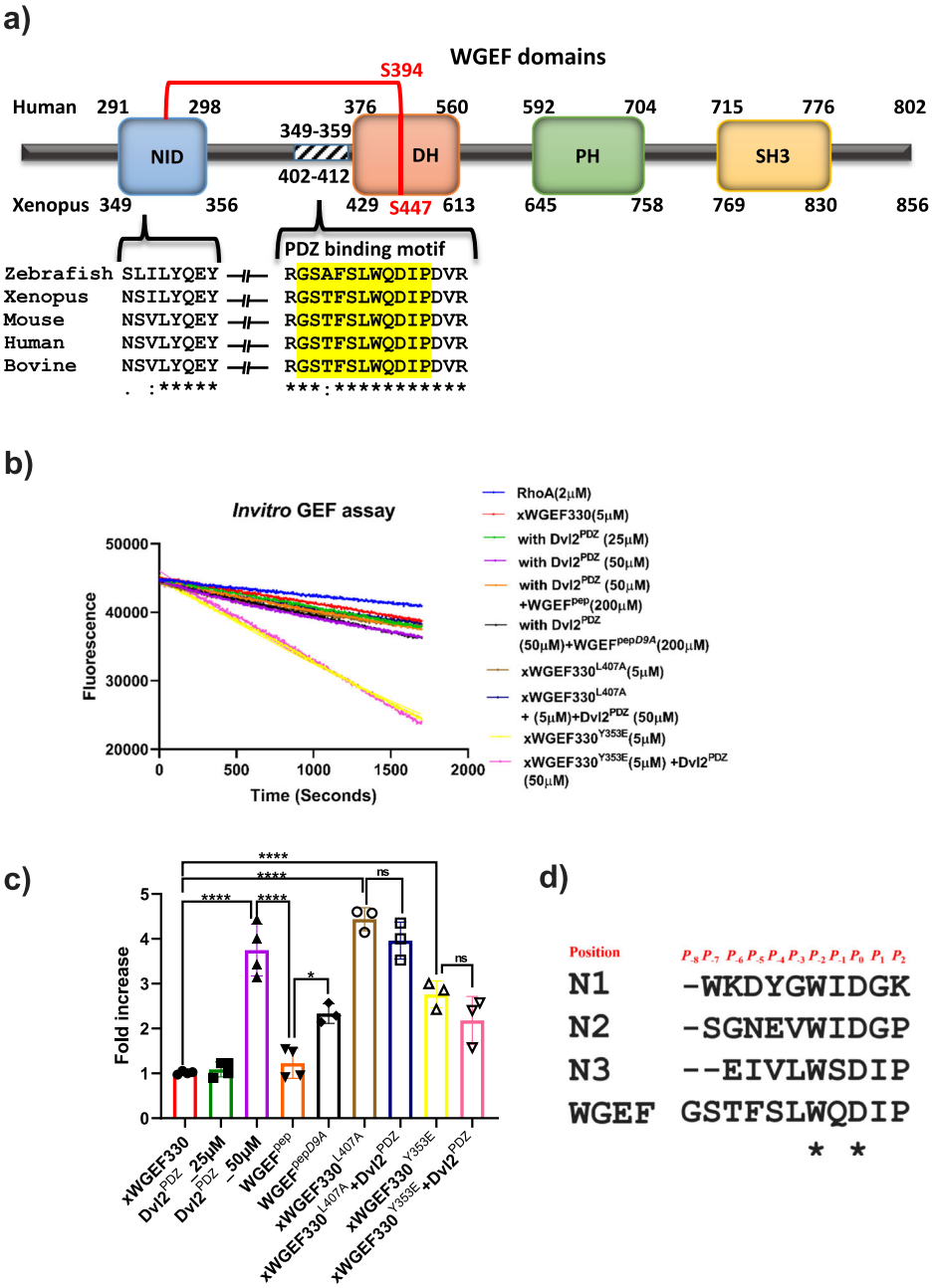


Fig. 1 | Activity studies of xWGEF330 and xWGEF330 mutants. a Schematic representation of WGEF domains (N-terminal Inhibitory: NID, Dbl homology: DH, pleckstrin-homology: PH and src Homology-3: SH3) and sequence alignment of the conserved PDZ binding motif of Zebrafish (Uniprot ID: E7EY470), Xenopus (Uniprot ID: A5X5J0), Mouse (Uniprot ID: Q8BWA8), Human (Uniprot ID: Q8IW93-1) and Bovine (Uniprot ID: E1BQ24) ahead of DH domain in WGEF. The amino acid residue numbers in the upper and lower parts of the diagram correspond to hWGEF and xWGEF, respectively. **b** GEF assay plots of xWGEF330, xWGEF330^{L407A} and xWGEF330^{Y353E}, in the presence and absence of Dvl2^{PDZ} and peptide variants **c** Bar graph showing relative GEF activities of xWGEF variants normalized against that of wild-type (autoinhibited xWGEF330) protein. The error bar and * represent standard deviation (±SD) and significance as $^{****}P < 0.0001$, $^{***}P < 0.001$, $^{**}P < 0.01$ and $^{*}P \leq 0.01$, respectively. **d** Sequence alignment of Dvl2^{PDZ} binding internal peptides and WGEF internal peptide sequence. Residue positions are labelled in red.

seem to affect its GEF activity. The higher GEF activity observed for xWGEF330^{L407A} could be due to the structural rearrangement caused by the mutation. However, from our experiments it is not clear how L407A substitution influences the autoinhibition of the GEF. Thus, this aspect requires further investigations.

Next, to investigate whether the WGEF–Dvl2^{PDZ} interaction promotes the GEF activity by releasing its autoinhibition or it follows some unknown activation mechanism, we generated a Y353E mutant of xWGEF330. This particular tyrosine (Y$^{353xWGEF/295hWGEF}$) of hWGEF is present in NID and locks the GEF by interacting with its DH domain. Furthermore, hWGEF^{Y295E} is shown to exist in autoinhibition-free form[21]. Interestingly, the xWGEF330^{Y353E} mutant exhibited a nearly threefold increase in the GEF activity compared to its wild-type form. Further, the presence of Dvl2^{PDZ} (50 μM) did not alter the GEF activity of the mutant (Fig. 1b, c). Thus, these results suggest that the Dvl2^{PDZ} binds to the WGEF^{pep} region of the GEF, lying between its inhibitory (NID) and the DH domains. Furthermore, this particular interaction between WGEF and Dvl2^{PDZ} activates the GEF by partially releasing it from its autoinhibitory state regulated by the NID. The mechanism through which the autoinhibition governed by the SH3 domain is released remains unknown.

Binding of WGEF^{pep} with Dvl2^{PDZ} is mediated by the residues at the P_{-2}, P_0 and P_2 positions of the binding motif

Although the WGEF^{pep} region is highly conserved in the WGEF family (Supplementary Fig. 1), the interaction between WGEF and Dvl2^{PDZ} appears to be feeble. This aligns with the observations made for other PDZ–internal peptide interactions[33,35]. Perhaps differential affinity of the Dvl with its effectors is one of the strategies adopted by Wnt-Frizzled pathways to exert diverse signalling modes[25,30]. Although our GEF assay-based studies have shown that, indeed, the residue at the P_0 position of WGEF^{pep} plays a crucial role in mediating WGEF–Dvl2^{PDZ} binding, it is not clear how variations in other residue positions influence the interaction. Therefore, to obtain the structural basis of Dvl2^{PDZ}–WGEF interaction, initially, we tried to crystallize this protein complex. Despite extensive efforts, crystallization trials met with no success. Hence, we resorted to crystallizing the WGEF^{pep} (peptide) in complex with Dvl2^{PDZ}. However, even this attempt did not yield

Table 1 | Data collection and refinement statistics for the structures of human Dishevelled2 PDZ domain, co-crystallized and fused with WGEF^{pep}, respectively

	hDvl2^{PDZ} crystallized with WGEF^{pep}	*hDvl2^{PDZ}* fused with WGEF^{pep}
Data collection		
Source	ID23-2 beamline of ESRF	BL21-PX beamline of Indus-2 synchrotron
Space group	$P4_32_12$	$I\,4_1$
a (Å)	45.98	59.49
b (Å)	45.98	59.49
c (Å)	78.42	58.25
α (°)	90	90
β (°)	90	90
γ (°)	90	90
Resolution limits (Å)	32.51–1.75	42.07–3.00
R_{merge}	0.067(0.731)	0.058(1.091)
I/s (I)	23.3(3.8)	27.8(2.7)
Number of reflections	157,526(7686)	30,769(5090)
Unique reflections	9023(482)	2071(336)
Completeness (%)	100(100)	99.9(100)
Multiplicity	17.5(15.9)	14.9(15.1)
CC(1/2)	0.999(0.915)	1(0.731)
Refinement		
Resolution limits (Å)	32.51–1.75	29.75–3.00
Number of reflections	8984	2067
Working set	8594	1951
Test set	390	116
R_{work}/R_{free}	0.227/0.252	0.299/0.335
Number of atoms		
Protein	1245	512
Water	20	
B factors		
Protein atoms (Å^2)	39.26	114.32
Water	37.25	
RMSD from ideal values		
Bond length (Å)	0.007	0.009
Bond angles (°)	1.089	1.361
Ramachandran plot		
Preferred (%)	97.53	87.32
Allowed (%)	2.47	12.68
PDB code	8WWR	8YR7

Single crystal was used for data collection. Values in parenthesis are for the highest-resolution shell.

the desired result, as there was no electron density for the bound peptide in the crystal structure of the WGEF^{pep}–Dvl2^{PDZ} complex (PDB ID: 8WWR) (Table 1). Furthermore, we tried to obtain the WGEF^{pep}–Dvl2^{PDZ} complex structure by fusing the peptide with the C-terminus of the PDZ domain. Even this approach failed to produce the WGEF^{pep}–Dvl2^{PDZ} complex structure (PDB ID: 8YR7) (Supplementary Fig. 2 and Table 1). Therefore, we systematically designed divergent peptides of WGEF^{pep} and biochemically probed their interaction with Dvl2^{PDZ}. The closest homologue of WGEF^{pep} is an engineered peptide, N3^{pep}, whose structure in complex with Dvl2^{PDZ} is available (PDB ID: 3CC0). Hence, by using this structure as template, we designed WGEF^{pep} variants by successively substituting its residues from P_{-3} to P_2 positions with alanine (Fig. 1d), as these residues are seen to be lying within 5 Å from Dvl2^{PDZ} residues and they are conserved between WGEF^{pep} and N3^{pep}. To assess the differential contributions from these residues of the peptide, we determined the dissociation constants of the WGEF^{pep} variants for Dvl2^{PDZ}, employing Microscale Thermophoresis (MST). In addition to the WGEF^{pep} variant peptides the minimal hWGEF (330–581) construct, obtained from our solubility screening, was used for the MST study (Fig. 2a, b, d and Supplementary Fig. 3). It is interesting to note that the affinities of WGEF^{pep} and N3^{pep} are nearly identical for Dvl2^{PDZ}, with dissociation constants (K_d) of 45 ± 5 and 54 ± 7 μM, respectively (Fig. 2a, d and Table 2). This further validates the choice of our model for delineating WGEF–Dvl2^{PDZ} interaction. Substitution of Gln at the P_{-1} (Fig. 1d) with Ala (WGEF^{pepQ8A}) enhances the affinity of the peptide by almost 6-fold, whereas Ile at P_1 to Ala (WGEF^{pepI10A}) does not seem to affect the WGEF^{pep}–Dvl2^{PDZ} interaction (Fig. 2b, d and Table 2). It is interesting to note that amongst these two residues, only the side chain of Ser at P_{-1} makes direct interaction with Dvl2^{PDZ} in the Dvl2^{PDZ}–N3^{pep} complex crystal structure. On the contrary, the substitution of Leu at the P_{-3} position (WGEF^{pepL6A}) significantly reduces the affinity of the peptide for Dvl2^{PDZ} (Fig. 2c, d and Table 2). Furthermore, the substitution of tryptophan (WGEF^{pepW7A}), aspartate (WGEF^{pepD9A}) and proline (WGEF^{pepP11A}) residues at the P_{-2}, P_0 and P_2 positions, respectively, with alanine completely abolishes the WGEF–Dvl2^{PDZ} interactions (Table 2 and Supplementary Table 1). Thus, MST studies suggest that the residues at P_{-2}, P_0 and P_2 positions (W, D & P) are crucial for the WGEF-Dvl2^{PDZ} binding, whereas leucine at the P_{-3} position may influence the peptide–Dvl2^{PDZ} interaction either by stabilizing the complex or by enhancing the solubility of the peptide.

Internal peptide ligands of Dvl2^{PDZ} exhibit divergent modes of interaction

From GEF activity and MST studies, we have identified the residues of the *internal peptide motif* that drive the WGEF–Dvl2^{PDZ} interaction. Furthermore, in both canonical and non-canonical (internal) peptides, the ligand adopts antiparallel β strand conformation and forms a β sheet by associating with the β2 strand of the PDZ domain[40]. Additionally, in the canonical binding mode, the C-terminus of the P_0 residue engages with the X-Φ-G-Φ motif of the carboxylate binding loop to stabilize the complex (Fig. 1a, d). However, in the non-canonical PDZ domain binding, the *carboxylate binding loop* may not adopt a closed conformation and instead engages with the internal peptide through a network of hydrogen bonds involving either the main chain or the side chain atoms of the peptide[35,37,41]. Due to the lack of structural information on the WGEF^{pep}–Dvl2^{PDZ} complex, their binding mode has remained unclear. Hence, to address this aspect, we performed a total of 9 μs molecular dynamics simulations of the Dvl2^{PDZ} structure, complexed with WGEF^{pep}, WGEF^{pepQ8A} and N3^{pep} peptides. Since for two of the former peptides, there was no crystal structure available, we modelled the respective Dvl2^{PDZ}–peptide complexes using the coordinates of the *apo* (PDB ID: 8WWR) (Table 1) and Dvl2^{PDZ}-N3^{pep} structures.

To understand the role of every residue in stabilizing the interaction, we plotted residue-wise peptide–PDZ domain interaction propensity (probability of the residue side-chain lying within 4 Å from the PDZ residue) (Fig. 3a–d). Propensity and the residue RMSF (Supplementary Fig. 4) plots obtained from MD simulations show that residues at positions P_{-2}, P_{-1}, P_0, P_1 and P_2 play significant roles in stabilizing the peptide–Dvl2^{PDZ} interaction. Amongst these, the aspartate residue of the P_0 position of both WGEF^{pep} and N3^{pep} forms maximum number of interactions with X-Φ-G-Φ motif of the *carboxylate binding loop* (Fig. 3a, b, e, f). Apart from Asp at the P_0 position, it was observed that residues at the positions P_{-1}, P_{-2} and P_{-3} also make interactions with the residues belonging to the β2-α2 region of Dvl2^{PDZ} (Fig. 3a, b, e, f). Interestingly, most interactions between WGEF^{pep}-Dvl2^{PDZ} are similar to those observed in the Dvl2^{PDZ}-N3^{pep} (Fig. 3a, b). However, subtle differences in the WGEF^{pep} residues at P_{-4} and P_{-5} exhibit divergence in their interaction with the PDZ domain, when compared with those of N3^{pep} interactions. In Dvl2^{PDZ}–N3^{pep} complex residues at P_{-4} and P_{-5} positions interact with β2 strand of Dvl2^{PDZ} domain, which is absent in WGEF^{pep}-Dvl2^{PDZ} simulated model (Fig. 3a, b, e, f). Perhaps that is why in

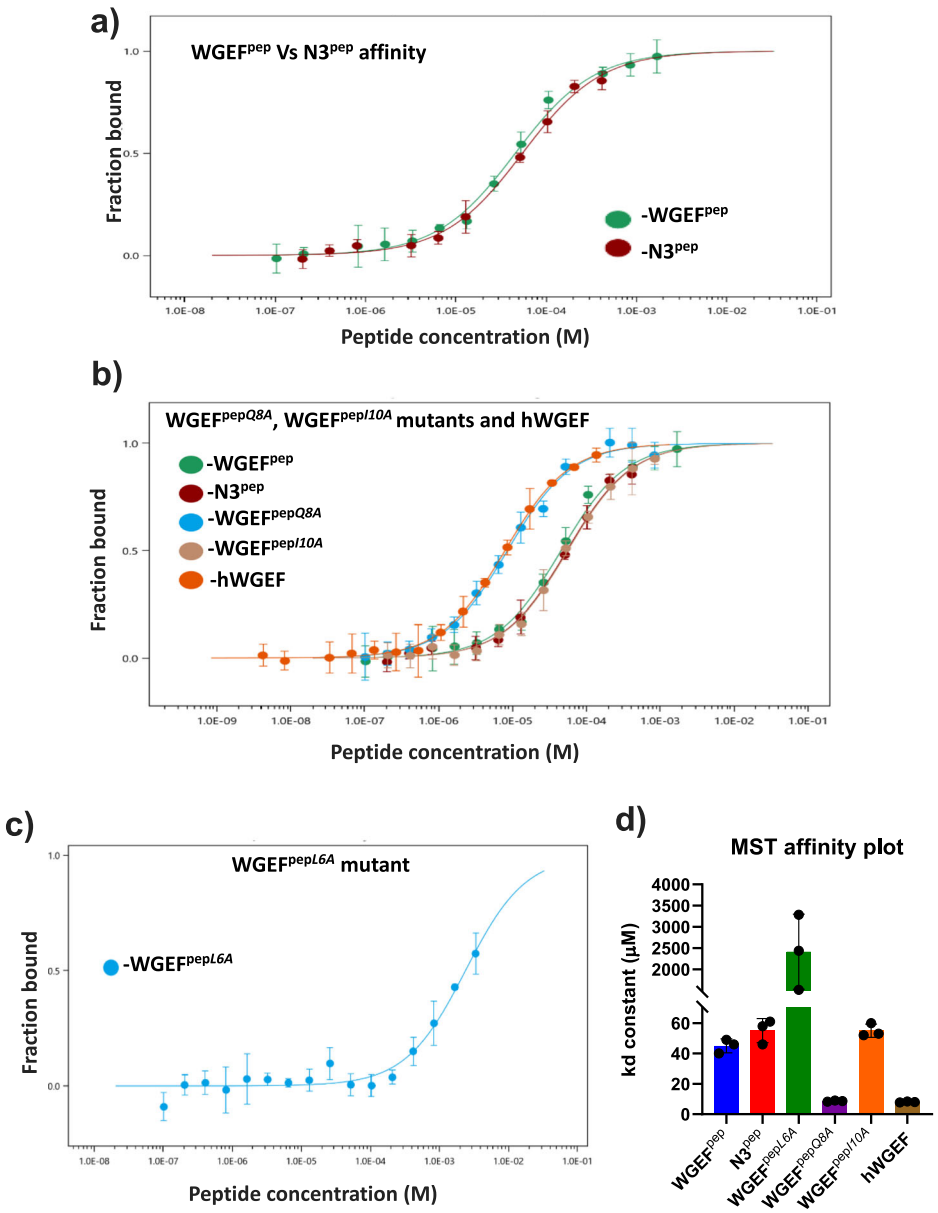


Fig. 2 | Binding studies of Dvl2^{PDZ} with different internal peptides and human WGEF. a Binding curve corresponding to the interaction of Dvl2^{PDZ} with WGEF^{pep} and N3^{pep}. **b** and **c** Binding curve corresponding to WGEF^{pepQ8A}, WGEF^{pepI10A} peptides, hWGEF$^{330-581}$, and WGEF^{pepL6A} peptide, respectively. **d** Bar plots showing dissociation constants between Dvl2^{PDZ} and different peptides. The error bars represent K_d confidence obtained from triplicates.

WGEF^{pep}-Dvl2^{PDZ} simulations, residues from P_{-4} onwards (towards N-terminus) exhibit higher RMSF values (Supplementary Fig. 4). Furthermore, the tryptophan at the P_{-2} position of N3^{pep} provides additional stacking interactions to stabilize the PDZ–peptide complex.

Our MST studies show that the substitution of Gln at P_{-1} with Ala (WGEF^{pepQ8A}) in WGEF^{pep} considerably increases the affinity of the mutant peptide towards Dvl2^{PDZ} by 6-folds (Fig. 2b, d). Interestingly, the MD simulations on the WGEF^{pepQ8A}–Dvl2^{PDZ} complex indicate a higher propensity for the Trp at the P_{-2} position to stay in the hydrophobic groove formed by Ile 280, Leu 278, Leu 337, and Val 341 residues at the β2-α2 region of the PDZ domain (Fig. 3c, d, g). However, the backbone conformation of this residue that favours the hydrophobic interaction with the PDZ domain is governed by the residue at the P_{-1} position. The presence of a residue with a longer side chain at P_{-1} could restrict the conformational space of Trp at P_{-2} and hence, may hinder the additional hydrophobic interactions between this residue and the PDZ domain (Fig. 3h). Thus, substitution of Gln with a shorter-side chain amino acid, like Ala, at P_{-1} position promotes WGEF^{pep}–Dvl2^{PDZ} interaction, as seen from MST studies. Thus, our MD studies provide insights into the Dvl2^{PDZ}–WGEF interaction and corroborate our biochemical studies on the Dvl2^{PDZ}–peptide complexes.

Furthermore, our MD studies and previously reported crystal structures of Dvl2-internal peptide complexes show no conserved mode of interaction between the internal peptides and Dvl2^{PDZ}. In the Dvl2^{PDZ}-N3^{pep} crystal structure, it is observed that N3^{pep} forms antiparallel β strand interaction with the β2 strand of the Dvl2^{PDZ} domain (Supplementary Fig. 5a). Thus, to assess whether there is conservation of β sheet formation of the internal peptides with the β2 strand of the Dvl2^{PDZ} domain, we calculated the secondary structural propensity of the peptide residues using their respective MD trajectories. In case of Dvl2^{PDZ} domain and internal peptide complexes (WGEF^{pep}, WGEF^{pepQ8A} and N3^{pep} peptides), it was observed that these ligands did not show consistency in forming β sheet with the β2 strand of the Dvl2^{PDZ} domain, as the ligand may not essentially adopt β strand conformation. Perhaps this loss of secondary structural conformation might aid the flexibility of the peptide and hence, enhance its interaction with Dvl2^{PDZ} (Supplementary Fig. 5b–d).

Dvl2^{PDZ} exhibits ligand-dependent conformational changes

Earlier studies have shown that certain PDZ domains like Erbin^{PDZ} and Syntrophin^{PDZ} exhibit limited flexibility during peptide interaction. Perhaps that is why Erbin^{PDZ} specifically accommodates C-terminal peptides with higher sequence specificity. However, Synrophin^{PDZ} can interact with both

Table 2 | Consolidated table showing dissociation constant (K_d) values for binding studies between different peptides, hWGEF protein and Dvl2^{PDZ}

Peptides	Sequence	K_d (μM)
WGEF^{pep}	GSTFSLWQDIP	45 ± 5
N3^{pep}	EIVLWSDIP	54 ± 7
WGEF^{pepL6A}	GSTFSAWQDIP	2400 ± 900
WGEF^{pepW7A}	GSTFSLAQDIP	ND
WGEF^{pepQ8A}	GSTFSLWADIP	8.5 ± 0.8
WGEF^{pepD9A}	GSTFSLWQAIP	ND
WGEF^{pepI10A}	GSTFSLWQDAP	56 ± 4.3
WGEF^{pepP11A}	GSTFSLWQDIA	ND
hWGEF	Residues from 330 to 581	7.9 ± 0.3

the *C-terminal* as well as the *internal peptides*, provided the peptide undergoes a conformational change to form a β finger, in order to be incorporated within the peptide binding loop, as seen in the structure of nNOS–Syntrophin^{PDZ} complex[42,43]. Another PDZ domain-containing protein, Par-6, displays allostery-mediated binding for C-terminal peptides, which exhibits an increase in affinity for the ligand by 10-folds upon binding of Cdc42 to its adjacent *Cdc42/Rac interactive binding* (CRIB) domain[44]. Par-6^{PDZ} domain binds to the internal peptide of Pals1, as well, through a conformational change in the PDZ domain, which is independent of allosteric alterations[37]. Thus, it is clear from previous studies that PDZ domains exhibit huge conformational dynamics to accommodate their binding motifs. Earlier efforts have tried to characterize these dynamics using MD simulations[42]. It is worth mentioning that the dynamical features of PDZ domains presented from different MD studies extensively vary[45]. This could be due to variations in the type of PDZ structures (*apo*, or bound with C-terminal and/or internal peptide) employed in the MD simulations and the simulation time[45,46]. Since we have elucidated the *apo* structure of human Dvl2^{PDZ} (PDB ID: 8WWR), we looked into the dynamics in its *apo* and different peptide-bound forms. It is important to mention that our hDvl2^{PDZ} *apo* structure differs from the reported *Xenopus* Dvl2^{PDZ} *apo* structures (PDB IDs: 3FY5 and 2F0A), with a substitution of just one amino acid (Human M^{323}/*Xenopus* I^{310}).

A comparison of available Dvl2^{PDZ} structures shows deviations with RMSD values in the range of 1.7–2 Å. Importantly, significant conformational variations are observed in the α2 helix and β2–β3 loop, which are part of the peptide binding pocket. This conformational variation is more prominent in the structures of human Dvl2^{PDZ}, bound with different peptides (Fig. 4a–c and Supplementary Fig. 5a). This opens up the question of whether the peptide binding pocket of Dvl2^{PDZ} is preformed or it exhibits peptide-induced conformational fit. To identify the amino acids that are dynamically coupled, we performed spectral decomposition of the *dynamic cross-coupling matrix (cij)* (Fig. 4d) obtained from all the MD simulations. As explained in the "Methods" section, the statistical coupling analysis (SCA) provides two main clusters of dynamically coupled amino acids, referred to as Sector 1 and Sector 2. These sectors can be further resolved into five *independent components* (ICs) (Supplementary Fig. 6). It is interesting to note that Sector 1 corresponding to all the simulations, which primarily consists of residues belonging to the peptide-binding pocket, show significant variations in terms of their residue composition and location of residues in the protein structure (Fig. 4e–h). Interestingly, Sector 1 of *apo* Dvl2^{PDZ} is composed of residues residing in β2 and α2 regions of the domain, whereas Sector 1 of Dvl2^{PDZ} internal peptide (WGEF^{pep}, WGEF^{pepQ8A} and N3^{pep}) complexes had residues from the β2 strand, β1–β2 and β2–β3 loops (Fig. 4e–h). These findings further validate the internal allostery of PDZ domains as elucidated by other studies[47]. Additionally, variations in Sector 1 composition highlight peptide-dependent conformational adaptation of Dvl2^{PDZ} domain. In contrast to the earlier study by Munz et al., we observe that the α2 helix of the *apo* Dvl2^{PDZ} structure has limited conformational flexibility rather than having the ability to explore all possible conformational space, specific to the ligand-bound forms. Perhaps, divergent observations of our studies from those published before could be due to the absence of a "true" *apo* structure of Dvl2^{PDZ}. Furthermore, the shorter duration of earlier MD simulation (200 ns)[42] might have exaggerated the conformational sampling of the PDZ domain. Thus, our structural and MD simulation studies suggest that Dvl2^{PDZ} adopts peptide-dependent conformational change due to its intrinsic flexibility. In short, we can infer that Dvl2^{PDZ} may not have a preformed binding pocket, rather it modulates the binding pocket based on the substrate available for binding. This adaptation is perhaps responsible for the peptide promiscuity of Dvl2^{PDZ}.

Discussion

Despite nearly three decades of studies on PDZ domains, their modes of interaction with their binding partners have remained elusive. It is quite intriguing to learn that despite having highly conserved domain structure, PDZ domains exhibit divergent modes of interaction with their binding partners[31,42,48]. Dishevelled, being a signalling hub, recognizes diverse binding partners through its different domains, especially involving its PDZ domain[49]. Therefore, as seen in other PDZ domains, Dvl2^{PDZ} exhibits significant promiscuity by recognizing binding partners having both *C-terminal* as well as *internal binding peptide* motifs[27,39,50]. Thus, it is expected that Dvl2^{PDZ} displays divergent binding modes with its partners[35]. In the context of Wnt-PCP signalling, Dvl2 assembles with the activated Frizzled receptors and triggers RhoA by activating its GEF, i.e., WGEF. The role of Dvl2 in this signal transduction event is to activate WGEF by releasing it from its autoinhibited state. This step involves interaction between the PDZ domain of Dvl2 and WGEF. However, the particular region of WGEF that mediates its interaction with Dvl2^{PDZ} was not known. Here, we have successfully identified an internal peptide motif of WGEF, lying between the N-terminal inhibitory and the DH domains, that facilitates the interaction of the protein with Dvl2 to activate the GEF. Indeed, this is one of the crucial mechanisms involved in releasing the autoinhibition of WGEF. Using MD simulations and comparative structural analysis, we show that Dvl2^{PDZ} exhibits diverse modes of ligand recognition, which is perhaps true of the WGEF^{pep}, as well. Compared to the other known Dvl2^{PDZ}–*internal peptide* interactions, in the WGEF-Dvl2^{PDZ} complex, only five residues at the P_2, P_0, P_{-1}, P_{-2} and P_{-3} positions appear to contribute significantly in stabilizing their interactions (Fig. 1d). Furthermore, in terms of consensus sequence, the substitution of P11A at the P_2 position of WGEF^{pep} makes the peptide homologues to the class III PDZ binding peptide. However, in terms of the mode of binding, they differ significantly. Class III peptide bound to nNOS^{PDZ} (PDB ID: 1B8Q)[51] happens to be relatively shifted towards the N-terminal of the α2 helix of the Dvl2^{PDZ} domain (Supplementary Fig. 7). Therefore, just based on the sequence of binding peptide mode of binding cannot be inferred. Additionally, from MD studies we observed that WGEF^{pep} and N3^{pep} peptides do not seem to adopt the stringent β strand conformation, as seen in case of the crystal structures of other PDZ-*internal peptide* complexes[35]. In short, the ligand (peptide) and the Dvl2^{PDZ} undergo conformational adaptation for mutual recognition. A genome-wide screen for internal PDZ binding motifs suggests that the ability of PDZ domains to bind internal peptides is much more prevalent than previously recognized[41]. To comprehend how Dvl2^{PDZ} binds to the identified binding motif within the structure of hWGEF, we looked at the AlphaFold predicted structure of hWGEF (Supplementary Fig. 8a). Clearly, major portion of the structure, including the helix that connects PDZ binding motif to the NID domain, is predicted with low confidence level (Supplementary Fig. 8b). Furthermore, the spatial disposition of different domains of the structure appears to be less reliable as the loops that connect these domains are predicted with low confidence level. However, to propose a hypothesis using this model, we superimposed the part of the hWGEF structure that is predicted with higher confidence on its closest structure homologue, Leukaemia-associated RhoGEF (PDB ID: 1X86). This comparison shows that in the absence of Dvl2^{PDZ} interaction, the GTPase

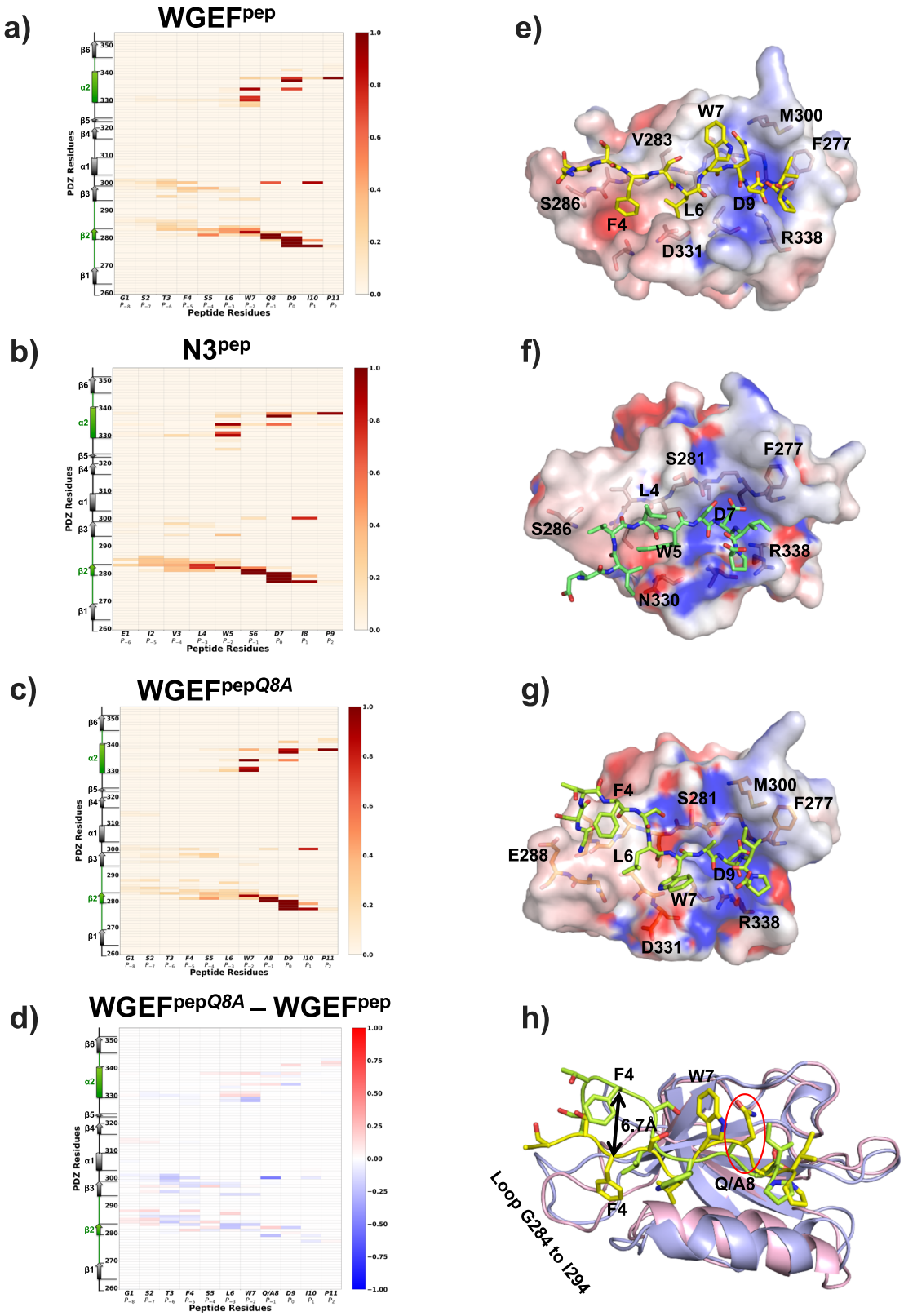


Fig. 3 | Interaction propensity of Dvl2^{PDZ} with peptide ligands. a Interaction propensity with WGEF^{pep}. **b** Interaction propensity with N3^{pep} and **c** interaction propensity with WGEF^{pepQ8A}. **d** Shows difference in the interaction propensity of WGEF^{pepQ8A} and WGEF^{pep}. **e–g** show the structure of Dvl2^{PDZ} (represented as surface) bound with peptide variants, WGEF^{pep}, N3^{pep} and WGEF^{pepQ8A} (represented in ball-and–sticks), belonging to one of the stable trajectories, respectively. **h** Superimposition of WGEF^{pep} (yellow peptide) and WGEF^{pepQ8A} (green peptide) structures.

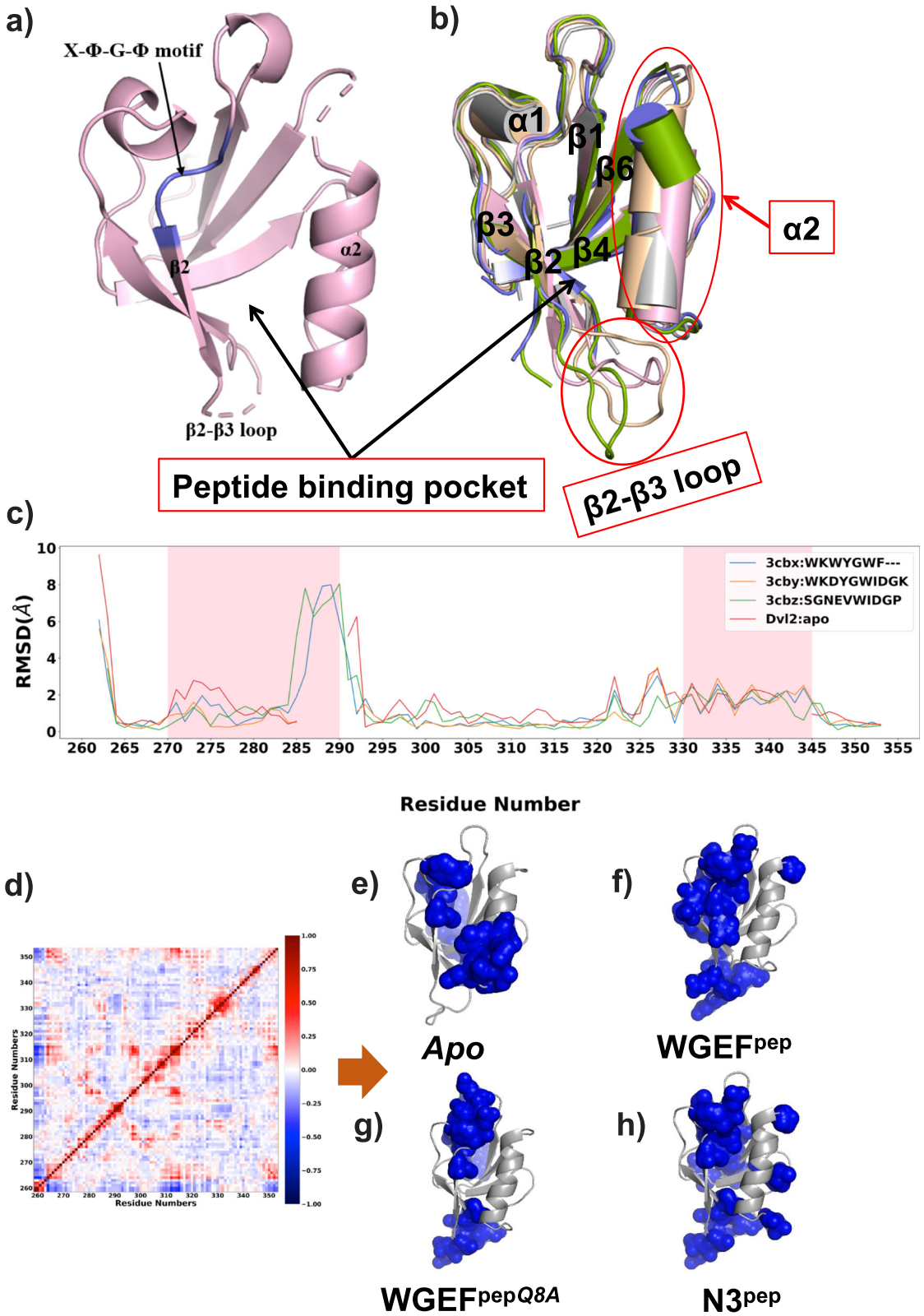


Fig. 4 | Conformational variations between the *apo* and ligand-bound Dvl2^{PDZ} structures. a *Apo* structure of human Dvl2^{PDZ}. **b** Superimposition of *apo* Dvl2^{PDZ} and peptide ligand bound Dvl2^{PDZ} structures colour coded as Dvl2^{PDZ} (PDB ID: 8WWR)-grey, C1^{pep} (PDB ID: 3CBX)-green, N1^{pep} (PDB ID: 3CBY)-blue, N2^{pep} (PDB ID: 3CBZ)-peach, N3^{pep} (PDB ID: 3CC0)-pink, major conformational deviations in the α2 helix and β2-β3 loop regions of Dvl2^{PDZ} are marked with a red circle. **c** Residue-wise RMSD plot between the *apo* and *internal peptide* ligand bound Dvl2^{PDZ} structures. **d** Representative plot of the *dynamic cross-coupling matrix* (*cij*) used for calculating the cluster of dynamically coupled amino acids in peptide bound and *apo* Dvl2^{PDZ} structures. **e–h** Dynamically coupled amino acids belonging to Sector 1 were shown as blue spheres on the *apo* and WGEF^{pep}, WGEF^{pepQ8A} and N3^{pep} bound Dvl2^{PDZ} structures, respectively.

(RhoA) binding pocket of hWGEF is blocked by the N-terminal inhibitory and C-terminal SH3 domains (Supplementary Fig. 8c, d). Thus, the intra-protein interactions could render the GEF into its inactive state (Supplementary Fig. 8e). When the Dvl2^{PDZ} domain engages with the "binding motif", it could introduce conformational changes in the NID, resulting in it moving away from the GTPase binding pocket of the GEF. Thus, the hWGEF–Dvl2^{PDZ} association could partially release the autoinhibitory state of the former (Supplementary Fig. 8f). Perhaps, some other unknown mechanisms might also disengage the SH3 domain from the DH domain, which would result in the complete activation of GEF (Supplementary Fig. 8g). Since the putative mechanism of hWGEF activation proposed here is based on the AlphaFold predicted structure, substantive GEF activation mechanism warrants further structural studies.

In light of this, studies reported here have thus provided invaluable insights on WGEF–Dvl2^{PDZ} interaction and augmented the existing repository of PDZ-*internal peptide* interaction modes. Our study may also pave the way for designing inhibitors to selectively block Wnt-frizzled signalling, which is dysregulated in many pathological conditions[52].

Methods

Peptides and constructs

Synthetic peptides such as WGEF^{pep}, WGEF^{pep} mutants and N3^{pep} were procured from Sai Biotech. Fused constructs of hDvl2^{PDZ}-WGEF^{pep} were commercially synthesized from Twist Biosciences (https://www.twistbioscience.com).

Constructs, cloning and mutagenesis

Human *Arhgef19* (encoding hWGEF 330–581) and *Xenopus Arhgef19* (encoding xWGEF 330–856) genes were PCR amplified and cloned into a modified p3E vector (with N-terminal cleavable GST tag). Human *RhoA* (encoding hRhoA 2–180) was cloned in a modified pET33b vector (with N-terminal double Strep-tag). Human Dishevelled2 PDZ (hDvl2 PDZ) was a gift from Nicola Burgess-Brown (Addgene plasmid # 38876; RRID: Addgene_38876); this vector was modified by introducing a stop codon after C354 residue of Dvl2^{PDZ} to eliminate the fused peptide, cloned at the C-termini of PDZ. This construct has an N-terminal cleavable 6xHis tag. Mutations were introduced into xWGEF (330–856) through site-directed mutagenesis method and confirmed by DNA sequencing.

Protein expression and purification

Recombinant hWGEF (330–581) and RhoA (2–180) were expressed into *Escherichia coli (E. coli)* strain BL21*(DE3), whereas xWGEF (330–856) was expressed into C41 (DE3) strain. Cells were grown at 37 °C in Luria Broth (LB) containing Ampicillin (100 µg/mL) followed by induction with 0.5 mM isopropyl β-D-thiogalactopyranoside (IPTG) for 16 h at 18 °C once the OD (optical density at 600 nm) of 0.6 is reached. After incubation, cells were pelleted at 4000 rpm for 20 min and stored at −80 °C. GST-tagged proteins were purified using glutathione (GSH) Sepharose-4B affinity purification followed by dialysis and GST tag cleavage using PreScission protease at 4 °C overnight. The GST tag and PreScission protease were removed by desalting, followed by a second GST affinity purification. To obtain stable and pure protein, gel filtration chromatography was performed using Hiprep 26/60 Sephacryl S-300 HR SEC column in a buffer containing 10 mM Tris–HCl pH 8, 150 mM NaCl, 2 mM DTT and 5% glycerol for xWGEF. Hiprep 16/60 Sephacryl S-200 HR SEC column with 10 mM MOPS pH 6.5, 150 mM NaCl, 2 mM DTT and 10% Glycerol buffer was used for gel filtration chromatography of hWGEF. RhoA was purified using streptactin affinity purification, after which the protein was subjected to gel filtration chromatography using Hiprep 16/60 Sephacryl S-200 HR SEC column in the buffer containing 10 mM Tris–HCl pH 8, 150 mM NaCl, 2mM EDTA, 2 mM DTT and 5% glycerol. 6xHis tagged hDvl2^{PDZ} and fused hDvl2^{PDZ}-WGEF^{pep} constructs were expressed in Rosetta (DE3) cells and purified using Ni^{2+}-nitrilotriacetate (Ni^{2+}-NTA). 6xHis tag is further cleaved by using TEV protease during overnight dialysis. TEV protease and His tag are eliminated through desalting and 2nd Ni^{2+}-NTA purification. Finally, pure protein is obtained after gel filtration using Hiprep 16/60 Sephacryl S-200 HR SEC column in buffer containing 10 mM Tris–HCl pH 8, 150 mM NaCl, 2 mM DTT and 5% glycerol for hDvl2^{PDZ} and 10 mM Tris–HCl pH 8, 350 mM NaCl, 2 mM DTT and 5% glycerol for hDvl2^{PDZ}-WGEF^{pep}. Mutations were introduced into xWGEF (330–856) through site-directed mutagenesis; expression and purification were carried out as described above for the GST-tagged proteins. All proteins were concentrated and stored at −80 °C for further studies.

Protein crystallization, data collection and structure determination

Crystals of Human Dvl2^{PDZ} (hDvl2^{PDZ}) were grown with sitting drop vapour diffusion method at 293 K by mixing 1:1 ratio of hDvl2^{PDZ} (10 mg/mL) with 6 M excess of WGEF^{pep} and reservoir buffer containing 0.1 M tri-sodium citrate (pH 5.6), 20% (v/v) Isopropanol, 20% (w/v) PEG4000. Crystals of Human Dvl2^{PDZ} fused with WGEF^{pep} (hDvl2^{PDZ}-WGEF^{pep}) were grown with sitting drop vapour diffusion method at 293 K by mixing 1:1 ratio of protein (10 mg/mL) and reservoir buffer containing 0.1 M Sodium acetate (pH 4.5) and 3 M Sodium chloride. The crystallization plate was set up using a Mosquito® crystallization robot (TTP Labtech, Royston UK). Crystals were frozen in cryoprotectants containing crystallization conditions supplemented with 28% glycerol. X-ray diffraction data was collected at 100 K on microfocus beamline ID23-2 of the European Synchrotron Radiation Facility (ESRF), Grenoble, France[53]. Crystals of the fused hDvl2^{PDZ}-WGEF^{pep} protein were diffracted at Indus-2 synchrotron on BL-21 PX beamline at RRCAT, Indore. All the data sets were integrated with XDS[54] and scaled using AIMLESS[55,56], implemented on CCP4 software[57]. The structure was solved by molecular replacement with PHASER[58]. Human Dvl2-PDZ structure (PDB entry: 2REY) served as a search model. The structure was further iteratively built using COOT[59] and refined using PHENIX[60,61]. The structure validation was performed using Molprobidity[62]. Coordinates and structure factors for hDvl2^{PDZ} and WGEF^{pep} fused hDvl2^{PDZ} are deposited in the Protein Data Bank under accession code: PDB IDs 8WWR and 8YR7, respectively.

In vitro guanine nucleotide exchange assay

Guanine nucleotide exchange assays were carried out using RhoA loaded with fluorescently labelled Mant-GDP[63,64]. Briefly, 150 µM of RhoA was incubated with 5 Molar excess of fluorescent label, i.e., Mant-GDP in a buffer containing 10 mM Tris pH 8, 150 mM NaCl, 2 mM DTT, 5 mM EDTA, 0.5 mM $MgCl_2$ and 5% Glycerol. The reaction was incubated on ice for 30 min, followed by the addition of 10 mM $MgCl_2$ to stop the reaction. Excess Mant-GDP was removed through buffer exchange and Mant-GDP loaded RhoA was concentrated and stored at −80 °C. For assays, 2 µM of Mant-GDP loaded RhoA was incubated with 5 µM of xWGEF in the presence or absence of hDvl2^{PDZ} (0 µM, 25 µM and 50 µM) for 30 min at 25 °C. To observe effect of the synthetic peptides WGEF^{pep} (GSTFSLWQDIP) and WGEF^{pepD9A} (GSTFSLWQAIP) on Dvl2^{PDZ}–WGEF interaction, these peptides were added to a final concentration of 200 µM in the respective reactions. Once the fluorescence was stable, excess GTP (10x of labelled GTPase) was added to start the exchange reaction. All the GEF assays were performed in a buffer containing 10 mM Tris, pH 8, 150 mM NaCl, 2 mM DTT, 10 mM $MgCl_2$ and 5% Glycerol. A decrease in fluorescence was recorded using a BioTek Cytation5 plate reader with excitation and emission of 360 and 440 nm, respectively. Fluorescence was normalized and decrease was fitted using a single exponential decay mode in GraphPad Prism. All the assays were performed in triplicates.

Microscale thermophoresis (MST)

MST experiments were performed using Monolith NT.115 instrument from Nanotemper technologies. Dvl2^{PDZ} domain was fluorescently labelled as per the instructions, using Monolith Redmaleimide 2nd generation-cysteine reactive label. The experimental condition consists of 10 mM Tris–HCl pH8, 150 mM NaCl, 2 mM DTT, 5% Glycerol and 0.05% Tween20. For Dvl2^{PDZ}-peptide affinity determination, synthetic WGEF

peptides were serially diluted (1:2) 16 times, starting from an approximate concentration of 3.4 mM. 50 nM of fluorescently labelled $Dvl2^{PDZ}$ domain was added to the serially diluted peptides, followed by incubation at 25 °C for 30 minutes. Thermophoresis signals were recorded at 20% of excitation power and 40% MST at 25 °C. Data was fitted using the K_d fit model in the MO. Affinity Analysis v2.3 software[65]. Like peptides, hWGEF (330–581) protein was serially diluted to obtain a protein concentration range of 138–0.0042 μM. Experimental conditions were the same as mentioned above for peptides. All MST experiments were done in triplicates.

Molecular dynamics (MD) simulations of $hDvl2^{PDZ}$ and interacting peptides

MD simulations were performed by using the Desmond module from the Schrödinger suite[66]. The crystal structure of *apo* human $Dvl2^{PDZ}$ (PDB code: 8WWR) was used as a template for MODELLER-based homology modelling in order to build the missing loops[67]. $WGEF^{pep}$, $WGEF^{pepQ8A}$ mutant peptide and $N3^{pep}$ peptides were modelled and docked on $Dvl2^{PDZ}$ by using the previously available internal ($N3^{pep}$) peptide structure (PDB code: 3CC0). These peptides, along with the modelled $Dvl2^{PDZ}$ structure, were further used for 1 μs MD simulation. *Apo* form of $Dvl2^{PDZ}$ was also simulated by adopting a similar methodology. All the structures were placed in a simulation box of 10 Å and further solvated using TIP3P water molecules along with Na^+ ions to neutralize the system[68]. Next, temperature equilibration was done at a constant temperature of 300 K using the Nosé–Hoover thermostat, and the pressure was equilibrated at 1.01 bar under NPT ensemble using Martyna-Tobias-Klein barostat[69–72]. Followed by temperature and pressure equilibration, MD runs were performed using OPLS_2005 force[73,74]. MD runs were performed in triplicates. Other parameters used for simulations are mentioned in Supplementary Table 2. RMSD plots for $Dvl2^{PDZ}$ and peptides are given in Supplementary Fig. 9.

Dynamical coupling analysis

The dynamical cross-correlation matrix (DCCM), dC_{ij}[75], was calculated as

$$dC_{ij} = \left\langle r_i r_j \right\rangle - \left\langle r_i \right\rangle \left\langle r_j \right\rangle$$

where r_i and r_j are the position vectors of the i^{th} and j^{th} Cα atoms, respectively of $Dvl2^{PDZ}$, at time t. The R implementation of the Bio3D programme (version 2.4-1)[76] was used for the DCCM calculations. The dynamical sectors were obtained from the spectral decomposition algorithm[77] using the methodology (statistical coupling analysis) reported earlier[64,78]. Trajectories from all the triplicate MD runs were merged for dynamical coupling analysis.

Statistics and reproducibility

Guanine nucleotide exchange assays were performed in triplicates. Fluorescence decrease was fitted using a single exponential decay mode in GraphPad Prism. All results are reported with standard deviation (±SD). Statistical significance was calculated using ordinary One-way ANOVA with multiple comparisons. * Represents adjust P value with statistical significance of $****P < 0.0001$, $***P < 0.001$, $**P < 0.01$ and $*P \leq 0.01$. MST assays were performed in triplicates. Data was fitted using the K_d fit model in the MO. Affinity Analysis v2.3 software. All results are reported with standard deviation (±SD). MD simulation runs for *apo* and peptide docked $Dvl2^{PDZ}$ were performed in triplicates. Plots reported from MD simulation studies are represented with standard deviation (±SD).

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The atomic model of *apo* human Dvl2 PDZ is available in the Protein Data Bank (PDB) under the accession code 8WWR. The atomic model of human Dvl2 PDZ fused with $WGEF^{pep}$ (no electron density for the peptide) is available in the Protein Data Bank (PDB) under the accession code 8YR7. Initial coordinates, simulation input files and final coordinates of the MD simulation runs are available at https://zenodo.org/records/10683731[79]. The source data for graphs in the manuscript can be found in Supplementary Data 1.

Received: 4 November 2023; Accepted: 15 April 2024;
Published online: 07 May 2024

References

1. Logan, C. Y. & Nusse, R. The Wnt signaling pathway in development and disease. *Annu. Rev. Cell Dev. Biol.* **20**, 781–810 (2004).
2. Liu, J. et al. Wnt/β-catenin signalling: function, biological mechanisms, and therapeutic opportunities. *Signal Transduct. Target. Ther.* **7**, 3 (2022).
3. Chae, W. J. & Bothwell, A. L. M. Canonical and non-canonical Wnt signaling in immune cells. *Trends Immunol.* **39**, 830–847 (2018).
4. Komiya, Y. & Habas, R. Wnt signal transduction pathways. *Organogenesis* **4**, 68–75 (2008).
5. Groenewald, W., Lund, A. H. & Gay, D. M. The role of WNT pathway mutations in cancer development and an overview of therapeutic options. *Cells* **12**, 990 (2023).
6. Patel, S., Alam, A., Pant, R. & Chattopadhyay, S. Wnt signaling and its significance within the tumor microenvironment: novel therapeutic insights. *Front. Immunol.* **10**, 486317 (2019).
7. Schlessinger, K., Hall, A. & Tolwinski, N. Wnt signaling pathways meet Rho GTPases. *Genes Dev.* **23**, 265–277 (2009).
8. Kjøller, L. & Hall, A. Signaling to Rho GTPases. *Exp. Cell Res.* **253**, 166–179 (1999).
9. Bos, J. L., Rehmann, H. & Wittinghofer, A. GEFs and GAPs: critical elements in the control of small G proteins. *Cell* **129**, 865–877 (2007).
10. Omble, A. & Kulkarni, K. GPCRs that Rhoar the Guanine nucleotide exchange factors. *Small GTPases* **13**, 84–99 (2022).
11. Schiller, M. R. Coupling receptor tyrosine kinases to Rho GTPases-GEFs what's the link. *Cell. Signal.* **18**, 1834–1843 (2006).
12. Ridley, A. Rho GTPases: Integrating integrin signaling. *J. Cell Biol.* **150**, F107–F109 (2000).
13. Wang, Y. et al. WGEF is a novel RhoGEF expressed in intestine, liver, heart, and kidney. *Biochem. Biophys. Res. Commun.* **324**, 1053–1058 (2004).
14. Kim, K., Lee, S. A. & Park, D. Emerging roles of ephexins in physiology and disease. *Cells* **8**, 87 (2019).
15. Hardy, K. M. et al. Non-canonical Wnt signaling through Wnt5a/b and a novel Wnt11 gene, Wnt11b, regulates cell migration during avian gastrulation. *Dev. Biol.* **320**, 391–401 (2008).
16. Tada, M., Concha, M. L. & Heisenberg, C. P. Non-canonical Wnt signalling and regulation of gastrulation movements. *Semin. Cell Dev. Biol.* **13**, 251–260 (2002).
17. Tanegashima, K., Zhao, H. & Dawid, I. B. WGEF activates Rho in the Wnt-PCP pathway and controls convergent extension in Xenopus gastrulation. *EMBO J.* **27**, 606–617 (2008).
18. Hu, D. J. K., Yun, J., Elstrott, J. & Jasper, H. Non-canonical Wnt signaling promotes directed migration of intestinal stem cells to sites of injury. *Nat. Commun.* **12**, 7150 (2021).
19. Sahin, M. et al. Eph-dependent tyrosine phosphorylation of ephexin1 modulates growth cone collapse. *Neuron* **46**, 191–204 (2005).
20. Yohe, M. E. et al. Auto-inhibition of the Dbl family protein Tim by an N-terminal helical motif. *J. Biol. Chem.* **282**, 13813–13823 (2007).
21. Yohe, M. E., Rossman, K. & Sondek, J. Role of the C-terminal SH3 domain and N-terminal tyrosine phosphorylation in regulation of Tim and related Dbl-family proteins. *Biochemistry* **47**, 6827–6839 (2008).
22. Kim, K. et al. Intermolecular steric inhibition of Ephexin4 is relieved by Elmo1. *Sci. Rep.* **7**, 4404 (2017).

23. Kim, K. et al. The intermolecular interaction of Ephexin4 leads to autoinhibition by impeding binding of RhoG. *Cells* **7**, 211 (2018).
24. Gao, C. & Chen, Y. G. Dishevelled: the hub of Wnt signaling. *Cell. Signal.* **22**, 717–727 (2010).
25. Sharma, M., Castro-Piedras, I., Simmons, G. E. & Pruitt, K. Dishevelled: a masterful conductor of complex Wnt signals. *Cell. Signal.* **47**, 52–64 (2018).
26. Wallingford, J. B. & Habas, R. The developmental biology of dishevelled: an enigmatic protein governing cell fate and cell polarity. *Development* **132**, 4421–4436 (2005).
27. Wong, H. C. et al. Direct binding of the PDZ domain of dishevelled to a conserved internal sequence in the C-terminal region of frizzled. *Mol. Cell* **12**, 1251–1260 (2003).
28. Paclíková, P., Bernatík, O., Radaszkiewicz, T. W. & Bryja, V. The N-terminal part of the dishevelled DEP domain is required for Wnt/β-Catenin signaling in mammalian cells. *Mol. Cell. Biol.* **37**, e00145–17 (2017).
29. Schwarz-Romond, T. et al. The DIX domain of dishevelled confers Wnt signaling by dynamic polymerization. *Nat. Struct. Mol. Biol.* **14**, 484–492 (2007).
30. Nielsen, C. P., Jernigan, K. K., Diggins, N. L., Webb, D. J. & MacGurn, J. A. USP9X deubiquitylates DVL2 to regulate WNT pathway specification. *Cell Rep.* **28**, 1074–1089 (2019).
31. Lee, H. J. & Zheng, J. J. PDZ domains and their binding partners: Structure, specificity, and modification. *Cell Commun. Signal.* **8**, 1–18 (2010).
32. Songyang, Z. et al. Recognition of unique carboxyl-terminal motifs by distinct PDZ domains. *Science* **275**, 73–77 (1997).
33. Ali, M. et al. Integrated analysis of Shank1 PDZ interactions with C-terminal and internal binding motifs. *Curr. Res. Struct. Biol.* **3**, 41–50 (2021).
34. Zhu, Y. et al. Deciphering the unexpected binding capacity of the third PDZ domain of whirlin to various cochlear hair cell partners. *J. Mol. Biol.* **432**, 5920–5937 (2020).
35. Zhang, Y. et al. Inhibition of Wnt signaling by dishevelled PDZ peptides. *Nat. Chem. Biol.* **5**, 217–219 (2009).
36. Aasland, R. et al. Normalization of nomenclature for peptide motifs as ligands of modular protein domains. *FEBS Lett.* **513**, 141–144 (2002).
37. Penkert, R. R., DiVittorio, H. M. & Prehoda, K. E. Internal recognition through PDZ domain plasticity in the Par-6–Pals1 complex. *Nat. Struct. Mol. Biol.* **11**, 1122–1127 (2004).
38. Skelton, N. J. et al. Origins of PDZ domain ligand specificity: structure determination and mutagenesis of the erbin PDZ domain. *J. Biol. Chem.* **278**, 7645–7654 (2003).
39. Lee, H. J., Shi, D. L. & Zheng, J. J. Conformational change of dishevelled plays a key regulatory role in the wnt signaling pathways. *Elife* **4**, e08142 (2015).
40. Ivarsson, Y. Plasticity of PDZ domains in ligand recognition and signaling. *FEBS Lett.* **586**, 2638–2647 (2012).
41. Mu, Y., Cai, P., Hu, S., Ma, S. & Gao, Y. Characterization of diverse internal binding specificities of PDZ domains by yeast two-hybrid screening of a special peptide library. *PLoS ONE* **9**, e88286 (2014).
42. Münz, M., Hein, J. & Biggin, P. C. The role of flexibility and conformational selection in the binding promiscuity of PDZ domains. *PLoS Comput. Biol.* **8**, e1002749 (2012).
43. Hillier, B. J., Christopherson, K. S., Prehoda, K. E., Bredt, D. S. & Lim, W. A. Unexpected modes of PDZ domain scaffolding revealed by structure of nNOS-syntrophin complex. *Science* **284**, 812–815 (1999).
44. Whitney, D. S., Peterson, F. C. & Volkman, B. F. A conformational switch in the CRIB-PDZ module of Par-6. *Structure* **19**, 1711–1722 (2011).
45. Lu, C., Knecht, V. & Stock, G. Long-range conformational response of a PDZ domain to ligand binding and release: a molecular dynamics study. *J. Chem. Theory Comput.* **12**, 870–878 (2016).
46. Amacher, J. F., Brooks, L., Hampton, T. H. & Madden, D. R. Specificity in PDZ-peptide interaction networks: computational analysis and review. *J. Struct. Biol. X.* **4**, 100022 (2020).
47. Stevens, A. O. & He, Y. Allosterism in the PDZ family. *Int. J. Mol. Sci.* **23**, 1454 (2022).
48. Harris, B. Z. & Lim, W. A. Mechanism and role of PDZ domains in signaling complex assembly. *J. Cell Sci.* **114**, 3219–3231 (2001).
49. Wharton, K. A. Runnin' with the Dvl: proteins that associate with Dsh/Dvl and their significance to Wnt signal transduction. *Dev. Biol.* **253**, 1–17 (2003).
50. Tonikian, R. et al. A specificity map for the PDZ domain family. *PLoS Biol.* **6**, e239 (2008).
51. Tochio, H., Zhang, Q., Mandal, P., Li, M. & Zhang, M. Solution structure of the extended neuronal nitric oxide synthase PDZ domain complexed with an associated peptide. *Nat. Struct. Biol.* **6**, 417–421 (1999).
52. Katoh, M. WNT/PCP signaling pathway and human cancer. *Oncol. Rep.* **14**, 1583–1588 (2005).
53. Flot, D. et al. The ID23-2 structural biology microfocus beamline at the ESRF. *J. Synchrotron Radiat.* **17**, 107–118 (2010).
54. Kabsch, W. et al. *XDS*. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 125–132 (2010).
55. Evans, P. R. An introduction to data reduction: space-group determination, scaling and intensity statistics. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **67**, 282–292 (2011).
56. Evans, P. R. & Murshudov, G. N. How good are my data and what is the resolution? *Acta Crystallogr. Sect. D Biol. Crystallogr.* **69**, 1204–1214 (2013).
57. Winn, M. D. et al. Overview of the CCP4 suite and current developments. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **67**, 235–242 (2011).
58. McCoy, A. J. et al. Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
59. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 486–501 (2010).
60. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 213–221 (2010).
61. Liebschner, D. et al. Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr. Sect. D Struct. Biol.* **75**, 861–877 (2019).
62. Williams, C. J. et al. MolProbity: more and better reference data for improved all-atom structure validation. *Protein Sci.* **27**, 293–315 (2018).
63. Kanie, T. & Jackson, P. Guanine nucleotide exchange assay using fluorescent MANT-GDP. *Bio-Protocol* **8**, e2795–e2795 (2018).
64. Roy, D., Sengupta, D. & Kulkarni, K. Substrate induced dynamical remodeling of the binding pocket generates GTPase specificity in DOCK family of guanine nucleotide exchange factors. *Biochem. Biophys. Res. Commun.* **631**, 32–40 (2022).
65. Asmari, M., Ratih, R., Alhazmi, H. A. & El Deeb, S. Thermophoresis for characterizing biomolecular interaction. *Methods* **146**, 107–119 (2018).
66. Bowers, K. J. et al. Scalable algorithms for molecular dynamics simulations on commodity clusters. In *SC'06: Proceedings of the 2006 ACM/IEEE Conference on Supercomputing.* (ed. Horner-Miller, B.) 43 (IEEE, 2006).
67. Šali, A. & Blundell, T. L. Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* **234**, 779–815 (1993).
68. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926–935 (1983).
69. Evans, D. J. & Holian, B. L. The Nose–Hoover thermostat. *J. Chem. Phys.* **83**, 4069–4074 (1985).

70. Nosé, S. A unified formulation of the constant temperature molecular dynamics methods. *J. Chem. Phys.* **81**, 511–519 (1984).
71. Martyna, G. J., Tobias, D. J. & Klein, M. L. Constant pressure molecular dynamics algorithms. *J. Chem. Phys.* **101**, 4177–4189 (1994).
72. Martyna, G. J., Tuckerman, M. E., Tobias, D. J. & Klein, M. L. Explicit reversible integrators for extended systems dynamics. *Mol. Phys.* **87**, 1117–1157 (1996).
73. Jorgensen, W. L., Maxwell, D. S. & Tirado-Rives, J. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.* **118**, 11225–11236 (1996).
74. Shivakumar, D. et al. Prediction of absolute solvation free energies using molecular dynamics free energy perturbation and the opls force field. *J. Chem. Theory Comput.* **6**, 1509–1519 (2010).
75. Karplus, M. & Ichiye, T. Comment on a "Fluctuation and cross-correlation analysis of protein motions observed in nanosecond molecular dynamics simulations". *J. Mol. Biol.* **263**, 120–122 (1995).
76. Grant, B. J., Rodrigues, A. P. C., Elsawy, K. M., Mccammon, J. A. & Caves, L. S. D. Bio3d: an R package for the comparative analysis of protein structures. *Bioinformatics* **22**, 2695–2696 (2006).
77. Rivoire, O., Reynolds, K. A. & Ranganathan, R. Evolution-based functional decomposition of proteins. *PLoS Comput. Biol.* **12**, e1004817 (2016).
78. Singh, S., Thulasiram, H. V., Sengupta, D. & Kulkarni, K. Dynamic coupling analysis on plant sesquiterpene synthases provides leads for the identification of product specificity determinants. *Biochem. Biophys. Res. Commun.* **536**, 107–114 (2021).
79. Omble, A., Mahajan, S., Bhoite, A. & Kulkarni, K. Dishevelled2 activates WGEF via its interaction with a unique internal peptide motif of the GEF [Dataset]. *Zenodo* https://zenodo.org/records/10683731 (2024).

Acknowledgements

K.K. would like to acknowledge funding from the Department of Biotechnology, India (BT/PR12502/BRB/10/ 1387/2015). We thank Prof. Dawid Igor for generously sharing various constructs of Dvl, Frizzled and WGEF. We acknowledge the European Synchrotron Radiation Facility (ESRF), ID23-2 beamline, for the provision of synchrotron radiation facilities and we would like to thank Dr. Christoph Mueller-Dieckmann for assistance and support in using the beamline. We acknowledge BL-21 PX beamline scientists Dr. Ravindra D. Makde, Dr. Ashwani Kumar and Dr. Biplab Ghosh at RRCAT, Indore, India, for helping with data collection. The authors acknowledge the Central Instrumentation Facility (CIF), Savitribai Phule Pune University (SPPU) for the MST facility. The authors acknowledge Dr. Durba Sengupta for their inputs on MD simulations. A.O. acknowledges the Council of Scientific & Industrial Research (CSIR), India for fellowship.

Author contributions

A.O. designed and performed all the experiments. S.M. helped with the crystallization studies. A.B. screened the initial constructs of WGEF for expression and solubility. K.K. conceptualized the problem, supervised overall work, analysed the data and wrote the paper with inputs from A.O.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s42003-024-06194-6.

Correspondence and requests for materials should be addressed to Kiran Kulkarni.

Peer review information *Communications Biology* thanks the anonymous reviewers for their contribution to the peer review of this work. Primary Handling Editors: Huijuan Guo, Tobias Goris, and David Favero. A peer review file is available.

SMALL GTPASES
2022, VOL. 13, NO. 1, 84–99
https://doi.org/10.1080/21541248.2021.1896963

Taylor & Francis
Taylor & Francis Group

REVIEW

GPCRs that *Rho*ar the Guanine nucleotide exchange factors

Aishwarya Omble [a,b] and Kiran Kulkarni [a,b]

[a]Division of Biochemical Sciences, CSIR-National Chemical Laboratory, Pune, India; [b]Academy of Scientific and Innovative Research (AcSIR), Ghaziabad, India

ABSTRACT

Cell migration, a crucial step in numerous biological processes, is tightly regulated in space and time. Cells employ Rho GTPases, primarily Rho, Rac, and Cdc42, to regulate their motility. Like other small G proteins, Rho GTPases function as biomolecular switches in regulating cell migration by operating between GDP bound 'OFF' and GTP bound 'ON' states. Guanine nucleotide exchange factors (GEFs) catalyse the shuttling of GTPases from OFF to ON state. G protein-coupled receptors (GPCRs) are the largest family of cell surface receptors that are involved in many signalling phenomena including cell survival and cell migration events. In this review, we summarize signalling mechanisms, involving GPCRs, leading to the activation of RhoGEFs. GPCRs exhibit diverse GEF activation modes that include the interaction of heterotrimeric G protein subunits with different domains of GEFs, phosphorylation, protein–protein interaction, protein–lipid interaction, and/or a combination of these processes.

ARTICLE HISTORY
Received 1 January 2021
Revised 15 February 2021
Accepted 24 February 2021

KEYWORDS
Cell migration; guanine nucleotide; exchange factors (GEFs); Rho GTPases; G protein-coupled receptors (GPCRs); cell signalling

Introduction

Cell migration is a fundamental event in various biological processes such as embryogenesis, immune surveillance, homoeostasis, wound repair, and tissue formation[1]. Cells can migrate as individuals or as a collective group of cells. Collective cell migration contributes to angiogenesis, lumen formation, and tumour cell invasion[2]. On the other hand, individual cell migration is exhibited by cells like fibroblasts (during wound healing) and leukocytes (during immune response)[1]. To initiate migration, cells receive a variety of migratory and invasive cues from their surroundings, which could be in the form of chemicals such as chemokines[3], growth factors (EGF, IGF-1) [4] or sometimes physical signals such as temperature and pressure [5,6]. The migration cycle of an individual cell includes polarization of the front and rear end of the cell, protrusive structure formation at the front end, adherence of the protrusions with the extracellular matrix (ECM), and retraction of the rear end due to contractile force generated by the cytoskeleton[7]. As compared to individual cell migration, collective cell migration is much more complex in terms of its execution and signalling[2]. Similar to individual cell migration, cells moving in a collective fashion also establish front and rear polarity. However, in the collective cell migration, cells at the front end polarize and produce protrusions, whereas, the ones present at the lateral and rear end maintain cell-cell adhesion to generate the required traction force[2]. Thus, the initial steps in all types of cell migration include polarization and formation of protrusions like lamellipodia and filopodia, which require coordinated cytoskeletal dynamics. As cell migration is a vital biological step, it is tightly regulated in space and time by employing a complex signalling network involving the Rho family of small G proteins[8]. These G proteins control cytoskeletal dynamics by regulating actin polymerization. Amongst the known 22 Rho GTPases family members, RhoA, Rac and Cdc42 are the members, most extensively studied for their role in cell migration[8]. RhoA is mostly involved in stress fibre formation and cell contraction [9,10] whereas, Rac controls the formation of lamellipodia, cell spreading, and membrane ruffling [10–13]. Similarly, Cdc42 has been shown to have important roles in filopodia formation and cell reshaping [11,13]. Thus, this Rho GTPase trio lies at the central stage of signal transduction events that control cell migration. Like all canonical G proteins, Rho GTPases exert signalling events by being biomolecular switches. They are active (ON) when bound to GTP and inactive (OFF) when bound to GDP. Two different classes of proteins, Guanine nucleotide exchange factors (GEFs) and GTPase activating proteins (GAPs) help to shuttle the GTPases between ON and OFF states. GEFs activate GTPases by exchanging their

CONTACT Kiran Kulkarni ka.kulkarni@ncl.res.in Academy of Scientific and Innovative Research (Acsir), Ghaziabad 201002, India

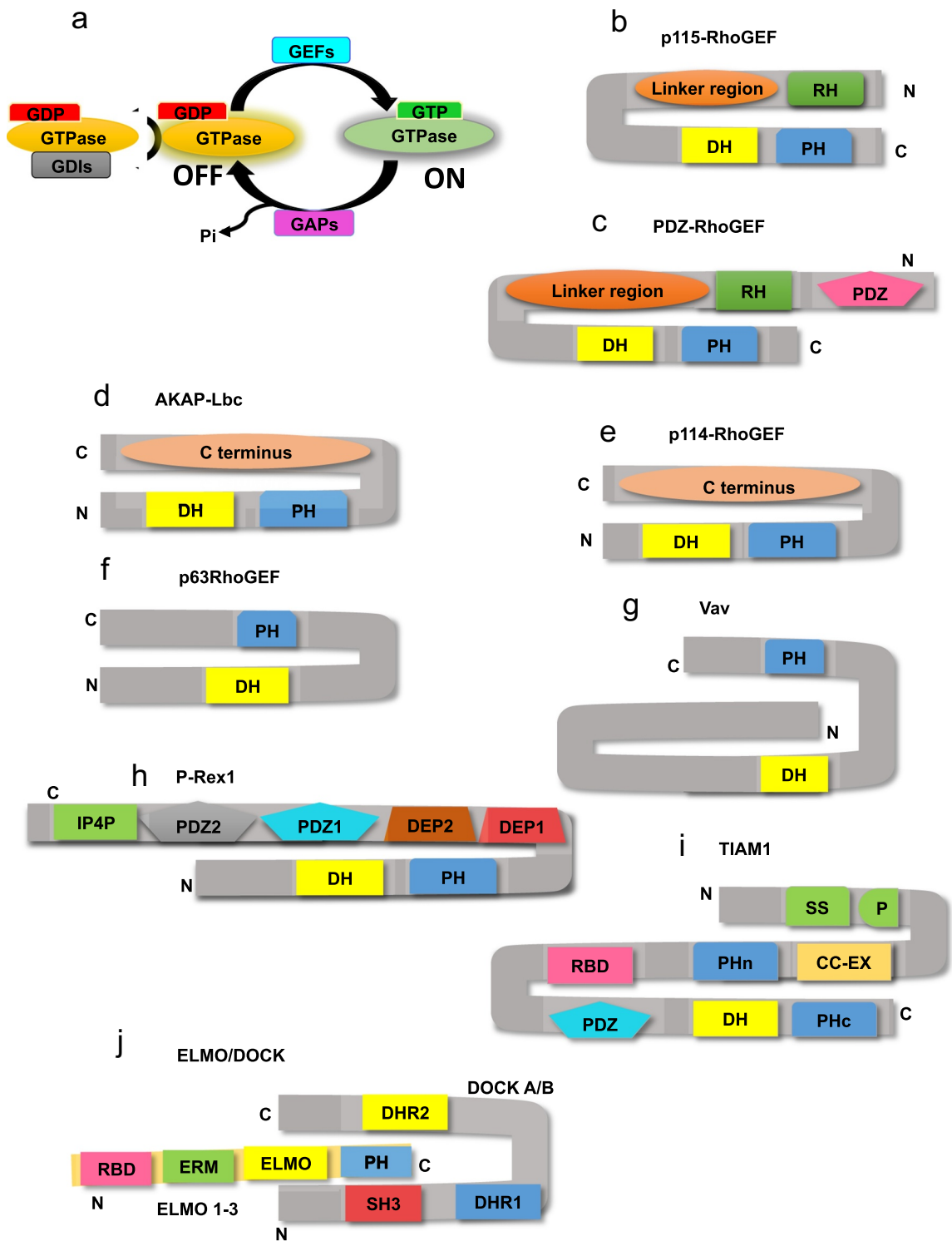


Figure 1. Schematic representation of some of the GEFs discussed in this review showing intra-domain interaction rendering them in autoinhibitory states. (a) GTPase switching mechanism between ON and OFF state. Canonical GEFs such as (b) p115-RhoGEF and (c) PDZ-RhoGEF are proposed to exhibit autoinhibition due to intra-domain interaction involving their RH domain. Similarly, auto-inhibition of (d) AKAP-Lbc and (e) p114-RhoGEF is due to intra-domain interaction involving its non-RH, C terminal region. (f) In p63RhoGEF and (g) Vav GEF, the PH domain is suggested to be responsible for its autoinhibition. (h) In P-Rex GEF, the C terminal domains like IPI4, PDZ1, and PDZ2 are responsible for intra-domain interactions. (i) In TIAM1, the initial 50 amino acid domain (N50), N terminal PH domain (PHn) and the coiled coil extension (CC-Ex) domain of TIAM1 interact with each other to render the GEF to its autoinhibitory state. (j) In non-canonical GEFs like DOCK proteins interaction between N terminal SH3 domain and C terminal domains and/or DHR2 domain is responsible for their autoinhibition.

bound GDP with GTP, whereas, GAPs bring them to OFF state by promoting their intrinsic GTP hydrolysis [14] (Figure 1a). For Rho GTPases, there exist another level of regulation in terms of Guanine nucleotide dissociation factors (GDIs), which sequester GDP bound GTPases in the cytosol and replenish them back when required by the cell [15] (Figure 1a). Many reviews excellently summarize the regulation of Rho GTPases [8,14,16–18] by GEFs, GAPs, and GDIs. However, how exactly the migratory cues propagate from various sources to the respective GEFs, GAPs, and GDIs is not well understood. In many instances, the migratory information propagates from the plasma membrane which is communicated largely through membrane-bound proteins such as G protein-coupled receptors (GPCR), Receptor tyrosine kinase (RTK), integrin, and cytokine receptors [16,19,20]. Amongst these, GPCRs are known to be the largest class of membrane receptors that play a key role in the regulation of GEFs. These seven transmembrane receptors are associated with heterotrimeric G proteins, comprising Gαβγ subunits and upon receiving the appropriate extracellular cues, they exert the signalling process by releasing the Gα and Gβγ subunits into the cytosol which in turn regulate their downstream effectors[21]. Aberration in these signalling events is implicated in various pathological conditions such as metabolic disorders, immunological disorders, neurodegenerative diseases, infectious diseases, and cancer [22–24]. A broad variety of signalling cascades initiated by GPCRs during cell migration is discussed earlier by Cotton *et al.* [25]. However, in this review, we will try to summarize how migratory signals from GPCRs specifically activate RhoGEFs, in the context of cell migration. The mode of propagation of these signals includes, binding of GPCR activated heterotrimeric G protein subunits with different domains of GEFs, phosphorylation of GEFs by various kinases downstream of GPCRs, protein–protein interactions, and protein–lipid interactions.

GPCR mediated heterotrimeric G protein subunit interaction with RH domains of GEFs

Cell polarization, protrusions, and focal adhesions at the site of leading edge formation in both, the individual cell migration as well as collective cell migration (in leader cells) are regulated by RhoA, Rac as well as Cdc42 GTPases [2,7]. Membrane proteins such as chemokine, cytokine, GPCRs, and adhesion receptors activate these GTPases through their GEFs[19]. RhoGEFs can be classified into two categories as canonical and non-canonical (Docker GEFs). Almost all canonical GEFs comprise catalytic Dbl homology domain (DH) and at least one pleckstrin homology (PH) domain[26]. PH domain is primarily involved in the regulation of catalytic activity and also promotes membrane localization of GEF by interacting with membrane lipids like Phosphotidylinositides (PIPs) [27–29]. In a few GEFs, other domains like Ras binding domain (RBD), additional PH domains, regulatory G protein signalling homology (RH) domain, and PDZ domain, are also known to regulate their catalytic activity [30–32]. Most of the canonical GEFs are known to exist in an autoinhibited state due to the interaction between domains at their N and C terminus (Figure 1b-J), making the catalytical domain inaccessible for the GTPase binding. Unlike canonical GEFs, in non-canonical GEFs the PH and DH domains are replaced by DHR-1 and DHR-2 domains (also known as Docker domains), which perform functions similar to their canonical counterparts. Some of the Dedicator of cytokinesis (DOCK) family members are also known to possess additional domains like SH3 or PH domains [33–35]. Similar to the canonical GEFs, DOCK family GEFs are known to exist in an autoinhibitory state[36]. For example, N terminal SH3 domain of DOCK180/DOCK1 is known to interact with its DHR2 domain, to render it inactive[37].

In both types of GEFs, activation happens upon release of their autoinhibitory state either through protein–protein interactions or structural changes on account of post-translation modifications like phosphorylation, which disrupt the intra-domain interactions in them [29,38,39]. However, for many GEFs, the activation mechanism is still not clear. In few instances, GPCRs are known to activate GEFs either by promoting interaction between the heterotrimeric G protein subunits with them or indirectly through pathways regulated by these subunits [40–42]. In some examples, GEFs are known to be activated due to their direct interactions with different domains of GPCRs[43]. In this section, we will discuss ligand-mediated signalling through GPCRs that activate canonical GEFs like leukaemia-associated Rho GEF (LARG), p115-RhoGEF, and PDZ-RhoGEF through heterotrimeric G protein subunit interaction with the regulator of G protein signalling homology (RH) domain of these GEFs.

Leukaemia-associated Rho GEF (LARG) is a canonical GEF, shown to be expressed in most of the cells including peripheral blood leukocytes, spleen, prostate, testis, ovary, small intestine, colon with a minimal expression in thymus [44] and specifically, activates RhoA (Table 1). In mouse embryonic fibroblasts (NIH-3T3), GPCRs, M1 muscarinic cholinergic receptor (M1-mAChR) and Mas related

Table 1. Overview of canonical and non-canonical GEFs discussed in this review, their downstream Rho GTPases, tissue-specific expression and activating GPCRs.

Guanine nucleotide exchange factors (GEFs)	GTPase specificity	Tissue Specificity		Activating GPCR	References
		High expression	Moderate/low expression		
GPCR mediated heterotrimeric G protein subunit interaction with RH domains of GEFs					
LARG	RhoA	Peripheral blood leukocytes, spleen, prostate, testis, ovary, small intestine and colon	Thymus	mAChR, Mas related receptor, G2A, $S1P_2$, PAR and LPA	40,44,45,46,47
PDZ-RhoGEF	RhoA and Cdc42 (through non-RH domain)	Brain, testis, heart, ovary, central nervous system (CNS) and placenta	Kidney, pancreas, spleen, prostate, colon, skeletal muscle, lung and liver	GRPR, LPA_1R, LPA_2 R and CXCR4	46,48,49,52,53,56,53,54
p115-RhoGEF	RhoA	Haematopoietic cells		LPAR, CXCR4, S1PR and β_2AR	50,55,56,56
Activation of GEFs involving interaction between heterotrimeric G protein subunits and non-RH domains of GEFs					
AKAP-Lbc	RhoA	Cardiac tissues		AT_1R and LPAR	54,55,57
p63RhoGEF	RhoA, RhoB, and RhoC	Heart and brain		M_3R and H_1R	64,65,67
p114-RhoGEF	RhoA and Rac1	Kidney and pancreas	Liver, skeletal muscle and testis	M_3R and LPAR	59–61,58
P-Rex1	Rac1/2/3, Cdc42 and RhoG	Neutrophils, macrophages, platelets, endothelial cells, neurons and brain tissues	Bone marrow, thymus, spleen, lymph nodes and lungs	LPA_1R, CXCR4 and fMLPR	74,87,59,60,61
Activation of GEFs through phosphorylation, protein-lipid and protein-protein interactions mediated by GPCRs					
αPIX	Rac1 and Cdc42	Heart, skeletal muscle and immune cells		fMLPR	13,82,62,63
βPIX	Cdc42 and Rac1	Ubiquitous		LPAR	13,76,82,64
Vav1	Rac1/2, RhoG, and Cdc42	Haematopoietic cells	Pancreas, the tooth enamel, and the trophoblast layer	CXCR4 and fMLPR	42,87,89,65,66
Vav2	RhoA, RhoG, Rac1, and Cdc42	Broadly expressed		CXCR4	42,88,89
Vav3	RhoA, RhoG, and Rac1	Broadly expressed, highest in brain and haematopoietic cells		fMLPR	87,89,67
TIAM1	Rac1	Brain, testis and epidermis		LPA_1R and $S1P_1R$	12,99
TIAM2	Rac1	Only restricted to neuronal cells		?	99,100,68
Activation of non-canonical GEFs through GPCR by phosphorylation, protein-lipid and protein-protein interactions					
DOCK180/ DOCK1	Rac1	Broadly expressed		CXCR4 and LPA	76,69,106,134
DOCK2	Rac1/2	Haematopoietic cells		fMLPR	110,70,71

receptor activate RhoA specific GEF LARG, through direct binding of $G\alpha_q$ subunit (Figure 2a). Similarly, G protein-coupled receptor 132 (G2AR), when triggered with its lipid ligand, activates LARG through $G\alpha_{13}$[40]. From co-immunoprecipitation studies, it has been shown that the RH domain of LARG interacts with $G\alpha_q$ and $G\alpha_{13}$ subunits of GPCRs complex to quell its autoinhibitory state[40]. Similarly, in mouse smooth muscle cells, activation of Sphingosine 1 phosphate receptor 2 ($S1P_2R$) is shown to promote their motility, which is induced by the interaction between LARG and $G\alpha_{12/13}$ subunit [45] (Figure 2b). Analogous LARG activation mechanism is seen in human prostate cancer cells (PC-3) and Human embryo kidney (HEK293T) cells where the Thrombin receptor (PAR) operates through PAR/$G\alpha_{13}$/LARG pathway. In mouse embryonic fibroblasts LARG is activated through Lysophosphatidic acid receptor (LPAR)/$G\alpha_{12/13}$ /LARG/RhoA pathway involving GPCR, LPAR [46,47] (Figure 2b). LARG is also found to localize with Microtubule organizing centres (MTOC), indicating that the extracellular signals through GPCRs might be involved in maintaining microtubule dynamics and cell polarity through the same $G\alpha_{12/13}$ /LARG/RhoA pathway[47].

PDZ-RhoGEF and p115-RhoGEF provide examples for $G\alpha_{13}$ subunit induced activation of GEF, where the heterotrimeric G subunit interacts with their RH domain to release their autoinhibitory state [38,48]. However, GPCRs that provide the heterotrimeric G subunit ($G\alpha_{13}$) for GEF activation differ from cell to cell. C-X-C chemokine receptor type 4 (CXCR4), LPAR and Gastrin-releasing peptide receptor (GRPR) activate RhoA in breast cancer cell lines (MCF7, MDA-MB -231), PC-3 cells and colon cancer cells (Caco-2), respectively, through $G\alpha_{13}$/PDZ-RhoGEF/RhoA pathway [38,46,49] (Figure 2c). As seen in PDZ-RhoGEF, p115-

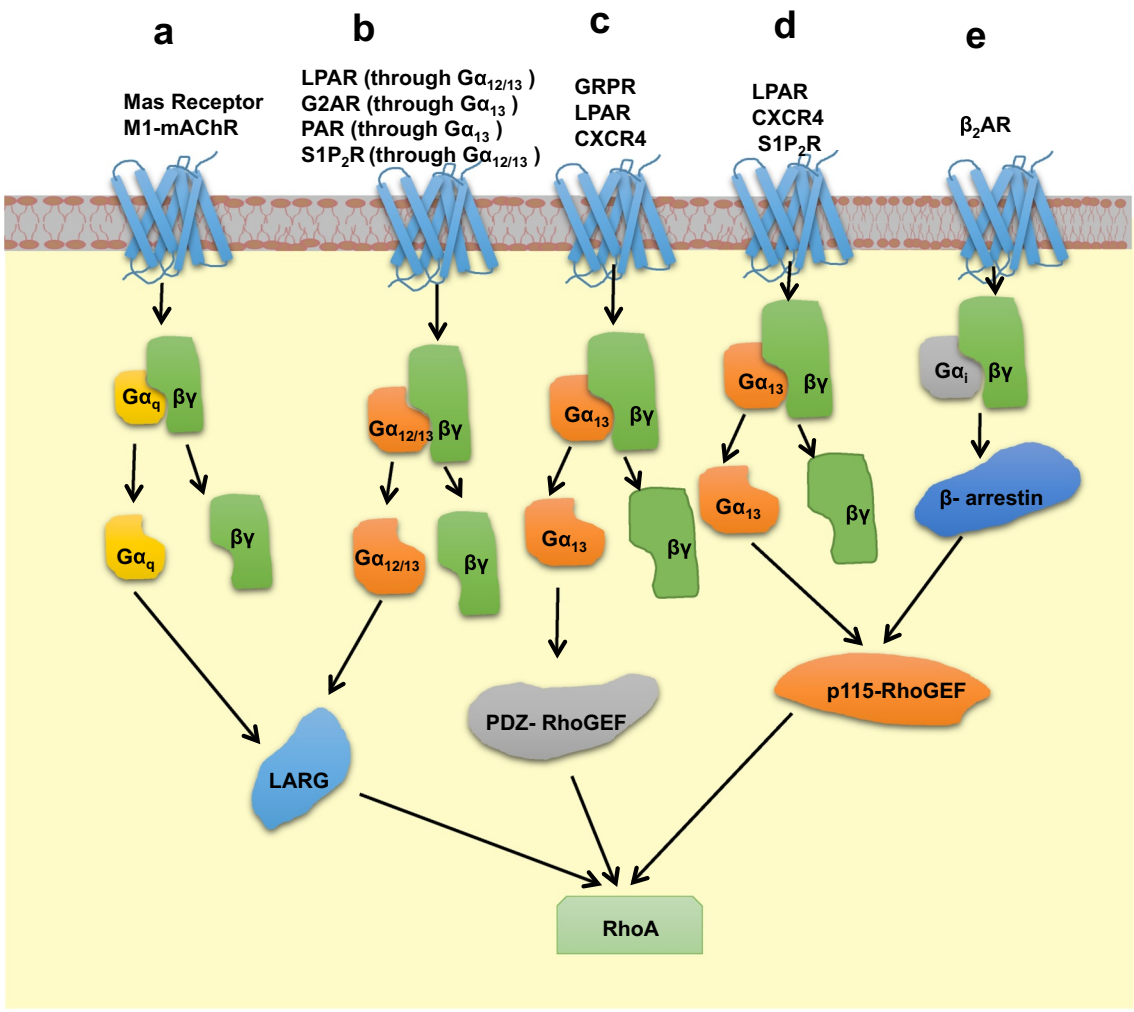


Figure 2. Schematic representation of signalling pathways which leads to activation of GEFs through interaction between heterotrimeric G protein subunits and RH domains of GEFs. (a) GPCRs, Mas receptor and M1-mAChR activate the GEF, LARG, through $G\alpha_q$ binding to its RH domain. (b) Activation of LARG through GPCRs, LPAR & $S1P_2R$ occur through direct interaction with $G\alpha_{12/13}$ whereas, GPCRs, G2A & PAR activate LARG through $G\alpha_{13}$ subunit. (c) PDZ-RhoGEF mediated RhoA activation is established by GRPR, LPAR, and CXCR4 triggered $G\alpha_{13}$ heterotrimeric subunit binding to the RH domain of PDZ-RhoGEF. (d) GPCRs, LPAR, CXCR4, and $S1P_2$ R activate p115-RhoGEF through interaction with $G\alpha_{13}$ heterotrimeric subunit. (e) Upon activation of β_2AR, p115-RhoGEF is triggered by β-arrestin activation through $G\alpha_i\beta\gamma$ signalling.

RhoGEF also activates RhoA signalling in lymphocytes through $G\alpha_{13}$/p115-RhoGEF/RhoA upon CXCR4, LPAR, and S1PR activation [50] (Figure 2d). These are the examples where G protein subunits, $G\alpha_{12/13}$, bind directly to the RH domain containing GEFs to activate them. An exception is seen for p115-RhoGEF in renal cell carcinoma 7 (RCC7) where activated GPCR, β2-adrenergic receptor (β2AR), triggers the GEF indirectly through βArrestin2[55] (Figure 2e). Furthermore, there are examples where PAR and LPA receptors in PC-3 cells are known to specifically activate LARG and PDZ-RhoGEF, respectively. However, the mechanism underlining their specificity is yet to be determined[46]. Recently it has been shown that prostaglandin receptor activated $G\alpha_s$ binds to the PH and DH domains as well as their linker region of PDZ-RhoGEF. This interaction of $G\alpha_s$ with PDZ-RhoGEF renders the GEF active towards Cdc42 [51,52] (Figure 3a).

Activation of GEFs involving interaction between heterotrimeric G protein subunits and non-RH domains of GEFs

G proteins of certain GPCRs are also known to activate GEFs in RH domain independent manner. Examples of these GEFs include A-kinase anchoring protein (AKAP)-Lbc, p114-RhoGEF, p63RhoGEF, Trio, and P-Rex. Details of activation mechanisms of these GEFs by GPCRs will be discussed further. Tissue-specific expression of aforementioned GEFs is given in

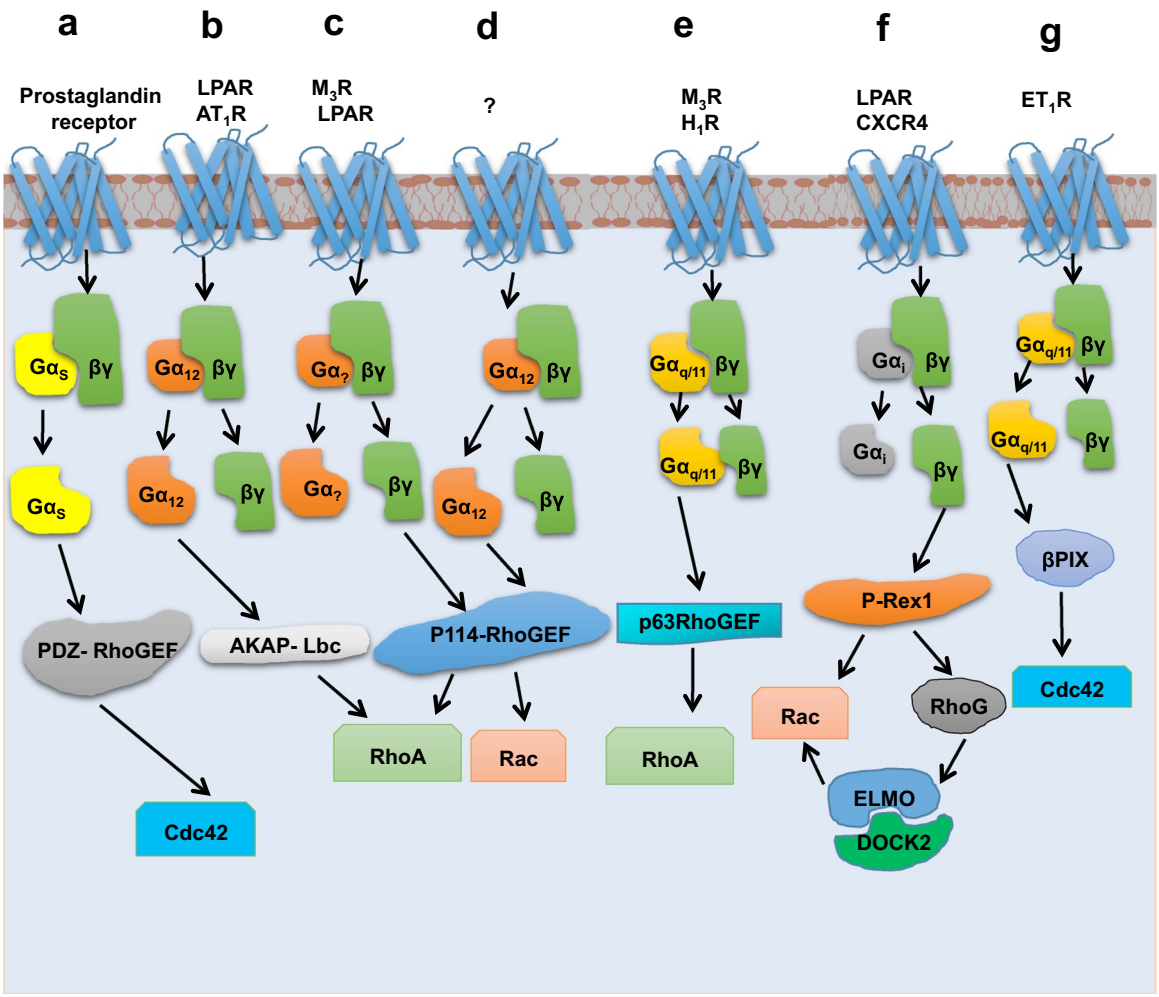


Figure 3. Schematic representation of signalling pathways which leads to activation of GEFs through interaction between heterotrimeric G protein subunits and non-RH domains of GEFs. (a) Stimulation of prostaglandin receptor activates $G\alpha_s$, which in turn binds to PDZ-RhoGEF through DH, PH domains, and their linker region of GEF, thus making it active towards Cdc42. (b) AKAP-Lbc dependent activation of RhoA, by LPAR and AT_1R, happens through the interaction of $G\alpha_{12}$ with the C terminus of AKAP-Lbc. (c) and (d) Activation of Gβγ (through M_3R and LPAR) and $G\alpha_{12}$ triggers p114-RhoGEF, where Gβγ interacts with the C terminus of DH-PH domain of p114-RhoGEF. (e) p63RhoGEF mediated triggering of RhoA upon activation of M_3R and H_1R receptors. This involves the interaction of $G\alpha_{q/11}$ towards the PH domain of p63RhoGEF. (f) Stimulated GPCRs, LPAR, and CXCR4 activate P-Rex1 through the interaction of the Gβγ subunit with the IP4P domain and the tandem PDZ domains of P-Rex1. This may further lead to the activation of ELMO/DOCK2 through P-Rex activated RhoG. (g) Activation of ET_1R triggers βPIX mediated Cdc42 through the interaction of $G\alpha_q$ with βPIX.

Table 1. Activation of LPA receptor and Angiotensin 1 receptor (AT_1R) promotes RhoA mediated signalling in HEK293 cells and adult ventricular fibroblasts (AVFs) through $G\alpha_{12}$/AKAP-Lbc/RhoA pathway [72,73] (Figure 3b). Here, the interaction of AKAP-Lbc with $G\alpha_{12}$ is independent of the RH domain, as this protein lacks this domain[74]. However, the binding of $G\alpha_{12}$ to the GEF depends on a 106 amino acid residue region, stretching from residues 2567–2672 towards the C terminus of AKAP-Lbc[41]. Studies in breast cancer cell lines (MDA-MB-435) show that, through a feedback loop signalling, α6β4 integrin regulates LPA induced AKAP-Lbc/$G\alpha_{12}$/RhoA pathway to promote activation of Rac GTPase[75]. AKAP-Lbc is predominantly localized in the cytoplasm where it is known to regulate spatial distribution of Protein kinase A (PKA) activity during cell migration [72,76].

The next example involving non-RH domain GEF interacting with heterotrimeric G protein includes p114-RhoGEF, activated by GPCRs, M3-muscarinic receptor (M_3R), and LPA receptor operating in HEK293 and NIH3T3 cells, respectively. In these pathways, the Gβγ subunit of G protein interacts with the DH-PH domain of p114-RhoGEF to bring about its activation, eventually leading to RhoA and Rac signalling [77,78] (Figure 3c). Apart from Gβγ, the $G\alpha_{12}$

subunit is also found to interact with p114-RhoGEF (Figure 3d), however, it binds to a region spanning residues from 686 to 791, towards the C terminus of the DH-PH domain [41,77,79]. Furthermore, some evidence suggest that p114-RhoGEF gets triggered both in GPCR dependent and independent manner. For example, it has been shown that tissue-specific apical cell constriction in invertebrates is brought about by non-conserved GPCRs such as Mist (mesoderm-invagination signal transducer) and Smog through the activation of p114-RhoGEF[80]. However, in vertebrates the apical cell constriction process, involving the same p114-RhoGEF, takes place in a GPCR independent manner. In this instance, another GEF DAPLE, which possess Gα-binding-and-activating (GBA) motif could directly activate $G\alpha_i\beta\gamma$ resulting in stimulation of p114-RhoGEF through the Gβγ subunit[81].

Another non-RH domain containing GEF, p63RhoGEF, preferentially expressed in heart and brain tissues, is known to interact with the $G\alpha_{q/11}$ subunit. Here the agonist-stimulated GPCRs like M3-cholinoceptor (M_3R) and the histamine H1 receptor (H_1R), triggers the GEF to activate $G\alpha_{q/11}$ /p63RhoGEF/RhoA pathway [82,83] (Figure 3e). Crystal structure of $G\alpha_{i/q}$ -p63RhoGEF-RhoA shows an amphipathic helical extension of PH domain (residues 471–485) of the GEF interacting with $G\alpha_{i/q}$ [83,84]. Thus, it can be proposed that the autoinhibitory state of the GEF due to the inter-domain interaction invloving PH domain [85,86] is released upon its engagement with $G\alpha_{i/q}$. Furthermore, in fibroblast and COS-7 cells, it is shown that M_3R activated p63RhoGEF competes with phospholipase C (PLCβ) for binding $G\alpha_q$[83]. Similar to p63RhoGEF, the Triple function domain protein (Trio) also contains an inhibitory C terminal extension beyond its PH domain[85]. Trio exhibits dual Rac/RhoA GEF activity[87]. $G\alpha_q$ subunit renders Trio active towards RhoA by binding the inhibitory PH domains and hence relieving its autoinhibition[85]. However, the GPCR which is involved in this pathway has not yet been identified.

In MCF-7 breast cancer cells, chemotactic GPCRs, CXCR4 and LPAR, activate Rac through the Gβγ/PIP3-dependent Rac exchanger1 (P-Rex1)/Rac pathway (Figure 3f). Here the $G\alpha_q$ or $G\alpha_{13}$ but not $G\alpha_s$ or $G\alpha_i$ form stable complex with Gβγ. Therefore, the activation of P-Rex1 could be through CXCR4 and LPAR, preferably via $G\alpha_i\beta\gamma$ dependent fashion [88–90]. From cryo-electron microscopy structure it has been suggested that Gβγ interaction with P-Rex1 is a multidomain assembly involving IP4P and tandem PDZ domains of the GEF[91]. Operating under the same pathway, P-Rex1 is also known to activate RhoG leading to the RhoG/ELMO/DOCK2 dependent activation of Rac [92] (Figure 3f).

Activation of GEFs through Phosphorylation, protein–lipid, and protein–protein interactions mediated by GPCRs

Phosphorylation is another means through which GPCRs activate GEFs, wherein these receptors activate downstream kinases that in turn trigger GEFs. One such example is p21-activated kinases interacting exchange proteins (PIX) which, exist in two isoforms αPIX and βPIX and are known to activate Cdc42 and Rac1 GTPases[13]. αPIX is predominantly expressed in haematopoietic and muscle cells whereas, βPIX is expressed ubiquitously. Splice variants of these proteins are majorly found in the brain [13,93]. It has been shown that activation of βPIX involves $G\alpha_i$ of LPA receptor in HeyA8 ovarian cancer cell lines, where $G\alpha_i$ recruits Src kinase, which phosphorylates Tyr-442 residue of βPIX to activate the GEF [94–96] (Figure 4a). Apart from Src, other kinases like p21-activated kinases (PAK) are also known to phosphorylate numerous serine and tyrosine residues of βPIX, especially the residues at its PH, GITI-binding, N-terminal spacer region (amino acid 61–99), and C terminal coiled-coil domains. Other than GEF activation, phosphorylation at many of these sites could have important roles in regulating βPIX interaction with other signal transduction proteins[97]. Another kinase, PKA, is shown to phosphorylate βPIX primarily at Ser-516 and Thr-526, when triggered through Endothelin 1 receptor (ET_1R) induced $G\alpha_s$ subunit[98]. Furthermore, βPIX is also known to interact with Endothelin 1 receptor (ET_1R) activated $G\alpha_q$ subunits to trigger Cdc42[99] (Figure 3g). Though these studies suggest that the direct binding of the $G\alpha_q$ subunit is required for βPIX activation[99]. But, the role of phosphorylation in this case is not clear.

Although relatively little is known about phosphorylation induced activation of αPIX by PAK, it has been suggested that PAK triggers this GEF by releasing it from its inactive dimer state [13,100]. This pathway involves activation of Gβγ through formyl-methionyl-leucyl-phenylalanine receptor (fMLPR), which further triggers a cascade of signalling events, propagating through p21-activated kinase 1 (PAK1). Further, PAK1 interacts with the SH3 domain of the GEF to release αPIX from its dimeric state to its active monomeric state which further triggers Cdc42 [13,100]. Activated Cdc42 is also known to enhance the affinity of this GEF towards Rac1 (Figure 4b). Thus, it appears that there is a mutual feedforward mechanism between these two GTPases. On one hand, binding of Cdc42 to one of the DH domains of the dimeric αPIX introduces

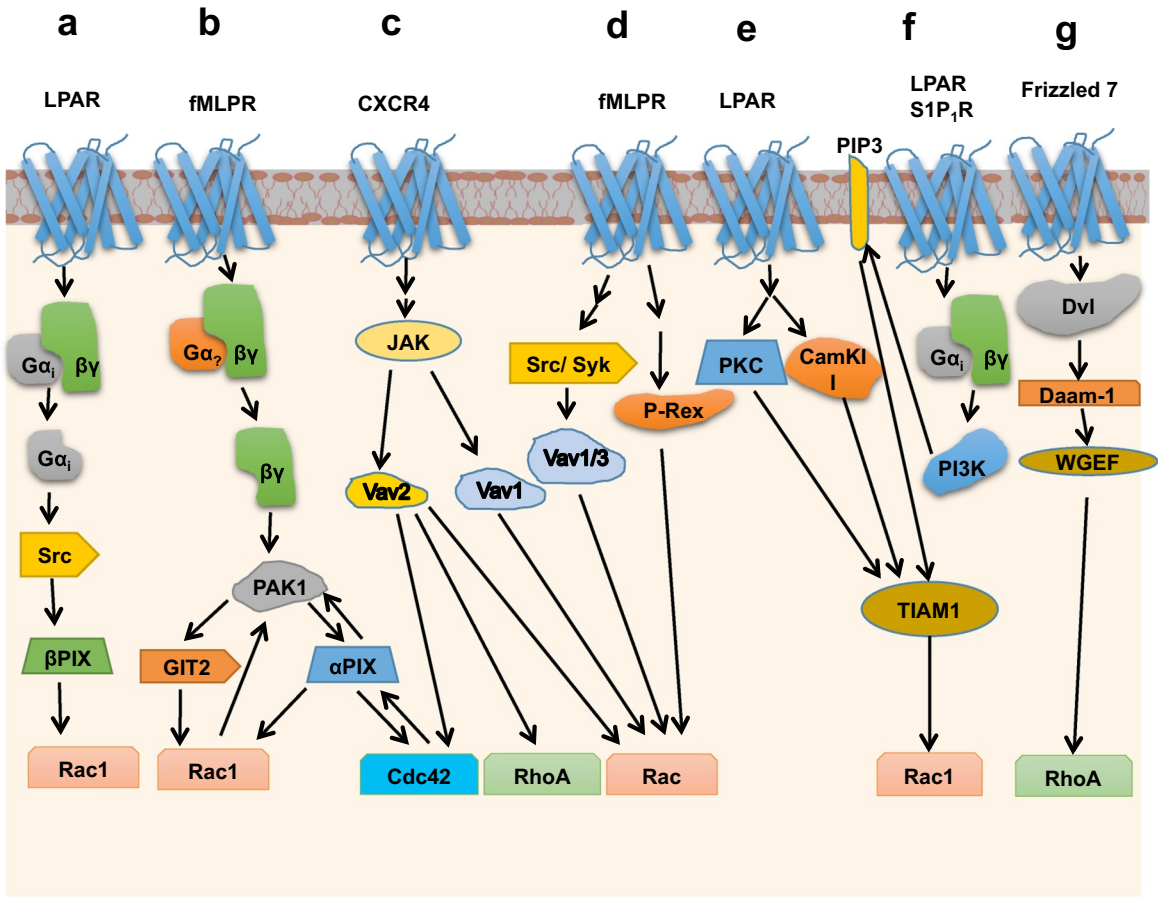


Figure 4. Activation of GEFs through Phosphorylation, protein-protein, and lipid-mediated interactions involving downstream of GPCR signalling. (a) Upon LPAR stimulation, activation of βPIX takes place through, $G\alpha_i$ mediated Src kinase activation, which further phosphorylates βPIX. (b) Activation of fMLPR triggers Cdc42 through Gβγ-PAK1-αPIX and Rac1 through Gβγ-PAK1-αPIX-GIT2 sequential complex formation, where activation of αPIX takes place through PAK1 binding. Cdc42 and Rac1 are involved in feedforward mechanism where αPIX and Rac1 enhance PAK1 activity. Here activation of Cdc42 enhances GEF activity of αPIX towards Rac1. (c) CXCR4 activates Vav1 and Vav2 through JAK signalling. (d) Activation of Rac through fMLPR stimulation involves concerted signalling between Vav1/3 and P-Rex GEFs. (e) LPAR triggers Rac1 through PKC and CamKII mediated phosphorylation of TIAM1. (f) Stimulation of GPCRs, LPAR and $S1P_1R$, activate TIAM through $G\alpha_i\beta\gamma$/PI3K/TIAM1/Rac1 pathway, where TIAM1 interacts with a PI3K lipid product, PIP3. (g) Protein–protein interaction mediated triggering of RhoA when activated Frizzled7 recruits Dishevelled (Dvl) and Daam-1 proteins which activate the Rho WGEF.

conformational changes in the GEF to enhance its activity towards Rac1[100][101], whereas, on the other hand, in a feedforward mechanism, αPIX enhances kinase activity of PAK1 to further activate Cdc42 [102]. Furthermore, αPIX is also known to trigger Rac1 in fMLPR controlled pathway. Here, fMLPR triggers to form a quaternary complex Gβγ-PAK1-αPIX-GIT2 in a sequential manner to recast its specificity for Rac1 (Figure 4b). During this pathway, the role of phosphorylation in the activation of αPIX is not clear, but it has been shown that activation of αPIX could take place through binding to phosphatidylinositol 3, 4, 5-trisphosphate (PIP3), a lipid product of PI3-kinase [103]. Similar to Cdc42, Rac1 is also suggested to operate in the feedforward cycle to further enhance the activity of PAK1 towards the GEF[104].

Vav family of GEFs provide another example for activation of GEFs, through phosphorylation relay triggered by GPCRs like CXCR4 and fMLPR [42,105]. This family of GEF consists of 3 members, namely, Vav1 (specific to Rac1/2, RhoG, and Cdc42), Vav2 (specific to Rac1, RhoA, RhoG, and Cdc42), and Vav3 (specific to Rac1, RhoG, and RhoA) [106,107] and their tissue-specific expression is shown in Table 1. Out of these GEFs, Vav1 and Vav2 are activated by pathways involving Janus kinase2 (JAK2), Janus kinase3 (JAK3) and the receptor involved in this particular signalling is CXCR4[42]. It is interesting to know that activation of this GEF through JAK is independent of GPCR-$G\alpha_i$ pathway [42,108] (Figure 4c). In P1V1 neutrophils, through an unknown mechanism, fMLPR leads to activation of P-Rex1 and Vav1/Vav3[105] GEFs (Figure 4d).

However, the role of GPCRs, CXCR4 and fMLPR, in the activation of Vav proteins is not clear. For Vav1 it has been surmised that fMLPR could activate Src or Syk kinases through downstream signalling to trigger this RacGEF [109,110].

The T-cell lymphoma invasion and metastasis inducing factor 1 (TIAM1) provides a good example of phosphorylation mediated activation of GEF. Through Small-angle X-ray scattering (SAXS) based model it was suggested that the initial 50 amino acid domain (N50), N terminal PH domain (PHn), and the coiled-coil extension (CC-Ex) domain of TIAM1 interact with each other to render the GEF inactive[32]. Release of this autoinhibition involves multiple steps including phosphorylation of TIAM1 by protein kinase C (PKCα) and Ca2+/calmodulin-dependent protein kinase II (CamKII) (Figure 4e). In Swiss 3T3 fibroblast cells, this phosphorylation pathway is triggered by the LPA receptor[111]. Based on various experimental studies it has been proposed that sequential phosphorylation of Ser29 and Ser33 of TIAM1 by atypical C kinase (aPKCγ)[32], followed by phosphorylation of Tyr829 by TrkB kinase[112], disrupt the intra-domain interactions of the GEF to make it active [32,113]. Apart from phosphorylation, there exist other modes of TIAM1 activation including protein–protein interactions involving PHn-CC-Ex domain of TIAM1 with Par3 in the Par polarity complex (Par3, Par6 and aPKC) [32,114,115].

It has also been proposed that lipid-mediated interactions too can activate TIAM1. For example GPCRs, LPA_1R and $S1P_1R$, induced $G\alpha_i\beta\gamma$/PI3K/TIAM1/Rac1 pathway promote binding of PI3K lipid product, phosphatidylinositol 3, 4, 5-trisphosphate (PIP3)[116], to the PH domain of the GEF to make it active [12,117] (Figure 4f). However, complete activation of TIAM1 requires additional signals from $G\alpha_i\beta\gamma$[12]. Although TIAM2/STEF is relatively less characterized compared to TIAM1, it has been suggested that the activation of this GEF is also a complex multistep process. Forster resonance energy transfer based biosensor studies on dibutyryl cAMP (dbcAMP) treated PC12D cells has revealed that the phosphorylation of Thr-749 by dbcAMP activated PKA at the plasma membrane enhances the GEF activity of TIAM2 towards Rac1[118].

Activation of GEFs through direct interaction of GPCR domains

Few reports suggest that GPCRs could activate GEFs by directly interacting with them, instead of utilizing the heterotrimeric G proteins. For example, PDZ-RhoGEF and LARG interact with the C terminus of LPA_1R and LPA_2R receptors through their PDZ domains (Figure 5a) to get activated[43]. Similarly, activation of Vav2 involves interaction of the C terminus of the Parathyroid hormone receptor (PTHR) with its DH and PH domains. It is interesting to note that Vav2 competes with PTHR to bind $G\alpha_q$ subunit and also interferes with the PTHR signalling[119] (Figure 5b). Other than these examples, there is not much information on how GPCRs could directly activate GEFs.

Non-canonical GPCRs activate GEFs through the scaffold protein Dishevelled (Dvl)

There are about 11 non-canonical GPCRs identified as Frizzled (FZD) class receptors. Other than smoothened (SMO), all other members of this family are activated by secreted glycoproteins, Wingless/Int1 (Wnt). There are 19 Wnts identified in humans. A combination of interactions between FZDs and Wnts is known to trigger diverse signalling pathways classified as canonical and non-canonical Wnt signalling pathways[120]. However, the exact mapping between specific Wnt-FZD interactions to a particular pathway is not clear. The common feature in the non-canonical FZD GPCR pathway is the recruitment of a scaffold protein dishevelled (Dvl) at the cytosolic end of the receptor upon binding of Wnts at their extracellular domain. It has been shown that Dvl forms higher oligomers to act as a signalling node, known as signalosome to further propagate the signal[121]. With reference to GEF activation, Frizzled 7 (FZD7) is known to activate RhoA through WGEF (Weakly similar to Rho GEF 5). In *Xenopus* embryos, it has been shown that activation of FZD7 leads to recruitment of Dvl-2, which subsequently interacts with Dishevelled-associated activator of morphogenesis 1 (Daam-1) protein. This signal further propagates to activate WGEF. However, the exact mechanism of activation is not clear[122] (Figure 4g). It is interesting to note that WGEF is found to interact directly with the PDZ domain of Dvl-2 and with the N terminal domain of Daam-1 leading to its activation. Thus, it has been suggested that FZD7 could operate through multiple modes to activate WGEF[122].

Activation of non-canonical GEFs through GPCRs

As mentioned previously, the non-canonical Dedicator of cytokinesis (DOCK) family GEFs, also known as Docker GEFs, contain DHR-1 and DHR-2 domains, in place of PH and DH domains of canonical GEFs.

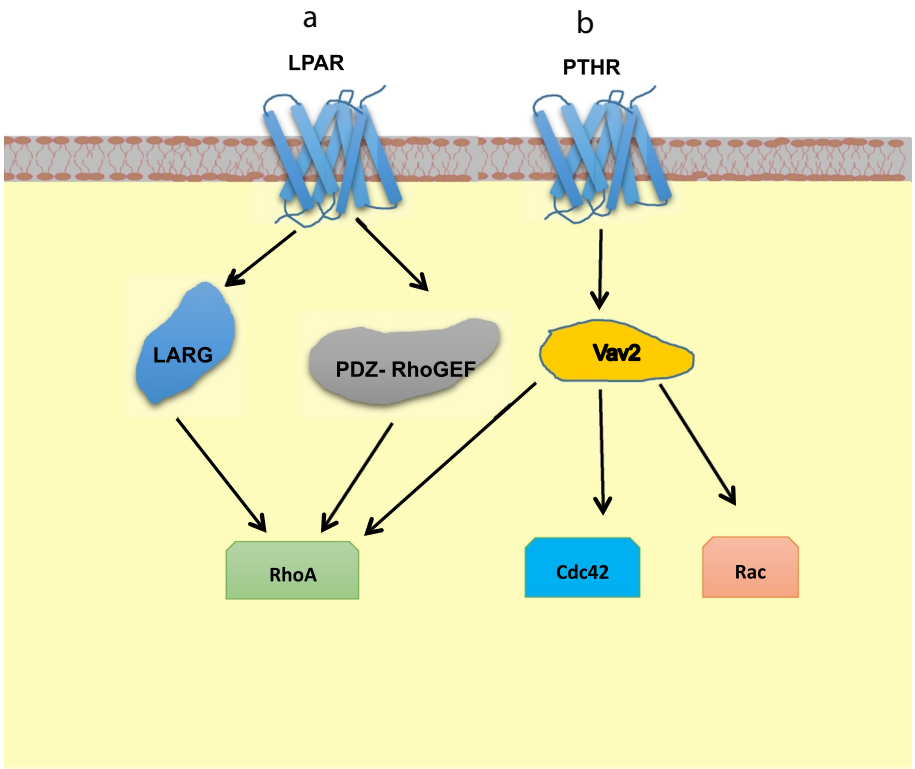


Figure 5. Schematic representation of activation of GEFs upon direct interaction with GPCR domains. (a) GEFs, LARG, and PDZ-RhoGEF activate RhoA signalling upon interacting with the C terminus of LPAR through their PDZ domains. (b) Similarly, Vav2 interacts with the C terminus of PTHR through its DH and PH domains.

Similar to canonical GEFs, Docker GEFs are known to be activated through phosphorylation, protein–protein interactions, and/or lipid-mediated interactions. In HL60 and HeLa cell lines, activation of the chemokine receptor, CXCR4, triggers activation of Rac through Gβγ/ELMO1/DOCK180/Rac1 pathway. Similarly, in MDA-MB-231 cell lines, CXCR4 activates Rac through $G\alpha_i$/ELMO1/DOCK180/Rac1 pathway[69][123], (Figure 6a). In both of these pathways Gβγ and $G\alpha_i$ subunits were found to interact with ELMO1, wherein $G\alpha_i$ interacts with N terminus of ELMO1,[69][123], leading to conformational changes in the ELMO-DOCK complex, resulting in the release of DOCK from its autoinhibitory state. Another mechanism for DOCK180 activation by LPA is through a sequential complexation of $G\alpha_{i2}$-src-p130Cas/ELMO/DOCK180[94] [124,125] (Figure 6b), in which Src kinase is shown to phosphorylate DOCK180 at Tyr1811, enhancing the interaction of DOCK180 with CrkII and p130Cas proteins, which activates Rac[126].

DOCK2, another member of the DOCK family is shown to be activated through lipid-mediated interaction. This process involves fMLPR induced association of DHR-1 domain of DOCK2 with phosphatidylinositol 3, 4, 5-trisphosphate (PIP_3) in a PI3K dependent manner [127,128], wherein interaction with PIP_3 triggers GEF activity of DOCK2 towards Rac (Figure 6c). It has been also suggested that activation of chemokine receptors triggers DOCK2 through $G\alpha_i$ subunit, but the exact mechanism is unclear[129]. In apoptotic cells, Brain-specific angiogenesis inhibitor 1 receptor (BAI1) is shown to trigger ELMO/DOCK/Rac signalling (Figure 6d). These signalling events depend on the interaction of cytosolic fragment of BAI1 (aa 1431–1582) with N terminal fragments of ELMO1 (aa 1–558) and ELMO2 (aa 1–520), which subsequently binds to DOCK1/DOCK2 [130,131]. Although the activation of non-canonical GEFs is similar to canonical GEFs, the pathways that operate are markedly different.

Summary

Activation of GTPases like RhoA, Cdc42, and Rac is a crucial step in cell migration, as they are the key players controlling the cytoskeletal rearrangement. This process is highly regulated in space and time by employing GEFs and GAPs. GEFs provide one of the important steps of regulation, wherein they switch the GTPases 'ON' by exchanging GDP with GTP. Based on sequence and structure GEFs can be classified as canonical and non-canonical. The common feature of both canonical and non-canonical GEFs is that they exist in

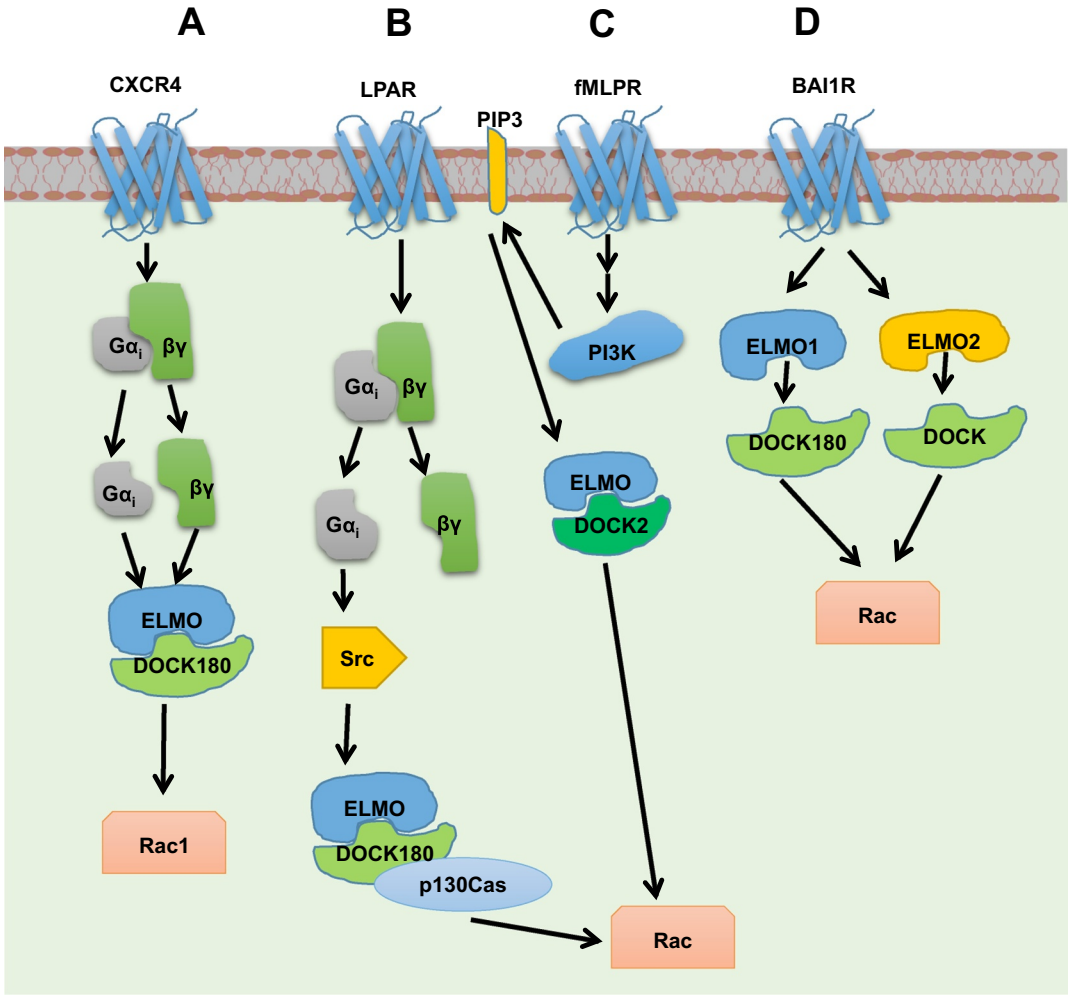


Figure 6. Schematic representation of activation of non-canonical GEFs by GPCRs. (a) CXCR4 activates ELMO1/DOCK180 through the interaction of ELMO1 with Gβγ and $G\alpha_i$ subunits. (b) Stimulation of LPAR leads to Rac activation through $G\alpha_{i2}$-Src-p130Cas/ELMO/DOCK180/Rac1 pathway during which Src is proposed to phosphorylate DOCK180 at Tyr1811. (c) ELMO/DOCK2 mediated Rac activation takes place through binding of DHR-1 domain of DOCK2 to lipid, PIP3 upon fMLPR activation. (d) ELMO1 and ELMO2 are shown to interact with the C terminal of the BAI1 receptor through their N-terminal fragments to trigger the DOCK/Rac pathway.

an autoinhibitory inactive state (Figure 1b-J) and release of this autoinhibitory state is essential for them to activate their cognate GTPases. The GEF activation process is done through multiple modes like phosphorylation of GEFs or interaction of GEFs with a different scaffold or activating proteins or through lipid interactions. In this review, we have consolidated how GPCRs play a crucial role in activating GEFs by exerting these activation mechanism. These events include the interaction of GEFs (through their RH and non-RH domains) with heterotrimeric G protein subunits activated by GPCRs, activation of kinases that phosphorylate GEFs, and protein/lipid interactions that release GEFs from their autoinhibited state. Complete understanding of these signalling events at the atomic level is crucial to decipher signalling processes, as this would be highly rewarding in addressing pathological conditions like cancer metastasis, immunological disorders, and neurodegenerative diseases.

Although the field of cell signalling is rapidly evolving, there are significant gaps in our understanding of signalling events that control cell migration, especially with reference to the role of GPCRs. Majority of the efforts, to date, are on mapping the key players involved in the signal transduction, which is largely driven by gene knockout/silencing and protein pull-down studies. However, the key questions like how in real-time the signals transverse from GPCRs to the downstream proteins? how concomitantly operating diverse pathways synchronize to exert cell migration?

and how few GPCRs autonomously operate to control divergent pathways? remain unaddressed. Perhaps, innovative real-time imaging techniques, involving bioluminescence resonance energy transfer and fluorescence resonance energy transfer would help to dissect the system. Although some efforts have been made in this direction [132–135], given the emerging complexity of the system, unravelling the signalling events that control cell migration remains challenging.

Acknowledgments

KK thank the ICGEB Research Grant [CRP/IND15] for the grant and AO would like to acknowledge the Council of Scientific & Industrial Research, India for fellowship.

Disclosure statement

The authors declare no conflicts of interest.

Funding

This work was supported by the International Center for Genetic Engineering and Biotechnology [CRP/IND15].

ORCID

Aishwarya Omble http://orcid.org/0000-0001-7091-4578
Kiran Kulkarni http://orcid.org/0000-0002-7857-8932

References

[1] Trepat X, Chen Z, Jacobson K. Cell migration. Compr Physiol. 2012;2(4):2369–2392.
[2] Zegers MM, Friedl P. Rho GTPases in collective cell migration. Small GTPases 2014; 5:e983869.
[3] Paoletti S, Petkovic V, Sebastiani S, et al. A rich chemokine environment strongly enhances leukocyte migration and activities. Blood. 2005;105 (9):3405–3412.
[4] Peplow PV, Chatterjee MP. A review of the influence of growth factors and cytokines in in vitro human keratinocyte migration. Cytokine. 2013;62(1):1–21.
[5] Coombs VA, Nissen BK, Marks R, et al. The influence of temperature on epidermal cell migration in vitro. Br J Exp Pathol. 1973;54(6): 673–677. PMID: 4798773.
[6] Su C, Zhang B, Liu W, et al. High extracellular pressure promotes gastric cancer cell adhesion, invasion, migration and suppresses gastric cancer cell differentiation. Oncol Rep. 2016;36(2):1048–1054.
[7] Ridley AJ, Schwartz MA, Burridge K, et al. Cell migration: integrating signals from front to back. Science. 2003;302(5651):1704–1709.
[8] Martin K, Reimann A, Fritz RD, et al. Spatio-temporal co-ordination of RhoA, Rac1 and Cdc42 activation during prototypical edge protrusion and retraction dynamics. Sci Rep. 2016;6(1):1–14.
[9] Ridley AJ, Hall A. The small GTP-binding protein rho regulates the assembly of focal adhesions and actin stress fibers in response to growth factors. Cell. 1992;70(3):389–399.
[10] Nobes CD, Hall A. Rho 1. nobes CD, hall A. Rho GTPases control polarity, protrusion, and adhesion during cell movement. J Cell Biol. 1999;144(6):1235–1244.
[11] Nobes CD, Rho HA. Rac, and Cdc42 GTPases regulate the assembly of multimolecular focal complexes associated with actin stress fibers, lamellipodia, and filopodia. Cell. 1995;81(1):53–62.
[12] Van Leeuwen FN, Olivo C, Grivell S, et al. Rac activation by lysophosphatidic acid LPA1 receptors through the guanine nucleotide exchange factor Tiam1. J Biol Chem. 2003;278(1):400–406.
[13] Manser E, Loo TH, Koh CG, et al. PAK kinases are directly coupled to the PIX family of nucleotide exchange factors. Mol Cell. 1998;1(2):183–192.
[14] Bos JL, Rehmann H, Wittinghofer A. GEFs and GAPs: critical elements in the control of small G proteins. Cell. 2007;129(5):865–877.
[15] Cho HJ, Kim BY, Baek KE, et al. Regulation of Rho GTPases by RhoGDIs in human cancers. Cells. 2019;8 (9):1–12.
[16] Buchsbaum RJ. Rho activation at a glance. J Cell Sci. 2007;120(7):1149–1152.
[17] Dovas A, Couchman JR. RhoGDI: multiple functions in the regulation of Rho family GTPase activities. Biochem J. 2005;390(1):1–9.
[18] Goicoechea SM, Awadia S, Garcia-Mata R. I'm coming to GEF you: regulation of RhoGEFs during cell migration. Cell Adhes Migr. 2014;8(6):535–549.
[19] Zheng Y, Quilliam LA. Activation of the Ras superfamily of small GTPases.Workshop on exchange factors. EMBO Rep. 2003;4(5):463–468.
[20] Hartmann S, Ridley AJ, Lutz S. The function of rho-associated kinases ROCK1 and ROCK2 in the pathogenesis of cardiovascular disease. Front Pharmacol. 2015;6:276.
[21] Hilger D, Masureel M, Kobilka BK. Structure and dynamics of GPCR signaling complexes. Nat Struct Mol Biol. 2018;25(1):4–12.
[22] Heng BC, Aubel D, Fussenegger M. An overview of the diverse roles of G-protein coupled receptors (GPCRs) in the pathophysiology of various human diseases. Biotechnol Adv. 2013;31:1676–1694.
[23] Du Z, Lovly CM. Mechanisms of receptor tyrosine kinase activation in cancer. Mol Cancer. 2018;17:1–13.
[24] Hamidi H, Ivaska J. Every step of the way: integrins in cancer progression and metastasis. Nat Rev Cancer. 2018;18(1):1–16.
[25] Cotton M, Claing A. G protein-coupled receptors stimulation and the control of cell migration. Cell Signal. 2009;21(7):1045–1053.
[26] Rossman KL, Der CJ, Sondek J. GEF means go: turning on Rho GTPases with guanine nucleotide-exchange factors. Nature Reviews Molecular Cell Biology. 2005;6(2):167–180.
[27] Viaud J, Lagarrigue F, Ramel D, et al. Phosphatidylinositol 5-phosphate regulates invasion through binding and activation of Tiam1. Nat Commun. 2014;5(1):1–17.

[28] Baumeister MA, Martinu L, Rossman KL, et al. Loss of phosphatidylinositol 3-phosphate binding by the C-terminal Tiam-1 pleckstrin homology domain prevents in vivo Rac1 activation without affecting membrane targeting. J Biol Chem. 2003;278 (13):11457–11464.
[29] Schmidt A, Hall A. Guanine nucleotide exchange factors for Rho GTPases: turning on the switch. Genes Dev. 2002;16(13):1587–1609.
[30] Consonni SV, Brouwer PM, Van Slobbe ES, et al. The PDZ domain of the guanine nucleotide exchange factor PDZGEF directs binding to phosphatidic acid during brush border formation. PLoS One. 2014;9(5):e98253.
[31] Fukuharaa S, Chikumi H, Silvio Gutkind J. RGS-containing RhoGEFs: the missing link between transforming G proteins and Rho? Oncogene. 2001;20 (13):1661–1668.
[32] Xu Z, Gakhar L, Bain FE, et al. The Tiam1 guanine nucleotide exchange factor is autoinhibited by its pleckstrin homology coiled-coil extension domain. J Biol Chem. 2017;292(43):17777–17793.
[33] Premkumar L, Bobkov AA, Patel M, et al. Structural basis of membrane targeting by the Dock180 family of Rho family guanine exchange factors (Rho-GEFs). J Biol Chem. 2010;285(17):13211–13222.
[34] Brugnera E, Haney L, Grimsley C, et al. Unconventional Rac-GEF activity is mediated through the Dock180-ELMO complex. Nat Cell Biol. 2002;4 (8):574–582.
[35] Kulkarni K, Yang J, Zhang Z, et al. Multiple factors confer specific Cdc42 and Rac protein activation by dedicator of cytokinesis (DOCK) nucleotide exchange factors. J Biol Chem. 2011;286(28):25341–25351.
[36] Chang L, Yang J, Jo CH, et al. Structure of the DOCK2 –ELMO1 complex provides insights into regulation of the auto-inhibited state. Nat Commun. 2020;11 (1):1–17.
[37] Lu M, Kinchen JM, Rossman KL, et al. A steric-inhibition model for regulation of nucleotide exchange via the Dock180 family of GEFs. Curr Biol. 2005;15(4):371–377.
[38] Chen Z, Guo L, Hadas J, et al. Activation of p115-RhoGEF requires direct association of Gα13 and the Dbl homology domain. J Biol Chem. 2012;287 (30):25490–25500.
[39] Patel M, Karginov AV. Phosphorylation-mediated regulation of GEFs for RhoA. Cell Adhes Migr. 2014;8 (1):11–18.
[40] Booden MA, Siderovski DP, Der CJ. Leukemia-associated Rho Guanine nucleotide exchange factor promotes Gαq-coupled activation of RhoA. Mol Cell Biol. 2002;22(12):4053–4061.
[41] Martin JW, Cavagnini KS, Brawley DN, et al. A Gα12-specific binding domain in AKAP-Lbc and p114RhoGEF. J Mol Signal. 2016;11:1–17.
[42] Bartolomé RA, Molina-Ortiz I, Samaniego R, et al. Activation of Vav/Rho GTPase signaling by CXCL12 controls membrane-type matrix metalloproteinase-dependent melanoma cell invasion. Cancer Res. 2006;66(1):248–258.
[43] Yamada T, Ohoka Y, Kogo M, et al. Physical and functional interactions of the lysophosphatidic acid receptors with PDZ domain-containing Rho guanine nucleotide exchange factors (RhoGEFs). J Biol Chem. 2005;280(19):19358–19363.
[44] Kourlas PJ, Strout MP, Becknell B, et al. Identification of a gene at 11q23 encoding a guanine nucleotide exchange factor: evidence for its fusion with MLL in acute myeloid leukemia. Proc Natl Acad Sci U S A. 2000;97(5):2145–2150.
[45] Medlin MD, Staus DP, Dubash AD, et al. Sphingosine 1-phosphate receptor 2 signals through leukemia-associated RhoGEF (LARG), to promote smooth muscle cell differentiation. Arterioscler Thromb Vasc Biol. 2010;30(9):1779–1786.
[46] Wang Q, Liu M, Kozasa T, et al. Thrombin and lysophosphatidic acid receptors utilize distinct rhoGEFs in prostate cancer cells. J Biol Chem. 2004;279 (28):28831–28834.
[47] Goulimari P, Knieling H, Engel U, et al. LARG and mDia1 Link Gα 12/13 to Cell Polarity and Microtubule Dynamics. Mol Biol Cell. 2008;19(1):30–40.
[48] Patel M, Kawano T, Suzuki N, et al. Gα13/PDZ-RhoGEF/RhoA signaling is essential for gastrin- releasing peptide receptor – mediated colon cancer cell migration. Molecular Pharmacology. 2014;86 (3):252–262.
[49] Struckhoff AP, Rana MK, Kher SS, et al. PDZ-RhoGEF is essential for CXCR4-driven breast tumor cell motility through spatial regulation of RhoA. J Cell Sci. 2013;126(19):4514–4526.
[50] Mathew D, Kremer KN, Torres RM. ARHGEF1 deficiency reveals Gα13-associated GPCRs are critical regulators of human lymphocyte function. J Clin Invest. 2019;129(3):965–968.
[51] Castillo-Kauil A, García-Jiménez I, Cervantes-Villagrana RD, et al. Gαsdirectly drives PDZ-RhoGEF signaling to Cdc42. J Biol Chem. 2020;295 (50):16920–16928.
[52] Slepak VZ, Pronin A. A Gs-RhoGEF interaction: an old G protein finds a new job. J Biol Chem. 2020;295 (50):16929–16930.
[53] Fukuhara S, Murga C, Zohar M, et al. A novel PDZ domain containing guanine nucleotide exchange factor links heterotrimeric G proteins to Rho. J Biol Chem. 1999;274(9):5868–5879.
[54] Scarlett KA, White ESZ, Coke CJ, et al. Agonist-induced CXCR4 and CB2 heterodimerization inhibits Gα13/RhoA-mediated migration. Mol Cancer Res. 2018;16(4):728–739.
[55] Ma X, Zhao Y, Daaka Y, et al. Acute activation of β2-adrenergic receptor regulates focal adhesions through βArrestin2- and p115RhoGEF protein-mediated activation of RhoA. J Biol Chem. 2012;287 (23):18925–18936.
[56] Chen Z, Guo L, Sprang SR, et al. Modulation of a GEF switch: autoinhibition of the intrinsic guanine nucleotide exchange activity of p115-RhoGEF. Protein Sci. 2011;20:(1):107–117.
[57] Appert-Collin A, Cotecchia S, Nenniger-Tosato M, et al. The A-kinase anchoring protein (AKAP)-Lbc-signaling complex mediates α1 adrenergic receptor-induced cardiomyocyte hypertrophy. Proc Natl Acad Sci U S A. 2007;104(24):10140–10145.

[58] Turton KB, Wilkerson EM, Hebert AS, et al. Expression of novel "LOCGEF" isoforms of ARHGEF18 in eosinophils. J Leukoc Biol. 2018;104 (1):135–145.
[59] Cervantes-Villagrana RD, Color-Aparicio VM, Reyes-Cruz G, et al. Protumoral bone marrow-derived cells migrate via Gβγ-dependent signaling pathways and exhibit a complex repertoire of RhoGEFs. J Cell Commun Signal. 2019;13(2):179–191.
[60] Jaiswal M, Dvorsky R, Ahmadian MR. Deciphering the molecular and functional basis of Dbl family proteins: a novel systematic approach toward classification of selective activation of the Rho family proteins. J Biol Chem. 2013;288(6):4486–4500.
[61] Welch HCE Regulation and function of P-Rex family Rac-GEFs. 2015; 6:49–70.
[62] Niggli V. Signaling to migration in neutrophils: importance of localized pathways. Int J Biochem Cell Biol. 2003;35(12):1616–1638.
[63] Mazaki Y, Hashimoto S, Tsujimura T, et al. Neutrophil direction sensing and superoxide production linked by the GTPase-activating protein GIT2. Nat Immunol. 2006;7(7):724–731.
[64] Premont RT, Perry SJ, Schmalzigaug R, et al. The GIT/PIX complex: an oligomeric assembly of GIT family ARF GTPase-activating proteins and PIX family Rac1/Cdc42 guanine nucleotide exchange factors. Cell Signal. 2004;16(9):1001–1011.
[65] Bustelo XR. The VAV family of signal transduction molecules. Crit. Rev. Oncog. 1996;7(1–2):65–88.
[66] Razidlo GL, Schroeder B, Chen J, et al. Vav1 as a central regulator of invadopodia assembly. Curr Biol. 2014;24(1):86–93.
[67] Fujikawa K, Miletic AV, Alt FW, et al. Vav1/2/3-null Mice Define an Essential Role for Vav Family Proteins in Lymphocyte Development and Activation but a Differential Requirement in MAPK Signaling in T and B Cells. J Exp Med. 2003;198(10):1595–1608.
[68] Shepherd TR, Hard RL, Murray AM, et al. Distinct ligand specificity of the Tiam1 and Tiam2 PDZ domains. Biochemistry. 2011;50(8):1296–1308.
[69] Wang Y, Xu X, Pan M, et al. ELMO1 directly interacts with Gβγ subunit to transduce GPCR signaling to Rac1 activation in chemotaxis. J Cancer. 2016;7(8):973–983.
[70] Kunimura K, Uruno T, Fukui Y. DOCK family proteins: key players in immune surveillance mechanisms. Int Immunol. 2020;32(1):5–15.
[71] Fukui Y, Hashimoto O, Sanui T, et al. Haematopoietic cell-specific CDM family protein DOCK2 is essential for lymphocyte migration. Nature. 2001;412 (6849):826–831.
[72] Diviani D, Soderling J, Scott JD. AKAP-Lbc anchors protein kinase a and nucleates Gα12-selective Rho-mediated stress fiber formation. J Biol Chem. 2001;276(47):44247–44257.
[73] Cavin S, Maric D, Diviani D. A-kinase anchoring protein-Lbc promotes pro-fibrotic signaling in cardiac fibroblasts. Biochim Biophys Acta Mol Cell Res. 2014;1843(2):335–345.
[74] Aittaleb M, Boguth CA, Tesmer JJG. Structure and function of heterotrimeric G protein-regulated Rho guanine nucleotide exchange factors. Mol Pharmacol. 2010;77(2):111–125.
[75] O'Connor KL, Chen M, Towers LN. Integrin α6β4 cooperates with LPA signaling to stimulate Rac through AKAP-LBc-mediated RhoA activation. Am J Physiol Cell Physiol. 2012;302:2–11.
[76] Paulucci-Holthauzen AA, Vergara LA, Bellot LJ, et al. Spatial distribution of protein kinase A activity during cell migration is mediated by A-kinase anchoring protein AKAP Lbc. J Biol Chem. 2009;284(9):5956–5967.
[77] Blomquist A, Schwörer G, Schablowski H, et al. Identification and characterization of a novel Rho-specific guanine nucleotide exchange factor. Biochem J. 2000;352(2):319–325.
[78] Niu J, Profirovic J, Pan H, et al. g protein βγ subunits stimulate p114rhogef, a guanine nucleotide exchange factor for RhoA and Rac1: regulation of cell shape and reactive oxygen species production. Circ Res. 2003;93 (9):848–856.
[79] Terry SJ, Elbediwy A, Zihni C, et al. Stimulation of cortical myosin phosphorylation by p114RhoGEF drives cell migration and tumor cell invasion. PLoS One. 2012;7(11):e50188.
[80] Philippe JM, Lellouch AC, Lecuit T, Garcia De Las Bayonas A. Distinct RhoGEFs activate apical and junctional contractility under control of G Proteins during Epithelial Morphogenesis. Curr Biol. 2019;29 (20):3370–3385.e7. .
[81] Marivin A, Morozova V, Walawalkar I, et al. GPCR-independent activation of G proteins promotes apical cell constriction in vivo. J Cell Biol. 2019;218 (5):1743–1763.
[82] Souchet M, Portales-Casamar E, Mazurais D, et al. Human p63RhoGEF, a novel RhoA-specific guanine nucleotide exchange factor, is localized in cardiac sarcomere. J Cell Sci. 2002;115(Pt 3):629–640. PubMed 11861769.
[83] Lutz S, Freichel-Blomquist A, Yang Y, et al. The guanine nucleotide exchange factor p63RhoGEF, a specific link between Gq/11-coupled receptor signaling and RhoA. J Biol Chem. 2005;280(12):11134–11139.
[84] Lutz S, Shankaranarayanan A, Coco C, et al. Structure of Gαq-p63RhoGEF-RhoA complex reveals a pathway for the activation of RhoA by GPCRs. Science. 2007;318(5858):1923–1927. (80-).
[85] Rojas RJ, Yohe ME, Gershburg S, et al. Gαq directly activates p63RhoGEF and trio via a conserved extension of the Dbl homology-associated pleckstrin homology domain. J Biol Chem. 2007;282(40):29201–29210.
[86] Guo X, Stafford LJ, Bryan B, et al. A Rac/Cdc42-specific exchange factor, GEFT, induces cell proliferation, transformation, and migration. J Biol Chem. 2003;278(15):13207–13215.
[87] Kruse K, Lee QS, Sun Y, et al. N-cadherin signaling via Trio assembles adherens junctions to restrict endothelial permeability. J Cell Biol. 2019;218(1):299–316.
[88] Cervantes-Villagrana RD, Adame-García SR, García-Jiménez I, et al. Gβγ signaling to the chemotactic effector P-REX1 and mammalian cell migration is directly regulated by $G\alpha_q$ and $G\alpha_{13}$ proteins. J Biol Chem. 2019;294(2):531–546.

[89] Sosa MS, Lopez-Haber C, Yang C, et al. Identification of the Rac-GEF P-Rex1 as an Essential Mediator of ErbB signaling in breast cancer. Mol Cell. 2010;40:(6):877–892.
[90] Carretero-Ortega J, Walsh CT, Hernández-García R, et al. Phosphatidylinositol 3,4,5-triphosphate-dependent rac exchanger 1 (P-Rex-1), a guanine nucleotide exchange factor for rac, mediates angiogenic responses to stromal cell-derived factor-1/chemokine stromal cell derived factor-1 (SDF-1/CXCL-12) linked to rac. Mol Pharmacol. 2010;77(3):435–442.
[91] Cash JN, Urata S, Li S, et al. Cryo-electron microscopy structure and analysis of the P-Rex1-Gβγ signaling scaffold. Sci Adv. 2019;5(10):eaax8855.
[92] Damoulakis G, Gambardella L, Rossman KL, et al. P-Rex1 directly activates RhoG to regulate GPCR-driven Rac signalling and actin polarity in neutrophils. J Cell Sci. 2014;127(11):2589–2600.
[93] Koh CG, Manser E, Zhao ZS, et al. β1PIX, the PAK-interacting exchange factor, requires localization via a coiled-coil region to promote microvillus-like structures and membrane ruffles. J Cell Sci. 2001;114(Pt 23):4239–4251. PubMed:11739656.
[94] Ward JD, Ha JH, Jayaraman M, et al. LPA-mediated migration of ovarian cancer cells involves translocalization of Gαi2 to invadopodia and association with Src and β-pix. Cancer Lett. 2015;356(2):382–391.
[95] Feng Q, Baird D, Yoo S, et al. Phosphorylation of the cool-1/β-pix protein serves as a regulatory signal for the migration and invasive activity of Src-transformed cells. J Biol Chem. 2010;285(24):18806–18816.
[96] Feng Q, Baird D, Peng X, et al. Cool-1 functions as an essential regulatory node for EGF receptor and Src-mediated cell growth. Nat Cell Biol. 2006;8(9):945–956.
[97] Mayhew MW, Jeffery ED, Sherman NE, et al. Identification of phosphorylation sites in βPIX and PAK1. J Cell Sci. 2007;120(22):3911–3918.
[98] Chahdi A, Miller B, Sorokin A. Endothelin 1 induces β1Pix translocation and Cdc42 activation via protein kinase A-dependent pathway. J Biol Chem. 2005;280(1):578–584.
[99] Chahdi A, Sorokin A. The role of β1Pix/caveolin-1 interaction in endothelin signaling through Gα subunits. Biochem Biophys Res Commun. 2010;391(3):1330–1335.
[100] Frank SR, Hansen SH. The PIX-GIT complex: a G protein signaling cassette in control of cell shape. Semin. 2008;19:234–244. Cell Dev. Biol.
[101] Zhou W, Li X, Premont RT. Expanding functions of GIT Arf GTPase-activating proteins, PIX Rho guanine nucleotide exchange factors and GIT-PIX complexes. J Cell Sci. 2016;129(10):1963–1974.
[102] Feng Q, Albeck JG, Cerione RA, et al. Regulation of the Cool/Pix proteins. Key binding partners of the Cdc42/Rac targets, the p21-activated kinases. J Biol Chem. 2002;277(7):5644–5650.
[103] Yoshii S, Tanaka M, Otsuki Y, et al. αPIX nucleotide exchange factor is activated by interaction with phosphatidylinositol 3-kinase. Oncogene. 1999;18(41):5680–5690.
[104] Lyda JK, Tan ZL, Rajah A, et al. Rac activation is key to cell motility and directionality: an experimental and modelling investigation. Comput Struct Biotechnol J. 2019;17:1436–1452.
[105] Lawson CD, Donald S, Anderson KE, et al., Welch HCE. P-Rex1 and Vav1 cooperate in the regulation of Formyl-Methionyl-Leucyl-Phenylalanine–Dependent Neutrophil Responses. J Immunol. 2011;186(3):1467–1476. .
[106] Heo J, Thapar R, Campbell SL. Recognition and activation of Rho GTPases by Vav1 and Vav2 Guanine nucleotide exchange factors †. Biochemistry. 2005;17(17):6573–6585.
[107] Bustelo XR. Regulatory and signaling properties of the Vav Family. Mol Cell Biol. 2000;20(5):1461–1477.
[108] Vila-Coro AJ, Rodríguez-Frade JM, De Ana AM, et al. The chemokine SDF-lα triggers CXCR4 receptor dimerization and activates the JAK/STAT pathway. Faseb J. 1999;13(13):1699–1710.
[109] Fumagalli L, Zhang H, Baruzzi A, et al. The Src family Kinases Hck and Fgr regulate Neutrophil responses to N Formyl-Methionyl-Leucyl-Phenylalanine. J Immunol. 2007;178(6):3874–3885.
[110] Aghazadeh B, Lowry WE, Huang XY, et al. Structural basis for relief of autoinhibition of the Dbl homology domain of proto-oncogene Vav by tyrosine phosphorylation. Cell. 2000;102(5):625–633.
[111] Fleming IN, Elliott CM, Buchanan FG, et al. Ca2+/Calmodulin-dependent Protein Kinase II regulates Tiam1 by reversible protein Phosphorylation. Biochemistry. 1999;274:12753–12758.
[112] Miyamoto Y, Yamauchi J, Tanoue A, et al. TrkB binds and tyrosine-phosphorylates Tiam1, leading to activation of Rac1 and induction of changes in cellular morphology. Proc Natl Acad Sci U S A. 2006;103:10444–10449.
[113] Minard ME, Kim LS, Price JE, et al. The role of the guanine nucleotide exchange factor Tiam1 in cellular migration, invasion, adhesion and tumor progression. Breast Cancer Res Treat. 2004;84(1):21–32.
[114] Wang S, Watanabe T, Matsuzawa K, et al. Tiam1 interaction with the PAR complex promotes talin-mediated Rac1 activation during polarized cell migration. J Cell Biol. 2012;199(2):331–345.
[115] Mertens AEE, Pegtel DM, Collard JG. Tiam1 takes PARt in cell polarity. Trends Cell Biol. 2006;16(6):308–316.
[116] Fleming IN, Gray A, Downes CP. Regulation of the Rac1-specific exchange factor Tiam1 involves both phosphoinositide 3-kinase-dependent and -independent components. Biochem J. 2000;351(1):173–182.
[117] Hoshino M, Sone M, Fukata M, et al. Identification of the stef gene that encodes a novel guanine nucleotide exchange factor specific for Rac1. J Biol Chem. 1999;274(25):17837–17844.
[118] Goto A, Hoshino M, Matsuda M, et al. Phosphorylation of STEF/Tiam2 by protein kinase A is critical for Rac1 activation and neurite outgrowth in dibutyryl cAMP-treated PC12D cells. Mol Biol Cell. 2011;22(10):1780–1790.

[119] Emami-Nemini A, Gohla A, Urlaub H, et al. The guanine nucleotide exchange factor Vav2 is a negative regulator of parathyroid hormone Receptor/Gq Signaling. Mol Pharmacol. 2012;82 (2):217–225.
[120] Schulte G, Wright SC. Frizzleds as GPCRs – more conventional than we thought! trends. 2018;39:828–842. Pharmacol Sci.
[121] Gammons MV, Renko M, Johnson CM, et al. Wnt Signalosome Assembly by DEP Domain Swapping of Dishevelled. Mol Cell. 2016;64(1):92–104.
[122] Tanegashima K, Zhao H, Dawid IB. WGEF activates Rho in the Wnt-PCP pathway and controls convergent extension in Xenopus gastrulation. Embo J. 2008;27 (4):606–617.
[123] Li H, Yang L, Fu H, et al. Association between Gαi2 and ELMO1/Dock180 connects chemokine signalling with Rac activation and metastasis. Nat Commun. 2013;4:1–12.
[124] Ward JD, Dhanasekaran DN. LPA Stimulates the Phosphorylation of p130Cas via Gαi2 in Ovarian Cancer Cells. Genes Cancer. 2012;3(9–10):578–591.
[125] Matsui H, Harada I, Sawada Y. Src, p130Cas, and Mechanotransduction in cancer cells. Genes Cancer. 2012;3(5–6):394–401.
[126] Feng H, Hu B, Liu KW, et al. Activation of Rac1 by Src-dependent phosphorylation of Dock180^{Y1811} mediates PDGFRα-stimulated glioma tumorigenesis in mice and humans. J Clin Invest. 2011;121 (12):4670–4684.
[127] Kunisaki Y, Nishikimi A, Tanaka Y, et al. DOCK2 is a Rac activator that regulates motility and polarity during neutrophil chemotaxis. J Cell Biol. 2006;174 (5):647–652.
[128] Wang J, Xu L, Shaheen S, et al. Growth of B Cell receptor microclusters is regulated by PIP2 and PIP3 Equilibrium and Dock2 recruitment and activation. Cell Rep. 2017;21(9):2541–2557.
[129] Thelen M, Stein JV. How chemokines invite leukocytes to dance. Nat Immunol. 2008;9:(9):953–959.
[130] Park D, Tosello-Trampont AC, Elliott MR, et al. BAI1 is an engulfment receptor for apoptotic cells upstream of the ELMO/Dock180/Rac module. Nature. 2007;450 (7168):430–434.
[131] Weng Z, Situ C, Lin L, et al. Structure of BAI1/ELMO2 complex reveals an action mechanism of adhesion GPCRs via ELMO family scaffolds. Nat Commun. 2019;10(1). 10.1038/s41467-018-07938-9
[132] Afm TI, Yue H, Scavello M, et al. The cAMP-induced G protein subunits dissociation monitored in live Dictyostelium cells by BRET reveals two activation rates, a negative effect of caffeine and potential role of microtubules. Cell Signal. 2018;48:25–37.
[133] Islam AFMT, Stepanski BM, Charest PG Studying chemoattractant signal transduction dynamics in Dictyostelium by BRET. In: Methods in Molecular Biology. 2016;1407: 63–77.
[134] Galés C, Rebois RV, Hogue M, et al. Real-time monitoring of receptor and G-protein interactions in living cells. Nat Methods. 2005;3(3):177–184.
[135] Xu X, Brzostowski JA, Jin T. Monitoring dynamic GP CR signaling events using fluorescence microscopy, FRET imaging, and single-molecule imaging. Methods Mol Biol. 2009;571:371–383.