Understanding Rho GTPase Interactions with Scaffolding Proteins and Ras Protein Shuttling Mechanisms

by

E. Sıla Özdemir

A Dissertation Submitted to the

Graduate School of Sciences and Engineering

in Partial Fulfillment of the Requirements for

the Degree of

Doctor of Philosophy

in

Chemical and Biological Engineering

September, 2018

Understanding Rho GTPase Interactions with Scaffolding Proteins and Ras Protein Shuttling Mechanisms

Koç University

Graduate School of Sciences and Engineering

This is to certify that I have examined this copy of a doctoral dissertation by

E. Sıla Özdemir

and have found that it is complete and satisfactory in all respects,

and that any and all revisions required by the final

examining committee have been made.

Committee Members:

_____

Prof. Özlem Keskin Özkaya (Advisor)

_____

Prof. Attila Gürsoy (Co-Advisor)

_____

Prof. Tarık Tihan

_____

Assoc. Prof. Seda Keskin Avcı

_____

Asst. Prof. Mert Gür

_____

Asst. Prof. Özlem Ulucan Açan

Date _____

To my mother Sezer, my father Ekrem and my fiancée Kemal...

Thank you for always being with me.

ABSTRACT

The Rho GTPase family and the Ras GTPase family are subfamilies of the Ras GTPase superfamily. The GTPases selectively interact with structurally and functionally diverse effectors, such as IQ motif–containing GTPase-activating proteins (IQGAP) and phosphodiesterase-δ (PDEδ). Rho GTPases are distinguished from other GTPases by the presence of an "insert loop," which is rich in charged residues in a helical motif. The Rho GTPases, Cdc42 and Rac1, in their GTP-bound active forms, interact with all three human IQGAPs. The IQGAP–Cdc42 interaction promotes metastasis by enhancing actin polymerization. However, despite their high sequence identity, Cdc42 and Rac1 differ in their interactions with IQGAP. The exact mechanism of Cdc42 and Rac1 binding to IQGAP is unclear. Using all atom molecular dynamics (MD) simulations, the detailed mechanisms of Cdc42 and Rac1 interactions with IQGAP2 were studied. It was observed that the Cdc42 "insert loop" is important for the interaction of the first Cdc42 and allows to the binding of the second Cdc42. By contrast, differences in Rac1 insert-loop sequence and structure precluded its interaction with IQGAP and only one Rac1 can bind. Ras GTPase subfamily protein, Ras is highly mutated in human cancers. Proper localization of Ras proteins at the plasma membrane (PM) is crucial for their functions. Their localization is mediated by PDEδ. GTP-bound Arf-like protein 2 (Arl2) is a regulator of PDEδ-mediated transport of farnesylated proteins; however, the exact mechanism of Arl2-assisted release of farnesylated proteins from PDEδ is still unclear. Using MD simulations, detailed mechanisms of Arl2-mediated release of KRas4B, the most abundant oncogenic Ras isoform, from PDEδ were investigated. The detailed analysis showed that allosteric changes (involving β6 of PDEδ and additional PDEδ residues) compress the hydrophobic PDEδ pocket and push the KRas4B out.

Single amino acid variation (SAV) distribution in protein-protein interaction (PPI) interfaces was also investigated. Interfaces have a small subset of residues called hot spots that contribute significantly to the binding energy, and they may form clusters called hot regions. Singlet hot spots are the single amino acid hot spots outside of the hot regions. Statistical and structural analyses of SAVs were performed with literature

curated experimental thermodynamics data, and demonstrated that SAVs which destabilize PPIs and cause diseases are more likely to be found in singlet hot spots rather than hot regions and than energetically less important interface residues. This part of the thesis demonstrates that SAVs in singlet hot spot residues have significant effects on protein stability and function.

This thesis study was supported by TUBITAK (Research Grant No: 114M196 and Scholarship: 2211-E).

ÖZETÇE

Rho GTPaz ailesi ve Ras GTPaz ailesi, Ras GTPaz superailesinin alt aileleridir. GTPazlar yapısal ve işlevsel olarak farklılıklar gösteren efektörlerle seçici olarak etkileşirler. IQ motif-containing GTPase-activating protein (IQGAP) ve phosphodiesterase-δ (PDEδ) bu efektörler arasındadır. Rho GTPazlar diğer GTPazlardan, yüklü amino asitlerce zengin bir heliks motifi olan "insert loop"un varlığıyla ayrılır. Rho GTPazlar, Cdc42 ve Rac1 GTP bağlı aktif formlarındayken üç insan IQGAP proteiniyle de etkileşim kurarlar. IQGAP–Cdc42 etkileşimi aktin polimerizasyonunu arttırarak metaztaza katkıda bulunurlar. Ancak, yüksek sekans benzerliğine ragmen Cdc42 ve Rac'ın IQGAP ile olan etkileşimi farklılık gösterir. Cdc42 ve Rac1'ın IQGAP'e bağlanma mekanizması kesin olarak bilinmemektedir. Moleküler dinamik simulasyonlarını kullanarak Cdc42 ve Rac1'ın IQGAP2 ile olan etkileşimlerinin detaylı mekanizması çalışıldı. Cdc42'nin "insert loop"u ilk Cdc42'nin bağlanmasında önem taşımakta ve ikinci Cdc42'nin bağlanmasına yardımcı olmaktadır. Aksine, Rac1'ın "insert loop" sekansındaki ve yapısındaki farklılılar onun IQGAP2 ile etkileşimini engellemiştir ve sadece bir Rac1 IQGAP2'ye bağlanabilmektedir. Ras GTPaz alailesi proteinlerinde Ras, insan kanserlerinde yüksek oranda mutasyona uğramış halde bulunmaktadır. Ras proteinlerinin plasma membranında (PM) uygun şekilde konumlanması, Ras fonksiyonu için kritik önem taşımaktadır. Ras proteinlerinin konumlandırılmasında PDEδ görev almaktadır. GTP-bağlı Arf-like protein 2 (Arl2), farnesilli proteinlerin PDEδ-aracılı taşınmasında bir regülatör görevi almaktadır, ancak farneillli proteinlerin PDE'dan Arl2-yardımlı salınımının kesin mekanizması bilinmemektedir. MD simülasyonlarını kullanarak Arl2-aracılı en çok rastlanan onkojenik Ras izoformu olan KRas4B'nin PDEδ'dan salınımının detaylı mekanizmasını araştırıldı. Detaylı analizler, PDEδ'nın β6 bölgesi ve bazı amino asitlerini kapsayan alosterik değişikliklerin hidrofobik PDEδ cebini sıkıştırarak KRas4B'yi dışarı ittiğini gösterdi.

Ayrıca, tekli amino asit varyasyonlarının (SAV) protein-protein etkileşim (PPI) ara yüzlerine dağılımını da incelendi. Ara yüzler, hot spot adı verilen ve bağlanma enerjisine

önemli katkıda bulunan az sayıda amino asit içerir, bu hot spotlar kümelenerek hot regionları oluşturur. Tekli hot spotlar ise hot regionlar dışında kalan hot spotlara denmektedir. Literatürden toplanmış deneysel thermodinamik dataya sahip SAVlarla, istatistiksel ve yapısal analizler gerçekleştirildi. Bunun sonucunda, PPIı destabilize eden ve hastalığa neden olan SAVların, hot regionlar ya da energy bakımında daha az önemli bölgelerde bulunmaktansa, tekli hot spotlarda bulunmayı tercih ettiğini gösterildi. Bu analizler tekli hot spotlara denk gelen SAVların, proteinin stabilitesine ve fonksiyonuna önemli etkisi olduğunu gösteriyor.

Bu tez çalışması TUBITAK tarafından desteklenmiştir (Araştırma Projesi No: 114M196 ve Burs: 2211-E).

ACKNOWLEDGEMENT

I would like to thank my thesis advisors Prof. Özlem Keskin and Prof. Attila Gürsoy, I really appreciate their guidance on my thesis project and graduate studies. I am also grateful to our collaborators Prof. Ruth Nussinov, Dr. Hyunbum Jang and Nussinov group, and Dr. David B. Sacks, Dr. Zhigang Li and Sacks group.

I appreciate the feedback offered by my thesis committee members, Prof. Tarık Tihan and Assoc. Prof. Seda Keskin. Advice and comments given by them have been a great help in the progress of my studies. I would also like to thank my other thesis jury members Asst. Prof. Mert Gür, and Asst. Prof. Özlem Ulucan Açan for their valuable time.

I would like to mention my dear friends and colleagues Derya Soner Cavga, Farideh Halakou, Ayşe Seda Yazgılı, Ali Tuğrul Balcı, Kayra Köseoğlu, Jackson Weako, Meltem Eda Ömür, Asma Hakouz, Damla Övek, Serena Muratcıoglu, Bilgesu Erdoğan, Efe Elbeyli and Yusuf İsmail Ertuna who have always been by my side during these four years both in social and academic life.

Once again I would like to thank and express my gratitude and love toward my parents Ekrem and Sezer, my fiancée Kemal and also my aunt Ümran. Thank you for your support and everything.

I acknowledge TUBITAK, the foundation financially supporting this project (Research Grant No: 114M196) and me (2211-E) during my doctoral studies.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

NOMENCLATURE

Arl2	Arf-like protein 2
Cdc42	cell division control 42
GAP	GTPase activating protein
GDP	Guanosine diphosphate
GEF	Guanine-nucleotide exchange factor
GRD2	GAP-related domain 2
GTP	Guanosine-5'-triphosphate
HVR	Hypervariable region
IQGAP2	IQ Motif Containing GTPase Activating Protein 2
PDB	Protein Data Bank
PDEδ	Phosphodiesterase-δ
PI3K	Phosphoinositide 3-kinase
PM	Plasma Membrane
PPI	Protein-Protein Interaction
Rac1	Ras-related C3 botulinum toxin substrate 1
Ras	Rat sarcoma
Rho	Ras homolog gene family
RhoGDI	Rho GDP dissociation inhibitor
SAV	Single amino acid variation
SKEMPI	Structural database of Kinetics and Energetics of Mutant Protein Interactions

Chapter 1

INTRODUCTION

Ras superfamily GTPases act as molecular switches in response to receptor-mediated extracellular signals. They regulate multiple cellular processes, such as growth, survival, cell invasion, cell motility, and vesicle trafficking, by interacting with effector molecules. Rho GTPases and Ras GTPases are two subfamilies of the Ras GTPase superfamily. Rho GTPases, Cdc42 and Rac1 are known as playing roles in metastasis by promoting actin polarization. Despite the effort, the exact mechanism associating Cdc42 and Rac1 with metastasis is not clear. Ras GTPase isoform KRas is mutated in approximately 20-30% of human cancers. However, these proteins are still "undruggable". Thus, community focus on elucidating mechanisms promoting Rho GTPase-mediated metastasis and therapeutically targeting oncogenic Ras.

Cdc42 and Rac1 interacts with the scaffolding proteins, IQGAP. Despite the high degree of sequence identity, Cdc42 and Rac1 bind to IQGAP1 and IQGAP2 in different ways and perform distinct cellular functions. Two molecules of Cdc42 can bind to IQGAP1 and IQGAP2 through their GAP-related domain (GRD1 and GRD2, for IQGAP1 and IQGAP2, respectively) [1] . The structure reveals two different Cdc42-binding sites in GRD2, the Ex-domain and the RasGAP sites. Because IQGAP2 dimerization through GRD2s is promoted by Cdc42 binding, the complex has dimerized GRD2s and four Cdc42s bound to these GRD2s [1, 2]. On the other hand, it was speculated that only one Rac1 molecule can bind to GRD, and there has been no evidence that it facilitates IQGAP dimerization. However, the exact mechanism of Rac1–GRD2 interaction and the differing stoichiometry have not been identified. In this thesis, the underlying mechanism of the different stoichiometry that is involved in the binding of Cdc42 and Rac1 to GRD2 and of IQGAP2 dimerization was investigated. These findings then suggested how the mechanism can help to elucidate the key question of the roles of Cdc42 and Rac1 in actin polymerization in metastasis.

Ras is a major drug target [3, 4], and to date, no drugs are available. Plasma membrane localization is essential for its proper function. Interfering with proper Ras localization can be one possible therapeutic approach. Localization is regulated by membrane trafficking/solubilizing factors. PDEδ has a role in binding to and shuttling of small GTPases [5, 6]. PDEδ interacts with Ras proteins (KRas4B and likely KRas4A) [7, 8] through their hypervariable regions (HVRs), embedding the prenyl group in the hydrophobic cavity [9, 10]. Studies show that PDEδ can also associate with two non-prenylated Arf-like proteins, Arl2 and Arl3. GTP-dependent interaction of Arl2/Arl3 with PDEδ indicates that PDEδ is an Arl2/Arl3 effector [5, 11, 12]. PDEδ stabilizes Arl2/Arl3 in their GTP-bound form by blocking GTP hydrolysis [11]. Arl2 and Arl3 are responsible for regulation of the release and/or uptake of farnesylated cargos and small inhibitory molecules from PDEδ [5, 13]. However, the exact mechanism behind the KRas4B release from PDEδ to facilitate PM anchorage is unknown and the role of Arl2/Arl3 in the release is unclear. In this thesis, a mechanism of the Arl2-GTP-mediated release of farnesylated KRas4B from the PDEδ hydrophobic pocket was suggested. Arl2-GTP binding promotes formation of fast-dissocating ternary complex and allosteric changes in PDEδ and mediated the release of farnesylated KRas4B, but not geranylgeranylated KRas4B.

Single amino acid variations (SAVs) in protein-protein interaction (PPI) sites play critical roles in diseases. Distribution of SAVs within residues of PPI sites (interface) may be related to their disease association. Interfaces have a small subset of residues called hot spots that contribute significantly to the binding energy, and they may form clusters called hot regions. Singlet hot spots are the single amino acid hot spots outside of the hot regions. Statistical and structural analyses of SAVs were performed with literature curated experimental thermodynamics data, and demonstrated that SAVs in singlet hot spot residues have significant effect on protein stability and function.

Chapter 2 gives a comprehensive review on Rho GTPase and Ras GTPase subfamilies, their signaling and significance, IQGAP interaction of Cdc42/Rac1 and membrane

shuttling of Ras. The chapter also reviews SAV and their disease-association, as well as studies related to their correspondence to the interface residues.

Chapter 3 provides a detailed explanation about the materials and methods used in this thesis.

Chapter 4 presents computational results for Cdc42/Rac1–IQGAP2 interactions, and their effect on IQGAP2 dimerization, allostery and metastasis. The chapter provides a detailed discussion of the results including the outcome of differences in Cdc42–IQGAP2 and Rac1–IQGAP2 interactions.

Chapter 5 focuses on Arl2-mediated release of KRas from PDEδ, the effect of active and inactive Arl2 on the release, and different release mechanisms of farnesylated and geranylgeranylated KRas. It also discusses possible therapeutic approaches to target shuttling mechanism of Ras proteins.

Chapter 6 presents results from single amino acid variations (SAVs) distribution analyses. It discuss the link between the binding free energy change and disease-association of SAVs and their location in interface residues.

Chapter 7 presents the conclusion and future perspectives

Chapter 2

LITERATURE REVIEW

2.1. Ras Superfamily GTPases

Guanosine-5'-triphosphateases (GTPases) are a large family of hydrolase enzymes. They bind to GTP to hydrolyze it to GDP [14]. GTPases have a common G-domain structure where the GTP binding and hydrolysis takes place. This domain is highly conserved among all GTPases. G-domain is also responsible from regulator and effector binding. One of the major GTPase superfamily is the Ras superfamily of GTPases. This superfamily participates in almost every aspect of cell biology together with their associated regulators and effectors [15]. The Ras superfamily comprise over 150 human members. The members are evolutionarily conserved within other species such as in *Drosophila, C. elegans, S.cerevisiae,* and plants [15]. The Ras superfamily is divided into five major subfamily based on sequence and functional similarities. These five subfamilies are Ras, Rho, Rab, Ran and Arf (Figure 2.1).

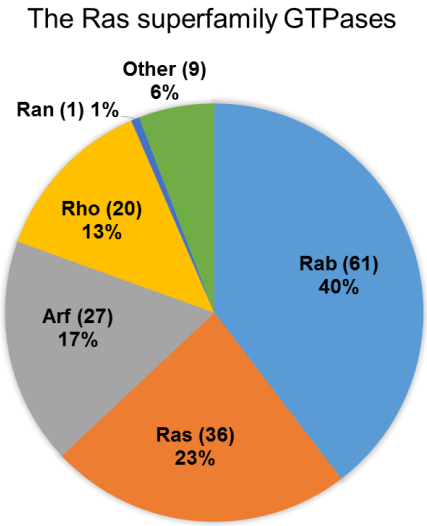


Figure 2.1. The five major Ras GTPases subfamilies. Number in the parentheses indicates the number of the member among over 150 GTPases.

2.1.1. Structures and functions of Ras superfamily GTPases

Acting as a binary molecular switch is a common biochemical mechanism among the GTPases [16].

The binary switch determines whether Rho GTPases are in the active or inactive state. The switch is on when GTP is bound and off when it is hydrolyzed to GDP [17]. Regulation of this switch cycle depends on two critical intrinsic reactions, GDP/GTP exchange and GTP hydrolysis. Guanine nucleotide exchange factors (GEFs) catalyze the exchange of GDP to GTP. GTPase-activating proteins (GAPs) accelerate the intrinsic GTP hydrolysis rate of the GTPases, which is not sufficient to GTP hydrolysis by itself [18-21]. Deregulation and alterations of the GTPases and the switch regulators may result in dysregulation of the cell cycle, which may affect cell morphogenesis, migration, adhesion, and oncogenesis [19, 22, 23] (Figure 2.2).

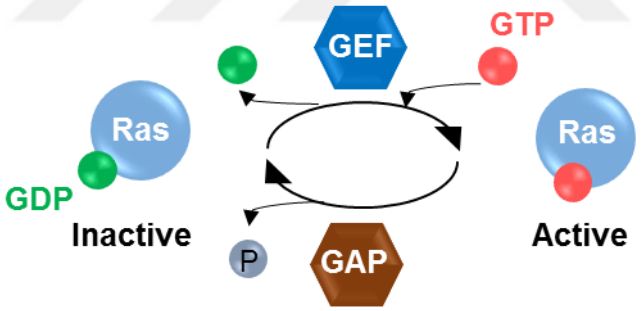


Figure 2.2. Binary switch cycle of the Ras superfamily GTPases. GEF and GAP proteins control the cycling between the GTP-bound active form and GDP-bound inactive form of Ras GTPases.

The GTPases can be located at the plasma membrane through their hypervariable region (HVR) and a lipid anchor after post-translational modifications (PTMs) [24]. Covalent PTM of C-terminal cysteine residues by attachment of a farnesyl or geranylgeranyl group is observed for most Ras GTPase superfamily members. Subcellular localization promoted through the PTMs influences effector binding or activation and regulatory protein interactions [25, 26]. Since functioning as molecular switches, the GTPases

share a set of conserved G box GDP/GTP-binding motif elements beginning at the N-terminus: G1, GXXXXGKS/T; G2, T; G3, DXXGQ/H/T; G4, T/NKXD; and G5, C/SAK/L/T [27]. G-domain is approximately 20kDa and consists of 166 residues, which is highly conserved by all Ras superfamily proteins.

Variations in structure, PTM-mediated specific subcellular locations, and their regulatory and effectors proteins allow the GTPases to function as sophisticated modulators of a remarkably complex and diverse range of cellular processes. They are responsible for cell motility, proliferation, polarity, transformation, intracellular trafficking, gene expression, growth, survival, cell invasion.

2.2. Rho GTPase Subfamily

The Rho GTPase family is a subfamily of the Ras GTPase superfamily. To date, the mammalian Rho GTPase family includes 22 members; Cdc42, Rac1, and RhoA are the most studied among them [28]. Rho GTPases are key regulators of actin reorganization. Each member of Rho GTPases have a distinct role in actin polymerization and regulation. RhoA promotes actin stress fiber formation and focal adhesion assembly; Rac1 promotes lamellipodium formation and membrane ruffling; and Cdc42 promotes filopodium formation. Therefore, they mainly regulate cellular processes, such as growth, survival, cell invasion, cell motility, vesicle trafficking, as well as in regulation of endocytosis and exocytosis by interacting with effector molecules [18, 29-31].

Rho GTPases are distinguished from other GTPases by the presence of an "insert loop," which is rich in charged residues in an α-helical motif (Figure 2.3). This α-helix is an important region for effector binding [32, 33]. Similar to other GTPases, Rho GTPases act as binary switches and anchor to the PM through the HVR. The GTP- and GDP-bound forms are referred to as the active and inactive states, respectively. Their activation upon GTP binding is accompanied by conformational changes in two regions, the switch I and II motifs (Figure 2.3). These regions serve as platforms for selective interactions with structurally and functionally diverse effectors, such as p21-activated kinase 1 (PAK1); p67phox, a member of the NADPH oxidase family; and IQ motif-containing GTPase-activating proteins (IQGAPs) [34, 35].

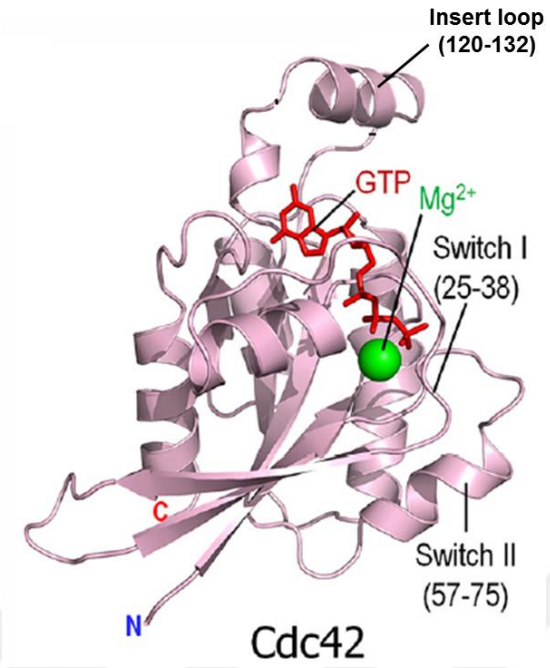


Figure 2.3. Structure of the Rho GTPase Cdc42. The insert loops in an α-helical (120-132), which is specific to Rho GTPases. Switch I (25-38) and switch II (57-75) are important regions for effector and GTP/GDP binding (PDB ID: 5CJP [1]).

2.2.1. Cdc42 and Rac1 proteins

Human homolog of S. cerevisiae cell-division-cycle protein (Cdc42) was first identified in 1990 [36] and Ras-related C3 botulinum toxin substrate 1 (Rac1) was isolated in 1989 [37]. Cdc42 and Rac1 have ~70% sequence identity, yet their functions in cellular processes differ [1, 38]. The sequence of their switch I and switch II region are almost identical, while the insert loop shows a remarkable sequential difference (Figure 2.4).

```
Cdc42 2    QTIKCVVVGDGAVGKTCLLISYTTNKFPSEYVPTVFDNYAVTVMIGGEPYTLGLFDTAGL  Cdc42 61
Rac1  2    QAIKCVVVGDGAVGKTCLLISYTTNAFPGEYIPTVFDNYSANVMVDGKPVNLGLWDTAGQ  Rac1  61
                                  switch I
Cdc42 62   EDYDRLRPLSYPQTDVFLVCFSVVSPSSFENVKEKWVPEITHHCPKTPFLLVGTQIDLRD  Cdc42 121
Rac1  62   EDYDRLRPLSYPQTDVFLICFSLVSPASFENVRAKWYPEVRHHCPNTPIILVGTKLDLRD  Rac1  121
           switch II
Cdc42 122  DPSTIEKLAKNKQKPITPETAEKLARDLKAVKYVECSALTQKGLKNVFDEAILAAL       Cdc42 177
Rac1  122  DKDTIEKLKEKKLTPITYPQGLAMAKEIGAVKYLECSALTQRGLKTVFDEAIRAVL       Rac1  177
           insert loop
```

Figure 2.4. Sequence alignment of Cdc42 and Rac1

Cdc42 plays a role in cytoskeletal rearrangement and cell motility. It facilitates the polarization of actin and microtubules and stimulates filopodium formation [39, 40]. Rac1 regulates the formation of cortical actin-containing membrane ruffles and lamellipodia [41]. Cdc42 regulates actin polymerization by interacting Wiskott–Aldrich syndrome protein (WASP) and N-WASP [42, 43]. WASP and N-WASP each contain an acidic C-terminus (VCA) and CRIB domain. It was observed that overexpression of N-WASP plus Cdc42 induces very long microspikes, like an exaggeration of Cdc42 activity [44], suggesting that these proteins may be involved in the formation of filopodia downstream of Cdc42. Cdc42 activates N-WASP by binding to CRIB domain and release N-WASP from it auto-inhibited state. Arp2/3 is recruited to VCA of N-WASP after the activation. The Arp2/3 complex binds to actin monomers and acts as a nucleation site for new actin polymerization (Figure 2.5) [45].

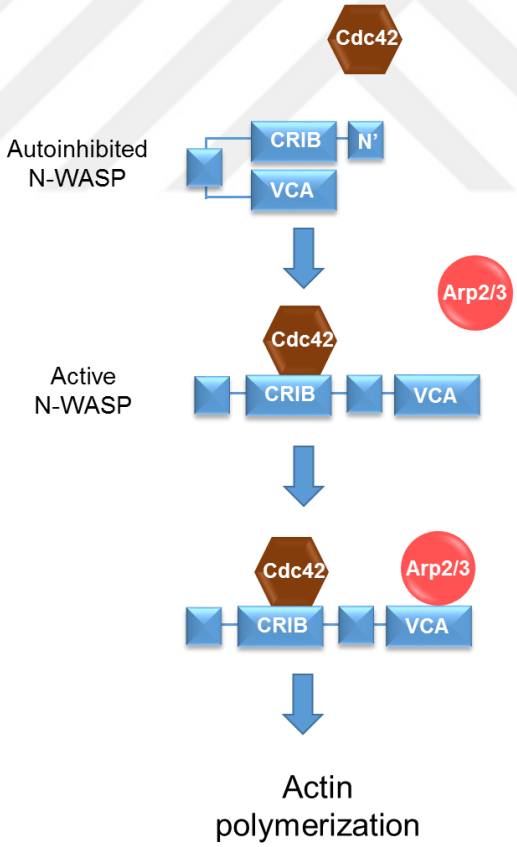


Figure 2.5. Cdc42-mediated actin polymerization

2.2.2. IQGAPs and Rho GTPases

IQGAP is a ubiquitously expressed eukaryotic scaffold protein that is found in many species ranging from yeast to humans. The first identified and most studied is human IQGAP1, a 189-kDa and 1657-amino acid–long protein [46]. IQGAP1 plays a role in cell motility, cell–cell attachment, protein trafficking, transcription, regulation of the cytoskeleton, and microbial pathogenesis [47, 48]. Cdc42 activity and intrinsic GTP hydrolysis are inhibited upon binding to IQGAPs, which stabilize its active form [49, 50]. Cdc42 liberation fuels its signaling. Consistent with this, overexpression of IQGAP1 in mammalian cells promotes metastasis by stimulating actin polymerization mediated by Cdc42 [20, 23, 49, 51]. In addition to Cdc42 and Rac1, IQGAPs interact with over 100 molecules, including calmodulin (CaM), mitogen-activated protein kinase pathway proteins such as extracellular signal regulated kinase 2 (ERK2), growth factors, chemokine receptors, actin, β-catenin, Yes-associated protein 1 (YAP1), small molecules such as Ca^{2+}, and more [46, 52-54]. These interactions occur through five major domains of IQGAPs (Figure 2.6): the calponin homology domain for actin and Ca^{2+}-CaM binding, the WW domain for ERK2 and phosphoinositide 3-kinase α binding, IQ motifs for CaM and myosin light chain binding, GRD for Cdc42 and Rac1 binding, and the RasGAP C-terminal domain for E-cadherin and β-catenin binding [55-58]. Some of the binding partners compete or interfere with each other for binding to IQGAPs. For example, CaM binding causes allosteric Cdc42 release from GRD and decreases actin binding to the calponin homology domain [57]. Free, active Cdc42 can then promote actin polymerization through Wiskott–Aldrich syndrome protein (WASP) and Arp2/3. IQGAP dimerization can also help to simultaneously bind to competing targets [59].

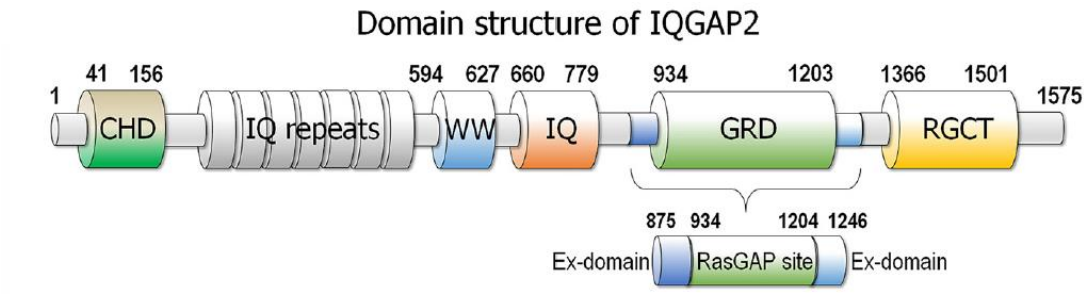


Figure 2.6. Domain structure of IQGAP2 (PDB entry 5CJP).

IQGAP1 is also known to form dimers, and dimerization is required for its function [2]. IQGAP1 is identified as overexpressed in several cancers and observed as playing a role in mammalian cell migration and invasion, therefore it appears to be an oncoprotein [60]. Other IQGAP isoforms, IQGAP2 and IQGAP3 shares considerable overall sequence identity with IQGAP1. They are more specialized for certain tissues [61]. IQGAP2, mainly expressed in the liver, kidney, and platelets. All three human IQGAPs have been reported to bind to Cdc42 and Rac1 in their GTP-bound forms [54]. The binding site for Cdc42 and Rac1 encompasses the RasGAP-related domains (GRDs) of IQGAP isoforms. Despite the high degree of sequence identity, Cdc42 and Rac1 bind to IQGAP1 and IQGAP2 in different ways and perform distinct cellular functions. Two molecules of Cdc42 can bind to IQGAP1 and IQGAP2. The crystal structure of the GRD of IQGAP2 in complex with Cdc42 has been solved recently [1]. The structure reveals two different Cdc42-binding sites in GRD2, the Ex-domain and the RasGAP sites (Figure 2.6). The Ex-domain is composed of the N- and C-terminal regions of GRD2. Because IQGAP2 dimerization through GRD2s is promoted by Cdc42 binding, the complex has dimerized GRD2s and four Cdc42s bound to these GRD2s [1, 2].

There is evidence that overexpressed IQGAP1 in cancer affects actin polymerization [62], and that it promotes migration of tumor cells [40]. GTP-bound Cdc42 and IQGAP1 regulate filopodia formation. IQGAP1 C-terminal half activates N-WASP [63]. Studies showed that Cdc42 and IQGAP1 bind to N-WASP and activate it synergistically [64]. Cdc42-IQGAP binding stabilizes its active form, which facilitates Cdc42-dependent N-WASP activation. Actin polymerization by Rac1 differs, with lamellipodia (mesh) formation induced by Rac1 interaction with WAVE, a WASP-family verprolin homologous protein, and is not facilitated by IQGAP1 [65].

2.3. Ras GTPase Subfamily

Ras genes were first identified and characterized as transduced oncogenes in the Harvey [66] and Kirsten [67] strains of acutely transforming retroviruses. The viral H-*ras* and K-*ras* isoforms were identified in the late 1970s by Scolnick and colleagues [68]. They showed that 21 kDa protein product of these human Ras genes can bind GDP/GTP [69-71]. They identified the localization of these proteins as the cytoplasmic surface of the

plasma membrane (PM) using electron microscopic immunocytochemistry [72]. A single missense mutation in codon 12 causes *HRAS* and *KRAS* gene activation in the EJ/T24 bladder carcinoma cell line, [73], and lung and colon tumor cells [74], respectively was discovered. A third transforming human *RAS* gene, *NRAS*, was discovered in neuroblastoma-derived DNA in 1980s [75, 76]. Active *RAS* genes were detected in majority of the human tumor cell lines [77]. The screening of the tumor cell lines identify mutations at codons 13 and 61, as well as the mutations at codon 12, in oncogenic *RAS* gene. Ras GTPases are membrane-associated proteins. Therefore, they are activated in response to various extracellular stimuli, such as growth factors, receptors, and hormones. They play a role in activation and regulation of multiple signaling pathways such as cell proliferation, migration, survival and differentiation [78, 79]. Ras proteins also act as molecular switches by alternating between inactive GDP-bound and active GTP-bound states [80-86]. Active Ras-GTP can bind and activate downstream effectors, including Raf kinase, phospho-inositol 3 kinase (PI3K), and Ral guanine nucleotide dissociation stimulator (RalGDS) [87-89].

The three human *RAS* genes encode four highly similar proteins: HRas, NRas, and KRas4A and KRas4B. The two KRas proteins arise from alternative splicing at the C terminus [90, 91]. HRas, NRas and KRas4A have 189 amino acids, while KRas4B has 188 amino acids. The G-domain (residues 1-166) of the four isoforms are highly conserved (~ 89%), while HVRs have extremely low sequence identify (~ 8%) (residues 167-189 for H-Ras, K-Ras4A, and N-Ras; 167-188 for K-Ras4B) (Figure 2.7) [28]. It indicates that HVR is significantly contribute differentiating functions of these *RAS* isoforms.

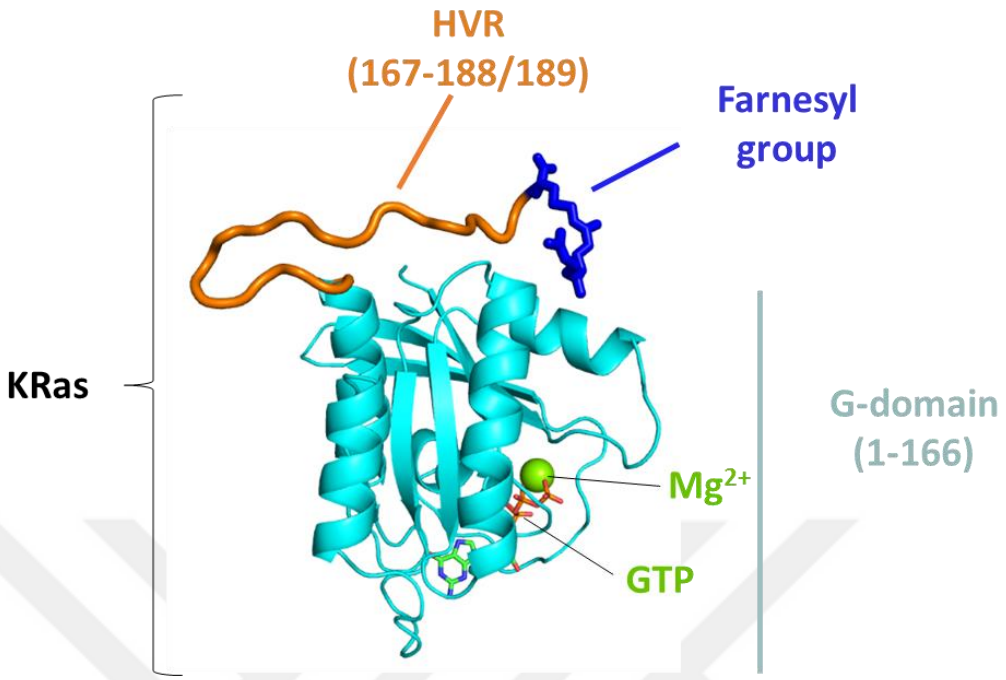


Figure 2.7. General structure of farnesylated full-length KRas4B (G-domain PDB ID: 3GFT, HVR and farnesyl is modelled [92])

2.3.1. Ras structure and post-translational modifications (PTMs)

H-Ras was the first Ras structure determined in complexed with GppNHp, a GTP analog, (PDB ID 5P21) [93, 94] and GDP (PDB ID 4Q21) [95]. G-domain consists of the functional P-loop (residues 10-17), switch I (residues 30-38), and switch II (residues 59-76) regions similar to the Rho GTPases constitute the active site for GTP/GDP binding and interaction sites for effector proteins, including Raf, PI3K, RalGDS and GAP (Figure 2.8) [96]. The first half of the proteins (residues 1-86) are called “effector lobe” and it consists of all the structural elements. The effecter lobe has 100% sequence identity across Ras isoforms [97, 98]. The second half is “allosteric lobe” (residues 87-166) containing a more variable sequence among Ras isoforms [99].

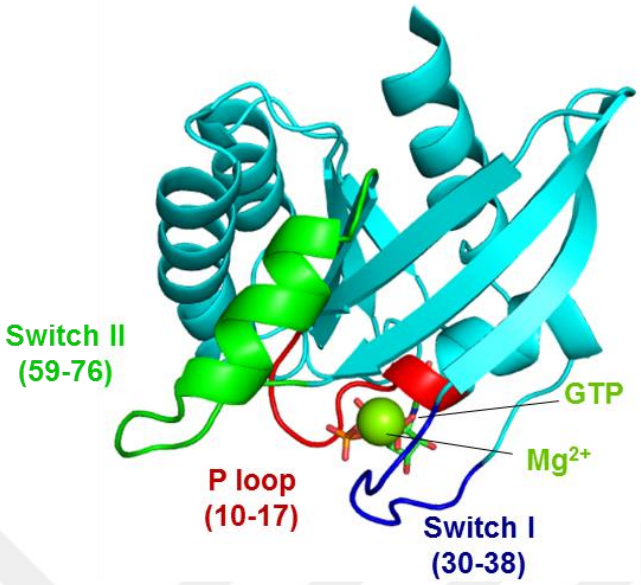


Figure 2.8. Structure of the KRas4B. Switch I (30-38) and switch II (59-76) are important regions for effector and GTP/GDP binding (PDB ID: 3GFT).

Variations across the HVR of Ras isoforms are probably responsible for the biological differences. HVRs are synthesized as soluble precursors and their C-terminal CAAX sequences (where C stands for cysteine, A is an aliphatic residue and X denotes any amino acids) undergo a series of post-translational modifications (PTMs) [100]. Prenylation (addition of prenyl group to the cysteine residue) of Ras proteins, involving farnesylation or geranylgeranylation, is a significant modification for trafficking and membrane insertion [101, 102]. Mutation of the CAAX motif in the HVR has been demonstrated to demolish PM localization and Ras signaling [103].

2.3.2. Ras signaling pathways

Two of the main Ras signaling pathways are mitogen-activated protein kinases (MAPK) and phosphoinositide-3 kinase (PI3K) pathways (Figure 2.9). MAPK and PI3K pathways are the cell's key mechanisms for controlling cell survival, proliferation, and motility [104-107]. Epidermal growth factor (EGF) binding to epidermal growth factor receptor (EGFR), results in dimerization and activation of EGFR. The activated EGFR recruits Ras-specific GEFs. GEF exchanges GDP by GTP and activates Ras. In the MAPK pathway, activated Ras recruits a MAP kinase kinase kinase (Raf) to the cell membrane and activates its kinase domain by promoting Raf side-to-side dimer

formation [108-110]. Active Raf then initiates a protein kinase cascade leading to an activation of a MAP kinase, ERK1 [111, 112]. PI3K, on the other hand, can be activated through a direct interaction with Ras [113]. Active Ras binds and activates PI3K that converts phosphatidylinositol (4,5)-bisphosphate (PIP_2) into phosphatidylinositol (3,4,5)-trisphosphate (PIP_3). PIP_3 serves as a plasma membrane docking site for the pleckstrin homology (PH) domain of Akt/PKB and its upstream activator PDK1 [114]. Phosphorylation of Akt at Thr308, and Ser473 by PDK1 and mTORC2, respectively, leads to the full activation of Akt [115].

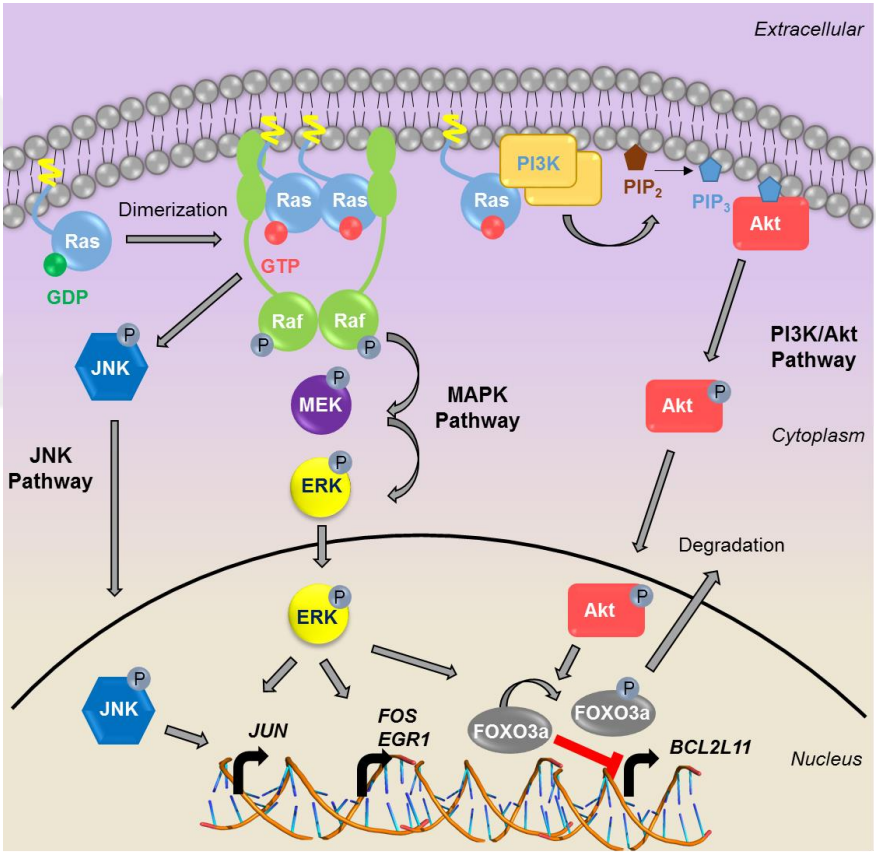


Figure 2.9. Ras signaling pathways. Schematic models illustrating MAPK- and PI3K/Akt-mediated gene regulation.

2.3.3. KRas4B and PM shuttling

The Catalog of Somatic Mutations in Cancer (COSMIC) [116] confirms that K-Ras is the most frequently mutated isoform in Ras-driven cancers (86%) [117]. K-Ras

mutations occur at very high frequency in pancreatic ductal adenocarcinoma, lung and colon tumors. Significantly, 98% of the oncogenic Ras mutations are found at amino acid residues G12, G13 and Q61, whose mutations impair intrinsic and GAP-mediated GTP hydrolysis, locking mutant Ras in the active Ras-GTP state.

Membrane anchorage is a common feature of the KRas proteins and PM attachment is essential for their function [118]. Farnesylation or geranylgeranylation, is a significant PTM for trafficking and membrane insertion [101, 102]. Guanosine nucleotide dissociation inhibitors (GDIs) such as RhoGDI are responsible for the shuttling of Rho proteins from the Ras superfamily of GTPases [119]. The high structural similarity of phosphodiesterase-δ (PDEδ) to RhoGDI suggests that PDEδ has a role in binding to and shuttling of small GTPases [5, 6]. It features a hydrophobic prenyl-binding pocket into which prenylated small G-proteins insert their farnesylated C-terminus. This facilitates the intracellular diffusion of these proteins by shielding the hydrophobic prenyl moiety from the cytosol. However, the prenyl binding pocket of PDEδ is much shorter than the RhoGDI binding pocket resulting in a weaker interaction between PDEδ and geranylgeranylated proteins [5, 120, 121]. PDEδ also lacks an N-terminal helix-loop-helix that can interact with the GTPase switch regions. Consequently, unlike RhoGDI binding, PDEδ binding to prenylated proteins is nucleotide-independent [5, 122]. PDEδ appears more promiscuous in terms of ligand specificity and can associate with various prenylated proteins including protein kinases, PDE subunits, and GTPases. Studies support the role of PDEδ in solubilizing and trafficking prenylated KRas, HRas, and Rheb [5, 122, 123]. PDEδ interacts with Ras proteins (KRas4B and likely KRas4A) [7, 8], through their hypervariable regions (HVRs), embedding the prenyl group in the hydrophobic cavity.[9, 10] Biochemical studies indicate that PDEδ binding to small GTPases is more favorable for farnesylated than geranylgeranylated proteins [6, 122]. PDEδ increases the kinetics of trapping and thereby counteracts the entropic tendency of the membranes to randomly distribute Ras over all membranes and maintain K-Ras4B localized enrichment at the PM [120, 124-126].

Studies show that PDEδ can also associate with two non-prenylated Arf-like proteins, Arl2 and Arl3. GTP-dependent interaction of Arl2/Arl3 with PDEδ indicates that PDEδ

is an Arl2/Arl3 effector [5, 11, 12]. PDEδ stabilizes Arl2/Arl3 in their GTP-bound form by blocking GTP hydrolysis [11]. Arl2 and Arl3 are responsible for regulation of the release and/or uptake of farnesylated cargos and small inhibitory molecules from PDEδ [5, 13]. Ras is a major drug target [3, 4], and to date, no drugs are the clinic. PM localization is essential for its proper function. Localization is regulated by membrane trafficking/solubilizing factors such as PDEδ and Arl2/Arl3, respectively [127]. Interfering with proper Ras localization can be one possible therapeutic approach [128, 129].

2.4. Single Amino Acid Variations and PPI Interfaces

Missense variations resulting in alterations in amino acid sequences are one of the main reasons of Mendelian diseases or complex genetic disorders (Figure 2.10) [130, 131]. Proteins are responsible for diverse functions in all cellular machineries. Most, if not all, biological processes such as proliferation, cell signaling and apoptosis are mediated by protein-protein interactions (PPIs) [132-134]. SAVs that disrupt PPIs most likely affect many protein functions, which may lead to diseases [135-138]. Recent studies on mechanistic effects of SAVs suggest that disease-causing variations prefer to be located on protein-protein interaction sites (interfaces) compared to other regions in the surface [139-144]. Also, it has been observed that significant physicochemical changes occur in proteins upon SAVs in interfaces compared to variations in non-interface residues, and a significant number of SAVs located in interfaces are predicted as disease-causing [145].

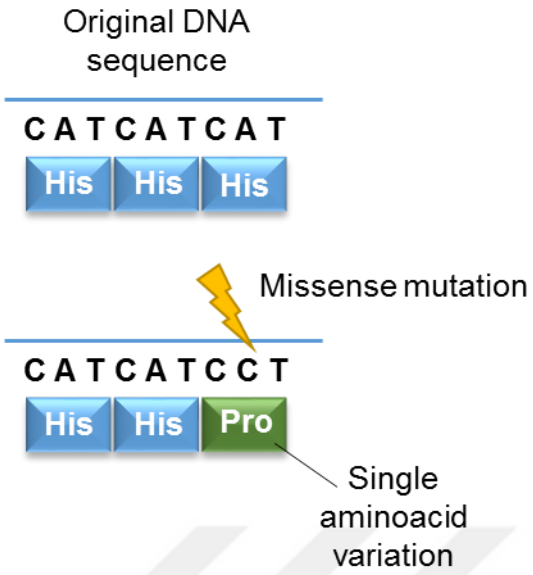


Figure 2.10. Schema of a missense mutation leading to SAV.

SAVs in interface residues can interfere with PPIs by affecting their stability and/or their binding properties [135, 146]. Effect of a SAV on protein binding can be measured with binding free energy changes (ΔΔG). Recent studies show that SAVs, especially disease-causing SAVs, destabilize proteins and their interactions by effecting the electrostatic component of binding energy [145, 147, 148]. However, disease-causing SAVs may also increase the stability of interaction of protein complexes, therefore the effect of stabilizing SAVs cannot be ignored [145, 149]. Indeed, SAVs on suppressors and activators can have opposing phenotypes [135, 150]. The changes in binding free energy upon SAVs serve as a good indicator for disease association [151].

The contribution of each amino acid in interfaces to the formation and stability of the interaction is not the same. Using experimental methods, it has been shown that even a single residue can provide large binding free energy change compared to other residues of an interface, which makes it crucial for the interaction. These critical residues that contribute more to the binding free energy than others are called “hot spots” [152, 153]. Hot spots are identified *in vitro* and *in silico* as those residues that cause at least 2.0 kcal/mol increase in the binding free energy upon changing them into alanine [152, 154]. Hot spots are not randomly distributed at interfaces and but rather tightly packed, they are clustered to form hot regions [155, 156]. Hot spots not involving in any hot regions are called singlet hot spots. David and Sternberg showed that disease-causing and

destabilizing SAVs have tendency to occur in hot spot residues [141]. As well as their energetic contribution, hot spots at PPI interfaces are observed to fluctuate less than the other residues therefore they are assumed to form ready-to-bind motifs prior to binding [157, 158]. Molecular dynamics simulations of a set of monomer and complex structures showed that hot spot residues in unbound monomers had significantly lower flexibility as compared to their environment residues [159].

Chapter 3

MATERIALS and METHODS

3.1. MD Simulations

3.1.1. Generating initial configurations of the Cdc42–IQGAP2 and Rac1–IQGAP2 Complexes

The simulation inputs were built using VMD version 1.9.2 [160]. Bonded parameters for the GTP were adopted from previous works [7, 161-163]. The crystal structure of the GRD2 dimer bound to four Cdc42s (PDB entry 5CJP) provided a base structure for this part of the thesis. In the structure, dimerization occurs between the Ex-domains of two GRD2s. Especially, the C-terminal portions (residues 1204 –1246) are mainly involved in the dimerization, and they intertwine with each other This structure including four Cdc42s bound to GRD2 was directly used to perform Ex-mode domain bound Cdc42 (hereafter referred to as Ex-mode Cdc42)–GRD2 (chain A and chain E) and RasGAP site bound Cdc42 (hereafter referred to as RasGAP mode Cdc42)–GRD2 (chain C and chain E) simulations. For Ex-mode Rac1–GRD2 and RasGAP mode Rac1–GRD2 complexes, the crystal structure of wildtype Rac1 (PDB entry 3TH5) was used. The structures were modeled using PRISM web server [164, 165]; the crystal structure of Cdc42–GRD2 was used as template, and initial configurations for Ex-mode Rac1–GRD2 and RasGAP Rac1-GRD2 complexes were obtained. GNP, a GTP analog bound to Rac1, was changed to GTP to be consistent with GTP-bound Cdc42–GRD2 complexes. Totally five initial configurations were prepared in an aqueous environment; one apo-GRD2, two Cdc42–GRD2 (Ex-mode Cdc42–GRD2 and RasGAP mode Cdc42–GRD2), two Rac1–GRD2 (Ex-mode Rac1–GRD2 and RasGAP mode Rac1–GRD2) (Figure 3.1). Then two additional initial configuration was prepared to introduce the selected mutation to the systems. Tyr1106 residue of GRD2 was in silico mutated to Ala using Discovery Studio version 4.1. Then, one RasGAP mode Cdc42-mutated

GRD2 (Y1106A) complex, and one RasGAP mode Rac1-mutated GRD2 (Y1106A) complex were built using VMD.

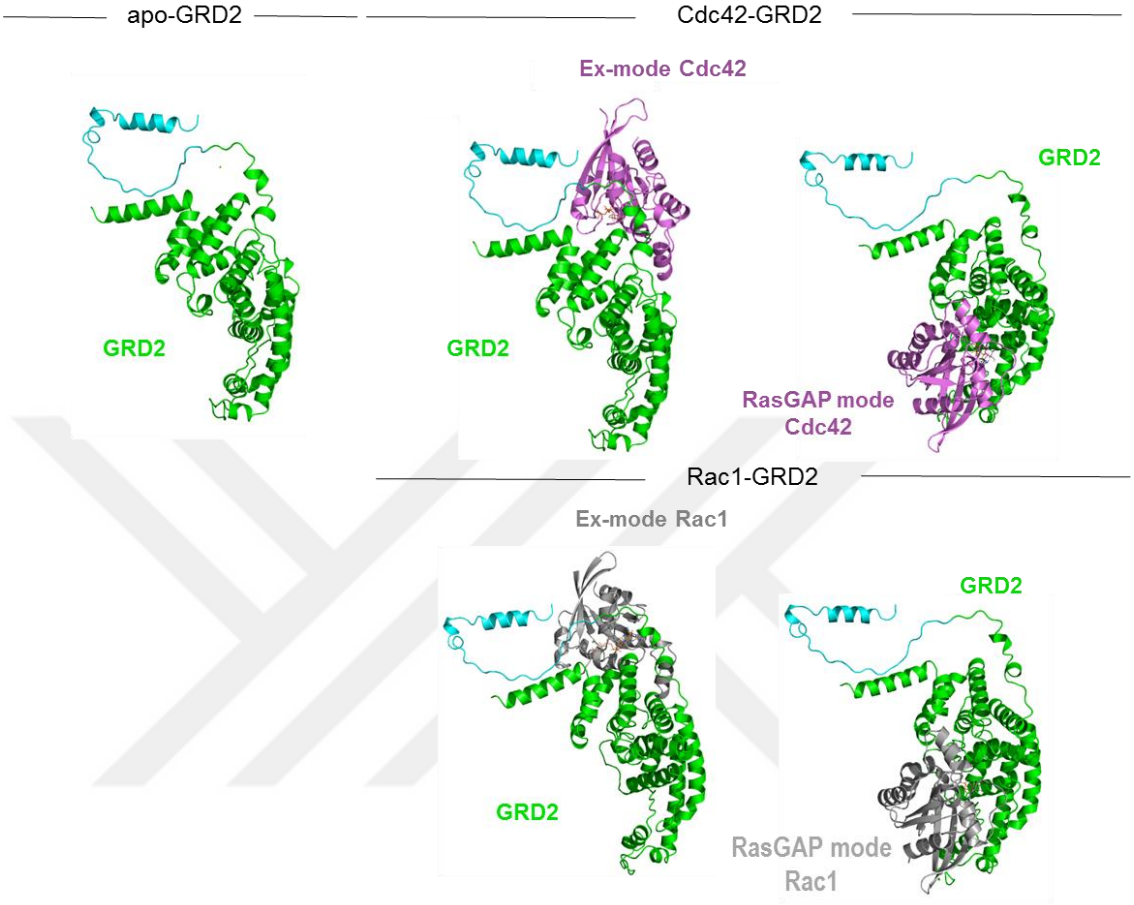


Figure 3.1. Initial configurations of the simulated Cdc42–GRD2 and Rac1–GRD2 systems. In the upper panel, apo-GRD2 and systems with Cdc42 are shown, and in the lower panel, systems with Rac1 are represented. Pink and grey molecules are Cdc42 and Rac1, respectively. The C-term of Ex-domain is represented by cyan, while the rest of the GRD2 is green (GRD2 and Cdc42 PDB ID: 5CJP [1] and Rac1 PDB ID: 3TH5 [166]).

3.1.2. Generating initial configurations of the Arl2–PDEδ–KRas4B HVR Complexes

The simulation inputs were built using CHARMM [167] c39b2 with the CHARMM 36 force field [168]. Bonded parameters for the GTP, the prenylated and geranylgeranylated cysteine were adopted from previous works. [7, 161-163]. PDEδ–KRas4B HVR complex structures were adopted from a previous work [9]. The

geranylgeranyl moiety is relatively longer than farnesyl moiety, the former is embedded deeper in PDEδ hydrophobic cavity resulting from the Phe133 aromatic ring flipped downward (state 1). In contrast, the latter causes a shallower cavity resulting from the Phe133 aromatic ring flipped upward (state 2). Two PDEδ structures were used with state 1 and state 2 Phe133 side chains (PDB code: 5E8F and 5F2U, respectively). The C-terminal KRas4B HVR (residues 167-185) was post-transitionally modified with the methylation and the farnesyl/geranylgeranyl groups. Since PDEδ interacting with geranylgeranylated HVR adopts state 1, only state 1 was modeled for PDEδ–geranylgeranylated HVR complex. PDEδ–farnesylated HVR complexes were modeled for both state 1 and state 2. In this thesis, Arl2–PDEδ crystal structures (PDB codes: 1KSH with Arl2-GDP and 1KSG with Arl2-GTP) were used to obtain Arl2–PDEδ–HVR complexes [5]. Arl2 sequences were modified into the human. For Arl2-GDP, the missing N-terminus portion ($Arl2_{1-16}$) was modeled with MODELLER [169] as an α-helix using Arl2-GTP as a template. The Ile residue was added to the C-terminus to avoid charged terminus. Totally eight initial structures (systems 1-8) were prepared in an aqeous environment (Table 3.1).

Table 3.1. Initial configurations of the simulated systems.

Systems	Arl2		KRas4B HVR	PDEδ	
	PDB	State (GDP/GTP)	Prenylation	PDB	Phe133 side chain / state
system 1	1KSH	GDP	-	-	-
system 2	1KSH	GDP	farnesyl	5E8F	Down / 1
system 3	1KSH	GDP	farnesyl	5F2U	Up / 2
system 4	1KSH	GDP	geranylgeranyl	5E8F	Down / 1
system 5	1KSG	GTP	-	-	-
system 6	1KSG	GTP	farnesyl	5E8F	Down / 1
system 7	1KSG	GTP	farnesyl	5F2U	Up / 2
system 8	1KSG	GTP	geranylgeranyl	5E8F	Down / 1

The first four systems (systems 1-4) were prepared using Arl2-GDP, while the last four systems (systems 5-8) include Arl2-GTP. Two Arl2–PDEδ complexes (systems 1 and

5), two Arl2–PDEδ (state 1)–farnesylated HVR complexes (systems 2 and 6), two Arl2–PDEδ (state 2)–farnesylated HVR complexes (systems 3 and 7) and two Arl2–PDEδ (state 1)–geranylgeranylated HVR complexes (systems 4 and 8) were obtained (Table 3.1 and Figure 3.2).

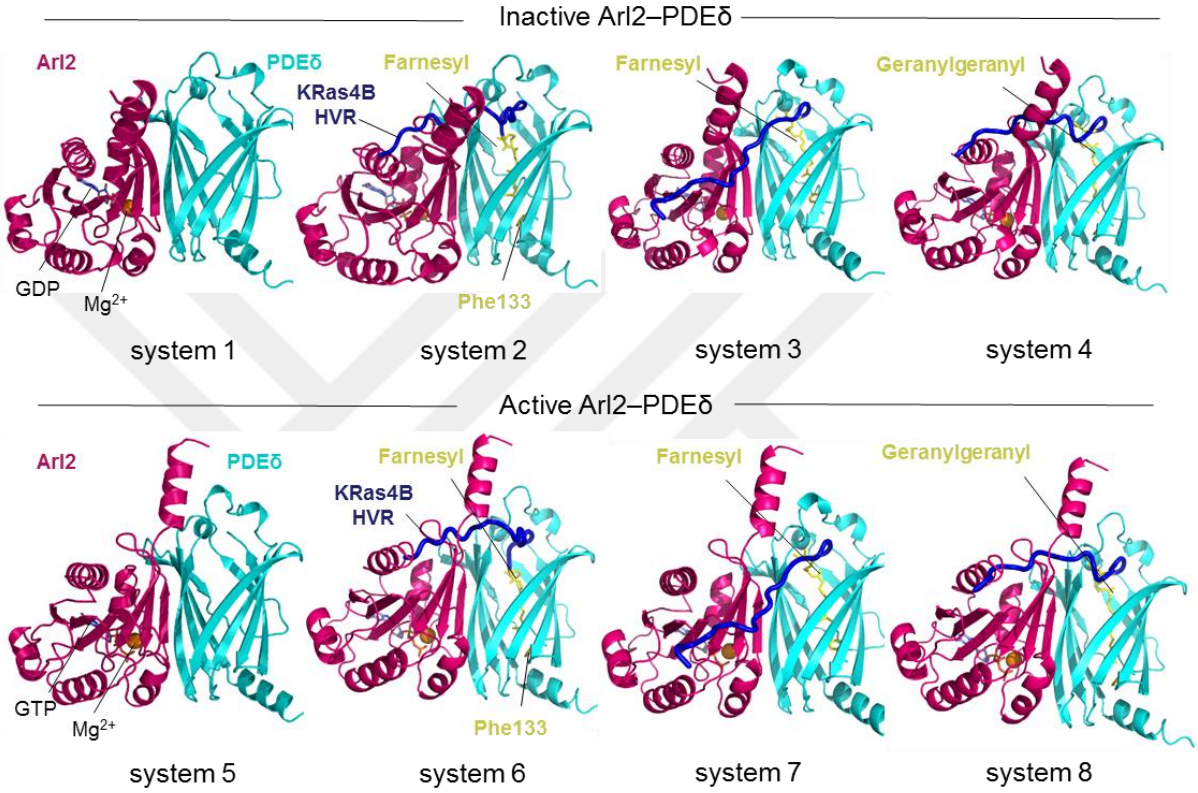


Figure 3.2. Initial configurations of the simulated Arl2–PDEδ–HVR systems. In the upper panel, systems with Arl2-GDP (systems 1-4) are shown, and in the lower panel, systems with Arl2-GTP (systems 5-8) are represented. Pink and cyan molecules are Arl2 and PDEδ, respectively. The KRas4B HVR is represented by dark blue tube, and HVR's prenyl groups and Phe133 residue of PDEδ are shown by yellow sticks. Both GDP and GTP are shown by blue sticks, and Mg^{2+} molecule is represented by orange sphere (Arl2 PDB ID: 1KSH/1KSG, PDEδ PDB ID: 5E8F/5F2U, HVR is modelled).

3.1.3. Atomistic MD simulations

Cdc42–GRD2 and Rac1–GRD2 complexes were solvated by the TIP3P water model and the isometric unit cell boxes containing the protein complexes were created. The unit cell box of each system contain around 80,000 atoms with ions, approximately 150 Na^+, 1 Mg^{2+}, and 130 Cl^-. Minimization was performed for 10,000 steps. In the

production runs, the Langevin temperature control maintained the constant temperature at 310 K, and the Nosé-Hoover Langevin piston pressure was kept at 1 atm. All simulations were performed in the NPT (constant number of atoms, pressure, and temperature) ensemble. A total of 3.5 μs of simulations for seven systems, each with 500 ns, were performed using the NAMD parallel-computing code [170] on a Biowulf cluster at the National Institutes of Health (Bethesda, MD) and on a yunus cluster at Koc University (Istanbul, Turkey). To exempt initial transients, the first 100 ns trajectories were removed, and thus averages were taken afterward. The simulated trajectories were analyzed using the CHARMM all- atom additive force field (version C36) programming package [171]. (Sample configuration and input files are presented in APPENDIX A)

The initial Arl2–PDEδ–KRas4B HVR complexes were solvated by the modified TIP3P water model [172] and gradually relaxed with the proteins held rigid. For both the farnesylated and geranylgeranylated Arl2–PDEδ–KRas4B HVR systems the unit cell box of 90 $Å^3$ contains around 75,000 atoms with ions, 35 Na^+, 1 Mg^{2+}, and 40 Cl^-, to satisfy a total cation concentration near 100 mM. The Langevin temperature control was used to maintain the constant temperature at 310 K, and Nosé-Hoover Langevin piston pressure control was used to maintain the pressure at 1 atm. A total of 5 μs simulation were performed for the ten Arl2–PDEδ–KRas4B HVR systems; each has 500 ns simulation with the constant temperature at 310 K. All simulations were performed in the NPT (constant number of atoms, pressure, and temperature) ensemble. The NAMD 2.10 [170] parallel computing code was employed in the production runs on a Biowulf cluster at the National Institute of Health.

To exempt initial transients, the first 100 ns trajectories were removed, and thus averages were taken afterward. The simulated trajectories were analyzed using the CHARMM all- atom additive force field (version C36) programming package [171].

3.1.4. MM-GBSA calculations

To compare the interaction strengths of the systems, binding energy calculations have been performed. To calculate the binding free energy of the complexes, the molecular mechanics energies combined with the Generalized Born and surface area continuum

solvation (MM-GBSA) method was used [173, 174]. The average binding free energy is derived from the Gibbs free energy,

$$\langle \Delta G_b \rangle = \langle \Delta G_{gas} \rangle + \langle \Delta G_{sol} \rangle - T\Delta S \qquad (1)$$

where $\langle \Delta G_b \rangle$ is the average binding free energy and $\langle \Delta G_{gas} \rangle$ is the gas phase contribution. The gas phase contribution to the binding free energy is a sum of the internal energy, the van de Waals interaction, and the electrostatic energy, $\Delta G_{gas} = \Delta H_{intra} + \Delta H_{vdW} + \Delta H_{elec}$. $\langle \Delta G_{sol} \rangle$ is the solvation energy contribution and it can be divided into the electrostatic contribution and the nonpolar contribution, $\Delta G_{sol} = \Delta G_{sol}^{elec} + \Delta G_{sol}^{nonpolar}$. The electrostatic contribution to solvation is calculated by solving the Generalized Born equation within the CHARMM program [175]. $G_{sol}^{nonpolar} = \gamma SASA + \beta$, where SASA denotes the solvent accessible surface area with a set of the surface tension parameter. $T\Delta S$ is the entropic contribution. The entropy term is divided into the translational, rotational, and vibrational contributions, $T\Delta S = T\Delta S_{trans} + T\Delta S_{rot} + T\Delta S_{vib}$. The translational and rotational entropies are evaluated from the calculation of principal moment of inertia, and the vibrational entropy is obtained from the quasiharmonic mode calculation in the VIBRAN module of the CHARMM program [175]. Angle brackets represent the average along the simulations. Overall, the average binding free energy was calculated as sum up all these contributions;

$$\Delta G = \Delta H_{intra} + \Delta H_{vdW} + \Delta H_{elec} + \Delta G_{sol}^{elec} + \Delta G_{sol}^{nonpolar} - (T\Delta S_{trans} + T\Delta S_{rot} + T\Delta S_{vib}) \qquad (2)$$

The change in binding free energy due to the complex formation is calculated as follows for Cdc42–GRD2 and Rac1–GRD2 complexes:

$$\Delta G_b = G_b^{complex} - \left(G_b^{\mathrm{Cdc42/Rac1}} + G_b^{GRD2} \right). \qquad (3)$$

3.1.5. Stability and interaction analyses

Equilibrium state of each system after the simulations and residue fluctuations were measured by RMSD and RMSF analyses, respectively. RMSD is a measurement indicating how two structures are different than each other by calculating root mean square distance between each corresponding atoms. RMSD is calculated as follows:

$$\mathrm{RMSD} = \sqrt{\frac{1}{\mathrm{N}}\sum_{\mathrm{i}=1}^{\mathrm{N}} \delta_{\mathrm{i}}^{2}} \qquad (4)$$

Where N is the number of atoms, δ_i is the distance between i[th] atoms from two different structures. RMSF is the measurement of average displacement of each residue throughput the simulations. RMSF indicates the more and less fluctuating residues and regions, it is a helpful measurement to observer overall behavior of specific residues and regions in the whole structures. RMSF is calculated as follows:

$$RMSF = \sqrt{\frac{1}{T}\sum_{t_j}^{T} (x_i(t_j) - \bar{\mathrm{x}})^{2}} \qquad (5)$$

Where T is the duration of the simulation in terms of time steps. $x_i(t_j)$ is the coordinates of i[th] residue at time t_j. $\bar{\mathrm{x}}$ is the mean coordinates for i[th] residue.

Inter- and intramolecular salt bridges and hydrogen bonds (H-bonds) established throughout the simulations were detected using VMD version 1.9.2. For salt bridge formation, atoms closer than 3.2 Å were considered forming salt bridges. In this thesis, salt bridges and H-bonds, which were detected over 20% of total time frames of the simulations were considered as stable interactions.

3.1.6. Allosteric pathway analysis

Allosteric paths induced by binding of Ex-domain Cdc42 to GRD2 and Arl2-GTP to PDEδ were determined by WISP algorithm [176]. The algorithm can identify the shortest primary communicating pathway through nodes that are protein residues of interest and edges connecting the nodes. 100 optimal pathways between selected residue pairs on Ex-mode Cdc42 and GRD2, and Arl2 and PDEδ were calculated. The residue pair selection was made by considering residues that are in those regions that are mostly

involved in the interactions and have high positive cross-correlation. The cross-correlation of the atomic fluctuations throughout the simulation was calculated. Positively correlated residues with a high correlation coefficient move in the same direction. WISP algorithm was compiled as a VMD plugin.

3.2. SAV Analyses

3.2.1. Mapping SAVs to PPI interface residues

SAV data used in this part of the thesis is taken from SKEMPI 2.0, which is not publicly available yet but can be obtained from SKEMPI team. The data format is the same with previous SKEMPI dataset (http://life.bsc.es/ pid/mutation database/) [177]. This data set includes experimentally measured affinities of wild-type complexes and affinities upon SAVs collected from scientific literature. In the SAV analysis, insertion and deletion variations and entries with multiple variations are excluded. Around one-third of SAVs is obtained from alanine scanning experiments [177].

PPI interfaces are extracted from a previous data set, Piface [178]. Piface is a structurally non-redundant protein-protein interface data set. This data set has been generated by structurally comparing and clustering 130209 protein-protein interfaces extracted from protein data bank [179]. This interface set has been also used in template-based modeling [180]. A PPI interface is described as the contact region between two interacting proteins. The distance between any two heavy atoms of the two residues from different chains should be less than the sum of their Van der Waals radii plus 0.5 Å in order to define them as contacting residues, nearby residues are any residues within the 6 Å distance from alpha carbon atoms of contacting residues on the same chain. The interface scaffold is built by combining both contacting and nearby residues [178, 181, 182]. Other residues which are not core and interface residues are classified as surface residues. Hot spot residues having smaller distance than 6 Å between their Cα atoms are considered as in the same hot region, at least three hot spot residues are required to form a hot region [156].

In this part of the thesis, SAVs from SKEMPI 2.0 are mapped to contacting residues of Piface interfaces of the proteins which have interfaces in Piface. However, the main

focus is SAVs in contacting residues. Since some of the interfaces have high sequence similarity, redundancy of interfaces and SAVs in contacting residues are checked after mapping. It is seen that more than 10% of the SAVs are coming from redundant sequences. These SAVs were not excluded, since they may be involved in different interactions.

The mapping procedure consists of four steps: 1) First PDB IDs obtained from SKEMPI 2.0 data set and PDB IDs of interfaces from Piface are compared and matching PDB IDs are listed. 2) Then, SAV residues are matched with contacting residues of PPI interfaces. 3) Using HotRegion [155], KFC2a [183] and Robetta [184] web servers, computational hot spots are found on contacting residues. Majority voting is performed using the hot spot predictions of these three servers to determine hot spots. For some interface residues, Robetta result could not be obtained, therefore KFC2b results are used in majority voting in such cases. Then, hot regions are determined using the approach explained above. 4) Finally, hot spot residues and hot regions are matched with SAV residues to identify singlet hot spot SAVs, hot region SAVs and non-hot spot SAVs.

3.2.2. Determining changes in binding free energy upon SAVs

In order to calculate the binding free energy changes in proteins upon SAVs, the experimental affinity data (K_d) from SKEMPI 2.0 is used. Binding free energy is calculated for each SAV (ΔG_{SAV}) and wild type (ΔG_{WT}) residue (*Eq. 6*) where R is the universal gas constant. Then, $\Delta\Delta G$ values are calculated with these ΔG values for experiments conducted at different temperatures (*Eq. 7*). For most of the cases, temperature is equal to 298K. However, for some cases, experimental temperature varies between 273 and 323 K. Temperature (T) in the following relations is changed according to different experimental temperatures.

$$\Delta G = \ln(K_d) RT \qquad (6)$$

$$\Delta\Delta G = \Delta G_{SAV} - \Delta G_{WT} \qquad (7)$$

where R=8.314⁄4184 kcal/K.mol and T = 273.15+25 K. There are some entries with several experimental data for the same SAV in SKEMPI 2.0, for these variations the average of the $\Delta\Delta G$ values are taken. If these several $\Delta\Delta G$ values for the same SAV

differ more than 1.5 kcal/mol, these outlier cases are removed before calculating the averages. Therefore, 2 cases are removed in this analysis. Removed cases are variation of aspartic acid to alanine in the 39th residue of D chain of 1BRS and variation of threonine to glutamic acid in the 17th residue of I chain of 1CHO.

A threshold is applied to categorize effect of SAVs as destabilizing, stabilizing and neutral. SAVs with $\Delta\Delta G$ values larger than 0.5 kcal/mol are considered as destabilizing while $\Delta\Delta G$ values smaller than -0.5 kcal/mol indicate stabilizing SAVs. Neutral SAVs have $\Delta\Delta G$ between -0.5 kcal/mol and 0.5 kcal/mol [185].

3.2.3. Determining functional effects of variations on proteins

PolyPhen2 web server is used to evaluate the disease-association of SAVs in contacting residues [186, 187]. PolyPhen2 evaluates a variation as probably damaging (more confident prediction), possibly damaging (less confident prediction) or benign (harmless) based on HumDiv model. HumDiv is compiled from all damaging alleles with known effects on the molecular function causing human Mendelian diseases from the UniProtKB database. In the analysis, protein structures which belong to other organisms rather than human are excluded and both probably and possibly damaging SAVs are considered as disease-causing.

HGMD [188], dSysMap [189], PinSNP [190], COSMIC [191] and human disease mutation data from Sahni et al. [144] data sets are used to match human SKEMPI 2.0 SAVs with known disease variations to strength the analyses utilizing PolyPhen2 predictions. In SKEMPI 2.0, PDB location of SAVs are provided, while in these data sets listed above the UniProt location of SAVs are given. Therefore, after converting PDB residues to UniProt residues, SKEMPI 2.0 SAVs are matched with variations in these data sets. In order to obtain energy distribution profile of disease-causing and benign SAVs, $\Delta\Delta G$ values of 298 disease-causing and 191 benign SAVs. Statistical analysis of the distributions is obtained by two tailed, unpaired t-test.

3.2.4. Calculating χ^2 and odds ratio

χ^2 and p-values are calculated using the number of destabilizing, stabilizing and neutral SAVs in singlet hot spot residues, hot regions and non-hot spot residues to test the dependence of change in binding free energy on SAV location. In order to find the preferences of SAVs to be located in the different regions or residues of proteins, odds ratios are calculated according to

$$OR_{ij} = \frac{x_i/(1-x_i)}{x_j/(1-x_j)} \quad (8) \quad \& \quad x_i = n_i/N_i \quad (9)$$

where x_i is the probability of observing a variant in the specific region *i*, N_i is the total residue number in the specific region *i* and n_i is the number of variation in the specific region *i*. Two-tailed p-values are calculated with Fisher's exact test to test statistical significance of the OR values using the statistical packages in R version 3.3.1 (https://www.r-project.org/).

Chapter 4

Cdc42–IQGAP2 and Rac1–IQGAP2 INTERACTIONS

4.1. Cdc42 Insert-Loop Binding to the Ex-domain of GRD2

In the crystal structure of the Cdc42–GRD2, there are four Cdc42s bound to GRD2 dimer. The dimerization occurs between the Ex-domains of two GRD2s. Especially, the C-terminal portions (residues 1204-1246) are mainly involved in the dimerization, and they intertwine with each other (Figure 4.1).

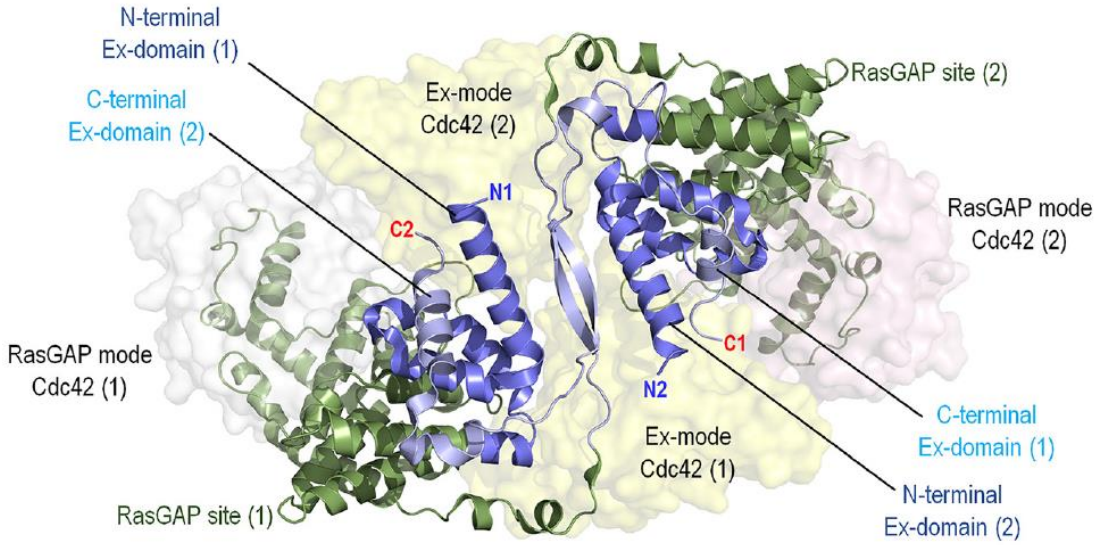


Figure 4.1. Crystal structure of GRD of IQGAP2 (GRD2) dimer with four Cdc42 molecules. Ex-domains form a dimerization interface, and the C-term of the Ex-domain intertwines with the counterpart C-term of the other GRD2. Each monomeric unit of GRD2 contains two Cdc42 molecules at the RasGAP site and the Ex-domain. The *number* in *parenthesis* denotes two monomeric units of GRD2 (PDB ID: 5CJP).

Atomistic MD simulations were performed on the Ex-mode Cdc42–GRD2 complex for 500 ns. The equilibrium conditions were tested checking RMSD and energy plots as a function of time (Figure B.1 and Figure B.2 in APPENDIX B). During the simulation, the bound C-term vigorously fluctuates compared with the C-term of apo-GRD2 (Figure 4.2, apo-GRD2 in the left panel and Ex-mode Cdc42–bound GRD2 in the right panel).

Root mean square fluctuation (RMSF) calculations show that the residues in C-term (residues 1204–1246, right panel in Figure 4.2) fluctuate much more than the other residues of GRD2 upon Ex-mode Cdc42 binding. Because of this fluctuation, this region undergoes conformational changes; it is released and freely moves (Figure 4.2). Throughout the simulation, the C-term points in different directions.

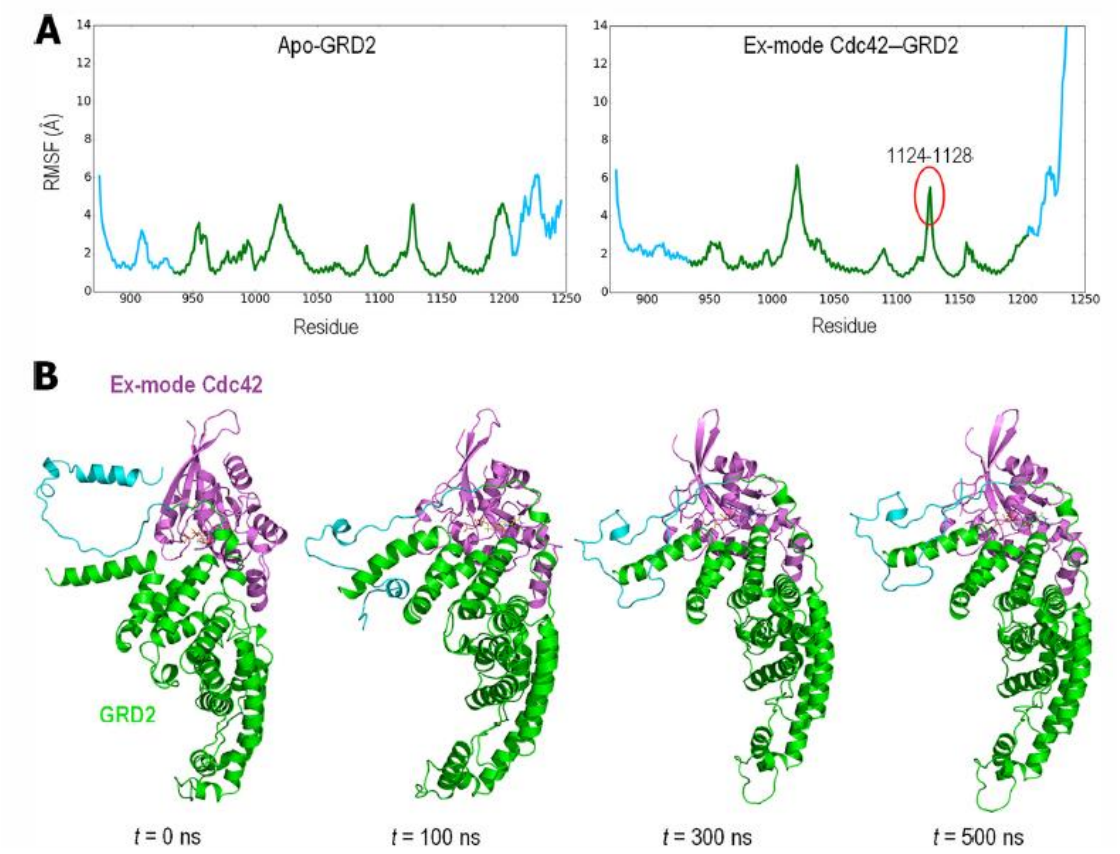


Figure 4.2. C-term fluctuations in GRD2 upon Ex-mode Cdc42 binding A) RMSF plots of apo-GRD2 (*left*) and GRD2 from Ex-mode Cdc42-GRD2 complex (*right*). *Green lines* cover the RasGAP site residues, and the remaining residues (*cyan lines*) belong to the Ex-domain. In the *right panel*, the *red ellipse* represents the residues between 1124 and 1128 that conformationally change upon Ex-mode Cdc42 binding. In the *right panel*, C-term residues show higher fluctuation than 10 Å (up to 20 Å); however, to clearly observe and compare fluctuation profiles of GRD2s from different complexes, the *y* axis limit is determined as 10 Å. **B)** Time series of snapshots of Ex-mode Cdc42-GRD2. *Pink molecules* are Cdc42s, and *green molecules* are GRD2,

whereas *cyan-highlighted residues* are the C-terms of the Ex-domain of GRD2s (PDB ID: 5CJP).

To confirm that the conformations with released C-term are highly populated throughout the Ex-mode Cdc42-GRD2 simulation, the clustering analysis was performed with respect to RMSD. The clustering analysis showed that C-term is released in 90% of the conformations (Figure 4.3). This indicates that C-term–released conformations are highly populated throughout Ex-mode Cdc42-GRD2 simulation. These findings suggest that Cdc42-induced release of the C-term provides a platform for interacting with the C-term of the second GRD2 and facilitates dimerization.

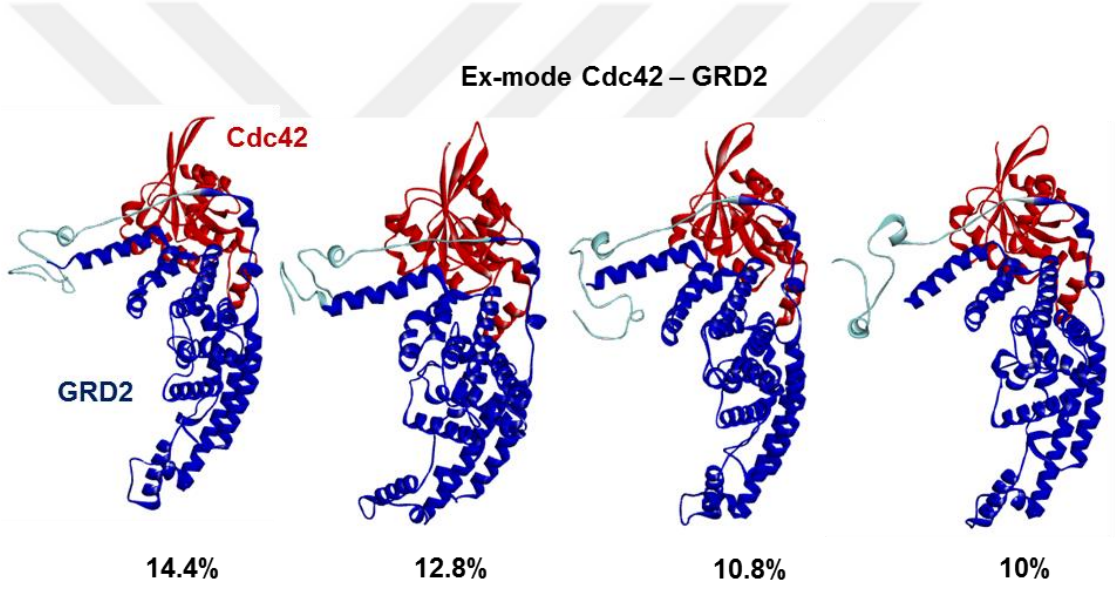


Figure 4.3. Clustering analysis for Cdc42–GRD2 interactions. Snapshots and populations of the representatives of four most populated conformational clusters from Ex-mode Cdc42–GRD2 complex (PDB ID: 5CJP).

In the crystal structure study, it was suggested that Ex-mode binding must precede Cdc42 binding to the RasGAP mode; however, how the binding of the first Cdc42 facilitates the binding of the second was unknown [1]. Throughout the Ex-mode Cdc42-GRD2 simulation, as well as the fluctuation of C-term, the changes in the RasGAP site that enable the RasGAP mode Cdc42 binding were observed. To investigate how the binding of the Ex-mode Cdc42 allosterically affects the conformation of the RasGAP site, a dynamical network analysis was conducted using weighted implementation of the

suboptimal paths (WISP) algorithm [176]. The algorithm can identify the shortest primary communicating pathway through nodes that are protein residues of interest and edges connecting the nodes. For this analysis, 100 optimal pathways between selected residue pairs on Ex-mode Cdc42 and GRD2 were calculated. The residue pair selection is made by considering residues that are in those regions that are mostly involved in the interactions and have high positive cross-correlation. The cross-correlation of the atomic fluctuations was calculated throughout the simulation. Positively correlated residues with a high correlation coefficient move in the same direction. Numerous residue pairs were chosen, but pathways were obtained between only five of them. All possible pathways obtained by WISP start with Ex-mode Cdc42 residue Asn-132 and terminate at GRD2 residues 1124–1128 (Figure 4.4), which have a high fluctuation profile (Figure 4.2 A, residues in red ellipse).

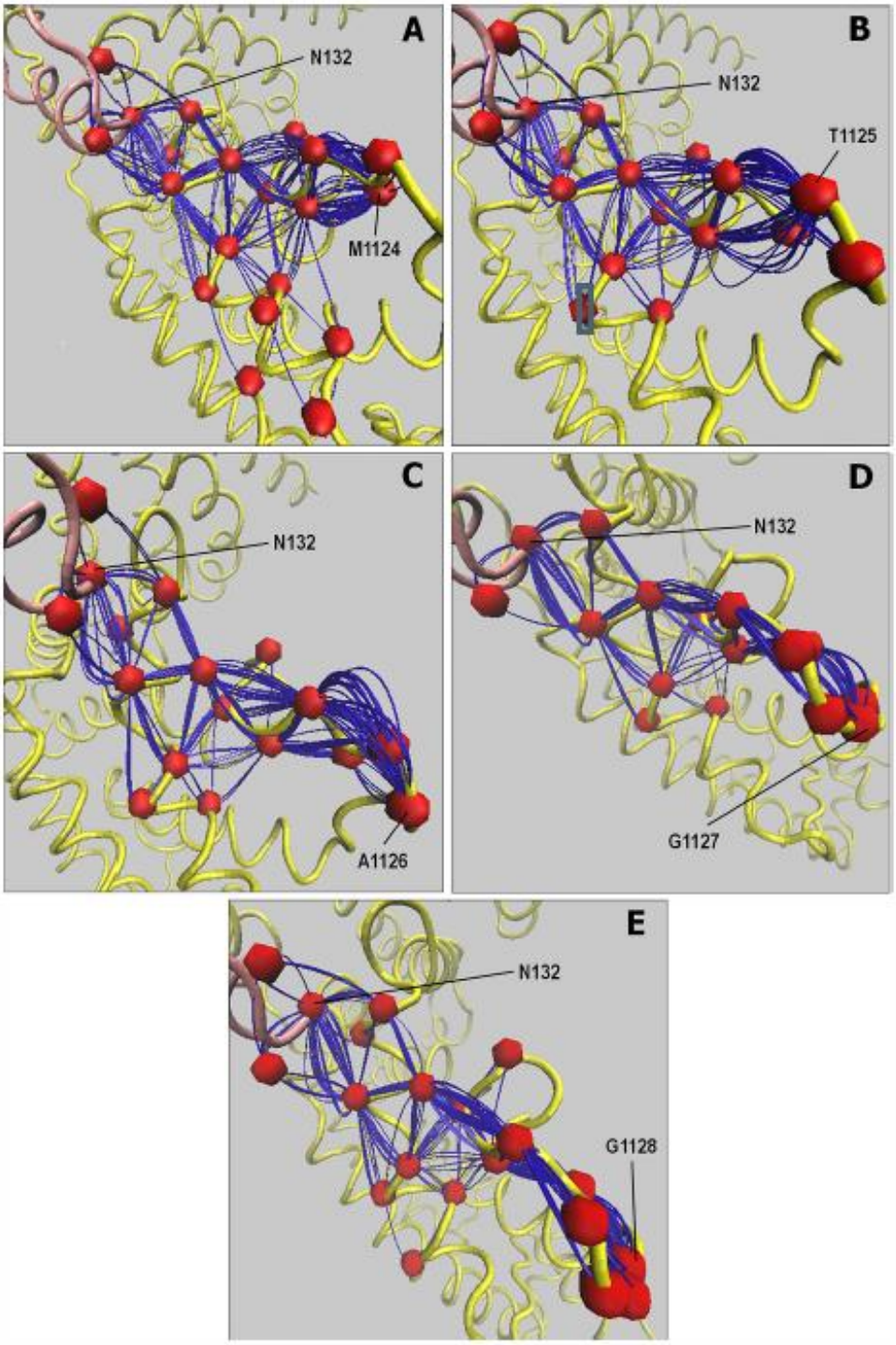


Figure 4.4. Allosteric pathways between Ex-mode Cdc42 and GRD2. A) Asn-132/Met-1124, **B)** Asn-132/Thr-1125, **C)** Asn-132/Ala-1126, **D)** Asn-132/Gly-1127, **E)** and Asn-132/Gly-1128 of Ex-mode Cdc42 and GRD2 obtained by the WISP method. *Pink structures* are Cdc42s, and *yellow structures* are GRD2s. The *red balls* represent the residues involve in allosteric pathways, and the *blue lines* represent the possible paths followed by allosteric pathways.

The frequently involved residues in each of the500 pathways (100 pathways for each five residue pairs) are Gln-983, Ile-1122, Arg-982, Asp-1123, and Ile-1121 of GRD2 (Table 4.1). Allosteric changes in residues 1124–1128 of GRD2, initiated by Asn-132 of Ex-mode Cdc42, prepare the RasGAP site for Cdc42 binding.

Table 4.1. Frequently involved residues in allosteric pathways

Residue	Gln-983	Ile-1122	Arg-982	Asp-1123	Ile-1121
Freq.(%)	14	14	14	11	9

Analysis of the interaction interface of Ex-mode Cdc42 with GRD2 identifies the regions and residues playing key roles. Stable salt bridges and H-bonds between Ex-mode Cdc42 and GRD2 generally involve insert-loop and switch I residues of Cdc42 (Figure 4.5).

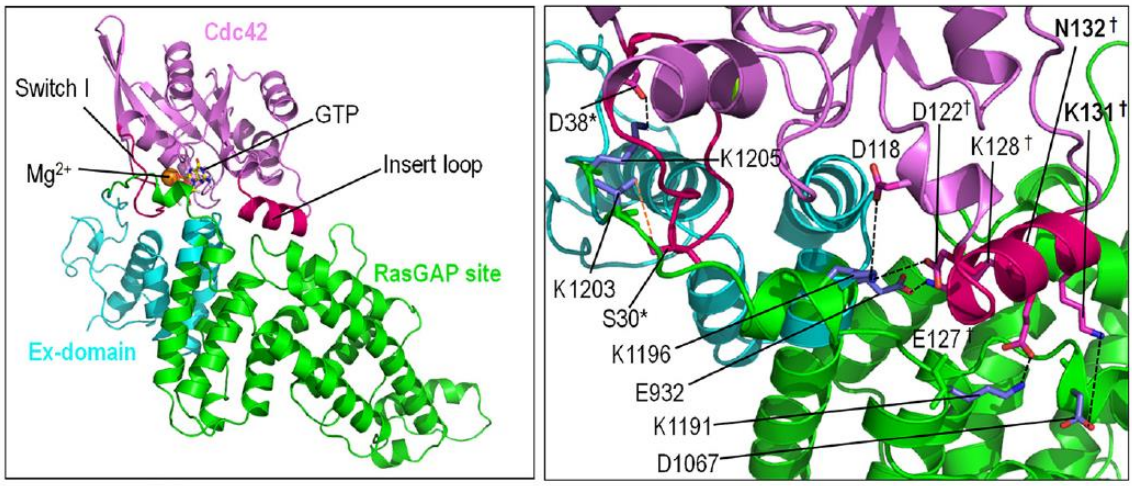


Figure 4.5. Interactions of Ex-mode Cdc42 with GRD2. On the *right-hand side*, stable salt bridges and stable H-bond formed between Ex-mode Cdc42 and GRD2 can be seen. *Green molecules* are GRD2, and pink molecules are Cdc42. *Dark pink* regions are insert loops and switch I of Cdc42. In the residue label, insert loop residues (†) and switch I residues (*) are indicated. *Black dashes*, salt bridges; *orange dashes*, H-bond. The residues Asn-132 and Lys-131 of Cdc42 are *highlighted* by *boldface type* (PDB ID: 5CJP).

Residue Asn-132 of the insert loop, which is identified as the starting point of the allosteric changes in the RasGAP site, forms a stable hydrophilic interaction throughout the simulation that confirms its importance. Previous experimental studies also showed that Asn-132 residue plays an important role in the stability of Cdc42–GRD binding [1, 32]. It has been shown that mutations in this residue significantly affect binding affinity of Cdc42 to GRD1 [32]. Also, it has been reported that Asn-132 is involved in stable interactions with GRD2 [1]. The insert loop is a Rho GTPase–specific α-helix structure, rich in both positively and negatively charged amino acids. This loop plays an important role in the interactions of Rho GTPases [32, 33]. Taken together, this analysis suggests that the binding mechanism of Ex-mode Cdc42 to GRD2 exploits the highly charged insert loop and switch I residues and that this interaction, especially that involving residue Asn-132, induces allosteric changes in the RasGAP site to facilitate the RasGAP mode Cdc42 binding. The Ex-mode Cdc42–GRD2 interaction then allosterically releases the C-term and increases its fluctuation to induce GRD2 dimerization.

4.2. Rac1 Binding to the Ex-domain of GRD2

The Ex-mode Rac1–GRD2 interaction was modelled using PRISM web server, which is a template-based protein-protein interaction prediction tool [164]. 500 ns long atomistic MD simulations were performed on the Ex-mode Rac1–GRD2 complex. The equilibrium conditions were tested checking RMSD and energy plots as a function of time (Figure B.1 and B.2 in APPENDIX B). First, it was observed that the orientation of the insert loop of Rac1 differs from that of Cdc42 (Figure 4.6, *left*). Because of this difference, fewer stable interactions formed between Rac1 and GRD2 compared with the Ex-mode Cdc42–GRD2 (Figure 4.6, *right*). Although some insert-loop residues are involved in salt-bridge and H-bond formation, stable Cdc42 interactions outnumber the stable Rac1 interactions.

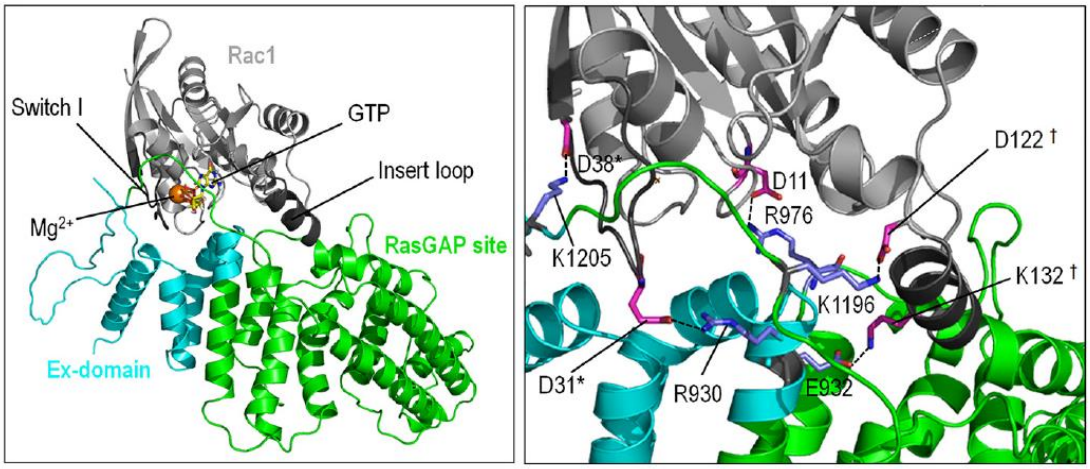


Figure 4.6. Interactions of Ex-mode Rac1 with GRD2. On the *right-hand side*, stable salt bridges and stable H-bond formed between Ex-mode Rac1 and GRD2 can be seen. *Green molecules* are GRD2, and grey molecules are Rac1. *Dark grey* regions are insert loops and switch I of Rac1. In the residue label, insert loop residues (†) and switch I residues (*) are indicated. *Black dashes*, salt bridges; *orange dashes*, H-bond (GRD2 PDB ID: 5CJP and Rac1 PDB ID: 3TH5).

When the structure and the sequence of Cdc42 and Rac1 were compared in detail, it was observed that there are some differences between the insert loops of Cdc42 and Rac1, both in sequence and structure. The sequence of the insert loop of Cdc42 is $_{120}$RDDPSTIEKLAKN^{132}, and that of Rac1 is $_{120}$RDDKDTIEKLKEK$_{132}$ (Figure 4.7). There is a 38% difference between the sequences of the insert loops of Cdc42 and Rac1 (5 of the 13 amino acids differ).

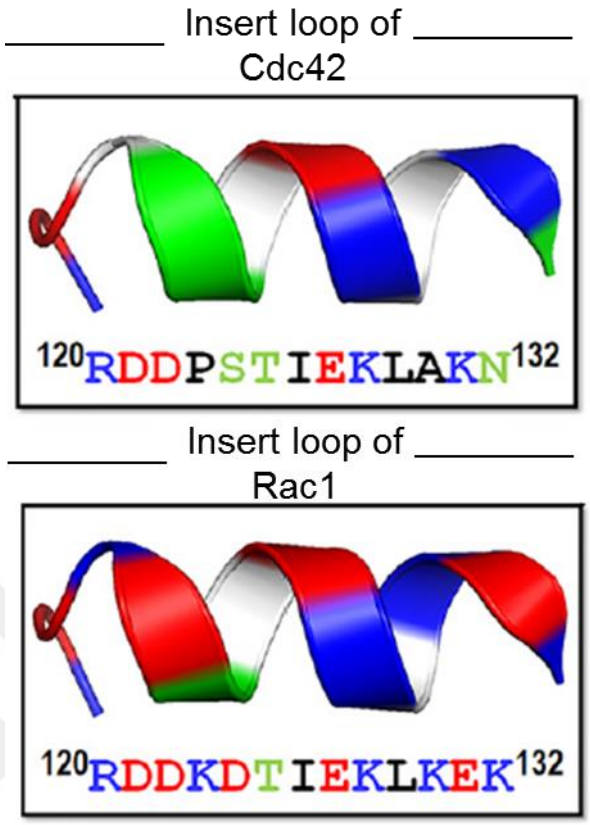


Figure 4.7. Amino acid sequences of insert loops of Cdc42 and Rac1

The polar Asn-132 residue of Cdc42 (Figure 4.5, *right*) was observed as an important residue in the analyses, inducing allosteric changes and involved in hydrophobic interaction; however, this residue is replaced with positively charged Lys-132 in Rac1. Previous experimental studies showed that mutating Asn-132 of Cdc42 to Lys decreases the affinity of Cdc42 for IQGAP1 [32]. Also, positively charged Lys-131 in Cdc42, which forms a salt bridge with Asp-1067 of GRD2 (Figure 4.5, *right*), is replaced by negatively charged Glu-131 in Rac1. These two changes result in disruption of the stable interactions between Ex-mode Cdc42 and GRD2. The five different amino acids affect Rac1 interactions and change the structure of its insert loop. The allosteric effects induced by residue 132 of Cdc42 cannot be observed in the RasGAP site of GRD2 after the Ex-mode Rac1 interaction. The changes in amino acid sequence in the Rac1 insert loop prevent it from inducing allosteric changes in the RasGAP site.

To compare the interaction strengths of these complexes, the binding free energy of Ex-mode Cdc42–GRD2 and Ex-mode Rac1–GRD2 were calculated using molecular

mechanics energies combined with the generalized Born surface area continuum solvation (MM-GBSA) method [192]. The calculated binding free energies for the Ex-mode Cdc42-GRD2 and the Ex-mode Rac1–GRD2 interactions are – 23.9 ± 9.2 and – 6.6 ± 8.2 kcal/mol, respectively (Figure 4.8). Compared with the Ex-mode Cdc42-GRD2 interaction, the Ex-mode Rac1–GRD2 interaction is significantly weaker, suggesting that because of the differences in the insert loop, Rac1 has a small number of residues involved in the intermolecular interaction with the Ex-domain of GRD2.

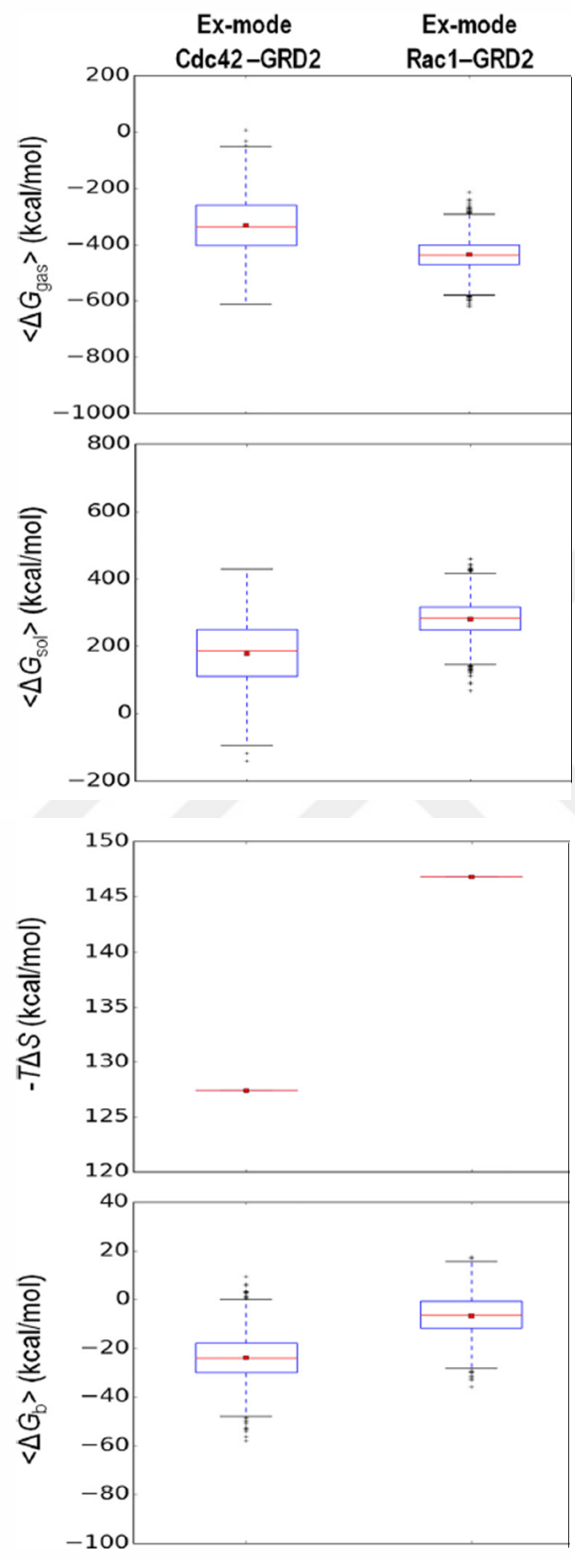


Figure 4.8. Binding free energies of Ex-mode Cdc42–GRD2 and Ex-mode Rac1–GRD2, in kcal/mol calculated by the MM-GBSA method. In the calculation, gas phase contribution, <Δ*G*gas>, the solvation energy contribution, <Δ*G*sol>, and the entropic

contribution, $-T\Delta S$, combine to give the average binding free energy, $\langle \Delta Gb \rangle$. In the box graphs, the *red dot* and *red line* denote the mean and median values, respectively. Whiskers above and below the box denote the 90th and 10th percentiles.

Consistent with these results, release of the C-term could not be observed throughout the simulation. Also, the fluctuation profile of GRD2 is similar to that of apo-GRD2 (Figure 4.9, *left and right*). Unlike the GRD2 from Ex-mode Cdc42-GRD2 complex (Figure 4.2, *right part of the right panel*), the C-term does not fluctuate vigorously (Figure 4.9, *right part* of *the right panel*). The C-term is not released but is folded inward, which closes the dimerization interface (Figure 4.9).

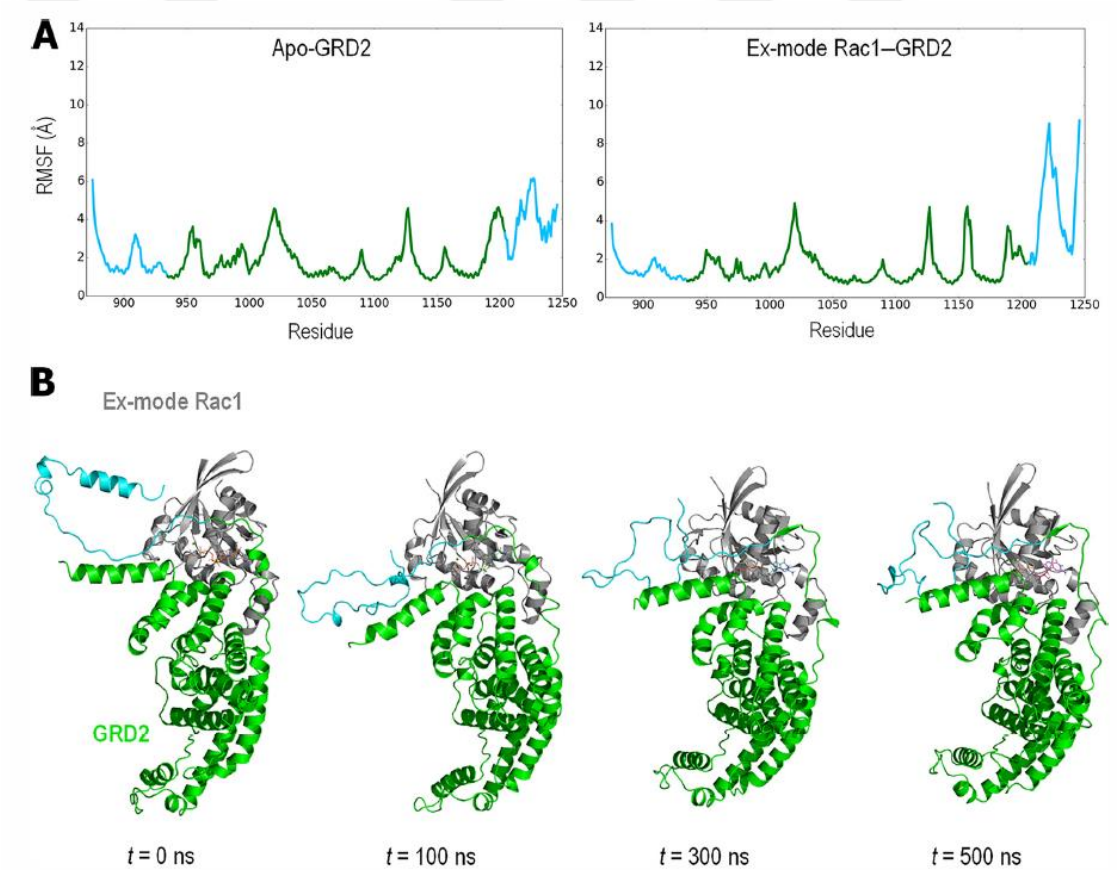


Figure 4.9. C-term fluctuations in GRD2 upon Ex-mode Rac1 binding A) RMSF plots of apo-GRD2 (*left*) and GRD2 from Ex-mode Rac1-GRD2 complex (*right*). *Green lines* cover the RasGAP site residues, and the remaining residues (*cyan lines*) belong to the Ex-domain. **B)** Time series of snapshots of Ex-mode Rac1-GRD2. *Gray molecules* are Rac1, and *green molecules* are GRD2, whereas *cyan-highlighted residues* are the C-terms of the Ex-domain of GRD2s (GRD2 PDB ID: 5CJP and Rac1 PDB ID: 3TH5).

Clustering analysis shows that the folded C-term conformation is highly populated throughout this simulation, and folded C-term populates 96.2% of all conformations (Figure 4.10).

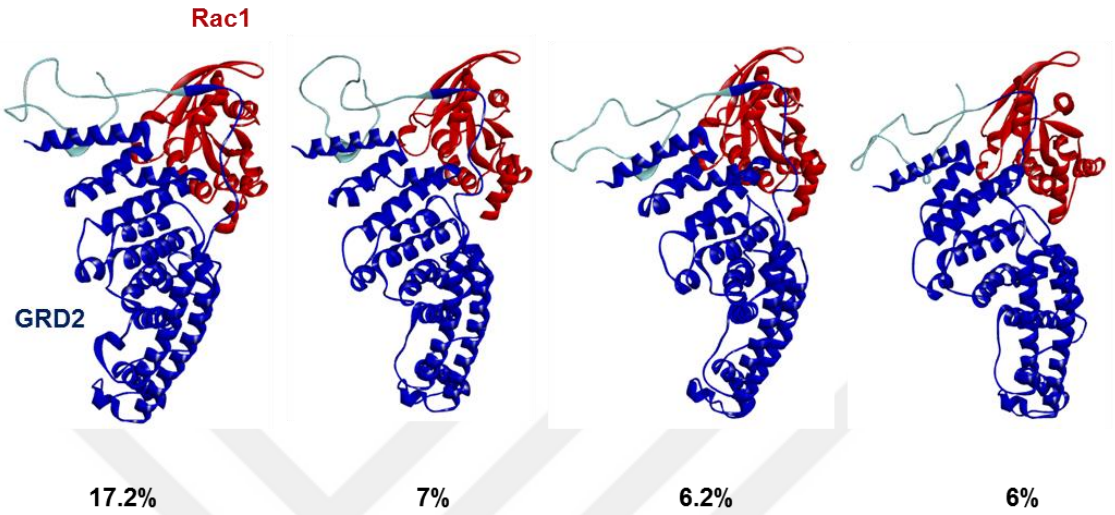


Figure 4.10. Clustering analysis for Rac1–GRD2 interactions. Snapshots and populations of the representatives of four most populated conformational clusters from Ex-mode Rac1–GRD2 complex (GRD2 PDB ID: 5CJP and Rac1 PDB ID: 3TH5).

As a result of this folding, exposure of the dimerization interface decreases, and the C-terms of the GRD2s cannot interact with each other. This finding also suggests that Rac1 cannot bind to the Ex-domain and cannot facilitate dimerization.

4.3. Cdc42 and Rac1 Binding to the RasGAP Site of GRD2

To observe the nature of RasGAP mode binding, RasGAP mode Cdc42–GRD2 and Rac1–GRD2 complexes were simulated for 500 ns. The equilibrium conditions were tested checking RMSD and energy plots as a function of time (Figure B.1 and B.2 in APPENDIX B). It was observed that the binding mechanisms of Cdc42 to the RasGAP site of GRD2 are significantly different from those of Ex-mode Cdc42. In the case of the Ex-mode Cdc42–GRD2 interaction, the insert loop plays an important role in stabilizing the interaction (Figure 4.5). However, the insert loop of Cdc42 is less involved in the RasGAP mode Cdc42–GRD2 interaction (Figure 4.11). Instead, the salt bridges and H-bonds established between RasGAP mode Cdc42 and GRD2 mostly engage switch I and switch II residues.

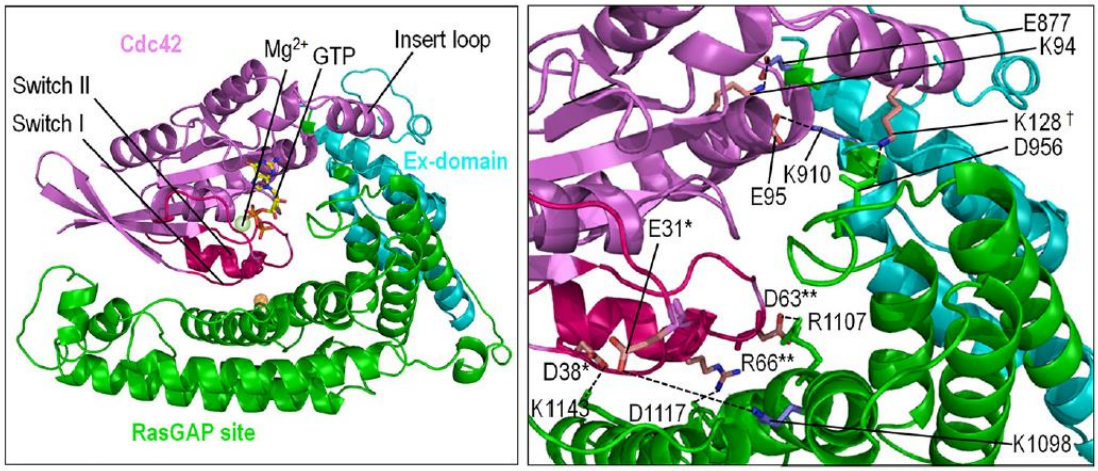


Figure 4.11. Interactions of RasGAP mode Cdc42 with GRD2. On the *right-hand side*, stable salt bridges formed between RasGAP mode Cdc42 and GRD2 can be seen. *Green molecules* are GRD2, and *pink molecules* are Cdc42. *Dark pink regions* are switch I and switch II of Cdc42. In the residue label, switch I residues (*) and switch II residues (**) are indicated. *Dashed lines*, salt bridges (PDB ID: 5CJP).

The binding mechanism of Rac1 to the RasGAP site of GRD2 shares similarities with the RasGAP mode Cdc42–GRD2 interaction, indicating that the RasGAP mode Rac1–GRD2 interaction also employs switch I and switch II residues (Figure 4.12). These findings also indicate that Rac1 can bind to the RasGAP site unlike the Ex-domain, because of stable interactions between RasGAP mode Rac1 and GRD2.

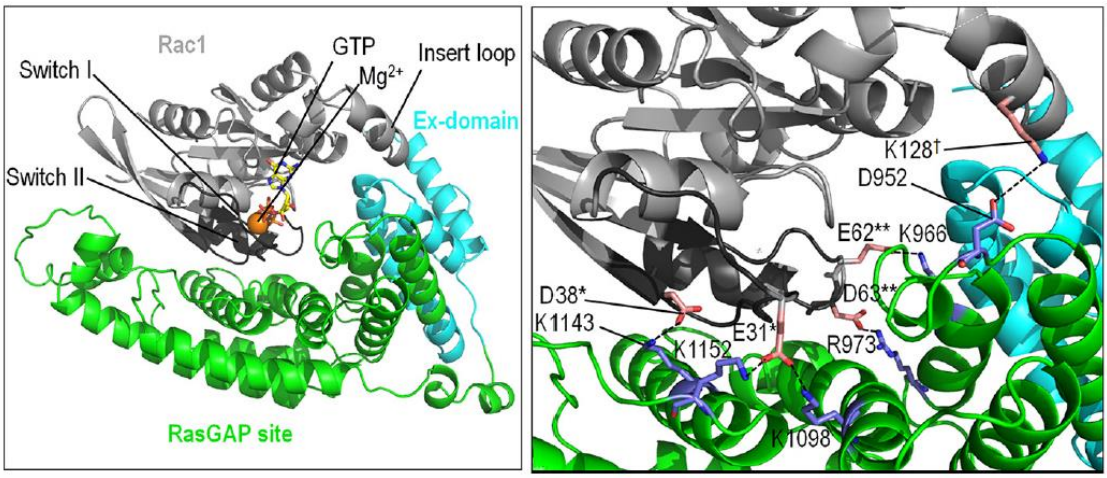


Figure 4.12. Interactions of RasGAP mode Rac1 with GRD2. On the *right-hand side*, stable salt bridges formed between RasGAP mode Rac1 and GRD2 can be seen. *Green*

molecules are GRD2, and grey *molecules* are Rac1. *Dark grey regions* are switch I and switch II of Rac1. In the residue label, switch I residues (*) and switch II residues (**) are indicated. *Dashed lines*, salt bridges (GRD2 PDB ID: 5CJP and Rac1 PDB ID: 3TH5).

Although binding free energy calculations imply that RasGAP mode Rac1 binding is weaker than RasGAP mode Cdc42 binding (–8.1 ± 10.2 kcal/mol *versus* –43.1 ± 14.1 kcal/mol, respectively; Figure 4.13), the GRD2 structure that was used in the simulations (PDB entry 5CJP, chain E) is already allosterically changed and preorganized for the RasGAP mode Cdc42 binding. Nonetheless, the RasGAP mode Rac1–GRD2 simulation still provides a framework for this interaction.

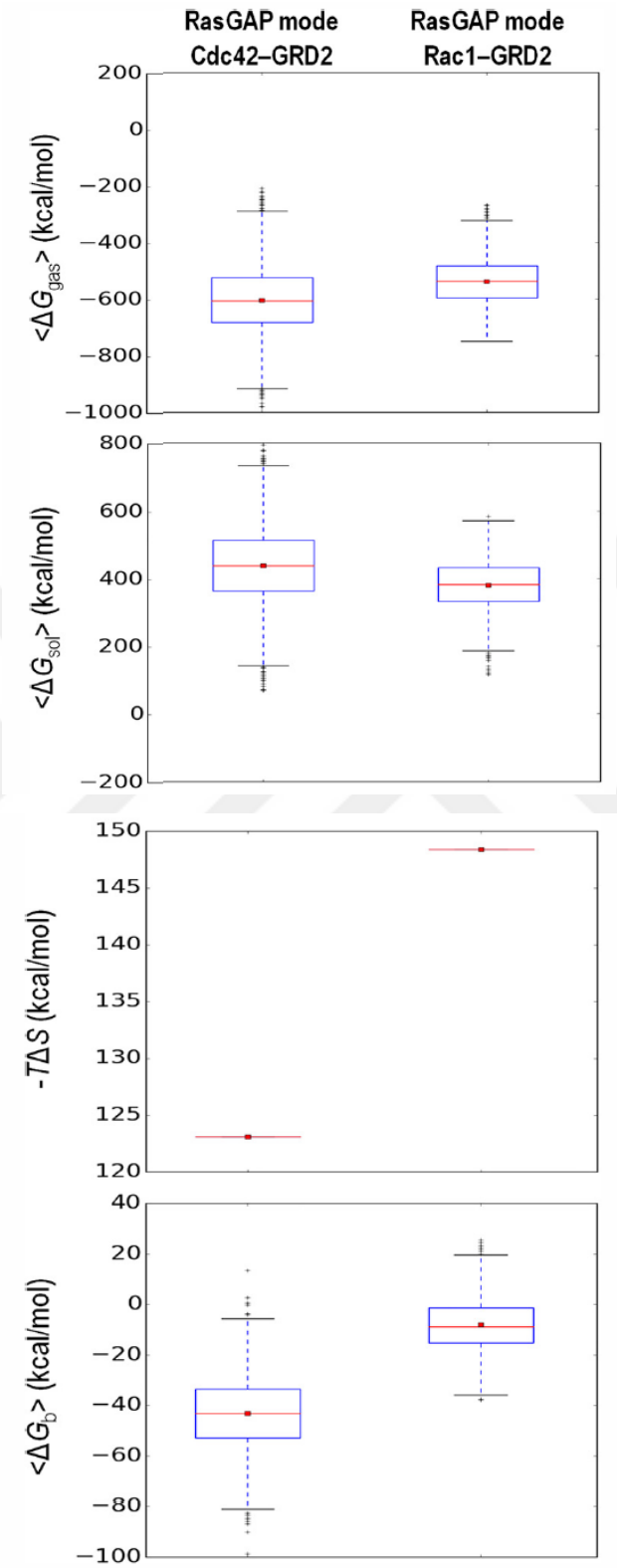


Figure 4.13. Binding free energies of RasGAP mode Cdc42–GRD2 and RasGAP mode Rac1–GRD2, in kcal/mol calculated by the MM-GBSA method. In the calculation, gas phase contribution, <Δ*G*gas>, the solvation energy contribution,

<ΔG_{sol}>, and the entropic contribution, $-T\Delta S$, combine to give the average binding free energy, <ΔG_b>. In the box graphs, the *red dot* and *red line* denote the mean and median values, respectively. Whiskers above and below the box denote the 90th and 10th percentiles.

Another similarity between RasGAP mode Cdc42 and Rac1 binding is the motion of the C-term. In both cases, the release of C-term was not observed, unlike in the case of Ex-mode Cdc42 binding, and a folded C-term appears throughout both simulations (Figure 4.14). This indicates that the RasGAP mode binding of neither Cdc42 nor Rac1 is sufficient to induce C-term release and facilitate GRD2 dimerization. It was concluded that Ex-mode Cdc42 binding is likely to be required for GRD dimerization, because this is the only case the release of the C-term was observed.

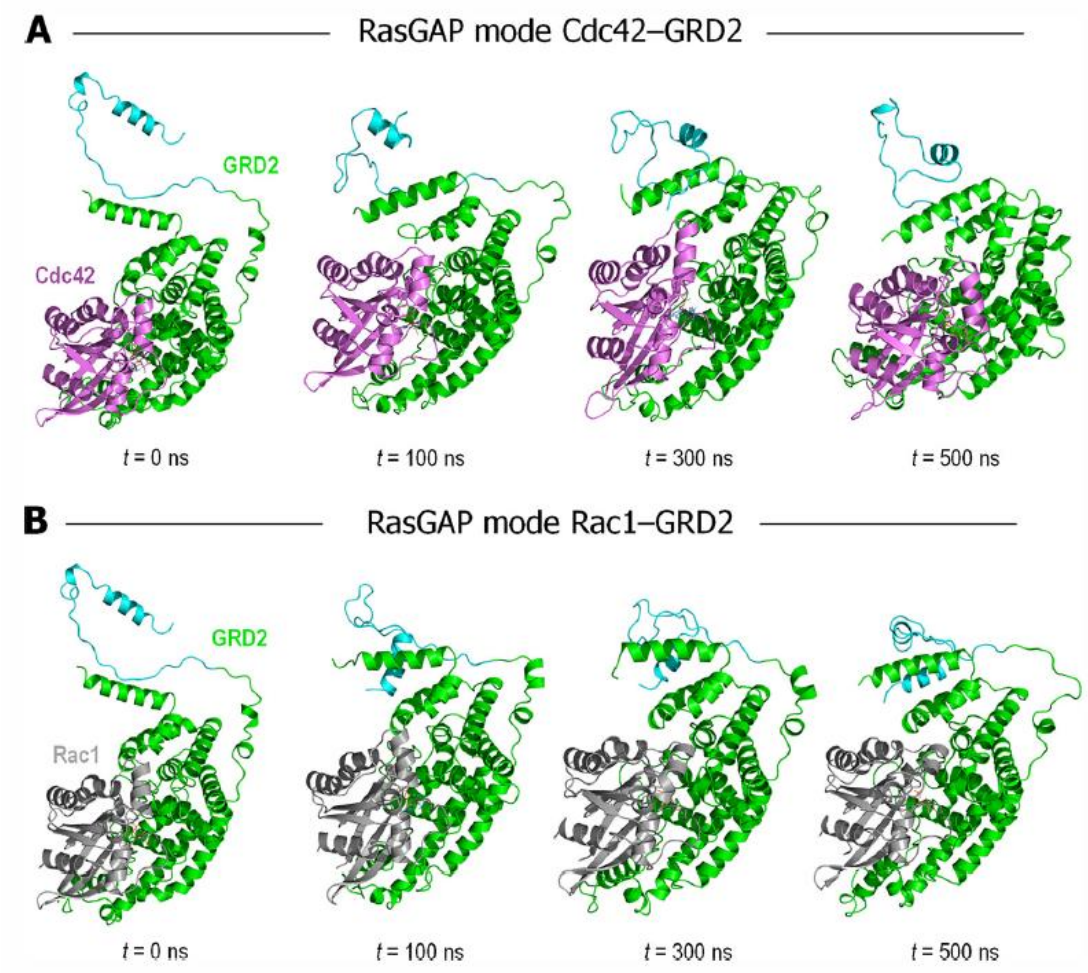


Figure 4.14. Time series of snapshots of RasGAP mode Cdc42 and Rac1 binding to GRD2. A) RasGAP mode Cdc42–GRD2 and **B)** RasGAP mode Rac1–GRD2. *Pink*

molecules are Cdc42, *gray molecules* are Rac1, and *green molecules* are GRD2, whereas *cyan-highlighted residues* are the C-terms of the Ex-domain of GRD2s (GRD2 PDB ID: 5CJP and Rac1 PDB ID: 3TH5).

4.4. Specific Mutation on RasGAP Site of GRD2

The energetically important residues in GRD2 were investigated to mutate and observe their effect on the Cdc42 and Rac1 binding. Ten hot spots, which are the residues contributing to the binding free energy most, were found in GRD2 using the HotRegion web server [193]. Tyr-1106 residue was chosen to mutate, because it has the highest pair potential among the other hot spot residues. Then MD simulations were performed for 500 ns after mutating *in silico* residue Tyr-1106 to alanine under the same conditions as the previous simulations of RasGAP mode Cdc42–GRD2 and RasGAP mode Rac1–GRD2 complexes. Because Tyr-1106 is a RasGAP site residue, RasGAP mode complexes were simulated. The equilibrium conditions were tested checking RMSD and energy plots as a function of time (Figure B.1 and Figure B.2 in APPENDIX B). *In silico* mutations were performed using the "mutate" function of Discovery Studio version 4.1. It was observed that the interaction of mutated GRD2 with Cdc42 and Rac1 is considerably weaker than the interaction of wild type GRD2 (Figure 4.15). The binding free energy for RasGAP mode Cdc42–GRD2 (wild type) interaction is –43.1 ± 14.1 kcal/mol; following mutation to alanine, it is –10.9 ± 11.5 kcal/mol. A similar decrease in binding strength is observed in the mutated RasGAP mode Rac1-GRD2 interaction; for the wildtype RasGAP mode Rac1–GRD2 interaction, it is –8.1 ± 10.2 kcal/mol, whereas for the mutant, it is 5.3 ± 11.8 kcal/mol.

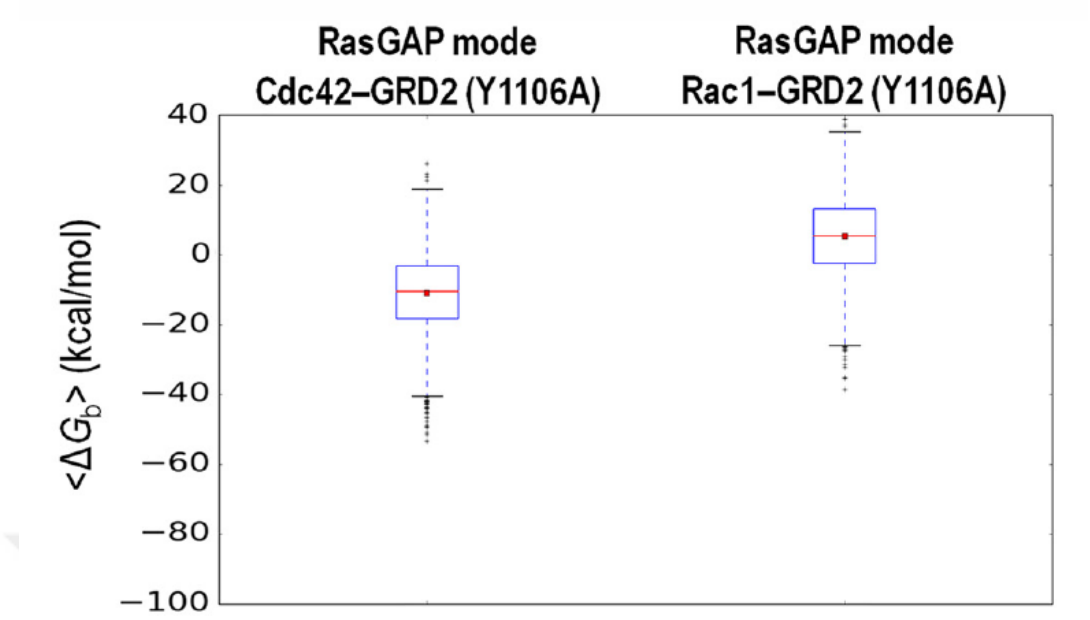


Figure 4.15. Binding free energies of RasGAP mode Cdc42-GRD2^{Y1106A} and RasGAP mode Rac1-GRD2^{Y1106A} in kcal/mol calculated by the MM-GBSA method

Also, in the mutated complexes, there are fewer stable salt bridges and H-bonds than in the wildtype interactions (Figure 4.16). Tyr-1106 forms hydrophilic interactions in wildtype Cdc42–GRD2 and Rac1–GRD2 complexes. Additionally, upon mutation, stable interactions of some of the nearby residues observed in the wild type simulations, such as Arg-1107 and Asp-956 (forming stable salt bridges with RasGAP Cdc42 residues (Figure 4.11)) and Asp-952 and Lys-966 (forming stable salt bridges with RasGAP Rac1 residues (Figure 4.12), are eliminated.

To validate the model, a construct comprising the C-terminal half of IQGAP1 in which Tyr-1193 (which corresponds to Tyr-1106 in GRD2) was changed to Ala was generated by our collaborators (David B. Sacks group, NIH). The abilities of the wildtype and the mutated IQGAP1 proteins to bind to purified Cdc42 and Rac1 *in vitro* were compared. GST-pulldown analysis showed that the wildtype IQGAP1 binds Cdc42 and Rac1. By contrast, mutating a single amino acid (Tyr-1193) in IQGAP1 caused a complete loss of the Rac1-GRD1 interaction and markedly attenuated binding of Cdc42 to GRD1. Together, these data argue for the importance of the Tyr-1106 residue in IQGAP2 and

provide a possible explanation for the complete loss of Cdc42 and Rac1 interactions with GRD2.

4.5. Discussion

This part of the thesis provides possible mechanisms for the interactions of Cdc42 and Rac1 with GRD2 in atomistic detail [1]. The C-term plays a role in dimerization of IQGAP via its GRD. In GRD2s, they intertwine and stabilize the dimer structure; their release upon Cdc42 binding to the Ex-domain induces dimerization. Through atomic-level conformations and allosteric pathway analysis, it was showed that upon binding of Ex-mode Cdc42, allosteric changes are induced in the RasGAP site. These allosteric changes create a suitable binding interface for the binding of Cdc42 to the RasGAP site.

When replacing Ex-mode Cdc42 with Rac1, with ~70% sequence identity, the C-term cannot be released; instead, it folds. The interaction interface for dimerization cannot be created when the C-term is folded; therefore, Rac1 cannot induce dimerization. Also, allosteric changes in RasGAP site is not observed upon Ex-mode Rac1–GRD2 interaction. This is consistent with the observation that Rac1 can bind to the RasGAP site of apo-GRD2 and does not require allosteric changes to bind this site. The binding affinity of Rac1 to the Ex-domain is considerably lower than the affinity of Cdc42 to the Ex-domain and there are fewer stable interactions between Ex-mode Rac1 and GRD2 than between Ex-mode Cdc42 and GRD2. Detailed analysis indicates that the insert loop, which is Rho GTPase–specific and important for effector binding, plays a role in Cdc42–Ex-domain interaction. There are some sequence differences between the insert loops of Cdc42 and Rac1, which may decrease the interactions between Rac1 and the Ex-domain. All of these findings imply that Rac1 cannot interact with the Ex-domain and that neither facilitated dimerization nor allosteric changes are induced by Rac1. These results support a 1:1 binding mode of Rac1 and GRD.

Although their interaction interfaces on the RasGAP site differ, the Cdc42 and Rac1 interactions with the RasGAP site share some similarities. These interactions mostly include switch I and switch II residues of Cdc42 and Rac1. Because these regions of Cdc42 and Rac1 are almost identical, these GTPases interact with the RasGAP site in a

similar manner. Inducing C-term folding but not release is the other similarity in the RasGAP mode Cdc42 binding and RasGAP mode Rac1 binding. The C-term release that is observed in the case of the Ex-mode Cdc42-GRD2 interaction cannot be observed in any of the RasGAP mode binding. This suggests that only Ex-mode Cdc42 binding induces dimerization by releasing the C-term. With these findings, the understanding of IQGAP dimerization is enhanced and an explanation is provided as to why Cdc42, but not Rac1, can induce the dimerization.

Thus, the mechanism of the interactions of Cdc42 and Rac1 with GRD2 can now be understood. Atomistic MD simulation data, combined with our collaborator's mutagenesis experiments, yield realistic models of Cdc42–GRD2 and Rac1–GRD2 interactions (Figure 4.16).

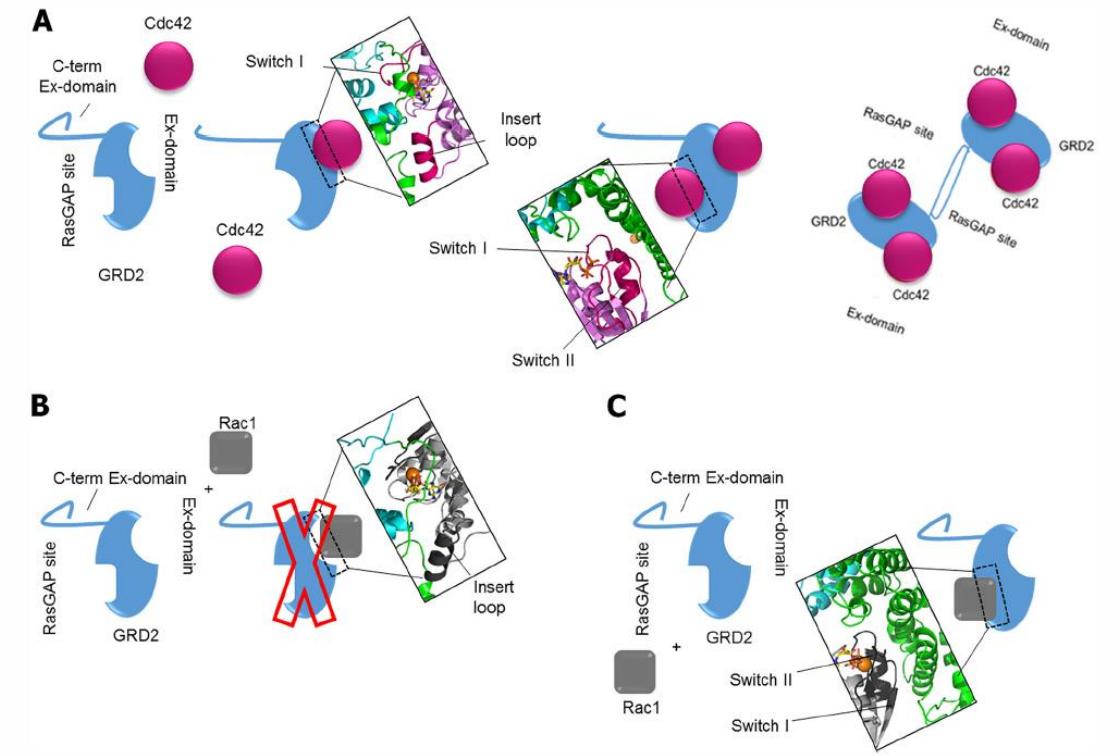


Figure 4.166. Interaction of GRD2 with Cdc42 and Rac1. A) *Schematic diagram* illustrating the release of C-term and allosteric change in RasGAP site upon Ex-mode Cdc42 binding. Allosteric change in the RasGAP site induces binding of the second Cdc42 to RasGAP site, and GRD2 with two Cdc42s can dimerize. **B)** Rac1 cannot bind to the Ex-domain and cannot induce allosteric change in the RasGAP site and release of

C-term. **C)** Rac1 can bind to the RasGAP site of apo-GRD2 but cannot induce the release of C-term. Therefore, Rac1 cannot facilitate GRD2 dimerization.

Cdc42 interaction with IQGAP is crucial because of their collective roles in cell motility [30, 31]. IQGAP binds and stabilizes Cdc42 in its GTP-bound active form, blocking GTP hydrolysis [49]. Little is known about the detailed mechanisms of Cdc42 and Rac1 in actin polymerization [64, 194], although progress is being made [195]. Their functions differ; Cdc42 enhances filopodia (containing parallel bundles of filamentous (F)-actin) formation, whereas Rac1 promotes lamellipodia [38, 196, 197]. There is evidence that IQGAP1 is overexpressed in cancer, that it affects actin polymerization [62], and that it promotes migration of tumor cells [40]. GTP-bound Cdc42 and IQGAP1 regulate filopodia formation. Cdc42 activates N-WASP by interacting with its Cdc42/Rac binding (CRIB) motif and releasing N-WASP's auto inhibited conformation; IQGAP1 C-terminal half activates N-WASP [63]. Studies showed that Cdc42 and IQGAP1 bind to N-WASP and activate it synergistically [64]. Activated N-WASP then recruits actin and Arp2/3, which nucleate filaments [30, 198]. Cdc42–IQGAP binding stabilizes its active form, which facilitates Cdc42-dependent N-WASP activation. Actin polymerization by Rac1 differs, with lamellipodia (mesh) formation induced by Rac1 interaction with WAVE, a WASP-family verprolin homologous protein, and is not facilitated by IQGAP1 [65]. Still, exactly how the different binding mechanisms of Cdc42 and Rac1 to IQGAP affect their actin related functions is unclear. One possibility explored here is through stabilizing and releasing active Cdc42. It is possible that the 2:1 stoichiometry of Cdc42 binding to IQGAP1 is more stable than the 1:1 stoichiometry for Rac1. Which factor induces Cdc42 release from IQGAP and the mechanism regulating IQGAP1 binding to Cdc42 and Rac1 are also still open questions.

To conclude, Rho GTPases, particularly Cdc42 and Rac1, are widely studied, because of their effects on the actin cytoskeleton and thus tumor cell metastasis. However, exactly how they influence actin polymerization has been enigmatic. The emergence of the insert loop distinguishing them from the Ras family was established as critical, but the mechanism, mediated by IQGAP scaffolding protein, has been obscure. This part of the thesis provides a deeper understanding at the conformational level. Having more

insight into the interaction mechanism of Cdc42 with IQGAP is likely to facilitate targeting these molecules for therapeutic purposes. The scaffolding protein IQGAP2 dynamics and allosteric response not only link key pathways; they can also control pathway cross-talk and signaling by fine-tuning responses, in this case to distinct members of the same family, with key consequences for actin polymerization.

Chapter 5

KRas4B RELEASE from PDEδ

5.1. Arl2-GDP and Arl2-GTP Binding to PDEδ

Atomistic MD simulations were performed on the Arl2-GDP– and Arl2-GTP–GRD2 complexes for 500 ns. The equilibrium conditions were tested checking RMSD and energy plots as a function of time (Figure B.3 and Figure B.4 in APPENDIX B).

PDEδ binds to Arl2 in a GTP-specific manner and inhibits the dissociation of GTP from Arl2. Arl2-mediated regulation of PDEδ-mediated transport of prenylated cargos is expected upon only Arl2-GTP binding. To compare the interaction strengths of the HVR and PDEδ in GTP- and GDP-bound Arl2, the binding free energies of PDEδ–HVR in systems 2-4 and systems 6-8 were calculated using the MM-GBSA method [192]. The calculated binding free energies of the PDEδ–HVR interaction are –69.9 ± 8.2 kcal/mol, –49.8 ± 6.0 kcal/mol, and –62.3 ± 8.6 kcal/mol for systems 2-4 with Arl2-GDP, respectively, and –46.5 ± 9.2 kcal/mol, –45.4 ± 8.3 kcal/mol, and –41.6 ± 5.0 kcal/mol for systems 6-8 with Arl2-GTP, respectively (Figure 5.1). PDEδ–HVR interaction in systems with Arl2-GTP (systems 6-8) is weaker than with Arl2-GDP (systems 2-4). This suggests that Arl2-GTP can mediate KRas4B release from PDEδ. The difference between the interaction strengths in system 3 and system 7 is not large compared to the other systems (system 2 vs. system 6 and system 4 vs. system 8). It was previously reported that the interaction of farnesylated HVR with PDEδ in state 1 causes some clashes [9]. PDEδ in state 2 is more favorable for farnesylated cargo binding and more stable interactions can be established between farnesylated proteins and PDEδ in state 2, making state 2 more challenging for Arl2-stimulated HVR release from PDEδ.

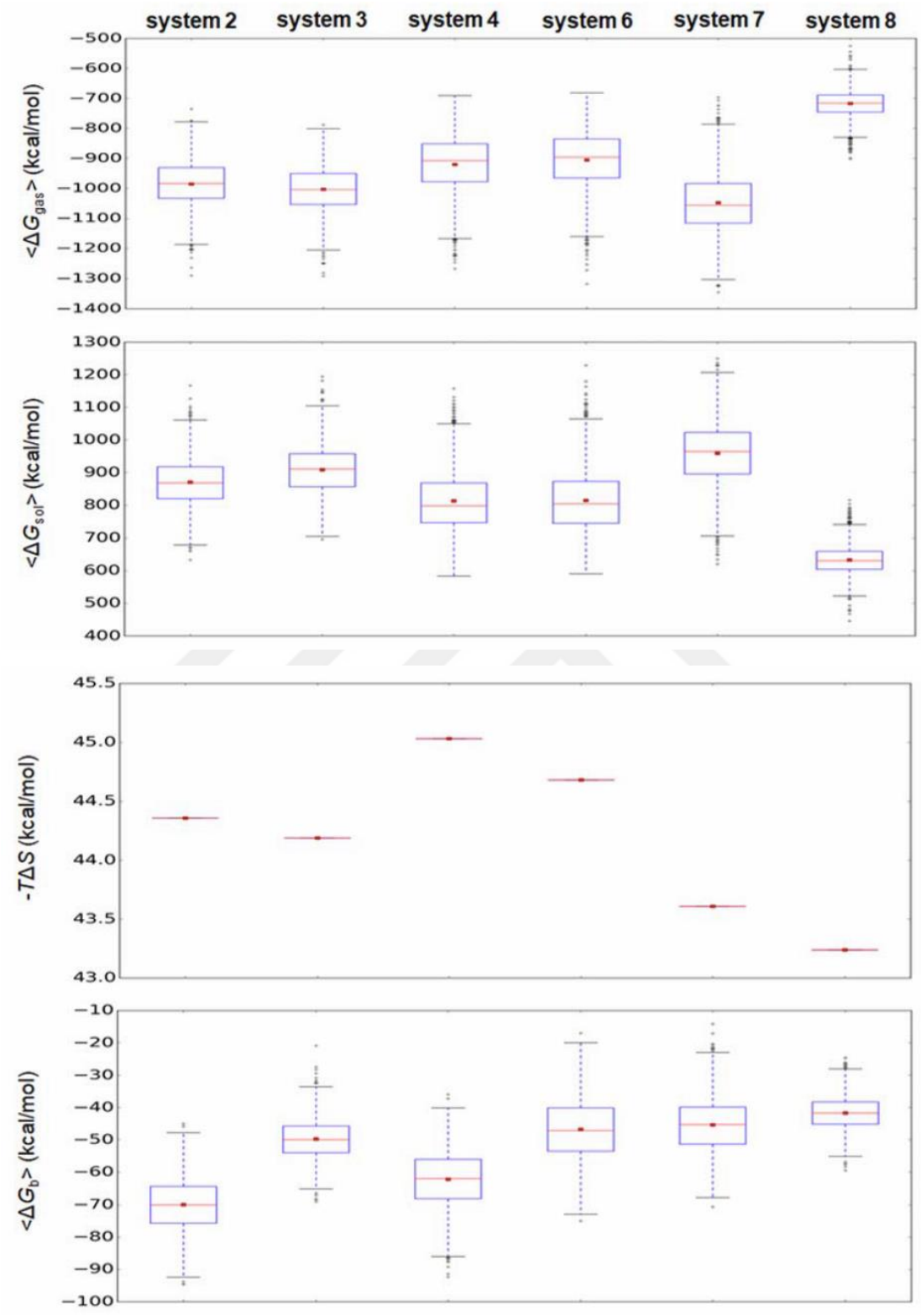


Figure 5.1. Binding free energies of PDEδ–HVR in systems 2-4 and systems 6-8 in kcal/mol calculated by MM-GBSA method. In the calculation, gas phase contribution, $\langle \Delta \boldsymbol{G_{gas}} \rangle$, the solvation energy contribution, $\langle \Delta \boldsymbol{G_{sol}} \rangle$, and the entropic contribution,

-$T\Delta S$, combines, the average binding free energy, $\langle \Delta \boldsymbol{G_b} \rangle$. In the box graphs, the red dot and red line denote the mean and median values, respectively.

Additional evidence that PDEδ–HVR interaction strength is reduced upon Arl2-GTP binding to PDEδ is the decrease in the number of stable interactions between PDEδ and the HVR. There are fewer stable salt bridges and hydrogen bonds (H-bonds) between PDEδ and HVR in systems 6 and 7 compared to systems 2 and 3, respectively (Figure 3). Previously, it has been reported that Glu88 of PDEδ plays a major role in KRas binding and inhibitors that can bind Glu88 impair oncogenic KRas-driven tumor growth [13]. In the systems 2 and 3, stable interactions between Glu88 and HVR residues are established; however, they are disrupted in systems 6 and 7 (Figure 5.2). On the other hand, in the geranylgeranylated systems (systems 4 and 8), even though the binding free energy calculations indicate that the PDEδ–geranylgeranylated HVR interaction is weakened by binding of Arl2-GTP, the number of stable interactions between PDEδ and HVR increases. This suggests that whereas farnesylated KRas4B release is mediated by Arl2, the mechanism of geranylgeranylated cargo release may differ. Since PDEδ binding to farnesylated HVR is more favorable than geranylgeranylated HVR, revealing the release mechanism of farnesylated KRas4B from PDEδ may provide insight into KRas4B trafficking to the PM.

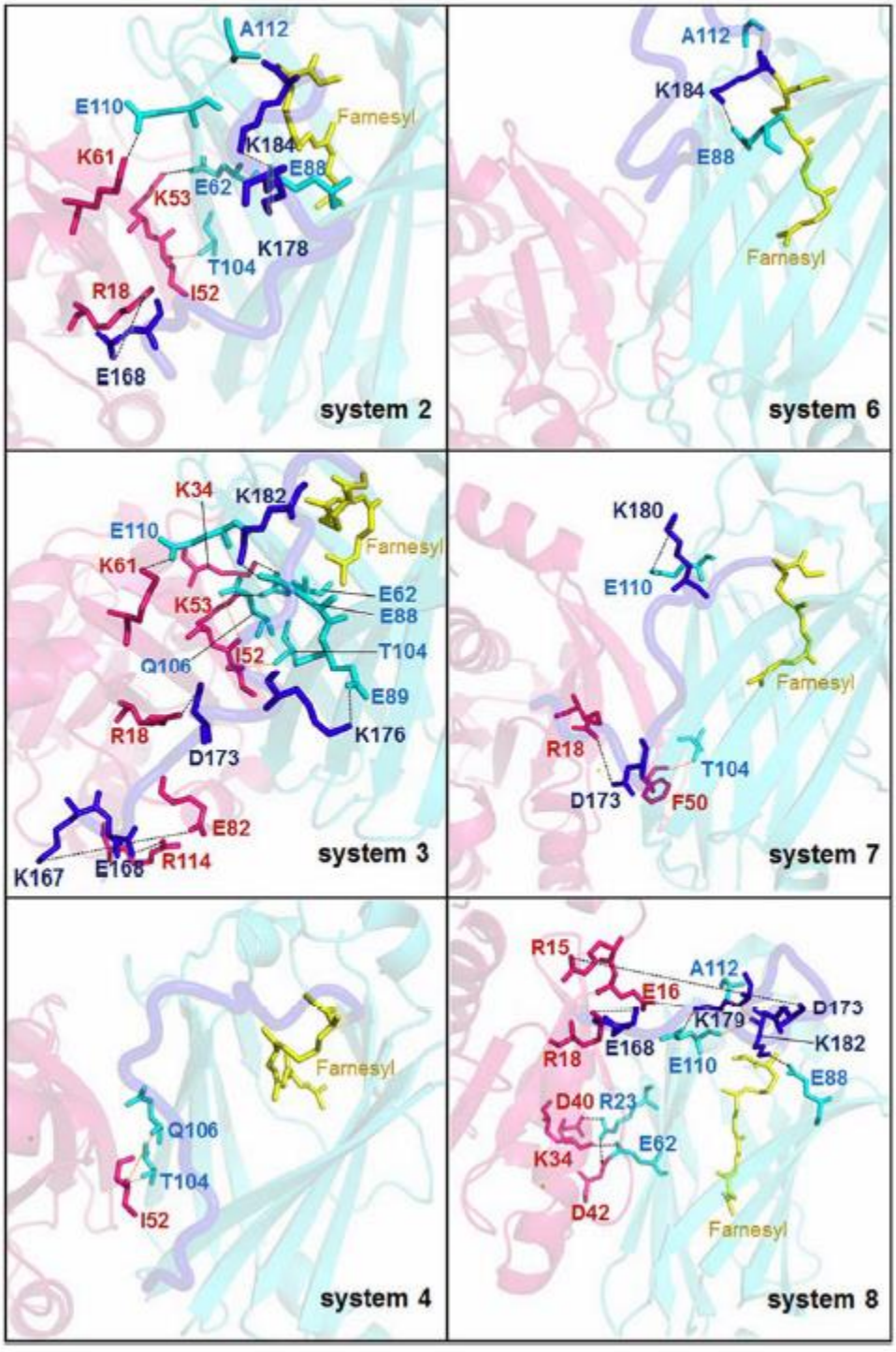


Figure 5.2. Interactions between Arl2, PDEδ, and HVR. Stable salt bridges and H-bond formed between Arl2, PDEδ, and HVR in systems 2-4 (left hand side) and systems 6-8 (right hand side) can be seen. Pink and cyan molecules/sticks show Arl2 and PDEδ molecules/residues, respectively, while dark blue molecules/sticks represent KRas4B

HVR molecules/residues. Black dashed lines show salt bridges, and orange dashed lines show H-bond.

5.2. Arl2-Mediated Slip out of KRas4B from PDEδ

KRas4B HVR should be released from the PDEδ pocket in order to bind the PM. To follow the release process, the initial and final positions of the farnesyl group inside the hydrophobic PDEδ pocket were analyzed. The distance between Cα atoms of Phe133 and the farnesyl/geranylgeranyl groups at 10 ns and 500 ns were measured. The distance is larger at 500 ns in each system except system 8 (Table 5.1 and Figure 5.3). The changes in the distance between systems 2-4 and systems 6-8 were compared. It was observed that upon Arl2-GTP binding to PDEδ, the farnesylated HVR slips out further compared to the Arl2-GDP (Table 5.1). In systems 6 and 7 with Arl2-GTP, the distance changes are 5.10 Å and 1.62 Å for PDEδ in states 1 and 2, respectively. However, the changes in the distance are smaller in systems 2 and 3 with Arl2-GDP, which are 3.86 Å and 0.64 Å for PDEδ in states 1 and 2, respectively. For PDEδ in state 1, both systems 2 and 6 with Arl2-GDP and Arl2-GTP, respectively, yield larger distance changes than the corresponding systems 3 and 7 with PDEδ in state 2. Since for PDEδ in state 2, Phe133 side chain is initially upward and already closer to the farnesyl groups in systems 3 and 7, smaller difference between Phe133 and farnesyl groups are expected. However, it was observed that downward Phe133 side chains in systems 2 and 6 flip upward during the simulations (Figure 5.3). Thus, even though Phe133 side chain moves closer to the farnesyl group in systems 2 and 6, the distance between Phe133 and farnesyl group at the end of the simulation is still larger. The movement of downward Phe133 to an upward position provides a repulsive force for HVR to be released from the PDEδ pocket. Also, the distance between Cα atoms of Phe133 and the farnesyl/geranylgeranyl groups throughout the simulations were measured (Figure B.5 in APPENDIX B). These graphs support that the observations from Figure 5.3 on the changes in the distance are valid not only for a single time frames, but during the simulations.

Table 5.1. The distances between Phe133 and CYF/CYG* at 10 ns, D_{10}, and at 500 ns, D_{500}, and the change in distance, $\Delta D = D_{500} - D_{10}$.

Systems	D_{10} (Å)	D_{500} ns (Å)	ΔD (Å)
system 2	17.00	20.86	3.86
system 3	20.31	20.95	0.64
system 4	18.99	20.69	1.70
system 6	15.51	20.61	5.10
system 7	19.23	20.85	1.62
system 8	20.77	20.37	-0.40

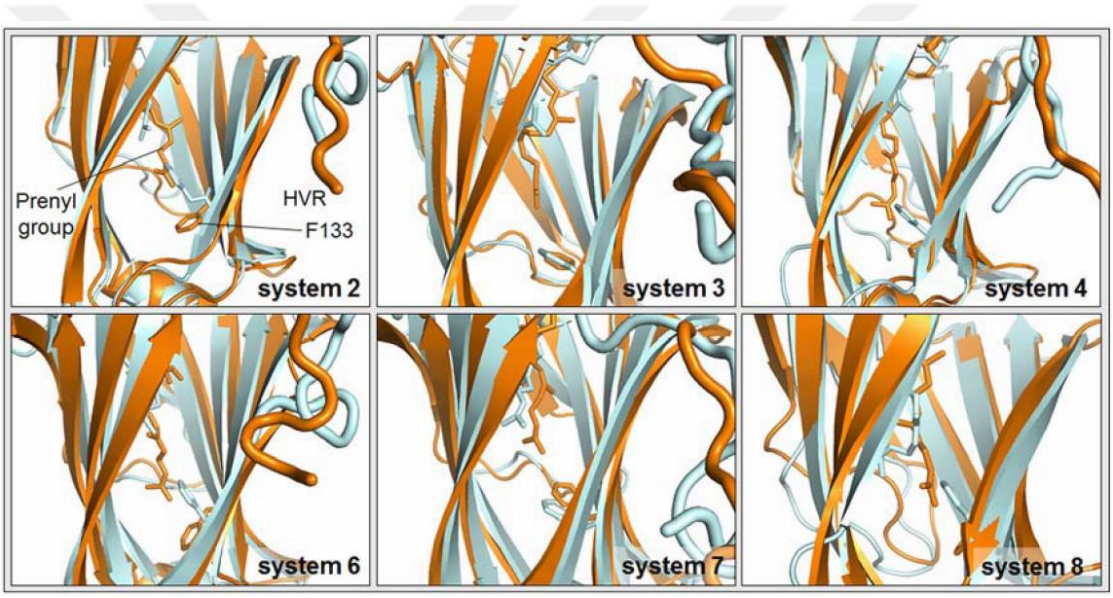


Figure 5.3. Initial and final positions of the KRas4B HVR. Initial and final positions of the farnesyl/geranylgeranyl group and Phe133 side chain in systems 2-4 and 6-8 (PDB ID: 5E8F/5F2U).

PDEδ residues Met20, Arg61, and Ile129 have been reported to move upon Arl2 binding, clashing with the farnesyl group inside of the pocket [199]. These changes facilitate farnesylated protein release. It was also observed conformational changes of these residues throughout the simulation. Residues which were initially positioned lower, shift upward and toward the inside of the pocket in the farnesylated systems (Figure 5.4). Like Phe133, these residues help HVR ejection from the hydrophobic pocket by compressing the pocket and pushing the HVR upward.

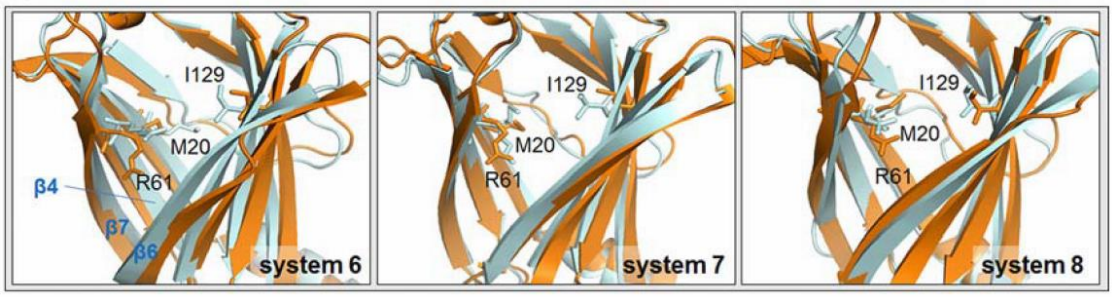


Figure 5.4. Initial and final positions of the some key residues. Initial and final positions of Met20, Arg61, and Ile129 of PDEδ in systems 6-8. Structures at 10 ns are represented by orange, while structures at 500 ns are shown by grey. HVR is removed to be able to clearly see the movement of the residues. β4/β7 and β6 interfaces are shown with blue labels (PDB ID: 5E8F/5F2U).

However, geranylgeranylated systems again show an opposite behavior. It has been shown that for system 8 with Arl2-GTP, the geranylgeranyl group of KRas4B HVR is intact in the hydrophobic PDEδ pocket (Table 5.1 and Figure 5.3). Also, the conformations of Met20, Arg61, and Ile129 do not change much compared to the change in farnesylated systems (Figure 5.4), in line with geranylgeranylated cargo release by a variant mechanism.

PDEδ open conformation is more favorable for the farnesylated cargo binding [199, 200]. The open conformation has a larger cavity volume compared to the closed PDEδ conformation. A closed conformation is induced by Arl2-GTP binding to PDEδ [199]. The conformational changes in PDEδ Met20, Arg61, Ile129, and Phe133 decrease the volume of the cavity and facilitate the shift to the closed conformation [199]. Arl2 interacts with β4 (residue 59-67) and mainly β7 (residues 101-109) at one side of the β-sandwich, and β6 (residues 88-97) at the other side [5], which is on the opposite PDEδ face (Figure 5.5). Formation of these interactions lead to movement of β4/β7 and β6 toward each other, which further promotes the closed conformation. During the simulations, it was also observed that as expected, new interactions are established between these β-strands and Arl2-GTP (Table 5.2). The β6 conformation changes, and the interaction between Glu88 on β6 and the HVR is disrupted upon Arl2-GTP binding

(Figure 5.2). Therefore, Arl2-GTP binding induces KRas4B release from PDEδ by changing PDEδ conformation to the unfavorable closed state. Because of this conformational change, the interactions between PDEδ and farnesylated HVR are also impaired, altogether resulting in the HVR slipping out from the pocket.

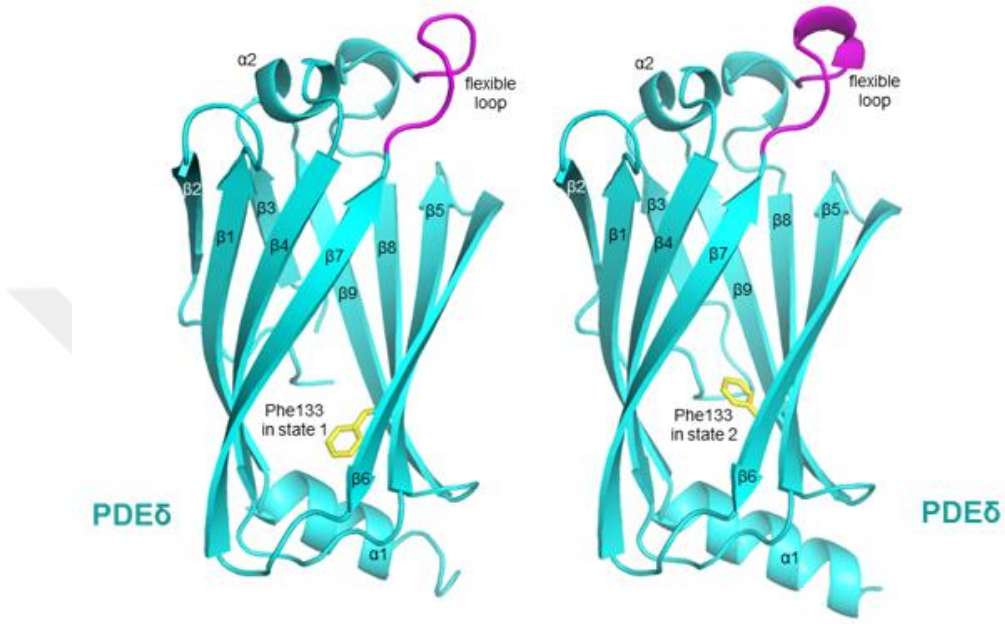


Figure 5.5. Structures and domains of PDEδ (PDB ID: 5E8F/5F2U).

Table 5.2. New interactions established between Arl2 and PDEδ upon Arl2-GTP binding.

Systems	Interactions*
system 5	Lys29-Asp42, Glu110-Lys13, Lys57-Glu16, Lys57-Glu56, Glu62**-Lys35, Glu88**-Lys9
system 7	Lys73-Glu82, Arg23-Asp42, Glu93**-Arg18
system 8	Lys57-Glu56, Glu62**-Lys34, Arg23-Asp40, Arg23-Asp42

*First residues in the residue pairs belong to PDEδ and the second ones belong to Arl2.
**PDEδ residues from β4, β7 or β6.

5.3. Allosteric Regulation of PDEδ by Arl2

Previous experimental studies showed that PDEδ binding sites to farnesylated cargos and Arl2 do not overlap [199, 200]. This suggests that Arl2 allosterically regulates PDEδ to promote the release of the farnesylated cargos. Phe94 and Ile98 of PDEδ are at the

Arl2–PDEδ interaction interface and undergo considerable conformational changes upon Arl2-GTP binding to PDEδ [199]. These residues are also identified as hot spots, which contribute to the binding free energy most [193], making them candidates for the interaction and allosteric regulation of PDEδ. Experimental studies showed that release of farnesylated Rheb from PDEδ is less efficient for PDEδ$^{F94A/I98A}$ double mutant than the wild type. In line with this, the ternary complex formed by Arl2–PDEδ–farnesylated cargos is more stable in the double mutant [199]. HVR position in systems 2 and 3 shows that when Arl2-GDP is bound, the HVR is folded, covering the β6 (residues 88-97) Arl2 interaction site of PDEδ, which involves Phe94 and Ile98. However, upon Arl2-GTP binding in systems 6 and 7, the HVR is released from β6 (Figure 5.6). Glu93 of PDEδ, which forms salt bridges with the HVR in systems with Arl2-GDP, establishes interactions with Arl2 and other PDEδ residues exposing the β6 interface in systems with Arl2-GTP. Differences in the HVR movement in the systems can be observed.

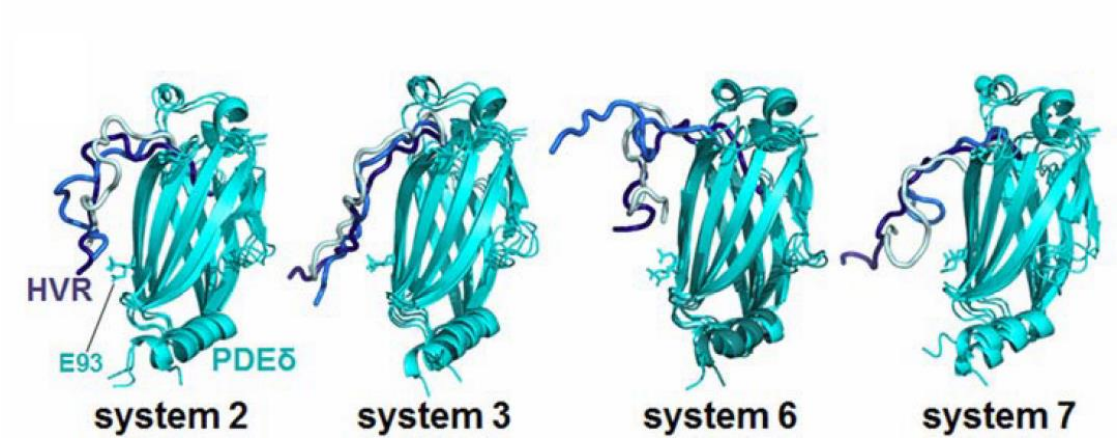


Figure 5.6. The position of KRas4B HVR in systems 2 and 3 with Arl2-GDP and in systems 6 and 7 with Arl2-GTP. Cyan molecules are superimposed PDEδs. Dark blue, blue, and light blue tubes represent the HVR conformations at 10 ns, 150 ns, and 500 ns, respectively (PDB ID: 5E8F/5F2U).

To reveal the allosteric mechanism of Arl2, a dynamical network analysis using WISP algorithm was conducted [176]. WISP can identify the shortest primary communicating pathway through nodes. Nodes refer to the protein residues of interest. 100 optimal paths between selected residue pairs on Arl2 and PDEδ were calculated. Residue pairs are

selected considering Arl2 residues interacting with Phe94$^{PDE\delta}$ and Ile98$^{PDE\delta}$, which are identified as Tyr76^{Arl2} and Tyr80^{Arl2}. PDEδ residue pairs are selected among the residues showing higher fluctuation profile upon Arl2-GTP binding (Figure 5.7), which comprise residues between 33-50 and 93-118. Residues 93-118 involve β6 and a flexible loop (residues 111-117), which are known to play a role in the farnesylated cargos interactions [199]. RMSF plots alone are not enough to determine allosteric change sites, since they show the average fluctuation of each residue through the simulations. Therefore, further analysis such as WISP and mutagenesis should be performed to validate allosteric changes. Among the selected residue pairs, only one pathway could be obtained in system 7 (Figure 5.7). This possible pathway obtained by WISP starts with Arl2 residue Tyr80 and ends in PDEδ residue Phe96. Frequently involved residues in the 100 different paths of this pathway are Gln70, Met71, Gly95, Val97, and Ile98 of PDEδ. Fluctuations observed in PDEδ residues 93-118 result from allosteric regulation induced by Arl2-GTP binding. Allosteric changes in β6 and the flexible loop, which are initiated by Tyr80^{Arl2} and involve Ile98$^{PDE\delta}$, facilitate KRas4B HVR release.

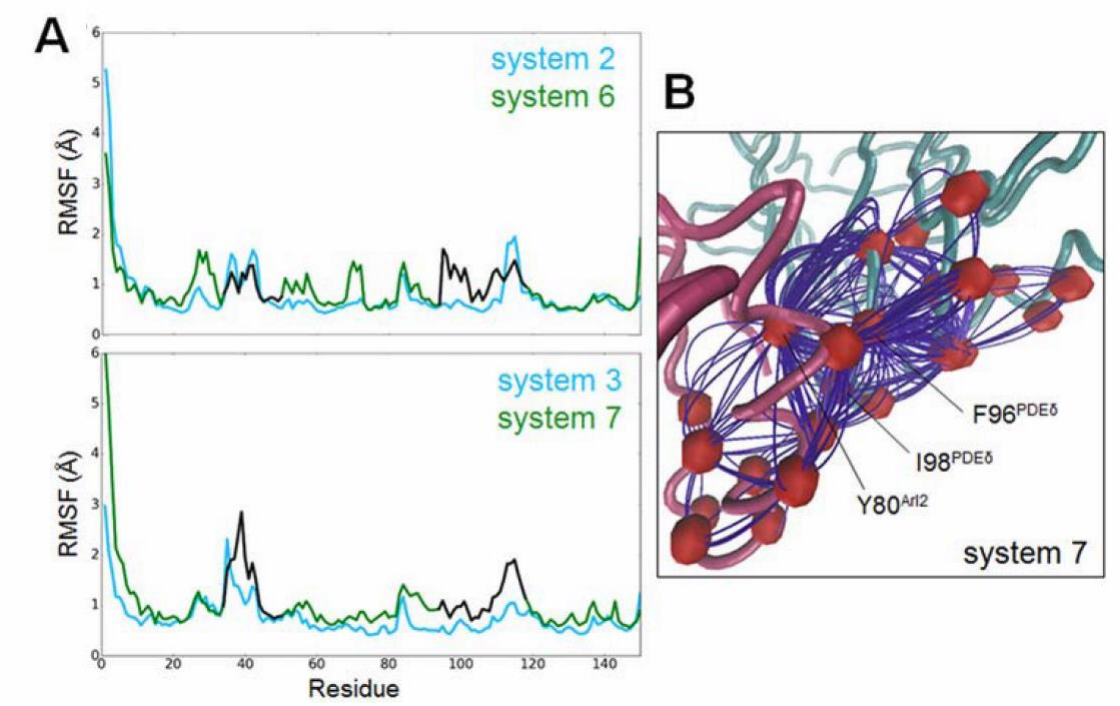


Figure 5.7. Allosteric changes in PDEδ induced by Arl2-GTP. A) RMSF plots of PDEδ in systems 2 and 6 (upper panel) and systems 3 and 7 (lower panel). Blue lines cover residues of systems 2 and 3, green lines cover residues of systems 6 and 7. The

residues between 33-50 and 93-118, which are used in WISP analysis, are represented with black lines in systems 6 and 7. **B)** Allosteric pathway between selected residue pair on Arl2 and PDEδ, $Tyr80^{Arl2}$/ $Ile98^{PDE\delta}$, in system 7 with Arl2-GTP.

To compare the KRas4B HVR release mechanism for the wild type and the double mutant structures, the systems 6 and 7 with double mutant $PDE\delta^{F94A/I98A}$ were simulated. It was observed that in mutant system 6 (systems 6^{m}), the HVR cannot slip out as much as in the wild type system 6 (Figure 5.8, the left side of the panel). The distance between Phe133 and the farnesyl group at 10 ns is 16.80 Å; at 500 ns it is 20.44 Å, a difference of 3.64 Å. This difference is smaller than between both system 2 (3.86 Å) and the wild type system 6 (5.10 Å) (Table 5.1). Therefore, in mutant system 6, the HVR is even more embedded in the hydrophobic pocket than in systems with Arl2-GDP. Also, the conformational changes observed in Met20, Arg61, and Ile129 in the wild type system 6 cannot be observed in mutant system 6 (Figure 5.8, the right side of the panel). Especially, Met20 and Ile129 are at almost the same position at the end of the 500 ns simulation as in the initial configuration.

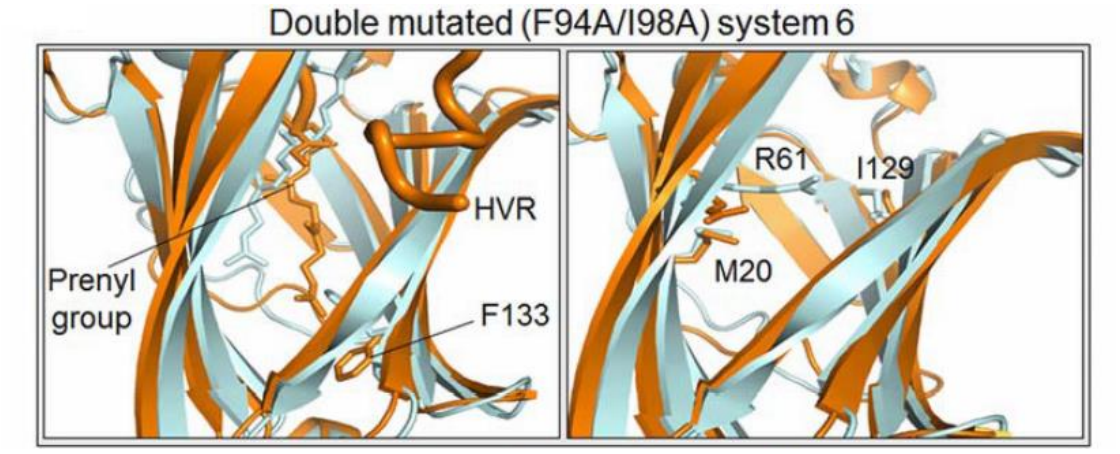


Figure 5.8. Effect of $PDE\delta^{F94A/I98A}$ double mutation on farnesylated KRas4B HVR release in systems 6^{m}. Initial and final positions of farnesyl group and Phe133 side chain (left panel), and some key residues (Met20, Arg61, and Ile129) of PDEδ (right panel) in system 6^{m}. Structures at 10 ns are represented by orange, while structures at 500 ns are shown by grey. HVR is removed in the right panels to be able to clearly see the movement of the residues (PDB ID: 5E8F).

The RMSF plots of system 2 and mutant system 6 were compared. In wild type system 6, some PDEδ residues (residues 33-50 and 93-118) fluctuate more than system 2, indicating allosteric changes in PDEδ conformation. However, mutant system 6 exhibits a fluctuation profile similar to system 2 (Figure 5.9), suggesting that allosteric regulation cannot occur. This further confirms that in mutant system 6, the HVR cannot slip out.

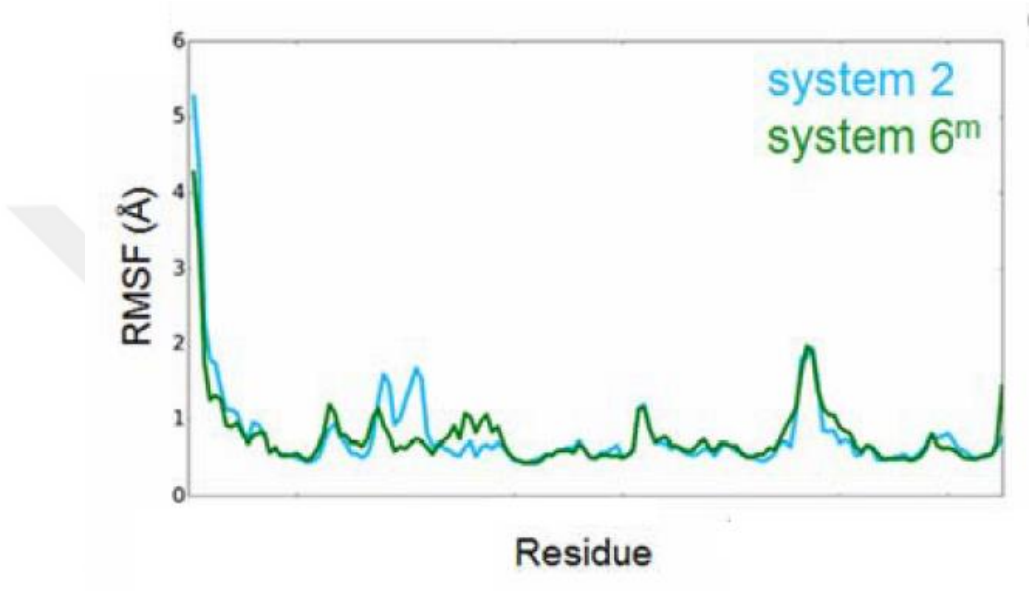


Figure 5.9. RMSF plots of PDEδ in systems 2 and system 6^m. Blue lines cover residues of systems 2, green lines cover residues of systems 6^m.

To compare the wild type and mutant system 7 (system 7^m), the HVR was monitored and observed that surprisingly, in system 7^m it can also be pushed upward (Figure 5.10, the left side of the panel). Met20 and Arg61 also undergo conformational changes and point upward, similar to wild type system 7 (Figure 5.10, the right side of the panel), however, Ile129 does not change considerably. The effect of mutations on system 7 is not as significant as that in system 6, possibly since in system 7 the double mutations also affect the favorable bonds between PDEδ and the farnesylated HVR. Thus, even in the presence of the mutations, HVR release from PDEδ is not affected in system 7 as much as in system 6.

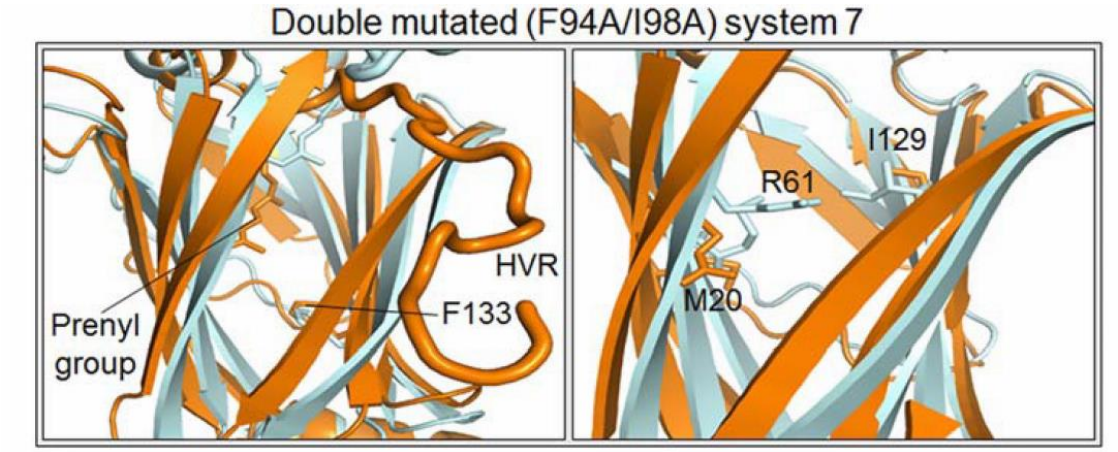


Figure 5.10. Effect of PDEδ$^{F94A/I98A}$ double mutation on farnesylated KRas4B HVR release in systems 7^{m}. Initial and final positions of farnesyl group and Phe133 side chain (left panel), and some key residues (Met20, Arg61, and Ile129) of PDEδ (right panel) in system 7^{m}. Structures at 10 ns are represented by orange, while structures at 500 ns are shown by grey. HVR is removed in the right panels to be able to clearly see the movement of the residues (PDB ID: 5F2U).

Fluctuation profiles of system 3 and system 7^{m} are similar (Figure 5.11); the fluctuations observed between residues 33-50 and 93-118 in the wild type system 7 cannot be seen in system 7^{m}. This indicates that the allosteric changes cannot be induced in system 7^{m}. The HVR bounce in system 7^{m} is not caused by allosteric changes, but by the disruption of favorable bonds between PDEδ and the HVR. Also, WISP analysis does not resulted in any allosteric paths unlike wild type system 7. Overall, allosteric changes initiated by Arl2-GTP binding and involving the Phe94, Ile98, Gln70, Met71, Gly95, and Val97 residues of PDEδ play an important role in farnesylated KRas4B release from PDEδ, and these allosteric changes are more efficient in the release of KRas4B HVR in system 6 compared to system 7.

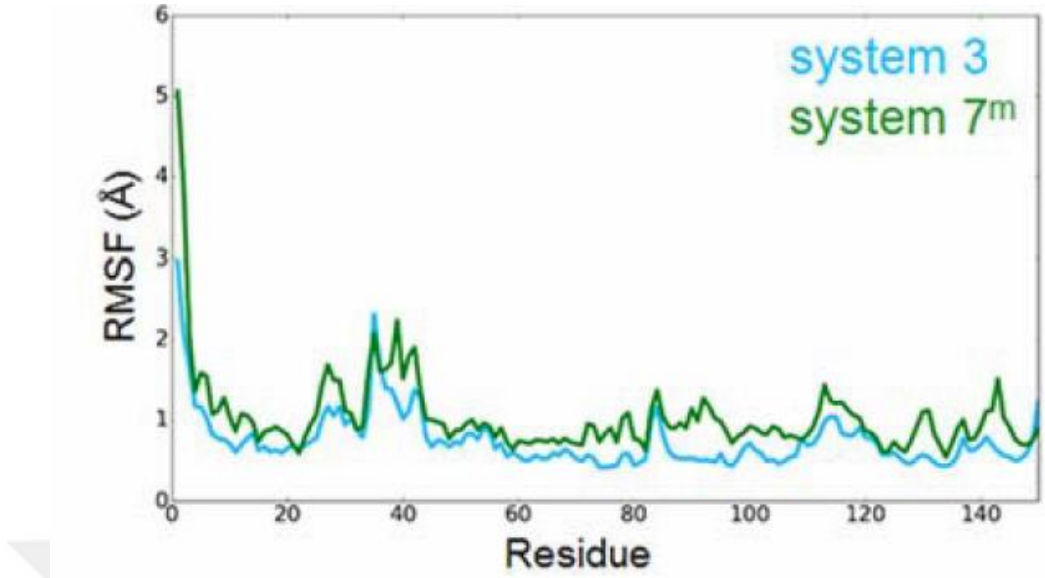


Figure 5.11. RMSF plots of PDEδ in systems 3 and system 7^{m}. Blue lines cover residues of systems 3, green lines cover residues of systems 7^{m}.

Furthermore, the number of stable interactions between PDEδ and HVR is higher in mutant systems 6 and 7 compared to wild type Arl2-GTP bound systems (Figure 5.12). In Figure 5.2, it was observed that Arl2-GDP bound systems have higher number of stable interactions, while Arl2-GTP binding induces dissociation of some interactions and leading fewer stable interactions. However, PDEδ$^{F94A/I98A}$ mutations, which inhibit allosteric changes upon Arl2-GTP binding, increase the number of stable interaction between PDEδ and HVR.

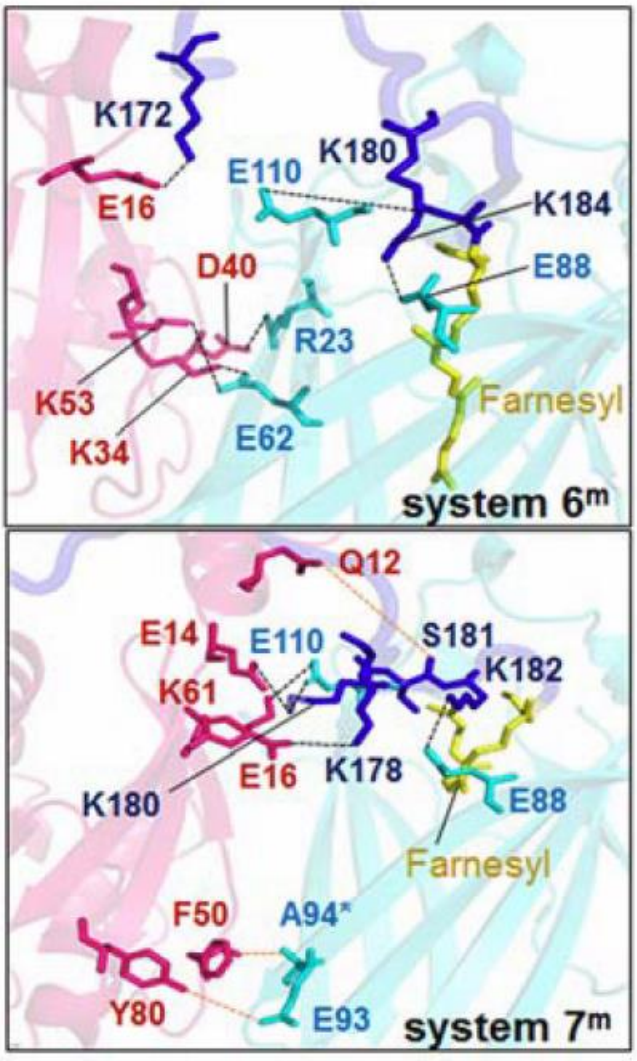


Figure 5.12. Interactions between Arl2, PDEδ, and HVR in systems 6^m and 7^m. Pink and cyan molecules/sticks show Arl2 and PDEδ molecules/residues, respectively, while dark blue molecules/sticks represent HVR molecules/residues. Black dashed lines show salt bridges, and orange dashed lines show H-bond.

5.4. Unstable Ternary Complex of Arl2-GTP, PDEδ and Farnesylated KRas4B

PDEδ binds to Arl2 in a GTP-dependent manner [5]. Arl2-GDP binding to free PDEδ is much weaker than Arl2-GTP binding to free PDEδ (Figure 5.13). The calculated binding free energies are 7.7 ± 5.9 kcal/mol for system 1 and -19.8 ± 12.4 kcal/mol for system 5. The steady energy profile of system 5 shows that the Arl2-GTP–PDEδ complex is stable throughout the simulation.

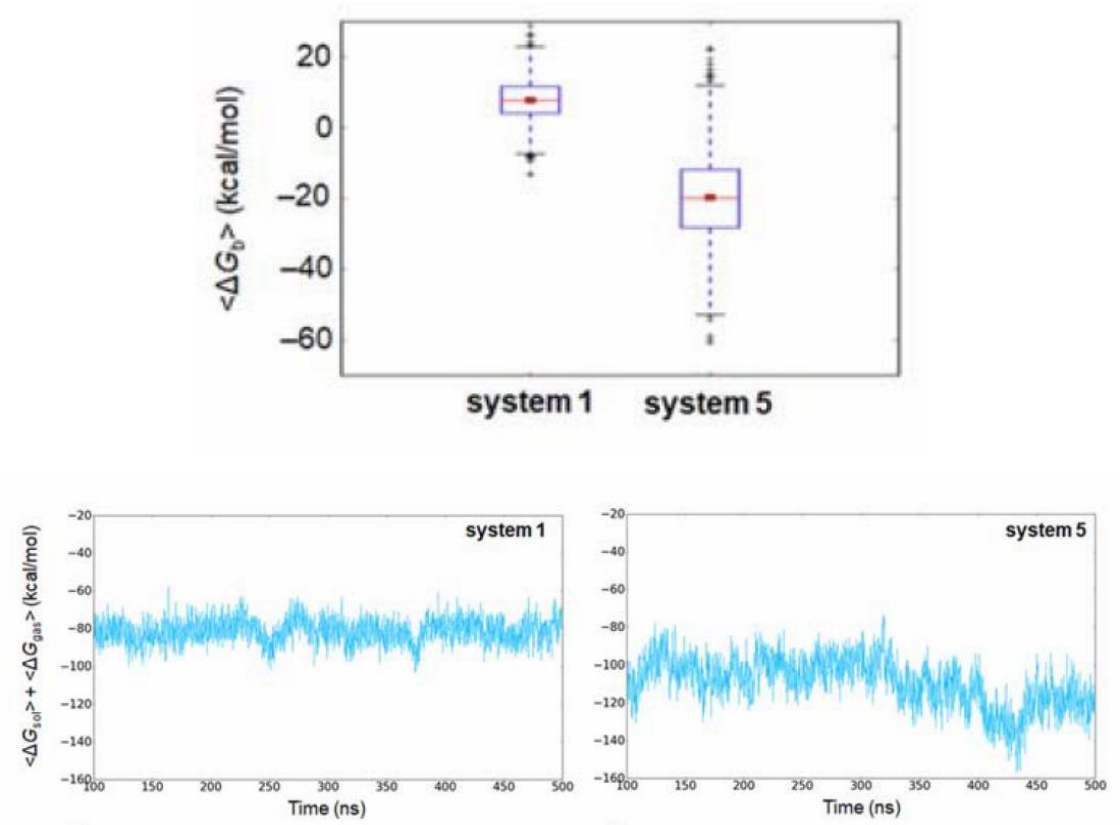


Figure 5.13. Binding energy profiles of Arl2–PDEδ complexes. Binding free energies of Arl2–PDEδ in systems 1 and 5 in kcal/mol calculated by MM-GBSA method (upper), energy profiles of Arl2–PDEδ over time (ns) in systems 1 and 5 in kcal/mol (lower). Only gas phase and solvation energy contributions are considered, and entropic contribution is excluded in the energy profiles.

The binding energy profile of Arl2 to PDEδ drastically changes in the presence of the farnesylated KRas4B HVR in the ternary complex. In Figure 5.14, the lower panels present step-like binding energy profiles between Arl2-GTP and PDEδ in systems 6 and 7. However, the energy profiles of systems 2 and 3 are smooth (Figure 5.14, upper panel), as in the case of the Arl2–free PDEδ interaction. This indicates that the interaction of Arl2-GTP with PDEδ is unsteady and hardly reaches equilibrium after binding of farnesylated KRas4B to PDEδ. Experimental studies showed that the binding sites for Arl2 and farnesylated cargos do not overlap and ternary complex formation is structurally feasible. However, it is unstable and mutually exclusive [199, 200]. Allosteric displacement of farnesylated cargo from PDEδ by Arl2 is mediated by the

intermediate ternary complex [199]. The Arl2–PDEδ–farnesylated cargo ternary complex dissociates into Arl2–PDEδ over time and the stability of Arl2–PDEδ interaction gradually increases (Figure 5.14, lower panel). This indicates that the Arl2–PDEδ complex is not steady and stable in the presence of the farnesylated KRas4B; however, as the release of KRas4B proceeds, the Arl2–PDEδ interaction is stabilized. In line with the experimental studies, it was observed that upon binding of Arl2-GTP and farnesylated HVR to their allosterically coupled PDEδ binding sites, a fast-dissociating ternary complex is formed. The previous observations showed that HVR affinity to PDEδ decreases and the HVR moves toward the outside of the pocket. Figure 5.14 shows that Arl2–PDEδ interaction strength increases during the HVR release.

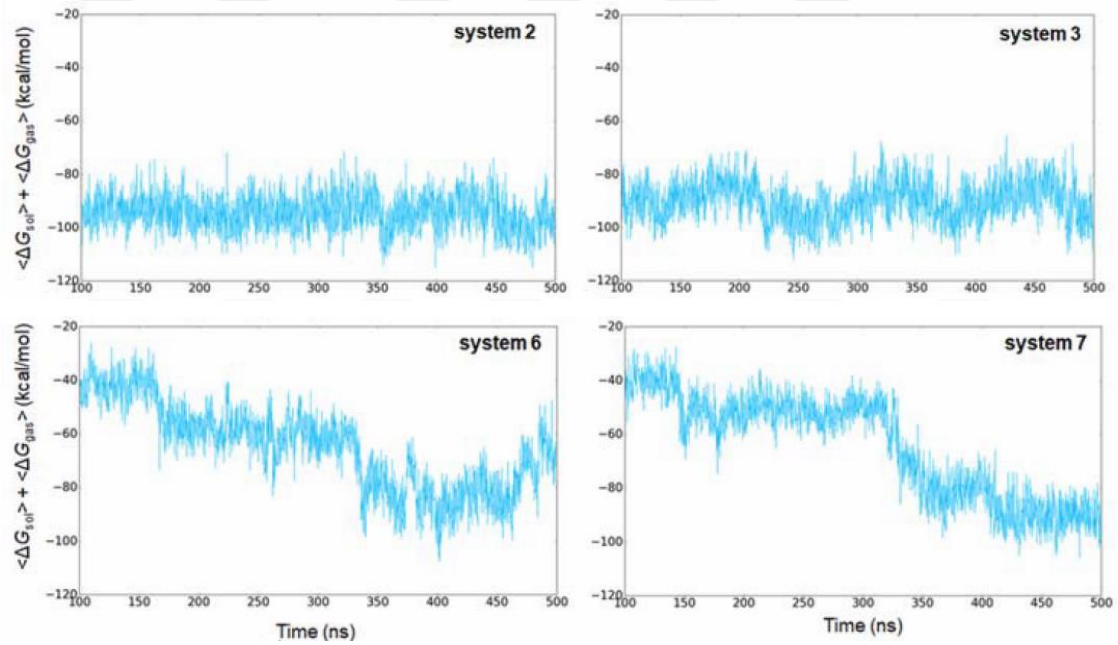


Figure 5.14. Binding energy profiles of Arl2–PDEδ in systems with farnesylated HVR complexes. Energy profiles of Arl2–PDEδ over time (ns) in systems 2 and 3 (upper panel) and in systems 6 and 7 (lower panel) in kcal/mol. Only gas phase and solvation energy contributions are considered, and entropic contribution is excluded in the energy profiles.

The energy profiles of systems 6 and 7 have been changed by PDEδ$^{F94A/I98A}$ double mutations. Instead of a step-like profile between Arl2-GTP and PDEδ in systems 6 and 7 (Figure 5.14, the lower panel), a smooth profile was observed in the corresponding

mutant systems (systems 6^m and 7^m) (Figure 5.15), similar to systems 2 and 3. Rather than fast-dissociating, the mutations stabilize the ternary complexes. The allosteric changes are necessary for the formation of the fast-dissociating intermediate complex.

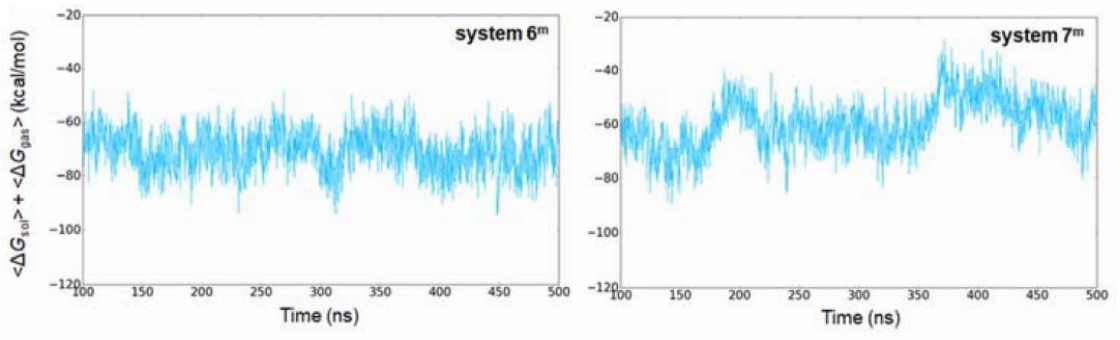


Figure 5.15. Binding energy profiles of Arl2–PDEδ in systems with mutations. Energy profiles of Arl2–PDEδ over time (ns) in the mutant systems (systems 6^m and 7^m) in kcal/mol. Only gas phase and solvation energy contributions are considered, and entropic contribution is excluded in the energy profiles.

In geranylgeranylated systems 4 and 8, the energy profiles of Arl2–PDEδ interaction are smooth and stable (Figure 5.16), supporting KRas4B geranylgeranylation regulation differing from the farnesylated Ras case. The Arl2–PDEδ interaction is not affected by geranylgeranylated KRas4B binding, in contrast to farnesylated KRas4B. The geranylgeranylated HVR did not slip out from the pocket as much as the farnesylated HVR did, and forms even more interactions with PDEδ upon Arl2-GTP binding, suggesting that its release is not coupled with Arl2. The ternary complex of Arl2–PDEδ–geranylgeranylated KRas4B is not fast-dissociating, but steady.

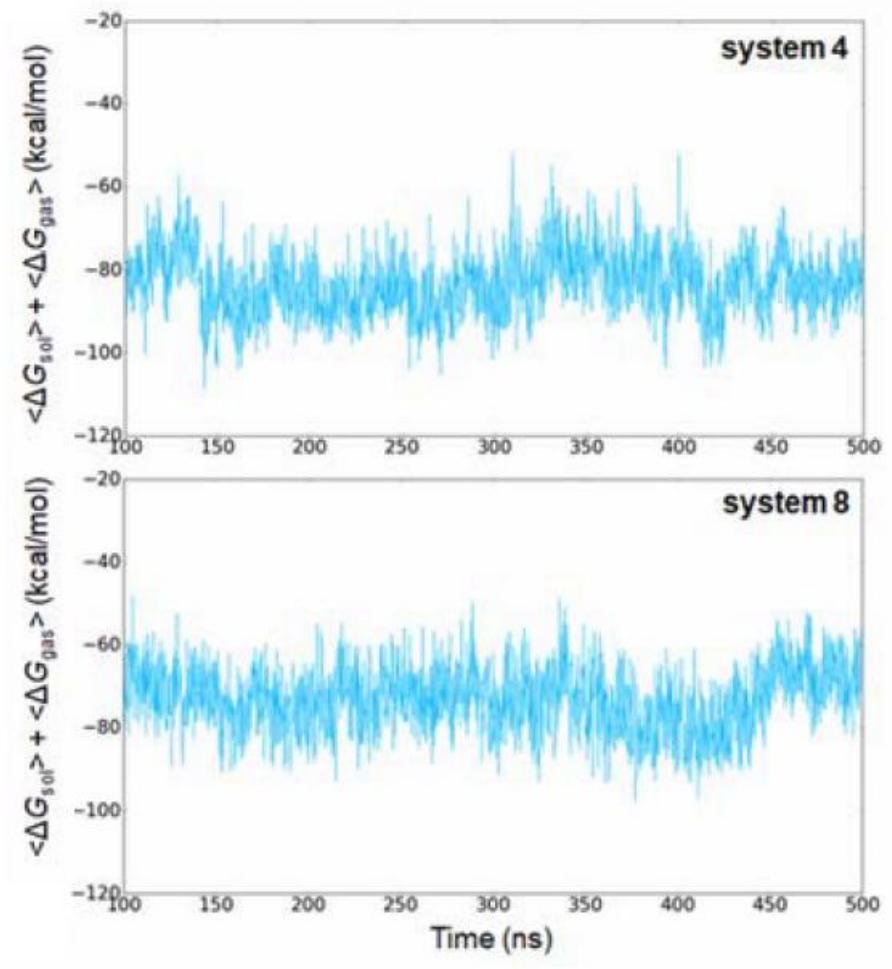


Figure 5.16. Binding energy profiles of Arl2–PDEδ in systems with geranylgeranylated HVR complexes. Energy profiles of Arl2–PDEδ over time (ns) in system 4 (upper panel) and in system 8 (lower panel) in kcal/mol. Only gas phase and solvation energy contributions are considered, and entropic contribution is excluded in the energy profiles.

5.5. Discussion

Subcellular localization critically affects prenylated Ras signaling capacity and activation [201, 202]. PM localization has been shown as increasing Ras activity through nanoclustering or dimerization [202-205] and playing a role in signaling selectivity of Ras isoforms [162, 201, 206]. Ras locates at the PM in the vicinity of receptors, effectors and nucleotide exchange factors, which facilitate its signaling [201, 205, 207]. Ras localization requires farnesylation-dependent interaction with PDEδ. PDEδ enriches its

localization at the PM and promotes its activity [124, 125]. Inhibiting Ras–PDEδ interaction may target oncogenic Ras [208].

In this part of the thesis, a possible mechanism for the release of KRas4B from PDEδ by Arl2 was described in atomistic detail. It was observed that binding of Arl2-GTP to PDEδ decreases KRas4B affinity to PDEδ, supporting KRas4B release being mediated by Arl2 in a GTP-dependent manner. The number of stable salt bridges and H-bonds between PDEδ and farnesylated KRas4B HVR decreases, and the farnesylated KRas4B HVR slips out from the hydrophobic PDEδ pocket upon Arl2-GTP binding. The detailed atomistic level analysis showed that some of the critical interactions between PDEδ and KRas4B HVR are disrupted by Arl2-GTP–PDEδ interaction. PDEδ Glu88 interaction with the HVR is abolished by Arl2-GTP binding. Disruption of Glu88-involved interactions can affect considerably the binding affinity of KRas4B to PDEδ, in line with this residue being a main drug target [13]. Disruption of these critical interactions is mostly caused by shifting of PDEδ β4/β7 and β6 [199]. The shift of these regions toward each other decreases the hydrophobic cavity volume. As have been shown here, the shift results in dissociating the critical interactions between β6 (involving Glu88) and KRas4B. It has been also showed that upon farnesylated KRas4B-bound systems interacting with Arl2-GTP, additional residues, Met20, Arg61, Ile129, and Phe133 of PDEδ shift toward the inside of the pocket and point upward. These changes facilitate the closed PDEδ conformation, which is not favorable for KRas4B binding. Moreover, once these residues point upward, the farnesylated KRas4B HVR is pushed outside of the PDEδ pocket. Together, these findings suggest that Arl2-GTP facilitates the release of farnesylated KRas4B HVR from PDEδ by decreasing the KRas4B HVR affinity to PDEδ.

Despite having distinct binding sites, binding of Arl2 and farnesylated cargos to PDEδ is mutually exclusive [199]. This, together with substantial conformational changes discussed above, imply that Arl2 binding allosterically regulates KRas4B release. The allosteric pathways initiated by Arl2-GTP residues and mostly comprised of β6 residues of PDEδ were identified. In Arl2-GDP bound systems, β6 residues are covered by the HVR; however, upon Arl2-GTP binding, the HVR swings and the β6 site is exposed.

Consequently, β6 residues fluctuate more upon Arl2-GTP binding, which also indicates the conformational changes. Among the β6 residues, PDEδ Phe94 and Ile98 were experimentally reported as playing a role in allosteric regulation inducing farnesylated Rheb release from PDEδ [199]. Ismail *et al*. experimentally constructed a PDEδ$^{F94A/I98A}$ double mutant and compared its affinity to Arl2 and Rheb with wild type PDEδ.[28] They concluded that after addition of Arl2, dissociation of Rheb from PDEδ$^{F94A/I98A}$ is not as efficient as in the wild type case. Based on this experimental study, it has been tried to validate whether these PDEδ residues are also crucial in KRas4B HVR release. It has been further confirmed the effect of these residues on allosteric regulation and KRas4B HVR release by analyzing double mutant (PDEδ$^{F94A/I98A}$) farnesylated systems. The allosteric changes observed in wild type farnesylated systems cannot be detected in the mutant systems and the HVR cannot slip out from the hydrophobic pocket in some systems. In some cases, allosteric changes can generate new binding sites and increase the number of binding partners [209], however in others, one of the binding partners can be released, as in the case of the Arl2–PDEδ–KRas4B HVR ternary complex.

Experimental studies demonstrated that Arl2-GTP binding induces the formation of a closed PDEδ conformation and farnesylated Rheb release from PDEδ by generating the fast-dissociating ternary complex [199, 200]. Also, double mutation (PDEδ$^{F94A/I98A}$) decreases the release efficiency of Rheb and increases the stability of the ternary complex [199]. The binding energy profiles of Arl2–free PDEδ and Arl2-GDP–PDEδ–KRas4B HVR systems were monitored. Both are steady throughout the simulations. However, the Arl2-GTP–PDEδ interaction is unsteady in farnesylated KRas4B HVR systems, but is gradually stabilized. This shows that formation of the fast-dissociating intermediate ternary complex is required for the release of the farnesylated KRas4B HVR; then, as release occurs, the Arl2–PDEδ system is equilibrated. These observations parallel experimental results suggesting formation of a fast-dissociating ternary complex required for farnesylated cargo release. PDEδ$^{F94A/I98A}$ mutant systems present similar behavior to Arl2-GDP bound systems. This indicates that without allosteric changes the fast-dissociating intermediate ternary complex cannot be formed. The ternary complex is stabilized, and the farnesylated KRas4B release is inhibited.

With respect to geranylgeranylated cargo shuttling and release, the analysis showed that KRas4B geranylgeranylated systems behave in an opposite manner. HVR dislocation, the stable interactions, and the binding energy profile of Arl2–PDEδ in geranylgeranylated systems all differ, with PDEδ farnesylated cargos being more favorable. Experimentally, it has been shown that PDEδ can interact with both farnesyl and geranylgeranyl groups [6]. However, the binding affinity between PDEδ and the geranylgeranyl group is much weaker [6, 121] implying those differences in the binding and release mechanisms. Releasing the long geranylgeranyl group from PDEδ is more likely to require more drastic conformational changes. Notably, in Rap1 members of the Ras subfamily of small GTPases, prenylation was recently observed to be required for high-affinity interactions with CAP1 in a geranylgeranyl-specific manner [210].

Thus, the mechanism of the Arl2-mediated release of farnesylated KRas4B from PDEδ can now be understood. Here, it has been revealed that Arl2-GTP binding to β4/β7 and β6 of PDEδ shifts these regions toward each other. This shifting decreases the PDEδ cavity volume and compresses the cavity, coupled with conformationally changes in the β6 residues. These changes disrupt some stable bonds between PDEδ and HVR, allosterically regulating HVR release. The compression of the cavity, which is unfavorable for farnesylated cargo binding, and allosteric changes in PDEδ residues push the HVR upward, toward the outside of the cavity. The HVR then dissociates from the fast-dissociating ternary complex and the Arl2–PDEδ complex is stabilized (Figure 5.17).

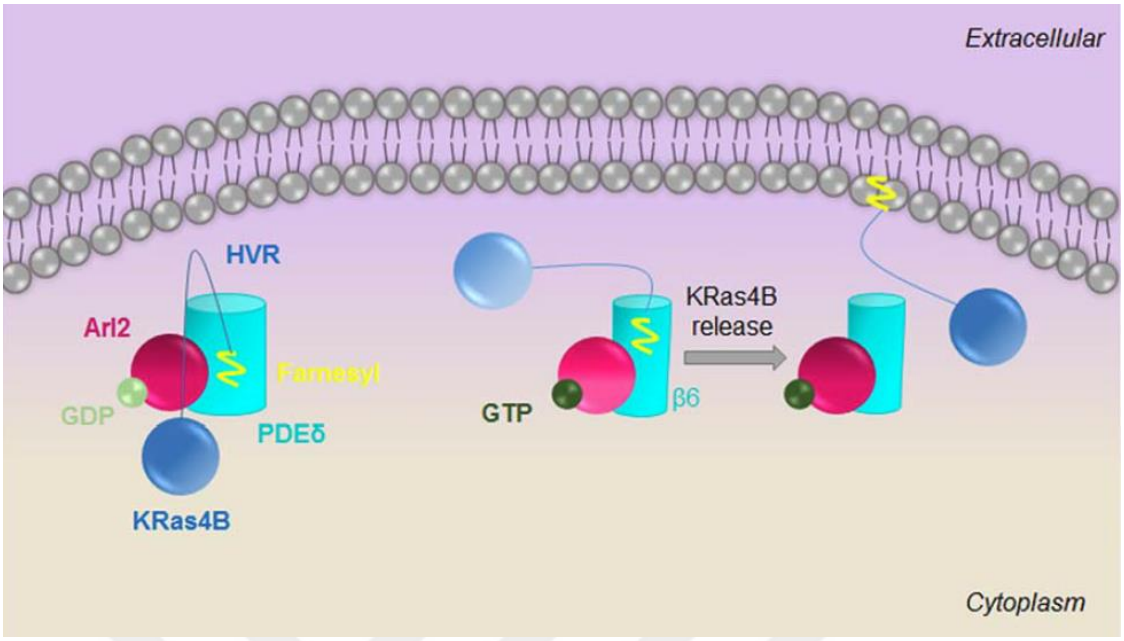


Figure 5.17. The release mechanism of KRas4B HVR from PDEδ. When Arl2-GDP is bound to PDEδ, the HVR is embedded in the hydrophobic PDEδ pocket and covers part of the Arl2 binding site preventing allosteric changes. However, upon Arl2-GTP binding, the allosteric site is revealed. PDEδ undergoes allosteric and conformational changes resulting in compressing the hydrophobic pocket and pushing the HVR outside. While HVR is bound, a fast-dissociation ternary complex is formed. Following HVR release, the Arl2–PDEδ complex is stabilized. In the middle panel, Arl2 and HVR are shaded with lighter colors to indicate the unsteady interactions.

To conclude, highly oncogenic Ras GTPases, particularly KRas, have drawn the attention of the community. However, their intracellular trafficking mechanisms have been enigmatic. The role of PDEδ in the farnesylated cargo shuttling was established as essential, but the exact release mechanism from PDEδ has been obscure. In this part of the thesis, this mechanism was studied at the conformational level. Disruption of membrane localization of KRas4B might be one therapeutic approach to target KRas-driven tumor growth. Unraveling the Arl2-mediated release mechanism of farnesylated KRas4B from PDEδ may facilitate therapeutic targeting. Targeting Arl2 interaction interface of PDEδ helps to impair PDEδ–Arl2 interaction and therapeutically aiming at allosteric sites of PDEδ may block formation of its closed form. These approaches might

be promising for drug design to prevent KRas4B release from PDEδ and inhibit proper membrane localization.

Chapter 6

SINGLE AMINO ACID VARIATIONS in INTERFACES

In this part of the thesis, a distinct study compared to the previous two chapters will be presented. The chapter covers statistical and structural analyses of distribution of literature curated experimental SAV data within interface residues. At the end of the chapter, an analysis specific to Ras GTPase proteins will be presented.

Distribution of SAVs effecting protein interaction stability and protein function within singlet hot spot, non-hot spot and hot region residues is analyzed. Figurer 6.1 explains the differences between hot spot, singlet hot spot, hot region and non-hot spot residues of a protein. Hot regions are hot spot clusters which meet the conditions described in Figure 6.1. Singlet hot spots are hot spot residues which are not part of any hot region. Hot spots include both singlet hot spots and hot regions. Residues which are not hot spots are classified as non-hot spots.

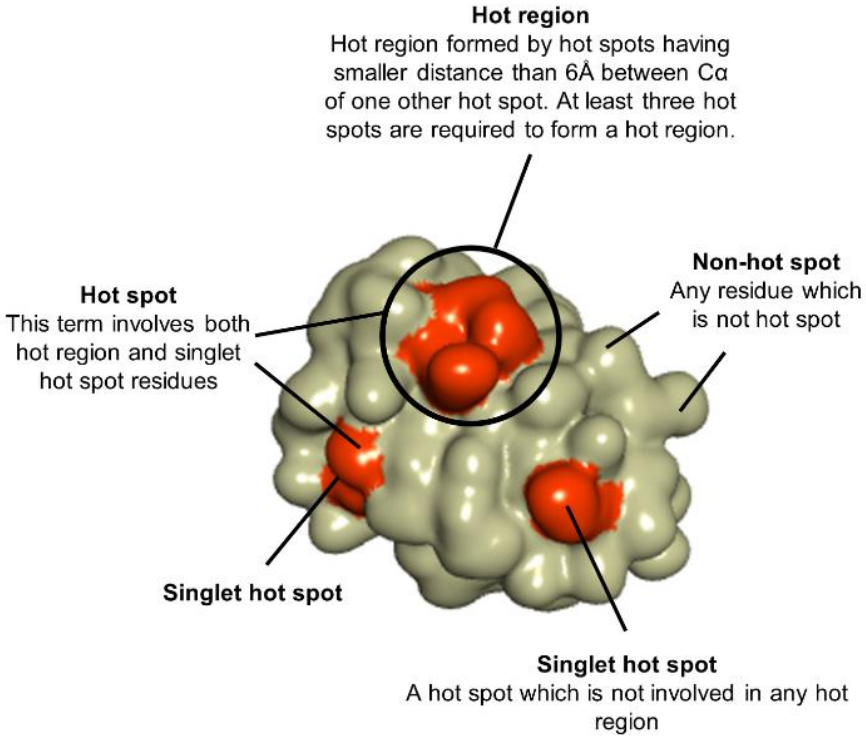


Figure 6.1. Schema of hot spots, hot region, singlet hot spots and non-hot spot residues Red residues are hot spots, however the residues inside the black circle are clustering and forming hot region. Other hot spot residues which cannot form hot region are singlet hot spots. Grey residues are non-hot spot residues.

The categories of SAVs which are used in the distribution analyses are summarized in Table 6.1. SAVs were classified in terms of their location, their effect on binding free energy and their disease-association.

Table 6.1. Classification of SAVs in interface residues

Classification	**Description**
Location	
Singlet hot spot SAVs	SAVs located in singlet hot spot residues
Hot region SAVs	SAVs located in hot regions
Non-hot spot SAVs	SAVs located in residues which are not hot spots.
Effect on binding energy	
Destabilizing SAVs	SAVs with $\Delta\Delta G$ values larger than 0.5 kcal/mol
Stabilizing SAVs	SAVs with $\Delta\Delta G$ values smaller than -0.5 kcal/mol
Neutral SAVs	SAVs having $\Delta\Delta G$ between -0.5 kcal/mol and 0.5 kcal/mol
Disease-association	
Benign SAVs	Human SAVs predicted as benign based on HumDiv model of PolyPhen2
Disease-causing SAVs	Human SAVs predicted as probably damaging (more confident prediction) and possibly damaging (less confident prediction) based on HumDiv model of PolyPhen2

6.1. Binding Free Energy Change Distribution of SAV within Interface Residues

Since SKEMPI 2.0 data set consists of mutations mostly found in interface residues [177, 211], the analysis is focused on the contacting residues of interfaces. Initially 4441 SAVs (with 5079 different ΔG entries) from 320 different PDB structures are collected from the SKEMPI 2.0 database and are mapped onto Piface interfaces [178]. It is

observed that 1643 out of 4441 SAVs correspond to contacting residues of 172 PDB structures consisting of 234 different chains. The number of SAVs in these 234 chains is varied between 1 and 114 with a median of 5. The maximum number of 114 SAVs comes from one chain (PDB ID: 3SGB, chain I) whose several residues were mutated to all possible other amino acids. In these chains, there are 679 hot spot SAVs and 356 of them are in hot regions (Table 6.2).

Table 6.2. General statistics on the data and number of SAVs, hot spots and hot spot SAVs on PPI interfaces

Single SAVs from SKEMPI	4441	Contacting residues in Piface interfaces	3612*
PDB structures / chains**	172/234	SAVs in contacting residues	1643
Hot spots in interfaces	1294	Hot spot SAVs	679
Hot regions clusters / hot region residues in interfaces	151/855	Hot region SAVs	356
Singlet hot spots in interfaces	439	Singlet hot spot SAVs	323

*This number obtained by counting interface residues in 234 chains of 172 structures which have at least one SAV in ** with SAV (s) in their interfaces

First, destabilizing, stabilizing and neutral SAVs are mapped to hot regions, non-hot spots and singlet hot spots of protein complexes to find the number of SAVs in each of these locations. The aim is to see whether there is a relationship between SAV localization and binding free energy change. Therefore, a contingency table was constructed (Table 6.3) and test the dependence of the change in binding free energy on SAV distribution within hot regions, non-hot spots and singlet hot spots with chi square (χ^2) test. The result of the χ^2 test indicates that the binding free energy changes depend on the distribution of SAVs within these regions (p-value < 2.2e-16).

Table 6.3. χ^2 test for testing dependence of binding free energy change distribution

	Hot Region	**Singlet Hot Spot**	**Non-hot Spot**	**p-value**
Destabilizing SAVs	273	291	542	< 2.2e-16
Stabilizing SAVs	30	15	86	
Neutral SAVs	53	17	336	

Therefore, in order to reveal the nature of this dependence, the correspondence of destabilizing, stabilizing and neutral SAVs to hot regions, non-hot spots and singlet hot spots is further analyzed with odds ratio (OR) calculations. Hot region is formed if at least three hot spots tightly cluster spatially together, an interface might contain one or a few hot regions. Some hot spot residues are not involved in any hot region, therefore they are considered as singlet hot spot residues. In the interfaces, there are totally 855 residues in 151 hot regions, therefore there are 5.66 residues in each hot region on average (Table 6.2). The maximum number of hot regions in an interface is 3, the average number of hot regions per interface is 1.52 (total number of hot regions / number of interfaces with hot regions), while some interfaces do not have any hot regions (Table S2). In Table 6.4, the distribution of these SAVs within hot spots, hot regions, singlet hot spots and non-hot spots is analyzed. It is observed that destabilizing SAVs are 2.53, 1.54 and 6.44 times more likely to occur in hot spots, hot regions and singlet hot spots rather than in non-hot spot residues. Moreover, destabilizing SAVs significantly prefer to be in singlet hot spots rather than hot regions (OR=4.19, $p < 0.05$). The preference of destabilizing SAVs to be located mostly in hot spots are expected as the definition of hot spots are based on the destabilizing energies ($\Delta\Delta G \geq 2.0$ kcal/mol). However, the preference to be in singlet hot spots is striking. It has been also shown that, although it is not statistically significant, stabilizing SAVs are distributed with almost the same preference over non-hot spots, hot regions and singlet hot spots. An opposite trend can be seen for neutral SAVs, they are more likely to occur in non-hot spot residues compared to hot region residues (OR=0.39, $p < 0.05$) and compared to singlet hot spots (OR=0.24, $p < 0.05$). These results indicate that any SAV that is destabilizing is more likely to be found in hot spots, especially in singlet hot spots, while a neutral SAV is probably located in non-hot spots. Overall, distribution profile of SAVs having different effects on the binding free energies shows that hot regions, singlet hot spots and non-

hot spots are enriched in different SAV profiles. Therefore, the distribution profile of SAVs within these residues was continued to be analyzed to differentiate them.

Table 6.4. Odds ratios of destabilizing, stabilizing and neutral SAVs for interface residue distribution*.

	Destabilizing SAVs		Stabilizing SAVs		Neutral SAVs	
Hot region vs Non-hot spot	273 542	1.54†	30 86	0.94	53 336	0.39†
Singlet hot spot vs Non-hot spot	291 542	6.44†	15 86	0.92	17 336	0.24†
Singlet hot spot vs Hot region	291 273	4.19†	15 30	0.97	17 53	0.61
Hot spot vs Non-hot spot	273+291 542	2.53†	30+15 86	0.94	53+17 336	0.34†

*The number of SAVs in corresponding groups are given in the second, fourth and sixth columns. The corresponding odd ratios are in the following columns. All calculations are based on the total number of residues in each group. Hot region residues = 855, Non-hot spot residues=2318, Singlet hot spot =439, Hot-spot=1294. †Statistically significant p-value (<0.05) was obtained for OR values. ORs are given in the second columns.

6.2. Disease-Association Distribution of SAV within Interface Residues

Disease-association of SAVs is found by PolyPhen2 web server [187]. PolyPhen2 prediction is performed for only human SAVs. After excluding SAVs from non-human organisms, 489 human SAVs out of 1643 SKEMPI 2.0 SAVs are left. Total hot region, singlet hot spot and non-hot spot residue numbers are recalculated excluding non-human PDBs. Totally, 181 hot region, 176 singlet hot spot and 766 non-hot spot residues are found and OR calculations are performed based on these numbers. First, in order to associate binding free energy change with disease-association of SAVs, binding free energy change distribution of disease-causing and benign SAVs is analyzed. Figure 6.2 shows distributions of ΔΔGs of disease-causing and benign SAVs. x-axis shows ΔΔG values, while y-axis presents frequencies of these observations for disease-causing (blue

bars) and benign (orange bars) SAVs. Distribution of ΔΔG values caused by disease-causing SAVs is broader and shifts towards more positive values indicating their destabilizing effect. On the other hand, ΔΔG values of benign SAVs show a skewer distribution indicating a more neutral effect. The mean and standard deviation of ΔΔGs caused by disease-causing SAVs are 1.42 kcal/mol and 1.35 kcal/mol, while the mean and standard deviation of ΔΔGs caused by benign SAVs 0.84 kcal/mol and 1.16 kcal/mol, respectively. The difference between these two distributions is statistically highly significant ($p<<0.05$).

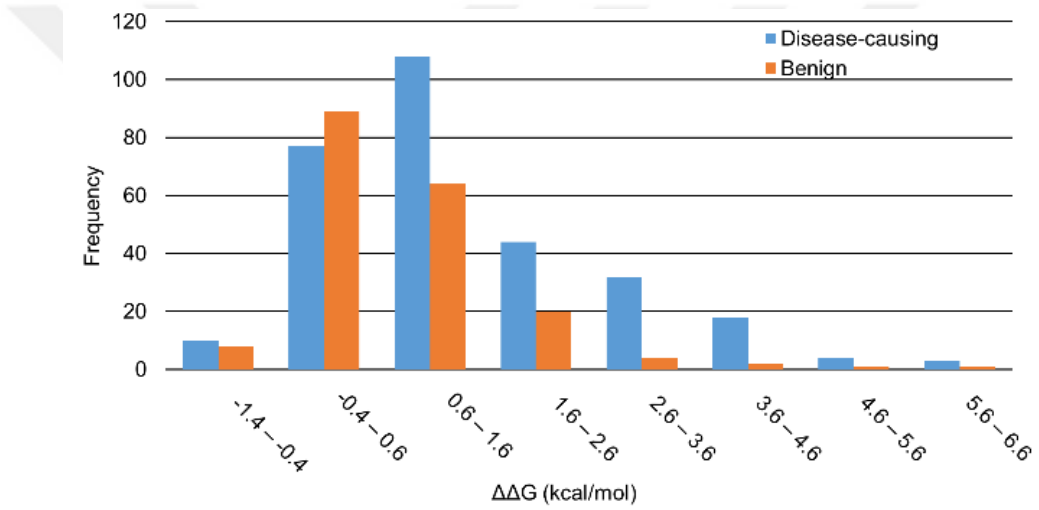


Figure 6.2. The distribution of binding free energy change in disease-causing and benign SAVs. ΔΔG values of disease-causing SAVs (blue bars) and benign SAVs (orange bars) are in x-axis and frequencies of these values are in y-axis.

Then, the dependence of disease-association on location of SAVs is tested and confirmed (hot region, singlet hot spots, non-hot spots) with χ^2 calculation by constructing a contingency table (Table 6.5, p-value=8.86e-05). Then, the distribution of disease-causing and benign SAVs within singlet hot spot, hot region or non-hot spot residues is further analyzed with OR calculations.

Table 6.5. χ^2 test for testing dependence of disease-association distribution

	Hot Region	**Singlet Hot Spot**	**Non-hot Spot**	**p-value**
Disease-causing SAVs	50	72	176	8.86e-05
Benign SAVs	29	18	144	

Table 6.6 shows how benign SAVs are distributed within hot region, singlet hot spots and non-hot spots. As expected, benign SAVs are 2.03 times more likely (OR=0.49, $p < 0.05$) to occur in non-hot spot residues compared to singlet hot spots. Disease-causing SAVs (both more confident and less confident PolyPhen2 predictions were considered as disease-causing) in Table 6.6 prefer to occur in singlet hot spot residues compared to non-hot spot residues (OR=2.32, $p < 0.05$). Also, singlet hot spot and hot regions comparison shows that disease-causing SAVs are 1.81 times more likely to be found in singlet hot spots (OR=1.81, $p < 0.05$) rather than in hot regions. Considering only more confident predictions, similar results with higher OR values are obtained. Disease-causing SAVs with more confident predictions prefer to be in singlet hot spot residues compared to both non-hot spot residues and hot regions (OR=2.61, $p < 0.05$ and OR=1.84, $p < 0.05$, respectively). This highlights that disease-causing SAVs specifically prefer to be located in singlet hot spots, which further differentiates SAV distribution on singlet hot spots from hot regions. This result can be explained by the epistasis of hot spots in the same hot region. If a SAV exists in a hot region, other hot spots in the same hot region may compensate for its effect. However, if a SAV corresponds to a singlet hot spot, its effect might be more severe, since their effect on binding free energy is additive.

Table 6.6. Odds ratios of benign and disease-causing SAVs for interface residue distribution for human proteins*.

	Benign SAVs		Disease-causing SAVs			
			More and less confident PolyPhen2 predictions		More confident PolyPhen2 predictions	
Hot region vs Non-hot spot	29 144	0.83	50 176	1.27	35 111	1.41
Singlet hot spot vs Non-hot spot	18 144	**0.49**†	72 176	**2.32**†	54 111	**2.61**†
Singlet hot spot vs Hot region	18 29	0.60	72 50	**1.81**†	54 35	**1.84**†
Hot spot vs Non-hot spot	29+18 144	**0.65**†	50+72 176	**1.74**†	35+54 111	**1.96**†

*All calculations are based on the total number of residues in each group. Hot region residues = 181, Non-hot spot residues=766, Singlet hot spot =176, Hot-spot=357
†statistically significant p-value (<0.05) was obtained for OR value. ORs are given in the second columns.

Since the classification as disease-causing and benign comes from PolyPhen2 predictions, the analyses in Table 6.6 are repeated, after matching human SKEMPI 2.0 SAVs with known disease variations from HGMD [188], dSysMap [189], COSMIC [191], PinSNP [190] which covers dbSNP, OMIM and Sahni et al data sets [144]. Thirty five human SKEMPI 2.0 SAVs match with disease variations from these data sets. Table C.1 gives information about disease types, gene names and data sets of these matched SAVs. According to the analysis with these 35 SAVs, disease-causing SAVs are more likely to be located in singlet hot spots rather than hot region with a higher OR in Table 6.6 (4.14, p-value=0.055. All calculations are based on the total number of residues in each group. Singlet hot spot residues = 23, hot region residues=27). Although these number is not statistically significant due to small numbers of observations, they provide some confirmations about the results obtained using PolyPhen2 predictions.

6.3. SAV Distribution within Interface Residues of GTPases

Five GTPases structures among the 172 different structures were identified. These are Rac1, HRas and Rho6 (two different structures are available for HRas and Rho6) (Figure 6.3). SAV distribution within interface residues of these GTPases were analyzed and parallel to the previous results destabilizing SAVs and SAVs predicted as disease-causing by PolyPhen2 preferred to be in singlet hot spot residues compared to both non-hot spot residues and hot regions (Table 6.7). The sample size for these analyses is small, therefore the results are not statistically significant. However, these results suggest that SAVs occurring in the singlet hot spot residues of GTPases are more prone to destabilize their interaction and lead the diseases.

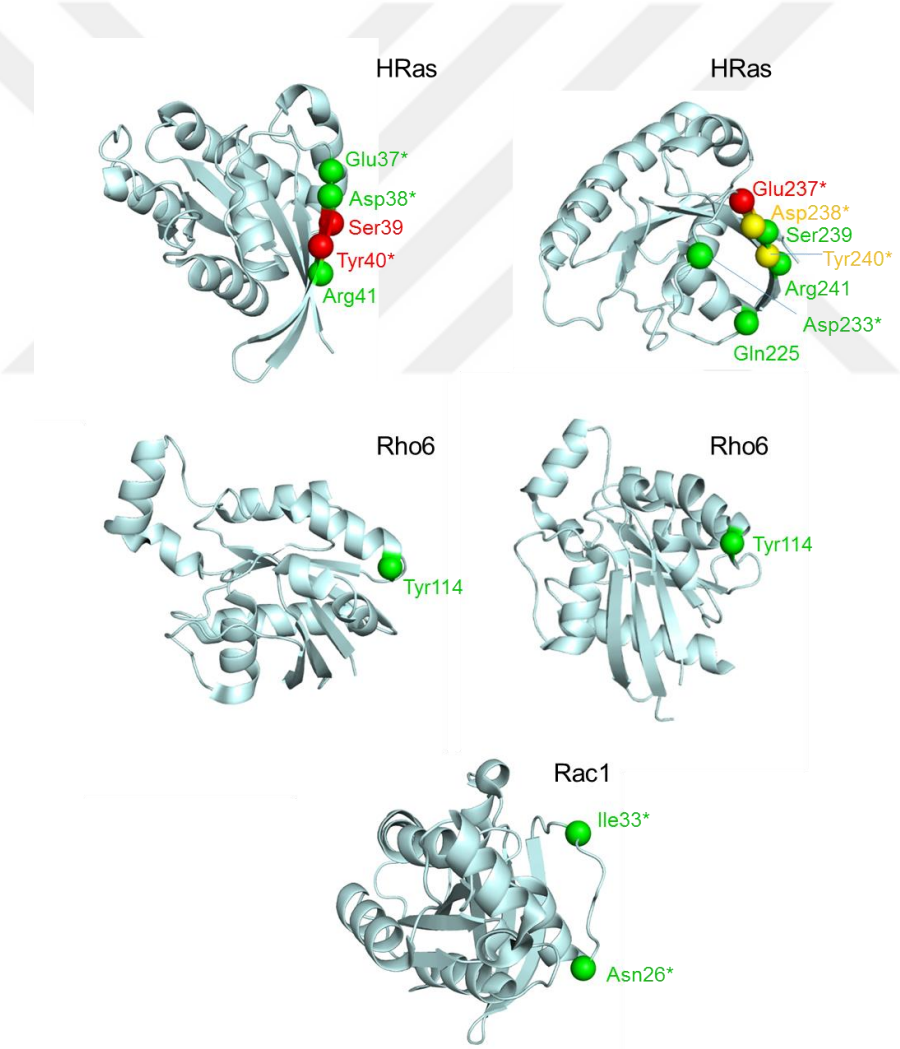


Figure 6.3. GTPase structures with SAVs. Singlet hot spot SAVs, hot region SAVs and non-hot spot SAVs are shown by yellow, red and green spheres. SAVs predicted as disease-causing are indicated with *. Residues are represented only by their Cα atoms.

Table 6.7. Odds ratios of disease-causing and destabilizing SAVs for interface residue distribution for GTPase proteins.

	Disease-causing SAVs		**Destabilizing SAVs**	
Singlet hot spot vs Non-hot spot	3 8	8.46	3 9	5.96
Singlet hot spot vs Hot region	3 3	9.05	3 4	7.81
Hot region vs Non-hot spot	3 8	0.97	4 9	1.22

*All calculations are based on the total number of residues in each group. Hot region residues = 13, Non-hot spot residues=34, Singlet hot spot =4. ORs are given in the second columns.

SAVs targeting singlet hot spots in Cdc42-GRD2 interface may impair Cdc42 binding to GRD2 in the case of IQGAP overexpression ans inhibit actin polymerization. Similarly, Arl2-PDEδ interaction may be terminated by singlet hot spot SAVs, consequently KRas4B-PDEδ complex is stabilized and oncogenic KRas4B attachment to the cellular membrane is inhibited.

6.4. Discussion

The detailed analysis of SAVs obtained from a large experimental data set provides a link between SAV distribution within the different interface residues and their effects. It has been shown that destabilizing SAVs are more likely to be in hot spot residues, specifically in singlet hot spots, while neutral SAVs tend to be located in non-hot spot residues. Since, hot spots are the major contributors to the binding free energy, they are considered to be crucial in establishing the stability and affinity of the protein complexes [154, 212]. These results show that SAVs being located in hot spots, especially in singlet hot spots are more likely to change binding affinity of PPI compared to SAVs non-hot spot residues. This indicates the importance of hot spots (singlet hot spots and hot regions) to provide stability to the protein complex. Hot spots are evolutionary

conserved residues. Therefore, it is expected that SAVs in these residues affect protein function and stability [212, 213].

The relationship between computationally predicted disease-association of SAVs and their distribution within non-hot spot, singlet hot spot residues and hot regions is also analyzed. It has been shown that benign SAVs significantly tend to occur in non-hot spot residues compared to hot regions and singlet hot spots. Also, it is observed observe that singlet hot spot residues are significantly enriched in disease-causing SAVs. This may be explained by additivity of singlet hot spots and epistatic effect of hot spots in hot regions on binding free energy. When a SAV occurs in one of the hot region other residues forming hot region may compensate the effect of this SAV on interaction, since they are epistatic. However, SAVs in singlet hot spots can directly affect PPIs due to their additive effect on binding free energy. Since the classification as disease-causing and benign comes from predictions, the findings are further confirmed by matching SKEMPI 2.0 SAVs with known disease variation data from various data sets. Similar results to our previous results are obtained; disease-causing SAVs are more likely to be found in singlet hot spots. Although the number of the matching SAVs is small, similar distribution trends in both prediction and experimental data are shown.

Singlet hot spots and hot regions are important in pharmacological studies. Designing drugs to target PPIs is a promising area for drug discovery and drug repurposing. Studies on this area reveal that there is a close relationship between PPI specific drugs and hot spots in interfaces. Also, computational methods confirm the relationship between hot spots and druggability [214]. Drugs targeting hot spots in interfaces increase their possibility to bind to their target interface and establish stable interaction. The major characteristics of druggable hot spots are about their location preferences, they have tendency to form clusters. Hot spots present a binding platform for PPI. These regions usually coevolve with hot regions of their binding partners which also give a critical insight for drug repurposing [215]. Therefore, it is crucial to establish a link between disease-association of SAVs and their distribution within contacting residues. The effect of SAVs in singlet hot spots cannot be easily compensated and destabilize the

interaction, therefore drugs designed to target singlet hot spots may be more effective to eliminate the interactions.

To conclude, revealing the relationship between disease-association of SAVs and their distribution within interface residues will provide important knowledge for targeting interfaces by therapeutic drugs. In this part of the thesis, with the use of experimentally obtained binding free energy data and PPI interfaces, a comprehensive analysis of SAV distribution is performed. The relationship between the effects of SAVs and their distribution within different interface residues is mainly studied. The analysis showed the importance of this relation to predict the impact of SAVs in PPIs and to evaluate their disease-associated effect. Predicting and determining these impacts are crucial to understand the nature of SAVs and their overall influence to biological processes.

Chapter 7

CONCLUSION

The main focuses of this dissertation are to reveal the detailed mechanism of interaction of Rho GTPases Cdc42 and Rac1 with the scaffolding protein IQGAP2 and understand the Arl2-mediated KRas4B release from the shuttling factor PDEδ. It also explains how SAVs are distributed within protein-protein interaction interface residue.

Ras superfamily GTPases regulate multiple cellular processes such as growth, intracellular trafficking, survival, cell invasion, cell motility, gene expression, cell proliferation, and transformation by interacting with effector molecules. To date, the mammalian Ras GTPase superfamily includes 5 main subfamily with more than 150 members; Ras, Cdc42, and Rac1 are the most studied among them. There are numerous mechanisms, which is not fully understood yet, related to Ras superfamily GTPases. Cdc42 and Rac1 binding mechanism and stoichiometry to IQGAP are among these unsolved mechanisms. Here, the underlying mechanism of the different stoichiometry that is involved in the binding of Cdc42 and Rac1 to the scaffolding protein IQGAP2 and facilitates its dimerization was studied. Cdc42 is known for playing a role in metastasis and this thesis study provides an insight to decipher how Cdc42 can stimulate actin polymerization in metastasis. It was revealed that insert loop and switch I of Cdc42 are mainly responsible for binding of the first Cdc42 to IQGAP2. Cdc42 binding to this site induces allosteric changes in the other site of IQGAP to enable the second Cdc42 to bind. This binding, also, releases the C-term residues of IQGAP2 and facilitates dimerization. However, the orientation of the insert loop of Rac1 does not allow two Rac1 to bind to IQGAP2; thus, the C-term cannot be released. Since Cdc42 and IQGAP2 interactions play an important role in actin metabolism and metastasis, this thesis study provides a help to elucidate the key question of the roles of Cdc42 and Rac1 in actin polymerization in metastasis.

Ras protein shuttling to the plasma membrane and its release from the shuttling factor in order to attach the membrane are required to be understood to develop related therapeutic approach. The releases mechanism of Ras GTPases KRas4B, which is the most frequently mutated Ras isoform in human cancer with currently no drugs in the clinics, from the shuttling factor PDEδ was studied. It was revealed the Arl2-mediated release mechanism of farnesylated KRas4B from PDEδ showing that Arl2 binding weakens the PDEδ–farnesylated KRas4B interaction. It was showed that allosteric changes (involving β6 of PDEδ and additional PDEδ residues) compress the hydrophobic PDEδ pocket and push the KRas4B out. Mutating PDEδ residues that mediate allosteric changes in PDEδ terminates the release process. This mechanistic account may inspire efforts to develop drugs suppressing oncogenic KRas4B release.

Finally, it was aimed to understand how SAVs are distributed within interface residues and how this distribution can be associated with their effect on protein function and binding. SAVs with literature curated experimental thermodynamics data were statistically and structurally analyzed and observed that destabilizing SAVs have a statistically significant tendency to occur in hot spots, specifically in singlet hot spots. Interestingly, analysis of computationally predicted disease-causing and benign SAVs showed that the disease-causing SAVs are more likely to occur in singlet hot spot residues.

To conclude, this thesis study enhances the understanding of interactions and shuttling mechanisms of Ras GTPases superfamily and approaches cancer in terms of both metastasis and cell proliferation, therefore, it may facilitate therapeutic targeting Ras GTPases.

APPENDIX A

A sample minimization configuration file (suitable for VMD-generated psf file)

```
#############################################################
## JOB DESCRIPTION                                         ##
#############################################################

# Minimization and Equilibration of
# Ubiquitin in a Water Box

#############################################################
## ADJUSTABLE PARAMETERS                                   ##
#############################################################

structure          5cjpBCE_autopsf.psf  # psf file generated by VMD
coordinates        5cjpCE_autopsf.pdb # pdb file generated by VMD

set temperature    310
set outputname     IQGAP2_BCEmin

firsttimestep      0

#############################################################
## SIMULATION PARAMETERS                                   ##
#############################################################

# Input
paraTypeCharmm          on
parameters          par_all36_ras_c39b2_hbj.inp
temperature         $temperature


# Force-Field Parameters
exclude             scaled1-4
1-4scaling          1.0
cutoff              12.0
switching           on
switchdist          10.0
pairlistdist        14.0

# Integrator Parameters
timestep            2.0  ;# 2fs/step
rigidBonds          all  ;# needed for 2fs steps
```

```
nonbondedFreq       1
fullElectFrequency  2
stepspercycle       10


# Constant Temperature Control
langevin            on    ;# do langevin dynamics
langevinDamping     1     ;# damping coefficient (gamma) of 1/ps
langevinTemp        $temperature
langevinHydrogen    off    ;# don't couple langevin bath to hydrogens

# Periodic Boundary Conditions
cellBasisVector1    107.7      0.   0.0
cellBasisVector2     0.0   120.61   0.0
cellBasisVector3     0.0    0      84.66
cellOrigin          -17.12   -2.15  -9.96

wrapAll             on

# PME (for full-system periodic electrostatics)
PME                 yes
PMEGridSpacing      1.0

#manual grid definition
#PMEGridSizeX        45
#PMEGridSizeY        45
#PMEGridSizeZ        48


# Constant Pressure Control (variable volume)
useGroupPressure      yes ;# needed for rigidBonds
useFlexibleCell       no
useConstantArea       no

langevinPiston        on
langevinPistonTarget  1.01325 ;#  in bar -> 1 atm
langevinPistonPeriod  100.0
langevinPistonDecay   50.0
langevinPistonTemp    $temperature


# Output
outputName          $outputname

restartfreq         500     ;# 500steps = every 1ps
dcdfreq             500
xstFreq             500
outputEnergies      100
```

```
outputPressure      100

#####################################################################
## EXECUTION SCRIPT                                 ##
#####################################################################

# Minimization
minimize            10000
reinitvels          $temperature

#run 10000 ;# 20ps
```

A sample product run configuration file (suitable for VMD-generated psf file)

```
#############################################################
## JOB DESCRIPTION                                         ##
#############################################################
# product RUN
#############################################################
## ADJUSTABLE PARAMETERS                                   ##
#############################################################
structure               arlgtppdehvf1_autopsf.psf  # psf file generated by VMD
Coordinates             arlgtppdehvf1_autopsf.pdb # pdb file generated by VMD
binCoordinates          arlgtppdehvf1_min.coor # output of minimization step
binVelocities           arlgtppdehvf1_min.vel # output of minimization step
extendedSystem          arlgtppdehvf1_min.xsc # output of minimization step

set temperature    310 # body temperature
set outputname     arlgdtppdehvf1_50ns_run

firsttimestep     0
#############################################################
## SIMULATION PARAMETERS                                   ##
#############################################################
# Input
paraTypeCharmm          on
parameters          par_all36_ras_c39b2_hbj.inp
#temperature         $temperature

# Force-Field Parameters
exclude             scaled1-4
1-4scaling          1.0
cutoff             12.0
switching           on
```

```
switchdist          10.0
pairlistdist        14.0

# Integrator Parameters
timestep            2.0  ;# 2fs/step
rigidBonds          all  ;# needed for 2fs steps
nonbondedFreq       1
fullElectFrequency  2
stepspercycle       10

# Constant Temperature Control
langevin            on    ;# do langevin dynamics
langevinDamping     2     ;
langevinTemp        $temperature
langevinHydrogen    off    ;# don't couple langevin bath to hydrogens

# Periodic Boundary Conditions
cellBasisVector1    76.40     0.   0.0
cellBasisVector2     0.0   87.76   0.0
cellBasisVector3     0.0    0      87.53
cellOrigin          16.28 9.75 23.41
wrapAll             on

# PME (for full-system periodic electrostatics)
PME                 yes
PMEGridSpacing      1.0
#manual grid definition
#PMEGridSizeX        45
#PMEGridSizeY        45
#PMEGridSizeZ        48
```

```
# Constant Pressure Control (variable volume)
useGroupPressure      yes ;# needed for rigidBonds
useFlexibleCell       no
useConstantArea       no
langevinPiston        on
langevinPistonTarget  1.01325 ;#  in bar -> 1 atm
langevinPistonPeriod  200.0
langevinPistonDecay   100.0
langevinPistonTemp    $temperature

# Output
outputName          $outputname
restartfreq         50000     ;# 500steps = every 1ps
dcdfreq             50000
xstFreq             50000
outputEnergies      50000
outputPressure      50000
#############################################################
## EXECUTION SCRIPT                                        ##
#############################################################
run 25000000 ;# 50ns
```

A sample psf generation CHARMM input file

```
* File: k12dgtw_water_mk_psf_salt.inp
* Generate combined coordinates for NAMD simulation
*
!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
! set directory
set type k12dgtw_water
set dir data/ozdemires/@type
!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
bomlev -2

! read topology
open read card unit 10 name
/data/ozdemires/charmm_input/top_all36_ras_c39b2_hbj.inp
read rtf card unit 10
close unit 10

! read parameters
open read card unit 10 name
/data/ozdemires/charmm_input/par_all36_ras_c39b2_hbj.inp
read param flex card unit 10
close unit 10

!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!

! Get k-ras (1-185) coordinates with farnesylated cys
read sequence card
* The sequence is RasK 1 to 185
*
185
MET THR GLU TYR LYS LEU VAL VAL VAL GLY ALA ASP GLY
VAL GLY LYS SER ALA LEU THR ILE GLN LEU ILE GLN ASN
HSD PHE VAL ASP GLU TYR ASP PRO THR ILE GLU ASP SER
TYR ARG LYS GLN VAL VAL ILE ASP GLY GLU THR CYS LEU
LEU ASP ILE LEU ASP THR ALA GLY GLN GLU GLU TYR SER
ALA MET ARG ASP GLN TYR MET ARG THR GLY GLU GLY PHE
LEU CYS VAL PHE ALA ILE ASN ASN THR LYS SER PHE GLU
ASP ILE HSD HSD TYR ARG GLU GLN ILE LYS ARG VAL LYS
ASP SER GLU ASP VAL PRO MET VAL LEU VAL GLY ASN LYS
CYS ASP LEU PRO SER ARG THR VAL ASP THR LYS GLN ALA
GLN ASP LEU ALA ARG SER TYR GLY ILE PRO PHE ILE GLU
THR SER ALA LYS THR ARG GLN GLY VAL ASP ASP ALA PHE
TYR THR LEU VAL ARG GLU ILE ARG LYS HSD LYS GLU LYS
MET SER LYS ASP GLY LYS LYS LYS LYS LYS LYS SER LYS
THR LYS CYF

generate RASK first NTER last CT1 setup
```

```
! Get GTP in crystal structure
read sequence GTP 1
generate GTP setup first none last none

! Get MG in crystal structure
read sequence MG 1
generate MAGN setup first none last none

open read card unit 1 name /@dir/coor/@type_add_salt_system.crd
read coor card unit 1
close unit 1

set N ?nres

open read card unit 1 name /@dir/coor/temp_blk1.crd
read sequence coor unit 1
generate BLK1 nodihedrals noangles first none last none
close unit 1
open read card unit 1 name /@dir/coor/temp_blk1.crd
read coor card unit 1 offset @N
close unit 1

set N ?nres

open read card unit 1 name /@dir/coor/temp_blk2.crd
read sequence coor unit 1
generate BLK2 nodihedrals noangles first none last none
close unit 1
open read card unit 1 name /@dir/coor/temp_blk2.crd
read coor card unit 1 offset @N
close unit 1

set N ?nres

open read card unit 1 name /@dir/coor/temp_blk3.crd
read sequence coor unit 1
generate BLK3 nodihedrals noangles first none last none
close unit 1
open read card unit 1 name /@dir/coor/temp_blk3.crd
read coor card unit 1 offset @N
close unit 1

set N ?nres

open read card unit 1 name /@dir/coor/temp_blk4.crd
read sequence coor unit 1
```

```
generate BLK4 nodihedrals noangles first none last none
close unit 1
open read card unit 1 name /@dir/coor/temp_blk4.crd
read coor card unit 1 offset @N
close unit 1

set N ?nres

open read card unit 1 name /@dir/coor/temp_sod.crd_mod
read sequence coor unit 1
generate SODI noangle nodihedral
close unit 1
open read card unit 1 name /@dir/coor/temp_sod.crd_mod
read coor card unit 1 offset @N
close unit 1

set N ?nres

open read card unit 1 name /@dir/coor/temp_cla.crd_mod
read sequence coor unit 1
generate CHLO noangle nodihedral
close unit 1
open read card unit 1 name /@dir/coor/temp_cla.crd_mod
read coor card unit 1 offset @N
close unit 1
!!!!!!!!!!!!!!!!!
! Stat for protein
coor stat sele segid RASK end
coor stat sele segid RASK .and. type CA end
coor stat sele segid GTP end

! Stat for salt
coor stat sele segid MAGN end
coor stat sele resn SOD end
coor stat sele resn CLA end
! Stat for water
coor stat sele segid BLK1 .and. type OH2 end
coor stat sele segid BLK2 .and. type OH2 end
coor stat sele segid BLK3 .and. type OH2 end
coor stat sele segid BLK4 .and. type OH2 end

!!!!!!!!!!!!!!!!!
!IOFOrmat NOEXtended
!!!!!!!!!!!!!!!!!

! Write psf file in charmm format
open write card unit 1 name /@dir/psf/@type_namd_salt_charmm.psf
write psf card unit 1
```

```
close unit 1

! Write psf file in charmm format
open write card unit 1 name /@dir/psf/@type_namd_salt_xplor.psf
write psf card xplor unit 1
close unit 1

!!!!!!!!!!!!!!!!!

! Get the final coordinates in pdb
open write form unit 1 name /@dir/coor/@type_namd_salt_charmm.pdb
write coor pdb unit 1
close unit 1

stop
```

A sample minimization configuration file (suitable for CHARMM-generated psf file)

```
#%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%#
#   FILE: arlgdppdehvf1_namd_preequ_mini.cnfg                  #
#   Minimization of solvent with protein fix                   #
#%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%#
#%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%#
#   Set variable values                                        #
#%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%#
set temp            310
set dir             /data/ozdemires/k12dgtw_water
set type            k12dgtw_water

#%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%#
#   Parameters                                                 #
#%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%#
paraTypeCharmm      on
parameters          /data/ozdemires/charmm_input/par_all36_ras_c39b2_hbj.inp

#%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%#
#   Input files                                                #
#%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%#
structure           ${dir}/psf/${type}_namd_salt_xplor.psf      ;# Name of structure file
coordinates         ${dir}/coor/${type}_namd_salt_charmm.pdb    ;# Name of input
pdb file

#%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%#
#   Restart files                                              #
#%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%#
# NOTE: Do not set the initial restart velocity if you have also specified a temperature!
if {0} {
velocities          ${dir}/vel/                    ;# Initial velocies in pdb file
binVelocities       ${dir}/vel/                    ;# Initial velocies in binary format
                                                   ;# Remove the "temperature" entry if you use
this!
binCoordinates      ${dir}/coor/                   ;# Initial coor in binary format
```

```
extendedSystem       ${dir}/xsc/                          ;# Input periodic cell para read
from
}

#%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%#
#   Simulation temperature                              #
#%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%#
# NOTE: Do not set the initial velocity temperature if you have also specified a .vel
restart file!
temperature          $temp

#%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%#
#   Output files                                        #
#%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%#
outputName           ${dir}/coor/${type}_namd_preequ_mini_pdb    ;# Name of final
output coor file
binaryoutput         no                                  ;# Save final coor in binary format
restartname          ${dir}/coor/${type}_namd_preequ_mini        ;# Name of
output(restart) coor & vel file
binaryrestart        yes                                 ;# Save restart file in binary format
restartfreq          10000                               ;# Freq saving coor & vel (1000 steps
= every 1ps)
restartsave          yes                                 ;# Append time step at end of file name
dcdfile              ${dir}/trj/${type}_namd_preequ_mini.dcd     ;# Name of trajectory
file
dcdfreq              100                                 ;# Freq saving trajectory
dcdUnitCell          yes                                 ;# Write unit cell data to dcd file
#veldcdfile    ${dir}/vel/${type}_namd_preequ_mini_vel.dcd       ;# Name of vel
trajectory file
#veldcdfreq          500                                 ;# Freq saving vel trajectory

#%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%#
#   Output configurations                               #
#%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%#
outputEnergies       100                                 ;# Freq for printing Energies
outputTiming         100                                 ;# Freq for printing Timing info
outputPressure       100                                 ;# Freq for printing Pressure

#%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%#
#   Periodic Boundary conditions                        #
```

```
#%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%#
# NOTE: Do not set the periodic cell basis if you have specified an .xsc restart file!
if {1} {
cellBasisVector1     100.0  0.   0.
cellBasisVector2       0.  100.0  0.
cellBasisVector3       0.   0.  100.0
cellOrigin             0.   0.   0.
}
#xstFile        ${dir}/xsc/${type}_namd_preequ_mini.xst     ;# Name of cell param trajectory file
#xstfreq        100                              ;# Freq for writing cell param trajectory
wrapWater       on                               ;# Wrap water around periodic boundary
wrapAll         on                               ;# Wrap all atoms around periodic boundary

#%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%#
#   Force-field parameters                                #
#%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%#
cutoff          12.                              ;# Local interaction dist to elec & VdW calculation
switching       on                               ;# Smoothing func for cutoff, otherwise truncated
switchdist      10.                              ;# Should be less than or equal to cutoff value
pairlistdist    14.                              ;# Cutoff for generating pair list,larger than cutoff
exclude         scaled1-4                        ;# Exclude pairs in nonbonding interactions
1-4scaling      1.0                              ;# Scaling factor for 1-4 interaction
COMmotion       no                               ;# No center of mass motion for entire system
splitPatch      hydrogen                         ;# Set hydrogen to allow faster distance check
hgroupCutoff    2.5                              ;# Use for group-based distance testing

#%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%#
#   Time step parameters                                  #
#%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%#
timestep        2.                               ;# 2 fs/step
firsttimestep   0                                ;# Starting time step value
stepspercycle   20                               ;# Freq updating nonbond list
```

```
#%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%#
#   Multiple time step parameters for long range elec             #
#%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%#
fullElectFrequency    1                              ;# Calc long range elec stat every
time step
nonbondedFreq         1                              ;# Freq for calc short range nonbond
interac
MTSAlgorithm          impulse                        ;# Integr method for long range
elec
longsplitting         c1                             ;# Method for split elec forces between
long and short

#%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%#
#   PME parameters                                   #
#%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%#
if {0} {
PME                   no                             ;# Use particle mesh Ewald for
electrostatics
PMETolerance          1.0e-6                         ;# Affect value of Ewald
coefficient
PMEInterpOrder        4                              ;# For cubic 4
PMEGridSizeX          92                             ;# Number of grid ponts in x
PMEGridSizeY          92                             ;# Number of grid ponts in y
PMEGridSizeZ          92                             ;# Number of grid ponts in z
}

#%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%#
#   Constant temperature control                     #
#%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%#
if {0} {
langevin              on                             ;# Use langeving dynam for temp
control
langevinTemp          $temp
langevinDamping       2.5                            ;# Friction coeff (gamma) (1/ps),
use a small value
langevinHydrogen      off                            ;# Don't couple langevin bath to
hydrogens
}

#%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%#
```

```
#  Temperature control for heating or cooling                    #
#%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%#
if {0} {
reassignFreq         50000                              ;# Frequency to reassign temperature
reassignTemp          10.0                              ;# Initial temperature
reassignIncr         30.0                              ;# Temperature increment or
decrement
reassignHold         $temp                              ;# Target temperature
}

#%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%#
#   Constant pressure control                              #
#%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%#
useGroupPressure      yes                              ;# Should use w/SHAKE, needed
for 2fs steps
useFlexibleCell       no                              ;# Allows dimensions of cell to
fluctuate
                                         ;# no for water box, yes for membrane
useConstantArea       no                              ;# x-y fixed and z dim can vary
                                         ;# no for water box, yes for membrane
if {1} {
LangevinPiston        on                              ;# Use Nose-Hoover Langeving
Piston Algorithm
LangevinPistonTarget  1.01325                              ;# Target pressure to run simu
at (bars)
LangevinPistonPeriod  200                              ;# Oscillation period
LangevinPistonDecay   100                              ;# Damping time scale
LangevinPistonTemp    $temp                              ;# Should be set equal to the
target temperature
SurfaceTensionTarget  0.0                              ;# Apply suface tension in xy
plane (dyn/cm)
}

#%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%#
#   Fix hydrogens                              #
#%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%#
if {1} {
rigidBonds           all                              ;# Fix h-bonds to none, water, or all h-
atoms
rigidTolerance        1.0e-8                              ;# Allowable bond-length error for
ShakeH
rigidIterations      100                              ;# Maximum ShakeH iterations
rigidDieOnError       on                              ;# Exit and report an error
```

```
useSettle           off                                 ;# Keep waters rigid
}

#%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%#
#   Harmonic constraints : constraint protein backbone with k=5    #
#%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%#
if {1} {
constraints          on                                 ;# Turn on the use of contraints
consexp              2                                  ;# Exponent to be use in harmonic
constraint energy
consref       ${dir}/cons/${type}_preequ_constr_atoms_ref_5.pdb    ;# A pdb file to
use for reference positions
conskfile     ${dir}/cons/${type}_preequ_constr_atoms_kfile_5.pdb  ;# A pdb file to
use for k values
conskcol             B                                  ;# Column of pdb file to use for
harmonic constraint
}

#%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%#
#   Fixed atoms constraint : fix protein                        #
#%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%#
if {1} {
fixedAtoms           on                                 ;# Turn on the use of contraints
fixedAtomsForces      on                                ;# Calculate forces between fixed
atoms
fixedAtomsFile       ${dir}/cons/${type}_preequ_fix_protein.pdb      ;# A pdb file to
use for reference positions
fixedAtomsCol        O                                  ;# Set O to read occupancy column
in pdb file
}

#%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%#
#   Minimization parameters                                   #
#%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%#
if {1} {
minimization         on                                 ;# Perform conjugate gradient energy
minimization
minTinyStep          1.0e-8                             ;# First initial step for line minimizer
minBabyStep          1.0e-2                             ;# Max initial step for line
minimizer
minLineGoal          1.0e-4                             ;# Gradient reduction factor for line
minimizer
```

```
}

#%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%#
#   Extra parameters                              #
#%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%#
# Put here any custom parameters that are specific to this job

#%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%#
#   Execution script                              #
#%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%#
#----------------------------#
minimize            10000                         ;# Minimization step
#----------------------------#
#----------------------------#
#run                10000                         ;# Dynamics step
#----------------------------#
#%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%#
#   End of script
#%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%#
```

A sample product run configuration file (suitable for CHARMM-generated psf file)

#%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%#

FILE: arlgdppdehvf1_namd_prod_10_ns.cnfg

10 ns dynamics and production

#%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%#

Set variable values

#%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%#

set temp 310

set dir /data/ozdemires/arlgdppdehvf1

set type arlgdppdehvf1

#%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%#

Parameters

#%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%#

paraTypeCharmm on

parameters /data/ozdemires/charmm_input/par_all36_ras_c39b2_hbj.inp

#%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%#

Input files

#%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%#

structure ${dir}/psf/${type}_namd_salt_xplor.psf ;# Name of structure file

coordinates ${dir}/coor/${type}_namd_preequ_ewald_last_pdb.coor ;# Name of input pdb file

#%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%#

Restart files

#%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%#

NOTE: Do not set the initial restart velocity if you have also specified a temperature!

if {1} {

#velocities ${dir}/vel/ ;# Initial velocies in pdb file

binVelocities ${dir}/vel/${type}_namd_preequ_ewald_last.50000.vel ;# Initial velocies in binary format

;# Remove the "temperature" entry if you use this!

binCoordinates ${dir}/coor/${type}_namd_preequ_ewald_last.50000.coor ;# Initial coor in binary format

extendedSystem ${dir}/xsc/${type}_namd_preequ_ewald_last.50000.xsc ;# Input periodic cell para read from

}

#%%#

Simulation temperature

#%%#

NOTE: Do not set the initial velocity temperature if you have also specified a .vel restart file!

#temperature $temp

#%%#

Output files

#%%#

outputName ${dir}/coor/${type}_namd_prod_10_ns_pdb ;# Name of final output coor file

binaryoutput no ;# Save final coor in binary format

restartname ${dir}/coor/${type}_namd_prod_10_ns ;# Name of output(restart) coor & vel file

binaryrestart yes ;# Save restart file in binary format

restartfreq 500000 ;# Freq saving coor & vel (1000 steps = every 1ps)

restartsave yes ;# Append time step at end of file name

dcdfile ${dir}/trj/${type}_namd_prod_merged_10_ns.dcd ;# Name of trajectory file

dcdfreq 5000 ;# Freq saving trajectory

```
dcdUnitCell         yes                         ;# Write unit cell data to dcd file
#veldcdfile            ${dir}/vel/${type}_namd_prod_10_ns_vel.dcd  ;# Name of vel trajectory file
#veldcdfreq         1000                        ;# Freq saving vel trajectory
#%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%#
#   Output configurations                                 #
#%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%#
outputEnergies      5000                        ;# Freq for printing Energies
outputTiming        5000                        ;# Freq for printing Timing info
outputPressure      5000                        ;# Freq for printing Pressure
#%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%#
#   Periodic Boundary conditions                          #
#%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%#
# NOTE: Do not set the periodic cell basis if you have specified an .xsc restart file!
if {0} {
cellBasisVector1    90.0  0.   0.
cellBasisVector2     0.  90.0  0.
cellBasisVector3     0.   0.  90.0
cellOrigin          0.   0.   0.
}
xstFile             ${dir}/xsc/${type}_namd_prod_10_ns.xst      ;# Name of cell param trajectory file
xstfreq             5000                        ;# Freq for writing cell param trajectory
wrapWater           on                          ;# Wrap water around periodic boundary
wrapAll             on                          ;# Wrap all atoms around periodic boundary

#%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%#
#   Force-field parameters                                #
```

```
#%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%#
cutoff              12.                        ;# Local interaction dist to elec & VdW calculation
switching           on                         ;# Smoothing func for cutoff, otherwise truncated
switchdist          10.                        ;# Should be less than or equal to cutoff value
pairlistdist        14.                        ;# Cutoff for generating pair list,larger than cutoff
exclude             scaled1-4                  ;# Exclude pairs in nonbonding interactions
1-4scaling          1.0                        ;# Scaling factor for 1-4 interaction
COMmotion           no                         ;# No center of mass motion for entire system
splitPatch          hydrogen                   ;# Set hydrogen to allow faster distance check
hgroupCutoff        2.5                        ;# Use for group-based distance testing
#%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%#
#   Time step parameters                       #
#%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%#
timestep            2.                         ;# 2 fs/step
firsttimestep       0                          ;# Starting time step value
stepspercycle       20                         ;# Freq updating nonbond list
#%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%#
#   Multiple time step parameters for long range elec        #
#%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%#
fullElectFrequency  1                          ;# Calc long range elec stat every time step
nonbondedFreq       1                          ;# Freq for calc short range nonbond interac
MTSAlgorithm        impulse                    ;# Integr method for long range elec
longsplitting       c1                         ;# Method for split elec forces between long and short
#%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%#
#   PME parameters                             #
#%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%#
```

```
if {1} {
PME                 yes                                 ;# Use particle mesh Ewald for electrostatics
PMEGridSpacing       1.0                    ;# NAMD assigns automatically the grid space
#PMETolerance        1.0e-6                              ;# Affect value of Ewald coefficient
#PMEInterpOrder       4                                  ;# For cubic 4
#PMEGridSizeX         92                                 ;# Number of grid ponts in x
#PMEGridSizeY         92                                 ;# Number of grid ponts in y
#PMEGridSizeZ         92                                 ;# Number of grid ponts in z
}
#%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%#
#   Constant temperature control                                #
#%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%#
if {1} {
langevin            on                                ;# Use langeving dynam for temp control
langevinTemp         $temp
langevinDamping      2.5                 ;# Friction coeff (gamma) (1/ps), use a small value
langevinHydrogen     off                   ;# Don't couple langevin bath to hydrogens
}
#%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%#
#  Temperature control for heating or cooling                   #
#%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%#
if {0} {
reassignFreq         50000                              ;# Frequency to reassign temperature
reassignTemp          10.0                              ;# Initial temperature
reassignIncr         30.0                              ;# Temperature increment or decrement
reassignHold         $temp                              ;# Target temperature
}
```

```
#%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%#
#   Constant pressure control                          #
#%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%#
useGroupPressure      yes                  ;# Should use w/SHAKE, needed for 2fs steps
useFlexibleCell      no                              ;# Allows dimensions of cell to fluctuate
                                                      ;# no for water box, yes for membrane
useConstantArea       no                              ;# x-y fixed and z dim can vary
                                                      ;# no for water box, yes for membrane
if {1} {
LangevinPiston       on                  ;# Use Nose-Hoover Langeving Piston Algorithm
LangevinPistonTarget  1.01325                    ;# Target pressure to run simu at (bars)
LangevinPistonPeriod  200                                ;# Oscillation period
LangevinPistonDecay   100                                ;# Damping time scale
LangevinPistonTemp    $temp               ;# Should be set equal to the target temperature
SurfaceTensionTarget  0.0                  ;# Apply suface tension in xy plane (dyn/cm)
}
#%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%#
#   Fix hydrogens                                    #
#%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%#
if {1} {
rigidBonds           all                   ;# Fix h-bonds to none, water, or all h-atoms
rigidTolerance       1.0e-8                ;# Allowable bond-length error for ShakeH
rigidIterations      100                             ;# Maximum ShakeH iterations
rigidDieOnError      on                               ;# Exit and report an error
useSettle            off                             ;# Keep waters rigid
}
#%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%#
```

Harmonic constraints

#%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%#

if {0} {

constraints on ;# Turn on the use of contraints

consexp 2 ;# Exponent to be use in harmonic constraint energy

consref ${dir}/cons/ ;# A pdb file to use for reference positions

conskfile ${dir}/cons/ ;# A pdb file to use for k values

conskcol B ;# Column of pdb file to use for harmonic constraint

}

#%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%#

Fixed atoms constraint

#%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%#

if {0} {

fixedAtoms on ;# Turn on the use of contraints

fixedAtomsForces off ;# Calculate forces between fixed atoms

fixedAtomsFile ${dir}/coor/ ;# A pdb file to use for reference positions

fixedAtomsCol O ;# Set O to read occupancy column in pdb file

}

#%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%#

Minimization parameters

#%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%#

if {0} {

minimization on ;# Perform conjugate gradient energy minimization

minTinyStep 1.0e-8 ;# First initial step for line minimizer

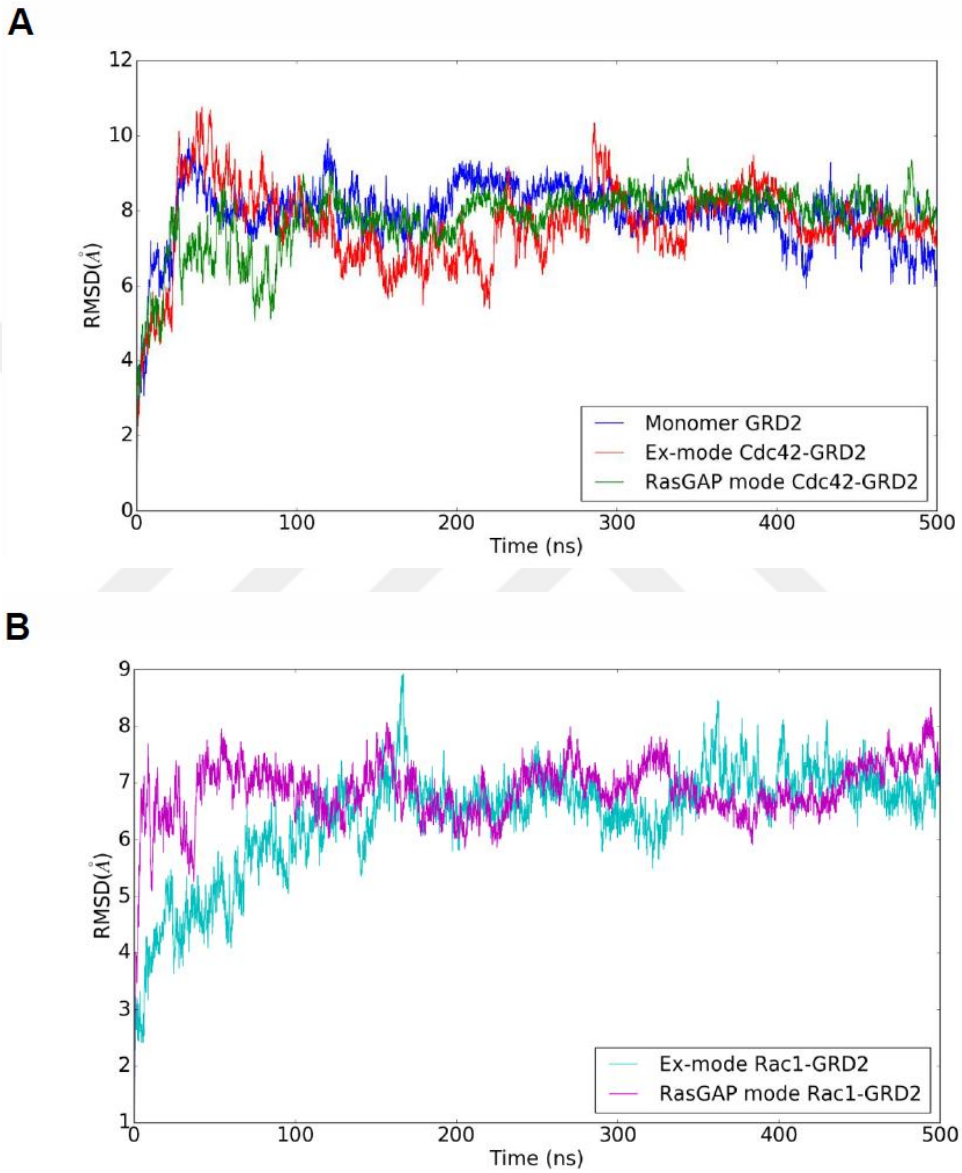
minBabyStep 1.0e-2 ;# Max initial step for line minimizer

minLineGoal 1.0e-4 ;# Gradient reduction factor for line minimizer

}

```
#%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%#

#   Extra parameters                                   #

#%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%#

# Put here any custom parameters that are specific to this job

#%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%#

#   Execution script                                   #

#%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%#

#----------------------------#

#minimize          5000                    ;# Minimization step

#----------------------------#

#----------------------------#

run             5000000                    ;# Dynamics step

#----------------------------#

#%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%#

#   End of script

#%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%#
```

APPENDIX B


A
RMSD(Å)
Time (ns)
Monomer GRD2
Ex-mode Cdc42-GRD2
RasGAP mode Cdc42-GRD2
B
RMSD(Å)
Time (ns)
Ex-mode Rac1-GRD2
RasGAP mode Rac1-GRD2

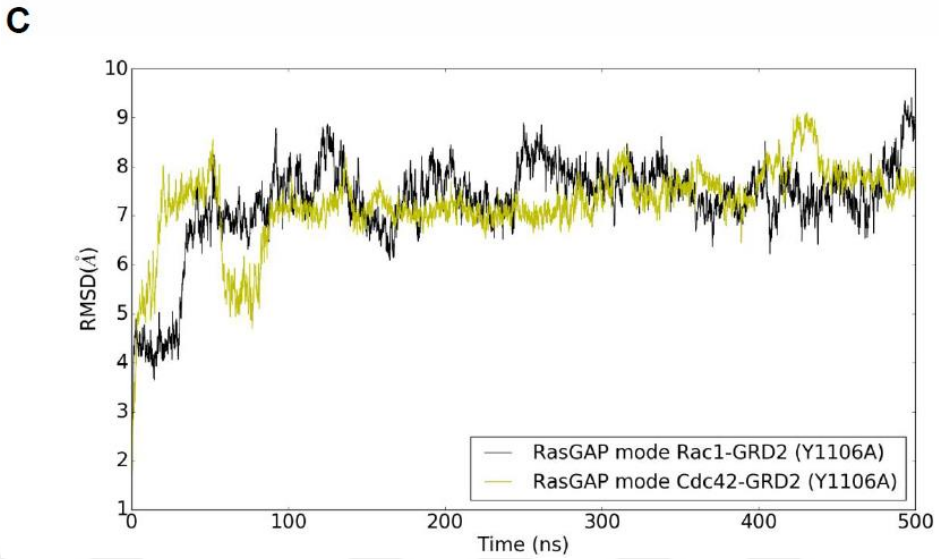
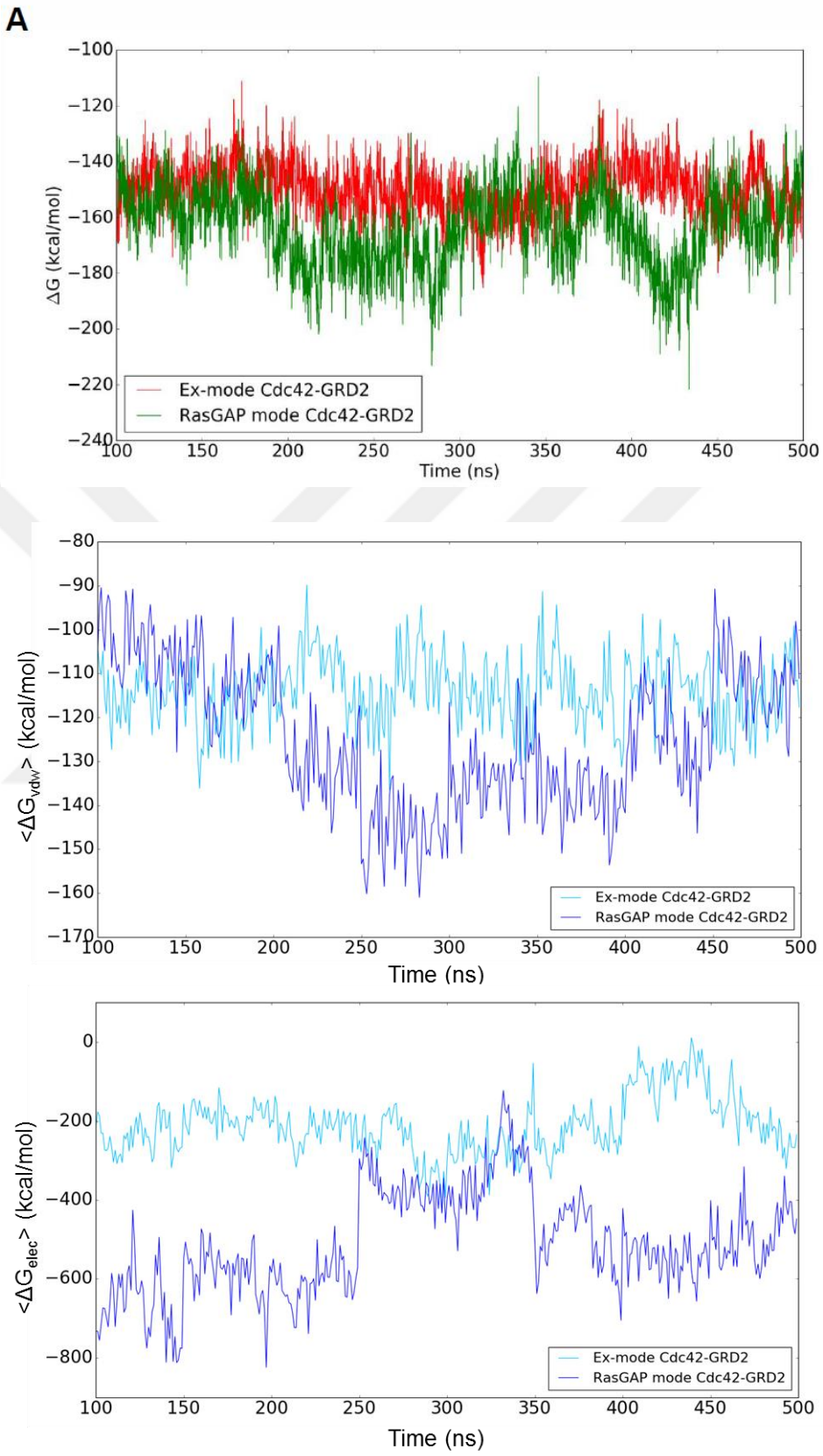
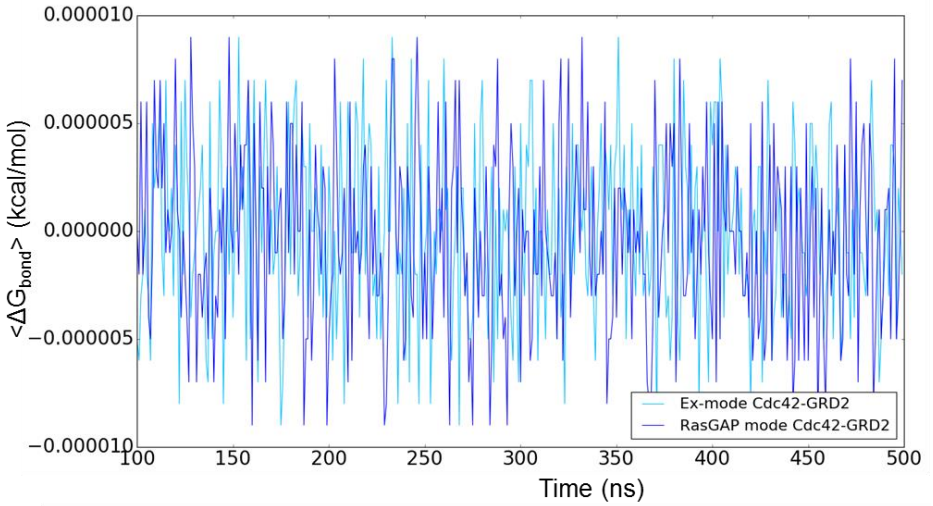
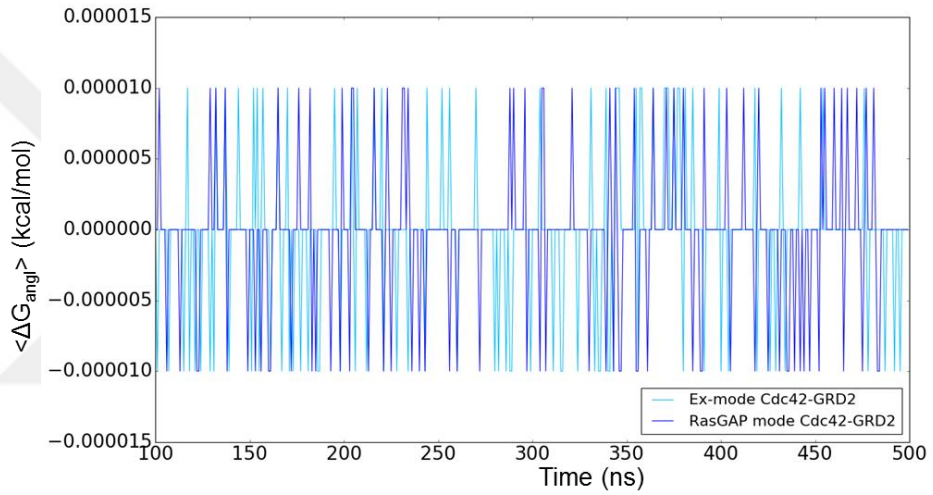
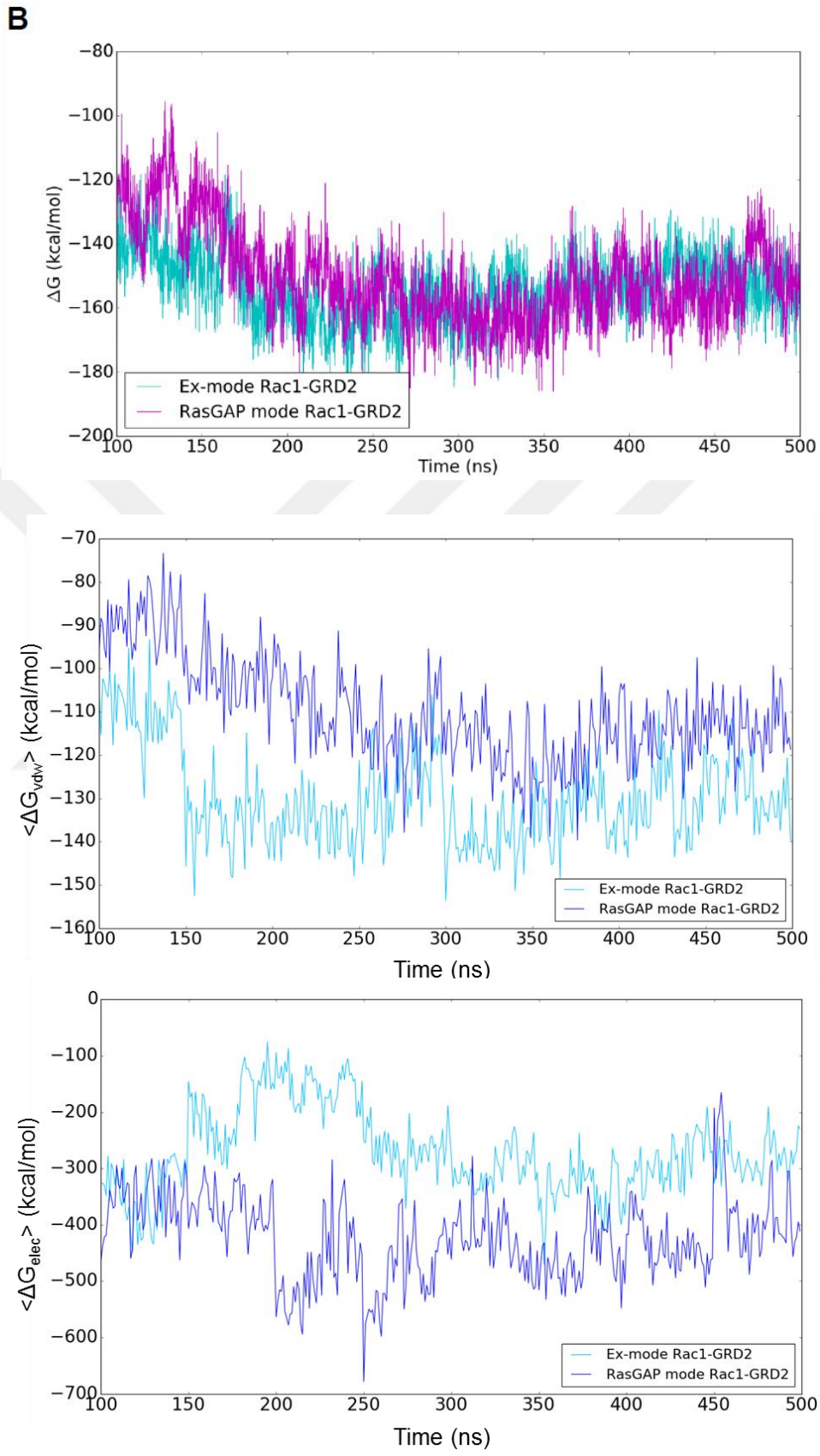


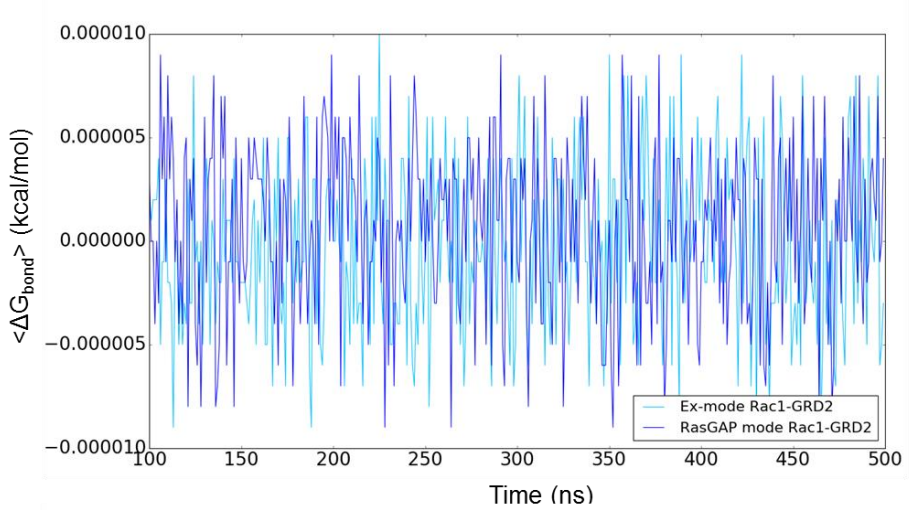
Figure B.1. RMSD plots of Cdc42/Rac1–IQGAP2 simulations. RMSD (Å) values of **A)** monomer GRD2, Ex-mode Cdc42–GRD2 complex and RasGAP Cdc42–GRD2 complex, **B)** Ex-mode Rac1–GRD2 complex and RasGAP Rac1–GRD2 complex, **C)** RasGAP Rac1–GRD2 (Y1106A) complex and RasGAP Cdc42–GRD2 (Y1106A) complex as a function of time.

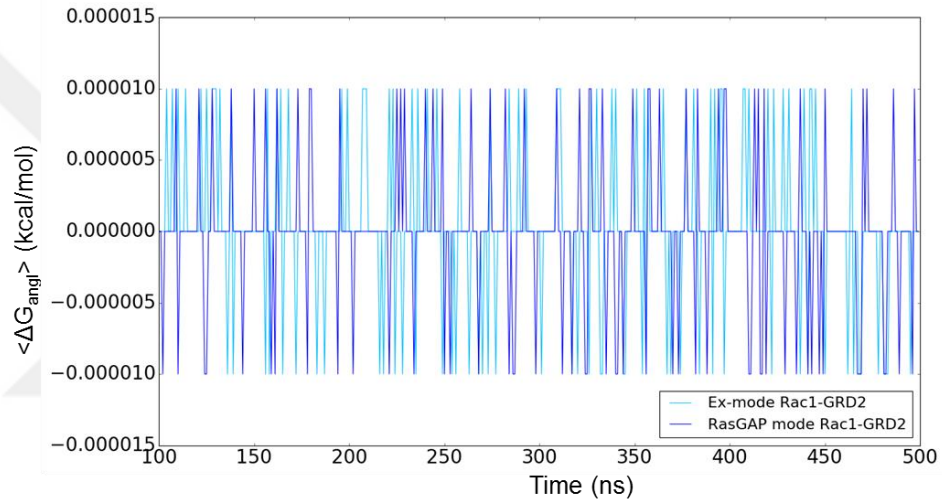

A
ΔG (kcal/mol)
Time (ns)
Ex-mode Cdc42-GRD2
RasGAP mode Cdc42-GRD2
<ΔG_vdW> (kcal/mol)
Time (ns)
Ex-mode Cdc42-GRD2
RasGAP mode Cdc42-GRD2
<ΔG_elec> (kcal/mol)
Time (ns)
Ex-mode Cdc42-GRD2
RasGAP mode Cdc42-GRD2


0.000010
0.000005
0.000000
−0.000005
−0.000010
<ΔG_bond> (kcal/mol)
100
150
200
250
300
350
400
450
500
Time (ns)
Ex-mode Cdc42-GRD2
RasGAP mode Cdc42-GRD2


0.000015
0.000010
0.000005
0.000000
−0.000005
−0.000010
−0.000015
<ΔG_angl> (kcal/mol)
100
150
200
250
300
350
400
450
500
Time (ns)
Ex-mode Cdc42-GRD2
RasGAP mode Cdc42-GRD2


B
ΔG (kcal/mol)
Time (ns)
Ex-mode Rac1-GRD2
RasGAP mode Rac1-GRD2
<ΔG_vdW> (kcal/mol)
Time (ns)
Ex-mode Rac1-GRD2
RasGAP mode Rac1-GRD2
<ΔG_elec> (kcal/mol)
Time (ns)
Ex-mode Rac1-GRD2
RasGAP mode Rac1-GRD2


0.000010
0.000005
0.000000
−0.000005
−0.000010
<ΔG_bond> (kcal/mol)
100
150
200
250
300
350
400
450
500
Time (ns)
Ex-mode Rac1-GRD2
RasGAP mode Rac1-GRD2


0.000015
0.000010
0.000005
0.000000
−0.000005
−0.000010
−0.000015
<ΔG_angl> (kcal/mol)
100
150
200
250
300
350
400
450
500
Time (ns)
Ex-mode Rac1-GRD2
RasGAP mode Rac1-GRD2

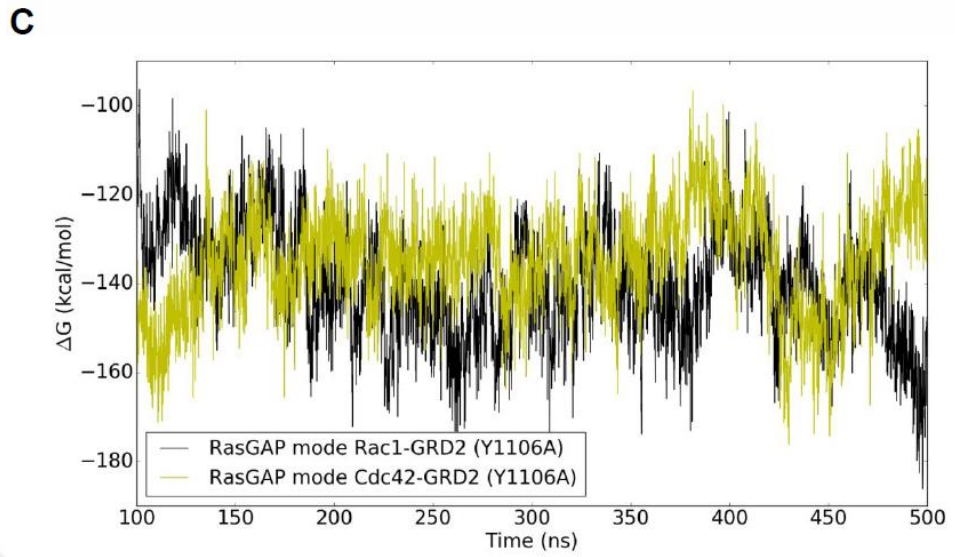
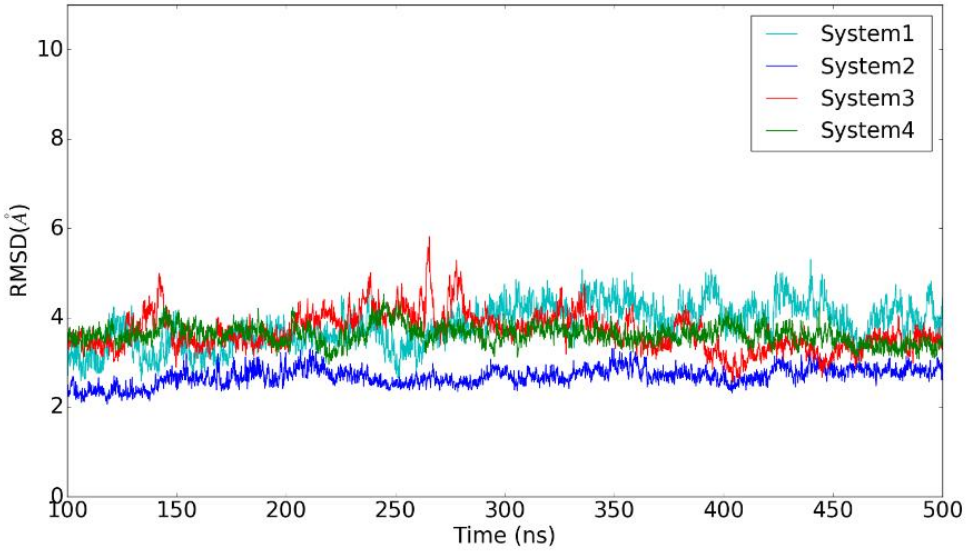
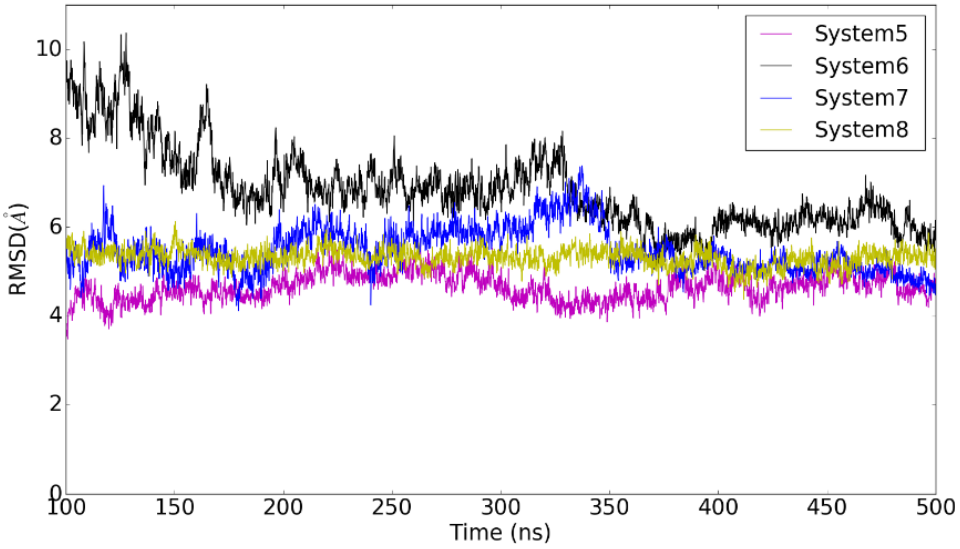


Figure B.2. Energy component plots as a function of time. A) Ex-mode Cdc42–GRD2 complex and RasGAP Cdc42–GRD2 complex, **B)** Ex-mode Rac1–GRD2 complex and RasGAP Rac1–GRD2 complex, **C)** RasGAP Rac1–GRD2 (Y1106A) complex and RasGAP Cdc42–GRD2 (Y1106A) complex. To exempt initial transients, the first 100 ns trajectories were removed.

A


System1
System2
System3
System4
RMSD(Å)
Time (ns)
0
2
4
6
8
10
100
150
200
250
300
350
400
450
500

B


System5
System6
System7
System8
RMSD(Å)
Time (ns)
0
2
4
6
8
10
100
150
200
250
300
350
400
450
500

C

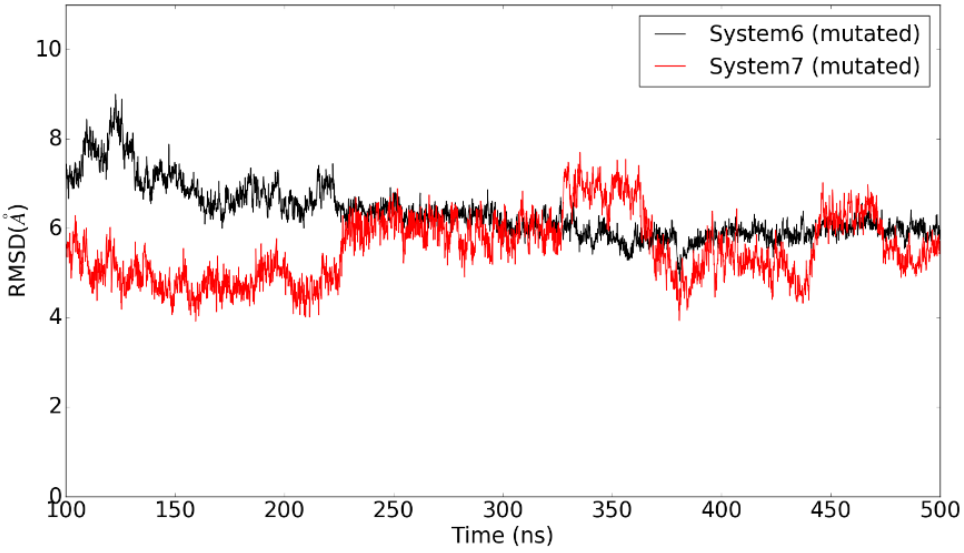
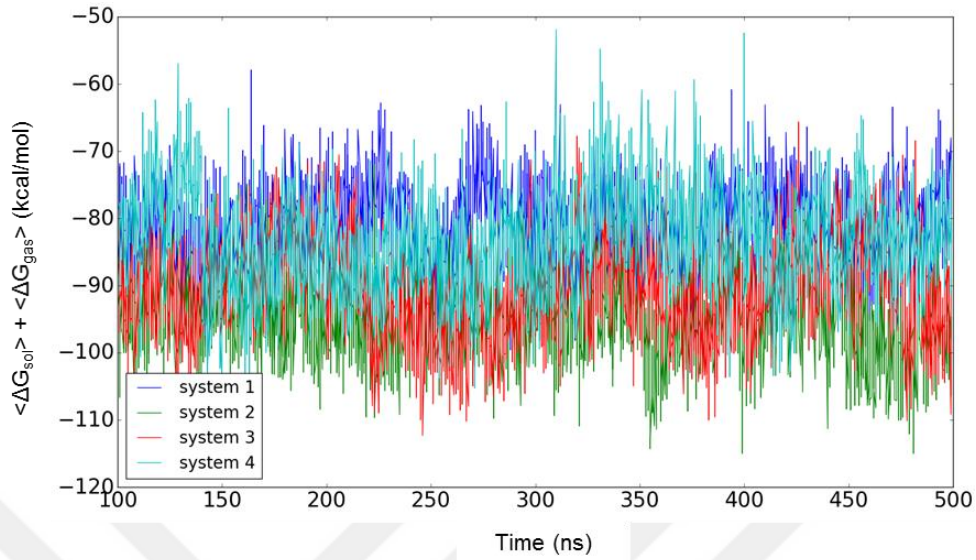
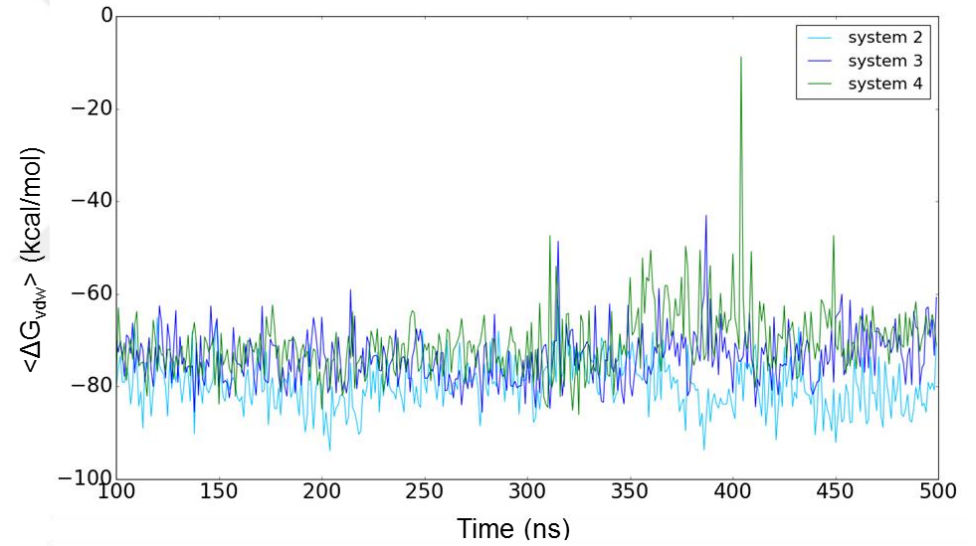
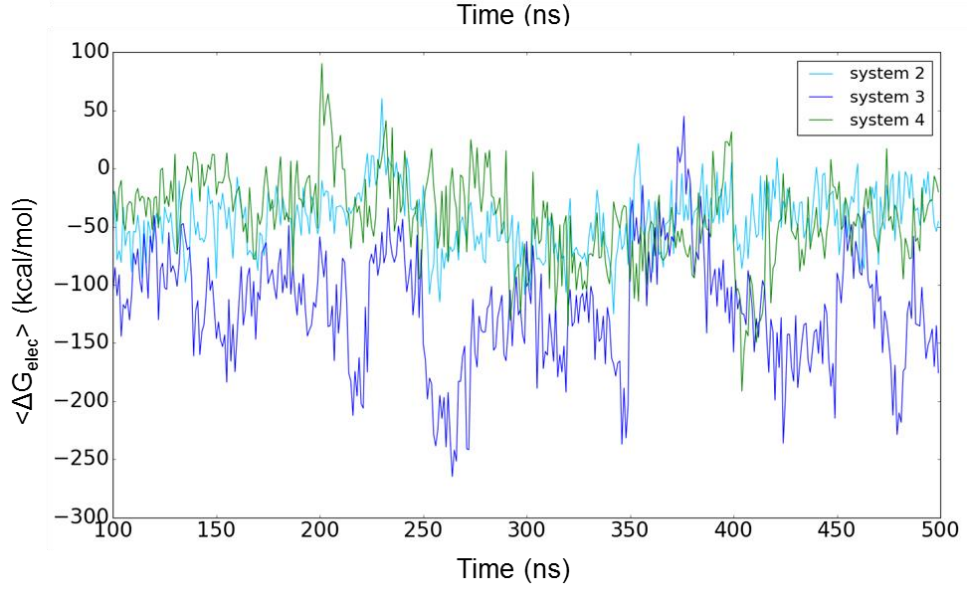


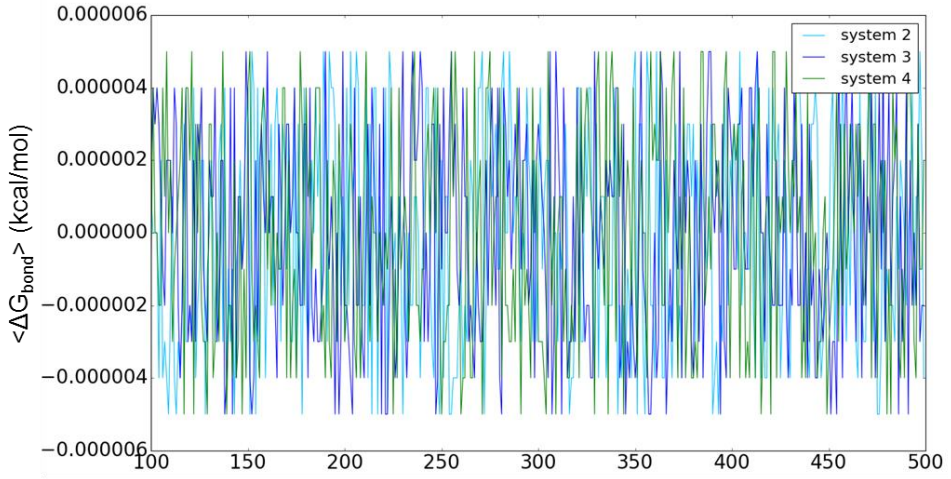
Figure B.3. RMSD plots of Arl2–PDEδ–KRas4B simulations. RMSD (Å) values of **A)** systems 1-4, **B)** systems 5-8, and **C)** system 6 and system 7 with double mutant (F94A/I98A) PDEδ as a function of time. To exempt initial transients, the first 100 ns trajectories were removed.

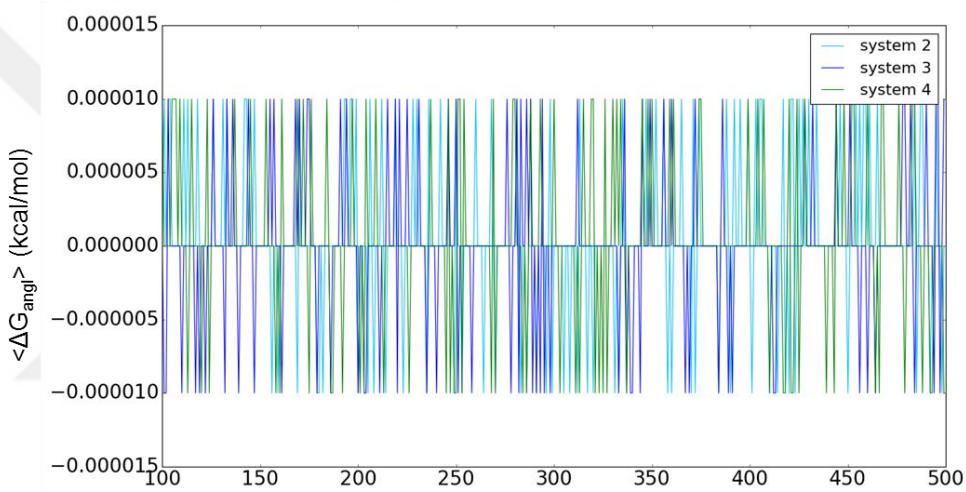
A


−50
−60
−70
−80
−90
−100
−110
−120
100
150
200
250
300
350
400
450
500
<ΔG_sol> + <ΔG_gas> (kcal/mol)
Time (ns)
system 1
system 2
system 3
system 4

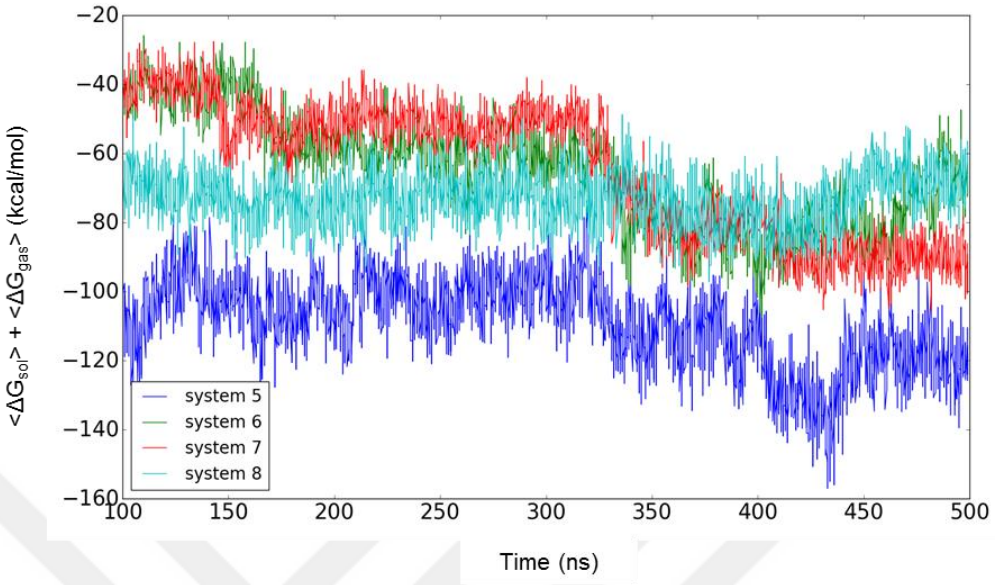

0
−20
−40
−60
−80
−100
100
150
200
250
300
350
400
450
500
<ΔG_vdW> (kcal/mol)
Time (ns)
system 2
system 3
system 4

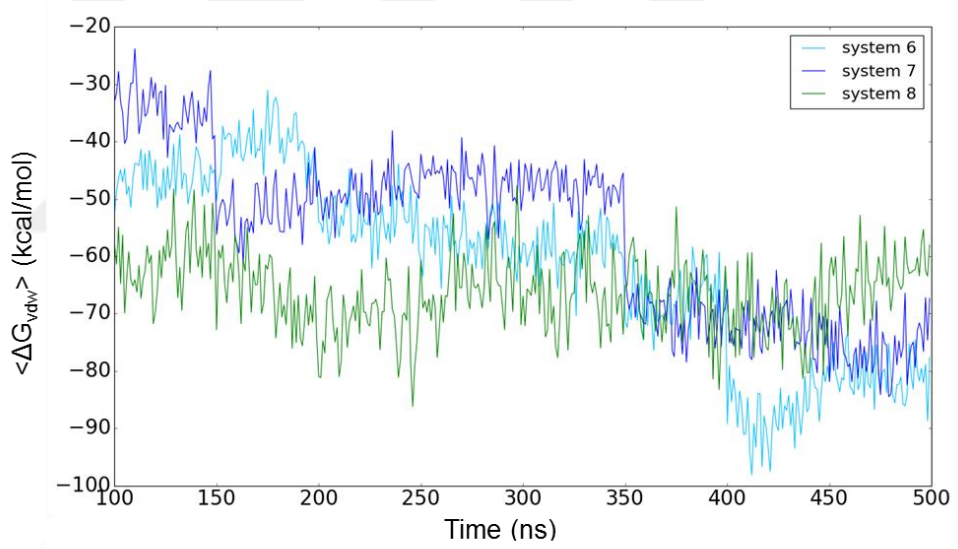

100
50
0
−50
−100
−150
−200
−250
−300
100
150
200
250
300
350
400
450
500
<ΔG_elec> (kcal/mol)
Time (ns)
system 2
system 3
system 4

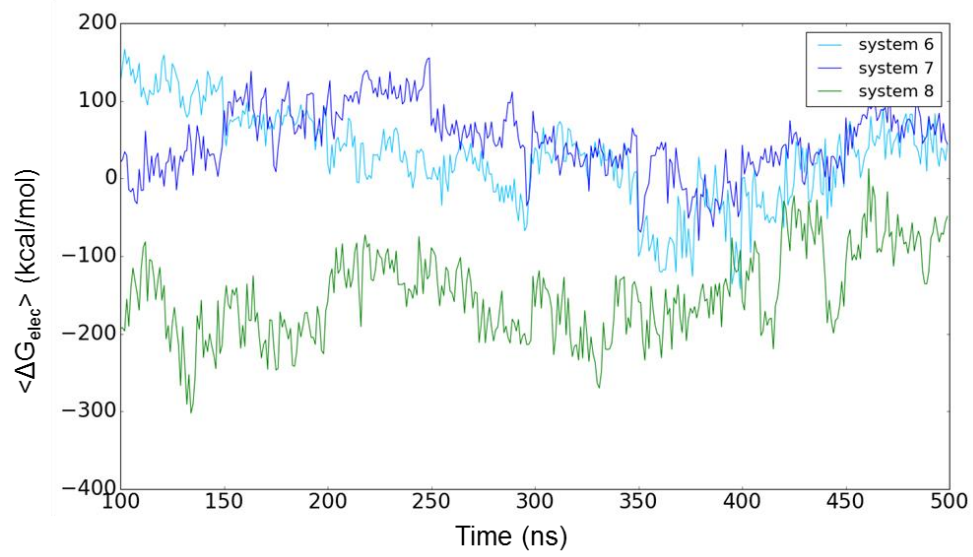

system 2
system 3
system 4
<ΔG_bond> (kcal/mol)
0.000006
0.000004
0.000002
0.000000
−0.000002
−0.000004
−0.000006
100
150
200
250
300
350
400
450
500

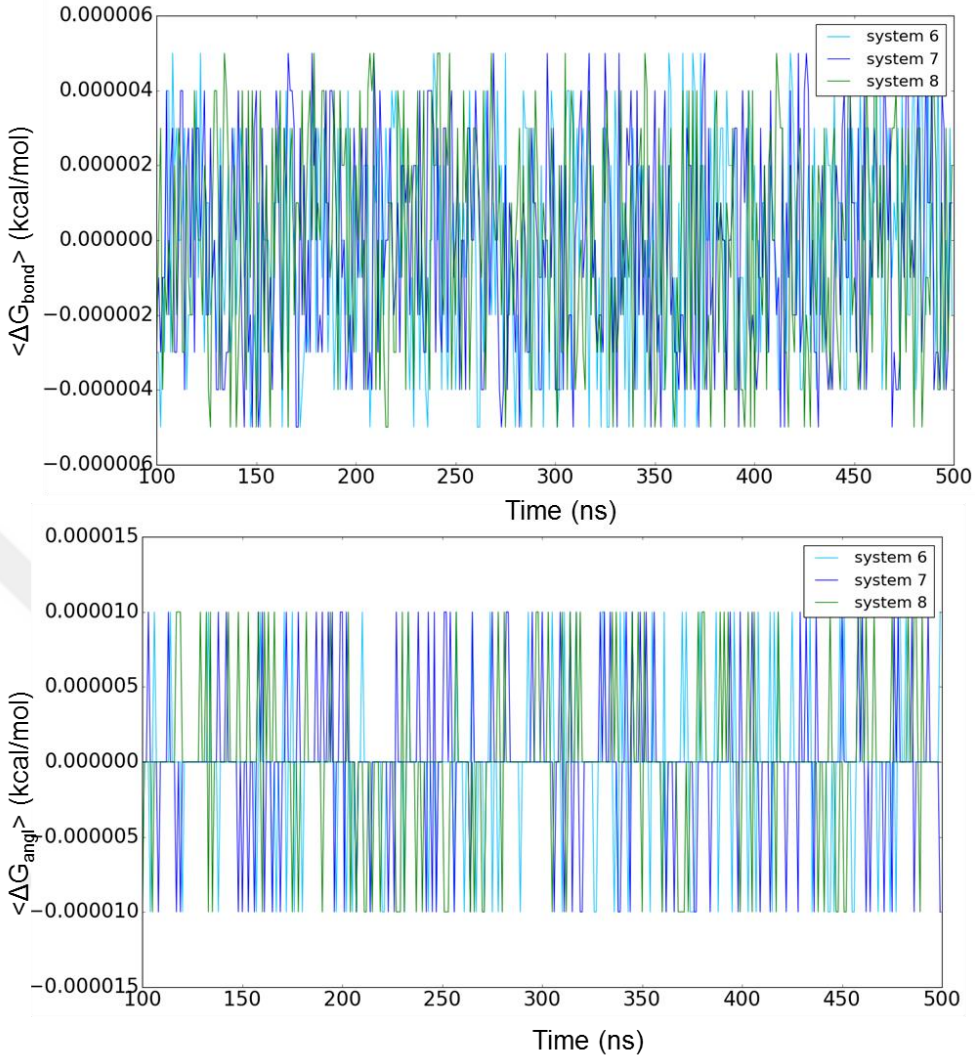

system 2
system 3
system 4
<ΔG_angl> (kcal/mol)
0.000015
0.000010
0.000005
0.000000
−0.000005
−0.000010
−0.000015
100
150
200
250
300
350
400
450
500

B


<ΔG_sol> + <ΔG_gas> (kcal/mol)
Time (ns)
system 5
system 6
system 7
system 8


<ΔG_vdw> (kcal/mol)
Time (ns)
system 6
system 7
system 8


<ΔG_elec> (kcal/mol)
Time (ns)
system 6
system 7
system 8


0.000006
0.000004
0.000002
0.000000
−0.000002
−0.000004
−0.000006
<ΔG_bond> (kcal/mol)
100
150
200
250
300
350
400
450
500
Time (ns)
system 6
system 7
system 8
0.000015
0.000010
0.000005
0.000000
0.000005
−0.000010
−0.000015
<ΔG_angl> (kcal/mol)
100
150
200
250
300
350
400
450
500
Time (ns)
system 6
system 7
system 8

C

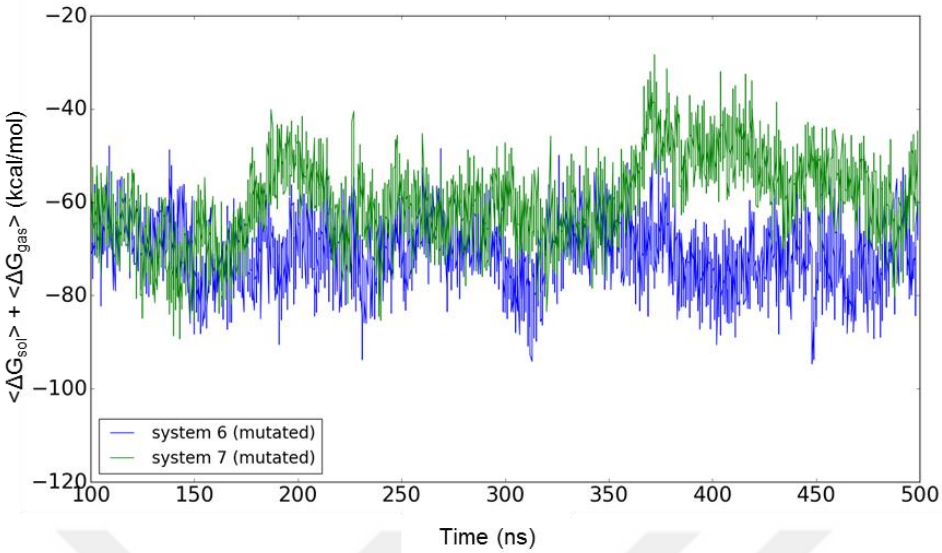
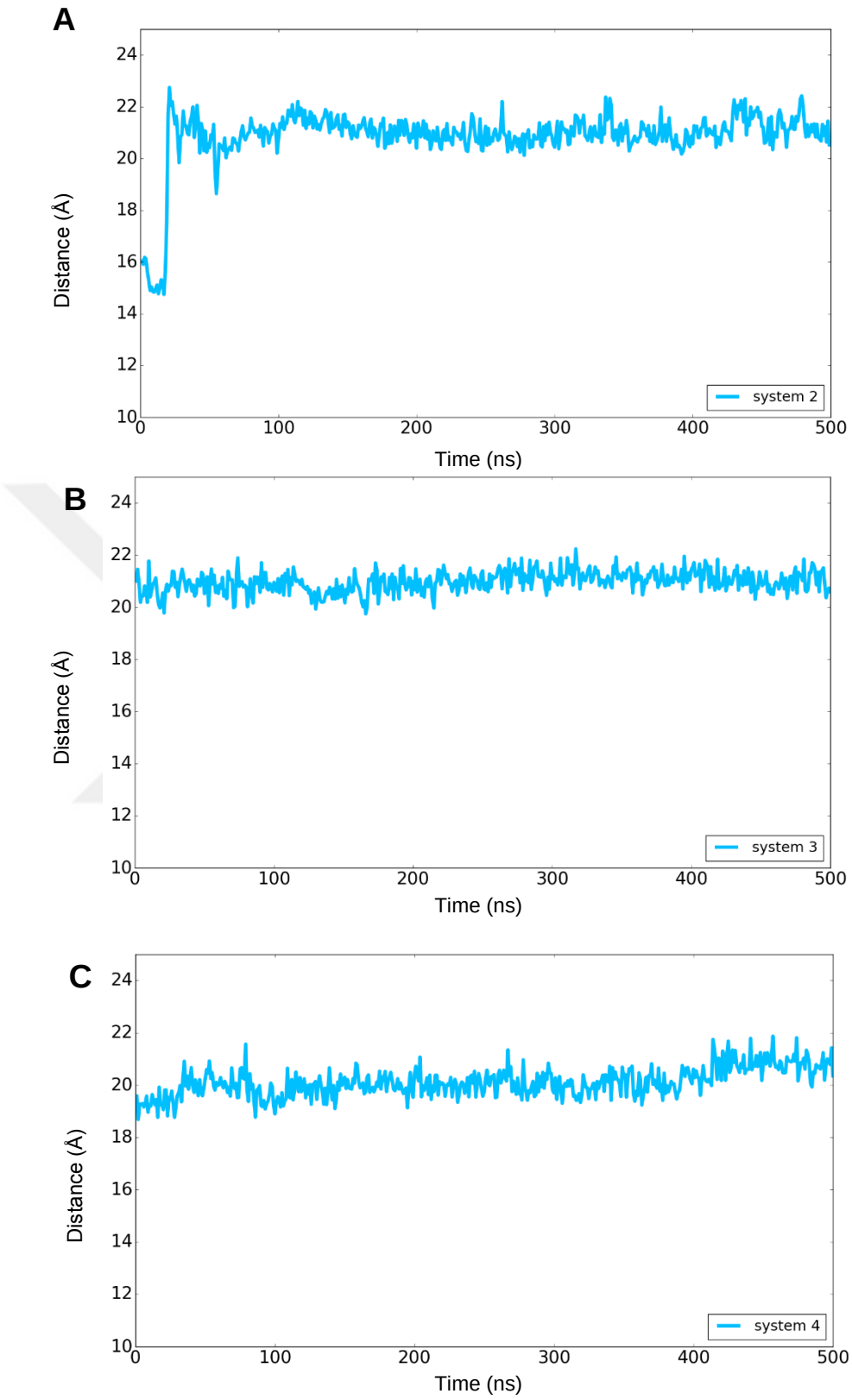
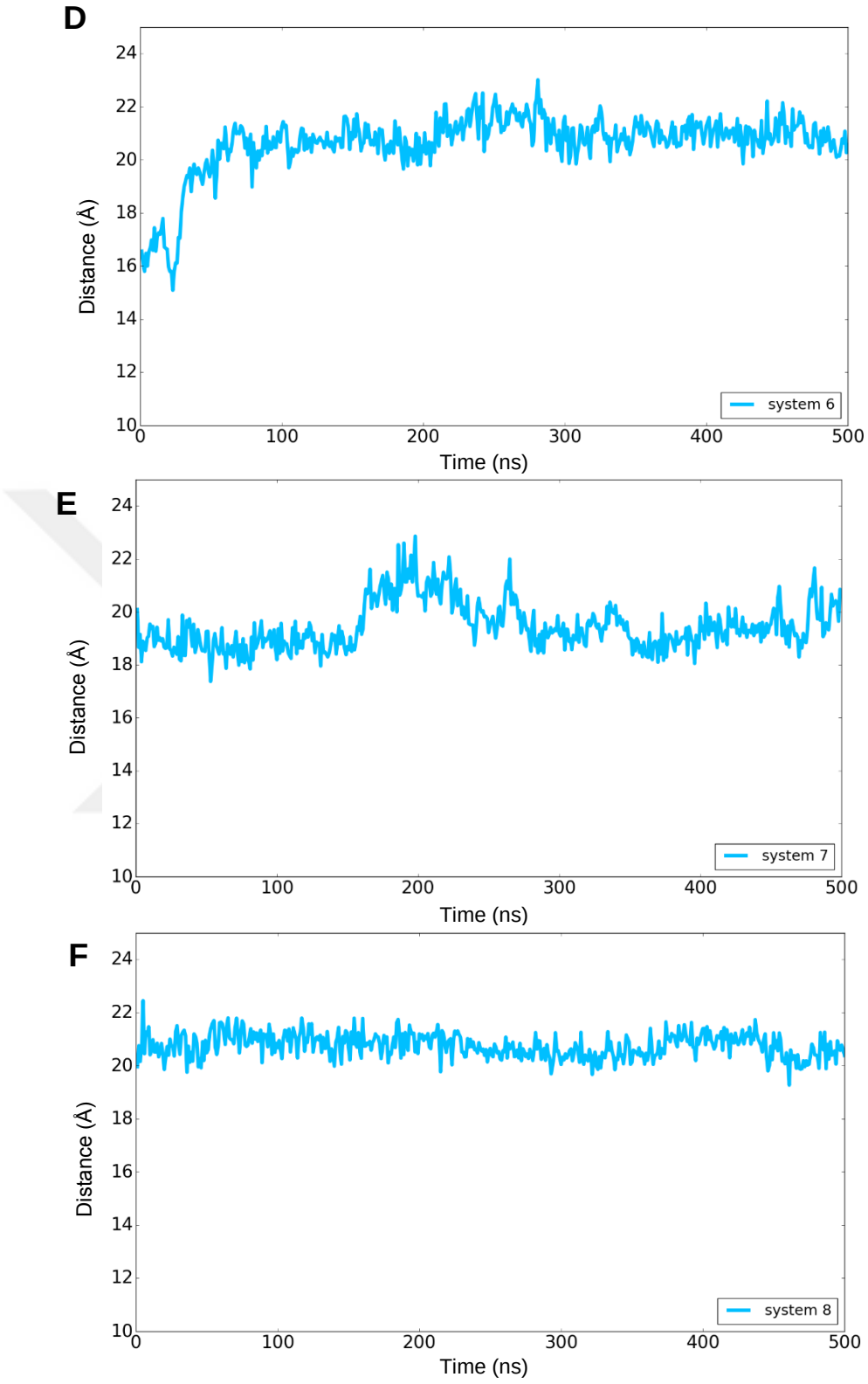


Figure B.4. Energy component plots as a function of time for **A)** Systems 1-4, **B)** systems 5-8, and **C)** system 6 and system 7 with double mutated (F94A/I98A) PDEδ. To exempt initial transients, the first 100 ns trajectories were removed. Energy values are calculated for interaction between Arl2 and PDEδ.


A
Distance (Å)
Time (ns)
system 2
B
Distance (Å)
Time (ns)
system 3
C
Distance (Å)
Time (ns)
system 4


D
Distance (Å)
Time (ns)
system 6
E
Distance (Å)
Time (ns)
system 7
F
Distance (Å)
Time (ns)
system 8

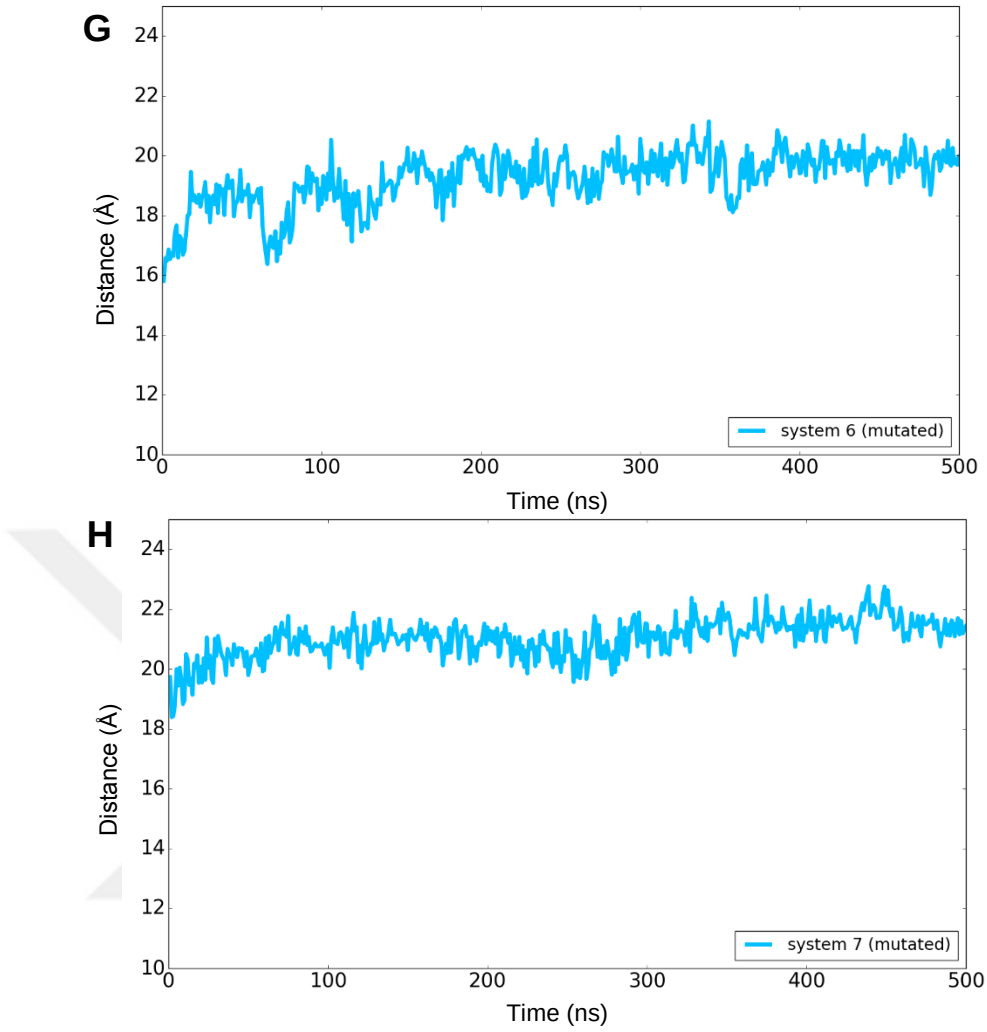


Figure B.5. Distances between Cα atoms of Phe133 (PDEδ) and farnesyl/geranylgeranyl groups in **A-C)** systems 2-4, **D-F)** systems 6-8, **G-H)** mutated system 6 and mutated system 7 throughout the 500 ns simulations.

APPENDIX C

Table C.1. Disease-causing variations from multiple data sets matching with SKEMPI 2.0 variations

PDB	Chain	GeneName	VarID	Disease	PDB Location	SKEMPI Variation	Hot Spot	Hot Region
1A22	A	GH1	COSM561947	Lung cancer	18	H18A	NH	NHR
1A22	A	GH1	CM104401	Growth hormone deficiency	25	F25A	NH	NHR
1A22	A	GH1	COSM3520868	Skin cancer	48	P48A	NH	NHR
1A22	A	GH1	COSM3520867	skin	56	E56A	NH	NHR
1A22	A	GH1	COSM4068663	stomach	64	R64A	H	NHR
1A22	A	GH1	COSM3958649	lung	68	Q68A	NH	NHR
1A22	A	GH1	VAR_015814	Growth hormone deficiency	175	T175A	H	NHR
1A22	A	GH1	COSM4686902	Large intestine	178	R178A	NH	NHR
1A22	A	GH1	COSM4936751	liver	178	R178A	NH	NHR
1A22	A	GH1	CM040745	Growth hormone deficiency	179	I179A	H	NHR
1A22	A	GH1	CM095352	Growth hormone deficiency	182	C182A	NH	NHR
1A22	B	GHR	121909358	Growth hormone receptor	243	R243A	H	NHR
1A22	B	GHR	VAR_002708	Growth hormone insensitivity	244	E244A	NH	NHR
1A22	B	GHR	VAR_002709	Laron syndrome	271	R271A	H	NHR
1A22	B	GHR	COSM3776716	Urinary tract cancer	271	R271A	H	NHR
1A22	B	GHR	COSM738739	Lung cancer	275	E275A	NH	NHR

1A22	B	GHR	CM104107	Laron syndrome	304	W304F	H	HR
1A22	B	GHR	CM950573	Laron syndrome	322	C322A	NH	NHR
1A22	B	GHR	COSM281363	Large intestine cancer	364	D364A	NH	NHR
1A22	B	GHR	CM930315	Laron syndrome	417	R417A	NH	NHR
1A4Y	B	ANG	VAR_044147, rs121909535	Amyotrophic lateral sclerosis	12	Q12A	NH	NHR
1A4Y	B	ANG	VAR_044151	Amyotrophic lateral sclerosis	31	R31A	H	NHR
1A4Y	B	ANG	VAR_044153, rs121909540	Amyotrophic lateral sclerosis	40	K40G, K40Q	H	NHR
1A4Y	B	ANG	CM126054	Amyotrophic lateral sclerosis	89	W89A	NH	NHR
1A4Y	B	ANG	COSM3494733	Skin cancer	114	H114A	NH	NHR
1A4Y	B	ANG	VAR_044157	Amyotrophic lateral sclerosis	114	H114A	NH	NHR
1DAN	T	F3	COSM4938858	Liver	68	K68A	NH	NHR
1E+96	A	RAC1	COSM3640040	Skin cancer	33	I33N	NH	NHR
1GC1	C	CD4	COSM3464492	Skin cancer	27	H27A	NH	NHR
1GC1	C	CD4	COSM468839	Kidney cancer	52	N52A	NH	NHR
1H9D	B	CBFB	COSM1189291	Lung cancer	4	V4A	NH	NHR
1H9D	B	CBFB	COSM1478979	Breast cancer	104	N104A	H	HR
1HE8	A	PIK3CG	COSM3631735	Skin cancer	223	K223V	NH	NHR
1IAR	A	IL4	COSM4187248	kidney	78	Q78A,Q78E	NH	NHR
2J0T	D	TIMP1	COSM189914	Large intestine	70	C70S	H	HR

BIBLIOGRAPHY

1. LeCour, L., Jr., et al., *The Structural Basis for Cdc42-Induced Dimerization of IQGAPs.* Structure, 2016. **24**(9): p. 1499-508.
2. Ren, J.G., et al., *Self-association of IQGAP1: characterization and functional sequelae.* J Biol Chem, 2005. **280**(41): p. 34548-57.
3. Lu, S., et al., *Drugging Ras GTPase: a comprehensive mechanistic and signaling structural view.* Chem Soc Rev, 2016. **45**(18): p. 4929-52.
4. Lu, S.Y., et al., *GTP Binding and Oncogenic Mutations May Attenuate Hypervariable Region (HVR)-Catalytic Domain Interactions in Small GTPase K-Ras4B, Exposing the Effector Binding Site.* J Biol Chem, 2015. **290**(48): p. 28887-28900.
5. Hanzal-Bayer, M., et al., *The complex of Arl2-GTP and PDE delta: from structure to function.* EMBO J, 2002. **21**(9): p. 2095-106.
6. Zhang, H.B., et al., *Photoreceptor cGMP phosphodiesterase delta subunit (PDE delta) functions as a prenyl-binding protein.* J Biol Chem, 2004. **279**(1): p. 407-413.
7. Nussinov, R., et al., *A New View of Ras Isoforms in Cancers.* Cancer Res, 2016. **76**(1): p. 18-23.
8. Chakrabarti, M., H. Jang, and R. Nussinov, *Comparison of the Conformations of KRAS Isoforms, K-Ras4A and K-Ras4B, Points to Similarities and Significant Differences.* J Phys Chem B, 2016. **120**(4): p. 667-79.
9. Muratcioglu, S., et al., *PDE delta Binding to Ras Isoforms Provides a Route to Proper Membrane Localization.* J Phys Chem B, 2017. **121**(24): p. 5917-5927.
10. Dharmaiah, S., et al., *Structural basis of recognition of farnesylated and methylated KRAS4b by PDEdelta.* Proc Natl Acad Sci USA, 2016. **113**(44): p. E6766-E6775.
11. Linari, M., M. Hanzal-Bayer, and J. Becker, *The delta subunit of rod specific cyclic GMP phosphodiesterase, PDE delta, interacts with the Arf-like protein Arl3 in a GTP specific manner.* FEBS Lett, 1999. **458**(1): p. 55-9.
12. Veltel, S., et al., *Specificity of Arl2/Arl3 signaling is mediated by a ternary Arl3-effector-GAP complex.* Febs Letters, 2008. **582**(17): p. 2501-2507.

13. Martin-Gago, P., et al., *A PDE6delta-KRas Inhibitor Chemotype with up to Seven H-Bonds and Picomolar Affinity that Prevents Efficient Inhibitor Release by Arl2.* Angew Chem Int Ed Engl, 2017. **56**(9): p. 2423-2428.
14. Scheffzek, K. and M.R. Ahmadian, *GTPase activating proteins: structural and functional insights 18 years after discovery.* Cell Mol Life Sci, 2005. **62**(24): p. 3014-38.
15. Colicelli, J., *Human RAS superfamily proteins and related GTPases.* Sci STKE, 2004. **2004**(250): p. RE13.
16. Vetter, I.R. and A. Wittinghofer, *The guanine nucleotide-binding switch in three dimensions.* Science, 2001. **294**(5545): p. 1299-304.
17. Rathinam, R., A. Berrier, and S.K. Alahari, *Role of Rho GTPases and their regulators in cancer progression.* Front Biosci (Landmark Ed), 2011. **16**: p. 2561-71.
18. Sahai, E. and C.J. Marshall, *RHO-GTPases and cancer.* Nat Rev Cancer, 2002. **2**(2): p. 133-42.
19. Barbacid, M., *ras oncogenes: their role in neoplasia.* Eur J Clin Invest, 1990. **20**(3): p. 225-35.
20. Casteel, D.E., et al., *Rho isoform-specific interaction with IQGAP1 promotes breast cancer cell proliferation and migration.* J Biol Chem, 2012. **287**(45): p. 38367-78.
21. Gideon, P., et al., *Mutational and kinetic analyses of the GTPase-activating protein (GAP)-p21 interaction: the C-terminal domain of GAP is not sufficient for full activity.* Mol Cell Biol, 1992. **12**(5): p. 2050-6.
22. Etienne-Manneville, S. and A. Hall, *Rho GTPases in cell biology.* Nature, 2002. **420**(6916): p. 629-35.
23. Zandvakili, I., et al., *Rho GTPases: Anti- or pro-neoplastic targets?* Oncogene, 2016.
24. Lam, B.D. and P.L. Hordijk, *The Rac1 hypervariable region in targeting and signaling: a tail of many stories.* Small GTPases, 2013. **4**(2): p. 78-89.
25. Porfiri, E., et al., *Prenylation of Ras Proteins Is Required for Efficient Hsos1-Promoted Guanine-Nucleotide Exchange.* Journal of Biological Chemistry, 1994. **269**(36): p. 22672-22677.

26. Williams, J.G., et al., *Elucidation of binding determinants and functional consequences of Ras/Raf-cysteine-rich domain interactions.* J Biol Chem, 2000. **275**(29): p. 22172-9.
27. Bourne, H.R., D.A. Sanders, and F. Mccormick, *The Gtpase Superfamily - Conserved Structure and Molecular Mechanism.* Nature, 1991. **349**(6305): p. 117-127.
28. Rojas, A.M., et al., *The Ras protein superfamily: evolutionary tree and role of conserved amino acids.* J Cell Biol, 2012. **196**(2): p. 189-201.
29. Boguski, M.S. and F. McCormick, *Proteins regulating Ras and its relatives.* Nature, 1993. **366**(6456): p. 643-54.
30. Ridley, A.J., *Rho family proteins: coordinating cell responses.* Trends in Cell Biology, 2001. **11**(12): p. 471-477.
31. Tapon, N. and A. Hall, *Rho, Rac and Cdc42 GTPases regulate the organization of the actin cytoskeleton.* Curr Opin Cell Biol, 1997. **9**(1): p. 86-92.
32. Owen, D., et al., *The IQGAP1-Rac1 and IQGAP1-Cdc42 interactions: interfaces differ between the complexes.* J Biol Chem, 2008. **283**(3): p. 1692-704.
33. Zong, H., K. Kaibuchi, and L.A. Quilliam, *The insert region of RhoA is essential for Rho kinase activation and cellular transformation.* Molecular and Cellular Biology, 2001. **21**(16): p. 5287-5298.
34. Lapouge, K., et al., *Structure of the TPR domain of p67phox in complex with Rac.GTP.* Mol Cell, 2000. **6**(4): p. 899-907.
35. Watanabe, T., S. Wang, and K. Kaibuchi, *IQGAPs as Key Regulators of Actin-cytoskeleton Dynamics.* Cell Struct Funct, 2015. **40**(2): p. 69-77.
36. Shinjo, K., et al., *Molecular-Cloning of the Gene for the Human Placental Gtp-Binding Protein-Gp (G25k) - Identification of This Gtp-Binding Protein as the Human Homolog of the Yeast Cell-Division-Cycle Protein-Cdc42.* Proceedings of the National Academy of Sciences of the United States of America, 1990. **87**(24): p. 9853-9857.
37. Didsbury, J., et al., *Rac, a Novel Ras-Related Family of Proteins That Are Botulinum Toxin Substrates.* Journal of Biological Chemistry, 1989. **264**(28): p. 16378-16382.

38. Hall, A., *Rho GTPases and the actin cytoskeleton.* Science, 1998. **279**(5350): p. 509-514.
39. Stowers, L., et al., *Regulation of the polarization of T cells toward antigen-presenting cells by Ras-related GTPase CDC42.* Proc Natl Acad Sci U S A, 1995. **92**(11): p. 5027-31.
40. Mataraza, J.M., et al., *IQGAP1 promotes cell motility and invasion.* J Biol Chem, 2003. **278**(42): p. 41237-45.
41. Jaffe, A.B. and A. Hall, *Rho GTPases: biochemistry and biology.* Annu Rev Cell Dev Biol, 2005. **21**: p. 247-69.
42. Aspenstrom, P., U. Lindberg, and A. Hall, *Two GTPases, Cdc42 and Rac, bind directly to a protein implicated in the immunodeficiency disorder Wiskott-Aldrich syndrome.* Curr Biol, 1996. **6**(1): p. 70-5.
43. Bishop, A.L. and A. Hall, *Rho GTPases and their effector proteins.* Biochemical Journal, 2000. **348**: p. 241-255.
44. Miki, H., et al., *Induction of filopodium formation by a WASP-related actin-depolymerizing protein N-WASP.* Nature, 1998. **391**(6662): p. 93-6.
45. Machesky, L.M. and K.L. Gould, *The Arp2/3 complex: a multifunctional actin organizer.* Curr Opin Cell Biol, 1999. **11**(1): p. 117-21.
46. Weissbach, L., et al., *Identification of a human rasGAP-related protein containing calmodulin-binding motifs.* J Biol Chem, 1994. **269**(32): p. 20517-21.
47. Briggs, M.W. and D.B. Sacks, *IQGAP proteins are integral components of cytoskeletal regulation.* EMBO Rep, 2003. **4**(6): p. 571-4.
48. Tanos, B.E., C. Yeaman, and E. Rodriguez-Boulan, *An emerging role for IQGAP1 in tight junction control.* Small GTPases, 2016: p. 1-9.
49. Swart-Mataraza, J.M., Z.G. Li, and D.B. Sacks, *IQGAP1 is a component of Cdc42 signaling to the cytoskeleton.* Journal of Biological Chemistry, 2002. **277**(27): p. 24753-24763.
50. Kuroda, S., et al., *Role of IQGAP1, a target of the small GTPases Cdc42 and Rac1, in regulation of E-cadherin- mediated cell-cell adhesion.* Science, 1998. **281**(5378): p. 832-5.

51. Brown, M.D. and D.B. Sacks, *IQGAP1 in cellular signaling: bridging the GAP.* Trends Cell Biol, 2006. **16**(5): p. 242-9.
52. Cheung, K.L., et al., *The Ras GTPase-activating-like protein IQGAP1 mediates Nrf2 protein activation via the mitogen-activated protein kinase/extracellular signal-regulated kinase (ERK) kinase (MEK)-ERK pathway.* J Biol Chem, 2013. **288**(31): p. 22378-86.
53. Jameson, K.L., et al., *IQGAP1 scaffold-kinase interaction blockade selectively targets RAS-MAP kinase-driven tumors.* Nature Medicine, 2013. **19**(5): p. 626-630.
54. Smith, J.M., A.C. Hedman, and D.B. Sacks, *IQGAPs choreograph cellular signaling from the membrane to the nucleus.* Trends Cell Biol, 2015. **25**(3): p. 171-84.
55. Brill, S., et al., *The Ras GTPase-activating-protein-related human protein IQGAP2 harbors a potential actin binding domain and interacts with calmodulin and Rho family GTPases.* Mol Cell Biol, 1996. **16**(9): p. 4869-78.
56. Roy, M., Z. Li, and D.B. Sacks, *IQGAP1 binds ERK2 and modulates its activity.* J Biol Chem, 2004. **279**(17): p. 17329-37.
57. Ho, Y.D., et al., *IQGAP1 integrates Ca2+/calmodulin and Cdc42 signaling.* J Biol Chem, 1999. **274**(1): p. 464-70.
58. Hart, M.J., et al., *IQGAP1, a calmodulin-binding protein with a rasGAP-related domain, is a potential effector for cdc42Hs.* EMBO J, 1996. **15**(12): p. 2997-3005.
59. Marianayagam, N.J., M. Sunde, and J.M. Matthews, *The power of two: protein dimerization in biology.* Trends Biochem Sci, 2004. **29**(11): p. 618-25.
60. Jadeski, L., et al., *IQGAP1 stimulates proliferation and enhances tumorigenesis of human breast epithelial cells.* J Biol Chem, 2008. **283**(2): p. 1008-17.
61. Hedman, A.C., J.M. Smith, and D.B. Sacks, *The biology of IQGAP proteins: beyond the cytoskeleton.* EMBO Rep, 2015. **16**(4): p. 427-46.
62. Bashour, A.M., et al., *IQGAP1, a Rac- and Cdc42-binding protein, directly binds and cross-links microfilaments.* Journal of Cell Biology, 1997. **137**(7): p. 1555-1566.

63. Bensenor, L.B., et al., *IQGAP1 regulates cell motility by linking growth factor signaling to actin assembly.* J Cell Sci, 2007. **120**(Pt 4): p. 658-69.
64. Le Clainche, C., et al., *IQGAP1 stimulates actin assembly through the N-WASP-Arp2/3 pathway.* Journal of Biological Chemistry, 2007. **282**(1): p. 426-435.
65. Heasman, S.J. and A.J. Ridley, *Mammalian Rho GTPases: new insights into their functions from in vivo studies.* Nature Reviews Molecular Cell Biology, 2008. **9**(9): p. 690-701.
66. Harvey, J.J., *An Unidentified Virus Which Causes the Rapid Production of Tumours in Mice.* Nature, 1964. **204**: p. 1104-5.
67. Kirsten, W.H. and L.A. Mayer, *Morphologic responses to a murine erythroblastosis virus.* J Natl Cancer Inst, 1967. **39**(2): p. 311-35.
68. Scolnick, E.M., et al., *Studies on the nucleic acid sequences of Kirsten sarcoma virus: a model for formation of a mammalian RNA-containing sarcoma virus.* J Virol, 1973. **12**(3): p. 458-63.
69. Shih, T.Y., et al., *Identification of a sarcoma virus-coded phosphoprotein in nonproducer cells transformed by Kirsten or Harvey murine sarcoma virus.* Virology, 1979. **96**(1): p. 64-79.
70. Scolnick, E.M. and W.P. Parks, *Harvey sarcoma virus: a second murine type C sarcoma virus with rat genetic information.* J Virol, 1974. **13**(6): p. 1211-9.
71. Scolnick, E.M., A.G. Papageorge, and T.Y. Shih, *Guanine nucleotide-binding activity as an assay for src protein of rat-derived murine sarcoma viruses.* Proc Natl Acad Sci U S A, 1979. **76**(10): p. 5355-9.
72. Willingham, M.C., et al., *Localization of the src gene product of the Harvey strain of MSV to plasma membrane of transformed cells by electron microscopic immunocytochemistry.* Cell, 1980. **19**(4): p. 1005-14.
73. O'Toole, C.M., et al., *Identity of some human bladder cancer cell lines.* Nature, 1983. **301**(5899): p. 429-30.
74. Capon, D.J., et al., *Activation of Ki-ras2 gene in human colon and lung carcinomas by two different point mutations.* Nature, 1983. **304**(5926): p. 507-13.
75. Shimizu, K., et al., *Three human transforming genes are related to the viral ras oncogenes.* Proc Natl Acad Sci U S A, 1983. **80**(8): p. 2112-6.

76. Hall, A., et al., *Identification of transforming gene in two human sarcoma cell lines as a new member of the ras gene family located on chromosome 1.* Nature, 1983. **303**(5916): p. 396-400.
77. Pulciani, S., et al., *Oncogenes in solid human tumours.* Nature, 1982. **300**(5892): p. 539-42.
78. Campbell, S.L., et al., *Increasing complexity of Ras signaling.* Oncogene, 1998. **17**(11 Reviews): p. 1395-413.
79. Reuther, G.W. and C.J. Der, *The Ras branch of small GTPases: Ras family members don't fall far from the tree.* Curr Opin Cell Biol, 2000. **12**(2): p. 157-65.
80. Downward, J., *The ras superfamily of small GTP-binding proteins.* Trends Biochem Sci, 1990. **15**(12): p. 469-72.
81. Grand, R.J. and D. Owen, *The biochemistry of ras p21.* Biochem J, 1991. **279 (Pt 3)**: p. 609-31.
82. Bourne, H.R., D.A. Sanders, and F. McCormick, *The GTPase superfamily: conserved structure and molecular mechanism.* Nature, 1991. **349**(6305): p. 117-27.
83. Wittinghofer, A. and E.F. Pai, *The structure of Ras protein: a model for a universal molecular switch.* Trends Biochem Sci, 1991. **16**(10): p. 382-7.
84. Wittinghofer, A. and I.R. Vetter, *Structure-function relationships of the G domain, a canonical switch motif.* Annu Rev Biochem, 2011. **80**: p. 943-71.
85. Lowy, D.R., et al., *Regulation of p21ras activity.* Trends Genet, 1991. **7**(11-12): p. 346-51.
86. Lamontanara, A.J., et al., *The SH2 domain of Abl kinases regulates kinase autophosphorylation by controlling activation loop accessibility.* Nat Commun, 2014. **5**: p. 5470.
87. Cherfils, J. and M. Zeghouf, *Regulation of small GTPases by GEFs, GAPs, and GDIs.* Physiol Rev, 2013. **93**(1): p. 269-309.
88. Geyer, M. and A. Wittinghofer, *GEFs, GAPs, GDIs and effectors: taking a closer (3D) look at the regulation of Ras-related GTP-binding proteins.* Curr Opin Struct Biol, 1997. **7**(6): p. 786-92.

89. Sprang, S.R., *G proteins, effectors and GAPs: structure and mechanism.* Curr Opin Struct Biol, 1997. **7**(6): p. 849-56.
90. Downward, J., *Ras signalling and apoptosis.* Curr Opin Genet Dev, 1998. **8**(1): p. 49-54.
91. Khosravi-Far, R., et al., *Increasing complexity of Ras signal transduction: involvement of Rho family proteins.* Adv Cancer Res, 1998. **72**: p. 57-107.
92. Muratcioglu, S., et al., *PDEdelta Binding to Ras Isoforms Provides a Route to Proper Membrane Localization.* J Phys Chem B, 2017. **121**(24): p. 5917-5927.
93. de Vos, A.M., et al., *Three-dimensional structure of an oncogene protein: catalytic domain of human c-H-ras p21.* Science, 1988. **239**(4842): p. 888-93.
94. Pai, E.F., et al., *Refined crystal structure of the triphosphate conformation of H-ras p21 at 1.35 A resolution: implications for the mechanism of GTP hydrolysis.* EMBO J, 1990. **9**(8): p. 2351-9.
95. Milburn, M.V., et al., *Molecular switch for signal transduction: structural differences between active and inactive forms of protooncogenic ras proteins.* Science, 1990. **247**(4945): p. 939-45.
96. Muegge, I., et al., *Electrostatic control of GTP and GDP binding in the oncoprotein p21ras.* Structure, 1996. **4**(4): p. 475-89.
97. Gorfe, A.A., B.J. Grant, and J.A. McCammon, *Mapping the nucleotide and isoform-dependent structural and dynamical features of Ras proteins.* Structure, 2008. **16**(6): p. 885-96.
98. Buhrman, G., et al., *Analysis of binding site hot spots on the surface of Ras GTPase.* J Mol Biol, 2011. **413**(4): p. 773-89.
99. Buhrman, G., et al., *Allosteric modulation of Ras positions Q61 for a direct role in catalysis.* Proc Natl Acad Sci U S A, 2010. **107**(11): p. 4931-6.
100. Lowy, D.R. and B.M. Willumsen, *Function and regulation of ras.* Annu Rev Biochem, 1993. **62**: p. 851-91.
101. Ray, A., N. Jatana, and L. Thukral, *Lipidated proteins: Spotlight on protein-membrane binding interfaces.* Prog Biophys Mol Biol, 2017. **128**: p. 74-84.
102. Cox, A.D. and C.J. Der, *Farnesyltransferase inhibitors: promises and realities.* Curr Opin Pharmacol, 2002. **2**(4): p. 388-93.

103. Willumsen, B.M., et al., *The p21 ras C-terminus is required for transformation and membrane association.* Nature, 1984. **310**(5978): p. 583-6.
104. Nandan, M.O. and V.W. Yang, *An Update on the Biology of RAS/RAF Mutations in Colorectal Cancer.* Curr Colorectal Cancer Rep, 2011. **7**(2): p. 113-120.
105. Courtney, K.D., R.B. Corcoran, and J.A. Engelman, *The PI3K pathway as drug target in human cancer.* J Clin Oncol, 2010. **28**(6): p. 1075-83.
106. Vivanco, I. and C.L. Sawyers, *The phosphatidylinositol 3-Kinase AKT pathway in human cancer.* Nat Rev Cancer, 2002. **2**(7): p. 489-501.
107. Castellano, E. and J. Downward, *Role of RAS in the regulation of PI 3-kinase.* Curr Top Microbiol Immunol, 2010. **346**: p. 143-69.
108. Rajakulendran, T., et al., *A dimerization-dependent mechanism drives RAF catalytic activation.* Nature, 2009. **461**(7263): p. 542-5.
109. Freeman, A.K., D.A. Ritt, and D.K. Morrison, *The importance of Raf dimerization in cell signaling.* Small GTPases, 2013. **4**(3): p. 180-5.
110. Lavoie, H., et al., *Inhibitors that stabilize a closed RAF kinase domain conformation induce dimerization.* Nat Chem Biol, 2013. **9**(7): p. 428-36.
111. Marais, R., et al., *Ras recruits Raf-1 to the plasma membrane for activation by tyrosine phosphorylation.* EMBO J, 1995. **14**(13): p. 3136-45.
112. Avruch, J., et al., *Ras activation of the Raf kinase: tyrosine kinase recruitment of the MAP kinase cascade.* Recent Prog Horm Res, 2001. **56**: p. 127-55.
113. Ong, S.H., et al., *Stimulation of phosphatidylinositol 3-kinase by fibroblast growth factor receptors is mediated by coordinated recruitment of multiple docking proteins.* Proc Natl Acad Sci U S A, 2001. **98**(11): p. 6074-9.
114. Cantley, L.C., *The phosphoinositide 3-kinase pathway.* Science, 2002. **296**(5573): p. 1655-7.
115. Alessi, D.R., et al., *Characterization of a 3-phosphoinositide-dependent protein kinase which phosphorylates and activates protein kinase Balpha.* Curr Biol, 1997. **7**(4): p. 261-9.
116. Forbes, S.A., et al., *COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer.* Nucleic Acids Res, 2011. **39**(Database issue): p. D945-50.

117. Karnoub, A.E. and R.A. Weinberg, *Ras oncogenes: split personalities.* Nat Rev Mol Cell Biol, 2008. **9**(7): p. 517-31.
118. Willumsen, B.M., et al., *Harvey murine sarcoma virus p21 ras protein: biological and biochemical significance of the cysteine nearest the carboxy terminus.* EMBO J, 1984. **3**(11): p. 2581-5.
119. Wu, S.K., et al., *Structural insights into the function of the Rab GDI superfamily.* Trends Biochem Sci, 1996. **21**(12): p. 472-476.
120. Zhang, H., et al., *Photoreceptor cGMP phosphodiesterase delta subunit (PDEdelta) functions as a prenyl-binding protein.* J Biol Chem, 2004. **279**(1): p. 407-13.
121. Chen, Y.X., et al., *Synthesis of the Rheb and K-Ras4B GTPases.* Angew Chem Int Ed Engl, 2010. **49**(35): p. 6090-5.
122. Nancy, V., et al., *The delta subunit of retinal rod cGMP phosphodiesterase regulates the membrane association of Ras and Rap GTPases.* J Biol Chem, 2002. **277**(17): p. 15076-84.
123. Hanzal-Bayer, M., M. Linari, and A. Wittinghofer, *Properties of the interaction of Arf-like protein 2 with PDE delta.* J Mol Biol, 2005. **350**(5): p. 1074-1082.
124. Chandra, A., et al., *The GDI-like solubilizing factor PDEdelta sustains the spatial organization and signalling of Ras family proteins.* Nat Cell Biol, 2011. **14**(2): p. 148-58.
125. Schmick, M., et al., *KRas Localizes to the Plasma Membrane by Spatial Cycles of Solubilization, Trapping and Vesicular Transport.* Cell, 2014. **157**(2): p. 459-471.
126. Weise, K., et al., *Dissociation of the K-Ras4B/PDEdelta complex upon contact with lipid membranes: membrane delivery instead of extraction.* J Am Chem Soc, 2012. **134**(28): p. 11503-10.
127. Vartak, N. and P. Bastiaens, *Spatial cycles in G-protein crowd control.* EMBO J, 2010. **29**(16): p. 2689-99.
128. Cox, A.D., C.J. Der, and M.R. Philips, *Targeting RAS Membrane Association: Back to the Future for Anti-RAS Drug Discovery?* Clin Cancer Res, 2015. **21**(8): p. 1819-27.

129. Gysin, S., et al., *Therapeutic strategies for targeting ras proteins.* Genes Cancer, 2011. **2**(3): p. 359-72.
130. Collins, F.S., L.D. Brooks, and A. Chakravarti, *A DNA polymorphism discovery resource for research on human genetic variation.* Genome Res, 1998. **8**(12): p. 1229-31.
131. Risch, N. and K. Merikangas, *The future of genetic studies of complex human diseases.* Science, 1996. **273**(5281): p. 1516-1517.
132. Keskin, O., et al., *Principles of protein-protein interactions: what are the preferred ways for proteins to interact?* Chem Rev, 2008. **108**(4): p. 1225-44.
133. Alexov, E., *Protein-protein interactions.* Curr Pharm Biotechnol, 2008. **9**(2): p. 55-6.
134. Jones, S. and J.M. Thornton, *Principles of protein-protein interactions.* Proceedings of the National Academy of Sciences of the United States of America, 1996. **93**(1): p. 13-20.
135. Yates, C.M. and M.J.E. Sternberg, *The effects of non-synonymous single nucleotide polymorphisms (nsSNPs) on protein protein interactions.* Journal of Molecular Biology, 2013. **425**(21): p. 3949-3963.
136. Stefl, S., et al., *Molecular mechanisms of disease-causing missense mutations.* J Mol Biol, 2013. **425**(21): p. 3919-36.
137. Kar, G., A. Gursoy, and O. Keskin, *Human cancer protein-protein interaction network: a structural perspective.* PLoS Comput Biol, 2009. **5**(12): p. e1000601.
138. Gao, M., H.Y. Zhou, and J. Skolnick, *Insights into Disease-Associated Mutations in the Human Proteome through Protein Structural Analysis.* Structure, 2015. **23**(7): p. 1362-1369.
139. David, A., et al., *Protein-protein interaction sites are hot spots for disease-associated nonsynonymous SNPs.* Human Mutation, 2012. **33**(2): p. 359-363.
140. Schuster-Bockler, B. and A. Bateman, *Protein interactions in human genetic diseases.* Genome Biology, 2008. **9**(1).
141. David, A. and M.J. Sternberg, *The contribution of missense mutations in core and rim residues of protein-protein interfaces to human disease.* J Mol Biol, 2015. **427**(17): p. 2886-98.

142. Jubb, H.C., et al., *Mutations at protein-protein interfaces: Small changes over big surfaces have large impacts on human health.* Prog. Biophys. Mol. Biol., 2016.
143. Butler, B.M., et al., *Conformational dynamics of nonsynonymous variants at protein interfaces reveals disease association.* Proteins, 2015. **83**(3): p. 428-35.
144. Sahni, N., et al., *Widespread macromolecular interaction perturbations in human genetic disorders.* Cell, 2015. **161**(3): p. 647-60.
145. Nishi, H., et al., *Cancer missense mutations alter binding properties of proteins and their interaction networks.* PLoS One, 2013. **8**(6): p. e66273.
146. Alexov, E. and M. Sternberg, *Understanding molecular effects of naturally occurring genetic differences.* J Mol Biol, 2013. **425**(21): p. 3911-3.
147. Petukh, M., T.G. Kucukkal, and E. Alexov, *On human disease-causing amino acid variants: statistical study of sequence and structural patterns.* Hum Mutat, 2015. **36**(5): p. 524-34.
148. An, O., et al., *Structural and functional analysis of perforin mutations in association with clinical data of familial hemophagocytic lymphohistiocytosis type 2 (FHL2) patients.* Protein Sci, 2013. **22**(6): p. 823-39.
149. Kucukkal, T.G., et al., *Structural and physico-chemical effects of disease and non-disease nsSNPs on proteins.* Curr Opin Struct Biol, 2015. **32**: p. 18-24.
150. Engin, H.B., J.F. Kreisberg, and H. Carter, *Structure-Based Analysis Reveals Cancer Missense Mutations Target Protein Interaction Interfaces.* PLoS One, 2016. **11**(4): p. e0152929.
151. Peng, Y. and E. Alexov, *Investigating the linkage between disease-causing amino acid variants and their effect on protein stability and binding.* Proteins, 2016. **84**(2): p. 232-9.
152. Bogan, A.A. and K.S. Thorn, *Anatomy of hot spots in protein interfaces.* Journal of Molecular Biology, 1998. **280**(1): p. 1-9.
153. Clackson, T. and J.A. Wells, *A hot spot of binding energy in a hormone-receptor interface.* Science, 1995. **267**(5196): p. 383-6.
154. Tuncbag, N., A. Gursoy, and O. Keskin, *Identification of computational hot spots in protein interfaces: combining solvent accessibility and inter-residue potentials improves the accuracy.* Bioinformatics, 2009. **25**(12): p. 1513-20.

155. Cukuroglu, E., A. Gursoy, and O. Keskin, *HotRegion: a database of predicted hot spot clusters.* Nucleic Acids Research, 2012. **40**(D1): p. D829-D833.
156. Keskin, O., B. Ma, and R. Nussinov, *Hot regions in protein--protein interactions: the organization and contribution of structurally conserved hot spot residues.* J Mol Biol, 2005. **345**(5): p. 1281-94.
157. Xia, J.F., et al., *APIS: accurate prediction of hot spots in protein interfaces by combining protrusion index with solvent accessibility.* BMC Bioinformatics, 2010. **11**: p. 174.
158. Jones, S. and J.M. Thornton, *Protein-protein interactions: a review of protein dimer structures.* Prog Biophys Mol Biol, 1995. **63**(1): p. 31-65.
159. Yogurtcu, O.N., et al., *Restricted mobility of conserved residues in protein-protein interfaces in molecular simulations.* Biophys J, 2008. **94**(9): p. 3475-85.
160. Humphrey, W., A. Dalke, and K. Schulten, *VMD: visual molecular dynamics.* J Mol Graph, 1996. **14**(1): p. 33-8, 27-8.
161. Chavan, T.S., et al., *High-affinity interaction of the K-Ras4B hypervariable region with the Ras active site.* Biophys J, 2015. **109**: p. 2602-13.
162. Jang, H., et al., *Mechanisms of membrane binding of small GTPase K-Ras4B farnesylated hypervariable region.* J Biol Chem, 2015. **290**(15): p. 9465-77.
163. Jang, H., et al., *Membrane-associated Ras dimers are isoform-specific: K-Ras dimers differ from H-Ras dimers.* Biochem J, 2016. **473**(12): p. 1719-32.
164. Baspinar, A., et al., *PRISM: a web server and repository for prediction of protein-protein interactions and modeling their 3D complexes.* Nucleic Acids Res, 2014. **42**(Web Server issue): p. W285-9.
165. Tuncbag, N., et al., *Predicting protein-protein interactions on a proteome scale by matching evolutionary and structural similarities at interfaces using PRISM.* Nat Protoc, 2011. **6**(9): p. 1341-54.
166. Krauthammer, M., et al., *Exome sequencing identifies recurrent somatic RAC1 mutations in melanoma.* Nat Genet, 2012. **44**(9): p. 1006-14.
167. Brooks, B.R., et al., *CHARMM: the biomolecular simulation program.* J Comput Chem, 2009. **30**(10): p. 1545-614.
168. Klauda, J.B., et al., *Update of the CHARMM all-atom additive force field for lipids: validation on six lipid types.* J Phys Chem B, 2010. **114**(23): p. 7830-43.

169. Webb, B. and A. Sali, *Protein structure modeling with MODELLER.* Methods Mol Biol, 2014. **1137**: p. 1-15.

170. Phillips, J.C., et al., *Scalable molecular dynamics with NAMD.* J Comput Chem, 2005. **26**(16): p. 1781-802.

171. Brooks, B.R., et al., *CHARMM: The Biomolecular Simulation Program.* Journal of Computational Chemistry, 2009. **30**(10): p. 1545-1614.

172. Durell, S.R., B.R. Brooks, and A. Bennaim, *Solvent-induced forces between two hydrophilic groups.* J. Phys. Chem., 1994. **98**: p. 2198–2202.

173. Wang, W. and P.A. Kollman, *Free energy calculations on dimer stability of the HIV protease using molecular dynamics and a continuum solvent model.* J Mol Biol, 2000. **303**(4): p. 567-82.

174. Zoete, V., M. Meuwly, and M. Karplus, *Study of the insulin dimerization: binding free energy calculations and per-residue free energy decomposition.* Proteins, 2005. **61**(1): p. 79-93.

175. Brooks, B.R., et al., *Charmm - a program for macromolecular energy, minimization, and dynamics calculations.* J. Comp. Chem., 1983. **4**(2): p. 187-217.

176. Van Wart, A.T., et al., *Weighted Implementation of Suboptimal Paths (WISP): An Optimized Algorithm and Tool for Dynamical Network Analysis.* J Chem Theory Comput, 2014. **10**(2): p. 511-517.

177. Moal, I.H. and J. Fernandez-Recio, *SKEMPI: a Structural Kinetic and Energetic database of Mutant Protein Interactions and its use in empirical models.* Bioinformatics, 2012. **28**(20): p. 2600-2607.

178. Cukuroglu, E., et al., *Non-redundant unique interface structures as templates for modeling protein interactions.* Plos One, 2014. **9**(1).

179. Berman, H.M., et al., *The Protein Data Bank.* Nucleic Acids Res, 2000. **28**(1): p. 235-42.

180. Muratcioglu, S., et al., *Advances in template-based protein docking by utilizing interfaces towards completing structural interactome.* Curr Opin Struct Biol, 2015. **35**: p. 87-92.

181. Li, X., et al., *Protein-protein interactions: hot spots and structurally conserved residues often locate in complemented pockets that pre-organized in the unbound states: implications for docking.* J Mol Biol, 2004. **344**(3): p. 781-95.
182. Keskin, O., et al., *Protein-protein interactions: organization, cooperativity and mapping in a bottom-up Systems Biology approach.* Phys Biol, 2005. **2**(2): p. S24-35.
183. Zhu, X. and J.C. Mitchell, *KFC2: a knowledge-based hot spot prediction method based on interface solvation, atomic density, and plasticity features.* Proteins, 2011. **79**(9): p. 2671-83.
184. Kim, D.E., D. Chivian, and D. Baker, *Protein structure prediction and analysis using the Robetta server.* Nucleic Acids Res, 2004. **32**(Web Server issue): p. W526-31.
185. Zhao, N., et al., *Determining effects of non-synonymous SNPs on protein-protein interactions using supervised and semi-supervised learning.* PLoS Comput Biol, 2014. **10**(5): p. e1003592.
186. Adzhubei, I.A., et al., *A method and server for predicting damaging missense mutations.* Nat Methods, 2010. **7**(4): p. 248-9.
187. Adzhubei, I., D.M. Jordan, and S.R. Sunyaev, *Predicting functional effect of human missense mutations using PolyPhen-2.* Curr Protoc Hum Genet, 2013. **Chapter 7**: p. Unit7 20.
188. Stenson, P.D., et al., *The Human Gene Mutation Database: towards a comprehensive repository of inherited mutation data for medical research, genetic diagnosis and next-generation sequencing studies.* Hum Genet, 2017. **136**(6): p. 665-677.
189. Mosca, R., et al., *dSysMap: exploring the edgetic role of disease mutations.* Nat Methods, 2015. **12**(3): p. 167-8.
190. Lu, H.C., J.H. Braga, and F. Fraternali, *PinSnps: structural and functional analysis of SNPs in the context of protein interaction networks.* Bioinformatics, 2016. **32**(16): p. 2534-2536.
191. Forbes, S.A., et al., *COSMIC: exploring the world's knowledge of somatic mutations in human cancer.* Nucleic Acids Res, 2015. **43**(Database issue): p. D805-11.

192. Sun, H., et al., *Assessing the performance of MM/PBSA and MM/GBSA methods. 4. Accuracies of MM/PBSA and MM/GBSA methodologies evaluated by various simulation protocols using PDBbind data set.* Phys Chem Chem Phys, 2014. **16**(31): p. 16719-29.

193. Cukuroglu, E., A. Gursoy, and O. Keskin, *HotRegion: a database of predicted hot spot clusters.* Nucleic Acids Res, 2012. **40**(Database issue): p. D829-33.

194. Etienne-Manneville, S., *Actin and microtubules in cell motility: Which one is in control?* Traffic, 2004. **5**(7): p. 470-477.

195. Kazanietz, M.G. and M.J. Caloca, *The Rac GTPase in Cancer: From Old Concepts to New Paradigms.* Cancer Research, 2017. **77**(20): p. 5445-5451.

196. Carlier, M.F. and S. Shekhar, *Global treadmilling coordinates actin turnover and controls the size of actin networks.* Nat Rev Mol Cell Biol, 2017. **18**(6): p. 389-401.

197. Burridge, K. and K. Wennerberg, *Rho and Rac take center stage.* Cell, 2004. **116**(2): p. 167-79.

198. Leung, T., et al., *Myotonic dystrophy kinase-related Cdc42-binding kinase acts as a Cdc42 effector in promoting cytoskeletal reorganization.* Mol Cell Biol, 1998. **18**(1): p. 130-40.

199. Ismail, S.A., et al., *Arl2-GTP and Arl3-GTP regulate a GDI-like transport system for farnesylated cargo.* Nat Chem Biol, 2011. **7**(12): p. 942-9.

200. Watzlich, D., et al., *The interplay between RPGR, PDE delta and Arl2/3 regulate the ciliary targeting of farnesylated cargo.* Embo Reports, 2013. **14**(5): p. 465-472.

201. Nussinov, R., C.J. Tsai, and H. Jang, *Oncogenic Ras Isoforms Signaling Specificity at the Membrane.* Cancer Res, 2018. **78**(3): p. 593-602.

202. Nussinov, R., C.J. Tsai, and H. Jang, *Is Nanoclustering essential for all oncogenic KRas pathways? Can it explain why wild-type KRas can inhibit its oncogenic variant?* Semin Cancer Biol, 2018: p. DOI: 10.1016/j.semcancer.2018.01.002.

203. Muratcioglu, S., et al., *GTP-Dependent K-Ras Dimerization.* Structure, 2015. **23**(7): p. 1325-35.

204. Nan, X.L., et al., *Ras-GTP dimers activate the Mitogen-Activated Protein Kinase (MAPK) pathway*. Proc Natl Acad Sci USA, 2015. **112**(26): p. 7996-8001.

205. Nussinov, R., C.J. Tsai, and H. Jang, *Oncogenic KRas mobility in the membrane and signaling response.* Semin Cancer Biol, 2018: p. DOI: 10.1016/j.semcancer.2018.02.009.

206. Erwin, N., et al., *Influence of isoform-specific Ras lipidation motifs on protein partitioning and dynamics in model membrane systems of various complexity.* Biol Chem, 2017. **398**(5-6): p. 547-563.

207. Nussinov, R., et al., *Principles of K-Ras effector organization and the role of oncogenic K-Ras in cancer initiation through G1 cell cycle deregulation.* Expert Rev Proteomic, 2015. **12**(6): p. 669-682.

208. Spiegel, J., et al., *Small-molecule modulation of Ras signaling.* Nat Chem Biol, 2014. **10**(8): p. 613-22.

209. Ozdemir, E.S., et al., *Unraveling the molecular mechanism of interactions of the Rho GTPases Cdc42 and Rac1 with the scaffolding protein IQGAP2.* J Biol Chem, 2018. **293**(10): p. 3685-3699.

210. Zhang, X., et al., *Cyclase-Associated Protein 1 (CAP1) is a prenyl-binding partner of Rap1 GTPase.* J Biol Chem, 2018: p. DOI: 10.1074/jbc.RA118.001779.

211. Jankauskaite, J., et al., *SKEMPI 2.0: An updated benchmark of changes in protein-protein binding energy, kinetics and thermodynamics upon mutation.* bioRxiv, 2018.

212. Agius, R., et al., *Characterizing changes in the rate of protein-protein dissociation upon interface mutation using hotspot energy and organization.* PLoS Comput Biol, 2013. **9**(9): p. e1003216.

213. Moreira, I.S., P.A. Fernandes, and M.J. Ramos, *Hot spots--a review of the protein-protein interface determinant amino-acid residues.* Proteins, 2007. **68**(4): p. 803-12.

214. Hall, D.R., et al., *Lessons from Hot Spot Analysis for Fragment-Based Drug Discovery.* Trends Pharmacol Sci, 2015. **36**(11): p. 724-36.

215. Li, X., et al., *Protein-protein interactions: Hot spots and structurally conserved residues often locate in complemented pockets that pre-organized in the unbound states: Implications for docking.* Journal of Molecular Biology, 2004. **344**(3): p. 781-795.