Selectivity and ranking of tight-binding JAK-STAT inhibitors using Markovian milestoning with Voronoi tessellations

Anupam Anand Ojha, Ambuj Srivastava, Lane William Votapka, and Rommie E. Amaro*

Department of Chemistry and Biochemistry, University of California San Diego, La Jolla, CA, USA

E-mail: ramaro@ucsd.edu

Abstract

Janus kinases (JAK) are a group of proteins in the non-receptor tyrosine kinase (NRTKs) family that play a crucial role in growth, survival, and angiogenesis. They are activated by cytokines through the Janus kinase - signal transducer and activator of transcription (JAK-STAT) signaling pathway. JAK-STAT signaling pathways have significant roles in the regulation of cell division, apoptosis, and immunity. Identification of the V617F mutation in the Janus homology 2 (JH2) domain of JAK2 leading to myeloproliferative disorders has stimulated great interest in the drug discovery community to develop JAK2-specific inhibitors. However, such inhibitors should be selective towards JAK2 over other JAKs and display an extended residence time. Recently, novel JAK2/STAT5 axis inhibitors (N-(1H-pyrazol-3-yl)pyrimidin-2-amino derivatives) have displayed extended residence times (hours or longer) on target and adequate selectivity excluding JAK3. To facilitate a deeper understanding of the kinase-inhibitor interactions and advance the development of such inhibitors, we utilize a multiscale Marko-

vian milestoning with Voronoi tessellations (MMVT) approach within the Simulation-Enabled Estimation of Kinetic Rates v.2 (SEEKR2) program to rank-order these inhibitors based on their kinetic properties and further explain the selectivity of JAK2 inhibitors over JAK3. Our approach investigates the kinetic and thermodynamic properties of JAK-inhibitor complexes in a user-friendly, fast, efficient, and accurate manner compared to other brute force and hybrid enhanced sampling approaches.

1 Introduction

Enzymes play a fundamental role in biology by increasing the rate constant of a chemical reaction and reducing the energy barrier between substrates and products.[1,2] A kinase is an enzyme essential for modulating key processes in organisms including but not limited to protein regulation, cell growth, cell signaling, and metabolism.[3,4] Kinases are responsible for catalyzing the addition of a phosphate group to a substrate (e.g., proteins or saccharides) from a high-energy molecule (e.g., adenosine triphosphate (ATP)). Referred to as phosphorylation, kinases fall under the family of phosphotransferases. A molecule's reactive ability can depend on its phosphorylation state, which may have implications for its binding ability. Depending on the substrate that kinases act on, they can be classified as lipid kinases, carbohydrate kinases, and protein kinases. Among these, protein kinases are the most thoroughly researched due to their influential roles in eukaryotic cells.[4,5] Protein kinases phosphorylate the targeted protein by attaching the γ-phosphate group from ATP to the hydroxyl group of threonine, serine, or tyrosine residues.[6,7] Over the past few decades, these kinases have been shown to influence the development of tumors.[3,8–10] More than 400 human diseases are found to be linked to protein kinases,[7] which have received increasing attention as vital drug targets,[5,11–13] especially in the treatment of cancer.[14]

Tyrosine kinases (TKs) are a family of proteins that catalyze the transfer of phosphate groups from ATP molecules to tyrosine residues of the target protein.[15,16] The TKs can be broadly divided into receptor and non-receptor tyrosine kinases. The receptor tyrosine

kinases (RTKs) are membrane-bound and pass the extracellular signal to the inside of the cells, while non-receptor tyrosine kinases (nRTKs) are mainly cytosolic and bind to ligands to activate downstream signaling.[17–19] nRTKs are involved in cell signaling, differentiation, proliferation, and apoptosis. Janus kinase (JAK) proteins are nRTK receptors involved in activating transcription and production of cytokines to recruit immune cells at the site of infections. The abnormalities and mutations in JAK proteins lead to neurological and immune system defects.

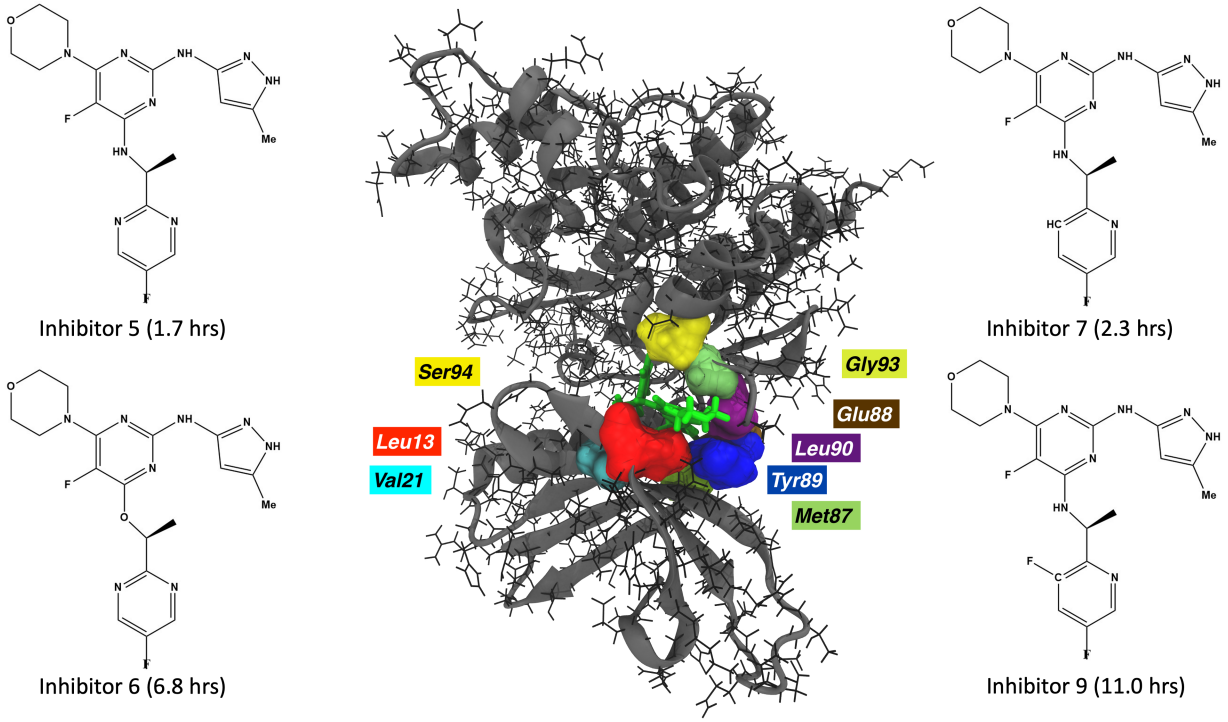

Figure 1: JAK2-inhibitor 9 complex with interacting residues within a cut-off distance of 2.5 Å (center). The inhibitors with large residence times for JAK2 proteins are displayed.

The JAK family consists of four members in mammals, namely Janus kinase 1 (JAK1), Janus kinase 2 (JAK2), Janus kinase 3 (JAK3), and Tyrosine kinase 2 (TYK2).[20,21] JAKs are crucial in signal transduction pathways and are moderated by the binding of cytokine ligands to their respective receptors.[14] Mutations in JAK1 and JAK3 are primarily known to cause severe combined immune deficiency (SCID) diseases. Cytokines are small secreted proteins that can have important effects on the modulation of cell differentiation, immune

development, and inflammatory responses.[14,22,23] A pathway involved in immune response uses the signal transducer and activator of transcription (STAT) proteins, known as the JAK-STAT pathway.[14] When a pro-inflammatory cytokine binds to its respective cytokine receptor, a conformational change occurs in the receptor, leading to the activation of JAK.[21] Activated JAKs trans-phosphorylate and phosphorylate tyrosine residues on the cytokine receptor, producing a docking site for a pair of STAT proteins.[24] Once the STATs gets phosphorylated, these proteins dissociate from the cytokine receptor and dimerize.[25] The resulting phosphorylated STAT dimer makes its way to the cell nucleus and binds to a specific DNA strand to perform gene transcription.[26] The JAK-STAT signaling pathway can have implications for human diseases, including but not limited to, rheumatoid arthritis (RA), inflammatory bowel diseases (IBD), multiple sclerosis (MS), Parkinson's disease (PD), and cancer.[25] While a functional overlap exists between the different kinases in the JAK family, it is eventually dependent on the specific cytokine ligand-receptor interactions and the specific activation of JAKs that can lead to the pathogenesis of a particular disease.

2 Methods

2.1 Simulation Enabled Estimation of Kinetic Rates v.2 (SEEKR2)

2.1.1 Markovian milestoning with Voronoi tessellations

A Voronoi tessellation is a subdivision of space into n regions or "Voronoi cells".[27,28] From a given set of points, $\mathbf{a} = \{a_1, a_2, a_3, ..., a_n\}$ and set of Voronoi cells, $\mathbf{V} = \{V_1, V_2, V_3, ..., V_n\}$ such that $a_1 \in V_1$, $a_2 \in V_2$, $a_3 \in V_3$, ... , $a_n \in V_n$ (Figure 2), let us define a distance metric, $d(a, b)$, that estimates the distance between the two points, a and b. According to the definition of a Voronoi tessellation, a point α will belong to the cell V_1 if and only if $d(a_1, \alpha) < d(a_i), \alpha)$ for $i \in \{2, 3, ..., n\}$.

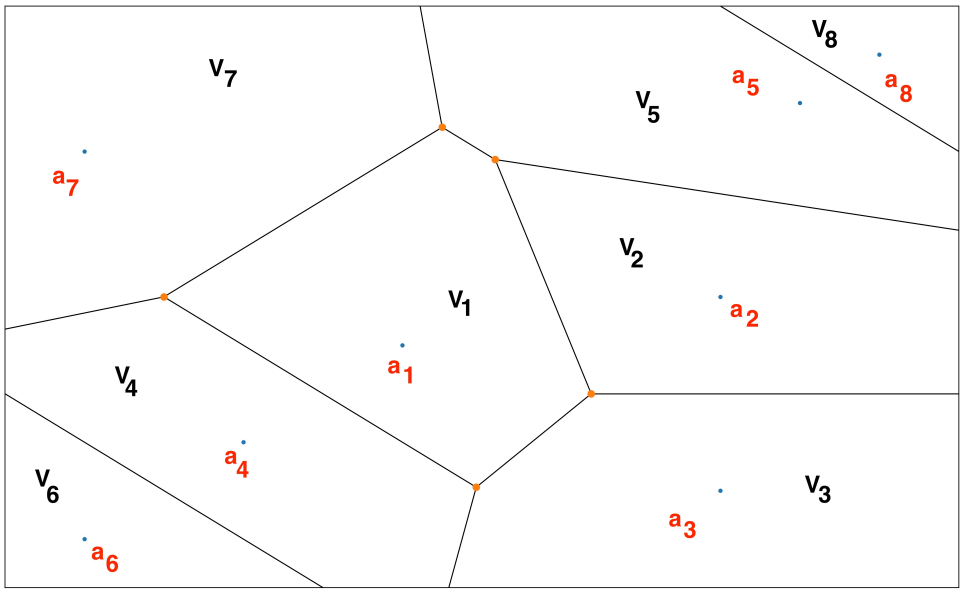

Figure 2: A representative Voronoi diagram where $V_1, V_2, V_3, ..., V_n$ represents the Voronoi cells and $a_1 \in V_1$, $a_2 \in V_2$, $a_3 \in V_3$, ... , $a_n \in V_n$.

SEEKR2 is an open-source software that automates the process of preparation, initiation, running, and analysis of milestoning calculations utilizing MD and Brownian dynamics (BD) simulations to estimate the kinetics and thermodynamics of receptor-ligand binding and unbinding.[29–31] MD simulations are run using the OpenMM simulation engine, while BD simulations are run using the Browndye software.[32] In the SEEKR2 multiscale milestoning approach, the phase space of the receptor-ligand complex is split into two regions, i.e., the MD and the BD region. This partition is based on a predefined collective variable (CV), i.e., the distance between the center of mass (COM) of the ligand and the COM of the receptor's binding site. In the region closer to the binding site, solvent effects and molecular flexibility must be included for describing molecular interactions, therefore, MD simulations are employed. The MD region is further partitioned into several Voronoi cells. Steered molecular dynamics (SMD) simulations are run to generate starting structures for SEEKR2 simulations.[33] SMD simulations pull the ligand slowly out of the binding pocket with a moving harmonic restraint, and a snapshot of the trajectory is saved for every Voronoi cell as it passes through them. Fully atomistic, flexible, and parallel MD simulations are

performed in each Voronoi cell with reflective boundary conditions. When the ligand is farther away from the binding site, i.e., in the BD region, rigid body BD simulations are adequate to describe the diffusional encounter of the ligand and the receptor.

The MMVT-SEEKR2 approach has shown to estimate binding and unbinding kinetic and thermodynamic properties for less complex receptor-ligand complexes with close accuracy, especially the model host-guest systems, i.e., β-cyclodextrin with guest ligands and the model protein system, i.e., the trypsin-benzamidine complex.[31] We thereby extend our efforts in exploring the capabilities of SEEKR2 in estimating kinetic and thermodynamic properties for more complex systems, specifically ligands which are strong binders and have large residence times.

2.1.2 Estimating residence times and binding free energies

According to the MMVT approach, the system evolves according to a continuous-time Markov jump between Voronoi cells.[34,35] Let the rate matrix associated with the evolution be $\mathbf{Q}$, N_{ij} be the number of transitions between milestones, i and j, and R_{i} be the time spent by the trajectory having last touched milestone i. The diagonal and the off-diagonal elements of the transition matrix, $\mathbf{Q}$ are represented by equations 1 and 2 respectively.

$$q_{\mathrm{ii}} = -\sum_{j \neq i} q_{\mathrm{ij}} \qquad (1) \qquad\qquad q_{\mathrm{ij}} = \begin{cases} \frac{N_{\mathrm{ij}}}{R_{\mathrm{i}}} & \text{if } R_{\mathrm{i}} \neq 0 \\ 0 & \text{if } R_{\mathrm{i}} = 0 \end{cases} \qquad (2)$$

MD simulations are run within the Voronoi cells until the convergence is reached. Reflective boundary conditions are employed at the boundaries to confine trajectories within the Voronoi cells. Consequently, velocities of the trajectories are reversed as they touch the edges of the adjacent Voronoi cells. For a Voronoi cell α, let N_{ij}^{α} be the number of trajectory collisions with an ith milestone after having last touched the jth milestone within anchor α, let R_i^{α} be the simulation time having last touched the i_{th} milestone within anchor α, let T_{α} be the total simulation time in cell α, let $N_{\alpha,\beta}$ be the total number of collisions within Voronoi cell α with the boundary shared with Voronoi cell β, and let T be the reciprocal

sum of time spent in all the cells as described by equation 3, then N_{ij} and R_{i} is represented by equations 4 and 5 respectively. The equilibrium probability, π is obtained by solving equations 6 and 7.

$$T = (\sum_{\alpha=1}^{n} \frac{\pi_\alpha}{T_\alpha})^{-1} \tag{3}$$

$$N_{\mathrm{ij}} = T \sum_{\alpha=1}^{n} \pi_\alpha \frac{{N_{\mathrm{ij}}}^\alpha}{T_\alpha} \tag{4}$$

$$R_{\mathrm{ij}} = T \sum_{\alpha=1}^{n} \pi_\alpha \frac{{R_{\mathrm{i}}}^\alpha}{T_\alpha} \tag{5}$$

$$\sum_{\beta=1,\beta\neq\alpha}^{n} \pi_\beta \frac{N_{\beta,\alpha}}{T_\beta} = \sum_{\beta=1,\beta\neq\alpha}^{n} \pi_\beta \frac{N_{\alpha,\beta}}{T_\alpha} \tag{6}$$

$$\sum_{\alpha=1}^{n} \pi_\alpha = 1 \tag{7}$$

With $\hat{\mathbf{Q}}$ as the N-1 by N-1 matrix obtained from the upper left corner of $\mathbf{Q}$, one can compute the mean first passage times (MFPT) or residence times for each milestone described by vector $\mathbf{T}^N$ by solving equation 8.

$$\hat{\mathbf{Q}}\mathbf{T}^N = -\mathbf{1} \tag{8}$$

where $\mathbf{1}$ is a vector of ones. Stationary probabilities obtained from the milestoning simulations are used to construct the free energy profile of unbinding of the receptor-inhibitor complexes with the bound-state milestone as a reference. Stationary probabilities, $\mathbf{p}$, are found by solving the eigenvalue equation 9.

$$\mathbf{Q}\mathbf{p} = \mathbf{p} \tag{9}$$

Let $\mathrm{k_B}$ be Boltzmann's constant, T be the temperature, p_i be the stationary probability of the $\mathrm{i_{th}}$ milestone and p_{ref} be the stationary probability of the bound-state or the reference milestone. The expression for estimating the free energy profile of the $\mathrm{i^{th}}$ milestone, i.e., ΔG_{i}

is given by equation 10.

$$\Delta G_{\mathrm{i}} = -k_{\mathrm{B}} T \ln(\frac{p_i}{p_{ref}}) \tag{10}$$

2.1.3 Ranking JAK-inhibitor complexes with SEEKR2

N-(1H-pyrazol-3-yl)pyrimidin-2-amino derivatives are ATP-competitive inhibitors of the JAK2-STAT5 pathway that are reported to display prolonged residence times on JAK2 and sufficient selectivity against JAK3, both at biochemical and cellular levels.[36,37] The residence times of four inhibitors with the JAK2 and JAK3 kinase, were experimentally determined using a rapid dilution enzymatic assay.[38] We present a computationally inexpensive and efficient application of the SEEKR2 program to predict and rank order the residence times of the JAK2 and JAK3 inhibitors.

System Preparation:

To estimate the residence times of the four inhibitors in the two JAK proteins, an all-atomistic MD simulation is performed in SEEKR2. The X-ray crystal structure of JAK2-JH1 domain in complex with inhibitor 6 (PDB ID: 3ZMM) is used as the reference structure for JAK2 SEEKR2 simulations.[36] For the preparation of JAK2 complex with inhibitors 5, 7, and 9 (Figure 1), the X-ray crystal structure of the JAK2 domain in complex with inhibitor 6 (PDB ID: 3ZMM) is used as a reference structure. Inhibitor 6 is modified to 5, 7, and 9 using the Maestro module of the Schrödinger software suite (Figure 1).[39] AMBER ff14SB forcefield is used to parameterize the protein, while the inhibitor is parameterized using the Antechamber module with the general Amber force field (GAFF) with the AM1-BCC charge model.[40–43] The protein-inhibitor complex is then explicitly solvated with the TIP4P-Ew water model and a salt (Na+/Cl-) concentration of 150 mM in a truncated octahedral periodic box with a 10 Å water buffer.[44] OpenMM MD engine is used to run the simulation at 300 K with a 2 fs timestep and a nonbonded cut-off radius of 9 Å.[45,46] The system is systematically heated from 0 K to 300 K in steps of 3 K of 20 ps each, followed by 20 ns each

of NPT and NVT equilibration simulations.

The X-ray crystal structure of JAK3-JH1 domain in complex with an indazole substituted pyrrolopyrazine (PDB ID: 3ZC6) is used as the reference structure for the JAK3 SEEKR2 simulations.[47] The inhibitor complexed with JAK3 is removed, and the structure is aligned to the JAK2 complexed with inhibitor 6. Inhibitor 6 is then placed at the ATP binding site of the JAK3 protein. Inhibitor 6 is modified to 5, 7, and 9 using the Maestro module of the Schrödinger software suite, and the same protocol is followed for JAK3 as performed for the JAK2 complexes for system solvation and equilibration.

Steered molecular dynamics and Voronoi cell definition: To define Voronoi cells, we described the CV as the distance between the COM of the inhibitor and the COM of α-carbons of the binding site[48] (supplementary table S1). The cut-off distance for the binding site of the inhibitor is defined as 3 Å. All the α-carbon atoms of the surrounding residues of the JAK protein within the cut-off distance of any of the atoms of the inhibitor is defined as the binding site for the receptor-inhibitor complex. Supplementary Table S1 displays the residues of each JAK-inhibitor complex selected for the COM calculation of the binding site. For JAK2-inhibitor complexes, CV-based milestones are defined as concentric spheres and are located at distances of 2.5, 3.0, 3.5, 4.0, 4.5, 5.0, 5.5, 6.0, 6.5, 7.0, 7.5, 8.0, 8.5, 9.0, 9.5, 10.0, 11.0, 12.0, 13.0, 14.0, 15.0, and 16.0 Å respectively from the COM of the binding site. Similarly, for the JAK3-inhibitor complexes, CV-based milestones are defined as concentric spheres and are located at a distance of 3.0, 3.5, 4.0, 4.5, 5.0, 5.5, 6.0, 6.5, 7.0, 7.5, 8.0, 8.5, 9.0, 9.5, 10.0, 11.0, 12.0, 13.0, 14.0, 15.0, and 16.0 Å respectively from the COM of the binding site. In the case of JAK3-inhibitor complexes, none of the residues of the JAK3 protein interacted with the inhibitor within the 2.5 Å radius, leading to the choice of the first milestone at 3.0 Å. This choice should not be problematic since the milestoning procedure would not be significantly sensitive to the choice of the number of milestones as long as each state and pathway is adequately represented in each milestoning model and the results are sufficiently converged. SMD simulations are employed to generate starting structures within

each Voronoi cell where the ligand bound to the complex is slowly pulled out of the binding site in such a way that there is no significant stress to the system and it stays in the local equilibrium. The ligand is slowly pulled from the bound state to the outermost Voronoi cell with a moving harmonic restraint of 50,000 kJ mol^{-1} nm^{-2} in equal steps over the course of 1 μs.

SEEKR2 molecular dynamics simulations: With the starting structures of each Voronoi cell obtained by SMD, MMVT simulations are employed with the same forcefield parameter files used during equilibration simulations. No harmonic restraint is applied during these simulations. Reflective boundary conditions are employed to retain the trajectories within individual Voronoi cells. A total of 400 ns of MD simulations is run within each Voronoi cell. In short, 21 independent and parallel MD simulations of 400 ns are run for each of the four JAK2-inhibitor complexes, and 20 independent and parallel MD simulations of 400 ns within each Voronoi cell is run for each of the four JAK3-inhibitor complexes. To improve the sampling and account for stochasticity, three replicas of SEEKR2 simulations are run for each JAK-inhibitor complex.

3 Results and Discussion

Estimating thermodynamic and kinetic parameters, such as the residence time and free energy of binding and unbinding, is challenging in cases of receptor-inhibitor complexes with an extended residence time.[49,50] A minor change in the structure of inhibitors sometimes leads to an enormous change in the residence time in the binding pockets of proteins. We estimated the residence time of four inhibitors in the binding pocket of JAK2 and JAK3 proteins. We showed that the trend of the residence time predicted by the SEEKR2 milestoning approach captures that of experimental methods. We showed that the trend of the residence time predicted by the SEEKR2 milestoning approach reproduces the experimental findings. Inhibitors 5 and 9 displayed the lowest and the highest residence times for the JAK2 protein,

respectively. Similarly, inhibitors 6 and 9 displayed the lowest and the highest residence times for the JAK3 protein, respectively. Long-scale MD simulations are performed to study the structural aspects of protein-ligand interactions, primarily focusing on these particular inhibitors to explain the discrepancy in their respective residence times.

3.1 Determination of kinetic and thermodynamic parameters from SEEKR2 simulations

Simulations in the majority of the Voronoi cells converged after 400 ns. The MFPT or residence time is calculated using equation 8. The residence times reported in Figure 3a and 3b are the means of the residence times obtained from three independent SEEKR2 simulations for each of the JAK-inhibitor complexes (Supplementary Table S2). Residence times for the novel series of inhibitors for the JAK2 and JAK3 estimated by the SEEKR2 program are in close agreement with the experimental studies (Figure 3a and 3b). SEEKR2 not only predicted the residence times correctly but also preserved the rank ordering of residence times for inhibitors in both the JAK2 and JAK3 complexes. It can be seen from Figure 3a and 3b that inhibitors 6 and 9 display extended residence times in the ATP-binding site of the JAK2 complexes.

ΔG_i is calculated for each of the milestones using equation 10. In case of the JAK2-inhibitor 5 complex, two energy barriers exist as the inhibitor dissociates with the receptor, one at milestone 4 and the other at milestone 11 (Figure 4a). Similarly, two energy barriers exist for the JAK2-inhibitor 9 complex, one at milestone 5 and the other at milestone 13 (Figure 4a). The energy barriers for inhibitor 9 for both the transitions are higher than that of inhibitor 5, indicating that inhibitor 9 is a stronger binder with a higher residence time. For the JAK3-inhibitor 6 complex, two energy barriers exist as the inhibitor dissociates with the receptor, one at milestone 5 and the other at milestone 11 (Figure 4b). Similarly, two energy barriers exist for the JAK3-inhibitor 9 complex, one at milestone 7 and the other at milestone 12 (Figure 4b). The energy barriers for inhibitor 9 for both the transitions are

higher than that of inhibitor 6, indicating that inhibitor 9 is a stronger binder with a higher residence time.

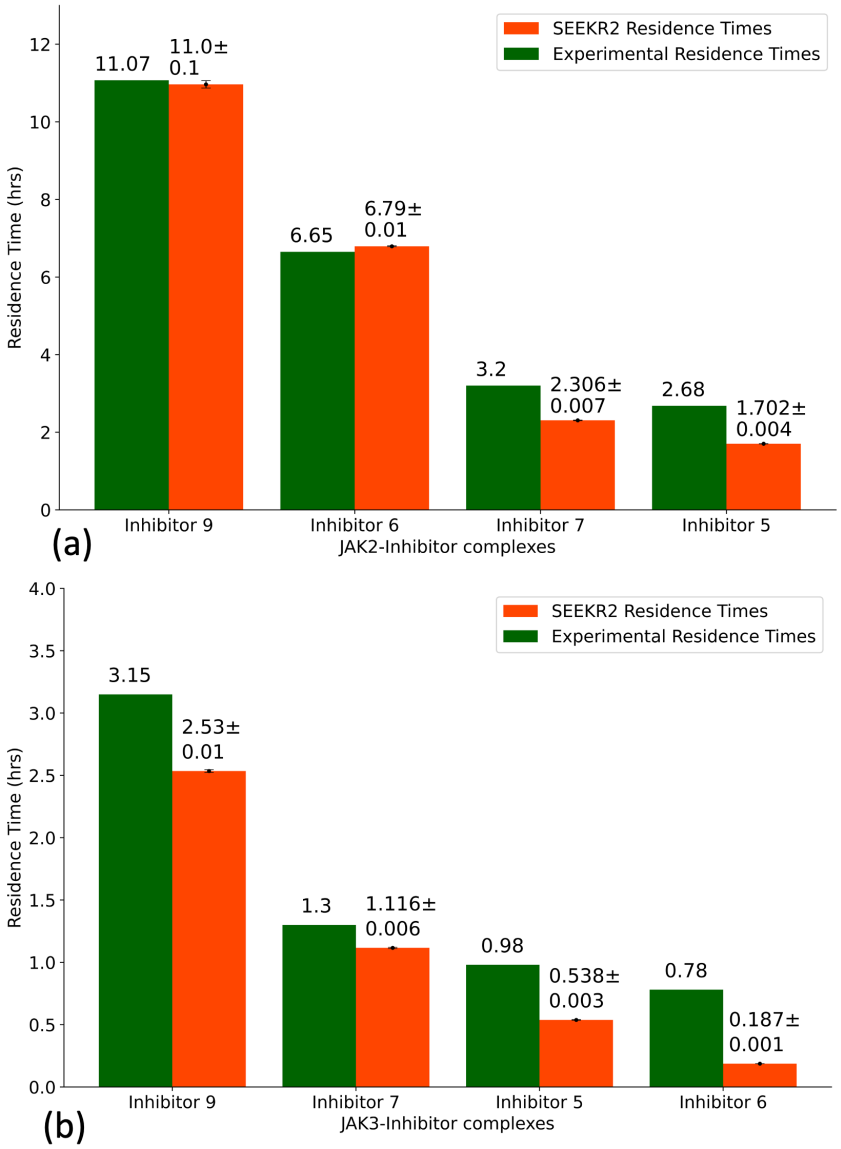

Figure 3: Residence times of JAK2 and JAK3 inhibitors as obtained from the experiments and the SEEKR2 milestoning method. The values of the residence times and the error bars for each of the JAK-inhibitor complex is the average of the three independent SEEKR2 calculations. (a) Residence times of the inhibitors for the JAK2 protein and (b) residence times of the inhibitors for the JAK3 protein are displayed in the figure. Error bars are present for the SEEKR2 residence time data, but they are sometimes too small be visible in the figure.

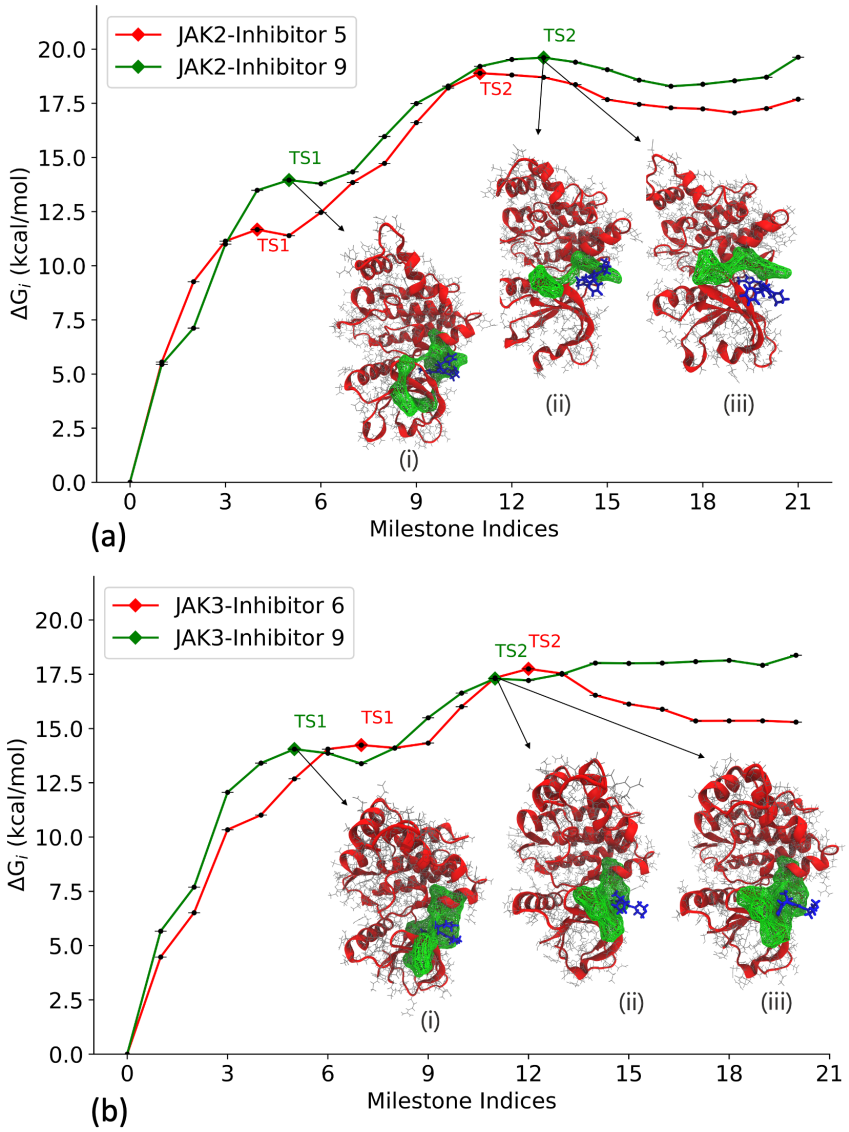

Figure 4: ΔG_i obtained from the SEEKR2 milestoning method for the JAK proteins complexed with the inhibitors. Also shown are the dominant poses of inhibitor 9 as it unbinds from the ATP binding site of JAK complexes. These poses are obtained from the SEEKR2 trajectories for milestones with the local maximum values of ΔG_i. ΔG_i values obtained for each of the JAK-inhibitor complex is the average of the three independent SEEKR2 calculations. (a) ΔG_i values for the JAK2 protein complexed with inhibitor 5 and inhibitor 9 along with (i) JAK2-inhibitor 9 complex at TS1 (ii) JAK2-inhibitor 9 complex at TS2 (pose 1) (iii) JAK2-inhibitor 9 complex at TS2 (pose 2) (b) ΔG_i values for the JAK3 protein complexed with inhibitor 6 and inhibitor 9 along with (i) JAK3-inhibitor 9 complex at TS1 (ii) JAK3-inhibitor 9 complex at TS2 (pose 1) (iii) JAK3-inhibitor 9 complex at TS2 (pose 2)

3.2 Long scale Molecular Dynamics Simulations

To understand and analyze critical aspects of binding and unbinding of the inhibitors at the ATP binding sites of JAK2 and JAK3 and to explain the discrepancy in the residence times of inhibitors and selectivity towards JAK2 over JAK3, three independent 2 μs MD simulations are run for each JAK-inhibitor complex. The starting structures in the first Voronoi cell for each receptor-inhibitor complex served as the starting structures for the long-timescale MD simulations. We used the same forcefield parameter files for the complexes as used in the SEEKR2 simulations. For each of the receptor-inhibitor complexes, a total of 6 μs of MD simulations are run at 300 K with a 2 fs timestep and a nonbonded cut-off radius of 9 Å using the OpenMM MD engine. Simulation trajectories are analyzed using the CPPTRAJ module of the Amber 22 package.[51–53] Analyses including and not limited to ligand-binding site distance analysis, minimum average distance analysis, principal component analysis (PCA), and root mean squared fluctuation (RMSF) analysis are performed to gain a deeper understanding of the binding behavior of these inhibitors.

3.2.1 Discrepancy in residence times: Structure of inhibitors and its interaction with JAKs

The inhibitors, namely 5, 6, 7, and 9, constitute a pyrazol-3-yl amine ring, a heteroaryl C-ring, and a morpholine ring (Figure 5). Different inhibitors are synthesized by substitutions at the heteroaryl C-ring. The pyrazol-3-yl amine ring forms multiple hydrogen bonds with the ATP binding pocket of the JAKs (Figure 6a and 6b), and these contacts are consistent with all the inhibitors. The solvent-exposed morpholine ring does not interact much with the residues in the binding region. Interestingly, a single substitution at the heteroaryl C-ring of the inhibitor leads to a significant difference in their residence times. In the case of inhibitor 9 with respect to inhibitor 5, one of the nitrogen atoms in the heteroaryl C-ring is substituted by a -CF group (Figure 1), leading to a five-fold increase in the residence time of inhibitor 9.

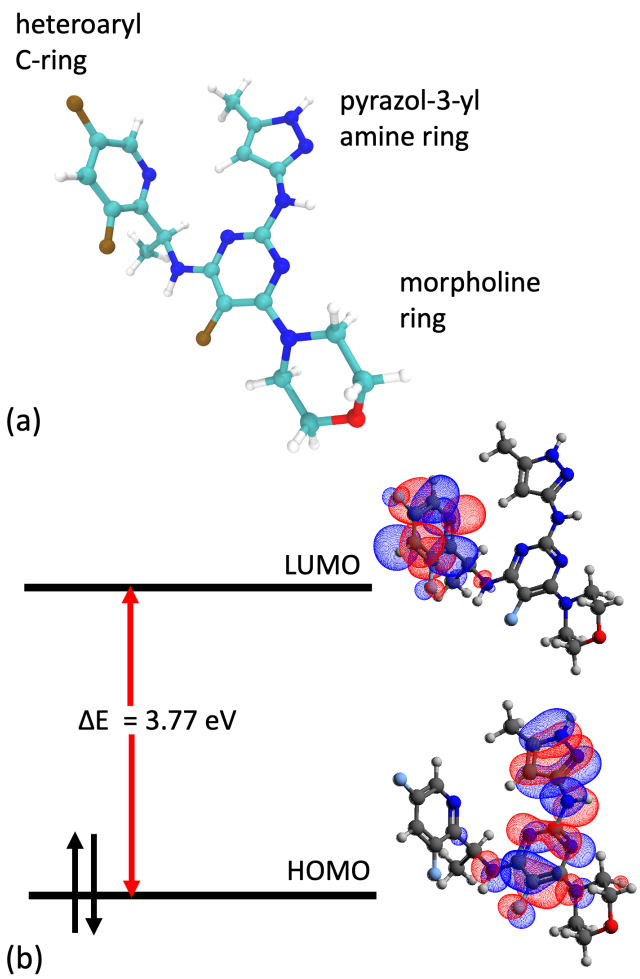

Figure 5: (a) Composition of inhibitor 9 and (b) Molecular orbitals of inhibitor 9

Inhibitor 9 displayed the highest residence time in both the JAK2 and JAK3 proteins. To investigate further the contributions of the heteroaryl C-ring towards the increased residence time and determine the donor-acceptor capabilities of the inhibitor, quantum mechanical (QM) calculations are run for inhibitor 5 and inhibitor 9 to determine the highest occupied molecular orbitals (HOMO) and lowest unoccupied molecular orbitals (LUMO). The Gaussian 16 suite of programs is used to carry out geometry optimization using Becke's three-parameter functional in combination with the Lee-Yang-Parr correlation functional (B3LYP) and 6-31G(d,p) basis set.[54–57] It is observed that the heteroaryl C-ring constitutes the LUMO (Figure 5) for all the inhibitors. The presence of fluorine in inhibitor 9 causes extra stabilization of the bound state since the substituted fluorine atom in the heteroaryl

C-ring interacts with the hydrogen of the β-carbon of the serine residue (Ser94), maintaining an average distance of 2.64 Å with a minimum distance of 2 Å (Figure 6b). In contrast, for inhibitor 5, this interaction is missing (Figure 6a). Further evidence is provided by the HOMO-LUMO energy calculations obtained from the QM calculations. It is observed that the HOMO-LUMO energy difference for inhibitor 9 (3.77 eV) is higher than that of inhibitor 5 (3.40 eV). The HUMO energies for the inhibitors 5 and 9 are nearly identical, but the LUMO energy for inhibitor 9 is higher than that for inhibitor 5. A higher energy LUMO suggests a more electron-deficient character of the heteroaryl C-ring leading to stabilization interaction with the serine (Ser94) residue of JAK2. In short, the electronegativity of F leads to the electrostatic pull of the hydrogen atom in the serine residue and is responsible for a higher residence time for inhibitor 9 than other inhibitors.

To gain additional insights into the dynamics of the receptor-inhibitor complex and to explain the discrepancy in residence times of inhibitor 5 and inhibitor 9 for the JAK2-inhibitor complex, PCA is implemented to the 3D positional coordinates obtained from the MD trajectories.[58–60] PCA explains the variance in the dataset by transforming the MD trajectories into a set of orthogonal vectors or principal components representing characteristic molecular internal motions. The first PC shows the maximum variance in the data, followed by the second PC, and so on. Although the first PC is extremely useful in gaining insights into the system dynamics, the actual motion of the system is the combination of all the PCs. Figure 6c and 6d shows the first PC obtained for the JAK2-inhibitor 5 and JAK2-inhibitor 9 complex respectively. Figure 6c shows a greater domain movement around the binding region of JAK2-inhibitor 5 complex. This motion may be attributed to a region of high instability around the binding site for inhibitor 5 leading to a lower residence time than inhibitor 9.

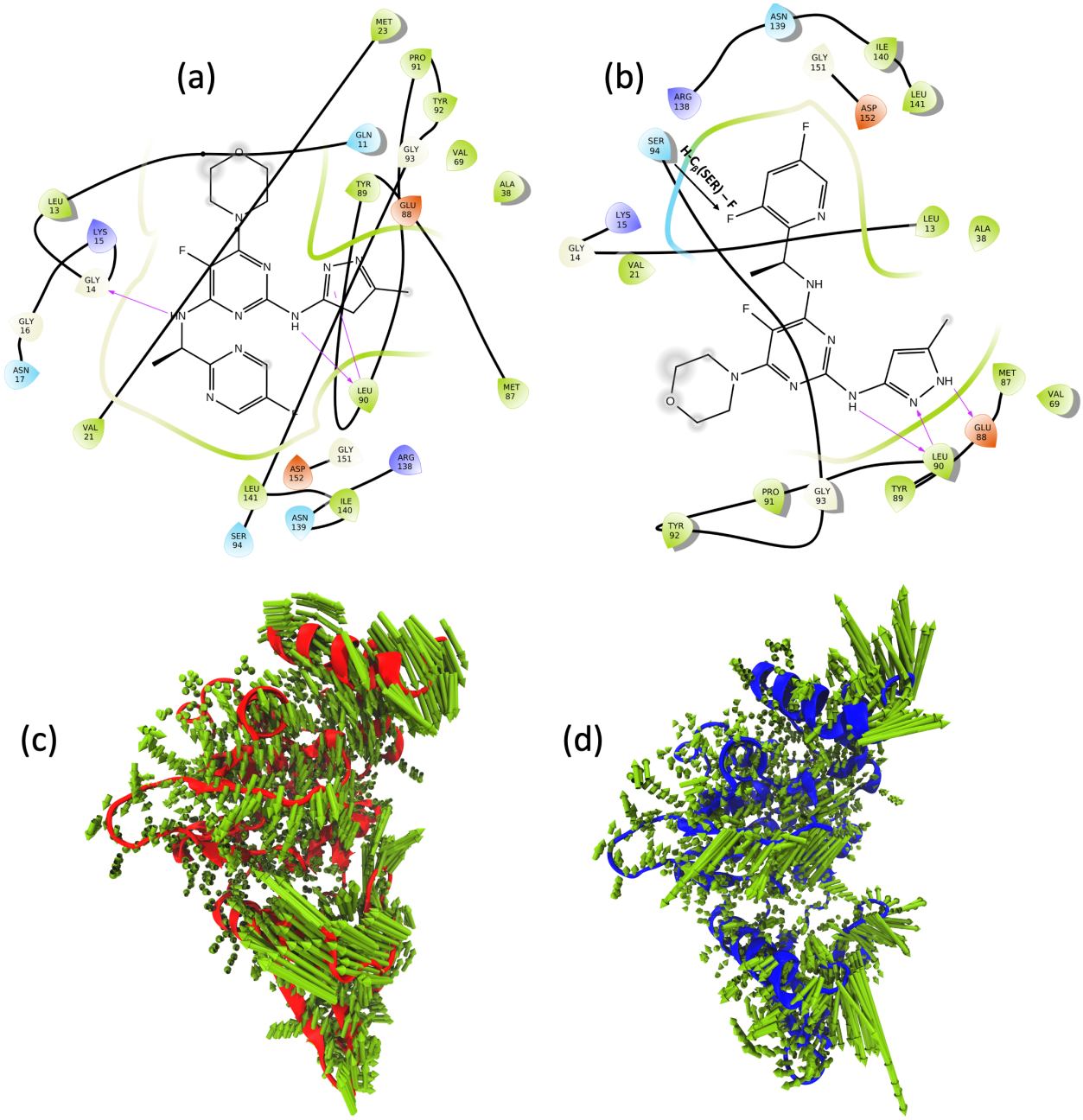

Figure 6: (a-b) Binding site of inhibitors for JAK2 complex showing important interactions with surrounding residues. (a) JAK2-inhibitor 5 complex (b) JAK2-inhibitor 9 complex (c-d) principal component analysis for JAK2 inhibitor complexes from 2 μs of MD simulation trajectory (c) First normal mode for JAK2-inhibitor 5 complex (d) First normal mode for JAK2-inhibitor 9 complex.

3.2.2 Selectivity of inhibitors towards JAK2 over JAK3

The inhibitors at the binding site of the JAK2 protein display higher residence times than the same series of inhibitors for the JAK3 protein. To corroborate these experimental findings, minimum average distance analysis is performed to obtain a detailed description of the binding pocket of the JAK-inhibitor complex. The minimum distance between any two atoms of the amino acid and the inhibitor averaged over the course of the 2 μs trajectory for all the residues is calculated for the JAK-inhibitor complexes. Figure 7a represents the binding pocket of JAK2-inhibitor 9 complex, while Figure 7b represents the binding pocket of JAK3-inhibitor 9 complex. Interacting residues described in the figure are chosen with a cut-off distance of 4 Å. Supplementary Table S3 shows the list of interacting residues for inhibitor 9 in complex with JAK2 and JAK3 proteins. It is evident from Figure 7a and 7b that inhibitor 9 interacts with more residues of JAK2 over JAK3. It is also observed that the binding site occupies a larger volume, and the inhibitor is placed deeper in the binding pocket of JAK2, explaining the selectivity of the same towards JAK2 over JAK3. Interestingly for JAK3, it has been observed that the substituted fluorine atom in the heteroaryl C-ring in inhibitor 9 does not interact with the hydrogen of the β-carbon or any other heavy atom of the serine residue (Ser93).

Root mean square fluctuations (RMSF) calculations are performed to identify important residues and domains associated with inhibitor binding and unbinding.[61] A root mean squared (RMS) fit to the average structure is performed to obtain the fluctuations without rotations and translations, and a mass-weighted averaging of atomic fluctuations for each residue is carried out for the entire simulation trajectory. As demonstrated in Figure 8a, the binding site flanking residues for JAK2, namely Gly14, Lys15, Phe18, Gly19, Ser45, Glu47, Asp52, Arg55, Glu56, and Arg80, have lower RMSF and stabilize upon inhibitor 9 binding as compared to inhibitor 5 suggesting their role in stabilizing the receptor-inhibitor complex. Similarly, in JAK3 proteins, however, residue fluctuations are mostly similar, though only a few of the binding site flanking residues, such as Phe21, Gly22, Gln46, Gly49, Pro50, Asp51,

Gln52, and Phe56, show a significant difference in fluctuations upon inhibitor 9 binding as compared to inhibitor 6 (Figure 8b). A higher number of residues in JAK2 contributing to the low fluctuations at the binding site may contribute to the selectivity of inhibitor 9 towards JAK2 over JAK3.

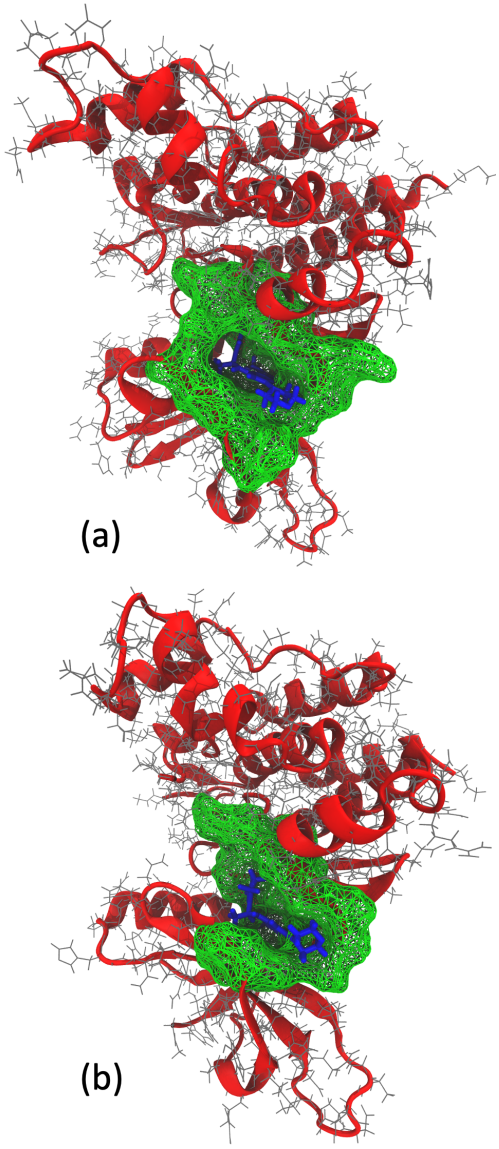

Figure 7: Binding site (green mesh) obtained from minimum average inhibitor-residue distances from three independent 2 μs of MD simulation trajectories. (a) JAK2-inhibitor 9 complex (b) JAK3-inhibitor 9 complex

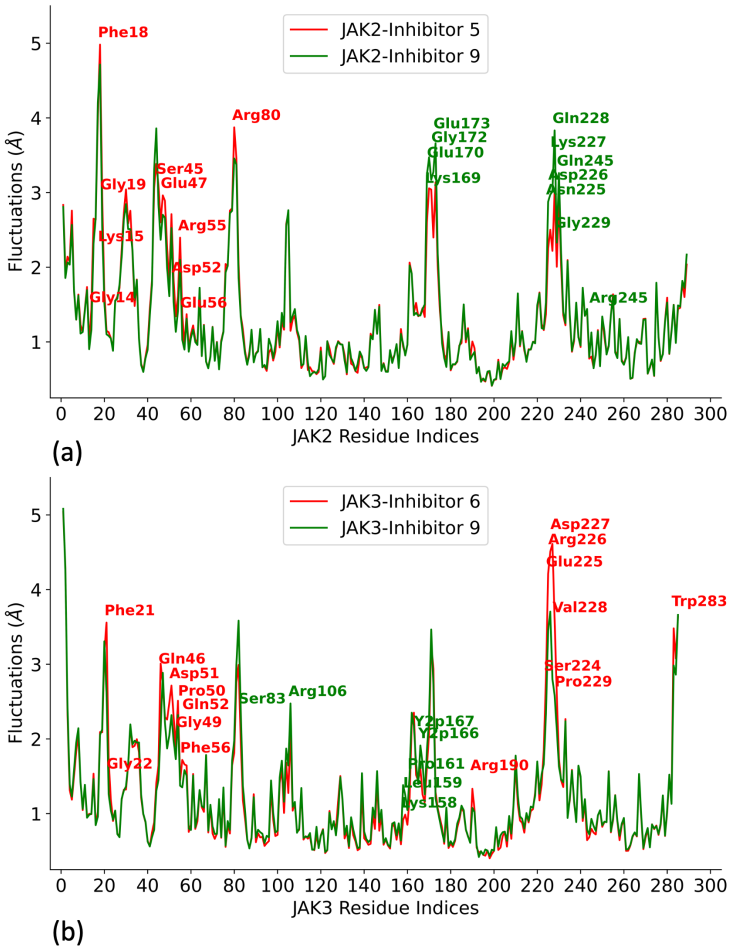

Figure 8: Residue fluctuation analysis for JAK2 and JAK3 inhibitor complexes obtained from three independent 2 μs of MD simulation trajectories. (a) JAK2-inhibitor 5 vs. JAK2-inhibitor 9 complex (b) JAK3-inhibitor 6 vs. JAK3-inhibitor 9 complex

The binding pocket volume of the JAKs is a direct consequence of residues interacting with the inhibitor at the ATP binding site. These pocket volumes are complementary to the shape of the inhibitors as well. To compare the binding pockets of different inhibitors in JAK2 and JAK3 proteins, POVME, a tool to analyze binding pocket volumes, was utilized.[62,63] POVME provides a grid-based pocket representation of the inhibitor binding site. The pocket volumes are calculated with a grid spacing of 0.1 Å and a distance cut-off of 1.09 Å. Deep pocket volumes are observed for the JAK2 inhibitors where these inhibitors are tightly bound to the interacting residues. Supplementary Figure S1 shows a distinct difference in the binding pocket volumes for JAK2 vs. JAK3 proteins, where the volumes associated with

inhibitors in the binding domain of JAK2 are significantly higher than those of JAK3.

Inhibitor-binding site distance analysis is performed for each receptor-inhibitor complex averaged over three independent MD simulation trajectories of 2 μs each. From the starting structure of the zeroth milestone of each JAK-inhibitor complex, residues encompassing the inhibitor within a cut-off radius of 4 Å defined the binding site. The distance between the center of masses of the inhibitors and the α-C atoms of the binding site are used to calculate the inhibitor-binding site distance. It has been observed for all four inhibitors that the inhibitor-binding site distance in the case of JAK2-inhibitor complexes is less than that of the JAK3-inhibitor complexes (Supplementary Figure S2) suggesting strong binding of the inhibitors to the JAK2 protein.

4 Conclusion

The SEEKR2 milestoning method proved efficient in estimating the experimental residence times for different JAK-inhibitor complexes. The trend in residence times for the set of inhibitors for the JAK2 and JAK3 proteins is also conserved. It becomes evident from the SEEKR2 milestoning approach and the experiments that the series of inhibitors display an extended residence time and bind stronger to JAK2 than to JAK3. Among the inhibitors, inhibitor 9 displayed the highest residence time in the JAK2 protein. The results are further supported by MD simulations where important binding residues have lower distances from the inhibitor and lesser fluctuation in the JAK2-inhibitor 9 complex. In addition, the QM calculations show a higher electron density on the fluorine groups in the heteroaryl C-ring of inhibitor 9, strengthening the binding with JAK2 and JAK3 proteins resulting in the highest residence time among all the inhibitors. SEEKR2 proves to be a useful tool in ranking ligands as binders within a reasonable computational time compared to some other enhanced sampling approaches. The SEEKR2 project is available at `https://github.com/seekrcentral/seekr2`. The structures of JAK2 and JAK3 proteins complexed with the

inhibitors, analysis scripts, and scripts for system preparations for SEEKR2 simulations are available at `https://github.com/anandojha/kinase_SEEKR`.

Acknowledgement

The authors acknowledge Benjamin Jagger and Shiksha Dutta for insightful and helpful discussions. A.A.O. acknowledges support the Molecular Sciences Software Institute (MolSSI) fellowship under NSF grant OAC-1547580. R.E.A. acknowledges support from NSF Extreme Science and Engineering Discovery Environment (XSEDE) CHE060063 and NIH GM132826. All simulations were performed using the Triton Shared Computing Cluster (TSCC) and Expanse at the San Diego Supercomputing Center (SDSC). The latter allocation is supported by NSF XSEDE CHE060063 to R.E.A.

Supporting Information Available

Supplementary information includes the list of interacting residues of JAKs with inhibitors, the pocket volume analysis, and the inhibitor-binding site analysis for different JAK-inhibitor complexes.

References

(1) Rogne, P.; Rosselin, M.; Grundström, C.; Hedberg, C.; Sauer, U. H.; Wolf-Watz, M. Molecular mechanism of ATP versus GTP selectivity of adenylate kinase. *Proceedings of the National Academy of Sciences* **2018**, *115*, 3012–3017.

(2) Klostermeier, D.; Rudolph, M. *Biophysical Chemistry*; CRC Press, Taylor & Francis Group, 2017.

(3) Matthews, D.; Gerritsen, M. *Targeting Protein Kinases for Cancer Therapy*; Wiley, 2011.

(4) Taylor, S. S.; Ilouz, R.; Zhang, P.; Kornev, A. P. Assembly of allosteric macromolecular switches: lessons from PKA. *Nature reviews Molecular cell biology* **2012**, *13*, 646–658.

(5) England, J. P.; Hao, Y.; Bai, L.; Glick, V.; Hodges, H. C.; Taylor, S. S.; Maillard, R. A. Switching of the folding-energy landscape governs the allosteric activation of protein kinase A. *Proceedings of the National Academy of Sciences* **2018**, *115*, E7478–E7485.

(6) Hunter, T. The Croonian Lecture 1997. The phosphorylation of proteins on tyrosine: its role in cell growth and disease. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* **1998**, *353*, 583–605.

(7) Patterson, H.; Nibbs, R.; McInnes, I.; Siebert, S. Protein kinase inhibitors in the treatment of inflammatory and autoimmune diseases. *Clinical & Experimental Immunology* **2014**, *176*, 1–10.

(8) Brugge, J. S.; Erikson, R. L. Identification of a transformation-specific antigen induced by an avian sarcoma virus. *Nature* **1977**, *269*, 346–348.

(9) Collett, M. S.; Erikson, R. Protein kinase activity associated with the avian sarcoma virus src gene product. *Proceedings of the National Academy of Sciences* **1978**, *75*, 2021–2024.

(10) Levinson, A. D.; Oppermann, H.; Levintow, L.; Varmus, H. E.; Bishop, J. M. Evidence that the transforming gene of avian sarcoma virus encodes a protein kinase associated with a phosphoprotein. *Cell* **1978**, *15*, 561–572.

(11) Ferguson, F. M.; Gray, N. S. Kinase inhibitors: the road ahead. *Nature reviews Drug discovery* **2018**, *17*, 353–377.

(12) Nussinov, R.; Tsai, C.-J. Allostery in Disease and in Drug Discovery. *Cell* **2013**, *153*, 293–305.

(13) Taylor, S. S.; Kim, C.; Cheng, C. Y.; Brown, S. H.; Wu, J.; Kannan, N. Signaling through cAMP and cAMP-dependent protein kinase: Diverse strategies for drug design. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* **2008**, *1784*, 16–26, Inhibitors of Protein Kinases (5th International Conference, IPK-2007) and Workshop Session on Molecular Design and Simulation Methods (Warsaw, Poland, June 23-27, 2007).

(14) Hochhaus, A. Protein Tyrosine Kinases. From Inhibitors to Useful Drugs. *Leukemia* **2006**, *20*, 1465–1465.

(15) Hunter, T.; Cooper, J. A. Protein-tyrosine kinases. *Annual review of biochemistry* **1985**, *54*, 897–930.

(16) Schlessinger, J. Cell signaling by receptor tyrosine kinases. *Cell* **2000**, *103*, 211–225.

(17) Pawson, T. Regulation and targets of receptor tyrosine kinases. *European journal of cancer* **2002**, *38*, S3–S10.

(18) Lemmon, M. A.; Schlessinger, J. Cell signaling by receptor tyrosine kinases. *Cell* **2010**, *141*, 1117–1134.

(19) Taniguchi, T. Cytokine signaling through nonreceptor protein tyrosine kinases. *Science* **1995**, *268*, 251–255.

(20) Ghoreschi, K.; Laurence, A.; O'Shea, J. J. Janus kinases in immune cell signaling. *Immunological Reviews* **2009**, *228*, 273–287.

(21) Yamaoka, K.; Saharinen, P.; Pesu, M.; Holt, V. E.; Silvennoinen, O.; O'Shea, J. J. The janus kinases (jaks). *Genome biology* **2004**, *5*, 1–6.

(22) Morris, R.; Kershaw, N. J.; Babon, J. J. The molecular details of cytokine signaling via the JAK/STAT pathway. *Protein Science* **2018**, *27*, 1984–2009.

(23) Zhang, J.-M.; An, J. Cytokines, inflammation and pain. *International anesthesiology clinics* **2007**, *45*, 27.

(24) Imada, K.; Leonard, W. J. The jak-STAT pathway. *Molecular immunology* **2000**, *37*, 1–11.

(25) Xin, P.; Xu, X.; Deng, C.; Liu, S.; Wang, Y.; Zhou, X.; Ma, H.; Wei, D.; Sun, S. The role of JAK/STAT signaling pathway and its inhibitors in diseases. *International immunopharmacology* **2020**, *80*, 106210.

(26) Gao, B. Cytokines, STATs and liver disease. *Cell Mol Immunol* **2005**, *2*, 92–100.

(27) Du, Q.; Faber, V.; Gunzburger, M. Centroidal Voronoi tessellations: Applications and algorithms. *SIAM review* **1999**, *41*, 637–676.

(28) Moller, J. *Lectures on random Voronoi tessellations*; Springer Science & Business Media, 2012; Vol. 87.

(29) Votapka, L. W.; Jagger, B. R.; Heyneman, A. L.; Amaro, R. E. SEEKR: simulation enabled estimation of kinetic rates, a computational tool to estimate molecular kinetics and its application to trypsin–benzamidine binding. *The Journal of Physical Chemistry B* **2017**, *121*, 3597–3606.

(30) Votapka, L. W.; Stokely, A. M.; Ojha, A. A.; Amaro, R. E. SEEKR2: Versatile multiscale milestoning utilizing the OpenMM molecular dynamics engine. *Journal of chemical information and modeling* **2022**, *62*, 3253–3262.

(31) Jagger, B. R.; Ojha, A. A.; Amaro, R. E. Predicting ligand binding kinetics using a Markovian milestoning with voronoi tessellations multiscale approach. *Journal of Chemical Theory and Computation* **2020**, *16*, 5348–5357.

(32) Huber, G. A.; McCammon, J. A. Browndye: a software package for Brownian dynamics. *Computer Physics Communications* **2010**, *181*, 1896–1905.

(33) Izrailev, S.; Stepaniants, S.; Isralewitz, B.; Kosztin, D.; Lu, H.; Molnar, F.; Wriggers, W.; Schulten, K. *Computational molecular dynamics: challenges, methods, ideas*; Springer, 1999; pp 39–65.

(34) Vanden-Eijnden, E.; Venturoli, M. Markovian milestoning with Voronoi tessellations. *The Journal of chemical physics* **2009**, *130*, 194101.

(35) West, A. M.; Elber, R.; Shalloway, D. Extending molecular dynamics time scales with milestoning: Example of complex kinetics in a solvated peptide. *The Journal of chemical physics* **2007**, *126*, 04B608.

(36) Guan, H.; Lamb, M. L.; Peng, B.; Huang, S.; DeGrace, N.; Read, J.; Hussain, S.; Wu, J.; Rivard, C.; Alimzhanov, M., et al. Discovery of novel Jak2–Stat pathway inhibitors with extended residence time on target. *Bioorganic & medicinal chemistry letters* **2013**, *23*, 3105–3110.

(37) Bian, M.; Ma, Q.-q.; Wu, Y.; Du, H.-h.; Guo-Hua, G. Small molecule compounds with good anti-inflammatory activity reported in the literature from 01/2009 to 05/2021: a review. *Journal of Enzyme Inhibition and Medicinal Chemistry* **2021**, *36*, 2139–2159.

(38) Copeland, R. A. *Evaluation of enzyme inhibitors in drug discovery: a guide for medicinal chemists and pharmacologists*; John Wiley & Sons, 2013.

(39) Madhavi Sastry, G.; Adzhigirey, M.; Day, T.; Annabhimoju, R.; Sherman, W. Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. *Journal of computer-aided molecular design* **2013**, *27*, 221–234.

(40) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general amber force field. *Journal of computational chemistry* **2004**, *25*, 1157–1174.

(41) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. *Journal of chemical theory and computation* **2015**, *11*, 3696–3713.

(42) Jakalian, A.; Bush, B. L.; Jack, D. B.; Bayly, C. I. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: I. Method. *Journal of computational chemistry* **2000**, *21*, 132–146.

(43) Jakalian, A.; Jack, D. B.; Bayly, C. I. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. *Journal of computational chemistry* **2002**, *23*, 1623–1641.

(44) Horn, H. W.; Swope, W. C.; Pitera, J. W.; Madura, J. D.; Dick, T. J.; Hura, G. L.; Head-Gordon, T. Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew. *The Journal of chemical physics* **2004**, *120*, 9665–9678.

(45) Eastman, P.; Swails, J.; Chodera, J. D.; McGibbon, R. T.; Zhao, Y.; Beauchamp, K. A.; Wang, L.-P.; Simmonett, A. C.; Harrigan, M. P.; Stern, C. D., et al. OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. *PLoS computational biology* **2017**, *13*, e1005659.

(46) Eastman, P.; Pande, V. OpenMM: A hardware-independent framework for molecular simulations. *Computing in science & engineering* **2010**, *12*, 34–39.

(47) Lynch, S. M.; DeVicente, J.; Hermann, J. C.; Jaime-Figueroa, S.; Jin, S.; Kuglstatter, A.; Li, H.; Lovey, A.; Menke, J.; Niu, L., et al. Strategic use of conformational bias and structure based design to identify potent JAK3 inhibitors with improved selectivity against the JAK family and the kinome. *Bioorganic & medicinal chemistry letters* **2013**, *23*, 2793–2800.

(48) Dodda, L. S.; Tirado-Rives, J.; Jorgensen, W. L. Unbinding dynamics of non-nucleoside inhibitors from HIV-1 reverse transcriptase. *The Journal of Physical Chemistry B* **2018**, *123*, 1741–1748.

(49) Ahmad, K.; Rizzi, A.; Capelli, R.; Mandelli, D.; Lyu, W.; Carloni, P. Enhanced-Sampling Simulations for the Estimation of Ligand Binding Kinetics: Current Status and Perspective. *Frontiers in molecular biosciences* **2022**, *9*.

(50) Dickson, A.; Tiwary, P.; Vashisth, H. Kinetics of ligand binding through advanced computational approaches: a review. *Current topics in medicinal chemistry* **2017**, *17*, 2626–2641.

(51) Case, D. A.; Cheatham III, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz Jr, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. The Amber biomolecular simulation programs. *Journal of computational chemistry* **2005**, *26*, 1668–1688.

(52) Salomon-Ferrer, R.; Case, D. A.; Walker, R. C. An overview of the Amber biomolecular simulation package. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2013**, *3*, 198–210.

(53) Roe, D. R.; Cheatham III, T. E. PTRAJ and CPPTRAJ: software for processing and analysis of molecular dynamics trajectory data. *Journal of chemical theory and computation* **2013**, *9*, 3084–3095.

(54) Frisch, M. e.; Trucks, G.; Schlegel, H.; Scuseria, G.; Robb, M.; Cheeseman, J.; Scalmani, G.; Barone, V.; Petersson, G.; Nakatsuji, H., et al. Gaussian 16. 2016.

(55) Raghavachari, K. Perspective on "Density functional thermochemistry. III. The role of exact exchange". *Theoretical Chemistry Accounts* **2000**, *103*, 361–363.

(56) Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Physical review B* **1988**, *37*, 785.

(57) Zhao, Y.; Schultz, N. E.; Truhlar, D. G. Design of density functionals by combining the method of constraint satisfaction with parametrization for thermochemistry, thermochemical kinetics, and noncovalent interactions. *Journal of chemical theory and computation* **2006**, *2*, 364–382.

(58) Galindo-Murillo, R.; Roe, D. R.; Cheatham III, T. E. Convergence and reproducibility in molecular dynamics simulations of the DNA duplex d (GCACGAACGAACGAACGC). *Biochimica et Biophysica Acta (BBA)-General Subjects* **2015**, *1850*, 1041–1058.

(59) Galindo-Murillo, R.; Roe, D. R.; Cheatham, T. E. On the absence of intrahelical DNA dynamics on the μs to ms timescale. *Nature communications* **2014**, *5*, 1–8.

(60) Abdi, H.; Williams, L. J. Principal component analysis. *Wiley interdisciplinary reviews: computational statistics* **2010**, *2*, 433–459.

(61) Fuglebakk, E.; Echave, J.; Reuter, N. Measuring and comparing structural fluctuation patterns in large protein datasets. *Bioinformatics* **2012**, *28*, 2431–2440.

(62) Durrant, J. D.; de Oliveira, C. A. F.; McCammon, J. A. POVME: an algorithm for measuring binding-pocket volumes. *Journal of Molecular Graphics and Modelling* **2011**, *29*, 773–776.

(63) Wagner, J. R.; Sørensen, J.; Hensley, N.; Wong, C.; Zhu, C.; Perison, T.; Amaro, R. E. POVME 3.0: software for mapping binding pocket flexibility. *Journal of chemical theory and computation* **2017**, *13*, 4584–4592.