UNCOVERING CRYPTIC POCKETS IN BIOLOGICALLY RELEVANT PROTEINS: AN IMPROVED COMPUTATIONAL METHODOLOGY

by

ABHISHEK KUMAR SONI
(20121084)

A MS thesis

submitted in partial fulfillment of the requirement
for the award of the dual degree of

Bachelor of Science-Master of Science

in

CHEMISTRY

DEPARTMENT OF CHEMISTRY
INDIAN INSTITUTE OF SCIENCE EDUCATION AND RESEARCH PUNE
PUNE, MAHARASHTRA – 411008

March 2017

CERTIFICATE

This is to certify that this dissertation entitled **"Uncovering Cryptic Pockets in Biologically Relevant Proteins: A Novel Methodology"** towards the partial fulfilment of the BS-MS dual degree programme at the Indian Institute of Science Education and Research, Pune represents the research carried out by **"Abhishek Kumar Soni"** at "**Bioinformatics Institute (BII), A*STAR/National University of Singapore (NUS)**" under the supervision of **"Dr. Chandra Verma (HOD), Dr. Peter J Bond (PI) of BII, A*STAR and Prof. Paul Matsudaira (HOD) of NUS"** during the academic year 2016-2017.

Abhishek Kumar Soni

Soni

Committee Member	Signature	Date
PETER BOND		15/03/17
CHANDRA VERMA		15/3/17
PAUL MATSUDAIRA		29/3/17

DECLARATION

I hereby declare that the matter embodied in the report entitled **"Uncovering Cryptic Pockets in Biologically Relevant Proteins: A Novel Methodology "** are the results of the investigations carried out by me at **Bioinformatics Institute (BII), A*STAR/National University of Singapore (NUS)**" under the supervision of **"Dr. Chandra Verma (HOD), and Dr. Peter J Bond (PI) of BII, A*STAR and Prof. Paul Matsudaira (HOD) of NUS** and the same has not been submitted elsewhere for any other degree.

Abhishek Kumar Soni

Soni

Committee Member	Signature	Date
PETER BOND		15/03/17
CHANDRA VERMA		15/3/'17
PAUL MATSUDAIRA		29/05/2017

ACKNOWLEDGEMENT

My journey in writing this MS thesis has been one of the most challenging process in my life and obviously it would not have finished without supports from many people that sincerely helped me in this “academic marathon”. When I let my thought drift back to a year ago, I seem to recall the selection process that I was determined to adhere to as I chose potential MS research project. Then my senses were flooded with the stunning psychedelic pictures splashed about the A*STAR and NUS, Singapore. From ties to jive, the diverse personalities of the members of this group are reflected in the information that they pass on to innocent bystanders like myself through the internet. The entire group has been a friendly and inspiring bunch, and I will very much miss the lab parties, informative group meetings, and, of course, the biology and chemistry.

In particular, I would like to thank my supervisors Dr. Chandra Verma and Dr. Peter Bond, for their suggestions and continuous encouragement as well as Dr. Jan Marzinek, for his thoughtfulness and insight, persistent motivation and for countless nitty-gritty discussions (without which I would remain lost in the quagmire of debugging). The support was so intense that I could be able to work on two projects in 10 months with a manuscript already in the stage of submission. I would like to thank A*STAR for providing amazing computing.

I would also like to express my deepest gratitude to Prof. Paul Matsudaira of NUS, Singapore for the one-on-one useful discussions and also because my research was kindly funded by him. I would like to thank Dr. Roland Huber for providing assistance to solve technical issues and provide scientific suggestions. I also acknowledge Kristy Y W Chang for guiding me through all the initial scientific procedures in my MS project.

I thank Dr. Arun Venkatnathan, Associate Professor, IISER Pune who kept me motivated towards my Master’s thesis project with his constant feedback and played an important part in making this possible. I thank IISER Pune for giving me an opportunity to work at Singapore and everyone who supported me to make this research a success.

Last but not the least, I fully dedicate this MS thesis to my beloved family, my dad, mom, and my brother, that have peacefully and happily coexisted at all the time with me. Never ending gratitude for always there for me. To my whole family: “It’s been emotional and I love you”.

CONTENTS

LIST OF FIGURES (F) AND TABLES (T)

ABSTRACT

Revealing and targeting novel cryptic pockets in a protein has become a major challenge in the field of drug design. Static experimental protein structures represent ensemble averaged lacking the dynamics of the real biological system. Molecular dynamics (MD) simulations serve as a “virtual microscope” to probe the dynamics of the biological structures based on Newton’s equations of motion. In this thesis I have demonstrated an improved simulation based technique that uses benzene molecules as probes and enables revealing cryptic pockets hidden in the experimental structures of proteins. In this method benzene rings contain additional dummy atom in its geometric center with an optimized Lennard-Jones potential. Initially the method has been benchmarked against experimentally well-defined wild-type and mutant K-Ras (G12D) protein, essential bio-substrate in development of many cancers. The application of the method resulted in the formation of “saddle” like benzene clusters around the pockets of protein. With the novel protocol I have shown that all pockets previously defined by experimental and *in silico* methods for both wild-type and mutant K-Ras have been observed with enlarged volume together with the exposure of a novel pocket. The method was shown to be efficient within only 100 ns of sampling. Subsequently the novel approach was applied on envelope (E) protein of all dengue virus (DENV) serotypes as well as Zika virus (ZIKV). The novel consensus pocket exposed was positioned in between two monomers and found in all dimers studied. This crucial region has a great potential for blocking dimer-to-trimer transition and viral infection process. Hence, it was a subject to subsequent virtual screening. Ten Food and Drug Administration (FDA) approved drugs (based on binding affinity and key interaction with both the monomers) have been shortlisted for the experimental validation. The most promising molecule named *Itraconazole* was shown to possess the highest affinity for ZIKV. In summary a general novel method for screening cryptic sites on protein structures which can be further used in drug design has been proposed.

1. INTRODUCTION

The identification of druggable sites on proteins has become an area of great interest in the search for new therapeutics[1]. Traditionally, sites where ligands are seen in crystal structures solved in water have been used as targets for screening in the search for new and more potent modulators. However, the observation of ligands bound to pockets that were occluded in apo states of the proteins suggested that cryptic pockets may be more prevalent than assumed. Careful analysis also showed that the binding of ligands resulted in the opening of pockets that were cryptic in the absence of ligands – this suggesting the possibility of allosteric modulators[2]. Various experimental and computational techniques were developed in order to locate such transient pockets which are discussed below[2–5].

1.1. Experimental and computational tools to locate cryptic pockets

Experimental tools include NMR based fragment screening and Hit validation of library of small molecules[5], fluorescence assays[6], mass spectrometry and structural analysis comprising of analysing protein crystal structures with the incorporation of side-chain flexibility[7]. Computational techniques include programs such as FTMap[4], SiteMap[8], MDpocket[9], GRID[10], and trj_cavity[11]. Methods such as FTMap and GRID consider only static experimental structures and hence are unable to detect transient sites which can be only characterized considering dynamics of the real biological system[12,13]. While classical molecular dynamics (MD) simulations did provide a picture of the flexible dynamic structures of proteins, nevertheless the accessible timescales seem insufficient to result in the exposure of cavities that can be deemed large enough to be druggable. Also most of the simulations of proteins performed in the water are unable to detect hydrophobic cryptic sites because these pockets are hidden in crystal structure which are solved in polar solvents. This resulted in the development of special methods to probe such cavities using hydrophobic molecules which include mixed-solvent MD and cosolvent-MD[7,14], site-identification by ligand competitive saturation (SILCS)[13], ligand mapping or probe based MD simulations (pMD)[7,12,14–16]. The last two methods introduce small probe molecules into protein-water systems prior to running MD simulations and were inspired by the fragment-based NMR[5] and multi solvent crystallographic (MSCS)[17] approaches. Commonly

used probe molecules included benzene[12], cyclohexane[4], and isopropanol[14], chosen based on their drug like features (for example, most of the drug contains the aromatic rings in their structure)[15]. Subsequently, the probe affinity maps are generated and compared with the starting experimental structure of the protein in order to examine the new pockets being cryptic in static experimental structures. This is followed by virtual screening of libraries of ligands based on the known structure of pockets[18].

1.2. The benchmark of new benzene mapping method: K-Ras protein

Mutant K-Ras protein is a cancerous protein and a major drug target that has been widely studied[19]. All of the druggable pockets are experimentally validated which was the motivation to choose this protein as benchmark testing of the novel method. Further aim would be to extend this protocol on envelope (E) protein serotypes of DENV and ZIKV to locate cryptic pockets in them. Wild-type K-Ras is a signalling protein that belongs to the family of GTPases which act as a molecular switch[19]. Somatic point mutations in K-Ras protein are implicated in~20 % of human cancers, making it an important drug target[20]. There is a rapid exchange between GTP bound (active) state to GDP bound (inactive) state in order to transmit signal[19]. The common point mutations such as G12D, Q61H and G12C lead to constitutive activation of the protein[20] converting it into oncogenic ones.

Structurally, the G-Domain (1-166) of K-Ras includes two lobes, namely lobe 1 (1-86) and lobe 2 (87-188)[21]. Lobe 1 is the GTP/GDP and effector binding region[19]. A significant conformational change in switch I (30-40) and switch II (58-76) regions of this lobe is promoted by the effector binding[21]. Lobe 2 is an allosteric binding region where binding of ligands alters the conformation of switch region which is responsible for functional specificity of Ras proteins[22]. Recently the main focus has been placed on locating cryptic pockets in various mutants of K-Ras and development of effective inhibitors that bind to vicinity of switch regions[22–25]. Previously, the research has been focussed on four experimentally well-characterized pockets named as P1-P4 that were found near the switch regions and the C-terminal region in the available K-Ras crystal structures (Figure 1.1 and Table 1.1). All experimentally validated 4 pockets (Table 1.1) were also found using *in silico* methods FTMap[18,26], ensemble docking[18]. More recently the pMD simulation method

Table 1.1. Experimentally and computationally revealed pockets details of K-Ras protein. The pMD method listed in the table refers to the work of Prakash *et al* (2015)[14]. All the pockets were discovered in our study and for all the methods employed, it is collectively named as letter "**P**".

Pockets	Location	Relevance	Computational Methods	Ligands bound
P1	Core β2-β3 sheet and outer switch II region	Inhibit Ras-SOS and Ras-Effector interaction	pMD, FTMap[26], Autodock[18]	Indole, DCAI, Kobe ligands, Benzimidazole[27].
P2	Between switch II and helix3	Control GTP binding affinity and disrupt Ras-effector interaction	FTMap[26], pMD	Cystine rich molecules bound to G12C[23], no covalent inhibitor found to bind to perform effective inhibition.
P3	Near C terminus of Helix 5 and loop7	Allosterically control the activity of switch I and stabilise the ras in inactive state.	FTMAP[26], Autoligand[18], Blind Dock[18], pMD	Metal Cyclens[22]
P4	Behind switch I, helix 1 and between nucleotide binding loop (covering the whole nucleotide binding region)	Disrupts the crucial Ras-Sos interaction and stabilise the Ras in inactive state (switch I moves away from GTP)	BlindDock[18], pMD	Andrographolide derivatives[25], Zn(II)-bis(2-picolyl) amines[24]
P5	Between helix 3 and helix 4	Binding of Ca^{2+} makes Q61 more ordered in active state and hence increment in GTP hydrolysis	pMD, FTMAP[18,26]	Small organic molecules such as dimethylformamide and glycerols (FTMap Study)[24,26]
P6	Near p3 and between β sheets and helix 5 and Loop3	Disrupts the allosteric communication between residues of lobe 1 and lobe 2 which alters the orientation of G-lobe relative to membrane,	Autoligand[18], pMD (only cluster of probe interaction spots, not captured by MDpocket, identified as sub-site)	No effective inhibitors till now found to bind.
P7	Between upper part of helix, switch I and β sheets	Mutations in some residues of the pocket disrupts the communication between lobe1 and lobe 2	No computational method	No effective inhibitors till now found to bind.
P8	Between helix 4 and helix 5	Disrupts the interaction between helix4 and switch I which may alter the orientation of G-domain with respect to membrane.	MD[24], pMD)	No effective inhibitors till now found to bind

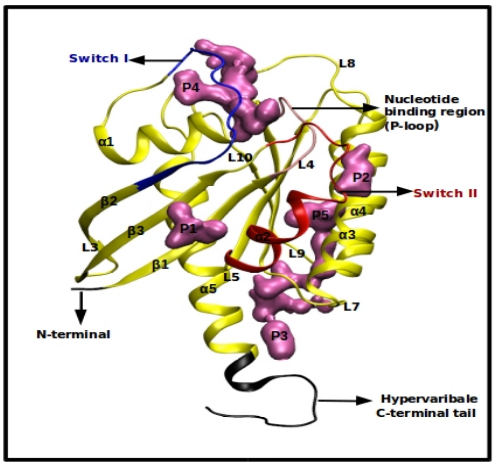


Figure 1.1. Pockets analysis on the crystal structure performed in this work**.** The structure of K-Ras protein (protein data bank entry: 4DSO[28]) showing the switch regions (cartoon representation in blue and red) and its important loops (L). The four experimentally tested pockets (P1-P4) and additional *in silico* defined P5 are shown in pink surface representation (located by MDpocket[5]).

where isopropanol was used as a probe molecule was employed to map the cryptic pockets in G12D mutant K-Ras[14]. They explored the following pockets: P1 (in fragments), P2 (opened only in < 50 % of simulation time, not observed by MDpocket), P3 and P4 (small subsites). However, P5-P8 except P7 have not been captured by MD pocket and were only confirmed by smaller probe-density clusters. P1-P5 have also been observed in the crystal structure of G12D mutant K-Ras (Figure 1). Pocket P5-P8 were identified as low consensus sites since they have been only found in a single study via FTMap or MSCS technique[26].

1.3. DENV and ZIKV as a drug target

The DENV and ZIKV flaviviruses are affecting hundreds of million people a year and result in tens thousands of deaths[29]. Currently there is no approved therapeutic against ZIKV or highly effective vaccine against all DENV serotypes (present day vaccine could be used by 9-45 years old only[30]). Therefore, it is of utmost importance and great interest to study the pathology of these viruses in order to develop highly effective therapeutics.

Mature flavivirus particle is ~50 nm in diameter with its outer surface covered with the 90 envelope (E) and membrane (M) protein dimers arranged in the icosahedral symmetry being embedded in the lipid bilayer (Figure 1.2). Inside the virus, there is a capsid (C) protein together with the RNA genome. Each of 180 monomers consists

of three domains - I, II, and III[31]. During the receptor mediated endocytosis, virus is internalized in the endosome and E protein undergoes a pH driven conformational change and dimer-to-trimer transition (fusogenic state)[31]. This leads to the exposure of fusion peptides (FPs) of the domain II of E protein which initiates the fusion of the virus and host cell membranes. There are four DENV serotypes which exhibit 60-80 % E proteins homology and a high sequence identity with ZIKV (>~45% with DENV1)[31,32].

Various computational and experimental studies have been performed to understand the conformational changes of E proteins as well as to reveal the druggable cryptic pockets on the viral envelope[31–34]. In particular, the latter could potentially help in designing small molecules which can efficiently bind and prevent the dimer-to-trimer transition, blocking the viral infection process. It has been studied that during the transition of the virus envelope, domain II rotates about a hinge at the domain I-II interface[31]. Modis *et al.* found a druggable, hydrophobic pocket occupied by the detergent *n*-octyl-β-D-glucose (β-OG) around this hinge (Figure 1.2-A) suggesting that binding a compound to this pocket might inhibit the rotation of Domain II and prevent the virus from fusing and infecting the host cell[35]. With the (β-OG) pocket as a target, most of the ligand-protein docking studies using small molecules[33,35–37] and linear peptides[38] were consequently conducted and a few compounds were found to experimentally inhibit viral activity such as flavionoids[37] and phytocompounds from methanol leaf[39]. However, the drug approval process for novel compounds is time consuming and expensive[40]. Many lead compounds are filtered out at the preclinical and clinical trial stages due to problems with efficacy and toxicity[40]. It is thus unsurprising that so far, there has been little to no updates on the development of these potential DENV and ZIKV inhibitors. A few lead compounds have also already been filtered out due to poor pharmocokinetic properties[36].

An alternative approach to drug discovery that will save time and cost is drug repurposing. Drug repurposing utilizes drugs approved by the U.S. Food and Drug Administration (FDA) to treat diseases or conditions other than its intended target[40]. The pharmacological and toxicity profiles of these drugs are known, thus drastically shortening the cost and approval process for new indications[40]. Current motivation of

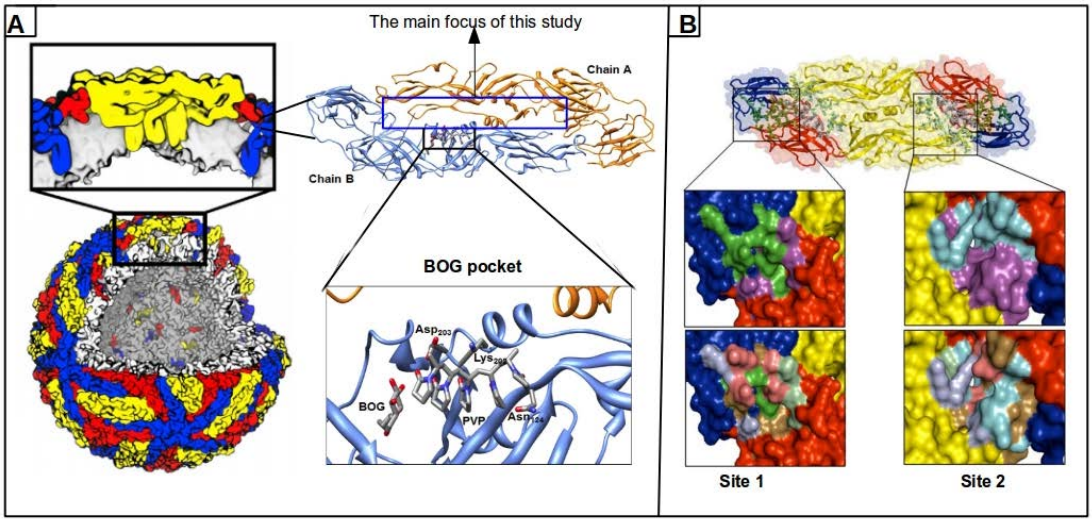


Figure 1.2. (A) Onion like structure of DENV showing the envelope protein containing two momomeric chains[32]. The most studied BOG (β-OG) pocket[35] in domain II of E protein is shown in cartoon representation[32]. The pockets between the interface of two chains is the main focus of this study **(B)** The two sites at the interface studied by Yenamalli *et al*[33]. Site 2 has not been targeted and is near the fusion peptide which is also the focus of this study (discussed later). Three domains are also shown with domain I (red), domain II (yellow) and domain III (blue) in cartoon representation.

this study lied in the work of Yennamalli *et al*[33] where they have targeted a site present at intradimeric region of E protein crystal structure of DENV2 and DENV3 (site 1) and performed virtual screening of FDA approved drug, experimental assays and predicted that it could interfere with the fusion process (Figure 1.2-B).

1.4 Purpose of the Study

The following computational protocol has been proposed to identify FDA-approved drugs for the inhibition of the flaviviruses E protein dimer-to-trimer transition:

- Mapping experimentally known and new druggable sites in wild-type and mutant K-Ras protein via sets of novel benzene mapping MD simulation methodology.
- Assessment of each method efficiency of in revealing the cryptic pockets.
- Finding the uncommon residues of wild-type and mutant K-Ras protein which can help designing novel phrmacophores that can selectively target the mutant.
- Application of the most efficient method on ZIKV and all serotypes of DENV.
- Perform virtual screening of library of FDA approved drug compound on the intradimeric pocket exposed in benzene mapping MD simulation.
- Searching for the set of common ligands against all the serotypes.

2. COMPUTATIONAL METHODOLOGY

2.1. Molecular Dynamics (MD) Simulations

Atomistic molecular dynamics simulation (also referred as “computational microscope”) based on Newton’s laws of motions gives the ability to look at the systems at an atomistic level and can be used to study the interactions within the molecules, self-assembly process of polymers and timescales of protein conformational changes[41]. MD simulation can provide a useful insight into the role of active site components of a protein and effect of mutations at those sites in their functioning[14,32,41]. The recent advances in computational resources make it possible to simulate microsecond time scales of systems of millions of atoms[32]. The atoms in the system are assigned random velocity via Maxwell-Boltzmann distribution. The potential (bonded and non-bonded) between the particles are dependent upon position. Thus, a new position will give rise to force on the particle (negative gradient of potential energy) which in turn produce acceleration in the system evolving the system at new position and velocity governed by classical Newton’s equation of motion. The accuracy of trajectory of particles is completely determined by the force-fields that comprised of bonded and non-bonded potential terms as follows:

$$\mathbf{V_{total}}(r) = V_{bonded}(r) + V_{non\text{-}bonded}(r) \quad (1.1)$$

$$\mathbf{V_{total}} = V_{bonds} + V_{angles} + V_{dihedrals} + V_{electrostatic} + V_{Van\ der\ Waals} \quad (1.2)$$

In this work, Amber-99SB-ILDN-Q force field[42] was used for both K-Ras and E protein. This force field includes the corrected side chain torsional potentials and refitting of residue charges. The expressions for the force-field potentials can be found elsewhere[43]. The van der Waals potentials are also termed as “Lennard-Jones” potentials. All the MD simulations have been carried out by GROMACS (GRonigen Machine for chemical simulations)[43]. GROMACS is a simulation package designed for simulating proteins, nucleic acids, lipids and soft matters.

2.2. Computational Docking and protein flexibility

Computational docking is an important tool in structure-based drug discovery[44]. It enables us to predict the three-dimensional orientation of a ligand in a binding site[45],

study ligand-protein interactions and find the most energetically favorable protein-ligand complex within a huge database of ligands via virtual screening[45]. Compared with standard experimental methods, computational docking is less time-consuming and is more cost-effective. To understand the rationale for docking the known ligand binders, a general understanding of scoring is needed. Scoring refers to the prediction of binding affinity between a docked ligand and receptor (score) via a mathematical method (scoring function)[46,47]. All the ligand hits are then ranked based on the scores of their top docking poses[47]. As a ligand will be docked to a protein regardless of its binding affinity, a score threshold is required to remove false positives from the database. Hence, the poorest score among the docked known ligand binders was set as the score threshold.

The Autodock Vina[48] is a fast docking software whose scoring function comprehensively accounts for both intermolecular and intramolecular interactions. Furthermore, the performance of Autodock Vina has been tuned using a database of binding affinities for experimentally determined protein-ligand complexes. However, no matter how robust a scoring function might seem to be, the accuracy of a scoring function is highly dependent on the target protein[49].

For docking to yield reasonably accurate results, it is important to consider protein flexibility in the input protein structure[50]. 85% of proteins in the PDB have four to seven flexible residues in the binding site that undergo side-chain rearrangements due to ligand binding. Furthermore, side-chain flexibility has led to more accurate results for the VEGFR-2 protein (vascular endothelial growth factor receptor 2) using Autodock Vina[51]. Hence, it is unsurprising that the ligand orientation is incorrectly predicted in 50-70% of cases when only one static protein structure is used for docking[52]. In contrast, when multiple protein conformations generated from clustering of MD trajectories are used, the precision of results in predicting ligand orientation generally improves[52]. However, to dock ligands to millions of protein conformations is computationally expensive[40]. In order to reduce the number of proteins and ensure that the conformations chosen maximally represent the entire conformational space sampled by the E-protein, one needs to perform RMSD-based cluster analysis[53]. All conformations were divided into groups of structurally-similar protein conformations (clusters), and the remaining unique conformations were deemed as noise.

2.3 Benzene Parameters

The partial charges for the benzene (C_6H_6) were taken from Tan *et al*[7] with -0.126e charge on the carbon atom and 0.126e charge on the hydrogen atom. In case of benzene with exclusions (K-Ras^{BEX} in Table 2.1), the non-bonded parameters, namely van der Waals and electrostatic interactions between the benzene molecules were switched off (their interactions with protein, explicit water and ions remained unchanged). Further modifications involved the introduction of a dummy atom with zero mass and charge placed at the centre of mass/geometry of each benzene molecule and was covalently bonded (simulated by using the LINCS[54] algorithm for constraining the distance to 1.4 Å) to three carbon atoms. The dummy atom had a single repulsive Lennard-Jones potential for interactions with the remaining dummy atoms of other benzene molecules only (its interactions with non-dummy atoms were switched off). ϵ of 0.002 kcal mol^{-1} was chosen (Table 2.1) to prevent any attractive interactions between the dummy atoms. Two values of σ were tested: 4.5 Å and 5.5 Å. Lennard-Jones potentials for both are shown in Figure 3.1.

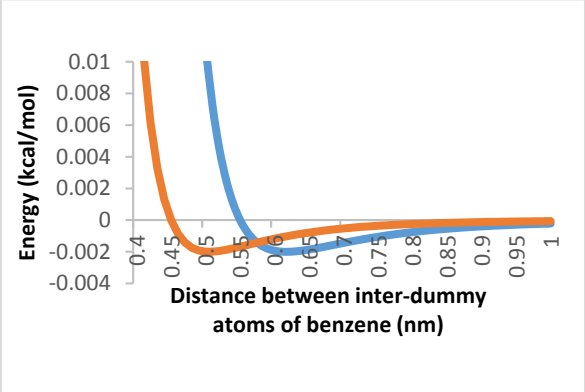


Figure 2.1. Lennard-Jones potential with respect to dummy atom – dummy atom distance with two sigma values i) 4.5 Å (red) ii) 5.5 Å (blue). The epsilon value was kept constant (0.002 kcal mol^{-1}) for the dummy atom in both the curves.

2.4. System Preparation for K-Ras protein

The initial starting structure of K-Ras (G12D) bound with GTP was obtained from RCSB protein data bank (PDB, ID: 4DSO[28], Resolution: 1.85 Å).The missing C-terminal lysine residues were added using Pymol [55] to complete the 180 amino acid long protein structure (sequence reference from PDB data bank[28]). Wild-type K-Ras was modelled by applying single point mutation (D12G) using Pymol. For simulations involving benzene molecules, 58 molecules were randomly placed around the protein (0.2 M concentration), while this number corresponded to 116 molecules at

0.4 M concentration. Each system was solvated with approximately 15,000 TIP3P[56] water molecules. The system was then neutralized with 2 and 1 chloride ions for wild-type and mutant respectively. The Amber-99SB-ILDN-Q force field[42] was used for protein simulations and the generalized AMBER force field[57] for the benzene probes. The force field for GTP was adapted from the work of Meagher *et al*[58]. All simulations were performed using GROMACS 5.1.1[43]. Initially each system was energy minimized via the steepest descent method[59]. Subsequently, the system was equilibrated in isothermal-isochoric (NVT) and isothermal-isobaric (NPT) ensembles for 2 ns each with position restraints on protein and benzene's heavy atoms (force constant of 1000 kJ mol^{-1} nm^{-2}). Unrestrained production runs were performed in the NPT ensemble for 300 ns. Equations of motion were integrated via the Verlet leap frog algorithm[60] with a 2 fs time step. The trajectory was saved every 10 ps for analysis. The cut-off was set at 12 Å for the short-range neighbour list and van der Waals interactions modelled by 12-6 Lennard-Jones potential (L-J). The particle mesh Ewald method[61] was applied for long range electrostatics with a 12 Å real-space cut-off. The velocity rescale thermostat with an additional stochastic term[62]

Table 2.1. The pMD simulation systems employed for both wild-type and mutant (G12D) K-Ras protein. Each simulation was carried out for 300 ns without restraints. The two Lennard-Jones sigma (σ) values used were **i)** 4.5 Å (marked by *) and **ii)** 5.5 Å. Sigma values control the distance of closest approach of two benzene molecules. The ϵ had the well depth ~0 kcal mol^{-1} so that no attractive interactions between the dummy atoms were present. The pockets are labelled as single primes and double primes for mutant and wild-type respectively since the residues surrounding the respective pockets in both were not all same.

Methods	Name	Concentration	σ of dummy atom (Å)	ϵ of dummy atom (kcal mol^{-1})	Pocket naming
K-Ras +water(classical MD)	$K\text{-}Ras^{W}$	-	-	-	w'/w"
K-Ras + water + Benzene	$K\text{-}Ras^{B0.2}$	0.2 M	-	-	b'/b"
	$K\text{-}Ras^{B0.4}$	0.4 M			B'/B"
K-Ras + water + Benzene with exclusions	$K\text{-}Ras^{BEX\ 0.2}$	0.2 M	-	-	e'/e"
	$K\text{-}Ras^{BEX0.4}$	0.4 M			E'/E"
K-Ras + water + Benzene with exclusions & dummy*	$K\text{-}Ras^{BEXD0.2*}$	0.2 M	4.5	0.002	d*'/d*"
	$K\text{-}Ras^{BEXD0.4*}$	0.4 M	4.5	0.002	D*'/D*"
K-Ras + water + Benzene with exclusions & dummy	$K\text{-}Ras^{BEXD0.2}$	0.2 M	5.5	0.002	d'/d"
	$K\text{-}Ras^{BEXD0.4}$	0.4 M	5.5	0.002	D'/D"

was maintained at a temperature of 310 K with a collision frequency of 10 ps^{-1}.The Parrinello-Rahman barostat[63] was used for the production runs, maintaining the pressure at 1 bar. Initial velocities were set according to Maxwell distribution at 310 K. Periodic boundary conditions were applied in all directions. The constraints on all the bond lengths involving hydrogen atoms were imposed via the LINCS algorithm[54].

2.5. System Preparation for DENV and ZIKV Envelope (E) protein

The initial starting structure of mature E protein dimer of DENV2 was obtained from RCSB protein data bank (PDB, ID: 3J27[28] , Resolution: 3.5 Å). Other structures were homology modelled based on the amino acid sequence since there were no full E protein structure available at the time of starting of the project. The same simulation setup was applied as for the K-Ras protein (section 2.4). The benzene molecules selected had dummy atom with sigma value 4.5 Å and epsilon 0.002 kcal mol^{-1}.

2.6. Pocket Analyses protocol

The analysis of the MD trajectories for the presence of pockets on the surface of K-Ras was performed using MDpocket[9]. Given various conformations of proteins generated from MD as an input, MDpocket is a robust and free open source tool which can easily detect large scale tunnel like protein binding cavities in them. Approximately 8000 conformations of the proteins from each simulation were subject to pocket analysis; the conformations were chosen after the total and hydrophobic SASA of the protein reached a plateau. An optimum iso-value of 2.9 was chosen which controls the size (mesh density) of the pockets in VMD[64]. All residues belonging to a given pocket were extracted based on a particular cut-off radius (2-3.5 Å) within the pocket which ensures the inclusion of all the residues surrounding the pockets despite with the variability in their sizes. The radius of the pocket mesh generated by MDpocket was kept constant for each pocket in both wild-type and mutant protein. Only pockets larger than 30 atoms were considered ensuring a significant pocket volume (>= 70 $Å^3$). Pocket meshes were splitted into individual pockets with explicitly defined residues, the volume and total SASA of each pocket was calculated for the 8000 frames (before and after 100 ns) via MDpocket and then averaged. The evolution of the volumes of the pockets with simulation time for the full trajectory have been recapitulated. For the flaviviruses, the same protocol was repeated.

2.7. Structural Analysis

Structural characterizations were carried out by examining the structural deviations (RMSD) and fluctuations (RMSF) for the protein backbone heavy atoms and Cα atoms respectively. SASA and RMSD based clustering was performed on C-α atoms for the full protein as well as for given individual pockets using Gromos method[65] with an RMSD cut-off distance of 1 Å for the full K-Ras protein. The C-terminus of K-Ras (168-180) was not included in the clustering due to its high structural fluctuations during the simulations; this region is located far from the cryptic pockets of the interest and therefore was not found to be important for the regions of interest, while for the flaviviruses full E protein was included. Results from the clustering analysis on the C-alpha atoms of the protein were used to align the generated pockets on to the most persistent confirmation of the protein. This could clearly suggest which regions of protein are important for pocket opening. For E protein of all the serotypes, top 3 cluster representatives along with the native crystal structure have been subjected to virtual screening. Per-residue propensity secondary structure assessment and time dependent secondary structure analysis were performed using DSSP[66]. All structural analysis were performed using the GROMACS 5.1.1 package. Trajectories were visualised via VMD.

2.8. Ligand and Receptor preparation and Virtual Screening

Input files in Autodock Vina[48] are required to be in pdbqt format, a modified protein data bank (pdb) format that has atomic charges. Other common molecular file formats used include mol2, sdf and pdb. 1985 FDA-approved molecules containing polar hydrogens were downloaded in pdbqt format from the ZINC database[67]. Gasteiger partial atomic charges were added to ligands in pdbqt format. All the rotatable bonds were set to be rotatable. Similarly, receptors (E protein) have been converted to pdbqt format with the non-polar hydrogens merged and polar hydrogens added according to neutral pH. All the docking runs were performed through Autodock Vina suite[48]. Pymol plugin of Vina was used to inspect the docked structures of ligand on each of the DENV and ZIKV protein clusters across all the serotypes obtained from pMD simulations. The binding site's x-y-z coordinates were determined by making a cubical box of 29 Å covering the whole binding site with center of box similar across all the serotypes. Mostly residues from 242-251 in the binding site (approximately four to five residues) were selected to be flexible

because of their conserved nature across all the serotypes. The pocket is located near those residues. The protein was treated with flexible side chains. These steps for the remaining protein conformations (for the three most dominant clusters) across all the serotypes have been repeated for the confirmation of results (Figure 2.2).

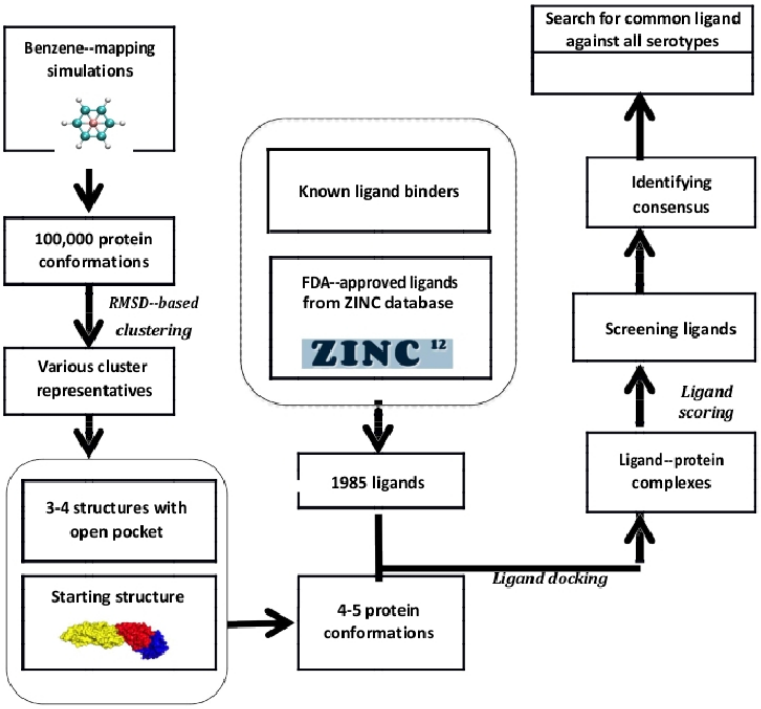


Figure 2.2. Systematic representation of the computational docking pipeline used to identify FDA-approved hits. These hits can potentially inhibit the DENV and ZIKV.

The clustering with cutoff of 1.3 Å was performed on all the atoms of amino acid residues (except hydrogen) of E protein with output as full protein ensuring that they surround the entire pocket mesh generated by MDpocket (cutoff radius ~6-7.5 Å). The interaction of docked ligand-protein complex were analyzed by LigPLot software[68]. Each ligand was docked to each protein conformation using a script in Autodock Vina. A standard number of ten docking poses of the ligand were generated and ranked based on their relative estimated binding affinities. For each protein conformation, a python script from Autodock Vina ranked the ligands based on the scores of their top docking poses. Finally, a detailed comparison was made across all protein conformations and identified top 100 consensus ligand hits with the highest scores and finally selected ten ligands which were common in all the serotypes with high affinity scores (-7.5 to -11 kcal mol^{-1}). The scores were well above the score of BOG score (-6.3 kcal mol^{-1}, a benchmark value).

3. RESULTS AND DISCUSSIONS

3.1. Wild-type and mutant (G12D) K-Ras Protein

3.1.1. Novel Modifications in Benzene probes

In the systems involving unmodified benzene (K-Ras^{B}), aggregation of benzene molecules was observed, both on the protein surface and in bulk solvent (Figure 3.1-A). It was previously reported in the SILCS[13] study that at higher concentrations (~2.7M) the probe molecules aggregated. In this study, in K-Ras^{BEX} the benzene molecules did not interact with each other, resulting in benzene molecules occupying the same space at 10 to 12 locations (7- 8 benzene molecules per cluster) (Figure

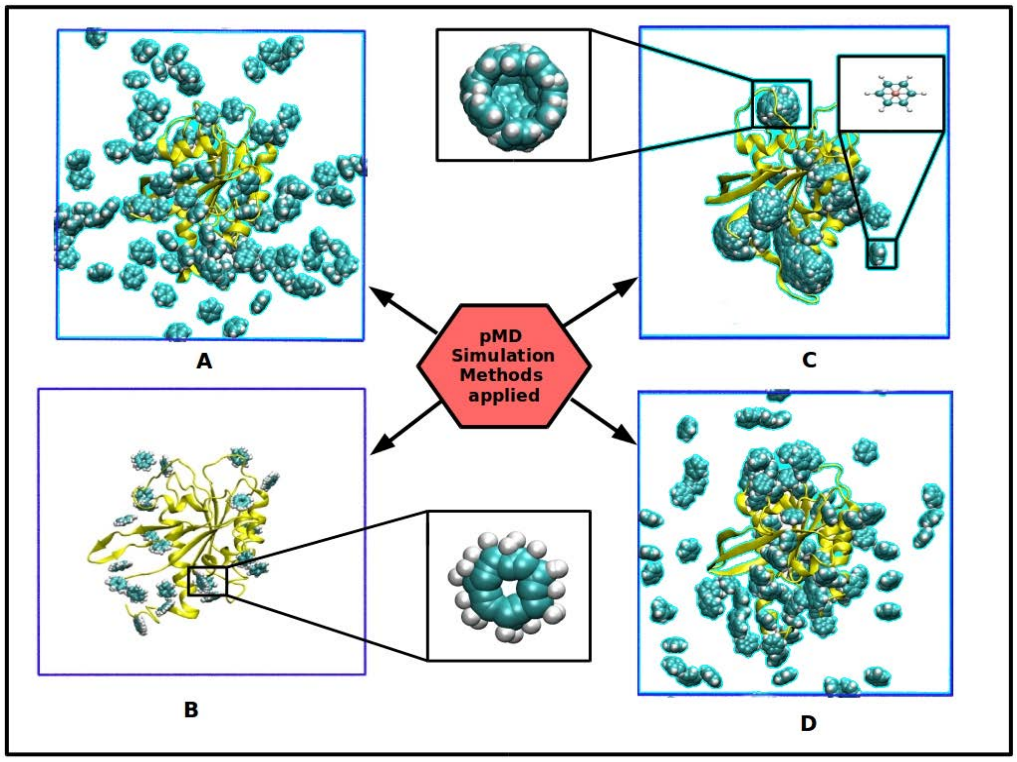


Figure 3.1. Various pMD methods involved in the new pocket mapping protocol based on K-Ras protein. **(A)** Benzene in water without modifications (K-Ras^{B}), **(B)** Benzene with exclusions - no interactions in between benzene molecules are present (K-Ras^{BEX}) **(C)** Benzene with exclusions and additional "dummy" atom inside the benzene ring of sigma value 4.5 Å which has a single repulsive potential remaining dummy atoms of benzene molecules (K-Ras^{BEXD*}), **(D)** Benzene with dummy atom of sigma value 5.5 Å (K-Ras^{BEXD}).

3.1-B). When dummy atom was introduced, with sigma value 4.5 Å and epsilon 0.002 kcal mol^{-1} (K-Ras^{BEXD*}), 8-9 "saddle-like" aggregated clusters (12-14 benzene molecules per cluster) of benzene were observed to form around the protein's

surface with hydrogen atom of one benzene in contact with the hydrogen atom of the benzene (Figure 3.1-C). However, when dummy atoms with sigma value of 5.5 Å and epsilon 0.002 kcal mol^{-1} together with exclusion maintained ($K\text{-}Ras^{BEXD}$) was used, benzene molecules were not observed to form any clusters around protein's surface and behaved like normal benzene molecules with greater aggregations than $K\text{-}Ras^{B}$ and occupying the similar regions of protein as observed in $K\text{-}Ras^{B}$ (Figure 3.1-D). Low concentration-0.2 M and 0.4 M of benzene probes were required which ensured the elimination of unwanted phase separation between benzene and water.

3.1.2. Structural Analysis of K-Ras Protein

The RMSD of the protein backbone with respect to the experimental structure varied from to 2 to 5 Å with fluctuations caused mainly by switch loops and the C-terminal region (168-180); upon excluding these regions, the RMSD reduced to 0.8-1.2 Å (Figure not shown). The results are similar to those reported by Prakash *et al*[14].

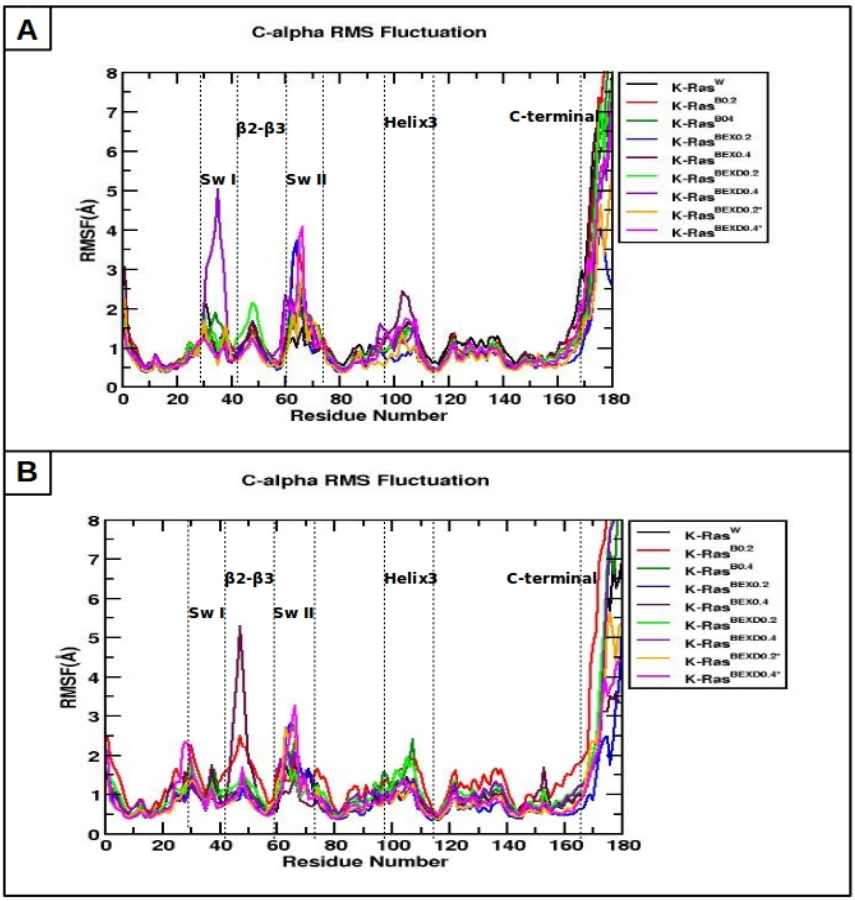


Figure 3.2. Root mean squared fluctuations (RMSFs) of K-Ras protein central carbon atoms averaged over 100 ns with respect to the experimental structure plotted for both **(A)** mutant and **(B)** the wild-type K-Ras. Sw I and Sw II corresponds to switch 1 region and switch 2 region respectively. Helix 3 corresponds to 94-104 region of protein.

The root mean squared fluctuations (RMSFs) of the Cα atoms were also similar to those reported by Prakash *et al*[14] when compared over same time scale (Figure 3.2). Unsurprisingly, the switch I (30-40) and switch II (60-75) regions have the highest fluctuations (~2-5 Å) in both wild-type and mutant proteins. The RMSFs graph plotted till 100 ns and 300 ns overlap with each other with negligible difference (Figure not shown). Large fluctuations (~2 Å) were also observed for residues 94-104 (Figure 3.2) which correspond to the helix 3 of the protein (Figure 1.1). A large pocket was found between helix 3 and helix 4 which has been characterized as a "small subsite" of low volume (36 $Å^3$) by previous pMD[14] and FTMap studies[18] (discussed in later sections). K-Ras$^{BEXD0.4}$ exhibited the highest RMSF value (~5 Å) in switch I region of the mutant as compared to other pMD methods employed here. Visualization revealed that after 45 ns, the benzene molecules quickly start aggregating near the switch I (residues D30-Y40) and moved it away from the P-loop (10-16) to a new minimum, facilitating the exposure of nearby pockets (Figure 3.3).

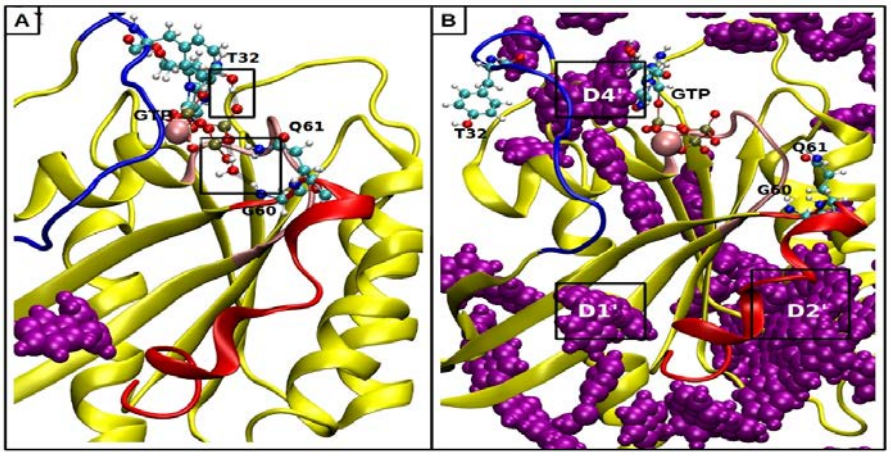


Figure 3.3. (A) The initial configuration (t=0 ns) of the production run of the system K-Ras$^{BEXD0.4}$ showing protein with its switch I (blue cartoon representation) and switch II (red cartoon representation) interacting with the GTP (CPK representation) binding P-loop (cartoon representation in pink color) through a bridging hydrogen bond (black dashed line) between G60 NH group and T32 carboxyl group with oxygen of γ phosphate of GTP maintaining K-Ras in active state. **(B)** t = 10 ns when benzene probes triggered the complete opening of the D4' and D2' and D1' pockets near the switch regions of protein in K-Ras$^{BEXD0.4}$ forcing the switch loops to move away from the GTP binding region (P-loop).

In the case of K-Ras$^{BEX0.4}$ the fluctuations of residues 40-50 (β2 and β3 region) significantly differ amongst wild-type and mutant proteins (~5.5Å and ~1.7 Å respectively). A significantly higher number of benzene molecules (5-7 overlapping benzene molecules) aggregated around the β2 and β3 sheet region of wild-type for a

greater sampling time as compared to the mutant (Figure not shown).

The secondary structure occupancy in both wild-type and mutant K-Ras is very similar and comparable with the experimental structural X-ray data in all the methods (Figure 3.4) with some variations such as the formation of b-bridge and 3_{10} helix at region formed by residues G60, S145, Q150 and F90-H95 in the crystal structure which may have resulted from force field related issues (time dependent DSSP analysis have also been performed for the conformation of result). All structural motifs are within < 7% difference compared to as observed in the crystal structure, across all the methods employed. Thus, none of the methods employed here resulted in any significant secondary structural changes or in any unfolding events.

SASA provides an initial comparison of the propensity of each method to expose the protein surface and any cryptic pocket. K-Ras^{W} exhibited the lowest hydrophobic SASA across all methods employed. The trend observed was decreasing over the

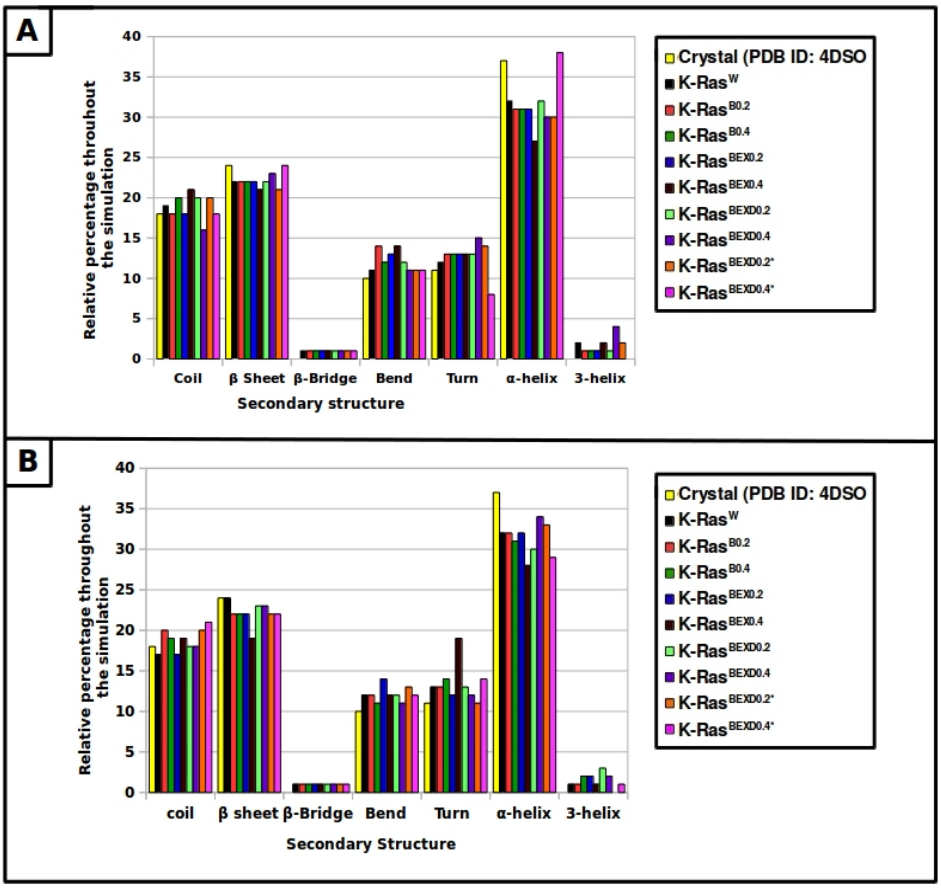


Figure 3.4. Secondary structure occupancy of K-RAS protein for **(A)** mutant and **(B)** wild-type averaged over the 300 ns of the simulation time across all methods studied. For the comparison the occupancy for crystal structure is also shown.

simulation time with destabilizing nature and reaching ~ 4400 Å ^{2} for the mutant at 300 ns or being stabilized at ~ 4600 Å ^{2} after 150 ns in case of the wild-type. There is a consistent and continuous increase (~6 % for K-Ras$^{B0.4}$, ~9 % K-Ras$^{BEX0.4}$ and ~20 % for K-Ras$^{BEXD0.4*}$) of SASA (total and hydrophobic) (when the benzene molecules are added, resulting clearly from inducing the exposure of pockets (Figure 3.5). Total SASA for K-Ras$^{BEXD0.4}$ is lower (~ 16 %) than that of K-Ras$^{BEXD0.4*}$ and comparable to the one observed for K-Ras$^{B0.4}$. This can be attributed to the large sigma value of the dummy atoms. The probe molecules come closer (the distance between dummy atoms of two benzene molecules recorded was 5.5 Å and overlap around the pockets in the proteins (Figure 3.3-B). The increase in the benzene concentration (0.2 M versus 0.4M) across each method resulted in increased SASA (~5-12 %) which may further indicate the more effective opening of cryptic pockets (Figure 3.5 and 3.6). The highest SASA amongst all methods was observed for K-Ras$^{BEXD0.4*}$ system (sigma value = 4.5 Å) which suggested the highest exposure of transient cryptic sites (Figure 3.5).

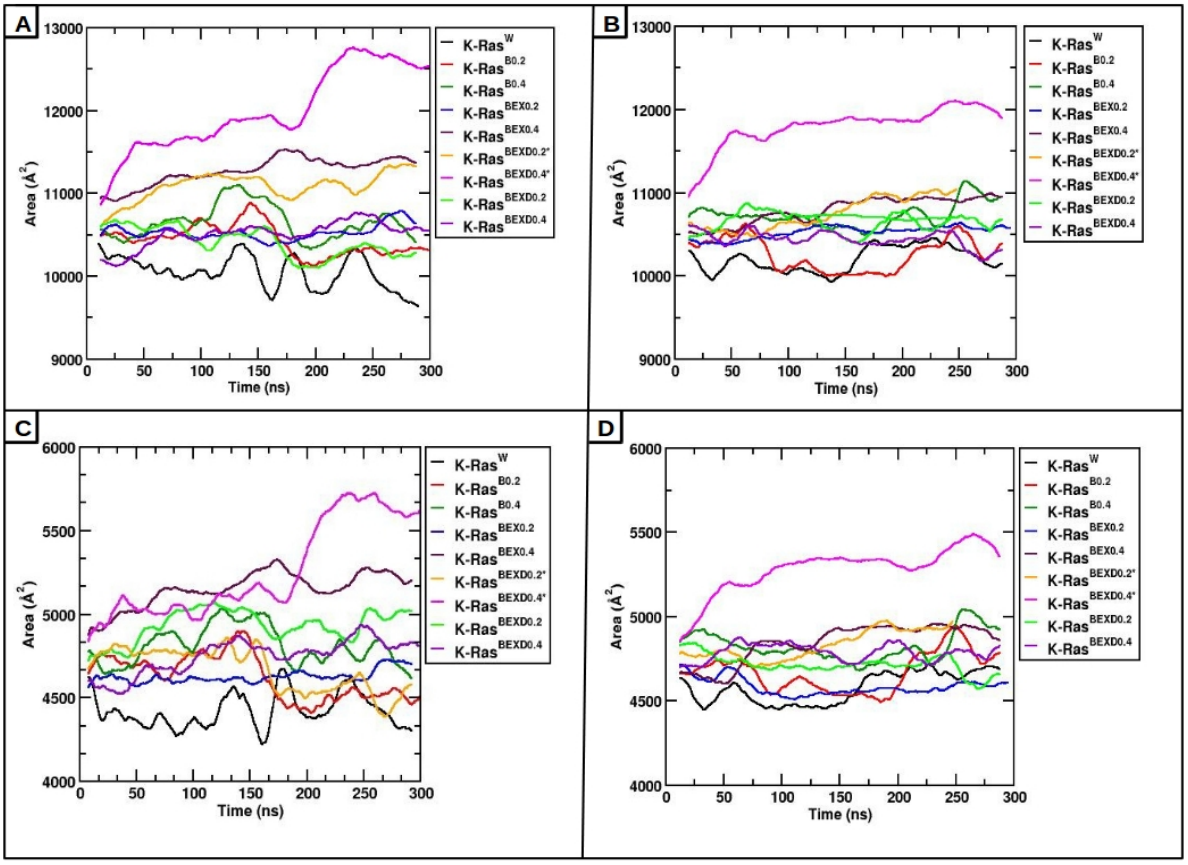


Figure 3.5. (A) Total SASA of mutant K-Ras **(B)** Total SASA of wild-type K-Ras **(C)** Hydrophobic SASA of mutant K-Ras **(D)** Hydrophobic SASA of wild-type K-Ras.

Clustering showed that the dominant cluster explored over 50% of the total simulation time in all the methods. 4 to 30 clusters were found across all systems with the maximum number corresponding to K-Ras$^{BEXD0.4*}$ in both wild-type and mutant. This indicates that this method resulted in the largest exploration of conformational space. The spread of clusters originated largely in the highly mobile switch regions and C-terminal regions. The RMSD between the most populated cluster and the starting/experimental structure was within 3 Å in all cases indicating that the conformations were stable, folded and close to the experimental conformations.

3.1.3 Pocket Analysis

All MD trajectories extracted after 100 ns generated in this work were a subject to pocket analysis. Pocket analysis of regions (8000 frames) across each simulation (where the total and hydrophobic SASA of full protein had plateaued)(Figure 3.5) showed that the four experimentally and computationally tested pockets (P1-P4)

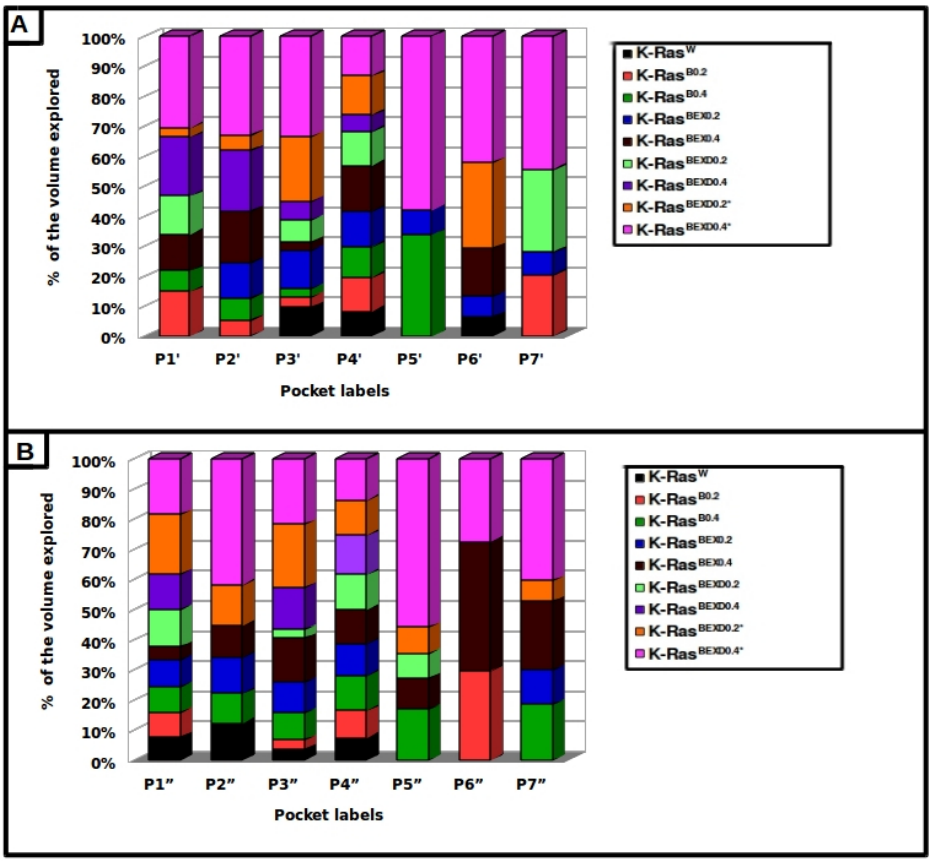


Figure 3.6. Pocket volume analysis. The histograms show the contribution of each method in exposing the specific pockets for both mutant and the wild-type of the K-Ras protein. The analysis performed on the plateau region of SASA after 100 ns (8000 frames) of each trajectory.

located in various studies[14,18,22,23,25,26,27] were well reproduced in both wild-type and mutant K-Ras.

Pocket volumes were greater than 100 Å^3 when the probe concentration was 0.4M, except for K-Ras $^{BEXD0.4}$ in wild-type where D2" is not exposed (refer to Table 2.1 for the abbreviations of pocket names). At the lower concentration of 0.2 M in each system, one of the pockets remained unexplored; some of them are b2" in K-Ras$^{B0.2}$ and e1' in K-Ras$^{BEXD0.2}$. The volume of the respective pockets compared across both the concentrations in each method revealed that higher concentration is triggering the enlarged opening of the pockets in most of the cases (Figure 3.6). In some of the cases, volume got decreased as well (b1' (270 Å^3)) and B1' (125 Å^3) at 0.4 M, but the number of such cases were lower to those where volume is increased on increasing the concentration (Figure 3.6). Four other pockets (P5-P8, refer to Table 1.1 for details) that have been described in the literature as "low consensus hot spots" using MSCS[26], FTMap[18,26] or grid energy based pocket analysis[18] have also been located by our novel pMD simulation methods with greater exposed volumes (refer to Figure 3.6 and will be discussed in detail in later sections) as compared to those reported in the literature.

Out of these, P6 (between alpha-5 and beta sheets) and P7 (located between alpha-5 and switch I and beta sheets and had not been previously explored by any computational methods) were extremely cryptic in nature and effectively opened by benzene with exclusions overlapping one behind each other (K-Ras$^{BEX0.4}$) and K-Ras $^{BEXD0.4*}$ with "saddle" like clusters only (Figure 3.1) with significantly exposed volume of 530 Å^3 (E6") and 220 Å^3 (E7") and 225 Å^3 (D*6'') and 380 Å^3 (D*7'') respectively in wild. In case of the wild-type (K-Ras$^{BEX0.4}$) the fluctuation of β2 and β3 regions as suggested by high RMSF value (~ 5 Å) in above sections signifies the opening of a pocket E6" with bigger volume where benzene molecules aggregated.

Among all the methods, K-Ras^{W} system of both wild-type and mutant protein were observed to expose the significantly lower or negligible pocket volumes. On the other hand, K-Ras$^{B0.4}$ exposed pocket B1'-B3' and B5' in mutant in the range of 100-250 Å^3 but located B4' with higher exposed volume of 1100 Å^3 which is greater as compared to the volume of the pocket in crystal (824 Å^3). In the wild-type B2" is not

exposed where in turn B7" is exposed with a volume of 180 $Å^3$. In case of K-Ras$^{BEX0.4}$ which provided higher exposure of E2", E3", E6" and E7" pockets in wild-type and E1', E2', E3', E4' and E6' in mutant comparison to K-Ras$^{B0.4}$ (Figure 3.6) the same is observed in wild-type with E2", E3", E6" and E7" were greater in volume as compared to respective pockets exposed by unmodified benzene molecules. This can be attributed to the higher cluster densities of overlapping benzene molecules (L-J potentials are switched off) around the cryptic pockets of protein as compared to that of unmodified benzene molecules.

On the other hand, the K-Ras$^{BEXD0.4*}$ system involving dummy atom of sigma value 4.5 Å exhibited the maximum number of the pockets (eight, including novel pocket P7'/P7") with the highest exposed volume and SASA (Figure 3.6 and 3.12). Each pocket has its biological relevance (refer to Table 1.1). In this work it has been observed that pockets captured by MDpocket were well mapped by the clusters formed by the benzene molecules for both wild-type and mutant K-Ras (Figure 3.7) for K-Ras$^{BEXD0.4*}$ system. This indeed indicates that benzene clusters interacts with hydrophobic residues of proteins and facilitates the exposure of pockets.

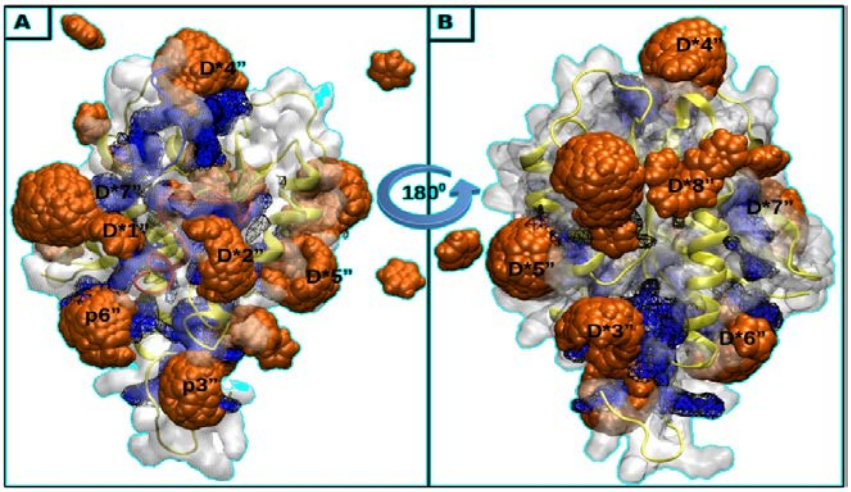


Figure 3.7. Final snapshots at 300 ns of the K-Ras$^{BEXD0.4*}$ from the (A) front and (B) back perspective. The eight probe density clusters of benzene molecules (brown van der Waals representation) together with the seven pockets (blue mesh). Pocket p8" (back side of protein) is only captured by probe interacting to that region. Protein is shown in cartoon representation (yellow) as well as surface representation (white).

The system (K-Ras $^{BEXD0.4*}$) significantly improved the volume of pockets such as D*2"-D*5" and D*7 in wild-type and D*1'-D*7' in mutant K-Ras. The transition of total and hydrophobic SASA at 175 ns (Figure 3.5) in the case of K-Ras$^{BEXD0.4*}$ (mutant) corresponded to the opening of one of the pocket near switch II region which is D*2'

and other pocket D*7' as confirmed by trajectory, volume and SASA of individual pockets (Figure 3.11 and 3.12). In K-Ras$^{BEXD0.4*}$, eight binding sites have been discovered out of which seven (D*1'/D*1"-D*7'/D*7") have been well opened by probe molecules and captured by MDpocket. Apart from this, one pocket (P8) which was found between helix4 (residues 126-138) and helix5 (residue 152-163) was opened only in < 50 % of the simulation time and could not be further recapitulated By MDpocket. This is confirmed by only 2-3 aggregates of benzene molecules involving 2-4 benzene molecules each sitting next to each other depending upon the

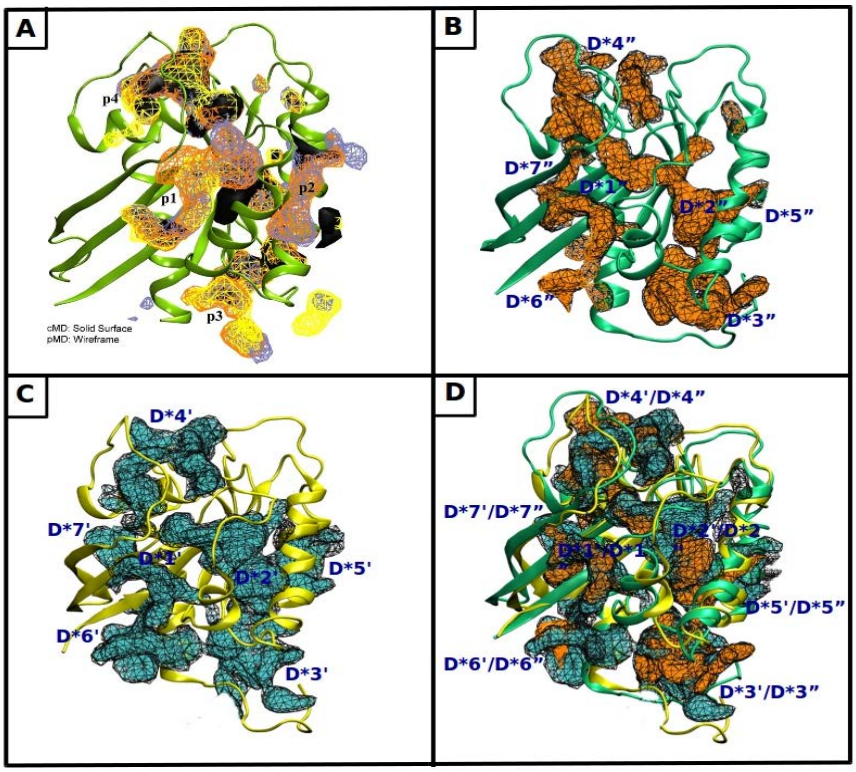


Figure 3.8. (A) Pockets shown by Prakash *et al* (Reference)[14]. **(B)** Pockets captured by K-Ras$^{BEXD0.4*}$ in wild-type K-Ras **(C)** Pockets captured by K-Ras$^{BEXD0.4*}$ in mutant K-Ras **(D)** Overlayed pockets in both wild-type (orange mesh) and mutant (grey mesh) on the most probable aligned conformation of the wild-type (green) and mutant (yellow) Ras protein.

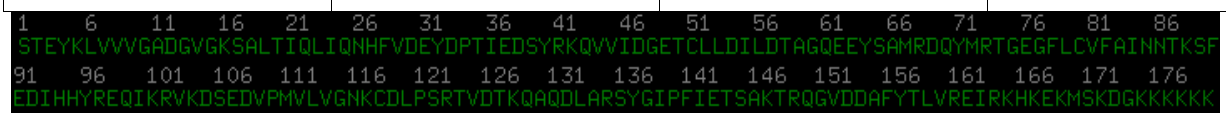
sigma value of dummy atom in K-Ras$^{BEXD0.4*}$ method surrounding that region of protein (Figure 3.7). The system K-Ras$^{BEXD0.4}$ with larger sigma value (5.5 Å) only exposed pockets D1' (350 Å^3), D2' (710 Å^3), D3' (260 Å^3) and D4' (590 Å^3) in mutant and D1"(320 Å^3), D3"(720 Å^3), and D4"(1510 Å^3) in wild-type K-Ras (Figure 3.6).In previous studies pockets P1, P3, P3, P5 and P8 have been derived using FTMap[40,53] which only considered the static experimental structure of the K-Ras mutant protein. In Figure 1.1, MDpocket analysis was performed on the crystal structure of mutant K-Ras mutant protein (PDB: 4DSO). Only five pockets (namely P1, P2, P3, P4 and P5

in pink surface) of which four were experimentally tested (namely P1, P2, P3 and P4) have been located in the crystal structure with significantly smaller volumes when compared to that exposed by the novel K-Ras$^{BEXD0.4*}$ system which facilitates the exposure of enlarged pocket volume (Figure 3.8). An extensive pocket volume comparison across all other methods with crystal structure have also been performed (Figure 3.6 and 3.10) which further suggested that probe molecules facilitates pocket opening in both mutant and wild-type K-Ras. Thus it was observed that various cryptic and weak binding sites become exposed when any protein dynamics is introduced. Those results are significantly improved by introducing benzene molecules, particularly by the use of our new method involving dummy atoms and exclusions. Surprisingly, the more the aggregations around the pocket is, higher the pocket volume exposure as observed in most promising K-Ras$^{BEXD0.4*}$ system. The useful aggregations totally can be attributed to optimum selection of the L-J parameter of the dummy atom involved in benzene molecules.

It was also shown that the experimentally or *in silico* defined pockets in previous studies [14,18,22,23,25,26,27] possess lower number of residues involved in particular cavity in comparison to our study of K-Ras$^{BEXD0.4*}$ as indicated in Table 3.1. This also indicates that volume of pockets (Figure 3.6 and 3.10) in our study are much greater than previous studies cited above (discussed in next section). Previously, P1, P2, P3 and P4 pockets have been observed in both the wild-type and mutant and were defined as high consensus sites in numerous computational and experimental studies as cited above. The pockets P5-P8 except P7 have mostly been studied by *in silico* studies in either of wild or mutant[14,18,26] (Table 1.1). In this work all the seven pockets namely D*1”/D*1’-D*7/D*7’ were found in both the wild-type and mutant K-Ras via K-Ras$^{BEXD0.4*}$. Those two sets of pockets are located in close proximity with different volumes in wild and mutant with different volumes and uncommon residues between both of them. This will open a window to design the ligands/inhibitors that can selectively target the mutant K-Ras impairing its cancerous activity which has not been considered before. However, in case of the residues surrounding pocket D*3’ and D*3’’ for wild-type and mutant respectively, they were found to be exactly same making it a challenging task when designing new selective inhibitors against the mutant. However, this region which allosterically control the switch I region of K-

Table 3.1. Residues surrounding the pockets explore by K-Ras$^{BEXD0.4*}$ in this study for both the wild-type and the mutant of K-Ras protein. Pockets were extracted after 100 ns time window where SASA shows plateau region. Results are compared with the previous work performed by Prakash *et al*[14] where pockets were captured by MDpocket .Bold text highlights the common residues across all studies. Pocket P8 residues has not been shown since it has not been captured by MDpocket by K-Ras$^{BEXD0.4*}$ Amino acids corresponding to each residue are shown in black box.

Pocket No.	**Mutant (G12D)**	**Wild-type**	**Residues reported by Prakash *et al.*[14] In G12D mutant**
D*1'/D*1"/p1	**5,** 6, **7**, 39, **54**, 55, **56**, 70, 71, 75, 76, 77, 78	**5**, **7,** 39, 40, 41, 55, **56**, 67, 70, 71 75, 76, 77	**5, 54**, 74, **7**, **56**, 67, 37, 38
D*2'/D*2''/p2	**9**, 10, 11, 12, 37, 35, 37, 38, 40, 56, 57, 58, 59, **60,** 61, 64, 65, 66, 67, 68, 69, 70, 71. 72, **78**, 92, 95, 96, **99**, 102, 103	**9,** 10, 11, 12, 36, 37, 38, 40, 56, 57, 58, 59, 62, 64, 65, 66, 67, 68, 69, 70, 71, 72, 95, 96, 102, 103	7, **9, 60**, **78,** 72, 99, 100
D*3'=D*3"/p3	**106, 107, 108**, 109, 110, **111**, **162, 165, 166,** 169, 170, 171, 172, 173, 174, 175, 176, 177, 178, 179, 180	**106, 107, 108,** 109, 110, **111, 162, 165, 166,** 169, 170, 171, 172, 173, 174, 175, 176, 177, 178, 179, 180	**105, 106, 107, 108, 111, 162, 165, 166**
D*4'/D*4"/p4	8, 11, 13, 14, 15, 16, 17, 18, 19, 20, **21**, 22, 23, 24, 28, 29, 32, **33,** 34, 35, **36, 38, 39, 40**, 56, 57, 59, 60, 116, 117, 119, 145, 146	9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 20, **21,** 22, 24, 25, 28, 29, **30**, 31, 32, 34, 35, **36**, 27, **38**, **40,** 56, 57, 58, 59, 60, 116, 117, 120, 145, 146	**30, 33, 38, 39, 40, 21, 36**
D*5'/D*5''/S4	9, 72, 78, 80, 90, 91, 93, 94, 95, 96, 98, 99, 101, 106, 107, 108, 109, 110, 111, 113, 130, 133, 134, 137, 139, 140	9, 72, 76, 80, 93, 94, 96, 97, 100, 101, 106, 107, 108, 109, 110, 111, 133, 134, 137, 138, 139	-
D*6'/D*6"	3, 4, 5, 16, 76, 163, 166, 167, 168	1, 2, 3, 4, 76, 163, 164, 167	-
D*7'/D*7"	6, 23, 24, 40, 41, 42, 44, 53, 54, 55, 153, 156, 157, 160	4, 6, 20, 22, 23, 44, 46, 53, 152, 153, 156, 157, 160	-



Ras can be a useful drug target in the cases where wild-type Ras is driving the oncogenic activity of the mutant Ras[24].

4.1.4 Time Dependent Efficiency of the Novel Method

As the pockets exposure is dependent on the sampling time. The volume of the pockets opened up by all the methods captured by MDpocket across two time windows of the MD trajectories: 50-100 ns and after 100 ns (220-300 ns) have been compared. The results from most promising method K-Ras$^{BEXD0.4*}$ shows that in the case of mutant all the pockets (D*1-D*7) with significant volumes were reproduced within 100 ns and have volume greater than that of pockets present in crystal except the case of pocket D*5' (Figure 3.10). The volume of D*5 before 100 ns is 105 Å^3 but in crystal it is 360 Å^3. However, on increasing the sampling time, it increase to 525 Å^3. In particular, within 50-100 ns pockets D*4' and D*6' explored a volume of 1575 Å^3 and 985 Å^3 and then decreased to 10 % and 35 % respectively (their final volume) at 100-300 ns (Figure 3.10).

Moreover, in case of the wild-type similar results were observed for all the pockets and also for D*5"when compared with the crystal (Figure 3.10). One exception corresponded to pocket D*7" which was not exposed in the first 100 ns also supported by its total SASA (Figure 3.12). This pocket volume started increasing after ~100 ns. In case of pockets D*1'/D*1"-D*3'/D*3" & D*5'/D*5" generated for the 100-300 ns time window, 14-70 % large volumes were observed as compared to that generated from 50-100 ns (Figure 3.10).

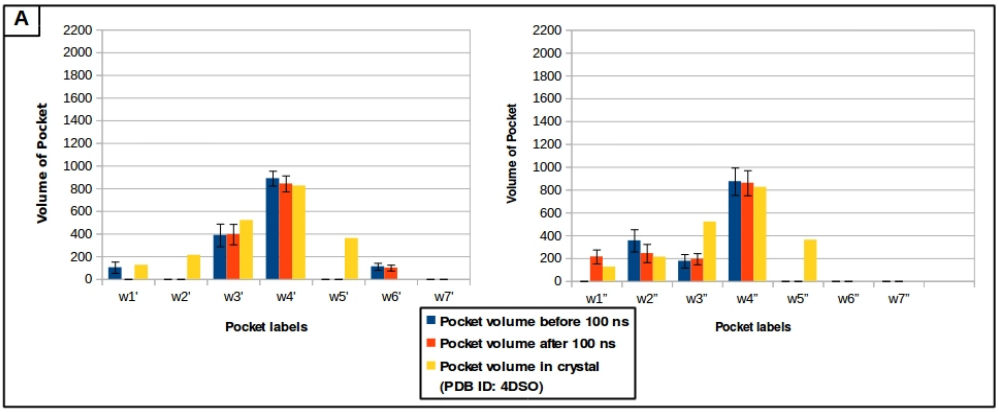


Figure 3.9 The volume of each pocket (Å^3) generated from the most efficient method, namely **(A)** K-Ras^W for trajectory taken before 100 ns and after 100 ns time windows for mutant (pockets labelled with single prime) and wild-type (pockets labelled with double prime).

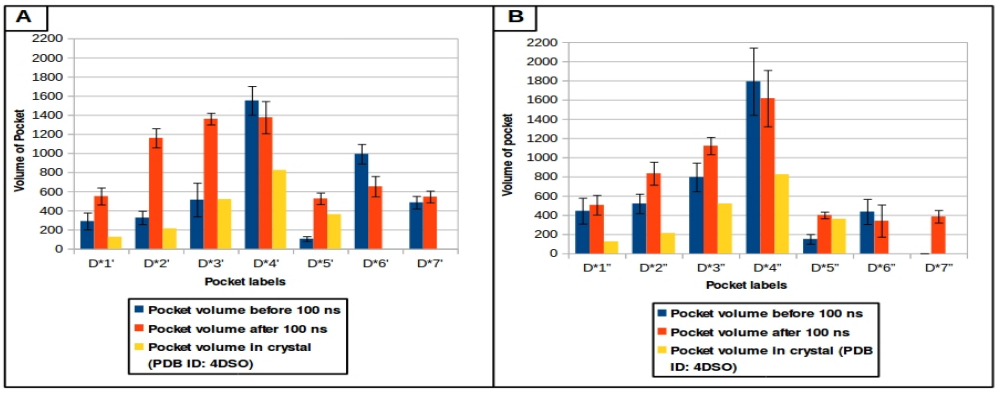


Figure 3.10. The volume of each pocket (Å^3) generated from the most efficient method, namely K-Ras$^{BEXD0.4*}$ for trajectory taken before 100 ns and after 100 ns time windows for **(A)** mutant and **(B)** wild-type K-Ras. Pockets labelled with single prime are for mutant and pockets labelled with double prime are for wild-type.

The tremendous change in the averaged volume of D*2' was observed after 100 ns (325 Å^3 before 100 ns and 1160 Å^3 after 100 ns). The volume of D*4' shows fluctuations after 115 ns in case of the mutant (Figure 3.11) which is possibly attributed to the simultaneous opening of pocket D*1' and D*2; which are present adjacent to D*4'in the switch region (Figure 3.8). Similar pattern were observed in case of pockets D*6' which is present near D*3'and may have been affected by its increase in volume. The aforementioned trend of these two pockets (D*4" and D*6") was also observed in wild-type (Figure 3.10 and 3.11).

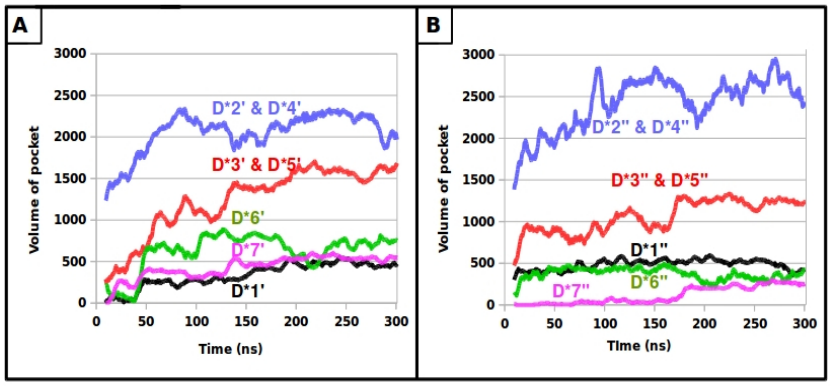


Figure 3.11 Time dependent evolution of each pocket volumes (in Å^3) explored in the K-Ras$^{BEXD0.4*}$ system for **(A)** mutant and **(B)** wild-type. Pocket D*2' & D*4'/D*2" & D*4" and D*3' & D*5'/ D*3" & D*5" were combined due to the consistent 2.9 isovalue used across all the method which resulted in joined volumes.

The SASA over the simulation time for most of the pockets converged within 60 ns of the simulation time in case of the wild type using K-Ras$^{BEXD0.4*}$ method. However, this was not the case for D*7" (a novel pocket) where SASA started increasing after 99

ns resulting in opening of the cryptic pocket triggered by penetration of benzene clusters (confirmed by trajectory analysis). The SASA values and volume for all the pockets in mutant except D*1' and D*4' reached higher values well after 50 ns as compared to the wild-type (Figure 3.11 and 3.12). This is further supported by more number of residues involved in the formation of pockets D*2', D*3' and D*5'-D*7' in mutant as compared to wild-type (Table 3.1). The initial values of SASA at 0 ns highlight the presence or absence of pockets in crystal (starting structure) which are significantly increasing over the simulation time (Figure 3.12). For example, D*6'/D*6" and D*7'/D*7" in both mutant and wild-type start from almost negligible volume and SASA which further increases to the range of 200-500 Å^2 due to interactions with benzene clusters which gets stabilized at the end of the simulation.

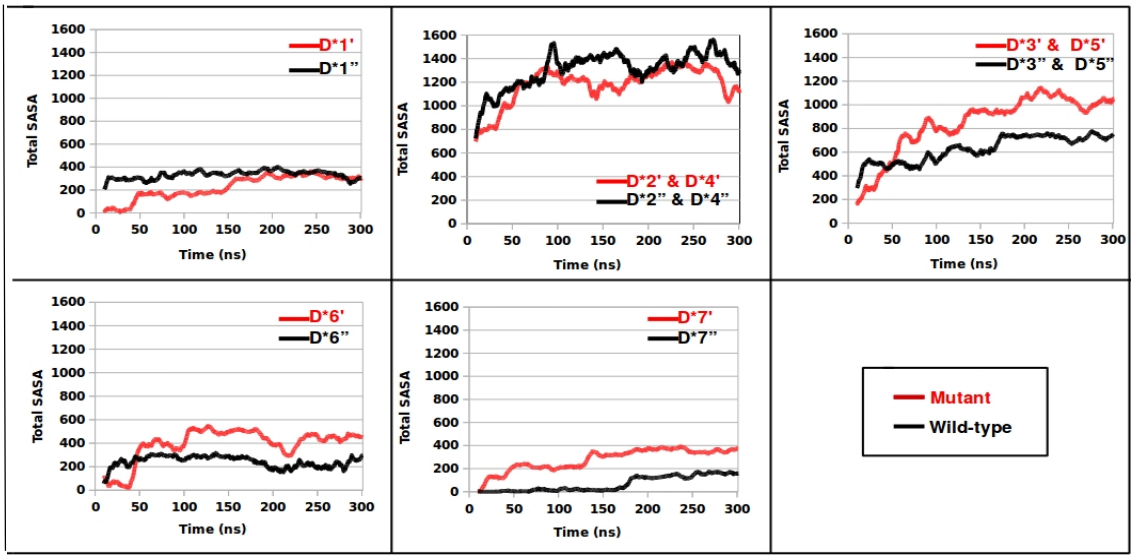


Figure. 3.12 Total solvent accessible surface area (SASA) of individual pockets (in Å^2) explored in K-Ras$^{BEXD0.4*}$ system over the simulation time.

The volume data extracted in this study was compared with the results derived from the FTMap study in literature[18], and found that pocket P1, P2, P3 and P4 showed the volume of 112, 143.6, 173, 139.9 Å^3 respectively by FTMap and also with experimental data where reported volume of P1 245 Å^3 and reported SASA for P2 was 432 Å^2 [23,28]. The volumes were much smaller compared to the volume of highly exposed pockets D*1'/D*1"-D*4'/D*4" derived from our method Kras$^{BEXD0.4*}$ within 100 ns only (Figure 3.10). Apart from this, D*5'/D*5"-D*7'/D*7" were also found within 100 ns to be of comparable volume to be identified as druggable sites and the volume further increased on increasing the time.

To better elucidate the efficiency of method **K-Ras$^{BEXD0.4*}$**, one needs to account for the efficiency of other pMD methods employed in this study. The average volume of the pockets generated across each method before and after 100 ns had been recapitulated. In K-Ras^{W}, w1', w3', w4' and w6' were located in mutant and w2"-w4" were found out in wild-type in the volume range of 100-800 Å^3 within 100 ns. Even after increasing the sampling time, there was not significant improvement in volume of the pockets observed (Figure 3.9). However, addition of benzene molecules improved the results. In K-Ras$^{B0.2}$ increment of sampling time after 100 ns increased the volumes of only b2' (by 33 %) and b1" (40%). Other cryptic pockets such as b7' and b7" get exposed with volume 250 Å^3 and 370 Å^3 on increasing the sampling time (Figure not shown). Increasing the concentration helped the enlarged exposure of B2', B1", B2" and B4" within 100 ns as compared to the respective pocket generated by K-Ras^{W} and K-Ras$^{B0.2}$. Pocket 4 volume decreased slightly (1%-15%) on increasing the sampling time though. Further enlargement in pocket volume has been achieved by K-Ras$^{BEX0.4}$ system. The pocket exposed E1'-E4' within 100 ns with 30 % increment in the volume of E4' as compared to that of e4' (Figure not shown). However, E4" has been decreased to 30 % in wild-type as compared to e4". Increasing the sampling time lead to expose E4" and e4" with almost similar volume of 1250 Å^3. High concentration (0.4 M) helped in the location of transient cryptic pocket E6" (absent in K-Ras$^{BEX0.2}$) and E7" (64% increment as compared to e7" within 100 ns). K-Ras^{BEXD} behaved similar to K-Ras^{B} system and helped locating experimentally validated pockets with volume comparable to K-Ras^{B} system and increasing the sampling time only helped in exposure of some of the pockets such as D2' and D3" (approximate 60% increment with respect to volume before 100 ns) (Figure not shown). Optimising sigma value of dummy atom improved the pocket searching efficiency i.e. in K-Ras $^{BEXD0.2*}$ system, all the experimentally tested pockets from d*1'/d*4"-d*4'/d*4" were successfully located with the significantly exposed pocket volume within 100 ns. However, cryptic pockets d*7', d*6" and d*7" have not been explored. Increasing the concentration of benzene molecules containing dummy atom in this system (K-Ras $^{BEXD0.4*}$) facilitated the opening of all the seven pockets within 100 ns (Figure 3.10).

All of the above results stress that the pMD method which involves benzene with exclusions and a dummy atom of sigma value 4.5 Å and epsilon 0.002 kcal mol^{-1} as

a probe molecules is therefore an extremely promising technique as compared to other pMD methods applied here in exposing most of the cryptic pockets in biologically relevant protein structures. The method located all the experimentally tested pockets and novel pocket within 100 ns with significant exposed volumes in both wild-type and mutant. However, for facilitation of enlarged pocket exposure, one needs to increase the sampling time (Figure 3.10). Only putting exclusions in benzene (K-Ras ^{BEX}) can also be extremely helpful in exposing cryptic pockets as well, like E*6'/E*6" and E*7'/E*7" (Figure 3.6) because they can occupy smaller space to penetrate those pockets. The method K-Ras $^{BEXD0.4*}$ showed better results in exposing all the pockets as compared to K-Ras ^{B}, K-Ras ^{BEXD} and K-Ras^{W}.

3.2. DENV and ZIKV Envelope (E) Protein

3.2.1. Benzene mapping MD simulation

After benchmarking the pockets explored in wild-type and mutant K-Ras protein via most promising method K-Ras$^{BEXD0.4*}$, the study had been extended to other druggable target i.e. dimeric E protein of DENV serotypes and ZIKV. The aim was to reveal cryptic pockets followed by extensive virtual screening to develop inhibitors binding to the selected pocket of interest. A 300 ns explicit solvent benzene mapping simulation with a dummy atom have been performed on all the four serotypes of the DENV as well as on the ZIKV E protein. E protein was not unfolding even upon addition of probes (confirmed by structural analysis). Additional simulation of each protein was also performed in water only. However, the pockets remains closed or very transient in only water system as observed in K-Ras system as confirmed by plotting total and hydrophobic SASA of the protein and individual pockets (Figure not shown). Upon the addition of benzene probes containing dummy atom, pockets at the interface of two monomers are stabilized over the simulation time. Various pockets have been located on all E protein studied (Figure 3.13). However, the interest of this work was limited to the interface of the two monomers with a potential of blocking the conformational change and dimer-to-trimer transition during the viral infection (Figure 3.13 red rectangle).

Hence, one pocket was selected in the antibody binding region[69] being located in

between the two monomers by the benzene mapping and binding of drug there could either potentially stabilize the dimer. The pocket was partially opened in crystal and was discussed in the work of Yenamalli *et al*[33]. This pocket has been located symmetrically on both sides of the dimer (Figure 3.15). Interestingly, this site was located near the fusion peptide loop in the dimeric protein supporting the potential

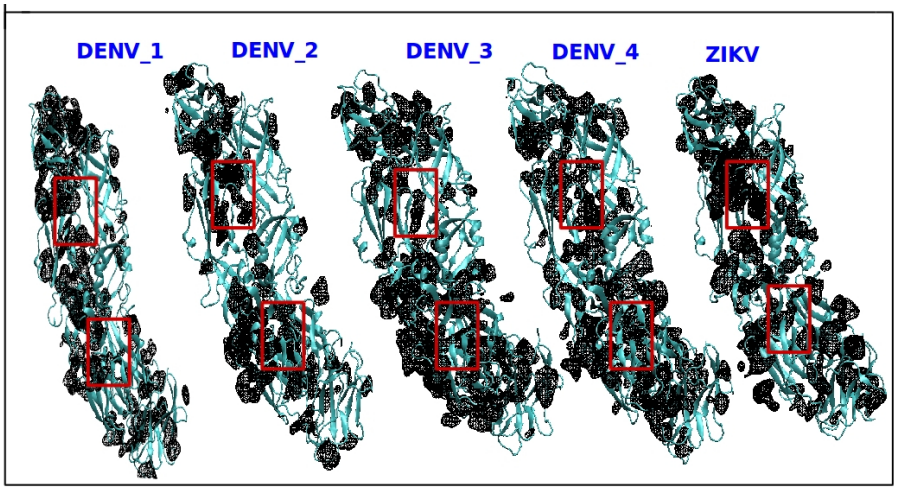


Figure 3.13. E protein of flaviviruses showing various pockets (black mesh) located by benzene mapping MD simulation with the benzene molecules containing a dummy atom of sigma 4.5 Å and epsilon of 0.002 kcal mol^{-1}. The red boxes show the region of interest for virtual screening. The red rectangular regions denote the monomer-monomer interface with a potential blocking of dimer-to-trimer transition. Protein in cartoon representation in shown in cyan.

for blocking the viral infection process. The cavity of the pocket had a volume of ~400-600 $Å^3$ and SASA of 2,200-2400 $Å^2$ across all the serotypes. The residues are covered with more than 35 residues across all the serotypes (Table 3.2). Interestingly, the pocket was closed in the crystal structure of DENV1, DENV2 and DENV4 crystal structure but got exposed upon addition of benzene molecules (Figure 3.15). However, in the case of DENV3 and ZIKV those pockets remained open in the crystal and top three most probable structure obtained from benzene mapping MD simulation. However, benzene mapping simulations revealed higher number of residues in comparison to the data from literature[33] suggesting the greater volume of the pocket than in the crystal structures (Table 3.2).

Residues 242-250 are common in all the regions surrounding the pocket. The sequence alignment of the residues among all the serotypes shows that the residues 242-248 are conserved for ~70-80 % across all the serotypes suggesting high potency for blocking all dengue serotypes and Zika. Thus, the flexible docking was performed on this binding region. Among these are 8-11 charged residues and 12-17

Table 3.2. Residues surrounding the pocket of interest across all the E protein serotypes. Blue residues corresponds to chain A while black one are the residues of chain B of a dimer. The literature study indicates the work of Yenamalli *et al.*[33] where they performed docking studies on crystal structure of DENV2. However, the pocket selected for docking in this study was not targeted by them.

E protein	Residues surrounding the pocket
DENV-1	5-8, 27, 28, 44-47, 96-98, 113-116, 152, 242-250, 273-274, 279-281, 317
DENV-2	6, 8, 26-29, 46-47, 95-99,102-103, 110, 240-247, 266, 277-278, 316-317
DENV-3	6- 9, 27-29, 95-98, 110-113, 240-249, 271, 273-279, 280-282
DENV-4	5-9, 26-30, 44-47, 70-72, 95-99, 113-115, 151-153, 239-251, 264-275
ZIKV	6, 27-28, 65-72, 97-99, 114-116, 246-254, 273-274, 285, 313
DENV-2 (literature)	1-8, 28, 30, 44, 151-155, 97-109, 244-247, 316

hydrophobic residues across all he serotypes (Table 3.2 and Figure 3.14). Previous study by Yenamalli *et al*[33] showed that this cavity was present in dimer form but not in the trimer form of the E protein of DENV2 and DENV3. Also, the cavity was found to be conserved in pre fusion structure of E proteins from Japanese encephalitis virus (JEV) and Tick-borne encephalitis (TBE)[33].

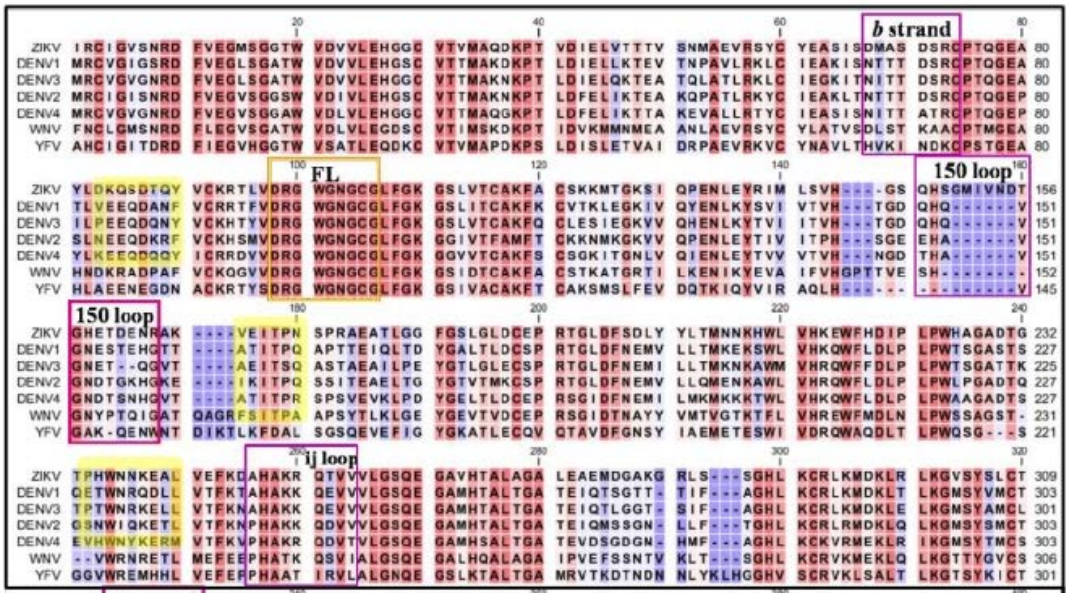


Figure 3.14. Sequence alignment of the E protein of various flaviviruses[69]. The yellow boxes show the residues which are highly conserved across all the serotypes and forming the pocket.

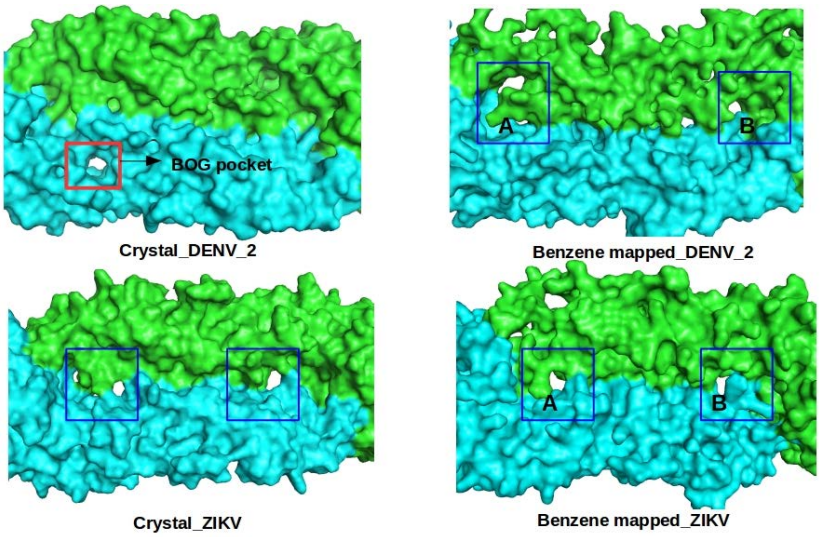
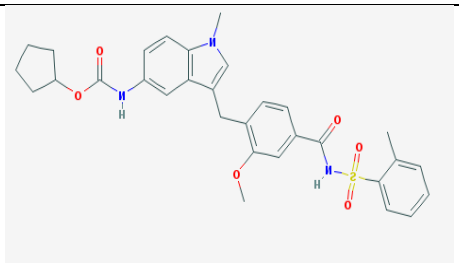
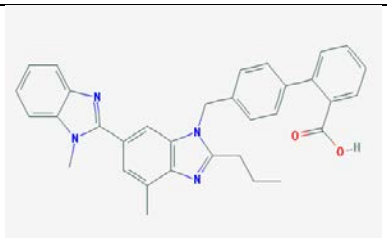
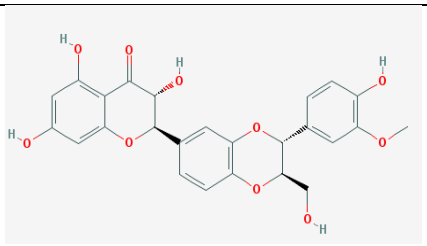
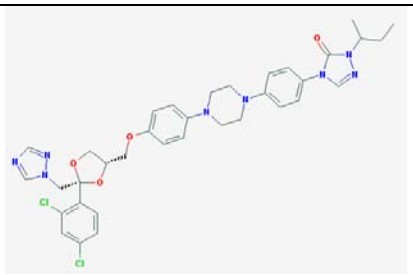
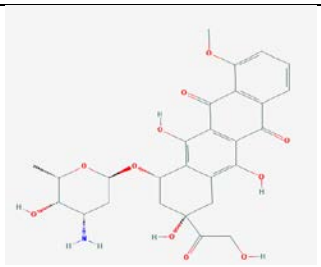


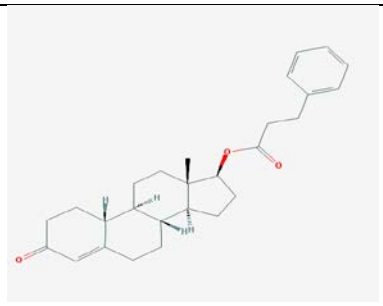
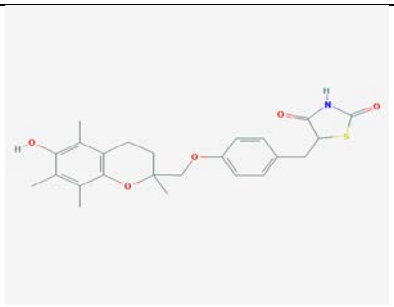
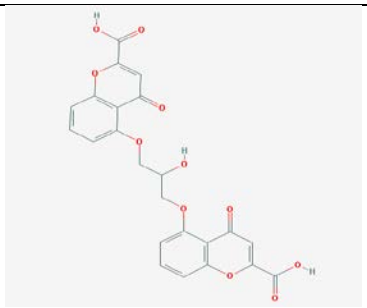
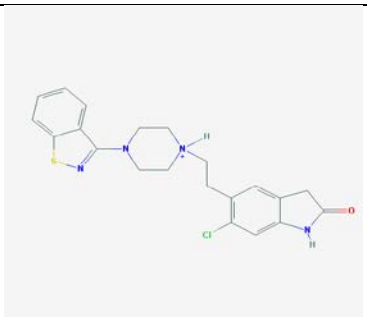
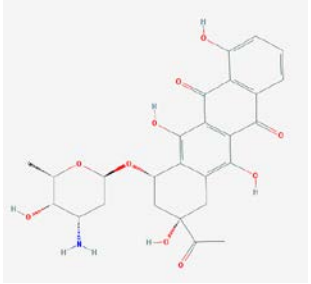
Figure 3.15. Intra dimeric pocket in crystal structure (left) and in most probable conformation of protein obtained from benzene mapping MD simulation. The pocket was not visible in DENV1 and DENV4 homology modelled structure but visible in DENV3. Surface representation of E protein colored in blue and green for each monomer.

4.2.2. Virtual Screening and Hit Identification

The pocket A (Figure 3.15) across all the serotypes has been targeted by virtual screening. Docking protocol was performed on top three most dominant clusters with selected flexible residues extracted from simulation of E protein for DENV and ZIKV which captured 50-70 % of all sampled conformations. To verify that Autodock Vina is reliable for our protein system, the compound "R1" was docked on the site1 (residues 1, 143-149, 156, 178, 295 (domain I) and residues 324, 353-358, and 359-360 (domain III)) of DENV2 explored by Yenamalli *et al*[33]*.* The predicted binding affinity corresponded to -9.6±0.1 kcal mol^{-1} which was comparable to the affinity reported in their work (-9.43 kcal mol^{-1} on DENV2). The interactions of drug with protein residues were also in agreement with the literature. Similarly the n-octyl- β-D-glucoside (β-OG) ligand was docked into the experimentally defined hydrophobic β-OG pocket of domain II in DENV2 with observed docking score of -6.3 kcal mol^{-1} and got an agreement with the affinity reported in literature (-5.9 kcal mol^{-1})[37]. This indicates that despite the limitations of scoring functions[45], Autodock Vina is still sufficiently accurate to be used for our protein system.

Table 3.5. Results from virtual screening using FDA approved drug library performed on the most dominant clusters of the E protein dimers extracted from benzene mapping protocol. Results were obtained for all dengue serotypes and Zika virus. Top ten highest affinity drugs were identified. The error bars on the docking scores fell in the range of ±0.5-0.9 kcal mol^{-1}. Figures for drug molecules have been extracted from PubChem[70].

Popular Names and functions of selected FDA approved drug molecules	2D Structure of FDA approved drug	Average Vina scores considered over top three most probable confirmation (kcal mol^{-1})				
		DEN V1	DENV 2	DENV 3	DENV 4	ZIKV
Zafirlukast (Cytochrome P450 2C9 Inhibitor)		-9.5	-10.8	-9.5	-8.0	-9.6
Telmisartan (Antihypertensiv-e agent)		-8.4	-9.3	-9.6	-8.8	-9.5
Silibinin (Protects liver cell from toxins-Anti-cancer effect)		-8.2	-9.0	-9.5	-8.9	-10.0
Itraconazole (Antifungal Agent)		-8.7	-9.0	-8.8	-8.7	-10.9
Doxorubicin (Anticancer agent**)**		-7.6	-9.0	-9.0	-9.4	-9.3

Nandrolone Phenpropionate (Androgen receptor agonist)		-7.6	-9.1	-9.3	-8.5	-8.8
Troglitazone (Antidiabetic and anti-inflammatory drug, now replaced with rosiglitazone)		-8.5	-10.3	-9.1	-8.7	-8.7
Cromolyn (Anti-allergic and anti-inflammatory agent)		-7.6	-9.1	-9.1	-9.2	-8.9
Ziprasidone Hydrochloride (antipsychotic agent to treat schizophrenia)		-8.3	-9.2	-9.3	-8.5	-8.9
Carminomycin (Antibiotic)		-7.6	-9.0	-9.3	-8.9	-9.6

Amongst all dengue serotypes and Zika virus approximately hundred consensus ligand hits were identified. Top ten ligand were sorted based on docking scores as well as their biological relevance and activity against particular disease and they are presented in Table 3.5. For some of the ligands, posing was given the preference which implied that those confirmation of ligand was selected which interacts with both the chains of E protein no matter whether it did not come to top rank based on affinity score out of ten generated poses. In addition all top ten ligands were docked on the crystal structures and homology modelled E protein structures prior any simulation. It was observed that the ligands were binding with almost similar or lowered affinities (difference in the range of 0.1 to 5.9 kcal mol^{-1}). However the visualization of binding poses showed that those ligands were sitting inside the cavity (ligand poses) exposed by MD simulation unlike the ones bound to crystal where they are exposed to the surface since the cavity was partially opened (Figure 3.16).

The difference in affinity scores for a particular drug binding to given dengue serotype or Zika virus arises due to its variable interaction with different residues of the pockets as well as their volume and size. Although similar location is observed for all E proteins studied here the binding modes are different (Figure 3.17 and Table 3.5). For example, in DENV3 presence of nitrogen of lysine 244 residue makes hydrogen bond with hydrogen of hydroxyl group of drug Carminomycin unlike in DENV2 where lysine was replaced by histidine. The drug resembled anticancer agent doxorubicin and have shown a good affinity in the range of -7.6 to -9.6 kcal mol^{-1} across all the E protein of dengue serotype and Zika. Similarly, nature of binding poses had also been recapitulated for other remaining nine selected drugs.

The predicted interactions of all the ligands with E protein are mostly governed by hydrophobic forces because of presence of hydrophobic and non-aromatic rings in drugs. In addition hydrogen bond formation (1-7 bonds) had also been recapitulated since the pocket contains 8-11 charged residues as well. All the drugs were most often contacting the 4-5 residues from 242-251 which are more than 80 % conserved across the Zika and all the serotypes of dengue. 8 out of 10 selected ligands were forming at least one hydrogen bond with any one of the residues from 242-251.

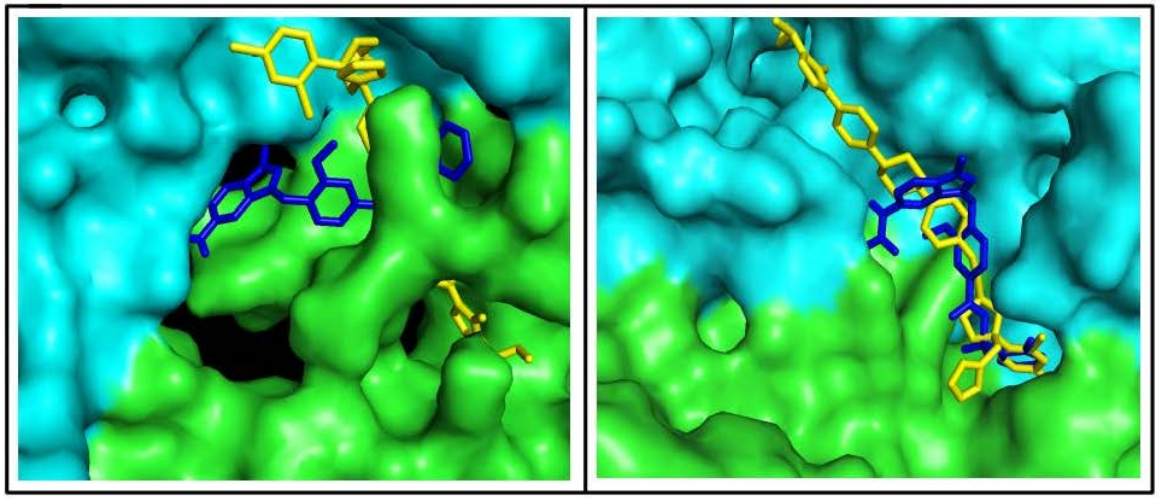


Figure 3.16. Binding of highest affinity drug Itraconazole (yellow) and Zafirlukast (blue) in a completely opened pocket conformation obtained from Benzene mapping simulation (left) and in a closed pocket in native crystal structure of DENV2 (PDB ID:3J27) (right). Similar studies have been also performed for other remaining drugs across all the E protein of Zika and all the serotypes of dengue.

However, hydrophobic interaction dominated in those regions. Apart from this the other conserved residues which interacted with the ligands were 44-45, 47-48, 97-99 and 278-279.

All ten highest affinity molecules containing non-polar aromatic or on aromatic cyclic rings have at least two hydrophobic interactions with at least two of the above mentioned conserved residues (for example, refer to the case of carminomycin and itraconazole shown in figure 3.17-3.19). However, one drug (Itrconazole, an antifungal agent[70]) with the highest affinity of ~11.0 kcal mol^{-1} toward ZIKV and comparable affinity toward other serotypes of DENV (-8.7 kcal mol^{-1} for DENV1 and 9.0 kcal mol^{-1} for DENV2, -8.8 kcal mol^{-1} for DENV3 and -8.7 kcal mol^{-1} for DENV4) had a binding mode being positioned in between two monomeric chains (Figure 3.18). The binding pose of itraconazole with the residues interacting is shown in Figure 3.18. In case of ZIKV its interactions with E proteins was dominated by three aromatic rings (Figure 3.18) as well as one polar interaction and the drug being extended and bound to the other nearby exposed pocket. Thus, its interaction with both the chains could suggest a way to inhibit the possible dimer-to-trimer transition and virus fusion process.

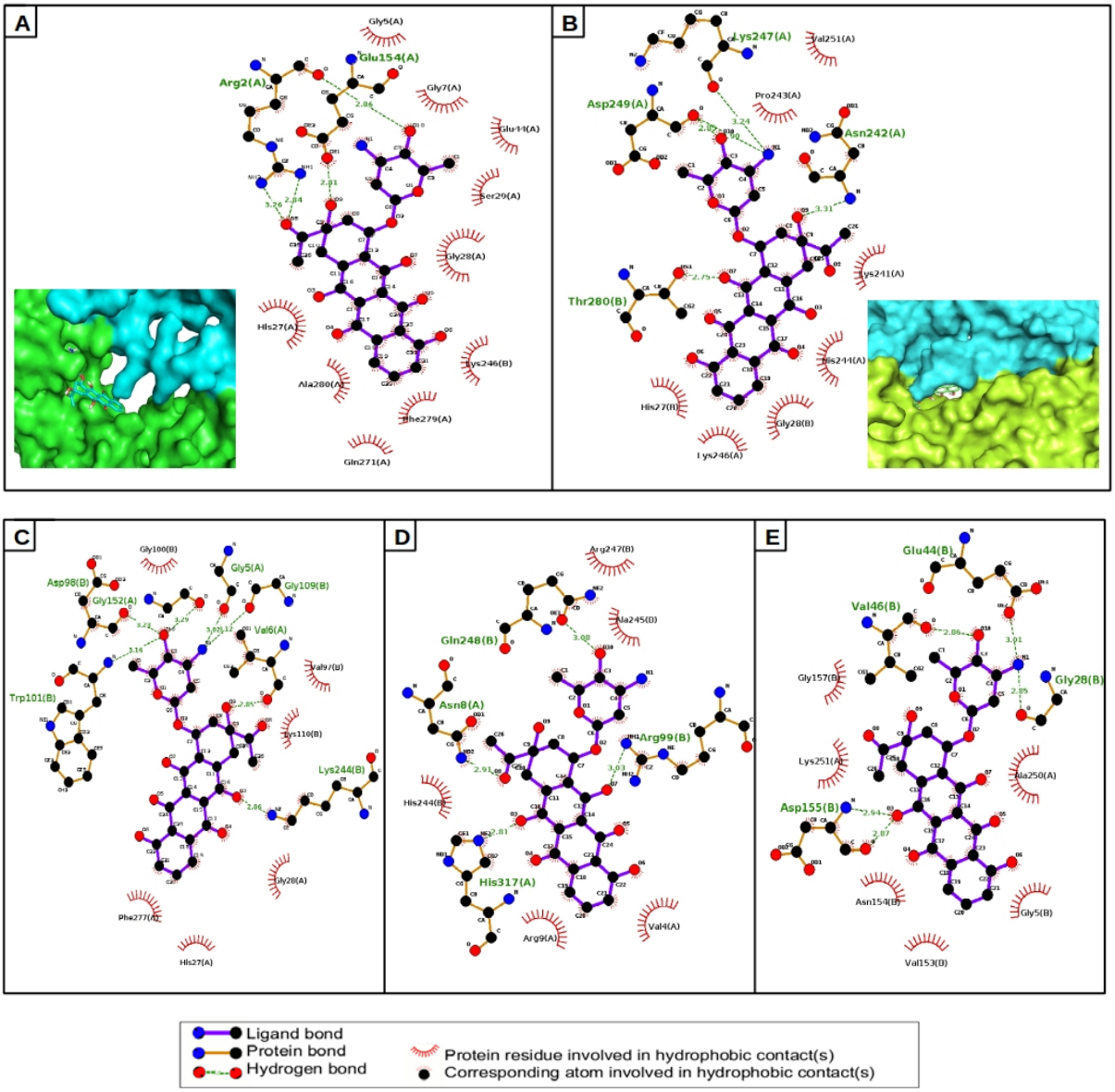


Figure 3.17. Binding poses of second highest affinity drug: Carminomycin (resembling with the anticancer drug doxorubicin) in **(A)** DENV1 **(B)** DENV2 **(C)** DENV3 **(D)** DENV4 and **(E)** ZIKV. The position of drug is shown for DENV1 and DENV2 in surface representation interacting between the two chains for which their respective residues are labelled as A and B. The number of hydrogen bonds varied from 4-7 together with hydrophobic interactions of the cyclic rings. Different residues interacting with drugs leads to different binding modes. The affinities are mentioned in Table 3.5.

Similar observations of interaction of drug with both chains were made in the case of dengue serotypes (Figure 3.19). The drug was found to interact with several of its sterioisomers in the top docking hits as well. Apart from this, drug silibinin, zafirlukast, and telmisartan were also large enough to interact well with both the chains possessing high affinity scores (data not shown). Interestingly, anticancer agent doxorubicin[70] also exhibited high affinity to all dengue serotypes

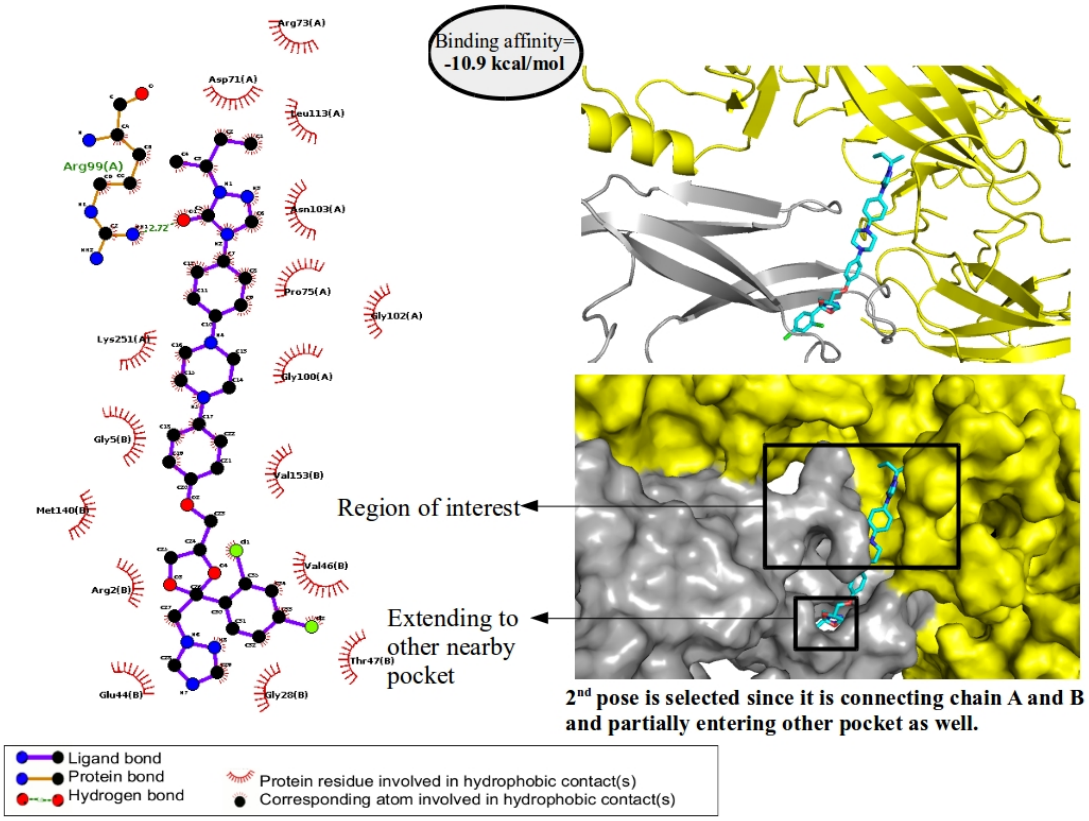


Figure 3.18. The highest affinity FDA approved ligand: Itraconazole bound with ZIKV with affinity ~11 kcal mol^{-1}. Most of the interactions are hydrophobic because of presence of various benzene rings, triazole and cyclohexane ring. The drug also forms a hydrogen bond with Arg 99 of E protein.

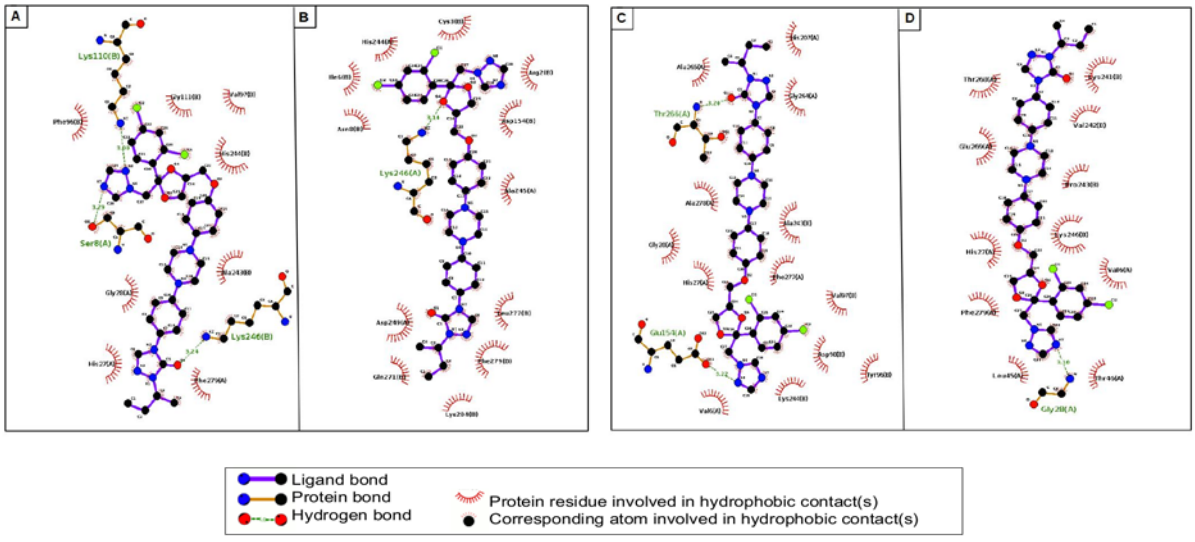


Figure 3.19. Binding poses of highest affinity drug: itraconazole in **(A)** DENV1 **(B)** DENV2 **(C)** DENV3 **(D)** DENV4. The position of drug is shown to interact with both the chains for which their respective residues are labelled as A and B. The number of hydrogen bonds varied from 1-2 together with the dominant hydrophobic interactions of the cyclic rings. Different residues interacting with drugs leads to different binding modes.

and Zika and may be a promising antiviral therapeutic. Also, the different binding

modes of a particular drug in all the serotype could certainly guide a way to modulate the chemical functionality and binding affinity of the drug considering conserved residues in the pockets across all the serotypes.

The benzene mapped most dominant confirmation of ZIKV E protein bound to itraconazole was aligned with the one bound to a broadly neutralizing antibody (PDB ID: 5JHL[69]) solved at 3 Å (Figure 3.20). This antibody was found to bind the ZIKV with high affinity, interacting with both the chains and was able to neutralize the ZIKA strains in mice and in vitro. It has been observed that the antibody was interacting with both of the chains near to the fusion loop where the drug itraconazole was found with its binding mode. This could guide a way in future to simulate the combination of antibody, itraconazole and other selected drugs bound with the E protein of all the serotypes and moving a step ahead in antiviral therapeutic development.

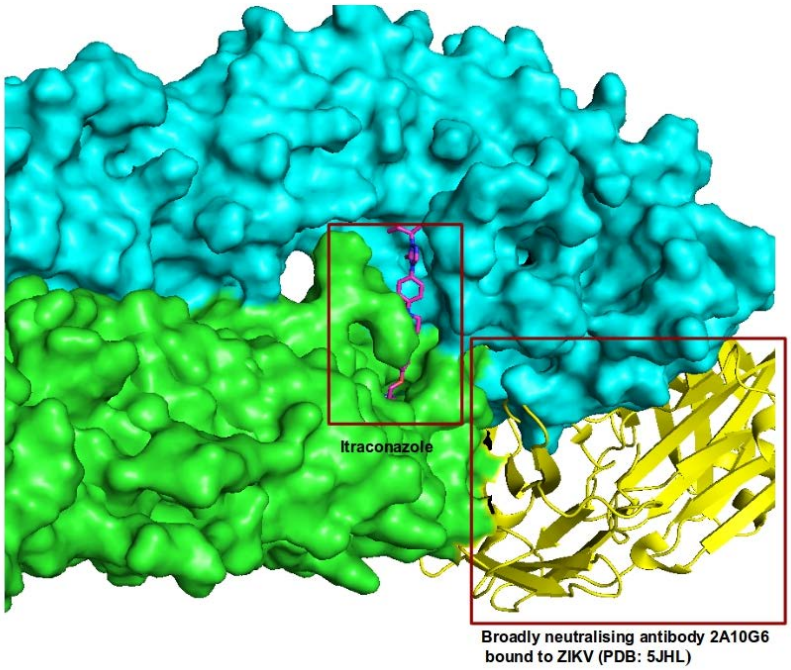


Figure 3.20. Surface representation of ZIKV envelope protein with open pocket obtained in our benzene mapping simulations aligned with the protein bound with antibody (yellow cartoon representation) (PDB ID; 5JHL[69]).

CONCLUSION AND FUTURE DIRECTIONS

In this thesis it has been successfully demonstrated the novel pMD simulation method which involves probe benzene molecules with additional dummy atom and optimized L-J parameters (σ = 4.5 Å and ϵ = 0.002 kcal mol^{-1}) in the geometric center of the benzene ring. The “saddle” like clusters of benzene molecules around the pockets helped reproducing the four experimentally known allosteric binding sites of Ras along with the exposure of novel putative binding sites. This can further guide the design of novel anti-cancer therapeutics selectively targeting mutant K-Ras. The sampling of the conformational states of K-Ras leading to the opening of a particular pocket had also been recapitulated in this study. **All the results have been compiled into a manuscript: Soni AK, Marzinek J, Chang YW, Verma C & Bond PJ. *Improved Benzene Mapping Method for revealing the cryptic pockets: The case of K-Ras protein*, 2017, to be submitted.**

This method was also able to expose cryptic pockets on the intradimeric region of E protein (between two monomers of the dimer) of DENV and ZIKV serotypes with a capability of blocking the low pH conformational change and dimer-to-trimer transition. The most dominant conformation of the intradimer pocket was a subject to docking and scoring tools using FDA approved library. In this work ten small drug molecules are proposed which can effectively bind all serotypes of dengue and Zika virus. In particular, an antifungal agent had shown highest affinity towards Zika (-11.0 kcal mol^{-1}) and other serotypes of dengue virus with its interaction with both the chains of the dimer. Also, the other several pockets generated at the interface of two monomers or at the interface of two dimers conserved across all the serotypes from our benzene mapping simulation bound with the modified chemical groups of the selected ligands and antibody.

In future, the selected FDA approved drugs should be experimentally tested along with the antibody bound which can lead to novel therapeutic pathway against all flaviviruses studied here. The pockets in K-Ras are of greater relevance in biology and characterization of these pockets computationally will open a way to perform experiments where the new drugs can be designed accordingly for the desired

inhibition against oncogenic mutant Ras. The probe density benzene clusters generated can be very well correlated to the “druggability” of the binding site which can be in future applied to other important proteins and biomolecules such as MDM2 and ion channels protein target for anticancer therapies. The general applicability of the developed benzene mapping method in finding the cryptic pockets will also be a subject of future research.

BIBLIOGRAPHY

1. Kozakov, D., Hall, D. R., Napolean, R. L., Yueh, C., Whitty, A. & Vajda, S. New Frontiers in Druggability. *Journal of Medicinal Chemistry* **58,** 9063–9088 (2015).
2. Feixas, F., Lindert, S., Sinko, W. & McCammon, J. A. Exploring the role of receptor flexibility in structure-based drug discovery. *Biophysical Chemistry* **186,** 31–45 (2014).
3. Le Guilloux, V., Schmidtke, P. & Tuffery, P. Fpocket: an open source platform for ligand pocket detection. *BMC Bioinformatics* **10,** 168 (2009).
4. Ngan, C. H., Bohnuud, C., Mottarella, S. E., Beglov, E., Miller, E. A., Hall, D. R., Kozakov, D. & Vajda, S. FTMAP: Extended protein mapping with user-selected probe molecules. *Nucleic Acids Res.* **40,** (2012).
5. Kobayashi, M., Retra, K., Figaroa, F., Hollander, J. G., Ab, E., Heetebrij, R. J., Irth, H. & Siegel, G. Target immobilization as a strategy for NMR-based fragment screening: comparison of TINS, STD, and SPR for fragment hit identification. *J. Biomol. Screen.* **15,** 978–89 (2010).
6. Scott, D. E., Bayly, A. R., Abell, C. & Skidmore, J. Small molecules, big targets: drug discovery faces the protein-protein interaction challenge. *Nat. Rev. Drug Discov.* **15,** 533–550 (2016).
7. Tan, Y. S., Spring, D. R., Abell, C. & Verma, C. S. The Application of Ligand-Mapping Molecular Dynamics Simulations to the Rational Design of Peptidic Modulators of Protein-Protein Interactions. *J. Chem. Theory Comput.* **11,** 3199–3210 (2015).
8. Halgren, T. New method for fast and accurate binding-site identification and analysis. *Chem. Biol. Drug Des.* **69,** 146–8 (2007).
9. Schmidtke, P., Bidon-Chanal, A., Luque, F. J. & Barril, X. MDpocket : Open Source Cavity Detection and Characterization on Molecular Dynamics Trajectories. *Bioinformatics* btr550 (2011). doi:10.1093/bioinformatics/btr550
10. Sciabola, S., Stanton, R. V., Mills, J. E., Flocco, M. M., Baroni, M., Cruciani, G., Parruccio, F. & Masson, J. S. High-throughput virtual screening of proteins using GRID molecular interaction fields. *J. Chem. Inf. Model.* **50,** 155–169 (2010).
11. Paramo, T., East, A., Garzón, D., Ulmschneider, M. B. & Bond, P. J. Efficient characterisation of protein cavities within molecular simulation trajectories: trj_cavity. *J Chem Theory Comput* 140328121610008 (2014). doi:10.1021/ct401098b
12. Tan, Y. S., Śledź, S., Lang, S, Stebbs, C. J., Spring, D. R., Abell, C. & Best, R. B. Using ligand-mapping simulations to design a ligand selectively targeting a cryptic surface pocket of polo-like kinase 1. *Angew. Chemie - Int. Ed.* **51,** 10078–10081 (2012).
13. Guvench, O. & MacKerell, A. D. Computational fragment-based binding site identification by ligand competitive saturation. *PLoS Comput. Biol.* **5,** (2009).
14. Prakash, P., Hancock, J. F. & Gorfe, A. A. Binding hotspots on K-ras: Consensus ligand binding sites and other reactive regions from probe-based molecular dynamics analysis. *Proteins Struct. Funct. Bioinforma.* **83,** 898–909 (2015).
15. Prakash, P., Sayyed-Ahmad, A. & Gorfe, A. A. pMD-Membrane: A Method for Ligand Binding Site Identification in Membrane-Bound Proteins. *PLoS Comput. Biol.* **11,** (2015).
16. Lama, D., Brown, C. J., Lane, D. P. & Verma, C. S. Gating by Tryptophan 73 Exposes a Cryptic Pocket at the Protein-Binding Interface of the Oncogenic eIF4E Protein. *Biochemistry* **54,** 6535–6544 (2015).
17. Mattos, C., Bellamacina, C. R., Peisach, A., Pereira, A., Vitkup, D., Petsko, G. A. & Ringe, D. Multiple solvent crystal structures: Probing binding sites, plasticity and hydration. *J. Mol. Biol.* **357,** 1471–1482 (2006).
18. Grant, B. J., Lukman, S. Hocker, H. J., Sayyah, J., Brown, J. H., McCammon, J. A. & Gorfe A. A., Novel allosteric sites on ras for lead generation. *PLoS One* **6,** (2011).
19. McCormick, F. K-Ras protein as a drug target. *Journal of Molecular Medicine* **94,** 253–258 (2016).
20. Prior, I. A., Lewis, P. D. & Mattos, C. A comprehensive survey of ras mutations in cancer. *Cancer Research* **72,** 2457–2467 (2012).
21. Cox, A. D. & Der, C. J. Ras history: The saga continues. *Small GTPases* **1,** 2–27 (2010).
22. Rosnizeck, I. C, Spoerner, M., Harch, T., Kreitner, S., Filchtinski, D., Herrmann, C., Engel, D., König, B. & Kalbitzer, H. R. Metal-bis(2-picolyl)amine complexes as state(T) inhibitors of activated ras protein. *Angew. Chemie - Int. Ed.* **51,** 10647–10651 (2012).
23. Ostrem, J. M., Peters, U., Sos, M. L., Wells, J. a & Shokat, K. M. K-Ras(G12C) inhibitors allosterically control GTP affinity and effector interactions. *Nature* **503,** 548–51 (2013).

24. Marcus, K. & Mattos, C. Direct attack on RAS: Intramolecular communication and mutation-specific effects. *Clinical Cancer Research* **21,** 1810–1818 (2015).
25. Hocker, H. J. *et al.* Andrographolide derivatives inhibit guanine nucleotide exchange and abrogate oncogenic Ras function. *Proc. Natl. Acad. Sci. U. S. A.* **110,** 10201–6 (2013).
26. Buhrman, G., O'Connor, C., Zerbe, B., Kearney, B. M., Napolean, R., Kovrigina, E. A., Vajda, S., Kozakov, D., Kovrigin, E. L. & Mattos, C. Analysis of binding site hot spots on the surface of Ras GTPase. *J. Mol. Biol.* **413,** 773–789 (2011).
27. McCarthy, M., Prakash, P. & Gorfe, A. A. Computational allosteric ligand binding site identification on Ras proteins. *Acta Biochimica et Biophysica Sinica* **48,** 3–10 (2015).
28. Maurer, T., Garrenton, L. S, Oh, A., Pitts, K., Anderson, D. J., Skelton, N. J., Fauber, B. P., Pan, B., Malek, S., Stokoe, D., Ludlam, M. J. C., Bowman, K. K., Wu, J., Gianetti, J. M., Starovasnik, M. A., Mellman, E., Jackson, P. K., Rudolph, J., Wnag, W. & Fang, G. Small-molecule ligands bind to a distinct pocket in Ras and inhibit SOS-mediated nucleotide exchange activity. *Proc. Natl. Acad. Sci. U. S. A.* **109,** 5299–304 (2012).
29. Murray, C. J. L. L. & Lopez, A. D. Measuring the global burden of disease. *N. Engl. J. Med.* **369,** 448–57 (2013).
30. Vannice, K. S., Durbin, A. & Hombach, J. Status of vaccine research and development of vaccines for dengue. *Vaccine* **34,** 2934–2938 (2016).
31. Mukhopadhyay, S., Kuhn, R. J. & Rossmann, M. G. A structural perspective of the flavivirus life cycle. *Nat. Rev. Microbiol.* **3,** 13–22 (2005).
32. Huber, R. G., Marzinek, J. K., Holdbrook, D. A. & Bond, P. J. Multiscale molecular dynamics simulation approaches to the structure and dynamics of viruses. *Progress in Biophysics and Molecular Biology* (2016). doi:10.1016/j.pbiomolbio.2016.09.010
33. Yennamalli, R., Subbarao, N., Kampmann, T., McGeary, R. P., Young, P. R. & Kobe, B. Identification of novel target sites and an inhibitor of the dengue virus E protein. *J Comput Aided Mol Des* **23,** 333–341 (2009).
34. Fuzo, C. A. & Degrève, L. New pockets in dengue virus 2 surface identified by molecular dynamics simulation. *J. Mol. Model.* **19,** 1369–1377 (2013).
35. Modis, Y., Ogata, S., Clements, D. & Harrison, S. C. A ligand-binding pocket in the dengue virus envelope glycoprotein. *Proc. Natl. Acad. Sci. U. S. A.* **100,** 6986–91 (2003).
36. Wang, Q. Y. & Shi, P. Y. Flavivirus Entry Inhibitors. *ACS Infect. Dis.* **1,** 428–434 (2016).
37. Ismail, N. A. & Jusoh, S. A. Molecular Docking and Molecular Dynamics Simulation Studies to Predict Flavonoid Binding on the Surface of DENV2 E Protein. *Interdiscip. Sci. Comput. Life Sci.* (2016). doi:10.1007/s12539-016-0157-8
38. Panya, A., Bangphoomi, K., Choowongkomon, K. & Yenchitsomanus, P.-T. Peptide Inhibitors Against Dengue Virus Infection. *Chem. Biol. Drug Des.* 148–157 (2014).
39. Chopade, A. R., Sayyad, F. J. & Pore, Y. V. Molecular Docking Studies of Phytocompounds from the Phyllanthus Species as Potential Chronic Pain Modulators. *Sci Pharm* **83,** 243–267 (2015).
40. Aubé, J. Drug repurposing and the medicinal chemist. *ACS Medicinal Chemistry Letters* **3,** 442–444 (2012).
41. Rapaport, D. C. Molecular dynamics simulation. *Comput. Sci. Eng.* **1,** 537–542 (1999).
42. Best, R. B., De Sancho, D. & Mittal, J. Residue-specific ??-helix propensities from molecular simulation. *Biophys. J.* **102,** 1462–1467 (2012).
43. Van Der Spoel, D. *et al.* GROMACS: Fast, flexible, and free. *Journal of Computational Chemistry* **26,** 1701–1718 (2005).
44. Cavasotto, C. N. & Orry, A. J. W. Ligand docking and structure-based virtual screening in drug discovery. *Curr. Top. Med. Chem.* **7,** 1006–1014 (2007).
45. Kitchen, D. B., Decornez, H., Furr, J. R. & Bajorath, J. Docking and scoring in virtual screening for drug discovery: methods and applications. *Nat. Rev. Drug Discov.* **3,** 935–49 (2004).
46. Trott, O. & Olson, A. J. Software news and update AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **31,** 455–461 (2010).
47. Tanchuk, V. Y., Tanin, V. O., Vovk, A. I. & Poda, G. A New Scoring Function for Molecular Docking Based on AutoDock and AutoDock Vina. *Curr. Drug Discov. Technol.* **12,** 170–8 (2015).
48. Seeliger, D. & De Groot, B. L. Ligand docking and binding site analysis with PyMOL and Autodock/Vina. *J. Comput. Aided. Mol. Des.* **24,** 417–422 (2010).
49. Schulz-Gasch, T. & Stahl, M. Binding site characteristics in structure-based virtual screening: Evaluation of current docking tools. *J. Mol. Model.* **9,** 47–57 (2003).

50. Carlson, H. a. Protein flexibility is an important component of structure-based drug discovery. *Curr. Pharm. Des.* **8,** 1571–1578 (2002).
51. Abreu, R. M. V, Froufe, H. J. C., Queiroz, M. J. R. P. & Ferreira, I. C. F. R. Selective Flexibility of Side-Chain Residues Improves VEGFR-2 Docking Score using AutoDock Vina. *Chem. Biol. Drug Des.* **79,** 530–534 (2012).
52. Totrov, M. & Abagyan, R. Flexible ligand docking to multiple receptor conformations: a practical alternative. *Current Opinion in Structural Biology* **18,** 178–184 (2008).
53. Torda, A. E. & van Gunsteren, W. F. Algorithms for clustering molecular dynamics configurations. *J. Comput. Chem.* **15,** 1331–1340 (1994).
54. Hess, B., Bekker, H., Berendsen, H. J. C. & Fraaije, J. G. E. M. LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.* **18,** 1463–1472 (1997).
55. DeLano, W. Pymol: An open-source molecular graphics tool. *CCP4 Newsl. Protein Crystallogr.* **700,** (2002).
56. Mark, P. & Nilsson, L. Structure and dynamics of the TIP3P, SPC, and SPC/E water models at 298 K. *J. Phys. Chem. A* **105,** 9954–9960 (2001).
57. Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. a & Case, D. a. Development and testing of a general amber force field. *J. Comp. Chem.* **25,** 1157–74 (2004).
58. Meagher, K. L., Redman, L. T. & Carlson, H. A. Development of polyphosphate parameters for use with the AMBER force field. *J. Comput. Chem.* **24,** 1016–1025 (2003).
59. Brereton, R. G. Steepest Ascent, Steepest Descent and Gradient Methods. *Comprehensive Chemometrics* **1,** 577–590 (2010).
60. Van Gunsteren, W. F. & Berendsen, H. J. C. A Leap-frog Algorithm for Stochastic Dynamics. *Mol. Simul.* **1,** 173–185 (1988).
61. Essmann, U., Perera, L., Berkowitz, M. L., Darden, T., Lee, H. & Pederson, L. G. A smooth particle mesh Ewald method. *J Chem Phys* **103,** 8577–8593 (1995).
62. Bussi, G., Donadio, D. & Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **126,** (2007).
63. Parrinello, M. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **52,** 7182 (1981).
64. Humphrey, W., Dalke, A. & Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graph.* **14,** 33–38 (1996).
65. Daura, X., Gademann, K., Jaun, B., Seebach, D., Van Gunsteren, W. F. & Mark, A. E. Peptide Folding: When Simulation Meets Experiment. *Angew. Chem. Int. Ed* **38,** 38: 236-240 (1999).
66. Martin, J. Letellier, G., Marin, A., Taly, J., de Brevern, A. G. & Gibret, J. Protein secondary structure assignment revisited: a detailed analysis of different assignment methods. *BMC Struct. Biol.* **5,** 17 (2005).
67. Irwin, J. J., Sterling, T., Mysinger, M. M., Bolstad, E. S. & Coleman, R. G. ZINC: A free tool to discover chemistry for biology. *Journal of Chemical Information and Modeling* **52,** 1757–1768 (2012).
68. Laskowski, R. A. & Swindells, M. B. LigPlot+: Multiple ligand-protein interaction diagrams for drug discovery. *J. Chem. Inf. Model.* **51,** 2778–2786 (2011).
69. Dai, L., Song, J., Lu*, X.*,Deng Y. Q., Musyoki, A. M., Cheng, H., Zhang, Y., Yuwan, Y., Song, H., Haywood, J., Xiao, H., Yan, J. & Schi, Y. Structures of the Zika Virus Envelope Protein and Its Complex with a Flavivirus Broadly Protective Antibody. *Cell Host Microbe* **19,** 696–704 (2016).
70. Canny, S. A., Cruz, Y., Southern, M. R. & Griffin, P. R. PubChem promiscuity: A web resource for gathering compound promiscuity data from PubChem. *Bioinformatics* **28,** 140–141 (2012).