



OPEN

Characterization of the binding of MRTX1133 as an avenue for the discovery of potential KRAS^{G12D} inhibitors for cancer therapy

Abdul Rashid Issahaku^{1,2}, Namutula Mukelabai², Clement Agoni¹, Mithun Rudrapal^{3✉}, Sahar M. Aldosari⁴, Sami G. Almalki⁴ & Johra Khan⁴

The Kirsten rat sarcoma (KRAS) oncoprotein has been on drug hunters list for decades now. Initially considered undruggable, recent advances have successfully broken the jinx through covalent inhibition that exploits the mutated cys12 in the switch II binding pocket (KRAS^{G12C}). Though this approach has achieved some level of success, patients with mutations other than cys12 are still uncatered for. KRAS^{G12D} is the most frequent KRAS mutated oncoprotein. It is only until recently, MRTX1133 has been discovered as a potential inhibitor of KRAS^{G12D}. This study seeks to unravel the structural binding mechanism of MRTX1133 as well as identify potential drug leads of KRAS^{G12D} based on structural binding characteristics of MRTX1133. It was revealed that MRTX1133 binding stabilizes the binding site by increasing the hydrophobicity which resultantly induced positive correlated movements of switches I and II which could disrupt their interaction with effector and regulatory proteins. Furthermore, MRTX1133 interacted with critical residues; Asp69 (− 4.54 kcal/mol), His95 (− 3.65 kcal/mol), Met72 (− 2.27 kcal/mol), Thr58 (− 2.23 kcal/mol), Gln99 (− 2.03 kcal/mol), Arg68 (− 1.67 kcal/mol), Tyr96 (− 1.59 kcal/mol), Tyr64 (− 1.34 kcal/mol), Gly60 (− 1.25 kcal/mol), Asp12 (− 1.04 kcal/mol), and Val9 (− 1.03 kcal/mol) that contributed significantly to the total free binding energy of − 73.23 kcal/mol. Pharmacophore-based virtual screening based on the structural binding mechanisms of MRTX1133 identified ZINC78453217, ZINC70875226 and ZINC64890902 as potential KRAS^{G12D} inhibitors. Further, structural optimisations and biochemical testing of these compounds would assist in the discovery of effective KRAS^{G12D} inhibitors.

Kirsten Rat Sarcoma (KRAS) viral oncogene is the most studied of the Ras family of proteins (HRAS, NRAS and KRAS) with the most mutations that result in a range of cancers including colorectal cancer (CRC), non-small cell lung cancer (NSCLC) and pancreatic adenocarcinoma (PDAC)¹. In these cancers, KRAS is responsible for approximately 45% of cases in CRC, 35% in NSCLC and 90% in PDAC with poor prognosis and resistance to standard-of-care chemotherapy, denoting a critically unmet medical need that requires novel therapies to target KRAS². However, unlike other oncogenic proteins such as Vascular Endothelial Growth Factor Receptor (VEGFR) and Epidermal Growth Factor Receptor (EGFR) that are easily druggable^{3,4}, drugging KRAS was hampered by the lack of a well-defined binding site and its high affinity towards an abundance of GDP/GTP in the cytosol until recently^{5,6}.

KRAS rotates between a GTP-bound and GDP-bound states representing an active and an inactive states respectively whose conversion is controlled by GEFs which produce the exchange from GDP to GTP, and GAP-which influence the hydrolysis of GTP^{7,8}. The activation by binding GTP perturbs the effector protein binding

¹Bio-computation and Drug Design Laboratory, School of Health Sciences, University of KwaZulu-Natal, Westville Campus, Durban 4001, South Africa. ²Department of Physiotherapy, School of Health Sciences, University of KwaZulu-Natal, Westville Campus, Durban 4001, South Africa. ³Department of Pharmaceutical Chemistry, Rasiklal M. Dhariwal Institute of Pharmaceutical Education and Research, Pune, Maharashtra 411019, India. ⁴Department of Medical Laboratory Sciences, College of Applied Medical Sciences, Majmaah University, Al-Majmaah 11952, Saudi Arabia. ✉email: rsmrpal@gmail.com

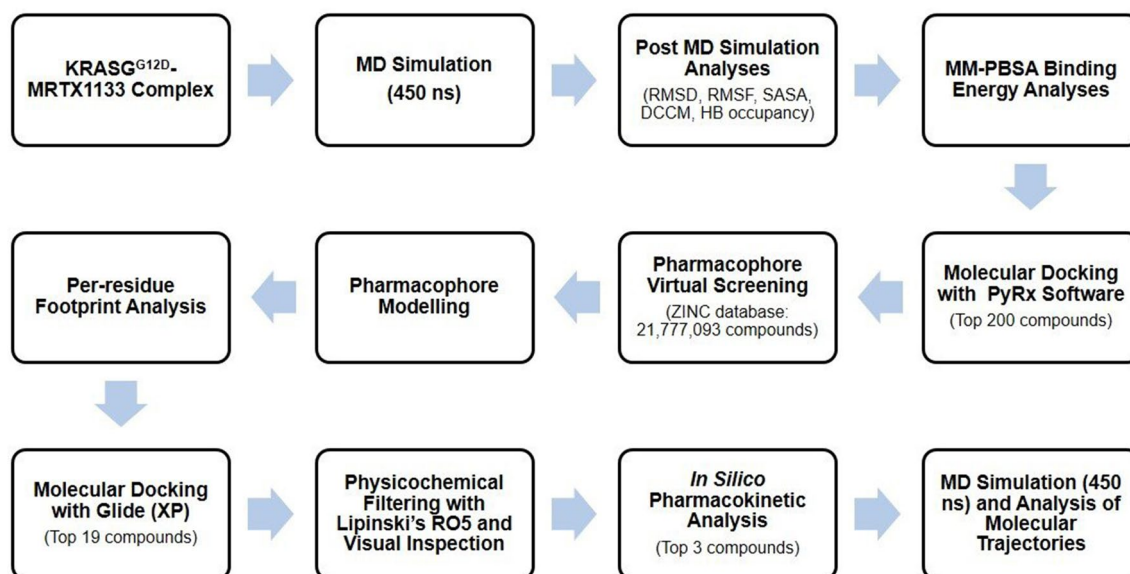


Figure 1. Schematic diagram of the approach applied to this study. First four sections indicate the characterization approach applied to MRTX1133, while rest of the sections depict the process of hit compounds identification.

regions of KRAS by altering their conformations⁹ which resultantly activate downstream signaling pathways^{10,11}. The mutations at codon 12 to other residue with a wide steric footprint hinders the binding of KRAS with GTPase activating proteins (GAPs) thereby interfering with GTP hydrolysis¹². This interference locks KRAS in its innately active GTP-bound state hence favoring KRAS effector pathways activation¹³. Being a cellular signal transducer, the activation of KRAS biologically affects three downstream signaling pathways such as the rapidly accelerated fibrosarcoma-mitogen-activated protein kinase-extracellular signal-regulated kinase (RAF-MEK-ERK)¹⁴, RAL guanine nucleotide dissociation stimulator-RAL (RALGDS-RAL)¹⁵ and phosphoinositide 3-kinase-protein kinase B-mammalian target of rapamycin (PI3K-AKT-Mtor)¹⁶. Responding to other stimuli and growth factors, KRAS together with its downstream pathways regulate cellular functions including survival, proliferation, migration, differentiation, adhesion and apoptosis which becomes corrupted upon mutations of KRAS resulting in over activation and cancer¹⁷. These mutations which mostly occur at codon 12, 13 and 61 (G12D, G12V, G12C, G13D and Q61R) accounting for 70% of cases with missense mutation at codon 12 and 13 limiting the interaction of the protein with GAPs thus preventing the switching off of KRAS¹⁸. Due to the reasons stated in the first stanza, it took several decades of research to finally drug KRAS. However, this breakthrough is limited to G12C wherein the inhibitory activity depends on an electrophilic warhead forming a stable covalent bond with the nucleophilic head of the mutated cysteine¹⁹ thus these inhibitors occupy the induced switch II pocket^{20,21}. This selective targeting of KRAS^{G12C} has left KRAS cancer patients driven by other mutations with unmet medical needs, spurring further search for inhibitors with novel mechanism of action to reversibly inhibit KRAS. KRAS^{G12D} is the most common occurring mutation among the KRAS driven tumors representing 33% of cases¹⁸ and has therefore been on the top list of drug hunters, however, due to the absence of a reactive residue within the binding pocket, developing inhibitors has been challenging wherein limited cellular activities were observed^{22–24}. Nonetheless, the constant perseverance through optimization and structure-based activity improvements by drug developers has yielded MRTX1133, a reversible, potent and selective KRAS^{G12D} inhibitor which has shown efficacy in KRAS^{G12D} mutant xenograft mouse tumor model²⁵.

In this current study, we sought to investigate and provide molecular insights and conformational dynamics bases for MRTX1133 evasion of KRAS^{G12D} resistance through molecular dynamics simulations and structural analysis (Fig. 1). These insights would be further exploited for the design of structure-based pharmacophore towards the identification of hit compounds with KRAS^{G12D} inhibitory potentials from the ZINC database.

Methodology

System preparation. The x-ray diffraction crystal structure with 1.3 Å resolution in complex with MRTX1133 and GDP was retrieved from the protein data bank (PDB ID: 7RPZ)²⁶. The retrieved protein was prepared using UCSF Chimera candidate version 1.13.1²⁷ by removing all nonstandard residues including water molecules. The prepared complex, which included the KRAS^{G12D}, MRTX1133 and the GDP was saved as a KRAS^{G12D}-MRTX1133 complex. The protein containing only the GDP was also saved as holo. The systems (KRAS^{G12D}-MRTX1133 and holo protein) were then prepared for MD simulation by adding amber charges to MRTX1133 and the GDP, while all atoms of the protein complexes were minimized using the protein preparation wizard of Schrödinger Maestro (*Schrödinger Suite 2017–4 Protein Preparation Wizard*, 2018).

Molecular dynamics (MD) simulations. The systems set up were taken through a long-range molecular dynamics simulation. The graphical processing unit version of Particle Mesh Ewald Molecular Dynam-

ics (PMEMD) engine in AMBER 18 package²⁸ was used to perform the simulation. The AMBER force field (FF14SB)²⁹ and the integrated pdb4amber program were used to parameterize and modify the proteins, respectively, while the GDP and MRTX1133 were parameterized using the antechamber module of AMBER18. This module was also used to create partial atomic charges for the ligands. Afterwards, the tLeap module of AMBER 18 was utilized to neutralize the systems by adding Na⁺ and Cl⁻ counter ions as well as solvating the systems by suspending them in Transferable Intermolecular Potential with 3 Points (TIP3P) water box of 10 Å. Input coordinates and topology files for the systems were then generated using the tLeap module. The systems were then initially minimized for 2500 steps with the steepest descent and then switched to conjugate gradient algorithm for further 5000 steps minimization using 500 kcal/molÅ⁻² restraint potential. Afterwards the systems were then heated gradually from 0 to 300 K for 5 ps in isothermal-isobaric NTP ensemble with Langevin thermostat at a pressure of 1 bar using Brendsen barostat³⁰. Equilibrations were then performed to relax the systems at 300 K for 1000 ps without energy restraint. Molecular dynamics production was then performed for 450 ns with the SHAKE algorithm³¹ used to restrain all hydrogen bonds at 2 fs time step.

Post MD analysis. The trajectories and coordinates generated from the MD simulations were all analyzed using the CPPTRAJ and PTRAJ modules³² of AMBER 18. Origin data tool³³ was used to plot all graphs while visualization of snapshots was performed with Discovery Studio (BIOVIA 2017) and UCSF Chimera^{37,34}.

Thermodynamics estimations. We utilized the molecular mechanics Poisson-Boltzmann surface area (MM-PBSA) method to predict the binding free energy of the complexes using the mmpbsa.py module³⁵ incorporated in AMBER 18. This method is widely used due to its efficiency and low cost. From each of the trajectories generated from the systems, the binding free energy was averaged over 30,000 snapshots using a time step of 50 ps. The energies calculated from this method are depicted as follows:

$$\Delta G_{\text{bind}} = G_{\text{complex}} - G_{\text{receptor}} - G_{\text{ligand}} \quad (1)$$

$$\Delta G_{\text{bind}} = E_{\text{gas}} + G_{\text{sol}} - T\Delta S \quad (2)$$

where ΔG_{bind} is taken to be the sum of the gas phase and solvation energy terms less the entropy ($T\Delta S$) term

$$E_{\text{gas}} = E_{\text{int}} + E_{\text{vdw}} + E_{\text{ele}} \quad (3)$$

where E_{gas} is the total of the AMBER force field internal energy terms. E_{int} (bond, angle and torsion), the covalent van der Waals (E_{vdw}) and the non-bonded electrostatic energy component (E_{ele}). The solvation energy is denoted by the equation:

$$G_{\text{sol}} = G_{\text{PB}} + G_{\text{non-polar}} \quad (4)$$

$$G_{\text{non-polar}} = \text{SASA} \quad (5)$$

The polar solvation contribution is denoted as G_{PB} and $G_{\text{non-polar}}$ represents the non-polar contribution energy and is computed from the solvent assessable surface area (SASA) which is obtained by the use of 1.4 Å water probe radius. Per-residue decomposition analyses were also performed to estimate individual energy contribution of the residues of the substrate pocket to the affinity and stabilisation of the compounds.

Dynamic cross correlation matrices (DCCM). Analysis of the dynamic cross correlation of MD simulation trajectory is used to investigate the correlated motion of residual-based fluctuations during the course of simulation. This investigation was performed on the two switches of the protein which are very critical for interaction with effector proteins. The cross-correlation coefficient C_{ij} for the pair of C- α atoms i and j were calculated based on the formula below:

$$C_{ij} = \frac{\langle \Delta r_i^* \Delta r_j \rangle}{\left(\langle \Delta r_i^2 \rangle \langle \Delta r_j^2 \rangle \right)^{1/2}} \quad (6)$$

where Δr_j and Δr_i are the displacement from the mean locations of the j th and i th atom respectively. The cross-correlation coefficient C_{ij} has a range of +1 to -1 wherein the lower and upper limits denote fully anti-correlated and correlated movements during the course of simulation. The CCPTRAJ module incorporated in AMBER 18 package was used to calculate the dynamic cross-correlation.

Principal component analysis (PCA). We used the PCA approach to evaluate changes in the compound's dynamics and conformations during the period of simulation in reference to the MRTX1133. This method comprises the eigenvalues and eigenvectors which respectively informs on the quantum and direction of motions. The two principal components were calculated by employing the CPPTRAJ module, whereas the conformational behaviours were projected along the directions of these two components (ev1/PC1 vs. ev2/PC2).

Pharmacophore modelling. The phase model was utilized to generate a pharmacophore model centred on PRED analysis of the KRAS^{G12D}-MRTX1133 complex. The residues that contributed energies above -1.0 kcal/

mol to the total free binding energy of MRTX1133 complexing were identified and their respective interactions with the moieties on MRTX1133 were used as the bases for the pharmacophore model.

Virtual screening and docking of hits. The pharmacophore model created was then used for virtual screening of the ZINC database³⁶ for inhibitors that possessed good fitness with RMSD values less than 0.9 Å. The retrieved inhibitors were then prepared for structure-based screening with PyRx software³⁷ with a grid box of center: $x = 0.48195$, $y = 3.92564$, and $z = -22.5689$ with size of: $x = 20.183$, $y = 17.1422$, and $z = 18.9123$. Compounds that showed binding energies lower than -8.5 kcal/mol were eliminated while those scoring -8.6 kcal/mol and above were further processed for docking with Glide³⁸. The KRAS^{G12D} protein was prepared using the protein preparation wizard suit incorporated in Schrödinger. For the binding site, a grid box centred at MRTX1133 was utilized to cater for maximum length of the ligand. OPLS3e force field³⁹ was applied to input partial charges and default van der Waals scaling was employed. The selected compounds were flexibly docked with extra precision (XP), with a maximum of writing out one pose per compound as well as post docking minimisation of the compounds. MRTX1133 was redocked and the docking conformation was realigned with its co-crystallized conformation to determine the root mean square deviation (RMSD) to validate the docking program.

ADME and toxicity prediction. The pharmacokinetics of the hit compounds were analysed with SwissADME⁴⁰, an online server that evaluates the hit compounds functionality inside the body. Evaluating the toxicity of compounds is also an important part of the drug development process. This allows for the determining of the toxic features of a compound through the estimated lethal dose (LD₅₀) in mg/kg weight. ProTox-II, an online server to determine the toxicity of compounds by incorporating fragment propensity, machine learning and molecular similarity was utilized⁴¹. The compounds are classified into categories ranging from class I to VI based on their safety according to the globally harmonized system of classification of labelling chemicals^{42,43}.

MD simulation of hit compounds. The simulation of selected hit compounds was performed using the same protocol used for the KRAS^{G12D}-MRTX1133 complex as explained above.

Results and discussion

The KRAS protein and its other family members (HRAS and NRAS) consist of two main components, a membrane-targeting hyper variable region (HVR) and the conserved catalytic domain⁴⁴. The conserved domain is made up of five alpha helices that surrounds six beta sheets and engages in interactions with regulators and effectors which is conditioned by conformational changes of two flexible regions. These flexible regions, switch I and Switch II cover residue 30–38 and 59–76 respectively⁴⁵. We therefore investigated the impact of MRTX1133 on these switches.

MRTX1133 binding stabilizes KRAS^{G12D} binding site. Binding allosterically to KRAS^{G12D}, MRTX1133 effectiveness as a small molecule inhibitor is to instigate a communication across the structure to a spatially distant site specifically the switches (I and II) whose conformational changes influence the catalytic domain's interactions with effector and regulatory proteins⁴⁶. This communication will depend on the immediate impact of the small molecule on the binding pocket of the protein. Thus, we first sought to determine the effect of MRTX1133 binding on the allosteric binding pocket residues by computing the deviations of the C- α atoms of the pocket residues over 450 ns simulation relative to the starting structure. The root-mean square deviation (RMSD) matrix was therefore utilized and the results displayed in Fig. 2. The occupied site showed an average RMSD values of 1.80 Å, while the unoccupied pocket displayed RMSD values of 2.61 Å. These results indicate MRTX1133 presence induced stability of the pocket residues, a deviation from the normal state. An analysis of the solvent accessibility surface area (SASA) of the pocket was also investigated and shown in Fig. 2. Average SASA results of 1169.66 Å² and 856.44 Å² were presented by the holo and MRTX1133 sites respectively. This matrix is informative of the surface area of the pocket that is exposed to solvent interaction and thus could also be indicative of the folding and/or realignment of the residues towards the hydrophilic or hydrophobic core of the protein. As observed, the presence of MRTX1133 reduced the surface area of the site that is available for external interaction in a holo state by 313.22 Å². Thus, the site residues favored the hydrophobic core of the protein in MRTX1133 presence. The hydrophobicity as presented in Fig. 2C estimates the degree of affinity between the residues side chains and water. This feature is critical for the stabilization and folding of the protein structure⁴⁷. As observed, the deeper the residues in the pocket the higher the hydrophobicity. This is depicted by the color bar wherein high hydrophilicity is denoted by blue color while hydrophobicity is brown. Taken together, the stability and the increase in hydrophobicity of the site residues could condition signaling to the protein–protein interface that influences its overall therapeutic effect.

Since proteins are sensitive to stimuli, these induced changes could be the initiation process for the inhibition of KRAS. To further ascertain the impact of these changes on distant critical domains, we investigated the residual response to these pocket site changes.

KRAS switches reacts to binding site changes through correlated residual movements. The variations in internal correlation movements of the residues of switch I and II were analyzed by cross-correlation matrices of the C- α atoms fluctuations. Figure 3 shows the DCCM plots of switch I and II. The most negative regions (blue–black) and the most positive regions (yellow–red) reflect strong anti-correlated movements and strong correlated motions of the specific residues respectively. Juxtaposing the movements of switch I in

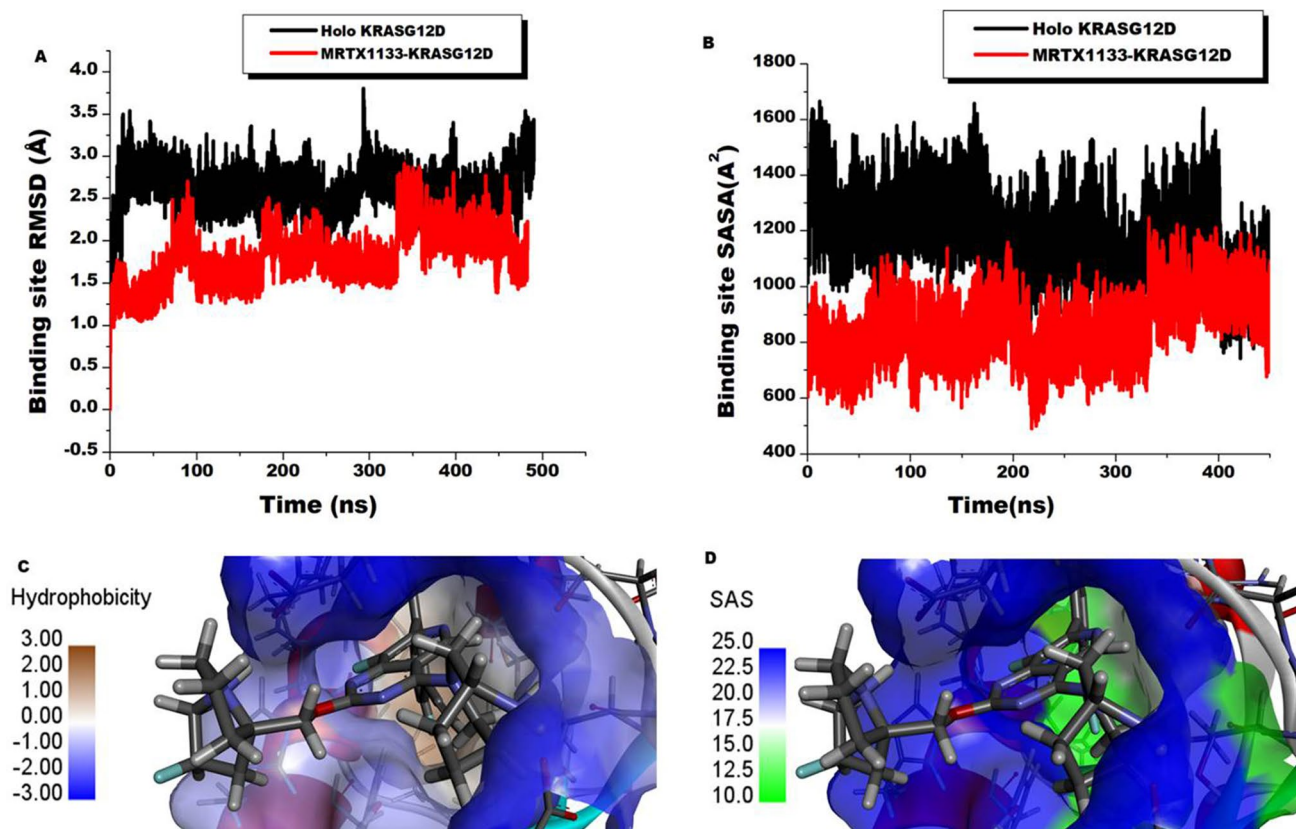


Figure 2. Comparative stability and solvent accessibility surface area of the binding site of KRAS^{G12D} upon the binding of MRTX1133 and the holo protein. (A) Shows the RMSD plots of the C- α atoms of the binding site residues indicating MRTX1133 binding stabilized the residues (red) relative to the holo (black). (B) Shows the SASA plots of the binding site residues indicating a reduction in surface area availability upon MRTX1133 binding (red) relative to the holo protein (black). (C) Shows the hydrophobicity of the binding site. (D) Shows the solvent accessible of area of the binding site.

a complexed state (Fig. 3A) and the holo (Fig. 3A1) shows a more correlated movement of the residues of the complexed switch than the holo as evidenced by the degree of red regions presented. This trend was observed in the comparison of switch II of the complexed state (Fig. 3B) and the holo switch II (Fig. 3B1). Generally, the binding of MRTX1133 presented strong correlated residual movements of Switch I and II than the unbound (holo) over the period of simulation. This is observed from the degree of positive (red) regions depicted by the Fig. 3. This differential observation could influence the switches interaction with effector and regulator proteins thus highlighting the basis of MRTX1133 inhibitory prowess.

Atomistic interactions underscoring the potency of MRTX1133. Upon attaining the impact of the binding of MRTX1133 on the structural dynamics of the protein, we further probed at the atomic level to ascertain the underlying interactions that accounts for the observable structural changes. Thus, six snapshots were taken from the 450 ns simulation trajectory for analysis. Trajectories generated during the process of simulation contain events that helps to unravel structural activities and atomistic insights that occurred at a specific time-frame⁴⁸. Snapshot visualization analysis revealed MRTX1133 formed varied bond types with the binding site residues. While some of these interactions endured though the 450 ns simulations, others were altered during the process as observed in Fig. 4. The sp^2 hybridized O atom of Gly60 was observed to form a conventional hydrogen bond with the HNO3 atom of MRTX1133 in all the observed snapshots. Asp69 and His 95 were also observed to form an enduring hydrogen bond with the compound. The sp^2 hybridized OD1 atom of Asp69 interacted with the H02 atom of MRTX1133, while the HE2 atom of His95 interacted with both the sp^2 hybridized NO1 and sp^3 hybridized F01 atoms of MRTX1133. Thr58 interacted with the HN03 atom of the compound through the sp^2 hybridized O atom, which was observed at 100 ns, 200 ns, and 300 ns. Arg68 was also observed to interact through conventional hydrogen bond through the HH21 atom and the sp^2 hybridized N03 atom of MRTX1133 at 200 ns and 300 ns, while Ser65 was observed to interact at only 300 ns through the H atom and the sp^3 hybridized O02 atom. To highlight the percentage of time occupied by these interactions in the 450 ns simulation period, the hydrogen bond occupancy was calculated and presented in Table 1. The hydrogen bond occupancy captures hydrogen interactions including those occurred outside the trajectory captured by the snapshots.

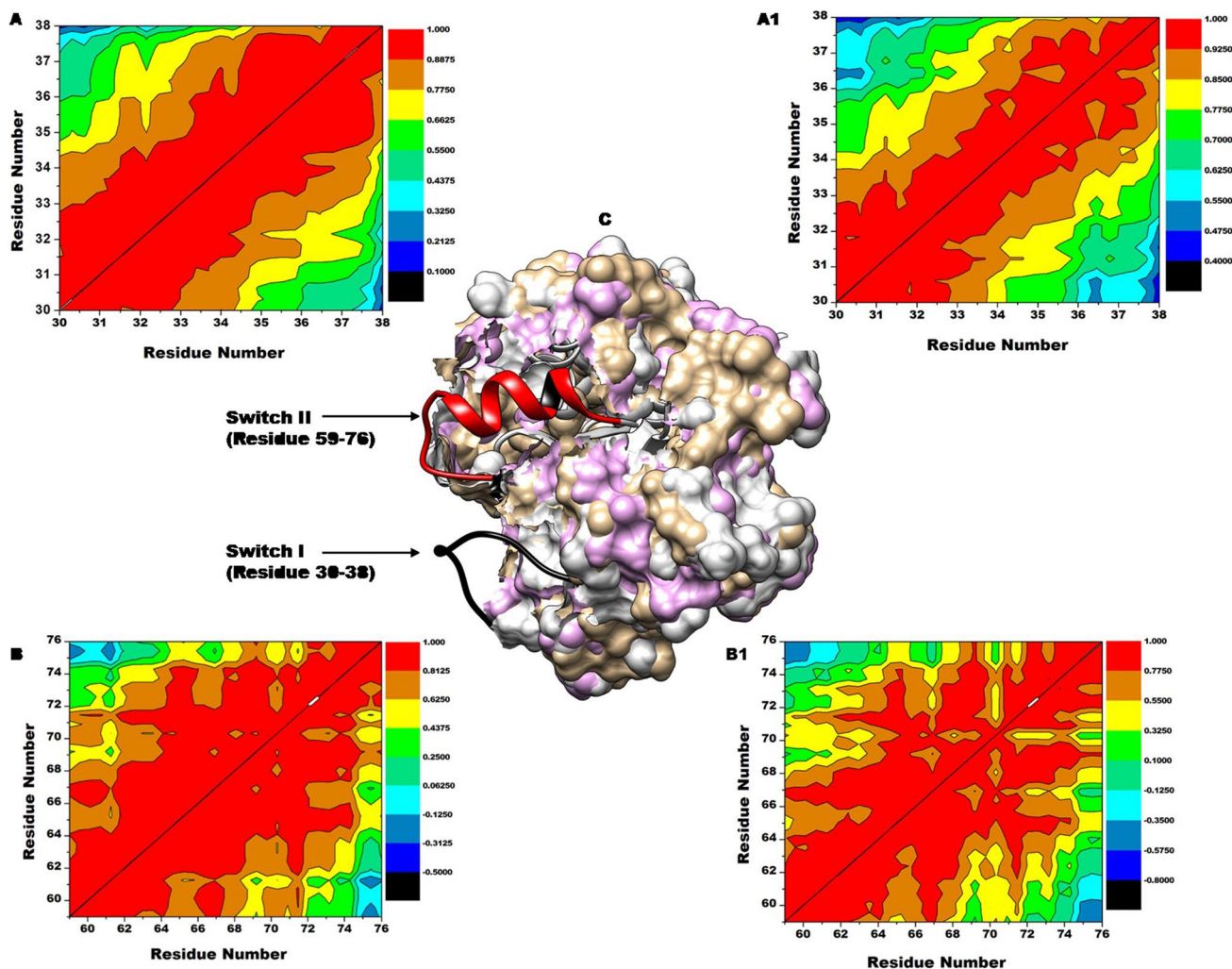


Figure 3. Cross-correlation matrices of the fluctuations of coordinates of C- α atoms of the switches around their mean positions during the 450 ns simulation. DCCM plots of switch I MRTX1133 complex (A), Switch I holo (A1), Switch II MRTX1133 complex (B), and (B1) Switch II holo protein. This indicates MRTX1133 complexing induced a more correlated movement. (C) KRAS structure showing switch I (black) and switch II (red).

Thermodynamics of MRTX1133. The binding of a ligand to proteins underpins a variety of recognition processes in biological systems. Among these range of processes is the binding energy which underscores the affinity of the ligand and receptor. The free binding energy therefore provides information on the spontaneity, ligand permeation and reaction kinetics⁴⁹. The free binding energy profile of MRTX1133 was therefore estimated using the MM-PBSA method as outlined earlier and presented in Table 2. MRTX1133 showed total free binding energy (ΔG) of -73.16 kcal/mol. This low binding energy is indicative of its spontaneity and permeation in to the hydrophobic core of the protein thus engaging in many interactions with the site residues which result in more release of energy. This observation could therefore explain the low pERK IC_{50} of 2 nM recorded experimentally for MRTX1133 thus its observed potency. The gas phase, electrostatic and van der Waals interaction were observed as prominent contributors to the total free binding energy.

Since the binding interactions between the ligand and the binding site residues is crucial to the quantum of energy released (total free binding energy), we decomposed the total free binding energy into individual residue contributions. This allowed us to identify the residues that individually contributed the most towards the complexing and stability of MRTX1133. The results are presented in Fig. 5. Asp69 (-4.54 kcal/mol), His95 (-3.65 kcal/mol), Met72 (-2.27 kcal/mol), Thr58 (-2.23 kcal/mol), Gln99 (-2.03 kcal/mol), Arg68 (-1.67 kcal/mol), Tyr96 (-1.59 kcal/mol), Tyr64 (-1.34 kcal/mol), Gly60 (-1.25 kcal/mol), Asp12 (-1.04 kcal/mol), and Val9 (-1.03 kcal/mol) were observed as residue with prominent contributions. Others include: Ile100 (-0.98 kcal/mol), Glu62 (-0.73 kcal/mol), Arg102 (-0.53 kcal/mol), Gly10 (-0.50 kcal/mol), Val103 (-0.48 kcal/mol) and Phe78 (-0.48 kcal/mol) as depicted in Fig. 4. We therefore monitored these residues interactions with MRTX1133 over 450 ns simulation period to ascertain the functional groups of the ligand they engaged with.

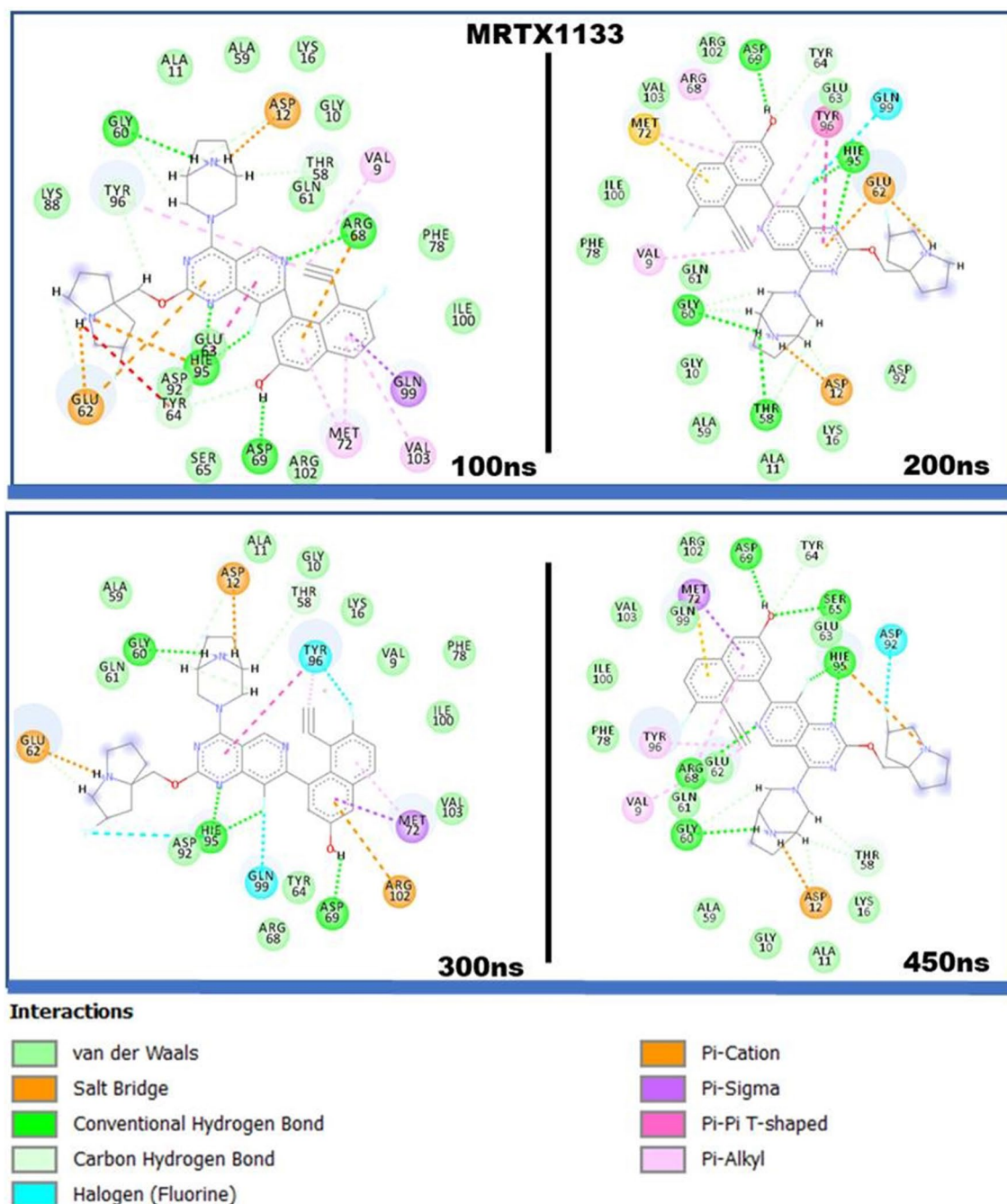


Figure 4. 2D interactions of MRTX1133 with KRAS^{G12D} binding site residues. Snapshots depicts interactions at 100 ns, 200 ns, 300 ns and 450 ns. MRTX1133 engages in varying interactions.

Pharmacophore model generation. The functional or structural components of a bioactive compound that possess biological activity is often referred to as pharmacophores. These elements define important, steric and electronic function-determining points necessary for an optimal interaction with a pharmacological target, thus they represent a common denominator of a class of compounds that exhibit a similar pharmacological profile and are therefore recognized by the same site of the target protein⁵⁰. A pharmacophore should therefore satisfy the criterion of interacting with crucial binding site residues. As such, we therefore identified the components of MRTX1133 that interacted with the residues that resulted in high energy contributions as crucial elements for pharmacophore modelling. Per-residue decomposition (PRED)-centered pharmacophore model was therefore generated from the 450 ns simulations performed for MRTX1133. Residues with contributions higher than -1 kcal/mol to the total free binding energy of MRTX1133 as presented above were monitored for the period of simulation to identify functional groups they engaged with. These crucial functional groups on MRTX1133 and their interactions is shown in Fig. 6. The functional groups were then used to create a pharmacophore model for compounds screening in ZINC database.

Compounds	H-acceptor	H-donor	Occupancy (%)	Average distance (Å)
MRTX1133	ASP69-OD1	MRTX1133-HO02	99.7	2.59
	ASP12-OD1	MRTX1133-HN02	65.4	2.80
	GLY60-O	MRTX1133-HN03	53.9	2.82
	GLU62-OE1	MRTX1133-H06	50.1	2.79
	THR58-O	MRTX1133-HN03	43.4	2.80
	MRTX1133-N01	HIS95-HE2	39.9	2.91
	MRTX1133-N03	ARG68-HH11	20.5	2.91
	MRTX1133-F01	GLN99-HE22	10.2	2.88
	GLU63-O	MRTX1133-H02	5.0	2.90
	MRTX1133-O02	SER65-H	2.5	2.74
	MRTX1133-N01	TYR64-HH	0.9	2.95
	ALA59-O	MRTX1133-HN03	0.7	2.87
	MRTX1133-N02	TYR96-HH	0.5	2.83
	MRTX1133-F02	LYS88-HZ3	0.2	2.90

Table 1. Dynamic hydrogen bond interactions of MRTX1133 at KRAS^{G12D} binding site over the 450 ns MD simulation.

System	Energy components (kcal/mol)				
	ΔE_{vdw}	ΔE_{ele}	ΔG_{gas}	ΔG_{sol}	ΔG_{bind}
KRAS ^{G12D} -MRTX1133	-62.40 ± 4.73	-633.03 ± 30.37	-695.43 ± 29.75	622.27 ± 29.75	-73.16 ± 5.15

Table 2. Binding energy profile of MRTX1133 compounds. ΔE_{ele} = electrostatic energy; ΔE_{vdw} = van der Waals energy; ΔG_{bind} = total binding free energy; ΔG_{sol} = solvation free energy ΔG = gas phase free energy.

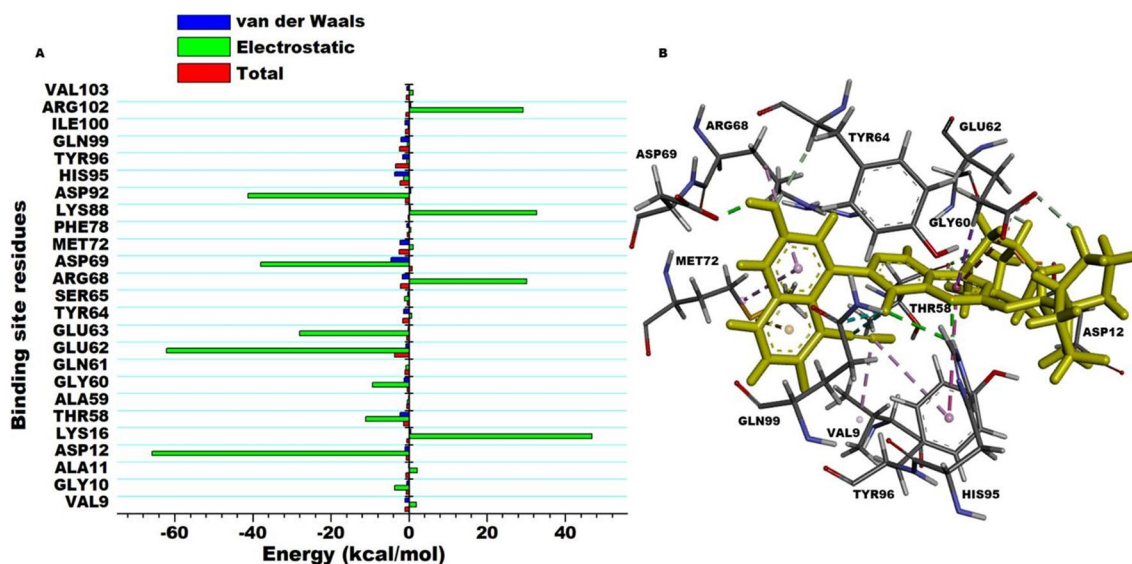


Figure 5. Per-residue decomposition analysis showing binding site residue energy contributions towards the complexing of MRTX1133. Electrostatic energy featured prominently towards the total free binding energy. (A) Shows the energy plots. (B) Shows the interacting residues.

Pharmacophore based virtual screening. The pharmacophore hypothesis generated was used as a 3D structural query with several combinations. The ZINC database was therefore screened for matching compounds. On zincpharmer, the 2 aromatic rings of the pyridopyrimidine, 2 hydrogen acceptors and 4 hydrophobic moieties were selected due to their interaction with the high energy contributing residues. Molecular weight ≤ 500 Da, rotatable bonds ≤ 10 were set as pre-requisite for the generation of ideal hits. A total of 21,777,093 compounds were screened. An initial 200 compounds were retrieved. These hit compounds were then docked into the allosteric binding site occupied by MRTX1133 using PyRx software and the compounds exhibiting high docking

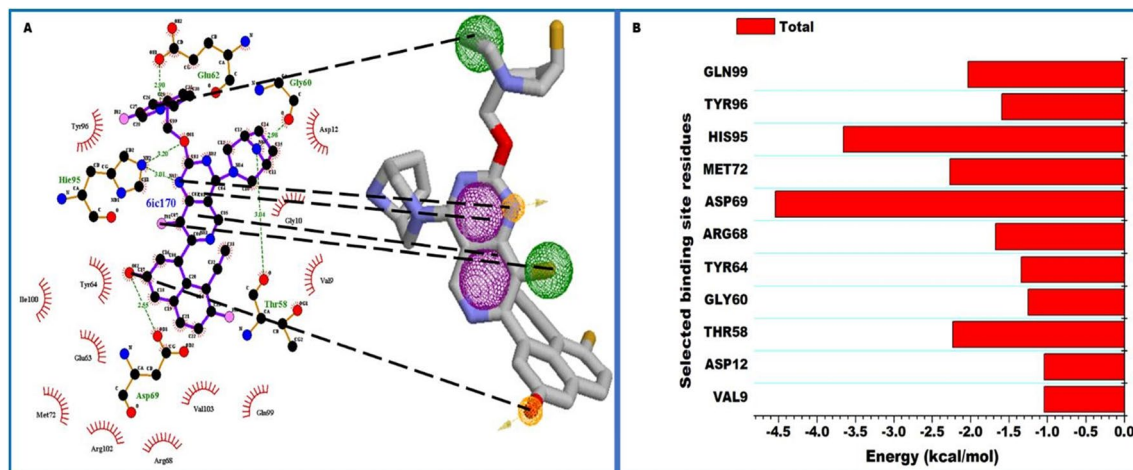


Figure 6. A pharmacophore model of MRTX1133. **(A)** 2D illustration of MRTX1133 interacting with binding site residues after 450 ns MD simulation. **(B)** The selected pharmacophoric moieties in MRTX1133, which interacted with the most negative energy contributing binding site residues. **(C)** PRED plot highlighting only the total binding energy contributed by the selected active site residues.

ZINC Compound ID	Pyrx docking score (kcal/mol)	Glide XP Score (kcal/mol)	2D Structure
ZINC78453217	-10.1	-9.77	
ZINC70875226	-10.6	-9.34	
ZINC64890902	-9.8	-10.0	
MRTX1133	-11.4	-10.92	

Table 3. Summary of binding energy scores obtained from PyRx and Glide XP for the identified top four compounds.

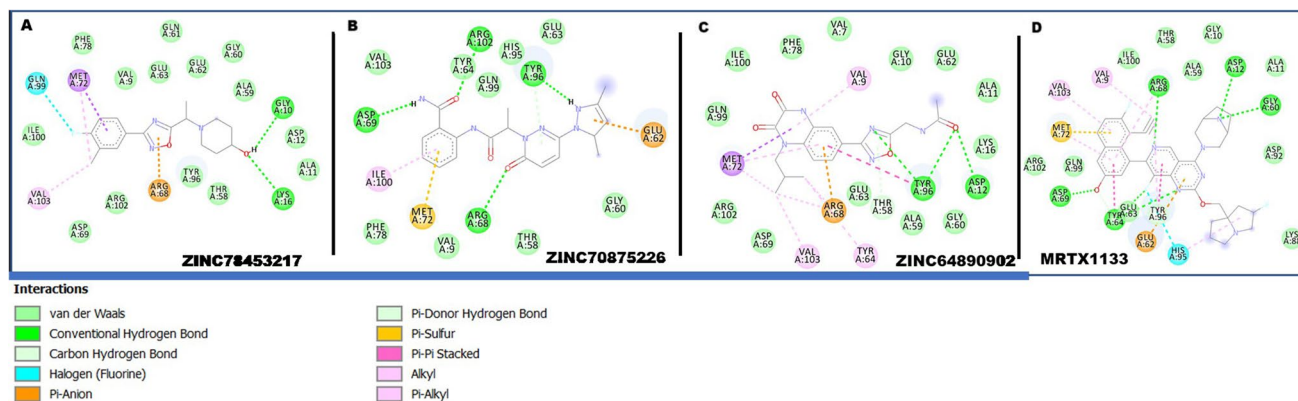


Figure 7. 2D molecular interactions of MRTX1133 and selected compounds within the binding site of KRAS allosteric binding site. The selected compounds and MRTX1133 show similar interactions with the binding site residues suggesting the selected compounds have the potential to elicit similar therapeutic effects.

scores were further selected. Nineteen (19) compounds were selected and further subjected to docking using GLIDE docking suit incorporated in Schrödinger with an extra precision (XP). Three compounds were then selected from the compounds for further analysis. The results of the compounds are presented in Table 3 and in Table S1.

Hit compound interactions within the binding pocket. After the docking of the compounds, the interactions that characterized their binding was further investigated. The type of interactions exhibited by a compound in the binding site of a protein is crucial in determining its therapeutic influence on the protein⁵¹. After docking the selected compounds, we sought to ascertain the interactions that feature between the compounds and the binding site residues. We thus visualized the complexes in comparison to MRTX1133. The known inhibitor, MRTX1133, was observed to engage in differential interactions with the site residues. Key among the interaction types are conventional and carbon hydrogen interactions, pi anion and alkyl interactions, van der Waals and halogen interactions. These interactions between the inhibitor and the residues varied depending on the functional groups. The selected compounds were also observed to interact with the site residues through similar interaction types as presented in Fig. 7.

Physicochemical properties and drug-likeness of the selected hits. To ascertain information into the drug-likeness of the hit compounds and their pharmacokinetic properties, a comparative analysis between the compounds and MRTX1133 was performed. The physicochemical properties of the three hit compounds and the other selected compounds were evaluated and presented in Tables 4 and S2 respectively. SwissADME, an online server was employed to profile their physicochemical properties.

Generally, an increase in the molecular weight of a therapeutic agent reduces its concentration at the intestinal epithelial surface and hence decreases absorption⁵². Studies have, therefore, reported that this could prevent passive diffusion of the agent across the bilayer membrane. Thus, it is suggested that the molecular weight of drug candidates should be less than 500 g/mol in accordance with the Lipinski's rule of five (RO5)^{53,54}. The selected compounds together with the hits all possessed molecular weight within the acceptable range for drug-like molecules as presented in Tables 4 and S2. Interestingly, these compounds outscored MRTX1133 regarding the molecular weight suggesting they would possess better intestinal absorption and cellular uptake.

Also, to determine the impact of molecular weight on bioactivity is the number of rotatable bonds, an excellent definition of molecular flexibility⁵⁵ (Lu et al. 2004). Verber et al.⁵⁶ reported that the number of rotatable bonds increases with an increase in molecular weight. According to the Lipinski's RO5, the number of rotatable bonds of a drug-like compound should be less than 10; this enhances absorption and bioavailability^{55,57,58}. As observed from our evaluation, all the compound possessed rotatable bonds below 10 including MRTX1133 which possessed 5 rotatable bonds. LogP is the logarithm of the coefficient of chemical compounds' permeation across n-octanol and water ($C_{\text{octanol}}/C_{\text{water}}$) and thus estimates the hydrophilicity of chemical compounds⁵⁹. High MLogP figure depicts a reduction in aqueous solubility and this could result in reduction of compound absorption. The hit compounds all showed lower MLogP values lower than MRTX1133 (Table 4). Again, LogS which is indicative of membrane permeability was used to evaluate the aqueous solubility of the selected and hit compounds in comparison to MRTX1133. An acceptable range of 0 to -6 for aqueous solubility accounts for 95% of approved drugs⁶⁰. The compounds all fell within the acceptable range and are therefore classified as soluble while MRTX1133 was classified as moderately soluble.

The topological polar surface area (TPSA) strings the oxygen and nitrogen (surface polar atoms) in relation with the hydrogen atoms that are attached in a compound. It helps in determining the potential of a compound to traverse the cells wherein a low TPSA value depicts the compound can permeate the cell^{61,62}. It also offers information on the volume and molecular size which controls a compound's physiological transport across the lipid bilayer including the blood-brain barrier (BBB) and gastrointestinal tract (GIT)⁶³. Elevated TPSA has been reported to interfere with the transportability of potential drugs which therefore reduces their bioactivity⁶⁴. As

Physicochemical properties	Compounds			
	ZINC78453217	ZINC70875226	ZINC64890902	MRTX1133
Formula	C16H20FN3O2	C19H20N6O3	C17H19N5O4	C33H33F3N6O2
MW (g/mol)	305.35	380.40	357.36	602.65
MLog P_{ow}	2.66	1.46	1.48	3.25
LogS (Ali) (mol/L)	- 3.42	- 3.13	- 2.71	- 5.44
Class	Soluble	Soluble	Soluble	Moderately soluble
TPSA (\AA^2)	62.39	124.90	122.88	109.42
Molar Refractivity	84.32	103.55	95.39	173.75
H-bond acceptors	6	5	6	8
H-bond donors	1	2	2	3
Rotatable bonds	3	6	6	5
Lipinski violations	No; 0	No; 0	No; 0	Yes; 1
Toxicity				
LD ₅₀ (mg/kg)	1000	200	649	2500
Toxicity class	4	3	4	5

Table 4. Physicochemical properties of the top 3 hit compounds relative to MRTX1133. ^aHBD = number of hydrogen bond donors (HBD ≤ 5); HBA = number of hydrogen bond donors (HBD ≤ 10); TPSA = total polar surface area; MLOGP = predicted octanol/water partition coefficient (MLOGP < 4.15); MW = molecular weight (g/mol) (MW ≤ 500); Rotatable bonds ≤ 8 ; TPSA = Topological polar surface area (≤ 140).

presented in Table 3, two compounds, ZINC70875226 and ZINC64890902 exhibited TPSA values higher than MRTX1133 while ZINC78453217 exhibited the least. All compounds however fell well with the acceptable range of TPSA of $\leq 140 \text{ \AA}^2$.

Hydrogen bond (HB) is one of the key factors that have been suggested to influence the solubility of therapeutic agents since they have to be severed to accelerate the transportation of the agents across the lipid bilayer membrane⁶⁵. Therefore, a high number of HB affects permeation by passive diffusion as a result of the reduction in partitioning from the aqueous environment into the lipid bilayer membrane. Hydrogen bond acceptors (HBA) and hydrogen bond donors (HBD) are therefore indicators of hydrogen bonding of a compound. According to the Lipinski's RO5, drug-like compounds should possess HBA ≤ 10 and HBD of ≤ 5 as the basis for oral bioavailability and activity⁵². The results computed in Tables 4 and S2 shows that all selected and hit compounds qualify for adequate hydrogen bond acceptance and donation.

Impact of hit compounds on the conformational dynamics of KRAS^{G12D}. The three hit compounds identified from the zinc database were further subjected to verification through molecular dynamics simulations. The four compounds were simulated for a period of 450 ns to reveal possible molecular events that underscore their binding to KRAS^{G12D}. Unlike molecular docking that only proffer analysis on static complexes, molecular dynamics provides the opportunity to assess the events of the complexes over a period of time mimicking real-life situations. The complexes involving the compounds; ZINC78453217, ZINC70875226, and ZINC64890902 were therefore simulated. The compounds stability within the binding site and their permeation into the hydrophobic core of the protein together with their impact on the stability of the binding site as well as the global conformation dynamics of KRAS^{G12D} upon ligation were investigated.

Hit compounds penetrates deeper into the hydrophobic core of KRAS^{G12D} with resultant stability. The solvent accessibility surface area (SASA) of the compounds was evaluated to ascertain the surface area available for aqueous interaction during the simulation period. This information provides insights into the level of penetration of the compounds into the hydrophobic core of the protein⁶⁶. Deeper penetration will therefore decrease the surface area availability of the compounds and vice versa. The compounds presented average SASA values of $34.89 \pm 15.37 \text{ \AA}^2$, $75.48 \pm 36.56 \text{ \AA}^2$, and $128.08 \pm 39.10 \text{ \AA}^2$ for ZINC78453217, ZINC70875226, and ZINC64890902 respectively. These results depict ZINC78453217 penetrated deeper into the hydrophobic core compared to the other compounds since it presented the least available solvent accessible surface area. Binding and permeation of a compound to a protein is conditioned by intermolecular interactions which consequently influences the stability of the compound within the binding pocket. As such, the stability of the compounds was estimated. The root mean-square deviation matrix was utilized to compute the atomic deviations of the compounds during the 450 ns period of simulation. This matrix is informative on the extent to which atomic positions vary over time which is indicative of the stability of the compounds. The compounds exhibited average RMSD values of $1.25 \pm 0.17 \text{ \AA}$, $1.55 \pm 0.20 \text{ \AA}$, and $1.84 \pm 0.28 \text{ \AA}$ for ZINC78453217, ZINC70875226, and ZINC64890902 respectively while MRTX1133 presented $0.97 \pm 0.12 \text{ \AA}$. These results indicate general stability of the compounds within the binding site. As observed from the values and Fig. 8, ZINC78453217 which exhibited the least SASA, is the most stable among the hit compounds suggesting it could possess higher therapeutic potentials. Afterwards we sought to determine the impact of these compounds on the binding pocket.

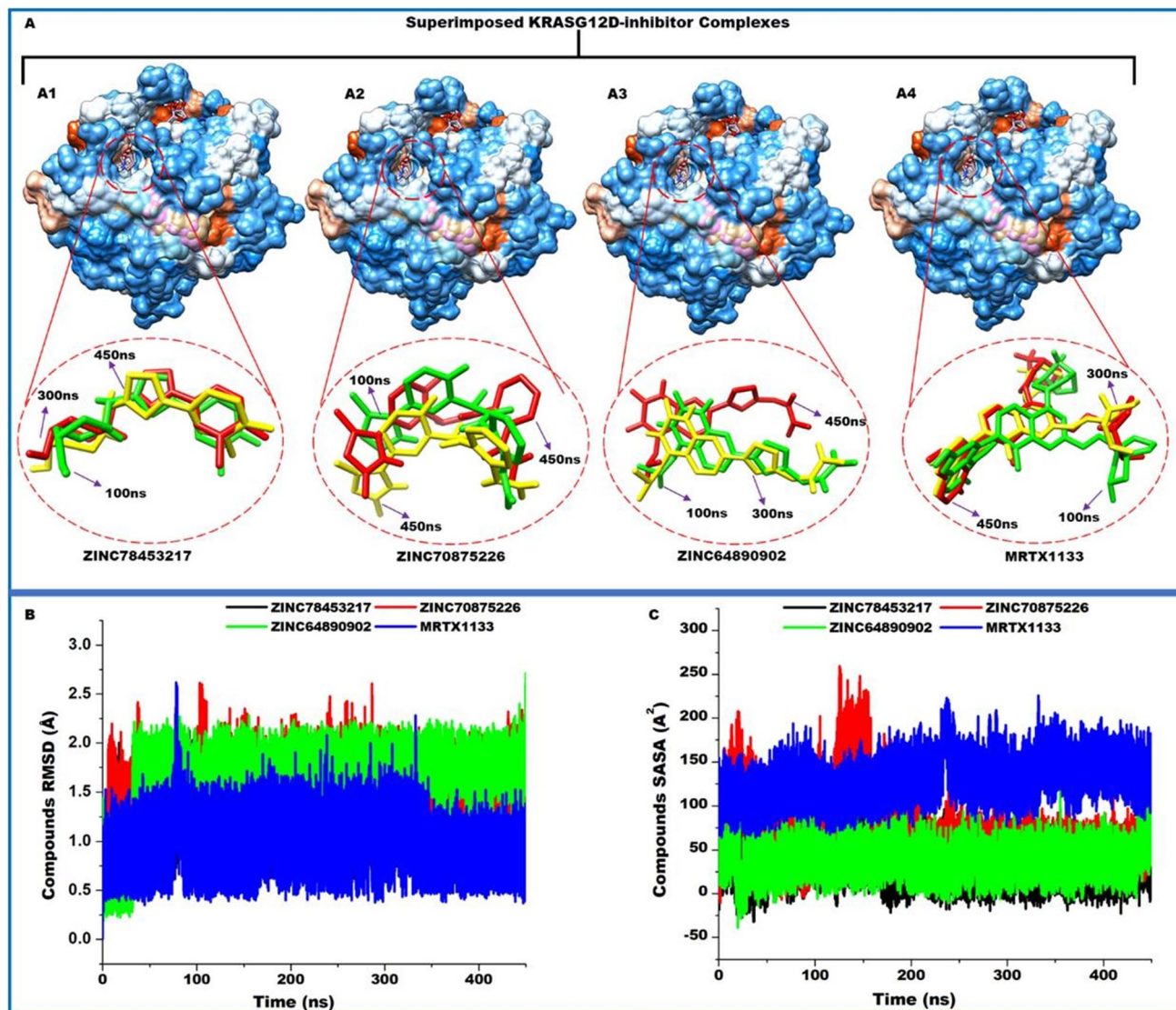


Figure 8. (A) Superimposition of KRAS^{G12D} complexed with ZINC78453217, ZINC70875226, ZINC64890902 and MRTX1133. Insert shows the binding orientation of each inhibitor at 100 ns, 300 ns, and 450 ns. (B) Comparative inhibitor stability plots using RMSD metrics for ZINC78453217, ZINC70875226, ZINC64890902 and MRTX1133. (C) Comparative inhibitor SASA plots of ZINC78453217, ZINC70875226, ZINC64890902 and MRTX1133.

The differential positioning and stability of the compounds relative to MRTX1133 was further ascertained by conducting the PCA of the compounds. We clustered the compounds movements in the direction of two principal components: PC1 and PC2. From this computation presented in Fig. 9, it is observed that the motions of ZINC64890902 over the simulation period is more dispersed than the other compounds including MRTX1133. ZINC78453217 was observed to display more compact motion during the simulation period while ZINC70875226 showed moderate dispersed motion relative MRTX1133. The compact motion exhibited by ZINC78453217 among the selected compounds corroborates with the RMSD and SASA findings wherein it showed the most stable with the least exposure to solvent environment among the compounds.

Hit compounds disrupt binding site conformations. The inhibitory or activation ability of bioactive compounds is often underscored by conformational changes they induce on the target protein. Conformational changes beyond or below a certain threshold result in the disruption of the basal functionalities of the protein⁶⁷. Thus, this phenomenon has been exploited by drug developers to contain certain abnormal protein activities. The effect of the compounds on the binding site of KRAS^{G12D} was therefore evaluated. Effects of compounds on the binding site and on the protein as a whole is often underscored by interactions that characterize the compound's binding. We, therefore, visualized snapshots from the simulation and the results presented in Fig. 10. This revealed the compounds engaged in varying interaction types during the course of simulation. The RMSD of the C- α atoms of the binding site residues were determined. As Fig. 11 discloses, the compounds exhibited vary-

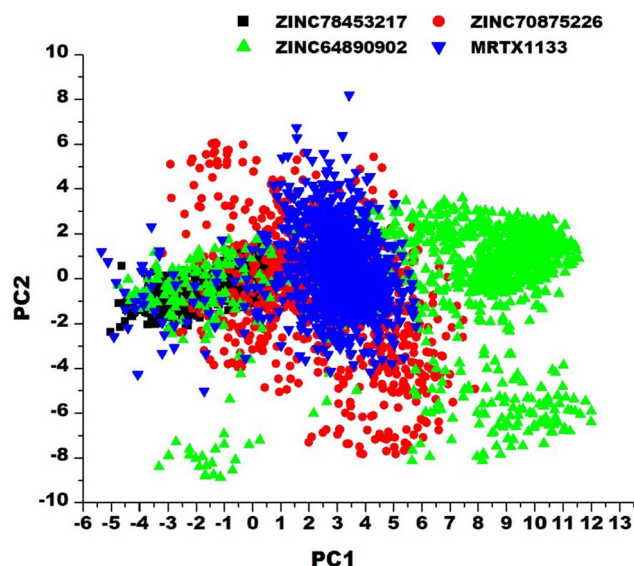


Figure 9. Principal component analysis of the motions of the selected compounds at the allosteric binding site of KRAS^{G12D}.

ing effects on the binding site. They presented average values of $2.12 \pm 0.18 \text{ \AA}$, $3.16 \pm 0.48 \text{ \AA}$, and $3.69 \pm 0.46 \text{ \AA}$, for ZINC78453217, ZINC70875226, and ZINC64890902, respectively. ZINC78453217 induced a stable binding site comparatively below MRTX1133 and the holo (Fig. 2) while the other compounds destabilized the binding site beyond the holo. These observations suggest the compounds elicit bioactivity on the KRAS^{G12D} binding pocket and therefore suggests their potentials as therapeutic agents.

Hit compounds induces global stability with reduced flexibility of KRAS^{G12D}. Proteins are sensitive to external stimuli and thus in most instances react through global conformational changes. These changes as earlier stated could influence the normal functions of the proteins. We, therefore, sought to determine the impact of the hit compounds on the entire KRAS^{G12D} through estimating the stability, flexibility and compactness. These was achieved through estimating the RMSD, the root mean-square fluctuation (RMSF) and the radius of gyration (RoG) of the C- α atoms of the protein over the 450 ns simulation period. While, the RMSD is informative of the general stability and system convergence over the period of simulation and therefore expresses the confidence in the ensuing deductions, the RMSF discloses the flexibility of the protein while RoG expresses the compactness of the protein. Higher RMSF values indicate more flexibility and vice versa while high RoG values indicate less compactness. As presented in Fig. 9, the systems presented average RMSD values of $1.38 \pm 0.14 \text{ \AA}$, $2.65 \pm 0.56 \text{ \AA}$, $3.00 \pm 0.46 \text{ \AA}$, $1.64 \pm 0.13 \text{ \AA}$ and $3.46 \pm 0.49 \text{ \AA}$ for the ZINC78453217, ZINC70875226, ZINC64890902, MRTX1133 and the holo respectively. The highest stability of the protein was observed with when ZINC78453217 was bound. This compound stabilized the protein below MRTX1133. Generally, the compounds induced elevated stability of the protein relative to the unbound. Probing the flexibility of KRAS^{G12D} showed a general reduction in the flexibility of the residues during the simulation period. Average RMSF values of $13.52 \pm 4.05 \text{ \AA}$, $13.65 \pm 4.16 \text{ \AA}$, $13.93 \pm 4.33 \text{ \AA}$, $13.73 \pm 4.11 \text{ \AA}$ and $14.11 \pm 4.36 \text{ \AA}$ for ZINC78453217, ZINC70875226, ZINC64890902, MRTX1133 and the holo protein, respectively. Zoning into the fluctuations that characterized the switches upon complexing with the compounds revealed ZINC78453217 and ZINC64890902 induced a reduction switch I residues fluctuations relative to the unbound. The other compounds together with MRTX1133 increased switch I fluctuation beyond the unbound. All the compounds however, reduced the fluctuations of Switch II residues below the unbound. The RoG also presented a similar trend of global protein behavior with the RMSF. A general reduction in the RoG values relative to the unbound was observed as presented in Fig. 11. This reduction indicates increased compactness and rigidity of KRAS^{G12D} upon the compounds binding corroborating the reduced fluctuations of the residues. Average RoG values of $15.34 \pm 0.05 \text{ \AA}$, $15.52 \pm 0.11 \text{ \AA}$, $15.40 \pm 0.07 \text{ \AA}$, $15.58 \pm 0.08 \text{ \AA}$, and $15.64 \pm 0.10 \text{ \AA}$ for ZINC78453217, ZINC70875226, ZINC64890902, MRTX1133 and holo protein respectively.

Binding energy landscape of hit compounds. The binding energy profile of the hit compounds were evaluated using the MM-PBSA method due to its popularity⁶⁸. The binding energies are pivotal to ligand-protein complexing by disclosing the spontaneity, folding, permeation and kinetic reactions of the ligation process⁵⁰. Higher binding affinity (most negative) indicates the strength of interactions involved in the ligation and thus reveals the potency of the compound. The hit compounds' energy profiles were therefore determined and presented in Table 5. Average total binding free energies (ΔG) of $-38.69 \pm 2.76 \text{ kcal/mol}$, $-41.80 \pm 7.67 \text{ kcal/mol}$, and $-35.24 \pm 3.99 \text{ kcal/mol}$ were estimated for ZINC78453217, ZINC70875226 and ZINC64890902 respectively.

These results suggest spontaneous compounds complexing with KRAS^{G12D} with ZINC70875226 showing the highest energy. The variations in the binding free energies presented by the compounds could be due to the

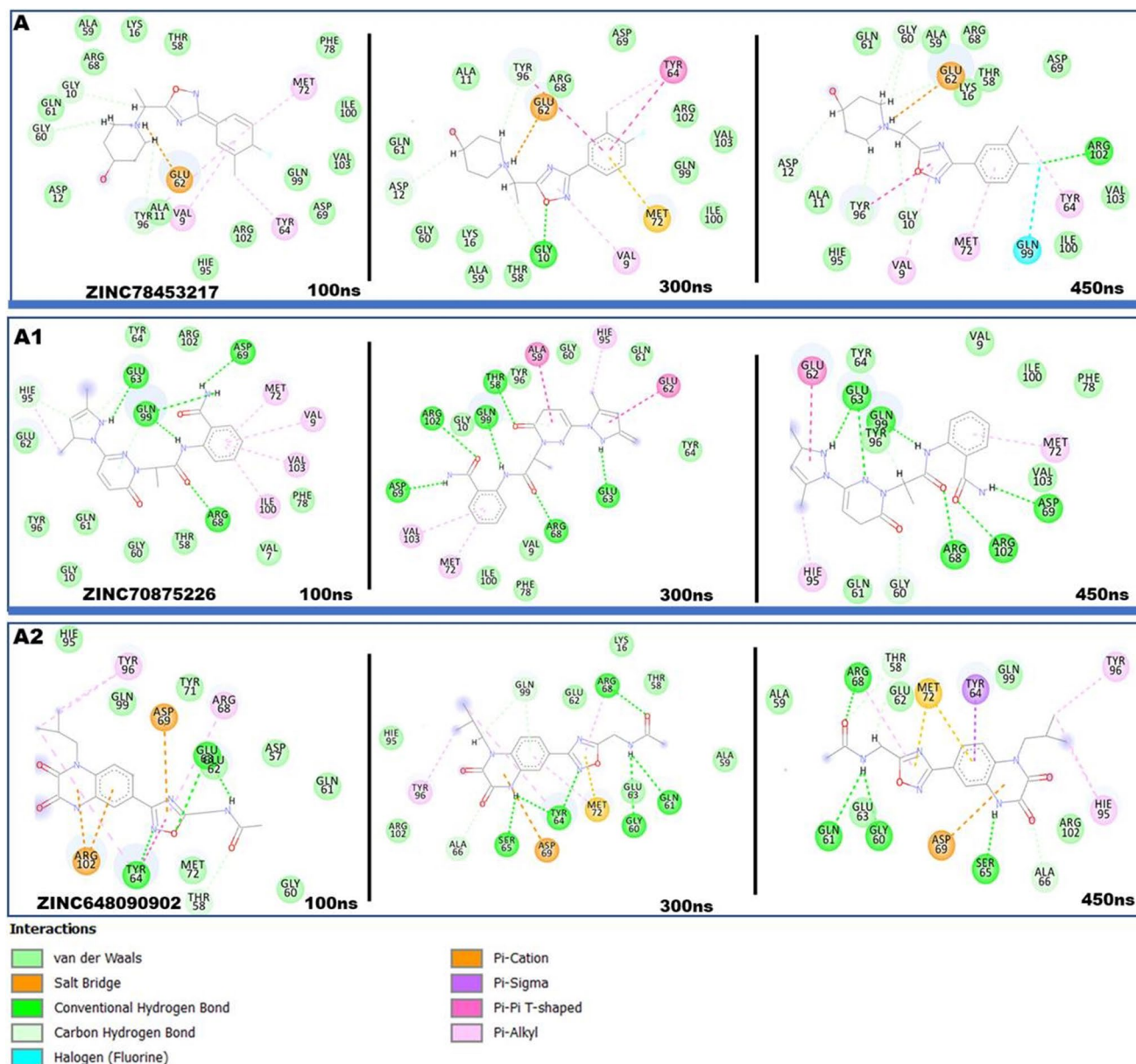


Figure 10. 2D interactions of hit compounds with KRAS^{G12D} binding site residues. Snapshots depicts interactions at 100 ns, 200 ns, 300 ns and 450 ns. (A) Shows ZINC78453217 interactions with the binding site residues. (A1) Shows ZINC70875226 interaction with the binding site residues. (A2) Shows ZINC64890902 interactions with the binding site interactions.

type of interactions that accompany their binding. Though these compounds generally presented lower total free binding energies compared to MRTX1133, nonetheless their impact on the protein indicates they elicit bioactive potentials and could therefore be optimized to improve their activity. Contributing significantly to the total free binding energies are electrostatic, van der Waals and the gas phase energies.

To summarize the entire study, ZINC78453217, ZINC70875226 and ZINC64890902 are identified as the most promising inhibitors of KRAS^{G12D} (Fig. 12).

Conclusion

Oncogenic KRAS has eluded drug developers for decades until covalent inhibition was strategically employed to target the mutated cysteine within the switch II binding pocket. Though this has achieved some success, cancers driven by other mutations has not been successful. The development of MRTX1133 to target KRAS^{G12D} has therefore been a sign of relief. In our quest to unravel the molecular events that highlight its potentials to inhibit KRAS^{G12D} as well as identify potential inhibitors based on its structural characteristics through PRED, we employed MD simulations and other computational techniques. The 450 ns simulation of MRTX1133 in complex with KRAS^{G12D} revealed MRTX1133 binding stabilized the binding site with a consequential reduction in the surface area of the site available for aqueous interaction. These changes resulted in a strong positive correlated

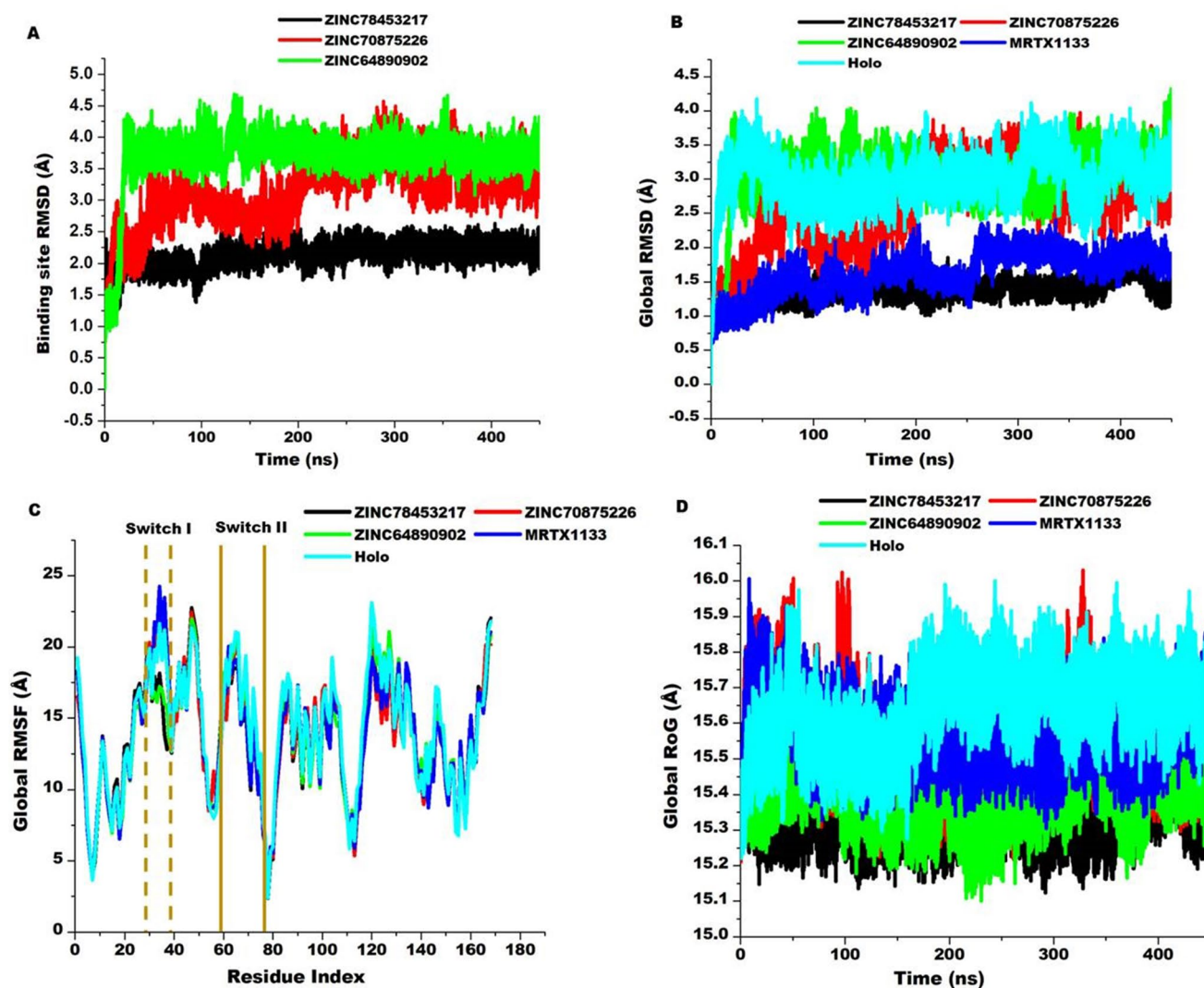


Figure 11. Comparative RMSD, RMSF and RoG plots of the hit compounds, MRTX1133 and unbound (holo) KRAS^{G12D} protein. (A) Shows the binding site RMSD elicited by the binding of the hit compounds. (B) Depicts the global stability of KRAS^{G12D} upon compounds binding. (C) Shows the residual fluctuations exhibited by the global protein residues during the simulation period. Highlights (deep yellow) show the switches regions. (D) Shows the RoG exhibited by the global protein. The bound conformations are contrasted to the unbound (Holo).

System	Energy components (kcal/mol)				
	ΔE_{vdw}	ΔE_{ele}	ΔG_{gas}	ΔG_{sol}	ΔG_{bind}
ZINC78453217	-41.23 ± 2.79	-296.46 ± 12.51	-337.69 ± 11.97	301.93 ± 11.97	-35.76 ± 3.14
ZINC70875226	-47.00 ± 5.87	-41.62 ± 9.77	-88.62 ± 13.51	46.82 ± 7.17	-41.80 ± 7.67
ZINC64890902	-40.35 ± 2.76	-262.97 ± 24.03	-303.33 ± 23.81	268.09 ± 21.92	-35.24 ± 3.99

Table 5. Binding energy profile of hit compounds. ΔE_{ele} = electrostatic energy; ΔE_{vdw} = van der Waals energy; ΔG_{bind} = total binding free energy; ΔG_{sol} = solvation free energy ΔG_{gas} = gas phase free energy.

movement of switch I and II which could disrupt KRAS^{G12D} interaction with effector and regulatory proteins. Again, the MRTX1133 was shown to exhibit strong binding affinity (-73.16 ± 5.15 kcal/mol) with KRAS^{G12D} corroborating experimental findings. This energy was then decomposed into individual residue contributions revealing Asp69, His95, Met72, Tyr64, Gly60, Asp12 and Val9 contributed the highest towards the binding energy. A pharmacophore was subsequently developed based on the prominent moieties of MRTX1133 that the high-energy contributing residues interacted. The pharmacophore model was then employed to screen the ZINC database towards identifying potential KRAS^{G12D} binders. Through molecular docking, MD simulation and ADMET

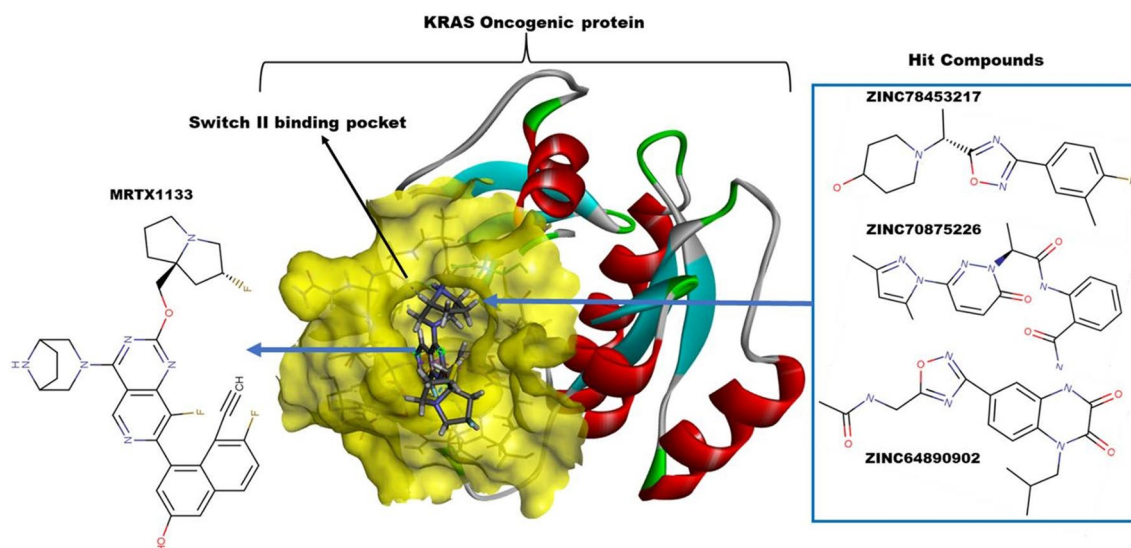


Figure 12. Most promising inhibitors of KRAS^{G12D}.

analysis, ZINC78453217, ZINC70875226 and ZINC64890902 were identified as the most promising binders of KRAS^{G12D}. Further structural analysis through MD simulation showed the identified compounds particularly ZINC78453217 exhibited similar impact on KRAS^{G12D} through the stabilization of the global protein, reduction in flexibility and increase in compactness of the global protein. The successful application of multidisciplinary computational drug discovery approaches has allowed the identification of diverse potential therapeutic agents laying the ground work for experimental exploration of the suggested compounds. Further optimisations and validation of these compounds could aid in the discovery of effective KRAS^{G12D} inhibitors.

Data availability

All data generated or analyzed during this study are included in this published article [and its supplementary information files].

Received: 23 July 2022; Accepted: 18 October 2022

Published online: 22 October 2022

References

- Huang, L. *et al.* KRAS mutation: From undruggable to druggable in cancer. *Signal Transduct. Target Therapy* **6**(1), 1–20. <https://doi.org/10.1038/s41392-021-00780-4> (2021).
- Herdeis, L. *et al.* Stopping the beating heart of cancer: KRAS reviewed. *Curr. Opin. Struct. Biol.* **71**, 136–147. <https://doi.org/10.1016/j.sbi.2021.06.013> (2021).
- Gounder, M. M. *et al.* Sorafenib for advanced and refractory desmoid tumors. *N. Engl. J. Med.* **379**, 2417–2428. <https://doi.org/10.1056/NEJMoa1805052> (2018).
- Soria, J.-C. *et al.* Osimertinib in untreated EGFR-mutated advanced non-small-cell lung cancer. *N. Engl. J. Med.* **378**, 113–125. <https://doi.org/10.1056/NEJMoa1713137> (2018).
- Cox, A. D. *et al.* Drugging the undruggable RAS: Mission possible?. *Nat. Rev. Drug Discov.* **13**, 828–851. <https://doi.org/10.1038/nrd4389> (2014).
- Simanshu, D. K., Nissley, D. V. & McCormick, F. RAS proteins and their regulators in human disease. *Cell* **170**, 17–33. <https://doi.org/10.1016/j.cell.2017.06.009> (2017).
- Sprang, S. R. G proteins, effectors and GAPs: Structure and mechanism. *Curr. Opin. Struct. Biol.* **7**, 849–856. [https://doi.org/10.1016/S0959-440X\(97\)80157-1](https://doi.org/10.1016/S0959-440X(97)80157-1) (1997).
- Cherfils, J. & Zeghouf, M. Regulation of small GTPases by GEFs, GAPs, and GDIs. *Physiol. Rev.* **93**, 269–309. <https://doi.org/10.1152/physrev.00003.2012> (2013).
- Wittinghofer, A. & Pai, E. F. The structure of Ras protein: A model for a universal molecular switch. *Trends Biochem. Sci.* **16**, 382–387. [https://doi.org/10.1016/0968-0004\(91\)90156-p](https://doi.org/10.1016/0968-0004(91)90156-p) (1991).
- Bourne, H. R., Sanders, D. A. & McCormick, F. The GTPase superfamily: A conserved switch for diverse cell functions. *Nature* **348**, 125–132. <https://doi.org/10.1038/348125a0> (1990).
- Downward, J. Targeting RAS signalling pathways in cancer therapy. *Nat. Rev. Cancer* **3**, 11–22. <https://doi.org/10.1038/nrc969> (2003).
- Hunter, J. C. *et al.* Biochemical and structural analysis of common cancer-associated KRAS mutations. *Mol. Cancer Res.* **13**, 1325–1335. <https://doi.org/10.1158/1541-7786.MCR-15-0203> (2015).
- Ostrem, J. M. L. & Shokat, K. M. Direct small-molecule inhibitors of KRAS: From structural insights to mechanism-based design. *Nat. Rev. Drug Discov.* **15**, 771–785. <https://doi.org/10.1038/nrd.2016.139> (2016).
- Wood, K. W. *et al.* ras mediates nerve growth factor receptor modulation of three signal-transducing protein kinases: MAP kinase, Raf-1, and RSK. *Cell* **68**, 1041–1050. [https://doi.org/10.1016/0092-8674\(92\)90076-o](https://doi.org/10.1016/0092-8674(92)90076-o) (1992).
- Hofer, F. *et al.* Activated Ras interacts with the Ral guanine nucleotide dissociation stimulator. *Proc. Natl. Acad. Sci. U.S.A.* **91**, 11089–11093. <https://doi.org/10.1073/pnas.91.23.11089> (1994).
- Brunn, G. J. *et al.* Direct inhibition of the signaling functions of the mammalian target of rapamycin by the phosphoinositide 3-kinase inhibitors, wortmannin and LY294002. *EMBO J.* **15**, 5256–5267 (1996).
- Hamarshah, S. *et al.* Immune modulatory effects of oncogenic KRAS in cancer. *Nat. Commun.* **11**, 5439. <https://doi.org/10.1038/s41467-020-19288-6> (2020).

18. Prior, I. A., Hood, F. E. & Hartley, J. L. The frequency of Ras mutations in cancer. *Can. Res.* **80**, 2969–2974. <https://doi.org/10.1158/0008-5472.CAN-19-3682> (2020).
19. Dunnett-Kane, V. *et al.* Mechanisms of resistance to krasg12c inhibitors. *Cancers* **13**, 1–14. <https://doi.org/10.3390/cancers13010151> (2021).
20. Fell, J. *et al.* Discovery of Tetrahydropyridopyrimidines as Irreversible Covalent Inhibitors of KRAS-G12C with In Vivo Activity. *ACS Med. Chem. Lett.* <https://doi.org/10.1021/acsmchemlett.8b00382> (2018).
21. Nagasaka, M. *et al.* KRAS G12C game of Thrones, which direct KRAS inhibitor will claim the iron throne?. *Cancer Treat. Rev.* **84**, 101974. <https://doi.org/10.1016/j.ctrv.2020.101974> (2020).
22. Maurer, T. *et al.* 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**(14), 5299–5304. <https://doi.org/10.1073/pnas.1116510109> (2012).
23. Sun, Q. *et al.* Discovery of small molecules that bind to K-Ras and inhibit Sos-mediated activation. *Angew. Chem. Int. Ed. Engl.* **51**, 6140–6143. <https://doi.org/10.1002/anie.201201358> (2012).
24. Kessler, D. *et al.* Drugging all RAS isoforms with one pocket. *Future Med. Chem.* **12**, 1911–1923. <https://doi.org/10.4155/fmc-2020-0221> (2020).
25. Wang, X. *et al.* Identification of MRTX1133, a noncovalent, potent, and selective KRASG12D inhibitor. *J. Med. Chem.* **65**, 3123–3133. <https://doi.org/10.1021/acs.jmedchem.1c01688> (2022).
26. Berman, H. M. *et al.* The protein data bank. *Nucl. Acids Res.* **28**, 235–242. <https://doi.org/10.1093/nar/28.1.235> (2000).
27. Pettersen, E. F. *et al.* UCSF Chimera - A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612. <https://doi.org/10.1002/jcc.20084> (2004).
28. Case, D. A. *et al.* 'Amber 18' (2018).
29. Maier, J. A. *et al.* ff14SB: Improving the accuracy of protein side chain and backbone parameters from ff99SB. *J. Chem. Theory Comput.* **11**, 3696–3713. <https://doi.org/10.1021/acs.jctc.5b00255> (2015).
30. Larini, L., Mannella, R. & Leporini, D. Langevin stabilization of molecular-dynamics simulations of polymers by means of quasi-symplectic algorithms. *J. Chem. Phys.* **126**, 104101 (2007).
31. Gonnet, P. P-SHAKE: A quadratically convergent SHAKE in O (n²). *J. Comput. Phys.* **220**, 740–750 (2007).
32. Roe, D. R. & Cheatham, T. E. PTRAJ and CPPTRAJ: Software for processing and analysis of molecular dynamics trajectory data. *J. Chem. Theory Comput.* **9**, 3084–3095. <https://doi.org/10.1021/ct400341p> (2013).
33. Seifert, E. OriginPro 9.1: Scientific data analysis and graphing software—software review. *J. Chem. Inf. Model.* **54**, 1552–1552. <https://doi.org/10.1021/ci500161d> (2014).
34. Rudrapal, M. *et al.* In silico screening of phytopolyphenolics for the identification of bioactive compounds as novel protease inhibitors effective against SARS-CoV-2. *J. Biomol. Struct. Dyn.* <https://doi.org/10.1080/07391102.2021.1944909> (2021).
35. Miller, B. R. *et al.* MMPBSA.py: An Efficient program for end-state free energy calculations. *J. Chem. Theory Comput.* **8**, 3314–3321. <https://doi.org/10.1021/ct300418h> (2012).
36. Koes, D. R. & Camacho, C. J. ZINCPharmer: Pharmacophore search of the ZINC database. *Nucleic Acids Res.* **40**, 409–414. <https://doi.org/10.1093/nar/gks378> (2012).
37. Dallakyan, S. & Olson, A. J. Small-molecule library screening by docking with PyRx. *Methods Mol. Biol.* **1263**, 243–250. https://doi.org/10.1007/978-1-4939-2269-7_19 (2015).
38. Repasky, M. P., Shelley, M. & Friesner, R. A. Flexible ligand docking with Glide. *Curr. Protoc. Bioinf.* <https://doi.org/10.1002/0471250953.bi0812s18> (2007).
39. Roos, K. *et al.* OPLS3e: Extending force field coverage for drug-like small molecules. *J. Chem. Theory Comput.* **15**, 1863–1874. <https://doi.org/10.1021/acs.jctc.8b01026> (2019).
40. Daina, A., Michielin, O. & Zoete, V. SwissADME: A free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Sci. Rep.* **7**, 1–13. <https://doi.org/10.1038/srep42717> (2017).
41. Banerjee, P. *et al.* ProTox-II: A webserver for the prediction of toxicity of chemicals. *Nucleic Acids Res.* **46**, W257–W263. <https://doi.org/10.1093/nar/gky318> (2018).
42. Drwal, M. N. *et al.* ProTox: A web server for the in silico prediction of rodent oral toxicity. *Nucleic Acids Res.* **42**, 53–58. <https://doi.org/10.1093/nar/gku401> (2014).
43. Banerjee, P., Dehnbostel, F. O. & Preissner, R. Prediction is a balancing act: Importance of sampling methods to balance sensitivity and specificity of predictive models based on imbalanced chemical data sets. *Front. Chem.* <https://doi.org/10.3389/fchem.2018.0036> (2018).
44. Pantzar, T. The current understanding of KRAS protein structure and dynamics. *Comput. Struct. Biotechnol. J.* **18**, 189–198. <https://doi.org/10.1016/j.csbj.2019.12.004> (2020).
45. Li, Y., Han, L. & Zhang, Z. Understanding the influence of AMG 510 on the structure of KRASG12C empowered by molecular dynamics simulation. *Comput. Struct. Biotechnol. J.* **20**, 1056–1067. <https://doi.org/10.1016/j.csbj.2022.02.018> (2022).
46. Gasper, R. & Wittinghofer, F. The Ras switch in structural and historical perspective. *Biol. Chem.* **401**, 143–163. <https://doi.org/10.1515/hsz-2019-0330> (2019).
47. Karyagina, A. *et al.* Analysis of conserved hydrophobic cores in proteins and supramolecular complexes. *J. Bioinform. Comput. Biol.* **4**, 357–372. <https://doi.org/10.1142/S0219720006001837> (2006).
48. Likhachev, I. V., Balabaev, N. K. & Galzitskaya, O. V. Available instruments for analyzing molecular dynamics trajectories. *Open Biochem. J.* **10**, 1–11. <https://doi.org/10.2174/1874091X01610010001> (2016).
49. Raha, K. & Merz, K. M. Calculating binding free energy in protein-ligand interaction. *Ann. Rep. Comput. Chem.* **1**, 113–130. [https://doi.org/10.1016/S1574-1400\(05\)01009-1](https://doi.org/10.1016/S1574-1400(05)01009-1) (2005).
50. Wermuth, C. G. *et al.* Glossary of terms used in medicinal chemistry (IUPAC Recommendations 1998). *Pure Appl. Chem.* **70**, 1129–1143 (1998).
51. Durrant, J. D. & McCammon, J. A. Molecular dynamics simulations and drug discovery. *BMC Biol.* **9**, 71. <https://doi.org/10.1186/1741-7007-9-71> (2011).
52. Renukuntla, J. *et al.* Approaches for enhancing oral bioavailability of peptides and proteins. *Int. J. Pharm.* **447**, 75–93. <https://doi.org/10.1016/j.ijpharm.2013.02.030> (2013).
53. Lipinski, C. A. *et al.* Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Deliv. Rev.* **46**, 3–26. [https://doi.org/10.1016/s0169-409x\(00\)00129-0](https://doi.org/10.1016/s0169-409x(00)00129-0) (2001).
54. Omran, Z. & Rauch, C. Acid-mediated Lipinski's second rule: Application to drug design and targeting in cancer. *Eur. Biophys. J.* **43**, 199–206. <https://doi.org/10.1007/s00249-014-0953-1> (2014).
55. Lu, J. *et al.* Influence of molecular flexibility and polar surface area metrics on oral bioavailability in the rat. *J. Med. Chem.* **47**, 6104–6107. <https://doi.org/10.1021/jm0306529> (2004).
56. Veber, D. *et al.* Molecular properties that influence the oral bioavailability of drug candidates. *J. Med. Chem.* **45**, 2615–2623. <https://doi.org/10.1021/jm020017n> (2002).
57. Di, L., Kerns, E. H. & Carter, G. T. Drug-like property concepts in pharmaceutical design. *Curr. Pharm. Des.* **15**, 2184–2194. <https://doi.org/10.2174/138161209788682479> (2009).
58. Vallianatou, T., Giaginis, C. & Tsantili-Kakoulidou, A. The impact of physicochemical and molecular properties in drug design: Navigation in the “drug-like” chemical space. *Adv. Exp. Med. Biol.* **822**, 187–194. https://doi.org/10.1007/978-3-319-08927-0_21 (2015).

59. Lipinski, C. A. Drug-like properties and the causes of poor solubility and poor permeability. *J. Pharmacol. Toxicol. Methods* **44**, 235–249. [https://doi.org/10.1016/s1056-8719\(00\)00107-6](https://doi.org/10.1016/s1056-8719(00)00107-6) (2000).
60. Wang, J. & Hou, T. Recent advances on aqueous solubility prediction. *Comb. Chem. High Throughput Screen.* **14**, 328–338. <https://doi.org/10.2174/138620711795508331> (2011).
61. Ertl, P., Rohde, B. & Selzer, P. Fast calculation of molecular polar surface area as a sum of fragment-based contributions and its application to the prediction of drug transport properties. *J. Med. Chem.* **43**, 3714–3717. <https://doi.org/10.1021/jm000942e> (2000).
62. Prasanna, S. & Doerksen, R. J. Topological polar surface area: A useful descriptor in 2D-QSAR. *Curr. Med. Chem.* **16**, 21–41. <https://doi.org/10.2174/092986709787002817> (2009).
63. Shityakov, S. *et al.* Analysing molecular polar surface descriptors to predict blood-brain barrier permeation. *Int. J. Comput. Biol. Drug Des.* **6**, 146–156. <https://doi.org/10.1504/IJCBD.2013.052195> (2013).
64. Daga, P. R. *et al.* Physiologically based pharmacokinetic modeling in lead optimization. 2. Rational bioavailability design by global sensitivity analysis to identify properties affecting bioavailability. *Mol. Pharm.* **15**, 831–839. <https://doi.org/10.1021/acs.molpharmaceut.7b00973> (2018).
65. Penniston, J. T. *et al.* Passive permeation of organic compounds through biological tissue: A non-steady-state theory. *Mol. Pharmacol.* **5**, 333–341 (1969).
66. Agoni, C. *et al.* Extended double bond conjugation in the chalcone framework favours MAO-B inhibition: A structural perspective on molecular dynamics. *Comb. Chem. High Throughput Screen.* **25**, 2059–2069. <https://doi.org/10.2174/1386207325666220214110717> (2022).
67. Schweizer, L. & Mueller, L. protein conformational dynamics and signaling in evolution and pathophysiology. In *Arey Pharmacology and Therapeutics* 209–249 (Academic Press, San Diego, 2014).
68. Wang, C. *et al.* Recent developments and applications of the MMPBSA method. *Front. Mol. Biosci.* <https://doi.org/10.3389/fmolb.2017.00087> (2018).

Acknowledgements

Authors wish to acknowledge the Centre for High Performance Computing, Cape Town, South Africa for providing computational resources.

Author contributions

Conceptualization: A.R.I., N.M., C.A., and M.R.; methodology: A.R.I., N.M., C.A., and M.R.; data curation: A.R.I., N.M., and C.A.; formal analysis: A.R.I., N.M., and M.R.; data analysis: A.R.I. and C.A.; investigation: S.M.A., S.G.A., and J.K.; project administration: A.R.I. and M.R.; resources: A.R.I., N.M., C.A., and M.R.; original draft: A.R.I. and N.M.; writing—review and editing: A.R.I., N.M., M.R., and J.K.; funding acquisition: S.M.A., S.G.A., and J.K. All authors read and approved the final version of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-22668-1>.

Correspondence and requests for materials should be addressed to M.R.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022