RESEARCH ARTICLE

PROTEINS WILEY

A shift of dynamic equilibrium between the KIT active and inactive states causes drug resistance

Sanjay K. Srikakulam[1,2,3] | Tomas Bastys[2,4] | Olga V. Kalinina[1,5]

[1]Helmholtz Institute for Pharmaceutical Research Saarland (HIPS), Helmholtz Centre for Infection Research (HZI), Saarbrücken, Germany

[2]Graduate School of Computer Science, Saarland University, Saarbrücken, Germany

[3]Interdisciplinary Graduate School of Natural Product Research, Saarland University, Saarbrücken, Germany

[4]Max Planck Institute for Informatics, Saarbrücken, Germany

[5]Medical Faculty, Saarland University, Homburg, Germany

Correspondence
Olga V. Kalinina, Helmholtz Institute for Pharmaceutical Research Saarland (HIPS), Helmholtz Centre for Infection Research (HZI), Campus E8.1, 66123 Saarbrücken, Germany and Medical Faculty, Saarland University, 66421 Homburg, Germany Saarbrücken, Germany.
Email: olga.kalinina@helmholtz-hips.de

Abstract

Tyrosine phosphorylation, a highly regulated post-translational modification, is carried out by the enzyme tyrosine kinase (TK). TKs are important mediators in signaling cascades, facilitating diverse biological processes in response to stimuli. TKs may acquire mutations leading to malignancy and are viable targets for anti-cancer drugs. Mast/stem cell growth factor receptor KIT is a TK involved in cell differentiation, whose dysregulation leads to various types of cancer, including gastrointestinal stromal tumors, leukemia, and melanoma. KIT can be targeted by a range of inhibitors that predominantly bind to the inactive state of the enzyme. A mutation Y823D in the activation loop of KIT is known to be responsible for the loss of sensitivity to some drugs in metastatic tumors. We used all-atom molecular dynamics simulations to study the impact of Y823D on the KIT conformation and dynamics and compared it to the effect of phosphorylation of Y823. We simulated in total 6.4 μs of wild-type, mutant and phosphorylated KIT in the active- and inactive-state conformations. We found that Y823D affects the protein dynamics differently: in the active state, the mutation increases the protein stability, whereas in the inactive state it induces local destabilization, thus shifting the dynamic equilibrium towards the active state, altering the communication between distant regulatory regions. The observed dynamics of the Y823D mutant is similar to the dynamics of KIT phosphorylated at position Y823, thus we hypothesize that this mutation mimics a constitutively active kinase, which is not responsive to inhibitors that bind its inactive conformation.

KEYWORDS
cancer, KIT tyrosine kinase, molecular dynamics simulations, mutation Y823D, secondary resistance, signal transduction

1 | INTRODUCTION

The human genome contains 518 genes encoding kinases, of which 90 encode protein tyrosine kinases (TKs).[1,2] TKs are enzymes that phosphorylate a tyrosine residue of a target protein or phosphorylate their own tyrosines, leading to conformational changes and, typically, activation of downstream signaling cascades. Thus, TKs function as "on" or "off" switches in many cellular processes. The 90 TKs can be grouped into two classes: 58 of them belong to receptor tyrosine kinases (RTKs) and 32 of them are non-receptor tyrosine kinases (NRTKs). The major difference between these two classes is that the NRTKs, in contrast to RTKs, do not have the extracellular domain that is responsible for binding of extracellular ligand molecules. From here on, we will focus on RTKs. RTKs are cell surface receptors for several

growth factors, hormones and cytokines. Beside the ligand-binding extracellular domain, RTKs contain a single transmembrane helix separating the intracellular TK domain from the extracellular domain.

The RTK activation involves binding of a ligand to the RTK monomer leading to the dimerization of the receptors. Upon binding of a ligand to the extracellular domain the receptor undergoes extensive conformational changes that induce and stabilize the receptor dimerization leading to the stimulation of the kinase activity in the intracellular domain, thus resulting in auto-phosphorylation of tyrosine residues in it. The intracellular TK domain has a bi-lobar structure, with an ATP binding cleft located between the N- and C-terminal lobes that catalyzes the auto-phosphorylation and trans-phosphorylation of tyrosine residues and a kinase insert domain.[3] The N-lobe (residues 582-692) is composed of antiparallel β sheets adjacent to the α-helix (αC-helix) and the C-lobe (residues 763-935) shows predominantly the α-helical structure (Figure 1). The C-lobe contains the G-helix that binds the kinase substrate and an activation loop (A-loop) (residues 810-835) that begins with a highly conserved DFG (residues 810-812) motif. The major autophosphorylation sites are in the juxtamembrane region (JMR, residues 547-581) and in the kinase domain (KD, residues 582-935).[3] The phosphorylated tyrosine residues of the activated receptor now act as a binding site for proteins containing the Src homology 2 (SH2) and phosphotyrosine binding (PTB) domains.

In its unphosphorylated state, the intracellular TK domain exists in its inactive auto-inhibited conformation (Figure 1). As part of the activation process upon ligand binding to the extracellular domain, the A-loop adjacent to the active site in the inactive state switches from its auto-inhibitory position to a more open form. During this process, the DFG motif at the N-terminus of the A-loop flips its side chain away from the ATP binding site, thus allowing for binding of the ATP and Mg^{2+} cofactors. Following these structural changes, the JMR unwinds from its buried position in the TK domain to a solvent-exposed position, and the αC-helix undergoes orientational changes breaking the contacts with the JM-B fragment of the JMR. Along with JM-B, the JM-Z fragment is also involved in blocking the αC-helix, which regulates the catalytic activity of the kinases[4] and prevents the A-loop from adopting an active conformation, restricting the inter-lobe flexibility.

In 1984, the first connection between a viral oncogene, a mutated RTK and human cancer was established.[5] Since then it is well-established that aberrant signaling by RTKs is critically involved in human cancer. The profuse knowledge of the structure and activation mechanism of RTKs and the variations of TK signal transduction pathways in proliferative disorders led to the idea that tyrosine kinase inhibitors (TKIs) could have anticancer effects.[2,6] As a result, the development of target-specific TKIs and new anticancer drug discovery has become a hot area of anticancer research.

Most of the TKIs are small hydrophobic compounds that can rapidly reach their specific intracellular targets and inhibit the activation of the related TKs.[2,7,8] Unfortunately, the patients who gained remarkable benefit from the TKI therapy started showing increasing evidence of acquired resistance.[9-11] It is primarily due to the acquired secondary drug-resistance mutations that change the protein conformation and alter the drug binding site, leading to the therapy failure and cancer relapse. These mutations may alter the tightly controlled
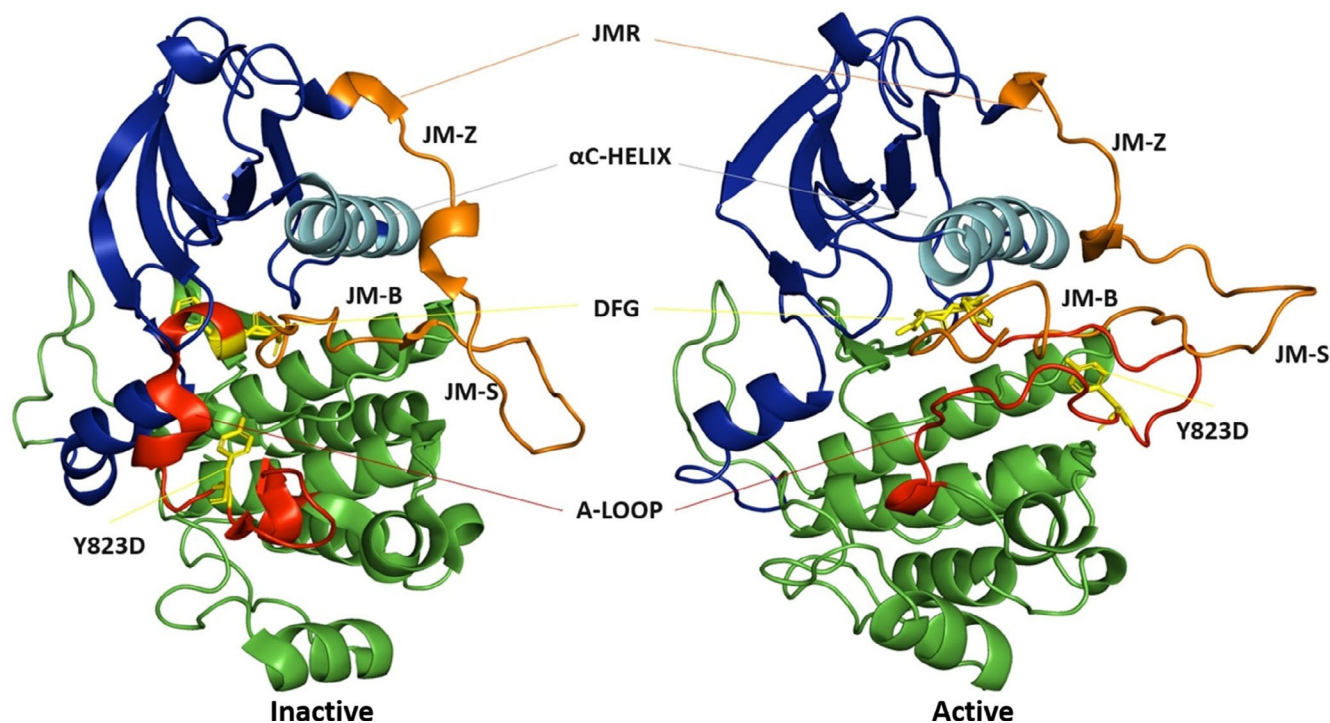

FIGURE 1 Structure of KIT cytoplasmic region. 3D structures of the native receptors in the inactive state (PDB id: 1T45) and in the active state (PDB id: 1PKG) are represented as cartoons. The N-lobe in blue, the C-lobe in green, the αC-helix in cyan, the A-loop in red, and the JMR in orange colors. The DFG motif and the mutation site Y823D are represented in sticks (yellow)

functions of the protein including ligand binding, conformational transitions, allosteric regulation, inducing resistance to several first and second line drugs.[2,3,12] However, the exact molecular mechanisms of the changes induced by such mutations are still not well understood.

KIT is a stem cell factor (SCF) receptor also known as proto-oncogene receptor tyrosine kinase, a member of the class III of RTKs. The activation of KIT proceeds similarly to other RTKs: a SCF molecule binds to a KIT monomer in the extracellular domain leading to dimerization, activation, and autophosphorylation of tyrosine residues in the KD.[13-16] During this process, the key regulatory regions in the TK domain undergo extensive conformational changes. A gain-of-function or an acquired secondary drug resistance mutation in the TK domain can activate the protein in the absence of the ligand. D816V is one of such mutations that has been studied extensively using the molecular dynamics (MD) simulations.[16-19] It has been shown that this mutation in the inactive state disrupts the communication between the A-loop and the JMR[17,18] and causes a partial folding of a conserved 3_{10} helices in the A-loop leading to the changes in the local H-bond network,[16] the global structural reorganization of the JMR, and constitutive activation of the protein.[20] KIT harboring this mutation is resistant towards several drugs including imatinib and sunitinib.[21-23] Y823D is also a similar gain-of-function or a secondary mutation which has been reported in several clinical cases.[24,25] Y823D can constitutively activate the PI3K/AKT, RAS-ERK, and JAK/STAT pathways leading to tumorigenesis by inducing cell proliferation, growth progression or migration.[21,26,27] This mutation is associated with various forms of cancer including gastrointestinal stromal tumors, testicular seminomas, melanogenesis, and in hematopoiesis.[20,21,28] Yet, very little is known about the impact of this mutation on the protein conformational changes crucial for activation and ligand binding.

It has been reported that Y823D can cause stabilization of the active conformation of the protein and that it is also responsible for loss of sensitivity to drugs in metastatic tumors.[29] This mutation is located in the A-loop of the KD and occurs at the only tyrosine residue in the A-loop. There exists literature evidence that this tyrosine residue in the A-loop is the last residue that undergoes autophosphorylation,[15] suggesting it might not play a role in the activation mechanism. However, it has been shown that this tyrosine residue is essential for the regulation of the kinase activity.[15,23]

In this present study, we explored the impact of the Y823D mutation and tyrosine phosphorylation on protein conformation using MD simulations. Specifically, we aimed to understand the differences in dynamics within and between the different states of the protein, which will enable us to identify the conformational changes that are crucial for the activation and ligand binding.

2 | MATERIALS AND METHODS

2.1 | Target selection

We retrieved the crystallographic structures of the wild-type (WT) inactive and active state of KIT protein from the Protein Data Bank (PDB id of inactive state 1T45 and active state 1PKG).[30] The water molecules in the crystal structures were retained for the MD simulations. The missing atoms in the crystal structure of the active state of the protein were added using MODELLER 9.17[31,32] and the phosphorylated tyrosine residues at position 568 and 570 were modeled back to their unphosphorylated state. In silico, substitution of Tyr (Y) to Asp (D) at position 823 was performed using MODELLER with the corresponding WT structures as the template for both active and inactive states. Tyr (Y) at position 823 was also phosphorylated in both monoanionic (TP1) and dianionic (TP2) states using CHARMM-GUI.[33-35] Generated models of the KIT, its MU Y823D, and two phosphorylated versions (TP1, TP2) are referred to as KIT-A^{WT}, KIT-A^{MU}, KIT-A^{TP1}, KIT-A^{TP2} and KIT-I^{WT}, KIT-I^{MU}, KIT-I^{TP1}, and KIT-I^{TP2} for active and inactive states, respectively.

2.2 | Preparation of the systems

MD simulations were performed using the GROMACS software package version 5.1.4[36] using two force fields, Amber99SB*-ILDN[37,38] and CHARMM36.[39] The molecular systems KIT-A^{WT} and KIT-A^{MU}, KIT-I^{WT} and KIT-I^{MU} were parameterized using the Amber99SB*-ILDN force field.[37,38] In a separate set, all generated models were parameterized using the CHARMM36 force field,[39] and these variants are further referred to as KIT-A*^{WT}, KIT-A*^{MU}, KIT-A*^{TP1}, KIT-A*^{TP2} and KIT-I*^{WT}, KIT-I*^{MU}, KIT-I*^{TP1}, KIT-I*^{TP2}. The molecules were centered in a cubic box with a 1.5 nm buffer, under periodic boundary conditions and the systems were explicitly solvated with TIP3P water molecules. Counterions[40] ClJ and NaJ for Amber99SB*-ILDN force field simulations and Cl and Na for CHARMM36 force field simulations were added when necessary to neutralize the overall charge (0.15 mol/L concentration). Energy minimization for each of the molecular systems was performed using the steepest descent algorithm. A maximum of 50 000 steps was performed until a maximum force of 1000.0 kJ mol^{-1} nm^{-1} was achieved. Following the energy minimization, each of the molecular systems was subjected to two consecutive steps of equilibration procedure. At first, each system was maintained at a temperature of 310 K during the NVT ensemble for 100 ps with a time step of 2 fs, followed by a 100 ps simulation in the NPT ensemble with a time step of 2 fs maintaining the pressure at 1 bar to equilibrate the system.

2.3 | Production of trajectories

After the NPT ensemble simulations, we performed a total of 64 100 ns-long production simulations with eight replicas for each of the systems parameterized by Amber99SB*-ILDN force field and four replicas for each of the systems parameterized by CHARMM36 force field. The coordinates were recorded at every 100 ps. The temperatures of solute and solvent were separately coupled to the velocity rescale thermostat (modified Berendsen thermostat)[41] at 310 K with a relaxation time of 0.1 ps. The pressure was maintained at 1 atm by

isotropic coordinate scaling with a relaxation time of 5 ps using Parrinello-Rahman barostat.[42] A time step of 2 fs was used to integrate the equations of motion based on the leap-frog algorithm.[43] Lennard-Jones interactions were set to a cut-off of 1.4 nm, and the Particle Mesh Ewald (PME) method[44] was used to treat long-range electrostatic interactions. All bonds were constrained using P-LINCS algorithm.[45]

2.4 | Analysis of the trajectories

The resulting trajectories from the simulations were analyzed using various tools including the tools from the GROMACS package. From the root mean square deviation (RMSD) profiles, the first 10 ns were considered to be a part of the equilibration process, therefore it was ignored, and the last 90 ns were retained for further analysis except for hydrogen bond (H-bond) calculations.

The secondary structure profiles were calculated using gmx do_dssp program available in GROMACS that uses the DSSP algorithm.[46] The calculation was performed over the concatenated trajectories of each system. The consensus secondary structure was defined as the type of secondary structure most prevalent at a given position over the whole simulation time (only α-helices and β-strands were considered) and visualized using Biotite.[47] The secondary structure elements such as 3_{10}-helix, pi-helix, bend, turn, bridge, and coil were jointly marked as coil.

To study the coupling between JMR and KD in KIT proteins, two characteristic distances were monitored over the MD simulations of each system. The distance d1 between the centroid of the JM-B region (residues 547-559 as C1) and the centroid of the residues in the N-lobe (residues 582-692 as C1') and the distance d2 between the centroid of the JM-S region (residues 560-570 as C2) and the centroid of the residues in the C-lobe (residues 763-935 as C2').

From the MD trajectories, the H-bonds between key residues were calculated using the program gmx hbond available in GROMACS. H-bonds were defined with a DHA angle cutoff of 120° and a donor-acceptor distance cutoff of 3.5 Å. The mean and standard error of the mean was calculated using R.[48]

For the molecular systems parameterized by the CHARMM36 force field, we performed analysis of RMSD, RMSF, secondary structure, H-bonds, and principal component analysis.

2.5 | Principal component analysis

Principal component analysis (PCA) is a method for reducing the dimensionality of a complex system. Before performing PCA, the concatenated trajectories from replica simulations were superimposed to minimize the variance over the ensemble.[49] The key idea of PCA is to identify the significant eigenvectors which define the dominant collective motions. The calculation was performed using the gmx covar module of GROMACS. The overlap between the first 10 modes of each trajectory was calculated using the gmx anaeig module of GROMACS. The perturbations of the systems can be described in terms of only a few principal components by ordering the eigenvalues of the diagonalized covariance matrix in a descending fashion. Thus, PCA helps to extract the large-scale motions from MD trajectories by isolating the dominant modes of internal motion.

2.6 | Mutual Information

Calculating mutual information (MI) between individual pairs of amino acids allows to identify whether the changes in the pairs of conformational distributions of these amino acids are correlated, linearly or nonlinearly. It is assumed that such correlated changes can cause perturbations to the energy landscape. Both backbone and side chain torsion angles (φ, ψ, and χ) were used to study the correlated motion between pairs of residues as they are assumed to contain the functionally relevant perturbations and conformational changes.[50] The MI between them was estimated using MutInf method[50] that allows us to capture significant concerted conformational changes of two residues as it focuses on torsion angles that are responsible for the low-frequency motions. We used a bin size of 24 degrees to obtain discrete distributions of the dihedral angles. The MI between pairs of residues are calculated as the individual sum of their entropies and subtracting it from their joint entropy over adaptive partitioning. Each MI value was then compared to the background distribution of all MI values for all pairs of amino acids in the trajectories of the WT and mutant structures. A P-value threshold of .01 was applied for filtering of significant correlations (for more details refer to the original article).[50]

To account for propagated error when comparing MI of WT and MU, bootstrap sets were created, following the procedure as described previously.[51] For this purpose for each dataset torsion angles were extracted from individual trajectories using the gmx g_chi command in GROMACS and sampled with replacement n frames from a simulation of length n. This procedure was repeated 10 times for each of the replicas while preserving the same order across the different dihedral angles and simulation runs. The mean and standard deviation, μ_{MIres} and sd_{MIres}, were then calculated for every pair of correlated residues MI_{res} from the 10 bootstrap sets. When comparing the MI in the wild-type and mutant complexes, MI_{resWT} and MI_{resMU}, only those residue pairs were retained, for which the following condition holds,

$$\left| \mu_{MI_{resMU}} - \mu_{MI_{resWT}} \right| > \sqrt{\frac{sd^2_{MI_{resMU}} + sd^2_{MI_{resWT}}}{N}} \qquad (1)$$

Where, N is the total number of independent replica simulations. To further eliminate the noise and weak interactions between correlated residues, a cutoff (MI >0.8 kT for the inactive state, and MI >1.7 kT for the active state) was chosen, so that pairs with a small absolute difference between mean values were disregarded. This step was necessary to select only the correlated residues that have largely different mean MI values when comparing the WT and MU simulations.

2.7 | Partial least-squares regression

Partial least-squares regression (PLS) was applied to identify the collective modes of internal dynamics associated with an external order parameter of functional interest using the functional mode analysis tool.[52] For this analysis, the coordinates of the backbone atoms, backbone atoms without JMR region and all protein atoms excluding hydrogen atoms were used as inputs to build the statistical models. The constants 0 and 1 corresponding to the trajectories in WT and MU simulations were used as the read-out variable to be estimated.

To validate the statistical models generated, we applied the following adapted k-fold cross-validation (CV) technique for WT and MU trajectories separately in the active and inactive states, again following the procedure described before.[51] The trajectories of WT and MU of the inactive state KIT protein were concatenated and superimposed to minimize the variance over the ensemble.[49] The resulting trajectory was divided into four equal parts. Three parts of the data with the labeled input containing equal parts from both WT and MU were then used in each iteration to train a model. Based on this we made predictions for the last part. The final number of PLS modes/components were chosen based on the Pearson correlation calculated between the actual and the predicted values, with a

compromise between the number of modes, the complexity and the quality of prediction using the "elbow method."[53,54]

3 | RESULTS

Structural features and internal dynamics of the KIT molecular systems were analyzed to investigate the differences between the conformations induced by the gain-of-function mutation Y823D and the two variants of phosphorylation of Y823.

3.1 | General observations

The Root Mean Square Deviations (RMSDs) were calculated for the backbone atoms and the key structural elements (αC-helix, JMR, and A-loop) (Figure 2A and Figure S1A) of the kinase domain (KD) to ascertain the time evolution of the structure. The backbone RMSD profiles of the inactive and active states show that they conform closely to their initial crystallographic structures. The RMSD mean values were in the range of 0.16 to 0.25 nm for inactive state and 0.36 to 0.52 nm for active state in trajectories resulting from simulations using both Amber99SB*-ILDN and CHARMM36 force fields.
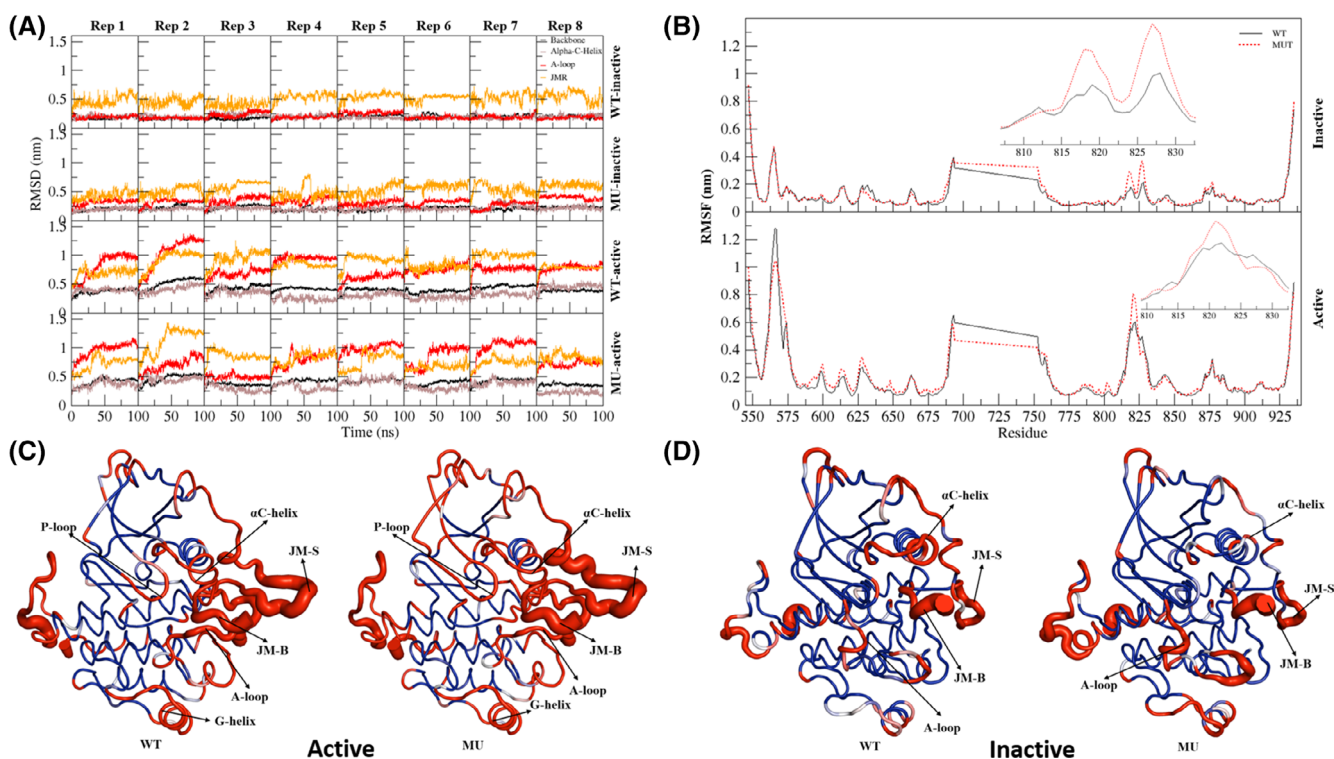
FIGURE 2 A, The RMSD values calculated for backbone atoms (black), αC-helix (brown), A-loop (red), and JMR (orange) from all eight trajectories of MD simulations of KIT[WT] and KIT[MU] in both the states (active and inactive). B, The RMSF values computed on the backbone atoms of the concatenated trajectory of MD simulations of KIT[WT] (black) and KIT[MU] (red). The RMSF values of the A-loop are given in the insert. The average conformation of the, C, active state KIT and D, inactive state KIT is represented as tubes. The highly flexible regions are shown in red, highly stable residues are shown in blue, and the intermediate flexible residues are shown in white. The size of the tube is proportional to the observed RMSF values in those regions, respectively

RMSD profiles show the conformational drift of the JMR in the inactive state of the protein is larger than the other KIT regions. In the active state, along with JMR, A-loop also was observed to have a large RMSD, whereas the αC-helix is observed to be rigid in both the states.

The protein flexibility was estimated by the root mean square fluctuations (RMSFs) averaged over time for every residue. The RMSF values of the backbone atoms (Figure 2B and Figure S1B) were comparable between different systems, ranging from 0.1 to 0.3 nm in the inactive state and 0.1 to 0.4 nm for the active state in trajectories resulting from simulations using both Amber99SB*-ILDN and CHARMM36 force fields. From the RMSF plot (Figure S1B) we can also observe how the MU and phosphotyrosines simulations converge, showing that the Y823D mutation mimics the presence of a phosphotyrosine. Both variants of the phosphotyrosine behave very similarly in the simulation. Further, we can also observe that the WT and MU simulations of both active and inactive states performed using two different force fields agree very well with each other (Figure S1B and S2).

3.2 | Secondary structure analysis

Upon analyzing the trajectories resulting from simulations using both Amber99SB*-ILDN and CHARMM36 force fields, we observe that the mutation does not significantly alter the secondary structure of JMR in both states of the protein (Figures S3 and S4) compared to its WT, but we observe slight changes in the secondary structure type during our simulations.

The mutation Y823D is located in the A-loop and induces local fluctuations in the inactive state of the protein. Unlike, other secondary mutations such as D816V/H/N/Y[17,20] in the A-loop, the mutation Y823D does not destabilize the 3_{10}-helix in the segment 817 to 819 in the trajectories resulting from Amber99SB*-ILDN force field (Figure S5), whereas in the active state almost all secondary structure in the A-loop is lost. The 3_{10}-helix is better retained in the MU simulation with the CHARMM36 force field, whereas in KIT-I*[TP1], KIT-I*[TP2] trajectories resulting from CHARMM36 force field (Figure S6), we observe very little secondary structure in the A-loop. The 3_{10}-helix in the A-loop is only observed in the inactive state of the protein and destabilization of this segment has been reported to be associated with the disruption of its integrity.[20] Such destabilization of key regions in the A-loop is associated with a major contribution to the loss of drug sensitivity.[20]

3.3 | Dynamic behavior of receptors

The effect of the mutation on the collective dynamic behavior of various KIT structural elements is markedly different in the active and inactive states. The computed scalar products between the first 10 PCA modes of KIT inactive state proteins indicate that the collective motions differ between them (Figure S7A). It is interesting to compare the dynamics of the WT and MU protein in each state.

When the trajectories of the KIT-I[WT] and KIT-I[MU] are projected into the subspace spanned by their first two PCs (Figure 3A), one can observe a large overlap of the WT and MU parts of the trajectories, and a considerable shift along the PC1 axis. Modes 1 and 2 of KIT-I[MU] display a slightly higher amplitude compared to KIT-I[WT] (Figure S7B), indicating increased flexibility of the protein upon mutation. These modes correspond to atomic motions (Figure 3B,C) of JMR coupled to deformations of the KD in the N-lobe, αC-helix, residues 626-631 preceding the αC-helix, orientational changes in the A-loop and the G-helix in the C-lobe. Principal modes of KIT-I[MU] describe atomic motions of JM-B coupled with a twist motion of the N-lobe and a displacement of the A-loop and G-helix in the C-lobe. This displacement/outward motion of the A-loop is not characteristic for the active conformation. The RMSF analysis along the individual modes (eigenvectors) reveals that the JMR, A-loop portions are dominant in the total backbone fluctuations along with kinase insert domain and minor perturbations of the αC-helix (data not shown).

Mapping KIT-A[WT] and KIT-A[MU] into the subspace spanned by the first two principal components of their concatenated trajectories shows that these trajectories occupy very different regions in it (Figure 3D). Analysis of the first eigenvectors of KIT in active state shows that there is a good agreement between the first principal mode of KIT-A[WT] and KIT-A[MU], but not so for the second mode (Figure S7C). Accordingly, the WT trajectory occupies a similar region as the MU trajectory along PC1, but differs much along PC2. The first eigenvalues for both WT and MU trajectories in the active state are much higher than those for the inactive state, indicating a generally higher flexibility of the protein in the active state (Figure S7D). In KIT-A[WT] the first mode is associated with displacement of JMR, A-loop, and unwinding twist motion of the N-lobe (Figure 3E,F), while the KIT-A[MU] experiences a conformational change in the JMR, its A-loop curtails from its elongated conformation, and an opposite twisted motion along an axis passing between the middle of the N- and C-lobe happens. The RMSF analysis along the first eigenvector shows that the JMR, kinase insert domain and the A-loop contribute the most to the backbone fluctuations in both KIT-A[WT] and KIT-A[MU] (data not shown).

PCA on the inactive state trajectories from CHARMM36 force field simulations confirm the similarity in the conformational distribution of the mutant and the phosphotyrosine simulations based on the projection onto the subspace defined by the first two PCs (PC1 and PC2; Figure S8).

3.4 | Coupling between JMR and KD in receptors

To check the possible coupling between JMR with KD in both states of the protein, their relative positions were characterized using two geometrical distances d1 and d2 (See Section 2). Monitoring of these distances over the course of the MD indicates that the d1 and d2 distributions (Table 1) in the inactive and active states of the KIT are very similar for both WT and MU. This demonstrates that the JMR is
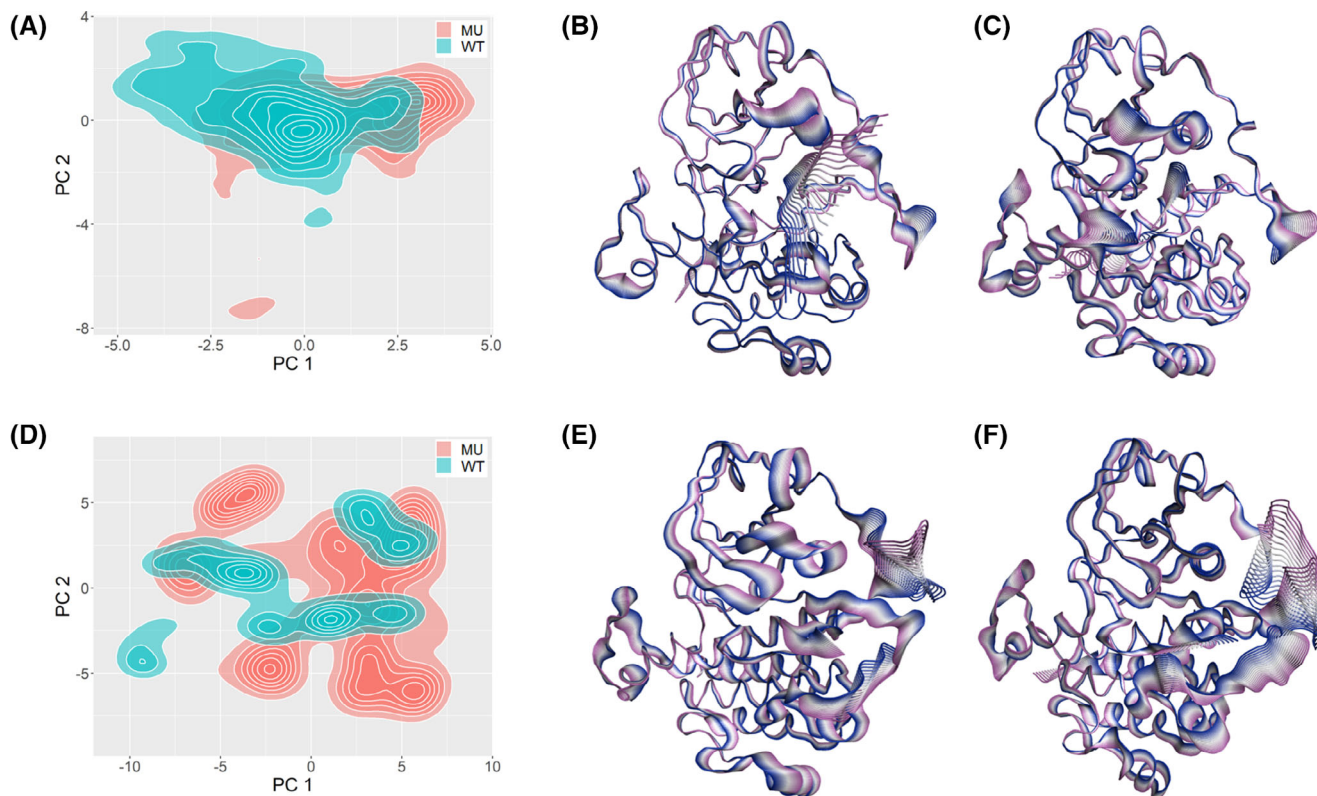
FIGURE 3 PCA of KIT inactive and active states. The atomic motions are shown as interpolation between the extremes along the first principal component of the, B,C, inactive state and, E,F, active state. The conformation distribution of different states (WT: cyan, MU: red) of KIT projected onto the subspace defined by the first two principal components (PC1 and PC2) of the concatenated WT and MU trajectories in the active, A, and in the inactive states, D

TABLE 1 Average distance between centroids of (d1) JM-B and N-lobe and (d2) JM-S and C-lobe

	KIT-I^{WT}	KIT-I^{MU}	KIT-A^{WT}	KIT-A^{MU}
d1	1.92 ± 0.01	1.92 ± 0.01	2.05 ± 0.04	2.08 ± 0.05
d2	2.3 ± 0.02	2.3 ± 0.02	3.05 ± 0.11	3.03 ± 0.09

Note: ± indicates standard error across independent simulations.

tightly coupled to the KD in both KIT conformations. The same can be observed in the PCA analysis we observed that the A-loop is responsible for the dominant fluctuations during the simulations of the active state (Figures 3E-F). Apart from these fluctuations, there are no major conformational changes observed and the computed d2 distance deviation between KIT-A^{WT} and KIT-A^{MU} is also negligible.

Frequency of H-bonds between key residues that maintain the conformation of the KIT protein and the residues that are known to be involved in establishing the allosteric communication between JMR and KD, such as H-bond between D792 and Y823[16,17,20,56] was measured (Tables 2 and 3). In KIT-I^{WT}, JMR binds to the C-lobe of KD through stable H-bonds V560...N787, K558...I789, Y568...F848 and Y570...Y846. These H-bonds are preserved in the mutated and phosphorylated structures of KIT indicating a strong coupling and stabilization of the JMR attachment to KD and of the internal JMR contacts. These data are consistent with the data from the coupling analysis between JMR and KD.

The signal from A-loop to the JMR is propagated through the C-loop, with Y823 as a key intermediate residue.[18] The corresponding H-bonds are affected upon the Y823D mutation and phosphorylation. The H-bond between residues Y823 and D792 is completely lost in KIT-I^{MU}, KIT-I^{*MU}, KIT-I^{*TP1}, and KIT-I^{*TP2} due to the changed properties of the residue at position 823. The interaction between D792 and N797 is also observed less frequently. However, the H-bond between D792 and H790 is not affected. Such modification in the local interaction network of the C-loop has been shown to be accompanied by a decreased communication to the distant KIT regulatory elements.[18] These observations suggest that the allosteric communication in the KIT inactive state between the A-loop and JMR is disrupted by the mutation Y823D in the A-loop. A beta-turn motif (residues 820-823) supported by a H-bond D820...N822 also disappears in the mutated and phosphorylated structures. It is known that this beta-turn motif is stabilized by the H-bond D820...N822 in the protein inactive state, and disruption of this H-bond results in the unfolding of the beta-turn motif.[18] Importantly, we also observe different χ angle distributions of D810 and F811 of the DFG motif between the KIT-I^{WT} and KIT-I^{MU} (Figure S10).

TABLE 2 Average H-bond number between residues in Amber99SB*-ILDN force field simulations

JMR...C-lobe	KIT-I^{WT}	KIT-I^{MU}	KIT-A^{WT}	KIT-A^{MU}
V560...N787	1.067 ± 0.01	1.08 ± 0.01	0.001 ± 0	0.008 ± 0.01
K558...I789	1.80 ± 0.03	1.87 ± 0.03	0	0.25 ± 0.25
Y568...F848	0.29 ± 0.6	0.35 ± 0.12	0.04 ± 0.04	0
Y570...Y846	1.23 ± 0.23	1.09 ± 0.24	0	0
E640...F811	0	0	0.71 ± 0.14	0.55 ± 0.13
JMS				
V569...E561	1.02 ± 0.06	1.21 ± 0.16	0	0
A-loop				
V824...L831	2.022 ± 0.01	1.9 ± 0.06	0	0
R815...Y823	0.89 ± 0.22	0.09 ± 0.04	0.32 ± 0.16	0.733 ± 0.3
D820...N822	2.68 ± 0.08	1.65 ± 0.21	0.97 ± 0.22	0.98 ± 0.23
A-loop...C-lobe				
Y823...R796	1.254 ± 0.23	1.1 ± 0.46	0	0
Y823...Y846	0	0	0.22 ± 0.08	0.42 ± 0.2
C-loop				
D792...H790	1.931 ± 0.01	1.93 ± 0.02	1.53 ± 0.12	1.7 ± 0.07
H790...N797	0	0.13 ± 0.1	0.15 ± 0.1	0.16 ± 0.1
D792...N797	2.088 ± 0.1	1.83 ± 0.18	1.4 ± 0.16	1.53 ± 0.22
C-loop...A-loop				
D792...Y823	1.20 ± 0.19	0	0	0
C-loop...C-lobe				
N797...C809	0.56 ± 0.12	0.87 ± 0.15	0.85 ± 0.14	0.55 ± 0.17
C-lobe				
L799...K807	2 ± 0	2 ± 0	2 ± 0	2 ± 0

Note: ± indicates standard error across independent simulations.

In the KIT active state the residues D792, H790, and N797 of the C-loop (catalytic loop) are involved in the formation of a local H-bond interaction network that is stable along the MD trajectories. Further, the H-bond analysis shows a stabilization of JMR, A-loop and C-loop suggesting that the mutation Y823D and the phosphorylation of Y823 cause stabilization of the active state. In particular, the H-bonds R815...Y(D)823 and Y(D)823...Y846 are observed more frequently in all simulations for the mutated and phosphorylated KIT in the active state. These H-bonds are located either in the A-loop or connect the A-loop and the C-lobe, probably stabilizing their mutual orientation. On the contrary, in the mutated and phosphorylated protein in the inactive state, the interactions of JMR with the KD is preserved, while the A-loop is displaced and destabilized.

3.5 | Changes in the correlated motions of amino acid residues

The correlation between the motions of all pairs of residues emerges from the various ways, in which they interact with one another. Here, we investigated how the mutation Y823D alters the concerted motion between residues, which in turn can affect the allosteric communication between distant regulatory regions of KIT. To do so, we computed the MI between trajectories of individual residues. MI reflects the degree to which two random variables are linked; high MI indicates a low uncertainty in one random variable given the information about the other. Interestingly, in both active and inactive states, most of the correlations are between distant residues (Figure 4A,D and Figure S9). We can observe that there are only a few residues in the mutated protein in the active state showing changes in the correlated motion and that almost all the residues involved in the interactions are between 8 and 21 Å apart from one another. Additionally, in the inactive state, we see that the changes in the pattern of correlated motions are much stronger in the mutant protein when compared to the WT.

Further, analysis of the correlated motion of residues in the inactive state and the active state revealed that certain residues act as hubs connected to several other residues (Figure 4B,E). However, the residues acting as hubs are not the same for active and inactive states, indicating that the correlated motions are different in different conformations of the same protein and that mutation Y823D affects these conformations differently. In the inactive state, the residues N797 and R888 act as hubs interacting with several key residues (D810, R815, R830, and M836) of the protein (Figure 4B), suggesting

TABLE 3 Average H-bond number between residues in CHARMM36 force field simulations

JMR...C-lobe	KIT-I*WT	KIT-I*MU	KIT-I*TP1	KIT-I*TP2	KIT-A*WT	KIT-A*MU	KIT-A*TP1	KIT-A*TP2
V560...N787	1.03 ± 0	1.11 ± 0.02	1.1 ± 0.05	1.1 ± 0.02	0	0	0.02 ± 0.02	0.05 ± 0.04
K558...I789	1.54 ± 0.07	2 ± 0.16	1.65 ± 0.13	1.34 ± 0.13	0	0.01 ± 0.01	0	0
Y568...F848	0.23 ± 0.09	0.59 ± 0.14	0.59 ± .028	0.74 ± 0.27	0	0	0	0
Y570...Y846	0.3 ± 0.12	0.16 ± 0.05	0.32 ± 0.06	0.06 ± 0.01	0	0	0	0
E640...F811	0	0	0	0	0.85 ± 0.14	0.73 ± 0.2	0.57 ± 0.13	0.84 ± 0.08
JMS								
V569...E561	1.13 ± 0.28	1.11 ± 0.09	1.11 ± 0.05	1.17 ± 0.14	0	0	0	0
A-loop								
V824...L831	2.02 ± 0.01	1.33 ± 0.08	1.94 ± 0.06	1.64 ± 0.22	0	0	0	0
R815...Y823	0.79 ± 0.39	0	1.14 ± 0.35	0.89 ± 0.46	0.05 ± 0.03	0.16 ± 0.04	0.31 ± 0.17	2.09 ± 0.66
D820...N822	2.24 ± 0.05	0.99 ± 0.22	0.79 ± 0.14	0.36 ± 0.09	0.34 ± 0.09	0.33 ± 0.08	0.49 ± 0.24	0.23 ± 0.08
A-loop...C-lobe								
Y823...R796	1.16 ± 0.11	0.57 ± 0.27	2.11 ± 0.54	2.87 ± 0.05	0	0	0	0
Y823...Y846	0	0	0	0	0.03 ± 0.02	0.1 ± 0.05	0.83 ± 0.45	0.08 ± 0.05
C-loop								
D792...H790	1.51 ± 0.02	1.35 ± 0.17	1.65 ± 0.1	1.32 ± 0.1	1.35 ± 0.13	1.11 ± 0.15	1.26 ± 0.15	1.53 ± 0.1
H790...N797	0	0	0	0	0.06 ± 0.06	0.18 ± 0.18	0.2 ± 0.11	0.03 ± 0.02
D792...N797	2.01 ± 0.08	1.71 ± 0.09	1.72 ± 0.08	1.73 ± 0.05	1.39 ± 0.22	1.29 ± 0.26	0.95 ± 0.2	1.49 ± 0.14
C-loop...A-loop								
D792...Y823	1.3 ± 0.1	0	0.57 ± 0.25	0	0	0	0	0
C-loop...C-lobe								
N797...C809	0.34 ± 0.04	0.33 ± 0.16	0.52 ± 0.12	0.2 ± 0.08	0.45 ± 0.11	0.61 ± 0.15	0.59 ± 0.35	0.4 ± 0.17
C-lobe								
L799...K807	2 ± 0	2 ± 0	2 ± 0	2 ± 0	2 ± 0	2 ± 0	1.99 ± 0.01	1.99 ± 0.01

Note: ± indicates standard error across independent simulations.

that any change in the motion of these residues will affect the correlated motion of the other residues. Similarly, in the active state, the residues E554 and the highly conserved residue R815 act as hubs and we also see that overall more residue pairs are involved in correlated motions in the WT than mutant protein (Figure 4E). Among the residues whose motion was most affected by the mutation in the inactive state are the residues of the inhibitor binding site and the conserved C-loop residues (residues 568, 797, 810, and 815), P+1 loop (the loop immediately following the A-loop, residues 830 and 836), critical residue of the DFG motif (residue 810) and F-helix (residue 861).

In KIT-I[WT], the signal from A-loop to the JMR is propagated through the C-loop residues, where Y823 acts as an intermediate residue.[18] From MI data, we can observe how mutation in this position has changed the communication between distant regulatory regions in the protein. Thus, the modification in the local interaction network of the C-loop that we observed in the H-bond analysis is accompanied by a reduction in the efficient communication to the distant KIT regulatory elements. In the active state of KIT we notice a different pattern, where none of the correlated motions below the threshold (1.7 kT) involving the key residues were affected by the mutation (Figure 5B).

3.6 | Partial least-squares regression

The functional mode analysis based on the partial least-squares (PLS) regression method aids in differentiating the major collective motions between the WT and MU proteins. As input, we used the coordinates of the backbone atoms as they proved to have the most predictive power when creating the statistical models for the inactive state (data for other input types not shown). For the active state, none of the input features were able to provide a satisfactory model that could be used to differentiate the collective motions between the WT and mutant proteins. Therefore, only the results for the inactive state of the protein are discussed.

We used fourfold cross validation (CV) technique to obtain the statistical models and measured the prediction quality through Pearson correlation score between the true and predicted label for a given input feature. We chose models composed of three PLS components to study the changes in the collective motions of the protein, which was sufficient in the CV scenario to achieve a correlation of .75 (Figure S11). Mapping the PLS results onto the backbone of the inactive state of KIT (Figure 5) showed that the mutation induced changes in the collective motions in the JM-B,
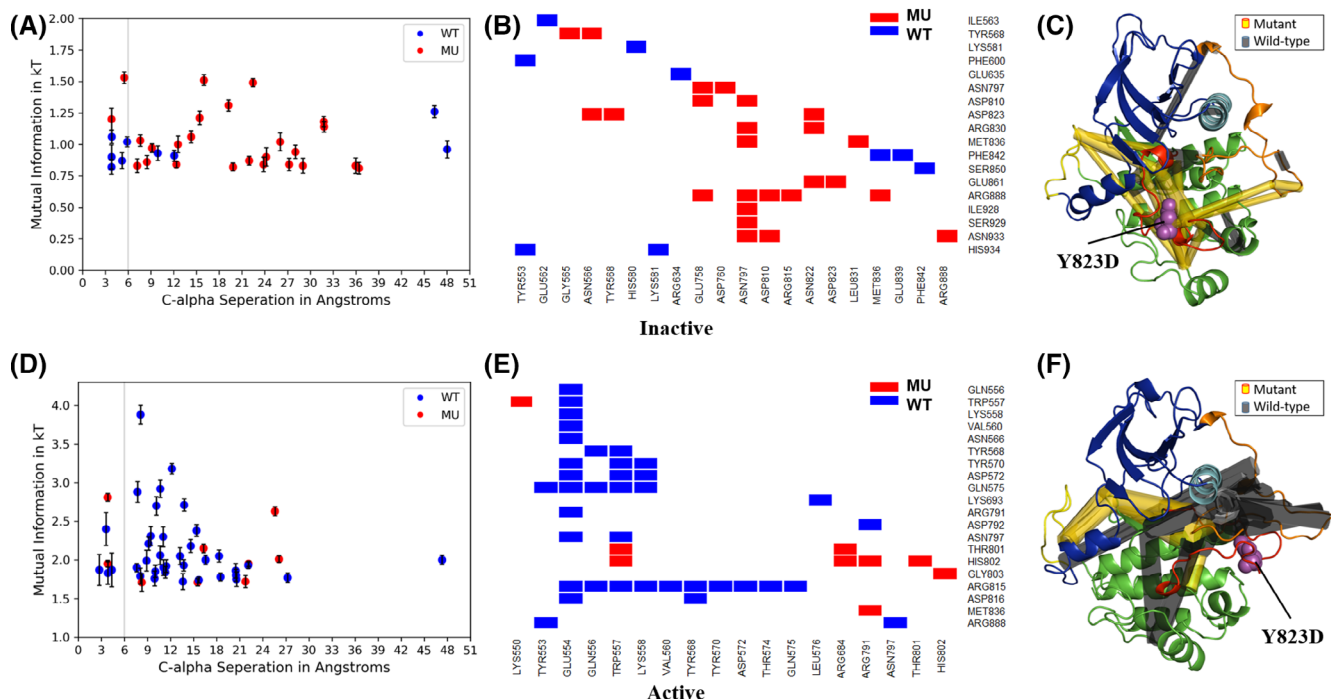
FIGURE 4 A,D, Distance between correlated residues (MI greater than 0.8 kT in the inactive state and greater than 1.7 kT in the active state). A gray line separates local (<6 Å) and long-range interactions. SD from the bootstrapped dataset is plotted as the error bars on top of the MI values from the actual (not bootstrapped) simulations. B,E, Correlations with MI greater than 0.8 kT for the inactive state and MI greater than 1.7 kT for active state are only shown. C,F, Cylinders connecting residues represent differences in MI between those residues in MU (yellow) and WT (gray), with the width of the cylinder proportional to the MI. Mutation site Y823D is represented as spheres in magenta
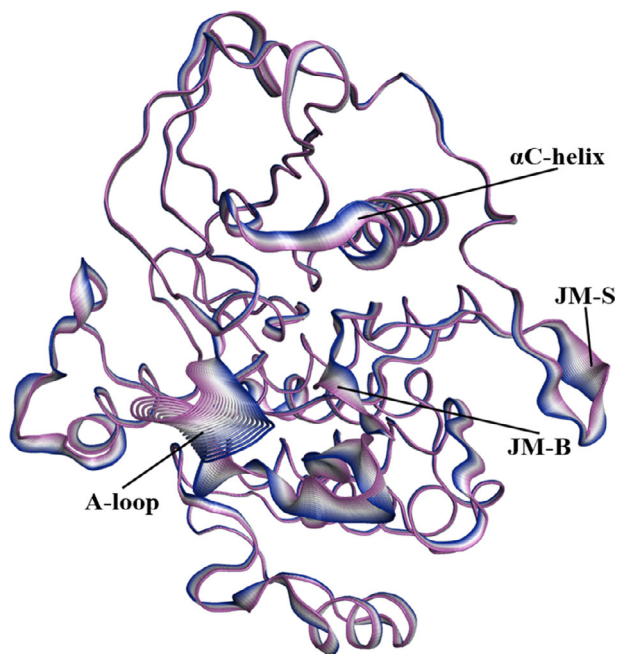

FIGURE 5 Interpolation between the extremes of the inactive state PLS models. Blue-to-magenta bands correspond to the interpolation along the mode that relates to the true label of simulation, WT and MU, to the underlying differences in the protein motions

JM-S, A-loop, and αC-helix. In agreement with this, we observe the same regions to undergo large shifts in the PCA analysis (see above).

4 | DISCUSSION

This study presents a first application of MD simulation aimed at understanding the effects of the mutation Y823D and phosphorylation of Y823, which confer resistance among others to imatinib and sunitinib[29,55-57] and is a secondary mutation in gastrointestinal stromal tumors,[29,57,58] both in the active and inactive states of KIT. Phosphorylation of Y823 is known to stabilize the active conformation of the A-loop most probably by strong electrostatic interactions of the phosphate group.[24,25] Mutations such as Y823D, which mimic Y823 phosphorylation, were suggested to stabilize the active conformation of the A-loop,[24-26] which was confirmed in our analysis. In the active state of the protein we observe that this mutation as well as phosphorylation of the tyrosine stabilizes the protein's active conformation by strengthening the H-bonds, particularly the one between R815 and Y823 which is known to stabilize the active conformation of the protein.[59] In line with our observations, mutations in the A-loop are known to disrupt the inactive conformation by introducing charged side chains into the pocket.[26,56,57]

The different analyses of the trajectories that we performed for both active and inactive conformations of the wild-type, mutant and phosphorylated KIT all demonstrate different dynamics in the same regions of the structure: the active site residues and other key residues in the A- and C-loops (Figure S12). These differences are prominent for the inactive state of KIT, but are much less significant for the active state. The dynamics of the mutant protein in all cases resembles the dynamics of the kinase phosphorylated at the mutation site

Y823. This can probably be explained by the fact that the substitution to a negatively charged aspartate in the position of Y823 mimics the effect of the phosphorylation at position Y823.

The biological significance of these structural regions was previously reported in the literature.[18,60-62] In particular, the C-loop is highly conserved in terms of structure and sequence among all kinases.[60] Another structural element, the F-helix, whose motion was also affected in our analysis, was shown to play an important role in protein kinase active structures, as it is considered as the central hub connecting key areas such as the substrate binding residues and the C-loop. Also the regulatory and catalytic spines are located at the N and C termini of the F-helix.[18,61,62] This puts the perturbations observed through RMSF analysis in the JM-B, JM-S, A-loop, and the residues preceding the αC-helix in the inactive state and the local destabilization of the A-loop through loss of H-bonds into a proper functional context. Most importantly, we observe the loss of allosteric communication, as evident from the changes in the hydrogen bond network reported above, between distant regulatory domains namely, JMR and the A-loop, which may be detrimental for kinase regulation. As observed from this analysis, the mutation Y823D and the phosphorylation of Y823 stabilize the active conformation of the protein by contributing to the formation of more stable H-bonds in the A- and C-loops.

The concerted alterations that we observe in all simulations of the inactive state may potentially indicate mechanisms that could affect the binding of drugs. Multiple lines of evidence presented here suggest that the changes in the structural conformation of the inactive state protein correspond to its destabilization upon mutation as well as phosphorylation and a shift of the dynamic equilibrium away from the inactive state, which is a primary target for many drugs. For example, imatinib specifically binds to the inactive state of KIT,[59,63] thus these changes in the key regions of the inactive state of the protein may alter the binding and/or the sensitivity to it and other drugs that specifically targets the inactive state of KIT.

5 | CONCLUSIONS

In our simulations, we found that the mutation Y823D as well as phosphorylation at this position in the A-loop of the protein TK KIT compromise the communication between key regulatory regions of the inactive state protein and induces local destabilization due to the loss and reduction of key H-bonds. Further, we also observed changes in the correlated motion of amino acids, especially between the mutation site and the active site residues and other key residues such as D792, R815, and Y823[16,20,59] in the A-loop and catalytic loop that are involved in drug binding and allosteric communication.

In summary, we collected multiple pieces of evidence demonstrating that the mutation Y823D shifts the conformational equilibrium towards the active state. This is likely to lead to activating the downstream signaling cascades and enhancing the expression of anti-apoptotic, cell proliferation, and growth expression genes. It is worth noting that Y823 in KIT is homologous to Y393 in the Abl kinase, where phosphorylation of this residue is known to destabilize the inactive conformation of the A-loop.[24] In our simulations, we confirm a similar trend for the mutation Y823D in KIT by conducting simulations with a phosphorylated tyrosine at this position. Thus, we suggest that an analogous mechanism of resistance evolved in a homologous kinase leading to similar pathologic consequences.

CONFLICT OF INTERESTS

TB is an employee of AstraZeneca and has stock ownership and/or stock options or interests in the company.

ORCID

Sanjay K. Srikakulam https://orcid.org/0000-0002-1752-5060
Olga V. Kalinina https://orcid.org/0000-0002-9445-477X

REFERENCES

1. Hubbard SR, Miller WT. Receptor tyrosine kinases: mechanisms of activation and signaling. *Curr Opin Cell Biol*. 2007;19(2):117-123. https://doi.org/10.1016/j.ceb.2007.02.010.
2. Chen Y, Fu L. Mechanisms of acquired resistance to tyrosine kinase inhibitors. *Acta Pharm Sin B*. 2011;1(4):197-207. https://doi.org/10.1016/j.apsb.2011.10.007.
3. Hubbard SR, Till JH. Protein tyrosine kinase structure and function. www.annualreviews.org. 2000
4. Li S, Covino ND, Stein EG, Till JH, Hubbard SR. Structural and biochemical evidence for an autoinhibitory role for tyrosine 984 in the juxtamembrane region of the insulin receptor. *J Biol Chem*. 2003;278 (28):26007-26014. https://doi.org/10.1074/JBC.M302425200.
5. Ullrich A, Coussens L, Hayflick JS, et al. Human epidermal growth factor receptor cDNA sequence and aberrant expression of the amplified gene in A431 epidermoid carcinoma cells. *Nature*. 1984;309:418-425. https://doi.org/10.1038/309418a0.
6. Arora A, Scholar EM. Role of tyrosine kinase inhibitors in cancer therapy. *J Pharmacol Exp Ther*. 2005;315(3):971-979. https://doi.org/10.1124/jpet.105.084145.
7. Sekhon N, Kumbla RA, Mita M. Chapter 1 - current trends in cancer therapy. In: Gottlieb RA, Mehta PK, eds. *Cardio-Oncology*. Boston: Academic Press; 2017:1-24. https://doi.org/10.1016/B978-0-12-803547-4.00001-X.
8. Broekman F, Giovannetti E, Peters GJ. Tyrosine kinase inhibitors: multi-targeted or single-targeted? *World J Clin Oncol*. 2011;2(2):80-93. https://doi.org/10.5306/wjco.v2.i2.80.
9. Rosenzweig SA. Acquired resistance to drugs targeting tyrosine kinases. *Adv Cancer Res*. 2018;138:71-98. https://doi.org/10.1016/bs.acr.2018.02.003.
10. Lovly CM, Shaw AT. Molecular pathways: resistance to kinase inhibitors and implications for therapeutic strategies. *Clin Cancer Res off J Am Assoc Cancer Res*. 2014;20(9):2249-2256. https://doi.org/10.1158/1078-0432.CCR-13-1610.
11. Engelman JA, Settleman J. Acquired resistance to tyrosine kinase inhibitors during cancer therapy. *Curr Opin Genet Dev*. 2008;18(1):73-79. https://doi.org/10.1016/j.gde.2008.01.004.
12. Barouch-Bentov R, Sauer K. Mechanisms of drug-resistance in kinases. *Expert Opin Investig Drugs*. 2011;20(2):153-208. https://doi.org/10.1517/13543784.2011.546344.

13. Roskoski R. Structure and regulation of KIT protein-tyrosine kinase—the stem cell factor receptor. *Biochem Biophys Res Commun*. 2005; 338(3):1307-1315. https://doi.org/10.1016/j.bbrc.2005.09.150.

14. Shi X, Sousa LP, Mandel-Bausch EM, et al. Distinct cellular properties of oncogenic KIT receptor tyrosine kinase mutants enable alternative courses of cancer cell inhibition. *Proc Natl Acad Sci*. 2016;113(33): E4784-E4793. https://doi.org/10.1073/pnas.1610179113.

15. Mol CD, Lim KB, Sridhar V, et al. Structure of a c-kit product complex reveals the basis for kinase transactivation. *J Biol Chem*. 2003;278 (34):31461-31464. https://doi.org/10.1074/jbc.C300186200.

16. Da P, Figueiredo S, Gomes C, et al. Differential effects of CSF-1R D802V and KIT D816V homologous mutations on receptor tertiary structure and allosteric communication. *PLOS One*. 2014;9(5):e97519. https://doi.org/10.1371/journal.pone.0097519.

17. Laine E, Chauvot de Beauchêne I, Perahia D, Auclair C, Tchertanov L. Mutation D816V alters the internal structure and dynamics of c-KIT receptor cytoplasmic region: implications for dimerization and activation mechanisms. Verkhivker GM. *PLoS Comput Biol*. 2011;7(6): e1002068. https://doi.org/10.1371/journal.pcbi.1002068.

18. Laine E, Auclair C, Tchertanov L. Allosteric communication across the native and mutated KIT receptor tyrosine kinase. Nussinov R. *PLoS Comput Biol*. 2012;8(8):e1002661. https://doi.org/10.1371/journal. pcbi.1002661.

19. Da Silva Figueiredo Celestino Gomes P, Chauvot De Beauchêne I, Panel N, et al. Insight on mutation-induced resistance from molecular dynamics simulations of the native and mutated CSF-1R and KIT. *PLoS ONE*. 2016;11(7):e0160165. https://doi.org/10.1371/journal. pone.0160165.

20. Chauvot de Beauchêne I, Allain A, Panel N, et al. Hotspot mutations in KIT receptor differentially modulate its allosterically coupled conformational dynamics: impact on activation and drug sensitivity. Nilges M, ed. *PLoS Comput Biol*. 2014;10(7):e1003749. https://doi. org/10.1371/journal.pcbi.1003749.

21. Abbaspour Babaei M, Kamalidehghan B, Saleem M, Huri HZ, Ahmadipour F. Receptor tyrosine kinase (c-kit) inhibitors: a potential therapeutic target in cancer cells. *Drug Des Devel Ther*. 2016;10: 2443-2459. https://doi.org/10.2147/DDDT.S89114.

22. Caenepeel S, Renshaw-Gegg L, Baher A, et al. Motesanib inhibits kit mutations associated with gastrointestinal stromal tumors. *J Exp Clin Cancer Res CR*. 2010;29(1):96. https://doi.org/10.1186/1756-9966-29-96.

23. Agarwal S, Kazi JU, Mohlin S, Pa ahlman S, Rönnstrand L. The activation loop tyrosine 823 is essential for the transforming capacity of the c-kit oncogenic mutant D816V. *Oncogene*. 2015;34(35):4581-4590. https://doi.org/10.1038/onc.2014.383.

24. Daub H, Specht K, Ullrich A. Strategies to overcome resistance to targeted protein kinase inhibitors. *Nat Rev Drug Discov*. 2004;3(12): 1001-1010. https://doi.org/10.1038/nrd1579.

25. Wang CM, Fu H, Zhao GF, et al. Secondary resistance to imatinib in patients with gastrointestinal stromal tumors through an acquired KIT exon 17 mutation. *Mol Med Rep*. 2009;2(3):455-460. https://doi.org/ 10.3892/mmr_00000121.

26. Agarwal S, Kazi JU, Rönnstrand L. Phosphorylation of the activation loop tyrosine 823 in c-kit is crucial for cell survival and proliferation. *J Biol Chem*. 2013;288(31):22460-22468. https://doi.org/10.1074/ jbc.M113.474072.

27. Maleddu A, Pantaleo MA, Nannini M, Saponara M, Lolli C, Biasco G. Mechanisms of secondary resistance to tyrosine kinase inhibitors in gastrointestinal stromal tumours (review). *Oncol Rep*. 2009;21(6): 1359-1366. https://doi.org/10.3892/or_00000361.

28. Debiec-Rychter M, Cools J, Dumez H, et al. Mechanisms of resistance to imatinib mesylate in gastrointestinal stromal tumors and activity of the PKC412 inhibitor against imatinib-resistant mutants. *Gastroenterology*. 2005;128(2):270-279. https://doi.org/10.1053/J.GASTRO. 2004.11.020.

29. Wakai T, Kanda T, Hirota S, Ohashi A, Shirai Y, Hatakeyama K. Late resistance to imatinib therapy in a metastatic gastrointestinal stromal tumour is associated with a second KIT mutation. *Br J Cancer*. 2004; 90(11):2059-2061. https://doi.org/10.1038/sj.bjc.6601819.

30. Berman HM, Westbrook J, Feng Z, et al.. The Protein Data Bank. http://www.rcsb.org/pdb/status.html. 2000.

31. Eswar N, Eramian D, Webb B, Shen M-Y, Sali A. Protein structure modeling with MODELLER. *Methods Mol Biol*. 2008;426:145-159. https://doi.org/10.1007/978-1-60327-058-8_8.

32. Martí-Renom MA, Stuart AC, Fiser A, Sánchez R, Melo F, Andrejšali AA. Comparative protein structure modeling of genes and genomes. www.annualreviews.org. 2000.

33. Jo S, Kim T, Iyer VG, Im W. CHARMM-GUI: a web-based graphical user interface for CHARMM. *J Comput Chem*. 2008;29(11):1859-1865. https://doi.org/10.1002/jcc.20945.

34. Jo S, Cheng X, Islam SM, et al. CHARMM-GUI PDB manipulator for advanced modeling and simulations of proteins containing nonstandard residues. *Adv Protein Chem Struct Biol*. 2014;96:235-265. https://doi.org/10.1016/bs.apcsb.2014.06.002.

35. Lee J, Cheng X, Swails JM, et al. CHARMM-GUI input generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM simulations using the CHARMM36 additive force field. *J Chem Theory Comput*. 2016;12(1):405-413. https://doi.org/10.1021/acs.jctc. 5b00935.

36. van der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJC. GROMACS: fast, flexible, and free. *J Comput Chem*. 2005;26:1701-1718. https://doi.org/10.1002/jcc.20291.

37. Hornak V, Abel R, Okur A, Strockbine B, Roitberg A, Simmerling C. Comparison of multiple AMBER force fields and development of improved protein backbone parameters. *Proteins*. 2006;65:712-725. https://doi.org/10.1002/prot.21123.

38. Martín-García F, Papaleo E, Gomez-Puertas P, Boomsma W, Lindorff-Larsen K. Comparing molecular dynamics force fields in the essential subspace. Roccatano D. *PLOS One*. 2015;10(3):e0121114. https:// doi.org/10.1371/journal.pone.0121114.

39. Huang J, Rauscher S, Nawrocki G, et al. CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nat Methods*. 2017;14(1):71-73. https://doi.org/10.1038/nmeth. 4067.

40. Joung IS, Cheatham TE. Determination of alkali and halide monovalent ion parameters for ese in explicitly solvated biomolecular simulations. *J Phys Chem B*. 2008;112(30):9020-9041. https://doi.org/10. 1021/jp8001614.

41. Bussi G, Donadio D, Parrinello M. Canonical sampling through velocity rescaling. *J Chem Phys*. 2007;126(1):014101. https://doi.org/10. 1063/1.2408420.

42. Parrinello M, Rahman A. Polymorphic transitions in single crystals: a new molecular dynamics method. *J Appl Phys*. 1981;52(12):7182-7190. https://doi.org/10.1063/1.328693.

43. Van Gunsteren WF, Berendsen HJC. A leap-frog algorithm for stochastic dynamics. *Mol Simul*. 1988;1:173-185. https://doi.org/10. 1080/08927028808080941.

44. Darden T, York D, Pedersen L. Particle mesh Ewald: an N·log (N) method for Ewald sums in large systems. *J Chem Phys*. 1993;98 (12):10089-10092. https://doi.org/10.1063/1.464397.

45. Hess B. P-LINCS: a parallel linear constraint solver for molecular simulation. *J Chem Theory Comput*. 2008 Jan;4(1):116-122. https://doi. org/10.1021/ct700200b.

46. Kabsch W, Sander C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*. 1983;22(12):2577-2637. https://doi.org/10.1002/bip. 360221211.

47. Kunzmann P, Hamacher K. Biotite: a unifying open source computational biology framework in python. *BMC Bioinform*. 2018;19(1):346. https://doi.org/10.1186/s12859-018-2367-z.

48. R Core Team. *R: a language and environment for statistical computing*. Vienna, Austria: R foundation for statistical computing; 2017 https://www.R-project.org/.

49. Gapsys V, de Groot BL. Optimal superpositioning of flexible molecule ensembles. *Biophys J*. 2013;104(1):196-207. https://doi.org/10.1016/j.bpj.2012.11.003.

50. McClendon CL, Friedland G, Mobley DL, Amirkhani H, Jacobson MP. Quantifying correlations between allosteric sites in thermodynamic ensembles. *J Chem Theory Comput*. 2009;5(9):2486-2502. https://doi.org/10.1021/ct9001812.

51. Bastys T, Gapsys V, Doncheva NT, Kaiser R, de Groot BL, Kalinina OV. Consistent prediction of mutation effect on drug binding in HIV-1 protease using alchemical calculations. *J Chem Theory Comput*. 2018;14(7):3397-3408. https://doi.org/10.1021/acs.jctc.7b01109.

52. Krivobokova T, Briones R, Hub JS, Munk A, De Groot BL. Partial least-squares functional mode analysis: application to the membrane proteins AQP1, Aqy1, and CLC-ec1. *Biophys J*. 2012;103:786-796. https://doi.org/10.1016/j.bpj.2012.07.022.

53. Tibshirani R, Walther G, Hastie T.. Estimating the Number of Clusters in a Data Set Via the Gap Statistic. https://statweb.stanford.edu/gwalther/gap.

54. Ketchen DJ, Shook CL. The application of cluster analysis in strategic management research: an analysis and critique. *Strateg Manag J*. 1996;17(6):441-458. https://doi.org/10.1002/(SICI)1097-0266(199606)17:6<441::AID-SMJ819>3.0.CO;2-G.

55. Zhao J, Quan H, Xu Y, Kong X, Jin L, Lou L. Flumatinib, a selective inhibitor of BCR-ABL/PDGFR/KIT, effectively overcomes drug resistance of certain KIT mutants. *Cancer Sci*. 2014;105:117-125. https://doi.org/10.1111/cas.12320.

56. Antonescu CR, Besmer P, Guo T, et al. Acquired resistance to imatinib in gastrointestinal stromal tumor occurs through secondary gene mutation. *Clin Cancer Res*. 2005;11(11):4182-4190. https://doi.org/10.1158/1078-0432.CCR-04-2245.

57. Nishida T, Kanda T, Nishitani A, et al. Secondary mutations in the kinase domain of the KIT gene are predominant in imatinib-resistant gastrointestinal stromal tumor. *Cancer Sci*. 2008;99(4):799-804. https://doi.org/10.1111/j.1349-7006.2008.00727.x.

58. Spitaleri G, Biffi R, Barberis M, et al. Inactivity of imatinib in gastrointestinal stromal tumors (GISTs) harboring a KIT activation-loop domain mutation (exon 17 mutation pN822K). *OncoTargets Ther*. 2015;8:1997-2003. https://doi.org/10.2147/OTT.S81558.

59. Mol CD, Dougan DR, Schneider TR, et al. Structural basis for the autoinhibition and STI-571 inhibition of c-kit tyrosine kinase. *J Biol Chem*. 2004;279(30):31655-31663. https://doi.org/10.1074/jbc.M403319200.

60. Huse M, Kuriyan J. The conformational plasticity of protein kinases. *Cell*. 2002;109(3):275-282. https://doi.org/10.1016/S0092-8674(02)00741-9.

61. Kornev AP, Taylor SS. Defining the conserved internal architecture of a protein kinase. *Biochim Biophys Acta BBA - Proteins Proteomics*. 2010;1804(3):440-444. https://doi.org/10.1016/J.BBAPAP.2009.10.017.

62. Kornev AP, Taylor SS, Eyck LFT. A helix scaffold for the assembly of active protein kinases. *Proc Natl Acad Sci*. 2008;105(38):14377-14382. https://doi.org/10.1073/pnas.0807988105.

63. Aleksandrov A, Simonson T. Molecular dynamics simulations show that conformational selection governs the binding preferences of imatinib for several tyrosine kinases. *J Biol Chem*. 2010;285(18):13807-13815. https://doi.org/10.1074/jbc.M110.109660.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.