Ultraefficient Unbiased Molecular Dynamics simulation of protein-ligand interactions: How profound yet affordable can it be?

Farzin Sohraby, Hassan Aryapour*, Mostafa Javaheri Moghadam, Masoud Aliyar

Department of Biology, Faculty of Science, Golestan University, Gorgan, Iran

*Corresponding author. Department of Biology, Faculty of Science, Golestan University, Gorgan, Iran. Tel No: +98-17-32254161; Fax No. +98-17-32245964; E-mail address: h.aryapour@gu.ac.ir

Abstract

Small molecules such as substrates, effectors and drugs play essential roles in biological processes. Knowing their interactions with biomacromolecular targets demands a deep understanding of binding and unbinding mechanisms. Dozens of papers have suggested that discovering of either binding or unbinding events by means of conventional UMD simulation urges a considerable amount of computational resources, therefore, only one who holds a supercomputer can afford such extensive simulations. Capabilities of full atomistic Unbiased Molecular Dynamics (UMD) simulation have been undervalued within the scientific community. Thus, myriads of researchers are impelled to be content with debatable biased MD simulations which seek validation for its preconceived framework. In this work, we present a stratagem to empower everyone to perform UMD simulations of protein-ligand binding and unbinding by typical desktop computers and achieve valuable and high-cost results in nanosecond time scale. Here, we have described kinetics of binding and unbinding of anticancer drug, dasatinib, to c-Src kinase in full atomistic details for the first time. We have attained multiple independent binding and unbinding events occurred in the nano-second timescale, even in times as little as 30 and 392.6 ns respectively, without presence of any biasing forces, an achievement that nobody has ever assumed to be possible.

Keywords: Ultraefficient Unbiased Molecular Dynamics (UUMD), Rational drug design, Ligand binding and unbinding pathway, Dasatinib, c-Src kinase

Introduction

Small molecule compounds are involved in nearly all cellular mechanisms and studying their roles can unravel secrets behind the scenes. These compounds can trigger cell signaling and metabolic pathways by interacting and binding to certain biomacromolecules like proteins and nucleic acids. Over the last few decades, sophisticated methods such as X-ray crystallography, NMR and electron microscopy revealed numerous structural details of many protein-ligand complexes. However, these complexes are just one or some static poses of a vivid system which its function is completely swayed by its movements and dynamics. Furthermore, in many molecular targets like androgen receptor (PDB ID: 2Q7I)[1], the binding pocket is buried deep inside the protein structure. The X-ray crystallographic structure doesn't reveal any details about the process of induction and the binding pathway, how the ligand makes an entrance into the protein and how it affects residues on its journey to reach the native binding pose. According to the Food and Drug Administration (FDA), small molecules make up the main proportion of approved drugs on the pharmaceutical market today. Thus, understanding the induction and binding mechanisms of small molecules to their molecular targets can immensely assist researchers to optimize and design much more specific and selective drugs accompanied by extremely low side-effects. Therefore, emergence of a complementary method which can take the advantages of conventional laboratory methods to determine the three-dimensional structure of biomacromolecular complexes is indispensable.

In the last several years, there have been many studies based on unbiased MD simulation which led to profound results and significantly altered our understanding of binding/unbinding mechanisms in molecular biology[2-11]. Unfortunately since this approach urges powerful supercomputers, it is not disseminated universally, and in fact it has convinced researchers to use biased approachs. Recent breakthroughs in the biased methods have permitted us to determine the binding/unbinding kinetics of a protein-ligand complex[12-19]. Although these advancements enhanced sampling by utilizing either modification of both potential energy surface and temperature or reassignment of probability weight factor, they result in uncertainty in total potential energy and substandard depiction of canonical ensemble, respectively. Besides, employment of biased methods requires mastery in precise selection of biasing variables and potentials, a crucial step that any mistakes in it will lead to false positive results. However, intrinsic drawbacks of biased methods encourage researchers to come up with initiatives.

Nevertheless, we believe this need can be addressed by Ultraefficient Unbiased Molecular Dynamics (UUMD, which is pronounced as WMD) simulation. UUMD is a unique approach that not only detracts the need of supercomputers but also is much more accurate than biased methods particularly in simulations of binding/unbinding event, a delicate process that even a slight difference in molecular charge can alter everything dramatically. In this experiment, we have introduced an optimized method that can encourage everyone to perform UMD simulation to model the binding pathway of protein-ligand complex and much more besides, in the nanosecond time scale. Moreover, thanks to UMD competency to reveal fine dynamic properties of the protein-ligand complex, further details are also achievable. Herein, we have presented new information about the sampling, induction, binding and also unbinding of dasatinib in complex with c-Src protein kinase by utilizing UUMD. Our team have decided on dasatinib/c-Src kinase complex, because UMD simulation of this complex has been conducted formerly by Shan et al.[2], so we could easily set their work as a benchmark for the present study. In spite of Shan et al work, we aimed to optimize UMD simulation so that it can be employed by any researchers in the field of structural biology and drug discovery. All of the computations of this work were performed on an Ubuntu desktop PC with a [dual-socket Intel(R) Xeon(R) CPU E5-2630 v3 + 2 NVIDIA GeForce GTX 1080] configuration.

The optimization of UMD simulation

There are underlying differences between our UUMD simulations and previous UMD simulations. Usually, small molecules tend to aggregate together to form a ball-like structure; even two ligands can aggregate inside the simulation box for an extensive amount of time. This detracts sampling effort significantly on one side and on the other side increases simulation time substantially. Besides, insertion of just one single ligand in the simulation box can reduce the probability of sampling and consumes an unaffordable computational time. This dilemma can probably be resolved by employment of repulsive forces among the ligands so that aggregation does not take place at all. Shan et al. applied a weak repulsive force among “the nitrogen atoms of the central amide group of dasatinib molecules" with a cut off of 2 nm after insertion of 6 molecules inside the simulation box, succeeding only one binding event after a total run time of 32 μs. Since this work, there have been many studies that employed UMD simulation to reconstruct the binding event[4,5,7-9] but none of them have employed repulsive forces among the

ligands yet. So, most of these studies have wasted their simulation time in the wake of ligands' aggregation, especially, when the ligands' concentration was high. Furthermore, the aggregation can limit the motions of ligands in the solvent and alter their behavior. In our UUMD simulations, we applied repulsive forces between Virtual Interaction Sites (VIS) which were placed evenly all over the molecule, almost between each two heavy atoms of the ligand (Fig1. e). The repulsive force was applied through the reassignment of the Lennard-Jones' parameters (Fig1. d). The values of the σ and ε parameters for VISs were set to 0.83 nm and 0.1 $kJ.mol^{-1}$, respectively. While our applied repulsive force prevents ligands' aggregation, its appliance preserves the native fluctuations and behavior of the ligands, even at considerably high concentrations. The results show that all of the 16 inserted ligands in the simulation box have demonstrated the native behavior of a single solvated ligand even when the concentration is 60 mM (Fig1. a, b, c). This optimization allows us to insert high concentrations of any ligands and simultaneously enhance sampling and probability of achieving the binding event, without any unwanted changes in protein structure (Ext Fig3).

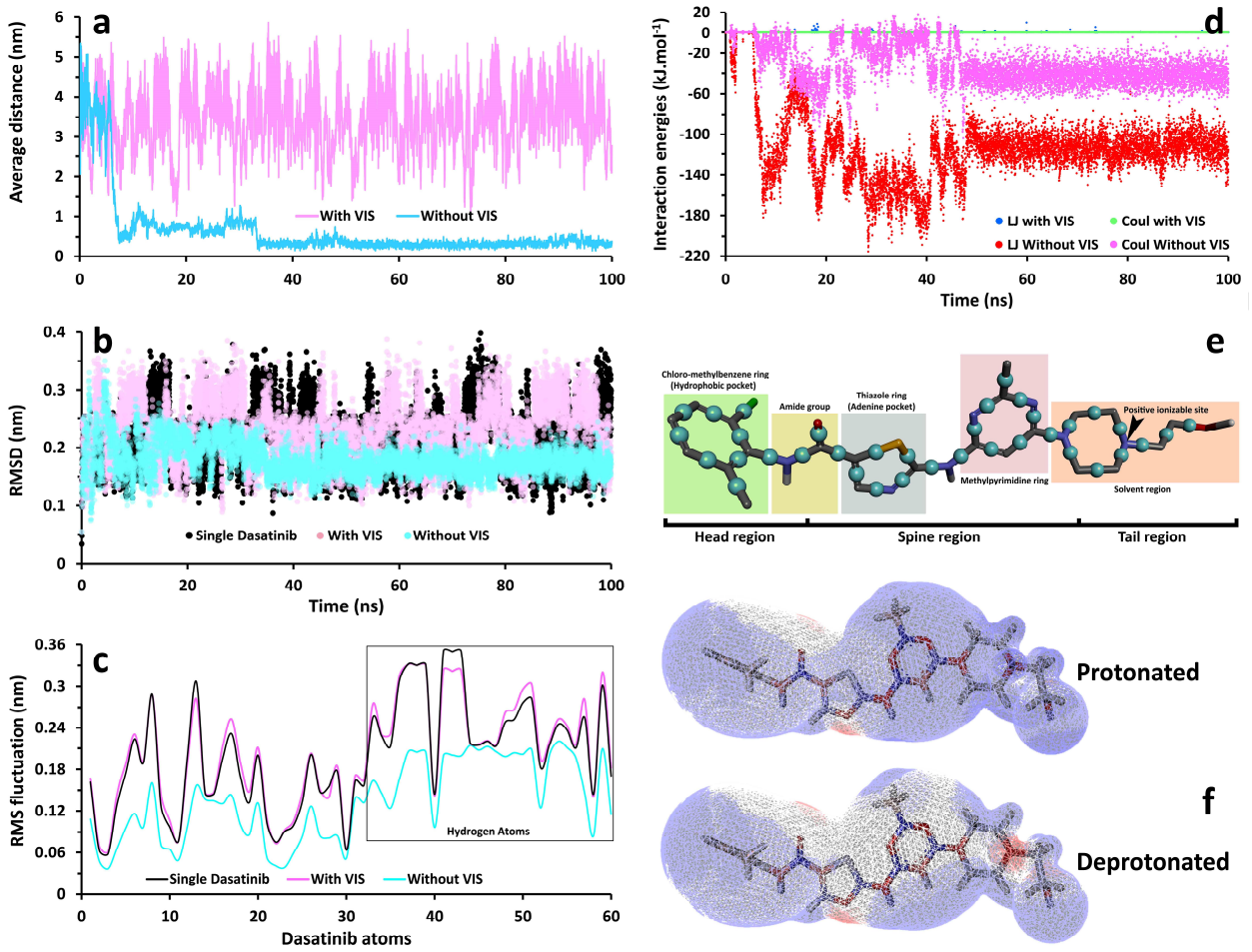


Figure 1. The influence of virtual interaction sites (VIS) on the molecular behavior of dasatinib in the simulation box which contains 16 molecules. a, The average distance between two randomly selected dasatinib molecules (herein numbers 1 and 3) with and without presence of VIS. **b,** The RMSD of a randomly selected dasatinib (number 1) compared to its own starting conformation during the simulation time with and without presence of VIS. In the absence of VIS, the RMSD of dasatinib 1 fell down after 35 ns and then fluctuated between 0.1 and 0.25 nm. In contrast, in the presence of VIS the RMSD fluctuated in the range of 0.1 to 0.4 nm which was similar to the control plot (a single dasatinib in the simulation box). **c,** Root mean square fluctuation (RMSF) of the dasatinib 1. The fluctuation of dasatinibs' atoms in the presence of VIS was very similar to the control plot (single dasatinib). On the contrary, in the absence of VIS the fluctuation of dasatinibs decreased significantly. **d,** Energies of Van der Waals (VdW) and Coulombic (Coul) interactions between dasatinib 1 and the other dasatinibs. In the absence of VIS, dasatinibs aggregated and their behavior were altered under the influence of strong attractive non-covalent forces (especially Van der Waals). The employment of the repulsive forces evened up the attractive forces. **e,** The fragmented form of dasatinib structure, the VISs are represented by dark cyan balls placed between heavy atoms of dasatinib to distribute repulsive energy evenly all over the molecule. **f,** The comparison of partial charge surfaces of two forms of dasatinib molecule. Atoms are colored based on their partial charges; red, blue and grey colors indicate negative, positive and neutral charges, resp.

In order to enhance the sampling of ligands from the surface of the protein even more, a series of short time (nanosecond time scale) simulations were utilized, instead of running long time (microsecond time scale) simulations (Fig2). In first attempt, for extracting the pathway of binding of the protonated dasatinib to the c-Src kinase, fifty independent simulations which had identical atomic coordinate but different velocities, called replicas, with time length of 10 ns were conducted while in each simulation 16 dasatinib molecules were presented inside the box. After analyzing all the trajectories, 4 runs (#5, #23, #40 and #50) with well-orientated dasatinibs were selected based on their orientations and conformations' closeness to the co-crystallized dasatinib for further rounds of simulations. In the next step, after removing all of the dasatinibs from the box except the best oriented one (according to the co-crystal structure), 5 replicas were generated from each of the 4 previously selected runs and then simulations were continued up to 100 ns. Out of these 20 replicas two runs, #23.3 and #40.5, demonstrated successful induction by reaching RMSD values below 2 Å, according to crystal structure. Then each of these runs, #23.3 and #40.5, was extended up to 500 ns. Moreover, out of the last round of simulations for run #5, one run, #5.5, showed a new binding pose which was extended for 1 μs (Fig5). For extracting the binding pathway of the deprotonated dasatinib, the similations were done based on the flowchart. This stratagem was also examined by Knapp et al. where they confirmed that short replicas are far more efficient than long-time runs in terms of protein folding[20].

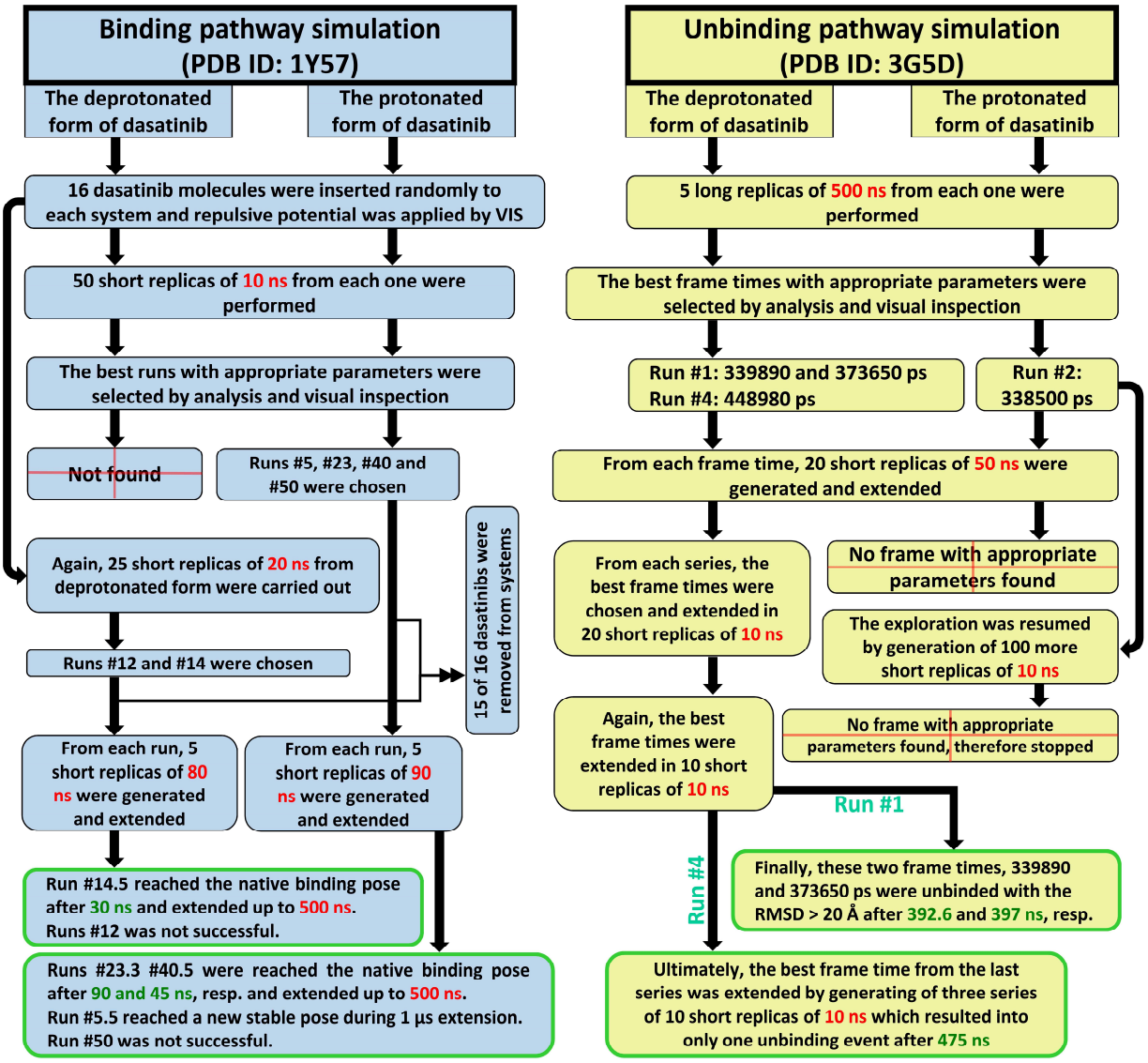


Figure 2. Ultraefficient Unbiased Molecular Dynamics flowchart. All replicas were identical except in their velocities.

Sampling of the binding pocket was more challenging when the deprotonated dasatinib was employed. In the first round of simulations (50 initial replicas), no successful run has been found. Thus, the second round of simulations was reorganized with 25 replicas at durations of 20 ns. As the result, two runs (#12 and #14) were detected as they were presenting well-orientated dasatinibs, which seemed to be good candidates for further simulations. So, these two runs were subjected for the next round which consisted of 5 replicas at durations of 80 ns. This finally

resulted to one successful run (#14.5) where dasatinib reached the native binding pose (crystal structure). By extending this run up to 500 ns, this achievement was approved.

Surprisingly, this method can significantly enhance sampling and the chance of achieving the native binding pose in the shortest time that ever reported. By applying present method described in this work, it took just 30 ns for the ligand (dasatinib) to reach its native binding pose while it took 2.3 μs for Shan et al.

Joyously this approach does not stop here, when it comes to the unbinding process, this method can be even more pleasing. By increasing the number of short replicas and then thoroughly inspecting and dissecting snapshots of the trajectories to decide about their competency for the next round of simulation, we achieved multiple unbinding events eventually, a promising stratagem which we are delightful to name it "rational sampling". Consider that the deprotonated dasatinib has fully unbounded from its receptor, c-Src kinase, only in 392.6 ns without utilizing any biasing forces.

The binding pathways of dasatinib

Full details of the dasatinib binding process have not been reported before, and herein, by analyzing all of the simulated trajectories, for the first time, we unravel some of the fine characteristics of the protein structure and also describe how and in what different modes dasatinib can bind to the c-Src kinase protein. The c-Src kinase domain has two main parts (Fig3.a); a big carboxyl-terminus lobe and a small amino-terminus lobe. The binding pocket exists in the middle of these two lobes which forms a cleft. These two main parts are joined by the αC-helix and a very important random-coil structure which is called "the hinge"[21]. This segment of the binding pocket plays a very critical role. Although the hinge sticks to the dasatinib and make it bind tightly to the binding pocket, it's able to let water molecules get inside the binding pocket so the unbinding can occur as the result of breakage of a few hydrogen bonds which will be described later.

Before dasatinib induces the binding pocket, the deep parts of the binding pocket are inaccessible as there are two evolutionary conserved salt bridges among three residues K295, E310 and D404, the indicators of the active form of the protein[22] (Fig3.b). Also, hydrophobic

residues on each side of the cleft interact with each other and can occasionally close the binding pocket. In the two simulation runs, #23.3 and #40.5, dasatinib induced the binding pocket when its head had been orientated towards the deep parts of the binding pocket (Fig3.c). As it was going down inside the pocket, hydrophobic interactions got stronger and the head of dasatinib interrupted the salt bridges, especially with its hefty chlorine substitute. In other words, the head of dasatinib could weaken the salt bridges and get to the deep parts of the binding pocket with ease in a short amount of time (Fig3.d). After that, the spine of dasatinib landed on the hinge and two hydrogen bonds formed between the backbone of M341 and two nitrogen atoms in the spine of dasatinib (Fig3.e.f). Based on the interaction energies of dasatinib with each residue in the binding pocket, the M341 has the most contribution in the stability of dasatinib (Ext Fig1). These two aforementioned hydrogen bonds encourage the hydrophobic interactions between the spine of dasatinib and the hydrophobic residues like the L273, V281, L393 and Y340 which somehow hug dasatinib and keep it away from the water molecules (Fig3.g). Meanwhile, the tail of dasatinib is almost solvated for most of the time.

In spite of the binding mode described above, a second binding pathway was also observed in the run #14.5. Although binding elements had not changed, the sequence of the incidents was different. In this run, the spine of dasatinib was the first region to land on the hinge and only one hydrogen bond formed between the oxygen atom of the M341 backbone and dasatinib, then, the hydrophobic interactions stabled the dasatinib afterwards (Fig3.h). Next, the head of dasatinib found its way into the deep parts of the binding pocket, underneath the conserved salt bridges, and simultaneously the second hydrogen bond formed between the nitrogen atoms of the M341 backbone and dasatinib so that it reached the native binding pose (Fig3.i). The estimated value of K_{on} for both binding forms of dasatinib was 7.56 $s^{-1}.\mu M^{-1}$, compare to the experimentally determined value (5 $s^{-1}.\mu M^{-1}$)[23].

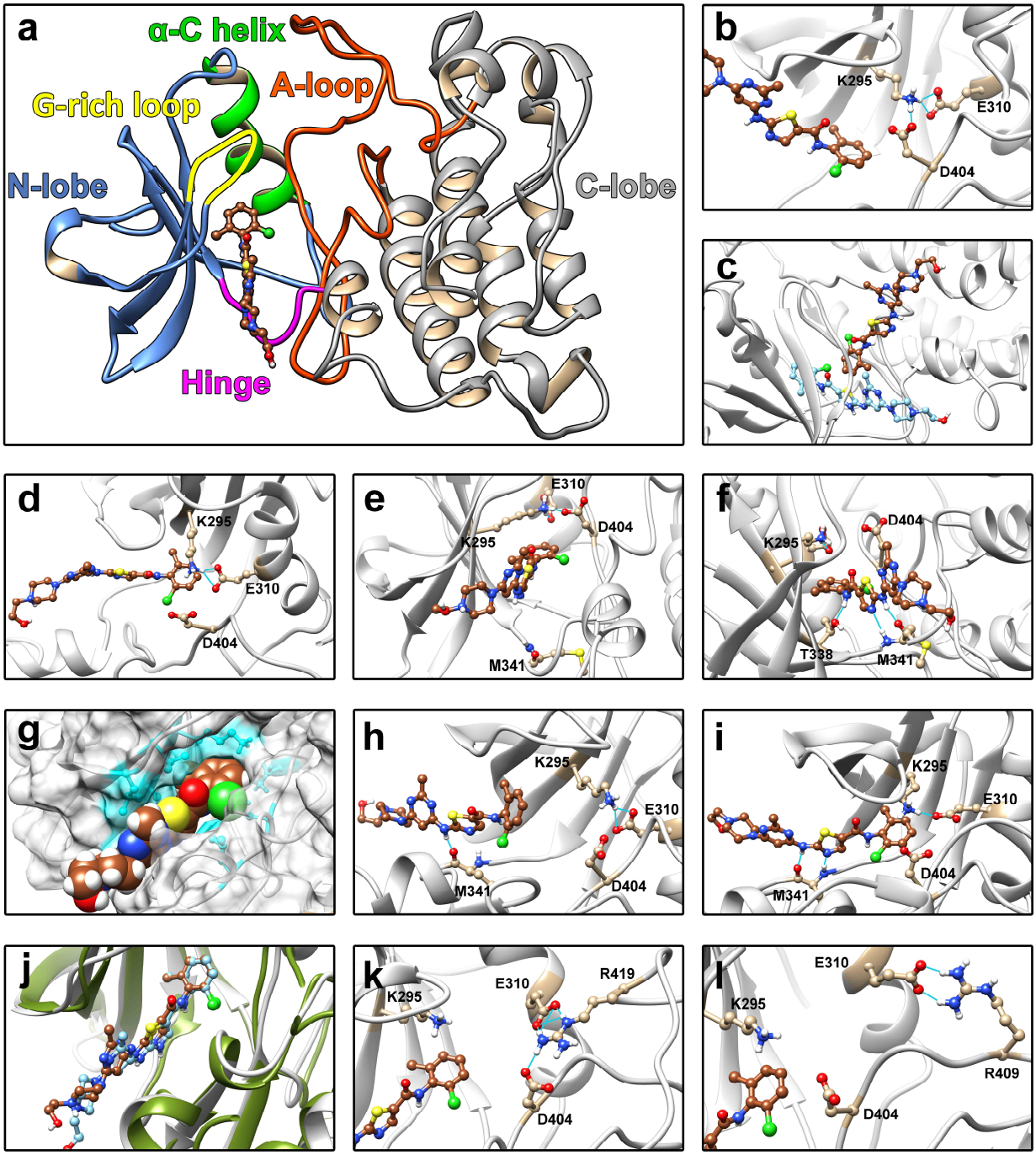


Figure 3. The atomic details of the dasatinib binding pathways. a, The entire structure of c-SRC kinase domain with its regions. **b,** Before the induction, K295, E310 and D404 stuck together and formed the salt bridges. **c,** The correct orientation of dasatinib in which the head pointed down towards the deep parts of the binding pocket, the X-ray structure of dasatinib is shown in cyan. Further selections (runs #5, #12, #14, #23, #40 and #50) were made based on this orientation. **d,** The interruption of the salt bridge by the head of dasatinib. It easily broke the salt bridge between K295 and D404 but the bond between K295 and E310 was stronger and maintained. **e,** Before the landing of the spine of dasatinib on the hinge. **f,** The landing of the spine and the formation of two hydrogen bonds between the backbone of M341 and the two Nitrogen atoms of dasatinib and also the formation of a hydrogen

bond between T338 and the head of dasatinib. **g,** The hydrophobic interactions between dasatinib (the spine and head) and the hydrophobic residues including L273, V281 and L393 which somehow hugged dasatinib, colored in cyan. **h,** The second binding mode, where only one hydrogen bond between the oxygen atom of the backbone of M341 and dasatinib formed at first. **i,** Then, the head of dasatinib found its way inside the binding pocket and got underneath the conserved salt bridges and a hydrogen bond formed between T338 and the amide group of dasatinib. **j,** The native binding pose achieved by UUMD simulation. **K,** The interference of the A-loop. The R419 broke the conserved salt bridges and confiscated both of the E310 and D404. **l,** The formation of a different salt bridge between R409 and E310 when the glutamate residue headed outwards. This could result in the αC-helix-out conformation.

The head of dasatinib requires a considerable amount of space and when this fragment is in the deep parts of the binding pocket, it weakens the evolutionary conserved salt bridges among three residues K295, E310 and D404. In fact, the presence of dasatinib inside the pocket is associated with the frequency of salt bridges' breakage. Surprisingly, these salt bridges in some way are the stabilizers of dasatinib's head. Because, the side chains of K295 and E310 can make Lennard-Jones (LJ) and coulombic interactions with the head of dasatinib while their absence makes dasatinib much more unstable (Fig4.a). Breakage of these salt bridges gives way to the water molecules to flow inside the binding pocket and get conserved in the deep parts of this pocket for a considerable amount of time and also causes higher fluctuations of both K295 and E310 which to some extent makes the head of dasatinib unstable.

The breakage of the salt bridges which trigger a cascade of significant conformational changes in the protein structure[24,25], can take place under the influence of three different factors; (i) kinetic energy of bonded residues and water molecules. A telling example is periodic motions of E310 to head outwards from the pocket for interacting with water molecules. (ii) Mechanical energy derived from the head fragment of dasatinib. (iii) Activation loop (A-loop). There are two key arginine residues in the A-loop which can interfere with these salt bridges, R419 and R409. R419 can bind to the both of D404 and E310 at the same time (Fig3.k). In the run #40.5, we found that when R419 confiscated these two acidic residues, dasatinib could get to the deep parts of the binding pocket much easier (Fig4.c). R409 could also bind to E310 when it was heading outwards, an incident that could result in formation of the αC-helix-out conformation[24,26,27] (Fig4.a-Fig3.j). It was also observed that these two arginine residues could even swap E310 with each other. Generally, the A-loop can bind to the revolutionary conserved E310 and D404 and transform the protein structure from active to inactive conformation by utilizing of R419 and

R409. It was previously reported that only R409 could trigger this conformational change[27] but in this work, we found that R419 can also be the cause and these conformational changes can be completely in the control of the A-loop which itself undergoes major conformational changes throughout this transformation.

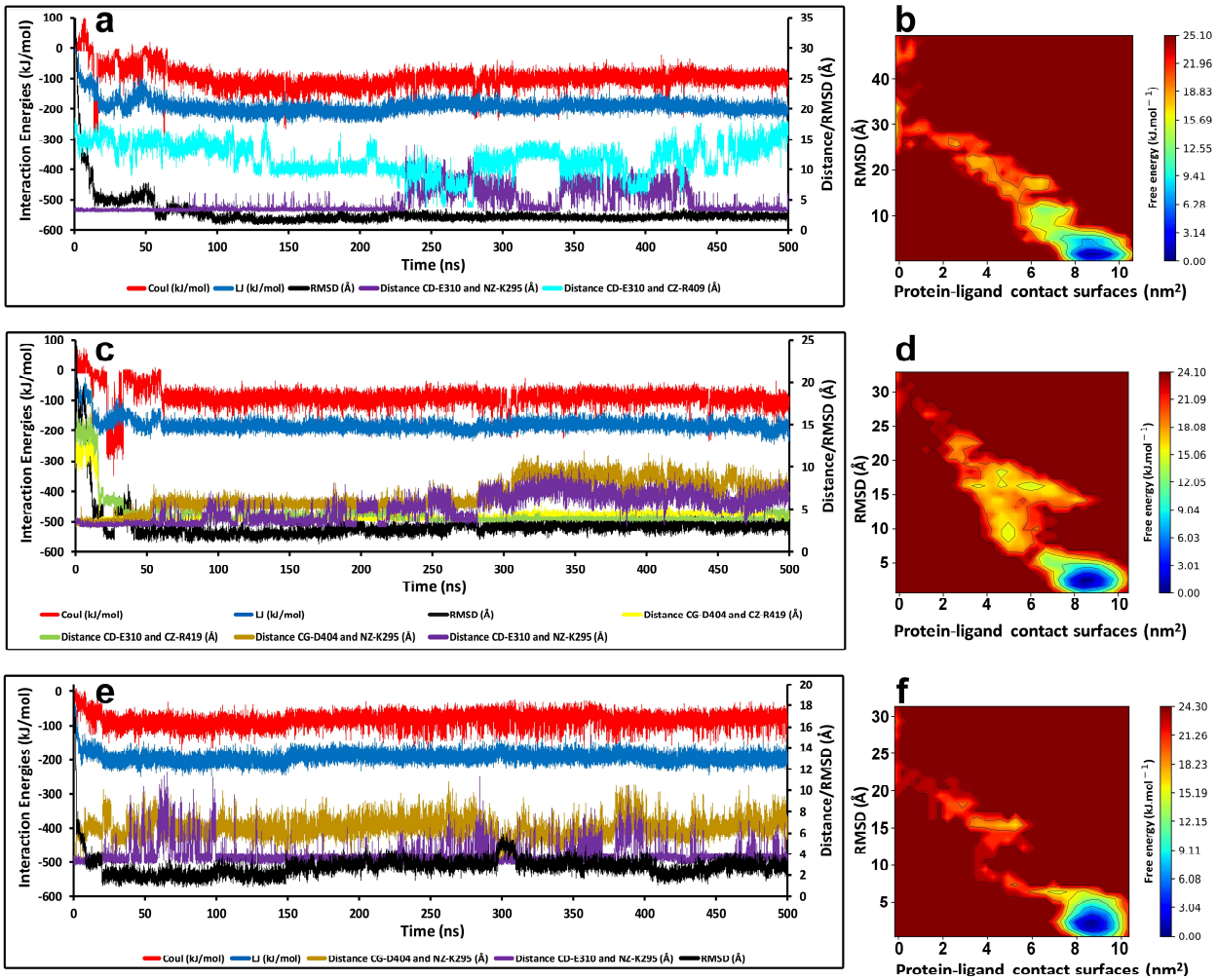


Figure 4. The Simulation of the binding pathway of protonated and deprotonated forms of dasatinib to Src-kinase. a, Run #23.3. Atomic distance between CD of E310 and NZ of K295 which represents the breakage and the reformation of this salt bridge throughout the simulation time. K295 and D404 did not give any salt bridge together during this run. The RMSD values, alongside with the interaction energies of the protein-ligand complex, converged to at least 0.9 Å just after 90 ns. Over a period, (around 230 to 450 ns), the aforementioned salt bridge used to be broken so that E310 was able to be engaged with R409 which resulted to the instability of dasatinib's head and a slight rise both in the interaction energies and the RMSD values. **b,** The Free Energy Landscape of the run #23.3 which was calculated by using "gmx sham". This landscape shows that in the binding process, there were no intermediate states and the binding event happened in a short amount of time. **c,** Run #40.5. Dasatinib reached the native binding pose in just 45 ns. The atomic distance of the salt bridges between K295 and each of E310/D404, and the distance of the salt bridges between R419 and each of E310/D404 showed that R419

confiscated both of D404 and E310, from 20 ns till the end of simulation, by breaking the anterior salt bridges, an effort that opens a way for dasatinib to bind. **d,** The free energy landscape of the run #40.5. **e,** Run #14.5. Dasatinib reached the native binding pose in just 30 ns. **f,** The free energy landscape of the run #14.5.

The T338 pulls the head of dasatinib toward itself by establishing an influential hydrogen bond with the amide group near the head of dasatinib so that water molecules cannot seep into this area. However, this hydrogen bond can be easily broken by either water mediation or the rotation of the T338 side-chain's dihedral angle. Although the methyl group of T338 can form Alkyl-Pi interaction with the head of dasatinib as well, its impact on captivation of the dasatinib's head is not as substantial as the hydrogen bond.

In this work, UUMD simulation enabled us to reveal different probable mechanisms for the binding of dasatinib to c-Src kinase. Moreover, this incredible method is even capable of predicting binding conformations and binding pockets that can be out of the reach of X-ray crystallography. In the run #5.5, we encountered an unusual binding conformation which had not been reported by anyone before. We found that the protonated form of dasatinib could bind in a 180-degrees flipped conformation compared to the native binding pose reported by X-ray crystallography (Fig5.e).

At the first step, R419 triggered the process by confiscating the E310 and D404, a phenomenon that is accompanied with breakage of salt bridges among the K295, D404 and E310 and thereafter expansion of the volume of the pocket (Fig5.b). This breakage made a way for F278 to turn inside the binding pocket and stabilize dasatinib by π-stacking interactions which as the result steered dasatinib to go deeper into the pocket (Fig5.c,d). Hydrophobic residues of the pocket also play a supportive role in the stability of this conformation (Fig5.e). The accumulation of these forces lowered the RMSD values to just above 3 Å.

Although the Van der Waals interaction energy scheme of this new binding pose (#5.5) is closely similar to VdW energy scheme of the native binding poses (#23.3, #40.5 and #14.5), its electrostatic interaction energy scheme is modestly different from the native ones. The electrostatic energy plot for the new binding pose is much more erratic due to the periodic formation and breakage of a salt bridge between the tail of dasatinib and D348. It is also evident that the overall electrostatic energy trend is more positive in comparison with the native ones

(Fig5.a). However, this conformation may not be as stable as the native conformation since during a micro-second simulation, it was observed that dasatinib could literally interact with the outer residues of the binding pocket and been pulled out occasionally.

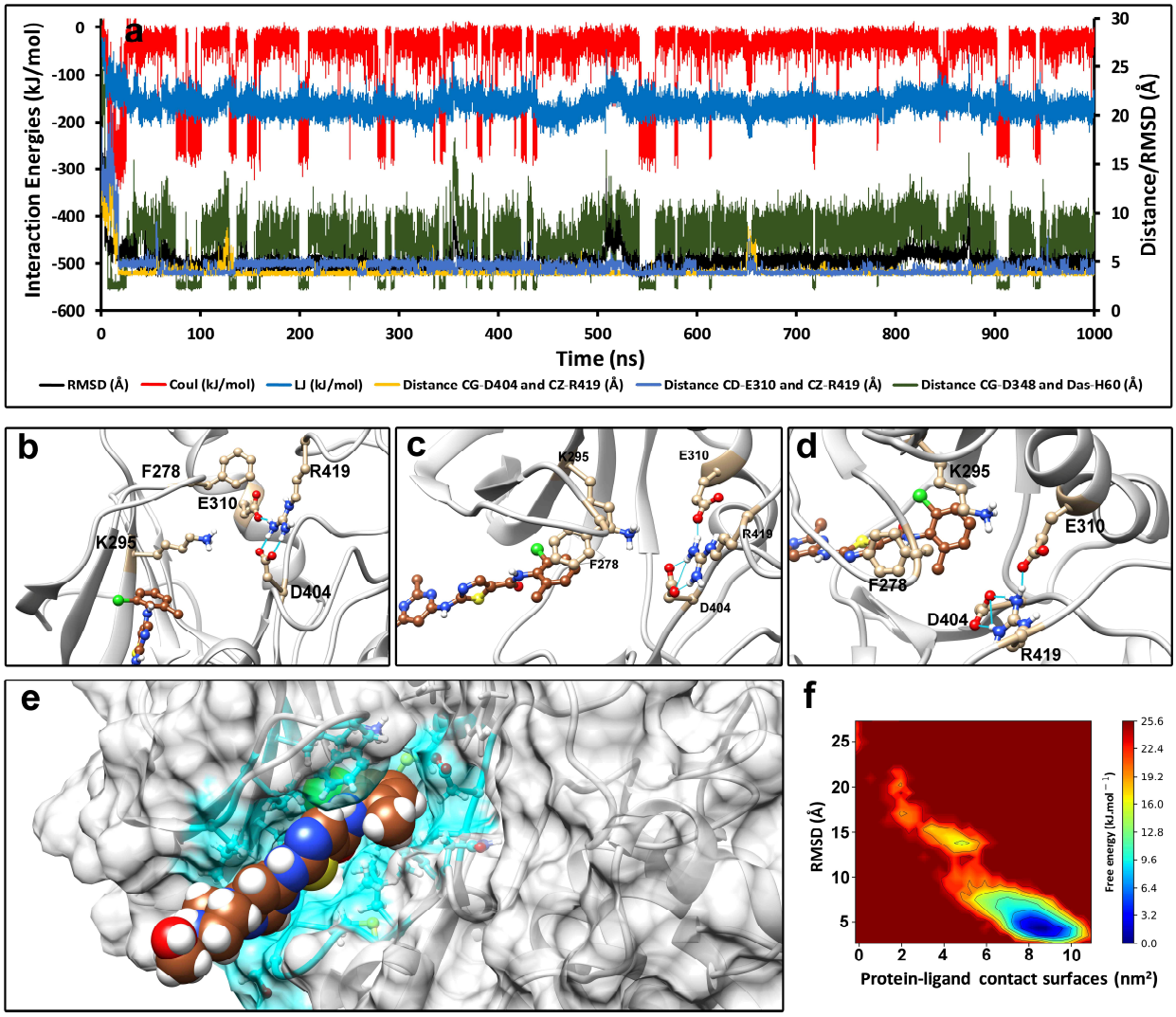


Figure 5. A overall scheme of the new binding mode for dasatinib in the run #5.5. a, The interaction energies of the protein-ligand complex and the distance between R419 and D404/E310 during 500 ns simulation. At the early stages of simulation, R419 started to confiscate both of E310 and D404 and preserved this conformation till the end of simulation. The Van der Waals interaction energy is relatively similar to the native bindings, but the electrostatic energy periodically dropped down in the wake of a salt bridge formation and breakage between the tail of dasatinib and D348. Surprisingly, RMSD for this novel binding mode reached to 3 Å. **b**, Salt bridge formation between R419 and both of E310 and D404. **c,** throughout this novel binding mode, the contribution of F278 to the binding is remarkable. F278, situated in the G-rich loop, can turn inside the binding pocket and stabilize the head of dasatinib by a π- stacking interaction. **d,** Dasatinib could get deep inside the binding pocket with the help of F278 and K295. **e,** The surface view of c-SRC kinase in complex with dasatinib. The run #5.5 shed a new light on our understanding about possible binding modes of dasatinib to the c-Src kinase by illustrating a novel binding mode

which had not been reported before. We found that dasatinib could bind to c-SRC kinase in a 180-degrees flipped conformation compared to the native binding mode in reference to the reported X-ray crystallography structures. **f,** The free energy landscape of dasatinib.

The unbinding pathways of dasatinib

For a good binder, just entering the binding pocket is not enough. After binding, it has to apply the brakes in order to stay in the binding pocket. Selectively entering into the binding pocket and do the binding must be accompanied with long residence times. Therefore, understanding the key elements of the unbinding process can be beneficial in designing of more effective drugs.

In order to find these key elements, we performed five long replicas with a duration time of 500 ns for each protonated and deprotonated forms of dasatinib in complex with c-Src kinase, which cumulated to the total run time of 5 μs (Fig8). By thoroughly analyzing and inspecting every single frame of the simulations and trajectories, we found that dasatinib's Achilles' heel is its own fluctuations. These fluctuations are due to the both protein intrinsic motions and dynamics, and reciprocal motions derived from water molecules. The more dasatinib jiggles and wiggles, the more interactions between binding pocket residues and dasatinib dismiss; thus, the more likely it is to be unleashed or unbound.

By and large, every protein has a particular magnitude of motions and dynamics. The c-Src kinase is no exception since it has reasonably normal motions; however, the conformations of the A-loop and the αC-helix are alterable and can eclipse the configuration of the binding pocket. When the A-loop is folded and the αC-helix is placed at the "in" conformation, more amino acids can interact with the ligand and shield it from water molecules so make it more stable (Fig6.a).

On the other hand, when the A-loop is unfolded and the αC-helix is placed at the "out" conformation, the binding pocket is fully exposed to the water molecules in a manner that allows water molecules to get inside the binding pocket and make the ligand unstable which results into the higher fluctuations (Fig6.b). As it was discussed before, the transiently breakage of the salt bridges between K295 and E310/D404, and also the interferences of R409 and R419 can result in the αC-helix-out conformation. So far, the cause of the A-loop transition from the folded to the

338 unfolded conformation is still unknown but it can play an important role in the binding and
339 unbinding processes[27].

340 By inspecting all of the simulations and trajectories frame by frame, we detected some
341 residues that can immensely alter the conformation and orientation of dasatinib inside the
342 binding pocket. One of these residues with the most pivotal role is T338. In the native binding
343 mode, the hydrogen bond between the amide group in the head of dasatinib and the OG atom of
344 the side chain of T338 pulls the head of dasatinib toward T338 and present a state that we refer
345 to it as the "forth state" hereinafter (Fig6.c).

346 On the contrary, the breakage of this hydrogen bond under the influence of various
347 factors causes the head of dasatinib to go farther away from T338, this state will be referred as
348 the "back state" (Fig6.d). This breakage is the trigger of the unbinding event. In fact, there are
349 two main transitions: (i) the transition from the "forth state" to the "back state" initiated by the
350 breakage of the hydrogen bond between T338 and the head of dasatinib, and (ii) the transition
351 from the "back state" to the unbound state provoked by the breakage of two hydrogen bonds
352 between the backbone of M341 and the spine of dasatinib.

353 Recently, Tiwary et al. have studied the nature of the binding pocket of c-Src[18]. They
354 were of the opinion that water molecules cannot flow inside the binding pocket as long as the
355 evolutionary conserved salt bridges are in place. Despite their statement, we found that water
356 molecules can flow either inside or outside of the binding pocket with the assistance of a number
357 of brilliantly positioned amino acids. Among these residues, M314 and M341 are the most
358 important ones which can hand over water molecules to the binding pocket at a certain rate.
359 M314 is located on the αC-helix and under the salt bridge between K295 and E310.

360 Normally, M314 can conserve one or two water molecules by a very weak hydrogen
361 bond, even in the presence of the salt bridges; it can hand over water molecules to the binding
362 pocket at a low rate (Fig6.g). But sometimes, it can channel water molecules alongside the αC-
363 helix (Fig6.e). However, this channel of water molecules only appears when the salt bridges are
364 broken and E310 is headed outwards. The consecutive presence of water molecules underneath
365 the head of dasatinib raises the chance of the breakage of the hydrogen bond between T338 and
366 dasatinib. Indeed, both M341 and M314 can deliver water molecules inside the pocket, but only
367 M314 can channel water in the deep parts of the binding pocket. By the collaboration of K401,

368 E339 and M341 a water molecule enters the binding pocket from either of the two sides of the
369 salt bridge between K401 and E339 and mediates the hydrogen bond between T338 and
370 dasatinib (Fig6.f). Additionally, the rotation of the dihedral angle of T338 side chain puts the OG
371 atom of T338 away from the amide group of dasatinib and close to the water molecules (Fig6.d).
372 All of these incidents lead to the breakage of the hydrogen bond between the OG of T338 and the
373 nitrogen atom of the amide group and consequently reposition of the head of dasatinib into the
374 "back state" (Fig6.h).

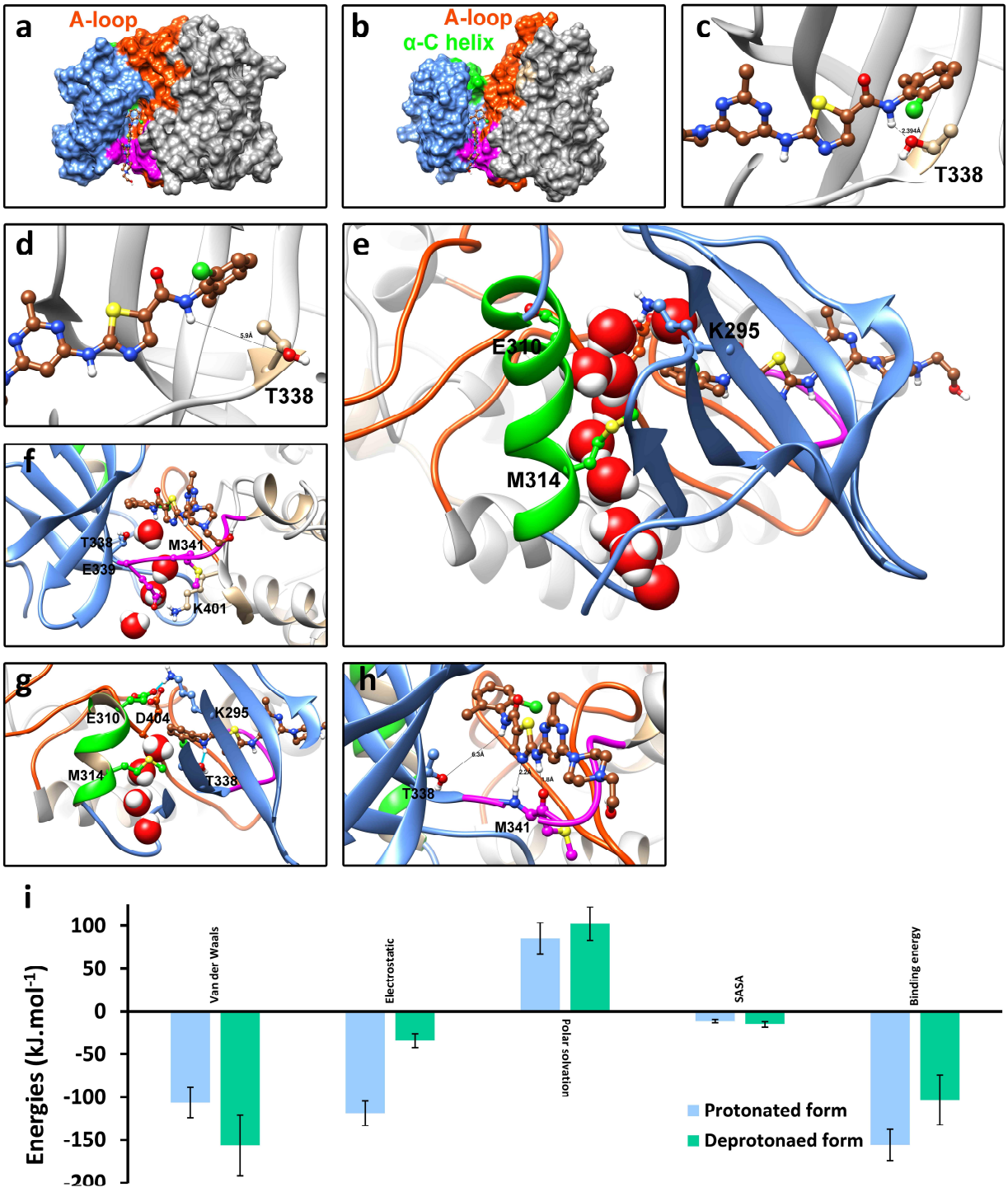


Figure 6. The atomic details of the dasatinib unbinding pathways. a, The conformation of c-Src kinase when A-loop is folded and α-C helix is located at the "in" conformation. **b**, The conformation of protein when A-loop is unfolded and the α-C helix is located at the "out" conformation so that the binding pocket is exposed to the solvent molecules. **c,** The hydrogen bond between OG atom of T338 and dasatinib keeps the head of dasatinib down in the "forth state". **d,** The breakage of hydrogen bond between T338 and dasatinib frees the ligand to get to the "back state". This breakage was undertaken by the rotation of the dihedral angle of T338 side-chain. **e,** The

water channel was created after the breakage of the salt bridges and when the E310 is headed outwards. This channel was managed by M314. **f,** Water-mediation of the hydrogen bond between T338 and dasatinib was engineered by a water molecule which was delivered into the pocket with the help of M341, E339 and K401. These residues act like a mouth and can swallow a single water molecule inside at a certain rate. **g,** The water delivery of M314 which acts as a suction machine and pull a single water molecule inside and conserve it in the deep parts of the binding pocket. **h,** The two hydrogen bonds formed between the backbone atoms of M341 and dasatinib which keep the spine of dasatinib down and help to hold the head in the back state. The breakage of these two hydrogen bonds is the rate-limiting step of the unbinding event. **i,** The average MMPBSA binding free energy of the protonated and the deprotonated dasatinib in complex with c-Src kinase protein in the 5 μs production runs. The values are the mean ± SD.

During the production runs, it was observed that dasatinib could be deviated from the native binding form as the head lifts up. However, after a while it goes down to the native binding form again. This gave us the idea that dasatinib motions in native pose work in a "back and forth" sort of manner and that's how we named these two states. Being at the "back state" while water molecules occupied the previous position (the "forth state"), is necessary for occurrence of a full unbinding event, but not suffice. There are two hydrogen bonds between the spine of dasatinib and the backbone of M341 which can hold the spine of dasatinib tightly and glue it to the binding pocket with the help of the other residues that hug the spine of dasatinib (Fig6.h). The back state was observed in 6 out of 10 long replicas which means that it is very probable (Fig8), however, the breakage of the two hydrogen bonds between M341 and dasatinib can be considered as a very rare event. In the 5 μs of production runs, by monitoring the length of these two hydrogen bonds, only a few short moments were spotted in which these hydrogen bonds were either broken or weakened (Fig7.a,b,c), and none of them persisted more than 2 ns. Thus, it crossed our mind that these two strong bonds evince a slow step throughout the unbinding event and the transition from the back state to the unbound phase or other states is the rate-limiting step of the entire unbinding pathway. Hence, the breakage of these hydrogen bonds was foregrounded from hereafter. The results of the free energy landscapes (Fig7.d,e,f) and the contribution of each residue of the binding pocket in the interaction energies (Ext Fig2) can also prove the fact that the breakage of these two hydrogen bonds with M341, which remain formed in the back state, is the rate-limiting step of the unbinding pathway.

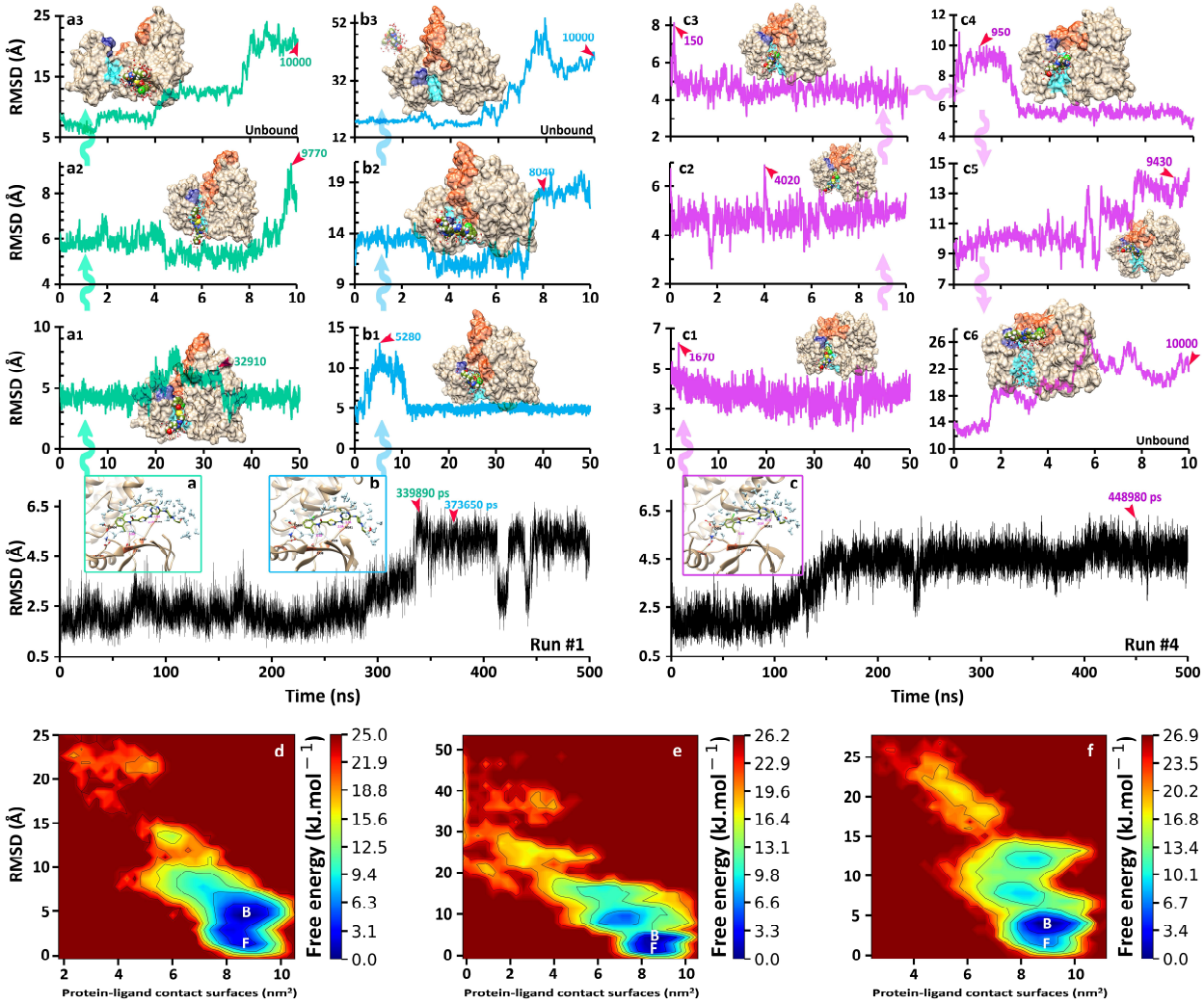


Figure 7. The simulations diagram for unbinding pathways of dasatinib. a, The frame time of 339890 ps from the run #1 was picked up. In this frame time dasatinib was in the back state and one of the established hydrogen bonds with M341 was weakened and the head of dasatinib was fully solvated. This snap shot was used as the starting point for a rational sampling attempt consisted of three tandem rounds of simulations (a1 to a3), we named it "A cascade". The best snapshots of the successful trajectories which were selected to be continued for the next rounds are indicated by red arrows on the RMSD plots. **b,** The frame time of 373650 ps from the run #1 was selected. In this frame time the dasatinib was in the back state and both of the established hydrogen bonds with M341 were weakened and the head of dasatinib was fully solvated. The "B cascade" (b1 to b3) was started from this snapshot and finished after the full unbound conformation was achieved. **c,** The frame time of 448980 ps from the run #4 was pinned. In this frame time dasatinib was in the back state position, both of established hydrogen bonds with M341 were broken and the spine was water mediated. This snapshot was also continued in 6 rounds of rational sampling (c1 to c6), the "C cascade". In the run #4, since the A-loop was folded and the α-C helix was located at the "in" conformation, it took 6 rounds of rational sampling to achieve the full unbound conformation. Nevertheless, all of these cascades were successful and we reached the full unbound conformation of dasatinib. **d,** The free energy landscape of the A cascade, **e,** the B cascade and, **f,** the C cascade. The letters "B" and "F" stand for the "Back state" and the "Forth state" respectively. The hydrogen bonds and distances are

presented by blue solid lines and pink dashed lines in a, b and c figures, resp. In surface view, cyan, blue and orange red colors indicate binding pocket, G-rich loop and A-loop, resp.

From the deprotonated dasatinib lane, two frame times of the run #1, 339890ps (Fig7.a) and 373650ps (Fig7.b), and one frame time of the run #4, 448950ps (Fig7.c) were selected for further simulations. All of them were extended in multiple short replicas (Fig2). Then, by utilizing the rational sampling, the good conformations were conveyed to the next rounds of simulations and so forth. Eventually, after thoughtful and careful samplings, we successfully achieved the first fully unbiased unbinding phase of dasatinib while RMSD values exceeded 20 Å.

While three rounds of simulations were sufficing to attain unbinding for the two snapshots of the run #1 (Fig7.a1-3, b1-3), it was more challenging for the snapshot of the run #4 as it demanded six rounds of samplings (Fig7.c1-6). In the run #4 unlike the run #1, the A-loop was folded so it was assumed that the unfolded conformation of the A-loop increases the hydration of the binding pocket which contributes to higher fluctuations of ligand. Thus, the frame time of the run #4 claimed twice the number of rounds of sampling in comparison with the run #1, a reaffirmation of this fact that the different conformations of the target protein must be fully clarified prior to drug design.

The main purpose of this sampling is making rare events more likely to happen. The best samples were extracted by analysis and inspection of every core aspects that were involved in the binding/unbinding pathways, therefore, we named it “rational sampling”. Our new findings also indicate that most of the key events in the binding and unbinding pathways occurred in very rare and short time intervals, which is why utilization of short replicas is much more effective for achieving these key events, rather than long-time simulations.

Stable states may not be palpable enough in the free energy landscapes of the three unbinding events (Fig7.d,e,f), which could be on account of the rational sampling and also because they were obtained by concatenation of all of the simulated trajectories of each route. However, each graph represents an unbinding route for dasatinib without implementation of any biasing forces, just like what happens in the real world.

459 Two different modes were observed through the unbinding. The sequence of the events
460 during the unbinding attempts was contrastive although all runs possessed similar elements, just
461 like the binding. In the B and C cascades, first, water molecules seeped underneath the spine
462 from the tail side and broke the two hydrogen bonds between the spine of dasatinib and M341,
463 then the tail and spine lifted up and solvated. Nonetheless, the head remained in contact with the
464 binding pocket until the last moments of the unbinding event. On the contrary, in the A cascade,
465 water molecules seeped underneath the spine of dasatinib from inside of the binding pocket and
466 broke just one of the two hydrogen bonds between the spine of dasatinib and M341. In this case,
467 the head of dasatinib was solvated first and then, after more samplings, the other hydrogen bond
468 with M341 was broken so that the ligand was released from the binding pocket and diffused on
469 the protein's surface.

470 To estimate the unbinding rate (K_{off}) of dasatinib, the mean residence time of 1.9 μs was
471 calculated from three independent unbinding events which was far from the experimental
472 observations, 18s[18]. We opine that this substantial difference is resulted since we accelerated the
473 unbinding process by utilization of the rational sampling. It is taken for granted that no current
474 MD approaches and computational resource are capable enough to simulate an unbinding event
475 in real-time. Therefore, it is inexorable to provide a statistical method for calculation of K_{off}.
476 Although several approaches and formulas have been developed for calculation of K_{off} by this
477 time[15,18,28], no one fits our case as in the rational sampling dozens of factors must be taken into
478 account. However, we aim to formulate K_{off} for the unbinding process by truly analyzing the
479 contribution of each factor involved in the rational sampling for our future studies.

480 The "back state" was observable at 22 percent of the overall 2.5 μs production run of the
481 deprotonated dasatinib, while it was 38 percent for protonated form in the same amount of run-
482 time (Fig8). It was evident that the inclination of protonated dasatinib for orienting to the back
483 state was almost 1.8 times more than the deprotonated form. The percentage of the back and
484 forth states was calculated by dividing the sum of RMSD values higher than 3.5 Å over the total
485 RMSD values. As it was mentioned before (Fig1.f), the positive formal charge on the tail of
486 protonated dasatinib had a dramatic impact on the entire characteristics of this molecule. This
487 extra formal charge made dasatinib to form much stronger electrostatic interactions with the
488 protein. The protonated form of dasatinib had a lower binding free energy and a higher affinity to

the pocket compared to the deprotonated form, presumably caused by an elevation in electrostatic interactions (Fig6.i). Despite all the efforts made to generate potent replicas from the protonated dasatinib when it was oriented at the back state, no unbinding was achieved. This can vouch for the validity and accuracy of these systems in which even a small biasing force can result into misconception.

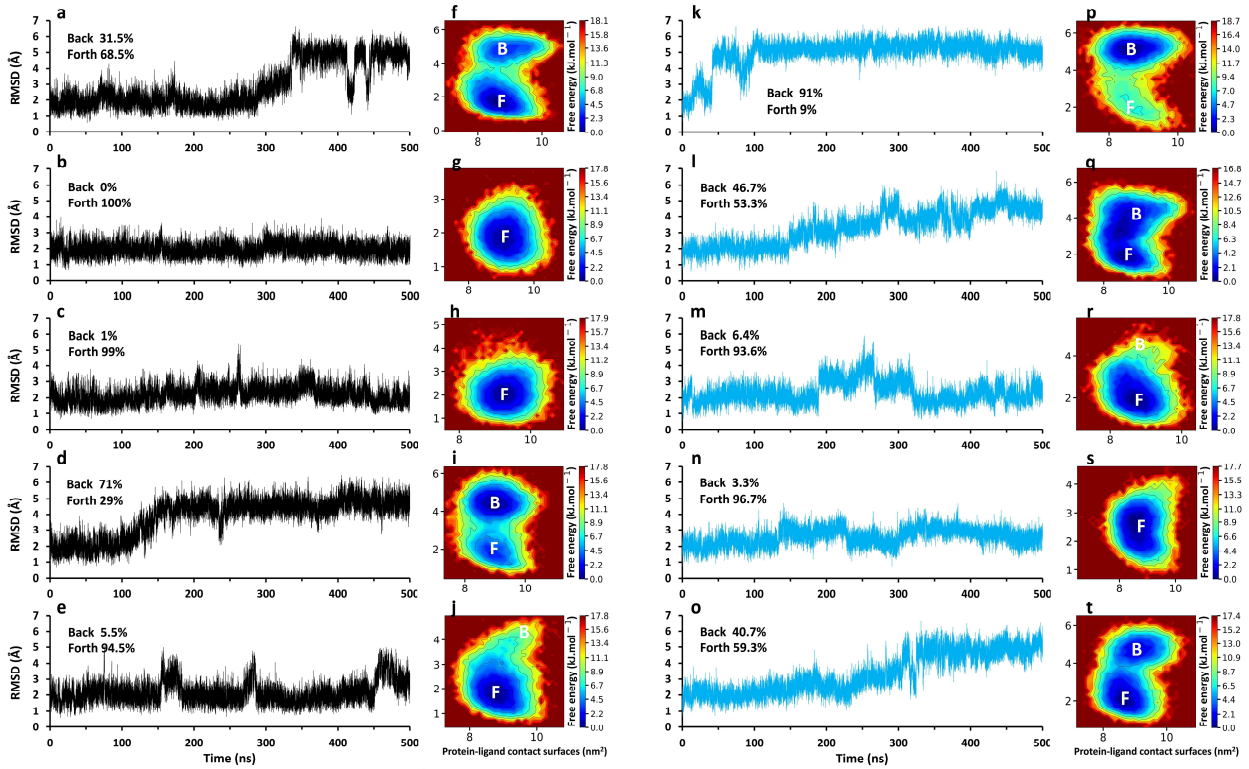


Figure 3. The RMSD plots and the free energy landscapes of the ten production runs of the unbinding. a-j, the deprotonated form of dasatinib and, k-t, the protonated form of dasatinib. The letters "B" and "F" stand for the "Back state" and the "Forth state" respectively.

It is also necessary to mention that all of the remarkable details that are presented in this work about the binding and unbinding process in dasatinib-c-Src kinase complex are just a tiny fraction of what is happening in cells. All we know so far about structural biology is only the tip of the iceberg. The details of the binding and unbinding processes are shown in nine videos which were included in the supplementary section.

Conclusion

Herein, we have introduced a stratagem to encourage researchers all around the world to simulate the iconic binding and unbinding events in atomistic details with any reasonable computational budget while no biasing force were implemented. By imposing repulsive forces among ligands and utilization of rational sampling the binding and unbinding processes were extremely speeded up.

In the binding section, the sampling of the target protein was enhanced by introducing a high concentration of ligands inside the simulation box on one side, and the ligands aggregation was prevented by imposing repulsive forces among them on the other side, meanwhile during all of this process, the rational sampling could be an extra help, especially for the discovery of new binding pockets and binding poses.

In the unbinding section, our magic wicker was the "rational sampling" which immensely accelerated searching for significant events by simply using a few rounds of short-time replicas instead of running extravagant long-time simulations. This is a promising stratagem that empowers researchers to perform enticing projects such as investigation of unbinding mechanisms. On this ground, the UUMD is a great help for researcher, who do not own supercomputers, to probe cellular mechanism in atomistic details by using a typical computer. Furthermore, the UUMD is advantageous in drug discovery as well.

Although we have achieved this great efficiency by employing OPLS force field, other force fields may be more efficient or challenging, a question that should be answered in future works. In drug discovery projects, on both commercial and academic grounds, efficiency is of immense significance. We are of the opinion that the UUMD is just initiated and there is still a long way for its development. Screening, identification and optimization of potential compounds based on the target macromolecule structure can be just some other appliances of UUMD. It is definitely more time-consuming than outdated methods like molecular docking, but it's extremely accurate and worth the costs.

Acknowledgment

We would like to thank the developers of Gromacs, UCSF chimera and VMD for their brilliant and flexible programs and also their incredible support. This task could not be done without these programs. This investigation was supported by the grant number 96-1206 from Golestan University, Gorgan, Iran.

Methods

Construction of systems

The X-ray crystallography of human c-Src kinase proteins with PDB ID: 1Y57[29] was obtained from Protein Data Bank, respectively. Missing side chains and atoms were added and refined using UCSF Chimera[30]. Then, the co-crystallized ligands and water molecules were removed and the apo-protein structure was simulated to reach equilibrium state for starting UMD studies. Hetero ligands (protonated and deprotonated dasatinib parameterized using ACEPYPE[31] with the default setting for assigning the partial charges and atom types. Dasatinib has one ionization state in the pH range of 6 to 8 and its ionization constants (pKa) was to be approximately 7.29[32]. So, at the physiological pH of the cell, there are two possible forms of dasatinib, protonated and deprotonated forms. Because both forms are present in the cell, we took both of them into account. The positive formal charge in the tail of dasatinib has a major effect on the entire characteristics of the compound. For determining the ligands binding pathways, a high concentration of ligands along with the corresponding protein was inserted to the simulation box. In the first round of simulations of each system (protonated and deprotonated), a large number of short runs with different velocities were carried out in order to find potential candidates for the binding pathway. The major criteria for the selection of candidates were the proper orientation and conformation of ligands compared to the co-crystallized ligand (PDB ID: 3G5D)[33]. The rest of the details are shown in the Flowchart.

MD simulation protocol and analyses

All binding and unbinding pathways simulations were initiated with pre-equilibrated state of the relevant apo-protein using the OPLS force field[34] in GROMACS 2018[35]. The related apo-protein was placed in the center of a triclinic box with a distance of 1.5 nm from all edges. First,

563 sixteen relevant ligands were inserted into each simulation box with random positions and
564 solvated with TIP3P water model[36]. Then, sodium and chloride ions were added to produce a
565 neutral physiological salt concentration of 150 mM. Each system was energy minimized, using
566 steepest descent algorithm, until the Fmax was found to be smaller than 10 $kJ.mol^{-1}.nm^{-1}$. All of
567 the covalent bonds were constrained using the Linear Constraint Solver (LINCS) algorithm[37] to
568 maintain constant bond lengths. The long-range electrostatic interactions were treated using the
569 Particle Mesh Ewald (PME) method[38] and the cut off radii for Coulomb and Van der Waals
570 short-range interactions was set to 0.9 nm for Dasatinib-c-Src systems. The modified Berendsen
571 (V-rescale) thermostat[39] and Parrinello–Rahman barostat[40] respectively were applied for 100 and
572 300 ps to keep the system in the stable environmental conditions (310 K, 1 Bar). Finally, UMD
573 simulations were carried out under the periodic boundary conditions (PBC), set at XYZ
574 coordinates to ensure that the atoms had stayed inside the simulation box, and the subsequent
575 analyses were then performed using GROMACS utilities, VMD[41] and USCF Chimera, and also
576 the plots were created using Daniel's XL Toolbox (v 7.3.2) add-in[42]. In addition, to estimate the
577 interaction energies we used both GROMACS utilities and the g_mmpbsa package[43]. The energy
578 plots were rendered using Matplotlib[44].
579

580 **References**

581 1 Askew, E. B., Gampe, R. T., Jr., Stanley, T. B., Faggart, J. L. & Wilson, E. M. Modulation of
582 androgen receptor activation function 2 by testosterone and dihydrotestosterone. *J Biol Chem*
583 **282**, 25801-25816, doi:10.1074/jbc.M703268200 (2007).
584 2 Shan, Y. *et al.* How does a drug molecule find its target binding site? *Journal of the American*
585 *Chemical Society* **133**, 9181-9183, doi:10.1021/ja202726y (2011).
586 3 Buch, I., Giorgino, T. & De Fabritiis, G. Complete reconstruction of an enzyme-inhibitor binding
587 process by molecular dynamics simulations. *Proc Natl Acad Sci U S A* **108**, 10184-10189,
588 doi:10.1073/pnas.1103547108 (2011).
589 4 Dror, R. O. *et al.* Pathway and mechanism of drug binding to G-protein-coupled receptors. *Proc*
590 *Natl Acad Sci U S A* **108**, 13118-13123, doi:10.1073/pnas.1104614108 (2011).
591 5 Stanley, N., Pardo, L. & Fabritiis, G. D. The pathway of ligand entry from the membrane bilayer
592 to a lipid G protein-coupled receptor. *Sci Rep* **6**, 22639, doi:10.1038/srep22639 (2016).
593 6 Gohlke, H. *et al.* Binding region of alanopine dehydrogenase predicted by unbiased molecular
594 dynamics simulations of ligand diffusion. *J Chem Inf Model* **53**, 2493-2498,
595 doi:10.1021/ci400370y (2013).
596 7 Mondal, J., Ahalawat, N., Pandit, S., Kay, L. E. & Vallurupalli, P. Atomic resolution mechanism of
597 ligand binding to a solvent inaccessible cavity in T4 lysozyme. *PLOS Computational Biology* **14**,
598 e1006180, doi:10.1371/journal.pcbi.1006180 (2018).

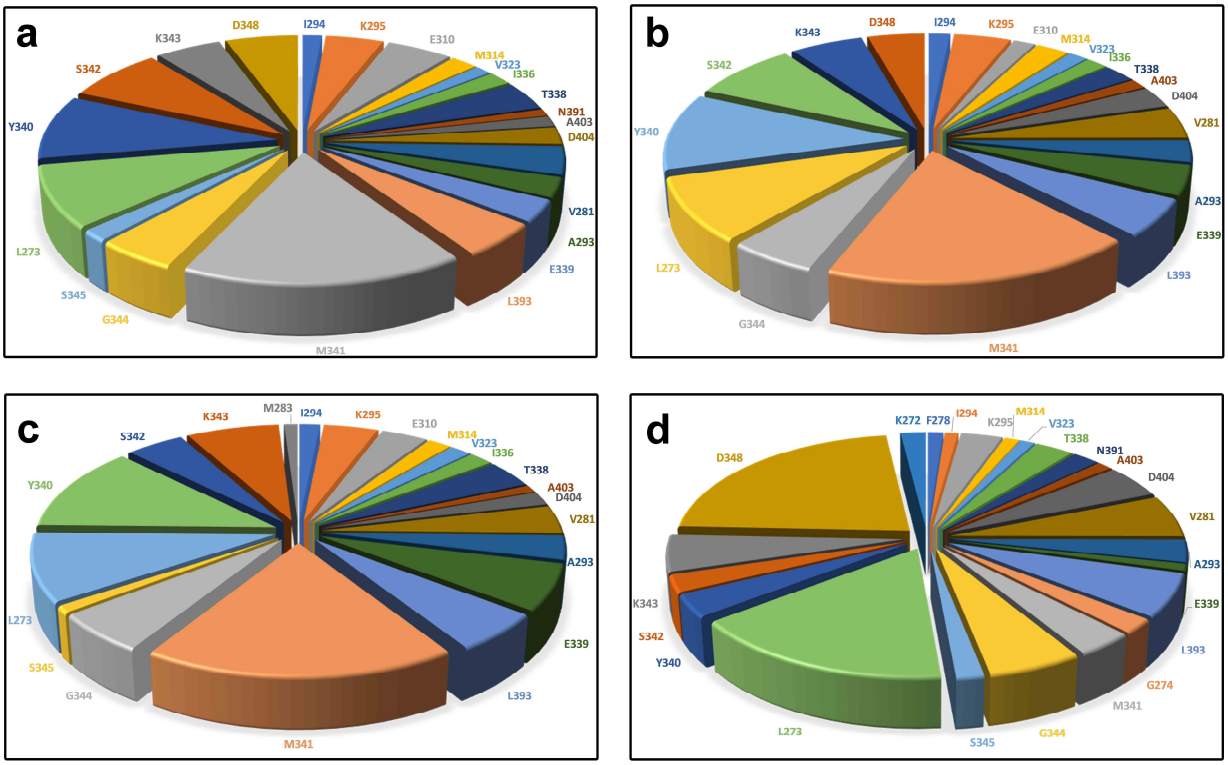
599 8 Schneider, S., Provasi, D. & Filizola, M. How Oliceridine (TRV-130) Binds and Stabilizes a mu-
600 Opioid Receptor Conformational State That Selectively Triggers G Protein Signaling Pathways.
601 *Biochemistry* **55**, 6456-6466, doi:10.1021/acs.biochem.6b00948 (2016).
602 9 Decherchi, S., Berteotti, A., Bottegoni, G., Rocchia, W. & Cavalli, A. The ligand binding
603 mechanism to purine nucleoside phosphorylase elucidated via molecular dynamics and machine
604 learning. *Nat Commun* **6**, 6155, doi:10.1038/ncomms7155 (2015).
605 10 Dickson, A. & Lotz, S. D. Ligand Release Pathways Obtained with WExplore: Residence Times and
606 Mechanisms. *J Phys Chem B* **120**, 5377-5385, doi:10.1021/acs.jpcb.6b04012 (2016).
607 11 Dickson, A. & Lotz, S. D. Multiple Ligand Unbinding Pathways and Ligand-Induced Destabilization
608 Revealed by WExplore. *Biophys J* **112**, 620-629, doi:10.1016/j.bpj.2017.01.006 (2017).
609 12 Casasnovas, R., Limongelli, V., Tiwary, P., Carloni, P. & Parrinello, M. Unbinding Kinetics of a p38
610 MAP Kinase Type II Inhibitor from Metadynamics Simulations. *J Am Chem Soc* **139**, 4780-4788,
611 doi:10.1021/jacs.6b12950 (2017).
612 13 Guo, D. *et al.* Molecular Basis of Ligand Dissociation from the Adenosine A2A Receptor. *Mol*
613 *Pharmacol* **89**, 485-491, doi:10.1124/mol.115.102657 (2016).
614 14 Motta, S., Callea, L., Giani Tagliabue, S. & Bonati, L. Exploring the PXR ligand binding mechanism
615 with advanced Molecular Dynamics methods. *Sci Rep* **8**, 16207, doi:10.1038/s41598-018-34373-
616 z (2018).
617 15 Tiwary, P., Limongelli, V., Salvalaglio, M. & Parrinello, M. Kinetics of protein-ligand unbinding:
618 Predicting pathways, rates, and rate-limiting steps. *Proc Natl Acad Sci U S A* **112**, E386-391,
619 doi:10.1073/pnas.1424461112 (2015).
620 16 Limongelli, V., Bonomi, M. & Parrinello, M. Funnel metadynamics as accurate binding free-
621 energy method. *Proc Natl Acad Sci U S A* **110**, 6358-6363, doi:10.1073/pnas.1303186110 (2013).
622 17 You, W. & Chang, C. A. Role of Molecular Interactions and Protein Rearrangement in the
623 Dissociation Kinetics of p38alpha MAP Kinase Type-I/II/III Inhibitors. *J Chem Inf Model* **58**, 968-
624 981, doi:10.1021/acs.jcim.7b00640 (2018).
625 18 Tiwary, P., Mondal, J. & Berne, B. J. How and when does an anticancer drug leave its binding
626 site? *Science Advances* **3** (2017).
627 19 Sinko, W., Miao, Y., de Oliveira, C. A. & McCammon, J. A. Population based reweighting of scaled
628 molecular dynamics. *J Phys Chem B* **117**, 12759-12768, doi:10.1021/jp401587e (2013).
629 20 Knapp, B., Ospina, L. & Deane, C. M. Avoiding False Positive Conclusions in Molecular
630 Simulation: The Importance of Replicas. *Journal of Chemical Theory and Computation* **14**, 6127-
631 6138, doi:10.1021/acs.jctc.8b00391 (2018).
632 21 Roskoski, R. A historical overview of protein kinases and their targeted small molecule inhibitors.
633 *Pharmacological Research* **100**, 1-23 (2015).
634 22 Meng, Y. & Roux, B. Locking the Active Conformation of c-Src Kinase through the
635 Phosphorylation of the Activation Loop. *Journal of Molecular Biology* **426**, 423-435 (2014).
636 23 Shan, Y. *et al.* A conserved protonation-dependent switch controls drug binding in the Abl
637 kinase. *Proc Natl Acad Sci U S A* **106**, 139-144, doi:10.1073/pnas.0811223106 (2009).
638 24 McClendon, C. L., Kornev, A. P., Gilson, M. K. & Taylor, S. S. Dynamic architecture of a protein
639 kinase. *Proceedings of the National Academy of Sciences* **111**, E4623-E4631,
640 doi:10.1073/pnas.1418402111 (2014).
641 25 Roskoski, R. Src protein-tyrosine kinase structure, mechanism, and small molecule inhibitors.
642 *Pharmacological Research* **94**, 9-25 (2015).
643 26 Lin, Y. L., Meng, Y., Jiang, W. & Roux, B. Explaining why Gleevec is a specific and potent inhibitor
644 of Abl kinase. *Proc Natl Acad Sci U S A* **110**, 1664-1669, doi:10.1073/pnas.1214330110 (2013).

645 27 Shukla, D., Meng, Y., Roux, B. & Pande, V. S. Activation pathway of Src kinase reveals
646 intermediate states as targets for drug design. *Nature Communications* **5**, 3397,
647 doi:10.1038/ncomms4397 (2014).
648 28 Mollica, L. *et al.* Kinetics of protein-ligand unbinding via smoothed potential molecular dynamics
649 simulations. *Sci Rep* **5**, 11539, doi:10.1038/srep11539 (2015).
650 29 Cowan-Jacob, S. W. *et al.* The crystal structure of a c-Src complex in an active conformation
651 suggests possible steps in c-Src activation. *Structure* **13**, 861-871, doi:10.1016/j.str.2005.03.012
652 (2005).
653 30 Pettersen, E. F. *et al.* UCSF Chimera--a visualization system for exploratory research and analysis.
654 *J Comput Chem* **25**, 1605-1612, doi:10.1002/jcc.20084 (2004).
655 31 Sousa da Silva, A. W. & Vranken, W. F. ACPYPE - AnteChamber PYthon Parser interfacE. *BMC*
656 *Research Notes* **5**, 367, doi:10.1186/1756-0500-5-367 (2012).
657 32 Gaulton, A. *et al.* The ChEMBL database in 2017. *Nucleic Acids Res* **45**, D945-D954,
658 doi:10.1093/nar/gkw1074 (2017).
659 33 Getlik, M. *et al.* Hybrid compound design to overcome the gatekeeper T338M mutation in cSrc. *J*
660 *Med Chem* **52**, 3915-3926, doi:10.1021/jm9002928 (2009).
661 34 Kaminski, G. A., Friesner, R. A., Tirado-Rives, J. & Jorgensen, W. L. Evaluation and
662 Reparametrization of the OPLS-AA Force Field for Proteins via Comparison with Accurate
663 Quantum Chemical Calculations on Peptides. *The Journal of Physical Chemistry B* **105**, 6474-
664 6487, doi:10.1021/jp003919d (2001).
665 35 Abraham, M. J. *et al.* GROMACS: High performance molecular simulations through multi-level
666 parallelism from laptops to supercomputers. *SoftwareX* **1-2**, 19-25,
667 doi:10.1016/j.softx.2015.06.001 (2015).
668 36 Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of
669 simple potential functions for simulating liquid water. *The Journal of chemical physics* **79**, 926-
670 935, doi:10.1063/1.445869 (1983).
671 37 Hess, B., Bekker, H., Berendsen, H. J. C. & Fraaije, J. G. E. M. LINCS: A linear constraint solver for
672 molecular simulations. *Journal of Computational Chemistry* **18**, 1463-1472,
673 doi:10.1002/(SICI)1096-987X(199709)18:12<1463::AID-JCC4>3.0.CO;2-H (1997).
674 38 Darden, T., York, D. & Pedersen, L. Particle mesh Ewald: An N·log(N) method for Ewald sums in
675 large systems. *The Journal of chemical physics* **98**, 10089-10092, doi:10.1063/1.464397 (1993).
676 39 Bussi, G., Donadio, D. & Parrinello, M. Canonical sampling through velocity rescaling. *The Journal*
677 *of chemical physics* **126**, 014101, doi:10.1063/1.2408420 (2007).
678 40 Parrinello, M. & Rahman, A. Polymorphic transitions in single crystals: A new molecular
679 dynamics method. *Journal of applied physics* **52**, 7182-7190, doi:10.1063/1.328693 (1981).
680 41 Humphrey, W., Dalke, A. & Schulten, K. VMD: Visual molecular dynamics. *Journal of Molecular*
681 *Graphics* **14**, 33-38, doi:10.1016/0263-7855(96)00018-5 (1996).
682 42 Kraus, D. Consolidated data analysis and presentation using an open-source add-in for the
683 Microsoft Excel® spreadsheet software. *Medical Writing* **23**, 25-28,
684 doi:10.1179/2047480613Z.000000000181 (2014).
685 43 Kumari, R., Kumar, R. & Lynn, A. g_mmpbsa—A GROMACS Tool for High-Throughput MM-PBSA
686 Calculations. *Journal of Chemical Information and Modeling* **54**, 1951-1962,
687 doi:10.1021/ci500020m (2014).
688 44 Hunter, J. D. Matplotlib: A 2D Graphics Environment. *Computing in Science & Engineering* **9**, 90-
689 95, doi:10.1109/MCSE.2007.55 (2007).

690

691 **Extended Figures**

692



693

694 **Ext Fig.1. The interaction energy contribution of each residue in the protein-ligand binding pathway. a,** run
695 #23.2. **b,** run #40.5. **c,** run #14.5. **d,** run #5.5.

696

697

698

699

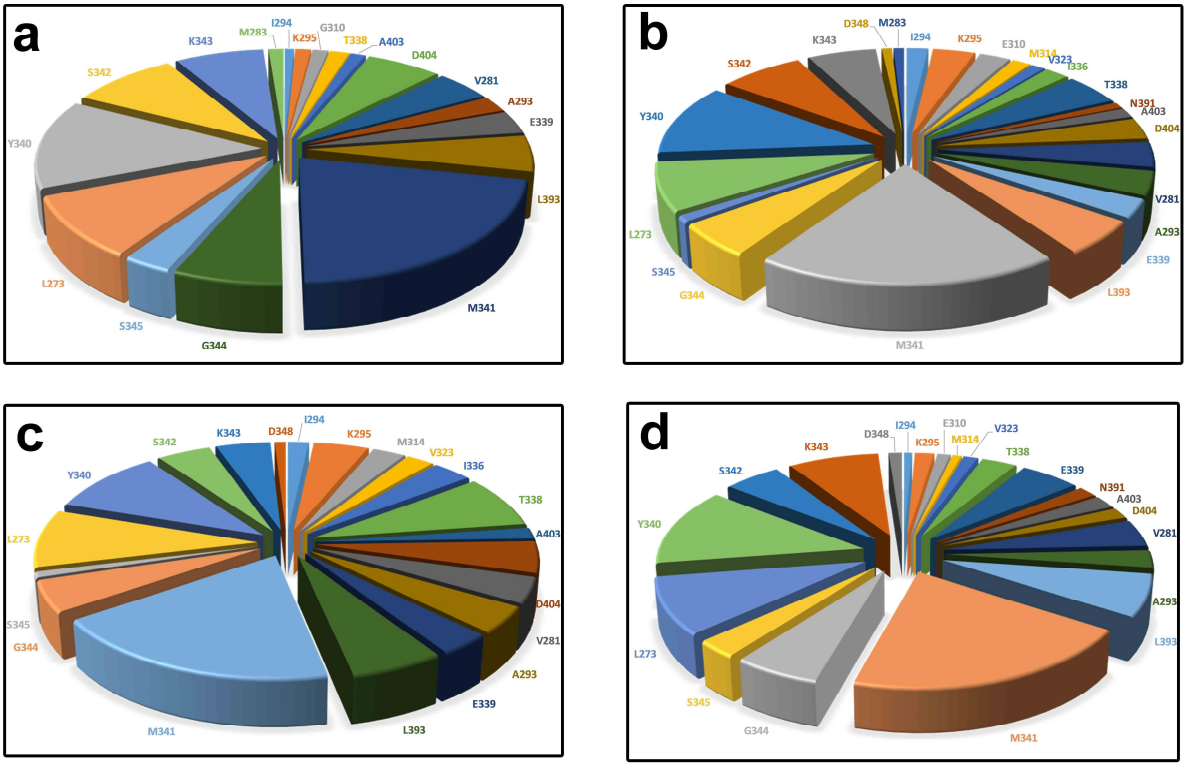
700

701

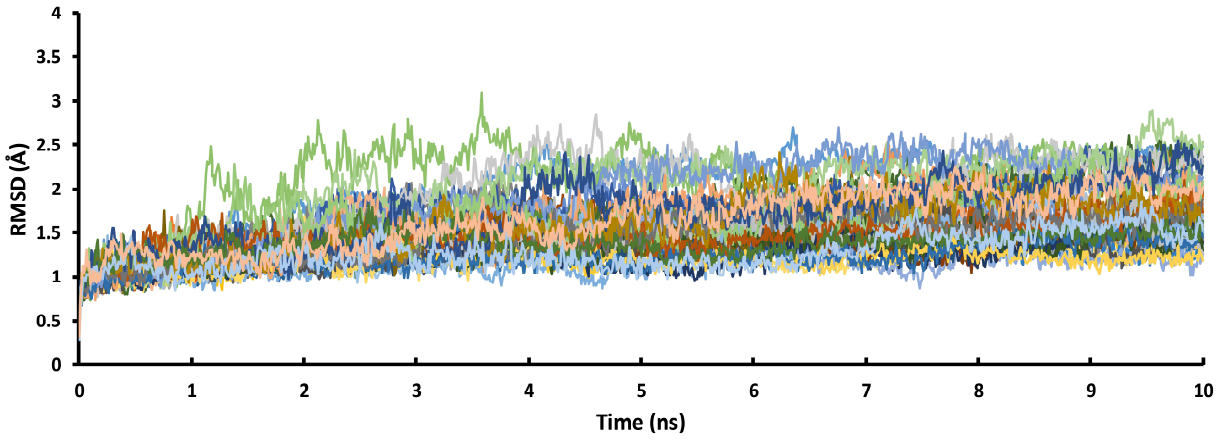
702

703

704



Ext Fig.2. The interaction energy contribution of each residue in the protein-ligand unbinding pathway in four out of ten 500 ns runs. a, Protonated dasatinib, run#1. **b,** Protonated dasatinib, run#5. **c,** Deprotonated dasatinib, run#2. **d,** Deprotonated dasatinib, run#4.



721

722 **Ext Fig.3. The RMSD values of the protein backbone in the 50 sampling runs with 16 ligands present in the**
723 **simulation box. The RMSD values show that the high concentration of the ligand has no effect on the folding of**
724 **the protein.**

725

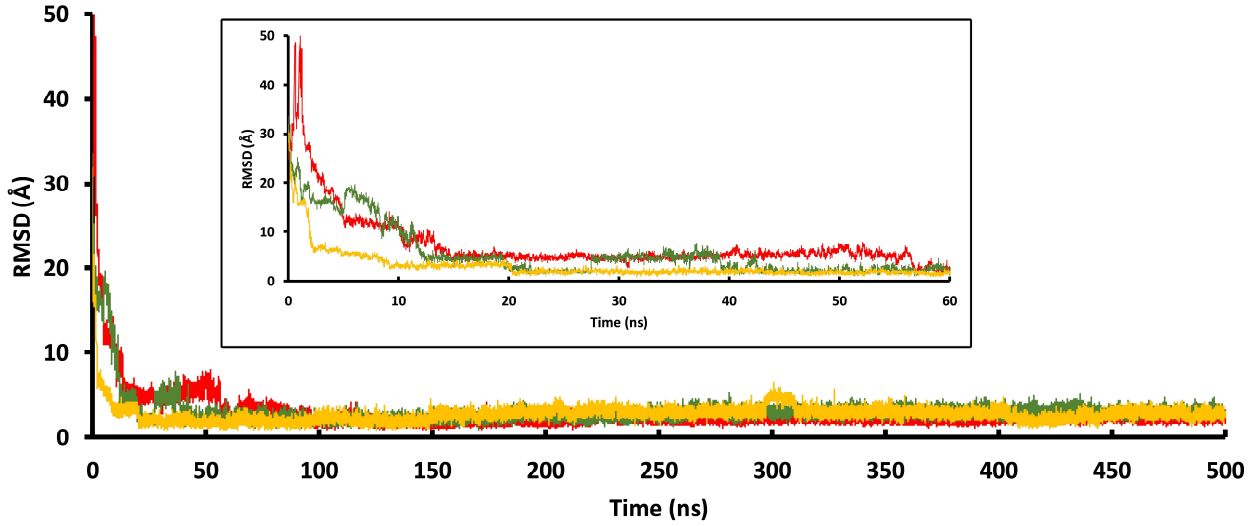
726

727

728

729

730



731

732 **Ext Fig.4. The RMSD values of dasatinib in the runs #23.3 (red), #40.5 (green) and #14.5 (yellow) that reached the**
733 **native binding pose.**