A Comprehensive Study of K-Ras Protein and Its Oncogenic Mutations:

A Dynamic Point of View

By

Sezen Vatansever

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

Biomedical Sciences and Engineering

KOÇ UNIVERSITY

August 21, 2017

A Comprehensive Study of K-Ras Protein and Its Oncogenic Mutations: A Dynamic Point of View

Koç University

Graduate School of Sciences and Engineering

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

Sezen Vatansever

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. Dr. Burak Erman (Advisor)

_____

Prof. Dr. Füsun Can

_____

Asst. Prof. Alper Uzun

_____

Prof. Dr. Kemal Yelekçi

_____

Assoc. Prof. Ebru Demet Akten

Date: August 21, 2017

ABSTRACT

K-Ras is the most frequently mutated protein in human cancers driving cancer initiation, progression and drug resistance, directly leading to nearly a million deaths per year. Sadly, there are still no drugs in that directly target mutant K-Ras in the clinic. Recent studies utilizing dynamics information show promising results for selectively targeting it. However, despite extensive characterization, the regulatory mechanisms of K-Ras dynamics remain elusive. Since protein function is related to its dynamics, understanding these mechanisms can present novel opportunities for identifying target sites on mutant K-Ras surface.

In this work, we investigate the regulation mechanisms of K-Ras dynamics and the effects of nucleotide binding and mutations on these mechanisms using extensive molecular dynamics (MD) simulations. We applied different MD simulation data analysis techniques to compare the dynamic characteristics of both active and inactive forms of wild-type K-Ras^{WT} and mutant K-Ras^{G12D}, the most recurrent mutant in cancer patients. Our results on K-Ras^{WT} showed excellent agreement with experimental data and served as a reference point for K-Ras^{G12D} analysis. Then, we demonstrated how G12D mutation induces structural and conformational changes that result in characteristic correlated motions in active K-Ras^{G12D}. Moreover, we developed a novel conditional time-delayed correlations (CTC) based approach to predict causal relationships in regulation of K-Ras dynamics. CTC analysis identified the regulatory sites that control K-Ras dynamics. Finally, we identified a novel drug target pocket in active K-Ras^{G12D} and screened a small molecule library against it using docking techniques. Our study draws a complete picture of the regulation of K-Ras dynamics. We anticipate that the identified regulatory sites on active K-Ras^{G12D} can present novel opportunities for direct targeting of K-Ras^{G12D} in future drug discovery efforts.

ÖZET

K-Ras kanserde en sık mutasyona uğrayan protein olup kanserin oluşumunu, ilerlemesini ve ilaç direncini tetikleyerek her yıl sayısı milyonu bulan insan ölümüne yol açar. Ne yazık ki kanser tedavisinde bu proteini doğrudan hedefleyen bir ilaç henüz yok. Protein dinamiği bilgisinden yararlanan son çalışmalar ümit vaad etse de K-Ras'taki dinamik süreçlerin nasıl düzenlendiği hala aydınlatılmış değil. Oysaki bir proteinin işlevi içsel dinamik süreçlerin düzenlenmesiyle ilgilidir ve bu düzenleyici mekanizmaları anlamamız mutant K-Ras üzerinde ilaç hedefi olabilecek bölgeleri ortaya çıkarmamız için yeni imkanlar yaratabilir.

Bu çalışmada, K-Ras'taki dinamik süreçlerin düzenlenmesi ile nükleotit bağlanması ve mutasyonların buna etkisini inceledik. Bu amaçla kanser hastalarında en sık rastlanan mutasyon olan G12D'yi seçip doğal tip ($K\text{-}Ras^{DT}$) ve mutant ($K\text{-}Ras^{G12D}$) formların hem aktif hem de inaktif durumlarının moleküler dinamik simulasyonlarını yaptık. K-Ras'ın bu farklı formlarının dinamik özelliklerini karşılaştırmak için simulayon verisini farklı yöntemleri birleştirerek analiz ettik. $K\text{-}Ras^{DT}$ analiznin sonuçları literatürdeki deney sonuçlarıyla mükemmel örtüştü ve mutant formun analizleri için iyi bir referans noktası oldu. Sonrasındaki $K\text{-}Ras^{G12D}$ analiziyle G12D mutasyonunun yapısal ve biçimsel değişimleri tetikleyerek proteinde karakteristik korele hareketlerin oluşumunu sağladığını gösterdik. Dahası, proteindeki korele hareketlerin nedensellik ilişkisini ortaya çıkarabilen yeni bir yaklaşım geliştirdik. Koşullu gecikmeli korelasyon hesabına dayalı bu yaklaşımla K-Ras'taki dinamik süreçleri kontrol eden bölgeleri aydınlatmayı başardık. Tüm analiz sonuçlarını birlikte yorumlayarak $K\text{-}Ras^{G12D}$ üzerinde yeni bir ilaç hedefi bölgesi bulduk ve buraya bağlanabilecek molekülleri yanaşma tekniğiyle taradık. Burada bulunan K-Ras dinamiklerini kontrol eden bölgelerin gelecek ilaç keşfi çalışmalarında yeni fırsatlar sunacağını umuyoruz.

ACKNOWLEDMENTS

This intensive work would not have been completed without the guidance and the help of numerous people who I am indebted to.

First and foremost, I would like to express my sincere gratitude to my thesis advisor Prof. Dr. Burak Erman for giving me the opportunity to pursue a Ph.D. study under his guidance. Many thanks for your insightful discussions, constructive suggestions, grateful encouragement and all other contributions that I cannot list here. It was an honor to work with you and be one of your Ph.D. students.

I would like to thank my thesis monitoring committee members Prof. Dr. Füsun Can and Asst. Prof. Alper Uzun for all of their guidance through this process. I would also like to thank my thesis committee members, Prof. Dr. Kemal Yelekçi and Assoc. Prof. E. Demet Akten for accepting to be a member of this committee and for their constructive comments on my thesis.

I would like to show my gratitude to our dear collaborator Asst. Prof. Zeynep H. Gümüş at Icahn School of Medicine at Mount Sinai for giving me the opportunity to conduct part of my research in her lab. It has been an instructive and exciting ten months for me in New York.

I acknowledge the financial support from the Scientific and Technological Research Council (TÜBİTAK) for the 2214A scholarship for my visit to Icahn School of Medicine at Mount Sinai.

Work would not have been as pleasant without my colleagues. I would like to thank my dear lab mate Aysima Hacısüleyman, for her excellent cooperation and friendship based on trust, empathy and humor. Special thanks to Dr. Ozlem Ulucan for her guidance and support whenever I need; your discussion, ideas and friendship have been absolutely invaluable. Also, I would like to thank my colleagues Nilay Karahan, Dr. Eda Süer and Pelin Erkoc for

broadening my perspective and enriching my doctoral experience. I also appreciate all the support of GSSE to overcome many administrative issues.

I am thoroughly grateful to my lovely friends, Burçin İçdem, Güven Korucu, Sümeyra Demir, Özge Güre, Inci Tezel, Saba Sağlıker and Gülay Gülpınar for their endless supports, patience and friendship. It would be impossible to finish this graduate work without your support and lifelong friendship.

My thanks and appreciations also go to my family, my parents and my lovely grandparents, for all their love and unlimited support. Deepest thanks to my sister, Sinem Vatansever, for being my best friend for life.

Saving the most important for last, I would like to deeply thank my spouse, Okay İnak, whose presence helped making the completion of my degree possible. He was the one who keeps me motivated me in tough times that I faced. Many thanks for his unconditional love, endless support and warm encouragement.

Thank you...

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1 INTRODUCTION

K-Ras is the most frequently mutated oncoprotein in human cancers[1-3]. Activating mutations in K-Ras protein drive tumor initiation, progression and drug resistance in patients with cancers of the pancreas, colon and lung[4-7], directly leading to nearly a million deaths per year [8]. Sadly, these mutations are predictive of very poor response to standard therapies, and often emerge and drive acquired resistance[9-11]. K-Ras is a small GTPase that is critical in the regulation of intracellular signaling networks in cellular growth, proliferation and differentiation[12]. To perform its cellular roles, it switches between its inactive GDP-bound and active GTP-bound states (Figure 1.1A) [13, 14]. The balance between the two states is regulated by guaninenucleotide exchange factors (GEFs), which bind to inactive K-Ras (K-Ras-GDP) and catalyze the exchange of GDP for GTP. Only active GTP-bound K-Ras (K-Ras-GTP) can bind and activate its downstream effector proteins, such as Raf kinase, phosphatidylinositol 3-kinase (PI3K), and Ral guanine nucleotide dissociation stimulator (RalGDS) (Figure 1.1A)[15-17]. To terminate the downstream signaling, K-Ras-GTP catalyzes the GTP hydrolysis and becomes inactive. In K-Ras-GTP, the P-loop (residues 10-17), switch I (SI, residues 25-40) and switch II (SII, residues 60-74) regions make up the active site whose well-ordered conformations allow GTP hydrolysis (Figure 1.2). Intrinsic GTPase hydrolysis of K-Ras-GTP can be accelerated by the binding of GTPase-activating proteins (GAPs)[18, 19]. However, oncogenic mutations in K-Ras impair the intrinsic GTPase function and GAP binding and thereby GTP hydrolysis. Unable to switch to its GDP-bound inactive state, mutant K-Ras remains continuously active, causing prolonged activation of downstream pathways associated with oncogenic cell growth[15, 20-22]. Oncogenic mutations in K-Ras are found in about 27% of human cancers[2]. These mutations in K-Ras

protein are recurrently observed at positions G12, G13 and Q61, where G12 is the most frequently mutated residue (89%), followed by G13 (9%) and Q61 (1%) (Figure 1.1B)[3, 15]. G12 most prevalently mutates to aspartate (G12D, 36%) followed by valine (G12V, 23%) and cysteine (G12C, 14%)[3, 15]. There is strong evidence that blocking mutant K-Ras activity can be very effective in cancer treatment[23, 24]. Yet, despite decades of research, there are still no drugs in the clinic today that can directly target these proteins[25, 26].

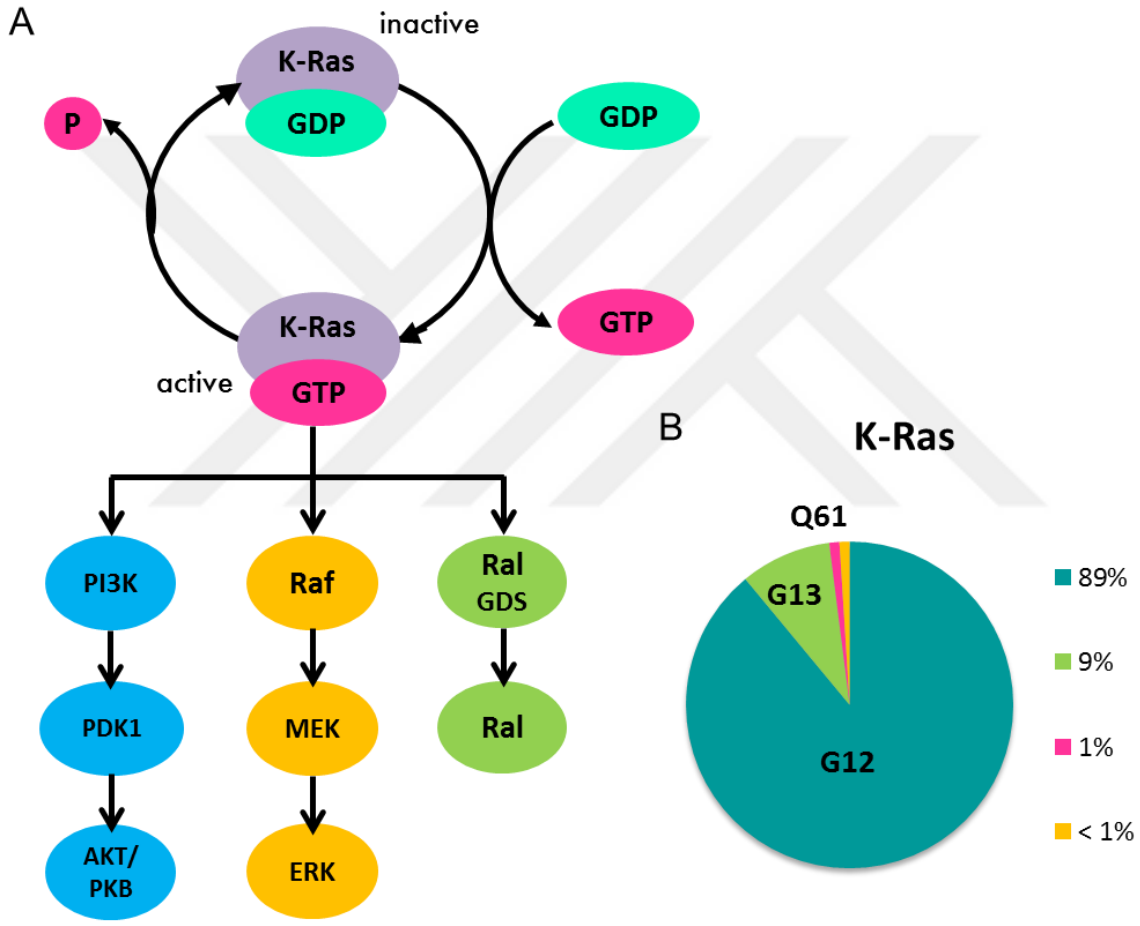


Figure 1.1 K-Ras GTP/GDP cycle, downstream effectors and mutation frequencies. A) Activation/inactivation cycle of K-Ras and downstream effector pathways implicated in K-Ras-mediated oncogenesis. Adapted from literature [27] B) Piechart graph of the frequencies of mutations in K-Ras.(Blue: G12, Green: G13, Magenta: Q61, Yellow: Others)

Part of the challenge in oncogenic K-Ras inhibitor design has been due to structure analyses that suggest a lack of well-defined druggable sites on its surface[28]. Accumulating studies suggest that K-Ras proteins are in dynamic and flexible states and their distinct characteristics cannot be identified by structural studies alone[24, 29-37]. Studies that have utilized protein dynamics data such as nuclear magnetic resonance (NMR) and mass spectrometry have identified binding pockets on specific K-Ras oncogenic mutants, attempted to stabilize their conformational states achieved promising results; however, these are limited to K-Ras^{G12C} mutant[30, 31, 35, 38], and the mechanisms that regulate K-Ras dynamics remain unknown. Therefore, we still need to clearly understand the intra-molecular regulatory relationships between residue pairs of K-Ras[26] and thereby investigate how the mutations trigger structural, conformational and dynamic changes in K-Ras that result in its constitutive activation. Understanding the mechanisms of dynamic regulation can present novel opportunities for identifying clinically actionable regulatory sites on mutant K-Ras surface.

In regulation of protein dynamics, correlated motions between protein residues are essential [39-41]. These motions enable the transfer of fluctuation information through the allosteric network[42], which inherently involves directionality, or "causality" of events[43]. If the motions of two residues are correlated, it would be valuable to identify whether the motions of one residue drive the motions of the other. However, while correlation calculations indicate interaction (which is necessary for allosteric transitions) they are symmetric and do not reveal the direction of information flow. A method for detecting causality between two sets of observables is *time-delayed correlation calculation*. It has been extensively used in causality analyses in economics since its inception[44], leading to the Nobel prize, but has not been widely adopted in protein dynamics.

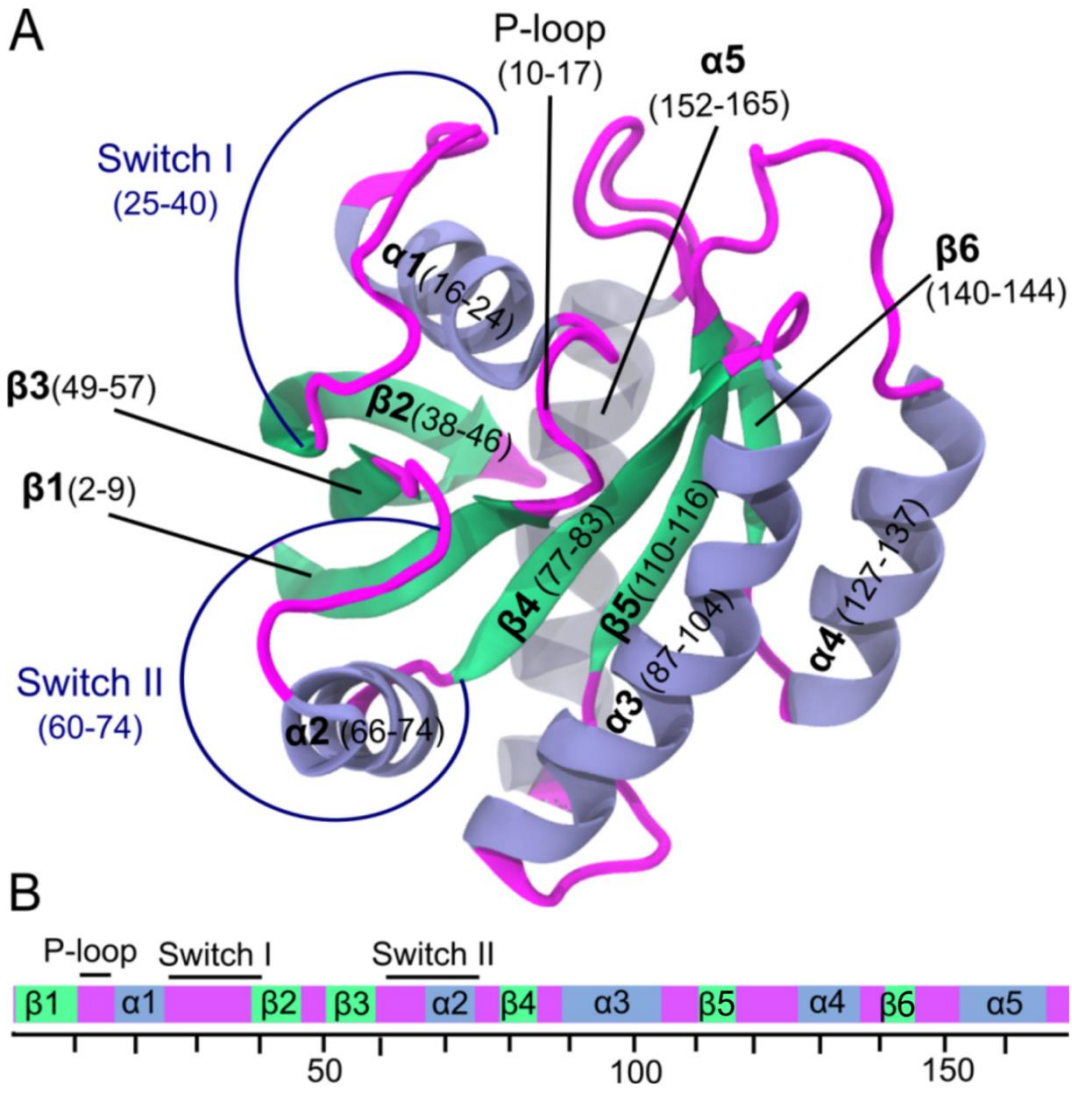


Figure 1.2 Crystal structure of wild-type K-Ras protein in GTP-bound state (PDB: 4OBE). A) K-Ras structure ribbon representation with secondary structures in blue for α-helices and green for β-sheets. B) Schematic of K-Ras sequences (residues 1-169). Functional regions are in same color used in K-Ras structure in A.

The major goal of this thesis is to study regulation mechanisms of K-Ras dynamics. To do this, we utilized several theoretical approaches for analyzing trajectories obtained from MD simulations of wild-type and mutant K-Ras proteins.

This thesis consists of eight chapters. In the present chapter, we give an overview of K-Ras protein as a drug target and the importance of dynamic characterization of it for direct

targeting. Chapter 2 reviewed the current literature on protein dynamics including theoretical aspects of MD simulations, causality relationships in protein motions, K-Ras in cancer, structure and dynamics of K-Ras and drug design efforts to target it. Chapters 3, 4, 5, and 6 are organized in a similar manner. Each of those chapters describes a complete study with the introduction, the methods used and the results obtained combined with the discussion section.

Chapter 3 presents a comprehensive study of intrinsic dynamics of K-Ras, including detailed analyses of causality between the motions of its residues. Development of new direct inhibitors that selectively bind to mutant K-Ras conformations while sparing those of WT K-Ras requires first understanding the dynamics of the WT protein in detail. To evaluate the GTP/GDP binding dependent changes in WT K-Ras dynamics, we analyzed data from MD simulations of K-Ras-GTP and K-Ras-GDP by combining several distinct methods. Here, we introduce a novel approach –CTC– that predicts causality relationships between residue pairs of a protein.

In Chapter 4, we probed the specific effects of the K-Ras mutations onto causality relationships in regulation of the protein dynamics. In that chapter, we improved our novel method CTC, introduced in Chapter 3, for faster and more accurate analysis of long time scale MD simulation data.

Chapter focuses on the most frequently observed K-Ras mutation, G12D. We investigated how the G12D mutation triggers structural, conformational and dynamic changes in K-Ras that result in its constitutive activation. For this purpose, we developed an integrated MD simulation data analysis approach that uses different computational metrics to quantify mutation-based changes in protein structure and its consequences.

Chapter 6 discusses the dynamic characteristics of active and inactive K-Ras^{G12D}. To design specific inhibitor drugs that selectively target constitutively active K-Ras^{G12D}-GTP, we need to identify differences between active and inactive K-Ras^{G12D}. Therefore, we performed detailed analysis of both active K-Ras^{G12D}-GTP and inactive K-Ras^{G12D}-GDP dynamics and provided comprehensive descriptions and comparison of these two states from extensive MD simulations.

In chapter 7, using the results from Chapter 5 and Chapter 6, which compare the dynamic characteristics of WT & mutant K-Ras and active & inactive mutant K-Ras respectively, we were seeking for a novel drug binding pocket that is specific to active K-Ras^{G12D}-GTP, the most recurrent mutant form. After identifying a novel drug binding pocket, we identified the small molecules that fit into novel binding site on K-Ras^{G12D} surface by applying docking techniques. In addition to K-Ras case, we applied docking methods into other drug design problems, too. We demonstrate three fruitful applications that we performed two of them in collaboration with experimental groups.

Finally, Chapter 8 discussed the main outcomes of this thesis. We conclude this thesis by emphasizing that understanding the regulation of protein dynamics is crucial for drawing a complete picture of protein function; identifying the molecular mechanisms of mutations; and development of better drugs.

2 LITERATURE REVIEW

2.1 Protein Dynamics

"Certainly no subject or field is making more progress on so many fronts at the present moment, than biology, and if we were to name the most powerful assumption of all, which

leads one on and on in an attempt to understand life, it is that all things are made of atoms, and that everything that living things do can be understood in terms of the jigglings and wigglings of atoms."

Richard P. Feynman

(*Feynman Lectures on Physics*, vol. 1, Ch. 3. 1963)

Proteins are dynamic macromolecules and essential components of the cell being responsible for a variety of biological functions. The function of a protein is related to its structure and dynamics. The physical basis of protein structure, dynamics and function has been extensively studied for several decades. Understanding of a protein function requires determination of protein structure at first[45]. Over the past half-century, improvements in structural biology have provided atomic-resolution models of many proteins [46].The experimental methods for the determination of the protein three-dimensional structures are X-ray crystallography and NMR spectroscopy. Numerous static protein structures have been determined by X-ray crystallography and NMR spectroscopy techniques[47] However, these techniques have their own limitations. In X-ray crystallography, the homogeneous crystal is needed while NMR spectroscopy technique is confined to structural solution of only small soluble proteins. Because of the limitations of experimental methods, computational methods for prediction of protein structure also have been developed[48]. These methods use an energy function for identifying the most stable conformation of a protein and a scoring function for evaluating the predicted models.

However, determination of the structure of a protein is not enough to understand the function of it. Proteins are dynamic entities where the substantial structural fluctuations occur[49]. These fluctuations and the resulting conformational changes are playing an essential role in

their functions[50]. That conformational flexibility and also its preservation through specific deformation modes are crucial for proteins function[51-53]. Therefore, a full understanding of the function of a protein requires not only determination of its statistic structure but even more importantly, analysis of its dynamic behavior[54].The methods for analysis of protein's dynamic behavior can be divided into two groups: High resolution structural approaches and theoretical approaches.

2.1.1 High resolution structural approaches

High resolution structural approaches involve X-ray crystallography and NMR spectroscopy[55]. Although protein motions have been frozen by these techniques, they provide indirect information on protein dynamics. For example, protein residues that can be found in different conformations often appear with partial occupancy in the X-ray structures. Moreover, each atom in the X-ray structures in Protein PDB has a B-factor value which is a rough measure of flexibility[56]. The B-factor calculated from the displacement of the atomic positions from an average value. The more flexible atoms have the larger B-factors distinguishing the rigid and flexible parts of a protein. Although they are widely used to estimate protein flexibility, B-factors are able to explain only harmonic motions, so that they systematically underestimate the magnitude of protein flexibility[57]. On the other hand, NMR spectroscopy also provides dynamic data giving more information than X-ray crystallography[58]. NMR experiments give dynamic information at different time scales: the pico-nanosecond scale from spin relaxation type experiments, the micro-millisecond scale in relaxation–dispersion measures and the second-minute regime in amide proton exchange saturation experiments[55]. Although these NMR experiments provide valuable dynamic information, we need to transform the observations into three-dimensional continuous images

to draw a complete picture of protein dynamics. However, theoretical methods for this transformation should be integrated very carefully [55, 59].

2.1.2 Theoretical approaches: MD Simulations

The limitations of the high resolution structural approaches for defining the protein dynamics lead improvement of theoretical approaches and thereby development of computational tools. Since computational tools deal with models, they are classified as theoretical approaches[60]. The computational tool preferred for describing the dynamics of a specific protein system depends on the systems size. For dynamic description of large protein systems coarse grained methods are used. Coarse-grained models simulate the dynamics of complex protein systems using their simplified representation. Since complex protein systems are too large to be studied in atomistic detail, coarse-grained modelling makes the representation of large protein systems possible by lowering the representation level from all-atom to coarse-grained[61]. In coarse-grained models, in order to reduce the degrees of freedom of the system protein atoms divided into groups which are represented by single beads and solvent molecules divided into clusters which are represented by particles[62-67]. Although coarse-grained modelling is highly efficient in terms of computational cost, the atomic details are lost in this technique.

A widely used computational method for studying protein dynamics is all-atom MD simulation. This method represents every atom in the system and evolves the positions and velocities of atoms according to the laws of classical physics. It computes the forces applied on the atoms using force fields. Those force fields are designed based on a combination of first-principles physics and parameter fitting to quantum mechanical calculations and experimental data[68]. A force field defines the total force on an atom as a sum of three types of component forces: bonded forces, van der Waals forces and electrostatic forces. Bonded

forces involve interactions between atoms connected by one or more covalent bonds. Van der Waals forces include the interactions between all atom pairs that fall off quickly with distance and therefore this type of interactions typically evaluated only for nearby pairs of atoms. Controversially, electrostatic forces include the interactions between all atom pairs that fall off quickly with distance. They are computed explicitly between nearby atom pairs; however, long-range electrostatic interactions are computed by using approximate methods which are more efficient than explicitly computing interactions between all distant atoms pairs.

In MD simulation algorithms, the equations of motion are solved numerically following the time evolution of the system, and they allow the derivation of kinetic and thermodynamic properties. This property has own significance since the user can apply different environmental conditions (e.g. temperature, pressure, pH) to the simulated system. Since the biologically important macromolecules and their environments can be studied, MD simulations have been accepted as computational experiments[69].

It has been 40 years since the first MD simulation of a macromolecule of biological interest was published[70, 71]. It was the simulation of the bovine pancreatic trypsin inhibitor whose X-ray crystal structure was available in 1975. Although this simulation lasted for only 9.2 ps, the results were instrumental in acceptance of the idea that proteins are dynamic systems and their internal motions are related to their function[71]. During the following years, MD simulation method has continued to evolve with increasing accuracy and predictive power that makes it a widely accepted tool all over the world. While the original simulation was less than 10 ps in length, current protocols allow the simulation of protein systems for the multi-nanosecond to microsecond time scale[55]. This rapid progress in MD simulation method and its popularization have been possible thanks to the advances in MD simulation softwares. The

most known softwares are GROMACS[72], GROMOS[73], AMBER[74], CHARMM[75] and NAMD[76] and they are freely available for the scientific community.

One of the current progresses in MD simulations is parallel-computing simulations. The two components of its algorithm are responsible for the computational cost of MD simulation[46]. The first component is the force calculation at each time step. This part requires substantial computation (e.g. one billion calculation steps for a system with one hundred thousand atoms). The second component is the need for the force calculation repeated many times. Since atomic vibrations are too fast, individual time steps of the simulation are limited to a few femtoseconds. Therefore, approximately one billion time steps are required for a microsecond simulation. It's impossible to complete such simulation on a single high-end processor core since it takes hundreds of years. To accelerate the simulations, people have started to develop softwares that parallelize molecular dynamics force calculations across multiple computer 22 years ago[77, 78]. Recent algorithmic innovations that reduce the communication requirements have improved the parallel computing scalability and efficiency. For example, as a part of Blue Gene project, IBM has developed Blue Matter application which can be scaled to thousands of cores for Blue Gene/L supercomputer[79]. Along with this, parallel computing performances of the most known MD simulation softwares such as GROMACS, AMBER and NAMD have made a marked progress.

The other substantial contribution to the current progresses in MD simulations has made by D. E. Shaw Research. Shaw and co-workers developed a software package, Desmond, for commodity clusters, that speeds up the simulations by an order of magnitude[80]. They could run 500 ns simulation per day with this software[81]. However, their greatest contribution to the advances in MD simulations is Anton supercomputer. In order to run very long time scale

simulations, they designed Anton computer specifically for MD simulations[46], Anton is composed of directly connected special-purpose chips. Each Anton chip contains a dozen processor cores which are customized for simulations[81]. As a result of their effort, Shaw and co-workers was able to complete the first millisecond atomistic MD simulation of a folded protein[82]. Their studies of protein folding simulations provided a basis for understanding the kinetics of protein folding[83-86]. In recent publications, they explained the other processes in protein dynamics like ion channels' mechanism, allosteric effects of ligand binding into membrane receptors and drug binding to target proteins by running long time scale simulations[87-90]. The extraordinary computational power of Anton allows extremely long simulations which can alter the interplay between experiment and theory.

In addition to parallel computing and the special purpose parallel supercomputer, usage of graphics processing units (GPUs) also has accelerated MD simulations. GPUs are able to perform large numbers of identical calculations in parallel on a single chip. Although GPUs had been designed for increasing the speed of rendering of three-dimensional graphics, they have widely used for scientific purposes. In the last few years, the most popular MD simulation programs, GROMACS, AMBER and NAMD optimized their codes for GPU support. Despite the fact that GPU accelerated MD simulations with these softwares could not reach the speed of Anton, they allow many scientists to perform the multi-microsecond simulations for small and medium-sized proteins since they are freely available for the scientific community[91, 92].

Along with the rapid progress in MD simulation softwares, people also have developed tools for visualization of output trajectories of MD simulations. The most popular molecular animation tool is Visual Molecular Dynamics (VMD). VMD is written in C++, using an

object-oriented design by the Group of Theoretical Biophysics from the University of Illinois[93]. It is a molecular graphics program designed for the display and analysis of molecular assemblies, in particular proteins, nucleic acids, and molecular systems on the basis of lipids (i.e. components of cell membranes). VMD has also been designed for animation and analysis of trajectories obtained MD simulations and it is extensively used by scientific community as molecular animation software. VMD supports variety of file formats of for biomolecules and output MD simulation trajectories. Its ability to process great amount of data and its options for visualization and rendering of images and animation make VMD an attractive tool for the community[94]. The program allows user to use various methods of visualization and coloring molecules. One of the advantages of the program is that it can operate on a remote computer. Moreover, it is integrated with NAMD software. Scientist with any background can easily operate VMD thanks to its user-friendly graphical user interface. VMD software, its source code and extensive documentation are freely available via anonymous ftp and through the World Wide Web.

MD simulation method has been accepted as a "virtual microscope"[95] to observe variety of biologic processes in molecular level thanks to the recent improvements in simulation softwares and trajectory visualization tools. More specifically, people utilized MD simulation method especially for their studies on conformational changes in proteins and the binding of ligands into proteins[60].

2.1.2.1 Conformational changes and ligand binding in MD simulations

Conformational changes in proteins have studied for many years, since proteins' function and regulation depend on the conformational changes[96]. Under physiological conditions, proteins obtain different conformational states. Conformational changes consist of significant

domain motions the thermal motions. Domain motions are the slow processes and they may assist ligand or effector binding. On the other hand, thermal motions such as side-chain rotations occurring at equilibrium are fast motions and they are precondition for protein activity[97]. Since MD simulations can capture these conformational changes composed of slow and fast motions, scientist use MD simulations to identify changes in conformational distribution resulting from mutations or ligand binding. Fast motions of proteins are very short time scale motions and they cannot be observed in experimental studies. Only molecular dynamic studies are able to record them even in femtosecond level. On the contrary, slow motions of proteins are long time scale motions and they can be observed only by experimental studies until recently. However, thanks to the recent progress in MD simulations (e.g. parallel-computing, special-purpose supercomputers, GPU acceleration etc.) the slow motions of proteins up to millisecond time scale can be recorded in MD simulations now[98]. In parallel to advances for calculation of long time scale motions, simulation data visualization tools also have been improved to deal with increasing amount of simulation data[99]. Therefore, scientists are able to perform detailed analysis and visualization of the protein conformational changes even from fast femtosecond motions to slow millisecond motions with MD simulations today.

In addition to conformational changes of proteins, ligand binding to proteins also has been extensively studied by utilizing MD simulations[46]. Variety of biological processes (e.g. enzyme reactions, molecular recognition, cellular signaling) result from the intermolecular interactions between proteins and ligand molecules. Moreover, those interactions between proteins and small-molecule ligands play a key role in the disease treatment where the ligands act as drugs. Ligands can alter protein activity by causing the protein to change its conformational state or inhibit protein activity by binding to the active site of a protein. All of

the changes in protein activity as consequences of ligand binding can be identified by MD simulations. Moreover, due to the recent improvement in accessible timescales, now scientists are able to perform MD simulations where the ligands bind to target proteins spontaneously even if they don't have prior knowledge of the binding site[100-102]

2.1.2.2 MD simulations in drug discovery

MD simulations have found applications in different areas, as they become faster and more accurate. Especially, they have been utilized as part of the drug development process Furthermore, it's speculated that biomolecular simulation may make an impact drug discovery area in the coming years[46]. It has been over 50 years when Feynman stated the importance of the 'jigglings and wigglings' of atoms for understanding the living things. There isn't any doubt that understanding these atomic motions of proteins is pertaining to drug discovery.

In 1894, Emil Fischer postulated the initial ligand binding theory –'*lock and key*' model- for the first time[103]. According to this analogy, the lock is the motionless target protein and the key is the ligand. Only the correctly sized ligand (*key*) can fit into the active site (*key hole*) of the target protein (*lock*) without causing any conformational changes. However, this rigid model couldn't explain all experimental evidence sufficiently. For this reason, the rigid model has been abandoned in favor of a dynamic model: '*Induced fit*'. In 1958, D. E. Koshland Jr. from the University of California, Berkley, recognized that proteins are flexible structures and proposed this new model[104]. This dynamic model states that the binding of a ligand to a target causes a change in the shape of the target protein [105]. This model assumes that target protein is partially flexible and only the proper ligand is able to induce the proper fitting into the active site. A further proposed model was "conformational selection". This model describes the stabilization of protein conformations by ligand binding. According to this

model, a protein can be found in variety of conformations from its unbound form to different interconvertible states. The binding of a ligand into the target protein triggers the conformational selection process where the ligand shows the highest binding affinity to a specific conformation of the protein and it preferentially binds to this conformation. During the conformational selection process the ligand does not create a new conformational state but causes a shift in the population in the favor of the binding competent state.

MD simulations have long been utilized the drug discovery studies where the target protein is assumed as a dynamic entity. MD simulations of target proteins are attractive because of they have the potential to discover novel drug binding sites. Structural studies are not enough to define target sites on the surfaces of such dynamic proteins. On the other hand MD simulations are capable to reveal novel pockets on the proteins that are not present in existing crystal structures[106-109]. In addition to their capacity to reveal novel drug target sites, MD simulations are able to refine low-resolution structural protein models that are utilized in structure-based drug design[110].

MD simulations of proteins have provided many insights into complex activity of proteins concerning the internal motions of them since the first MD simulation of a protein was performed 40 years ago. Thanks to the advances in the simulation tools and the computational power we can carry out MD simulations of larger and more complex systems and longer time scales. These advances allow us to observe the atomic motions that are related to protein function for longer time scales. We can also obtain detailed information that is not obtained from experiments. The success of simulation tools makes clear that the applications of MD simulations will play an even more important role for our understanding of protein systems in the future.

2.2 Causality in protein dynamics

Understanding the regulation of protein dynamics is crucial for drawing a complete picture of the protein function. In this regulation, the dynamics of different parts of the protein may be coupled by their long-range interactions and correlated motions. Such correlated motions are essential for the allosteric regulation which can render the protein dynamics productive in terms of biological function (e.g. catalysis, signal transduction). Realizing the importance of correlated motions, people have proposed several analysis methods for MD simulation data that aimed to define correlated motions[111, 112].

One of the popular analysis methods for revealing the correlated motions is the calculation of the Pearson coefficient from MD trajectories. Ichiye and Karplus applied this method to the analysis of the bovine pancreatic trypsin inhibitor dynamics for the first time[113, 114]. They could identify the regions of the protein that move in a correlated manner. As illustrated by the first application, the Pearson coefficient quantifies the correlation between pairs of residues that indicates allosteric coupling of their motions. Although it's a very popular method, Pearson coefficient calculation has a disadvantage: Coefficient values are symmetric. Therefore, one residue cannot be differentiated from another in a correlated pair. Nonetheless, it's known that correlated motions of residue pairs often imply a causal directionality. The most known two examples of causality relationships in protein motions are hemoglobin and GroEl shaperon. In hemoglobin dynamics, the binding of oxygen atom to the heme group drives the correlated motions among heme, helix F, and the rest of the protein. Specifically, oxygen binding induces the motions of heme group and then the motions of heme group are responded by the motions of helix F and the rest of the protein[115, 116]. On the other hand, in the GroEL chaperone dynamics, the binding of ATP molecule to the equatorial domain

drives the correlated motions between the equatorial, intermediate, and apical domains that results in that lead the opening of the cis cavity[117, 118].

For determining the causality relationships between protein residue pairs transfer entropy calculation has been used[119, 120]. Schreiber proposed an information theoretic measure which is called transfer entropy[121]. This theoretic metric is able to identify asymmetry in the correlated pairs and distinguish driving and responding elements. It was used mostly in neuroscience to detect the neuronal networks[122, 123]. Then transfer entropy was applied to MD simulation data of transcription factor Ets-1 in order to reveal the information flow in its dynamics[42] Analysis showed that binding of DNA to H1 helix causes the correlated motions of the H1 helix with HI1 by means of a relay helix. Another application is the analysis of the autoactivation of extracellular signal-regulated kinases 1 and 2[124]. This analysis elucidated that helix-C at N-domain drives the activation lip that results in activation. Although these studies show that transfer entropy can be a useful measure to understand the causality in protein dynamics, the statistical significance of calculated transfer entropy values is not clear in these studies[42, 124]. Resulting transfer entropy values are often too small, and thus it is critical to apply well-defined statistical significance test.

As we know from the literature, there is a method that can identify causation between time-series variables is the calculation of the time-delayed correlation. This method first proposed by Sir Clive William John Granger[44], a British econometrician and recipient of the 2003 Nobel Prize in Economics[125]. According to this method, for two time-series variables, X and Y, time-delayed correlation quantifies the information flow from the past of X to the future of Y. It has been extensively used in causality analyses in economics and in other fields, for example neuroscience[126], cell biology[127], human physiology[128, 129]. But to our knowledge, no applications to MD simulation data have been reported.

2.3 K-Ras protein as a drug target

2.3.1 K-Ras and cancer

The linkage between mutationally activated human Ras genes and human cancer was identified as retroviral oncogenes transduced from the rodent genome in 1982[130]. This investigation provoked intensive research for understanding the structure, biochemistry and biology of wild type and mutant Ras proteins to obtain information to be utilized in the discovery of drug molecules that can block mutant Ras function in cancer.

The pioneer study for identification of Ras was a retrovirus study which isolated rapidly transforming retroviruses from animals such as chickens, turkeys, rats, mice, cats and monkeys. In that extensive study, the oncogenic viruses cause acute formation of sarcomas in infected animals and potent transformation of cells in culture. After that study the potently oncogenic Harvey murine sarcoma virus and the Kirsten murine sarcoma virus were identified in 1964[131] and in 1967[132], respectively. Identification of these viruses leaded to discoveries of the oncogenic genetic elements in human including the human H-Ras and K-Ras oncogenes.

Between the late 1970s to the early 1980s, serial studies of Scolnick and colleagues elucidated origin of the viral H-Ras and K-Ras genes[133, 134]. Then they determined that the 21 kDa proteins are encoded by these genes[135] and also these proteins make complex with nucleotides (GDP and GTP)[136] that are bound to the cell membrane[137]. Furthermore, they revealed that overexpression of these proteins can transform cells[138] and their preferential binding to GTP is key for transformation[136, 139]. Scolnick and colleagues could establish the basics of Ras as GTP-binding membrane-associated proteins and their studies served as cornerstone for the further subsequent studies on Ras.

In 1982, three groups identified the same Ras genes in Kirsten and Harvey sarcoma viruses in NIH/3T3 DNA transfection assays[140, 141]. They could also reveal the molecular basis of H-Ras gene activation in the EJ/T24 bladder carcinoma cell line[142]. Then the activation of K-Ras gene in human colon and lung carcinomas was identified, too[143]. Today, thanks to advances in sequencing technology we know that K-Ras mutations are the biggest oncogene mountains in particularly in human colon and pancreatic cancer, which are two deadly cancers in human. In contrast to K-Ras, N-Ras mutations are associated with hematopoietic tumors. As The Catalog of Somatic Mutations in Cancer(COSMIC) confirms that K-Ras is the most frequently mutated Ras isoform in Ras driven cancers present in 86% of all tumors analyzed compared to 11% for N-Ras and 3% for H-Ras[3]. Likewise, each isoform has a different point mutation frequency. For K-Ras, 89% of the mutations are observed at residue 12, while very few mutations occur at codon 61. Controversially, 60% of N-Ras mutations are found at residue 61, whereas 35% of them are observed at codon 12. However, H-Ras mutations exhibit an intermediate pattern with 50% frequency at residue 12 and 40% frequency at residue 61[2].

2.3.2 Structure and dynamics of Ras

The first Ras crystal structure was determined 30 years ago[144]. This finding fueled the research on Ras structure and function. Ras proteins are low molecular weight so called small GTPases (~20-25 kD) and their structure is composed of two domains. First part is the G-domain (or catalytic domain consisting approximately 166 residues), which forms the bulk of the protein. It is responsible for the catalytic activity. The other part is the hyper-variable region. This region consists of a linker domain (a linkage between and catalytic domain and C-terminus; residues 167–179) and the C-terminal part (residues 180–186). C-terminal part undergoes farnesylation and its farnesyl tail targets the proteins to specific compartments of

plasma membranes[145]. The catalytic domain is composed of 5 α -helices, 6 β-strands, and 10 loops[146]. The catalytic domain also can be divided into two lobes: Lobe 1 (residues 1-86) and Lobe 2 (residues 86-166)[147]. Lobe1 is called effector lobe since it interacts with effector proteins and it spans the N-terminal part of the protein. Lobe 2 is the allosteric lobe that forms interactions with the membrane [148].

The three canonical members of RAS family in human are H-Ras, K-Ras and N-Ras. These isoform proteins show 90% sequence identity. The catalytic domain is structurally conserved between the isoforms while the hyper-variable region differs from each other. In detail, in the catalytic domain effector lobe is 100% conserved while the allosteric lobe exhibit small changes in the sequence. Earlier, it was considered that highly homologous catalytic domain serves the same function in all the isoforms. Within years, however, functional specificity of each isoform was investigated by experiments[149]. This functional specificity was considered to be a result of the different localizations of the each isoform on the membrane and subcellular compartments[150-152]. The differences in hyper-variable region of isoforms lead to distinct posttranslational modifications of them. They are localized to different domains on the plasma membrane by direction of these modifications. In addition to localization specificity, the role of the catalytic domain in selective interaction with effector and regulator proteins was suggested to contribute to the functional specificity[153, 154]. An example for selective interaction with regulator proteins is that RAS-GRF (a regulatory protein) activates only H-Ras[155] while RAS-GRP2 (another regulatory protein) activates N-Ras and K-Ras but not H-Ras[156]. Similarly, there are examples for selective interaction with effector proteins, too. K-Ras is a more potent activator of Raf-1 among the other isoforms. Moreover, H-Ras can activate PI3K more efficiently is activated[157].

After the first H-Ras crystal structure was determined in 1988, 130 H- Ras and 54 K-Ras crystal structures have been deposited in the PDB until today. These crystal structures are mostly GDP-bound forms or they are bound to GTP analogues (i.e. GppNHp). All of the structures have only the catalytic domain except recently reported K-Ras structures[158].

X-ray crystal structures of Ras proteins bound with GDP and GppNHp display major conformational changes in the SI and SII regions[144, 159]. Binding of GTP to Ras allows the switch regions to obtain a conformation that Ras can bind to its effectors, e.g. Raf and PI3K. In addition to relative conformations of the two forms, each GDP and GppNHp bound form of Ras also fluctuates in solution. However, crystal packing of the protein restricts those significant fluctuations where crystal packing forces freeze out the protein conformation under the crystallization. Therefore, rigid crystal structures are not enough for understanding the mechanism of Ras-effector protein interactions[160]. To capture the dynamic changes of the protein in solution, NMR spectroscopy is the most suitable technique providing detailed dynamic characteristics of the protein that are mostly lost in crystal structures.

The first Ras structure determined with NMR spectroscopy is catalytic domain of H-Ras bound to GDP has been (PDB ID 1CRP and 1CRR)[161]. Totally 40 determined NMR structures showed that folding topology of the solvated H-Ras structure is the same with the folding topology determined by X-ray crystallization. Moreover, the secondary structure elements -five α-helices and six β-strands- are identical with the crystal structures. NMR structures of H-Ras-GDP reveal also show that the SII region and to a lesser extent the SI region are flexible parts in solution. Furthermore, the superimposition of the averaged NMR structures (PDB ID 1CRQ) and the X-ray crystal structure of H-Ras-GDP (PDB ID 4Q21) shows the similarity of the structures excluding the SI regions.

On the other hand, determination of the wild-type Ras and GppNHp complex structures with

NMR spectroscopy was challenging because of the time scales of NMR experiments. On the NMR time scale, chemical exchange processes at intermediate rates causes dilatation of the residues in the P-loop and SI and SII regions[162]. To overcome to this dilatation, threonine at position 35 is substituted with serine and this mutation successfully could eliminate the slow conformational exchange process[162]. As a result of this, 20 NMR structures of GppNHp bound form of mutant H-Ras^{T35S} were deposited in PDB (PDB ID 2LCF)[163]. Those NMR structures showed the same conformational characteristics with the crystal structures of Ras- GppNHp complexes. Superimposition of those NMR structures indicated that fluctuations of SI region are remarkable while those of SII region are intermediate. These resulting conformational characteristics of Ras-GppNHp are quite different then the NMR structure of Ras-GDP where fluctuations of the switch regions showed the opposite characteristics.

2.3.3 MD simulations of Ras

The first all-atom MD simulation of solvated GTP-bound Ras was done by Foley and colleagues in 1992[164]. They performed the 200 ps-simulation of the solvated crystallographic H-Ras structure of Pai et al[159]. Although this study and other earliest simulations of Ras proteins were very short (max 500 ps) by the current simulation standards, the flexible nature of the nucleotide binding site could be observed[164-168]. Those short-time scale MD simulations helped distinguish the GTP and GDP bound Ras states displaying the distinct conformations of nucleotide binding site residues. Next, a longer (1 ns) simulation of oncogenic G12V mutant of H-Ras exhibit that the fluctuations of nucleotide binding site increased in the mutant form compared to wild-type form[169]. Further studies compared the simulations of Ras isoforms each other using homology models of the N- and K-Ras catalytic domains and they showed the differences between mutant and wild-type isoforms[147]. One

of the difference between K-Ras and H-Ras dynamics is the more flexible character of the nucleotide binding site and helix3-loop7 region in K-Ras than H-Ras that is similar to mutant H-Ras[170]. The flexibility of the loop 7 was also implicated by a following crystallographic study [171]. After that Gorfe and colleagues described the nucleotide and isoform-dependent dynamical features of Ras proteins. They found out isoform-specific allosteric communication pathways between the nucleotide binding region (lobe 1, the effector lobe) and the C-terminus (lobe 2, allosteric lobe)[147]. The allosteric communication pathways connecting the two lobes and their controlling by the nucleotide binding were further identified in the subsequent MD simulationstudies of Ras proteins [172-174]. These studies also revealed transition states between the active and inactive states of Ras [173]. Along with the detailed analyses of the monomer structure of Ras, the dynamics of GAP bound Ras complexes were also analyzed[175, 176]. Resat et. al. showed that binding of GAP to Ras induced the structural reorganization within the nucleotide binding site of Ras that accelerates the GTP hydrolysis[175]. The dynamics of effector-Ras complexes were analyzed by many groups[177-179]. These studies indicated the crucial role of the Ras dynamics in effector binding.

2.4 Direct targeting of K-Ras activity

As a central point of carcinogenesis, K-Ras is an urgent therapeutic target in human cancers. This section describes the strategies for direct targeting of K-Ras that were designed at various levels:

The first strategy is to interrupt the K-Ras-GTP interaction. Nonetheless, the design of competitive inhibitors of K-Ras-GTP binding is considered as not feasible since intracellular concentrations of GTP are very high (millimolar range) and also binding affinity of K-Ras for

GTP is very strong (picomolar range). Since it's not feasible to target GTP binding site, Ostrem and colleagues identified another target site which is adjacent to the active site on K-Ras^{G12C} [30] based on a former strategy for distinguishing G12C mutant from other mutants and WT Ras[180]. Consequently, a specific and irreversible inhibitor of K-Ras^{G12C} has developed[35]. This inhibitor binds to inactive K-Ras^{G12C}-GDP, stabilizes inactive form of K-Ras^{G12C} and thereby blocks the downstream signaling. Along with this, another inhibitor that binds to K-Ras^{G12C}-GDP was discovered by Patricelli and colleagues[29]. This inhibitor blocks the SOS-mediated activation of K-Ras^{G12C} where SOS (son of sevenless) is a GEF protein catalyzes K-Ras GTP/GDP exchange. Actually, this inhibitor of K-Ras^{G12C}-GDP-SOS interaction was developed based on the knowledge of well-defined binding pockets in previous studies[181, 182]. In conclusion, the molecules which can bind to K-Ras and successfully inhibit its activation are limited to K-Ras^{G12C} inhibitors. However, G12C mutation less frequent then the other K-Ras mutations. Thus, we still need inhibitors for the most frequent mutations, G12D and G12C, in fight against cancer.

The second strategy is design of GTP analogs. Ahmadian and collegues designed a modified GTP which can be hydrolyzed by the mutant K-Ras. They discovered that G12 mutants can hydrolyze a GTP analog, Diamino-benzophenone-phosphoroamidate-GTP, more effectively than WT protein[183]. Unfortunately, none of the molecules which are produced following this strategy has passed in vitro studies.

Another strategy for direct targeting of K-Ras activity is *gene therapy*. In order to inhibit K-Ras expression, people developed small antisense nucleotide sequences called small interfering RNA (siRNA) to silence the K-Ras gene directly. Chen and colleagues discovered an oligonucleotide that inhibits K-Ras expression efficiently in vitro[184], but this

oligonucleotide showed high toxicity and could not pass the preclinical tests. After that to reduce the toxicity, a biodegradable polymer was developed and carried by a local siRNA delivery system[185]. However, this system also could not pass the preclinical tests[186].

3 INTRINSIC DYNAMICS OF WILD-TYPE K-RAS: COMPARISION OF ACTIVE AND INACTIVE K-RAS

3.1 INTRODUCTION

K-Ras is the most frequently mutated oncogene in human cancers, but there are still no drugs that directly target it in the clinic. Recent studies utilizing dynamics information show promising results for selectively targeting mutant K-Ras. However, despite extensive characterization, the mechanisms by which K-Ras residue fluctuations transfer allosteric regulatory information remain unknown. Understanding the direction of information flow can provide new mechanistic insights for K-Ras targeting.

In summary, we present a comprehensive study of intrinsic K-Ras dynamics, including detailed analyses of causality between the motions of its residues. We first provide detailed, quantitative descriptions of both GTP- and GDP-bound K-Ras from extensive MD simulations. We use a statistical thermodynamics interpretation of fluctuation correlations to quantify K-Ras 'stiffening' upon activation. Stiffening changes protein dynamics. More importantly, using stiffness calculations jointly with measurements of reduced relative fluctuations, we define protein stability and show that K-Ras is more stable in active conformation. To characterize correlated motions that are persistent within the MD simulations of GTP- and GDP bound K-Ras, we map the correlated motion patterns within

their residues individually, and then compare and discuss their correlation decay time differences in detail. Our results show that inactive K-Ras is marked by a pronounced decrease in correlated motions of residues for shorter periods, while active K-Ras correlations have longer decay times. We analyze the ensuing events at the atomic scale. Finally, to enable a deeper understanding of K-Ras dynamics, we introduce the first causality calculations for K-Ras and predict specific driver/follower residue pairs during protein simulations. We apply a novel approach –CTC – using the motions of all residue pairs of a protein to predict directionality in the allosteric regulation of the protein fluctuations. CTC is the correlation between two time series subject to the condition that fluctuations of the first trajectory are correlated with later fluctuations of the second and thereby predict how past fluctuations of one trajectory affects the future fluctuations of the second. In some cases, CTC function of two trajectories may be asymmetric, with one affecting the other more strongly. We then predict that the fluctuations of a given residue control and modify the fluctuations of the delayed one. Results show the direction of information flow during allosteric modulation of its nucleotide-dependent intrinsic activity: active K-Ras SII region motions drive SI region motions, while α-helix-3L7 motions control both. Our results provide novel insights for strategies that directly target mutant K-Ras.

3.2 METHODS

3.2.1 MD Simulations

We performed all-atom MD simulations for both Mg^{+2}GDP- and Mg^{+2}GTP-bound K-Ras. Crystal structure of K-Ras-GDP was retrieved from PDB (PDB ID: 4OBE) and K-Ras-GTP structure was obtained by changing GDP to GTP by adding a phosphate group to GDP using Discovery Studio 4.5 software, (DS)[187]. Then, the complex was optimized with Clean

Geometry tool of DS. This geometry optimization tool uses a fast, Dreiding-like forcefield to improve the geometry of the selected atoms and results in an approximate three-dimensional structure. We solvated each protein in a TIP3P water box with 12Å buffering distance. We applied periodic boundary conditions and added ions to neutralize the system. We used a 2fs time-step with a 12Å cutoff for Van der Waals interactions and full particle-mesh Ewald electrostatics with 1Å grid spacing and direct space tolerance of 10^{-6}. We carried out all computations in N, P, T dynamics procedure. System temperature was kept constant at the physiological value of 310K using Langevin dynamics with a damping coefficient of $2ps^{-1}$. Constant pressure of 1atm was maintained by The Nose-Hoover Langevin piston method with a 200fs piston period and 100fs decay time. We used NAMD 2.11[76] with AMBER ff99SB[188] and general amber force fields (GAFF)[189]. We obtained parameters of GTP and GDP following the procedure in the next section. The initial system energy was first minimized for 10,000 steps, followed by 10,000 steps for equilibration. After equilibration, we performed 300ns simulations. Atomic coordinates $\hat{R}$ of all atoms were saved every 1ps. To eliminate all rotational and translational motions, we aligned the trajectories to the initial structure by using VMD software 1.9.2[93]. We visualized trajectories using VMD. Additionally, to test whether WT KRAS-GTP complex approached the active (close) state, we monitored the first 100 ns-trajectories of both GTP- and GDP-bound complexes with VMD. We observed that the active site residues obtain relatively close conformations in WT KRAS-GTP. We calculated distance distributions of several residue pairs that flank GTP and observed that they were in support of the relatively close active site conformation.

3.2.2 Parametrization of GTP and GDP

The procedure consisted of two parts: (i) charge derivation; and (ii) preparation of the coordination and parameter files for the complexes. Charge derivation was divided into three sequential steps. First, we optimized the initial GTP and GDP structures by using QM program Gaussian 03 with base 6-31G* and level HF. Second, we performed Molecular Electrostatic Potential (MEP) calculations using Gaussian 03 with base cc-pVTZ and the Density Functional Theory (DFT) method B3LYP in continuum solvent. The third step was Restrained Electrostatic Potential (RESP) charge derivation[190]. We used RESP program available under AMBER[191]. We automated the calculations using the programs RED-vIII, Ante_RED and Antechamber[191-193]. The RESP ESP charge Derive program (R.E.D.) sequentially executed these three steps by interfacing Gaussian and RESP programs, and allowed the automatic derivation of RESP and ESP charges for GTP and GDP. Ante_RED was useful to prepare input files for Gaussian. We used RED-vIII to perform MEP calculations and RESP charge derivations. Antechamber wrote out additional force field files (frcmod file) of molecules with missing parameters. For basic model building and Amber coordinate and parameter/topology file creation we used the LEaP module of AMBER 14 package[74].

3.2.3 Calculation of fluctuations

MD trajectories describe the time evolution of a system and are defined by the position vector $R_i(t)$ *of* every atom *i* at every time point *t* where $1 \leq t \leq N_T$ and N_T is the total number of *t* time points of the trajectory. The time average $\bar{R}_\iota$ position of each atom *i* is then defined as $\bar{R}_i = \sum_t^{N_T} R_i\,(t)/N_T$. The instantaneous fluctuation $\Delta R_i\,(t)$ of the position vector of the *i*th atom is defined as $\Delta R_i(t) = \Delta R_i(t) - \bar{R}_i$. The root mean squared fluctuation (RMSF) of a

residue is $\left\langle \left(\Delta R_i\right)^2 \right\rangle^{0.5}$, where the angular brackets show the time average of the quantity enclosed. The RMSF of each atom is proportional to its Debye-Waller (B-factor) measured experimentally.

3.2.4 Stiffness

We quantified nucleotide-bound K-Ras stiffness using a statistical thermodynamics interpretation of fluctuation correlations[194]. We assumed that the interaction between two fluctuating residues *i* and *j* can be represented by a spring, where the spring constant follows from the Gaussian Network Model (GNM)[195]: $k_{ij} = \frac{k_\beta T}{\langle(\Delta R_i)^2\rangle - 2\langle(\Delta R_i \Delta R_j)\rangle + \langle\left(\Delta R_j\right)^2\rangle}$ where ΔR_i is the instantaneous fluctuation of one end of the rod, ΔR_j is the fluctuation of the other end, k_B is the Boltzmann constant and *T* is the absolute temperature. The spring constant has dimensions of force/length. In GNM spring definition, each residue *i* is attached to *N-1* other residues via *N-1* springs[195]. Thus, how stiffly a residue *i* is attached to a protein can be quantified by $\bar{k}_i = \sum_j k_{ij}/N - 1$ where $\bar{k}_i$ is the mean spring constant for a residue *i*. For stiffness estimates of the complete complexes, we define an overall stiffness parameter $k_{overall}$ by the expression $k_{overall} = \sum_{j>i} k_{ij}$. To estimate the stiffness differences in active versus inactive K-Ras, we calculated $\bar{k}_i$ for each residue and $k_{overall}$ for the protein for both states.

3.2.5 Stability

We defined the stability of an interacting system of residues as the joint state of reduced RMSF and increased interaction stiffness. RMSF relates to the magnitude of fluctuations of individual residues and stiffness relates to the distance between two residues and therefore

they are two independent quantities. A small RMSF and a high stiffness denote increased stability.

3.2.6 Distance distributions between residue pairs

We calculated the distance between two residues (*i,j)* as $R_{ij} = \sqrt{(R_i(t) - R_j(t))^2}$. Residue pair distance distributions $W(R_{ij})$ were calculated by dividing the maximum distance between the pair into small bins and counting the number of observed distances in each bin. All distributions were normalized.

3.2.7 Time independent correlations (cross-correlation coefficient map)

Correlations intrinsic to K-Ras structure are defined by the cross-correlation coefficient map, $C(\Delta R_i, \Delta R_j)$:

$$C_{ij} = C(\Delta Ri, \Delta Rj) = \frac{\langle \Delta R_i(t) \cdot \Delta R_j(t) \rangle}{\langle (\Delta R_i(t))^2 \rangle^{1/2} \langle (\Delta R_j(t))^2 \rangle^{1/2}} = \frac{\sum_{t=1}^{N_t - \tau} \Delta R_i(t) \cdot \Delta R_j(t)}{\left[\sum_{t=1}^{N_t - \tau} (\Delta {R_i}^2(t))\right]^{1/2} \left[\sum_{t=1}^{N_t - \tau} (\Delta {R_j}^2(t))\right]^{1/2}}$$

where · denotes the dot product. Correlation varies between -1 and 1. If motions of two atoms are independent, then $\langle \Delta R_i(t) \cdot \Delta R_j(t) \rangle = 0$ and C_{ij}= 0. If the atoms always move in parallel in the same direction, then they are perfectly positively correlated, and C_{ij} = 1. If they always move in parallel in opposite directions, they are perfectly negatively correlated, and C_{ij} = -1. Cross-correlation coefficients lie in the range of -1 ≤ C_{ij} ≤1.

3.2.8 Time delayed correlations, mobility and causality

Time-delayed correlation of two fluctuations is defined by:

$$C_{ij}(\tau) = \frac{\langle \Delta R_i(0) \cdot \Delta R_j(\tau) \rangle_\chi}{\langle (\Delta R_i)^2 \rangle^{1/2} \langle (\Delta R_j)^2 \rangle^{1/2}}$$

where χ denotes that the average is a conditional average calculated according to $\langle \Delta R_i(0) \cdot \Delta R_j(\tau) \rangle_\chi = \sum_k p(\Delta R_j(t_k + \tau) | \Delta R_i(t_k)) P(\Delta R_i(t_k)) \Delta R_i(t_k) \Delta R_j(t_k + \tau)$.

Here, $p(\Delta R_j(t_k + \tau) | \Delta R_i(t_k))$ denotes the conditional probability of observing $\Delta R_j(t_k + \tau)$ at time $t_k + \tau$, given $\Delta R_i(t_k)$ at time t_k.

Similarly, if indices are exchanged, then $C_{ji}(\tau)$ represents the correlations of ΔR_i at time $t + \tau$ with earlier ΔR_j values at time t. If the fluctuations of residue i drive the fluctuations of residue j, then $C_{ij}(\tau) > C_{ji}(\tau)$. If $C_{ji}(\tau) > C_{ij}(\tau)$, residue j drives residue i because the fluctuation ΔR_j at time t is correlated with future fluctuations of ΔR_i. However, at τ=0, the equality $C_{ij}(0)$=$C_{ji}(0)$ holds.

Note that time-delayed autocorrelation $C_{ii}(\tau)$ is the correlation of the trajectory with its own past and future coordinates. If autocorrelation is large, it can correspond to a specific form of "persistence", a tendency for a system to remain in the same state from one observation to the next.

3.3 RESULTS AND DISCUSSION

3.3.1 Comparison of stiffness changes in active and inactive K-Ras

GTP binding increases K-Ras stiffness. To understand how nucleotide binding affects K-Ras dynamics, we quantified changes in its 'stiffness' – a metric that inversely correlates with

residue pair fluctuations - upon GTP vs. GDP binding. For this purpose, we represented the interaction between two fluctuating residue pairs (*i* and *j)* as a spring with a constant k_{ij}, and related its magnitude to the mean square fluctuations of residues *i* and *j* and to their cross-correlations using the Gaussian Network Model (GNM). GNM is a coarse grained model at the residue level but it has been used widely for predicting protein behavior[195]. Plotting this spring constant for every residue pair in both GTP- (Figure 3.1A) and GDP-bound (Figure 3.1B) K-Ras, we observe strong coordination in the fluctuations of GTP phosphate groups with those of K-Ras (Figure 3.1A).

To zoom in on and directly compare the effects of nucleotide binding on K-Ras stiffness, we calculated the differences in spring constant values between GTP- and GDP- bound K-Ras. In the following paragraphs we show that the spring constants calculated in this way are in agreement with experimental findings. In Figure 3.1C, red dots indicate that the differences are largely due to the stiffening effects of GTP-binding on residue pair fluctuations. Notice that Regions 1-3 in Figure 3.1C that correspond to secondary structures show significant increase in k_{ij} when GTP-bound. Furthermore, Region 1 corresponds to strong coordination of β2 and β3 motions, while Regions 2 and 3 correspond to increased stiffness of β4-α3 and α4.

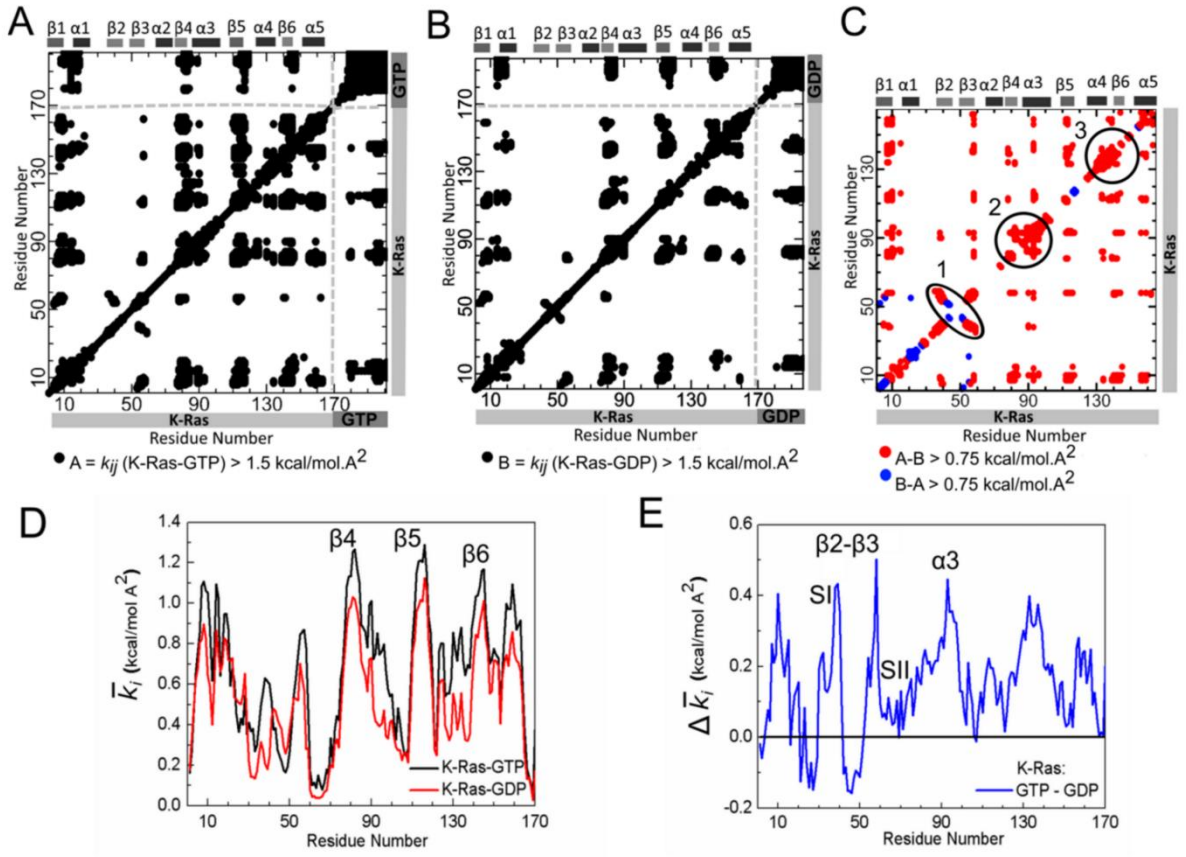


Figure 3.1 Stiffness results for GTP- and GDP-bound K-Ras and their difference. In panels A and B, both axes marks 1-169 represent the residue Cα atoms of K-Ras and marks 170-on represent GDP and GTP nucleotide heavy atoms, respectively with k_{ij} >1.5 kcal/mol·A^2. (A) k_{ij} for K-Ras-GTP. Atoms 170-181 are the γ, β, α-phosphate groups and 182-201 are the guanine atoms of GTP. (B) k_{ij} for K-Ras-GDP. Atoms 170-178 are the β and α phosphate groups and 178-197 are the guanine atoms of GDP. (C) Difference between active and inactive K-Ras k_{ij} values. Red regions are stiffer in K-Ras-GTP (k_{ij} values of K-Ras-GTP > K-Ras-GDP by at least 0.75 kcal/mol·A^2) and blue regions are stiffer in K-Ras-GDP (k_{ij} values of K-Ras-GDP > K-Ras-GTP by at least 0.75 kcal/mol·A^2). (D) Mean spring constants $\bar{\boldsymbol{k}}_{\boldsymbol{i}}$ for GTP and GDP bound states. (E) Mean spring constant differences $\boldsymbol{\Delta \bar{k}_i}$ for GTP and GDP bound states. Positive values correspond to larger mean stiffness in K-Ras-GTP.

Nucleotide binding affects spring constant of α2 (SII). We next investigated the effects of nucleotide binding on the spring constant of α2 (SII), because previous studies have shown that stiffness increases when SII refolds into an α-helical conformation through GTP binding[196]. We calculated the spring constants of the two terminal residues of α2 (A66 and T74), which were 0.10 kcal/mol·A^2 (69.91 pN/nm) for active and 0.04 kcal/mol·A^2 (27.78

pN/nm) for inactive K-Ras. Previous studies have utilized various experimental methods that have all led to spring constants within ~0.09-1.15 kcal/mol·A^2 (60–80 pN/nm) for helices[197, 198]. Our results for both K-Ras forms are on the same order of magnitude. Note that for active K-Ras the α2 spring constant is equalent to the characteristic spring constant of α-helices, while it is lower in inactive form. Hence, our results validate and quantify earlier, qualitative observations of Noe *et. Al.*[196] that the α2 spring constant reaches to the level of an α-helix spring constant during GTP binding.

Overall spring constant is higher in active complex. To estimate global changes in stiffness in response to nucleotide binding, we calculated the overall spring constants $k_{overall}$ *(details in Methods)* of nucleotide-K-Ras complexes, which were 0.70 kcal/mol·A^2 (481.75 pN/nm) for GTP-bound, and 0.55 kcal/mol·A^2 (385.12 pN/nm) for GDP-bound K-Ras. Both are of the same order of magnitude with an experimental study for another protein, myoglobin, which has an overall spring constant of ~300 pN/m[199, 200] pointing to an order of magnitude agreement of overall stiffnesses of proteins in general. In conclusion, GTP-binding increases the overall K-Ras stiffness. In other words, GTP-binding decreases pairwise residue fluctuations of K-Ras overall, making the protein more rigid.

Secondary structure motions show the strongest coordination with the rest of the protein. Quantifying the spring constant based on fluctuations allows for analyzing how, analogous to a virtual spring, the fluctuations of a specific residue are coupled with fluctuations of rest of the protein. To discover residues whose fluctuations are in strong coordination with K-Ras fluctuations and how they change between the two states, we compared the mean spring constant $\bar{k}_i$ of each residue *i*, for both active and inactive K-Ras (Figure 3.1D) as described in *Methods*. A large $\bar{k}_i$ value indicates that the motions of residue *i* are stiffly coupled with

protein motions; while a small $\bar{k}_i$ value indicates that the motions of the *i*th residue and the protein are flexibly coupled. For simplicity, we categorized the significant mean spring constant $\bar{k}_i$ values as highest, high and smallest (For details please see Table 3.1). In both states, the highest $\bar{k}_i$ values are of β-strand residues β4, β5 and β6, showing the strongest coordination of their motions with K-Ras motions. Next, high $\bar{k}_i$ values of β1, the P-loop and α5 residues indicate that their fluctuations are also strongly coupled with those of the protein. On the other hand, the smallest $\bar{k}_i$ values belong to SII region in active and SI and SII regions in inactive K-Ras which show that their residue fluctuations are not correlated with the rest of the protein (Table 3.2). Since we have defined the stiffness metric as a signifier of a decrease in residue fluctuations, we provide a second line of proof that increased stiffness stabilizes dynamic fluctuations in both forms of K-Ras by using RMSF graph (Figure 3.2). Clearly, the residues with the smallest mean spring constant $\bar{k}_i$ values from Figure 3.1D have the highest RMSF values in Figure.3.2 and vice versa.

Table 3.1 Categorization of $\bar{k}_i$ values

$\bar{k}_i$ RANGE	$\bar{k}_i$ VALUES (kcal/mol·A^2)		CATEGORY
	K-Ras-GTP	K-Ras-GDP	
%100	1.28 (Max)	1.08 (Max)	The Highest
%90	≥ 1.16	≥ 0.97	
%80	≥1.04	≥ 0.86	High

%10 %100	≤ 0.19 Min 0.08	≤ 0.14 Min 0.03	The Smallest

Table 3.2 Categorization of residues according to their mean spring constant $\bar{k}_i$ values for K-Ras-GTP (left) and K-Ras-GDP (right).

	K-Ras-GTP		K-Ras-GDP	
	Region	Residues	Region	Residues
The Most Rigidly Attached	β4	G77-I84	β4	G77-I84
	β5	P110-N116	β5	P110-N116
	β6	P140-T144	β6	P140-T144
Rigidly Attached	β1	L6-G10	β1	L6-G10
	P-loop	V14-G15	P-loop	V14-G15
	α5	F156-V160	α5	F156-V160
The Most Flexibly Attached	SII	G60-M67	SI	D30-P34
			SII	A59-R68

As indicated in previous studies where NMR and Atomic Force Microscopy were used, protein stiffness depends on secondary structure [199, 200], where loops contribute to structural flexibility and show large fluctuations, while β-strands and α-helices provide mechanical stability and show small fluctuations[199]. Our K-Ras results are consistent with these general observations. In addition, we observe stiff coupling of the fluctuations of the P-loop and the protein. This observation is important since the P-loop is the phosphate binding site of K-Ras and connects β1 and α1. Although loops are often flexible regions of proteins and show higher fluctuations, in K-Ras, motions of the P-loop residues are stiffly coupled to those of the protein, especially in active state ($\bar{k}_i$ =1.08 kcal/mol·A^2 for K-Ras-GTP, $\bar{k}_i$=0.85 kcal/mol·A^2 for K-Ras-GDP).

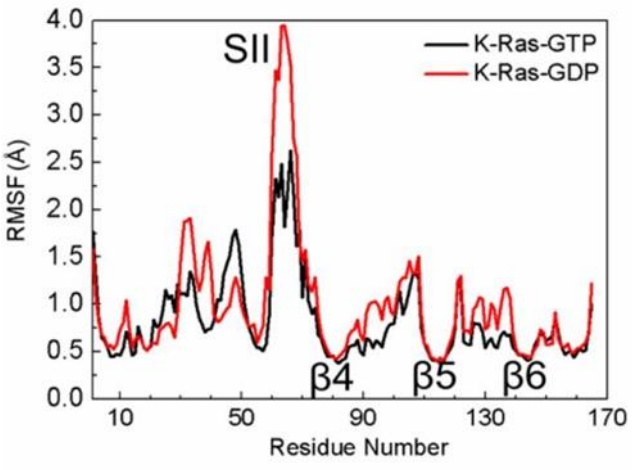


Figure 3.2 Comparison of RMSF values of K-Ras motion in active form with inactive form. RMSF values of active form (black) and inactive form (red).

The mean spring constant values of residues in β2, β3, α3 and switch regions –especially SI- are higher in active K-Ras than in inactive K-Ras. Finally, we calculated mean spring constant differences between GTP-bound active and GDP-bound inactive K-Ras, $\Delta \bar{k}_i$ ($\bar{k}_{i\ \text{K-Ras-GTP}}$ - $\bar{k}_{i\ \text{K-Ras-GDP}}$). Figure 2E shows that the fluctuations of β2 and β3 terminal (D38 and D57) and α3 center (D92-I93) residues are in stronger coordination with those of active K-Ras

(vs. inactive K-Ras) ($\Delta\bar{k}_i$ > 0.43 kcal/mol·A^2). Our results also indicate that although residues of switch regions have the smallest $\bar{k}_i$ values in both forms, some of their $\bar{k}_i$ values increase significantly in active form. In Figure 2E, $\Delta\bar{k}_i$ ranges between 0.20-0.36 kcal/mol·A^2 for residues in SI (D30-R41) and 0.02-0.19 kcal/mol·A^2 for residues in SII (G60-T74). These $\Delta\bar{k}_i$ values show stiffer coupling of the motions of GTP-K-Ras with the motions of switch residues, especially SI (vs GDP-K-Ras). This result is important as SI includes the binding site to effector proteins which only bind to GTP-bound K-Ras when SI flexibilty is reduced[201]. Earlier studies that used NMR spectra and RMSF calculation also support our results that GTP binding reduces the flexibility of both SI and SII, especially SI[202, 203]. Our results improve on this information by showing that fluctuations of switch regions – notably SI- are more stiffly coupled with K-Ras-GTP fluctuations (Figure 3.1E).

3.3.2 Comparison of residue pair correlations for active and inactive K-Ras

To identify if the fluctuations of one residue are related to fluctuations of another residue, we calculated the correlations of all residue-residue pairs in both GTP- vs GDP bound K-Ras complexes. As expected, cross-correlation coefficient maps of K-Ras-GTP (Figure 3.3A) and K-Ras GDP (Figure 3.3B) exhibit different correlation characteristics. The most remarkable differences between Figure 3A and 3B belong to two parts: (i) the correlation of α1-SI with L10-α5 and (ii) the correlations between β2 and β3. Positive correlation patterns within these two parts are evident in K-Ras-GTP simulations, but absent in K-Ras GDP simulations. To provide comprehensive information on nucleotide-dependent K-Ras dynamics, we present these two remarkable results from correlation analyses (Figure 3.3) as well as sources of correlated motions (i.e. H-bonds) together in the following sections.

The correlation of α1-SI with L10-α5 in active K-Ras motions is due to three specific H-bonds. MD simulations show that the correlation between α1-SI and L10-α5 in the active form results from GTP binding to active site residues, which also form specific H bonds with other K-Ras residues and water. Based on the average number of H-bonds each residue forms throughout the simulation, we estimated that the nucleotides remain bound to active site residues S17, D30, D119 and K147, and that GTP-binding (vs. GDP) is more stable for S17 and D30 (Table 3.3). Furthermore, correlated motions of α1-SI and L10-α5 in GTP-bound K-Ras originate specifically from three H-bonds: (i) A146-Q22, (ii) D30-GTP, (iii) D30-a water molecule. We observed a sustained H-bond between A146-Q22 during active but not in inactive complex simulation. This suggests that A146-Q22 interaction causes a strong relationship between L10α5 (A146-D154) and α1 (L19-I24) in active K-Ras (Figure 3.3A) with a correlation coefficient of 0.75, and a weak correlation coefficient of 0.28 for inactive K-Ras. At the same time, the active site residue D30 forms an H-bond with the nucleotide in both active and inactive K-Ras, while it also binds to a water molecule only in the active form. However, the H-bond in the active form between D30(O)-GTP(O2A) is more permanent than the H-bond in the inactive form between D30(O)-GDP(O2').

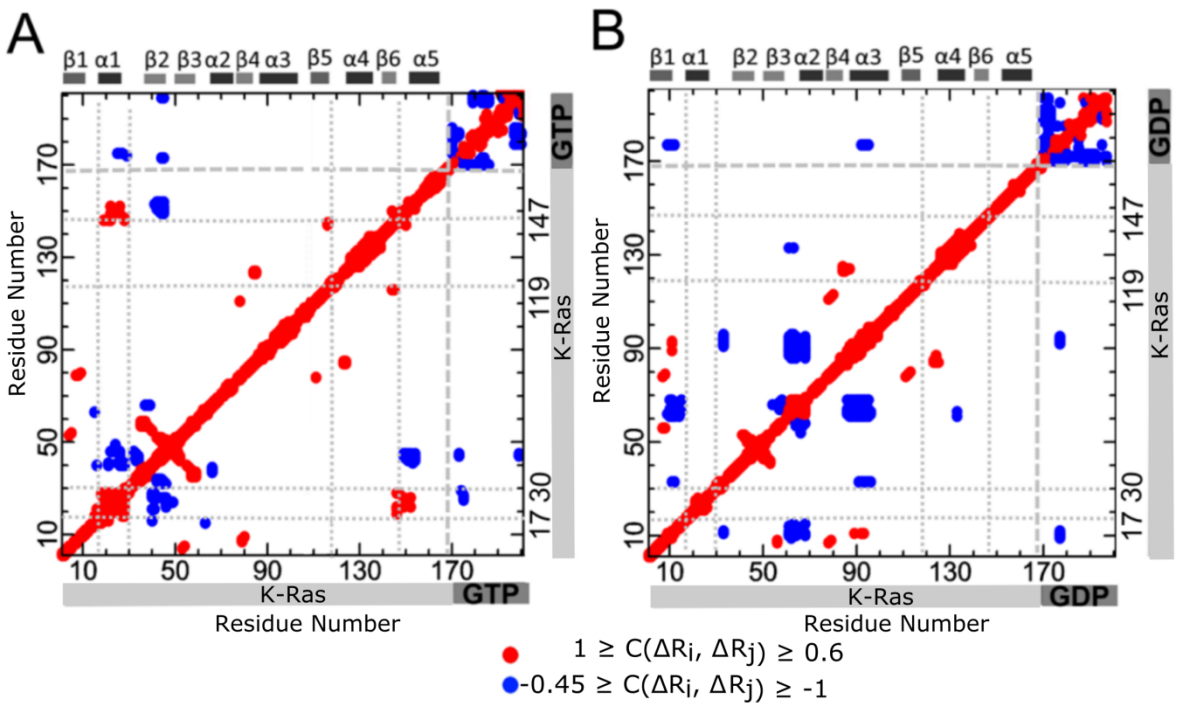


Figure 3.3 Cross-correlation coefficient maps for GTP and GDP bound states. Red dots show positive correlations ($1 \geq C(\Delta Ri, \Delta Rj) \geq 0.6$) and blue dots show negative correlations ($-0.45 \geq C(\Delta Ri, \Delta Rj) \geq -1$). Residues indices 1-169 refer to K-Ras. (A) Correlated fluctuations of K-Ras-GTP. Indices between 170-201 refer to GTP heavy atoms (182-201 are guanine atoms). (B) Correlated fluctuations of K-Ras-GDP. Indices between 170-197 refer to GDP heavy atoms (178-197 are guanine atoms).

Since H-bond of D30-GTP is effective throughout the full trajectory, the D30-GTP distance is invariant and the fluctuation correlations of the D30-GTP have longer decay times during K-Ras-GTP simulation. We next combined cross-correlation results with the distance distribution of D30 and nucleotides and quantified the decay times of their correlations during MD simulations. In addition to more permanent binding of D30(O)-GTP(O2A), nucleotide-D30 distance distribution pattern is close to the normal distribution curve with a mean of a smaller value in active K-Ras (Figure 3.4A), with a correlation coefficient of 0.97. To quantify decay time of this correlation in both complexes, we first defined two "connectivity vectors", $\Delta R_{30\text{-}GTP}$ and $\Delta R_{30\text{-}GDP}$, between D30(O) and nucleotides. As illustrated in Figure 3.4B, $\Delta R_{30\text{-}GTP}$ connects the starting point of fluctuation vector of $\Delta R_{D30(O)}$ to end point of

negative $\Delta R_{GTP(O2A);}$ $\Delta R_{30\text{-}GDP}$ starts from $\Delta R_{D30(O)}$ to negative $\Delta R_{GDP(O2')}$. We then calculated time-delayed autocorrelations of each connectivity vector throughout the MD simulations. The autocorrelation plot in Figure 3.4B summarizes the correlation of connectivity vectors at various time delays, where vector correlation coefficients are plotted with 1 ns delays at a time; slow decay of correlations in active K-Ras is clearly observed. Correlations decay to 1/e in about 3 ns for K-Ras-GDP (red line), vs. to ~10 ns for K-Ras-GTP (black line). One reason for this slow correlation decay is the H-bond, which binds D30 to a water molecule in active K-Ras. The O atom of D30 establishes an H-bond with the nearest water during 28% of the trajectory while it does not make any contact with waters when K-Ras is inactive.

Table 3.3 Comparison of the average number of hydrogen bonds formed throughout the simulation between the nucleotide and K-Ras.

	Average number of H bonds throughout the simulations	
Residue of K-Ras	GTP	GDP
SER17	0.45	0.20
ASP30	1.00	0.61

A continuously acting H-bond stabilizes β2-β3 distance and promotes longer decay times for β2-β3 correlations during K-Ras-GTP simulation. β2 and β3 are two parallel β strands located between SI and SII regions (Figure3.5A). Due to the presence of a persistent H-bond between R41(β2)-D54(β3) in K-Ras-GTP simulation, the peak value of $R_{41\text{-}54}$ distribution decreases (Figure 3.5B) and fluctuations of β2 and β3 become correlated (Figure 3.5A). Time-delayed autocorrelations of the vector $\Delta R_{38\text{-}57}$ between their terminal residues D38 and D57

are presented in Figure 3.5C showing that $\Delta R_{38\text{-}57}$ correlation decays much more slowly in active K-Ras.

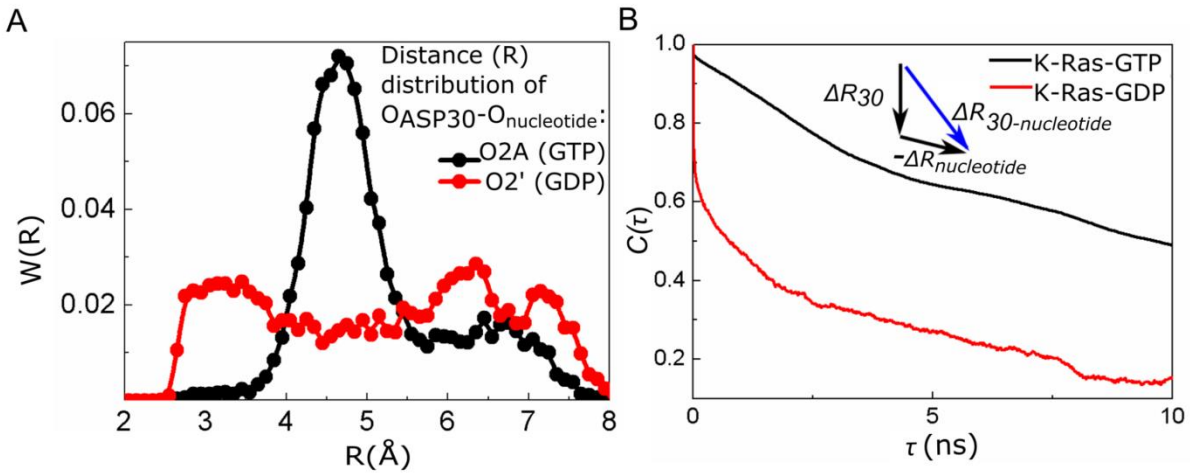


Figure 3.4 D30-GTP distance is more stable than that of D30-GDP. Fluctuations of D30(O) to GTP(O2A) "connectivity vector" are persistently correlated. (A) Distance distribution between D30 and connecting O atoms of GTP (black) and GDP (red) (B)Time delayed autocorrelations for the vector connecting Oxygen atom of D30 to O2A of GTP (black curve) and O2' of GDP (red curve). X-axis is the time delay (τ) and Y-axis is the time delayed autocorrelation of the vector for τ.

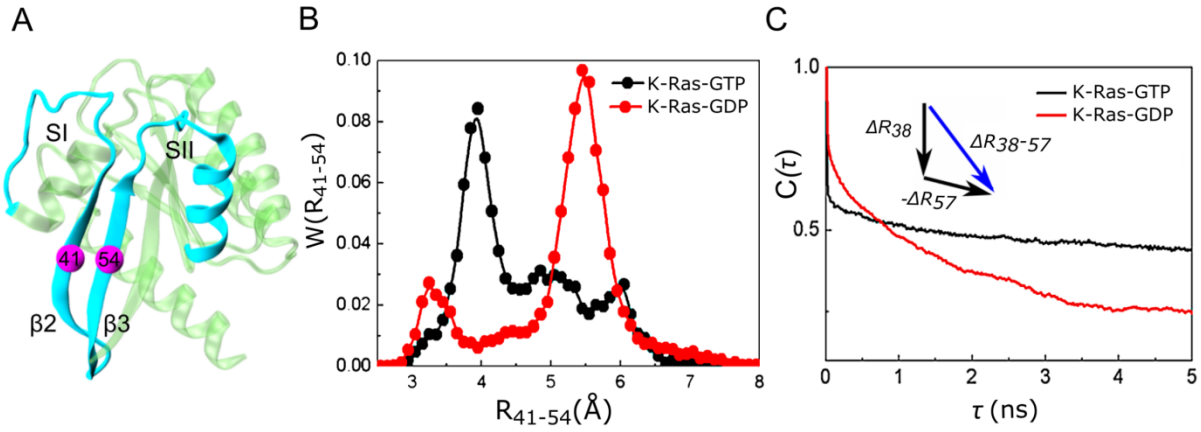


Figure 3.5 Correlation of β2 and β3 fluctuations is persistent in active K-Ras. (A) Locations of R41 (β2) and D54 (β3) relative to SI & SII. (B) Distance distribution between Cα atoms of R41 and D54 in K-Ras-GTP (black) and K-Ras-GDP (red). Distance values between R41 (β2) and D54 (β3) populate at 3.90 Å during GTP binding; they populate at 5.46 Å for GDP-bound K-Ras. (C) Time delayed autocorrelations for the fluctuations of the vector from D38 (Cα) to D57 (Cα).

3.3.3 Causality of Correlated Motions

Correlated motions of proteins often have a direction or causal relationship[42]. Correlations in the fluctuations of two residues indicate interaction, which is necessary for allosteric transitions. However, this is not sufficient for understanding the dynamic phenomenon completely since these symmetric correlations do not contain information on driver and follower relationships. To deduce causality, CTCs need to be analyzed. Our observation is supported by recent work[42] that identified causality in correlated motions from MD simulations using an information theory measure of transfer entropy. This work, in turn, was built on a study by Schreiber, who introduced the *entropy transfer* concept for fluctuating environments[121]. We follow up on these ideas and introduce a new method to dissect dynamic correlations of all residue pairs of a protein to identify driver and follower residues. For this purpose, we evaluate strong time-delayed (τ=5ns) correlations between residue pairs. The strongest causal relations are as follows (Figure 3.6):

SII motions drive SI in active K-Ras. SI-SII relationship is better understood by examining residues that drive their motions throughout the trajectory. Our causality calculations show that SI is driven by SII (Figure 3.6A and 3.7). We present *CTC* plots of R68(SII) with V29(SI) (Figure 3.7A) and with P34(SI) (Figure 3.7B) for active K-Ras. Red curve shows that the fluctuations of R68 at time t affect the fluctuations of V29 at time $t+\tau$. Fluctuation decay of K-Ras residues is in the order of 1ns. The red curve persists for time periods that are an order of magnitude longer. The reverse does not show a significant correlation: V29 does not correlate with later fluctuations of R68. Previously, a study reported that SI loop at residues 29–34 swings into the water using V29 and P34 as hinges during Ras inactivation[196]. We improved on this information by calculating time-delayed correlations and identified that SII

residues - especially R68 and D69- sustain active state conformation of SI by driving the motions of hinge residues V29 and P34. Another study also assessed the conformational transition of Ras from inactive to active state[147], where displacement of SII triggers the active state transition and SI follows SII after a lag time of multiple nanoseconds. Dominance of SII region motions was also observed in several studies[204, 205]. The nucleotide-bound form behavior is regulated by the relative arrangement of the two switches, rather than their individual conformations. We quantified this by verifying that SI fluctuations follow SII fluctuations in K-Ras-GTP. Since from information theoretic point of view correlations are regarded as information sources, we conclude that information flows from SII to SI. The directionality originates from the differences in the characteristic decay times. The problem is one of dynamics within few nanosecond time periods. Disruption of this flow is expected to interfere with the switch mechanism function, which is the basis of K-Ras activity.

α3 and Loop 7 (L7) motions drive switch region (SI & SII) motions in active K-Ras (Figure 3.6B). Fluctuations of the helical dimer interface residues of α3, E98 and R102[206] drive fluctuations of A66 (α2; SII), as shown in Figures3. 7C and 3.7D. Additionally, helical dimer interface residue S106 (L7) drives the motion of Y71 (α2; SII) (Figure3. 7E). On the other hand, fluctuations of R102 (α3) and S106 (L7) drive SI residues N26, D30, Y32 (Figures 3.7F to 3.7H).

Correlated motions of α2 and α3-L7 have been described in other studies, which also emphasized the necessity of understanding their effect on protein function[33, 204, 207]. We contribute to this knowledge by identifying their cause and effect relations. Furthermore, in previous studies, starting from the allosteric interaction between α2 and α3-L7, a novel ligand binding pocket, termed p3, which includes residues of L7 was defined and targeted for lead

generation[33, 208, 209]. It was reported that ligand binding to p3 pocket weakens effector protein binding by allosterically stabilizing Ras effector binding site (SI). Another proposed allosteric mechanism is that ligand binding to p3 pocket changes the switch region conformation. Our results suggest that allosteric modulation of ligand binding may freeze the fluctuations of L7 and stabilize SI motions. This is based on our finding that motions of effector binding site (D30-Y32) are driven by S106 (L7).

β2 and β3 both drive and follow residue motions in active K-Ras. Causality calculations suggest the following information flow in fluctuations: ILE21-GLN22 (α1) drives β2-β3 (Figure 3.6 and 3.8); which drives Y157 (α5), Q61 (SII) and T74 (SII) (Figure 3.6 and 3.8, with details in Table 3.4). Specifically, the differences in the characteristic decay times in Figure 3.8A-B demonstrate that information flows from β2-β3 to Y157 (α5). These findings improve on the previous observations of Abankwa et al. where they defined β2–β3 and α5 as a novel conformational switch[210]. Most importantly, we showed that Q61 (SII) motions follow E49 motions (β2-β3) (Figure 3.8D). Abankwa et al. also observed that mutations in D47-E49 cause hyperactive Ras. Our findings support this too by showing that fluctuations of E49 of the wild type cause fluctuations of the catalytic residue Q61 within SII, whose proper positioning is essential for effective catalysis[162]. Based on these results, we suggest that mutations in D47-E49 region may alter E49 fluctuations that cause improper Q61 fluctuations. Therefore, GTP catalysis is disrupted which results in constituently active K-Ras.

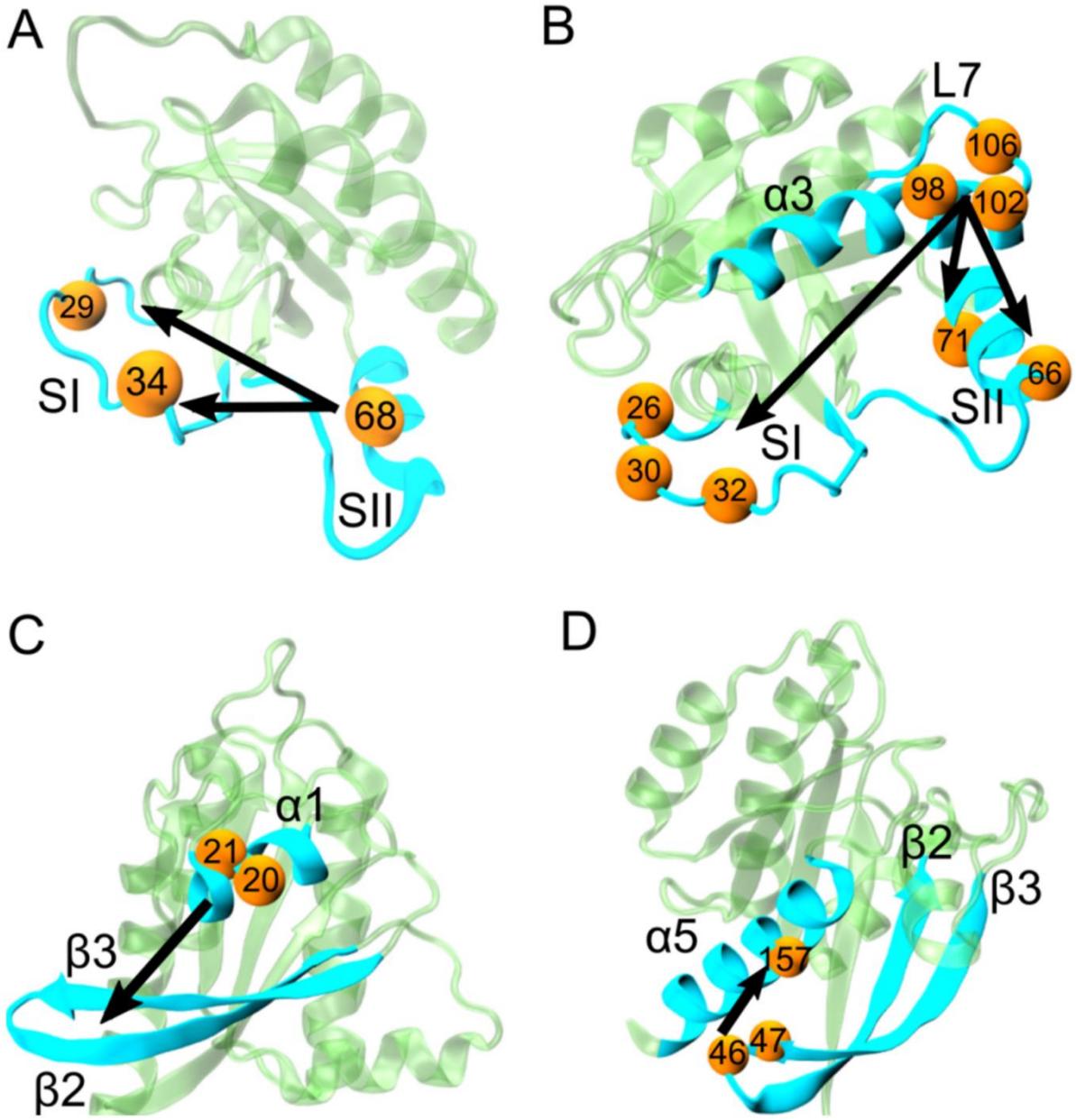


Figure 3.6 Causality relations in active K-Ras motions. Directionality in causal relationships is illustrated with arrows. Arrows start from driver residues and end at follower residues. Both residue types are represented with yellow spheres and marked with their residue numbers. The secondary structures they belong to are in turquoise. A) R68 (SII) drives V29 and P34(SI). B) E98 and R102 (α3) drive A66 (α2; SII). S106 (L7) drives Y71 (α2; SII). R102 (α3) drives N26 and Y32(SI). S106 (L7) drives D30. C) ILE21-GLN22 (α1) drives β2-β3. D) I46 and D47 (β2-β3) drive Y157 (α5).

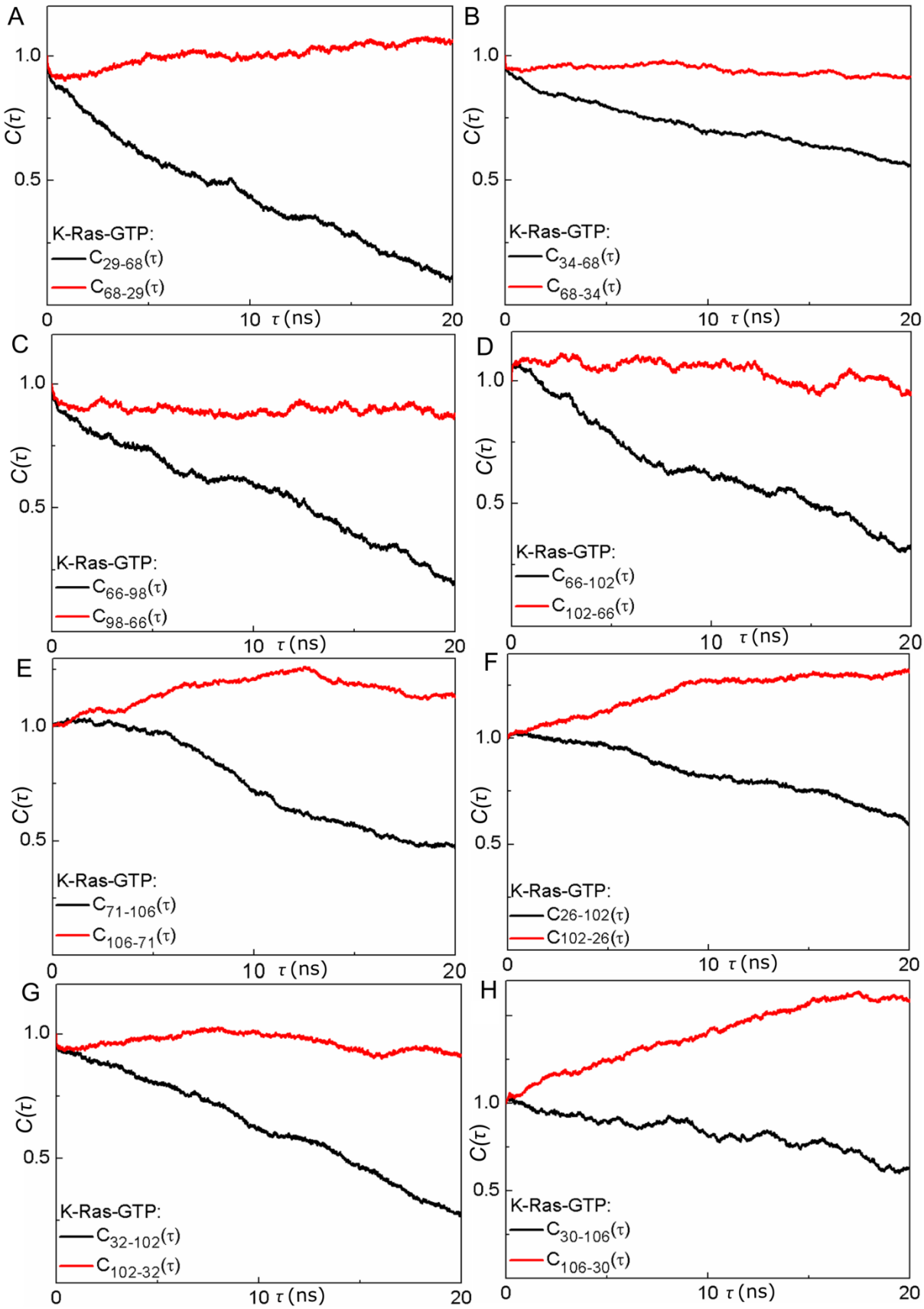

A
1.0
0.5
C(τ)
K-Ras-GTP:
C29-68(τ)
C68-29(τ)
0
10 τ (ns)
20
B
1.0
0.5
C(τ)
K-Ras-GTP:
C34-68(τ)
C68-34(τ)
0
10 τ (ns)
20
C
1.0
0.5
C(τ)
K-Ras-GTP:
C66-98(τ)
C98-66(τ)
0
10 τ (ns)
20
D
1.0
0.5
C(τ)
K-Ras-GTP:
C66-102(τ)
C102-66(τ)
0
10 τ (ns)
20
E
1.0
0.5
C(τ)
K-Ras-GTP:
C71-106(τ)
C106-71(τ)
0
10 τ (ns)
20
F
1.0
0.5
C(τ)
K-Ras-GTP:
C26-102(τ)
C102-26(τ)
0
10 τ (ns)
20
G
1.0
0.5
C(τ)
K-Ras-GTP:
C32-102(τ)
C102-32(τ)
0
10 τ (ns)
20
H
1.0
0.5
C(τ)
K-Ras-GTP:
C30-106(τ)
C106-30(τ)
0
10 τ (ns)
20

Figure 3.7 SII fluctuations drive SI fluctuations; α3-L7 motions drive switch region (SI & SII) motions in K-Ras-GTP. Red curves for $\langle \Delta \boldsymbol{R_i}(\boldsymbol{t}) \Delta \boldsymbol{R_j}(\boldsymbol{t}+\boldsymbol{\tau}) \rangle$ show that the fluctuations of residue *i* at time *t* affect the fluctuations of residue *j* at a later time *t*+*τ*. All correlations ($\boldsymbol{C}(\boldsymbol{\tau})$) are normalized with respect their value at zero ($\boldsymbol{C}(\boldsymbol{0})$). (A) R68 (SII) drives V29(SI). (B) R68 drives P34(SI). (C) E98(α3) drives A66 (α2; SII). (D) R102 (α3) drives A66. (E) S106 (L7) drives Y71 (α2; SII). (F) R102 (α3) drives N26(SI). (G) R102 drives Y32(SI). (H) S106 (L7) drives D30(SI).

Table 3.4 The β2-β3 region acts as both the driver and the follower in active K-Ras.

DRIVER RESIDUE(S)		FOLLOWER RESIDUE(S)	
RESIDUE(S) NAME	REGION	RESIDUE(S) NAME	REGION
I21-GLN22	α1	Q43-L53	β2- β3
I46-CYS51	β2- β3	Y157	α5
I46-GLY48	β2- L3	T74	α2; SII
E49	β3	Q61	SII

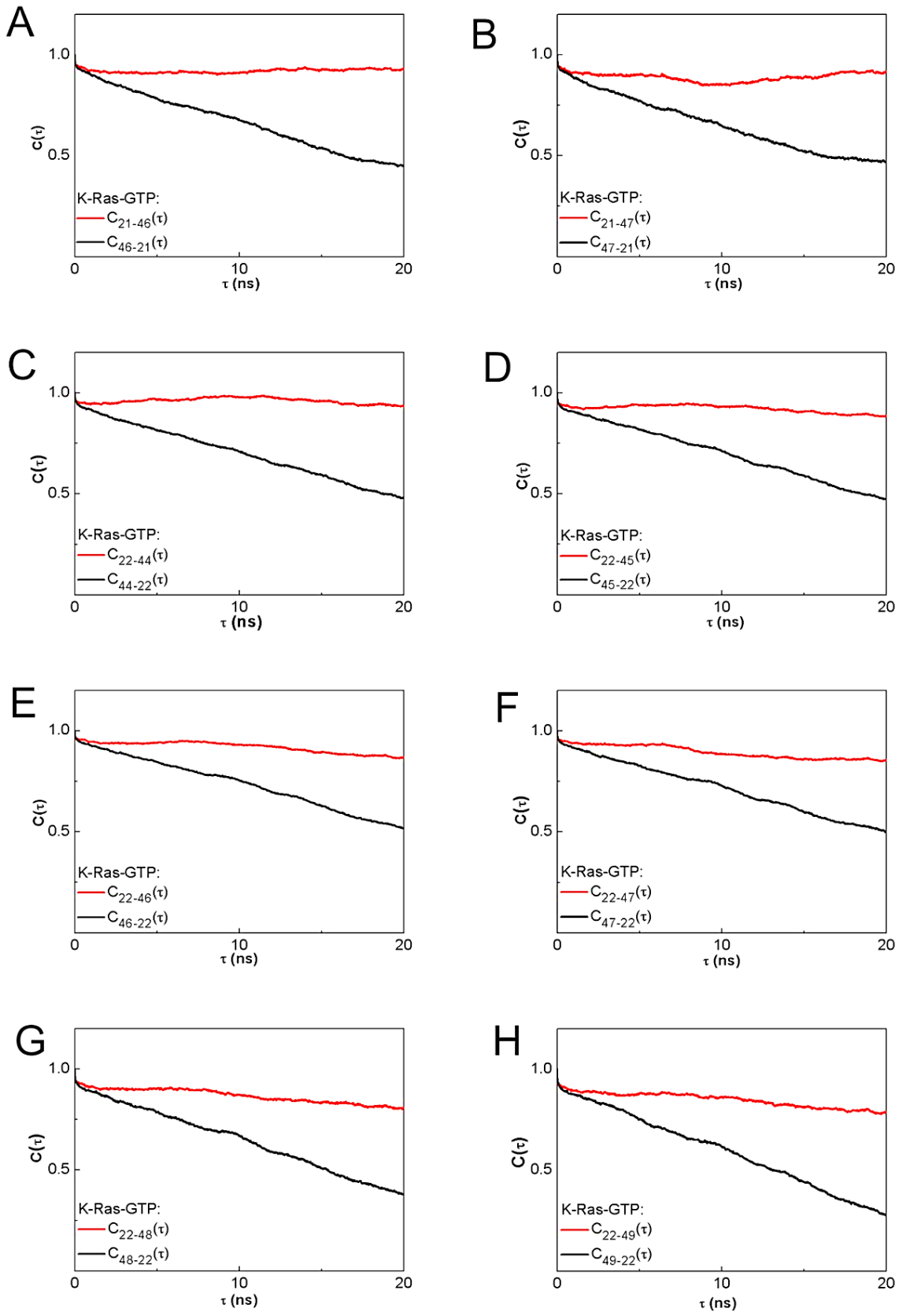

A
B
C
D
E
F
G
H
1.0
0.5
C(τ)
0
10
20
τ (ns)
K-Ras-GTP:
C21-46(τ)
C46-21(τ)
C21-47(τ)
C47-21(τ)
C22-44(τ)
C44-22(τ)
C22-45(τ)
C45-22(τ)
C22-46(τ)
C46-22(τ)
C22-47(τ)
C47-22(τ)
C22-48(τ)
C48-22(τ)
C22-49(τ)
C49-22(τ)

Figure 3.8 Fluctuations of ILE21-GLN22 (α1) drive the fluctuations of β2-β3. The red curves are for $\langle \Delta Ri(t) \Delta Rj(t+\tau) \rangle$ and show that the fluctuations of residue *i* at time *t* affect the fluctuations of residue *j* at a later time $t+\tau$. (A), (B) I21 drives I46 and D47, respectively. (C)-(H) Q22 drives V44, V45, I46, D47, G48 and E49, respectively.

4 CAUSALITY RELATIONSHIPS BETWEEN RESIDUE PAIRS OF MUTANT K-RAS PROTEINS

4.1 INTRODUCTION

In order to identify the effects of mutations to the protein dynamics, we analyzed mutant K-Ras dynamics in depth and compare the dynamics of WT and mutant proteins. For this purpose, we performed long time scale MD simulations of the most frequently observed mutant K-Ras proteins in cancer patients; G12V, G12D, G12C. We performed 900 ns MD simulations of the both active and inactive forms of each of these mutant K-Ras proteins. We analyzed mutant K-Ras dynamics in depth by applying our novel method CTC to the data from MD simulations of mutant K-Ras proteins and thereby identify driver-responder residue pairs in mutant K-Ras proteins' motions. Then we compared the results for different mutant K-Rases to reveal effects of the mutations onto causal relationships in K-Ras dynamics.

In the previous chapter, we applied CTC method to WT K-Ras data and obtained the results which are consistent with the experimental data. In this chapter, we developed a customized script for faster and more accurate calculation of the CTC in analysis of long time scale MD simulations. Our updated CTC method can quickly identify the causality in protein dynamics. We demonstrate the simplicity of computing CTC functions in studying protein dynamics by applying it to understand K-Ras motions. While correlations between the fluctuations of

residue pairs have already been shown in several Ras protein studies[203, 207, 211], despite extensive literature on K-Ras, there has been little attention on the role of causality (or directionality) in correlation dynamics of its residues.

We anticipate that our CTC method can quantify the changes due to the mutations or drug binding in protein dynamics. This computational tool will be easily applicable to the analysis of simulation data from different proteins to understand causality in their dynamics, which can then similarly be utilized in drug discovery. From this perspective, our method has the potential to set a novel paradigm for drug design by directing attention to changes in protein dynamics. The latter is in close relation to changes in protein function whose restoration to normal is the target of all drug design activities.

4.2 METHODS

4.2.1 MD simulations of mutant proteins

We have performed all-atom MD simulations of Mg^{+2}GTP-bound K-Ras and Mg^{+2}GDP-bound K-Ras for G12C, G12D and G12V mutant forms. We obtained the structures of K-Ras-GTP^{WT} and K-Ras-GDP^{WT} from Chapter 3, with the final K-Ras-GTP^{WT} structure in active (close) state and the final K-Ras-GDP^{WT} structure in inactive (open) state after 300ns-simulation. We constructed the mutant K-Ras^{G12C}, K-Ras^{G12D}, K-Ras^{G12V} structures based on K-Ras-GTP^{WT} and K-Ras-GDP^{WT} by respectively mutating glycine to cysteine, aspartate, at position 12 using DS 4.5 software. We optimized the protein complex with Clean Geometry tool of DS. We used NAMD 2.11 with AMBER ff99SB and GAFF[189]. We performed energy minimization of the initial model after we introduced mutation in K-Ras. Then, we run the MD simulations of each complex by following the same MD simulation protocol in the

previous study[212]. We applied minimization for 10,000 steps and equilibration for 500,000 steps. After equilibration, we performed 1 microsecond MD simulations, and saved atomic coordinates $\hat{R}$ of all atoms every 10 ps. We used the last 900 ns of the simulation trajectories in all computations. To eliminate all rotational and translational motions, we aligned the to the first frame using VMD software 1.9.2[93].

4.3 Identification of causality relationships between protein residue pairs

To predict causality between correlated residue pairs of a protein, we follow the six steps below for all pairs (*ij*) of its residues. (The algorithmic steps of our method are presented in Figure 4.1):

Step I: Correlation Calculation

We predict the correlation between two residues (C_{ij}) by calculating Pearson correlation coefficient (C_{ij} (ΔR_i, ΔR_j)) of residue *i* fluctuations (ΔR_i) with residue *j* fluctuations (ΔR_j). Pearson correlation coefficients lie in the range of $-1 \leq C_{ij} \leq 1$ and they are symmetric ($C_{ij} = C_{ji}$). If motions of two residues are independent, $C_{ij} = 0$. If the two residues always move in parallel in the same direction, then they are perfectly positively correlated, and $C_{ij} = 1$. If they always move in parallel in opposite directions, they are perfectly negatively correlated, and $C_{ij} = -1$.

Step II: Selection of Correlated Residue Pairs

In order to estimate whether there is a significant correlation within a residue pair, we aim to determine a cutoff value for C_{ij} between 0 and ±1. Since correlation is an effect size, we describe the significance of the correlations on system basis. First, we sort all C_{ij} values in the

system in a descending order and then separate negative and positive ones. Then, we draw log-log plots of C_{ij} values and find the convergence point for both negatives and positives. Finally, we define the C_{ij} value at the convergence point as "C_{ij} – cutoff". We select the residue pairs of which C_{ij} values are between C_{ij} – cutoff and ±1 and accept them as "correlated residue pairs" in the protein.

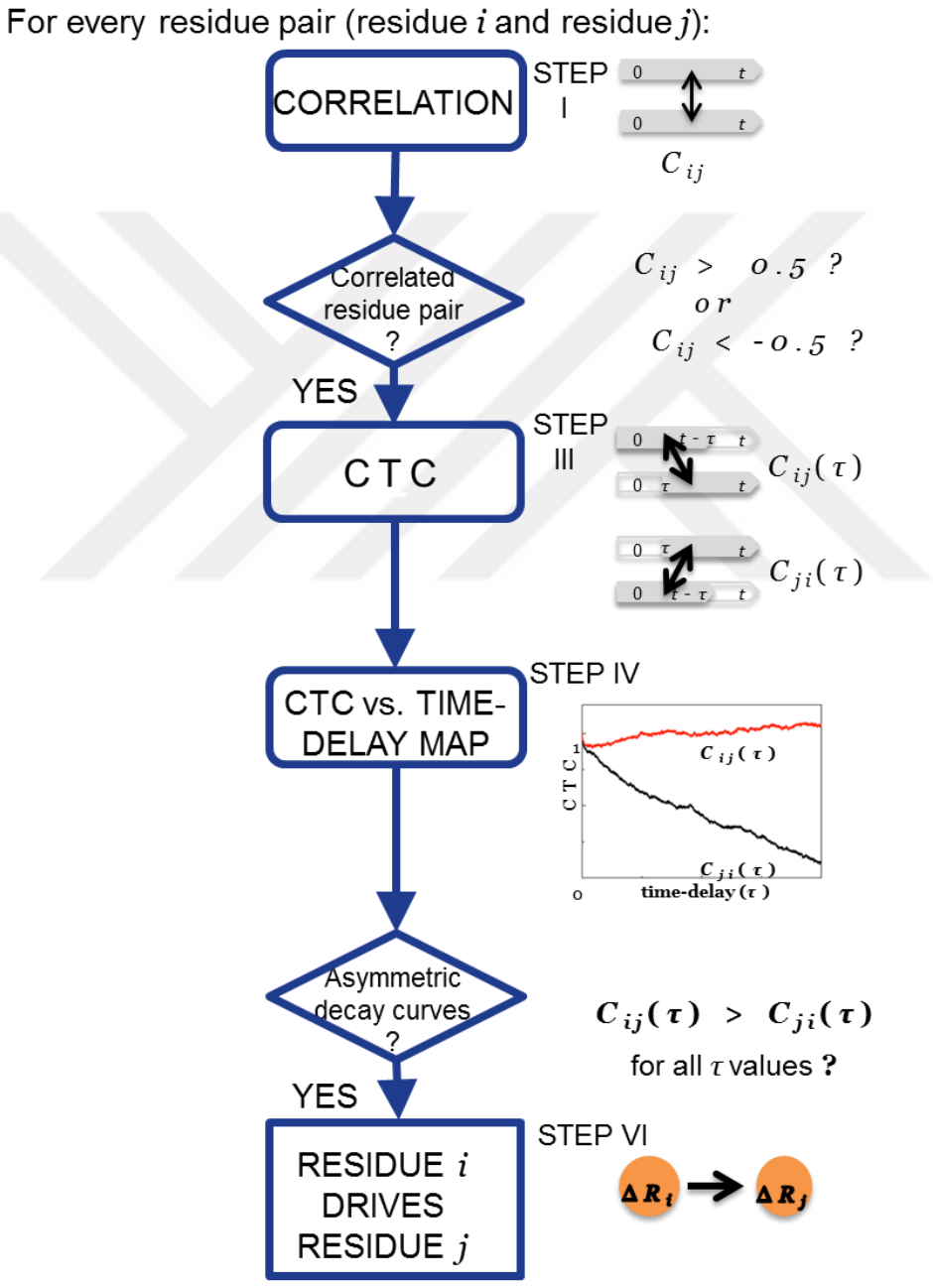


Figure 4.1 The flow diagram of CTC analysis.

Step III: CTC Calculation

We calculated CTC of fluctuations of the residues $C_{ij}(\tau)$ with the same formula in Chapter 4. $C_{ij}(\tau)$ represents the correlations of ΔR_j at time $t + \tau$ with earlier ΔR_i values at time t, where τ is the time-delay, while $C_{ji}(\tau)$ represents the reverse. If the fluctuations of residue i drive the fluctuations of residue j, then $C_{ij}(\tau) > C_{ji}(\tau)$. If $C_{ji}(\tau) > C_{ij}(\tau)$, residue j drives residue i because the fluctuation ΔR_j at time t is correlated with future fluctuations of ΔR_i. However, at τ =0, the equality $C_{ij}(0)=C_{ji}(0)$ holds.

Step IV: Time-Delay vs. CTC Map

We calculate CTC of a residue pair ($C_{ij}(\tau)$ and $C_{ji}(\tau)$) for different time-delay (τ) values and draw both $C_{ij}(\tau) - \tau$ and $C_{ji}(\tau) - \tau$ curves in the time-delay vs. CTC map to assess the decay of residue correlations by time-delay. In those maps, X-axis is time delay (τ) and Y axis is CTC of i and j for τ ($C_{ij}(\tau)$). All $C_{ij}(\tau)$ and $C_{ji}(\tau)$ values are normalized with respect their value at zero ($C_{ji}(0)$).

Step V: Selection of Residue Pairs which have Asymmetric Decay Curves in Time-Delay vs. CTC Map

We determine if decay curves of $C_{ij}(\tau) - \tau$ and $C_{ji}(\tau) - \tau$ are asymmetric and $C_{ij}(\tau)$ is greater than $C_{ji}(\tau)$ for all τ values. To quantify the difference between the two decay curves, first we calculate the area of the region bounded above by y = $C_{ij}(\tau)$ and bounded below by y= e^{-1}. We defined this area as "A($C_{ij}(\tau)$)". Similarly, we calculate "A($C_{ij}(\tau)$)", too. Then we calculate the absolute value of the difference between "A($C_{ij}(\tau)$)" and "A($C_{ji}(\tau)$)" and define it as "$\Delta C_{ij}(\tau)$" and use it in the next step.

Step VI: Identification of Driver-Responder Residue Pairs

We calculate all $\Delta C_{ij}(\tau)$ values for all *ij* pairs, we sorted them in a descending order and then drew log-log plots to find the convergence of $\Delta C_{ij}(\tau)$ values. The convergence point is the threshold. We then assumed that all *ij* pairs correspond to $\Delta C_{ij}(\tau)$ values over the threshold are driver-responder candidates.

4.4 RESULTS

4.4.1 Causality relationships in K-Ras^{G12C}-GTP

SII motions drive Loop 7 (L7) in active K-Ras^{G12C}. Our causality calculations show that L7 is driven by SII (Figure 4.9A-B). We present *CTC* plots of 62 (SII) with 104 (L7) (Figure 4.9A) and with 108 (L7) (Figure 4.9B) for active K-Ras. Red curve shows that the fluctuations of 62 at time *t* affect the fluctuations of 104 and 108 at time $t+\tau$.

SII motions drive SI in active K-Ras^{G12C}. Our causality calculations show that SI is driven by SII (Figure 4.9C). We present *CTC* plots of 65(SII) with 35(SI) (Figure 4.9C). Red curve shows that the fluctuations of 65 at time *t* affect the fluctuations of 35 at time $t+\tau$.

P loop motions drive SII. We present *CTC* plots of 17(P loop) with 65(SII) (Figure 4.9D). Red curve shows that the fluctuations of 17 at time *t* affect the fluctuations of 65 at time $t+\tau$.

β4 motions drive SII. We present *CTC* plots of 80(β4) with 65(SII) (Figure 4.9E). Red curve shows that the fluctuations of 80 at time *t* affect the fluctuations of 65 at time $t+\tau$.

Residue 116 motions drive SII. Fluctuations of 116 at time *t* affect the fluctuations of 65 at time $t+\tau$, as shown in Figure 4.9F.

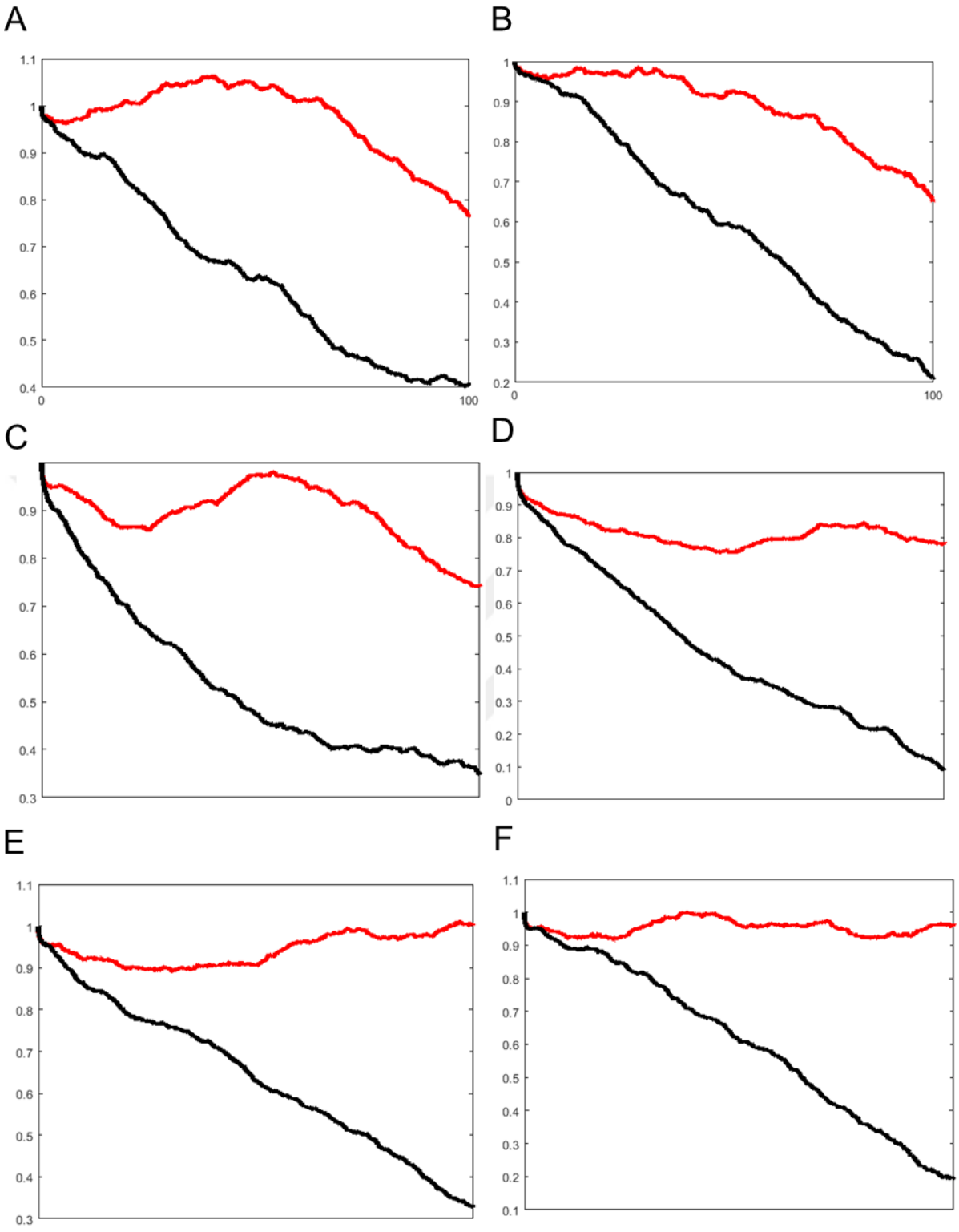


Figure 4.2 Causality relations in K-Ras^{G12C}-GTP. Red curves for $\langle \Delta \boldsymbol{R_i}(\boldsymbol{t}) \Delta \boldsymbol{R_j}(\boldsymbol{t}+\boldsymbol{\tau}) \rangle$ show that the fluctuations of residue *i* at time *t* affect the fluctuations of residue *j* at a later time *t*+τ. X axes is for *t* values from 1 ns to 100ns. All correlations ($\boldsymbol{C}(\boldsymbol{\tau})$) are normalized

with respect their value at zero ($\boldsymbol{C(0)}$) and shown in Y axes. (A) 62 (SII) drives 104(L7). (B) 62 drives 108(L7). (C) 65(SII) drives 35(SI). (D) 17 (P loop) drives 65. (E) 80 (β4) drives 65 (SII). (F) Residue 116 drives 65(SII).

4.4.2 Causality in K-Ras^{G12C}-GDP

SI and β4 motions drive SII motions in inactive K-Ras^{G12C} (Figure 4.10). Fluctuations of the SI residue E31 drive fluctuations of Q61 (SII), as shown in Figure 4.10A. Additionally, β4 residue F78 drives the motion of Q61 (SII) (Figure 4.10B).

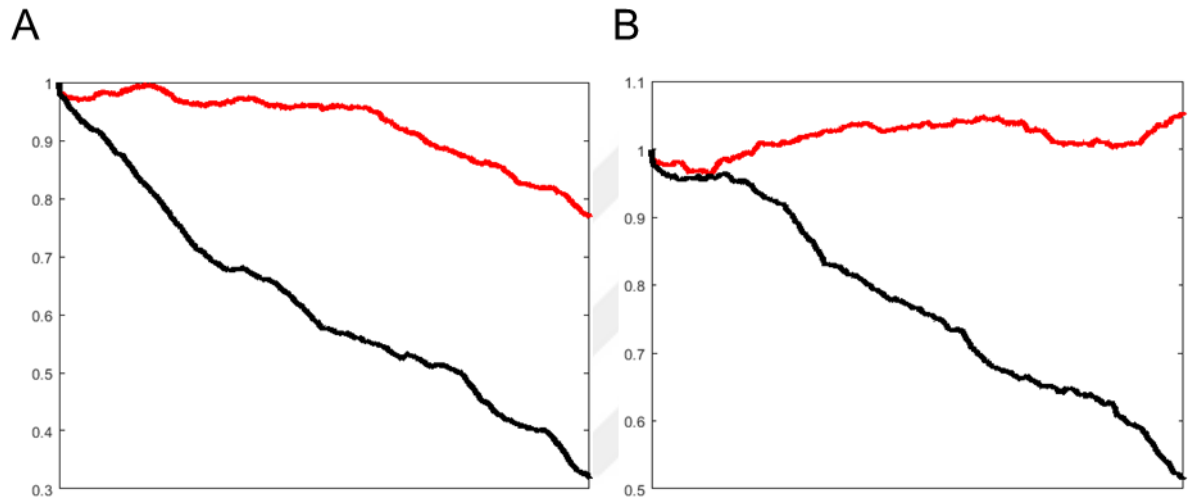


Figure 4.3 Causality relations in K-Ras^{G12C}-GDP. Red curves for $\langle \Delta \boldsymbol{R_i}(\boldsymbol{t}) \Delta \boldsymbol{R_j}(\boldsymbol{t}+\boldsymbol{\tau}) \rangle$ show that the fluctuations of residue *i* at time *t* affect the fluctuations of residue *j* at a later time *t*+*τ*. X axes is for *t* values from 1 ns to 100ns. All correlations ($\boldsymbol{C(\tau)}$) are normalized with respect their value at zero ($\boldsymbol{C(0)}$) and shown in Y axes. (A) 31 (SI) drives Q61 (SII). (B) F78 (β4) drives Q61 (SII).

4.4.3 Causality in K-Ras^{G12V}-GTP

P loop and β4 drive α3 motions in active K-Ras^{G12V}. Figures 4.11A-B show G12-G13 (P loop) and residues 83 and 88 of α3.

SII motions are driven by β3 and α4 motions. Figure 4.11C shows residue 52 drives Q61 and Figure 4.11D shows residue 135 drives 66.

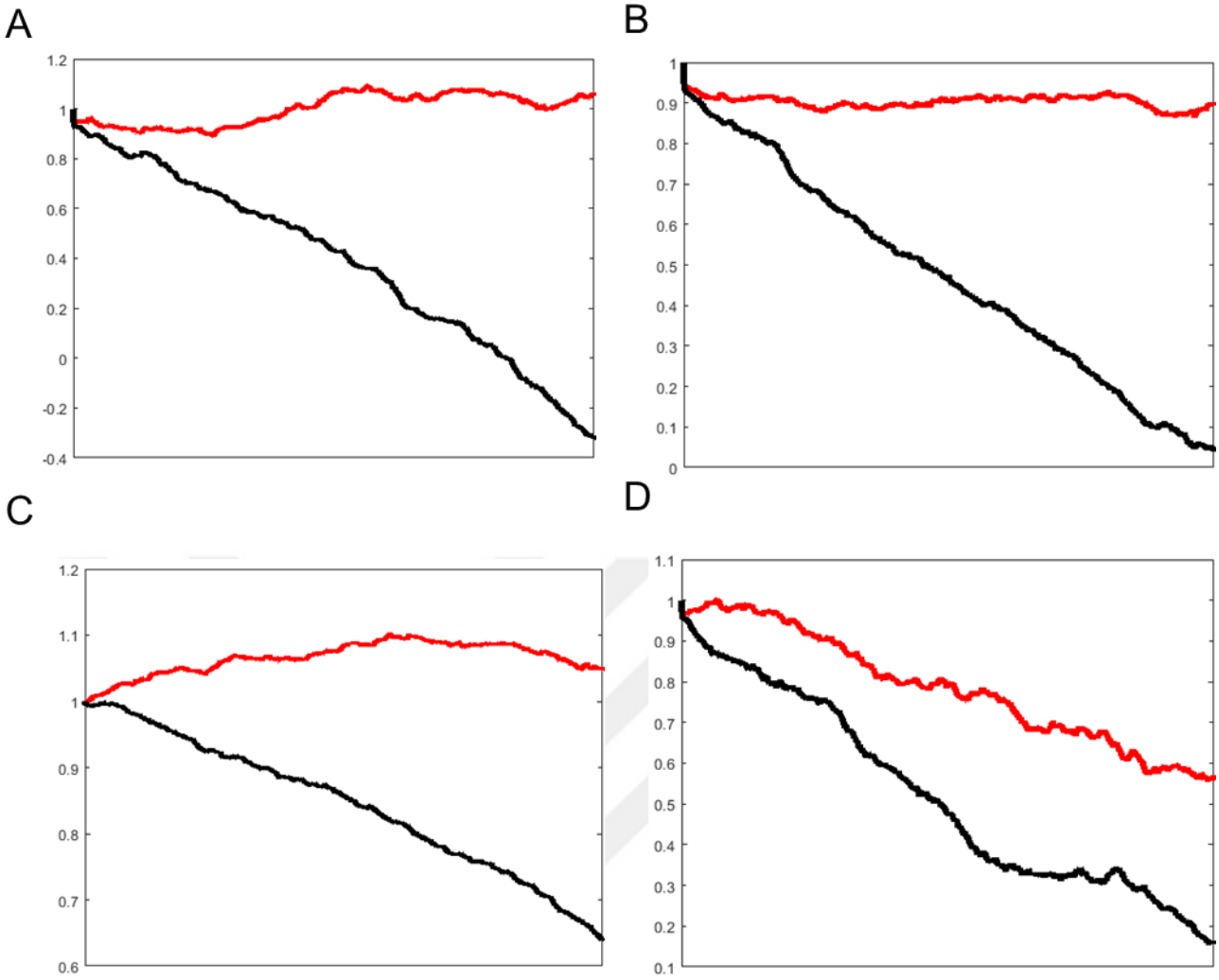


Figure 4.4 Causality relations in K-Ras^{G12V}-GTP. Red curves for $\langle \Delta \boldsymbol{R_i}(\boldsymbol{t}) \Delta \boldsymbol{R_j}(\boldsymbol{t}+\boldsymbol{\tau}) \rangle$ show that the fluctuations of residue *i* at time *t* affect the fluctuations of residue *j* at a later time *t*+τ. X axes is for *t* values from 1 ns to 100ns. All correlations ($\boldsymbol{C}(\boldsymbol{\tau})$) are normalized with respect their value at zero ($\boldsymbol{C}(\boldsymbol{0})$) and shown in Y axes. . (A) G12 (P loop) drives residue 83(α3). (B) G13 (P loop) drives residue 88(α3). (C) Residue 52(β3) drives Q61(SII). (D) Residue 135 (α4) drives 66.

4.4.4 Causality in K-Ras^{G12V}-GDP

We found that causality relations in active K-RAS^{G12V} motions disappear in the inactive form and new relations doesn't occur.

4.4.5 Causality in K-Ras^{G12D}-GTP

SII motions are main drivers in active K-Ras^{G12D}. Our causality calculations show that SII drives both the P-loop and β3 motions. We present *CTC* plots of 70(SII) with 9(P loop) (Figure 4.12A) and 69(SII) with 54(β4) (Figure 4.12B) for active K-Ras^{G12D}.

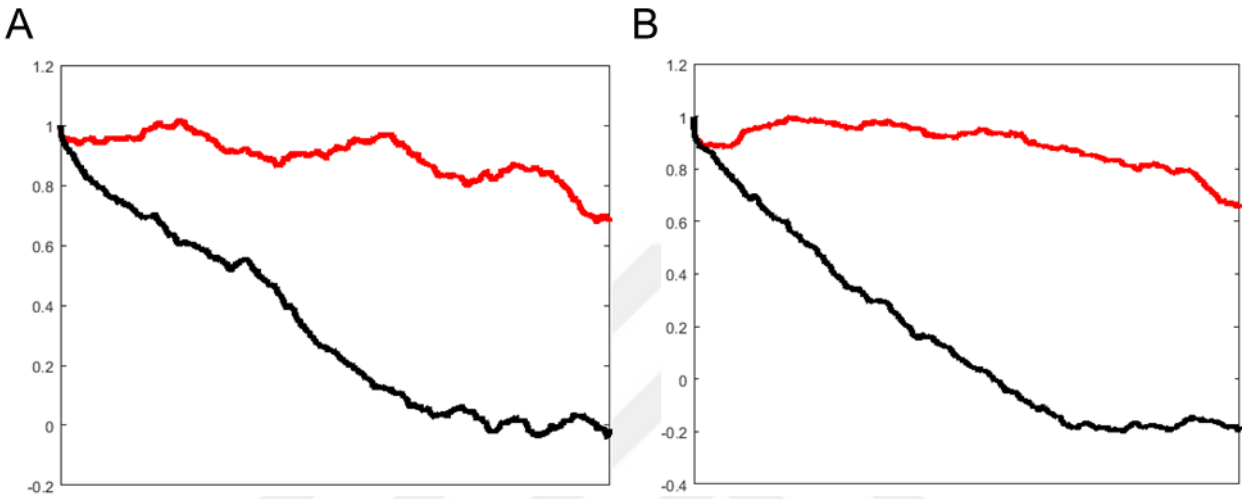


Figure 4.5 Causality relations in K-Ras^{G12V}-GTP. Red curves for $\langle \Delta \boldsymbol{R_i}(\boldsymbol{t}) \Delta \boldsymbol{R_j}(\boldsymbol{t}+\boldsymbol{\tau}) \rangle$ show that the fluctuations of residue *i* at time *t* affect the fluctuations of residue *j* at a later time *t*+*τ*. X axes is for *t* values from 1 ns to 100ns. All correlations ($\boldsymbol{C}(\boldsymbol{\tau})$) are normalized with respect their value at zero ($\boldsymbol{C}(\boldsymbol{0})$) and shown in Y axes. (A) Residue 70 (SII) drives residue 9(P loop). (B) Residue 69 (SII) drives residue 54(β4).

4.4.6 Causality in K-Ras^{G12D}-GDP

C-terminal motions are driven by switch region (SI & SII) motions in inactive K-Ras^{G12D}. Fluctuations of residue 162 (C-terminal) are driven by fluctuations of residues 36(SI) and 57, 76 (SII) as shown in Figures 4.13A, B, C, respectively.

SII motions are driven by β4 and α3 motions. Figure 4.13D shows residue 79(β4) drives 65(SII) and Figure 4.13E shows residue 95(α3) drives 69(SII).

SI motions drives β3. Fluctuations of residue 32 at time *t* affect the fluctuations of 41 at time $t+\tau$, as shown in Figure 4.13F.

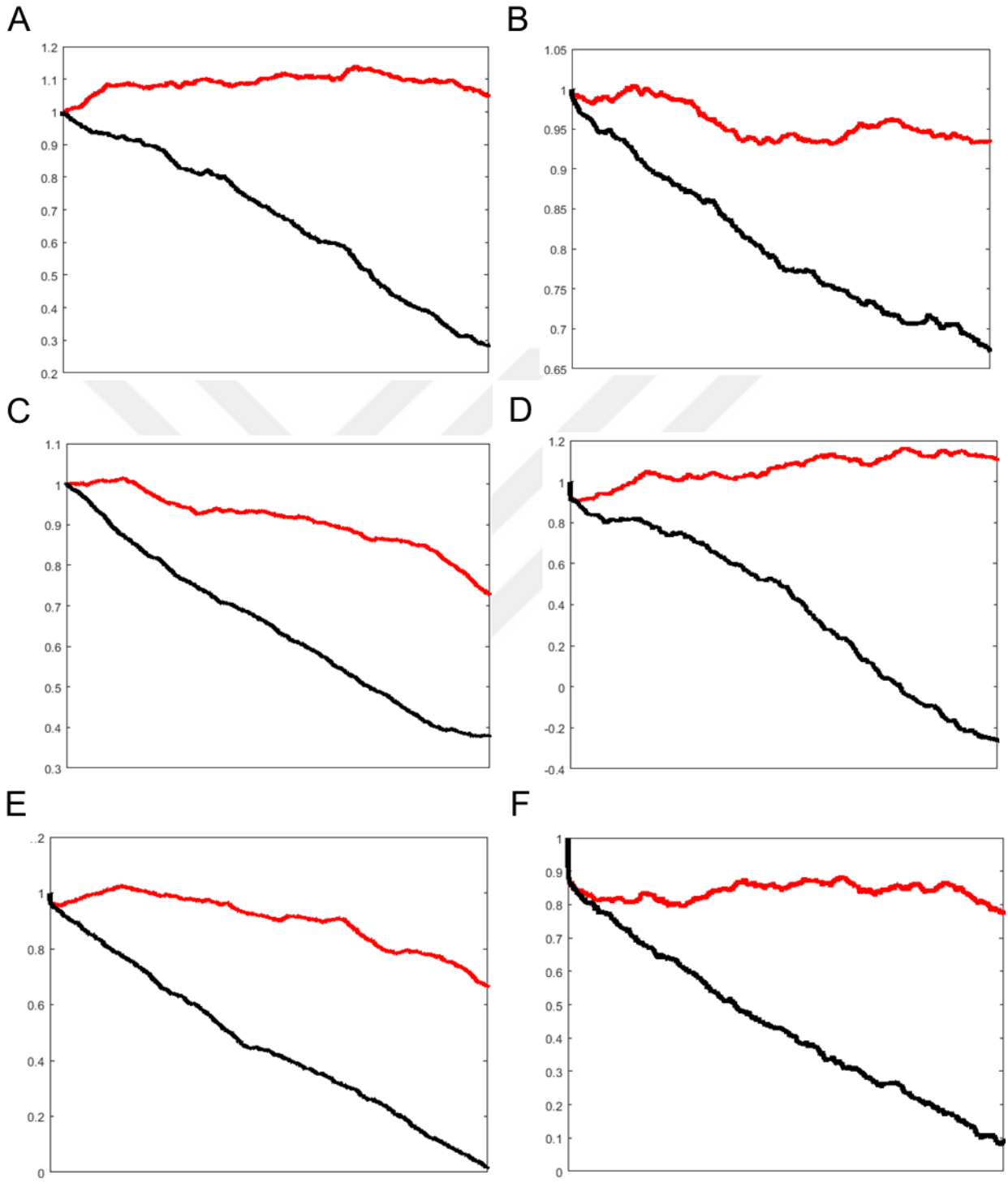


Figure 4.6 Causality relations in K-Ras^{G12D}-GDP. Red curves for $\langle \Delta \boldsymbol{R_i}(\boldsymbol{t}) \Delta \boldsymbol{R_j}(\boldsymbol{t}+\boldsymbol{\tau}) \rangle$ show that the fluctuations of residue *i* at time *t* affect the fluctuations of residue *j* at a later

time $t+\tau$. X axes is for t values from 1 ns to 100ns. All correlations ($\boldsymbol{C}(\boldsymbol{\tau})$) are normalized with respect their value at zero ($\boldsymbol{C}(\mathbf{0})$) and shown in Y axes. (A) 36 (SI) drives 162(C-terminal). (B) 57(SII) drives 162. (C) 76(SII) drives 162. (D) 17 (P loop) drives 65. (E) 79 (β4) drives 65 (SII). (F) Residue 95 drives 69(SII).

5 THE DYNAMIC BASIS OF ONCOGENIC K-RAS MUTATION: COMPARISION OF WILD TYPE AND MUTANT K-RAS

5.1 INTRODUCTION

In order to obtain more detailed information on mutant K-Ras dynamics, we focus on the most frequently observed K-Ras mutation, G12D. Here, we present how the most recurrent K-Ras oncogenic mutation, G12D, leads to structural, conformational and dynamical changes that impair K-Ras function using a new integrated MD simulation data analysis approach we have developed to quantify the changes in K-Ras in response to G12D mutation. We analyzed the MD-simulation data of K-Ras^{WT} and K-Ras^{G12D}. We first studied the structural changes in K-Ras upon G12D mutation, and discovered salt bridges that are either formed or destroyed upon mutation. Second, we evaluated the changes in the pair-wise distances between residues and quantified the local volume changes to identify changes in protein conformation. Third, we identified changes in protein dynamics though a multi-step process where we (i) quantified the residue fluctuations; (ii) evaluated correlation of residue fluctuations and identified lost or newly formed correlations upon mutation; (iii) calculated the characteristic decay times of residue fluctuations; and (iv) identified residue pairs that are causally related (i.e., residue pairs that show driver-responder behavior). Finally, we related the observed structural

changes to conformational and dynamical ones, which enabled us to identify the important changes that affect protein function. Overall, our study identifies regulatory sites on K-Ras^{G12D}, which can enhance our understanding of its dynamics and assist in the development of direct inhibitors.

5.2 METHODS

5.2.1 MD simulations

We performed all-atom MD simulations for Mg^{+2}GTP-bound K-Ras^{WT}. We obtained the K-Ras-GTP^{WT} structure from the final frame of the 300 ns simulation of active (close) state protein described in Chapter 3. We we performed 1 microsecond MD simulations following the protocols from Chapter 3. We saved atomic coordinates $\hat{R}$ of all atoms every 10 ps and used the last 900 ns of the simulation trajectories in all computations. To eliminate all rotational and translational motions, we aligned the trajectories to the first frame using VMD software 1.9.2[93]. We visualized the trajectories with VMD. To identify salt bridges formed in the protein during the MD simulations, we used Salt Bridges Plugin, Version 1.1, of VMD.

5.2.2 Pairwise distance calculations

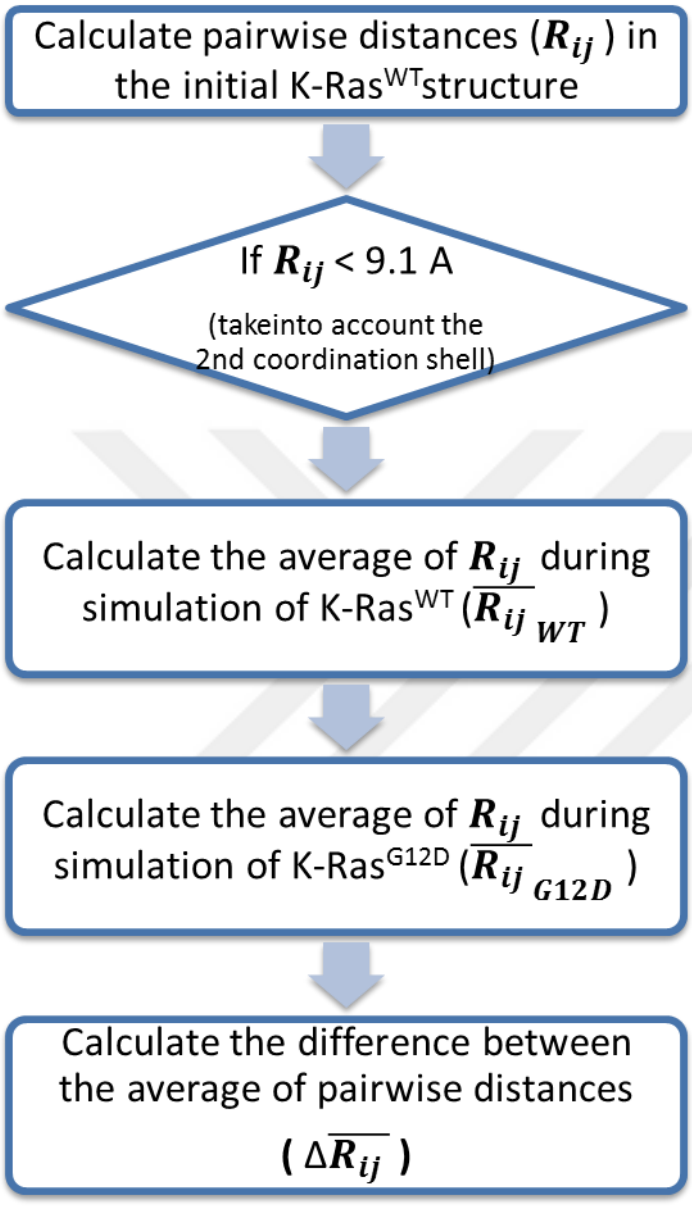
To quantify the effect of the G12D mutation on the distances between residue pairs, we developed a new computational algorithm. We detail our algorithm in Figure 5.1. Briefly; we first assumed K-Ras^{WT} as the initial state and K-Ras^{G12D} as the final state. Then, we calculated the distances between Cα atoms of two residues (*i, j)* as we described in Chapter 3. GNM studies in the literature typically assume the maximum Cα-Cα distance for the separation between two contacting residues ~7.2Å[195, 213], and label it the 'first coordination shell'[214, 215]. We followed this same protocol and determined the first coordination shell

around a selected residue by choosing its Cα as the center of a volume V with a radius of r1 ~7.2 Å[195]. However, because the contribution of non-bonded pairs to higher order coordination shells may also be significant[215, 216], we also studied residue pairs that are within their 'second coordination shell' in K-Ras^{WT} structure. We defined this second coordination shell at twice the volume of the first, with a radius of ~9.1 Å[215].

For every residue pair (*i*,j) where *j* is in the second coordination shell of *i*, we first calculated their time-averaged distances in K-Ras^{WT} ($\bar{R}_{ij\ WT}$) and in K-Ras^{G12D} ($\bar{R}_{ij\ G12D}$). We then calculated the difference ($\Delta\bar{R}ij$) between $\bar{R}_{ij\ WT}$ and $\bar{R}_{ij\ G12D}$, where $\Delta\bar{R}ij = \bar{R}_{ij\ G12D} - \bar{R}_{ij\ WT}$. The magnitude of the difference is the degree of distortion resulting from the G12D mutation. We present $\Delta\bar{R}ij$ values in the pairwise distances map (Figure 5.2) where a positive value indicates that a residue pair moves apart upon G12D mutation, while a negative value indicates that the pair gets closer. Then, we identified the residue pairs (*ij*) which are distorted by G12D mutation significantly. For this purpose, we selected the residue pairs that have the greatest (positive and negative) $\Delta\bar{R}ij$ values. We assumed that the residue pairs whose $\Delta\bar{R}ij$ values are greater than 2.75 or smaller than -1.35 showed the most significant distance changes. For those identified residue pairs, we also drew the distribution graphs $W(R_{ij})$ of their distances (R_{ij}) during the simulations of K-Ras^{WT} and K-Ras^{G12D}.

Additionally, to quantify the changes in local volumes upon G12D mutation, for each residue *i* we calculated the average of all $\Delta\bar{R}_{ij}$ values according to formula of $\langle\Delta\bar{R}i\rangle = \sum_j \Delta\bar{R}ij/N_n$, where N_n is the number of residues *j* which are in the second coordination shell of residue *i*. In detail, for a residue *i*, we chose it as the center of a volume V with a radius of 9.1 Å (the second coordination shell) and defined the residues *j* within this volume V as the *neighbors* of residue *i*. Then, we calculated the total change in the distance between residue *i* and its

neighbors, $\sum_j \Delta\bar{R}ij$, and divided it by number of neighbors $\sum_j \Delta\bar{R}ij/N_n$. The resulting $\langle\Delta\bar{R}i\rangle$ value is a measure of the change in volume around residue *i* due to G12D mutation.



.

Figure 5.1 The algorithm for calculating the change of distances between residue pairs ($\Delta\bar{R}ij$) upon G12D mutation.

5.2.3 Pairwise correlation calculations

We calculated the correlations between the fluctuations of residue pairs (C_{ij}) as described in Chapter3. For any residue pair, this value ranges from -1 to 1, where C*ij*= 0 if the residue

fluctuations are independent; Cij=1 for perfectly positively correlated residue pairs, and Cij=-1 for perfectly negatively correlated residue pairs.

5.2.4 Time delayed correlations and characteristic decay time calculations

We calculated time-delayed correlation of two fluctuations $C_{ij}(\tau)$ and autocorrelations $C_{ii}(\tau)$ of each protein residue as described in Chapter 3. For each residue i we calculated the time-delay value (τ) where the autocorrelation decay curve of the residue i reaches to 0.5 and accepted this time-delay value as the characteristic decay time of the residue i. We adopted the 0.5 criterion instead of the commonly used 1/e decay criterion because the criterion we adopted was a better indicator of short time relaxation differences whereas the latter led to values too close to each other.

5.3 RESULTS

5.3.1 Structural Changes

Close-Range Electrostatic Interactions in K-Ras^{WT} and K-Ras^{G12D}

Substitution of a non-polar, non-charged amino acid –glycine- with a negatively-charged amino acid –aspartate- triggers several conformational and dynamical changes in K-Ras^{G12D}. With the plausible assumption that the negatively-charged residue D12 may cause new electrostatic interactions within the protein and those interactions can be the sources of conformational changes upon G12D mutation, we compared the close-range electrostatic interactions (i.e. salt bridges) in K-RAS^{WT} and K-RAS^{G12D}. In the mutated structure, D12 forms salt bridges with K16 (P-loop) and K88. Additionally, K16 forms a salt bridge with D57 in K-RAS^{G12D}. However, those interactions are not present in K-RAS^{WT}. To identify the

effects of the new electrostatic interactions caused by G12D mutation, we then investigated the conformational changes in K-Ras.

5.3.2 Conformational Changes

Pairwise Distance Calculations

We compared the distances between residue pairs within their second coordination shell in K-Ras^{WT} and K-Ras^{G12D}. Figure 5.2 shows $\Delta\bar{R}_{ij}$ values for all residue pairs where K-Ras^{WT} is the reference and K-Ras^{G12D} is the final structure. As seen from the abundance of positive $\Delta\bar{R}_{ij}$ values in Figure 5.2, the dominant distortion of the protein upon mutation is expansion. The SII region (T58-T74) moves away from the phosphate binding loop (P-loop, G10-S17), SI (Q25-Y40), α3 (T87-K104) in K-Ras^{G12D}. The SI region also moves away from the P-loop. In detail, all the P-loop residues move away from T58-R68 (SII) residues as a block. Secondly, the residues in the segment P34-Y40 of SI move away from N-terminal residues of SII (D57-Q61). Thirdly, the α2 helix of SII moves away from α3, D92-R102. Fourth, SI residues move away from the P-loop residues. On the other hand, SI residues (F28-Y32, Y32-I36) obtain a closer conformation and an H-bond between D33 and I36 is established. This causes the T35-E37 part of SI to move away from the P-loop residues, G13-K16.

The most significant changes in distance were between residue pairs Q61 (SII)-A11, G12D, G13 (P-loop), E37(SI)-D57 (SII) and S65 (SII)-H95(α3). Residues R68, M72 and Y96 were also more distant in K-Ras^{G12D}. On the other hand, some residue pairs significantly approached in K-Ras^{G12D}, including E63-R68, M72-G75 and R73-K104, demonstrated by their negative $\Delta\bar{R}_{ij}$ values in Figure 5.2.

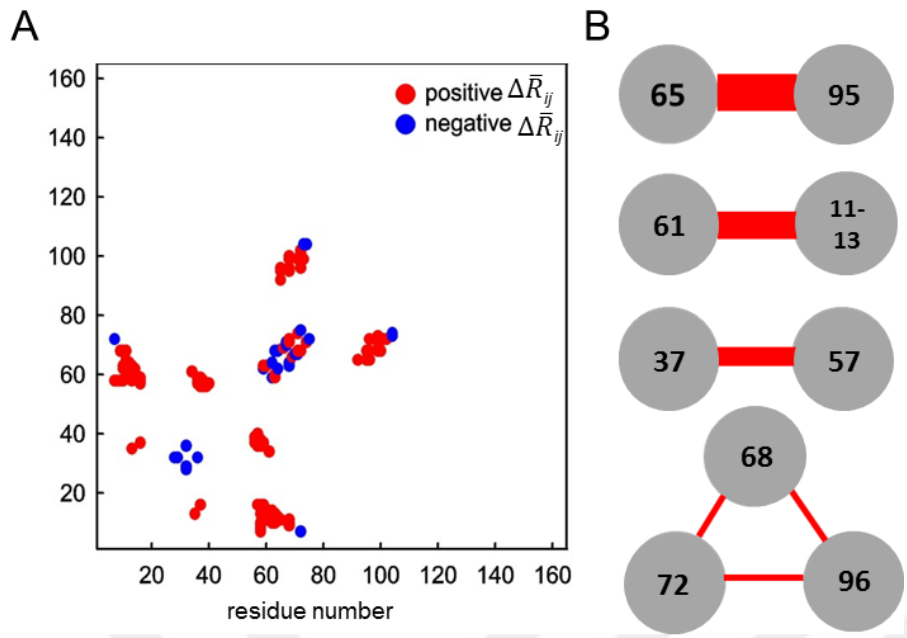


Figure 5.2 Difference of the averaged pairwise distances ($\Delta\bar{R}$ij) where K-Ras^{WT} is the initial and K-Ras^{G12D} is the final state. A) Red dots (positive $\Delta\bar{R}$ij values) show that pairs become distant and blue dots (negative $\Delta\bar{R}ij$ values) show that pairs become closer in K-Ras^{G12D}. B) Sample scheme for the most significant results. The width of bars shows the magnitude of the $\Delta\bar{R}ij$ values.

5.3.3 Distance distributions

To uncover the conformational differences between wild type and mutant K-Ras and to thereby better understand the effects of the G12D mutation, we calculated the probability distributions of the distances between pairs of residues that exhibited the largest changes in our distance calculations. We observed that K-Ras^{WT} is characterized by Gaussian-shaped distribution curves with stable distance values. We use the term 'stable' in the sense that the distribution is sharply peaked with small dispersion around the peak. However, mutant K-Ras shows significant deviations from the Gaussian, except for the residue pair E37 (SI)-D57 (SII).

Distance distributions of the residue pairs which underwent the largest change due to G12D mutation are shown in Figure 5.3. These are the distances between the alpha carbons of

residue pairs Q61-A11, Q61-G12D and Q61-G13, respectively, as shown in Fig. 5.3A-C. Their distribution patterns show marked differences between WT and mutant K-Ras. In WT K-Ras, they exhibit narrow distance distribution with smaller peak values compared to the mutant form. Residues A11, G12 and G13 reside at the P-loop. The P-loop has an omega shape and it forms a turn at the C-terminal neighbouhood of G12. As shown in Figure 5.3A-B, A11 and G12 are located within H-bond distance from Q61, while G13 is more distant from Q61 in WT. Furthermore, we observed that G12 (side chain H atom) forms an H-bond with G60 (backbone O atom), the neighbor of G60, in WT. However, when G12 is mutated to D12, this H-bond disappears because of the bulkier side chain of D12. In the absence of this H-bond, D12 and its neighbors move away from Q61. The broadened distance distribution values of the mutated system are indicators of lost hydrogen bonds.

We observed an H-bond between D38 (SI) and D57 (SII) in K-Ras^{WT} simulations. However, this H-bond dissappers in K-Ras^{G12D} simulations that results in moving away of D57 from D38 and more remarkably E37. The absence of this H-bond in K-Ras^{G12D} is indicated in Figure 5.3D where the peak distance value is greater in the K-Ras^{G12D}.

Among all pairs, the S65-H95 pair undergoes the largest conformational change upon mutation. S65 –part of the α2- moves away from H95 –part of the α3- in K-Ras^{G12D} (Figure 5.3E). K-Ras simulations show that this significant change results from the brakeage of the salt bridge between α2 and α3 helices. These two helices interact through the salt bridge between R68 and D92 in K-Ras^{WT}. However, G12D mutation breaks the R68-D92 salt bridge and causes the α2 and α3 helices to be distant from each other. The distant conformation of α2 and α3 in K-Ras^{G12D} can be implied by Figure 5.3E: The distance distribution curve of S65 (α2)-H95 (α3) has a larger distance peak value.Distances between the residues R68, M72 and Y96 are also more stable in K-Ras^{WT} as shown in Figure 5.3F-H.

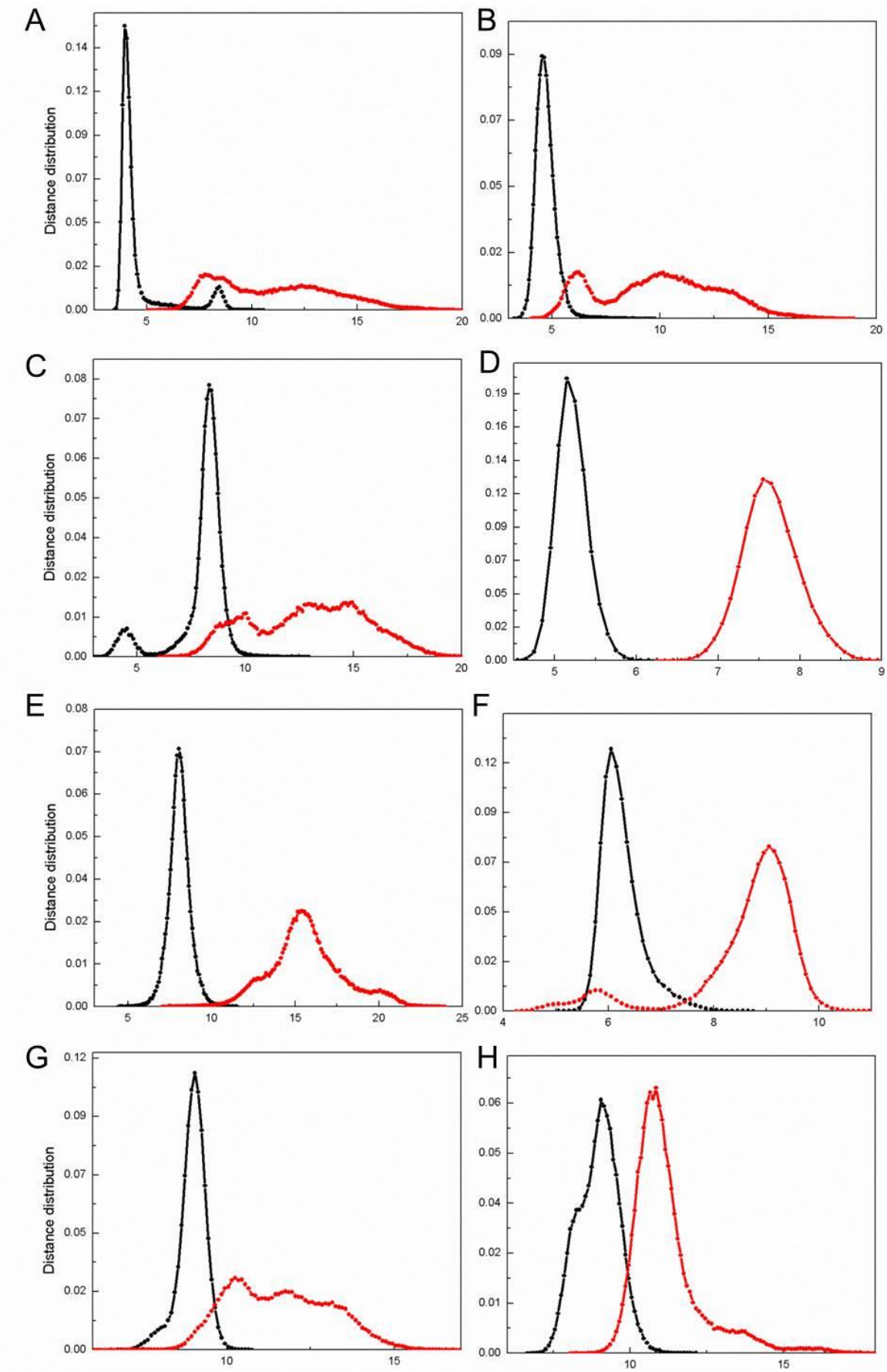

A
B
C
D
E
F
G
H
Distance distribution

Figure 5.3 Distribution of distances between residue pairs in K-Ras^{WT} (black) and K-Ras^{G12D} (red). (A) Distance distribution of residue pair Q61-A11 (B) Distance distribution of residue pair Q61G-12D (C) Distance distribution of residue pair Q61-G13 (D) Distance distribution of residue pair E37-D57. (E) Distance distribution of residue pair S65-H95. (F) Distance distribution of residue pair R68-M72. (G) Distance distribution of residue pair M72-Y96. (H) Distance distribution of residue pair R68-Y96.

Figure 5.4 displays the distance distributions of the residue pairs which significantly approached in K-Ras^{G12D}. We observed that in K-Ras^{G12D}, E63-R68 pair was in a closer conformation during the simulations (Figure 5.4A). This conformation may result from the fact that residues adjacent to E63, Y64 and S65, which make H-bond with R68 and D69, respectively. Those H-bonds are established between the backbone H atoms of Y64 & S65 and the backbone O atoms of R68 & D69 only K-Ras^{G12D}. Additionally, during the K-Ras^{G12D} simulations R73-K104 pair also has a closer conformation which may result from the H-bonds between R73-D105 and G75-K104 (Figure 5.4B).

On the other hand, Figure 5.4C shows that the residue pair of M72-G75 switches between two conformations in K-Ras^{WT}, one is closer conformation around 5.5 Å and the other one is distant conformation around 8.5 Å. However, their distance distribution is single peaked at 5.5 Å in K-Ras^{G12D} that is similar to the closer conformation in K-Ras^{WT}. We observed that M72 forms H-bond with G75 only in K-Ras^{G12D} simulations. This H-bond between M72 and G75 may cause the single peaked distance distribution of them.

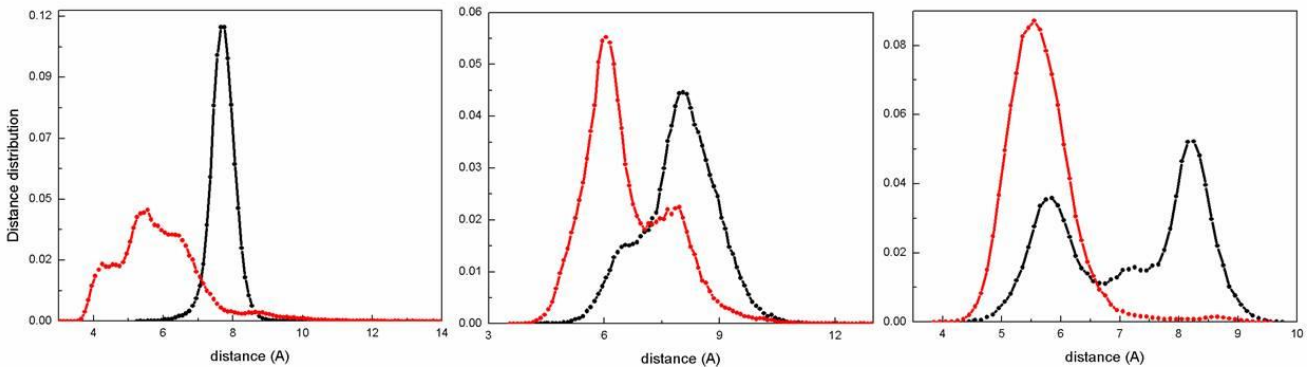


Figure 5.4 Distance distributions of residue pairs in K-Ras^{WT} (black) and K-Ras^{G12D} (red). (A) Distance distribution of residue pair E63-R68. (B) Distance distribution of residue pair R73-K104. (C) Distance distribution of residue pair M72-G75.

Changes in Local Volume upon Mutation

Fluctuations of a residue take place within the first or second coordination shell around it. In this volume, there are several other residues, which are either near neighbors along the chain or neighbors of spatially distant residues. Residues will have different number of neighbors as has been shown by GNM[195]. A residue with a smaller number of neighbors will show larger fluctuations than another reside with a larger number of neighbors. Therefore, the neighborhood of a given residue has significant effect on its fluctuations, and as we will show here there is also a significant effect on the dynamics. In order to understand which parts of K-Ras move away from their neighbors and which parts approach upon mutation, we calculated the average of $\Delta\bar{R}_{ij}$ values for each residue, $\langle\Delta\bar{R}i\rangle$. Figure 5.5 shows that most of the protein parts, especially the P-loop and SII, move away from their neigbours after G12D mutation.

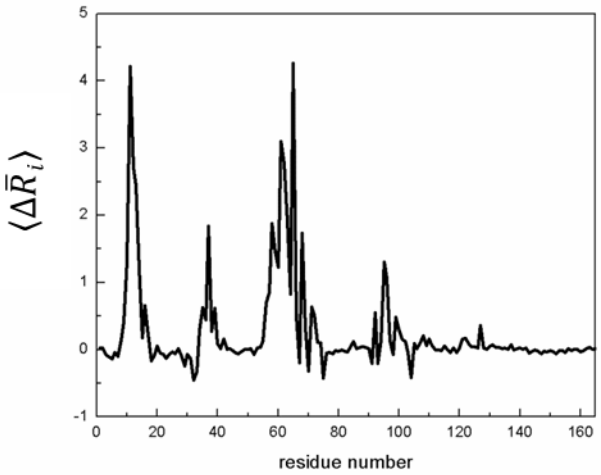


Figure 5.5 $\langle\Delta\bar{R}i\rangle$ values. K-RasWT is the initial and K-RasG12D is the final state. Positive values indicate the residues which move away from their neighbors and negative values indicates the residues which approach to their neighbors.

We identified the residue pairs which underwent the largest distance change due to G12D mutation above. Furthermore, we aimed to show how much each residue in the identified pairs deviates from its neighbours. For this purpose, we compared the individual $\sum_j \Delta \bar{R} ij$ values of the residues in the identified pairs (Table 5.1). We found that residues which become distant in K-Ras^{G12D} also move away from their neighbors. However, the residues become closer in K-Ras^{G12D} obtain different conformations relative to their neighbors.

Table 5.1 Sum of $\Delta \bar{R}$ij values in K-Ras

Residue Pair	**Conformation in K-Ras^{G12D}**	**Residue Number**	**Sum of $\Delta \overline{R_{ij}}$ (Å)**
11-61	distant	61	24.72
		11	33.70
12-61	distant	12	24.47
13-61	distant	13	24.35
37-57	distant	37	11.02
		57	9.45
65-95	distant	65	25.59
		95	10.42

68-72-96	distant	68	15.57
		72	6.23
		96	9.58
63-68	closer	63	17.23
73-104	closer	73	1.03
		104	- 2.50
72-75	closer	72	6.23
		75	-4.19

5.3.4 Dynamic changes

Residue Fluctuations

To predict the effects of the G12D mutation on flexibility of K-Ras, we calculated the RMSF, a measure of the average atomic fluctuations, value of each residue in K-Ras^{WT} and K-Ras^{G12D}. Figure 5.6 shows that the fluctuations of central residues of SII are increased in K-Ras^{G12D}.

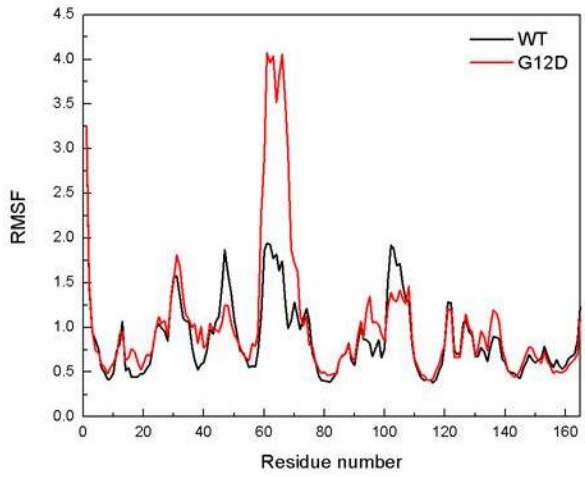


Figure 5.6 RMSF values of K-Ras^{WT} (black) and K-Ras^{G12D} (red) residues.

Pairwise Correlation Calculations of Residue Fluctuations

Regulation of protein dynamics is strictly coordinated by the correlations of residue fluctuations. Figure 5.7 presents pairwise correlations of residue fluctuations (C_{ij}), where we show the results for K-Ras^{WT} and K-Ras^{G12D} in the left and right panels, respectively.

Our results showed that K-Ras^{G12D} is marked by an increase in correlated motions of residues in comparison with K-Ras^{WT}. Specifically, β3-SII residues become negatively correlated with the residues of the P-loop, SI, and β4-α3.

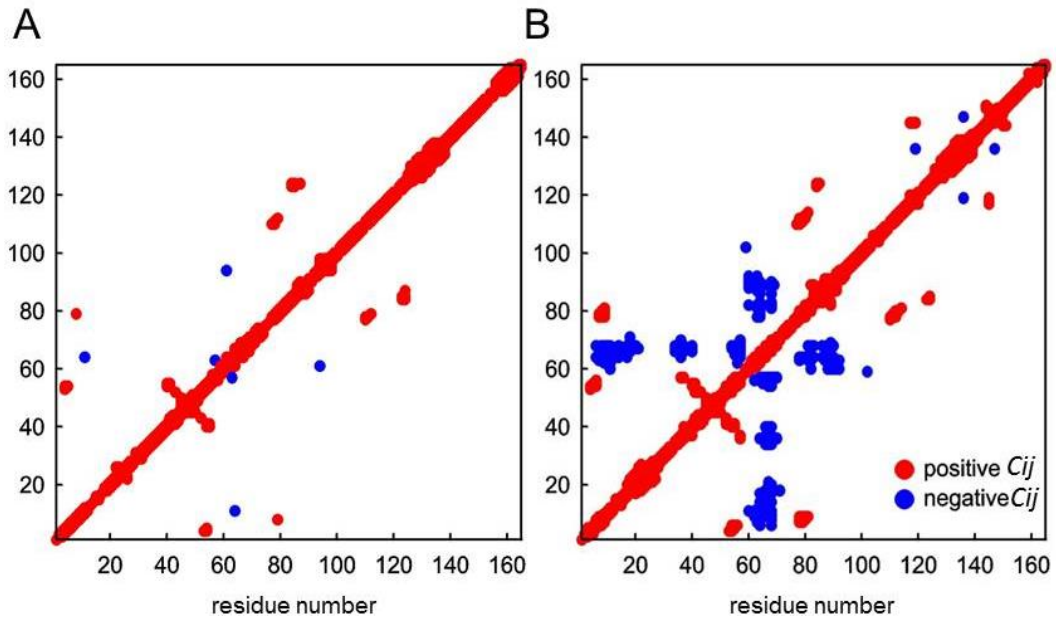


Figure 5.7 Correlation coefficient maps of K-Ras^{WT} and K-Ras^{G12D}. Red dots show positive correlations and blue dots show negative correlations. Residue indices 1-165 refer to K-Ras.

Characteristic Decay Times of Residue Fluctuations

Fluctuations of protein residues have characteristic decay times Figure 5.8 shows the characteristic decay times of residues in the P-loop, SI and SII are longer in K-Ras^{G12D}. Moreover, SII residues show the longest correlation decay times within the K-Ras^{G12D} residues.

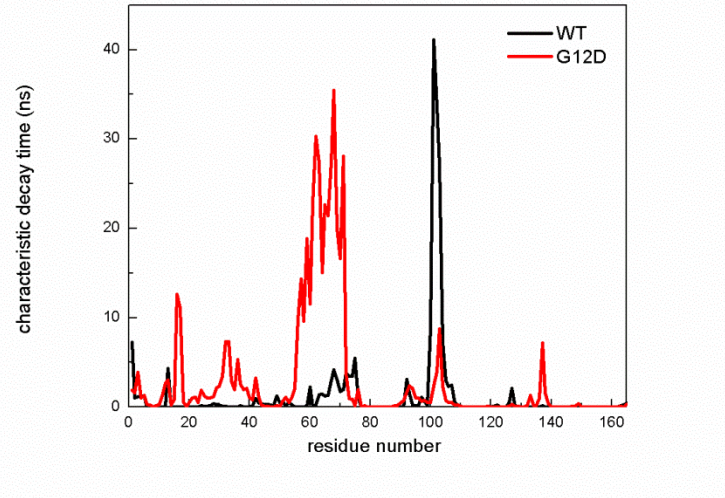


Figure 5.8 Characteristic decay times of residue fluctuations in K-Ras^{WT} (black) and K-Ras^{G12D} (red).

Causality between the Residue Fluctuations in K-Ras^{G12D}

We analyzed mutant K-Ras dynamics in depth by applying time delayed correlation analysis to MD simulation data of K-Ras^{G12D} and identified driver-responder residue pairs in K-Ras^{G12D} motions in Chapter 4. Our causality calculations show that SII drives both the P-loop and β3 motions.

5.3.5 Relations between the changes in K-Ras structure-conformation-dynamics

Relation between Structural and Conformational Changes in K-Ras^{G12D}

Formation of new salt bridges results in distortion of backbone Cα atoms in K-Ras^{G12D}. Salt bridges are formed between the sidechain atoms of the residues D12 (P-loop)-K16 (P-loop) and K16-D57 (SII) upon G12D mutation. We also discovered that SII region becomes distant from the P-loop and that they both move away from their neighbors in K-Ras^{G12D}. MD simulations of K-Ras^{G12D} show that sidechains of K16 -D57 pair approach each other and form salt bridges, while their backbone Cα atoms move away from each other as shown in Figure 5.9. Therefore, we may conclude that the new salt bridges formed between the P-loop and SII residues cause more distant conformation of these two regions and moving away of both of them from their neighbors. As a consequence of these structural and conformational changes in SII region, this region also becomes distant from SI and α3 regions.

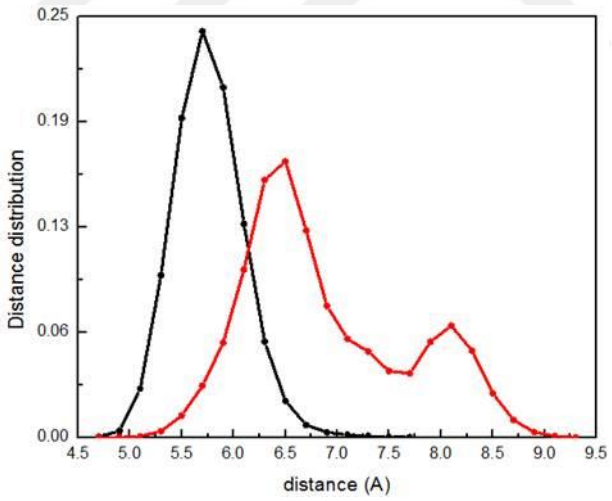


Figure 5.9 Distance distribution of residue pair K16 (Cα)-D57 (Cα) in K-Ras^{WT} (black) and K-Ras^{G12D} (red).

Relation between Structural and Dynamical Changes in K-Ras^{G12D}

Our K-Ras^{G12D} simulations showed the formation of salt bridge between of D12 (P-loop) and K16 (P-loop). After observed this salt bridge which occurs after mutation, we aimed to determine if the approaching of D12 and K16 via salt bridge causes changes in distant regions of K-Ras^{G12D}. For this purpose, we first described a connectivity vector, $\Delta R_{12\rightarrow16}$, between D12 and K16 following the same definition in Chapter 3. We then described connectivity vectors between the correlated residue pairs in K-Ras^{G12D}. Finally, we calculated the correlations of $\Delta R_{12\rightarrow16}$ with the other connectivity vectors. We identified that $\Delta R_{12\rightarrow16}$ is significantly correlated with $\Delta R_{60\rightarrow70}$, $\Delta R_{61\rightarrow75}$ and $\Delta R_{60\rightarrow82}$. Consequently, approaching of D12-K16 pair as a result of salt bridge formation between them affects the dynamics of distant residue pairs such as G60-Q70, Q61- G75 and G60-F82.

Relation between Conformational and Dynamic Changes in K-Ras^{G12D}

Fluctuations of dilated regions become negatively correlated in K-Ras^{G12D}. We observed that negative correlations occur between the regions which become distant from each other and also from their neighbors after G12D mutation. Combination of distance and correlation calculations gives us the relation between conformational and dynamic changes in the protein as a result of the mutation in its structure (Figure 5.10).

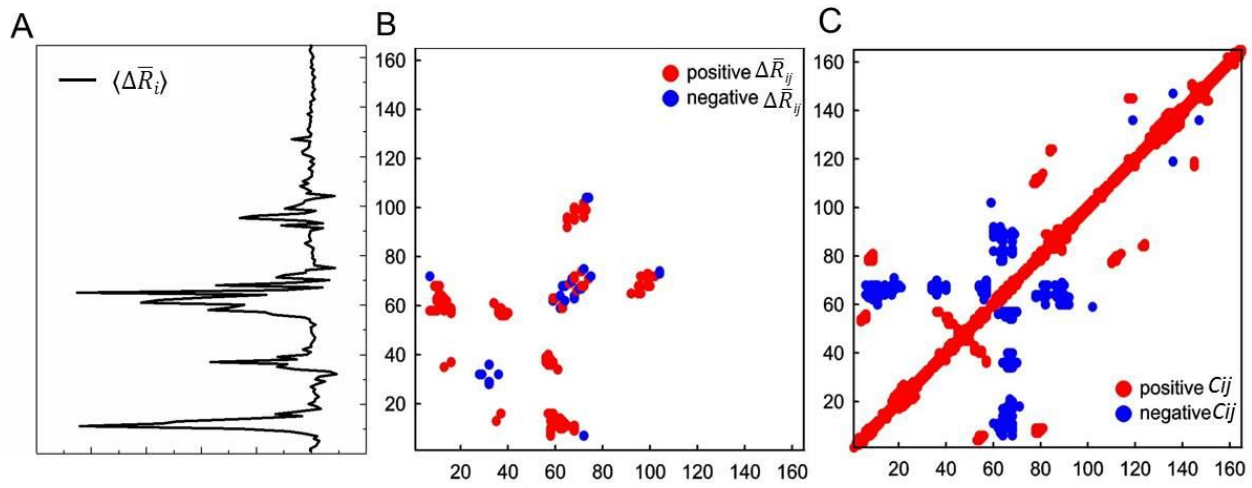


Figure 5.10 Comparison of the distance and correlation changes due to G12D mutation. (A) $\langle \Delta \bar{R}i \rangle$ values (B) Pairwise correlation coefficients (C) Difference of the averaged pairwise distances ($\Delta \bar{R}_{ij}$)

Relation between Negative Correlations and Characteristic Decay Times in K-Ras^{G12D}

We showed that fluctuations of SII region become negatively correlated with fluctuations of the P-loop, SI and α3 regions. Moreover, fluctuations of these regions have the longest correlation decay times in the protein (Figure 5.11). We wanted to see whether increased negative correlations between the residue fluctuations slow down the autocorrelation decay of the residue fluctuations. Therefore we investigated the relationship between the extent of correlations and decay rates. To identify the relation between the negative correlation values and the autocorrelation decay times, we calculated the average of negative correlation values for each residue of K-Ras^{G12D}. Figure 5.12 exhibits that the residues whose fluctuations more

negatively correlated with the other parts of the protein has longer characteristic decay times.

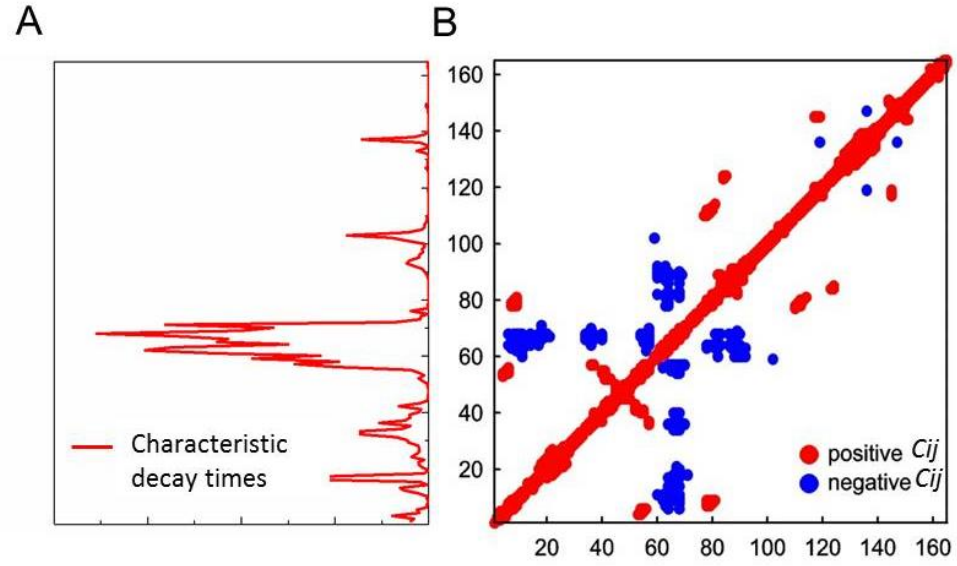


Figure 5.11 Comparison of characteristic decay times and pairwise correlations in K-Ras^{G12D}. (A) Characteristic decay times of K-Ras^{G12D} residues. (B) Pairwise correlations in K-Ras^{G12D}.

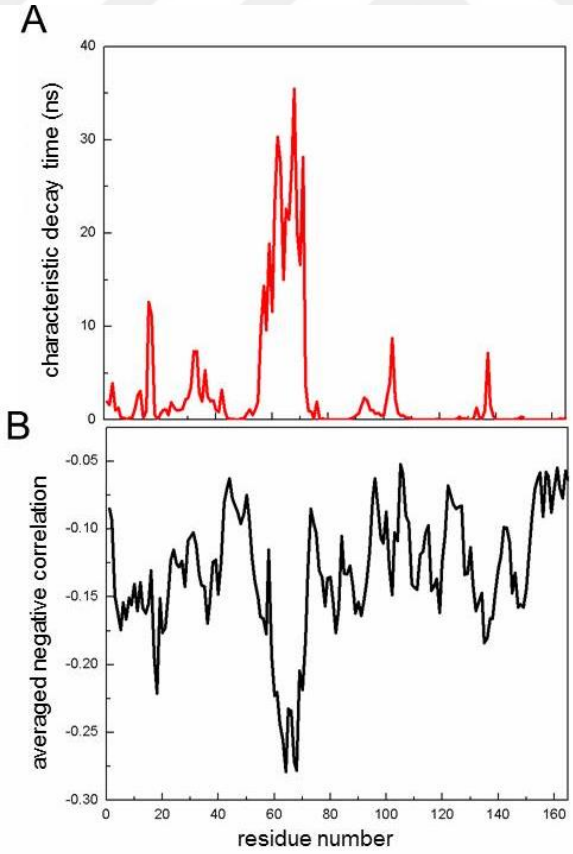


Figure 5.12 Comparison of characteristic decay times and averaged negative correlations in K-Ras^{G12D}. (A) Characteristic decay times of K-Ras^{G12D} residues. (B) Averaged negative correlation values per residue.

Relation between RMSF Values and Characteristic Decay Times in K-Ras^{G12D}

Autocorrelations of the residue fluctuations tend to decay slowly as fluctuation magnitudes of the residue increase. As identified by our calculations, residues of SII region have the longest characteristic decay times and the greatest RMSF values. To determine if there is a relation between RMSF values and the autocorrelation decay times, we drew the plot of RMSF vs. decay times per residue. Figure 5.13 displays that autocorrelation decay times become longer as the fluctuation values increase. The straight line is a least squares line through the points.

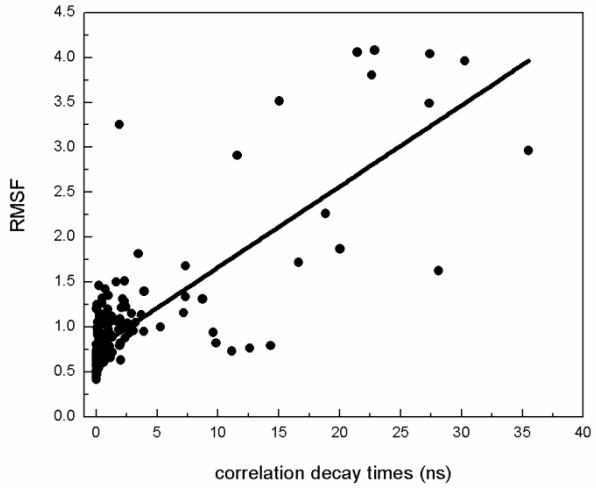


Figure 5.13 Relation between residue fluctuations and their decay times.

Relations between Negative Correlations, Characteristic Decay Times and Causality in K-Ras^{G12D}

Causality relationships occur within the negatively correlated residue pairs. We identified the driver-responder residue pairs in K-RAS^{G12D} dynamics. All of those pairs correspond to regions of the protein which exhibit negative correlations during MD simulations.

In causality relationship between K-Ras^{G12D}residue pairs, fluctuations of the driver residues show greater autocorrelation decay times compared to those of the responder residues. To understand the source of the causality relations in K-Ras^{G12D}, we compared the characteristic decay time of each residue in the driver-responder residue pairs. The driver residues have longer decay times than the responders. Slow decay of autocorrelation of residue fluctuations by time-delay corresponds to a specific form of "persistence". Therefore, we drew the autocorrelation decay curves of each residue in the driver-responder pairs (Figure 5.14). Our results show that the autocorrelation curves with slow decay correspond to driver residues.

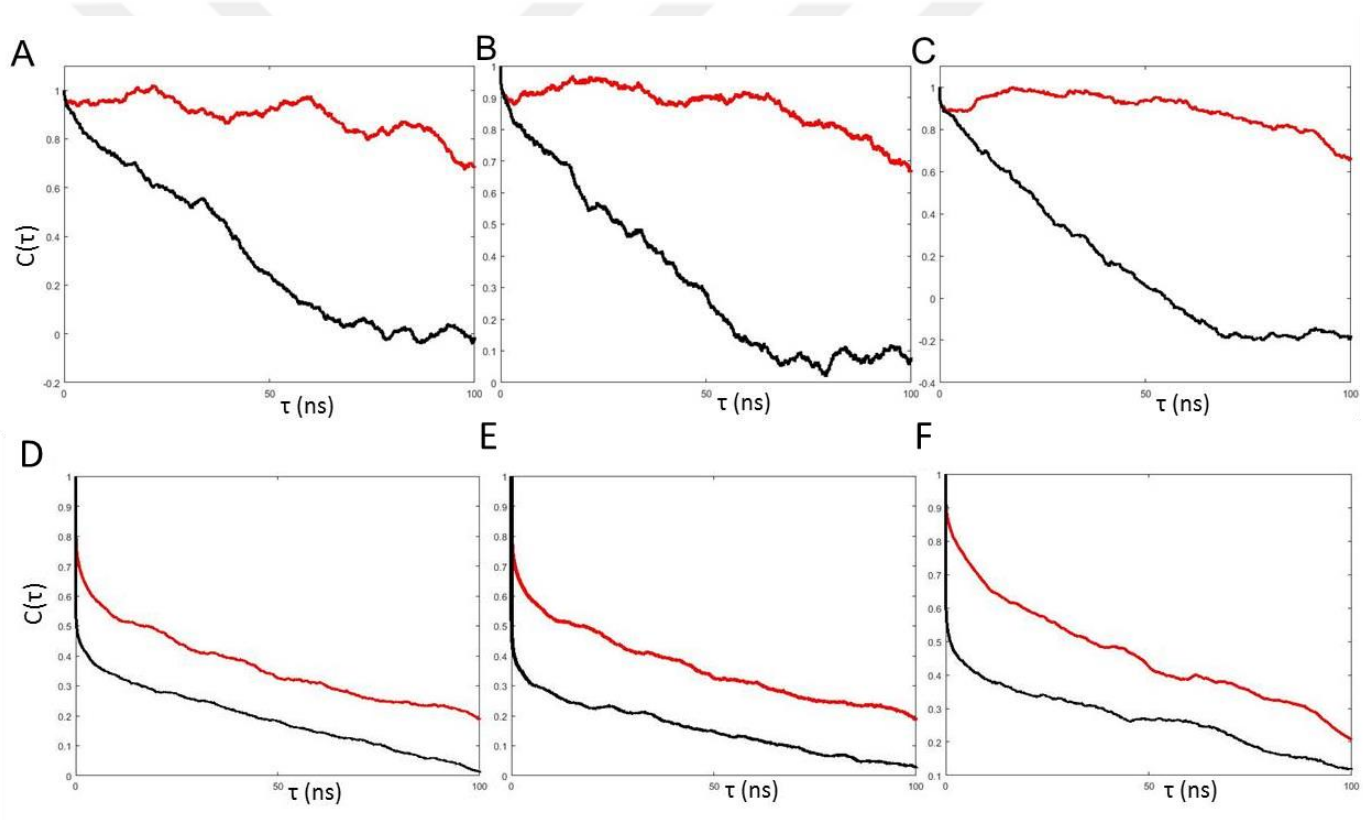


Figure 5.14 Autocorrelations for the fluctuations of the residues that are drivers or responders in K-Ras^{G12D}. A to C**:** Red curves for $\langle \Delta \boldsymbol{R_i}(\boldsymbol{t}) \Delta \boldsymbol{R_j}(\boldsymbol{t}+\boldsymbol{\tau}) \rangle$ show that the fluctuations of residue *i* at time *t* affect the fluctuations of residue *j* at a later time *t*+*τ*. X axes is for $\boldsymbol{t}$ values from 1 ns to 100ns. All correlations ($\boldsymbol{C}(\boldsymbol{\tau})$) are normalized with respect their value at zero ($\boldsymbol{C}(\boldsymbol{0})$) and shown in Y axes. **D to F:** Curves for $\langle \Delta \boldsymbol{R_i}(\boldsymbol{t}) \Delta \boldsymbol{R_i}(\boldsymbol{t}+\boldsymbol{\tau}) \rangle$ show that fluctuations of residue *i* at time *t* affect its fluctuations at a later time *t*+*τ*. Red curves

correspond to autocorrelations with slow decay. X axes is for ***t*** values from 1 ns to 100ns. (A) Q70 (SII) fluctuations drive V9 (P-loop) fluctuations. (B) Residue Q70 (SII) fluctuations drive C80 (β4) fluctuations. (C) D69 (SII) fluctuations drive D54 (β3) fluctuations. (D) Autocorrelations of residues Q70(driver)-V9(responder). Red line is autocorrelation decay curve of residue Q70, black line is autocorrelation decay curve of residue V9. (E) Autocorrelations of residues Q70(driver)-C80(responder). Red line is autocorrelation decay curve of residue Q70, black line is autocorrelation decay curve of residue C80. (F) Autocorrelations of residues D69(driver)-D54(responder). Red line is autocorrelation decay curve of residue D69, black line is autocorrelation decay curve of residue D54.

5.4 DISCUSSION

K-Ras is a small GTPase protein and its intrinsic GTPase function is accelerated by GAP proteins. In active K-Ras-GTP, the P-loop, SI and SII parts are bound to the phosphate groups of GTP and are responsible for the GTPase function of K-Ras. However, G12D mutation in the P-loop impairs GTP hydrolysis and freezes K-Ras in its active state[14], causing uncontrollable cellular growth and evasion of apoptotic signals[217-219]. In order to quantify conformational and dynamical changes caused by G12D mutation, we combined different MD data analysis methods. Indeed, we could identify the relation between the conformational changes of residues and the alterations in their dynamic correlations. Furthermore, we could also show the causality between the residue pairs and its relationship with the persistency of residue fluctuation in K-Ras^{GTP} for the first time.

The broad distribution of Q61-P-loop distances has an effect on GTP hydrolysis function of K-Ras^{G12D}. In G12D mutation, glycine residue is substituted by aspartate with bulkier side group and this structural change of the P-loop affect the conformations of other active site regions, SI and SII, as observed in the previous studies[218, 220-222]. Our distance calculations show that after G12D mutation, the P-loop residues 11, 12, 13 move away from SII residue Q61. Moreover, Q61-A11, Q61-G12D and Q61-G13 distances display broad distributions (Figure 5.3). Since Q61 is known as a catalytic residue playing a critical role in

both intrinsic and GAP-mediated GTP hydrolysis[15], the highly variable character of the P-loop-Q61 distance may affect the GTP hydrolysis in K-Ras^{G12D}. Furthermore, our distance calculations also show that the residues in SII move away from some residues in SI, α3 and distances between those residue pairs also display broad distributions in K-Ras^{G12D} (Figure 5.3). Considering the increased fluctuations of SII, we may say that the broad distance distributions with larger peak values between SII and the other parts of the protein arise from higher flexibility of SII in K-Ras^{G12D}.

Moving away of E37 from D57 in K-Ras^{G12D}affects the Raf effector binding. E37 and D57 been identified as hot-spot residues for the interaction of K-Ras with Raf protein[223]. Moreover, impairment of the interaction of K-Ras- Raf by G12D mutation has been observed in the previous studies[224, 225]. Our distance analyses exhibit that these two hot-spot residues move away from each other and also from their neighbors in K-Ras^{G12D}. These conformational changes of E37- D57 pair may disrupt the binding of K-Ras^{G12D} to Raf protein.

Approaching of residues R73 and K104 upon mutation affects the PI3Kγ effector binding. Pairwise distance calculations indicated that the distances of residue pairs 63-68, 72-75 and 73-104 significantly decreased upon G12D mutation. Among the all pairs, 73-104 has own significance since R73 is critical for interaction with the effector protein PI3Kγ[226], which is preferably activated by K-Ras^{G12D} with higher binding affinity[224, 225]. Additionally, the backbone carbonyls of R73 interact with the amino sidechain of K104[227]. Therefore, the decreased distance between R73 and K104 in K-Ras^{G12D} (Figure 5.3) may affect the interaction of these residues that allows PI3Kγ binding with higher affinity.

Distant conformations of active site residues may disturb the GTPase function in K-Ras^{G12D}. K-Ras undergoes conformational changes when it binds to GTP. The P-loop, SI and

SII constitute the active site of the protein that binds to phosphate groups of GTP and participate in GTP hydrolysis[15]. SI and SII are also responsible for controlling the binding to effector molecules. Conformational changes in the active site affect K-Ras interactions with the GAPs which amplify the GTPase activity of K-Ras[228]. Our distance calculations showed that the active site residues of K-Ras^{G12D} move away from their neighbors (Figure5.5) in consequence of the structural change in the P-loop. Moreover, SII fluctuations increase in the mutant form (Figure 5.6). These observations are consistent with the previous studies[218]. Since conformations of the active site play important roles in GTPase activity of K-Ras and its binding ability to GAPs, we postulate that the deviation of active site residues may impair the GTP hydrolysis and also GAP binding ability which leads to the constitutively active K-Ras^{G12D}.

G12D mutation augments the correlations between SII and the regions by increasing the flexibility of SII. As seen in the pairwise correlation maps, K-Ras^{G12D}differs markedly from K-Ras^{WT} and display increased levels of fluctuation correlations between SII residues and the other parts of the protein. This is consistent with Figure 12 which demonstrates that the amplitude of the SII fluctuations is larger than K-Ras^{G12D}. These results have been observed in a previous study, too[218].

6 INTRINSIC DYNAMICS OF MUTANT K-RAS: COMPARISON OF ACTIVE AND INACTIVE MUTANT K-RAS

6.1 INTRODUCTION

K-Ras, the most frequently mutated oncoprotein in human cancers[1-3], is found in inactive GDP-bound and active GTP-bound states, as we stated ın Chapter 1[13, 14]. It can bind and activate its downstream effector proteins only in its active state[15]. However, oncogenic K-Ras mutations lock the protein into a constitutively activate state in which K-Ras prolongs downstream signaling associated with oncogenic cell growth[15, 20-22]. In Chapter 5, we investigated how G12D mutation, the most prevalent mutation in cancer patients, alters the regulation of K-Ras dynamics that results in constitutively active K-Ras^{G12D}. However, for selectively inhibit the constitutively active K-Ras^{G12D} by drug binding, we need to identify differences between active and inactive K-Ras^{G12D}. Since structural studies are not enough to identify pockets which are found only on active K-Ras^{G12D} surface that inhibitor drugs can selectively bind to, we compared active and inactive K-Ras^{G12D} dynamics to identify such pockets. We performed detailed analysis of both active K-Ras^{G12D}-GTP and inactive K-Ras^{G12D}-GDP dynamics and provided quantitative, comprehensive descriptions of these two states from extensive MD simulations.

We characterized persistent correlated motions within the MD simulations of active K-Ras^{G12D}-GTP and inactive K-Ras^{G12D}-GDP, we mapped the correlated motion patterns within their residues individually, and then compared them each other and with the K-Ras^{WT}. Our

results showed that active K-Ras^{G12D}-GTP is marked by an increase in correlated motions of residues comparing to both inactive K-Ras^{G12D}-GDP and K-Ras^{WT}-GTP.

We, then, aimed to understand the source of correlated motions in active active K-Ras^{G12D}-GTP. For this purpose, we identified the residue pairs which have the most significant correlated motions in active K-Ras^{G12D}-GTP. Then we counted H-bonds between those residues and the protein during the simulations of active and inactive K-Ras^{G12D}-GDP.

In addition to H-bond calculations, we compared the distribution of distances between those residue pairs for active and inactive K-Ras^{G12D}-GDP. In order to identify which residues' motions cause the changes in distance distributions, we also compared the distribution of residue fluctuations.

We used a statistical thermodynamics interpretation of residue motions to quantify the deformation in K-Ras^{G12D} that is caused by the mutation and inactivation. We describe the deformation as "*strain*", in terms of relative displacement of residues in the protein. We calculated the pairwise strain values by comparing active K-Ras^{G12D}-GTP with inactive K-Ras^{G12D}-GDP and K-Ras^{WT}-GTP, respectively.

6.2 METHODS

For calculations of fluctuations, cross-correlations and pairwise distances and their distribution, we used the same methods in Chapter 4.

6.2.1 Calculation of pairwise strain values

We assumed that the strain as a description of deformation in terms of relative displacement of residues in the protein. We made different expressions of a strain values depending on whether it is defined with respect to the reference or the final configuration of the protein. We

defined the strain between two residues ($\boldsymbol{strain_{ij}}$) by the formula of $\boldsymbol{strain_{ij}} = \frac{\overline{R_{ij}}_{FINAL} - \overline{R_{ij}}_{REFERENCE}}{\overline{R_{ij}}_{REFERENCE}}$, where $\overline{R_{ij}}$ is the distance of the two residues. We calculate the strain values between residue pairs following the algorithm in Figure 6.1.

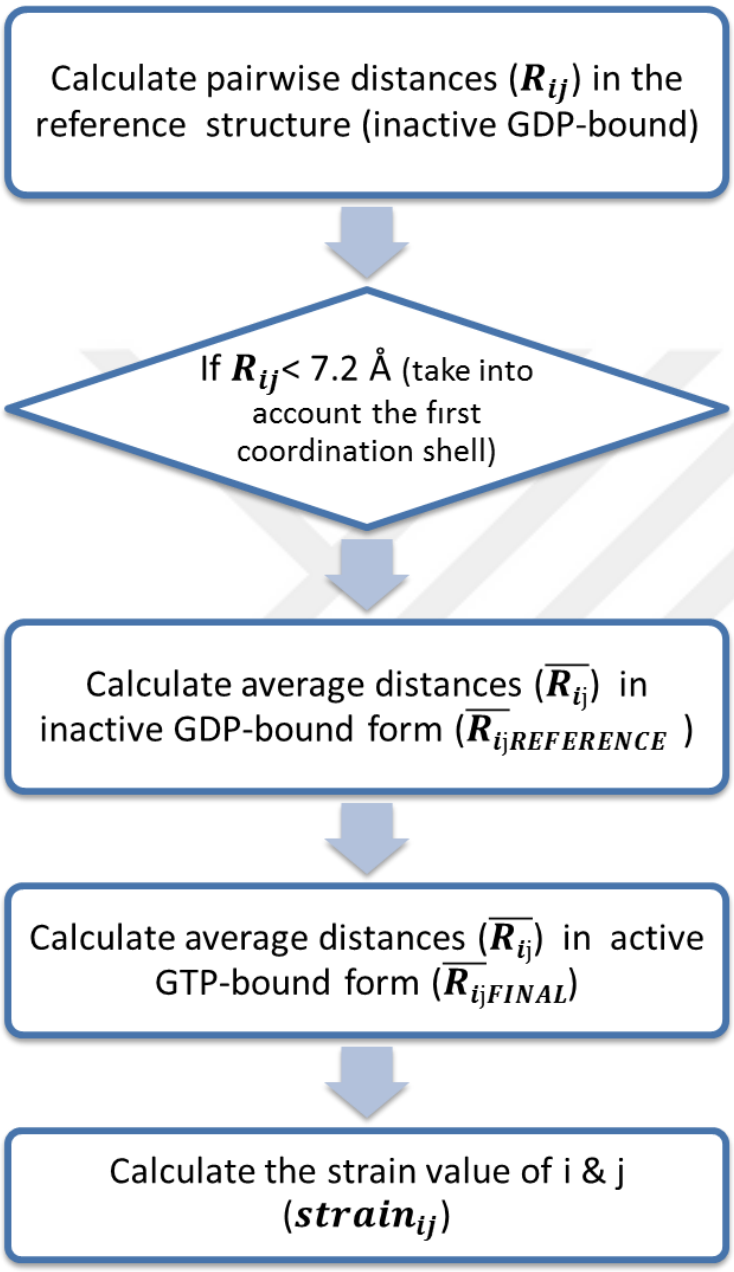


Figure 6.1 Algorithm of the pairwise strain value calculation.

6.3 RESULTS AND DISCUSSION

6.3.1 Comparison of residue pair correlations

To identify if the fluctuations of one residue are related to fluctuations of another residue, we calculated the correlations of all residue-residue pairs in both active and inactive K-Ras^{G12D}. As shown in Figure 6.2, correlation coefficient maps of active and inactive K-Ras^{G12D} exhibit different patterns. Active K-Ras^{G12D} shows increased correlations between residue pairs.

We, then, identified residue pairs that have the most significant correlations in active K-Ras^{G12D}. We found out two groups of correlated residues, where residues 64 and 67 act as the cores of the correlations as shown in Figure 6.3. In the first group, residue 67 motions are correlated with residues 17, 36 and 66; in the second group residue 67 motions are correlated with residues 14, 82, 90, 100 and 125.

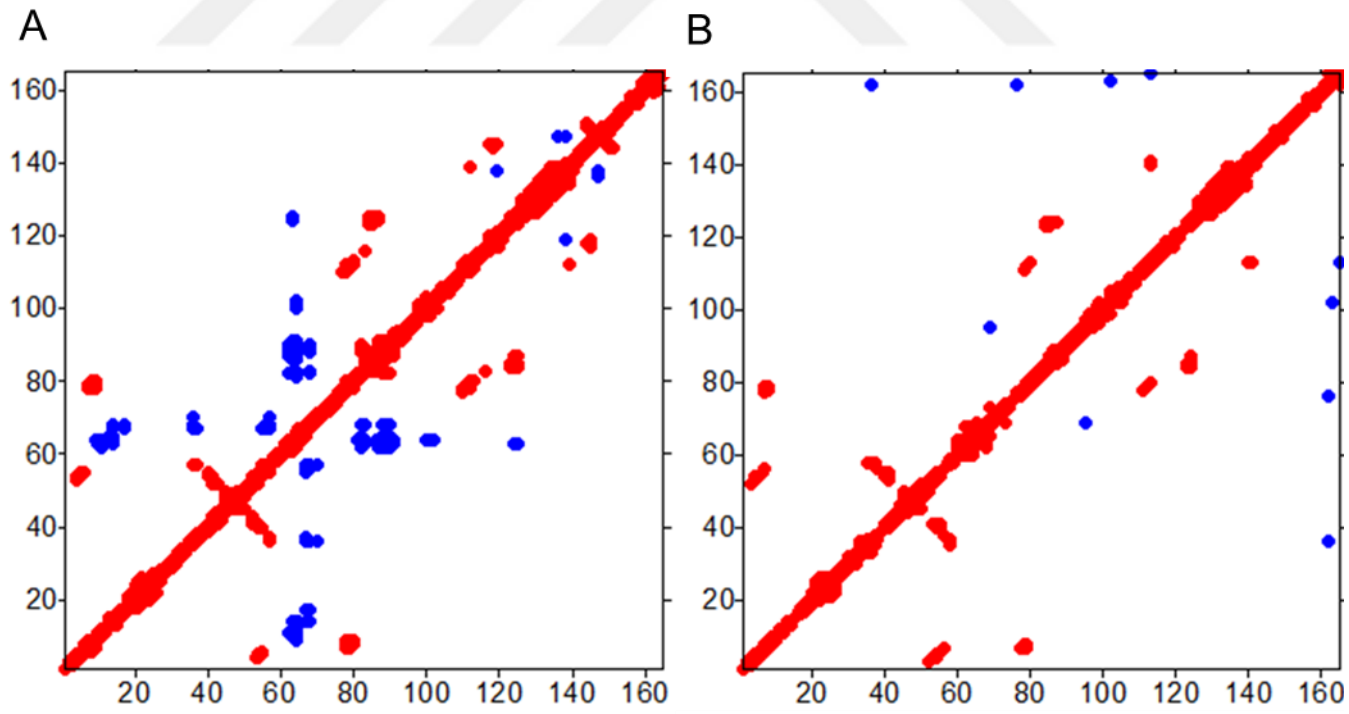


Figure 6.2 Cross-correlation coefficient maps for active and inactive K-Ras^{G12D}. Red dots show positive correlations ($1 \geq C(\Delta Ri, \Delta Rj) \geq 0.6$) and blue dots show negative correlations

(-0.45 ≥ C(ΔRi, ΔRj) ≥ -1). Residues indices 1-165 refer to K-Ras^{G12D}. (A) Correlated fluctuations of K-Ras^{G12D}-GTP (B) Correlated fluctuations of K-Ras^{G12D}-GDP

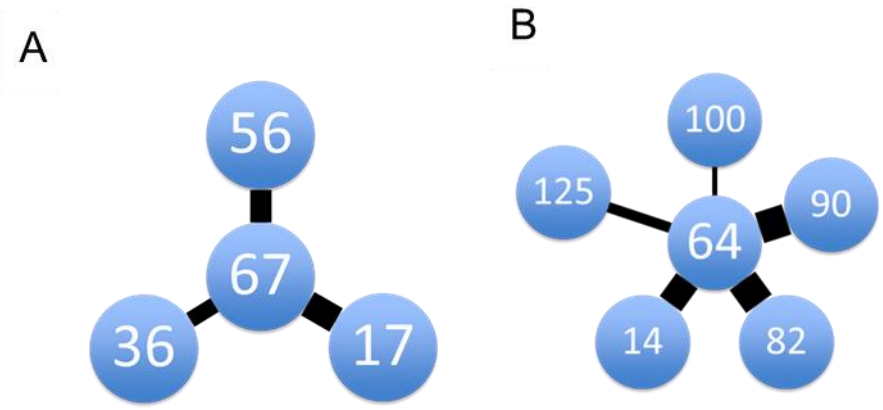


Figure 6.3 Scheme of the two groups of residue pairs which have the most significant correlations in active K-Ras^{G12D}. (A) Residue 67 motions are correlated with residues 17, 36 and 66. (B) Residue 67 motions are correlated with residues 14, 82, 90, 100 and 125.

6.3.2 H-bonds calculation

To understand the sources of increased correlated motions in active K-Ras^{G12D}, we calculated the average number of H-bonds that are established between each correlated residues and the protein throughout the simulation. We compared the average number of H-bonds in active and inactive forms; however, there is no significant difference between them.

6.3.3 Comparison of distance and fluctuation distributions of correlated residue pairs

After identification of the residue pairs whose fluctuations become significantly correlated in active K-Ras^{G12D}, we compared the distribution of distances between those residue pairs for active and inactive K-Ras^{G12D} (Figures 6.4-6.11). Distance distribution graphs exhibit that inactive K-Ras^{G12D} is characterized by Gaussian-shaped distribution curves with stable distance values. However, distribution curves for active K-Ras display significant deviations from the Gaussian. Although average of the distance values between the residue pairs are

almost similar, those distances are quite changeable in active K-Ras^{G12D} while they are almost stable in inactive K-Ras^{G12D}. Consequently, activation of K-Ras^{G12D} causes correlated motions of residue pairs by destabilizing their relative conformations. We, then, aimed to understand which residues' motions cause those highly changeable residue pair distances in active K-Ras^{G12D}. Considering that irregular individual fluctuations of residues may cause the changes in distances, we calculated the fluctuations of each residue in the identified residue pairs for every time step of the active and inactive K-Ras^{G12D} simulation. Then we compared the distribution patterns of individual residue fluctuations for active and inactive K-Ras^{G12D}. Fluctuation distribution curves in Figures 6.4-6.11 revealed that fluctuations of residues 64 and 67 increase and also become irregular upon activation of K-Ras^{G12D}. Interestingly, fluctuations of the same residues become correlated with the others acting as the cores of the correlated fluctuations in active K-Ras^{G12D}. However, fluctuations of the other residues, which are correlated with the residues 64 and 67, are stable around the similar values in both active and inactive K-Ras^{G12D}. We may conclude that increased and changeable fluctuations of the residues 64 and 67 in active K-Ras^{G12D} allow them become correlated different parts of the protein. Knowing the fluctuation value is a measure of flexibility[56], our results indicate that activation of K-Ras^{G12D} increases the flexibility of SII at residues 64 and 67. Moreover, thanks to the flexible nature of residues 64 and 67, these residues can move with the other parts of the protein in correlation. Those correlated motions which arose from the flexible SII may prolong the active state of K-Ras^{G12D}.

Residues 67&17

In Figure 6.4A, distance distributions show that the distance of 67-17 is not stable in active K-Ras^{G12D} and this instability is caused by the increased fluctuations of residue 67, as shown in Figure 6.4B and 6.4C.

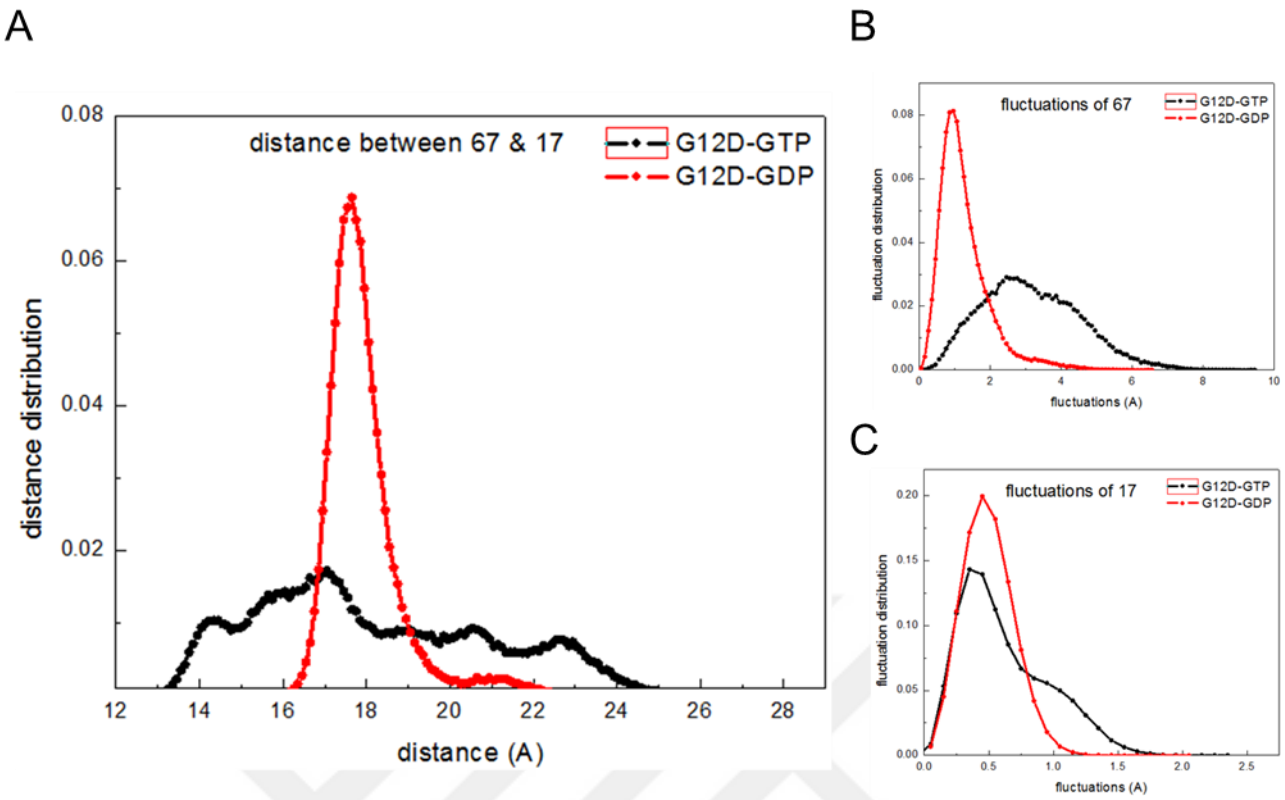


Figure 6.4 Distance and fluctuation distributions of residues 67 and 17 in K-Ras^{G12D}. (A) Distance distribution between and residues 67-17 in active (black) and inactive K-Ras^{G12D} (red). (B) Fluctuation distribution of residue 67 (C) Fluctuation distribution of residue 17.

Residues 67&36

In Figure 6.5A, distance distributions show that the distance of 67-36 is not stable in active K-Ras^{G12D} and this instability is caused by the increased fluctuations of residue 67, as shown in Figure 6.5B and 6.5C.

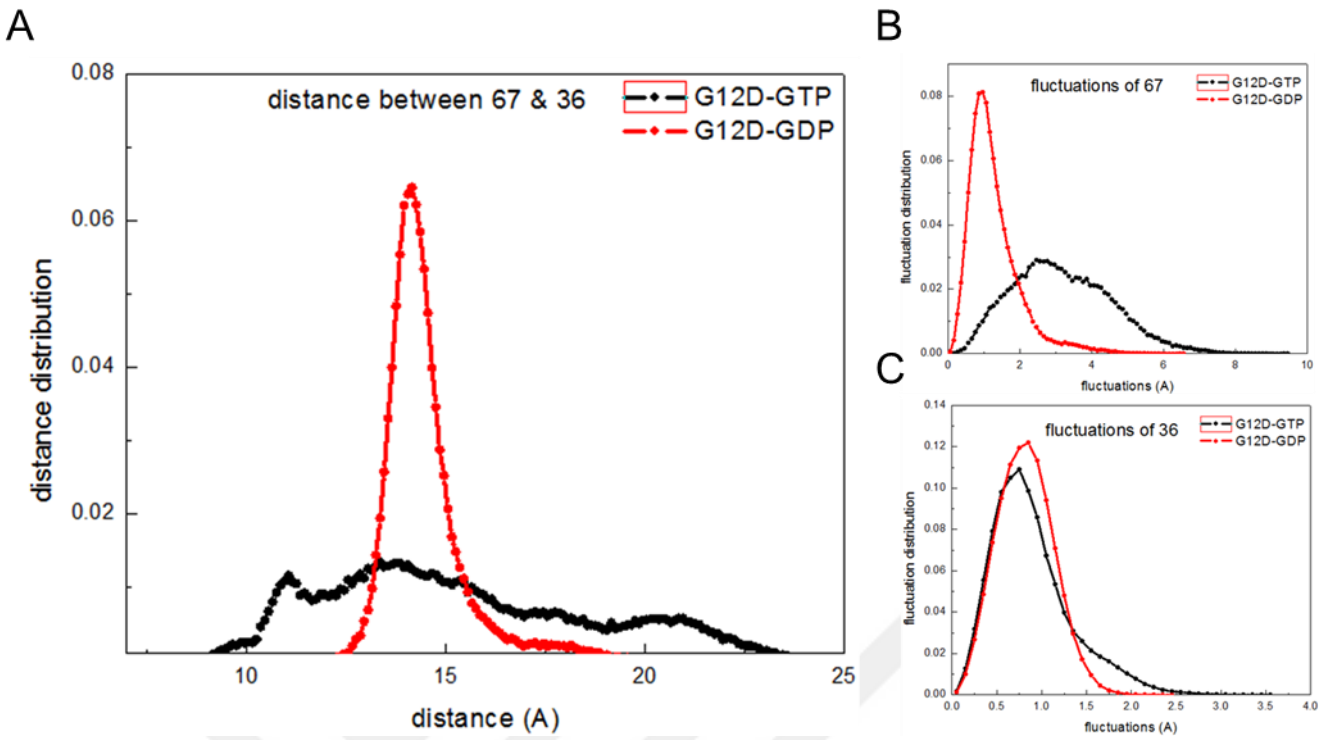


Figure 6.5 Distance and fluctuation distributions of residues 67 and 36 in K-Ras^{G12D}. (A) Distance distribution between and residues 67-36 in active (black) and inactive K-Ras^{G12D} (red). (B) Fluctuation distribution of residue 67. (C) Fluctuation distribution of residue 36.

Residues 67&56

In Figure 6.6A, distance distributions show that the distance of 67-56 is not stable in active K-Ras^{G12D} and this instability is caused by the increased fluctuations of residue 67, as shown in Figure 6.6B and 6.6C.

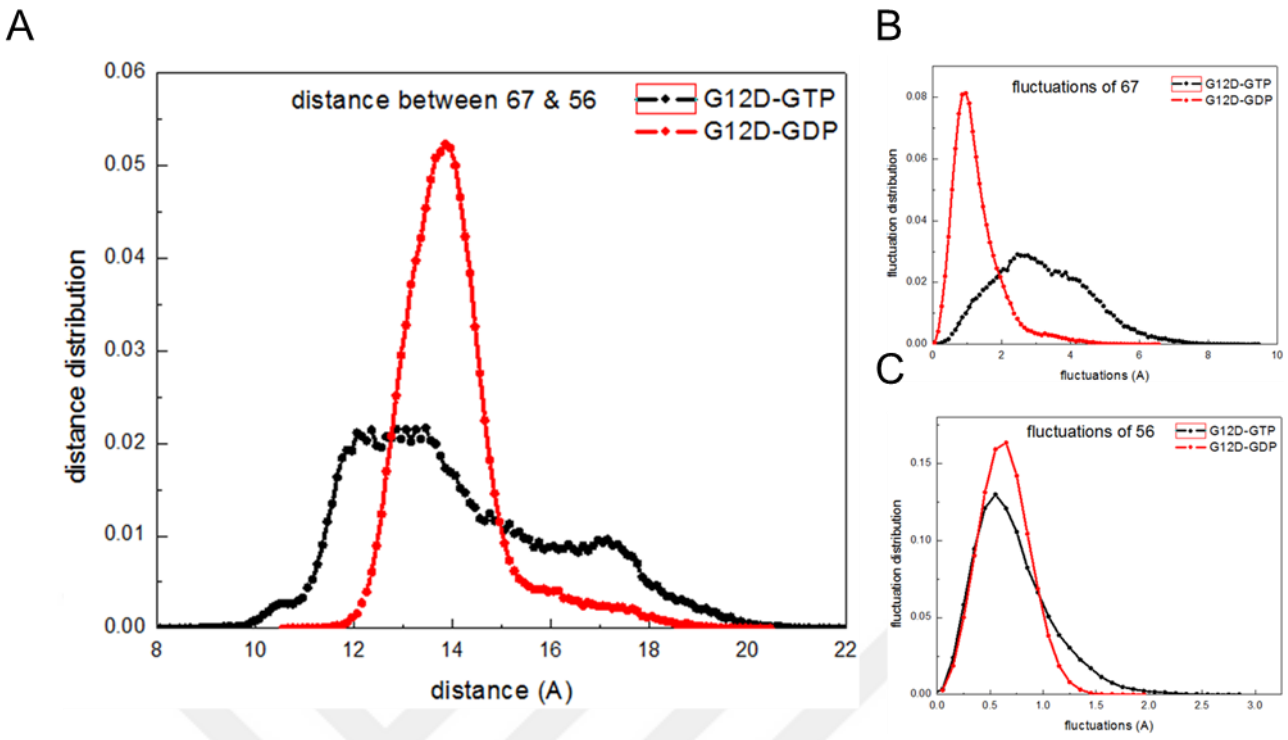


Figure 6.6 Distance and fluctuation distributions of residues 67 and 56 in K-Ras^{G12D}. (A) Distance distribution between and residues 67-56 in active (black) and inactive K-Ras^{G12D} (red). (B) Fluctuation distribution of residue 67. (C) Fluctuation distribution of residue 56.

Residues 64&14

In Figure 6.7A, distance distributions show that the distance of 64-14 is not stable in active K-Ras^{G12D} and this instability is caused by the increased fluctuations of residue 64, as shown in Figure 6.7B and 6.7C.

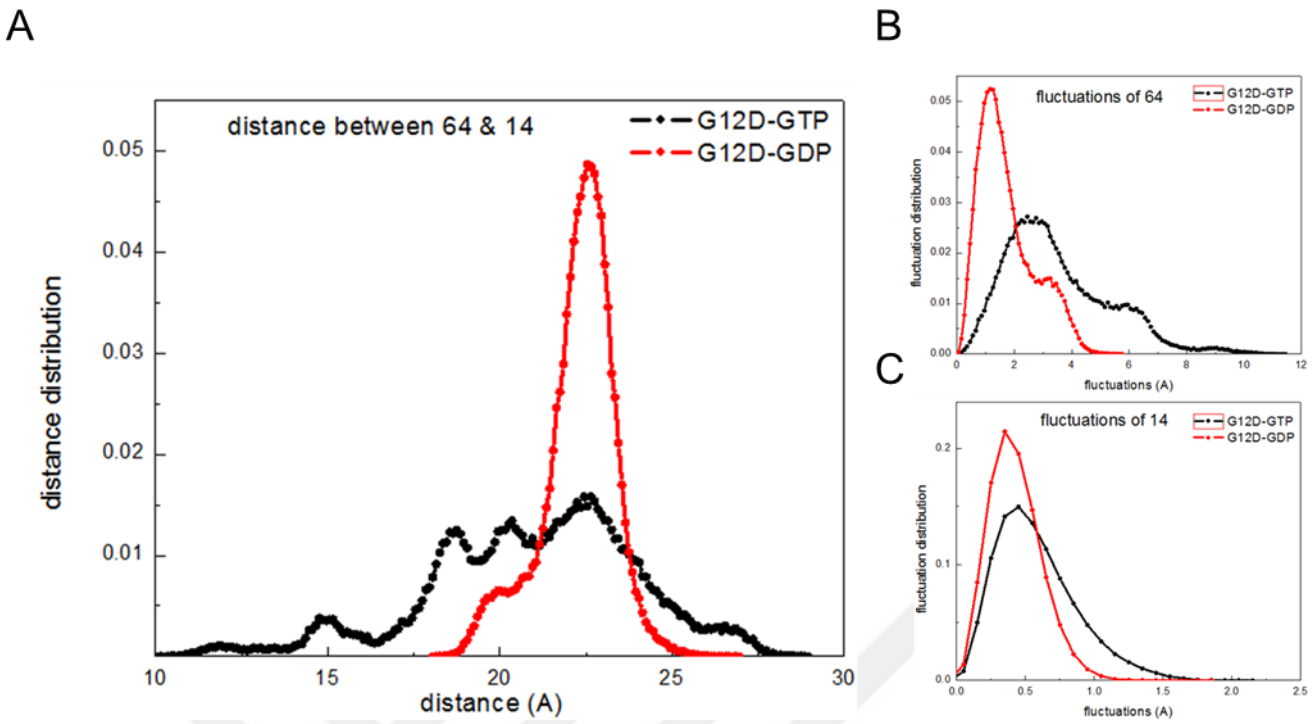


Figure 6.7 Distance and fluctuation distributions of residues 64 and 14 in K-Ras^{G12D}. (A) Distance distribution between and residues 64-14 in active (black) and inactive K-Ras^{G12D} (red). (B) Fluctuation distribution of residue 64. (C) Fluctuation distribution of residue 14.

Residues 64&82

In Figure 6.8A, distance distributions show that the distance of 64-82 is not stable in active K-Ras^{G12D} and this instability is caused by the increased fluctuations of residue 64, as shown in Figure 6.8B and 6.8C.

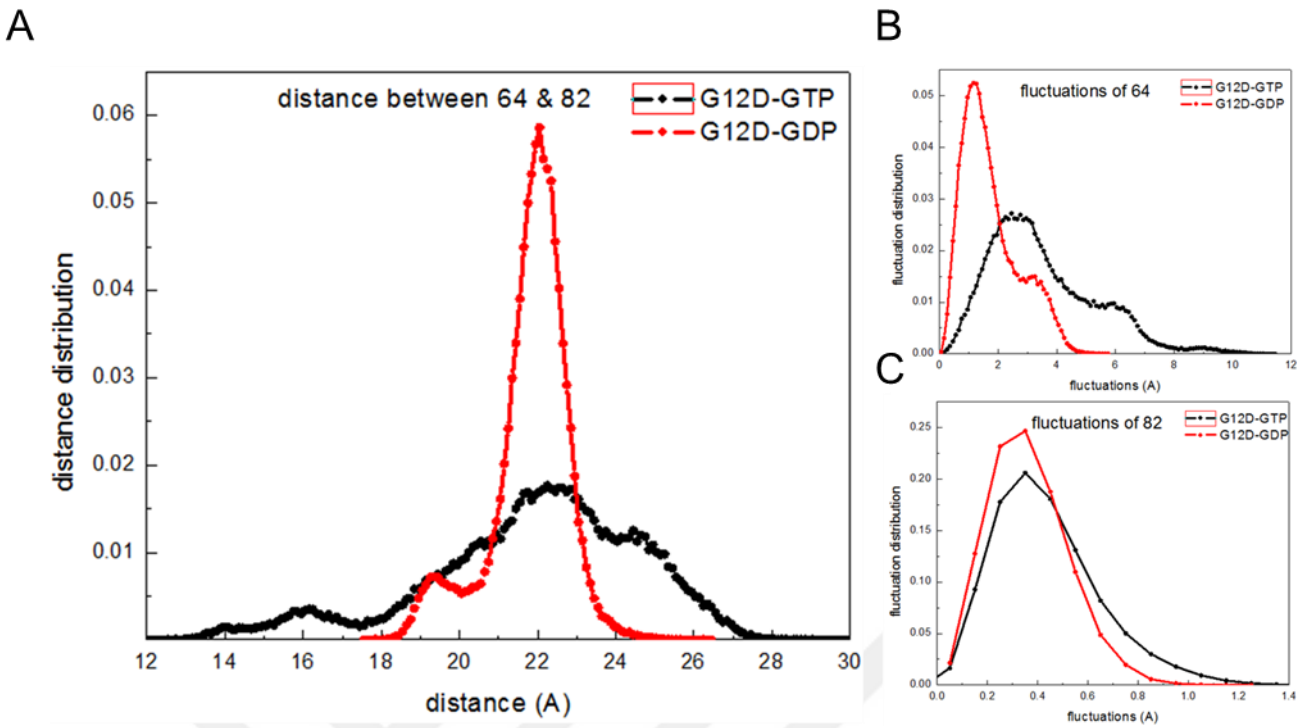


Figure 6.8 Distance and fluctuation distributions of residues 64 and 82 in K-Ras^{G12D}. (A) Distance distribution between and residues 64-82 in active (black) and inactive K-Ras^{G12D} (red). (B) Fluctuation distribution of residue 64. (C) Fluctuation distribution of residue 82.

Residues 64&90

In Figure 6.9A, distance distributions show that the distance of 64-90 is not stable in active K-Ras^{G12D} and this instability is caused by the increased fluctuations of residue 64, as shown in Figure 6.9B and 6.9C.

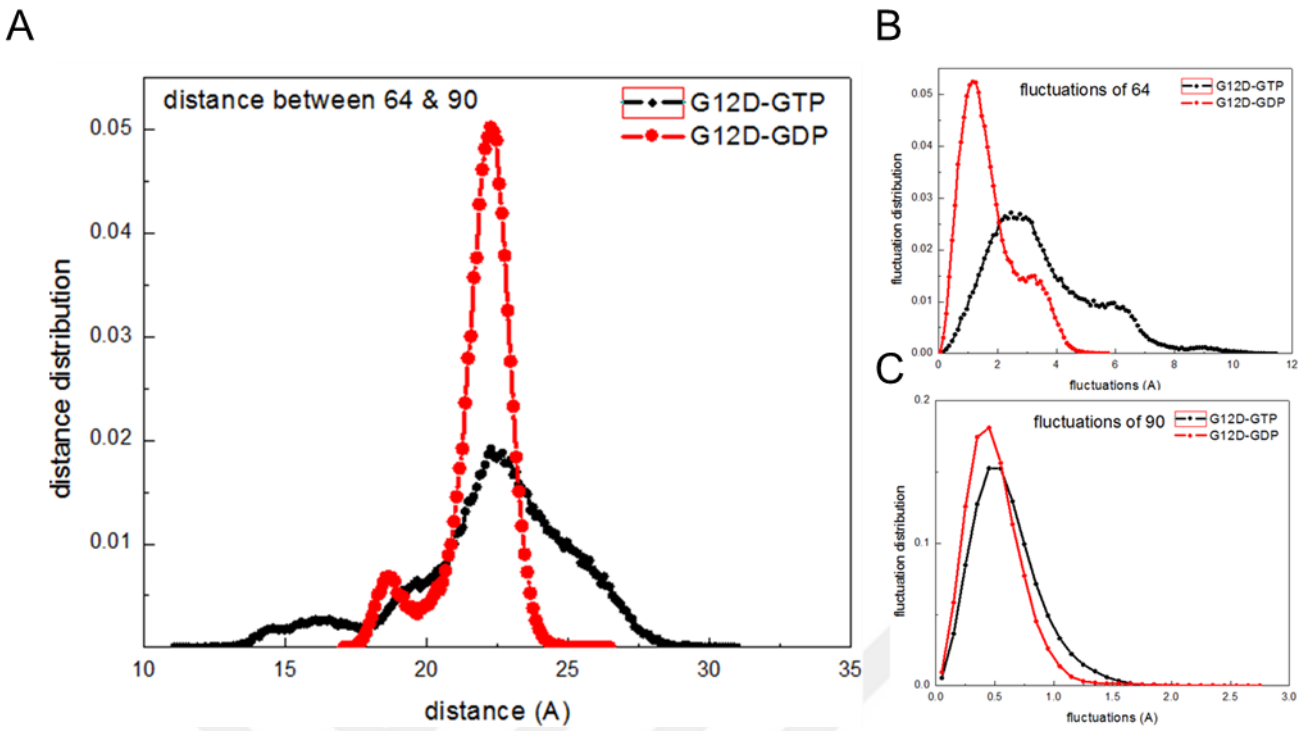


Figure 6.9 Distance and fluctuation distributions of residues 64 and 90 in K-Ras^{G12D}. (A) Distance distribution between and residues 64-90 in active (black) and inactive K-Ras^{G12D} (red). (B) Fluctuation distribution of residue 64. (C) Fluctuation distribution of residue 90.

Residues 64&100

In Figure 6.10A, distance distributions show that the distance of 64-100 is not stable in active K-Ras^{G12D} and this instability is caused by the increased fluctuations of residue 64, as shown in Figure 6.10B and 6.10C.

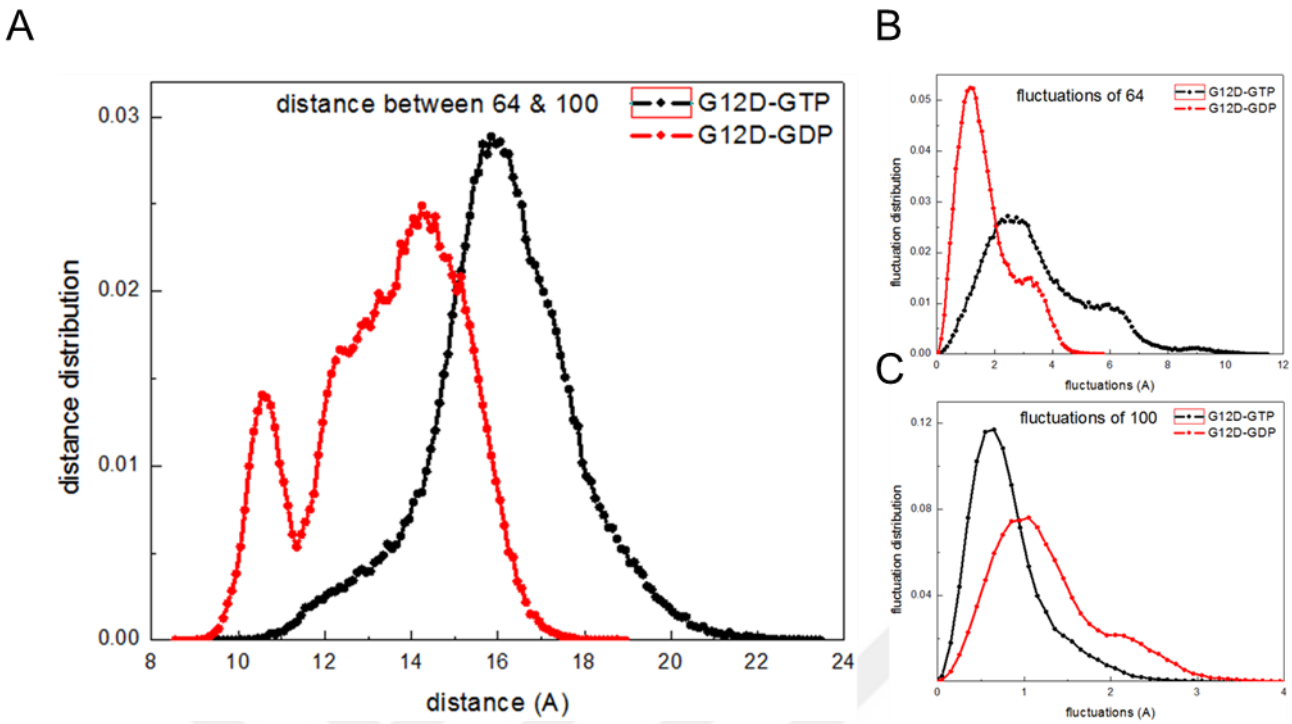


Figure 6.10 . Distance and fluctuation distributions of residues 64 and 100 in K-Ras^{G12D}. (A) Distance distribution between and residues 64-100 in active (black) and inactive K-Ras^{G12D} (red). (B) Fluctuation distribution of residue 64. (C) Fluctuation distribution of residue 100.

Residues 64&125

In Figure 6.11A, distance distributions show that the distance of 64-125 is not stable in active K-Ras^{G12D} and this instability is caused by the increased fluctuations of residue 64, as shown in Figure 6.11B and 6.1C.

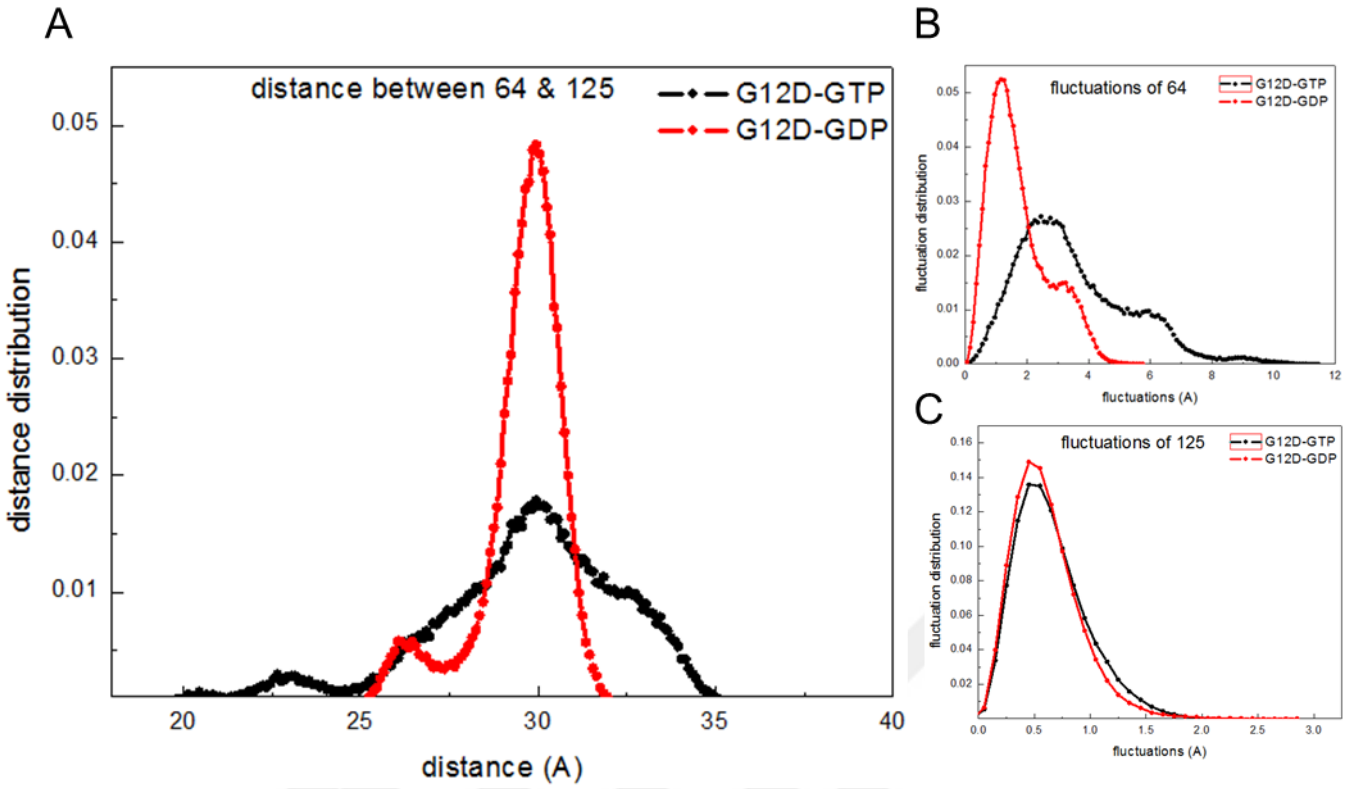


Figure 6.11 Distance and fluctuation distributions of residues 64 and 125 in K-Ras^{G12D}. (A) Distance distribution between and residues 64-100 in active (black) and inactive K-Ras^{G12D} (red). (B) Fluctuation distribution of residue 64. (C) Fluctuation distribution of residue 125.

6.3.4 Comparison of pairwise strain values

In order to quantify the relative displacement between the residue pairs as a result of activation, we calculated the strain of all residue-residue pairs where the reference state is inactive K-Ras^{G12D} and the final state is active K-Ras^{G12D}. Figure 6.12 shows the pairwise strain values. Since the strain of a residue pair implies the dilatation rate, positive strain values indicate the local dilatation and negative strain values indicate the local compression in the protein. As seen in Figure 6.12A, SI and the P-loop, β3 become dilated after activation of K-Ras^{G12D}. On the other hand, residues within α2 and α3 become compressed.

We, then, calculated the strain of all residue-residue pairs while the reference state is K-Ras^{WT} and the final state is K-Ras^{G12D}, to quantify the relative displacement between the residue pairs as a result of mutation. As seen in Figure 6.12B, SI and the P-loop, β3 become dilated after activation of K-Ras^{G12D}. On the other hand, residues within α2 become compressed.

Considering the similarities between two graphs in Figure 6.12 such as dilatation of SI and the P-loop, β3 and compression of α2, G12D mutation have the same effect with the activation of the protein on these regions.

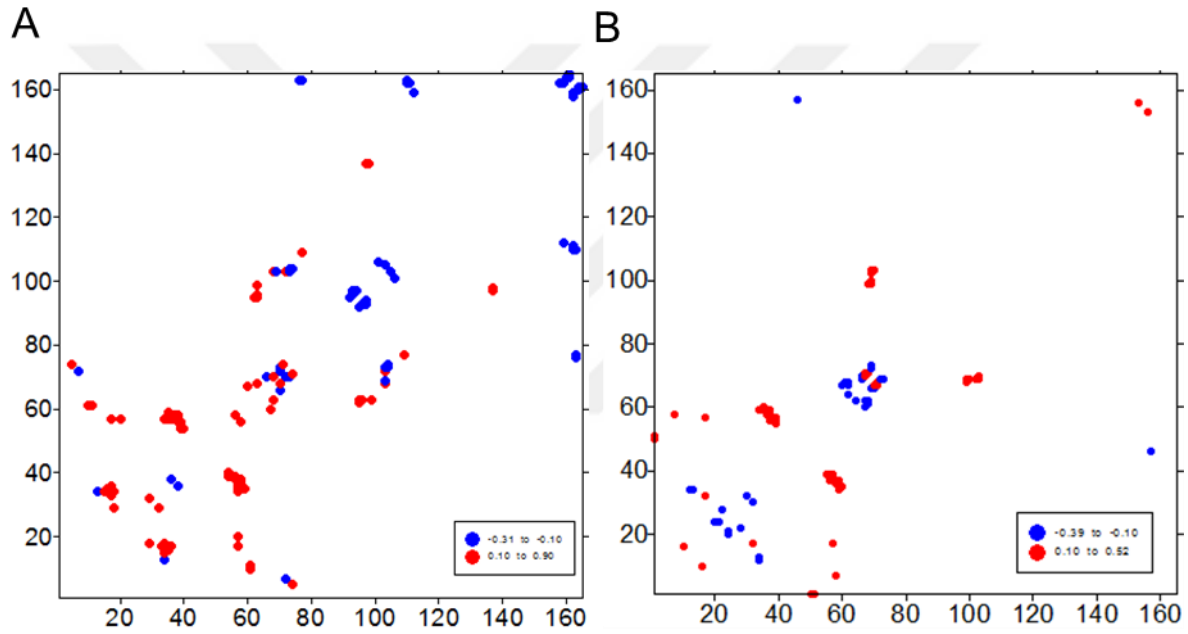


Figure 6.12 Pairwise strain values in K-Ras. Positive values are red and negative values are blue. (A) The reference state is inactive K-Ras^{G12D} and the final state is active K-Ras^{G12D}. (B) The reference state is WT K-Ras and the final state is mutant K-Ras.

7 DRUG BINDING INTO THE NOVEL TARGET POCKET IN MUTANT K-RAS

7.1 INTRODUCTION

Despite more than three decades of extensive research, we still have no drugs that no effective direct inhibitors of K-Ras in the clinic, leading perception that the most frequently mutated oncoprotein in human cancers, K-Ras, is "undruggable". Since structural studies alone are not enough to reveal drug target pockets on protein surface, only the studies that combine structural and dynamical analysis have been able to show promising results. Unfortunately, those studies include only G12C mutant while the dynamic regulation of the other mutants remain unknown. Since protein activity is related to its dynamics, we considered that knowledge of the dynamic regulation of mutant K-Ras can be utilized for identifying drug target sites on K-Ras surface. Therefore, we aimed to understand the dynamic regulation to the most frequently mutated K-Ras, G12D. However, identification of the effects of oncogenic mutation G12D on dynamic regulation of K-Ras requires a deep understanding of wild-type K-Ras dynamics. Thus, we first performed comprehensive analysis of intrinsic K-Ras dynamics that serve as a reference point for mutant K-Ras analysis. We, then, analyzed K-Ras^{G12D} dynamics and compared the results for K-Ras^{G12D} and K-Ras^{WT}. We observed that the residue 12 can make a salt bridge with K16 when it's mutated to aspartic acid. In K-Ras^{G12D}, residues D12 and K16 become closer via this salt bridge and the convergence of this pair establishes correlations with the conformations of the other residue pairs such as G60-Q70 (α2) and G60-F82 (β4). Moreover, among the residues of these pairs Q70 (α2) acts as the driver of the β4 region including F82. Along with the comparison of wild-type and mutant

protein, we also compared the active and inactive forms of the protein to identify possible drug target sites which are specific to the active form of mutant K-Ras^{G12D}. Detailed analysis and comparison of active and inactive K-Ras^{G12D} dynamics support the specific importance of residue F82 for active K-Ras^{G12D} dynamics. Our analysis showed that fluctuations of F82 are highly correlated with fluctuations of Y64. These results provided guidance for prediction of drug target sites on mutant K-Ras surface.

Finally, using detailed analysis of K-Ras^{G12D} dynamics and comparing them with the wild-type and inactive forms, we identified a novel drug target pocket in K-Ras^{G12D} that has not been shown in the previous experimental and computational studies. This pocket is observed in only active K-Ras^{G12D} and it disappears in K-Ras^{WT} (Figure 7.1).

In order to identify the small molecules that fit into novel binding site on K-Ras^{G12D} surface, we systematically tested all the drugs in the experimental molecules catalog (SHARDS) of ZINC database[229] by applying docking methodologies. We prioritized drug candidates based on their predicted binding affinities to this pocket and predicted success in blocking the activity of K-Ras^{G12D}. We selected the drug candidate which has the best score for further analysis. In order to determine that if it selectively binds to K-Ras^{G12D}, we docked this molecule to K-Ras^{WT}. Finally, we modified the selected molecule to allow new ligand-protein interactions. We anticipate that our novel approach will provide new insights to understanding mutant K-Ras dynamics and identify therapy candidates for the treatment of mutant K-Ras driven cancers.

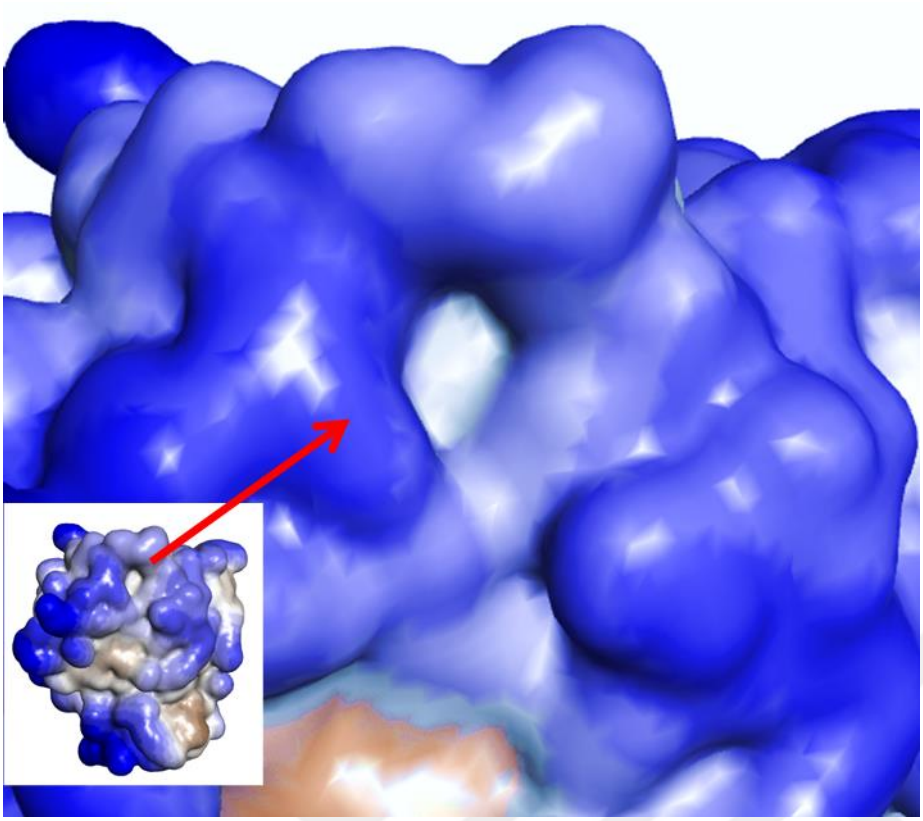


Figure 7.1 A novel drug target pocket in active K-Ras^{G12D}

7.2 METHODS

Molecular docking

From known drugs libraries, we found out those that bind to the novel pocket by following the steps below:

- We selected drugs from ZINC database - experimental molecules catalog (SHARDS) - to identify those that show highest binding affinity to the identified drug target pocket of K-Ras^{G12D}.
- We docked the selected drugs into the identified drug binding pocket by using Molegro molecular docking software (MVD 6.0)[230]. In MVD 6.0.software, MolDock Optimizer search algorithm was chosen. Number of runs was set as 50. Results of the virtual screening run were ranked by using the rerank score. The rerank

score in MVD provides an estimate of the strength of interaction between the protein and the ligand. Affinity is the strength of binding of a single molecule to its ligand and the lowest score shows highest binding affinity.

- We rank-ordered these drugs based on their binding affinities since the best docking score means strongest protein- ligand interaction. We chose the drugs with the top 30 scores. We then focused only on these 30 drugs that show high binding affinity.
- We visualized these drugs by MVD to determine the common poses of the drugs. Then we identified the best scored ligand among the most common poses as drug candidate for further analysis.

Molecular Modification

We added new fragments to the identified drug candidate to create new interactions between the drug molecule and K-Ras^{G12D} by using DS software.

7.3 RESULTS

Drug candidates that show highest binding affinity to K-Ras^{G12D}

Table 7.1 shows the docking scores of the molecules with the top 30 rerank scores (Rerank scoring system of MVD quantifies the strength of protein-ligand interaction). We determined the common poses among these 30 molecules by visualizing them with MVD. The first two molecules in Table 7.1 showed quite different binding modes compared to the binding modes of other molecules. Thus, we excluded them in the next steps. We identified that the best scored ligand among the cluster of most common poses is ZINC82876140. Figure 7.1 shows the binding mode of the molecule ZINC82876140 which has the highest binding affinity.

Table 7.1 Docking scores of the drugs with the top 30 scores

Rank	Ligand	Docking score
1	ZINC57218755	-67.1863
2	ZINC61844075	-67.0736
3	ZINC82876140	-66.7022
4	ZINC72206331	-66.5293
5	ZINC82873889	-66.5194
6	ZINC36876291	-66.3662
7	ZINC82876113	-65.9373
8	ZINC77266257	-65.8855
9	ZINC19404492	-65.6665
10	ZINC79436139	-65.6045
11	ZINC82870781	-65.4727
12	ZINC82873884	-65.3918
13	ZINC19263862	-65.3676
14	ZINC71500171	-64.9284
15	ZINC77274969	-64.857

16	ZINC11627993	-64.7142
17	ZINC41062015	-64.5531
18	ZINC50283449	-64.5405
19	ZINC36870965	-64.2502
20	ZINC52595550	-64.1224
21	ZINC50283874	-63.9558
22	ZINC83341005	-63.5133
23	ZINC62779868	-63.467
24	ZINC62780028	-63.4094
25	ZINC41062030	-63.3795
26	ZINC16943550	-62.5627
27	ZINC50284402	-61.495
28	ZINC37295030	-60.3184
29	ZINC82635551	-55.2547
30	ZINC82635442	-55.1782

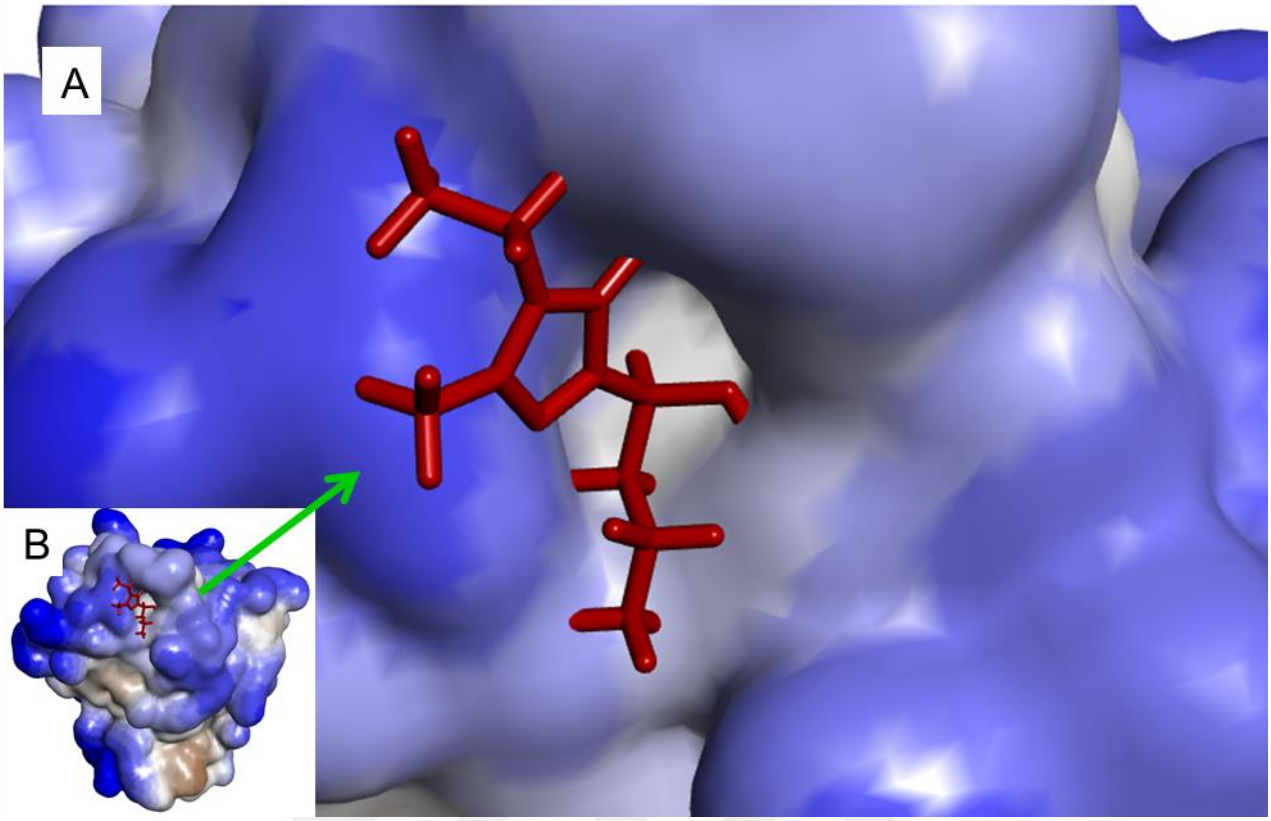


Figure 7.2 The molecule (red) which has the best score is bound to the novel drug target pocket of K-Ras^{G12D}. (A) Surface representation of the novel pocket and the bound ligand. (B) Zoom-out display of K-Ras^{G12D}- ligand complex.

Binding selectivity of the best scored ligand for K-Ras^{G12D}

We tested the selectivity of this ligand for binding to K-Ras^{G12D}. For this purpose we docked ZINC82876140 to K-Ras^{WT}. It gave a very bad binding score in K-Ras^{WT} complex (-35.7233) compared to K-Ras^{G12D} complex (-66.7022) that indicated the very low binding affinity of ZINC82876140 to K-Ras^{WT} . Consequently, results for the ligand- K-Ras^{WT} docking show the selective binding affinity of ZINC82876140 to K-Ras^{G12D}.

Modification of the best scored ligand

We added fragments to ZINC82876140 to allow additional protein-ligand interactions. Figure XA and B display the interactions within the K-Ras^{G12D}- ZINC82876140 complex and K-Ras^{G12D}- ZINC82876140 (modified) complex, respectively. New fragments added to the

original ligand cause new interactions with the protein. Although we were able to obtain new interactions by adding fragments, this can be a start point for advance drug design efforts.

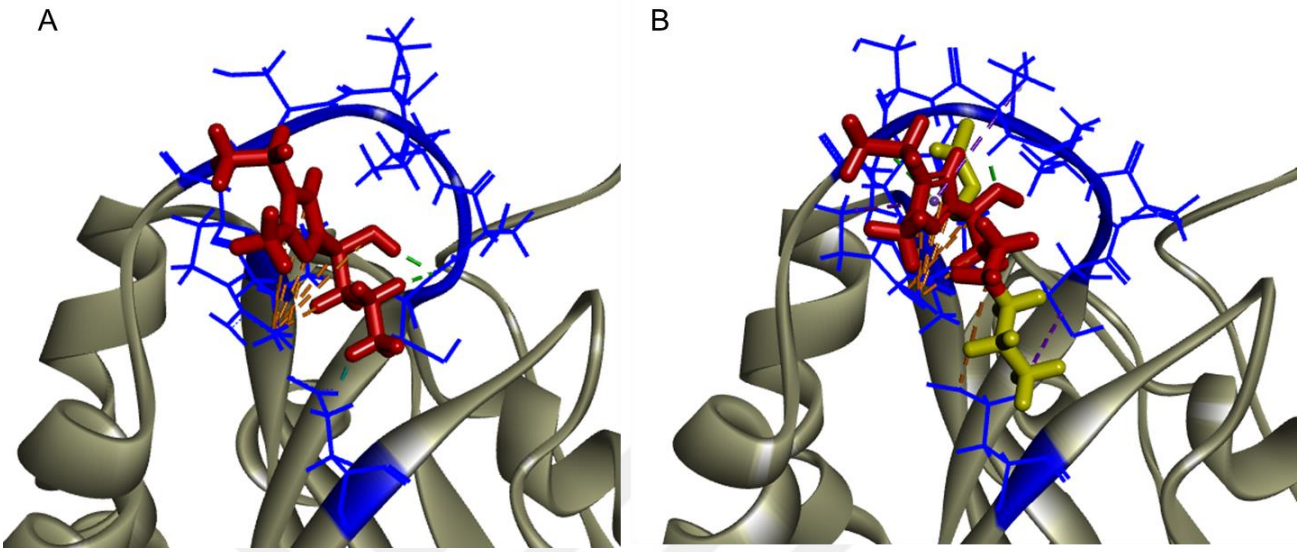


Figure 7.3 Protein(K-Ras^{G12D}) -ligand (ZINC82876140) interactions. Interacting protein atoms are blue and ligands are red. H-bonds are green, salt bridges are cyan, hydrophobic interactions are purple and the other electrostatic interactions are orange. and protein-ligand interactions are green. (A) K-Ras^{G12D}- ZINC82876140 complex. (B) K-Ras^{G12D}- ZINC82876140 (modified) complex. Modifications are indicated by yellow.

7.4 OTHER APPLICATIONS OF MOLECULAR DOCKING

7.4.1 Fight against Crimean Congo Hemorrhagic Fever Virus: Do we have any alternative drug?

7.4.1.1 Introduction

Crimean Congo Hemorrhagic Fever (CCHF) is the most important tick-borne viral disease of humans, causing increasing numbers of fatal outbreaks or sporadic cases across wide geographic area [231, 232]. A sudden onset of high fever and hemorrhagic manifestations are characteristics of the disease. The causative agent of the illness is Crimean Congo Hemorrhagic Fever Virus (CCHFV), a nairovirus in the family Bunyaviridae. World Health

Organization classifies CCHFV as biosafety level 4 (BSL4) pathogen. It is considered as a dangerous public health threat and a potential bio-terrorism agent in the wider world[233, 234]. In contrast with severity of the virus, currently no FDA-approved (The Food and Drug Administration) specific antiviral therapy is available. The only treatment for this neglected illness is postexposure administration of ribavirin, and the efficacy of this prophylaxis is in doubt[235]. Moreover, the requirement to study with it under BSL4conditions hinders essential experiments. Therefore, discovery of effective drugs for treatment of CCHF disease is now a priority for both public health and biodefense agencies.

Considering drug development stages, drug repositioning is an advantageous strategy to discover new and innovative indications of existing drug compounds for infectious and neglected diseases.[236, 237] Because it bypasses many time consuming steps, drug repositioning technique accelerates antiviral drug development process. A structural or functional protein of a virus can be a target for repositioning drugs.

Viral capsid proteins of encapsidated viruses are emerging as interesting targets for the development of new potent antiviral agents.(8) Capsid proteins tightly encloses viral genetic material to constitute ribonucleoprotein (RNP) complexes. Only in RNP form viral genome is replicated, transcribed and packaged into new progeny particles. Therefore, capsid proteins accepted as new potential targets of therapeutic agents against to HIV, coronaviruses[238] and other enveloped viruses[239]; influenzaviruses[240] and other negative-sense single-stranded RNA (ss (-)RNA) viruses[241]. Infectivity of bunyaviruses also depends on RNP formation[242]. Their three-segmented genome consists of small (S),medium (M), and large (L) RNA segments, which encode a viral nucleocapsid protein, glycoprotein precursor, and polymerase proteins respectively[243]. CCHFV nucleocapsid proteins are synthesized as monomers then they establish RNPs by oligomerization and RNA binding. It is recently

revealed that CCHFV nucleocapsid protein displays a distinct metal-dependent DNA-specific endonuclease activity in monomeric form[243].

The aim of this study is to discover new antiviral drugs that substitute ribavirin in CCHFV treatment. Drug repositioning strategy was followed in design of the research. Due to this strategy, molecular docking techniques were used for discovering possible drug candidates. This study is unique for its computational approach to CCHF treatment within the scope of drug repositioning manner. Resultant compounds will open new horizons on viral therapeutic development. Outputs may be used in therapies of all fatal viral hemorrhagic fevers.

7.4.1.2 Methods

Drug repositioning

Drug repositioning is simply identifying and developing new uses for existing drugs[244]. Therefore, this method skips time consuming steps of traditional drug discovery process such as Lipinski's rule of five, preclinical and clinical trials. Considering that any antiviral drug or even any FDA approved drug can be candidate for CCHF therapy, drug repositioning method was followed in the current study to accelerate the discovery of new inhibitor drugs for CCHFV. For identification of the possible CCHFV inhibitors from existing drugs that can bind to the viral capsid protein molecular docking programs were used in drug repositioning manner. Figure 7.4 summarizes the design of the study and following paragraphs explain the details of each step of the study.

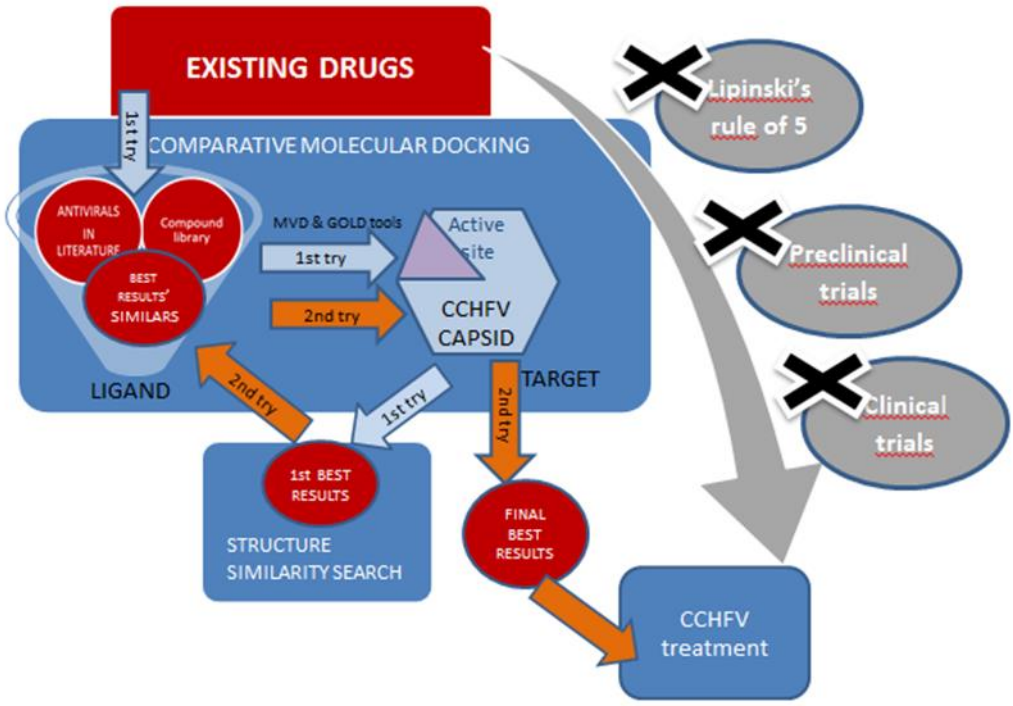


Figure 7.4 Scheme of the study design

Molecular docking

CCHFV capsid protein was selected as the target. Crystal structure of nucleocapsid protein was obtained from PDB (ID: 4AKL)[245]. Crystal structure of nucleocapsid protein was modified with DS 4.0 by saving only monomer form after removing the water molecules. MVD was used for virtual screening of protein. Electrostatic surface charge distribution was scanned. A large positively charged pocket on the protein was inspected. Constituent residues of this pocket were identified. In addition to identified pocket residues, the active site residues of nucleocapsid, which are responsible for protein function, were identified by literature-based search. CCHFV nucleocapsid protein distinct endonuclease activity was discovered by Guo et al[243]. Their mutational analyses revealed that R384, E387, K411, H453 and Q457 residues participated in a featured domain that served as viral endonuclease. Consequently, combination of silico inspections and literature data revealed that the largest pocket of the

protein includes the endonuclease activity site residues. Therefore, this endonuclease activity site was defined as target for molecular docking.

Drug molecules for molecular docking part were collected from ZINC Drug Databese (Zdd) library and Indofine Natural Products (INP) catalog. Zdd includes all drugs that have been approved for use in human and are commercially available as pure compounds. Along with Zdd, INP was chosen for its new promising natural compounds. In addition to these two libraries, antiviral compounds in the literature were collected by using the following steps: (i) It was considered that the antiviral drugs which disturb the other RNA viruses' lifecycles may arrest CCHFV's action through the capsid protein, since the functional mechanism of the target nucleocapsid protein is unclear. (ii) Therapeutic molecule names of well- known RNA viruses' infections (HIV, HCV, influenza, SARS) and other viral hemorrhagic fevers (Lassa, Ebola, Dengue,Rift Valley) was obtained from abstracts of published articles by PUBMED advance search tool[246]. (iii) Moreover, all approved and investigational antiviral drugs in Drug Bank database were collected[247]. Three library sets -Zdd, INP and the selected antivirals- were compiled as Mol2 file for molecular docking part.

In molecular docking part, two different molecular docking softwares were used for virtual screening and identifying the correct binding mode of molecules in a comparative manner. The first software was MVD 6.0. In this program, MolDock Optimiser search algorithm was chosen. Number of runs was set as 20. Results of the virtual screening run were ranked by using the rerank score. The rerank score in MVD provides an estimate of the strenght of interaction between proteins and ligands. The lowest score shows highest binding affinity. The second software was GOLD. In this program, ChemPLP (Piecewise Linear Potential) scoring function was selected. Other parameters were set as default. Docking results were

arranged by their PLP Fitness Scores. In GOLD software, the higher fitness score means the better docked pose.

In analysis of docking results, the results from each docking tool were ranked by their scores where the best docking score means strongest protein- ligand interaction. Compounds that took placed near the top of these two groups compared with each other.

As an additional step, structure-based discovery of candidates and re-docking were also performed. It was observed that some best scored candidates shared similar structure. After identification of the ligands that had higher binding affinity, those ligands' SMILE forms uploaded ZINC structure-based search tool. Compounds retrieved from search results were docked in the same way of previous ligands. Re-docking results were analyzed to find the best-scored molecule among the similar structures.

7.4.1.3 Results

Darunavir, saquinavir, and sofosbuvir are best-scored antivirals (Table 7.2) and their molecular structures are shown in Figure 7.5. Darunavir and saquinavir are HIV-1 protease inhibitors that block the proteolytic processing of precursor polypeptides Gag and Gag-Pol. (Cleavage of Gag and Gag-Pol is required for the formation of an infectious virion). Darunavir and saquinavir have been formally approved for clinical use in the treatment of AIDS. On the other hand, sofosbuvir is a hepatitis C virus nucleotide analog NS5B polymerase inhibitor. It has been approved for the treatment of chronic hepatitis C infection

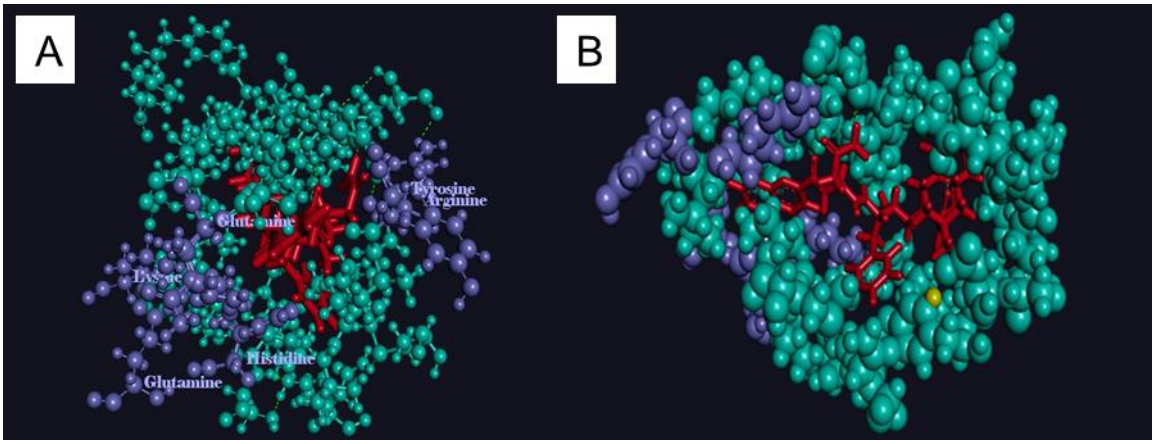


Figure 7.5 Active site residues of CCHFV nucleocapsid protein and bound saquinavir. (A) Endonuclease activity site residues (R384, E387, K411, H453,Q457) are represented by violet ball and sticks. Saquinavir is represented by red sticks. (B) Saquinavir (red) is bound to drug target pocket. The residues make up the binding site pocket are represented by cyan ball and sticks.

Table 7.2 Best scored antiviral candidates. ZINC numbers, names, rerank scores.

ZINC NUMBER	MOLECULE NAME	RERANK SCORE
39552192	darunavir	-129.409
2698553	saquinavir	-115.266
64526913	sofosbuvir	-105.553

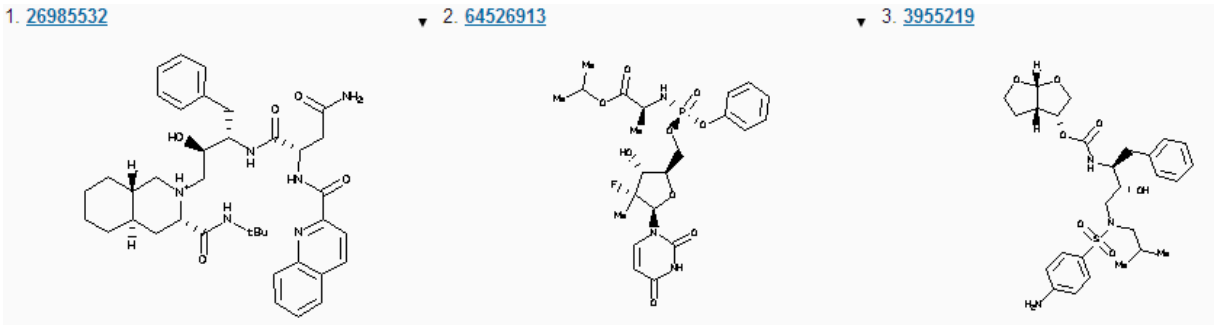


Figure 7.6 Structures of the best scored antivirals.

Among the FDA approved drugs cephalosporin antibiotics got highest rerank scores. Cephalosporins are bactericidal β-lactam antibiotics. They inhibit enzymes in the cell wall of susceptible bacteria, disrupting cell synthesis. We, then, collected all the cephalosporin in ZINC database and used for re-docking. Docking results showed that cefditroen pivoxil,cefteram pivoxil, cefpodoxime proxetil, cefpirom sulfate and ceftiofur (used in veterinary medicine) have high rerank scores similar to cefpiramide and cefonicid. We display list of the cephalosporin molecules and their scores in Table 7.3.

Table 7.3 Best scored cephalosporin antibiotics

ZINC NUMBER	MOLECULE NAME	RERANK SCORE
3916973	Cefditoren Pivoxil	-144.156
3871961	Cefteram Pivoxil	-139.778
15449382	Cefpodoxime Proxetil	-129.883
21985848	Cefpirome Sulfate	-126.131
4215267	Ceftiofur	-125.724
4215275	Cefuzonam Sodium	-125.053
26892366	Cefmenoxime Hydrochloride	-125.052
3830428	Cefonicid Sodium	-107.253

Itraconazole - a triazole antifungal agent for fungal infections- showed highest binding affinity after cephalosporins among the FDA approved drugs. The other azol derivatives posaconazole and ketoconazole also show high binding affinity. We also obtained the

molecules whose structures similar to itraconazole from ZINC database and docked them into target CCHFV nucleocapsid protein. Among the similar structures ravuconazole –a phase I/II antifungal- also proved high binding affinity. The list of the azole derivative drugs and their structures are shown in Figure 7.7 and Table 7.4.

Table 7.4 Best scored azol derivatives. ZINC numbers, names, rerank scores.

ZINC NUMBER	MOLECULE NAME	RERANK SCORE
4097343	Itraconazole	-123.313
3938482	Posaconazole	-114.915
643138	Ketoconazole	-106.072
600547	Ravuconazole	-105.065

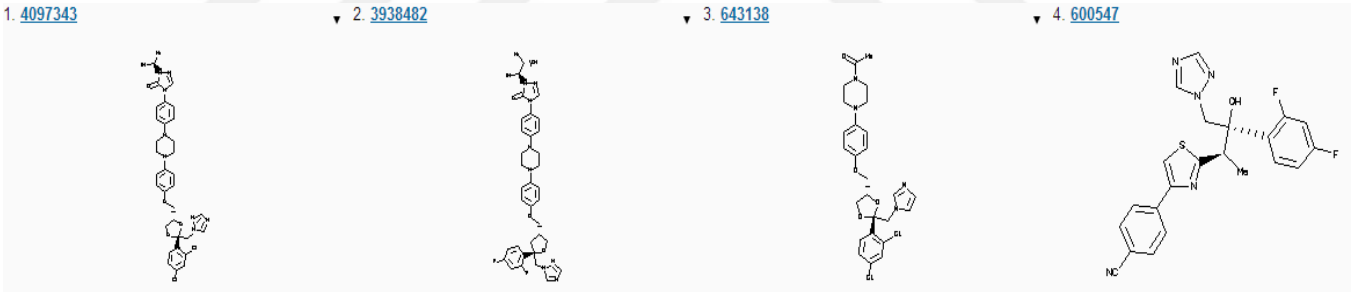


Figure 7.7 Structures of the best scored azol derivatives.

In conclusion this molecular docking study gave three promising results that will be used for treatment of CCHFV, a dangerous but neglected disease. Because drug repositioning technique was used, there is no need to evaluate druglikness of the resultant molecules. Although computational results were great, real effects of them on the CCHFV can only be seen by in vitro and in vivo researches.

7.4.1.4 Discussion

Nucleocapsid protein of CCHFV mechanism of function has not been proved yet. The major challenge for this study was defining target residues that have certain important roles in nucleocapsid function. Information from the literature was not enough to support in silico analyses of the nucleocapsid protein. It possesses a racket- shaped overall structure with two major parts: a “head” and a “stalk” domain. Carter et al. defined possible RNA binding site of head domain[245]. However, the study of Guo et al. demonstrated the weakness of CCHFV nucleocapsid protein RNA binding affinity. They revealed endonuclease activity domain (R384, E387, K411, H453,Q457). Carter et al. also worked on this domain. They presented this domain as possible dimer interface and considered that it could mediate the RNP formation. In silico analyses of the N protein in this study proved that endonuclease activity domain constituted most significant cleft of the whole protein. Because visual data match up with literature- based data only for endonuclease activity pocket, the pocket selected as the target for compound binding.

Itraconazole has very high score that makes it an attractive drug candidate. Other azole derivatives also have great binding affinity. Especially itraconazole perfectly matches with the binding cavity that is presented obviously in virtual screening results(Fig.7.8A). Recent studies showed that azole derivatives are promising HIV non-nucleoside reverse transcriptase inhibitors[248, 249]. These compounds bind to the reverse transcriptase and stop the viral replication. After modifications azoles can mimics nucleic acids[250]. Therefore, modified azoles may be beneficial for CCHF treatment.

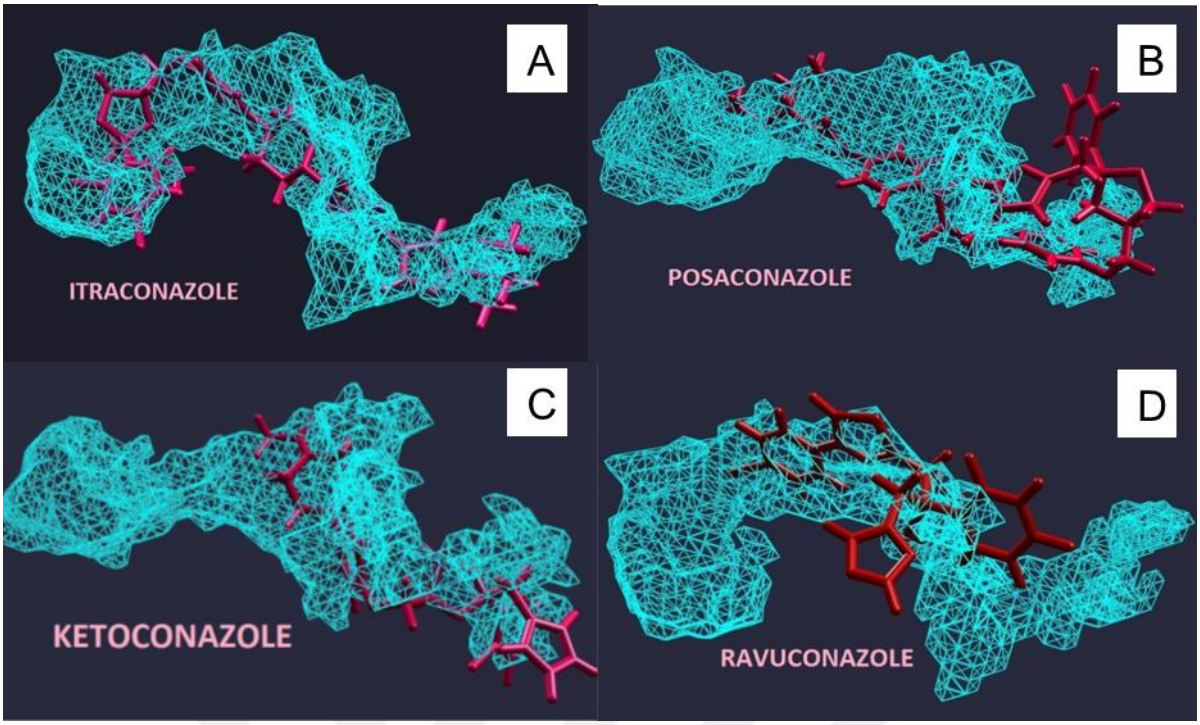


Figure 7.8 Azole derivatives (red) are in the binding cavity (cyan) of CCHFV nucleocapsid protein. (A) Binding pose of itraconazole into the binding cavity. (B) Posaconazole in the binding cavity. (C) Ketoconazole in the binding cavity. (D) Ravuconazole in the binding cavity.

7.4.2 Antiviral activity of an N-allyl acridone against dengue virus

7.4.2.1 Introduction

Dengue virus (DENV), a member of the family Flaviviridae, is at present the most widespread causative agent of a human viral disease transmitted by mosquitoes. Despite the increasing incidence of this pathogen, there are no antiviral drugs or vaccines currently available for treatment or prevention. In a previous screening assay, we identified a group of N-allyl acridones as effective virus inhibitors. Here, the antiviral activity and mode of action targeted to viral RNA replication of one of the most active DENV-2 inhibitors was further characterized.

The compound 10-allyl-7-chloro-9(10H)-acridone, designated 3b, was active to inhibit the in vitro infection of Vero cells with the four DENV serotypes, with effective concentration 50% (EC50) values in the range 12.5-27.1 μM, as determined by virus yield inhibition assays. The compound was also effective in human HeLa cells. No cytotoxicity was detected at 3b concentrations up to 1000 μM. Mechanistic studies demonstrated that virus entry into the host cell was not affected, whereas viral RNA synthesis was strongly inhibited, as quantified by real time RT-PCR. The addition of exogenous guanosine together with 3b rescued only partially the infectivity of DENV-2.

The acridone derivative 3b selectively inhibits the infection of Vero cells with the four DENV serotypes without a direct interaction with the host cell or the virion but interfering specifically with the intracellular virus multiplication. The mode of antiviral action for this acridone apparently involves the cellular enzyme inosine-monophospahe dehydrogenase together with another still unidentified target related to DENV RNA synthesis.

In this study, we performed an extensive docking modeling between the compound **3b** and the postulated target protein IMPDH to explore if an interaction is possible at the molecular level.

7.4.2.2 Methods

Molecular docking

Crystal structures of the proteins were obtained from PDB. In DS program, proteins were modified by removing hetatoms from the structures. Ligand structure was checked with MarvinSketch and its geometry optimised by DS. PDB file formats of the proteins and mol2 or sd file formats of the ligands were uploaded to all currently used softwares. The active sites of the enzymes were determined by combining literature data with virtual screening observations. DS was used as virtual screening tool.

MVD and GOLD molecular docking softwares were used together in a comparative manner. The searching space was specified that includes the active site properly. All explicit hydrogens were added into the structures. MolDock Score were used in MVD docking setup with the following settings: for the local search, MolDock Simplex Evolution algorithm was applied with a maximum of 2000 iterations per search. Fifty independent docking runs were carried out. Positions of the hydrogens for any hydrogen donors (both in 3b and in the enzyme) were optimised. The Ignore Similar Poses option was used to avoid reporting to similar poses. All poses returned from the runs were clustered according to the root-mean square deviation (RMSD) criterion and similar poses were removed (keeping the best-scoring one). Different poses of the compounds were ranked by their rerank scores and the best docked conformation was selected. More negative score indicates higher binding affinity. In Gold docking setup, Gold Genetic Algorithm was used with the slow option. ChemPLP method was chosen as the scoring function and ChemScore was the rescoring function. Search options of Genetic Algorithm were automatically configured by the programme. Resulting poses were ranked by their PLR fitness functions and Chemscore ΔG values. In GOLD higher fitness score means better docked pose.

7.4.2.3 Results and Discussion

In order to explore if an interaction is possible at the molecular level, an extensive docking modeling was performed between the compound 3b and the postulated target protein IMPDH. The three-dimensional structure of human IMPHD was obtained from Protein Data Bank (Accession Number: 1NF7). Molegro and GOLD molecular docking softwares were used together for effective pose prediction and virtual screening. For determining the best pose, Molegro's rerank scoring function and GOLD's ChemPLP fitness score were used. Opposite of the rerank scoring, higher ChemPLP fitness score means better docked pose. Comparative

docking results showed that the compound 3b occupied the active site of IMPHD with a rerank score of -80.6446 and ChemPLP score of 51.0029, Chemscore ΔG of -25.0379. Two hydrogen bond interactions were identified with the backbone amino acid residues SER327 and GLY328. Moreover, 26 hydrophobic interactions with 11 residues including active site -CYS331- were observed (Figure 7.9A).

Next we analyzed how 3b binds to native ligand´s interaction sites within the enzyme by using Molegro. Nicotinamide adenine dinucleotide (NAD) and inosine monophosphate (IMP) are chemically active native ligands of IMPDH which is an enzyme that catalyzes the NAD-dependent oxidation of IMP[251, 252]. Redocking was applied to NAD and IMP for fair and accurate comparison. Complex form of IMPHD bound to its native ligands is stored in Protein Data Bank (Accession Number: 1NFB). After determining the residues which interact with NAD and IMP separately, ligands were removed from the 1NFB structure. Then NAD and IMP were docked back into the protein. 3b-IMPHD docking was done by targeting the interaction regions of natural substrates.

It is known from the crystallographic studies that NAD binding site includes 11 structural amino acids[253] (Figure 7.9B). Redocking of NAD gave -117.426 rerank score, conversely, docking of 3b to this NAD binding site gave -67.4364 rerank score. Relatively low binding score of 3b revealed that this site does not fit for it (Figure 7.9C). As for IMP, the substrate binds to the enzyme through CYS331 and the other 11 amino acids (Figure 7.9D). IMP's docking back into IMPHD gave score of -112.119. When the compound 3b was docked in the IMP activity region, it made a hydrogen bond with the side chain residue SER68 by resulting rerank score of -82.2138 (Figure 7.9E). Then, the results implied that 3b showed higher binding affinity to IMP activity site than NAD activity site.

In addition to 3b-NAD/IMP activity site interaction analysis, we investigated 3b interactions with the inhibitory regions of two IMPDH antagonists: RIB monophosphate and C2-Mycophenolic Adenine Dinucleotide (MYD). RIB monophosphate and MYD show their antagonistic activity by blocking access to active site of IMPHD. Following the previous part's method, the position of 3b when it interacts with each binding region of those two inhibitors was figured out. The three-dimensional structure of IMPHD- RIB monophosphate, MYD complex was obtained from Protein Data Bank (Accession number: 1NF7). As presented in the previous crystallographic study, RIB monophosphate and MYD bind to their inhibitory pockets by making hydrogen bonds and hydrophobic interactions[254] (Figure 7.9F,H). At first, they were docked back into their pockets. Best-fitting poses' rerank scores were -116.695 for RIB monophosphate and -130.851for MYD. Later, 3b was docked into each drug's inhibitory pocket. 3b presented high affinity to RIB monophosphate binding region by resulting rerank score of 82.2913. (Figure 7.9G). In contrast, it showed low affinity to MYD inhibitory cavity (Figure 7.9I).

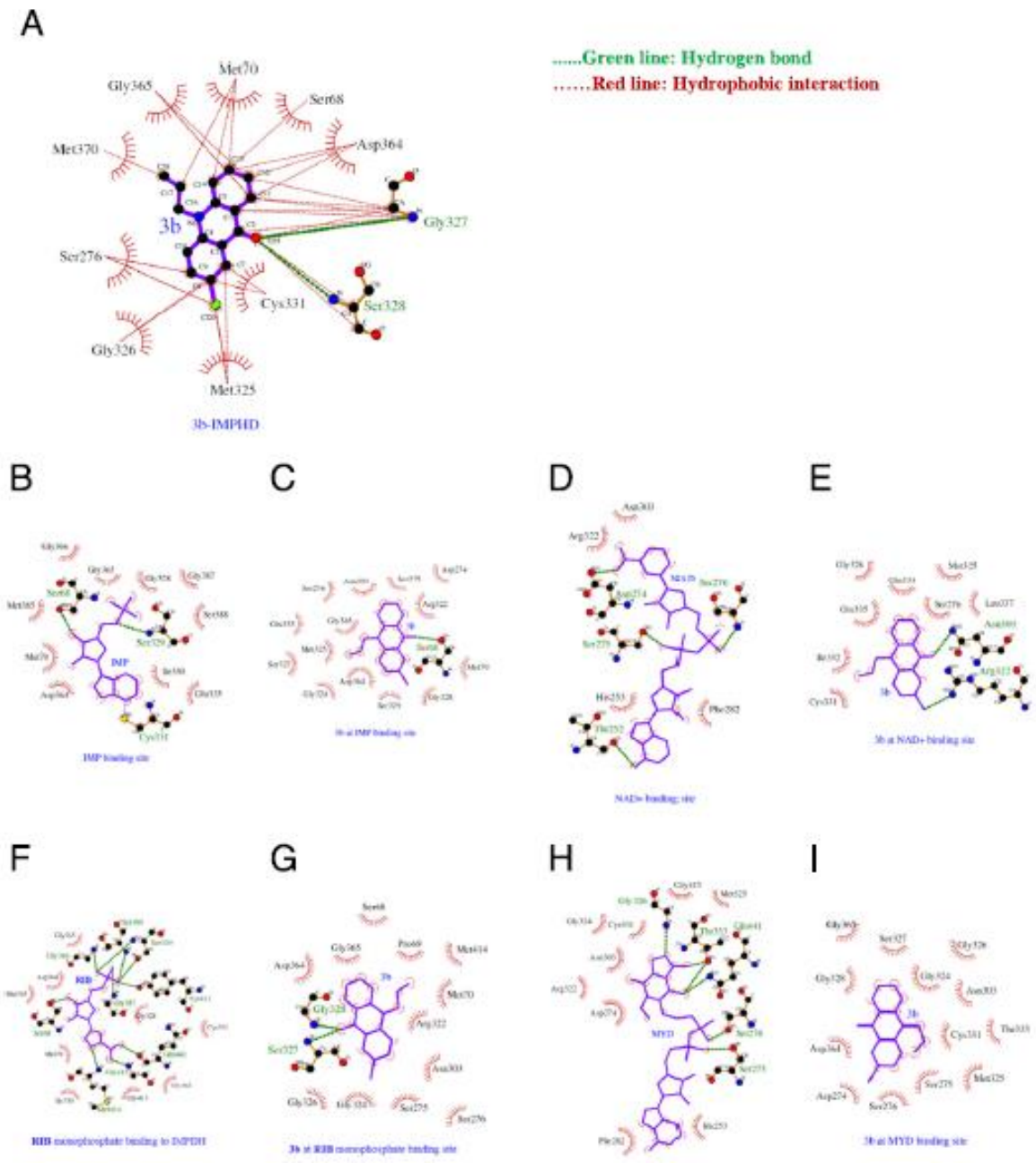


Figure 7.9 Docking modeling between the compound 3b and the postulated target protein IMPDH. (A) 3b binds to the active site of the IMPHD enzyme through 2 hydrogen bonds and hydrophobic interaction. (B) IMP interaction cavity consists of SER68, MET70, GLY328, SER329, ILE330, CYS331, GLU335, ASP364, GLY365, GLY366, GLY387, SER388. (C) 3b made a hydrogen bond with SER68 and hydrophobic interactions with SER68, MET70, ASP274, SER275, SER276, ASN303, ARG322, GLY324, MET325, SER327, GLY328, SER329, GLU335, ASP364, GLY365. (D) NAD interaction pocket residues are THR45, THR252, HIS253, ASP274, SER275, SER276, PHE282, ASN303, ARG322, HIS466, GLN469. (E) Two hydrogen bonds existed between 3b and ASN303, ARG322. 3b also made hydrophobic interactions with SER276, ASN303, ARG322, MET325,

GLY326, CYS331, ILE332, GLN334, LEU337. (F) RIB binding site includes SER68, MET70, ARG322, GLY328, SER329, ILE330, CYS331, ASP364, GLY365, GLY366, GLY387, SER388, TYR411, GLY413, MET414, GLY415, GLN441. (G) 3b made hydrogen bonds with SER327, GLY328 and hydrophobic interactions with PRO69, MET70, SER275, ASN303, ARG322, GLY324, GLY326, SER327, GLY328, ASP364, GLY365.(H) HIS253, ASP274, SER275, SER276, PHE282, ASN303, ARG322, GLY324, MET325, GLY326, CYS331, THR333 involves in MYD activity cavity. (I) 3b made hydrophobic interactions with ASP274, SER275, SER276, ASN303, GLY324, MET325, GLY326, SER327, GLY328, ASP364.

The molecular docking studies performed provide useful information about 3b-IMPHD interaction. The binding orientation of 3b to human IMPHD was found in order to predict the affinity and activity of the compound. As the docking results presented, 3b showed the best-fit orientation when it was docked into active site, IMP and RIB monophosphate binding site of IMPHD (Figure 7.10).

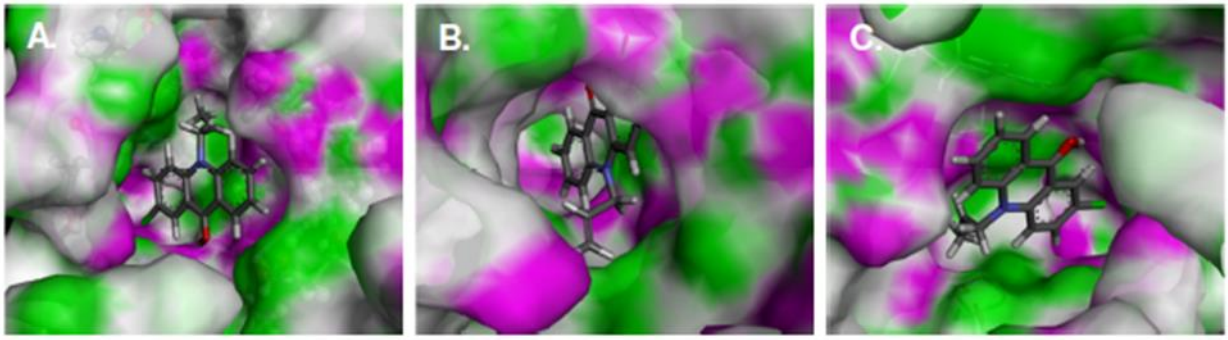


Figure 7.10 Prediction affinity about 3b-IMPHD interaction (A) Binding of 3b to IMPDH active site. (B) Binding of 3b to RIB monophosphate inhibitory site. (C) Binding of 3b to IMP activity site.

7.4.3 The interplay between viperin antiviral activity, lipid droplets and Junín mammarenavirus multiplication

7.4.3.1 Introduction

Junín arenavirus infections are associated with high levels of interferons in both severe and fatal case cases. Upon Junín virus (JUNV) infection a cell signaling cascade initiates, that

ultimately attempts to limit viral replication and prevent infection progression through the expression of host antiviral proteins. The interferon stimulated gene (ISG) viperin has drawn our attention as it has been highlighted as an important antiviral protein against several viral infections. The studies of the mechanistic actions of viperin have described important functional domains relating its antiviral and immune-modulating actions through cellular lipid structures. In line with this, through silencing and overexpression approaches here we identified viperin as an antiviral ISG against JUNV. In addition, we found that lipid droplets structures are modulated during infection suggesting its relevance for proper virus multiplication. The confocal microscopy images also revealed possible viperin-JUNV protein interactions that might be participating in this antiviral pathway at lipid droplet level.

In this study, we aimed to gain more insight on the viperin anti-JUNV mechanism. For this purpose, we investigated nucleocapsid (N) protein of JUNV and viperin protein interaction by using protein-protein docking tools in a comparative manner. Our results will help to understand the factors mediating innate immunity in arenavirus infection and may lead to the development of pharmacological agents that can boost their effectiveness thereby leading to new treatments for this viral disease.

7.4.3.2 Methods

Protein-protein docking

The structure of JUNV N protein was retrieved from Protein Data Bank (PDB, ID: 4K7E)[255]. It is the crystal structure of the C-terminal domain (residues 341-564) with 2.2 Å resolution. Missing residues were added by Chimera software [256] . The crystal structure with 1.7 Å resolution of viperin protein was also retrieved from PDB (ID:5VSM) [257]. Viperin was docked into N protein utilizing two different protein-protein docking softwares,

ClusPro [258] and ZDock [259], in a comparative manner. These programs produced different complex models and the most common viperin–N interaction model was selected.

7.4.3.3 Results and Discussion

We investigated N and viperin protein interaction by using protein-protein docking tools in a comparative manner. In the docking part, we obtained N-viperin protein complex structure. The resulting model showed that C-terminal domain of JUNV N protein interact with viperin (Figure 7.11). In particular, the first group of predicted interactions is between α4 and β5 residues of N and α4 residues of viperin; the second group of them is between α5 and α6 residues of N and α3 residues of viperin.

According to predicted models from docking studies, the C-terminal of N protein would be involved in the interaction with α3 and α4 residues on viperin. Previous studies determined that C-terminal domain of JUNV N (residues 341-564) might possess other important functions besides the exonuclease activity. Moreover, it has been indicated that amino acid residues between positions 394 and 502 (a part of N C-terminal) could be involved in zinc binding [260]. Viperin α3 belongs to the conserved central domain characteristic for the S-adenosylmethionine (SAM) enzyme family [261]. Moreover, this in silico interaction was confirmed with co-transfection experimental approach. The arenaviral N is the most abundantly expressed viral protein in infected cells. N forms a complex with the viral polymerase to transcribe viral mRNAs and synthesize viral genomic RNAs during viral replication [262]. Interestingly, it was observed a significant inhibition in RNA transcription in viperin transfected-JUNV infected cell cultures. Our results suggest that JUNV RTC might be recruited to these lipid structures and use them as platforms for viral transcription, where viperin might be interacting with N in order to counteract this essential function.

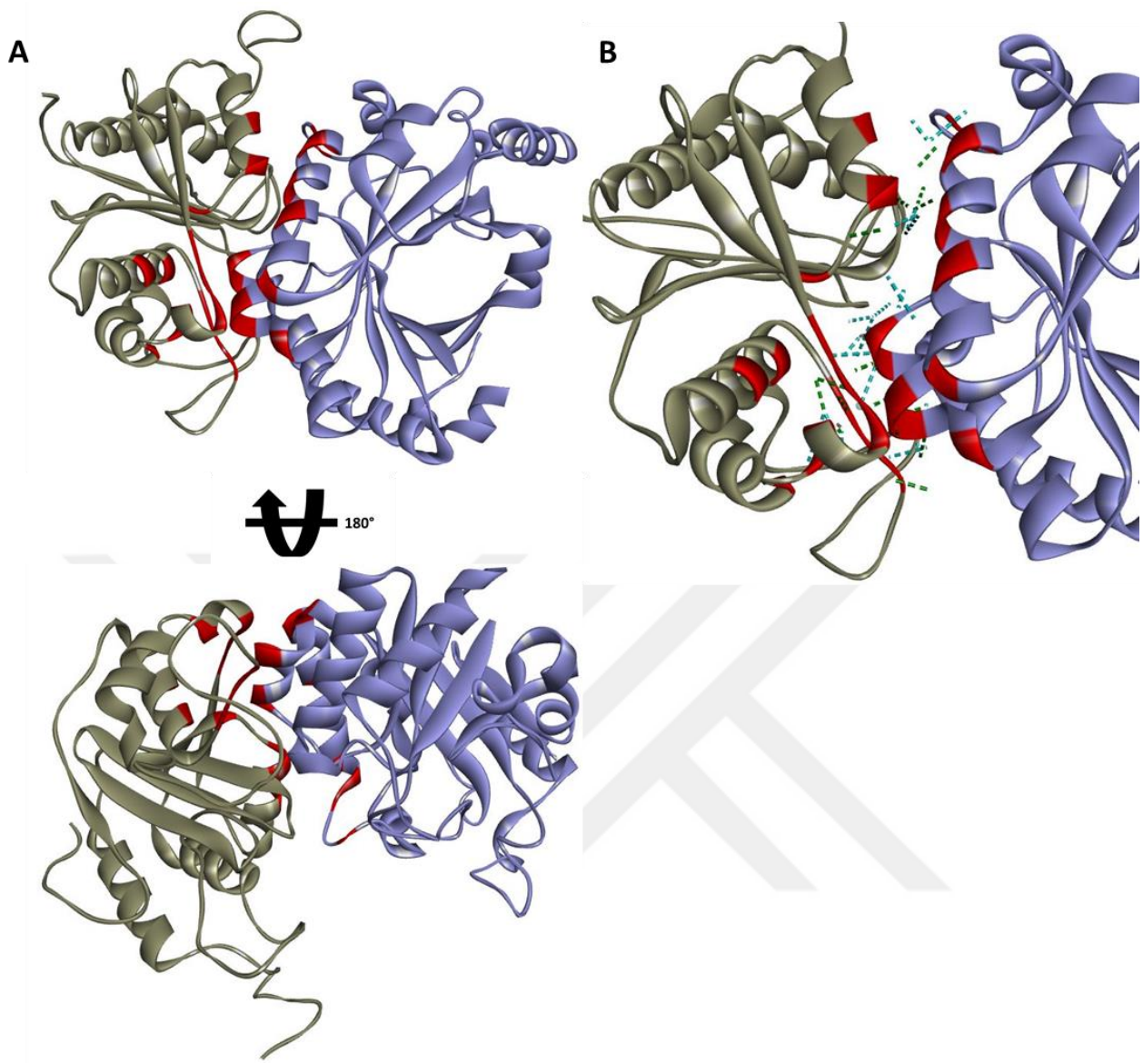


Figure 7.11 Interaction of viperin and JUNV N proteins (A) Theoretical analysis of VIP-JUNV interaction. Ribbon schematic binding complex of viperin (blue) and JUNV N C-terminal domain (grey). Residues involved in interaction site are shown as red ribbon. Reverse view of the complex showing other interacting residues. (B) Detailed projection of interacting residues of viperin-JUNV N complex. H-bonds are represented with green sticks and other electrostatic intaractions are represented by cyan sticks. Both interactions showed in (A) and (B) were predicted using ClusPro and ZDock docking softwares.

8 CONCLUSION

In this thesis, we used different MD simulation data analysis techniques to study the regulation mechanisms of K-Ras dynamics and to understand the effects of nucleotide binding and mutations on regulation of protein dynamics. Using MD simulations of K-Ras we compared the dynamic characteristics of active and inactive WT proteins (Chapter 3), WT and mutant proteins (Chapter 5) and active and inactive mutant proteins (Chapter 6). Developing a novel method that predicts causality between residue pairs of a protein we identified the causality relationships in regulation of K-Ras dynamics (Chapter 3 and 4). Lastly, combining the results from the Chapters 3-6, we identified a novel drug target pocket on K-Ras^{G12D}-GTP complex that is important for dynamic regulation of K-Ras. We screened a library of compounds against this novel drug target pocket by using docking techniques (Chapter 7). Additionally, we demonstrated the applications of docking techniques into three other drug design problems, too.

Oncogenic K-Ras is a high priority drug target in cancer treatment. To develop new direct inhibitors that selectively bind to mutant K-Ras conformations while sparing those of WT K-Ras, it is necessary to first understand the dynamic activity of the WT protein in detail.

To evaluate the nucleotide binding dependent changes in K-Ras stability, we used stiffness and RMSF calculations and proved that GTP binding rigidifies and hence stabilizes K-Ras motions. These results are in agreement with previous experimental and computational K-Ras studies[203, 263, 264]. Our calculations that use stiffness, RMSF and correlation graphs (Figure 1-2, S1) confirm that GTP-binding increases K-Ras stiffness and thereby decreases fluctuation amplitudes, leading to distinct correlation patterns. These striking changes in GTP-

bound-K-Ras dynamics enable its GTPase activity. Note that this nucleotide exchange is the first step in active to inactive transition[221, 265-271]. Overall, our results support the well-established allosteric nature of K-Ras activation[268], which has been suggested to play an important role in GTPase activity[272]. Although correlated fluctuations are necessary for allosteric information flow, their longer correlation decay times are also of crucial importance for complete allosteric transition. We calculated time-dependent autocorrelations of fluctuation vectors between residue pairs and discovered that correlations of K-Ras-GTP are stronger and persist for longer correlation times during simulations. Their persistency may allow complete allosteric information flow in K-Ras-GTP.

We broadened our analysis to quantify causality in allosteric regulation of K-Ras function. The most important results from our study are on causality. We applied a simple but powerful method that we defined as *CTC* into protein dynamics. To understand K-Ras dynamics, we investigated whether fluctuations of any residue caused fluctuations of another. Our results revealed the information flow in K-Ras switch mechanism and that SII fluctuations drive SI fluctuations. This prediction is an essential validation of our approach since the dominance of SII motions over SI motions was observed in previous experimental and computational studies[204, 205]. Surprisingly, in addition to the canonical switch mechanism, our algorithm also revealed causality relations in the novel switch mechanism that includes β2–β3 and α5, where β2–β3 motions drive α5. Moreover, fluctuations of α3-L7 drive fluctuations of SI and SII. Interestingly, previous studies reported that Ras effector binding site (SI) is allosterically stabilized by ligand binding into a novel pocket that includes L7[33, 208, 209]. Our results explain the allosteric effect of ligand binding on SI motions by showing the information flow from L7 to SI.

After characterized WT K-Ras dynamics in detail, we aimed to understand the mechanisms by which oncogenic mutations disturb WT K-Ras dynamics and thereby its function. Despite decades of research, the relationship between oncogenic K-Ras mutations and changes in K-Ras protein conformation and dynamics remain to be understood. We showed how the most recurrent K-Ras oncogenic mutation, G12D, leads to structural, conformational and dynamical changes that impair its function. We have used a new integrated MD simulation data analysis approach we have developed to quantify the changes in K-Ras. Our results show that D12 makes a salt bridge with K16 (P loop) which also makes a salt bridge with D57 (SII) after G12D mutation. These salt bridges cause conformational changes in the active site where SII moves away from the P-loop, SI (Q25-Y40) and α3. Furthermore, negative correlations occur between the fluctuations of these divergent parts: SII fluctuations become negatively correlated between the P-loop, SI (Q25-Y40) and α3 fluctuations. These negative correlations slow down the characteristic decay times of the residue fluctuations. The fluctuations of SII residue D57 show the highest negative correlations with other parts of the protein and also the longest characteristic decay times among the K-Ras^{G12D} residues. On the other hand, causality relations occur between the negatively correlated residue pairs and the residues with longer characteristic decay times act as the drivers in the driver-responder pairs. Overall, our study identifies regulatory sites on K-Ras^{G12D}, which can enhance our understanding of its dynamics and assist in the development of direct inhibitors.

We successfully investigated how G12D mutation, the most recurrent K-Ras mutation, affects the regulation of K-Ras dynamics that results in constitutively active K-Ras^{G12D}. However, for selective targeting of active K-Ras^{G12D}, we need to define the differences between active and inactive K-Ras^{G12D}. We found that the activation of K-Ras^{G12D} causes correlated motions of residue pairs by destabilizing their relative conformations. In active K-Ras^{G12D} dynamics,

residues 64 and 67 act as the cores of the correlated motions, however this correlation patterns disappear in the inactive form. Additionally, our strain calculations showed that SI and the P-loop, β3 become dilated after activation of K-Ras^{G12D}. On the other hand, residues within α2 and α3 become compressed.

All of these results guided in identification of drug target sites on K-Ras^{G12D}-GTP surface. Using detailed analysis of K-Ras^{G12D} dynamics and comparing them with the WT and inactive forms, we identified a novel target pocket on K-Ras^{G12D}-GTP surface that has not been shown in the previous experimental and computational studies. This novel pocket is shown in only K-Ras^{G12D}-GTP complex while it disappears in K-Ras^{WT}-GDP. Using molecular docking method, we screened a library of small molecules against this novel pocket.

We also demonstrated other successful applications of molecular docking. In the first application, we identified drug candidates that can bind into the capsid protein of CCHFV. In the second application, we demonstrated the possible interaction between a target protein and a drug whose inhibitory effects was shown experimentally. In the last application, we predicted the interaction between two proteins that bind each other in the experiments.

The regulatory mechanisms of K-Ras dynamics and the novel drug binding pocket that we present may help in drug discovery efforts to develop direct inhibitors of mutant K-Ras for treatment of cancer in the future. The computational tools we introduce are also easily applicable to the analysis of simulation data from different proteins to understand causality in their dynamic regulations, which can then similarly be utilized in drug discovery. From this perspective, our approach has the potential to set a novel paradigm for drug design by directing attention to changes in protein dynamics. We conclude by emphasizing that understanding the regulation of protein dynamics is crucial for having a complete in-depth picture of protein function; identifying the molecular mechanisms of mutations; and development of better drugs.

Note: Chapter 3[212] and Part 7.4.2[273] published in peer-reviewed journals. The studies in Chapters 4-7 will be submitted.

BIBLIOGRAPHY

1. A.G. Stephen, D. Esposito, R.K. Bagni, and F. McCormick, *Dragging ras back in the ring.* Cancer Cell, 2014. **25**(3): p. 272-81 DOI: 10.1016/j.ccr.2014.02.017.

2. S.A. Forbes, D. Beare, P. Gunasekaran, K. Leung, N. Bindal, H. Boutselakis, M. Ding, S. Bamford, C. Cole, S. Ward, C.Y. Kok, M. Jia, T. De, J.W. Teague, M.R. Stratton, U. McDermott, and P.J. Campbell, *COSMIC: exploring the world's knowledge of somatic mutations in human cancer.* Nucleic Acids Research, 2015. **43**(Database issue): p. D805-11 DOI: 10.1093/nar/gku1075.

3. I.A. Prior, P.D. Lewis, and C. Mattos, *A comprehensive survey of Ras mutations in cancer.* Cancer Res, 2012. **72**(10): p. 2457-67 DOI: 10.1158/0008-5472.CAN-11-2612.

4. R.J.C. Slebos, R.E. Kibbelaar, O. Dalesio, A. Kooistra, J. Stam, C.J.L.M. Meijer, S.S. Wagenaar, R.G.J.R.A. Vanderschueren, N. Vanzandwijk, W.J. Mooi, J.L. Bos, and S. Rodenhuis, *K-Ras Oncogene Activation as a Prognostic Marker in Adenocarcinoma of the Lung.* New England Journal of Medicine, 1990. **323**(9): p. 561-565 DOI: Doi 10.1056/Nejm199008303230902.

5. B. Vogelstein, E.R. Fearon, S.R. Hamilton, S.E. Kern, A.C. Preisinger, M. Leppert, Y. Nakamura, R. White, A.M. Smits, and J.L. Bos, *Genetic alterations during colorectal-tumor development.* N Engl J Med, 1988. **319**(9): p. 525-32 DOI: 10.1056/NEJM198809013190901.

6. B. Bournet, C. Buscail, F. Muscari, P. Cordelier, and L. Buscail, *Targeting KRAS for diagnosis, prognosis, and treatment of pancreatic cancer: Hopes and realities.* Eur J Cancer, 2016. **54**: p. 75-83 DOI: 10.1016/j.ejca.2015.11.012.

7. D.P. Ryan, T.S. Hong, and N. Bardeesy, *Pancreatic adenocarcinoma.* N Engl J Med, 2014. **371**(11): p. 1039-49 DOI: 10.1056/NEJMra1404198.

8. F. McCormick, *KRAS as a Therapeutic Target.* Clin Cancer Res, 2015. **21**(8): p. 1797-801 DOI: 10.1158/1078-0432.CCR-14-2662.

9. W. Pao, T.Y. Wang, G.J. Riely, V.A. Miller, Q.L. Pan, M. Ladanyi, M.F. Zakowski, R.T. Heelan, M.G. Kris, and H.E. Varmus, *KRAS mutations and primary resistance of*

lung adenocarcinomas to gefitinib or erlotinib. Plos Medicine, 2005. **2**(1): p. 57-61 DOI: ARTN e17

10.1371/journal.pmed.0020017.

10. R.G. Amado, M. Wolf, M. Peeters, E. Van Cutsem, S. Siena, D.J. Freeman, T. Juan, R. Sikorski, S. Suggs, R. Radinsky, S.D. Patterson, and D.D. Chang, *Wild-type KRAS is required for panitumumab efficacy in patients with metastatic colorectal cancer.* J Clin Oncol, 2008. **26**(10): p. 1626-34 DOI: 10.1200/JCO.2007.14.7116.

11. A. Lievre, J. Bachet, V. Boige, B. Landi, J. Emile, J. Cote, G. Tomasic, P. Rougier, F. Penault-Llorca, and P. Laurent-Puig, *KRAS mutation status is predictive of response to cetuximab therapy in colorectal cancer.* Annals of Oncology, 2006. **17**: p. 42-42.

12. C. Herrmann and N. Nassar, *Ras and its effectors.* Prog Biophys Mol Biol, 1996. **66**(1): p. 1-41.

13. M.V. Milburn, L. Tong, A.M. Devos, A. Brunger, Z. Yamaizumi, S. Nishimura, and S.H. Kim, *Molecular Switch for Signal Transduction - Structural Differences between Active and Inactive Forms of Protooncogenic Ras Proteins.* Science, 1990. **247**(4945): p. 939-945 DOI: DOI 10.1126/science.2406906.

14. I.R. Vetter and A. Wittinghofer, *Signal transduction - The guanine nucleotide-binding switch in three dimensions.* Science, 2001. **294**(5545): p. 1299-1304 DOI: 10.1126/science.1062023.

15. S. Lu, H. Jang, S. Muratcioglu, A. Gursoy, O. Keskin, R. Nussinov, and J. Zhang, *Ras Conformational Ensembles, Allostery, and Signaling.* Chem Rev, 2016. **116**(11): p. 6607-65 DOI: 10.1021/acs.chemrev.5b00542.

16. S.K. Fetics, H. Guterres, B.M. Kearney, G. Buhrman, B.Y. Ma, R. Nussinov, and C. Mattos, *Allosteric Effects of the Oncogenic RasQ61L Mutant on Raf-RBD.* Structure, 2015. **23**(3): p. 505-516 DOI: 10.1016/j.str.2014.12.017.

17. S.R. Sprang, *G proteins, effectors and GAPs: structure and mechanism.* Curr Opin Struct Biol, 1997. **7**(6): p. 849-56.

18. T.M. Glennon, J. Villa, and A. Warshel, *How does GAP catalyze the GTPase reaction of Ras?: A computer simulation study.* Biochemistry, 2000. **39**(32): p. 9641-9651 DOI: DOI 10.1021/bi000640e.

19. K. Scheffzek, M.R. Ahmadian, W. Kabsch, L. Wiesmuller, A. Lautwein, F. Schmitz, and A. Wittinghofer, *The Ras-RasGAP complex: Structural basis for GTPase activation and its loss in oncogenic Ras mutants.* Science, 1997. **277**(5324): p. 333-338 DOI: DOI 10.1126/science.277.5324.333.

20. S. Schubbert, K. Shannon, and G. Bollag, *Hyperactive Ras in developmental disorders and cancer.* Nat Rev Cancer, 2007. **7**(4): p. 295-308 DOI: 10.1038/nrc2109.

21. M.J. Smith, B.G. Neel, and M. Ikura, *NMR-based functional profiling of RASopathies and oncogenic RAS mutations.* Proc Natl Acad Sci U S A, 2013. **110**(12): p. 4574-9 DOI: 10.1073/pnas.1218173110.

22. A.A. Adjei, *Blocking oncogenic Ras signaling for cancer therapy.* J Natl Cancer Inst, 2001. **93**(14): p. 1062-74.

23. F. McCormick, *K-Ras protein as a drug target.* J Mol Med (Berl), 2016. **94**(3): p. 253-8 DOI: 10.1007/s00109-016-1382-7.

24. S. Lu, H. Jang, S. Gu, J. Zhang, and R. Nussinov, *Drugging Ras GTPase: a comprehensive mechanistic and signaling structural view.* Chem Soc Rev, 2016. **45**(18): p. 4929-52 DOI: 10.1039/c5cs00911a.

25. J.M. Ostrem and K.M. Shokat, *Direct small-molecule inhibitors of KRAS: from structural insights to mechanism-based design.* Nat Rev Drug Discov, 2016 DOI: 10.1038/nrd.2016.139.

26. K. Marcus and C. Mattos, *Direct Attack on RAS: Intramolecular Communication and Mutation-Specific Effects.* Clinical Cancer Research, 2015. **21**(8): p. 1810-1818 DOI: 10.1158/1078-0432.ccr-14-2148.

27. G.A. Repasky, E.J. Chenette, and C.J. Der, *Renewing the conspiracy theory debate: does Raf function alone to mediate Ras oncogenesis?* Trends in Cell Biology, 2004. **14**(11): p. 639-647 DOI: 10.1016/j.tcb.2004.09.014.

28. F. Zhang and J.K. Cheong, *The renewed battle against RAS-mutant cancers.* Cell Mol Life Sci, 2016. **73**(9): p. 1845-58 DOI: 10.1007/s00018-016-2155-8.

29. M.P. Patricelli, M.R. Janes, L.S. Li, R. Hansen, U. Peters, L.V. Kessler, Y. Chen, J.M. Kucharski, J. Feng, T. Ely, J.H. Chen, S.J. Firdaus, A. Babbar, P. Ren, and Y. Liu, *Selective Inhibition of Oncogenic KRAS Output with Small Molecules Targeting the Inactive State.* Cancer Discov, 2016. **6**(3): p. 316-29 DOI: 10.1158/2159-8290.CD-15-1105.

30. J.M. Ostrem, U. Peters, M.L. Sos, J.A. Wells, and K.M. Shokat, *K-Ras(G12C) inhibitors allosterically control GTP affinity and effector interactions.* Nature, 2013. **503**(7477): p. 548-+ DOI: 10.1038/nature12796.

31. S.M. Lim, K.D. Westover, S.B. Ficarro, R.A. Harrison, H.G. Choi, M.E. Pacold, M. Carrasco, J. Hunter, N.D. Kim, T. Xie, T. Sim, P.A. Janne, M. Meyerson, J.A. Marto,

J.R. Engen, and N.S. Gray, *Therapeutic targeting of oncogenic K-Ras by a covalent catalytic site inhibitor.* Angew Chem Int Ed Engl, 2014. **53**(1): p. 199-204 DOI: 10.1002/anie.201307387.

32. H. Singh, D.L. Longo, and B.A. Chabner, *Improving Prospects for Targeting RAS.* Journal of Clinical Oncology, 2015. **33**(31): p. 3650-+ DOI: 10.1200/Jco.2015.62.1052.

33. B.J. Grant, S. Lukman, H.J. Hocker, J. Sayyah, J.H. Brown, J.A. McCammon, and A.A. Gorfe, *Novel Allosteric Sites on Ras for Lead Generation.* PLoS One, 2011. **6**(10) DOI: ARTN e25711

10.1371/journal.pone.0025711.

34. E.S. Leshchiner, A. Parkhitko, G.H. Bird, J. Luccarelli, J.A. Bellairs, S. Escudero, K. Opoku-Nsiah, M. Godes, N. Perrimon, and L.D. Walensky, *Direct inhibition of oncogenic KRAS by hydrocarbon-stapled SOS1 helices.* Proceedings of the National Academy of Sciences of the United States of America, 2015. **112**(6): p. 1761-1766 DOI: 10.1073/pnas.1413185112.

35. P. Lito, M. Solomon, L.S. Li, R. Hansen, and N. Rosen, *Allele-specific inhibitors inactivate mutant KRAS G12C by a trapping mechanism.* Science, 2016. **351**(6273): p. 604-608 DOI: 10.1126/science.aad6204.

36. H. Jang, S.J. Abraham, T.S. Chavan, B. Hitchinson, L. Khavrutskii, N.I. Tarasova, R. Nussinov, and V. Gaponenko, *Mechanisms of membrane binding of small GTPase K-Ras4B farnesylated hypervariable region.* J Biol Chem, 2015. **290**(15): p. 9465-77 DOI: 10.1074/jbc.M114.620724.

37. H. Jang, A. Banerjee, T.S. Chavan, S. Lu, J. Zhang, V. Gaponenko, and R. Nussinov, *The higher level of complexity of K-Ras4B activation at the membrane.* Faseb Journal, 2016. **30**(4): p. 1643-55 DOI: 10.1096/fj.15-279091.

38. A.G. Taveras, S.W. Remiszewski, R.J. Doll, D. Cesarz, E.C. Huang, P. Kirschmeier, B.N. Pramanik, M.E. Snow, Y.S. Wang, J.D. delRosario, B. Vibulbhan, B.B. Bauer, J.E. Brown, D. Carr, J. Catino, C.A. Evans, V. Girijavallabhan, L. Heimark, L. James, S. Liberles, C. Nash, L. Perkins, M.M. Senior, A. Tsarbopoulos, A.K. Ganguly, R. Aust, E. Brown, D. Delisle, S. Fuhrman, T. Hendrickson, C. Kissinger, R. Love, W. Sisson, E. Villafranca, and S.E. Webber, *Pas oncoprotein inhibitors: The discovery of potent, ras nucleotide exchange inhibitors and the structural determination of a drug-protein complex.* Bioorganic & Medicinal Chemistry, 1997. **5**(1): p. 125-133 DOI: Doi 10.1016/S0968-0896(96)00202-7.

39. N.M. Goodey and S.J. Benkovic, *Allosteric regulation and catalysis emerge via a common route.* Nat Chem Biol, 2008. **4**(8): p. 474-82 DOI: 10.1038/nchembio.98.

40. D. Kern and E.R. Zuiderweg, *The role of dynamics in allosteric regulation.* Curr Opin Struct Biol, 2003. **13**(6): p. 748-57.

41. A.J. Wand, *On the dynamic origins of allosteric activation.* Science, 2001. **293**(5534): p. 1395 DOI: 10.1126/science.293.5534.1395a.

42. H. Kamberaj and A. van der Vaart, *Extracting the Causality of Correlated Motions from Molecular Dynamics Simulations.* Biophysical Journal, 2009. **97**(6): p. 1747-1755 DOI: 10.1016/j.bpj.2009.07.019.

43. E. Guarnera and I.N. Berezovsky, *Structure-Based Statistical Mechanical Model Accounts for the Causality and Energetics of Allosteric Communication.* PLoS Comput Biol, 2016. **12**(3): p. e1004678 DOI: 10.1371/journal.pcbi.1004678.

44. C.W.J. Granger, *Investigating Causal Relations by Econometric Models and Cross-spectral Methods.* Econometrica, 1969. **37**(3): p. 424-438 DOI: 10.2307/1912791.

45. P.E. Wright and H.J. Dyson, *Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm.* J Mol Biol, 1999. **293**(2): p. 321-31 DOI: 10.1006/jmbi.1999.3110.

46. R.O. Dror, R.M. Dirks, J.P. Grossman, H. Xu, and D.E. Shaw, *Biomolecular simulation: a computational microscope for molecular biology.* Annual Review of Biophysics, Vol 41, 2012. **41**: p. 429-52 DOI: 10.1146/annurev-biophys-042910-155245.

47. K. Henzler-Wildman and D. Kern, *Dynamic personalities of proteins.* Nature, 2007. **450**(7172): p. 964-972 DOI: 10.1038/nature06522.

48. J.S. Fetrow, A. Giammona, A. Kolinski, and J. Skolnick, *The protein folding problem: a biophysical enigma.* Curr Pharm Biotechnol, 2002. **3**(4): p. 329-47.

49. *Protein dynamics- J A McCammon.* 1984.

50. B. Brooks and M. Karplus, *HARMONIC DYNAMICS OF PROTEINS - NORMAL-MODES AND FLUCTUATIONS IN BOVINE PANCREATIC TRYPSIN-INHIBITOR.* Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences, 1983. **80**(21): p. 6571-6575 DOI: 10.1073/pnas.80.21.6571.

51. K.A. Henzler-Wildman, V. Thai, M. Lei, M. Ott, M. Wolf-Watz, T. Fenn, E. Pozharski, M.A. Wilson, G.A. Petsko, M. Karplus, C.G. Hubner, and D. Kern, *Intrinsic motions along an enzymatic reaction trajectory.* Nature, 2007. **450**(7171): p. 838-44 DOI: 10.1038/nature06410.

52. M. Karplus and J. Kuriyan, *Molecular dynamics and protein function.* Proc Natl Acad Sci U S A, 2005. **102**(19): p. 6679-85 DOI: 10.1073/pnas.0408930102.

53. E.Z. Eisenmesser, O. Millet, W. Labeikovsky, D.M. Korzhnev, M. Wolf-Watz, D.A. Bosco, J.J. Skalicky, L.E. Kay, and D. Kern, *Intrinsic dynamics of an enzyme underlies catalysis.* Nature, 2005. **438**(7064): p. 117-21 DOI: 10.1038/nature04105.

54. L.Q. Yang, P. Sang, Y. Tao, Y.X. Fu, K.Q. Zhang, Y.H. Xie, and S.Q. Liu, *Protein dynamics and motions in relation to their functions: several case studies and the underlying mechanisms.* J Biomol Struct Dyn, 2014. **32**(3): p. 372-93 DOI: 10.1080/07391102.2013.770372.

55. M. Orozco, *A theoretical view of protein dynamics.* Chem Soc Rev, 2014. **43**(14): p. 5051-66 DOI: 10.1039/c3cs60474h.

56. H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, and P.E. Bourne, *The Protein Data Bank.* Nucleic Acids Research, 2000. **28**(1): p. 235-42.

57. A. Kuzmanic, N.S. Pannu, and B. Zagrovic, *X-ray refinement significantly underestimates the level of microscopic heterogeneity in biomolecular crystals.* Nat Commun, 2014. **5**: p. 3220 DOI: 10.1038/ncomms4220.

58. I. Bahar, C. Chennubhotla, and D. Tobi, *Intrinsic dynamics of enzymes in the unbound state and relation to allosteric regulation.* Curr Opin Struct Biol, 2007. **17**(6): p. 633-40 DOI: 10.1016/j.sbi.2007.09.011.

59. S. Esteban-Martin, R.B. Fenwick, and X. Salvatella, *Synergistic use of NMR and MD simulations to study the structural heterogeneity of proteins.* Wiley Interdisciplinary Reviews-Computational Molecular Science, 2012. **2**(3): p. 466-478 DOI: 10.1002/wcms.1093.

60. *Molecular Dynamics Review.*

61. S. Kmiecik, D. Gront, M. Kolinski, L. Wieteska, A.E. Dawid, and A. Kolinski, *Coarse-Grained Protein Models and Their Applications.* Chem Rev, 2016. **116**(14): p. 7898-936 DOI: 10.1021/acs.chemrev.6b00163.

62. M. Baaden and S.J. Marrink, *Coarse-grain modelling of protein-protein interactions.* Curr Opin Struct Biol, 2013. **23**(6): p. 878-86 DOI: 10.1016/j.sbi.2013.09.004.

63. S. Riniker, J.R. Allison, and W.F. van Gunsteren, *On developing coarse-grained models for biomolecular simulation: a review.* Physical Chemistry Chemical Physics, 2012. **14**(36): p. 12423-30 DOI: 10.1039/c2cp40934h.

64. C. Clementi, *Coarse-grained models of protein folding: toy models or predictive tools?* Curr Opin Struct Biol, 2008. **18**(1): p. 10-5 DOI: 10.1016/j.sbi.2007.10.005.

65. V. Tozzini, *Coarse-grained models for proteins.* Curr Opin Struct Biol, 2005. **15**(2): p. 144-50 DOI: 10.1016/j.sbi.2005.02.005.

66. S.C. Kamerlin, S. Vicatos, A. Dryga, and A. Warshel, *Coarse-grained (multiscale) simulations in studies of biophysical and chemical systems.* Annual Review of Physical Chemistry, Vol 64, 2011. **62**: p. 41-64 DOI: 10.1146/annurev-physchem-032210-103335.

67. J.F. Dama, A.V. Sinitskiy, M. McCullagh, J. Weare, B. Roux, A.R. Dinner, and G.A. Voth, *The Theory of Ultra-Coarse-Graining. 1. General Principles.* Journal of Chemical Theory and Computation, 2013. **9**(5): p. 2466-80 DOI: 10.1021/ct4000444.

68. P.H. Nguyen, M.S. Li, and P. Derreumaux, *Effects of all-atom force fields on amyloid oligomerization: replica exchange molecular dynamics simulations of the Abeta(16-22) dimer and trimer.* Physical Chemistry Chemical Physics, 2011. **13**(20): p. 9778-88 DOI: 10.1039/c1cp20323a.

69. K.I.G. Ramachandran, D.; Krishnan, N., *Computational Chemistry and Molecular Modeling*. 1 ed. 2008: Springer-Verlag Berlin Heidelberg. XXII, 398.

70. J.A. McCammon, B.R. Gelin, and M. Karplus, *Dynamics of folded proteins.* Nature, 1977. **267**(5612): p. 585-90.

71. M. Karplus and J.A. McCammon, *Molecular dynamics simulations of biomolecules.* Nature Structural Biology, 2002. **9**(9): p. 646-52 DOI: 10.1038/nsb0902-646.

72. B. Hess, C. Kutzner, D. van der Spoel, and E. Lindahl, *GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation.* Journal of Chemical Theory and Computation, 2008. **4**(3): p. 435-447 DOI: 10.1021/ct700301q.

73. W.R.P. Scott, P.H. Hunenberger, I.G. Tironi, A.E. Mark, S.R. Billeter, J. Fennen, A.E. Torda, T. Huber, P. Kruger, and W.F. van Gunsteren, *The GROMOS biomolecular simulation program package.* Journal of Physical Chemistry A, 1999. **103**(19): p. 3596-3607 DOI: DOI 10.1021/jp984217f.

74. V.B. D.A. Case, J.T. Berryman, R.M. Betz, Q. Cai, D.S. Cerutti, T.E. Cheatham, III, T.A. Darden, R.E. Duke, H. Gohlke, A.W. Goetz, S. Gusarov, N. Homeyer, P. Janowski, J. Kaus, I. Kolossváry, A. Kovalenko, T.S. Lee, S. LeGrand, T. Luchko, R. Luo, B. Madej, K.M. Merz, F. Paesani, D.R. Roe, A. Roitberg, C. Sagui, R. Salomon-Ferrer, G. Seabra, C.L. Simmerling, W. Smith, J. Swails, R.C. Walker, J. Wang, R.M. Wolf, X. Wu and P.A. Kollman. *AMBER 14*. 2014.

75. B.R. Brooks, C.L. Brooks, 3rd, A.D. Mackerell, Jr., L. Nilsson, R.J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caflisch, L. Caves, Q. Cui, A.R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R.W. Pastor, C.B. Post, J.Z. Pu, M. Schaefer, B. Tidor, R.M. Venable, H.L. Woodcock, X. Wu, W. Yang, D.M. York, and M. Karplus, *CHARMM: the biomolecular simulation program.* J Comput Chem, 2009. **30**(10): p. 1545-614 DOI: 10.1002/jcc.21287.

76. J.C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R.D. Skeel, L. Kale, and K. Schulten, *Scalable molecular dynamics with NAMD.* Journal of Computational Chemistry, 2005. **26**(16): p. 1781-1802 DOI: 10.1002/jcc.20289.

77. S. Plimpton and B. Hendrickson, *Parallel Molecular-Dynamics Algorithms for Simulation of Molecular-Systems.* Parallel Computing in Computational Chemistry, 1995. **592**: p. 114-132.

78. S. Plimpton, *Fast Parallel Algorithms for Short-Range Molecular-Dynamics.* Journal of Computational Physics, 1995. **117**(1): p. 1-19 DOI: DOI 10.1006/jcph.1995.1039.

79. B.G. Fitch, A. Rayshubskiy, M. Eleftheriou, T.J.C. Ward, M. Giampapa, Y. Zhestkov, M.C. Pitman, F. Suits, A. Grossfield, J. Pitera, W. Swope, R.H. Zhou, S. Feller, and R.S. Germain, *Blue matter: Strong scaling of molecular dynamics on blue Gene/L.* Computational Science - Iccs 2006, Pt 2, Proceedings, 2006. **3992**: p. 846-854.

80. K.J. Bowers, R.O. Dror, and D.E. Shaw, *The midpoint method for parallelization of particle simulations.* Journal of Chemical Physics, 2006. **124**(18): p. 184109 DOI: 10.1063/1.2191489.

81. D.E. Shaw, R.O. Dror, J.K. Salmon, J.P. Grossman, K.M. Mackenzie, J.A. Bank, C. Young, M.M. Deneroff, B. Batson, K.J. Bowers, E. Chow, M.P. Eastwood, D.J. Ierardi, J.L. Klepeis, J.S. Kuskin, R.H. Larson, K. Lindorff-Larsen, P. Maragakis, M.A. Moraes, S. Piana, Y.B. Shan, and B. Towles, *Millisecond-Scale Molecular Dynamics Simulations on Anton.* Proceedings of the Conference on High Performance Computing Networking, Storage and Analysis, 2009.

82. D.E. Shaw, P. Maragakis, K. Lindorff-Larsen, S. Piana, R.O. Dror, M.P. Eastwood, J.A. Bank, J.M. Jumper, J.K. Salmon, Y. Shan, and W. Wriggers, *Atomic-level characterization of the structural dynamics of proteins.* Science, 2010. **330**(6002): p. 341-6 DOI: 10.1126/science.1187409.

83. K. Lindorff-Larsen, S. Piana, R.O. Dror, and D.E. Shaw, *How fast-folding proteins fold.* Science, 2011. **334**(6055): p. 517-20 DOI: 10.1126/science.1208351.

84. S. Piana, K. Lindorff-Larsen, and D.E. Shaw, *Protein folding kinetics and thermodynamics from atomistic simulation.* Proc Natl Acad Sci U S A, 2012. **109**(44): p. 17845-50 DOI: 10.1073/pnas.1201811109.

85. A. Dickson and C.L. Brooks, 3rd, *Native states of fast-folding proteins are kinetic traps.* J Am Chem Soc, 2013. **135**(12): p. 4729-34 DOI: 10.1021/ja311077u.

86. K.A. Beauchamp, R. McGibbon, Y.S. Lin, and V.S. Pande, *Simple few-state models reveal hidden complexity in protein folding.* Proc Natl Acad Sci U S A, 2012. **109**(44): p. 17807-13 DOI: 10.1073/pnas.1201810109.

87. S. Piana, K. Lindorff-Larsen, and D.E. Shaw, *Atomic-level description of ubiquitin folding.* Proc Natl Acad Sci U S A, 2013. **110**(15): p. 5915-20 DOI: 10.1073/pnas.1218321110.

88. M.O. Jensen, V. Jogini, D.W. Borhani, A.E. Leffler, R.O. Dror, and D.E. Shaw, *Mechanism of voltage gating in potassium channels.* Science, 2012. **336**(6078): p. 229-33 DOI: 10.1126/science.1216533.

89. A. Arkhipov, Y. Shan, R. Das, N.F. Endres, M.P. Eastwood, D.E. Wemmer, J. Kuriyan, and D.E. Shaw, *Architecture and membrane interactions of the EGF receptor.* Cell, 2013. **152**(3): p. 557-69 DOI: 10.1016/j.cell.2012.12.030.

90. J. Ostmeyer, S. Chakrapani, A.C. Pan, E. Perozo, and B. Roux, *Recovery from slow inactivation in K+ channels is controlled by water molecules.* Nature, 2013. **501**(7465): p. 121-4 DOI: 10.1038/nature12395.

91. A.W. Gotz, M.J. Williamson, D. Xu, D. Poole, S. Le Grand, and R.C. Walker, *Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 1. Generalized Born.* Journal of Chemical Theory and Computation, 2012. **8**(5): p. 1542-1555 DOI: 10.1021/ct200909j.

92. I. Buch, T. Giorgino, and G. De Fabritiis, *Complete reconstruction of an enzyme-inhibitor binding process by molecular dynamics simulations.* Proc Natl Acad Sci U S A, 2011. **108**(25): p. 10184-9 DOI: 10.1073/pnas.1103547108.

93. W. Humphrey, A. Dalke, and K. Schulten, *VMD: Visual molecular dynamics.* Journal of Molecular Graphics & Modelling, 1996. **14**(1): p. 33-38 DOI: Doi 10.1016/0263-7855(96)00018-5.

94. I.V. Likhachev, N.K. Balabaev, and O.V. Galzitskaya, *Available Instruments for Analyzing Molecular Dynamics Trajectories.* Open Biochem J, 2016. **10**: p. 1-11 DOI: 10.2174/1874091X01610010001.

95. X.P. Chen, W.H. Xu, D.F. Xu, S.M. Fu, and Z.C. Ma, *G12V Kras mutations in cervical cancer under virtual microscope of molecular dynamics simulations.* European Journal of Gynaecological Oncology, 2016. **37**(1): p. 69-74.

96. *Dynamics and function of proteins: The search for general concepts.* 1998.

97. T.D. Romo, J.B. Clarage, D.C. Sorensen, and G.N. Phillips, *Automatic Identification of Discrete Substates in Proteins - Singular-Value Decomposition Analysis of Time-Averaged Crystallographic Refinements.* Proteins-Structure Function and Genetics, 1995. **22**(4): p. 311-321 DOI: DOI 10.1002/prot.340220403.

98. F. Noe, *Beating the millisecond barrier in molecular dynamics simulations.* Biophys J, 2015. **108**(2): p. 228-9 DOI: 10.1016/j.bpj.2014.11.3477.

99. V.K.L. Stone J.E., Schulten K., *Immersive out-of-core visualization of large-size and long-timescale molecular dynamics trajectories.* , in *Lecture Notes in Computer Science*. 2011, Springer: Berlin, Heidelberg. p. 1-12.

100. D.P. Hurst, A. Grossfield, D.L. Lynch, S. Feller, T.D. Romo, K. Gawrisch, M.C. Pitman, and P.H. Reggio, *A Lipid Pathway for Ligand Binding Is Necessary for a Cannabinoid G Protein-coupled Receptor.* Journal of Biological Chemistry, 2010. **285**(23): p. 17954-17964 DOI: 10.1074/jbc.M109.041590.

101. R.O. Dror, A.C. Pan, D.H. Arlow, D.W. Borhani, P. Maragakis, Y. Shan, H. Xu, and D.E. Shaw, *Pathway and mechanism of drug binding to G-protein-coupled receptors.* Proc Natl Acad Sci U S A, 2011. **108**(32): p. 13118-23 DOI: 10.1073/pnas.1104614108.

102. Y.B. Shan, E. Kim, M.P. Eastwood, R.O. Dror, M.A. Seeliger, and D.E. Shaw, *How Does a Drug Molecule Find its Target Binding Site? (vol 133, pg 9181, 2011).* Journal of the American Chemical Society, 2014. **136**(8): p. 3320-3320 DOI: 10.1021/ja500545u.

103. E. Fischer, *Einfluss der Configuration auf die Wirkung der Enzyme.* Berichte der deutschen chemischen Gesellschaft, 1894. **27**(3): p. 2985-2993 DOI: 10.1002/cber.18940270364.

104. D.E. Koshland, *Application of a Theory of Enzyme Specificity to Protein Synthesis.* Proc Natl Acad Sci U S A, 1958. **44**(2): p. 98-104.

105. S.J. Teague, *Implications of protein flexibility for drug discovery.* Nat Rev Drug Discov, 2003. **2**(7): p. 527-41 DOI: 10.1038/nrd1129.

106. J.R. Schames, R.H. Henchman, J.S. Siegel, C.A. Sotriffer, H. Ni, and J.A. McCammon, *Discovery of a novel binding trench in HIV integrase.* J Med Chem, 2004. **47**(8): p. 1879-81 DOI: 10.1021/jm0341913.

107. M.A. Seeliger, P. Ranjitkar, C. Kasap, Y. Shan, D.E. Shaw, N.P. Shah, J. Kuriyan, and D.J. Maly, *Equally potent inhibition of c-Src and Abl by compounds that recognize inactive kinase conformations.* Cancer Res, 2009. **69**(6): p. 2384-92 DOI: 10.1158/0008-5472.CAN-08-3953.

108. H.M. Senn and W. Thiel, *QM/MM methods for biomolecular systems.* Angew Chem Int Ed Engl, 2009. **48**(7): p. 1198-229 DOI: 10.1002/anie.200802019.

109. Y. Shan, E.T. Kim, M.P. Eastwood, R.O. Dror, M.A. Seeliger, and D.E. Shaw, *How does a drug molecule find its target binding site?* J Am Chem Soc, 2011. **133**(24): p. 9181-3 DOI: 10.1021/ja202726y.

110. R. McGreevy, A. Singharoy, Q. Li, J. Zhang, D. Xu, E. Perozo, and K. Schulten, *xMDFF: molecular dynamics flexible fitting of low-resolution X-ray structures.* Acta Crystallogr D Biol Crystallogr, 2014. **70**(Pt 9): p. 2344-55 DOI: 10.1107/S1399004714013856.

111. O.F. Lange and H. Grubmuller, *Generalized correlation for biomolecular dynamics.* Proteins-Structure Function and Bioinformatics, 2006. **62**(4): p. 1053-1061 DOI: 10.1002/prot.20784.

112. P.H. Hunenberger, A.E. Mark, and W.F. Vangunsteren, *Fluctuation and Cross-Correlation Analysis of Protein Motions Observed in Nanosecond Molecular-Dynamics Simulations.* Journal of Molecular Biology, 1995. **252**(4): p. 492-503 DOI: DOI 10.1006/jmbi.1995.0514.

113. T. Ichiye and M. Karplus, *Collective Motions in Proteins - a Covariance Analysis of Atomic Fluctuations in Molecular-Dynamics and Normal Mode Simulations.* Proteins-Structure Function and Genetics, 1991. **11**(3): p. 205-217 DOI: DOI 10.1002/prot.340110305.

114. M.M. Teeter and D.A. Case, *Harmonic and Quasiharmonic Descriptions of Crambin.* Journal of Physical Chemistry, 1990. **94**(21): p. 8091-8097 DOI: DOI 10.1021/j100384a021.

115. J. Baldwin and C. Chothia, *Haemoglobin: The structural changes related to ligand binding and its allosteric mechanism.* Journal of Molecular Biology, 1979. **129**(2): p. 175-220 DOI: http://dx.doi.org/10.1016/0022-2836(79)90277-8.

116. M.F. Perutz, A.J. Wilkinson, M. Paoli, and G.G. Dodson, *The stereochemical mechanism of the cooperative effects in hemoglobin revisited.* Annual Review of

Biophysics and Biomolecular Structure, 1998. **27**: p. 1-34 DOI: DOI 10.1146/annurev.biophys.27.1.1.

117. H.R. Saibil and N.A. Ranson, *The chaperonin folding machine.* Trends in Biochemical Sciences, 2002. **27**(12): p. 627-632 DOI: Pii S0968-0004(02)02211-9

Doi 10.1016/S0968-0004(02)02211-9.

118. H. Grallert and J. Buchner, *Review: A structural view of the GroE chaperone cycle.* Journal of Structural Biology, 2001. **135**(2): p. 95-103 DOI: 10.1006/jsbi.2001.4387.

119. A. Hacisuleyman and B. Erman, *Entropy Transfer between Residue Pairs and Allostery in Proteins: Quantifying Allosteric Communication in Ubiquitin.* PLoS Comput Biol, 2017. **13**(1): p. e1005319 DOI: 10.1371/journal.pcbi.1005319.

120. A. Hacisuleyman and B. Erman, *Causality, transfer entropy, and allosteric communication landscapes in proteins with harmonic interactions.* Proteins, 2017. **85**(6): p. 1056-1064 DOI: 10.1002/prot.25272.

121. T. Schreiber, *Measuring information transfer.* Physical Review Letters, 2000. **85**(2): p. 461-464 DOI: DOI 10.1103/PhysRevLett.85.461.

122. A. Buehlmann and G. Deco, *Optimal Information Transfer in the Cortex through Synchronization.* Plos Computational Biology, 2010. **6**(9): p. e1000934 DOI: 10.1371/journal.pcbi.1000934.

123. B. Gourevitch and J.J. Eggermont, *Evaluating information transfer between auditory cortical neurons.* J Neurophysiol, 2007. **97**(3): p. 2533-43 DOI: 10.1152/jn.01106.2006.

124. D. Barr, T. Oashi, K. Burkhard, S. Lucius, R. Samadani, J. Zhang, P. Shapiro, A.D. MacKerell, and A. van der Vaart, *Importance of domain closure for the autoactivation of ERK2.* Biochemistry, 2011. **50**(37): p. 8038-48 DOI: 10.1021/bi200503a.

125. W.J. Granger. *Facts*. 2014; Available from: http://www.nobelprize.org/nobel_prizes/economic-sciences/laureates/2003/granger-facts.htmlClive.

126. Y. Chen, S.L. Bressler, and M. Ding, *Frequency decomposition of conditional Granger causality and application to multivariate neural field potential data.* J Neurosci Methods, 2006. **150**(2): p. 228-37 DOI: 10.1016/j.jneumeth.2005.06.011.

127. R. Gorelik and A. Gautreau, *Quantitative and unbiased analysis of directional persistence in cell migration.* Nature Protocols, 2014. **9**(8): p. 1931-1943 DOI: 10.1038/nprot.2014.131.

128. A.J. Lin, K.K.L. Liu, R.P. Bartsch, and P.C. Ivanov, *Delay-correlation landscape reveals characteristic time delays of brain rhythms and heart interactions.* Philosophical Transactions of the Royal Society a-Mathematical Physical and Engineering Sciences, 2016. **374**(2067) DOI: ARTN 20150182

10.1098/rsta.2015.0182.

129. E. Nelson-Wong, S. Howarth, D.A. Winter, and J.P. Callaghan, *Application of autocorrelation and cross-correlation analyses in human movement and rehabilitation research.* J Orthop Sports Phys Ther, 2009. **39**(4): p. 287-95 DOI: 10.2519/jospt.2009.2969.

130. M.J. Williams and K. Rottner, *Introduction to Small GTPases.* Small GTPases, 2010. **1**(1): p. 1 DOI: 10.4161/sgtp.1.1.12245.

131. J.J. Harvey, *An Unidentified Virus Which Causes the Rapid Production of Tumours in Mice.* Nature, 1964. **204**: p. 1104-5.

132. W.H. Kirsten and L.A. Mayer, *Morphologic responses to a murine erythroblastosis virus.* J Natl Cancer Inst, 1967. **39**(2): p. 311-35.

133. E.M. Scolnick 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.

134. E.M. Scolnick, E. Rands, D. Williams, and W.P. Parks, *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.

135. T.Y. Shih, M.O. Weeks, H.A. Young, and E.M. Scholnick, *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.

136. E.M. Scolnick, 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.

137. M.C. Willingham, I. Pastan, T.Y. Shih, and E.M. Scolnick, *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.

138. E.H. Chang, M.E. Furth, E.M. Scolnick, and D.R. Lowy, *Tumorigenic transformation of mammalian cells induced by a normal human gene homologous to the oncogene of Harvey murine sarcoma virus.* Nature, 1982. **297**(5866): p. 479-83.

139. J.B. Gibbs, I.S. Sigal, M. Poe, and E.M. Scolnick, *Intrinsic GTPase activity distinguishes normal and oncogenic ras p21 molecules.* Proc Natl Acad Sci U S A, 1984. **81**(18): p. 5704-8.

140. E. Santos, S.R. Tronick, S.A. Aaronson, S. Pulciani, and M. Barbacid, *T24 human bladder carcinoma oncogene is an activated form of the normal human homologue of BALB- and Harvey-MSV transforming genes.* Nature, 1982. **298**(5872): p. 343-7.

141. L.F. Parada, C.J. Tabin, C. Shih, and R.A. Weinberg, *Human EJ bladder carcinoma oncogene is homologue of Harvey sarcoma virus ras gene.* Nature, 1982. **297**(5866): p. 474-8.

142. C.M. O'Toole, S. Povey, P. Hepburn, and L.M. Franks, *Identity of some human bladder cancer cell lines.* Nature, 1983. **301**(5899): p. 429-30.

143. D.J. Capon, P.H. Seeburg, J.P. McGrath, J.S. Hayflick, U. Edman, A.D. Levinson, and D.V. Goeddel, *Activation of Ki-ras2 gene in human colon and lung carcinomas by two different point mutations.* Nature, 1983. **304**(5926): p. 507-13.

144. A.M. de Vos, L. Tong, M.V. Milburn, P.M. Matias, J. Jancarik, S. Noguchi, S. Nishimura, K. Miura, E. Ohtsuka, and S.H. Kim, *Three-dimensional structure of an oncogene protein: catalytic domain of human c-H-ras p21.* Science, 1988. **239**(4842): p. 888-93.

145. I.R. Vetter and A. Wittinghofer, *The guanine nucleotide-binding switch in three dimensions.* Science, 2001. **294**(5545): p. 1299-304 DOI: 10.1126/science.1062023.

146. M.V. Milburn, L. Tong, A.M. deVos, A. Brunger, Z. Yamaizumi, S. Nishimura, and S.H. Kim, *Molecular switch for signal transduction: structural differences between active and inactive forms of protooncogenic ras proteins.* Science, 1990. **247**(4945): p. 939-45.

147. A.A. Gorfe, 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. 885-896 DOI: 10.1016/j.str.2008.03.009.

148. G. Buhrman, C. O'Connor, B. Zerbe, B.M. Kearney, R. Napoleon, E.A. Kovrigina, S. Vajda, D. Kozakov, E.L. Kovrigin, and C. Mattos, *Analysis of binding site hot spots on the surface of Ras GTPase.* J Mol Biol, 2011. **413**(4): p. 773-89 DOI: 10.1016/j.jmb.2011.09.011.

149. E. Castellano and E. Santos, *Functional specificity of ras isoforms: so similar but so different.* Genes Cancer, 2011. **2**(3): p. 216-31 DOI: 10.1177/1947601911408081.

150. A. Wolfman, *Ras isoform-specific signaling: location, location, location.* Sci STKE, 2001. **2001**(96): p. pe2 DOI: 10.1126/stke.2001.96.pe2.

151. J.F. Hancock, *Ras proteins: different signals from different locations.* Nat Rev Mol Cell Biol, 2003. **4**(5): p. 373-84 DOI: 10.1038/nrm1105.

152. J.F. Hancock and R.G. Parton, *Ras plasma membrane signalling platforms.* Biochem J, 2005. **389**(Pt 1): p. 1-11 DOI: 10.1042/BJ20050231.

153. A. Ehrhardt, M.D. David, G.R. Ehrhardt, and J.W. Schrader, *Distinct mechanisms determine the patterns of differential activation of H-Ras, N-Ras, K-Ras 4B, and M-Ras by receptors for growth factors or antigen.* Mol Cell Biol, 2004. **24**(14): p. 6311-23 DOI: 10.1128/MCB.24.14.6311-6323.2004.

154. A. Ehrhardt, G.R. Ehrhardt, X. Guo, and J.W. Schrader, *Ras and relatives--job sharing and networking keep an old family together.* Exp Hematol, 2002. **30**(10): p. 1089-106.

155. M.K. Jones and J.H. Jackson, *Ras-GRF activates Ha-Ras, but not N-Ras or K-Ras 4B, protein in vivo.* J Biol Chem, 1998. **273**(3): p. 1782-7.

156. J. Clyde-Smith, G. Silins, M. Gartside, S. Grimmond, M. Etheridge, A. Apolloni, N. Hayward, and J.F. Hancock, *Characterization of RasGRP2, a plasma membrane-targeted, dual specificity Ras/Rap exchange factor.* J Biol Chem, 2000. **275**(41): p. 32260-7 DOI: 10.1074/jbc.M006087200.

157. J. Yan, S. Roy, A. Apolloni, A. Lane, and J.F. Hancock, *Ras isoforms vary in their ability to activate Raf-1 and phosphoinositide 3-kinase.* J Biol Chem, 1998. **273**(37): p. 24052-6.

158. T. Maurer, L.S. Garrenton, A. Oh, K. Pitts, D.J. Anderson, N.J. Skelton, B.P. Fauber, B. Pan, S. Malek, D. Stokoe, M.J. Ludlam, K.K. Bowman, J. Wu, A.M. Giannetti, M.A. Starovasnik, I. Mellman, P.K. Jackson, J. Rudolph, W. Wang, and G. Fang, *Small-molecule ligands bind to a distinct pocket in Ras and inhibit SOS-mediated nucleotide exchange activity.* Proc Natl Acad Sci U S A, 2012. **109**(14): p. 5299-304 DOI: 10.1073/pnas.1116510109.

159. E.F. Pai, U. Krengel, G.A. Petsko, R.S. Goody, W. Kabsch, and A. Wittinghofer, *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.

160. K.R. Acharya and M.D. Lloyd, *The advantages and limitations of protein crystal structures.* Trends Pharmacol Sci, 2005. **26**(1): p. 10-4 DOI: 10.1016/j.tips.2004.10.011.

161. P.J. Kraulis, P.J. Domaille, S.L. Campbell-Burk, T. Van Aken, and E.D. Laue, *Solution structure and dynamics of ras p21.GDP determined by heteronuclear three- and four-dimensional NMR spectroscopy.* Biochemistry, 1994. **33**(12): p. 3515-31.

162. Y. Ito, K. Yamasaki, J. Iwahara, T. Terada, A. Kamiya, M. Shirouzu, Y. Muto, G. Kawai, S. Yokoyama, E.D. Laue, M. Walchli, T. Shibata, S. Nishimura, and T. Miyazawa, *Regional polysterism in the GTP-bound form of the human c-Ha-Ras protein.* Biochemistry, 1997. **36**(30): p. 9109-19 DOI: 10.1021/bi970296u.

163. M. Araki, F. Shima, Y. Yoshikawa, S. Muraoka, Y. Ijiri, Y. Nagahara, T. Shirono, T. Kataoka, and A. Tamura, *Solution structure of the state 1 conformer of GTP-bound H-Ras protein and distinct dynamic properties between the state 1 and state 2 conformers.* J Biol Chem, 2011. **286**(45): p. 39644-53 DOI: 10.1074/jbc.M111.227074.

164. C.K. Foley, L.G. Pedersen, P.S. Charifson, T.A. Darden, A. Wittinghofer, E.F. Pai, and M.W. Anderson, *Simulation of the solution structure of the H-ras p21-GTP complex.* Biochemistry, 1992. **31**(21): p. 4951-9.

165. J.F. Diaz, B. Wroblowski, and Y. Engelborghs, *Molecular dynamics simulation of the solution structures of Ha-ras-p21 GDP and GTP complexes: flexibility, possible hinges, and levers of the conformational transition.* Biochemistry, 1995. **34**(37): p. 12038-47.

166. J.M. Chen, G. Lee, R.B. Murphy, R.P. Carty, P.W. Brandt-Rauf, E. Friedman, and M.R. Pincus, *Comparison of the computed structures for the phosphate-binding loop of the p21 protein containing the oncogenic site Gly 12 with the X-ray crystallographic structures for this region in the p21 protein and EFtu. A model for the structure of the p21 protein in its oncogenic form.* J Biomol Struct Dyn, 1989. **6**(5): p. 859-75 DOI: 10.1080/07391102.1989.10506518.

167. L.V. Mello, D.M. van Aalten, and J.B. Findlay, *Comparison of ras-p21 bound to GDP and GTP: differences in protein and ligand dynamics.* Protein Eng, 1997. **10**(4): p. 381-7.

168. D.C. Dykes, F.K. Friedman, S.L. Dykes, R.B. Murphy, P.W. Brandt-Rauf, and M.R. Pincus, *Molecular dynamics of the H-ras gene-encoded p21 protein; identification of flexible regions and possible effector domains.* J Biomol Struct Dyn, 1993. **11**(3): p. 443-58 DOI: 10.1080/07391102.1993.10508009.

169. N. Futatsugi and M. Tsuda, *Molecular dynamics simulations of Gly-12-->Val mutant of p21(ras): dynamic inhibition mechanism.* Biophys J, 2001. **81**(6): p. 3483-8 DOI: 10.1016/S0006-3495(01)75979-6.

170. S. Lukman, B.J. Grant, A.A. Gorfe, G.H. Grant, and J.A. McCammon, *The distinct conformational dynamics of K-Ras and H-Ras A59G.* PLoS Comput Biol, 2010. **6**(9) DOI: 10.1371/journal.pcbi.1000922.

171. G. Buhrman, G. Holzapfel, S. Fetics, and C. Mattos, *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 DOI: 10.1073/pnas.0912226107.

172. F. Raimondi, A. Felline, G. Portella, M. Orozco, and F. Fanelli, *Light on the structural communication in Ras GTPases.* J Biomol Struct Dyn, 2013. **31**(2): p. 142-57 DOI: 10.1080/07391102.2012.698379.

173. F. Raimondi, M. Orozco, and F. Fanelli, *Deciphering the deformation modes associated with function retention and specialization in members of the Ras superfamily.* Structure, 2010. **18**(3): p. 402-14 DOI: 10.1016/j.str.2009.12.015.

174. F. Raimondi, G. Portella, M. Orozco, and F. Fanelli, *Nucleotide binding switches the information flow in ras GTPases.* PLoS Comput Biol, 2011. **7**(3): p. e1001098 DOI: 10.1371/journal.pcbi.1001098.

175. H. Resat, T.P. Straatsma, D.A. Dixon, and J.H. Miller, *The arginine finger of RasGAP helps Gln-61 align the nucleophilic water in GAP-stimulated hydrolysis of GTP*. Proceedings of the National Academy of Sciences of the United States of America, 2001. **98**(11): p. 6033-6038 DOI: 10.1073/pnas.091506998.

176. T.A. Soares, J.H. Miller, and T.P. Straatsma, *Revisiting the structural flexibility of the complex p21(ras)-GTP: the catalytic conformation of the molecular switch II.* Proteins, 2001. **45**(4): p. 297-312.

177. D.L. Ensign and L.J. Webb, *Factors determining electrostatic fields in molecular dynamics simulations of the Ras/effector interface.* Proteins, 2011. **79**(12): p. 3511-24 DOI: 10.1002/prot.23095.

178. C. Kiel, D. Aydin, and L. Serrano, *Association rate constants of ras-effector interactions are evolutionarily conserved.* PLoS Comput Biol, 2008. **4**(12): p. e1000245 DOI: 10.1371/journal.pcbi.1000245.

179. J. Baussand and J. Kleinjung, *Specific Conformational States of Ras GTPase upon Effector Binding.* Journal of Chemical Theory and Computation, 2013. **9**(1): p. 738-749 DOI: 10.1021/ct3007265.

180. J. Rudolph and D. Stokoe, *Selective inhibition of mutant Ras protein through covalent binding.* Angew Chem Int Ed Engl, 2014. **53**(15): p. 3777-9 DOI: 10.1002/anie.201400233.

181. Q. Sun, J.P. Burke, J. Phan, M.C. Burns, E.T. Olejniczak, A.G. Waterson, T. Lee, O.W. Rossanese, and S.W. Fesik, *Discovery of small molecules that bind to K-Ras and inhibit Sos-mediated activation.* Angew Chem Int Ed Engl, 2012. **51**(25): p. 6140-3 DOI: 10.1002/anie.201201358.

182. M.C. Burns, Q. Sun, R.N. Daniels, D. Camper, J.P. Kennedy, J. Phan, E.T. Olejniczak, T. Lee, A.G. Waterson, O.W. Rossanese, and S.W. Fesik, *Approach for targeting Ras with small molecules that activate SOS-mediated nucleotide exchange.* Proceedings of the National Academy of Sciences of the United States of America, 2014. **111**(9): p. 3401-3406 DOI: 10.1073/pnas.1315798111.

183. M.R. Ahmadian, T. Zor, D. Vogt, W. Kabsch, Z. Selinger, A. Wittinghofer, and K. Scheffzek, *Guanosine triphosphatase stimulation of oncogenic Ras mutants.* Proceedings of the National Academy of Sciences of the United States of America, 1999. **96**(12): p. 7065-7070 DOI: DOI 10.1073/pnas.96.12.7065.

184. G. Chen, S. Oh, B.P. Monia, and D.W. Stacey, *Antisense oligonucleotides demonstrate a dominant role of c-Ki-RAS proteins in regulating the proliferation of diploid human fibroblasts.* J Biol Chem, 1996. **271**(45): p. 28259-65.

185. E.Z. Khvalevsky, R. Gabai, I.H. Rachmut, E. Horwitz, Z. Brunschwig, A. Orbach, A. Shemi, T. Golan, A.J. Domb, E. Yavin, H. Giladi, L. Rivkin, A. Simerzin, R. Eliakim, A. Khalaileh, A. Hubert, M. Lahav, Y. Kopelman, E. Goldin, A. Dancour, Y. Hants, S. Arbel-Alon, R. Abramovitch, A. Shemi, and E. Galun, *Mutant KRAS is a druggable target for pancreatic cancer.* Proceedings of the National Academy of Sciences of the United States of America, 2013. **110**(51): p. 20723-20728 DOI: 10.1073/pnas.1314307110.

186. T. Golan, E.Z. Khvalevsky, A. Hubert, R.M. Gabai, N. Hen, A. Segal, A. Domb, G. Harari, E. Ben David, S. Raskin, Y. Goldes, E. Goldin, R. Eliakim, M. Lahav, Y. Kopleman, A. Dancour, A. Shemi, and E. Galun, *RNAi therapy targeting KRAS in combination with chemotherapy for locally advanced pancreatic cancer patients.* Oncotarget, 2015. **6**(27): p. 24560-24570 DOI: 10.18632/oncotarget.4183.

187. D.S. BIOVIA, *Discovery Studio Modeling Environment*. 2015: San Diego: Dassault Systèmes.

188. V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg, and C. Simmerling, *Comparison of multiple amber force fields and development of improved protein backbone parameters.* Proteins-Structure Function and Bioinformatics, 2006. **65**(3): p. 712-725 DOI: 10.1002/prot.21123.

189. J.M. Wang, R.M. Wolf, J.W. Caldwell, P.A. Kollman, and D.A. Case, *Development and testing of a general amber force field.* Journal of Computational Chemistry, 2004. **25**(9): p. 1157-1174 DOI: DOI 10.1002/jcc.20035.

190. C.I. Bayly, P. Cieplak, W.D. Cornell, and P.A. Kollman, *A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints for Deriving Atomic Charges - the Resp Model.* Journal of Physical Chemistry, 1993. **97**(40): p. 10269-10280 DOI: DOI 10.1021/j100142a004.

191. F.Y. Dupradeau, A. Pigache, T. Zaffran, C. Savineau, R. Lelong, N. Grivel, D. Lelong, W. Rosanski, and P. Cieplak, *The R.ED. tools: advances in RESP and ESP charge derivation and force field library building.* Physical Chemistry Chemical Physics, 2010. **12**(28): p. 7821-7839 DOI: 10.1039/c0cp00111b.

192. J.M. Wang, W. Wang, and P.A. Kollman, *Antechamber: An accessory software package for molecular mechanical calculations.* Abstracts of Papers of the American Chemical Society, 2001. **222**: p. U403-U403.

193. E. Vanquelef, S. Simon, G. Marquant, E. Garcia, G. Klimerak, J.C. Delepine, P. Cieplak, and F.Y. Dupradeau, *R.E.D. Server: a web service for deriving RESP and ESP charges and building force field libraries for new molecules and molecular fragments.* Nucleic Acids Research, 2011. **39**(Web Server issue): p. W511-7 DOI: 10.1093/nar/gkr288.

194. B. Erman, *Effects of ligand binding upon flexibility of proteins.* Proteins, 2015. **83**(5): p. 805-8 DOI: 10.1002/prot.24785.

195. T. Haliloglu, I. Bahar, and B. Erman, *Gaussian dynamics of folded proteins.* Physical Review Letters, 1997. **79**(16): p. 3090-3093 DOI: DOI 10.1103/PhysRevLett.79.3090.

196. F. Noe, F. Ille, J.C. Smith, and S. Fischer, *Automated computation of low-energry pathways for complex Rearrangements in proteins: Application to the conformational switch of ras p21.* Proteins-Structure Function and Bioinformatics, 2005. **59**(3): p. 534-544 DOI: 10.1002/prot.20422.

197. I. Adamovic, S.M. Mijailovich, and M. Karplus, *The elastic properties of the structurally characterized myosin II S2 subdomain: a molecular dynamics and normal mode analysis.* Biophys J, 2008. **94**(10): p. 3779-89 DOI: 10.1529/biophysj.107.122028.

198. J. Howard, *Mechanics of Motor Proteins and the Cytoskeleton*. First ed. 2001: Sinauer Associates, INC.

199. F. Rico, A. Rigato, L. Picas, and S. Scheuring, *Mechanics of proteins with a focus on atomic force microscopy.* J Nanobiotechnology, 2013. **11 Suppl 1**: p. S3 DOI: 10.1186/1477-3155-11-S1-S3.

200. G. Zaccai, *How soft is a protein? A protein dynamics force constant measured by neutron scattering.* Science, 2000. **288**(5471): p. 1604-7.

201. M. Spoerner, C. Hozsa, J.A. Poetzl, K. Reiss, P. Ganser, M. Geyer, and H.R. Kalbitzer, *Conformational States of Human Rat Sarcoma (Ras) Protein Complexed with Its Natural Ligand GTP and Their Role for Effector Interaction and GTP Hydrolysis.* Journal of Biological Chemistry, 2010. **285**(51): p. 39768-39778 DOI: 10.1074/jbc.M110.145235.

202. F. Shima, Y. Ijiri, S. Muraoka, J.L. Liao, M. Ye, M. Araki, K. Matsumoto, N. Yamamoto, T. Sugimoto, Y. Yoshikawa, T. Kumasaka, M. Yamamoto, A. Tamura, and T. Kataoka, *Structural Basis for Conformational Dynamics of GTP-bound Ras Protein.* Journal of Biological Chemistry, 2010. **285**(29): p. 22696-22705 DOI: 10.1074/jbc.M110.125161.

203. A. Kapoor and A. Travesset, *Differential dynamics of RAS isoforms in GDP- and GTP-bound states.* Proteins, 2015. **83**(6): p. 1091-106 DOI: 10.1002/prot.24805.

204. R. Clausen, B.Y. Ma, R. Nussinov, and A. Shehu, *Mapping the Conformation Space of Wildtype and Mutant H-Ras with a Memetic, Cellular, and Multiscale Evolutionary Algorithm.* Plos Computational Biology, 2015. **11**(9) DOI: ARTN e1004470

10.1371/journal.pcbi.1004470.

205. B.J. Grant, A.A. Gorfe, and J.A. McCammon, *Ras Conformational Switching: Simulating Nucleotide-Dependent Conformational Transitions with Accelerated Molecular Dynamics.* Plos Computational Biology, 2009. **5**(3) DOI: ARTN e1000325

10.1371/journal.pcbi.1000325.

206. S. Muratcioglu, T.S. Chavan, B.C. Freed, H. Jang, L. Khavrutskii, R.N. Freed, M.A. Dyba, K. Stefanisko, S.G. Tarasov, A. Gursoy, O. Keskin, N.I. Tarasova, V. Gaponenko, and R. Nussinov, *GTP-Dependent K-Ras Dimerization.* Structure, 2015. **23**(7): p. 1325-1335 DOI: 10.1016/j.str.2015.04.019.

207. B.J. Grant, A.A. Gorfe, and J.A. McCammon, *Ras conformational switching: simulating nucleotide-dependent conformational transitions with accelerated molecular dynamics.* PLoS Comput Biol, 2009. **5**(3): p. e1000325 DOI: 10.1371/journal.pcbi.1000325.

208. M. Spoerner, T. Graf, B. Konig, and H.R. Kalbitzer, *A novel mechanism for the modulation of the Ras-effector interaction by small molecules.* Biochemical and Biophysical Research Communications, 2005. **334**(2): p. 709-713 DOI: 10.1016/j.bbrc.2005.06.144.

209. I.C. Rosnizeck, T. Graf, M. Spoerner, J. Trankle, D. Filchtinski, C. Herrmann, L. Gremer, I.R. Vetter, A. Wittinghofer, B. Konig, and H.R. Kalbitzer, *Stabilizing a Weak Binding State for Effectors in the Human Ras Protein by Cyclen Complexes.* Angewandte Chemie-International Edition, 2010. **49**(22): p. 3830-3833 DOI: 10.1002/anie.200907002.

210. D. Abankwa, M. Hanzal-Bayer, N. Ariotti, S.J. Plowman, A.A. Gorfe, R.G. Parton, J.A. McCammon, and J.F. Hancock, *A novel switch region regulates H-ras membrane orientation and signal output.* Embo Journal, 2008. **27**(5): p. 727-735 DOI: 10.1038/emboj.2008.10.

211. S. Lukman, B.J. Grant, A.A. Gorfe, G.H. Grant, and J.A. McCammon, *The Distinct Conformational Dynamics of K-Ras and H-Ras A59G.* Plos Computational Biology, 2010. **6**(9) DOI: 10.1371/journal.pcbi.1000922.

212. S. Vatansever, Z.H. Gumus, and B. Erman, *Intrinsic K-Ras dynamics: A novel molecular dynamics data analysis method shows causality between residue pair motions.* Sci Rep, 2016. **6**: p. 37012 DOI: 10.1038/srep37012.

213. I. Bahar, A.R. Atilgan, and B. Erman, *Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential.* Folding & Design, 1997. **2**(3): p. 173-181 DOI: Doi 10.1016/S1359-0278(97)00024-2.

214. I. Bahar, A.R. Atilgan, M.C. Demirel, and B. Erman, *Vibrational dynamics of folded proteins: Significance of slow and fast motions in relation to function and stability.* Physical Review Letters, 1998. **80**(12): p. 2733-2736 DOI: DOI 10.1103/PhysRevLett.80.2733.

215. A.R. Atilgan, P. Akan, and C. Baysal, *Small-world communication of residues and significance for protein dynamics.* Biophys J, 2004. **86**(1 Pt 1): p. 85-91 DOI: 10.1016/S0006-3495(04)74086-2.

216. L.V. Woodcock, *Entropy difference between the face-centred cubic and hexagonal close-packed crystal structures.* Nature, 1997. **385**(6612): p. 141-143 DOI: DOI 10.1038/385141a0.

217. J. Downward, *Targeting ras signalling pathways in cancer therapy.* Nature Reviews Cancer, 2003. **3**(1): p. 11-22 DOI: 10.1038/nrc969.

218. C.C. Chen, T.K. Er, Y.Y. Liu, J.K. Hwang, M.J. Barrio, M. Rodrigo, E. Garcia-Toro, and M. Herreros-Villanueva, *Computational analysis of KRAS mutations: implications for different effects on the KRAS p.G12D and p.G13D mutations.* PLoS One, 2013. **8**(2): p. e55793 DOI: 10.1371/journal.pone.0055793.

219. S. Lu, A. Banerjee, H. Jang, J. Zhang, V. Gaponenko, and R. Nussinov, *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. 28887-900 DOI: 10.1074/jbc.M115.664755.

220. S.Y. Lu, H. Jang, R. Nussinov, and J. Zhang, *The Structural Basis of Oncogenic Mutations G12, G13 and Q61 in Small GTPase K-Ras4B.* Scientific Reports, 2016. **6** DOI: ARTN 21949

10.1038/srep21949.

221. P. Prakash, J.F. Hancock, and A.A. Gorfe, *Binding hotspots on K-Ras: consensus ligand binding sites and other reactive regions from probe-based molecular dynamics analysis.* Proteins-Structure Function and Bioinformatics, 2015. **83**(5): p. 898-909 DOI: 10.1002/prot.24786.

222. S.Y. Lu, A. Banerjee, H. Jang, J. Zhang, V. Gaponenko, and R. Nussinov, *GTP Binding and Oncogenic Mutations May Attenuate Hypervariable Region (HVR)-Catalytic Domain Interactions in Small GTPase K-Ras4B, Exposing the Effector Binding Site.* Journal of Biological Chemistry, 2015. **290**(48): p. 28887-28900 DOI: 10.1074/jbc.M115.664755.

223. J. Baussand and J. Kleinjung, *Specific Conformational States of Ras GTPase upon Effector Binding.* Journal of Chemical Theory and Computation, 2013. **9**(1): p. 738-749 DOI: 10.1021/ct3007265.

224. *Biochemical and Structural Analysis of Common Cancer-Associated KRAS Mutations.* 2015.

225. N.T. Ihle, L.A. Byers, E.S. Kim, P. Saintigny, J.J. Lee, G.R. Blumenschein, A. Tsao, S.Y. Liu, J.E. Larsen, J. Wang, L.X. Diao, K.R. Coombes, L. Chen, S.X. Zhang, M.F. Abdelmelek, X.M. Tang, V. Papadimitrakopoulou, J.D. Minna, S.M. Lippman, W.K. Hong, R.S. Herbst, I.I. Wistuba, J.V. Heymach, and G. Powis, *Effect of KRAS Oncogene Substitutions on Protein Behavior: Implications for Signaling and Clinical Outcome.* Journal of the National Cancer Institute, 2012. **104**(3): p. 228-239 DOI: 10.1093/jnci/djr523.

226. E. Castellano and J. Downward, *RAS Interaction with PI3K: More Than Just Another Effector Pathway.* Genes Cancer, 2011. **2**(3): p. 261-74 DOI: 10.1177/1947601911408079.

227. G. Yin, S. Kistler, S.D. George, N. Kuhlmann, L. Garvey, M. Huynh, R.K. Bagni, M. Lammers, C.J. Der, and S.L. Campbell, *A KRAS GTPase K104Q Mutant Retains Downstream Signaling by Offsetting Defects in Regulation.* J Biol Chem, 2017. **292**(11): p. 4446-4456 DOI: 10.1074/jbc.M116.762435.

228. S. Lu, H. Jang, J. Zhang, and R. Nussinov, *Inhibitors of Ras-SOS Interactions.* ChemMedChem, 2016. **11**(8): p. 814-21 DOI: 10.1002/cmdc.201500481.

229. J.J. Irwin, T. Sterling, M.M. Mysinger, E.S. Bolstad, and R.G. Coleman, *ZINC: a free tool to discover chemistry for biology.* J Chem Inf Model, 2012. **52**(7): p. 1757-68 DOI: 10.1021/ci3001277.

230. R. Thomsen and M.H. Christensen, *MolDock: A new technique for high-accuracy molecular docking.* Journal of Medicinal Chemistry, 2006. **49**(11): p. 3315-3321 DOI: 10.1021/jm051197e.

231. M. Keshtkar-Jahromi, J.H. Kuhn, I. Christova, S.B. Bradfute, P.B. Jahrling, and S. Bavari, *Crimean-Congo hemorrhagic fever: current and future prospects of vaccines and therapies.* Antiviral Res, 2011. **90**(2): p. 85-92 DOI: 10.1016/j.antiviral.2011.02.010.

232. Ö. Ergönül, *Crimean-Congo haemorrhagic fever.* The Lancet Infectious Diseases, 2006. **6**(4): p. 203-214 DOI: 10.1016/s1473-3099(06)70435-2.

233. L. Borio, T. Inglesby, C.J. Peters, A.L. Schmaljohn, J.M. Hughes, P.B. Jahrling, T. Ksiazek, K.M. Johnson, A. Meyerhoff, T. O'Toole, M.S. Ascher, J. Bartlett, J.G. Breman, J.E.M. Eitzen, M. Hamburg, J. Hauer, D.A. Henderson, R.T. Johnson, G. Kwik, M. Layton, S. Lillibridge, G.J. Nabel, M.T. Osterholm, T.M. Perl, P. Russell, K. Tonat, and B. for the Working Group on Civilian, *Hemorrhagic Fever Viruses as Biological Weapons.* Jama, 2002. **287**(18): p. 2391 DOI: 10.1001/jama.287.18.2391.

234. M. Zimatore, M. Danova, E. Vassallo, C. Porta, M.C. Macaluso, L. Schittone, A. Riccardi, and S. Palmeri, *Weekly taxanes in metastatic breast cancer (review).* Oncol Rep, 2002. **9**(5): p. 1047-52.

235. O. Ergonul, *Treatment of Crimean-Congo hemorrhagic fever.* Antiviral Res, 2008. **78**(1): p. 125-31 DOI: 10.1016/j.antiviral.2007.11.002.

236. L.B. Tari and J.H. Patel, *Systematic drug repurposing through text mining.* Methods Mol Biol, 2014. **1159**: p. 253-67 DOI: 10.1007/978-1-4939-0709-0_14.

237. J. Agniswamy, C.H. Shen, Y.F. Wang, A.K. Ghosh, K.V. Rao, C.X. Xu, J.M. Sayer, J.M. Louis, and I.T. Weber, *Extreme multidrug resistant HIV-1 protease with 20 mutations is resistant to novel protease inhibitors with P1'-pyrrolidinone or P2-tris-tetrahydrofuran.* J Med Chem, 2013. **56**(10): p. 4017-27 DOI: 10.1021/jm400231v.

238. M.H. Hou, S.Y. Lin, C.L. Liu, Y.M. Chang, J.C. Zhao, and S. Perlman, *Structural basis for the identification of the n-terminal domain of coronavirus nucleocapsid protein as an antiviral target.* Protein Science, 2015. **24**: p. 284-284.

239. K. Klumpp and T. Crepin, *Capsid proteins of enveloped viruses as antiviral drug targets.* Current Opinion in Virology, 2014. **5**: p. 63-71 DOI: 10.1016/j.coviro.2014.02.002.

240. C. Cianci, S.W. Gerritz, C. Deminie, and M. Krystal, *Influenza nucleoprotein: promising target for antiviral chemotherapy.* Antivir Chem Chemother, 2012. **23**(3): p. 77-91 DOI: 10.3851/IMP2235.

241. Y. Sun, Y. Guo, and Z. Lou, *A versatile building block: the structures and functions of negative-sense single-stranded RNA virus nucleocapsid proteins.* Protein Cell, 2012. **3**(12): p. 893-902 DOI: 10.1007/s13238-012-2087-5.

242. J.W. Rodgers, Q. Zhou, T.J. Green, J.N. Barr, and M. Luo, *Purification, crystallization and preliminary X-ray crystallographic analysis of the nucleocapsid protein of Bunyamwera virus.* Acta Crystallogr Sect F Struct Biol Cryst Commun, 2006. **62**(Pt 4): p. 361-4 DOI: 10.1107/S1744309106006397.

243. Y. Guo, W. Wang, W. Ji, M. Deng, Y. Sun, H. Zhou, C. Yang, F. Deng, H. Wang, Z. Hu, Z. Lou, and Z. Rao, *Crimean-Congo hemorrhagic fever virus nucleoprotein reveals endonuclease activity in bunyaviruses.* Proc Natl Acad Sci U S A, 2012. **109**(13): p. 5046-51 DOI: 10.1073/pnas.1200808109.

244. T.T. Ashburn and K.B. Thor, *Drug repositioning: identifying and developing new uses for existing drugs.* Nat Rev Drug Discov, 2004. **3**(8): p. 673-83 DOI: 10.1038/nrd1468.

245. S.D. Carter, R. Surtees, C.T. Walter, A. Ariza, E. Bergeron, S.T. Nichol, J.A. Hiscox, T.A. Edwards, and J.N. Barr, *Structure, function, and evolution of the Crimean-Congo hemorrhagic fever virus nucleocapsid protein.* J Virol, 2012. **86**(20): p. 10914-23 DOI: 10.1128/JVI.01555-12.

246. N.R. Coordinators, *Database Resources of the National Center for Biotechnology Information.* Nucleic Acids Research, 2017. **45**(D1): p. D12-D17 DOI: 10.1093/nar/gkw1071.

247. V. Law, C. Knox, Y. Djoumbou, T. Jewison, A.C. Guo, Y. Liu, A. Maciejewski, D. Arndt, M. Wilson, V. Neveu, A. Tang, G. Gabriel, C. Ly, S. Adamjee, Z.T. Dame, B. Han, Y. Zhou, and D.S. Wishart, *DrugBank 4.0: shedding new light on drug metabolism.* Nucleic Acids Research, 2014. **42**(Database issue): p. D1091-7 DOI: 10.1093/nar/gkt1068.

248. T. Fraczek, A. Paneth, R. Kaminski, A. Krakowiak, and P. Paneth, *Searching for novel scaffold of triazole non-nucleoside inhibitors of HIV-1 reverse transcriptase.* J Enzyme Inhib Med Chem, 2016. **31**(3): p. 481-9 DOI: 10.3109/14756366.2015.1039531.

249. W.L. Jorgensen, *Azoles and related derivatives as non-nucleoside reverse transcriptase inhibitors (NNRTIs) in antiviral therapy (HIV)*. 2013, Google Patents.

250. A.H. El-Sagheer, A.P. Sanzone, R. Gao, A. Tavassoli, and T. Brown, *Biocompatible artificial DNA linker that is read through by DNA polymerases and is functional in Escherichia coli.* Proc Natl Acad Sci U S A, 2011. **108**(28): p. 11338-43 DOI: 10.1073/pnas.1101519108.

251. G.W. Crabtree and J.F. Henderson, *Rate-limiting steps in the interconversion of purine ribonucleotides in Ehrlich ascites tumor cells in vitro.* Cancer Res, 1971. **31**(7): p. 985-91.

252. R.C. Jackson, G. Weber, and H.P. Morris, *IMP dehydrogenase, an enzyme linked with proliferation and malignancy.* Nature, 1975. **256**(5515): p. 331-3.

253. D.S. Risal, M.D.; Goldstein, B.M., *Crystal structure of human inosine monophosphate dehydrogenase type II complexed with the MPA/NAD analog C2-MAD.* To be published.

254. D.S. Risal, M.D.; Goldstein, B.M., *The conformation of NAD bound to human inosine monophosphate dehydrogenase type II.* To be published.

255. Y.J. Zhang, L. Li, X. Liu, S.S. Dong, W.M. Wang, T. Huo, Y. Guo, Z.H. Rao, and C. Yang, *Crystal structure of Junin virus nucleoprotein.* Journal of General Virology, 2013. **94**: p. 2175-2183 DOI: 10.1099/vir.0.055053-0.

256. E.F. Pettersen, T.D. Goddard, C.C. Huang, G.S. Couch, D.M. Greenblatt, E.C. Meng, and T.E. Ferrin, *UCSF Chimera--a visualization system for exploratory research and analysis.* J Comput Chem, 2004. **25**(13): p. 1605-12 DOI: 10.1002/jcc.20084.

257. M.K. Fenwick, Y. Li, P. Cresswell, Y. Modis, and S.E. Ealick, *Structural studies of viperin, an antiviral radical SAM enzyme.* Proc Natl Acad Sci U S A, 2017. **114**(26): p. 6806-6811 DOI: 10.1073/pnas.1705402114.

258. S.R. Comeau, D.W. Gatchell, S. Vajda, and C.J. Camacho, *ClusPro: an automated docking and discrimination method for the prediction of protein complexes.* Bioinformatics, 2004. **20**(1): p. 45-50.

259. B.G. Pierce, K. Wiehe, H. Hwang, B.H. Kim, T. Vreven, and Z. Weng, *ZDOCK server: interactive docking prediction of protein-protein complexes and symmetric multimers.* Bioinformatics, 2014. **30**(12): p. 1771-3 DOI: 10.1093/bioinformatics/btu097.

260. M.A. Tortorici, P.D. Ghiringhelli, M.E. Lozano, C.G. Albarino, and V. Romanowski, *Zinc-binding properties of Junin virus nucleocapsid protein.* Journal of General Virology, 2001. **82**(Pt 1): p. 121-8 DOI: 10.1099/0022-1317-82-1-121.

261. G. Shaveta, J.H. Shi, V.T.K. Chow, and J.X. Song, *Structural characterization reveals that viperin is a radical S-adenosyl-L-methionine (SAM) enzyme.* Biochemical and Biophysical Research Communications, 2010. **391**(3): p. 1390-1395 DOI: 10.1016/j.bbrc.2009.12.070.

262. D.D. Pinschewer, M. Perez, and J.C. de la Torre, *Role of the virus nucleoprotein in the regulation of lymphocytic choriomeningitis virus transcription and RNA replication.* J Virol, 2003. **77**(6): p. 3882-7.

263. F. Raimondi, G. Portella, M. Orozco, and F. Fanelli, *Nucleotide Binding Switches the Information Flow in Ras GTPases.* Plos Computational Biology, 2011. **7**(3) DOI: ARTN e1001098

10.1371/journal.pcbi.1001098.

264. J.F. Diaz, B. Wroblowski, and Y. Engelborghs, *MOLECULAR-DYNAMICS SIMULATION OF THE SOLUTION STRUCTURES OF HA-RAS-P21 GDP AND GTP COMPLEXES - FLEXIBILITY, POSSIBLE HINGES, AND LEVERS OF THE CONFORMATIONAL TRANSITION.* Biochemistry, 1995. **34**(37): p. 12038-12047 DOI: 10.1021/bi00037a047.

265. P. Prakash, A. Sayyed-Ahmad, and A.A. Gorfe, *The Role of Conserved Waters in Conformational Transitions of Q61H K-ras.* Plos Computational Biology, 2012. **8**(2): p. e1002394 DOI: 10.1371/journal.pcbi.1002394.

266. L. Zhang, S. Bouguet-Bonnet, and M. Buck, *Combining NMR and Molecular Dynamics Studies for Insights into the Allostery of Small GTPase–Protein Interactions.* Methods in molecular biology (Clifton, N.J.), 2012. **796**: p. 235-259 DOI: 10.1007/978-1-61779-334-9_13.

267. G. Buhrman, C. O'Connor, B. Zerbe, B.M. Kearney, R. Napoleon, E.A. Kovrigina, S. Vajda, D. Kozakov, E.L. Kovrigin, and C. Mattos, *Analysis of Binding Site Hot Spots on the Surface of Ras GTPase.* Journal of Molecular Biology, 2011. **413**(4): p. 773-789 DOI: 10.1016/j.jmb.2011.09.011.

268. P. Prakash and A.A. Gorfe, *Lessons from computer simulations of Ras proteins in solution and in membrane.* Biochimica et biophysica acta, 2013. **1830**(11): p. 10.1016/j.bbagen.2013.07.024 DOI: 10.1016/j.bbagen.2013.07.024.

269. B.J. Grant, S. Lukman, H.J. Hocker, J. Sayyah, J.H. Brown, J.A. McCammon, and A.A. Gorfe, *Novel Allosteric Sites on Ras for Lead Generation.* Plos One, 2011. **6**(10): p. e25711 DOI: 10.1371/journal.pone.0025711.

270. M.M. Edreira, S. Li, D. Hochbaum, S. Wong, A.A. Gorfe, F. Ribeiro-Neto, V.L. Woods, and D.L. Altschuler, *Phosphorylation-induced Conformational Changes in Rap1b: ALLOSTERIC EFFECTS ON SWITCH DOMAINS AND EFFECTOR LOOP.*

The Journal of Biological Chemistry, 2009. **284**(40): p. 27480-27486 DOI: 10.1074/jbc.M109.011312.

271. A. Banerjee, H. Jang, R. Nussinov, and V. Gaponenko, *The disordered hypervariable region and the folded catalytic domain of oncogenic K-Ras4B partner in phospholipid binding.* Curr Opin Struct Biol, 2016. **36**: p. 10-7 DOI: 10.1016/j.sbi.2015.11.010.

272. B.J. Grant, J.A. McCammon, and A.A. Gorfe, *Conformational Selection in G-Proteins Lessons from Ras and Rho.* Biophysical Journal, 2010. **99**(11): p. L87-L89 DOI: 10.1016/j.bpj.2010.10.020.

273. M.B. Mazzucco, L.B. Talarico, S. Vatansever, A.C. Carro, M.L. Fascio, N.B. D'Accorso, C.C. Garcia, and E.B. Damonte, *Antiviral activity of an N-allyl acridone against dengue virus.* J Biomed Sci, 2015. **22**: p. 29 DOI: 10.1186/s12929-015-0134-2.

VITA

SEZEN VATANSEVER
Koc University, School of Medicine,
Rumelifeneri Yolu, 34450, Sarıyer, Istanbul, Turkey
sezenvatansever@gmail.com +1 536 675 9323

EDUCATION:

PhD., Koç University, Institute of Health Sciences, Istanbul- TURKEY Biomedical Sciences and Engineering	2013 – 2017 (Expected)
Associate Researcher, Mount Sinai Medical School, Institute of Genetics and Genomic Sciences, NYC, NY- USA	2016 - 2017
M.D., Ankara University, Faculty of Medicine, Ankara – TURKEY Medicine	2004 – 2011

WORK EXPERIENCE:

Research & Teaching Assistant Koç University, Graduate School of Sciences and Engineering, Istanbul- TURKEY	2013 – 2016
Ministry of Health Şarköy TSM, Physician, Tekirdağ – TURKEY Worked as a general practitioner in various positions (i.e. district health director, emergency care, family medicine, occupational health and safety)	2011 – 2013
Seconda Universita di Napoli, Medical Intern, ITALY	04.2011 – 08.2011
Ankara University, Medical Intern, TURKEY	05.2010 - 04.2011

TEACHING EXPERIENCE:

Drug Design, Thermodynamics, Public Health, Organic Chemistry for Health Sciences, Anatomy

CERTIFICATES:

IEEE Photoshop Training Certificate /2015
T.C SAĞLIK BAKANLIĞI Certificate of Turkish Sign Language /2013
T.C. ÇSGB General Directory of Occupational Health and Safety Occupational Physician Certificate /2012

PUBLICATIONS:

Vatansever S., Gumus Z. and Erman B., Intrinsic K-Ras dynamics: A novel molecular dynamics data analysis method shows causality between residue pairs. Nature Scientific Reports, 2016. 6: p. 37012

Mazzucco M.B., Talarico L.B., Vatansever S., et. al, Antiviral activity of an N-allyl acridone against dengue virus. Journal of Biomedical Sciences, 2015. 22: p. 29