Computational Modeling of the Janus Kinases: Investigation of the Conformational Plasticity and Molecular Basis for Inhibitions

Ph.D. Thesis

By

Md Fulbabu Sk



DEPARTMENT OF BIOSCIENCES & BIOMEDICAL ENGINEERING

INDIAN INSTITUTE OF TECHNOLOGY INDORE

JULY 2022

Computational Modeling of the Janus Kinases: Investigation of the Conformational Plasticity and Molecular Basis for Inhibitions

A THESIS

Submitted in partial fulfillment of the
requirements for the award of the degree
of
DOCTOR OF PHILOSOPHY

by

Md Fulbabu Sk



DEPARTMENT OF BIOSCIENCES & BIOMEDICAL ENGINEERING
INDIAN INSTITUTE OF TECHNOLOGY INDORE
JULY 2022

INDIAN INSTITUTE OF TECHNOLOGY INDORE

CANDIDATE'S DECLARATION

I hereby certify that the work which is being presented in the thesis entitled **Computational Modeling of the Janus Kinases: Investigation of the Conformational Plasticity and Molecular Basis for Inhibitions** in the partial fulfillment of the requirements for the award of the degree of **DOCTOR OF PHILOSOPHY** and submitted in the **Department of Biosciences and Biomedical Engineering, Indian Institute of Technology Indore**, is an authentic record of my own work carried out during the time period from **JUNE 2017** to **JUNE 2022** under the supervision of **Dr. Parimal Kar,** Associate Professor, Indian Institute of Technology Indore.

The matter presented in this thesis has not been submitted by me for the award of any other degree at this or any other institute.

Signature of the Student
(Md Fulbabu Sk)

This is to certify that the above statement made by the candidate is correct to the best of my knowledge.

Signature of Thesis Supervisor
(Dr. Parimal Kar)

Md Fulbabu Sk has successfully given his Ph.D. Oral Examination held on **19-09-2022**.

Signature of Thesis Supervisor
(Dr. Parimal Kar)

ACKNOWLEDGEMENTS

The five years since 2017 were my life's most valuable and unforgettable time. I got acquainted with many extraordinary and bright people. I have had several fascinating experiences at this institute. The highs and lows in these five years have been bearable and, at times, enjoyable due to the grace of ***Almighty Allah****. I am grateful to that supreme power for the good health and strength which I needed to complete this thesis work.*

I thank my advisor, ***Dr. Parimal Kar,*** *for offering me the greatest opportunity of my life by letting me pursue my* ***Doctor of Philosophy*** *in his laboratory. His knowledge, wisdom, and advice throughout the years have endowed me with the power to tackle obstacles for the rest of my career. I thank him for his understanding and encouragement, especially during the early days of my Ph.D. when I made little progress and was hesitant to face him. I thank him for being supportive all the time. I'm extremely grateful for having him as my guide. I am also grateful for his advice on improving my skills in scientific writing, presentation, and communication. Out of all the people who have mentored me in my life,* ***Dr. Kar*** *is one of the few who have changed my life for the better and allowed me to achieve my lifelong aspiration to contribute to the progress of scientific knowledge and its dissemination somewhat.*

I thank my research progress evolution committee, ***Prof. Amit Kumar, Dr. Sharad Gupta, Dr. Debasis Nayak,*** *and* ***Dr. Anjan Chakrabarty****. I thank them for their guidance and suggestions and for helping me keep on track. I am grateful to* ***Prof. Amit Kumar (IIT Indore), Dr. Sushabhan Sadhukhan (IIT Palakkad), Dr. Debasis Nayak (IISER Bhopal), and Dr. Hem Chandra Jha (IIT Indore),*** *who always encouraged and guided me during research leading to great collaborative work.*

I owe gratitude to all faculty members of my alma mater ***Aligarh Muslim University,*** *especially* ***Dr. Shahid Hussain*** *(Professor of Physics, Aligarh Muslim University,*

Aligarh), who helped me gain vision into the world of research through his knowledge and expertise and made me capable enough to pursue Ph.D.

I wish to express my sincere thanks to the ***Director****, IIT Indore, for providing the necessary amenities involved in the research. I would like to thank* ***IIT Indore*** *for providing the laboratory, computational facilities, and beautiful residential complexes. I want to extend my thanks to the* ***Department of Science and Technology****, the* ***Government of India****, for providing me with handsome financial assistance during the entire tenure of my Ph.D.*

I would also like to acknowledge the past and present members of the ***KAR Research lab****. This lab has been like a second family over the years to me. Everyone has been part of my life in so many aspects. I would like to thank* ***Dr. Nisha Amarnath Jonniya*** *for helping me immensely at the beginning of my Ph.D. study, patiently answering my silly questions, and offering suggestions for new directions in my research. Special thanks go to* ***Dr. Rajarshi Roy*** *for always being there when I needed technical and scientific discussions done. My deep appreciation goes to* ***Mr. Sayan Poddar, Mr. Suman Koirala, Mr. Kapil Ursal, Ms. Subhasmita Mahapatra, Mr. Naveen Kumar****, and* ***Mr. Sunanda Samanta*** *for being the most incredible friends in this world. Thank you for opening your hearts and for being there every step of the way. I love you guys! Your friendship means everything to me.*

I would like to thank all my dear friends at IIT Indore for always encouraging me and keeping my spirits up. ***Dr. Atreyee Ghosh****, and* ***Dr. Shweta Jakhmola****, another two gems of confidants, were my home away from home. Work, food, and fun were quadrupled with them, and eventually, Ph.D. became much easier. Mr.* ***Sumit Kundu****, Mrs.* ***Chanchal****, little sweet angel* ***Avni****, Mr.* ***Omkar****, Ms.* ***Deeksha****, Ms.* ***Kanchan****, Mr.* ***Satyam****, Mrs.* ***Bhavna Joshi****, Mr.* ***Amarnath Jonniya****, Mrs.* ***Neha Jain****, Mr.* ***Uma Sankar****, Mr.* ***Mrinal,*** *Mr.* ***Sandeep,*** *Mr.* ***Ankit Jaiswal*** *were some crucial names in my friend list to*

acknowledge here. Their friendships have helped me through many a rough time. I will never forget these great times we spent together.

I would like to devote a special thought to ***my family members*** *for their never-ending support. They are my constant source of inspiration and energy. If I achieve anything in my life, it is all due to them. Thank you very much, my dear* ***Dadi*** *(grandmother) and* ***Dado*** *(grandfather), for ensuring my education and constantly inspiring me to achieve my goals. A special thanks to my all-time mentors-cum-inspirations, Mr.* ***Mainul Kabir*** *&* *Mr.* ***Kawsar Ali,*** *for their constant support and encouragement that gave me an extra photon.*

Finally, I am most grateful to my beloved wife, ***Rijia Sultana,*** *and my big bundle of cuteness, my sweetheart* ***Amreen Falak Sinthia****. Rijia is the only one in the world who knows deeply about me and my real struggles. Her pleasant characteristics and firm and persistent will have given me extra calories to accomplish my Ph.D. With her company, I am confident to face difficulties and pursue excellence for the rest of my life.*

I also place on record my sense of gratitude to one and all who, directly or indirectly, have played a role in this venture.

-------- ***Md Fulbabu Sk***

Dedicated to

the memory of

my loving sweetheart

Dadi (grandmother)

SYNOPSIS

Introduction

Protein kinases represent a large and diverse class of proteins that play critical roles in intracellular signal transmission. All protein kinases typically catalyze the reaction of phosphorylation in eukaryotic cells by transferring the γ-phosphate of ATP (adenosine triphosphate) onto the substrate's hydroxyl moiety. Modulating protein kinase activity is vitally important, and their malfunction leads to many diseases.

The Janus kinase/signal transducer and activator of transcription (JAK/STAT) signaling pathway is regarded as one of the central communication nodes in the cell function[1]. More than 50 cytokines and growth factors have been identified in the JAK/STAT signaling pathway, such as hormones, interferons (IFN), interleukins (ILs), and colony-stimulating factors. The JAK/STAT signaling pathway has profoundly influenced the current understanding of human health and disease. Many papers have reported the importance of this pathway in malignancies[2] and autoimmune diseases[3,4]. Thus, inhibiting the JAK/STAT pathway is promising for treating various diseases[5,6].

Activation of the JAK/STAT signaling pathway: (1) cytokines and growth factors bind to their corresponding receptors, (2) leading to receptor dimerization and recruitment of related JAKs; (3) JAK activation leads to tyrosine phosphorylation of the receptors and formation of docking sites for STAT; (4) STATs are phosphorylated by tyrosine; (5) STATs dissociate from the receptor to form homodimers or heterodimers; (6) STAT dimers enter the nucleus, bind to DNA, and regulate transcription (see **Figure 1**). The Suppressor of Cytokine Signaling (SOCS) family of proteins are negative-feedback inhibitors of signaling induced by cytokines that act via the JAK/STAT pathway. SOCS proteins can act as ubiquitin ligases by recruiting Cullin5 to ubiquitinate signaling components. However, SOCS1, the most potent member of the family, can also inhibit JAK directly.

The JAK/STAT signaling and human diseases: The JAK/STAT signaling pathway has profoundly influenced the current understanding of human health and diseases. Many research papers have reported the importance of this pathway in malignancies and autoimmune diseases[1,4,7]. The JAK/STAT pathway is a central signaling hub in human immune responses and cancer[6]. It is critical for meeting the diverse challenges faced by the immune system, from resisting infection to maintaining immune tolerance, enforcing barrier functions, and guarding against cancer. The JAK/STAT signaling has been implicated in almost every autoimmune disease[8]. Therefore, a delicate balance must be reached to allow for efficient inductions of the JAK/STAT pathway when the immune system is called to action and proper diminution when the instigating threat subsides. All of the known STAT loss-of-function mutations can be associated with particular types of pathogens[3]. The recent identification of patients with gain-of-function STAT3 mutations highlights the central role of this pathway in promoting autoimmune disease. Over the past 30 years, the link between JAK/STAT signaling and oncogenesis has become a major thread in cancer biology[4].

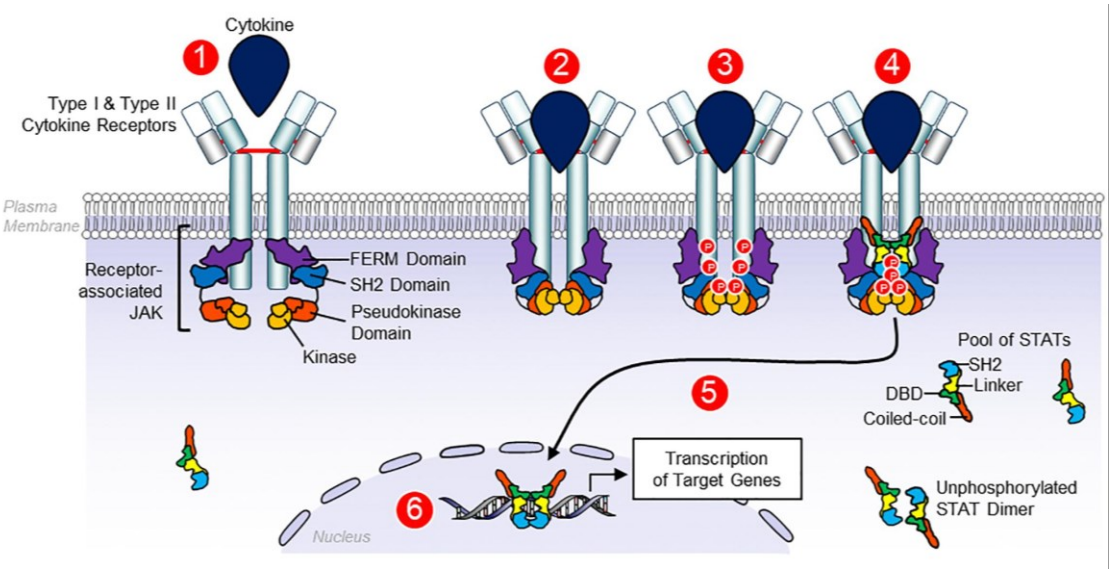


Figure 1: Activation and negative regulation of JAK/STAT signaling pathways. The figure is reprinted from Perez-Jeldres et al. (Ref[9]).

Objectives of work

Although involvements of the JAK/STAT-signaling pathway in regulating autoimmune disorders have been identified, a comprehensive mechanistic understanding of the conformational dynamics of proteins associated with the path is still lacking. How do the post-translational modifications of proteins (phosphorylation) impact the structure and dynamics of JAK? Moreover, the molecular basis of inhibition is still unclear due to the lack of complex crystal structures with small molecules. There is no comprehensive study on the specificity of allosteric inhibitors toward JAK. Besides, the mechanistic insights of the allosteric inhibitor-induced active to inactive state in JAK kinase are not reported yet. Hence, the following aims have been broadly addressed in this thesis:

1. **Unveiling the mysterious contributions of molecular recognition in the treatment of autoimmune disorders**
    - To understand the binding mechanism of second-generation competitive inhibitors and their specificity against JAK1.
2. **Atomistic insights into the impact of post-translational modifications (PTM) on JAK structure and dynamics**
    - Phosphorylation-induced conformational diversity of JAK1 and molecular basis of SOCS1 bindings
3. **Theoretical exploration of the active and inactive conformations of JAK through the enhanced sampling techniques**
    - Deciphering and characterizing the intrinsic conformational dynamics of phosphorylated JAK2 bound to type I and type II inhibitors
4. **Computer-aided drug discovery for autoimmune diseases**
    - Finding potential lead molecules by targeting the ATP-binding pocket of JAK1.
    - Identifying inhibitor-induced inactive conformations for allosteric drug designing.
    - Elucidating thermodynamics of inhibitor binding/unbinding pathway

Chapter 1: The JAK/STAT pathway and its involvement in autoimmune disorders and neuroinflammation pathogenesis.

This chapter describes a detailed introduction of JAK kinases and the JAK/STAT signaling pathway and how the dysfunction of the JAK/STAT signaling pathway takes part in the pathogenesis of auto-immune and neuro-inflammation diseases. We also discussed the features of active and inactive JAK kinases and their role in structure-based drug development and in enhancing future drug candidates' specificity and affinity.

The JAK/STAT pathway is a highly conserved pathway of signal transduction. Dysregulation of the JAK/STAT pathway is associated with various diseases. Chronic inflammation is a typical symptom of aging and age-related diseases. The JAK/STAT pathway frequently mediates abnormally elevated cytokines to induce gene transcription. However, few studies have been performed to identify the molecular mechanism of the JAK/STAT pathway regulating age-related diseases. This potentially is an important insight to extending the life span.

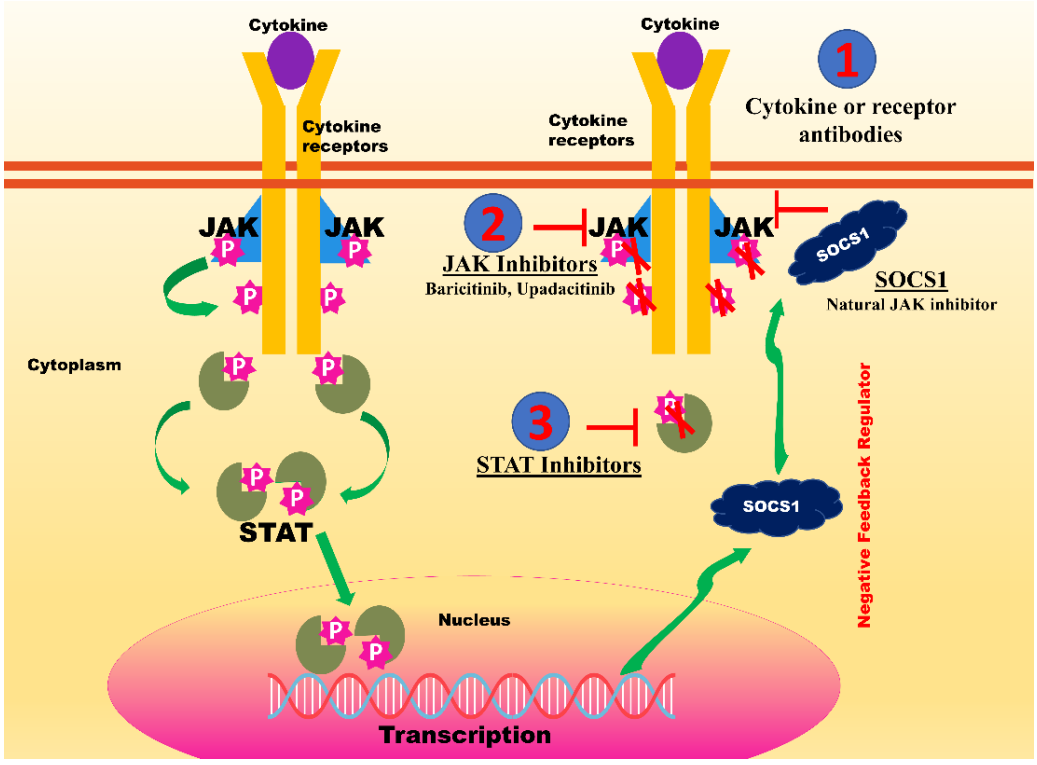


Figure 2: Therapeutic targets of the JAK/STAT signaling pathway. **(1)** Recombinant cytokines, cytokine antibodies, and receptor antibodies are designed to target cytokines or receptors; **(2)** small molecule and peptide inhibitors are designed to target JAKs kinase domain and block ATP and substrate binding, respectively **(3)** small-molecule inhibitors are designed to target the STAT proteins and inhibits the dimerization of STATs.

Medications that target the JAK/STAT pathway can be classified into three types: (1) cytokine or receptor antibodies, (2) **JAK inhibitors**, and (3) STAT inhibitors (**Figure 2**). In this thesis, we are interested in JAK inhibitors and their inhibition mechanisms. JAK inhibitors are a group of small-molecule inhibitors with different chemical structures. The therapeutic effects of JAK inhibitors are based on two factors. First, JAKs mediate various cellular activities. Inhibition of JAK function can cause immunosuppression and decrease abnormally elevated serum proinflammatory cytokines mediated by the JAK/STAT signaling pathway. Second, in some diseases, such as myeloproliferative diseases and cancers, identifying gain-of-function JAK mutants enables treatment through their inhibition. Various JAK inhibitors are being studied in preclinical and clinical studies. Tofacitinib, baricitinib, and upadacitinib are the first orally available JAK inhibitors approved for treating RA and other autoimmune diseases.

Chapter 2: Description of the methodology used for the study

This chapter deals with the methodology used in the present study. Here, a detailed description of molecular dynamics simulations (MD) has been provided. This chapter elaborates on how MD simulations can be used to provide insights into the structure-function relationship, conformational dynamics, and thermodynamic properties of biological macromolecules of interest at the atomistic level. Along with that, we have explained the components of the force field, different water models, simulations packages, basic workflow of MD, binding free energy calculations *via* molecular mechanics Poisson-Boltzmann surface area (MM-PBSA)[10] method, as well as other enhanced simulations techniques, such as accelerated molecular dynamic simulations (aMD), and Gaussian accelerated molecular dynamic simulations (GaMD)[11], and its benefits in conformational sampling with a reduced computational cost. Further, we have provided insights into the different post-simulation analysis techniques of the MD trajectories to explain the molecular mechanism and binding free energy of small molecules.

Chapter 3: Unveiling the mysterious contributions of molecular recognition in the treatment of autoimmune disorders

In this chapter, we have described the inhibition mechanisms of JAK-inhibitors (JAKinibs) and the role of non-covalent interactions in binding these JAKinibs. Here we explored the conformation of the JAK1 protein because it has received significant attention after being identified as a promising target for developing anti-RA therapeutics. Currently, no crystal structure is available for JAK1 in complex with the next-generation anti-RA drugs. In the current study, we investigated the mechanism of binding of baricitinib, filgotinib, itacitinib, and upadacitinib to JAK1 using a combined method of molecular docking, molecular dynamics simulation, and binding free energy calculation *via* the molecular mechanics Poisson–Boltzmann surface area (MM-PBSA) scheme. We found that the calculated binding affinity decreases in the order upadacitinib > itacitinib > filgotinib > baricitinib. Due to the increased favorable intermolecular electrostatic contribution, upadacitinib is more selective to JAK1 compared to the other three inhibitors. The dynamics cross-correlation (DCC) and principal component analyses (PCA) showed that different inhibitor bindings significantly affect the binding site dynamics of JAK1. Furthermore, our studies indicated that the hydrophobic residues and hydrogen bonds from the hinge region (Glu^{957} and Leu^{959}) of JAK1 played an essential role in stabilizing the inhibitors. The protein structural network (PSN) analysis reveals that the total number of links and hubs in JAK1/baricitinib (354, 48) is more significant than those in apo (328, 40) and the other three complexes. The JAK1/baricitinib complex shows the highest probability of the highest-ranked community, indicating a compact network of the JAK1/baricitinib complex, consistent with its higher stability than the rest of the four systems. Overall, our study may be crucial for the rational design of JAK1-selective inhibitors with better affinity.

Chapter 4: Atomistic insights into the impacts of post-translational modifications (PTM) of protein on JAK structure and dynamics

This chapter mainly focuses on the impact of phosphorylation states on the structure and dynamics of JAK1 kinase and how the suppressors of cytokine signaling 1 (SOCS1) inhibit the JAK activity by negative feedback regulations. As we know, the dysfunction of the JAK/STAT (Janus kinase/signal transducers and activators of transcription) pathway results in several pathophysiological conditions, including autoimmune disorders. The

negative feedback regulators of the JAK/STAT signaling pathway, SOCS, act as a natural inhibitor of JAK and inhibit aberrant activity. SOCS1 is the most potent member of the SOCS family, whose kinase inhibitory region targets the substrate-binding groove of JAK with high affinity and blocks the phosphorylation of JAK kinases. Overall, we performed an aggregate of 13 μs molecular dynamics simulations on the activation loop's three different phosphorylation (double and single) states. Results from our simulations show that the single Tyr^{1034} phosphorylation could stabilize the JAK1/SOCS1 complex as well as the flexible activation segment. The phosphate-binding loop (P-loop) shows conformational variability at dual and single phosphorylated states. Principal component analysis and protein structure network (PSN) analysis revealed that the different phosphorylation states and SOCS1 binding induce intermediate inactive conformations of JAK1, which could be a better target for future JAK1 selective drug design. PSN analysis suggests that the com-pY1034 system is stabilized due to higher values of network hubs than the other two complex systems. Moreover, the binding free energy calculations suggest that $pTyr^{1034}$ states show a higher affinity toward SOCS1 than the dual and $pTyr^{1035}$ states. We believe that the mechanistic understanding of JAK1/SOCS1 complexation will aid future studies related to peptide inhibitors based on SOCS1.

Chapter 5: Theoretical explorations of the active and inactive conformations of JAK through the enhanced sampling techniques

This chapter mainly comprises the conformational dynamics induced by type-I and type-II inhibitors and their free energy calculations. Although, accounting for the impact of possible structure rearrangements on the binding of different types of kinases inhibitors is complicated by the extensive conformational variability of their catalytic kinase domain (KD). The dynamic KD contains mainly four prominent mobile structural motifs, from N-lobe, the phosphate-binding loop (P-loop), and the αC helix; from C-lobe, the Asp-Phe-Gly (DFG) motif and the activation loop (A-loop). These distinct structural orientations imply a complex signal transmission path regulating the A-loop's flexibility and conformational preference for optimal JAK function. However, the exact dynamical features of the JAK by different types of inhibitor-induced structures remain unclear. We performed microsecond-long Gaussian accelerated MD (GaMD) simulations of three

different phosphorylated JAK2 systems: apo KD, ATP-competitive inhibitor (CI) bound KD in the catalytically active DFG-*in* conformation, and the allosteric inhibitor bound KD in the catalytically inactive DFG-*out* conformation. Our results indicate significant conformational variations upon inhibitor binding in the A-loop and αC helix motions. Our studies also reveal that the DFG-*out* inactive conformation is characterized by a close A-loop rearrangement, open catalytic cleft of N and C-lobe, the outward movement of the αC helix, and open P-loop states. Moreover, the outward sign of αC helix impacts the hallmark salt bridge formation of Lys^{882} and Glu^{898} in an inactive conformation. Finally, we compared their binding poses and free energy by the MM/PBSA approach. The free energy calculations suggested that the AI's binding affinity is higher than CI against JAK2 due to an increased favorable contribution from the total non-polar interactions and the involvement of the αC helix. Overall, our study provides the structural and energetic details to develop more promising type I/II JAK2 inhibitors for treating JAK-related diseases.

Chapter 6: Finding inhibitors and deciphering inhibitor-induced conformational plasticity in the Janus Kinase

The Janus kinase (JAK) is a master regulator of the JAK/STAT pathway. Dysregulation of this signaling cascade causes neuroinflammation and autoimmune disorders. Thus, JAKs have been considered an attractive target for developing anti-inflammatory drugs. The main issue with the recently developed inhibitors is the lack of specificity against JAK isoforms. Hence, designing efficient, effective, and specific targeted therapeutics without being cytotoxic has gained interest. This chapter focuses on the conventional systematic method of high-throughput virtual screening of a large phytochemicals library, followed by molecular dynamics simulation studies in conjunction with free energy calculations. We have also discussed inhibitor-induced conformational dynamics of the kinase. We performed virtual screening of natural products in combination with pharmacological and pharmacokinetic analyses. Subsequently, we performed unbiased molecular dynamics simulations to study the stability of the ligand-bound complexes and ligand-induced inactive conformations.

Notably, inactive kinases display remarkable conformational plasticity; however, ligand-induced molecular mechanisms of these conformations are still poorly understood. Herein,

we performed a free energy landscape analysis to explore the conformational plasticity of the JAK1 kinase. Our MD simulations show that Leonurine, STOCK1N-68642, STOCK1N-82656, and STOCK1N-85809 bound JAK1 exhibited a smooth transition from an active conformation (αC-in) to a completely inactive conformation (αC-out). We also found that ligand binding induces disorders in the αC-helix. Subsequently, the molecular mechanics Poisson Boltzmann surface area (MM/PBSA) calculation suggested three phytochemicals, namely STOCK1N-68642, Epicatechin, and STOCK1N-98615, have higher binding affinity compared to other ligand molecules. The ligand-induced conformational plasticity revealed by our simulations differs significantly from the available crystal structures, which might help design new rational and allosteric drugs. However, further molecular and pharmacokinetics are required to test the selectivity and affinity *in vitro* and *in vivo*.

Chapter 7: Conclusions and future scope of this thesis

The modern algorithms and recent computational techniques allow us to explore the protein structure and dynamics at atomic resolution in detail. The experimental 3D-protein structures provide great structural details of molecular events as a static photograph. However, these events are very dynamic and evolve over time. In this aspect, MD and other simulation methods are crucial in predicting protein's temporal and spatial dynamics in their relevant timescale. Here, this thesis provides a complete description of small molecule recognition mechanisms by JAK to inhibitor-induced conformational dynamics, along with the impact of phosphorylation states on the structural dynamics. Firstly, we have extensively studied the molecular recognition of second-generation FDA-approved JAK inhibitors to know the exact molecular mechanism of binding and specificity of these inhibitors. After that, we looked at the impact of phosphorylation states and the molecular basis of SOCS1 binding on the JAK1. It explained how the phosphorylation of the autophosphorylation site (Tyr^{1034}) significantly stabilizes the JAK1 kinase. We could also distinguish the primary autophosphorylation site (Tyr^{1034}) and observed that the hallmark salt bridge interaction was broken (Lys^{908} and Glu^{925}) in the second Tyr^{1035} phosphorylation state. In addition, we have also explored the critical aspects of a competitive inhibitor (CI),

an allosteric inhibitor (AI) bound active and inactive JAK2 conformations using the recently developed GaMD enhanced sampling.

Further, we could show that AI could be a better choice regarding selectivity and affinity against JAK2. Finally, we were motivated by the above results and performed an *in-silico*-based systematic virtual screening approach on the large phytochemical database to find a promising future drug candidate to treat immune-related diseases. This thesis opens up a new door in the structure-based drug development field, where we can improve the drug affinity and selectivity toward a particular target enzyme in conjunction with experimental structural biology and medicinal chemistry.

References

(1) Hu, X.; Li, J.; Fu, M.; Zhao, X.; Wang, W. The JAK/STAT Signaling Pathway: From Bench to Clinic. *Sig Transduct Target Ther* **2021**, *6* (1), 1–33. https://doi.org/10.1038/s41392-021-00791-1.

(2) Owen, K. L.; Brockwell, N. K.; Parker, B. S. JAK-STAT Signaling: A Double-Edged Sword of Immune Regulation and Cancer Progression. *Cancers* **2019**, *11* (12), 2002. https://doi.org/10.3390/cancers11122002.

(3) O'Shea, J. J.; Schwartz, D. M.; Villarino, A. V.; Gadina, M.; McInnes, I. B.; Laurence, A. The JAK-STAT Pathway: Impact on Human Disease and Therapeutic Intervention. *Annu Rev Med* **2015**, *66*, 311–328. https://doi.org/10.1146/annurev-med-051113-024537.

(4) Banerjee, S.; Biehl, A.; Gadina, M.; Hasni, S.; Schwartz, D. M. JAK–STAT Signaling as a Target for Inflammatory and Autoimmune Diseases: Current and Future Prospects. *Drugs* **2017**, *77* (5), 521–546. https://doi.org/10.1007/s40265-017-0701-9.

(5) Lashgari, N.-A.; Roudsari, N. M.; Momtaz, S.; Sathyapalan, T.; Abdolghaffari, A. H.; Sahebkar, A. The Involvement of JAK/STAT Signaling Pathway in the Treatment of Parkinson's Disease. *Journal of Neuroimmunology* **2021**, *361*, 577758. https://doi.org/10.1016/j.jneuroim.2021.577758.

(6) 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. https://doi.org/10.1016/j.intimp.2020.106210.

(7) Villarino, A. V.; Kanno, Y.; O'Shea, J. J. Mechanisms and Consequences of Jak-STAT Signaling in the Immune System. *Nat Immunol* **2017**, *18* (4), 374–384. https://doi.org/10.1038/ni.3691.

(8) T. Virtanen, A.; Haikarainen, T.; Raivola, J.; Silvennoinen, O. Selective JAKinibs: Prospects in Inflammatory and Autoimmune Diseases. *BioDrugs* **2019**, *33* (1), 15–32. https://doi.org/10.1007/s40259-019-00333-w.

(9) Pérez-Jeldres, T.; Tyler, C. J.; Boyer, J. D.; Karuppuchamy, T.; Yarur, A.; Giles, D. A.; Yeasmin, S.; Lundborg, L.; Sandborn, W. J.; Patel, D. R.; Rivera-Nieves, J. Targeting Cytokine Signaling and Lymphocyte Traffic via Small Molecules in Inflammatory Bowel Disease: JAK Inhibitors and S1PR Agonists. *Frontiers in Pharmacology* **2019**, *10*.

(10) Kollman, P. A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W.; Donini, O.; Cieplak, P.; Srinivasan, J.; Case, D. A.; Cheatham, T. E. Calculating Structures and Free Energies of Complex Molecules: Combining Molecular Mechanics and Continuum Models. *Acc. Chem. Res.* **2000**, *33* (12), 889–897. https://doi.org/10.1021/ar000033j.

(11) Miao, Y.; Feher, V. A.; McCammon, J. A. Gaussian Accelerated Molecular Dynamics: Unconstrained Enhanced Sampling and Free Energy Calculation. *J. Chem. Theory Comput.* **2015**, *11* (8), 3584–3595. https://doi.org/10.1021/acs.jctc.5b00436.

LIST OF PUBLICATIONS

Publications from Thesis:

(A) Manuscript published

1. **Sk, M. F.,** Jonniya, N. A., Roy, R., & Kar, P*. (2022), Unraveling the molecular mechanism of recognition of selected next-generation anti-rheumatoid arthritis inhibitors by Janus Kinase 1, ***ACS Omega,*** 7 (7), 6195-6209.
2. **Sk, M. F.**; Jonniya, N. A.; Roy, R.; Kar, P*. (2022), Phosphorylation-induced conformational dynamics and inhibition of Janus kinase 1 by suppressors of cytokine signaling 1. ***Journal of Physical Chemistry B,*** 126 (17), 3224-3239.
3. **Sk, M. F.***; Kar, P*. (2022), Finding inhibitors and deciphering inhibitor-induced conformational plasticity in the Janus kinase via multiscale simulations. ***SAR and QSAR in Environmental Research***

(B) Manuscript under revision

4. **Sk, M. F*.**; Kar, P*. (2022), Exploration of conformational dynamics of phosphorylated active and inactive Janus kinase 2 using Gaussian accelerated molecular dynamics simulations. ***Journal of Chemical Information and Modeling***. [**In Revision**].

Publications Apart from Thesis:

(A) Manuscript Published

5. Indari, O[†]; **Sk, M. F.**[†]; Jakhmola, S.; Jonniya, N. A.; Jha, H*; Kar, P*. (2022), Decoding the host-parasite protein interactions involved in cerebral malaria through glares of molecular dynamics simulations. ***Journal of Physical Chemistry B***. *126(2)*, 387-402. [**equal contribution].**
6. Jakhmola, Shweta[*†]; **Sk, M. F.**[†]; Chatterjee, A; Jain, K; Kar, P*, Jha, H. C.* (2022), Myelin protein antigenic epitopes presented by major histocompatibility complexes, a driver of multiple sclerosis. ***Computers in Biology and Medicine***. 148, 105856.
7. Jakhmola, Shweta; Jonniya, N.A.; **Sk, M. F.;** Rani, A; Jain, K; Kar, P*, Jha, H. C.* (2022), Identification of potential inhibitors against Epstein-Barr virus Nuclear Antigen 1 (EBNA1): An insight from docking and molecular dynamic simulations. ***ACS Chemical Neuroscience***. 12(16), 3060-3072.
8. Roy, R; Jonniya, N. A.; **Sk, M. F.**; Kar, P*. (2022) "Comparative structural dynamics of isoforms of Helicobacter pylori adhesin BabA bound to Lewis[b] hexasaccharide via multiple replica molecular dynamics simulations. ***Frontiers in Molecular Biosciences***, 9, 852895.
9. Jonniya, N. A.; **Sk, M. F.**; Kar, P*. (2021), Characterizing an allosteric inhibitor-induced inactive state in with-no-lysine kinase 1 using Gaussian accelerated molecular dynamics simulations. ***Physical Chemistry Chemical Physics***, *23(12)*, 7343-7358.
10. **Sk, M. F.**; Haridev, S. R.; Roy, R; Kar, P*. (2021), Investigating potency of TMC-126 against wild-type and mutant variants of HIV-1 protease: A molecular dynamics and free energy study. ***SAR and QSAR in Environmental Research***, *32(11)*, 941-962.
11. **Sk, M. F.**, Jonniya, N. A.; Roy, R; Poddar, S; Kar, P*. (2020), Computational investigation of structural dynamics of SARS-CoV-2 methyltransferase-stimulatory factor heterodimer nsp16/nsp10 bound to the cofactor SAM. ***Frontiers in Molecular Biosciences***, *7*, 353.

12. **Sk, M F.**; Roy, R; Jonniya, N. A.; Poddar, S; Kar, P*. (2021), Elucidating biophysical basis of binding of inhibitors to SARS-CoV-2 main protease by using molecular dynamics simulations and free energy calculations. ***Journal of Biomolecular Structure and Dynamics***, *39(10)*, 3649-3661.
13. Jain, N*; **Sk, M. F.**; Mishra, A; Kar, P; Kumar, A*. (2022), Identification of novel Efflux pump inhibitors for Neisseria gonorrhoeae via multiple ligand-based pharmacophores, e-pharmacophore, molecular docking, density functional theory, and Molecular dynamics approach. ***Computational Biology and Chemistry*****,** 98, 107682.
14. Jonniya, N. A.; **Sk, M. F.**; Roy, R; Kar, P*. (2022), Discovery of potential competitive inhibitors against With No Lysine Kinase 1 for treating hypertension by virtual screening, inverse pharmacophore-based lead optimization, and molecular dynamics simulations. ***SAR and QSAR in Environmental Research***. *33(2)*, 63-87.
15. **Sk, M. F.**; Roy, R; Kar, P*. (2021), Exploring the potency of currently used drugs against HIV-1 protease of subtype D variant by using multiscale simulations. ***Journal of Biomolecular Structure and Dynamics***, *39(3)*, 988-1003.
16. Jonniya, N. A.; **Sk, M. F.;** Kar, P*. (2019), Investigating phosphorylation-induced conformational changes in WNK1 kinase by molecular dynamics simulations. ***ACS Omega***, *4(17)*, 17404-17416.
17. **Sk, M. F.;** Jonniya, N. A.; Kar, P*. (2021), Exploring the energetic basis of binding of currently used drugs against HIV-1 subtype CRF01_AE protease *via* molecular dynamics simulations. ***Journal of Biomolecular Structure and Dynamics***, *39(16)*, 5892-5909.
18. Roy, R; Poddar, S; **Sk, M. F.**; Kar, P*. (2022), Conformational preferences of triantennary and tetraantennary hybrid *N*-glycans in aqueous solution: Insight from 20 microsecond long atomistic molecular dynamic simulations. ***Journal of Biomolecular Structure and Dynamics*****. [In Press]**
19. Roy, R; Jonniya, N. A.; Poddar, S; **Sk, M. F.**; Kar, P*. (2021), Unraveling the molecular mechanism of recognition of human interferon-stimulated gene product 15 by coronavirus papain-like proteases: A multiscale simulation study. ***Journal of Chemical Information and Modeling***, *61(12)*, 6038-6052.

20. Singh, S; **Sk, M. F.**; Sonawane, A; Kar, P; Sadhukhan, S*. (2021), Plant-derived natural polyphenols as potential antiviral drugs against SARS-CoV-2 via RNA dependent RNA polymerase (RdRp) inhibition: an in-silico analysis. ***Journal of Biomolecular Structure and Dynamics***, *39(16)*, 6249-6266.
21. Jonniya, N. A.; **Sk, M. F.**; Kar, P*. (2022), A comparative study of structural and conformational properties of WNK kinase isoforms bound to an inhibitor: insights from molecular dynamic simulations. ***Journal of Biomolecular Structure and Dynamics***, *40(3)*, 1400-1415.
22. Jonniya, N. A.; **Sk, M. F.;** Kar, P*. (2021), Elucidating specificity of an allosteric inhibitor WNK476 among With-No-Lysine kinase isoforms using molecular dynamic simulations." ***Chemical Biology & Drug Design***, *98(3)*, 405-420.
23. Roy, R, **Sk, M. F.**, Jonniya, N. A.; Poddar, S.; Kar, P*. (2022), Finding potent inhibitors against SARS-CoV-2 main protease through virtual screening, ADMET, and molecular dynamics simulation studies. ***Journal of Biomolecular Structure and Dynamics***, 40(14), 6556-6568.
24. Shankar, U; Jain, N; Mishra, S. K.; **Sk, M. F**; Kar, P; Kumar, A*. (2022), Mining of Ebola virus genome for the construction of multi-epitope vaccine to combat its infection. ***Journal of Biomolecular Structure and Dynamics***, *40(11)*, 4815-4831.
25. Masand, V. H*; **Sk, M. F.**; Kar, P; Vesna, R.; Magdi E.A. Z*. (2021), Identification of Food Compounds as inhibitors of SARS-CoV-2 main protease using molecular docking and molecular dynamics simulations. ***Chemometrics and Intelligent Laboratory Systems***, *217*, 104394.
26. Pyasi, S; Jonniya, N. A.; **Sk, M. F.**; Nayak, D*; Kar, P*. (2021), Finding potential inhibitors against RNA-dependent RNA polymerase (RdRp) of bovine ephemeral fever virus (BEFV): an in-silico study. ***Journal of Biomolecular Structure and Dynamics***, 1-19.

27. Roy, R†., **Sk, M. F.**†, Tanwar, O., Kar, P*. (2022), Computational studies indicated the effectiveness of human metabolites against SARS-CoV-2 main protease. ***Molecular Diversity.*** 1-16.**[In Press]**

Table of Contents

LIST OF FIGURES ..... **xxv**

LIST OF TABLES ..... **xxxvii**

ACRONYMS ..... **xli**

NOMENCLATURE ..... **xliv**

Chapter 1 ..... **3**

Introduction and Literature Review **3**

1.1. Protein tyrosine kinase 3

1.2. Kinase function 4

1.3. The JAK/STAT signaling: from the cell surface to the nucleus 5

1.4. Janus kinase family proteins 7

1.4.1. JAK family architecture 7

1.4.2. Active and inactive JAK structure 13

1.4.3. Mechanisms of JAK activation 15

1.5. The JAK/STAT signaling and human diseases 16

1.5.1. Inflammatory disorders 16

1.5.2. Neoplastic disorders 17

1.6. Therapeutic target molecules in the JAK/STAT signaling 17

1.7. JAK/STAT signaling pathway inhibitors 19

1.7.1. Negative feedback inhibitor 19

1.7.1.1. Suppressor of cytokine signal 19

1.7.2. JAK inhibitors 19

1.7.2.1. First-generation JAK inhibitors 20

1.7.2.2. Second-generation JAK inhibitors 20

1.8. Scope and aim of the thesis 21

1.9. Organization of the thesis 23

Chapter 2 ..... **27**

Theoretical Background **27**

2.1. Molecular modeling of Proteins 27

2.1.1. Quantum mechanics method 28

2.1.2. Molecular mechanics method 29

2.2. Force fields 30
2.2.1. Popular force fields 32
2.2.2. AMBER protein force field 33
2.3. Molecular Dynamics (MD) Simulations 34
2.3.1. Fundamentals of MD simulations 35
2.3.2. Simulation time-step 38
2.3.3. Solvent model 38
2.3.3.1. Implicit solvent model 38
2.3.3.2. Explicit solvent model 40
2.3.4. Periodic boundary conditions 41
2.3.5. Statistical ensembles 42
2.3.6. Temperature and pressure coupling 43
2.3.6.1. Popular Thermostats 43
2.3.6.2. Popular barostat 44
2.3.7. Long-range interactions 45
2.4. Main steps of an MD simulation 46
2.4.1. Preparation of system topology 46
2.4.2. Minimization 46
2.4.3. Heating 47
2.4.4. System equilibration 47
2.4.5. Production run 48
2.5. Enhanced sampling method 48
2.5.1. Accelerated molecular dynamic simulations (aMD) 49
2.5.2. Gaussian accelerated molecular dynamic simulations (GaMD) 51
2.6. Molecular docking and its algorithms 54
2.6.1. Search algorithm 55
2.6.2. Scoring functions 56
2.7. Trajectory analysis techniques 56
2.7.1. Stability and flexibility analyses 57
2.7.2. Dynamic cross-correlation matrix (DCCM) 58
2.7.3. Principal component analysis (PCA) 59
2.7.4. dihedral principal component analysis (dPCA) 60
2.7.5. Potential of mean force and free energy landscapes 60

2.7.6. Protein structure network (PSN) analysis 62
2.8. Binding free energy estimations 62
2.8.1 Free energy perturbation (FEB) 63
2.8.2 Thermodynamics integration (TI) 64
2.8.3 Molecular mechanisms (generalized Born) Poisson Boltzmann surface area (MM/(GB)PBSA) 65
2.8.4. Conformational entropy calculation 67
Chapter 3 71
Unraveling the Molecular Mechanism of Recognition of Selected Next Generation Anti-rheumatoid Arthritis Inhibitors by Janus Kinase 1 71
3.1. Introduction 72
3.2. Materials and methods 74
3.2.1. Molecular docking and systems preparation 74
3.2.2. Molecular dynamics (MD) simulations 75
3.2.3. Dynamic cross-correlation map (DCCM) 76
3.2.4. Principal component analysis and Free energy landscape 76
3.2.5. Protein structure networks (PSNs) 76
3.2.6. MM-PBSA calculations and energy decomposition analysis 77
3.3. Results and discussion 77
3.3.1. Structural stability and flexibility analysis of JAK1 78
3.3.2. Ligand dynamics 80
3.3.3. Correlation and anti-correlation motions 82
3.3.4. PCA and dynamics of free energy landscapes (FEL) 83
3.3.5. Protein structure network (PSN) analysis 85
3.3.6. Energetics of inhibitor binding with JAK1 88
3.3.7. Essential residues for inhibitors binding 90
3.3.8. Hydrogen bonds and hydrophobic interactions 93
3.3.9. Atomic contact at the active site and solvent accessibility 95
3.4. Conclusions 96

Chapter 4 101

Phosphorylation-Induced Conformational Dynamics and Inhibition of Janus Kinase 1 by Suppressors of Cytokine Signaling 1 101

4.1. Introduction 102

4.2. Materials and Methods 105

4.2.1. System preparations 105

4.2.2. Simulation protocol 106

4.2.3. Post-dynamics simulations trajectory analysis 107

4.2.4. Essential dynamics analysis and free energy surface (FES) 107

4.2.5. Protein structure network (PSN) analysis 107

4.2.6. Molecular mechanics Poisson-Boltzmann surface area (MM/PBSA) calculations 108

4.3 Results and Discussions 108

4.3.1. Stability and dynamics of phosphorylated conformation of JAK1 bound and unbound to SOCS1 108

4.3.2. Conformational diversity of hydrophobic spines. 113

4.3.3. P-loop conformational dynamics 114

4.3.4. Collective motions of JAK1 and SOCS1 115

4.3.5. Conformational variability of A-loop 117

4.3.6. Characterizations of global and local concerted motion in JAK1/SOCS1 complexes 119

4.3.7. Structural protein network and community 120

4.3.8. P-loop, A-loop, and KIR region interaction mechanisms 121

4.3.9. Phosphorylated induced interaction pattern of JAK1 and SOCS1 123

4.3.10. Buried interface surface area: double versus mono phosphorylation states 124

4.3.11. Identification of Key interacting residues 125

4.3.12. Hydrogen bonds within JAK1, salt-bridges, and secondary structure propensity of A-loop 127

4.4. Conclusion 127

Chapter 5 131

Exploration of conformational dynamics of phosphorylated active and inactive Janus kinase 2 using Gaussian accelerated molecular dynamics simulations 131

5.1. Introduction 132

5.2. Materials, method, and protocols 135

5.2.1. Preparation of apo and type I/II inhibitor bound forms of JAK2 135

5.2.2. Conventional molecular dynamics (cMD) simulations 136

5.2.3. Gaussian accelerated molecular dynamics (GaMD) simulations 136

5.2.4. Simulation trajectories analysis 138

5.3. Results and discussion 139

5.3.1. Structural stability and fluctuation analysis of active and inactive JAK2 139

5.3.2. The Hydrophobic spines (C-spine and R-spine) 142

5.3.3. The kinase intradomain motion 143

5.3.4. Local motions and conformational free energy surfaces 144

5.3.5. Conformational states of the HRD motif 145

5.3.6. Conformational diversity in αC helix and DFG motif 147

5.3.7. Comparison of A-loop conformational variability 150

5.3.8. Salt-bridge interactions control the conformational switch of the A-loop 152

5.3.9. The Tyr^{1007}-Tyr^{1008} region of A-loop 153

5.3.10. Community network analysis of kinase domain 154

5.3.11. Comparative binding modes 156

5.4. Conclusions 159

Chapter 6 163

Finding Inhibitors and Deciphering Inhibitor-induced Conformational Plasticity in the Janus Kinase via Multiscale Simulations 163

6.1. Introduction 164

6.2. Materials and methods 165

6.2.1. Target protein preparation and receptor grid generation 165

6.2.2. Ligand preparation and receptor-based virtual screening 166

6.2.3. Pharmacological properties and toxicity profiles of hits compounds 167

6.2.4. Density function theory and receptor-ligand based e-pharmacophore analysis 167
6.2.5. Molecular dynamics simulations and trajectory analysis 168
6.2.6. Ligand binding affinity calculation and energy decomposition 168
6.3. Results and Discussions 169
6.3.1. Systematic virtual screening of phytochemical dataset 169
6.3.2. Predicting the pharmacological and toxicity properties of hit compounds 169
6.3.3. Local structure and binding pose refinement 170
6.3.4. Ligand chemical structural similarity and clustering 171
6.3.5. Density functional theory (DFT) and e-pharmacophore analysis 171
6.3.6. Molecular dynamics simulation and binding conformational stability analysis 173
6.3.7. Essential dynamics of protein and Free energy landscape 178
6.3.8. The conformational landscape of αC-helix and its relationship to ligand binding 179
6.3.9. Binding free energy analysis of phytochemicals against JAK1 181
6.3.9. Key residues involved in the recognition process 182
6.3.10. Hydrogen bonds pattern between the phytochemicals and JAK1 184
6.4. Conclusions 186
Chapter 7 191
Conclusions and Future Perspective 191
7.1. Overall conclusions 191
10.2. Outlook and future directions 193
Appendix A 197
Appendix B 217
Appendix C 232
Appendix D 244
Bibliography 267

LIST OF FIGURES

Chapter 1

Figure 1.1. The schematic representation of the function of a protein kinase. Here, P represents the phosphoryl group (PO_3^{2-}) group. 4
Figure 1.2: The JAK/STAT signaling transduction pathway by type I and type II cytokines. Figure reprinted from Perez-Jeldres et al. (Ref[9]). 6
Figure 1.3: (A) Schematic of JAK structural domains. There are four main domains and seven JAK homology (JH) in the JAKs structure. The conserved tyrosine phosphorylation sites in JAK1 are Y^{1034}/Y^{1035}; the conserved tyrosine phosphorylation sites in JAK2 are Y^{1007}/Y^{1008}; the conserved tyrosine phosphorylation sites in JAK3 are Y^{980}/Y^{981}; the conserved tyrosine phosphorylation sites in Tyk2 are Y^{1054}/Y^{1055}. (B) Schematic of a cell-surface, ligand-induced cytokine receptor-JAK dimer. This figure is inspired by Glassman et al. (Ref[55]). 8
Figure 1.4: Structure of a typical protein kinase (PDB ID code 3EYG) domain displaying ATP binding site and conserved elements around it. 9
Figure 1.5: Alignment and identity table of amino acid sequences of three kinds of JAKs (whole sequences and conserved regions). The amino acid sequences of JAK1 (3EYG), JAK2 (4D0X), and JAK3 (3LXK) were compared using the CLUSTALW[56]. 10
Figure 1.6: Overview of the conformational variability of the catalytic domain (kinase domain). In **e),** the active conformation (based on PDB structure 2F4J, residues 235–498) is shown. The flexible structural elements are shown in color: the P-loop is shown in blue, the αC is shown in orange, and the A-loop is shown in green. The DFG motif is shown in the "licorice" representation and colored in red (Asp), violet (Phe), and white (Gly). The highlighted structural elements can adopt different conformations: **b)** The P-loop is mainly in the kinked (blue) or extended (cyan) conformation. **c)** αC conformations range between αC-in (orange) and αC-out (yellow) positions. **d)** The DFG motif is mainly in the DFG-in (or simply in, ball-and-stick representation) or the DFG-out (or simply out, "licorice" representation) state. **a)** The A-loop adopts many conformations that range between closed, inactive (emerald green) and extended, active (lime green). Figure reprinted from Fabian Paul et al. (Ref[73]). 14
Figure 1.7: Therapeutic targets of the JAK/STAT signaling pathway. . **(1)** Recombinant cytokines, cytokine antibodies, and receptor antibodies are designed to target cytokines or receptors; **(2)** JAK inhibitors are designed to target JAKs; and peptide inhibitors including JAKs natural inhibitor SOCS, **(3)** small-molecule inhibitors are designed to target the STAT proteins and inhibits the dimerization of STATs. 18

Chapter 2

Figure 2.1: Schematic representation of the spatio-temporal resolution of various molecular modeling techniques. The figure is inspired by Aminpour el al[111]..... 28
Figure 2.2: The stick representation of a molecules along with all bonded and nonbonded interactions component of the force field. Where $\mathbf{V_{LJ}}$, $\mathbf{V_{dih}}$ and $\mathbf{V_{Coul}}$ represented van der Waals, dihedrals, and electrostatics potentials, respectively. 32
Figure 2.3: Time evolution of protein forcefields. For simplicity only AMBER, CHARMM and OPLS are listed. 33
Figure 2.4: TIP3P widely used 3-point water model including key parameters, i.e., bond distance and HOH angle. The figure is inspired by ref.[149,150]..... 40
Figure 2.5: Schematic 2D representation of periodic boundary conditions with the simulation box (grey) surrounded by image boxes. The velocity of every atom is depicted by its corresponding vector. Short-range interactions are also defined by r_{cut}. 41
Figure 2.6: The schematic representation of the aMD method. When the original potential $V(r)$ (solid line) falls below threshold value E, a boost potential is added. The modified potential $V*(r)$ in color representation (dashed line) with different values of the acceleration factor α. The larger the value of α, it is closer to the original surface. Largest α in green color, and smaller α in brown color. It is adopted from Hamelberg et al. [182]. 50
Figure 2.7: Scheme illustration of GaMD. When the threshold energy is set to the maximum potential (E= V_{max}), the system potential energy surface is smoothened by adding a harmonic boost potential that follows the Gaussian distribution. The coefficient $\kappa 0$ in the range of 0-1 determines the magnitude of the applied boost potential. With greater $\kappa 0$, higher boost potential is added to the original energy surface in conventional MD, which provides an enhanced sampling of biomolecules across decreased energy barriers. It is adapted from Miao et al. [11]..... 51

Chapter 3

Figure 3. 1: (**A)** Typical active Janus kinase 1 (JAK1) structure with different vital segment color-coded, **(B)** schematical image corresponding JAK1 structure, and (C-F) 3-dimensional structures of inhibitors with each heavy atoms name: Baricitinib, Filgotinib, Itacitinib, and Upadacitinib, respectively. 73
Figure 3. 2: Evolution of JAK1 (A) protein backbone, (B) ligand-binding pocket backbone atoms root-mean-square deviations (RMSDs) over simulation time along with the apo system. The residual root-mean-square fluctuations (RMSFs) are shown in (C). Differents binding areas are highlighted with a dotted line rectangle box. 80
Figure 3. 3: The time evolution of ligand heavy atoms RMSD, (A); and potential of mean force (PMF) for all four inhibitors (B), their RMSD, took as a reaction coordinate. The PMF is obtained by calculating the free energy profile along with their respective RMSD values of inhibitors..... 82
Figure 3.4: Cross-correlation matrices of the coordinates' fluctuations for Cα atoms around their mean positions after the equilibrium of production simulation and shows the relative

motions of protein residues. The extent of correlated motions and anti-correlated motions are color-coded. (A) the apo system, (B) JAK1/Baricitinib, (C) JAK1/Filgotinib, (D) JAK1/Itacitinib, and (E) JAK1/Upadacitinib. 84
Figure 3. 5: Free energy landscape (FEL) analysis of five systems. (A) the apo system, (B) JAK1/Baricitinib, (C) JAK1/Filgotinib, (D) JAK1/Itacitinib, and (E) JAK1/Upadacitinib. The FELs were obtained using the projections of JAK1 C_{α}-atoms position vectors onto the first two principal components as reaction coordinates. As indicated by the color bar, free energy values calculated from the density of distribution are given in kcal/mol. 86
Figure 3.6: The hubs and community structures of residue contact networks in (**A**, **F**) unbound Apo, (**B**, **G**) Baricitinib, (**C**, **H**) Filgotinib, (**D**, **I**) Itacitinib and (**E**, **J**) Upadacitinib bound with Jak1, respectively. The hubs are represented as spheres centered on the Cα atoms on the Jak1 protein. Community structure describes a fundamental feature of residue contact networks. Colored according to the community nodes, the largest community-ranked 1 is red. 87
Figure 3.7: Decomposition of the binding free energy for the (A) JAK1/Baricitinib, (B) JAK1/Filgotinib, (C) JAK1/Itacitinib, and (D) JAK1/Upadacitinib, respectively. Residues with a free energy of more than 1 kcal/mol are labeled here, and the color diagram also shows their actual position in the JAK1 structure. 91
Figure 3.8: Time evolution of the number of atomic contacts within 3.5 Å of each ligand during the last 200-ns simulations (left panel). The solvent-accessible surface area (SASA) of the residues within 5 Å of each competitive inhibitor (right panel). 96

Chapter 4

Figure 4.1: The JAK/STAT pathway mechanisms and the production and action of SOCS proteins. The green line indicates promotion, the red line denotes inhibition, and P (in the pink star) indicates phosphorylated. **(B)** The X-ray crystal structure of JAK1 complexed with potent SOCS1 (PDB ID: 6C7Y) with their different essential regions. The glycine-rich loop is colored in light green, the hinge region in dark cyan, the catalytic domain in yellow, the activation loop in light cyan, αC-helix in dark orange, GQM motif in dark red, and the bound ADP is presented in a ball-and-stick representation with carbon atoms in gray, oxygen atoms in red, nitrogen atoms in blue and phosphorous atoms in orange. The SOCS1's KIR region is colored in deep pink, the ESS region in royal blue, and the BC loop in teal color. 103
Figure 4.2: **(A)** The distribution of the backbone RMSD of three complexes corresponding to the three different phosphorylation states. The backbone RMSD's distribution of JAK1 and SOCS1 in complex form ($Jak1_{com}$/$SOCS1_{com}$) and apo form ($Jak1_{apo}$/$SOCS1_{apo}$) for three different A-loop phosphorylated systems, **(B)** double phosphorylated, 2pY, **(C)** single phosphorylated, pY1034, and **(D)** single phosphorylated, pY1035, respectively. The backbone RMSD's distribution of αC-helix, DFG motif, glycine-rich P-loop, GQM motif, regulatory spine (R-spine), and activation loop of three different phosphorylated states of JAK1 in complex with and without SOCS1 (apo) forms. **(E-G)** complex, phosphorylated systems and **(H-J)** apo system, respectively. 109

Figure 4. 3: The free energy landscape of JAK1/SOCS1 complex protein measured by the backbone RMSD of JAK1 and SOCS1 for three different phosphorylated states. (A) One dimensional free energy (potential mean force, PMF) for three complexes and JAK1 RMSD as reaction coordinates. (B, C, and D) Two-dimensional free energy landscape of three phosphorylated complexes. 111
Figure 4.4: The conformational landscape of pY1035 JAK1 complex with SOCS1. **(A)** FES as a function of the pseudo torsion ζ(Gly1020) and distance R between K^{908}:NZ and E^{925}:CD. ζ(Gly1020) is defined using the CA atoms of I^{1019}, G^{1020}, D^{1021}, and F^{1022}. Horizontal and vertical lines are plotted at 120° and 4.5 Å to define cut-off values. (B) The αC-helix in/out structure is shown in blue and red color, and the starting crystal structure is shown in turquoise color. (**C-E**) Representative snapshots represented the FES minima. K^{908}, E^{925}, D^{1021}, and F^{1022} are shown. 112
Figure 4.5: The time evolution backbone RMSD of regulatory spines (R-spine) of com-2pY, com-pY1034 and com-pY1035, respectively. The hydrophobic catalytic spine (C-spine) and R-spine shown atom sphere model in dark orange and red colour in the corresponding active (R-spine aligned) and inactive (R-spine broken) conformations of JAK1. 114
Figure 4.6: The two-dimensional free energy landscape of three phosphorylated states JAK1/SOCS1 complexes regarding RMSD of P-loop and center of mass distance between conserved HRD motif and P-loop region. The different conformations and their corresponding representative structures are shown in (**A, D**) com-2pY; (**B, E**) com-pY1034, and (**C, F**) com-pY1035, respectively. 116
Figure 4.7: The 2D free energy landscape of com-2pY, com-pY1034 and com-pY1035 system is shown in panel **(A-C)** obtained from the A-loop dihedral PCA and **(D-F)** from the reaction coordinates of A-loop backbone RMSD and C.o.M distance between A-loop and KIR region, respectively. The local minima conformations were highlighted, and their representative structures were shown on their right panel. The crystal structure of $pTyr^{1035}$ conformation is also shown in red color. 118
Figure 4.8: The network hubs and community maps of residue contact networks in (**A**) com-2pY, (**B**) com-pY1034, and (**C**) com-pY1035, respectively. The hubs are represented as spheres centered on the Cα atoms on the Jak1 protein. Community structure describes a fundamental feature of residue contact networks. Coloured according to the community nodes, the largest community-ranked 1 is red. 122
Figure 4.9: The sandwiched-like structure and **(B)** the probability density function of P-loop (His^{885}), A-loop (Pro^{1044}), and KIR (His^{54}) angle for three phosphorylated JAK1/SOCS1 complexes. 125
Figure 4.10: (**A-C**) The free energy contribution of each residue to the JAK1/SOCS1 binding. (**D-F**) Details of the interactions between JAK1 and SOCS1 in the final production simulation. Black dashed lines indicate hydrogen bonds. 126

Chapter 5

Figure 5.1: The three-dimensional structures (ribbon representation) of (**A**) CP-690550 (ATP-competitive inhibitor, CI; PDB Id; 3FUP) bound active and (**B**) BBT594 (allosteric inhibitor, AI; PDB Id: 3UGC) bound inactive JAK2. Locations of the hinge, P-loop, αC-helix, DFG-motif, catalytic loop, and A-loop are indicated by color-coded. 132
Figure 5.2: (A) The backbone atoms RMSD probability density of three systems; (B-C) the probability density distributions of the activation loop (A-loop) and DFG motif of JAK2 protein. (D-F) The backbone Cα-atoms average RMSF of three sets of MD trajectories (black) and their respective standard deviation (SD) of each fluctuation (red line). 140
Figure 5.3: The time-evolution of backbone atoms RMSD values of hydrophobic (A) catalytic spine and (B) regulatory spines of apo, CI-bound, and AI-bound JAK2, respectively. Their corresponding structural alignment is also shown below. 142
Figure 5.4: (A-E) The pseudo-dihedral-based free energy surfaces of HRD motif Arg^{975}. Here we considered the last 500 ns trajectories from every three replicas for all systems to calculate the free energy. **(F)** salt-bridge between Arg^{975} (HRD) and Asp^{1004} (A-loop) interactions in apo, CI-bound, and AI-bound JAK2. 147
Figure 5.5: (A-B) The initial conformations of JAK2 bound with CI and AI. (C) The conformational space is sampled by three systems (red: apo; yellowgreen: CI and darkslategrey: AI) along with the pseudo-torsional angle (ζ) and distance between β3 (Lys^{882}) and αC-helix (Glu^{898}) residues. (D-I) The corresponding representative conformations of active and inactive JAK2 structures. 149
Figure 5.6: Two-dimensional free energy surfaces underlying the dynamics of critical structural motifs. The X-axis represents the allosteric pocket opening/closing by the distance between Phe^{995}@CZ (DFG motif) and Leu^{902}@CA (αC) residues. The Y-axis represents the closeness of DFG-Phe^{995} to N-lobe β-sheet strands by the distance between Phe^{995}@CZ (DFG motif) and Lys^{882}@CA (β3) residues..... 150
Figure 5.7: The two-dimensional free energy surface (FES) obtained from the dihedral principal component analysis (dPCA) of the A-loop for (**A**) apo and (**B, C**) inhibitors (CI/AI) bound JAK2, respectively. Each local minimum is highlighted and numbered (C1 to C3) with increasing energy. The K-means cluster analysis shows their representative structures on the middle panels. 151
Figure 5.8: The conformational space sampled by three systems and the dihedral angles (φ/ψ) from phosphorylated Tyr^{1007} and Tyr^{1008} and their corresponding snapshots. 154
Figure 5.9: Network hubs and community of residue contact network in (**A, D**) apo; (**B, E**) CI-bound, and (**C, F**) AI-bound system, respectively. The hubs are indicated as spheres centered on the Cα-atoms on the protein. The community network describes a fundamental feature of residual contact networks. Each color represents an individual community, and lines connecting different color spheres visualize the inter-community connections, and the thickness of these lines is proportional to the corresponding edge connectivity. 156
Figure 5.10: (**A**) Average per-residue energy contribution and the location of these residues in the JAK2 structure; highlighted residues are contributed lower than -1.0 kcal/mol. (**B-C**) JAK2 and essential areas are shown color-coded with CI and AI bound

states. **(D)** The binding free energy contribution of some important competitive and allosteric inhibitor binding sites. 159

Chapter 6

Figure 6.1: Workflow of identifying best lead molecules against JAK1 via systematic structure-based virtual screening, ADMET profiles, and molecular dynamics simulations. 166
Figure 6.2: (A) The probability of density function (PDF) plots of the backbone RMSD of target receptor protein JAK1 complexed with the phytochemicals. (B) The three-dimensional crystal structure of receptor protein, JAK1, highlighted the crucial binding residues, including P-loop, Hinge, αC-helix, catalytic, and activation loop (A-loop). (C) The probability of the center of the mass distance between the ATP-binding pocket and phytochemicals during two replica MD simulations. (D) The residue-wise C_α-atom RMSF of JAK1. 174
Figure 6.3: The dynamic cross-correlation map of JAK1 receptor (Cα-atoms): **(A)** JAK1/Epicatechin, **(B)** JAK1/Leonurine, **(C)** JAK1/STOCK1N-68642, **(D)** JAK1/STOCK1N-82656, **(E)** JAK1/STOCK1N-85809, and **(F)** JAK1/STOCK1N-98615. The blue (-1 to 0) and red (1 to 0) colors correspond to the anti-correlated and correlated motions of the atoms, and the white color suggests no correlation map among the residues. The red and blue lines in the 3D structures represent the strong correlation and anticorrelation motion, respectively. 176
Figure 6.4: The two-dimensional free energy landscapes generated by projecting the first two principal components, PC1, and PC2, for **(A)** JAK1/Epicatechin, **(B)** JAK1/Leonurine, **(C)** JAK1/STOCK1N-68642, **(D)** JAK1/STOCK1N-82656, **(E)** JAK1/STOCK1N-85809, and **(F)** JAK1/STOCK1N-98615. The representative structures of corresponding minima compared to the X-ray structure are shown on the left panel. 177
Figure 6.5: The conformational landscape of JAK1 complexed with **(A)** Epicatechin, **(B)** Leonurine, **(C)** STOCK1N-68642, **(D)** STOCK1N-82656, **(E)** STOCK1N-85809, and **(F)** STOCK1N-98615. Free energy landscape (FES) as a function of the pseudo torsion ζ(Gly1020) and distance R between Lys^{908}:NZ and Glu^{925}:CD. ζ(Gly1020) is defined using the CA atoms of Ile^{1019}, Gly^{1020}, Asp^{1021}, and Phe^{1022}. Vertical lines are plotted at 4.5 Å to define cut-off values for αC-helix in/out structure. Representative snapshots represented the FES minima. Lys^{908}, Glu^{925}, Asp^{1021}, and Phe^{1022} are shown. 180
Figure 6.6: The probability of helical content of αC-helix residues calculated in the combined trajectory for all six complexes and their corresponding representative structures. 183
Figure 6.7: The highlighted residues are the crucial residues of JAK1 from the P-loop, hinge, and catalytic area, which take part in the phytochemical binding affinity. The shaded surface represents the electrostatic surface area of JAK near the ATP-competitive inhibitor binding pocket, and phytochemicals are described in the ball and stick model in the binding cavity of JAK1. 184

Figure 6.8: Percentage of hydrogen bond occupation with the essential residues in the hinge, P-loop, and catalytic loop region of JAKs. The data were derived from the last 300 ns of the two simulations. 185

Appendix A

Figure A1: The position of each regulatory spine, catalytic spine, and R-shell in JAK1 structure. R-spine residues are Tyr^{940}, Leu^{929}, Phe^{1022}, His^{1001} and D^{1063}; R-shell residues are Leu^{954}, Met^{956} (gatekeeper residue) and Val^{938}; C-spine residues are Val^{889}, Ala^{906}, Leu^{964}, Val^{1009}, Leu^{1010}, Val^{1011}, Thr^{1070} and Leu^{1074}. 203
Figure A2: Evolution of salt bridge distance K908@NZ and E925@CD over simulation time. 203
Figure A3: The time evolution of backbone RMSD of the JAK1/ligand complexes, including the apo system. The data are derived from the replica simulation (**Run2**). ... 204
Figure A4: The superimposed structures of initial and final conformations of (A) Baricitinib, (B) Filgotinib, (C) Itacitinib, and (D) Upadacitinib complexed with JAK1.204
Figure A5: Time dependence root-mean-square deviations (RMSDs) of C_α atoms relative to their energy minimized and equilibrated structures of (A) P-loop, (B) hinge region, (C) catalytic region, (D) activation segment, (E) C-spine residues, (F) R-spine residues. ... 205
Figure A6: The time evolution of the radius of gyration of the four complexes during MD simulation. 205
Figure A7: The time evolution of the solvent accessible surface area of (A) protein/ligand and (B) ligands. 206
Figure A8: Superposition of JAK1 (light blue) and JAK2 (light pink, PDB:6VN8) JH1 domain complexed with Baricitinib. The RMSD of Baricitinib is 1.18 Å. 206
Figure A9: (Left) Two-dimensional free energy landscape of ligand movements. The ligand RMSD and radius of gyration of each ligand are taken as reaction coordinates. (Right) Ligand movements scatter plot of 20000 snapshots from MD simulation along RMSD and radius of gyration of ligand and plotted against simulation time. 207
Figure A10: Eigenvalues as a function of the eigenvectors. Here we calculated first 20 eigenvectors and its corresponding eigenvalues of covariance matrices of PCA. 207
Figure A11: Porcupine plots of the first PC motion modes of (A) apo, (B) JAK1/Baricitinib, (C) JAK1/Filgotinib, (D) JAK1/Itacitinib, and (E) JAK1/Upadacitinib models. The protein is shown in tube representation; the dark red color corresponds to αC helix, light green is the P-loop region, dark orange is the catalytic loop, the blue color corresponds to the hinge region, and dark magenta represents activation segment. The prominent motions are shown in the yellow color corresponding to the direction of the eigenvector, and the magnitude of length represents the eigenvalues. 208
Figure A12: PCA scatter plot of last 17000 snapshots from MD simulation along with first two principal components and plotted against simulation time for (A) apo, (B) JAK1/Baricitinib, (C) JAK1/Filgotinib, (D) JAK1/Itacitinib and (E) JAK1/Upadacitinib, respectively. 209

Figure A13: Energy components (kcal/mol) for the binding of inhibitors to the JAK1. $\Delta E_{vdW,}$ van der Waals energy; $\Delta E_{ele,}$ electrostatics energy in the gas phase; $\Delta G_{pol,}$ polar solvation energy; $\Delta G_{np,}$ non-polar solvation energy; TΔS, configurational entropy contribution. 210
Figure A14: The key residue interactions for JAK1 kinase with four inhibitors. (A) The total binding contributions (ΔG_{Total}); (B) the total polar contributions ($\Delta G_{ele+pol}$); (C) the total non-polar contributions (ΔG_{vdW+np}). 211
Figure A15: Energy decompositions of the different important regions of JAK1 and inhibitors complexes. (**A**) The glycine-rich, P-loop; (**B**) hinge region; (**C**) catalytic loop; (**D**) activation segment. 212
Figure A16: The final (200 ns) snapshot of each system (A) JAK1/Baricitinib, (B) JAK1/Filgotinib, (C) JAK1/Itacitinib, and (D) JAK1/Upadacitinib, the corresponding inhibitors are shown in ball and stick representations. (E) The hydrogen bond interaction pattern of JAK2 with Baricitinib. The protein is in new cartoons and with essential regions, and hydrogen-bonded residues are labeled, and bonds are shown in the black dotted line. 213
Figure A17: The hydrogen-bonded distance calculations from the simulations between the JAK1 and corresponding ligands, i.e., (A) Baricitinib, (B) Filgotinib, (C) Itacitinib, and (D) Upadacitinib, respectively..... 214
Figure A18: The inter-molecular interaction of JAK1 with (A) Baricitinib, (B) Filgotinib, (C) Itacitinib, and (D) Upadacitinib. 215
Figure A19: The hydrophobicity surface of the binding pocket of JAK1 for four different inhibitors. (A) JAK1/Baricitinib; (B) JAK1/Filgotinib; (C) JAK1/Itacitinib; and (D) JAK1/Upadacitinib. 216

Appendix B

Figure B1: The time evolution of backbone RMSD of JAK1/SOCS1 complexes with three different phosphorylated states. 219
Figure B2: The probability of density function (PDF) plots based on backbone RMSD of catalytic, hinge and catalytic spine (C-spine) of JAK1 in complexed with SOCS1 and apo conformation. 220
Figure B3: The probability of density function (PDF) plots based on backbone RMSD of KIR, ESS, BC-loop, αA-helix, EF-loop, αB-helix, and BG-loop SOCS1 in complexed with JAK1 and apo conformation. 220
Figure B4: Residue-wise C_α RMSF of apo and complex form of JAK1, is plotted for apo (magenta) and complex (light blue). A scatter plot of C_α RMSF of residues in complex is plotted against that of the equivalent residues in apo in **(A**) the double phosphorylated A-loop, 2pY, **(B)** single phosphorylated A-loop, pY1034, and **(C)** single phosphorylated A-loop, pY1035. The blue diagonal indicates unity slope, across which the points are expected to be equally distributed if there is no systematic difference in Cα RMSF of residues in complex and apo Jak1 (Chi-square test, p-value < 0.05)..... 221

Figure B5: Residue-wise C_α RMSF of the apo and complex form of SOCS1 is plotted for apo (magenta) and complex (light blue). A scatter plot of C_α RMSF of residues in the complex is plotted against that of the equivalent residues in apo in **(A**) the double phosphorylated A-loop, 2pY, **(B)** single phosphorylated A-loop, pY1034, and **(C)** single phosphorylated A-loop, pY1035. The blue diagonal indicates a unity slope, across which the points are expected to be equally distributed if there is no systematic difference in Cα RMSF of residues in complex and apo SOCS1 (Chi-square test, p-value < 0.05). **(D)** the three-dimensional structure of SOCS1 and **(E)** RMSF values of KIR region residues.. 222
Figure B6: The superimposed final snapshots of three phosphorylated JAK1/SOCS1 complex in three replicas, **(A)** run1, **(B)** run2 and **(C)** run3, respectively..... 223
Figure B7: Dynamic cross-correlation matrices (DCCM) of the fluctuations of Cα atoms in com-2pY **(A)**, com-pY1034 **(B)** and com-pY1035 **(C)** JAK1/SOCS1 complexes. The negative, blue color bar represents the negative correlation, and the positive, dark pink color bar indicates positively correlated motions in protein..... 223
Figure B8: Porcupine illustration of the PC1 and PC2 and percentage of the total variance captured by each eigenvector, with numbers indicating the cumulative variance accounted for all preceding eigenvectors for three phosphorylated states of JAK1/SOCS1 complex; (**A - C**) for com-2pY, (**D - F**) for com-pY1034 and (**G - I**) for com-pY1035, respectively. The blue arrow represents the direction of motion; the length characterizes the amplitude of motion. The highlighted areas are: red (αC-helix), magenta (activation segment), pink (hinge loop), yellow (catalytic domain), black (GQM motif), green (P-loop), and orange (SOCS1)..... 224
Figure B9: Two-dimensional free energy landscape from MD trajectories for **(A)** com-2pY, **(B)** com-pY1034, and **(C)** com-pY1035 JAK1/SOCS1 along with the lowest energy conformations at 300 K. The first two principal components (PC1 and PC2) were used as reaction coordinates, and free energy values are given in kcal/mol as indicated by the colour bar. The time evolution of the first two principal components for **(D)** com-2pY, **(E)** com-pY1034, and **(F)** com-pY1035 is also shown on its right panel. 225
Figure B10: The 2D free energy potential surface based on pseudo-dihedral angles, θ1 and θ2. Where, θ1 and θ2 defined separately for H54 (θ1: N, C_α, C_β, Cγ; θ2: C_α, C_β, C_γ, ND1), H885 (θ1: N, C_α, C_β, Cγ; θ2: C_α, C_β, C_γ, ND1) and P1044 (θ1: O, C, C_α, N; θ2: C, C_α, N, C_γ, ND1) amino acid. Residue H54, H885, and P1044 belong to SOCS1's KIR, JAK1's P-loop, and A-loop regions, respectively. 226
Figure B11: The probability density and time evolution of the buried surface area of JAK1/SOCS1 complexes..... 226
Figure B12: Protein-protein interaction diagrams for the JAK1/SOCS1 complexes. The upper panel corresponds to SOCS1, and the lower panel corresponds to JAK1 in the complex. The plot was generated with the dimplot module of LigPlot+. Hydrogen bonds are shown as a green dotted line, and hydrophobic residues are represented in red..... 227
Figure B13: The probability density function of HBs distances for Cys^{111}, Lys^{109}, Phe^{58}, Pro^{83} and Thr^{57} with JAK1 obtained by analysis of combined 3 μs trajectory. (**A**) com-2pY; (**B**) com-pY1034 and (**C**) com-pY1035, respectively. 228

Figure B14: The probability density function of HBs distances for Asp^{105} and Phe^{55} (BC-loop and KIR of SOCS1) with JAK1 obtained by analysis of combined 3 μs trajectory. (**A & D**) com-2pY; (**B & E**) com-pY1034 and (**C & F**) com-pY1035, respectively..... 229
Figure B15: The probability density function of HBs distances for Arg^{56} and Arg^{59} (KIR of SOCS1) with JAK1 obtained by analysis of combined 3 μs trajectory. (**A & D**) com-2pY; (**B & E**) com-pY1034 and (**C & F**) com-pY1035, respectively..... 230
Figure B16: The time evolution of total number of internal hydrogen bonds formed by the amino acid residues of JAK1 and SOCS1. The probability density of the salt-bridges formed by phosphorylated two tyrosine residues shown in right panel..... 231
Figure B17: The time-dependent behavior of secondary structure prediction of A-loop segment of JAK1. The probability density of compactness (radius of gyration) of the A-loop segment. 231

Appendix C

Figure C1: Backbone atoms root mean square deviation (RMSD) plot of the apo, competitive inhibitor (CI) bound, and allosteric inhibitor (AI) bound complexes. The data are derived from the three independent simulations with different initial velocities. 234
Figure C2: The probability of RMSD distributions of (A) P-loop, (B) αC helix backbone atoms of all residues. Their probable RMSD values are shown in dotted lines with their corresponding color code. 234
Figure C3: The one-dimensional potential of mean force (PMF) of (**A**) catalytic and (**B**) regulatory hydrophobic spines. Here the PMFs are as a function of the RMSD of regulatory and catalytic spines relative to their energy minimized structures. All heavy atoms in the spines are used in this calculation of RMSD. The "On" and "Off" state corresponding to the hydrophobic spines formed and broken conformations. 235
Figure C4: The dynamic cross-correlation residue motion of three systems. The upper triangle matrix corresponds to CI and AI bound JAK2, and the lower triangle matrix corresponds to apo-JAK2. The dotted line box denotes the N-terminal residues, and the second box represents the A-loop residues. 235
Figure C5: The two-dimensional free energy surfaces generated by projecting the first two principal components (PC1 and PC2) as a reaction coordinate during simulations, shown in left panels. The low energy conformations are highlighted (C1, C2, C3), and their superimposed representative structures are shown on middle panels by the K-means clustering analysis. The PCA porcupines of eigenvector 1 map on the 3D-structures shown in the right panels. (A-B) for apo, (C-D) CI-bound, and (E-F) AI-bound JAK2 kinase. 236
Figure C6: (**A**) The one-dimensional potential of mean force (PMF) as a function of HRD motif heavy atoms RMSD. (**B**) The probability density of the backbone bond angle (N-Cα-C) of Arg^{975}. (**C**) The probability density of the distance between His^{974} (HRD motif) amide hydrogen atom and Asp^{1036} (αF helix) side-chain oxygen atom. (**D**) Probability density of the distance between Arg^{975} amide hydrogen atom and Asp^{1036} side-chain oxygen atom. Since the two-oxygen atom in the sidechain of Asp1036 is equivalent chemically, distances

from the hydrogen atom to both oxygen atoms are estimated, and the smaller one is used to make these probability density plots. **(E)** The representative structure of hydrogen bond formation. 237
Figure C7: Time dependence of the dihedral angle of Arg^{975} in apo, CI-bound, and AI-bound JAK2 kinase. 238
Figure C8: The two-dimensional free energy surfaces as a function of backbone atoms RMSD and radius of gyration (RoG) of A-loop segment of JAK2 in (A) apo, (B) CI-bound, and (C) AI-bound inhibitors, respectively. Local minima are highlighted in C1, C2, and C3. 238
Figure C9: Interactions enhanced the particular rearrangement of the A-loop of JAK2 upon binding of type I/II inhibitors. (A) salt-bridge between Asp^{994} (DFG) and Lys^{999} (A-loop). 239
Figure C10: The interaction pattern and residues involved in the open/extended conformations of A-loop segment in active-like apo and CI-bound JAK2. The hydrogen bonds and salt bridges are shown as dotted lines. 239
Figure C11: The residual interactions pattern of activation loop and its nearby residues to maintain the unique conformations of closed A-loop in inactive-like JAK2. These conformations are obtained from the K-means clustering analysis of dPCA eigenvectors of inactive DFG-out AI-bound JAK2 kinase during GaMD simulations. The hydrogen bonds and salt bridges are shown as dotted lines. 240
Figure C12: The time evolution of the dihedral angles of phosphorylated residues (Tyr^{1007}/Tyr^{1008}), the radial distribution functions (RDF, g(r)) of water accessibility around the residues and their solvent accessible surface area (SASA). 241
Figure C13: Network comparison of apo, CI-bound, and AI-bound JAK2. (A-C) shows the number of hubs; (D-F) the number of links in the structure; (G-I) the number of linked nodes in the systems. The first row represents the difference between the apo and CI-bound complex, the middle row for apo and AI-bound JAK2, and the below row is CI-bound and AI-bound JAK2. 242
Figure C14: The per-residue binding free energy decomposition of (A) CI-bound and (B) AI-bound JAK2 systems of three independent replica runs. 243

Appendix D

Figure D1: Superimposing the redocked ligand with co-crystallized ligand MI1. The co-crystallized ligand is drawn in magenta. The ligands as docked with Glide are in green. 253
Figure D2: The interaction pattern of the top 10 hit molecules with the ATP-binding pocket of the JAK1 receptor. All plots were generated by a widely used LigPlot++ software. 254
Figure D3: The similarity scores between compounds calculated in the Similarity Workbench interface of ChemMine Tools, calculating the atom pair and maximum common substructure (MCS) similarities with the Tanimoto coefficient. 255

Figure D4: Multidimensional (2D) scaling of the chemical structures of the dataset reveals high structural diversity. The axes (V1 and V2) represent chemical space in two directions. 256
Figure D5: Frontier molecular orbitals (highest occupied molecular orbitals, HOMO and lowest unoccupied molecular orbitals, LUMO) surface of six lead phytochemical molecules. The conformers were analyzed via the Lee-Yang-Parr correlation function (B3LYP) and Beche's three-parameter exchange potential using 6-31G** basic set level. The color-coding region was red, indicating the most negative potential region, and blue indicating the most positive region of the molecules..... 257
Figure D6: Structure-based e-pharmacophores generated using JAK1 complex with phytochemicals. The phytochemicals are shown in green sticks (orange ring: aromatic ring, blue ball: hydrogen bond donor, pink ball: hydrogen bond acceptor). 258
Figure D7: The time evolution of the root mean-square deviations (RMSD) of backbone atoms of each JAK1 complexed with the top six lead molecules. 259
Figure D8: The ligand and protein (binding pocket residues only) center of mass distance with respect to simulation time. 260
Figure D9: The RMSD distribution of (A) ligand heavy atoms, backbone atoms of (B) P-loop, (C) Hinge region, (D) Catalytic loop, (E) Activation loop, (F) DFG motif, (G) catalytic spines and (H) regulatory spines. 261
Figure D10: The solvent accessible surface area probability density distribution of six complexes during the combined two replicas. 262
Figure D11: The eigenvalues plotted against the corresponding eigenvector indices from the Cα atoms covariance matrix constructed form the last 500 (2x250) ns trajectory of all six systems. 262
Figure D12: The per-residue energy decomposition plot of phytochemicals and JAK1 complexes. Here, we only considered the ATP-binding area of the JAK kinase and colour shaded areas represented the different important binding area. 263
Figure D13: Number of hydrogen bonds formed between phytochemicals and JAK1 along 1.0 μs (2X500 ns) simulation. 264

LIST OF TABLES

Chapter 1

Table 1.1: Important residues in the human Janus kinases. This table is reprinted from Ref[42]. 11
Table 1.2: Spine and shell residues of human JAK JH1 domains. This table is reproduced from Ref[42]. 12

Chapter 3

Table 3.1: The average RMSD during the last 100 ns production simulations (**Run1 & Run2**) of the JAK1-inhibitor complexes and standard deviations (SD) in the parentheses. All values are in the Å unit. 79
Table 3.2: Energetic components of the binding free energy for JAK1-inhibitors complexes in kcal/mol.($^{a}\Delta G_{solv} = \Delta G_{np} + \Delta G_{pol}$, $^{b}\Delta G_{pol\ +\ elec} = \Delta E_{elec} + \Delta G_{pol,}$ $^{c}\Delta E_{internal} = \Delta E_{vdW} + \Delta E_{elec}$, $^{d}\Delta S_{=}$ configuration entropy, $\Delta G^{'e} = \Delta E_{internal} + \Delta G_{solv}$.)..... 88
Table 3.3: The hydrogen bonds formed between JAK1 and inhibitor and the corresponding average distance and percent determined using the simulation trajectories of run1 and run2. 94

Chapter 4

Table 4.1: List of simulated systems and simulation lengths. 106
Table 4.2: Energetic components of the total binding free energy of SOCS1 bound to the JAK1. All values are given in kcal/mol. The standard error of the mean is provided in parentheses. 124

Chapter 5

Table 5.1: The energetics components of the total binding free energy for CI-bound and AI-bound JAK2 complexes in kcal/mol. 157

Chapter 6

Table 6.1: The six lead molecules' electrostatic potential (ESP), solvation energy, HOMO, and LUMO energy were generated by density functional theory (DFT) analysis in the Jaguar module of the Schrodinger suite. 172
Table 6.2: Average energetic components of the binding free energy for phytochemicals bound to the JAK1. All values are given in kcal/mol. 182

Appendix A

Table A1: The average backbone RMSD values (in Å unit) at four different segments of the last 100 ns trajectory (four equal segments of 25 ns trajectory) along their respective standard deviation (SD) values. 197
Table A2: The average RMSD during the last 100 ns production simulations (**Run1 & Run2**) of the JAK1-inhibitor complexes and standard errors of the mean (SEM) in the parentheses. 197
Table A3: Average radius of gyration (R_g), protein solvent accessible surface area ($SASA_{protein}$), and ligand SASA ($SASA_{ligand}$) for last 100 ns trajectories **(Run1 & Run2**). Standard errors of the mean are provided in the parentheses. 198
Table A4: Average network properties of Jak1 unbound and bound states with inhibitors generated through PSNs. 198
Table A5: Decomposition of binding free energy into contributions from individual residues[a] 198
Table A6: The energy contributions of the different parts of the receptor toward the inhibitor bindings. All are in kcal/mol units. 199
Table A7: Types of interactions involved in recognition of inhibitors on JAK1 200

Appendix B

Table B1: The average values of backbone RMSD, the radius of gyration (Rg), and solvent accessible surface area (SASA) for all seven systems. The statistical error was estimated based on the deviation between block averages and provided in the parentheses. 217
Table B2: Average network properties of three phosphorylated JAK1 bound with SOCS1, generated through PSNs. 217
Table B3: Hydrogen bond (HB) bonds formed between JAK1 and SOCS1 and the corresponding average distance and percentage of HB occupancy determined using the 3 μs trajectories in the MD simulations. 218

Appendix C

Table C1: The mean value of the distance between donor hydrogen and acceptor. 232
Table C2: Average network properties of three systems, generated through PSNs. 232
Table C3: Energetic components of the binding free energy for phytochemicals bound to the JAK1. All values are given in kcal/mol. 232
Table C4: Total energy decomposed into their residual levels for CI and AI bound systems*. 233
Table C5: Hydrogen bond (HB) bonds formed between JAK2, and inhibitors (type-I/type-II) and the corresponding average distance and percentage of HB occupancy determined using the 3 μs trajectories in the MD simulations. 233

Appendix D

Table D1: List of top hit compounds with their Glide XP-Docking scores in kcal/mol, obtained from the virtual screening workflow and ADMETlab 2.0. 244
Table D2: Prediction of essential ADME properties of all lead compounds obtained from QikProp. 246
Table D3: The drug-likeliness properties calculated using the SwissADME webserver. Lipinski's rule of 5 helps in distinguishing between drug-like and non-drug-like molecules. It predicts a high probability of success or failure due to drug-likeness for molecules complying with two or more rules. 247
Table D4: Toxicity prediction of all screened molecules using ProTox-II webserver. . 248
Table D5: Energetic components of the total binding free energy of natural compounds bound to the JAK1 for 1000 frames of 100 ns trajectory. All values are given in kcal/mol. The standard error of the mean is provided in parentheses. 250
Table D6: Energetic components of the binding free energy for phytochemicals bound to the JAK1. All values are given in kcal/mol. 251
Table D7: Hydrogen bonds formed between JAK1, and top six lead molecules and the corresponding average distance and percentage determined using the simulation trajectories of run1 and run2. 252

ACRONYMS

EPKs	Eukaryotic kinases
ATP	Adenosine triphosphate
ADP	Adenosine diphosphate
TPKs	Tyrosine protein kinases
STPKs	Serine/Threonine protein kinases
MAK	Mitogen-activated protein kinases
PKA	Protein kinase A
JAK	Janus Kinase
STAT	Signal Transducers and Activators of Transcription
P-loop	Phosphate binding loop
G-loop	Glycine-rich loop
A-loop	Activation loop
C-Spine	Catalytic spine
R-Spine	Regulatory spine
FDA	Food and Drug Administration
FF	Forcefield
DFT	Density Functional Theory
PME	Particle mesh Ewald

LJ	Lennard-Jones
AMBER	Assisted Model Building with Energy Refinement
CHARMM	Chemistry at HARward Macromolecular Mechanics
OPLS	Optimized Potentials for Liquid Simulations
GROMOS	GROningen MOlecular Simulations
GAFF	Generalized Amber Force Field
SANDER	Simulated Annealing with NMR-Derived Energy Restraints
MD	Molecular Dynamics Simulations
PBC	Periodic Boundary Condition
TIP3P	Transferable Intermolecular Potential with three-Point charges
SASA	Solvent-Accessible-Surface-Area
PB	Poisson-Boltzmann
GB	Generalized Born
aMD	Accelerated molecular dynamic simulations
GaMD	Gaussian accelerated molecular dynamics
CV	Collective Variables
RMSD	Root-mean-square-deviations
RMSF	Root-mean-square-fluctuations
DCCM	Dynamic Correlation matrix

PCA Principal component analysis

dPCA Dihedral Principal Component Analysis

FES Free energy surface

NAPS Network Analysis of Protein Structures

FEP Free Energy Perturbation

TI Thermodynamic Integration

MM-PBSA Molecular Mechanics Poisson-Boltzmann/Generalized-Born Surface Area

LCPO Linear Combination of Pairwise Overlap

PDB Protein Data Bank

PCN Protein Contact Network

PMF Potential of Mean Force

PSN Protein Structure Network

NOMENCLATURE

ns	nanosecond
μs	microsecond
ps	picosecond
nM	nanomolar
NVT	number, volume, temperature
NPT	number, pressure, temperature
C_α	carbon alpha atom
C_β	carbon beta atom
PC	principal components

Chapter 1

Chapter 1

Introduction and Literature Review

1.1. Protein tyrosine kinase

The eukaryotic kinases (EPKs)[1] were first discovered in 1943, but the elucidation of their primary function[2,3] of phosphorylation was identified in 1956. The discovery of protein kinases opens up a new avenue and allows a better understanding of various cell-signaling pathways[4,5], replication[6], DNA transcription[7], organ trafficking[8,9], cell differentiation[10,11], smooth muscle contraction[12,13], and different neurological regulations[14–16] and transmission[17], etc. The human kinase is essential, as evident from its occurrence in the eukaryotic genome, which acquired 2% to 3% of the genes. Approximately 2% of the human genes are encoded by 518 kinases[18,19], along with 106 pseudogenes in the human genome. Because of their overexpression and genetic alterations, including mutations and translocations, the dysfunction of protein kinase activity is involved in the pathogenesis of various diseases[20] such as autoimmune, inflammatory, cardiovascular, and nervous, as well as many types of malignancies[21]. Consequently, this enzyme family has become one of the most crucial therapeutic targets in the 21st century[22,23]. One-quarter of the drug discovery efforts in the world target protein kinases. Therefore, understanding their mechanism of action and structural characterization can significantly impact the development of more potent and selective drugs[18]. An atomic-level structure is tremendously helpful and typically generates substantial insight into how the biomolecule functions[24]. Atoms in a biomolecule are in constant motion, however, and both molecular function and intermolecular interactions depend on the dynamics of the molecules involved. One would like not just a static snapshot but the ability to watch these biomolecules in action, to perturb them at the atomic level, and to see how they respond[24]. Unfortunately, watching the motions of individual atoms and perturbing them in the desired fashion is difficult. An attractive alternative is to work with an atomic-level computer simulation of the relevant biomolecules[24]. Therefore, this thesis mainly comprises several *in-silico* techniques, including conventional molecular dynamics (MD)[25] simulation and enhanced sampling approaches to scrutinize the kinase conformational plasticity and inhibition mechanisms.

1.2. Kinase function

Protein kinases can modify the function of a protein in almost every conceivable way. Various post-translational modifications[26], including phosphorylation[27], glycosylation, methylation, ubiquitination, acylation, nitrosylation, lipidation, and proteolysis, increase or decrease enzyme activity and can alter other biological activities, such as transcription and translation, which affect normal and pathophysiological aspects of life[28–30].

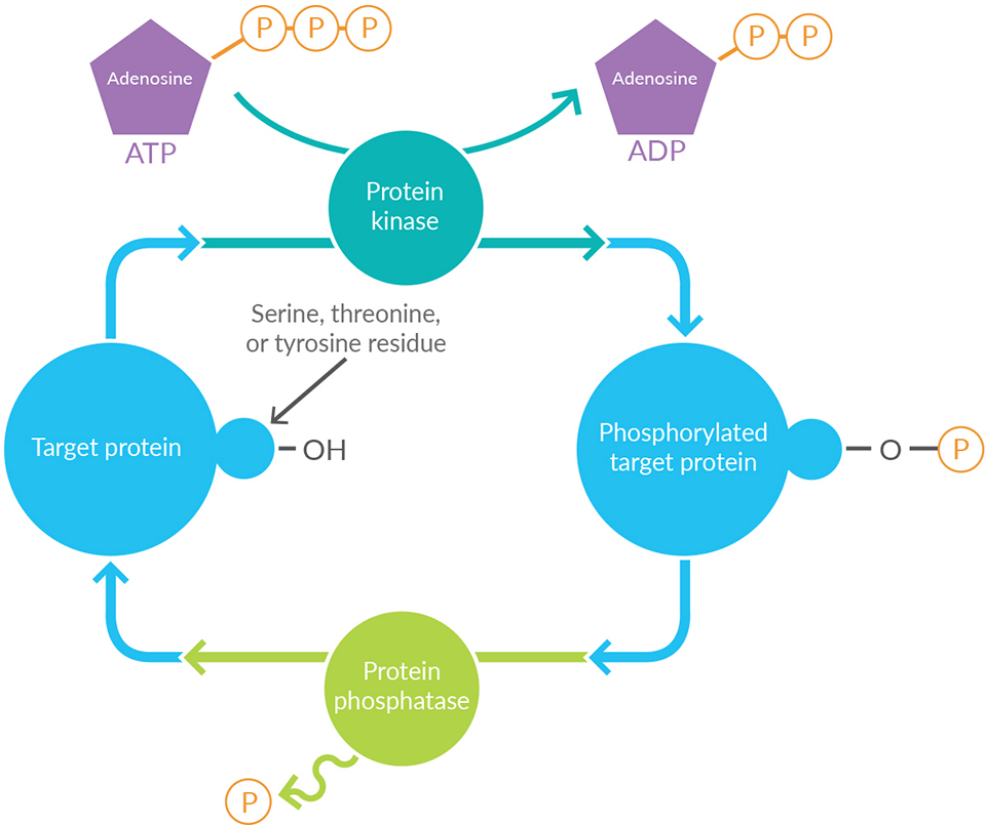


Figure 1.1. The schematic representation of the function of a protein kinase. Here, P represents the phosphoryl group (PO_3^{2-}) group.

Protein kinases are enzymes that catalyze the following phosphorylation reaction[31]:

$$MgATP + protein\text{-}O{:}\ H \rightarrow protein\text{-}O{:}\ PO_3^{2-} + MgADP + H^+$$

Note that the phosphoryl group (PO_3^{2-}) and not the phosphate (OPO_3^{2-}) group is transferred from ATP (adenosine triphosphate) to the protein substrate. Divalent cations such as Mg2+ or Mn2+ are required for the reaction, although Mg2+ is a physiological ion owing to its

significantly greater intracellular content. The kinase catalyzes the phosphorylation reaction by transferring γ phosphate of ATP onto the substrate's hydroxyl groups, including sugars, lipids, and amino acids, and releases ADP byproducts (**Figure 1.1**). The process of phosphorylation can be controlled and reversed by the protein phosphatases.

Based upon the identity of the phosphorylated -OH group, these protein kinases are categorized as protein-serine/threonine kinases (STPKs), protein-tyrosine kinases, or tyrosine-kinase-like proteins[27,29,32]. There are 385 serine/threonine kinases, 90 tyrosine kinases, and 43 tyrosine-kinase-like proteins. Among protein-tyrosine kinases, 58 are extracellular transmembrane, and 32 are intracellular non-receptor, including the **Janus kinase (JAK) family**[33,34]. A small group of proteins, including MEK1 and MEK2, catalyze the phosphorylation of both threonine and tyrosine residues on target proteins. They are called dual-specificity kinases[35]. Protein kinases themselves are also regulated by phosphorylation. Aberrant phosphorylation of cellular proteins has been associated with various disorders such as cancer, inflammatory diseases, cardiovascular disorders, diabetes, hypertension, cellular growth, and survival[20,21,26,27,30].

1.3. The JAK/STAT signaling: from the cell surface to the nucleus

The JAK/STAT (Janus kinase/signal transducer and activator of transcription) pathway[36] was largely uncovered in a series of ground-breaking studies from James Danell, George Stark, and Ian Kerr's laboratories three decades ago[37,38]. It transduces extracellular signals from various cytokines, growth factors, and hormones to the nucleus and is the leading player in the expression of thousands of protein-encoding genes. Each JAK protein binds to the juxtamembrane region of a specific cytokine receptor[38]. The genes specify each high-affinity cytokine receptor homo- or heterodimer. The receptors for type I cytokines share a common amino acid motif (WSxWS) in the extracellular part adjacent to the cell membrane[31]. Receptor type II cytokines are similar to type I cytokine receptors, consisting of an extracellular domain, transmembrane segment, and intracellular domain lacking catalytic activity but binding to specific Janus kinases.

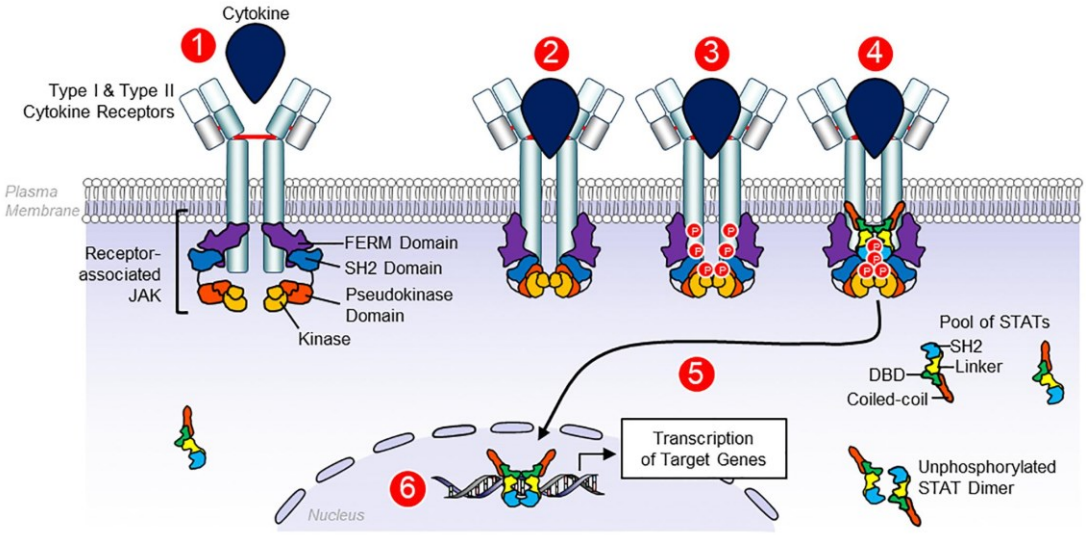


Figure 1.2: The JAK/STAT signaling transduction pathway by type I and type II cytokines. Figure reprinted from Perez-Jeldres et al. (Ref[39]).

The JAK/STAT cascade is an elegantly simple signaling cascade in which the cytokine requires only three components (receptor, JAK, and transcription factor) to elicit a response (see **Figure 1.2**)[39]. It involves several steps in which the extracellular signal is converted into a transcriptional response. First, cytokine binds to and induces the dimerization of the corresponding cytokine receptors, allowing JAK auto-activation by transphosphorylation. Once activated, JAKs phosphorylate the intracellular tails of the receptors on specific tyrosines, which act as docking sites for members of the signal transducer and activator of the transcription (**STAT**) family of transcription factors (see **Figure 1.2**). Then the receptor-localized STATs are phosphorylated by JAK, which leads to their disassociation from the receptor and activates to form a dimer. Finally, the dimerized STATs in the cytoplasm are transferred into the nucleus to regulate the expression of cytokines-responsive genes by combining with the specific DNA elements[40]. Several proteins attenuate cytokine signaling at multiple levels of the pathway to ensure that signaling is switched off appropriately. Notably, the suppressors of cytokines signaling (SOCS)[41–43] family are negative feedback inhibitors of the signaling cascade.

1.4. Janus kinase family proteins

The Janus kinase (JAK)[36,40] family of protein-tyrosine kinases comprises four members: JAK1, JAK2, JAK3, and TYK2 (tyrosine kinase 2). The discovery of JAKs happened in 1989–1994. In 1989, Wilks et al. [44]. they have identified that a tyrosine kinase has a recognizable kinase domain and a pseudokinase domain. In 1991, they found a second tyrosine kinase with this feature. Wilks called them JAK1 and JAK2[45]. The other two JAKs, tyrosine kinase 2 (TYK2) and JAK3, were identified in 1990 and 1994[46]. The JAK kinase has a modular architecture comprising seven JAK homology (JH1-JH7)[40] modules, 1100 amino acid residues, and a molecular weight of 120-140 kDa. These have been reorganized into four main functional domains (see **Figure 1.3**). The JH7-JH6 module corresponds to a FERM (F for 4.1 protein, E for ezrin, R for radixin, and M for moesin) domain as we go from the N-terminus to the C-terminus. This is a ubiquitous protein module of ~350 residues and involves targeting proteins to the plasma membrane. The JH5-JH3 module corresponds to an atypical SH2 (SRC kinase homology) domain. The rest of the two modules corresponds to the pseudokinase domain (JH2) and catalytically active protein-tyrosine kinase domain (JH1). All four members of the JAK family (JAK1, JAK2, JAK3, and TYK2) associate with the membrane-proximal regions of cytokine receptor ICDs through two distinct conserved motifs in the receptor: a proline-rich segment termed "Box1" and a hydrophobic segment called "Box2". Recently, Glassman *et al.* [47] presented a full-length JAK structure, which provides a structural roadmap for understanding the regulatory mechanisms that govern JAK activity and the promise of new therapeutic approaches.

1.4.1. JAK family architecture

The first visualized three-dimensional crystallographic structure was obtained for the **cyclic-AMP dependent protein kinase**, also known as **protein kinase A (PKA),** in 1991. It is a serine/threonine kinase and one of the best-characterized members and serves as a paradigm for the kinase family. Typically, the conserved catalytic core of a **Janus kinase domain (TK)** comprises two lobes: **amino-terminal (N-lobe) and carboxy-terminal (C-lobe)**. A typical JAK1 three-dimensional crystal structure is shown in **Figure 1.4**. The small N-lobe consists of an antiparallel β sheet (β_1-β_5) and a prominent α-helix called **αC-helix**. The large C-lobe is mainly composed of α-helices (αD-αI). The C-lobe contains short

conserved β-strands (β_6-β_9) containing catalytic residues. Followed by the αE-helix are the β_6-strand, the **catalytic loop**, the β_7 and β_8-strands, and the **activation segment (A-loop)** with the β_9-strand. In most protein kinases, the activation segment starts with **DFG** (Asp-Phe-Gly) motif and ends with **the APE** (Als-Pro-Glu) motif. JAK isoform has sequence identity from CLUSTALW in a range of 46–55% and sequence similarity of 61–71%, and the conserved regions have a sequence identity of 65–78% (see **Figure 1.5**).

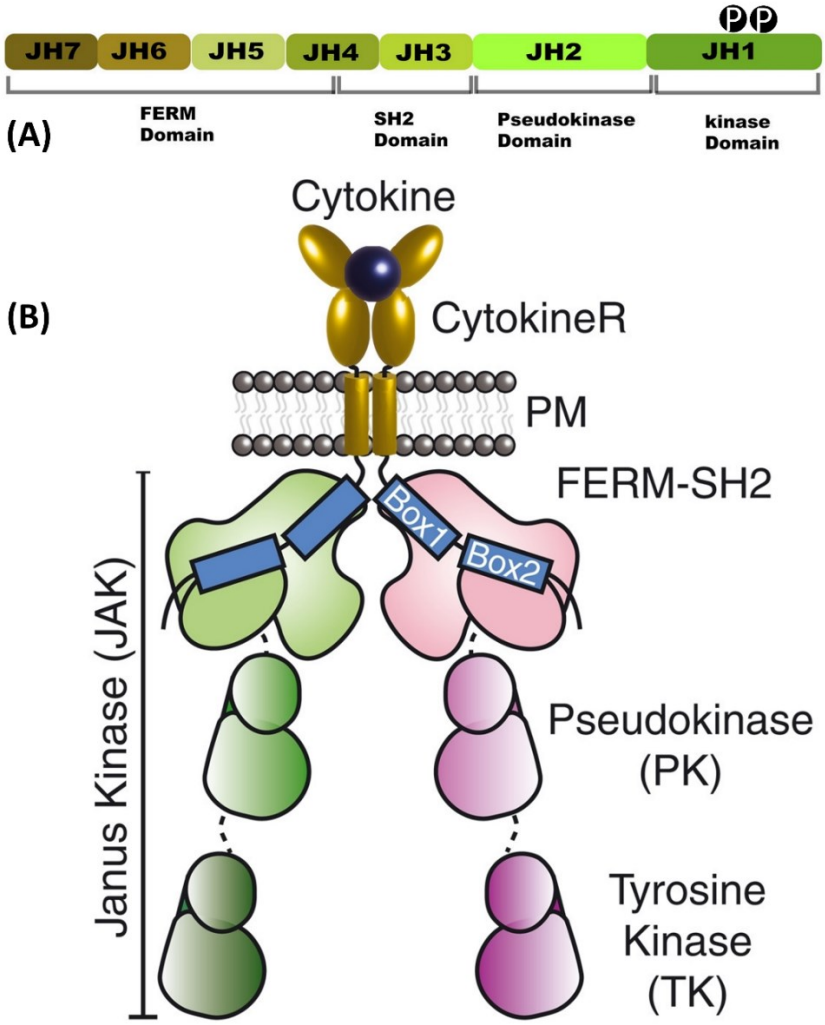


Figure 1.3: (A) Schematic of JAK structural domains. There are four main domains and seven JAK homology (JH) in the JAKs structure. The conserved tyrosine phosphorylation sites in JAK1 are Y^{1034}/Y^{1035}; the conserved tyrosine phosphorylation sites in JAK2 are Y^{1007}/Y^{1008}; the conserved tyrosine phosphorylation sites in JAK3 are Y^{980}/Y^{981}; the conserved tyrosine phosphorylation sites in Tyk2 are Y^{1054}/Y^{1055}. (B) Schematic of a cell-surface, ligand-induced cytokine receptor-JAK dimer. This figure is inspired by Glassman et al. (Ref[47]).

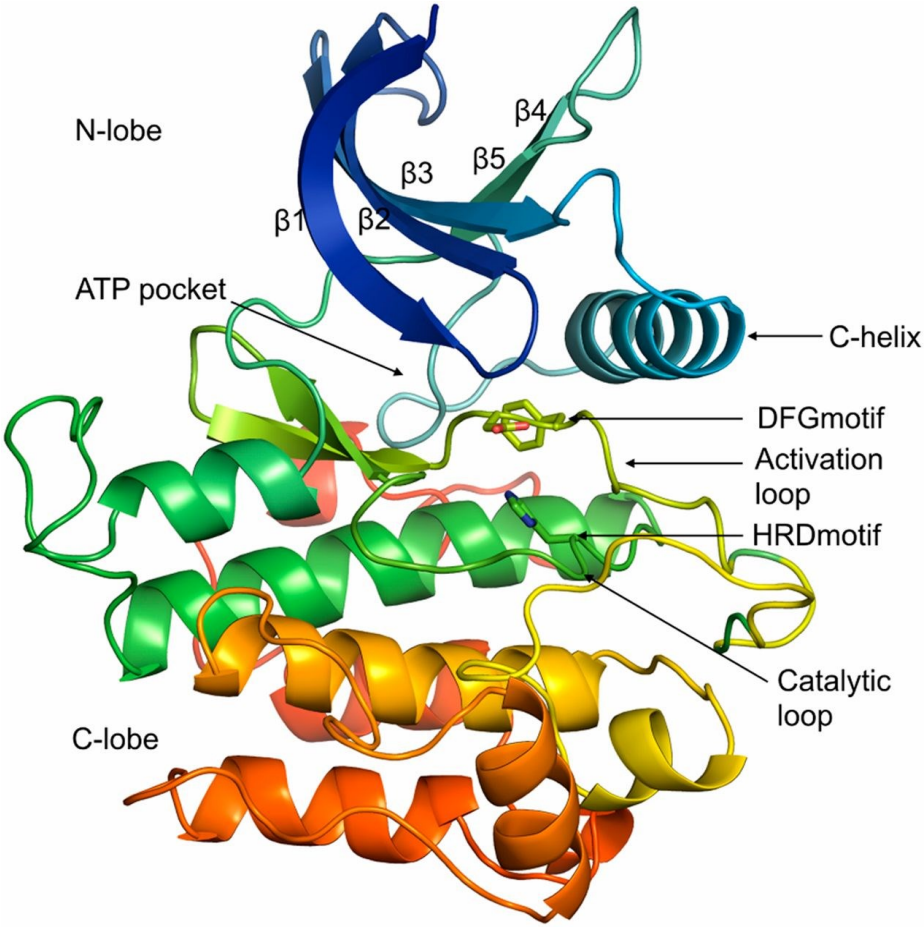


Figure 1.4: Structure of a typical protein kinase (PDB ID code 3EYG) domain displaying ATP binding site and conserved elements around it.

The ATP occupies the deep cleft between N- and C-lobes and the connecting loop of β_1 and β_2 strands. This connecting loop, also known as the phosphate-binding loop (**P-loop)**, contains a conserved glycine-rich motif (GxGxφG), where x refers to any amino acid and *φ* is a hydrophobic residue (tyrosine or phenylalanine). This loop is also known as the **glycine-rich loop (G-loop)**. These glycine residues allow the loop to interact and coordinate with the phosphates of ATP via backbone interactions. The P-loop contains a conserved **valine** residue in the β_2 strand that forms a hydrophobic contact with the adenine of ATP. In the absence of ATP, glycine residue makes the P-loop very flexible, facilitating small inhibitors' binding. The peptide substrate binds in the front end of the nucleotide-binding pocket and is close to the γ-phosphate of ATP.

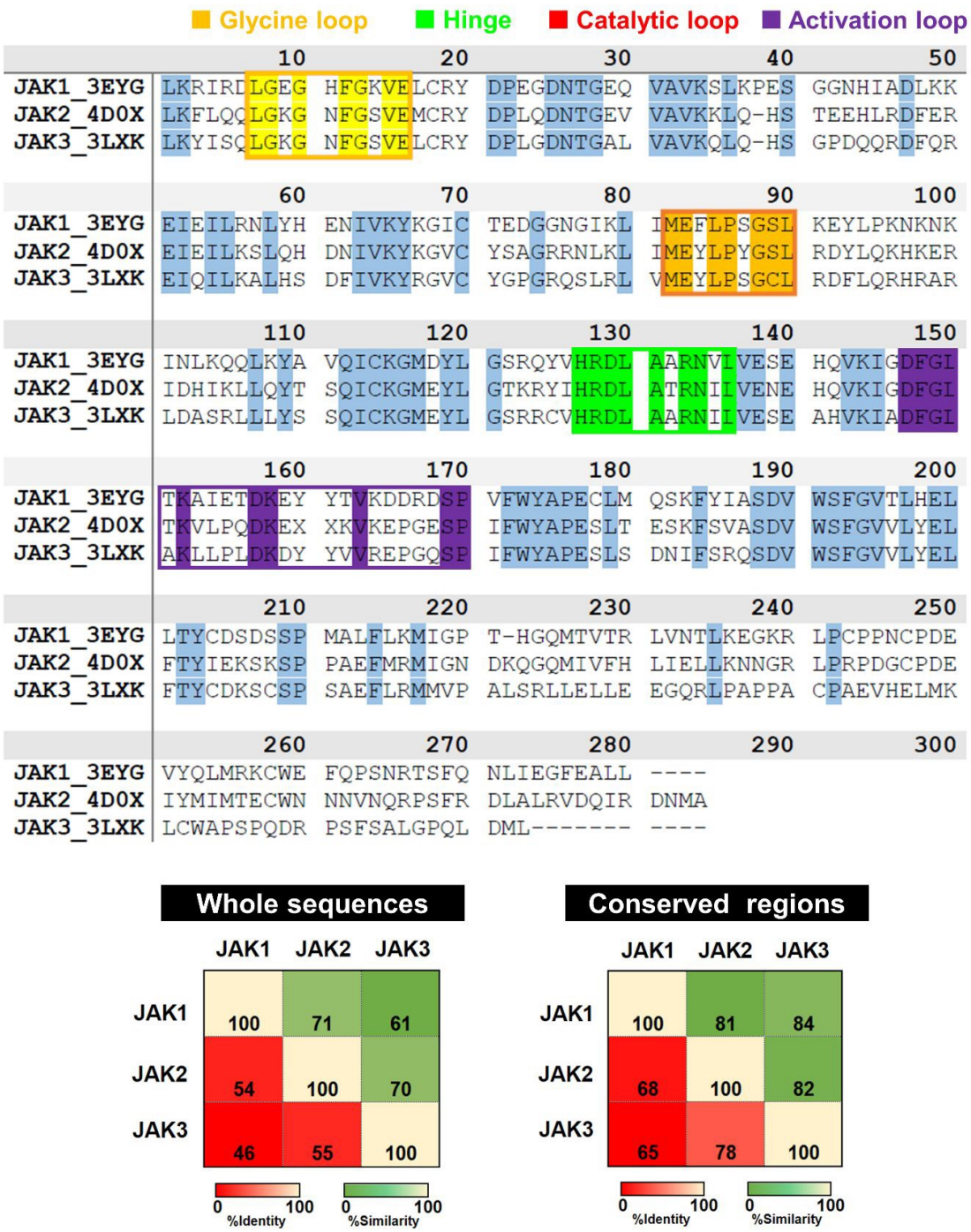


Figure 1.5: Alignment and identity table of amino acid sequences of three kinds of JAKs (whole sequences and conserved regions). The amino acid sequences of JAK1 (3EYG), JAK2 (4D0X), and JAK3 (3LXK) were compared using the CLUSTALW[48].

The β_3-strand typically contains an AXxx-K sequence, the lysine (K), which couples the α- and β-phosphates of ATP to the αC-helix. Conserved glutamate (E) locates near the center of αC-helix in protein kinases. The β_3-Lys form a stable hallmark salt-bridge with αC-Glu in active protein kinases—the location of these residues is listed in **Table 1.1**. The catalytic loop surrounding the site of phosphoryl transfer consists of HRDLAARN in most tyrosine kinases, including JAK1/3 and TYK2 but HRDLATRN in JAK2. The HRD-Asp in JAK serves as a base that accepts a proton from the tyrosyl-OH group.

The first residue of A-loop and DFG-motif, aspartate (Asp), is a member of the **K/E/D/D signature** of JAK and plays an essential role in the Mg^{2+} bindings, in turn, coordinates the α- β- and γ-phosphates of ATP. The second Mg2+ binds with the help of the catalytic loop, **Asn (N),** and coordinates the α- and γ-phosphates groups of ATP. The A-loop influences both the protein-substrate binding and catalytic efficiency. This loop in JAK contains two consecutive phosphorylatable tyrosines (see **Table 1.1**). In most protein kinases, the A-loop phosphorylation controlled the kinase activation mechanisms. The A-loop of JAK is stabilized by hydrophobic interactions between two residues N-lobe to the HRD-H of the catalytic loop and seven residues C-lobe to the DFG-D within the A-loop. All the functionally important human JAK JH1 residues are listed in **Table 1.1.**

Table 1.1: Important residues in the human Janus kinases. This table is reprinted from Ref[31].

Area/residue(s)	**JAK1**	**JAK2**	**JAK3**	**TYK2**
N-lobe				
P-loop residues	^{882}GEGHFG^{887}	^{856}GKGNFG^{861}	^{829}GKGNFK^{834}	^{904}GEGHFG^{909}
β_3-Lys (K)	K^{908}	K^{882}	K^{855}	K^{930}
αC-Glu (E)	E^{925}	E^{898}	E^{871}	E^{947}
Hinge residues	^{957}EFLPG^{961}	^{931}YKPYG^{936}	^{903}EYLPSG^{908}	^{979}EYVPLG^{984}
C-lobe				
αE-A-loop	Y^{999}-A^{1027}	Y^{972}-V^{1000}	C^{945}-L^{973}	Y^{1019}-A^{1047}
HRD motif	H^{1001}-D^{1003}	H^{974}-D^{976}	H^{947}-D^{949}	H^{1021}-D^{1023}
Catalytic (Asn)	N^{1008}	N^{981}	N^{954}	N^{1028}
A-loop residues	1021-1051	994-1024	967-997	1041-1071
DFG motif	D^{1021}-G^{1023}	D^{994}-G^{996}	D^{967}-G^{969}	D^{1041}-G^{1043}
Phosphorylation	**Y^{1034}/Y^{1035}**	**Y^{1007}/Y^{1008}**	**Y^{980}/Y^{981}**	**Y^{1054}/Y^{1055}**
APE motif	A^{1049}-E^{1051}	A^{1022}-E^{1024}	A^{995}-E^{997}	A^{1067}-E^{1071}
UniProt KB ID	P23458	O60674	P52333	P29597

Table 1.2: Spine and shell residues of human JAK JH1 domains. This table is reproduced from Ref[31].

Location	Symbol	JAK1	JAK2	JAK3	TYK2
Regulatory spine					
β_4-starnd (N-lobe)	RS4	Y^{940}	Y^{913}	Y^{886}	Y^{962}
αC-helix (N-lobe)	RS3	$\mathbf{L^{929}}$	$\mathbf{L^{902}}$	$\mathbf{L^{875}}$	$\mathbf{L^{951}}$
A-loop (N-lobe)	RS2	$\mathbf{F^{1022}}$	$\mathbf{F^{995}}$	$\mathbf{F^{968}}$	$\mathbf{F^{1042}}$
Catalytic (N-lobe)	RS1	$\mathbf{H^{1001}}$	$\mathbf{H^{974}}$	$\mathbf{H^{947}}$	$\mathbf{H^{1021}}$
αF-helix (N-lobe)	RS0	D^{1063}	D^{1036}	D^{1009}	D^{1083}
R-shell					
Gatekeeper-2	Sh3	L^{954}	L^{927}	L^{900}	L^{976}
Gatekeeper	Sh2	$\mathbf{M^{956}}$	$\mathbf{M^{929}}$	$\mathbf{M^{902}}$	$\mathbf{M^{978}}$
αC-β_4 loop	Sh1	V^{938}	V^{911}	V^{884}	I^{960}
Catalytic spine					
β_2-starnd (N-lobe)	CS8	V^{889}	V^{863}	V^{836}	V^{911}
β_3-starnd (N-lobe)	CS7	A^{906}	A^{880}	A^{853}	A^{928}
β_7-starnd (C-lobe)	CS6	L^{1010}	L^{983}	L^{956}	L^{1030}
β_7-starnd (C-lobe)	CS5	V^{1011}	V^{984}	V^{957}	L^{1031}
β_7-starnd (C-lobe)	CS4	V^{1009}	I^{982}	I^{955}	V^{1029}
αD-helix (C-lobe)	CS3	L^{964}	L^{937}	L^{910}	L^{986}
αF-helix (C-lobe)	CS2	T^{1070}	V^{1043}	S^{1012}	T^{1090}
αF-helix (C-lobe)	CS1	L^{1074}	L^{1047}	V^{1016}	L^{1094}

Local spatial pattern (LSP) alignment of EPK structures led to the discovery of two hydrophobic "spines" connecting the N-lobe with the C-lobe. The regulatory spine (R-spine) is the hallmark signature of an active kinase and is dynamically assembled as part of the regulation. In contrast, the catalytic spine (C-spine) is completed by binding the adenine ring of ATP. Both spines are anchored to the hydrophobic αF-helix. JAK has four hydrophobic residues (RS4, RS3, RS2, and RS1) in R-spine along with an anchor residue (RS0, Asp) at αF-helix (see **Table 1.2**). The R-spine contains a residue from the αC-helix and A-loop, which are important in determining the active and dormant states. The backbone of HRD-His (RS1) is anchored to the very hydrophobic αF-helix by a hydrogen bond to the conserved aspartate residue (RS0).

Meharena *et al*.[49] identified three residues in the protein kinase domain that stabilized the R-spine, labeled as Sh1, Sh1, and Sh3, where Sh refers to the shell. The Sh2 residue corresponds to the gatekeeper residue. The name gatekeeper signifies this residue's role in controlling access to the back cleft. The back cleft is sometimes called the back pocket or hydrophobic pocket II. Moreover, the protein kinase C-spines are made up of residues from the N- and C-lobes; all the C-spines residues are listed in **Table 1.2**. Small molecules protein kinase therapeutic antagonists generally interact with residues that make up C-spine and sometimes the R-spine and shell residues[35].

1.4.2. Active and inactive JAK structure

Eukaryotic kinases are among the biological molecules that are highly dynamic and flexible[50,51]. Protein kinases act like a molecular switch, transitioning between "active" and "inactive" states[52,53]. The "active" state has maximal activity, while the "inactive" state reflects minimal activity. All protein kinases perform the same phosphorylation of the substrate by transferring the γ phosphate from ATP. Thus, almost all kinases' active form resembles structurally[54–56]. On the other hand, the inactive state, representing the deactivated form, is not a constraint to be the same in all kinases. Therefore, a different class of kinases and the same kinase can adopt distinct "inactive" states. The active kinase tertiary structures are more or less similar conformations, while Huse and Kuriyan noted that an inactive kinase's crystal structure reveals distinct conformations[51]. Thus, the active and inactive protein kinases have different conformations. The different conformational dynamics of the various regions, such as A-loop, P-loop, αC-helix, etc., determine kinases' active or inactive states. A large conformational rearrangement occurs in the protein upon phosphorylation, ATP bound, or inhibitors bound state. Hence, a mechanistic understanding of these significant conformational changes will help develop novel targeted therapeutics.

The activation segment controls both catalysis and substrate binding[49]. Nolan *et al*. reported that the autophosphorylation residues from the activation segment are required for the activation of many kinases[57]. However, few kinases required no such phosphorylation for its activation, say EGFR[58–60]. A-loop, a part of the activation segment, plays the most critical role in regulating the conformational changes in many kinases. Tournier *et al*.[60]

stated that phosphorylation of Thr^{221} alone (out of dual phosphorylation site; Thr^{221} and Tyr^{223}) could activate JNK in the active form. JNK is expressed in the neuronal cells and is a widely studied target for small drugs for treating neurological disorders. pThr^{221} forms new interactions with the basic residues of the A-loop (Arg^{227} and Arg^{188}) and puts the protein in the active open state. The mechanism provided by Xie *et al.*[61] and Mishra *et al.*[62] suggested that basic residues contribute to the conformational rearrangements of the A-loop in JNK3. The behavior of kinases differs with the unphosphorylated/phosphorylated or ATP-bound conformations. It regulates the different open/closed states of a kinase. For example, in uJNK3 and similar kinases like CDK5, a large-scale movement was reported during the open-to-closed transition[63]. A similar mechanism was observed for the Abl kinase 1 and Src kinase, in which αc-helix, DFG, and A-loop coordinate in a manner to attain open-to-closed conformations[64] (see **Figure 1.6**).

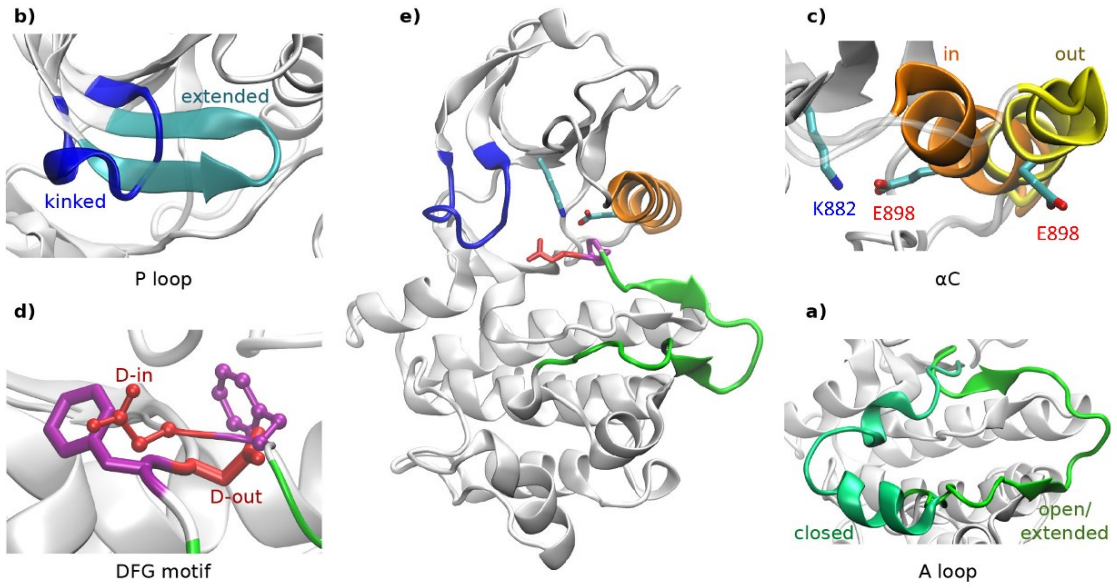


Figure 1.6: Overview of the conformational variability of the catalytic domain (kinase domain). In **e),** the active conformation (based on PDB structure 2F4J, residues 235–498) is shown. The flexible structural elements are shown in color: the P-loop is shown in blue, the αC is shown in orange, and the A-loop is shown in green. The DFG motif is shown in the "licorice" representation and colored in red (Asp), violet (Phe), and white (Gly). The highlighted structural elements can adopt different conformations: **b)** The P-loop is mainly in the kinked (blue) or extended (cyan) conformation. **c)** αC conformations range between αC-*in* (orange) and αC-*out* (yellow) positions. **d)** The DFG motif is mainly in the DFG-*in* (or simply in, ball-and-stick representation) or the DFG-*out* (or simply out, "licorice" representation) state. **a)** The A-loop adopts many conformations that range between closed,

inactive (emerald green) and extended, active (lime green). Figure reprinted from Fabian Paul *et al.* (Ref[65]).

Furthermore, Shan and co-workers studied the DFG-flip from the DFG-in to the DFG-out in Abl kinase in a long timescale MD simulations (2.2 μs)[66]. In addition, Meng and Roux observed that the phosphorylation of the Tyr^{416} of the A-loop of c-Src stabilizes several structural features of the kinase and locks them into catalytically competent conformations[67]. Similarly, Musavizadeh *et al.*[68] observed that single phosphorylation at Thr288 in Aurora-A kinase stabilizes the A-loop in open conformations. At the same time, paradoxically, the presence of second phosphorylation at Thr^{287} decreases such stability and makes the A-loop more flexible. It indicates that the conformational distributions of Aurora-A with both single and double phosphorylation modifications are different from the unphosphorylated state. Similarly, studies suggested that the αC-helix could also undergo conformational transition, like αC-helix-in or αC-helix-out conformations, and plays a crucial role in mediating the interplay between the active to the inactive state of different kinases[63,69]. Therefore, the fundamental difference between active/inactive states of kinases is determined by inter-lobe motions, the position of αC-helix, and the A-loop conformations.

1.4.3. Mechanisms of JAK activation

Protein kinases are stringently regulated and have not evolved to continuously catalyze the phosphorylation of thousands of molecules per minute, like hexokinase, a general metabolic enzyme. The JAK is activated by its cytokine receptor. The primary phosphorylated product is the enzyme itself mediated by the trans-phosphorylation of one JAK protein by another. Phosphorylation of the cytokines receptor and STATs protein also occurs. The interaction of a cytokine ligand with its receptor triggers a change in the receptor conformation leading to the repositioning of the associated JAKs. This rearrangement promotes the *trans*-phosphorylation of activation segment tyrosine residues. Previously, Feng et al.[70] showed the importance of these tyrosines in activating JAK2, Zhou et al.[71] in JAK3, and Gauzzi et al.[72] in TYK2 activity. These studies also indicate that the phosphorylation of the first tyrosine residue of the A-loop increases catalytic activity.

1.5. The JAK/STAT signaling and human diseases

The JAK/STAT signaling pathway has profoundly influenced recent understanding attained of human health and disease. Many papers have reported the importance of this pathway in malignancies and autoimmune diseases[73–78]. The JAK/STAT pathway is a central signaling hub in human immune responses and cancer[40]. It is critical for meeting the diverse challenges faced by the immune system, from resisting infection to maintaining immune tolerance, enforcing barrier functions, and guarding against cancer. The JAK/STAT signaling has been implicated in almost every autoimmune disease[79]. Therefore, a delicate balance must be reached to allow for efficient induction of the JAK/STAT pathway when the immune system is called to action and proper diminution when the instigating threat subsides. All known STAT loss-of-function mutations can be associated with particular types of pathogens[36]. The recent identification of patients with gain-of-function STAT3 mutations highlights the central role of this pathway in promoting autoimmune disease. Over 30 years, the link between JAK/STAT signaling and oncogenesis has become a significant thread in cancer biology[73]. The following is a brief overview of how JAK/STAT dysfunction impacts human disease.

Our immune systems sometimes work aberrantly and attack their own cell; these conditions are commonly called auto-immune disorder conditions. Typically, we divided the JAK/STAT-related human diseases into two major section[31]: **inflammatory disorders** and **neoplastic disorders**.

1.5.1. Inflammatory disorders

Rheumatoid arthritis (RA)[80] is an auto-immune disorder and chronic immune-related condition, primarily of joints, and is highly disabling and painful and reduces the quality of life. The illness is characterized by synovitis of multiple small joints, pain, and morning stiffness. In the Western population[81], the estimated prevalence of RA is 1-2%, and worldwide[82] majority is believed to be 1% of adult humans. The occurrence and development of RA are closely related to the intracellular JAK/STAT signaling. Revealing the role of this signaling pathway in the pathogenesis and further understanding the occurrence and development of RA may provide new therapeutic strategies and targets for RA treatment. Recently, small-molecule inhibitors' inhibition of JAK activation has

received significant attention. In the next section, we will discuss the details of these small molecule inhibitors.

Psoriasis[83] is a common inflammatory illness that affects ~2-4% of the population[31], and it occurs in men and women with an equal frequency. It is a chronic autoimmune disorder[84] like RA and occurs at any age, and is typified by patches of abnormal skin. Still, there is no proper cure, but JAK appears to be a promising therapeutic option. Thus many of the molecules with an active role in the pathogenesis of psoriasis, like IL-22, IL-23, IL-15, and IFNγ, employ the JAK/STAT pathway to mediate their function. Therefore, inhibition of the JAK/STAT pathway with specific JAK inhibitors has the potential to modulate pain in patients with psoriasis. Other inflammatory disorders include neuroinflammation (Parkinson's disorder, multiple sclerosis), inflammatory bowel diseases (IDB), Canine dermatitis, ulcerative colitis, etc. Naturally occurring mutations in cytokine receptors, JAKs, and STATs lead to immunodeficiency and myeloproliferative disorders in humans[37,85].

1.5.2. Neoplastic disorders

The JAK/STAT signaling pathway is associated with a dozen human neoplastic disorders, including myelofibrosis, polycythemia vera, and primary thrombocythemia. The JAK/STAT pathway is extensively present in various tissues and cells, and its overactivation is closely related to multiple diseases. Myelofibrosis is a rare idiopathic disorder that results from the proliferation of hematopoietic stem cells in the bone marrow and replaces these cells with scar tissue[86]. Dysfunction of the JAK/STAT pathway is the central pathogenic mechanism in neoplasm disorders. Polycythemia vera is a clonal neoplastic stem cell disorder characterized by excessive red blood cell production; this condition may also result in the overproduction of white blood cells and platelets. Primary thrombocythemia (thrombocytosis) is another clonal disorder involving a multipotent hematopoietic progenitor cell manifested clinically by the overproduction of blood platelets without the erythrocyte overproduction.

1.6. Therapeutic target molecules in the JAK/STAT signaling

The inhibiting of the JAK/STAT pathway is promising for treating various diseases. Medications that target the JAK/STAT pathway can be classified into three types: (1)

cytokine or receptor antibodies, (2) **JAK inhibitors**, and (3) STAT inhibitors (see **Figure 1.7**). In this thesis, we are interested in **JAK inhibitors** and their inhibition mechanisms. JAK inhibitors are a group of small-molecule inhibitors with different chemical structures. The therapeutic effects of JAK inhibitors are based on two factors. First, JAKs mediate various cellular activities. Inhibition of JAK function can cause immunosuppression and decrease the abnormally elevated serum proinflammatory cytokines mediated by the JAK/STAT signaling pathway. Second, in some diseases, such as myeloproliferative diseases and cancers, identifying gain-of-function JAK mutants enables treatment through their inhibition. Various JAK inhibitors are being studied in preclinical and clinical studies. Tofacitinib, baricitinib, and upadacitinib are the first orally available JAK inhibitors to be approved for treating RA and other autoimmune diseases.

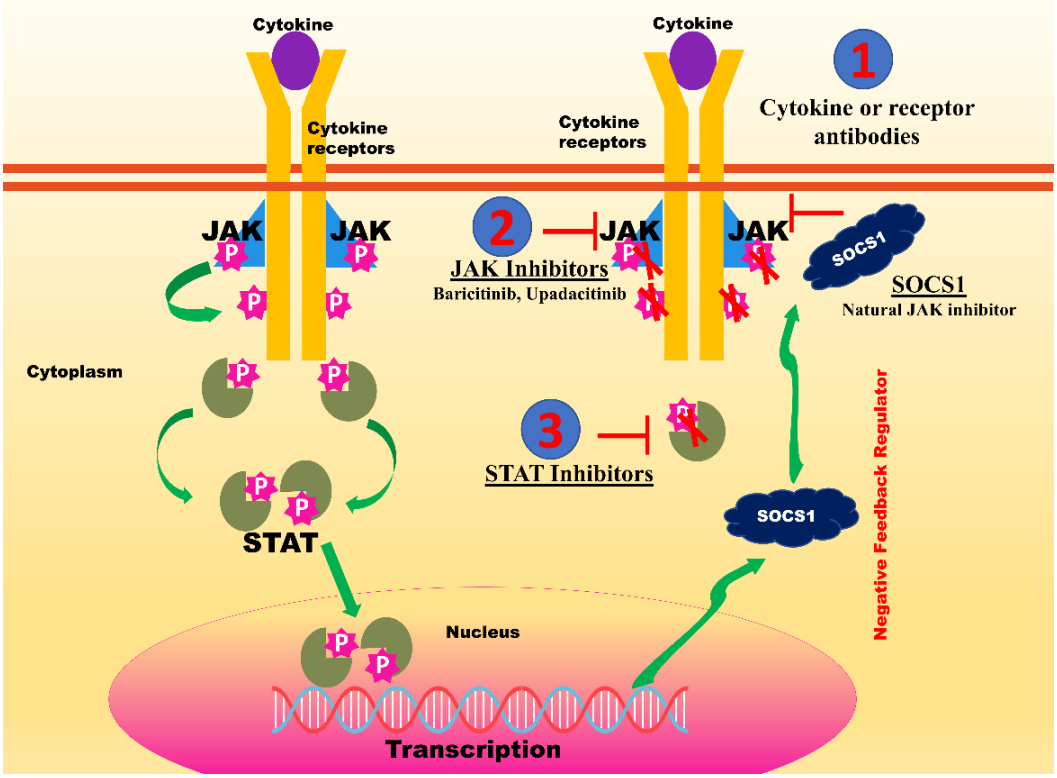


Figure 1.7: Therapeutic targets of the JAK/STAT signaling pathway. **(1)** Recombinant cytokines, cytokine antibodies, and receptor antibodies are designed to target cytokines or receptors; **(2)** JAK inhibitors are designed to target JAKs; and peptide inhibitors including SOCS, **(3)** small-molecule inhibitors are designed to target the STAT proteins and inhibits the dimerization of STATs.

1.7. JAK/STAT signaling pathway inhibitors

The JAK/STAT signaling pathway promotes cytokine-mediated cell activation efficiently. As I already discussed, the dysfunction and pathogenesis of various diseases are involved in the JAK/STAT pathway. Therefore, we have controlled the aberrant functioning by blocking the activation of JAK by natural or small molecule inhibitors. The JAK/STAT signaling pathway inhibitors are typically two types, natural negative feedback inhibitors, and JAK inhibitors. The details of these inhibitors are discussed in next following section

1.7.1. Negative feedback inhibitor

In JAK/STAT pathway, some negative regulatory factors have been identified, such as suppressor of cytokine signaling (SOCS), protein inhibitors of activated STAT (PIAS), and protein tyrosine phosphatase (PTP). In this thesis, we are very interested in SOCS and discussed the details of SOCS inhibitors in the following subsection and Chapter 4.

1.7.1.1. Suppressor of cytokine signal

Signaling via the JAK-STAT signaling pathway is a dynamic process that involves the rapid transmission of the signal from the cell membrane to the nucleus, followed by a highly organized response and subsequent controlled downregulation and attenuation of the initial signal[87]. SOCS proteins are the primary drivers of signal attenuation; they are induced by cytokine exposure (via STAT) and act as negative-feedback inhibitors to switch off the signaling cacade[43]. Sometimes it is also known as STAT-induced STAT inhibitory proteins (SSI)[88]. There are eight members of the SOCS family: cytokine-inducible SH2-containing protein (CIS) and SOCS1-7[89]. Out of these, eight CIS and SOCS1-3 perform a negative feedback loop on the cytokine signaling via the JAK/STAT pathway, while SOCS4-7 primarily regulates the growth factor receptor signaling[90,91]. SOCS proteins can be used as the ligand of ubiquitin E3 and degrade the proteins by binding to N-terminus[92]. The kinase inhibitory region (KIR) of SOCS1 and SOCS3 block the sublayer-binding groove of JAK1 and JAK2, respectively, and hinders the association with its substrate in the biochemical studies[93].

1.7.2. JAK inhibitors

Research on kinase-targeting drug development has gained significant attention over the last 30 years in treating many diseases[20,35]. In recent years, JAK inhibitors have shown

promising results in curing inflammation and immune disorders diseases[31,36,73]. As of 2022, the United States Food and Drug Administration (FDA) has approved only 5 JAK inhibitors[94]. Nearly all approved JAK antagonists are ATP-competitive inhibitors targeting the ATP binding pocket. The highly similar ATP binding site across the whole kinome renders a challenge in maintaining the selectivity for the desired kinase. JAK inhibitors can be subdivided into two different categories: *Pan-JAK* inhibitors and *Selective* inhibitors[95]. As suggested by the name, the first group includes those inhibitors, competitive with the ATP-cleft of various JAK family members. In contrast, selective inhibitors distinguish the kinase activity domain among JAK family members. To highlight the clinical course of this kind of inhibitor, *pan-JAK* and *selective* inhibitors can be labeled as First Generation or Second-generation JAK inhibitors, respectively.

1.7.2.1. First-generation JAK inhibitors

The first-generation inhibitors exploit the high homology degree present in the ATP-binding pocket to target different JAKs. These inhibitors showed lower specificity and demonstrated activity against three or even all four of the JAK family members. To date, five first-generation JAK inhibitors are present: Tofacitinib[96], Peficitinib[31], Oclacitinib[35], Ruxolitinib[36], and Baricitinib[97]. Tofacitinib and Ruxolitinib are non-selective inhibitors, and Baricitinib is a JAK1 selective inhibitor. Similarly, Peficitinib is a Pan-JAK inhibitor, and Oclacitinib is selective to both JAK1 and JAK2.

1.7.2.2. Second-generation JAK inhibitors

To improve the selectivity and increase the affinity of JAK inhibitors, another set of inhibitors was also introduced by many researchers and called the second-generation JAK inhibitors. These inhibitors are significant, leading to the emergence of a new class of therapeutics targeting JAKs. Moreover, these are highly specific to a particular enzyme and have more negligible side effect profiles linked to their mechanisms of action. Thus, there is a strong interest in designing more selective drugs, especially in long-term use for immune-related diseases.

To date, there are several second-generation inhibitors reported, namely, Upadacitinib[98], Decernotinib, Filgotinib[99], Fedratinib, Momelotinib, Lestaurtinib, Gandotinib, Pacritinib, Itacitinib[94], and Solcitinib. Upadacitinib, Filgotinib, Itacitinib, Solcitinib, Momelotinib,

and Gandotinib are selective inhibitors against JAK1. The Pacritinib and Fedratinib are specific to JAK2, while Decernotinib is selective to JAK3.

1.8. Scope and aim of the thesis

This thesis aims to investigate the detailed characterization of conformational dynamics of JAK and explore the details molecular basis for inhibition with the help of modern computational methods and algorithms. The JAK/STAT pathway is a primary signaling cascade in the human immune system. The JAK kinases are the master regulator of this pathway. Previous experimental studies showed that the dysfunction of this pathway played a pivotal role in the pathogenesis of many immune-related disorders and inflammations. The involvement of JAK in this pathway is crucial and a promising drug target for controlling the aberrant functioning of the JAK/STAT pathway. During the last decade, the drug known as JAK inhibitors or JAKinibs, blocking one or more of the molecules involved in this pathway, have been developed and tested in clinical trials for many different indications.

Although the involvement of the JAK/STAT signaling pathway in regulating immune responses has been identified, a comprehensive mechanistic understanding is still lacking. Moreover, due to the lack of complex crystal structures with next-generation small molecules and less characterized protein targets of the pathway, the discovery of novel specific small molecules has been relatively slow. The primary challenges in structure-based drug discovery (SBDD) are kinase domain structural conservation and lack of specificity toward the target. Therefore, we need a unique conformation in SBDD to enhance the affinity and specificity of inhibitors toward a particular target. Thus, understanding conformational plasticity is essential to uncovering novel conformations of kinases. However, exploring these conformational dynamics in an atomic-level resolution using experimental techniques is difficult.

In continuation of the above challenge in the kinase drug discovery, we have overcome these challenges with the help of advanced computational algorithms, and powerful graphics processing units (GPU) enabled MD studies to explore the more extended timescale dynamics and break the conformational energy barrier by using enhanced

sampling method like Gaussian accelerated MD approach. Protein kinases are flexible and dynamic molecules and undergo significant structural changes. However, there is a lack of comprehensive experimental or theoretical on the conformational plasticity of JAKs due to phosphorylation states, protein-protein interactions, and small molecule binding to the ATP-pocket and allosteric sites. Broadly, the following aims have been covered in this thesis:

***Aim 1*:** First, we need to understand the molecular basis of inhibition mechanisms of JAK and the role of non-covalent interactions in the inhibition mechanism. There are some next-generation JAK inhibitors specific to JAK1. However, the previous literature fails to explain its specificity. The molecular interaction spectra are also unknown in the absence of the experimentally determined crystal structure of the protein-ligand complexes. Here, we have investigated the molecular details of inhibition mechanisms using molecular docking and multiple replica MD simulations. We have successfully explored the binding mechanisms and specificity of these JAK inhibitors.

Aim 2: Here, we have discovered the impact of phosphorylation states on the conformational dynamics of JAK using multiple replica long microsecond MD simulations. Side by side, we also investigate the protein-protein interaction mechanisms of JAK1 and a natural negative feedback inhibitor, SOCS1 protein. Therefore, this study offers a details understanding of the conformational rearrangement of JAK, which can be a promising drug target for the SBDD industry for specific drug design.

Aim 3: In this objective, we explored the ligand-induced conformational dynamics of JAK2 and their characterization using an enhanced sampling method, GaMD, and free energy landscape techniques. Furthermore, these results give an important message to modern drug discovery and provide meaningful information about possible target conformation for inhibitors of any kinase family member.

Aim 4: In this aim, we have used the information from our previous aims and tried to provide some suitable specific inhibitors from the natural phytochemical database. Here, we have explored the extensive phytochemical database using modern systematic virtual

screening protocol, pharmacokinetics, and in-silico studies to select some promising future drug candidates to inhibit the activation mechanisms of JAK.

Altogether, my thesis presents the atomistic details of inhibition mechanisms of JAKs and the conformational plasticity of the kinase domain of JAK due to the phosphorylation and inhibitor binding. Understanding these mechanisms is essential to balance the detrimental outcome of uncontrolled inflammatory responses that could lead to chronic diseases.

1.9. Organization of the thesis

Based on the objectives stated above, the thesis has been organized into seven chapters. The introduction and review of relevant literature leading to the objectives constitute the first chapter. **Chapter 2** describes the theoretical basis of molecular dynamics simulations and protocols used during the investigation, while **Chapter 3** describes the contribution of non-covalent interactions in the small molecules' recognition by JAK1. **Chapter 4** describes the effect of phosphorylation on the structural stability and conformational dynamics of JAK1 and inhibition mechanisms by SOCS1. In **Chapter 5,** we have discussed how the binding of various types of inhibitors changes the conformational dynamics of JAK2. Further, we have deciphered why type II allosteric inhibitors are more specific to particular JAKs than a competitive inhibitor. Our simulation results agree with the experiments. We have provided detailed mechanistic insights into an allosteric inhibitor-induced conformational change in JAK2 from an active to an inactive state using the Gaussian accelerated molecular dynamics (GaMD) simulations. Further, **chapter 6** addresses the discovery of novel competitive inhibitors against JAK1 for treating auto-immune disorders and neuroinflammation. The challenges are involved in the ATP-competitive inhibitor design against the highly conserved isoform families. Here, through the conventional method of high-throughput virtual screening (HTVS), standard docking, and multiple replicas molecular dynamics simulations, we could identify the probable phytochemicals that may serve as prospective inhibitors in the novel anti-autoimmune therapies. Finally, **chapter 7** summarizes my dissertation and sketches future directions.

Chapter 2

Chapter 2

Theoretical Background

"…, it is that all things are made of atoms, and that everything that living things do can be understood in terms of the jigglings and wigglings of atoms."

-Richard Feynman

2.1. Molecular modeling of Proteins

Molecular modeling is a very well-known word in the world scientific community and in which we can mimic the behavior of molecules and molecular systems virtually (computer modeling). Proteins are dynamic polymers; in most cases, their functions are linked with their conformational flexibility. In structural biology, there are three commonly known tools for determining the three-dimensional structures of protein: nuclear magnetic resonance (NMR)[100] spectroscopy, X-ray crystallography[101], and Cryo-electron microscopy[102]. All these experimental tools have provided only static snapshots of native state or time-average data. However, understanding the biological process requires scrutiny at spatial and temporal resolutions that can be challenging for the current experimental techniques[24]. Over the past several decades, molecular modelings have become an indispensable tool for understanding the dynamics of biomolecules in atomic precision at the relevant spatio-temporal scales[24,103,104]. According to Klaus Schulten (2015 national lecture), these modeling techniques have served as "**computational microscopy**," providing information previously unattainable by experimental laboratory techniques. Some of the popular molecular modeling techniques are Quantum mechanics (QM) methods (e.g., *ab initio* molecular orbital or density functional theory (DFT) calculations)[105], **Molecular dynamics (MD) simulations,** and Coarse-grained MD[106]. **Figure 2.1** depicts the relationship between various temporal biological phenomena with their suitable molecular modeling techniques.

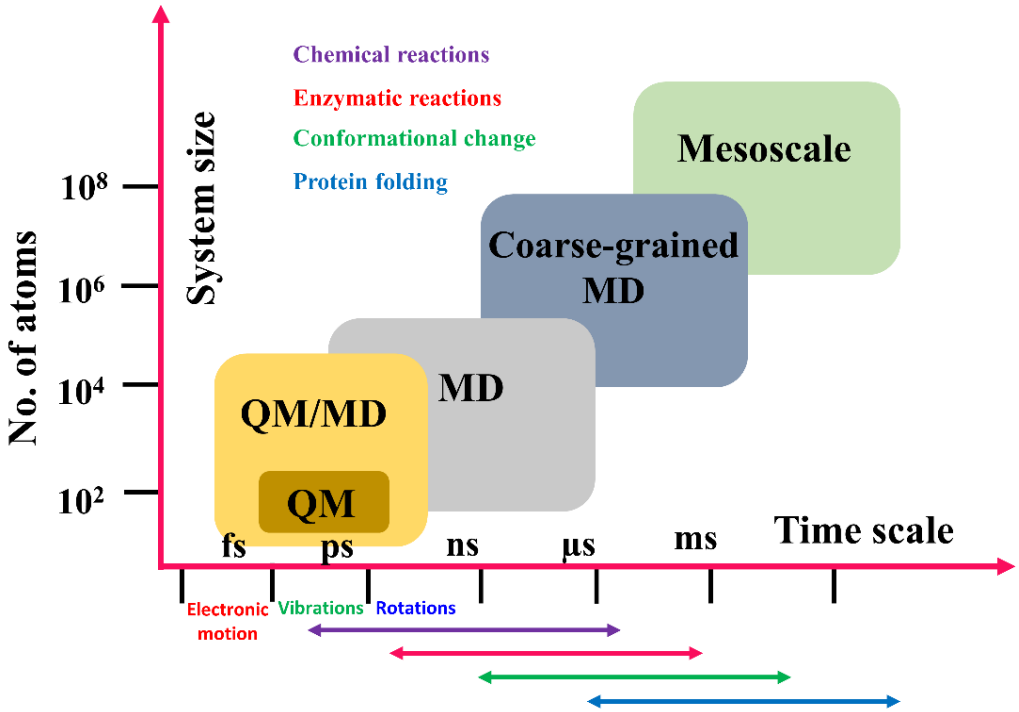


Figure 2.1: Schematic representation of the spatio-temporal resolution of various molecular modeling techniques. The figure is inspired by Aminpour et al. [107].

2.1.1. Quantum mechanics method

In the early twentieth century, physicists were able to describe the motion of small particles such as electrons and nuclei of atoms with the help of some set of rules/equations called **quantum mechanics** (QM)[108]. The QM method can be rigorously applied to calculate the thermodynamics (heat capacity) and molecular properties (molecular geometries, diploe moments, barrier to internal movement, conformational energies) of molecules in chemistry. QM provides the most accurate description of the system at its atomic level. However, applying QM for larger systems of interest is still difficult.

QM calculations depend on the solutions of the time-dependent **Schrödinger's** equation[109] of the entire system. For a single particle 1-dimensional system, the future state vector of a system forms its present state vector is defined as:

$$-\frac{\hbar\, d\psi(x,t)}{i\, dt} = -\frac{\hbar^2\, d^2\Psi(x,t)}{2m\, d^2x} + V(x,t)\psi(x,t) \qquad [2.1]$$

Where ψ is a representation of the state function of the wave function, $\hbar$ is the Planck's constant divided by 2π, m is the mass of the particle, t is time, x is the coordinate of the particle, and V is the total potential energy function of the system. For atoms of higher atomic number, one way to find an accurate wave function is by first applying the Hartree-Fork method[110] and approximate electron correlation with Møller-Plesset (MP) perturbation theory[111], coupled-cluster theory[112], etc. For polyatomic molecules (for example, biomolecules), more crude calculations become necessary, including density-functional, semi-empirical, etc. However, these methods are still time-consuming and computationally impractical to describe the several nanoseconds to microsecond time evolutions of larger biomolecular systems.

2.1.2. Molecular mechanics method

QM method rigorously resolved the electronic structure problem but failed to tackle the too large atoms systems, as mentioned in **Section 2.1.1**. The molecular mechanics (MM) is quite different from the previously mentioned QM method since it does not deal with the electronic Hamiltonian's wave function or electron density. MM method ignores the electronic degree of freedom and gives a simple model, where molecular behavior is predicted mechanistically based on their atomic coordinates. The cost of doing MM calculations scales as $O(N^2)$, where N is the system's atoms. More specifically, MM provides a reasonable trade-off between the wanted accuracy and the computer time. MM methods are less complicated, fast, and reliable in treating large systems, including enzymes. The drawback of MM is that parameters are derived for ground state systems and cannot accurately represent geometries involved in bond breaking and making processes[103]. The calculations use Born-Oppenheimer approximation[113], which describes the energy of molecules based on the nuclear positions, called potential energy surface, derived from classical mechanics. MM dominates in the molecular docking study of receptor-ligand bonding, where primarily the conformational dynamics play a significant role.

Furthermore, the hybrid **QM/MM** approach can combine the strength of *ab initio* QM calculations (accuracy)[114] and MM (speed) approaches. Thus, these methods have mainly

evolved to explore reactions in biological systems and the characterization of transition states. QM/MM methods are primarily used in reactivity and mechanism elucidation and rarely in docking. Warshel and Levitt's initiated the field of **QM/MM**[115]. Here, the accuracy of QM may be used for the difficult-to-model catalytic active site[116], and that involves substrates or directly participating amino acids to describe the chemistry taking place, while the unparalleled speed of MM may be applied to the parts that have a negligible chemical impact, i.e., for the remaining part of the enzyme, which does not participate directly in the reaction but involves many atoms.

2.2. Force fields

The MM method is sometimes called the **force field** (**FF**) method because it relies almost entirely on the classical mechanic force field. The FF is the functional form for calculating the system's potential energy (V) for a set of atoms as a function of their spatial coordinates. Utilizing the most optimized parameters for the empirical calculations in a mathematical model can achieve the accuracy of the quantum mechanical method. FF accurately describes the system's free energy, and its reliable algorithm helps explore the protein conformational space efficiently and quickly. Force field can be based on ***all-atom*, *united-atom*, or *coarse-grained***. Generally, these models are used for complex large biomolecules like proteins, nucleic acids, *etc.,* to achieve high computational efficiency.

In molecular mechanics, the functional form of potential energy (V_{total}) comprises **bonded terms** related to the covalently linked atoms and the **non-bonded terms** consisting of the long-range and short-range interactions of electrostatics, and van der Waals forces, respectively and is given below:

$$V_{total} = V_{bonds} + V_{angles} + V_{dihedrals} + V_{improper} + V_{vdW} + V_{elec} \quad [2.2]$$

where,

$$\boldsymbol{V}_{bonds} = \sum_{bonds} K_r(r - r_{eq})^2 \quad [2.3]$$

$$\boldsymbol{V}_{angles} = \sum_{angles} K_{\theta}(\theta - \theta_{eq})^2 \quad [2.4]$$

$$\boldsymbol{V}_{improper} = \sum_{improper} \frac{1}{2} K_{\psi}(\psi - \psi_0)^2 \quad [2.5]$$

$$\boldsymbol{V}_{dihedrals} = \sum_{dihedrals} \frac{V_n}{2}[1 + \cos(n\chi - \gamma)] \quad [2.6]$$

$$\boldsymbol{V}_{vdW} = 4\epsilon_{ij} \sum_{LJ,i<j} \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}}\right) \quad [2.7]$$

$$\boldsymbol{V}_{elec} = \sum_{Coul.,i<j} \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \quad [2.8]$$

where r is the bond length, θ is the angle formed between three bonded atoms, χ is the dihedral formed between four bonded atoms, r_{eq}, K_r, θ_{eq}, K_{θ}, K_{ψ}, V_n, γ, n are constants for bonded terms. ϵ_{ij} is the well depth, A_{ij} and B_{ij} are Lennard-Jones parameters and r_{ij} is the distance between atoms *i* and *j*, and ϵ_0 is the vacuum permitivity.

The schematic representation of all the potential components is shown in **Figure 2.2**. Equations (**2.2** to **2.6**) related to the intramolecular contributions to the total energy. It includes bond stretching, bending of angle, proper and improper torsional terms. Equations **2.7** and **2.8** include intermolecular terms, i.e., van der Waals **(6-12 Lennard Jones potential**) and electrostatic interactions. Apart from these key terms, the force field may include a few other terms to minimize the deviations from experimental observation and the treatment of hydrogen bonding interactions.

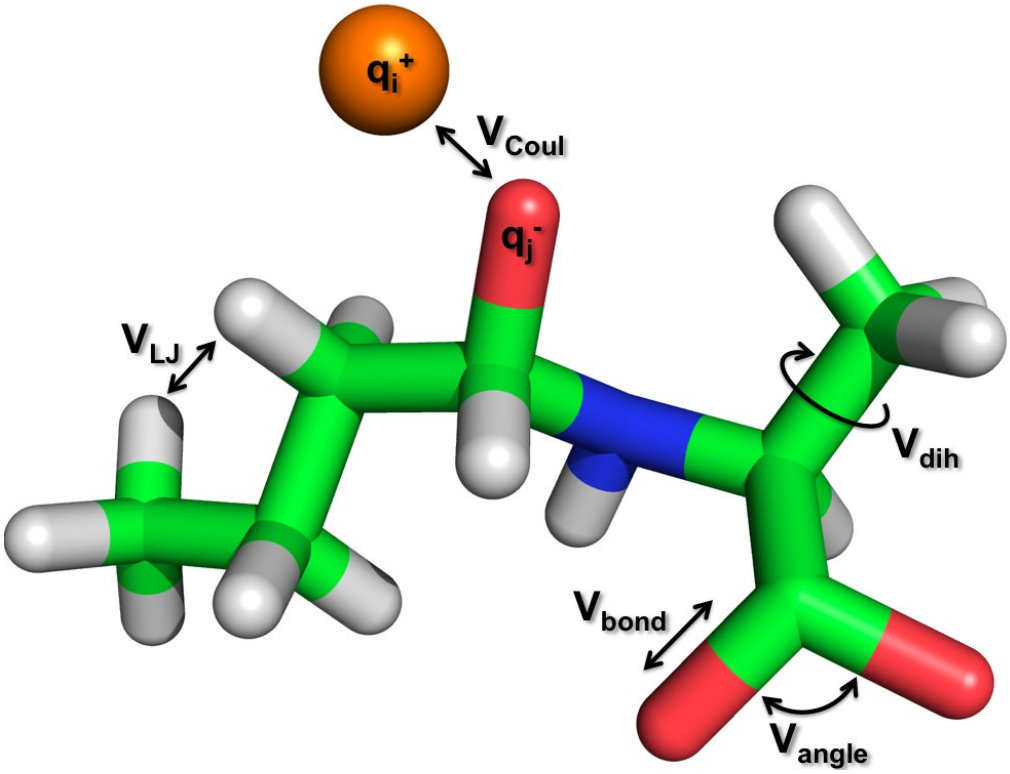


Figure 2.2: The stick representation of a molecules along with all bonded and nonbonded interactions component of the force field. Where $\mathbf{V_{LJ}}$, $\mathbf{V_{dih}}$ and $\mathbf{V_{Coul}}$ represented van der Waals, dihedrals, and electrostatics potentials, respectively.

2.2.1. Popular force fields

Some of the popular FFs, which are commonly used in biomolecular simulation are

- Assisted Model Building with Energy Refinement (**AMBER**)[117,118]
- Chemistry at HARvard Macromolecular Mechanics (**CHARMM**)[119]
- Optimized Potentials for Liquid Simulations (**OPLS**)[120]
- GROningen MOlecular Simulations (**GROMOS**)[121]

Choosing a better FF is a crucial task in any kind of molecular modeling technique. Although, there was no such kind of unified FF for all types of molecules. Therefore, we used system-specific FF to get higher accuracy in biomolecular modeling. Here, I listed the widely used available FF for different biomolecular systems.

- Proteins [ff19SB[122], **ff14SB**[123], ff99SB[124], **OPLS3**[125]**,** CHARMM36m[126]]
- Phosphorylated Protein [phosaa19SB, phosaa14SB**, phosaa10**]
- Nucleic Acids [OL15[127], OL21[128], CHARMM36 DNA[129], CHARMM36 RNA[130]]
- Carbohydrates [GLYCAM_06j[131]]
- Lipids [lipids21, CHARMM36 lipids[132]]
- Ligands [**GAFF2**[133]**,** OPLS-AA[134], CGenFF[135]]

Among these, in my thesis, I have used the protein (**ff14SB** & **phosaa10**) and ligands (**GAFF2**) FF available in AMBER packages. So, the **AMBER** protein force fields are discussed in detail in the subsequent section.

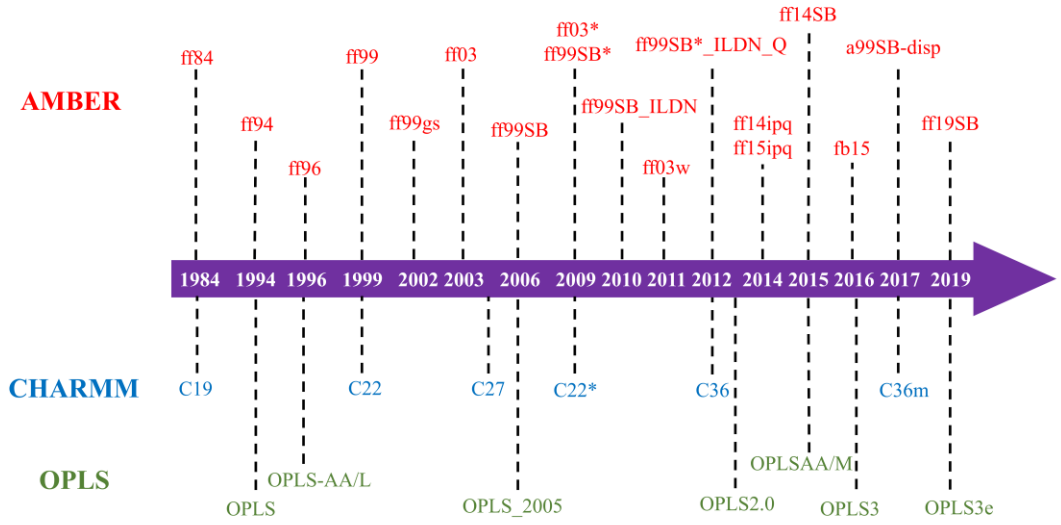


Figure 2.3: Time evolution of protein forcefields. For simplicity only AMBER, CHARMM and OPLS are listed.

2.2.2. AMBER protein force field

AMBER, CHARMM, and OPLS are the main widely-used major families of protein FF. They have very similar functional forms, while the differences mostly come from specific parameters. The time evolution of the protein FF is shown in **Figure 2.3**. For my study, I used **AMBER18** and **AMBER20** molecular dynamic suite and **ff14SB** for protein and updated generalized Amber force field (**GAFF2)** for small molecules or inhibitors. The

AMBER force field comprises five energy terms; covalent bond energy, angle-bending energy, the energy barrier for rotating a bond, van der Waals energy, and electrostatic energy. Hence, the total potential energy of the system follows:

$E_{AMBER} = E_{bond} + E_{angle} + E_{torsional} + E_{vdW} + E_{electrostatic}$

$$= \sum_{bond} \frac{1}{2} k_b (r - r_o)^2 + \sum_{angle} \frac{1}{2} k_a (\theta - \theta_o)^2 + \sum_{torsion} \frac{1}{2} V_n [1 + \cos(n\omega - \gamma)] + \sum_{j=1}^{N-1} \sum_{i=j+1}^{N} \left(4\varepsilon_{i,j} \left[\left(\frac{\delta_{i,j}}{r_{i,j}} \right)^{12} - \left(\frac{\delta_{i,j}}{r_{i,j}} \right)^{6} \right] \right) + \sum_{j=1}^{N-1} \sum_{i=j+1}^{N} \left(\frac{q_i q_j}{4\pi\varepsilon_o r_{i,j}} \right) \quad [2.9]$$

In the case of Amber FF, charges are fixed, which makes it less accurate than the polarizable force fields, such as **AMOEBA**[136] and **DRUDE**[137]. In **equation (2.9),** the vibrational mode for a typical bond is represented by a bond-stretching term. That is based on the assumptions that bonds cannot break in simulation. Thus, this model cannot be applied to chemical reaction studies where covalent bonds can break and form.

2.3. Molecular Dynamics (MD) Simulations

Molecular dynamics (MD) is an important computational tool for understanding the physical basis of the structure, the dynamic evolution of the system, and the function of biological macromolecules. Over the past half-century, MD simulation provided detail in the arena of structural biology by analyzing various dynamics of biological macromolecules in atomic resolution. The experimental techniques, including X-ray, NMR, and Cryo-EM, provide great structural details of molecular events as a static snapshot. However, the events are dynamic and evolve along with time. It is like predicting a football match from a single snapshot, whereas the real fun hides in its video recording. As I mentioned earlier, computer simulation works as a “computational microscope” and can give enormous crucial information at the structural and thermodynamical levels. Therefore, in the next section, I will discuss the mathematical basics of MD simulation.

2.3.1. Fundamentals of MD simulations

In this section, I will briefly discuss the fundamentals of MD simulations, starting from Newton's equations of motions and how these equations involve in the molecular modeling of biological macromolecules.

For a given particle *i* in a system of *N* particles:

$$\boldsymbol{F}_i = m_i \boldsymbol{a_i} \tag{2.10}$$

where $\boldsymbol{a_i}$ and m_i are the acceleration and mass of the i^{th} particle, respectively and $\boldsymbol{F}_i$ is a force working upon i^{th} particle due to the interaction with the rest of the particles present in the system. Here, **bold** letters correspond to a vector quantity and otherwise scalar.

We know that force can also be estimated from the gradient of the potential, as shown below:

$$\boldsymbol{F}_i = -\frac{\delta V(\boldsymbol{r}^N)}{\delta \boldsymbol{r_i}} \tag{2.11}$$

In general, Hamiltonian's equations of motions, the **Hamiltonian** $\boldsymbol{\mathcal{H}}$ relate the position and momenta of the particles to the system's total energy. By which the instantaneous positions ($\boldsymbol{r}_i$) and momenta ($\boldsymbol{p}_i$) of the i^{th} particle can be described as follow:

$$\boldsymbol{p}_i = -\frac{\delta \mathcal{H}}{\delta \boldsymbol{r_i}} \tag{2.12}$$

$$\boldsymbol{r}_i = -\frac{\delta \mathcal{H}}{\delta \boldsymbol{p_i}} \tag{2.13}$$

$$\mathcal{H}(\boldsymbol{p}_i, \boldsymbol{r}_i) = \sum_{i=1}^{N} \frac{\boldsymbol{p}_i^2}{2m_i} + U(\boldsymbol{r_i}) \tag{2.14}$$

Also, for a closed system, the **Hamiltonian** can be written as a sum of **kinetic energy** (E_K) and **potential energy** (V) as a function of coordinates and momenta.

$$\boldsymbol{r} = (\boldsymbol{r_1}, \boldsymbol{r_2}, \ldots\ldots, \boldsymbol{r_N}) \quad [2.15]$$

$$\boldsymbol{p} = (\boldsymbol{p_1}, \boldsymbol{p_2}, \ldots\ldots, \boldsymbol{p_N}) \quad [2.16]$$

$$\mathcal{H}(\boldsymbol{r}, \boldsymbol{p}) = E_K(\boldsymbol{p}) + V(\boldsymbol{r}) \quad [2.17]$$

The kinetic energy can be expressed as:

$$E_K = \sum_{i=1}^{N} \frac{1}{2m_i}(p_{ix}^2 + p_{iy}^2 + p_{iz}^2) \quad [2.18]$$

where m_i is the mass of the i^{th} particle and p_{ix}, p_{iy}, and p_{iz} are the momenta of the same particle along *x*, *y,* and *z* directions.

So, now it is possible to integrate the entire trajectory over time as we have the potential function. However, most cases cannot be solved for an N-body problem. Instead, an approximation of the solution is obtained using several finite-difference integration methods. All are based on the **Taylor series** expansion, as shown below:

$$\boldsymbol{r}(t + \Delta t) = \boldsymbol{r}(t) + \Delta t\mathbf{v}(\mathrm{t}) = \frac{1}{2}\Delta t^2\boldsymbol{a}(t) + \frac{1}{6}\Delta t^3\boldsymbol{b}(t) + \cdots \quad [2.19]$$

$$\boldsymbol{v}(t + \Delta t) = \boldsymbol{v}(t) + \Delta t\mathbf{a}(\mathrm{t}) + \frac{1}{2}\Delta t^2\boldsymbol{b}(t) \quad [2.20]$$

$$\boldsymbol{a}(t + \Delta t) = \boldsymbol{a}(t) + \Delta t\mathbf{b}(\mathrm{t}) \quad [2.21]$$

Here, $\boldsymbol{r}$, $\mathbf{v}$, $\boldsymbol{a}$, and $\boldsymbol{b}$ are the position, velocity, acceleration, and third derivative of position, respectively. Several algorithms are there to perform these calculations, which are needed to be computationally efficient and conserve energy and momentum. Among these, **Verlet algorithm**[138] is one of the well-established and reliable algorithms for these integrations. Here, the **Taylor series** expansions for the positions at time t and $t - \Delta t$ are combined.

$$\boldsymbol{r}(t + \Delta t) = \boldsymbol{r}(t) + \Delta t\mathbf{v}(\mathrm{t}) + \frac{1}{2}\Delta t^2\boldsymbol{a}(t) + \cdots \quad [2.22]$$

$$\boldsymbol{r}(t-\Delta t) = \boldsymbol{r}(t) - \Delta t\mathbf{v}(\mathrm{t}) + \frac{1}{2}\Delta t^2\boldsymbol{a}(t) - \cdots \quad [2.23]$$

Adding equations **2.22** and **2.23** yields:

$$\boldsymbol{r}(t+\Delta t) = 2\boldsymbol{r}(t) - \boldsymbol{r}(t-\Delta \mathrm{t}) + \Delta t^2\boldsymbol{a}(t) \quad [2.24]$$

So, by estimating the positions at current and previous timestep and accelerations at the current timestep, the **Verlet algorithm** can predict the new positions of the system. The velocity of the same system can be computed using the following formula.

$$\boldsymbol{v}(t) = \frac{\boldsymbol{r}(t+\Delta t) - \boldsymbol{r}(t-\Delta t)}{2\Delta t} \quad [2.25]$$

Basically, the **Verlet algorithm** is a two-step method as it calculates the position at two different times. Additional computer memory is required for running the same as storing positions from three consecutive timesteps. So, there is an improvement of the same algorithm, termed as the **velocity Verlet algorithm,** where the velocity is calculated as step $n + \frac{1}{2}$ and then the coordinates at step $n + 1$. So, in this algorithm, positions and velocities are calculated using the formula stated below.

$$\boldsymbol{r}(t+\Delta t) = \boldsymbol{r}(t) + \Delta t\boldsymbol{v}(t) + \frac{1}{2}\Delta t^2\boldsymbol{a}(t) \quad [2.26]$$

$$\boldsymbol{v}(t+\Delta t) = \boldsymbol{v}(t) + \frac{1}{2}\Delta t[\boldsymbol{a}(t) + \boldsymbol{a}(t+\ \Delta t)] \quad [2.27]$$

Here, the position of the next time step is estimated using the equation and which is further used to estimate the acceleration at the new time. Also, the velocity is obtained from the positions and acceleration using equation **2.27**.

Another modification of the **Verlet algorithm** is the **leapfrog algorithm**[139]. Here, the velocities $\boldsymbol{v}$ are calculated at time $t + \frac{1}{2}\Delta \mathrm{t}$ which are used to calculate the positions at time $t + \Delta t$. So, velocities leap over the positions and the positions leap over the velocities give the advantage of explicitly calculating the velocity, not in sync with the positions.

$$\boldsymbol{r}(t+\Delta t) = \boldsymbol{r}(t) + \boldsymbol{v}(t+\frac{1}{2}\Delta \mathrm{t})\Delta t \quad [2.28]$$

$$\boldsymbol{v}\left(t+\frac{1}{2}\Delta \mathrm{t}\right) = \boldsymbol{v}\left(t-\frac{1}{2}\Delta \mathrm{t}\right) + \mathbf{a}(\mathrm{t})\Delta \mathrm{t} \quad [2.29]$$

The advantage of the leapfrog algorithm over the Verlet algorithm is that the previous one set of positions and velocity needs to store where the next three consecutive positions are required for one set of velocities.

2.3.2. Simulation time-step

One of the most key components found and discussed in all previous algorithms is the timestep, Δt. The numerical stability and accuracy in conserving the energy and integrating the equations of motion depend on the choice of the timestep. One typically needs to pick a time step that is at least one order of magnitude lower than the fastest time scale in the system, which for all-atom simulations is usually bond vibrations (~ 10^{-13} sec). Too large timesteps can cause an MD simulation to become unstable, with E (energy) increasing rapidly with time. However, all bonds involving hydrogen atoms (high frequency bonds) are constrained using the LINCS[140], SHAKE[141] algorithm. This helps us to use a larger time step of 2 fs. Overall, in practice, 1-2 fs timestep is reasonable for biomolecular simulations along with the restraint on bonds involving hydrogen atoms.

2.3.3. Solvent model

The solvent model plays an important role in dictating the stability, flexibility, and interactions of chemical and biomolecular systems. There are two different solvent models available in MD simulations of biomolecules. One is the continuum solvent approach or **implicit solvent** model, and the other is the **explicit solvent** model. The explicit solvent model offers a more accurate and details description of a biomolecular system. Therefore, in the next section, I will discuss the implicit and explicit solvent models in detail.

2.3.3.1. Implicit solvent model

The implicit model characterized the solvent as a continuous medium surrounding the solute molecule, in which solvents are not considered atomistically but rather treated as a

continuum dielectric medium that considers the average bulk properties of the solvent. In this model, the solvent energy component, known as the solvation free energy ΔG_{solv} comprises of three terms: energy that is required to create a cavity in the solvent by the solute (ΔG_{cav}); the electrostatic interaction energy between solute and solvent (ΔG_{elec}); the short-term van der Waals interaction energy between them (ΔG_{vdW}).

$$\Delta G_{solv} = \Delta G_{cav} + \Delta G_{elec} + \Delta G_{vdW} \quad [2.30]$$

There are different continuum models that have been used in these calculations, including Solvent-Accessible-Surface-Area (SASA), Poisson-Boltzmann Equation (PB/PBE), Generalized Born (GB), and Generalized Born Surface area (GBSA). The PBE is an accurate model but computationally very expensive. Therefore, other approximate methods, such as GB approximation[142–144], are used. In this model, an electrostatic interaction term is added between each solute and solvent atom in which distance is replaced by a parameter termed Born radius (f_{GB}). It scrutinizes whether the solute atom is buried or solvent expoxed. The mathematical formulation of this electrostatic term is shown below:

$$E_{GB} = -\frac{1}{8\pi\epsilon_0}\left(1-\frac{1}{\epsilon}\right)\sum_{i,j}^{N}\frac{q_i q_j}{f_{GB}} \quad [2.31]$$

where ϵ_0 is the permittivity of free space, and ϵ is the permittivity of the solvent, q_i and q_j are the charges on atoms i and j, respectively. This method is computationally very fast. However, this model ignores several key solute-solvent interactions. As a result, it hovers between computational accuracy and speed. In this research study, I have used PB(GB) combined with the SASA to calculate the polar and non-polar solvation energy for estimating the binding free energy.

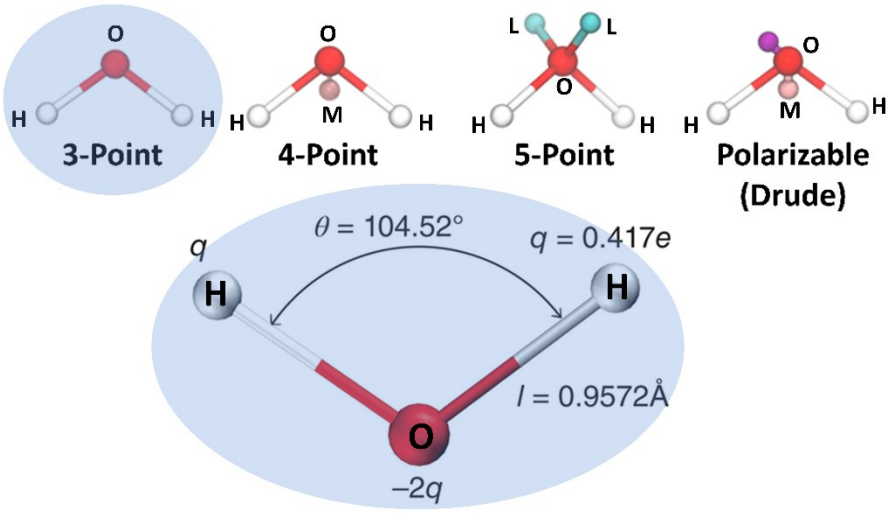


Figure 2.4: TIP3P widely used 3-point water model including key parameters, *i.e.,* bond distance and HOH angle. The figure is inspired by ref.[145,146].

2.3.3.2. Explicit solvent model

In the context of biomolecular simulation, we mostly considered the water molecules as a solvent. In the explicit solvent model, the solvent viscosity effect and hydrogen bond bridges are considered. Thus it is more accurate compared to the implicit model. In my study, I also choose **water** as an explicit solvent. A wide range of water as solvent models are available. These models could be **3-point**, **4-point**, **5-point,** and **polarizable**, as shown in **Figure 2.4.** The widely used **TIP3P**[147]**, TIPS**[148]**, SPC/E**[149]**,** and **SPC**[150] models are based on the 3-site approach, and the main difference among them is the HOH bond angle. For instance, SPC keeps 109° to enforce a tetrahedral bond angle, while TIP3P chooses the actual observed value for water (105°). Among these 3-site solvent models, TIP3P reports reproducing many of the thermodynamics properties.

The 3-point water model underestimates the viscosity effect of water; therefore, diffusion-based phenomena occur more quickly compared to reality. To overcome this issue, the 4-point water model **TIP4P**[151] is introduced, in which a dummy atom is placed behind the oxygen atom. This atom is massless and only carries an electrostatic charge which should assign to the oxygen atom. By this estimation, the polarity of the water molecules becomes close to reality, and the diffusion properties are also improved. A few more water models are available, including **TIP5P**[152,153], and **TIP6P**[154]**,** which increase the accuracy of water models with a high computational cost.

2.3.4. Periodic boundary conditions

Periodic boundary conditions allow us to estimate bulk properties more accurately. In a real physiological condition, along with solute, many solvent molecules are also involved in the systems. Due to the limitation of computational cost and neglecting the surface effect of solvent molecules, the number of atoms in the simulation system is limited. Hence, the periodic boundary condition (PBC) is used to encounter these issues, which creates an illusion of a bulk environment in a simulation. In PBC, the solute and solvent molecules are enclosed in a box replicated in all three Cartesian dimensions to give a periodic array completely filling the space.

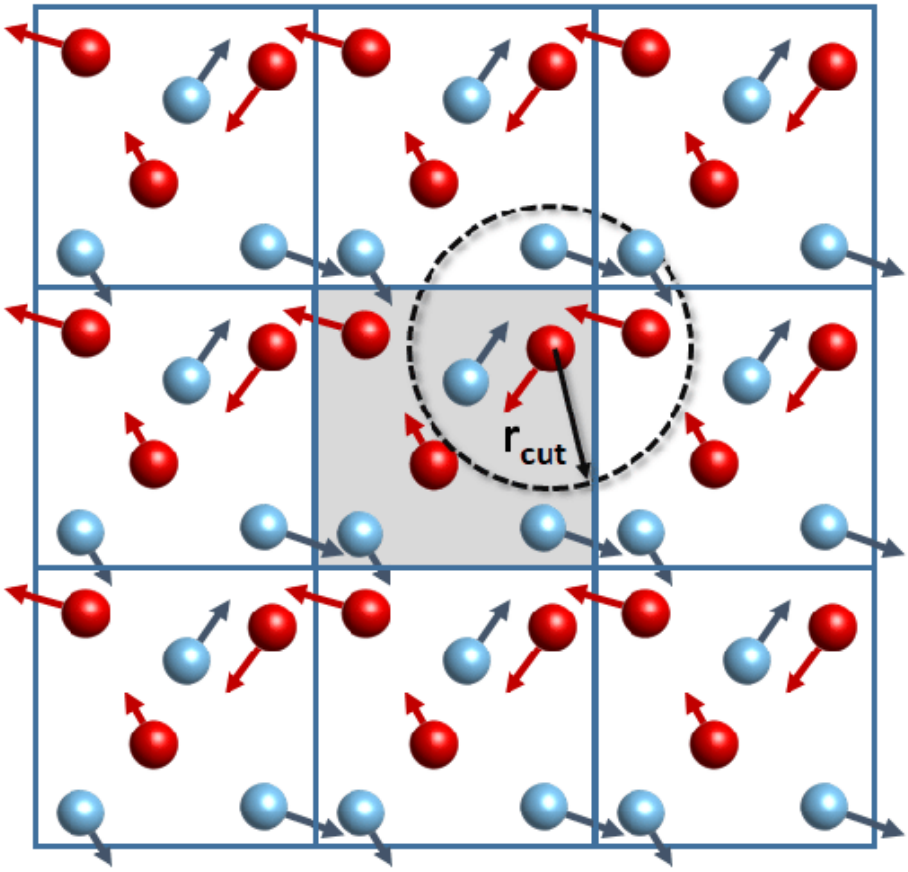


Figure 2.5: Schematic 2D representation of periodic boundary conditions with the simulation box (grey) surrounded by image boxes. The velocity of every atom is depicted by its corresponding vector. Short-range interactions are also defined by r_{cut}.

In PBC calculation, the unit cell of a system is surrounded by an infinite number of periodic images in all directions (**Figure 2.5**). Therefore, the system's atoms do not fall to the surface, which eventually reduces the surface effect[108]. The total number of atoms in the simulation box stays the same at any given time. When any atom leaves the unit cell, a duplicate of it enters from the opposite side with an identical velocity. In the PBC condition, atoms in the simulation box also interact with their images in the surrounding box. This way, the interaction between particles at or beyond a finite length can be neglected. This maximum length must be equal to or less than half of the simulations' box length (L). A potential cutoff sphere or radius (r_{cut}) must be less than L/2 and is defined as when a particle is separated by a distance equal to or larger than r_{cut}; two particles do not interact with each other. It will help to avoid expensive force calculations.

2.3.5. Statistical ensembles

One of the main objectives of MD simulations is to estimate/predict the thermodynamic behavior of real systems as observed in the laboratory. Typically, this means we are interested in macroscopic systems consisting of 10^{23} particles or more. So, the "**ensembles**" can be defined as a collection of a number of macroscopically identical but essentially independent **systems**. Here, the term macroscopically identical means as each of the systems constituting an ensemble satisfies the same macroscopic conditions, like **volume** (*V*), **energy** (*E*), **pressure** (*P*), **temperature** (*T*), and the **total number of particles** (*N*), etc. Statistical mechanics provides a link between microscopic details of the system and its macroscopic observables. The MD simulation with the macroscopic number of particles is computationally unrealistic as per the current development. So, the involvement of the statistical mechanical method to extract information is mandatory. A macroscopic observable consists of several distinct microscopic states. An ensemble is all these possible microstates with a certain set of constant parameters. For example, any isolated systems belong to the **microcanonical** ensemble (*NVE*), where the total **number of particles** (*N*), **volume** (*V*), and **total energy** (*E*) remain constant. In this ensemble, the energy of each microstate needs to remain constant. For that, exploration of potential energy surface by a system is restricted in this ensemble due to higher free energy barriers respective to the total energy of the system. Alternative to the microcanonical ensemble is the **canonical ensemble** (*NVT*) and **isobaric-isothermal** (*NPT*) ensemble, where volume and

temperature or pressure and temperature remain constant. The *NVT* ensemble is mostly used in the heating phase as it resembles a closed system, where heat can be transferred from an external source. The production run of MD simulation mostly uses the *NPT* ensemble to represent the real situation. However, in these two ensembles, the system's energy is not fixed, indicating that systems can spontaneously overcome the large potential energy barrier by fluctuations in total energy.

2.3.6. Temperature and pressure coupling

MD simulations are used to observe and glean properties of interest from some systems of study. Typically, thermodynamics properties of interest are measured under open-air conditions in the laboratory. To emulate laboratory conditions (exposed to the surroundings), sampling in the **canonical** (constant *T*) or **isobaric isothermal** (constant *P*) ensemble is desired. Therefore, some thermostat or barostat algorithms will be employed to control the temperature or pressure during the MD simulation. The temperature of MD simulation is basically measured in terms of the time average of kinetic energy (equipartition theorem).

$$< E_k >= \frac{1}{2}\sum_{i=1}^{N} m_i v_i^2 = \frac{3}{2} N k_B T \qquad [2.32]$$

Therefore, the calculated time-averaged temperature is referred to as the instantaneous temperature. The instantaneous temperature will not always be equal to the target temperature; in fact, in the NVT ensemble, the instantaneous temperature should undergo fluctuations around the target temperature, T.

2.3.6.1. Popular Thermostats

There are several thermostats available to maintain a constant *T* during MD simulations, like **Berendsen**[155], **Andersen**[156], **Langevin**[157]**,** and **Nose-Hoover**[158] thermostats. For example, Langevin and Anderson's thermostats mainly work by constraining a system variable to a pre-set distribution function. In Anderson thermostat, one of the first thermostats developed for MD studies, assigns a random particle a new velocity from a Boltzmann distribution at the pre-selected temperature by a collision event. In my thesis, I used a Langevin thermostat. It uses a general equation of the form,

$$F = F_{interaction} + F_{friction} + F_{random} \quad [2.33]$$

Where $F_{interaction}$ is the standard interactions calculated during the simulation, $F_{friction}$ is the damping used to tune the "viscosity" of the implicit bath, and F_{random} effectively gives random collisions with the solvent molecules. The frictional and random forces are coupled through a user-defined friction damping parameter.

2.3.6.2. Popular barostat

Like thermostats, there are several barostats available to maintain the constant pressure (P) during the simulation in the *NPT* ensemble. These include **Berendsen**[155], **Andersen**[156], **Parrinello-Rahman**[159], and **Martyna-Tuckerman-Tobias-Klein**[160,161] (MTTK) barostat. Barostat algorithms control pressure alone, not temperature. Therefore, if the target ensemble is *NPT*, then a barostat must be applied with a thermostat. In this thesis, I used the Berendsen barostat and the Langevin thermostat to control the pressure and temperature, respectively. The popular Berendsen barostat couples the pressure with a pressure bath, which is shown below;

$$\lambda = 1 - \kappa \frac{\delta t}{\tau_P}(P - P_{bath}) \quad [2.34]$$

$$\boldsymbol{r}_i = \lambda^{\frac{1}{3}} \boldsymbol{r}_i \quad [2.35]$$

Here, r_i is the rescaled coordinates, τ_P is the time constat, and κ is the isothermal compressibility that mainly governs the strength of coupling to the pressure bath. The isothermal compressibility is also related to the volume fluctuations.

$$\kappa = \frac{1}{V}\left(\frac{\delta V}{\delta P}\right) T = \frac{1}{k_B T} \frac{< V^2 > - < V >^2}{< V^2 >} \quad [2.36]$$

There are several options for scaling the pressure, like isotropic pressure scaling (simultaneously scaled all dimensions) and semi-isotropic pressure scaling (any given dimensions are scaled).

2.3.7. Long-range interactions

The electrostatic (Coulombic) interactions in the molecular systems between charged particles fall off as *1/r*, where *r* is the distance separating the particles. The atom-atom interactions are thus long-range compared to other interactions in these systems (which fall off as $1/r^3$ of faster). This means atoms or molecules separated by considerable distances can still have strong electrostatic interactions. However, this depends on the degree of shielding of the intervening medium (or its relative permittivity of dielectric constant). This means that the strength of the interaction is 80 times less in water than in a vacuum.

The number of bonded interactions remains unchanged during the MD simulation and is less exhaustive to measure. While the non-bonded interactions are generally the most time-consuming step of molecular mechanics or classical energy calculations, the strength and importance of non-bonded interactions vary substantially as a simulation proceeds. The non-bonded interactions play a significant role in the stability of proteins' tertiary/quaternary structure. They are the prominent forces in maintaining the structural integrity of proteins. These long-range interactions may be relevant but determining pairwise distances is an expensive computation that grows with the square of the number of atoms involved. The main idea here is to neglect certain interaction pairs to increase the simulation efficiency and maintain the authenticity of the produced results. Both Lennard-Jones (LJ) and Coulombic interactions can be ignored beyond a certain distance as their effect is negligible compared to the error of the simulations. For that reason, a cut-off scheme is used to ignore all long-range interactions beyond a certain distance. The intermolecular electrostatic interaction persists over a longer range than LJ interaction, which hinders the above method's usage. For that, the **Particle Mesh Ewald** (PME)[162] scheme is most widely used for calculating the long-range electrostatics in the periodic system during MD simulation.

To understand this PME technique, consider the relationship between the charge distribution and the Coulombic potential written in the differential form (the Poisson equation):

$$\nabla^2 \varphi(\boldsymbol{x}) = -(1/\varepsilon)\rho(\boldsymbol{x}) \qquad [2.37]$$

Where $\varphi(\boldsymbol{x})$ is the potential at point $\boldsymbol{x}$, $\rho(\boldsymbol{x})$ is the charge density at the same $\boldsymbol{x}$ point, and ε is the permittivity of the medium. To solve this equation, consider ρ and φ are smooth functions, and due to point charge, ρ is a delta function. Moreover, the electrostatic charges are distributed as a grid from which potentials are calculated. Finally, the force acted upon each particle is calculated based on its position in the grid space.

2.4. Main steps of an MD simulation

While every system studied will present unique challenges and considerations, the process of performing an MD simulation generally follows these steps:

2.4.1. Preparation of system topology

System preparation is arguably the most critical stage of a simulation and, in many cases, receives the least attention; if your system preparation is flawed, such flaws may prove fatal. It is mainly focused on preparing the starting structure of the desired system for input to an appropriate simulation package. The preparation MD simulation includes building an initial structure (from X-ray, NMR, Cryo-EM, or modeled), solvating (if necessary), applying an appropriate recent force field, and the charged residue protonation states assignment system charge neutrality, salt concentrations, etc. To mimic the biological condition, solvent molecules (water) are added to the systems. The number of added solvent molecules is directly proportional to the computational cost and effect on the pair-wise long-range interactions calculation also. Therefore, the solvent box size is crucial for simulation, and typically, the solvent box length should be twice the long-range interaction cut-off distance.

2.4.2. Minimization

Energy minimization is a widely used method in molecular modeling and integral to conformational search procedures. The purpose of minimization or relaxation is to find a local energy minimum of the starting structure so that the molecular dynamics simulation does not immediately "**blow up**" (i.e., the forces on any one atom are not so large that the atoms move an unreasonable distance in a single timestep). In addition, this step helps to relieve any steric clashes or unfavorable interactions in the system's initial configuration.

This involves standard minimization algorithms, including the steepest descent, conjugate gradient, Newton-Raphson, etc. These methods find the nearest local minima and are indispensable in correcting structural anomalies like short contacts or bad stereochemistry, and longer iteration steps are recommended. Basically, in most of the simulation protocols, minimization is conducted in two stages.

- **First Step**: Solute molecules are restrained to minimize the solvent molecules
- **Second Step**: Both solvent and solute are minimized without restraining

In this way, it prevents the distortions of the conformation due to the high energetic interaction between solvent and solute.

2.4.3. Heating

There is no temperature in the initial conformation, which means there is no energy in the system. In this stage, we gradually increase the kinetic energy to reach the desired temperature for a given system. The temperature should increase gradually over a proper timeframe or step-wise to avoid a sudden increase in the kinetic energy, making the simulation unstable. In the *NVE* ensemble, energy cannot be added, and in the *NPT* ensemble, the increasing kinetic energy expands the system to maintain the constant pressure. For the above reason, the ensemble for the heating step is *NVT*. Then, velocities are reassigned from a new Maxwell-Boltzmann distribution at each integration step, increasing the temperature. Heating stages reduce the chances of the system blowing up by equilibrating at each temperature. Alternatively, by using weight restraints, one can control the heating.

2.4.4. System equilibration

As the production run is conducted under the NPT ensemble, a buffer time is required between the heating and production run to switch the ensemble. The equilibration stage aims to evolve the system from its initial configuration to the equilibrium state. Here, certain thermodynamic properties, such as energy, temperature, pressure, etc., are monitored. Therefore, the equilibration is performed in the *NVT* ensemble, which allows equilibrating the kinetic and potential energies. That is, to distribute the kinetic energy

provided into the system during heating among all degrees of freedom. It usually implies that the potential energy "lags on" and must be equilibrated with the kinetic energy. The equilibration stage is completed as soon as the potential energy reaches a plateau. After that, the initial rapid decrease of potential energy fluctuates near a constant level, and other key properties no longer change with time. This behavior suggests that the system is equilibrated on the specified time scale.

2.4.5. Production run

Once equilibration is completed, we may begin collecting data for analysis. Typically, this phase is called the "production" phase. The main difference between equilibration and production is that we plan to retain and analyze the collected data in the production simulation. However, parameters are the same in both the equilibration and production stages. The production simulation must always be preceded by equilibration appropriate for the target production ensemble, and production data should never be collected immediately after a change in conditions (such as rescaling a box size, energy minimizing, or suddenly changing the temperature or pressure) except in very specific applications where this is the goal. Sometimes, we choose to discard some initial production data as additional equilibration. Usually, this is done to allow additional equilibration time after a change in the protocol (such as a switch from NVT to NPT or vice versa). The production simulation time ranges from a few nanoseconds to several microseconds. With the availability of the powerful GPU (Graphical Processing Unit), the microsecond long production stage is now a routine practice[24].

2.5. Enhanced sampling method

The atomistic MD simulation is one of the most widely used techniques in computational chemistry and biology because of its simplicity and ability to accurately sample the conformational space of a biomolecular system[25]. Due to high free energy barriers, many interesting biological phenomena, such as protein-ligand binding and protein conformational change, take place at timescales that are well beyond the capacity of

present-day supercomputers with conventional MD simulation. These processes are referred to as **rare events**[163–165]. Several techniques are available to enhance this transition probability, which helps us to sample more conformational spaces in the same time duration. These techniques are commonly called "enhanced sampling." Some of the popular enhanced techniques are umbrella sampling[166,167], metadynamics[168,169], adaptive biasing force (ABF)[170,171], steered MD (sMD)[172], replica exchange MD (REMD)[173,174], conformational flooding[175,176], parallel tempering[177], accelerated molecular dynamics (aMD)[178], Gaussian accelerated MD (GaMD)[179,180] and Ligand Gaussian accelerated MD (LiGaMD)[181], etc. In my study, I have used aMD and GaMD simulations for the conformational plasticity analysis of JAK kinases and protein-ligand complexes.

2.5.1. Accelerated molecular dynamic simulations (aMD)

The general idea behind the accelerated molecular dynamics scheme is depicted in **Figure 2.6**. A continuous non-negative bias boost potential function[178], $\Delta V(\boldsymbol{r})$ is defined such that when the original potential, $V(\boldsymbol{r})$ is below a certain value E (threshold energy), then the simulation is performed on the modified potential, *V*(r)*, represented using the dashed lines. When *V(r)* is greater than E, the simulation is performed on the original potential. Then this modified potential, $V*(r)$ can be written as,

$$V*(\boldsymbol{r}) = \begin{Bmatrix} V(\boldsymbol{r}) & ; & V(\boldsymbol{r}) \geq E \\ V(\boldsymbol{r}) + \Delta V(\boldsymbol{r}); & & V(\boldsymbol{r}) < E \end{Bmatrix} \quad [2.38]$$

The boost potential can be calculated from the following equations,

$$\Delta V(\boldsymbol{r}) = \frac{(E_p - V(\boldsymbol{r}))^2}{(\alpha_P + E_P - V(\boldsymbol{r}))} + \frac{(E_D - V_D(\boldsymbol{r}))^2}{(\alpha_D + E_D - V_D(\boldsymbol{r}))} \quad [2.39]$$

$$E_D = E_{D'} + \alpha_1 N_{res}/5 \quad [2.40]$$

$$\alpha_D = \alpha_1 x N_{res}/5 \quad [2.41]$$

$$E_P = E_{P'} + \alpha_2 N_{atom} \quad [2.42]$$

$$\alpha_P = \alpha_2 x N_{atom} \qquad [2.43]$$

Here, $V(\boldsymbol{r})$ is the actual potential, $V_D(\boldsymbol{r})$ is the actual dihedral potential, N_{res} is the number of residues, N_{atom} is the total number of atoms, $E_{P'}$ and $E_{D'}$ are the average potential and average dihedral angles, respectively. Using equation [**2.40, 2.42**], threshold energies (E_D, E_P) are calculated. The aMD process can be applied in two ways, i.e., it can boost only dihedral potentials or dual boost, where it boosts whole potential and dihedral potential combined. Reweighting is required to recover the true potential from the modified potential, which can be done by exponential average, Maclaurin series expansion, or cumulant expansion. Therefore, aMD is a robust bias potential simulation technique that helps us explore and overcome the high energy barriers without any advance knowledge of the location of the potential energy wells or saddle points.

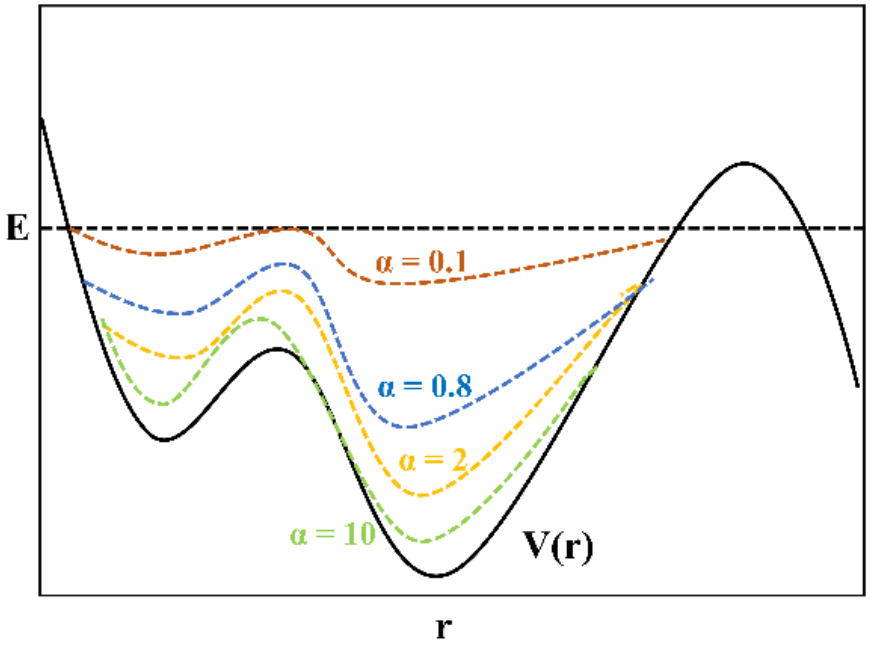


Figure 2.6: The schematic representation of the aMD method. When the original potential $V(r)$ (solid line) falls below threshold value E, a boost potential is added. The modified potential $V^*(r)$ in color representation (dashed line) with different values of the acceleration factor α. The larger the value of α, it is closer to the original surface. Largest α in green color, and smaller α in brown color. It is adopted from Hamelberg *et al.* [178].

2.5.2. Gaussian accelerated molecular dynamic simulations (GaMD)

Gaussian accelerated molecular dynamics (GaMD) is a robust computational method for simultaneous unconstrained enhanced sampling and free energy calculations of biomolecules. It works with the same theoretical background as aMD[178]. GaMD greatly accelerates biomolecular simulations by orders of magnitude. Like aMD, no predefined collective variable (CV) is needed for GaMD simulations. Furthermore, the added harmonic boost potential in GaMD exhibits a Gaussian distribution (**Figure 2.7**), which enables us to accurately recover the original biomolecular free energy landscapes by Gaussian approximation cumulant expansion to the second order[182]. This useful scheme substantially reduces the statistical noise, thereby overcoming the limitations of the early aMD technique.

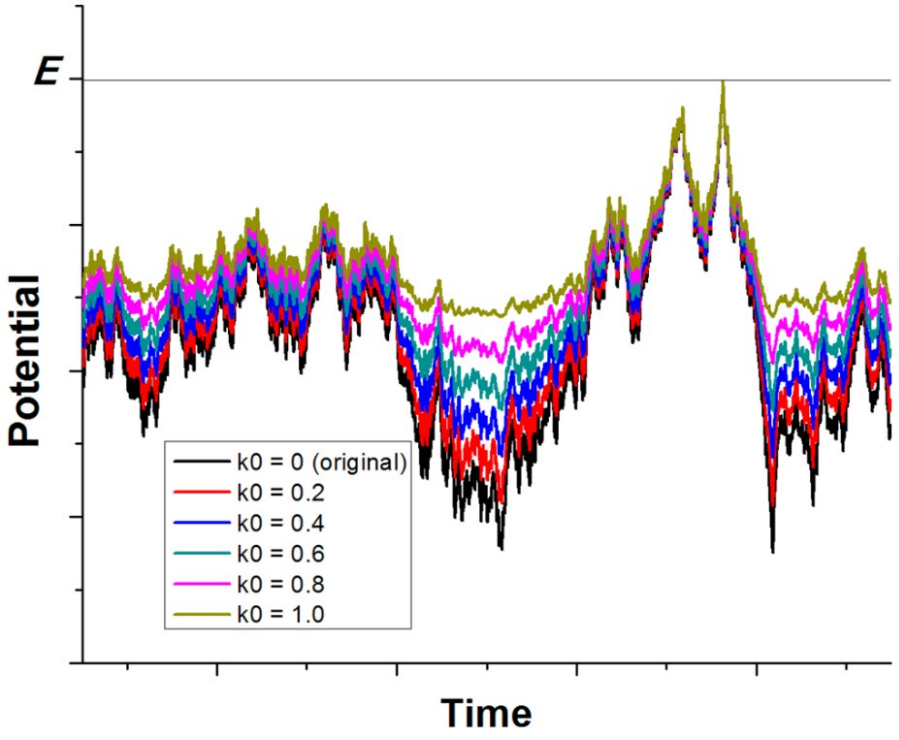


Figure 2.7: Scheme illustration of GaMD. When the threshold energy is set to the maximum potential ($E = V_{max}$), the system potential energy surface is smoothened by adding a harmonic boost potential that follows the Gaussian distribution. The coefficient κ_0 in the range of 0-1 determines the magnitude of the applied boost potential. With greater κ_0, higher boost potential is added to the original energy surface in conventional MD, which provides an enhanced sampling of biomolecules across decreased energy barriers. It is adapted from Miao *et al.* [179].

A brief mathematical formulation of GaMD technique[179] is shown below:
Assume that a system having N atoms at position $\boldsymbol{r} = \{\boldsymbol{r}_1,...\boldsymbol{r}_n\}$, when the system potential $V(\boldsymbol{r})$ is lower than the threshold energy (E), then the modified potential $V^*(\boldsymbol{r})$ and the harmonic boost potential $\Delta V(\boldsymbol{r})$ are calculated as follows.

$$V^*(\boldsymbol{r}) = V(\boldsymbol{r}) + \frac{1}{2}k(E - V(\boldsymbol{r}))^2, if\ V(\boldsymbol{r}) < E \quad [2.44]$$

$$\Delta V(\boldsymbol{r}) = \frac{1}{2}k(E - V(\boldsymbol{r}))^2, if\ V(\boldsymbol{r}) < E \quad [2.45]$$

Where k is a force constant. The boost potential $\Delta V(\boldsymbol{r})$ will be equal to zero when system potential $V(\boldsymbol{r})$ becomes higher than reference energy (E). The applied boost potential follows subsequent steps to smoothening the potential energy surface (PES).

For any two arbitrary potential energy ($V_1(\boldsymbol{r}), V_2(\boldsymbol{r})$ found on the original energy surface; where $V_1(\boldsymbol{r}) < V_2(\boldsymbol{r})$ and the ΔV potentials satisfy $V_1^*(\boldsymbol{r}) > V_2^*(\boldsymbol{r})$, then equation [2.44] can be expressed as follows:

$$E < \frac{1}{2}(V_1(\boldsymbol{r}) + V_2(\boldsymbol{r})) + \frac{1}{k} \quad [2.46]$$

and if $V_1(\boldsymbol{r}) < V_2(\boldsymbol{r})$ and the difference in modified potential energy surface should be smaller than the original energy surface, that is $V_2^*(\boldsymbol{r}) - V_1^*(\boldsymbol{r}) < V_2(\boldsymbol{r}) - V_1(\boldsymbol{r})$ and the equation [2.44] can be modified as.

$$E > \frac{1}{2}(V_1(\boldsymbol{r}) + V_2(\boldsymbol{r})) \quad [2.47]$$

Combining equation [2.46] and [2.47] and using the relationship, $V_{min} \leq V_1(\boldsymbol{r}) < V_2(\boldsymbol{r}) \leq V_{max}$, the threshold energy E follows the range given below.

$$V_{max} \leq E \leq V_{min} + \frac{1}{k} \quad [2.48]$$

where V_{max} and V_{min} are the maximum and minimum potential energies. To ensure that equation [2.48] is valid, $V_{max} \leq V_{min} + 1/k$ and k can be written as;

$$\frac{1}{k} \leq \frac{1}{V_{max} - V_{min}} \quad [2.49]$$

where k is defined as.

$$k \equiv k_0 \left(\frac{1}{V_{max} - V_{min}}\right), 0 < k_0 \leq 1 \quad [2.50]$$

As illustrated in **Figure 2.7**, k_0 determines the magnitude of the applied boost potential. With higher k_0, larger boost potential is added to the PES, which facilitates enhanced sampling of biomolecules across decreased energy barriers.

For accurate reweighting of potential energy values, standard deviation of the $\Delta V(\sigma_{\Delta V})$ must be significantly small (i.e., narrow distribution)[182].

$$\sigma_{\Delta V} = k(E - V_{avg})\sigma_V \leq \sigma_0 \quad [2.51]$$

where V_{avg} and σ_V are the average and standard deviation of the potential energies where σ_0 is an user specified upper limit for the accurate reweighting.
If $E = V_{max}$, From equation [2.48], we can obtain the following relation.

$$k_0 \leq \frac{\sigma_0}{\sigma_V} . \frac{V_{max} - V_{min}}{V_{max} - V_{avg}} \quad [2.52]$$

and if $E = V_{min} + \frac{1}{k}$, then we obtain the following relation.

$$k_0 \geq (1 - \frac{\sigma_0}{\sigma_V}) \frac{V_{max} - V_{min}}{V_{max} - V_{avg}} \quad [2.53]$$

The boost potential can be calculated for a given E and k_0 using the below formula.

$$\Delta V(\boldsymbol{r}) = \frac{1}{2} k_0 \frac{1}{V_{max} - V_{min}} (E - V(\boldsymbol{r}))^2, if\ V(\boldsymbol{r}) < E \quad [2.54]$$

Like aMD, GaMD also provides a dual boost or total potential and total dihedral boost option for simulation. Initially V_{max} and V_{min} are taken from the prior equilibrium MD simulation. Mainly we used dual boost potential for better acceleration.

In addition, GaMD has been combined with replica exchange molecular dynamics simulation (REMD)[183] to improve conformational sampling and free energy calculations further. More recently developed "selective GaMD" algorithms, including Ligand GaMD (LiGaMD)[181] and Peptide GaMD (Pep-GaMD)[184], have enabled unprecedented microsecond simulations to capture repetitive binding and dissociation of small-molecules ligands and highly flexible peptides. Accurate ligand/peptide binding free energies and kinetics rate constants are thus calculated through the selective GaMD simulations.

2.6. Molecular docking and its algorithms

Molecular docking is one of the key computational tools in structural molecular biology and computer-assisted drug design. The goal of ligand-protein or protein-protein docking is to predict the predominant binding mode(s) of a ligand/protein with a protein of known three-dimensional structure. Docking can be used to perform virtual screening on large libraries of compounds, rank the results, and propose structural hypotheses of how the ligands inhibit the target, which is invaluable in lead optimization. The docking protocols include searching algorithms and scoring functions. Initially, the searching algorithm predicts how ligands can fit into the pocket by searching different protein-ligand pairs' orientations. This is followed by evaluating and ranking the conformations obtained from the searching algorithm and estimating each complex's binding affinity[185,186]. Primarily, the methods employ the relative flexibility of receptor and ligand as binding leads to conformational changes described in the "induced-fit" mechanism.

2.6.1. Search algorithm

The ligand's different conformations could be assessed by calculating the root-mean-square deviations (RMSD) and comparing them with the native form. The smaller the RMSD, the closer it will be to the reference pose. For any searching algorithm, its accuracy primarily depends upon the number of degrees of freedom of conformational search. It can be reduced by:

- Considering the implicit solvent in the scoring function.
- Restricting ligand and protein flexibility (considering only ligands' flexibility while keeping the protein as rigid), thereby allowing large sampling of ligands' conformational space.

Ligand searching algorithm includes **systematic, stochastic**, and **simulation** methods.

Systematic method: It samples all degrees of freedom by exhaustive, fragmentation, and ensemble methods. The exhaustive method involves all possible orientations of the rotatable bonds of ligands. As the number of rotatable bonds increases, more conformations will be assessed. Hence, the docking process can be computationally effective by imposing geometric constraints. Further ligand conformations undergo refinement. The Glide[187] and FRED[188] use this method. DOCK[189], LUDI[190], and FlexX[191] use the fragmentation method. It simulates the ligand flexibility by the "place and join" methodology—several ligand fragments are docked into the binding pocket, which is linked with covalent bonds. For example, the ensemble method, for e.g., FLOG libraries of 3D coordinates and finds the best ligand for the receptor.

Stochastic method: The method employs a change in the ligand's population, either accepting or rejecting it based on probability. It can be categorized based on Monte Carlo (MC) or a Genetic algorithm (GA). The MC approach uses Boltzmann probability criteria for acceptance. In contrast, the GA applies the evolution theory for docking, as seen in AutoDock[192,193] and GOLD[194]. Initial ligand poses define specific phenotypes and genotypes for each ligand pose. The genotypes refer to the ligand's translation and orientation concerning the target protein, and phenotypes refer to the ligand's atomic coordinates.

2.6.2. Scoring functions

Reliable docking calculations require a robust scoring system, wherein numerical values are assigned to each candidate of the protein-ligand system generated by the sampling algorithm. The scoring functions allow the ranking of ligands by calculating the relative binding affinity between a ligand and the target. Several scoring functions for docking include **force field** (FF), empirical, and knowledge-based scoring functions.

The **FF-based scoring** function involves the total internal energies of the ligand and between ligand and protein. The FF-based scoring systems have some disadvantages, including their inability to estimate the long-range interactions and entropic contributions. They are also computationally expensive. However, other recent scoring functions, such as AutoDock and Gold score, have been developed considering these terms. In addition, other molecular mechanics FFs, namely AMBER and OPLS, can also be used for scoring. The **empirical-based scoring** function takes experimental data for deriving the coefficients in predicting the binding affinity between protein and ligand, for example, ChemScore[195], Score[196], and X-score[197]. The **Knowledge-based scoring** functions rely on predicting the binding affinity based on the poses (or interatomic interactions between the components by computing the statistical occurrence) that resemble the experimental geometries is provided the highest ranking. The method has a limitation that what kind of information is experimentally available in X-ray or NMR; for example, protonation states information is missing in these methods. The DrugScore[198] and SMoG[199] Score are two examples of the knowledge-based scoring function. These three scoring methodologies described above serve as a starting point for several strategies to enable more accurate scoring and ranking of docked poses.

2.7. Trajectory analysis techniques

Molecular dynamics simulations return highly complex data. The Cartesian positions of each atom of the system (thousands or even millions) are recorded at every time step of the trajectory; this may again be thousands to millions of steps in length. Therefore, further analysis is needed to extract useful information from the data.

In this section, we illustrate some analytical tools able to investigate conformational changes by analysis of typical short/long scale biomolecular simulations.

2.7.1. Stability and flexibility analyses

The structural stability of biomolecular simulation is mainly defined by its **root-mean-squared-deviations** (RMSD). RMSD is a statistical measure of finding similarities between two sets of values. RMSD measures the target coordinate's deviation from the reference coordinates. It calculates the average distance between the reference structure and selected atoms. RMSD is defined as:

$$RMSD = \sqrt{\frac{\sum_{i=0}^{N}[m_i.(X_i - Y_i)^2]}{M}} \quad [2.55]$$

Where N is the number of atoms, m_i is the mass of atom i, X_i is the target coordinate atom i, Y_i is the coordinate for the reference atom i, and M is the total mass. For non-mass weighted RMSD, the value of $m_i = 1$ and $M = N$.

Another important quantity is **root-mean-squared-fluctuations** (RMSF) to explore residual flexibility. Like RMSD, it measures the deviation between the position of particle i and some reference positions:

$$RMSF_i = \sqrt{\frac{\sum_{t_j=1}^{T}\left[x_{i(t_j)} - \langle x_i \rangle\right]^2}{T}} \quad [2.56]$$

Where T is the duration of the simulation (time steps) and $x_{i(t_j)}$ are the coordinates of the atom x_i at time t_j and $\langle x_i \rangle$ is the time-averaged position for the same particle. Hence, RMSF indicates the positional differences for the entire structure over time. RMSF is frequently compared with the B-factor values to check the quality of the simulation. B-factors measure the thermal motion of proteins determined at the crystallization phase, and it is provided in the PDB file alongside the *x*, *y*, and *z* coordinates. Atoms with high B-factors could identify the most flexible part of a protein, e.g., loops, and should be reasonably aligned with the calculated RMSF values. The B-factor is defined as:

$$B = \frac{8}{3N}\pi^2(RMSF)^2 \quad [2.57]$$

We also measured the compactness of the simulated system by the **radius of gyration (R_g),** which can be calculated as:

$$R_g = \left(\frac{\sum_i |r_i|^2 m_i}{\sum_i m_i}\right)^2 \quad (2.58)$$

Where m_i is the mass of atom i and r_i is the atom i position with respect to the center of mass of the molecule.

2.7.2. Dynamic cross-correlation matrix (DCCM)

The extent of correlations in a system can be examined by the magnitude of cross-correlation coefficients for each pair. It gives a graphical representation of a matrix form for all atom-pairs, known as a **dynamical cross-correlation matrix** (DCCM). This analysis has been extensively used to quantify the correlation motions between atoms. The DCC between i^{th} and j^{th} atoms are defined by:

$$DCC(i,j) = \frac{<\Delta r_i(t).\Delta r_j(t)>_t}{\left(\sqrt{<|\Delta r_i(t)|^2>_t}\right)\sqrt{<|\Delta r_j(t)|^2>_t}} \quad [2.59]$$

Where $r_i(t)$ denotes the the i^{th} atom's coordinates as a function of time t, $<.>_t$ denotes the time ensemble average and $\Delta r_i(t) = r_i(t) - < r_i(t) >_t$. The DCC has been used to calculate the dynamics of cross-correlation within the molecule. The DCC produces an $N \times N$ heatmap, where N is the number of (mainly alpha carbon) atoms in the system, and each element of the matrix corresponds to the dynamic cross-correlation between i and j atom. The correlation values vary between -1 and $+1$, where $1 =$ **correlation**; $-1 =$ **anti-correlation**; $0 =$ no correlation. The completely correlated fluctuations are in the same period and the same phase, while anti-correlated fluctuations are in the same period but opposite phase. A line along the diagonal reflects the **strong correlation**. The high diagonal value occurs when $i = j$, i.e., $DCC(i,j)$ = 1.00. The positive correlation along the diagonal reflects the correlations of contiguous residues within a secondary structure.

Off-diagonal positive and negative correlations indicate correlations of non-contiguous residues.

2.7.3. Principal component analysis (PCA)

The principal component analysis (PCA)[200,201] simplifies a dataset by reducing its dimensions to represent **principal components** (eigenvector with its corresponding high eigenvalue), contributing the most variance in the dataset. PCA takes trajectory from the molecular dynamics simulations to identifying configurational space containing only a few degrees of freedom. In Cartesian PCA, the overall rotational and translational motions are eliminated by the least-squares fit algorithm. A $3N \times 3N$ **covariance matrix** is formed for cartesian coordinates. Diagonalization of the covariance generates a set of eigenvectors and their associated eigenvalues. The eigenvectors describe the direction of motions, while the eigenvalue gives the amplitude of motions.

Consider a covariance matrix C and the elements C_{ij} of the matrix is defined as:

$$C_{ij} = \langle (x_i - \langle x_i \rangle)(x_j - \langle x_j \rangle) \rangle \qquad [2.60]$$

Where x_i and x_j are coordinates of the i^{th} or j^{th} atom, and $< x_i >$ and $< x_j >$ are the mean average coordinated of the i^{th} or j^{th} atom over the ensemble. For three dimensions, the covariance matrix for (x, y), (x, z), and (y, z) coordinates are carried out, and a covariance matrix C generates the matrix of $3N \times 3N$, where N denotes the number of atoms.

The covariance matrix is then diagonalized to get the eigenvalues[201]:

$$A^T C\, A = \lambda \qquad [2.61]$$

Where A is the eigenvectors and λ is the eigenvalues. The arrangement of eigenvectors is made in decreasing order of their corresponding eigenvalues. Finally, the eigenvector with the highest value of eigenvalue represents the first **principal component (PC)** and so on. The principal components provide new coordinates; hence trajectories are fitted on eigenvectors to give PCs:

$$q(t) = A^T M^{\frac{1}{2}}(x(t) - \langle x \rangle) \quad [2.62]$$

The trajectories can be filtered along PCs to isolate motions from other principal components. Generally, only the first few PCs of the total number of eigenvectors describe the relevant motions from simulation trajectories and are used for the analysis.

2.7.4. dihedral principal component analysis (dPCA)

The PCA is an extensively used technique in reducing a complex system's dimensions. However, it has been shown by Mu *et al*.[202] that instead of Cartesian coordinates, internal coordinates (backbone dihedral angles) can be employed in PCA of molecular dynamic simulations, especially for complex biomolecules containing large degrees of freedom. This may lead to avoiding mixing overall motions with the internal motions. The main advantage of using the backbone dihedral angle is that it will naturally undergo more changes than internal coordinates like bond length and bond angle for a molecule. Therefore, another method of PCA known as dihedral angle principal component analysis (dPCA)[202], which is based on internal coordinates like the dihedral angle (ϕ_n, ψ_n) of the peptide backbone may help to separate the internal motions from overall dynamics.

However, angular data always poses a PCA problem or analyses other statistical processes due to their circular nature. For example, the arithmetic means of 10° and 350° is $(10^{\circ} + 350^{\circ})/2 = 180^{\circ}$, instead of 0°. To circumvent such difficulties due to the circularity of variables, the space of dihedral angle (φ_n) is transformed to the linear metric coordinate space using the trigonometric functions $\sin \varphi_n \; and \; cos\varphi_n$.

2.7.5. Potential of mean force and free energy landscapes

The potential of mean force (**PMF**)[203] provides a free energy profile along a preferred coordinate, be it a geometric or energetic coordinate, such as the distance between two atoms or the torsional angle of a molecule. This free energy profile describes the average force of all possible configurations of a given system (the ensemble average of the force)

on particles of interest. The simplest representation of the PMF is to use the reaction coordinate. The PMF is related to the radial distribution function using the following expression for the Helmholtz free energy[108],

$$A(r) = -k_B \mathrm{T} ln[g(r)] + const. \quad [2.63]$$

The constant is chosen so that the most probable distribution corresponds to the free energy of zero. The radial distribution function means a relatively small change in the free energy.

Free energy landscape **(FEL)**[204,205] theory provides a framework for describing the thermodynamics and kinetics of complex systems as well as the probability of observing a particular event. The potential energy function of a multibody system such as a protein is a multidimensional and often very complex surface. At nonzero temperature, entropic contributions become relevant, and therefore the free-energy landscape governs the thermodynamics and kinetics. A common way to investigate and display the free energy involved in biomolecular isomerization and protein folding is to study it as a function of one or more order parameters, i.e., suitably chosen macroscopic quantities that distinguish the different states of the system. States are associated with local free-energy minima of the projected landscape. The depth of the minima is considered proportional to the stability of the states associated with them, and the barriers between different minima indicate activation energies between states[206]. Also, the transition among different states is more probable if the energy barriers that divide the states are low. Due to the complexity of the process and a large number of degrees of freedom involved, order parameters are often arbitrarily chosen.

The free energy landscape can be drawn based on the following equation:

$$G_i \approx -k_B T ln(\frac{N_i}{N_m}) \quad [2.64]$$

Where k_B is the Boltzmann constant, and T is the absolute temperature, N_i is the population of the i^{th} bin and N_m is the population of the most populated bin. A two-dimensional FEL was drawn in this study using various reaction coordinates, such as PC1 vs. PC2, dPC1 vs. dPC2, RMSD vs. distance, *etc*.

2.7.6. Protein structure network (PSN) analysis

To visualize protein structures other than secondary structure and fold arrangements is to represent the interactions between residues in a network. The network representation of complex three-dimensional structures provides insight into the structure-function relationship. In the present work, we have used the WebPSN[207] and the Network Analysis of Protein Structures (NAPS)[208] portals to construct protein networks. The webserver provides an interactive visualization of inter-residual interactions in modeled protein and MD simulation trajectories. The network can be built based on C_α, C_β, atom pairs, centroid networks, or interaction-strength networks. Here, the node is created based on the C_α atom of an amino acid residue, and an unweighted edge is constructed if the paired residue C_α-C_α distance lies within the threshold distance (R_C) of ~ 7 Å. NAPS provides user selection of nodes based on centrality measurement, physiochemical properties, or clusters of connected residues. NAPS provides relevance of nodes/edges based on various features, such as centrality measurement, shortest path, betweenness, hubs, links, community, *etc.* Overall network view supports various analyses for identifying functional residue, predicting coevolving residues, understanding the mechanism of protein-protein interaction or domain-domain for understanding the communications between them.

2.8. Binding free energy estimations

Free energy calculations have proven useful for several topics in computational biology, such as drug design and protein structure determination. Free energies can be expressed in two forms: the **Helmholtz** (F) and **Gibbs free energy** (G).

$$F = U - TS \qquad [2.65]$$

$$G = H - TS \qquad [2.66]$$

Where U is the internal energy, T is the temperature, H is the enthalpy, and S is the entropy of the system. Where H is the enthalpy of the system, as given by $\boldsymbol{H = U + PV}$, where P is the pressure and V is the volume of the system. It signifies free energies depend on the system's internal (potential) energy and entropy. Therefore, the affinity and stability of a

system depend on the interatomic interactions and changes in entropy. Primarily, solvent entropy also plays a significant role during protein hydrophobic core or core membrane formation.

The Gibbs free energy provides the binding affinities between the protein-ligand and protein-protein interactions. In terms of statistical mechanics, the probability of state x is directly related to its free energy:

$$p(x) \propto e^{-G(x)/k_B T} \quad [2.67]$$

Where k_B is the Boltzmann constant and $e^{-G(x)/k_B T}$ is called the Boltzmann factor. Hence, the free energy difference for two states A and B can be estimated by:

$$\frac{p(A)}{p(B)} = e^{-(G(A)-G(B))/k_B T} = e^{-\Delta G/k_B T} \quad [2.68]$$

This way, free energies can be estimated from simulations that are sufficiently converged. The difference in free energy ΔG can be estimated directly by evaluating the populations in A and B.

The free energy calculations have become an integral part of computational biochemical research due to their wide applicability in predicting binding affinities of small candidate compounds as well as evaluating their relative stabilities with biomolecular structures. The free energy estimation can be broadly divided into two groups: **rigorous methods** and **approximate methods**. The rigorous methods include free energy perturbation (FEP) and thermodynamic integration (TI). The primarily used approximate method is the molecular mechanics-Generalized Born/Poisson Boltzmann Surface Area **(MM-GB/PBSA)** scheme.

2.8.1 Free energy perturbation (FEB)

Free energy perturbation (**FEP**)[209–211] is a statistical mechanics-based method for estimating free energy via intermediate states between the reference (starting) and perturbation (final) states[212,213]. The mathematical formulation of FEP is stated below[214]:

$$\Delta F = F_1 - F_0 = -kTln\langle exp\left[\frac{-\langle E_1 - E_0\rangle}{kT}\right]\rangle \quad [2.69]$$

Here the free energy difference between an initial (reference) and a final (target) state of a system to an average of a function of their energy difference is evaluated by sampling for the initial state. Equation further modified as power series for obtaining an expression for free energy change. Here, E_1 is the energy of a target state, E_0 is the energy of the reference state, and V is a small perturbing potential. The following equations are obtained by expanding the above equation to second order.

$$E_1 = E_0 + V \quad [2.70]$$

$$\Delta F = F_1 - F_0 = \langle V\rangle - \frac{1}{2kT}(\langle V^2\rangle - \langle V\rangle^2) \quad [2.71]$$

This free energy method is more relevant for systems where the fluctuations are mainly governed by a Gaussian probability distribution.

2.8.2 Thermodynamics integration (TI)

Another well-established and key method is the thermodynamic integration (TI)[215–218]. The relative binding affinity of two ligands, L1 and L2, is calculated by considering an alchemical transformation between them connected through intermediate states defined by introducing a coupling parameter λ. This coupling parameter is estimated so that at λ = 0, the system corresponds to ligand L1 (initial state), and at λ = 1, the system corresponds to ligand L2 (final state). The total energy (V) of the system can be defined as

$$V(\lambda, x) = (1 - \lambda)V_1(\lambda, x) + \lambda V_2(\lambda, x) \quad [2.72]$$

Where V_1 and V_2 are the potential energies of ligands L1 and L2. The derivative of the energy with respect to λ is used to compute the free energy difference as follows

$$\Delta G_{alch} = \int_0^1 \langle\frac{\partial V(\lambda, x)}{\partial \lambda}\rangle_\lambda \, d\lambda \quad [2.73]$$

Where ⟨⋯⟩λ denotes an ensemble average for the state λ. In most cases, such an ensemble at each λ window is generated using a single MD trajectory.

The thermodynamic cycle approach is employed to calculate the relative binding affinities ΔΔG between these two ligands associating with a protein using the following equation

$$\Delta\Delta G = \Delta G_1 - \Delta G_2 = \Delta G_{alch}^{aq} - \Delta G_{alch}^{bound} \quad [2.74]$$

where ΔG_1 and ΔG_2 are the binding free energies of ligands L1 and L2, respectively. ΔG_{alch}^{aq} and ΔG_{alch}^{bound} are the free energy differences associated with the alchemical transformation of ligand L1 into L2 in free and bound states, respectively.

2.8.3 Molecular mechanisms (generalized Born) Poisson Boltzmann surface area (MM/(GB)PBSA)

The molecular mechanics Poisson Boltzmann (generalized Born) surface area (MM/PB(GB)SA) surface area is one of the most popular techniques for estimating the binding free energy between protein-ligand complexes[219,220]. Here both classical force fields and continuum solvation models work together to estimate the binding free energies of large molecular systems. Different alchemical methods such as exponential averaging and thermodynamics integration (TI) may give better accurate results, whereas MMPB/(GB)SA has a good trade-off between computational speed and accuracy[221,222].

Here is a brief description of the MMPB(GB)SA method regarding protein (P) and ligand (L) complex (PL). The overall binding free energy is shown below[223,224];

$$\Delta G_{bind} = G(PL) - G(P) - G(L) \quad [2.75]$$

The free energy of protein, ligand, and protein-ligand complexes are given as follows (represented by X):

$$G(X) = E_{MM}(X) + G_{solv}(X) - TS(X) \quad [2.76]$$

$$E_{MM}(X) = E_{bonded}(X) + E_{nonbonded}(X) \quad [2.77]$$

where E_{MM} is total molecular mechanics energy, G_{solv} is the solvation free energy. T denotes the temperature on the absolute scale, and S is the conformational entropy. E_{MM} is further composed of bonded ($E_{bonded}(X)$) and nonbonded term ($E_{nonbonded}(X)$).

$$E_{bonded}(X) = E_{bond}(X) + E_{angle}(X) + E_{dihedral}(X) \quad [2.78]$$

$$E_{nonbonded}(X) = E_{electrostatics}(X) + E_{vdW}(X) \quad [2.79]$$

Furthermore, the bonded term includes the contribution from bonds (E_{bond}), angles (E_{angle}) and dihedrals ($E_{dihedral}$) and the nonbonded term includes electrostatics ($E_{electrostatics}$) and van der Waals (E_{vdW}) energies. Combing all gives:

$$\Delta G_{bind} = \Delta G_{MM} - \Delta G_{solv} - T\Delta S \quad [2.80]$$

where;

$$\Delta E_{MM} = \Delta E_{bonded} + \Delta E_{electrostatics} + \Delta E_{vdW} \quad [2.81]$$

$$\Delta G_{solv} = \Delta G_{cav} + \Delta G_{vdW} + \Delta G_{elec} = \Delta G_{SASA} + \Delta G_{PB/GB} \quad [2.82]$$

$$\Delta G_{SASA} = \gamma . SASA + \beta \quad [2.83]$$

The solvation free energy consists of three components, i.e. G_{cav} which is the energy term for generating cavity in the solvent for accommodating the solute, van der Waals interaction term (G_{vdW}) and the electrostatic interaction energy (G_{elec}). The first two components combined to give G_{SASA}. This parameter can be estimated by combining the solvent accessible surface area (SASA) and two empirical parameters (γ, β), taken from a linear regression of the solvation free energy of a set of small apolar molecules in water. Where γ, the surface tension proportionality and was set to $0.00542\ kcal.mol^{-1}$Å^{-2} and b is the offset value set to $0.92\ kcal.mol^{-1}$. The SASA value is measured from the linear combination of pairwise overlap (LCPO)[225] algorithm having a probe radius of 1.4 Å.

Estimation of the electrostatic contribution can be obtained by employing the PB or GB model.

According to the PB model, the electrostatic potential can be estimated using the PB equation[226],

$$\nabla[\varepsilon(\mathbf{r})\nabla\varphi(\mathbf{r})] = -4\pi\rho(\mathbf{r}) - 4\pi\lambda(\mathbf{r})\sum_i z_i c_i exp\left(-\frac{z_i\varphi(\boldsymbol{r})}{k_B T}\right) \quad [2.84]$$

where $\varepsilon(\mathbf{r})$ is the dielectric constant and $\varphi(\mathbf{r})$ is the total electrostatic potential, $\lambda(\mathbf{r})$ is the Stern layer masking function. However, the PB equation cannot be solved analytically, and the numerical solutions of the same are still computationally expensive.

According to the GB model, the same is given by[144];

$$\Delta G_{GB} = -\frac{1}{2}(1-\frac{1}{\varepsilon})\sum_{i,j}\frac{q_i q_j}{\sqrt{r_{ij}^2 + R_i R_j exp(-\gamma\frac{r_{ij}^2}{R_i R_j})}} \quad [2.85]$$

Here, ε is the dielectric constant of the bulk solvent, q_i is the atomic charge of particle i and r_{ij} is the effective Born radius of atom. Several GB models are present, which differ mainly in estimating effective Born radii[143,227,228].

2.8.4. Conformational entropy calculation

The configurational entropy was calculated using the **normal mode analysis (NMA)**[229–231] approach. NMA was applied to calculate the vibrational entropy (S_{vib}) using the following equation[232]

$$S_{vib} = R\left[\frac{x}{e^x-1} - \ln(1-e^{-x})\right]; x = \frac{h\vartheta}{kT} \quad [2.86]$$

Where R denotes the universal gas constant, k is the Boltzmann constant, and T is the absolute temperature. The plank constant and the vibrational frequency are denoted by h and ϑ, respectively.

Chapter 3

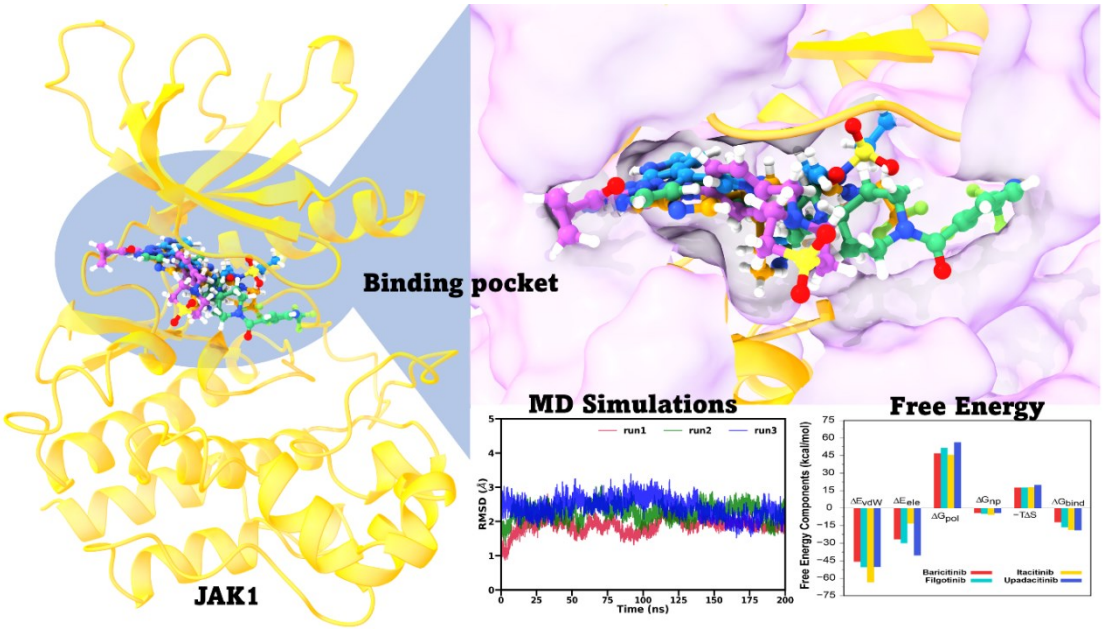
Chapter 3

Unraveling the Molecular Mechanism of Recognition of Selected Next Generation Anti-rheumatoid Arthritis Inhibitors by Janus Kinase 1

AIM: To understand the small molecules recognition by Janus kinase protein and their residual interactions energy spectra.

This chapter is reproduced from our original article:

Sk, M.F., Jonniya, N. A, Roy, R., & Kar, P. (2022). "Unraveling the Molecular Mechanism of Recognition of Selected Next Generation Anti-rheumatoid Arthritis Inhibitors by Janus Kinase 1" *ACS Omega*, 7(7), 6195-6209.



3.1. Introduction

Cytokines are essential for host defense and immunoregulation[233,234]. They are also significant players in the immunopathogenesis of autoimmune diseases, such as Rheumatoid arthritis (RA)[80]. Like other inflammatory diseases, RA is a chronic immune-related condition (primarily of joints), highly disabling and painful, and reduced life quality. In Western society, the estimated prevalence of RA is 1-2% [81] and is believed to be 1% worldwide[82]. Over the last 30 years, many therapies, including anti-IL-6 [235], anti-CD20, anti-TNF[236], and CTLA4-Ig[237] agents, have been developed to treat RA. Although these significantly improve care, there are some limitations to maximizing these therapies [238–241]. Despite their success, anti-IL-6, anti-CD20, anti-TNF, and CTLA4-Ig agents that target key cytokines are not entirely effective in all patients. Also, they can cause several potentially life-threatening side effects, including liver and lung damage and decreased ability to fight off infections. Recently, inhibition of Janus kinases (JAKs)- related cytokine signaling pathways using small molecules (inhibitors) has received significant attention nowadays[97]. JAKs are essential growth factors and intercellular parts of cytokine signaling pathways. The JAK kinase phosphorylates and activates the signal transducers transcription of STATs, which gather in the nucleus and rein in the transcription of effector genes[242,243].

The JAK family belongs to non-receptor cytoplasmic protein tyrosine kinases (PTKs), including four basic mammalian types, JAK1, JAK2, JAK3, and TYK2 (Tyrosine kinase 2)[243]. JAK1 has received significant attention after being identified as a promising target for anti-RA drugs. Besides, other inflammations are closely related to the JAK/STAT signaling pathway[244]. The human JAK1 protein JH1 (JAK Homology 1) domain comprises 290 (842-1132) residues. Similar to all protein kinases, JAK1 is divided into a large C-terminal lobe (residues 957-1132) and a small N-terminal lobe (residues 842-956), which were first observed for PKA (Protein Kinase A)[245]. The small N-terminal lobe is comprised of five-stranded anti-parallel β-sheet (β1-β5)[33], 3/10A-helix, and conserved α-helix (residues 919-931). The conserved ATP-phosphate binding glycine-rich loop (P-loop) (^{879}RDLGEGHFGKV^{889}) is located between its β1 and β2 sheets (**Figure 3. 1(A,B))**. The C-terminal lobe starts from a conserved hinge region (residue 956-968). Moreover, the large C-lobe of JAK1 contains two significant catalytic areas (^{1001}HRDLAARN^{1008}) and an

activation segment (residue 1021-1051). JAK1 is activated through the phosphorylation of Tyr^{1034} and Tyr^{1035}, located in the activation segment. The structural adjustment induced by the phosphorylation of the activation segment occurs in the ASP-PHE-GLY (DFG) motif. Kornev et al.[246,247] first uncovered eight hydrophobic residues as a catalytic or C-spine and another four hydrophobic residues that constitute a regulatory or R-spine occurring in the small N-terminal lobe and the large C-terminal lobe (see in **Appendix A, Figure A1**). These hydrophobic clusters also define the active and inactive states of JAKs.

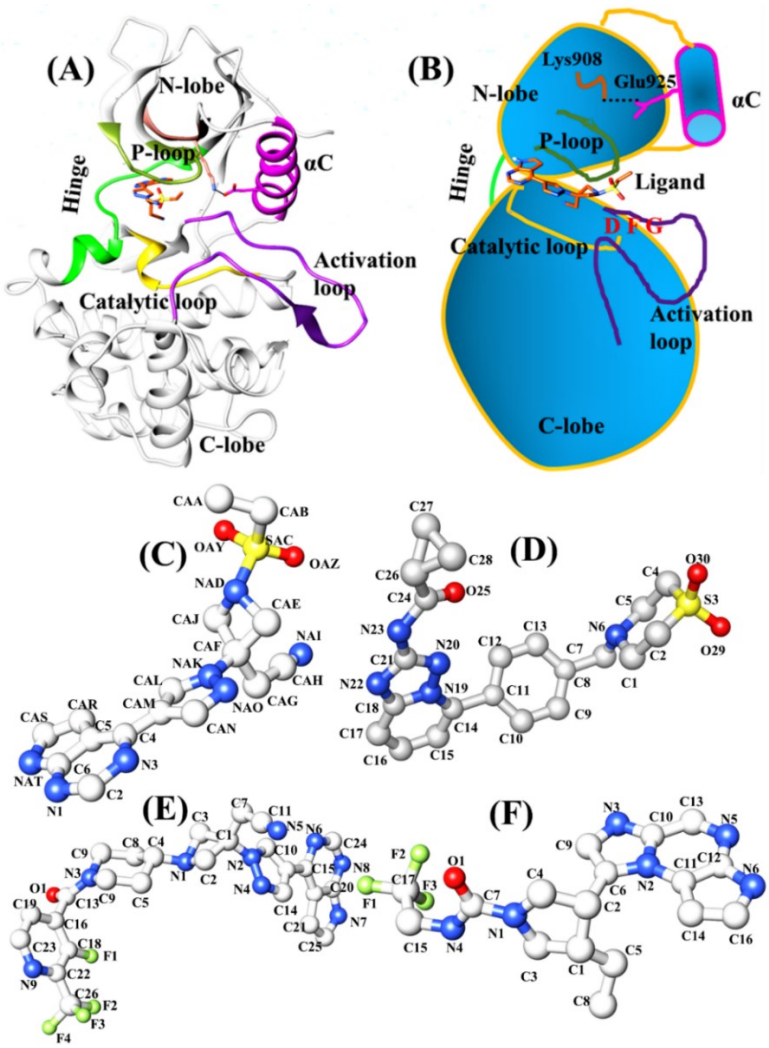


Figure 3. 1: (**A)** Typical active Janus kinase 1 (JAK1) structure with different vital segment color-coded, **(B)** schematical image corresponding JAK1 structure, and (C-F) 3-dimensional structures of inhibitors with each heavy atoms name: Baricitinib, Filgotinib, Itacitinib, and Upadacitinib, respectively.

The discovery of JAK kinase inhibitors opens an era of treatment for various autoimmune diseases. Tofacitinib (CP-690550) is a type-I pan-JAK inhibitor that the FDA first approved to treat RA[33,234,238]. In the face of extensive kinome selectivity of tofacitinib, it has limited selectivity within the JAK family of kinases. It leads to the discovery of JAK1-specific type-I inhibitors, such as Baricitinib (INCB028050)[234], Upadacitinib (ABT-494)[234], Filgotinib (GLPG0634)[99,248], and Itacitinib (INCB039110)[79,249]. Baricitinib is used to treat moderately to severely active RA in adults[97]. Upadacitinib is recommended to treat RA and other autoimmune diseases[98]. It is highly specific to JAK1 and shows 58-fold and 74-fold (*in vitro*) selectivity for JAK1 over JAK3 and JAK2, respectively[79]. Currently, Filgotinib and Itacitinib are in the phase-III clinical trial. Itacitinib shows > 20-fold selectivity for JAK1 over JAK2 and > 100-fold selectivity over JAK3 and TYK2[79].

Unfortunately, the mechanism of recognizing these drug molecules by JAK1 is not known entirely due to the absence of crystal structures of the complex. Therefore, it is necessary to explore the binding mechanisms of these anti-rheumatoid arthritis inhibitors by JAK1. Herein, we sought to investigate the mechanism of binding of type-I inhibitors, such as Baricitinib, Filgotinib, Itacitinib, and Upadacitinib to JAK1 by employing comprehensive computational protocols involving molecular docking, molecular dynamics simulations, and free energy calculations.

3.2. Materials and methods

3.2.1. Molecular docking and systems preparation

When this work was started, there were no complex X-ray crystallographic structures of JAK1 with Baricitinib, Filgotinib, Itacitinib, and Upadacitinib in Protein Data Bank (PDB) (https://www.rcsb.org). Therefore, we used the rapid Lamarckian Genetic algorithm (GA 4.2) search method in the AutoDock 4.2 software suite to prepare the initial complex structures. For the target receptor, JAK1, three-dimensional coordinates were taken from the PDB database (PDB entry: 6AAH; resolution, 1.8 Å)[250]. We removed the crystal waters, and the missing residues (913-916) were built using the UCSF Chimera[251] plugin MODELLER[252]. Polar hydrogen atoms and Kollmann charges were added by AutoDock Tools (ADT), while non-polar hydrogen atoms were merged. The ligand atomic coordinates were collected from the PubChem[253] database. The Gasteiger charges and

rotatable bonds were assigned to the ligands, and the grid spacing was set at 0.375 Å. AutoDock generated results in the pdbqt format, and optimal geometric conformation with the best binding energy (docking energy ranges between -9.0 to -12.1 kcal/mol) was selected from all possible conformations. Chimera[251] was used to prepare the protein and ligand complex for further molecular dynamics simulations study. The Cartesian coordinates of the docked structures are provided in Supporting Information. Each ionizable amino acid protonation state was characterized by PROPKA 3.0[254].

3.2.2. Molecular dynamics (MD) simulations

All the MD simulations were carried out using the AMBER18 MD suite[255]. The force fields used for the ligands and proteins were the updated Generalized Amber Force Field (GAFF2)[256] and ff14SB[257], respectively. The AM1-BCC[258] atomic charges were calculated for ligands with the antechamber module[259]. For the phosphorylated tyrosine residues, the Phosaa10 force field[260] was used. Each system was solvated using an explicit TIP3P[261] water model, and an octahedron periodic box with a distance cut-off of 10 Å was used. An appropriate number of counterions was added to neutralize the solvated system and maintain a physiological pH of 0.1 M.

Initial minimization of each solvated complex was achieved by performing 500 cycles of steepest descent, followed by another 500 cycles of the conjugate gradient with a harmonic restraint of 2.0 kcal mol^{-1} $Å^{-2}$. The second minimization stage was conducted without restraining the solute for 100 cycles of the steepest descent, followed by another 900 cycles of the conjugate gradient. Subsequently, all complexes were gradually heated from 0 K to 300 K for 50 ps at the NVT ensemble with a restraint force constant of 2.0 kcal mol^{-1} $Å^{-2}$ on the solute.

We performed the density equilibration of each complex with a restraint force of 2.0 kcal mol^{-1} $Å^{-2}$ under the NVT ensemble (300 K, 50 ps). Each system was then equilibrated at the NPT ensemble for 1.0 ns without any restraint. Finally, the production simulation was run for 200 ns with a time-step of 2.00 fs under the NPT ensemble. We performed 2 x 200 ns simulations for each system, starting with a different initial velocity. The coordinates were saved for every 10 ps, leading to 20,000 snapshots. During simulations, all bond lengths were constrained, including the hydrogen atom using the SHAKE algorithm[262].

Under the periodic boundary conditions (PBCs), the particle mesh Ewald (PME)[263] method was used to treat the long-range electrostatic interactions with a cut-off of 10.0 Å. The temperature was maintained at 300 K using the Langevin thermostat[264] with a collision frequency of 2.0 ps^{-1}, and the pressure was controlled by the Berendsen Barostat[265]. The post-processing of trajectories was done using the *Cpptraj* module[266] of AmberTools19.

3.2.3. Dynamic cross-correlation map (DCCM)

The pairwise cross-correlation coefficients, or dynamic cross-correlation maps[267], depict the correlation between pairs of atoms. We considered only the Cα atomic coordinates of each residue. The positive and negative values in the map represented the correlated and anti-correlated behavior (see **chapter 2**, **section 2.7.2.**). The last 140 ns trajectories of both runs were used for this analysis using *Cpptraj*.

3.2.4. Principal component analysis and Free energy landscape

In order to gain about the complex motions of all systems of JAK1, the principal component analysis (PCA)[268] was performed (see **chapter 2, section 2.7.3.** for details). It is a dimensional reduction method and could extract the dominant modes from the whole trajectory. Similar to the DCCM analysis, only Cα atomic coordinates were used for this analysis. The MD trajectories of both runs were employed for the PCA analysis. The cosine content of the first few PCs was obtained via GROMACS (g_covar, g_anaeig, and g_analyze modules)[269]. It is used to ascertain the convergence of each trajectory.

We used the first two principal components (PC1 and PC2) as reaction coordinates to draw the free energy landscape (FEL). In 3-dimensional FEL, the low energy "valleys" denote metastable conformational states, while "hills" correspond to the energetic barriers connecting those valleys.

3.2.5. Protein structure networks (PSNs)

We used a mixed protein structure network (PSN)[270] and elastic network model-normal mode analysis (ENM-NMA) based method to investigate structural communications in the

form of network in proteins, which are otherwise not easily detectable. The PSN analysis enumerates network features, like nodes, edges, hubs, links, community, etc., and the shortest path communication from stable states of molecular dynamics (MD) trajectories. In PSN, nodes correspond to the residue, and a hub represents a highly linked residue that generally refers to the degree of a node. Community or modularity is the region in a network where nodes are more connected to each other. Here, the ligand unbound, apo, and ligand-bound four JAK1 complexes network differences were explored by using WebPSN v2.0[207]. All the analyses of the hubs and network communities were visualized with VMD[271].

3.2.6. MM-PBSA calculations and energy decomposition analysis

To evaluate the binding free energy of all the four inhibitors against JAK1, the widely used molecular mechanics Poisson-Boltzmann surface area (MM-PBSA) method[272–276] was employed. In this study, a total of 7000 structural frames in an interval of 2 ps from the last 140 ns trajectory were used for the MM-PBSA calculation. Details of the MM-PBSA methodology are discussed in our previous works[277–281], and here we followed the same protocol. A brief description of the method was provided in chapter 2, **section 2.8.3.**. The configurational entropy ($-T\Delta S$) was calculated using the normal mode analysis (NMA)[230] method (see chapter 2, **section 2.8.4.**). We considered only 130 configurations from the last 140 ns trajectory for the entropy calculation due to the high computational cost. Further, we evaluated the significance of each residue in the drug binding by performing the per-residue decomposition of the total binding free energy using the molecular mechanics generalize- Born surface area (MM-GBSA) scheme.

3.3. Results and discussion

We combined molecular dynamics simulations and energetic analysis to elucidate the mechanism underlying the binding of Baricitinib, Filgotinib, Itacitinib, and Upadacitinib to the JAK1 kinase. The production simulations of these systems were stable in terms of the total and potential energies (data not shown) and the root-mean-square deviation (RMSD) from the corresponding initial structure. Here, we divided the last 100 ns trajectory into four equal segments of 25 ns trajectory and calculated the average RMSD

for each segment (see **Appendix A, Table A1**). We found that the average RMSD for successive segments lies within the standard deviation. This suggests the convergence of each simulation during the last 100 ns. The same trend was observed for replica simulations also. We also evaluated the stability of the hallmark salt bridge between K908@NZ and E925@CD (see **Appendix A, Figure A2**).

3.3.1. Structural stability and flexibility analysis of JAK1

The temporal root means square deviations (RMSD) of backbone atoms of the JAK1 kinase with respect to the corresponding initial equilibrated structure for all systems were calculated and shown in **Figure 3.2A**. A similar trend was also observed for the replica simulation (see **Appendix A, Figure A3**). The average RMSD values are reported in **Table 3.1**. The last 100 ns of both runs were considered for the calculation of the mean. It is evident from **Figure 3.2A** that all complex systems, including apo, reached equilibrium within the initial 100 ns of the simulation. The average RMSD value of the apo system was estimated to be 2.10 Å, and a similar value was obtained for all the complex systems. The superimposed initial and final conformations of each complex are also shown in **Appendix A, Figure A4**. Next, we investigated the structural stability of the binding pocket by evaluating the time-series RMSDs of the binding pocket (5 Å radius around the inhibitor) residues and displayed in **Figure 3.2B**. The corresponding average RMSD values are reported in **Table 3.1**. In the cases of JAK1/Baricitinib and JAK1/Filgotinib, the pocket was most stable (~ 0.64 Å), while a slightly larger deviation was obtained for JAK1/Itacitinib (1.25 Å).

We also calculated the temporal RMSD of different important regions of JAK1 complexed with these four inhibitors, as shown in **Appendix A, Figure A5**. The corresponding average values are recorded in **Appendix A, Table A2**. It is evident from **Appendix A, Figure A5** and **Appendix A, Table A2** that the P-loop region drifted similarly for all cases except JAK1/Itacitinib. In the case of JAK1/Itacitinib, the lowest deviation was obtained (1.37 Å) while varying between 1.97-2.17 Å for all other cases, including apo. The catalytic (~ 0.40 Å) and hinge (0.29-0.57 Å) regions were stable throughout the simulation for all systems. The same trend was observed for both C-spine (0.45-0.79 Å) and R-spine (0.50-0.77 Å) regions. In the cases of JAK1/Itacitinib and JAK1/Upadacitinib, a marginally lower

deviation (1.5 Å) was obtained for the activation region compared to the apo system (2.14 Å). Overall, no significant structural rearrangements compared to the apo system were obtained due to the ligand binding.

Table 3.1: The average RMSD during the last 100 ns production simulations (**Run1 & Run2**) of the JAK1-inhibitor complexes and standard deviations (SD) in the parentheses. All values are in the Å unit.

RMSD/Systems	Protein residues	Pocket residues	Ligand heavy atoms
JAK1/apo	2.10 (0.15)	0.97 (0.11)	-
JAK1/Baricitinib	2.12 (0.27)	0.70 (0.12)	1.46 (0.18)
JAK1/Filgotinib	2.26 (0.17)	0.64 (0.14)	0.44 (0.30)
JAK1/Itacitinib	1.96 (0.14)	1.25 (0.14)	2.77 (0.16)
JAK1/Upadacitinib	2.10 (0.18)	0.91 (0.15)	1.64 (0.13)

Subsequently, we measured the residual flexibility of each complex system by estimating the root means square fluctuations (RMSFs) of C_α atoms and compared them with the apo system, as shown in **Figure 3.2C**. It is evident from **Figure 3.2C** that all four complexes shared similar trends in RMSF patterns. The non-binding site regions, including C and N terminals and the activation segment, were more dynamic than other areas. The P-loop, catalytic loop, and hinge areas were very stable and displayed less flexibility for all complexes. The JAK1/Filgotinib system showed higher fluctuations around residues 910-922, 1038, and 1096 compared to other systems.

We also studied the structural compactness of each complex by calculating the radius of gyration (R_g) from the entire production simulation, and the temporal evolution of R_g is shown in Appendix A, **Figure A6**. The average value of R_g for all systems was found to vary between 19.63 and 19.98 Å (see **Appendix A, Table A3**). This suggests no significant change in R_g occurred due to the ligand binding. In addition, we also calculated the solvent-accessible surface area (SASA) for each system, including ligands (see **Appendix A, Figure A7(A, B)),** and the corresponding mean SASA values are displayed in **Appendix A, Table A3.** The average SASA values for all systems range from 13613.88 to 14241.07 $Å^2$. The JAK1/Baricitinib complex system showed a relatively higher SASA value

(14241.07 Å^2) than the other three complexes, including the apo JAK1 kinase. Further, it is evident from Table S3 that the JAK1/Itacitinib complex displayed the lowest SASA value. On the other hand, the lowest SASA was obtained for Upadacitinib (55.63 Å^2), while the highest SASA value was observed for Filgotinib (174.50 Å^2).

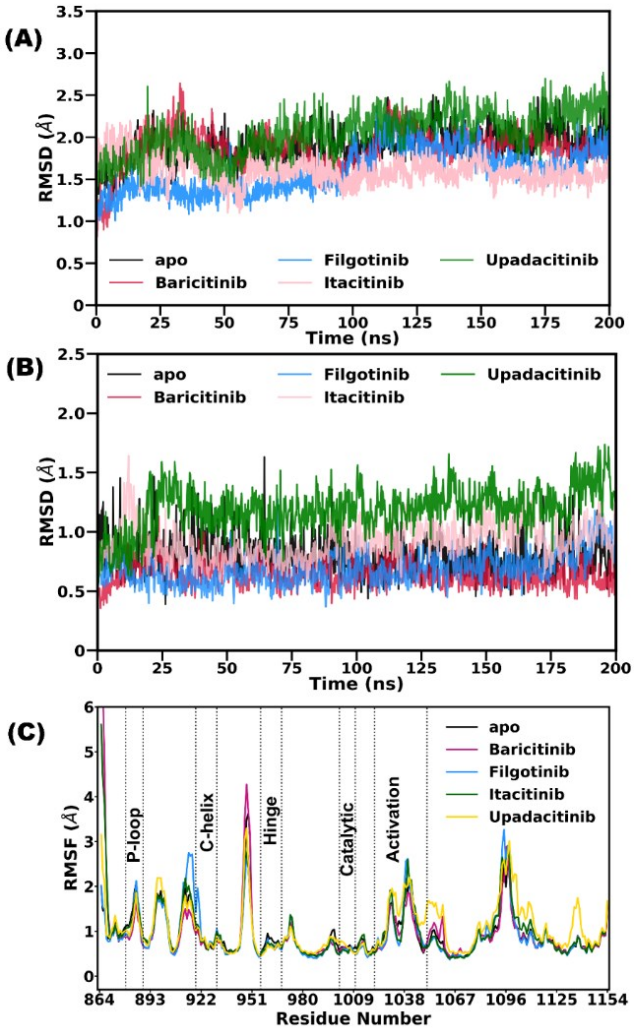


Figure 3.2: Evolution of JAK1 (A) protein backbone, (B) ligand-binding pocket backbone atoms root-mean-square deviations (RMSDs) over simulation time along with the apo system. The residual root-mean-square fluctuations (RMSFs) are shown in (C). Differents binding areas are highlighted with a dotted line rectangle box.

3.3.2. Ligand dynamics

We investigated the conformational dynamics of each ligand in the JAK1-bound state by calculating the corresponding RMSD of heavy atoms, as shown in **Figure 3.3(A)**. We aligned the protein with C_α atoms and then calculated the ligand RMSD. The average value

of RMSD is listed in **Table 3.1**. Baricitinib and Upadacitinib displayed a similar trend in RMSD, and the mean RMSD is almost the same for these two ligands. Itacitinib showed the maximum positional deviation inside the binding pocket during the simulation, and the average deviation was found to be 2.77 Å. On the other hand, Filgotinib showed the least deviation (0.44 Å) during the simulation. Moreover, we compared the JAK1/Baricitinib complex structures with the crystal structure of JAK2/Baricitinib (PDB ID: 6VN8, resolution: 1.9 Å) and an RMSD value of 1.18 Å for Baricitinib (see **Appendix A, Figure A8**).

Subsequently, we generated the potential of mean force (PMF) for each ligand regarding the corresponding ligand RMSD, as shown in **Figure 3.3(B)**. The PMF is constructed to determine the relatively low energy states sampled during the production simulation. It is evident from **Figure 3.3(B)** that the PMF profile is characterized by a single narrow peak for Baricitinib, Filgotinib, and Upadacitinib. However, the PMF profile of Itacitinib was wide and characterized by a single global minimum and two additional secondary minima. As shown in **Figure 3.3B**, the free energy minimum was observed at ~1.5 Å, ~0.4 Å, ~2.9 Å, and ~1.5 Å for JAK1/Baricitinib, JAK1/Filgotinib, JAK1/Itacitinib, and JAK1/Upadacitinib, respectively. In the case of the JAK1/Itacitinib complex, two additional secondary local minima were found at ~1.0 Å and ~1.7 Å.

Next, to understand the ligand movement inside the binding cavity, we generated the free energy landscape (FEL) for each ligand, shown in **Appendix A, Figure A9**. We considered the ligand RMSD and radius of gyration (R_g) as reaction coordinates for the FEL construction. We identified each minimum energy conformation with the K-mean[282] clustering algorithm, as shown in **Appendix A, Figure A9**. As evident from **Appendix A, Figure A9**, Baricitinib and Filgotinib exhibited two probable conformations in the cavity, and the probability of these conformations were 68% and 32% for Baricitinib while 83% and 17% for Filgotinib. Similarly, Itacitinib and Upadacitinib exhibited the three most probable structures, and the sampling spaces were relatively more expansive than the other two. The probability of occurring these three conformations was estimated to be 56%, 22%, and 22% for Itacitinib and 78%, 11%, and 11% for Upadacitinib. Overall, our study

suggests that Itacitinib and Upadacitinib are more flexible inside the binding cavity than the other two ligands.

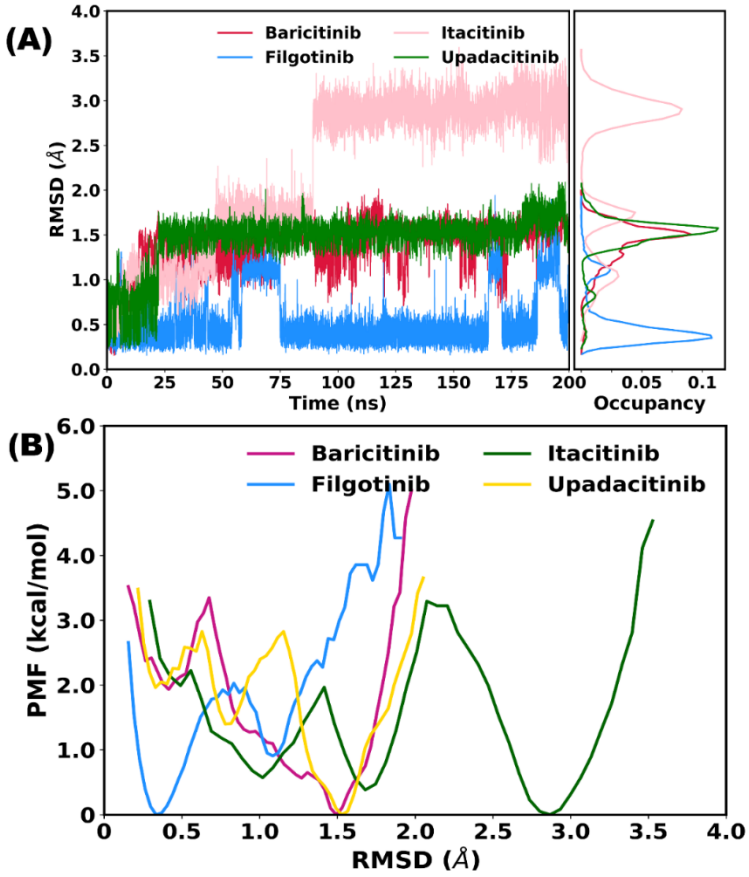


Figure 3.3: The time evolution of ligand heavy atoms RMSD, (A); and potential of mean force (PMF) for all four inhibitors (B), their RMSD, took as a reaction coordinate. The PMF is obtained by calculating the free energy profile along with their respective RMSD values of inhibitors.

3.3.3. Correlation and anti-correlation motions

In order to study the internal dynamics of JAK1, the dynamic cross-correlation maps (DCCM) for all systems were generated from the corresponding MD trajectory and displayed in **Figure 3.4(A-E)**. The DCCM analysis uncovers the relationships between residues and distinct areas by quantifying their relative movement. Highly correlated movement regions or highly positive movement regions are described by yellow to green color. In contrast, red to sky blue indicates the anti-correlation movement regions or highly negative movement regions and zero strength region on the color bar corresponding to no correlation movement regions. The positive and negative values are associated with the

same and opposite displacements of residues, respectively. It is evident from **Figure 3.4**(**A-D**) that R1, R2, and R3 regions displayed a notable strong negative correlation in apo, JAK1/Baricitinib, JAK1/Filgotinib, and JAK1/Itacitnib complexes, and a moderate negative and positive correlation was observed for the JAK1/Upadacitinib complex (**Figure 3.4**(**E)**). The catalytic and activation segment displayed a negative correlation concerning the P-loop region for JAK1 complexed with Baricitinib, Filgotinib, and Itacitinib (**Figure 3.4**(**B-D**)).

In contrast, the JAK1/Upadacitnib complex showed a moderate negative or anti-correlation movement (**Figure 3.4(E)).** The notable strong anti-correlation in apo, JAK1/Baricitinib, JAK1/Filgotinib, and JAK1/Itacitinib complexes demonstrate the direct impact on JAK1 on overall structure, further enhancing the flexibility of the overall conformation. In region R5 and its surroundings, the positively correlated motion increases for Upadacitinib, and the hinge region or the active site area displays no correlation.

The diagonal movements are highly positive for each system, and off-diagonals show different directional movements after the inhibitor binding. Jonniya et al.[277] also reported that the anti-correlated movement increases for the WNK-ligand complexes after the inhibitor binding. However, different inhibitors produced different motion modes, which is clear from this study. In the R4 region, the anti-correlation motions are present in JAK1/Itacitinib, while the anti-correlated motions get diminished in JAK1/Baricitinib, JAK1/Itacitinib, and JAK1/Upadacitinib complexes. Overall, compared to the apo system, all complex systems display a decrease in anti-correlated movements.

3.3.4. PCA and dynamics of free energy landscapes (FEL)

To further characterize the structural variations of the JAK1-inhibitor complexes, principal component analysis (PCA) was performed for all systems to extract the principal modes contributing to the complexes' functional dynamics. The MD trajectories of JAK1 from its free simulation and complexes were employed in PCA by taking the C_α atom into account. The PCA results in terms of eigenvalues (3N, 3x290=870) and eigenvectors were obtained by diagonalizing the covariance matrix for all five systems and displayed in decreasing order (see **Appendix A**, **Figure A10)**. The first 10 PCs contained 76%-86%, while the first two PCs depicted 35%-57% of the overall motion for all systems. The PC1 porcupine plot

of all systems is shown in **Appendix A, Figure A11(A-E)**, where the arrowhead indicates the direction of motion, and the length indicates the amplitude of the motion modes.

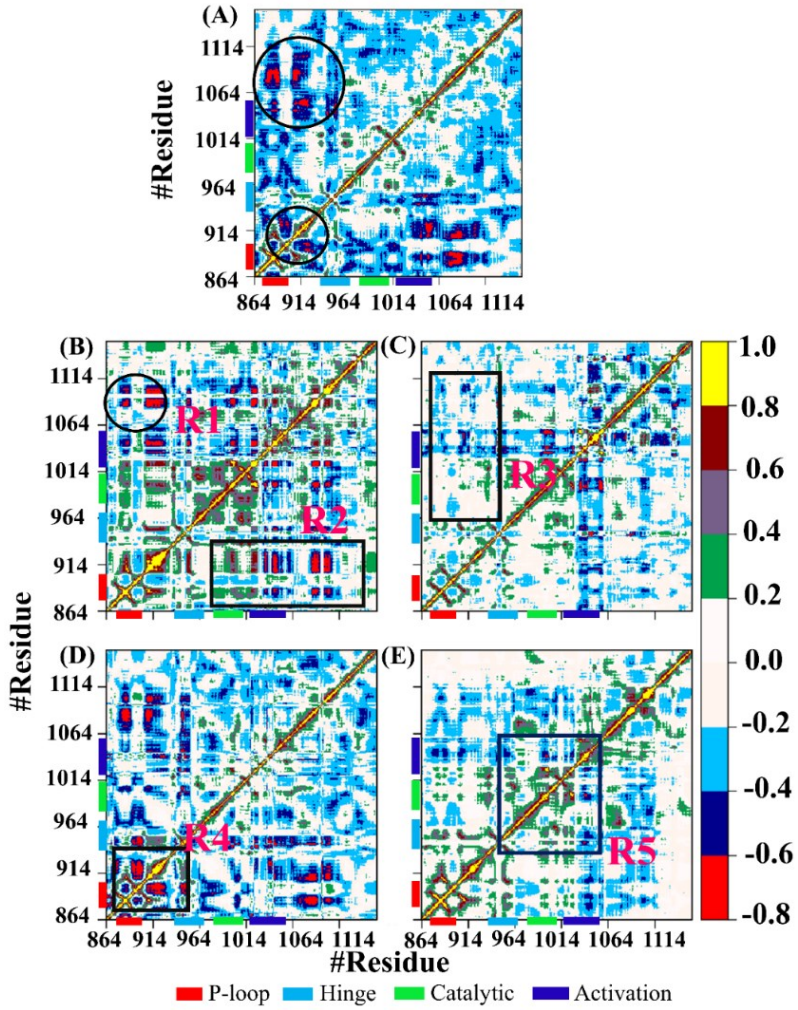


Figure 3.4: Cross-correlation matrices of the coordinates' fluctuations for Cα atoms around their mean positions after the equilibrium of production simulation and shows the relative motions of protein residues. The extent of correlated motions and anti-correlated motions are color-coded. (A) the apo system, (B) JAK1/Baricitinib, (C) JAK1/Filgotinib, (D) JAK1/Itacitinib, and (E) JAK1/Upadacitinib.

Nowadays, the free energy landscape (FEL) is extensively used to explore the structure-function relationship of proteins through the protein lower energy minima conformations during the molecular interaction due to different inhibitor bindings. Qualitative reconnoiter of this FEL can proclaim the energy barriers across the conformational basins. The K-

means[282] cluster analysis was performed in each basin to investigate the relationship between structure and FEL further. The center of the most populated cluster was chosen as the representative structure of each FEL's minimum basin. Herein, the FEL is drawn to understand the conformational dynamics of JAK1 after different types of inhibitor bindings. The FEL was generated utilizing the first two principal components (PC1, PC2) as collective variables and shown in **Figure 3.5(A-E)**. It depicted changes in the free energy minima in apo JAK1 and bound with Baricitinib, Filgotinib, Itacitinib, and Upadacitinib. The noteworthy difference is significantly showing the transition pathway during the entire simulation trajectories and the time variations of PC1 and PC2, as shown in **Appendix A, Figure A12(A-E)**, which gives the idea of sampled space of each system.

As shown in **Figure 3.5(B)**, a single global minimum is observed at the bottom of the FEL of JAK1/Baricitinib and shows three distinct local minima, separated by ~ 1-2 k_BT. The three distinct local minima correspond to the thermodynamic ground state and encompass different configurations. The representative structure (63%) of the most populated basin was extracted and shown in **Figure 3.5(B**). Similarly, the other three systems show a significantly different transition pathway (see **Appendix A, Figure A12(C-E)**) than JAK1/Baricitinib. There is a noticeable difference among them based on free energy minima. The JAK1 bound with Filgotinib, Itacitinib, and Upadacitinib showed two distinct deep energy minima separated by a shallow energy barrier, as shown in **Figure 3.5(C-E)**. The corresponding structures from each system's most populated basin were extracted and displayed in **Figure 3.5(A-E)**. Therefore, the unequal landscape reflects the multiple native states of protein and different possible binding geometries with the inhibitors.

3.3.5. Protein structure network (PSN) analysis

A PSN depicts a network of nodes and their edges. Amino acids represent nodes, and edges are represented as interactions among the nodes; it could be short-range or long-range interactions. Hubs are mainly defined as highest degree nodes, and the network community represents the more connected region of nodes.

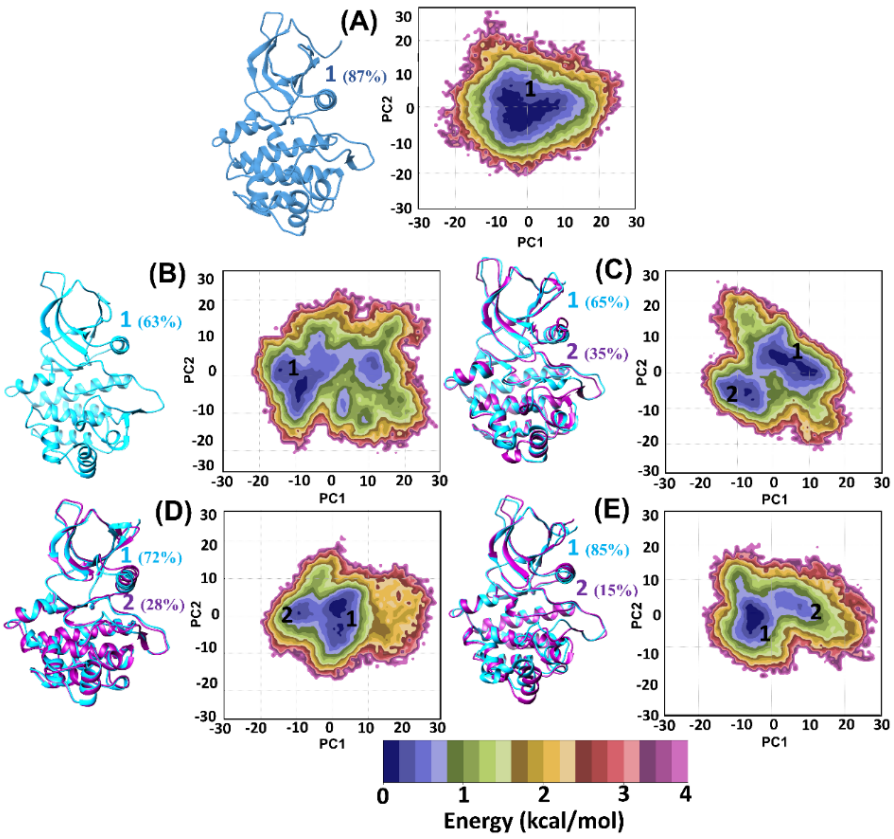


Figure 3.5: Free energy landscape (FEL) analysis of five systems. (A) the apo system, (B) JAK1/Baricitinib, (C) JAK1/Filgotinib, (D) JAK1/Itacitinib, and (E) JAK1/Upadacitinib. The FELs were obtained using the projections of JAK1 C_{α}-atoms position vectors onto the first two principal components as reaction coordinates. As indicated by the color bar, free energy values calculated from the density of distribution are given in kcal/mol.

Network properties such as the number of nodes, links, hub, and community were analyzed for both apo and ligand-bound JAK1. The average network properties are shown in Supporting Information **Appendix A, Table A4,** and hubs and communities are shown in **Figure 3.6(A-J)** for all systems. Interestingly, we found a significant difference in the total number of hubs and communities among all five systems. The apo structure possesses 40 hubs, whereas the inhibitor bound structures, JAK1/Baricitinib, JAK1/Filgotinib, JAK1/Itacitinib, and JAK1/Upadacitinib, consist of 48, 33, 29, and 29 hubs, respectively. As expected, the total number of links and hubs in JAK1/Baricitinib (354, 48) is more significant than apo (328, 40) and the other three bound structures (see **Figure 3.6(B)** and **Appendix A, Table A4**). The Jak1-Baricitinib complex shows the highest probability of

the highest-ranked community (C1, red), indicating a compact network of the JAK1/Baricitinib complex, consistent with its higher stability than the rest of the four systems. The highest-ranked C1 community of apo, Baricitinib, Filgotinib, Itacitinib and Upadacitinib complexes (see **Figure 3.6(F-J)**) stabilizes the binding pocket and C-terminal regions, which reduces the flexibility in those areas and correlated well with the RMSF results (see **Figure 3.2D**). Hubs near the binding site area are assumed to be essential for the inhibitor bindings. Our previous findings (**Figure 3.2C and 3.2D**) suggest that the active site (mainly hinge and catalytic region) of Itacitinib and Upadacitinib shows higher stability and lower fluctuations, further verified by hubs and community PSN analysis. Moreover, the rearrangement in the network properties suggests that each system has global conformations, which define their sensitivity and selectivity towards inhibitors.

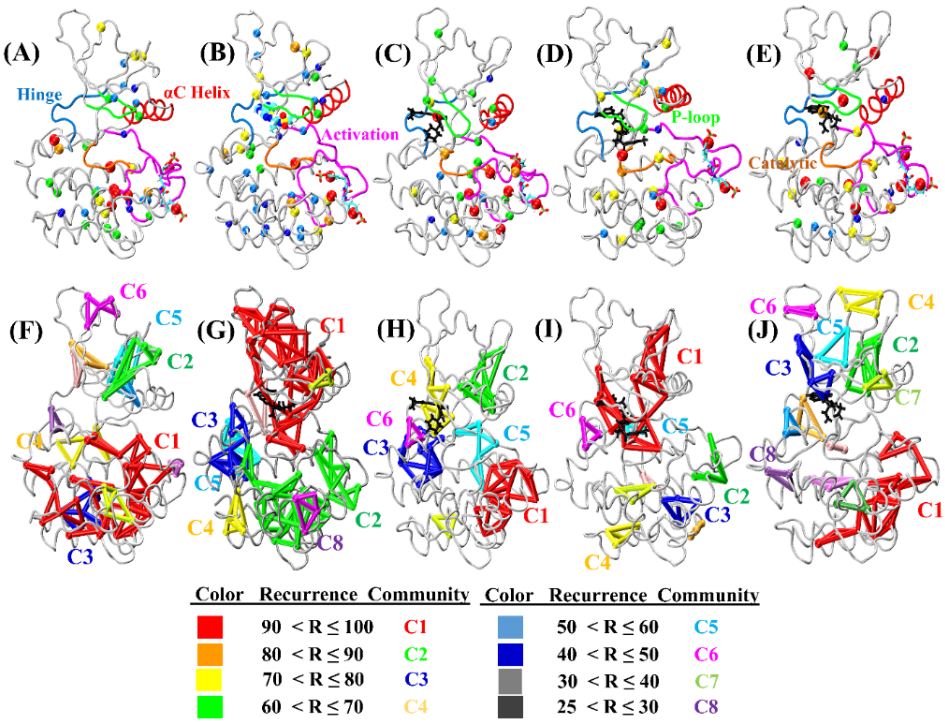


Figure 3.6: The hubs and community structures of residue contact networks in (**A**, **F**) unbound Apo, (**B**, **G**) Baricitinib, (**C**, **H**) Filgotinib, (**D**, **I**) Itacitinib, and (**E**, **J**) Upadacitinib bound with Jak1, respectively. The hubs are represented as spheres centered on the Cα atoms on the Jak1 protein. Community structure describes a fundamental feature of residue contact networks. Colored according to the community nodes, the largest community-ranked 1 is red.

The community analysis suggests that the activation segment of apo, JAK1/Filgotinib, and JAK1/Upadacitinib complexes were involved in the first largest community, and JAK1/Baricitinib and JAK1/Itacitinib in the second-largest community. The inhibitor-bound structures network community suggests that inhibitors induced perturbations in the network connectivity. These comparative PSN results provide an insight into the conformational dynamics that could be essential guidance towards JAK1 specific novel drug developments to treat rheumatoid arthritis-like autoimmune diseases.

3.3.6. Energetics of inhibitor binding with JAK1

The binding free energy (ΔG_{Bind}) of four anti-rheumatoid arthritis inhibitors against the JAK1 kinase was estimated using the MD/MM-PBSA method. The details of the different components of binding affinity of JAK1-inhibitors are summarized in **Table 3.2** and graphically shown in **Appendix A, Figure A13**. It is evident from **Table 3.2** that van der Waals (ΔE_{vdW}), electrostatic interactions (ΔE_{elec}), and non-polar desolvation energy (ΔG_{np}) favor the complex formation, while polar desolvation

Table 3.2: Energetic components of the binding free energy for JAK1-inhibitors complexes in kcal/mol. ([a]$\Delta G_{\text{solv}} = \Delta G_{\text{np}} + \Delta G_{\text{pol}}$, [b]$\Delta G_{\text{pol + elec}} = \Delta E_{\text{elec}} + \Delta G_{\text{pol}}$, [c]$\Delta E_{\text{internal}} = \Delta E_{\text{vdW}} + \Delta E_{\text{elec}}$, [d]$\Delta S$ = configuration entropy, [e]$\Delta G' = \Delta E_{\text{internal}} + \Delta G_{\text{solv}}$.)

Components	ΔE_{vdW}	ΔE_{elec}	ΔG_{pol}	ΔG_{np}	ΔG_{solv} [a]	$\Delta G_{\text{pol + elec}}$ [b]	$\Delta E_{\text{internal}}$ [c]	$-T\Delta S$ [d]	$\Delta G'$ [e]	ΔG_{Bind} [sim]
Baricitinib										
Run1	**-45.72 ± 0.03**	**-26.55 ± 0.06**	**46.86 ± 0.07**	**-4.11 ± 0.0**	**42.76 ± 0.07**	**20.31 ± 0.09**	**-72.27 ± 0.06**	**17.49 ± 0.59**	**-29.51 ± 0.59**	**-12.02 ± 0.59**
Run2	**-49.91 ± 0.05**	**-30.91 ± 0.09**	**53.88 ± 0.09**	**-4.24 ± 0.0**	**49.64 ± 0.09**	**22.98 ± 0.13**	**-80.81 ± 0.10**		**-31.17 ± 0.16**	**-13.68 ± 0.61**
Avg.	**-47.82 ± 0.06**	**-28.73 ± 0.11**	**50.37 ± 0.11**	**-4.17 ± 0.0**	**46.21 ± 0.11**	**21.65 ± 0.16**	**-76.54 ± 0.12**		**-30.34 ± 0.61**	**-12.85 ± 0.85**
Filgotinib										
Run1	**-50.53 ± 0.03**	**-29.93 ± 0.08**	**51.47 ± 0.08**	**-4.79 ± 0.0**	**46.68 ± 0.08**	**21.54 ± 0.11**	**-80.46 ± 0.04**	**17.53 ± 0.65**	**-33.78 ± 0.11**	**-16.25 ± 0.65**
Run2	**-50.32 ± 0.03**	**-29.92 ± 0.09**	**51.51 ± 0.08**	**-4.76 ± 0.0**	**46.75 ± 0.08**	**21.59 ± 0.12**	**-80.24 ± 0.09**		**-33.49 ± 0.14**	**-15.96 ± 0.66**
Avg.	**-50.43 ± 0.04**	**-29.93 ± 0.12**	**51.49 ± 0.11**	**-4.78 ± 0.0**	**46.72 ± 0.11**	**21.56 ± 0.16**	**-80.35 ± 0.13**		**-33.64 ± 0.21**	**-16.11 ± 0.68**
Itacitinib										

Run1	**-63.30 ± 0.04**	**-12.95 ± 0.06**	**45.53 ± 0.05**	**-5.74 ± 0.0**	**39.79 ± 0.05**	**32.58 ± 0.08**	**-76.25 ± 0.07**	**17.88 ± 0.72**	**-36.46 ± 0.11**	**-18.58 ± 0.73**
Run2	**-52.62 ± 0.06**	**-23.02 ± 0.12**	**43.98 ± 0.12**	**-5.24 ± 0.0**	**38.74 ± 0.12**	**20.96 ± 0.17**	**-75.64 ± 0.13**		**-36.88 ± 0.21**	**-19.02 ± 0.75**
Avg.	**-57.96 ± 0.07**	**-17.99 ± 0.13**	**44.76 ± 0.13**	**-5.49 ± 0.0**	**39.26 ± 0.13**	**26.77 ± 0.18**	**-75.94 ± 0.13**		**-36.68 ± 0.24**	**-18.8 ± 0.76**
Upadacitinib										
Run1	**-50.24 ± 0.03**	**-40.37 ± 0.07**	**56.21 ± 0.06**	**-4.16 ± 0.0**	**52.05 ± 0.06**	**15.84 ± 0.09**	**-90.61 ± 0.07**	**19.70 ± 0.71**	**-38.56 ± 0.11**	**-18.86 ± 0.72**
Run2	**-49.05 ± 0.04**	**-41.21 ± 0.06**	**56.12 ± 0.05**	**-4.10 ± 0.0**	**52.02 ± 0.05**	**14.91 ± 0.08**	**-90.26 ± 0.07**		**-38.24 ± 0.11**	**-18.54 ± 0.72**
Avg.	**-49.65 ± 0.05**	**-40.79 ± 0.09**	**56.16 ± 0.08**	**-4.13 ± 0.0**	**52.03 ± 0.08**	**15.38 ± 0.12**	**-90.43 ± 0.10**		**-38.40 ± 0.15**	**-18.70 ± 0.72**

energy (ΔG_{pol}) and the configurational entropy ($-T\Delta S_{\text{conf}}$) penalty disfavor the protein-ligand binding. The favorable intermolecular electrostatic interaction is overcompensated by the desolvation of polar solvation free energy. Consequently, the total polar contribution ($\Delta G_{\text{pol+elec}}$) of each system is unfavorable to the protein-ligand complexation. It means the primary binding interaction is the van der Waals interaction.

Further, the predicted binding free energy ($\Delta G_{\text{Bind}}^{\text{sim}}$) (see **Table 3.2)** reveals that Upadacitinib (-18.70 ± 0.72 kcal/mol) and Itacitinib (-18.80 ± 0.76 kcal/mol) have the highest binding affinity toward JAK1, followed by Filgotinib (-16.11 ± .68 kcal/mol) and Baricitinib (-12.85 ± 0.85 kcal/mol). It is encouraging to note that the rank of the experimental affinities[79,238,243,248] of all the complexes are well-matched with that of our calculated binding free energies. The total polar contributions ($\Delta G_{\text{pol + elec}}$) for the Upadacitinib are less unfavorable than the other three inhibitors. The non-polar desolvation energy ranges from -4.13 to -5.49 kcal/mol for each complex. **Table 3.2** also suggests that ΔE_{MM} (molecular mechanics) and total non-polar energy (vdW + np) favor the complexation.

In general, biological complexations are opposed by reducing the binding partners' configurational entropy[283]. Similarly, the inhibitor binding leads to entropy loss, reducing the binding affinity between protein and inhibitors. For JAK1-inhibitor systems, the configurational entropy varies from 17.5 to 19.7 kcal/mol. However, Upadacitinib and Itacitinib's highest affinity compared to the other two is due to the increased favorable

electrostatic interactions and van der Waals interactions, respectively. Upadacitinib is more potent against JAK1 compared to Baricitinib. This is because in the case of JAK1/Upadacitinib, ΔE_{vdW} (-49.65 kcal/mol) and ΔE_{elec} (-40.79 kcal/mol) are more favorable than JAK1/Baricitinib (ΔE_{vdW} = -47.82 kcal/mol, ΔE_{elec} = -28.73 kcal/mol). Further, the net polar energy ($\Delta G_{pol+elec}$ = 15.38 kcal/mol) is less unfavorable in the case of JAK1/Upadacitinib compared to JAK1/Baricitinib ($\Delta G_{pol+elec}$ = 21.65 kcal/mol). On the other hand, Itacitinib displays a better binding affinity than Baricitinib and Filgotinib due to an increased contribution of ΔE_{vdW} in Itacitinib (ΔE_{vdW} = -57.96 kcal/mol) relative to Baricitinib (ΔE_{vdW} = -47.82 kcal/mol) or Filgotinib (ΔE_{vdW} = -50.43 kcal/mol). Overall, our study suggests that among these four inhibitors, Upadacitinib and Itacitinib are the most potent inhibitor against JAK1. Mainly, the complexation is driven by the van der Waals interactions, implying that the hydrophobic amino acids in the binding pocket played a vital role in the inhibitor binding. Hui Gao and his coworker[244] also theoretically investigated the binding mechanisms of pyrrolotriazin derivative with JAK2 and obtained a similar result as we observed for JAK1 in the current study.

3.3.7. Essential residues for inhibitors binding

To get a more in-depth insight into the JAK1-inhibitors interaction patterns, the total binding energy was decomposed at the residual level using the MM-GBSA scheme for all complexes and shown in **Figure 3.7(A-D)**. It reflects the average energetic contribution of each residue obtained from both MD runs of each complex to the total binding free energy. Each contributing residue (larger than $|\Delta G| \geq 1.0$ kcal/mol) along with its energetic components (elect, vdW, pol, np) are reported in **Appendix A, Table A5**. As shown in **Figure 3.7(A-D)**, predominantly hydrophobic residues play a significant role in the binding and recognizing inhibitor by JAK1. Further, it is revealed from **Figure 3.7(A-D)** that the overall significant attractive binding contributions come from the P-loop, hinge loop, catalytic segment, and the starting part of the activation segment (A-loop). It can further be noted that the inhibitor-JAK1 interaction spectra of JAK1/Baricitinib, JAK1/Filgotinib, and JAK1/Upadacitinib are similar. However, the interaction spectrum is slightly different for JAK1/Itacitinib. This may be due to the asymmetric nature of the inhibitor. K. Sanachai et al. observed similar trends for the binding of Tofacitinib to JAK1/2/3[284]. In the case of JAK1/Upadacitinib, one of the HRD motif residues, $Arg^{1007,}$ contributes disfavorably to

the binding, which is in contrast to JAK1/Baricitinib, JAK1/Itacitinib, and JAK1/Filgotinib. The same trend was observed for JAK1/Tofacitinib[284].

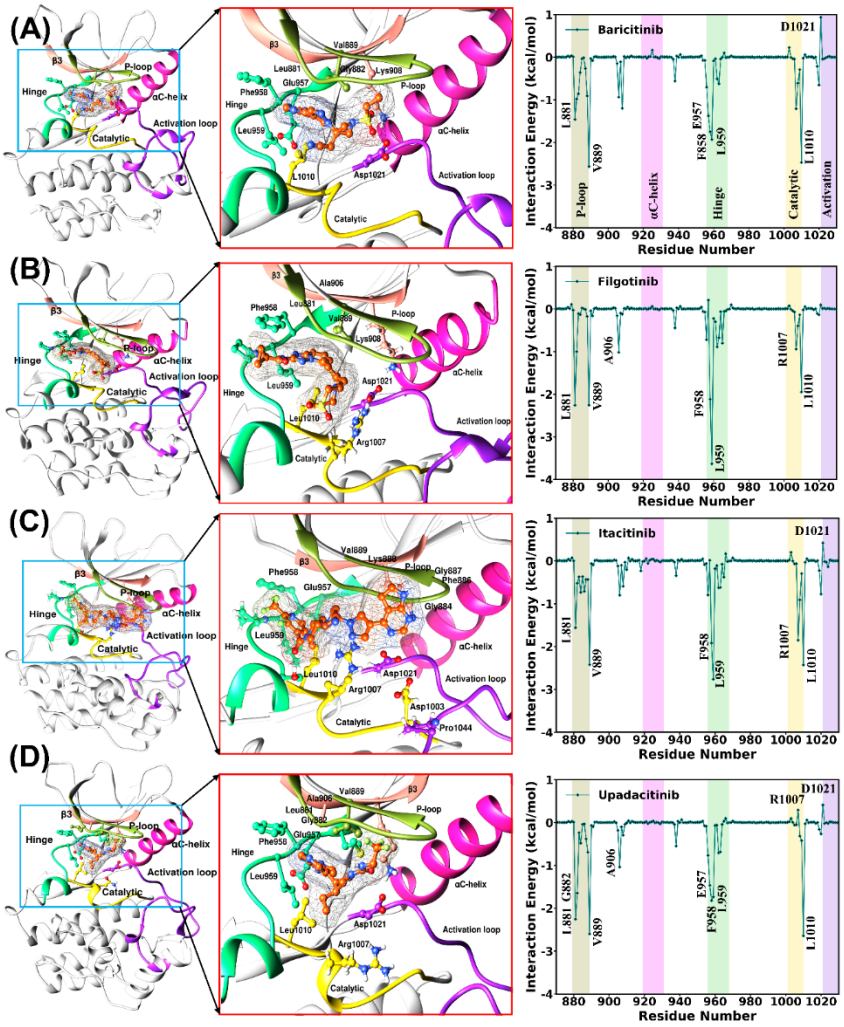


Figure 3.7: Decomposition of the binding free energy for the (A) JAK1/Baricitinib, (B) JAK1/Filgotinib, (C) JAK1/Itacitinib, and (D) JAK1/Upadacitinib, respectively. Residues with a free energy of more than 1 kcal/mol are labeled here, and the color diagram also shows their actual position in the JAK1 structure.

It is evident from **Appendix A, Table A5** and **Figure 3.7(A)** that in the case of the binding of Baricitinib to JAK1, the hotspot residues are Leu[1010], Val[889], Leu[959], Phe[958], Leu[881], Glu[957,] and Gly[882]. The catalytic side residue Leu[1010] and hinge residue Val[889] contribute to the binding by -2.59 and -2.44 kcal/mol, respectively. For JAK1/Filgotinib, the most prominent residues are Leu[959], Leu[881], Val[889], Phe[958], Leu[1010], Arg[1007,] and Ala[906] (see

Appendix A, Table A5 and **Figure 3.7B)**. The energetic contribution of the first three hinge residues is -3.60, -2.27, and -2.26 kcal/mol, respectively. Similarly, for JAK1/Itacitinib and JAK1/Upadacitinib, contributions mainly come from Leu^{959}, Val^{889}, Leu^{1010}, Arg^{1007}, Phe^{958}, Leu^{881}, and Phe^{886} for Itacitinib (**Figure 3.7C**) and Leu^{1010}, Val^{889}, Leu^{881}, Phe^{958}, Leu^{959}, Gly^{882}, Glu^{957} and Ala^{906} for Upadacitinib (**Figure 3.7D**). The contributions toward binding of residues Leu^{959}, $Val^{889,}$ and Leu^{1010} are -2.78, -2.65, and -2.50 kcal/mol, respectively (see **Appendix A, Table A6**). Similarly, residues Leu^{1010}, $Val^{889,}$ and Leu^{881} contributed -2.65, -2.61, and -2.28 kcal/mol toward the binding of Upadacitinib, respectively (see **Appendix A, Table A5**).

Furthermore, the detailed comparison of each complex's per-residual energy components for each complex's seven most highly contributing residues was analyzed and shown in **Appendix A, Figure A14**. The total energy, total polar energy, and the total non-polar energy of the predicted highly contributing residues for each complex are also shown in **Appendix A, Figure A14(A-C)**, respectively. As suggested by **Appendix A, Figure A14A**, the highest total energy contributing residues are Leu^{881}, Val^{889}, Phe^{958}, $Leu^{959,}$ and Leu^{1010}. The total polar contribution (see **Appendix A, Figure A14B**) of residues Leu^{881}, Phe^{886}, and Leu^{1010} disfavor the ligand-binding for all cases. Residue Glu^{957} contributes favorably to Baricitinib and Upadacitinib and disfavors of Filgotinib and Itacitinib binding. The total non-polar contributions of these residues contribute favorably to the ligand-binding (see **Appendix A, Figure A14C**).

We also calculated the energetic contribution of the four main loops (P-loop, hinge loop, catalytic loop, and activation loop) for all four complexes (see **Appendix A, Figure A15(A-D)** and **Table A6**). It is evident from **Appendix A, Figure A15(A)** that in the case of the P-loop, the total non-polar (vdW+np), side-chain, and backbone contributed favorably to the binding. However, the total polar (ele+pol) interactions disfavor the ligand-binding. Overall, P-loop is highly co-operating to the Itacitinib binding in the pocket. **Appendix A, Figure A15(B)** suggests that in the case of the hinge loop, all the energetic components, total polar (ele+pol), total non-polar (vdW+np), and both side-chain and backbone, contributed favorably to the ligand binding. However, in the case of Filgotinib, the total polar contributions disfavor the binding. Finally, the energetic

contribution of the hinge loop is relatively high for Filgotinib. In the case of the catalytic loop (see in **Appendix A, Figure A15(C)**), we found that the total polar (ele+pol) part opposes the ligand binding. Overall, P-loop and catalytic loop display a similar type of energetic contribution. The activation loop (**Appendix A, Figure A15(D)**) yielded different results from the P-loop, hinge, and catalytic loop. Only the total non-polar interactions (vdW+np) contributed favorably, while other energetic components disfavored the binding. Overall, this loop's total energetic contribution is unfavorable to the ligand binding.

3.3.8. Hydrogen bonds and hydrophobic interactions

Hydrogen bonds (h-bond) play a critical role in biochemical reactions, such as protein folding, protein-ligand, protein-lectin, or protein-protein interactions[285,286]. To complement the energetic analysis, we also investigated the h-bond interactions between protein-inhibitor for all four complexes using the corresponding MD trajectories (see **Appendix A, Figure A16(A-E**)). The percentage occupancy of h-bond obtained from the simulations determined the strength and stability of the h-bond interactions. In **Table 3.3,** the average h-bond occupancy obtained from run1 and run2 is recorded. The time evolution of h-bonds is shown in **Appendix A, Figure A17(A-D)**. In our simulations, we observed that both residues Glu^{957} and Leu^{959} formed h-bond interactions with inhibitors. Glu^{957} formed a strong hbond with baricitinib and upadicitinib, but Leu^{959} strongly bonded with filgotinib and itacitinib. N. K Williams et al.[33] also experimentally found these two direct h-bonds in the case of the pan-JAK inhibitor tofacitinib.

In JAK1/Baricitinib, we observed two h-bonds interactions (Glu^{957}@O – Lig@NAT, and Lig@N1 – Leu^{959}@N) with an occupancy of 90.0% and 21.75%, respectively (see **Appendix A**, **Figure A17A** and **Table 3.3**). In an earlier experimental study, it was found that Baricitinib formed two h-bonds with JAK2[287] (see **Appendix A, Figure A16(E)**). Further, it can be seen from **Appendix A**, **Figure A17A** that initially, Arg^{1007} had a strong h-bond interaction with Baricitinib, but after 10 ns, this h-bond broke due to positional deviations**. Appendix A**, **Figure A18(A)** shows that h-bonds and other interactions are also involved in the Baricitinib and JAK1. The gatekeeper residue M^{956} interacts with Baricitinib with a Pi-Sulfer bond (see **Appendix A, Figure A18,** and **Table A7**). Similarly,

residues Leu[881], Val[889], Ala[906,] and Leu[1010] were involved in Pi-Alkyl or Pi-Sigma hydrophobic interactions with Baricitinib. The hydrophobic surface of the binding pocket for all complexes is shown in **Appendix A, Figure A19(A-D)**. The residues that participate in the van der Waals interactions are Val[938], Lys[908], Gly[887], Lys[888], Gly[884], Glu[883], Asp[1021], Gly[882], Arg[1007], Asn[1008], Ser[963], Gly[962,] and Phe[958].

Table 3.3: The hydrogen bonds formed between JAK1 and inhibitor and the corresponding average distance and percent determined using the simulation trajectories of run1 and run2.

Binding couples		Molecular Dynamics			
Acceptor	Donor	Distance (Å)[a]	Occupancy (%)[b]		
JAK1/Baricitinib			Run1	Run2	Avg.
Glu[957]@O	Lig@NAT	2.82	88.92	91.08	90.0
Lig@N1	Leu[959]@N	2.94	22.03	21.46	21.75
JAK1/Filgotinib			Run1	Run2	Avg.
Leu[959]@O	Lig@N23	2.80	87.63	86.42	87.02
Lig@N22	Leu[959]@H	2.94	81.70	85.03	83.36
JAK1/Itacitinib			Run1	Run2	Avg.
Leu[959]@O	Lig@N7	2.84	64.78	64.59	64.68
Lig@N8	Leu[959]@N	2.93	50.54	25.09	37.82
JAK1/Upadacitinib			Run1	Run2	Avg.
Glu[957]@O	Lig@N6	2.82	88.80	87.91	88.36
Asp[1021]@OD1	Lig@N4	2.85	20.27	22.71	21.49

[a]The hydrogen bonds are determined by the acceptor…donor distance of ≤ 3.5 Å and acceptor…H-donor angle of ≥ 120°.
[b]Occupancy of the hydrogen bond is defined as the percentage of simulation time that a specific hydrogen bond exits.

In the case of the Filgotinib bound system, we noticed two strong h-bonds interactions (Leu[959]@O - Lig@N23 and Leu[959]@H - Lig@N22) with an occupancy of higher than 43.36% (see **Appendix A, Figure A17(B**)). Residues such as Val[889], Leu[1010], Ala[906], and Leu[881] are involved in Pi-Alkyl or Pi-Sigma hydrophobic interactions (see **Appendix A**, **Table A7**), and the hydrophobicity surface is shown in **Appendix A, Figure A19(B)**. Filgotinib also interacts with JAK1 via van der Waals interactions and Ser[963], Gly[882],

Met^{956}, Val^{938}, Glu^{957}, Phe^{958}, Gly^{962}, Pro^{960}, Ser^{961}, Lys^{970}, Arg^{1007}, Gly^{883}, Gly^{884}, Glu^{966,} and Lys^{965} play an essential role in this interaction (see **Appendix A**, **Figure A18(B)**).

As is evident from **Appendix A**, **Figure A16(C)** and **Table 3.3**, the Leu^{959} residue has two stable h-bonds with Itacitinib (Leu^{959}@O – Lig@N7 and Leu^{959}@N – Lig@N8) with the occupancy of 64.68% and 37.82%, respectively. As we can see from **Appendix A**, **Figure A17(C)**, the temporal evolution of these h-bonds is stable and within the 3.5 Å distance. On the other hand, His885 forms a new h-bond with Itacitinib after 175 ns simulation. In this case, Leu^{881}, Ala^{906}, Leu^{1010}, and Val^{889} are involved in hydrophobic (Pi-Alkyl or Pi-Sigma) interactions while Phe^{886}, Lys^{965}, Gly^{962}, Phe^{958}, Met^{956}, Lys^{908}, Gly^{1020}, Asp^{1021}, Val^{1009}, Gly^{882}, Gly^{884,} are involved in the van der Waals interactions.

The important h-bonds between JAK1 and Upadacitinib are reported in **Table 3.3** and shown in **Appendix A, Figure A16(D),** and **Figure A17(D)**. The gatekeeper residue Met^{956} and Glu^{883} participate in the Pi-Sulfur and halogen interactions, and residues Val^{938}, Val^{889}, Ala^{906}, Leu^{1010} are involved in Pi-Alkyl or Pi-Sigma hydrophobic interactions (see **Appendix A, Figure A18(D)**). Similarly, residues Phe^{958}, Gly^{1020}, Ser^{963}, Asn^{1008}, Val^{1009}, Leu^{964}, Arg^{1007}, Gly^{884}, Gly^{882}, and Gly^{881} are very crucial for the van der Waals and simple carbon-hydrogen bonding interactions, respectively. Overall, the van der Waals interactions stabilize JAK1-inhibitor complexes.

3.3.9. Atomic contact at the active site and solvent accessibility

We further explored the number of atomic contacts within a sphere of 3.5 Å for each inhibitor during the entire 200 ns simulations, and the same is shown in **Figure 3.8** (left panel). The average #atoms contacts, calculated from the last 60 ns MD simulations, are 17.7 ± 2.7, 23.9 ± 3.4, 14.1 ± 2.3, and 24.5 ± 3.7 for JAK1/Baricitinib, JAK1/Filgotinib, JAK1/Itacitinib, and JAK1/Upadacitinib, respectively. This means that Upadacitinib showed higher #atoms contacts, followed by Filgotinib and Baricitinib. It is known that the binding pocket's water molecules are replaced by inhibitors and could affect protein-ligand interactions[288]. Consequently, to provide insights into the water accessibility of the JAK1 binding pocket, the solvent-accessible surface area (SASA) was calculated by considering a sphere of 3.5 Å around each inhibitor as shown in **Figure 3.8** (right panel). The average SASA values were found to be 476.2 ± 55.7, 874.8 ± 66.8, 617.4 ± 117.8, and 407.1 ± 54.8

Å^2 for JAK1/Baricitinib, JAK1/Filgotinib, JAK1/Itacitinib, and JAK1/Upadacitinib complexes, respectively. The SASA calculation indicates that in the case of JAK1/Filgotinib, more solvent molecules can access the drug-binding pocket. On the other hand, Upadacitinib has less water molecule accessibility, indicating that the binding efficiency of Upadacitinib is more significant than the other three inhibitors. Overall, the number of atomic contacts and active site SASA values consistently support the calculated total binding affinity (**Table 3.2)**.

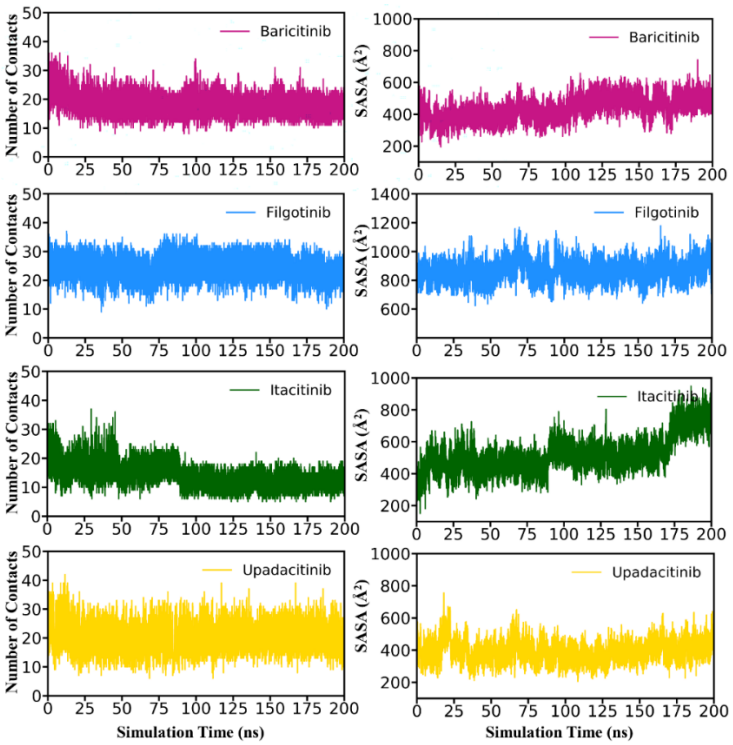


Figure 3.8: Time evolution of the number of atomic contacts within 3.5 Å of each ligand during the last 200-ns simulations (left panel). The solvent-accessible surface area (SASA) of the residues within 5 Å of each competitive inhibitor (right panel).

3.4. Conclusions

Herein, we studied the structural and molecular mechanisms of binding of type-I second-generation competitive inhibitors against the active JAK1 kinase at the atomic level by employing comprehensive protocols, including molecular docking, molecular dynamics simulations, and free energy measurements using the MM/PB(GB)SA method. First of all, the RMSD results suggest that all systems are in good equilibrium during the 200 ns production simulations. The stability of the binding pocket is increased after the inhibitor

binding with JAK1. The overall flexibility of residual motions was affected by inhibitor binding, and correlated movements were increased in Upadacitinib compared to the other three inhibitors complex. The PCA analysis suggests that the structural rearrangement has occurred after inhibitors bind with JAK1, and these structural changes decrease the hydrogen bond occupancy and affect the distance-dependent interactions. In addition, the PSN analysis on the quaternary structure of JAK1 suggests structural and network changes after inhibitor binding. The MM-PBSA and hydrogen bonds results revealed that the inhibitor's affinity decreases in the order Upadacitinib > Itacitinib > Filgotinib > Baricitinib, in good agreement with experimental findings ranks. The selectivity of the inhibitor Upadacitinib toward the JAK1 arises mainly due to an increase in total internal energy and a decrease in the size of total polar energy to other inhibitors. The main interacting profile was observed for hydrogen-bonding interactions with hinge region residues Glu^{957} and Leu^{959} and some strong hydrophobic interactions with residues Leu^{881}, Val^{889}, Ala^{906}, Val^{938}, Leu^{1010}, Met^{956}, Phe^{958}, and Leu^{959}. Most of the critical binding residues provided the inhibitor stabilizations through the van der Waals interactions, such as Phe^{886}, Lys^{965}, Gly^{962}, Phe^{958}, Met^{956}, Lys^{908}, Gly^{1020}, Asp^{1021}, Val^{1009}, Gly^{882}, Gly^{884}. Therefore, the structural and energetic study from MD simulations provides significant insights and knowledge to experiments for designing new inhibitors and analogous compounds against JAK1 to treat rheumatoid arthritis.

Chapter 4

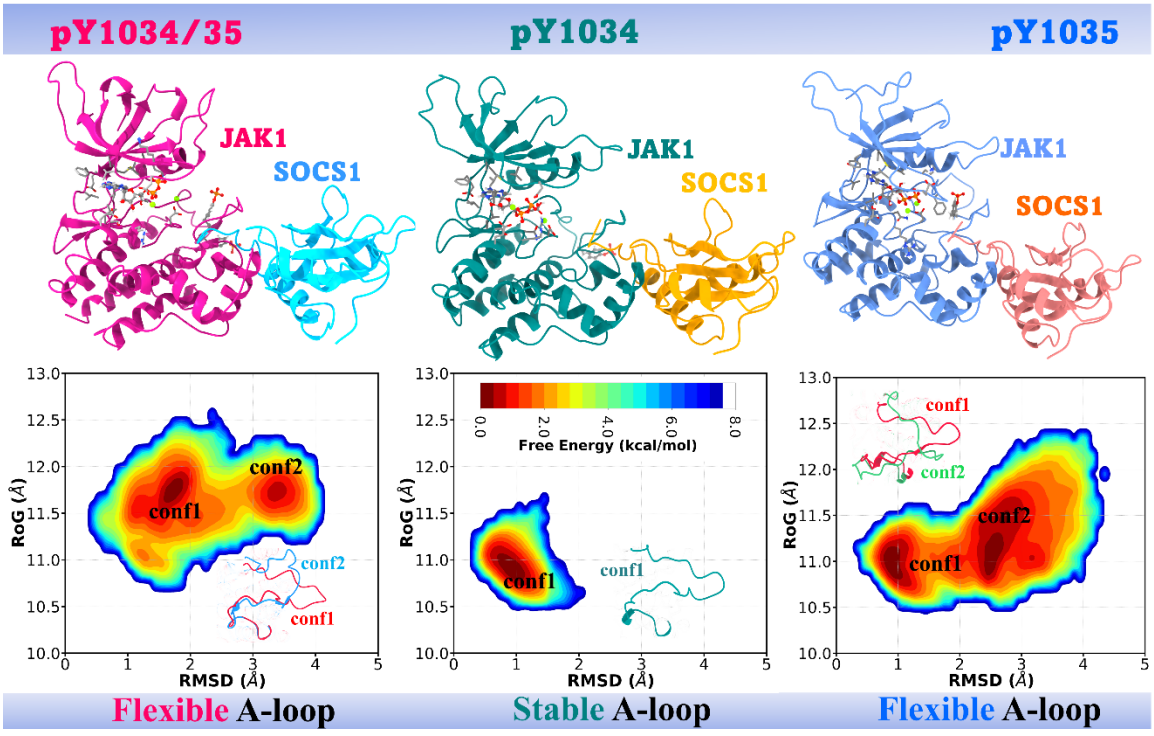
Chapter 4

Phosphorylation-Induced Conformational Dynamics and Inhibition of Janus Kinase 1 by Suppressors of Cytokine Signaling 1

AIM: To elucidate the effect of phosphorylation states on the Janus kinase structure and inhibition mechanisms by SOCS1.

This chapter is reproduced from our original article:

Sk, M.F., Jonniya, N. A., Roy, R., & Kar, P. (2022). "Phosphorylation-induced conformational dynamics and inhibition of Janus kinase 1 by suppressors and cytokine signaling 1". *J. Phys. Chem. B,* 126(17), 3224-3239.



4.1. Introduction

The expected life activities of cells depend upon the appropriate activation of adaptive and innate immunity. The action of specific cytokines essentially controls the immune response, growth, and differentiation of cells. However, excess cytokines or dysregulation of cytokine signaling can cause various diseases, including inflammation, autoimmune diseases, allergies, and cancer-like deadly diseases. Most cytokines utilize the Janus kinase-signal transducers and activators of the transcription (JAK/STAT) pathway to communicate biological responses[289]. In this pathway, JAK kinases are activated *via* the binding of cytokines to their respective receptors. The JAK family has four mammalian members (JAK1, JAK2, JAK3, and TYK2) and homologs in fish, chicken, and *Drosophila*. The activated JAKs phosphorylate the cytoplasmic receptor domain, creating docking sites for STATs proteins or SRC homology 2 (SH2)-containing signaling proteins. Subsequently, STATs are also get phosphorylated and bind to each other, forming a dimer. Then, the dimer form of STAT moves toward the nucleus, binds to DNA and causes the transcription of genes (See **Figure 4.1(A)**). Perverse functioning of the JAK/STAT pathway accelerates autoimmune disorders and tumorigenic disease states[290]. Consequently, the JAK/STAT signaling pathway has received significant attention for targeted therapy[291–293].

The activity of JAK is regulated in several ways to prevent aberrant functioning. The suppressors of cytokine signaling (SOCS) are primary negative regulators of the JAK family, whose expression is induced by the JAK/STAT activation[43,294–297]. Now, these SOCS go into a negative feedback loop and work as a natural inhibitor of the JAK/STAT signaling cascade (see **Figure 4.1(A)**)[298]. Typically, SOCS proteins are divided into eight proteins (SOCS 1-7 and cytokine-inducible SH2) in the human genome. These protein structures mostly share identical domain structures, with highly conserved C-terminal and low sequence conservation N-terminal[294,299]. SOCS protein mainly blocks the STATs protein binding to phosphorylated tyrosine residues of the receptor cytoplasmic domain. Moreover, SOCS1 and SOCS3 directly block JAK and reduce the chance of phosphorylation of STATs *via* the kinase inhibitory region (KIR) and the interaction of SH2 domains[41,300–302]. The interacting surface of SOCS1 and COCS3 with JAK is almost wholly conserved[303], although overall, their sequence similarity is 33%. The most potent

member of the SOCS family is SOCS1[43,295–297] and the principal regulator of many cytokines involved in cell signaling and immune response, like the IFNγ cytokine. Recently, R. Starr *et al.*[304] showed that deletion of the SOCS1 gene in mice led to premature death due to lymphocyte deficiencies, essential organ monocyte infiltration, and systemic inflammatory disease. SOCS1 also works as a tumor suppressor[305]. The downregulation of SOCS1 can play a crucial role in the progression of human cancer. Many researchers found it is silenced in primary tumors in >50% of hepatocellular carcinoma[306], 75% of melanoma[307], 44% of gastric carcinoma[308], and 40% of hepatoblastoma cases.

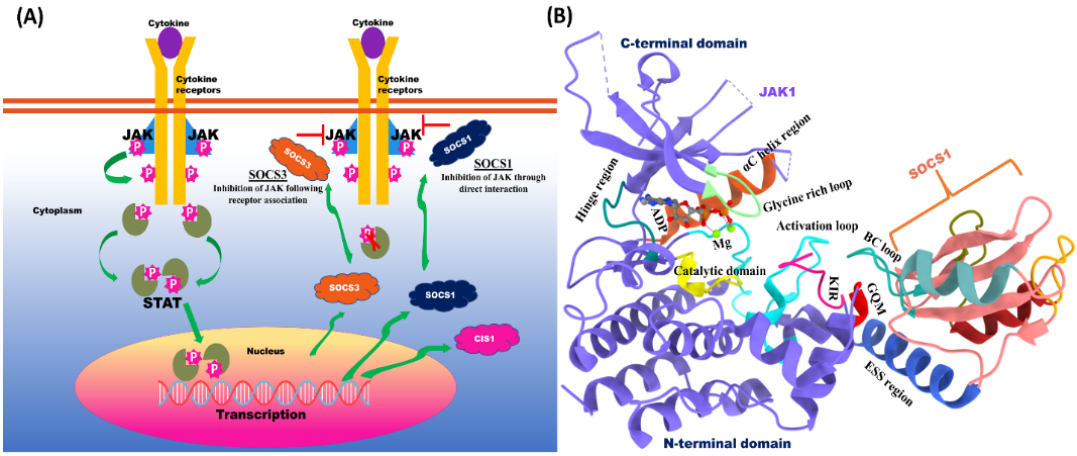


Figure 4. 1: The JAK/STAT pathway mechanisms and the production and action of SOCS proteins. The green line indicates *promotion, the red line denotes inhibition, and P (in the pink star) indicates phosphorylated.* ***(B)*** *The X-ray crystal structure of JAK1 complexed with potent SOCS1 (PDB ID: 6C7Y) with their different essential regions. The glycine-rich loop is colored in light green, the hinge region in dark cyan, the catalytic domain in yellow, the activation loop in light cyan, αC-helix in dark orange, GQM motif in dark red, and the bound ADP is presented in a ball-and-stick representation with carbon atoms in gray, oxygen atoms in red, nitrogen atoms in blue and phosphorous atoms in orange. The SOCS1's KIR region is colored in deep pink, the ESS region in royal blue, and the BC loop in teal color.*

Recently, Liau *et al.*[309] demonstrated the JAK1 inhibition by SOCS1 using X-ray crystallography (PDB ID: 6C7Y, **Figure 4.1(B)**). The kinase domain of JAK1 (JH1; Janus kinase homology 1) adopts a bilobate conformation: a smaller N-terminal domain and a larger C-terminal domain connected by a hinge region (residue 956 - 965), which is one of competitive inhibitor and ATP binding site (see **Figure 4.1(B)**). The N-terminal contains mainly 5-stranded antiparallel β-sheets, an essential regulatory αC-helix region (residues

919 - 932), and a critical phosphate-binding glycine-rich loop region (i.e., P-loop, residues 881 - 890). The C-terminal lobe consists of eight α-helices, three 3/10 helices, three pairs of antiparallel β-strands, a catalytic domain (residues 1001 - 1010), and the crucial activation loop (i.e., A-loop, residues 1021 - 1051). The catalytic domain has HRD (His-Arg-Asp, residues 1001 - 1003) motif, which is stable and conserved across the JAK family. The A-loop of JAK1 starts from the highly conserved DFG (Asp^{1021}-Phe^{1022}-Gly^{1023}) motif in the proximity of the ATP-binding site and ends with the Ala-Pro-Glu motif (i.e., APE motif, residues 1049 - 1051). The catalytic efficiency and protein substrate binding are influenced by A-loop[49], containing two phosphorylable tyrosine residues (Tyr^{1034} and Tyr^{1035}). These tyrosine residues play an essential role in defining the inactive and active conformation of the JAK family. In the C-terminal, there is a conserved motif GQM (Gly^{1097}-Gln^{1098}-Met^{1099}) at αG-helix that plays an important role in the interactions. The SOCS1 protein has three main interacting domains, a KIR region (residues 54 - 63), an extended SH2 subdomain (i.e., ESS, residues 64 - 78), and a BC loop (residues 105 - 113). The KIR region of SOCS1 targets the substrate-binding groove of JAK1 with greater potency and thereby blocks the phosphorylation cascade of STATs proteins (see **Figure 4.1(A)**). A study by Y. Wei *et al.*[310] suggested that the mutation of any residue of the KIR region of the SOCS1 motif leads to a significant decrease in the binding affinity against JAK1.

Only a few theoretical studies[309,310] have been conducted exploring the mechanisms of inhibition of JAKs by SOCS1 at the atomistic level. However, still, several questions remain unclear: (a) as it is well known that the JAK/STAT pathway is phosphorylation-dependent, so what is the effect of different phosphorylation states in the inhibition mechanisms of JAK/SOCS; (b) the impact of SOCS1 binding on the structural changes at the different subdomains and ATP (adenosine triphosphate) binding pocket of JAK1; (c) how the activation segment of JAK1 influences the interaction spectra. An understanding of these fundamental questions has particular significance in structure-guided inhibitors designed for JAKs. Herein, we choose molecular dynamics (MD) simulations to explore the influence of JAK1 activation loop phosphorylation on the JAK1/SOCS1 complexation. We know that a single MD simulation can easily get biased towards a local minima

conformation[311]. Therefore, we used multiple replica MD simulations with 1.0 μs independent production simulation to avoid this false positive conclusion. Overall, we performed an aggregate of 13 μs MD simulations considering three different phosphorylation states (double and singles) of the activation loop. We characterized the structural plasticity of the activation loop and glycine-rich P-loop, identifying several inactive conformations with the help of principal component analysis (PCA), two-dimensional free energy landscape profiles, and protein structural networks.

Further, we explored the conformational changes in the different domains of JAK1 upon SOCS1 binding and compared them with its SOCS1 unbound apo form. Finally, we elucidated the effect of phosphorylation on interaction spectra of JAK1/SOCS1 with the help of the widely used molecular mechanics Poisson-Boltzmann surface area (MM-PBSA) approach[272,273,312].

4.2. Materials and Methods

4.2.1. System preparations

The X-ray crystallographic structure of the JAK1^{pY1035}/SOCS1 complex (PDB ID: 6C7Y)[309] was employed in the current study. The crystal structure was solved for the mono-phosphorylation state (pTyr^{1035}). The missing residues were added using the Modeller[313] plugin of Chimera[251]. The co-crystallized water and other molecules were removed except ADP (Adenosine diphosphate) and Magnesium. The *LEaP* module [314] of AmberTools19[255] was used to add all the missing hydrogen atoms. We constructed the double phosphorylated (pTyr$^{1034/1035}$), and one of the single phosphorylated (pTyr^{1034}) states manually from the experimental structure of JAK1^{pY1035}/SOCS1. Finally, we prepared three complex systems, namely double phosphorylated A-loop (both Tyr^{1034} and Tyr^{1035}; com-2pY), Tyr^{1034} (com-pY^{1034}), and Tyr^{1035} (com-pY^{1035}) phosphorylated JAK1 complexed with ADP and SOCS1. The unphosphorylated A-loop was not considered in this study due to the physiological condition of SOCS1. Similarly, we also prepared JAK1 (complexed with ADP) at three different A-loop phosphorylated states (apo-2pY, apo-pY^{1034}, and apo-pY^{1035}) and apo form of SOCS1 (apo-SOCS1) only for comparison study. Overall, we constructed seven systems for further molecular dynamics (MD) simulations study.

Table 4.1: List of simulated systems and simulation lengths.

Systems		Total Atoms	Water molecules	Replica runs	Simulation length	Total Simulation
Com-2pY	**JAK1^{2pY} + SOCS1**	61488	18323	3	1 µs	3 µs
Apo-2pY	**JAK1^{2pY}**	39156	11500	1	1 µs	1 µs
Com-pY1034	**JAK1^{pY1034} + SOCS1**	61466	18316	3	1 µs	3 µs
Apo-pY1034	**JAK1^{pY1034}**	39137	11494	1	1 µs	1 µs
Com-pY1035	**JAK1^{pY1035} + SOCS1**	61487	18323	3	1 µs	3 µs
Apo-pY1035	**JAK1^{pY1035}**	39164	11503	1	1 µs	1 µs
Apo-SOCS 1	**SOCS1**	17151	5096	1	1 µs	1 µs
All trajectories						13 µs

4.2.2. Simulation protocol

All ionizable amino acid's protonation state was determined using the PROPKA3.1[254] server corresponding to pH 7. The DFG-motif Asp[1021] was protonated (Ash). The histidine (His) was at the Nε2-protonated (Hie) τ tautomer sate. AMBER ff14SB[257] was employed to model the protein, while the updated general Amber force field (GAFF2)[256] was used to describe ADP. The phosaa10[260] force field was used for phosphorylated tyrosine residues. Each system was enclosed on a periodic truncated octahedron TIP3P water box cell with 10 Å from the walls, and Na^+ and Cl^- ions were added to maintain the system's charge neutrality. MD simulations were performed using the *pmemd.cuda* module of AMBER18[315]. The system temperature was maintained at 300 K using the Langevin thermostat[316]. The long-range electrostatic interactions were calculated using the particle-mesh Ewald (PME)[263] scheme. The non-bonded cut-off was set to 10 Å. The energy of all systems was minimized using the steepest descent and conjugate gradient algorithms restraining the solute. After the energy minimization, the system was gently heated from 0 K to 300 K at the canonical (NVT) ensemble with positional constraints on to the protein (spring constant of 2 kcal/ (mol. $Å^2$)). Subsequently, solvated systems were equilibrated in an isothermal-isobaric (NPT) ensemble for 1.0 ns, and the pressure was controlled by Berendsen barostat[265]. Finally, all the complex systems underwent 3 x 1 µs production run under an NPT ensemble, while a single run of 1 µs was realized for the remaining systems.

Covalent bonds involving hydrogen atoms were constrained at their equilibrium distances with the SHAKE algorithm[317]. A 2.0 fs time step was used for all simulations, and the coordinates were recorded every 10.0 ps for further analysis. Overall, 300,000 and 100,000 frames were generated for the three complexes and four apo systems, respectively. All the systems and simulation details are reported in **Table 4.1**.

4.2.3. Post-dynamics simulations trajectory analysis

All analyses, including root-mean-square deviation (RMSD), root-mean-square fluctuations (RMSF), the radius of gyration (Rg), and solvent accessible surface area (SASA), were performed using the *Cpptraj* module[266] of AmberTools19. For hydrogen-bond studies, a distance of ≤ 3.5 Å and an angle cut-off of ≥120º were used[318–322]. All scientific plots were generated using in-house *Python* scripts, and the 3-D protein structures were visualized and created using ChimeraX[323].

4.2.4. Essential dynamics analysis and free energy surface (FES)

To identify the low energy structures from the high dimensional phase space trajectories, we applied the principal component analysis (PCA) technique[268] using the *Cpptraj* module of AmberTools19. The details mathematical background of PCA is discussed in **Chapter 2, section 2.7.3**.

4.2.5. Protein structure network (PSN) analysis

We used a mixed protein structure network (PSN)[270] and an elastic network model-normal mode analysis (ENM-NMA) 324-based method to investigate protein networks in three different phosphorylated structures of the JAK1/SOCS1 complex, which are otherwise not easily detectable. PSN analysis enumerates network features, such as nodes, edges, hubs, links, community, etc., and the shortest path communication from stable states of molecular dynamics (MD) trajectories. In PSN, a node corresponds to a residue, and a hub represents a highly linked residue that generally refers to the degree of a node. Community or modularity is the region in a network where nodes are more connected to each other. We explored the residual communication changes in three phosphorylated JAK1/SOCS1 complexes using WebPSN v2.0[207]. All the hubs and network communities were visualized using VMD[271].

4.2.6. Molecular mechanics Poisson-Boltzmann surface area (MM/PBSA) calculations

The binding free energy of JAK1/SOCS1 was calculated using the widely used molecular mechanics Poisson-Boltzmann surface area (MM/PBSA) scheme[272,273,312,324,325] in the AMBER18 suite. The details of the mathematical background were discussed in **Chapter 2, section 2.8.3.**

4.3 Results and Discussions

4.3.1. Stability and dynamics of phosphorylated conformation of JAK1 bound and unbound to SOCS1

All-atom molecular dynamics simulations in triplicate were conducted to evaluate the structural stability and dynamics of three different phosphorylated states of JAK1/SOCS1 complexes. First, we calculated the backbone atoms RMSD of the solute concerning the initial structure, and the time evolution of RMSD for each system is depicted in **Appendix B, Figure B1.** The corresponding RMSD distribution for these three complex systems is shown in **Figure 4.2(A)**. The average RMSD value was calculated by block averaging and reported in **Appendix B, Table B1**. It is evident from **Figure B1** that JAK1 bound to SOCS1 displayed less stability and more significant RMSD fluctuations compared to SOCS1. This agrees with the experimental data[309] and computational study[310]. Finally, the time evolution of RMSD of complexes for three different phosphorylation states is similar to the respective RMSD profile of JAK1 bound to SOCS1, implying that all protein fluctuations are mainly caulescent from JAK1[310].

Figure 4.2(A) suggests that the RMSD of com-pY1034 is quite restricted compared to the other two complexes. The distinctive peaks from the distribution correspond to a system's most probable conformational states. Herein, we observe that com-2pY has a distinct most probable peak at 2.31 Å and com-pY1034 has at 2.30Å with a shoulder peak at 1.71 Å. Similarly, com-pY1035 has the most probable peak at 2.00 Å and a shoulder at 3.23 Å. Moreover, **Table B1** suggests that the complex conformations of JAK1 and SOCS1 are more stable than the respective apo conformations, which is expected. **Figure 4.2(B, C, D)** shows the RMSD distribution and corresponding most probable conformation in complex

and apo form. We see a drastic change in the RMSD density distribution of apo (at ~ 3.15 Å) and complex (at ~ 1.25 Å) form of SOCS1 protein.

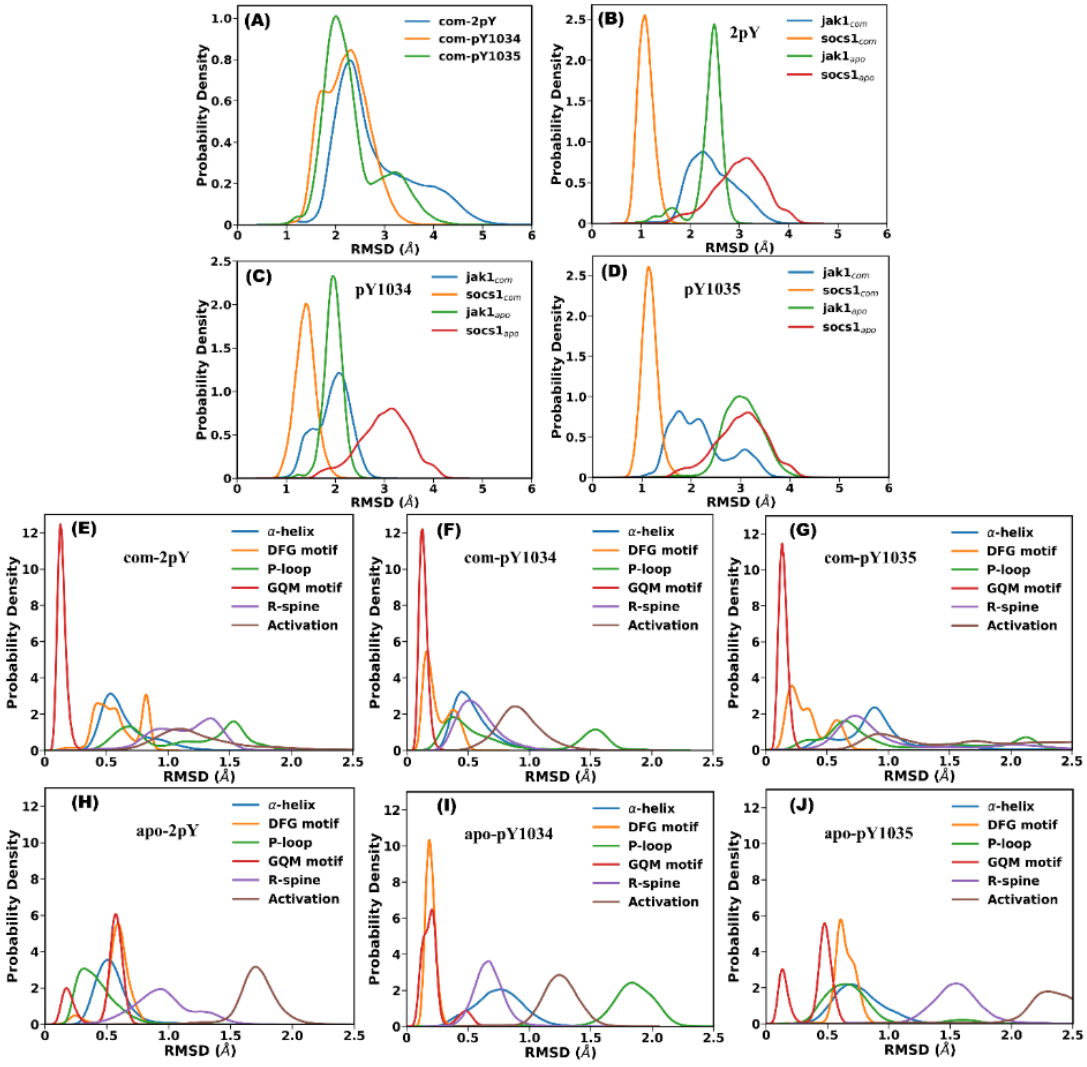


Figure 4.2: (A) The distribution of the backbone RMSD of three complexes corresponding to the three different phosphorylation states. The backbone RMSD's distribution of JAK1 and SOCS1 in complex form ($Jak1_{com}$/$SOCS1_{com}$) and apo form ($Jak1_{apo}$/$SOCS1_{apo}$) for three different A-loop phosphorylated systems, **(B)** double phosphorylated, 2pY, **(C)** single phosphorylated, pY1034, and **(D)** single phosphorylated, pY1035, respectively. The backbone RMSD's distribution of αC-helix, DFG motif, glycine-rich P-loop, GQM motif, regulatory spine (R-spine), and activation loop of three different phosphorylated states of JAK1 in complex with and without SOCS1 (apo) forms. **(E-G)** complex, phosphorylated systems and **(H-J)** apo system, respectively.

Subsequently, we calculated and compared the RMSD distribution of three phosphorylated states of JAK1 and found that the com-pY1034 phosphorylated system is more stable and deviates less from the initial conformation than com-2pY or com-pY1035. The distribution of RMSD for com-2pY and com-pY1035 complex is wider (~ 0.92 Å to ~ 4.14 Å) compared to com-pY1034 (~ 0.86 Å to ~ 2.85 Å). We projected the conformational space onto the one-dimensional, and two-dimensional free energy surfaces spanned by the backbone RMSD of JAK1 and SOCS1 proteins (see **Figure 4.3(A, B, C, D)**). It is evident from **Figures 4.3A** and **4.3C** that the com-pY1034 complex conformational sampling space is restricted with a single global free energy basin, center at (1.80 Å, 1.32 Å), which corresponds to its native active conformation. For com-2pY and com-pY1035, the analysis of free energy landscapes indicates that the conformational sampling spaces are broader (see **Figures 4.3A, 4.3B,** and 4.**3D**). It is noteworthy that com-pY1034 shows two distinct conformations (active and inactive), one global minimum (active), center at (1.83 Å, 1.16 Å), and a local minimum (inactive) at (3.85 Å, 1.16 Å); corresponding active and inactive conformation is shown in **Figure 4.3D**.

Further, we investigated the active and inactive conformations using two important reaction coordinates (the pseudo torsion ζ(Gly1020) and distance *R* between K^{908}:NZ and E^{925}:CD) suggested by Tsai *et al.*[326], and free energy surface (FES) based on these reaction coordinates shown in **Figure 4.4A**. As shown in **Figure 4.4A**, conf1 and conf2 are active conformations due to DFG in and αC-helix in structure. Similarly, conf3 is an inactive conformation and moderately DFG out and complete αC-helix out conformation, shown in **Figure 4.4B**. The representative snapshot of conf1, conf2, and conf3 structures and K^{908}, E^{925}, D^{1021}, and F^{1022} are shown in **Figure 4.4(C-E).** Hence, RMSD and FES analysis suggests that com-pY1034 is more stable and less flexible throughout our simulations. Previously, Jonniya et al.[277] reported that the WNK1 kinase single $pSer^{382}$ phosphorylation state is more stable than doubly phosphorylated ($pSer^{382}$ and $pSer^{378}$) conformation, which is consistent with our JAK1 results.

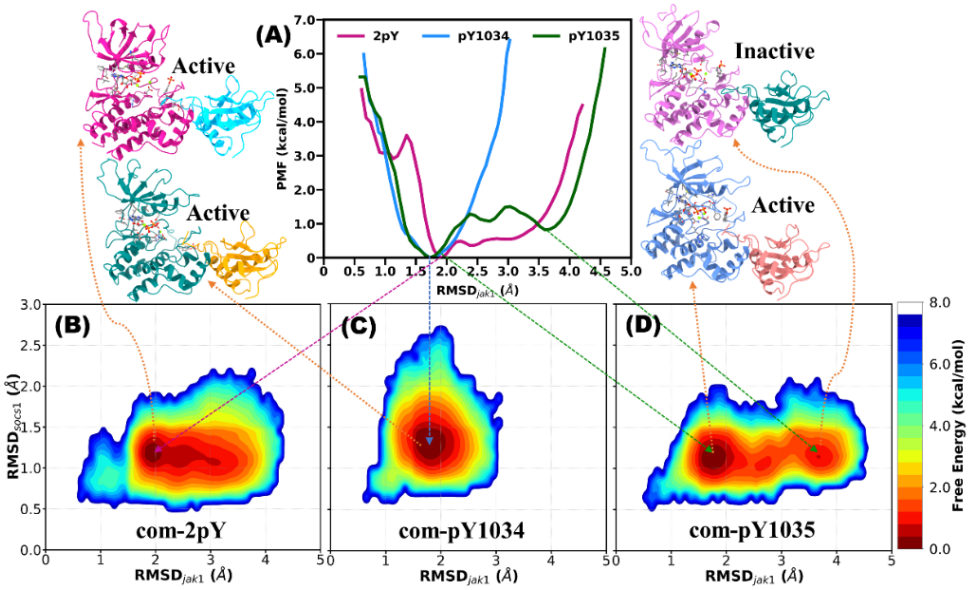


Figure 4. 3: The free energy landscape of JAK1/SOCS1 complex protein measured by the backbone RMSD of JAK1 and SOCS1 for three different phosphorylated states. (A) One dimensional free energy (potential mean force, PMF) for three complexes and JAK1 RMSD as reaction coordinates. (B, C, and D) Two-dimensional free energy landscape of three phosphorylated complexes.

We also calculated the backbone RMSD and estimated the corresponding PDF of the different meaningful sub-domain of JAK1 and SOCS1, see **Figure 4.2** and **Appendix B Figure B2** and **Figure B3**. The RMSD based probability density function suggests that N-terminal αC-helix, hinge area, catalytic spine (C-spine), and the catalytic region shows the lowest fluctuations with RMS deviation within 0.5 Å. We see some dynamic structural rearrangements of the P-loop as indicated by the probability density function, and com-pY1035 shows a higher value of RMSD of this region. The GQM motif of JAK1 directly interacts with SOCS1 and shows strong stability and the lowest RMS deviation after complex formation for all three phosphorylated states. The JAK1 complexes show two distinct RMSD probabilities, corresponding to two DFG motif conformations, but all apo-JAK1 are restricted with a single conformation (see **Figure 4.2(E-G)**). The regulatory spine (R-spine) of JAK1 is well aligned in all apo-JAK1, but in the complex form with SOCS1, these hydrophobic clusters (R-spine) are disordered in the case of com-2pY and com-pY1035, as suggested by **Figure 4.2E** and **4.2G**.

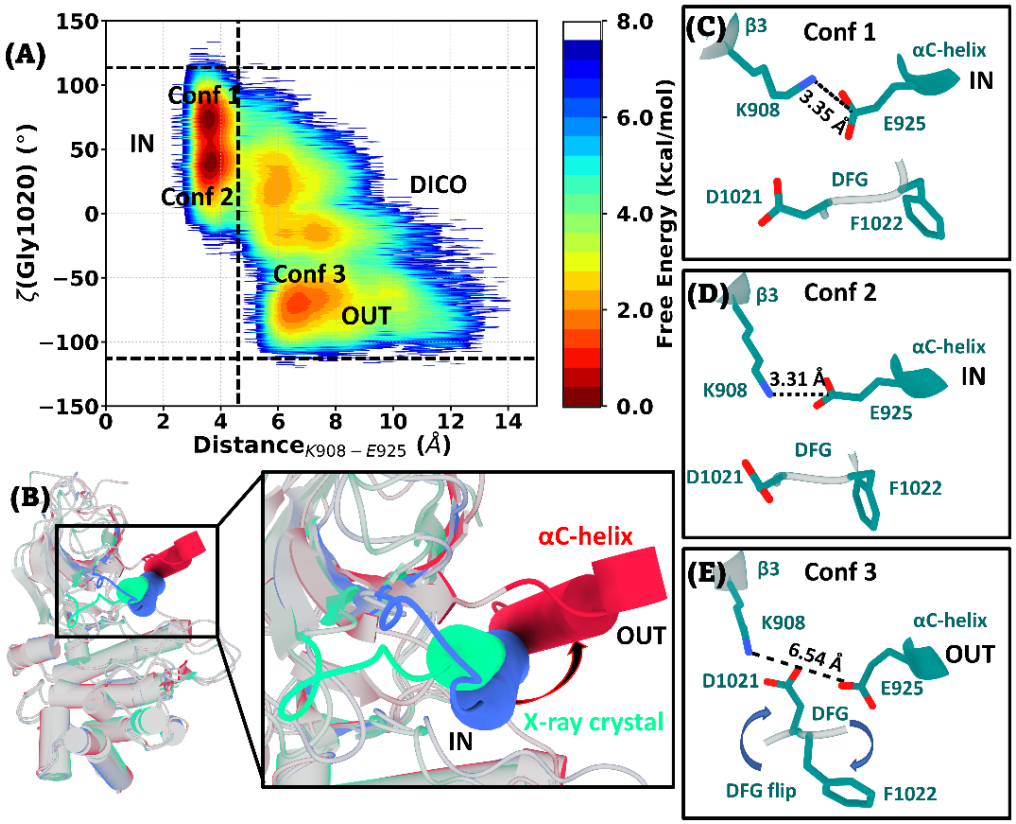


Figure 4.4: The conformational landscape of pY1035 JAK1 complex with SOCS1. **(A)** FES as a function of the pseudo torsion ζ(Gly1020) and distance R between K^{908}:NZ and E^{925}:CD. ζ(Gly1020) is defined using the CA atoms of I^{1019}, G^{1020}, D^{1021}, and F^{1022}. Horizontal and vertical lines are plotted at 120° and 4.5 Å to define cut-off values. (B) The αC-helix in/out structure is shown in blue and red color, and the starting crystal structure is shown in turquoise color. (**C-E**) Representative snapshots represented the FES minima. K^{908}, E^{925}, D^{1021}, and F^{1022} are shown.

Moreover, we calculated the RMSD based PDF of phosphorylated residues containing the A-loop subdomain of JAK1. The RMSD values of three different phosphorylated states of JAK1 are quite dissimilar in the complex and apo, indicated by **Figure 4.2**. The double, $pTyr^{1034/1035}$ and $pTyr^{1035}$ phosphorylation show higher deviation from their initial conformations in both complex and apo systems. Finally, we measured the RMSD based PDF of different sub-domain of SOCS1 in bound and unbound with JAK1, shown in **Appendix B, Figure B3**. The conserved αA and αB helices deviate marginally from their respective native initial conformation in both bound and unbound states (see **Figure B3**).

However, we notice drastic conformation changes at JAK1 interacting subdomain KIR, ESS, and BC-loop in the apo and complex states. As we see from **Figure B3,** after the complex formation with JAK1, the interacting portion of SOCS1 gets more stabilized than the off-interacting part, consistent with the experimental study by Liau *et al*.[309].

4.3.2. Conformational diversity of hydrophobic spines.

The previous study by Kornev *et al.*[245,246] suggested eight hydrophobic residues as a catalytic spine (C-spine) and four non-consecutive hydrophobic residues that constitute a regulatory (R-spine) spines to play an essential role in the protein kinase skeletal assembly. The R-spine of JAK1 consists of four hydrophobic residues (Tyr^{940}, Leu^{929}, $Phe^{1022,}$ and His^{1001}) along with an anchored residue, Asp^{1063} of hydrophobic αF-helix, which serves as the base of the R-Spine. The R-spine of an active JAK1 is nearly aligned or linear and least exposed to the solvent, while in inactive structure is broken by the αC-helix (Leu^{929}) and A-loop (Phe^{1022}). The time evolution of backbone RMSD of R-spine (see **Figure 4.5**,) residues suggests that double (Tyr^{1034} and Tyr^{1035}) and single $pTyr^{1035}$ phosphorylation show two different conformations of R-spine, aligned (active conformation) and broken (inactive conformation) due to the displacement of Phe^{1022} (DFG motif) in com-2pY and both Leu^{929} (αC-helix) and Phe^{1022} (DFG motif) in com-pY1035. In the case of com-pY1034, we observe that the R-spine maintains a linear alignment throughout the simulation, suggesting that the $pTyr^{1034}$ phosphorylation causes slightly higher stabilization to the R-spine and is consistent with the WNK1 kinase[277].

The protein kinase C-spines comprises the residues (Val^{889}, Ala^{906}, Leu^{1010}, Val^{1011}, Val^{1009}, Leu^{964}, $Thr^{1070,}$ and Leu^{1074}) from the small N-terminals and large C-terminals; the residues Val^{889} and Ala^{906} form an interaction with the adenine component of ATP. The backbone RMSD of the C-spine suggests that the phosphorylation does not affect the structural dynamics of the C-spine, which is in contrast to the R-spine.

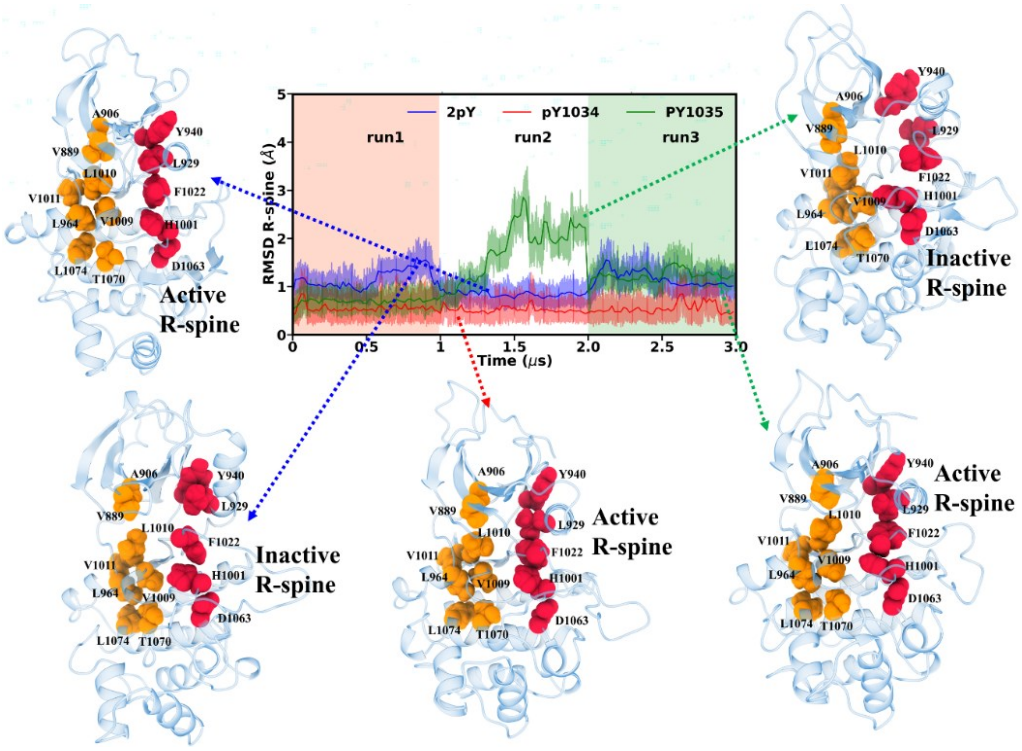


Figure 4 5: The time evolution backbone RMSD of regulatory spines (R-spine) of com-2pY, com-pY1034 and com-pY1035, respectively. The hydrophobic catalytic spine (C-spine) and R-spine shown atom sphere model in dark orange and red colour in the corresponding active (R-spine aligned) and inactive (R-spine broken) conformations of JAK1.

4.3.3. P-loop conformational dynamics

It is well known that the phosphate-binding loop (P-loop) has a crucial role in the ATP and competitive inhibitors binding and controlling the water accessibility of the binding pocket of all kinds of tyrosine kinases[327,328]. Therefore, the conformational dynamics of the P-loop is fundamental to exploring rigorously to design selective inhibitors[329,330]. Recently, *Collie et al*. suggested that two different conformations are critical for competitive inhibitors, unusual folded and extended conformation for c-Met kinases[331]. Herein, we discuss the conformation dynamics of JAK1's P-loop for three different phosphorylated states and complexed with SOCS1. Here, we generated the 2D free energy landscape of the P-loop based on the P-loop RMSD value and center of the mass distance between the P-loop and the conserved HRD motif (**Figure 4.6(A-F)**). It is evident from **Figure 4.6** that the three complexes have different conformational dynamics of the P-loop and agree with the experimental results[309]. In the cases of com-2pY and com-pY1034, we obtain three

conformations, in which conf1 is unusual, folded conformation, conf2 is the semi-extended conformation (inter-mediate), and conf3 is the native usual extended conformation, see **Figure 4.6**(**A, D**) and (**B, E**). However, com-pY1035 shows diverse conformational dynamics, one folded or 'kinked like' (conf1), which is usually found in c-Abl, c-Met, and Aurora-a kinases[329,332], one semi-extended (conf2), and two extended (conf3 and conf4) conformations. According to Xie *et al.,* these folded and semi-extended conformations are inactive conformations of kinases[330]. Moreover, in these folded conformations, the distance between the HRD motif and P-loop increases, suggesting that the P-loop is opening, and the volume and water accessibility of the ADP binding pocket increases.

4.3.4. Collective motions of JAK1 and SOCS1

We performed a residue-based RMSF calculation for three complex systems and compared it with the respective apo-system (see **Appendix B, Figure B4,** and **Figure B5**). This RMSF plot shows the flexible subdomain of the protein. It is evident from **Figure B4** and **Figure B5** that the loop regions are relatively more flexible, and the evolutionarily conserved part shows high rigidity, as expected and consistent with the previous theoretical[68,333], and experimental[309] studies of tyrosine kinases. It is evident from **Figures B4** (**A, B,** and **C**) that after the complex formation with SOCS1, the overall flexibility of JAK1 increases except for the SOCS1 binding residues and the GQM motif. Moreover, the N-terminal of JAK1 displays more flexibility than the apo system. However, the C-terminal of JAK1 gets stabilized after the complexation (see **Appendix B, Figure B6**). To understand the residue-wise mobility correlation between the complex and apo form of JAK1, we performed the statistical correlation analysis (Chi-square test, *P*-value <0.05), as shown in **Figure B4**. The correlation analysis suggests that pY1034 phosphorylation of JAK1 is highly correlated with apo (r = 0.93), which agrees well with our RMSD analysis and free energy landscapes (see **Figure 4.3)**.

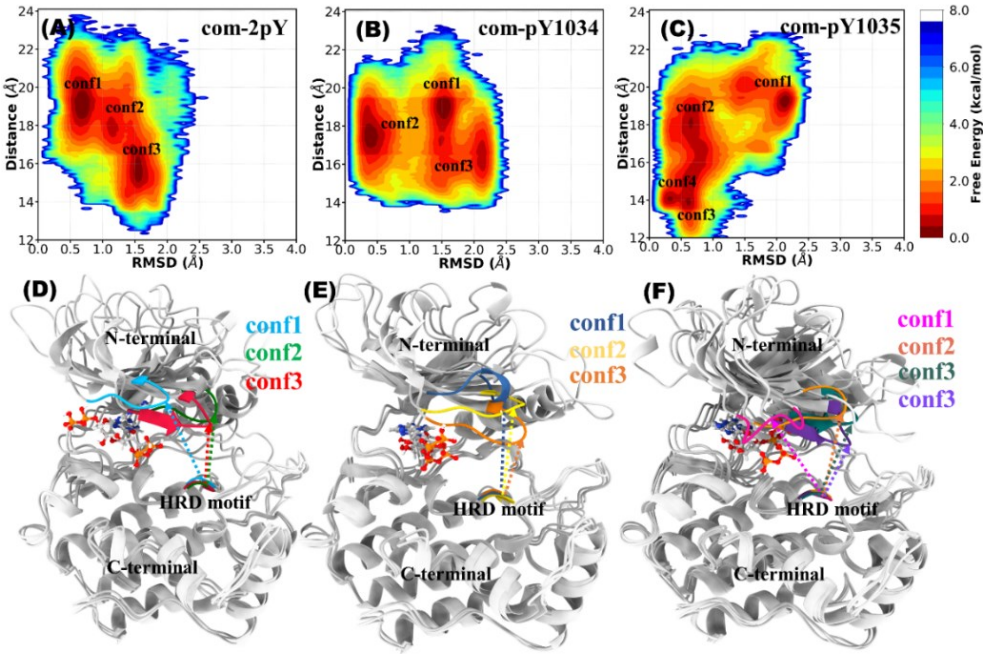


Figure 4.6: The two-dimensional free energy landscape of three phosphorylated states JAK1/SOCS1 complexes regarding RMSD of P-loop and center of mass distance between conserved HRD motif and P-loop region. The different conformations and their corresponding representative structures are shown in (**A, D**) com-2pY; (**B, E**) com-pY1034, and (**C, F**) com-pY1035, respectively.

Subsequently, we compared the residual flexibility of SOCS1 between apo and bound states (see **Appendix B, Figure B5(A-C)**). It is evident from **Figure B5** that relatively more significant flexibility is observed in the bound state compared to apo. However, the KIR, ESS sub-domain, and BC-loop (see **Figure B5D**) remain relatively stable and agree with the experimental study[309]. The flexibility of the KIR domain (His^{54} to Arg^{59}) is significantly reduced due to the complexation (See **Figure B5E**), which agrees with the earlier study by Wei *et al.*[310]. Overall, the binding of SOCS1 with JAK1 yields stable interactions near the binding site.

Next, we computed the dynamic cross-correlation map (DCCM) of these three complex systems, representing protein's correlated and anti-correlated motions (see **Appendix B, Figure B7**). It is evident from **Figure B7** that com-2pY shows high positive correlations involved in the diagonal part and their neighboring groups, which move together, and the off-diagonal part displays negatively correlated motions. Moreover, for each case, the N-terminal and C-terminal of JAK1 are negatively correlated, as shown in the R1 region

(**Figure B7),** along with little non-correlated motions in the com-pY1034 system. The R2 region displays a highly correlated motion, corresponding to the C-terminal of JAK1 and the A-loop domain. However, double phosphorylation (com-2pY) increases anti-correlated motions in SOCS1. In contrast, the single phosphorylation (com-pY1034 and com-pY1035) reduces both correlated and anti-correlated motions in SOCS1 (see region R3); rather, it generates random motions in most of the sub-domain of SOCS1. Overall, the DCCM analysis suggests that the double phosphorylation induces strong positively and negatively correlated motions in JAK1 and SOCS1, respectively. However, a single phosphorylation state reduces the correlated (or anti-correlated) motion in the JAK1/SOCS1 complex. A similar decrease in anti-correlated motions for a single phosphorylation state was reported for WNK[277] and JNK3 kinases[62].

4.3.5. Conformational variability of A-loop

Next, we extensively explored the conformational dynamics of the A-loop, located at the C-terminal, which is the primary phosphorylation (Tyr^{1034} and Tyr^{1035}) sites. It is well established that A-loop plays the most critical role in regulating conformational changes in many kinases[57]. Herein, we explored the conformational dynamics of the A-loop with the help of two different types of collective variables; dihedral (phi-psi) angles based principal components (dPC1 and dPC2), to elucidate the effect of phosphorylation states and global collective variables (RMSD of A-loop and the center of mass distance between A-loop and SOCS1's KIR segment) to explore the impact of SOCS1 binding on A-loop with the help of the two-dimensional free energy landscapes, shown in **Figure 4.7(A-F)**.

It is evident from **Figure 4.7A** that the dually phosphorylated A-loop exhibits three central minima energy basins showing specific conformations, of which, highly probable conf1 is one of the inactive conformations due to the closure behavior of the A-loop[333] and the other two minima correspond to the active conformation (see **Figure 4.7A**, the right panel). Moreover, in the P+1 loop, these conformations are extended, intermediate, and closed for conf1, conf2, and conf3, respectively. Surprisingly, from **Figure 4.7B,** we find that due to the Tyr^{1034} residue phosphorylation, dPCA suggests a single global conformation for A-loop, which means $pTyr^{1034}$ stabilizes the A-loop structure, as is evident from our RMSD analysis (**Figure 4.3**). Similarly, the single phosphorylation of Ser^{382}, $Thr^{221,}$ and Thr^{288}

residues in WNK1[277,334], JNK[60,62], and Aurora-A[68] stabilizes the flexible A-loop. Similarly, **Figure 4.7C** suggests four low-energy basins corresponding to four different conformations, in which conf1 is active and well-matched with its native X-ray crystal. In contrast, conf2 corresponds to an extended active conformation, while conf3 is an intermediate structure. Remarkably, conf4 is stable and extended toward C-lobe-like inactive conformation[333] with a high free energy barrier (~ 4.0 kcal/mol) with the native crystal structure.

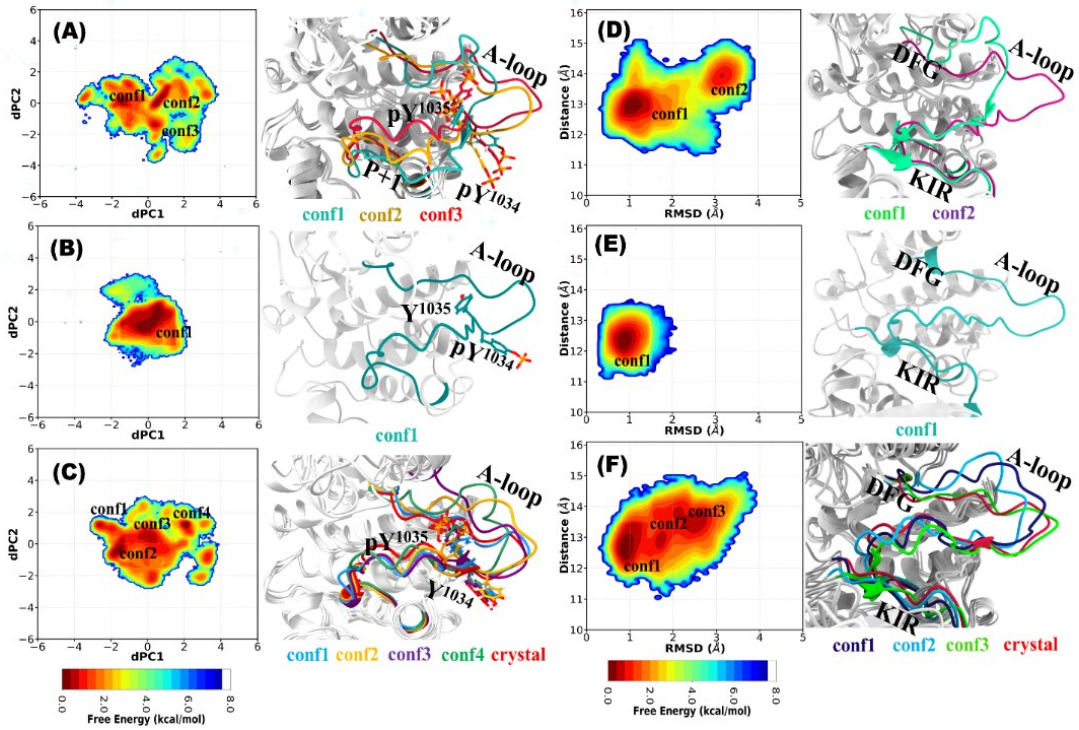


Figure 4.7: The 2D free energy landscape of com-2pY, com-pY1034 and com-pY1035 system is shown in panel **(A-C)** obtained from the A-loop dihedral PCA and **(D-F)** from the reaction coordinates of A-loop backbone RMSD and C.o.M distance between A-loop and KIR region, respectively. The local minima conformations were highlighted, and their representative structures were shown on their right panel. The crystal structure of pTyr[1035] conformation is also shown in red color.

Further, we explored the SOCS1 complexation effect at A-loop conformation. To know the structural intermediates, we generated 2D free energy potential surface of the A-loop, shown in **Figure 4.7D** to **F**. From **Figure 4.7D,** it is clear that the potential surface is characterized by two distinct free energy minima (conf1 and conf2) and a transition between an active (conf2) to inactive conformation (conf1) of A-loop. This is the same

inactive conformation, as discussed above (dPCA analysis), and the energy gap between these structures is around ~ 4.0 kcal/mol. In com-pY1034, the global free energy minimum is restricted with a single conformation (see **Figure 4.7E)**. Similarly, **Figure 4.7F** depicts that the com-pY1035 complex exhibits three minimum energy basins with different specific conformations and a conformational transition, active (conf3) conformation to inactive (conf1) via an intermediate (conf2) structure. In agreement with the dPCA results, this inactive (conf1) is very similar to the dPCA inactive conformation (conf4). The above results suggest that the distributions of the conformational intermediate of JAK1 for double and single phosphotyrosine are remarkably distinct. As indicated by Shukla *et al*., these intermediate conformations might be the best targets for specific drug design[335].

4.3.6. Characterizations of global and local concerted motion in JAK1/SOCS1 complexes

We characterized the JAK1/SOCS1 complex and detected the conformational variations upon phosphorylation and complex formation with SOCS1. Herein, we performed PCA of combined MD trajectories of each complex to analyze the simulations' conformational space sampled. The PCA results of three phosphorylated systems in terms of principal components (PC1 and PC2) and eigenvalues after the diagonalization of the covariance matrix are shown in **Appendix B, Figure B8.** The corresponding porcupine plots of PC1 and PC2 modes and percentage of each eigenvector variance with cumulative sums were also plotted. The first two eigenvectors (collective motions of the local fluctuations) suggest that the N-terminal of JAK1 shows more significant motion intensity of atomic fluctuations and agrees with the RMSF and DCCM analyses. Comparing the porcupine plots of three systems indicates that the com-2pY and com-pY1035 systems have a higher magnitude of the eigenvalues than com-pY1034, see **Figure B8**. The first 15 eigenvectors describe the collective motions fully, and the first 10 eigenvectors account for ~88.4%, ~80.1%, and ~87.1% overall motion in com-2pY, com-pY1034, and com-pY1035, respectively. The subsequent eigenvectors quickly decrease their proportion of variance of motions. Similarly, it is clear from the proportion of variance that the first two PCs (eigenvectors) capture ~68.8%, 48.7%, and 59.4% of the total motions in com-2pY, com-pY1034, and com-pY1035, respectively. In addition, the motion intensity of the P-loop (green color, **Figure B8**) has outward tendencies of movement and agrees with the RMSF

results. Remarkably, hinge, catalytic, and αC-helix have weak motion intensity due to the ADP binding and a strong salt bridge between β3 (Lys^{908}) and αC-helix (Glu^{925}). These results reasonably explain why the com-pY1034 system has higher correlated local motion and agree with our previous cross-correlated motion analysis results.

Furthermore, we considered the first two PCs (PC1 and PC2) as reaction coordinates to construct the free energy landscape (FEL) (**Figure B9(A**, **B**, **C**)) to identify the global and local low energy conformations of these complexes and the time evolution of these two PCs shown in **Figure B9 (D**, **E**, **F)**. The pattern of low energy minimum on the FEL and PCs atomic fluctuations are different in the three systems, as is evident from **Figure B9**. In the case of com-2pY, the conformational space shows a widely dispersed into one global minimum (conf1) and a local low energy conformation (conf2) (**Figure B9A, D**). In contrast, the single Tyr^{1034} phosphorylation shows a single broad, stable global minimum confined within a particular basin (**Figure B9B, E**). However, the single phosphorylation at Tyr^{1035} completely breaks the original conformational space and induces three low-energy basins, which makes com-pY1035 span into three conformational subspaces. Moreover, conf1 and conf2 are equally probable conformations with a narrow energy barrier of ~0.8 kcal/mol, which is room temperature accessible, and conf3 has an energy barrier of more than 1.5 kcal/mol. These results and our previous study[277] suggest that phosphorylation exerts a significant effect on JAK kinases and induces conformational diversity of JAK1/SOCS1 complexes. It is also telling that for the first single Tyr^{1034} phosphorylation, the JAK1 structure gets stabilized, which agrees with the previous RMSD and RMSF analysis.

4.3.7. Structural protein network and community

In the protein structure network (PSN) theory, amino acids represent nodes, and edges are represented as interactions among the nodes; it could be short-range or long-range interactions. Hubs are mainly defined as highest degree nodes, and the network community represents the more connected region of nodes. To get more detailed insight into the protein stability or protein function, we further performed the PSN analysis in terms of network hubs and community map, shown in **Figure 4.8(A-C)**. Previously, various studies have

reported that the protein-structure function relies on the interaction between residues (amino acids)[336–338].

Network properties such as the number of nodes, links, hub, and community were analyzed for the three phosphorylated systems and reported in **Appendix B, Table B2**. It is evident from **Figure 4.8(A-C)** and **Table B2** that the overall total number of hubs and communities are significantly different in the three systems. The double phosphorylated com-2pY system possesses 37 hubs, whereas the single phosphorylated com-pY1034 and com-pY1035 consist of 54 and 47 hubs, respectively. As expected, the total number of links and hubs in com-pY1034 (391,54) is significantly higher than com-2pY and com-pY1035, suggesting relatively higher structural stability and rigidity of com-pY1034.

Further, comparing the PSN community map of three systems, we concluded that com-pY1034 shows the highest probability of the highest-ranked community (C1, red), indicating a compact system of the systems due to Tyr^{1034} residue phosphorylation than Tyr^{1035} or double (Tyr^{1034} and Tyr^{1035}). The highest-ranked network community, C1, primarily stabilizes the C-terminal of each system (see **Figure 4.8**, right panel), although the links and hubs involved are higher for the case of the com-pY1034 system (see **Table B2**). Due to higher values of the C1 community, the residual collective motions are highly correlated and show lower residual fluctuations and agree with previous DCCM and RMSF results. These comparative PSN results provide insight into the conformational dynamics that could guide JAK1-specific novel drug developments.

4.3.8. P-loop, A-loop, and KIR region interaction mechanisms

To structurally characterize the effect of phosphorylation on the mechanisms of SOSC1 inhibition of JAK1 activity, we traced one important hydrophobic interaction among His^{885} (P-loop), Pro^{1044} (A-loop), and the first residue, His^{54} of KIR sub-domains. Pro^{1044} is highly conserved across all tyrosine kinases and plays a crucial role in substrate selectivity[339], and provides the platform upon which a substrate tyrosine sits during catalysis. Indeed, His^{54} of KIR occupies the substrate-binding pocket[339] and blocks the peptide substrate access of JAK1 by mimicking a substrate tyrosine.

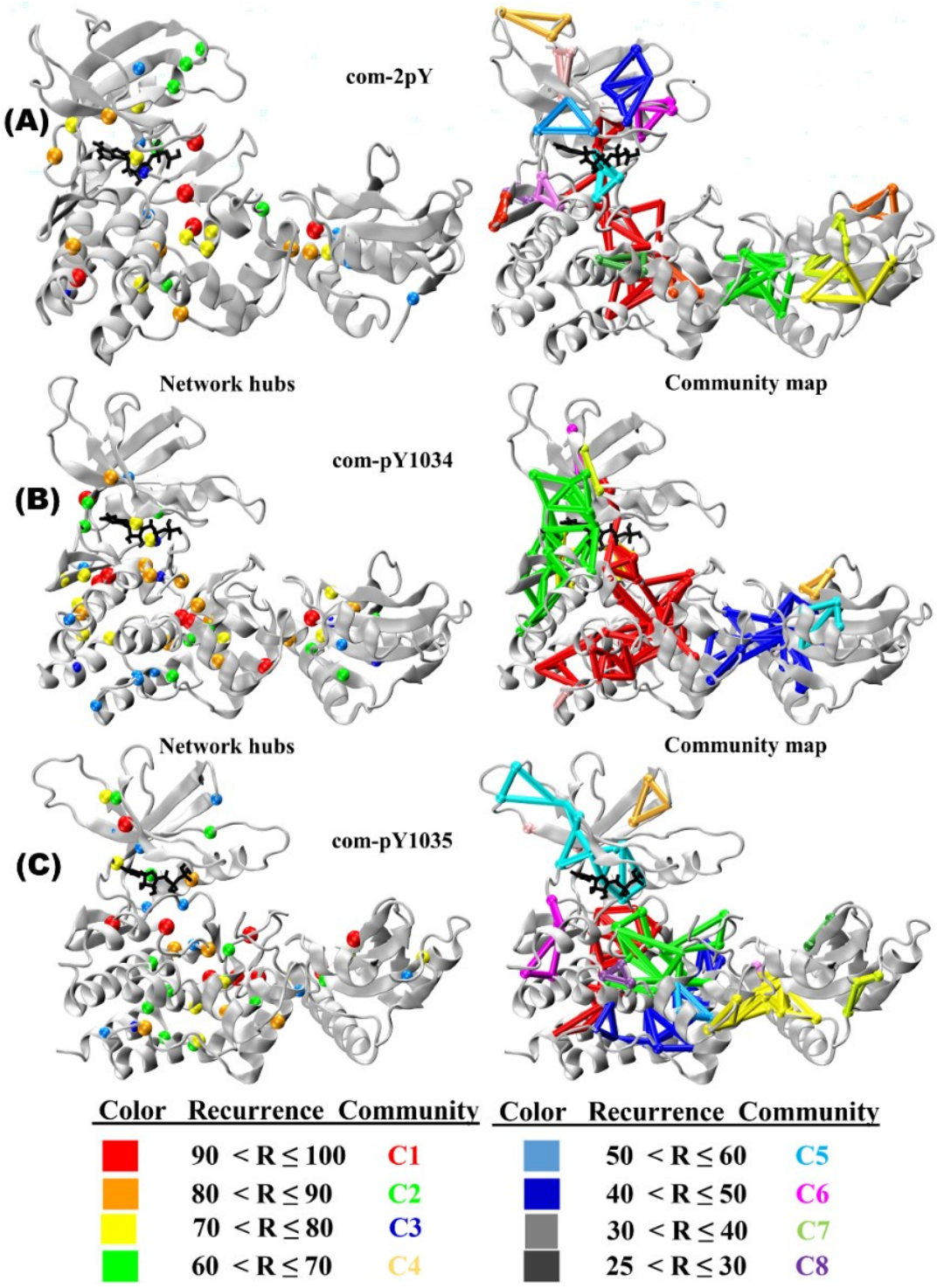


Figure 4.8: The network hubs and community maps of residue contact networks in (**A**) com-2pY, (**B**) com-pY1034, and (**C**) com-pY1035, respectively. The hubs are represented as spheres centered on the Cα atoms on the Jak1 protein. Community structure describes a fundamental feature of residue contact networks. Coloured according to the community nodes, the largest community-ranked 1 is red.

According to Liau et al.[309], His^{54} is sandwiched between the two planner sidechains from JAK1, the upper sidechain is from P-loop (His^{885}), and the lower sidechain is from A-loop (Pro^{1044}), as shown in **Figure 4.9A**. First, we explored their structural variations based on 2D free energy landscapes with the help of pseudo-dihedral (θ1 and θ2) angles (see **Appendix B Figure B10).** It is evident from **Figure B10** that His^{885} and His^{54} exhibit a large number of minima showing different specific structures. In contrast, Pro^{1044} shows highly structural conservation with a single global minimum and agrees with the experimental studies[339]. Moreover, His^{885} (P-loop) is more dynamic than His^{54} (KIR), previously confirmed by our P-loop conformation dynamics and RMSF results.

Next, we calculated the angle among His^{885}-Pro^{1044}-$His^{54,}$ considering Pro^{1044} as a reference point, and the distribution of the corresponding angles for all three complexes are shown in **Figure 4.9B**. As suggested by **Figure 4.9B**, com-2pY shows a bimodal distribution with equiprobable peaks at 110º and 133º, whereas a single peak at 145º characterizes the distribution for com-pY1034. However, in the case of com-pY1035, the most probable peak is located at 148º, with two additional peaks at 85º and 112º. The PCA analysis supports this observation. This stable hydrophobic interaction in com-pY1034 could be the reason for the stabilized A-loop (see **Figure 4.7 B** & **E**).

4.3.9. Phosphorylated induced interaction pattern of JAK1 and SOCS1

To provide further insights into the mechanism of recognition of SOCS1 by JAK1 (three phosphorylation states), we carried out a MM/PBSA-based free energy calculation. A summary of all individual energy contributions is shown in **Table 4.2**. It is evident from **Table 4.2** that the total binding free energies (enthalpy) have values ranging from -98.3 and -102.8 kcal/mol for three phosphorylated complexes. The single phosphorylation (com-pY1034) displays a moderately higher binding affinity (-102.8 ± 1.8 kcal/mol) than the second single (com-pY1035) and double phosphorylation states of A-loop, in good agreement with our previous RMSD and free energy surface based structural dynamics analyses. The total nonpolar solvation energy (ΔG_{np}) varies from -12.4 to -13.1 kcal/mol. As found in the previous study[310], the binding favorable intermolecular electrostatic interaction energy is overcompensated by the desolvation of polar groups. The sum of the

desolvation polar groups and intermolecular electrostatic interactions ($\Delta G_{\text{pol,elec}}$) is opposed to the complex formation, and the energy ranges between 51.9 to 54.6 kcal/mol. We observed that the binding is mainly favored by the van der Waals interactions, as we have already shown in our previous protein-protein interactions study[275].

4.3.10. Buried interface surface area: double versus mono phosphorylation states

As we see from our above structural analyses, proteins experience a wide variety of conformational dynamics during the complex formation at different phosphorylation states, which can be crucial for facilitating their diverse functions. A previous study reported that a protein-protein interface's buried and accessible surface controlled its intrinsic flexibility and interaction patterns[340]. Therefore, we measured the buried surface area of the interaction interface, shown in **Appendix B, Figure B11(A, B, C)**. It is evident from **Figure B11A** that the single phosphorylation at pY^{1034} and pY^{1035} yields a larger buried surface area than the double phosphorylation state. **Figure B11B** suggests that com-pY1034 shows a more stable buried surface than the other two phosphorylation states. The majority of the buried surface area of the JAK1/SOCS1 interface is composed of A-loop and αG-helix (GQM motif) and is in agreement with the experimental study[302,309]. Overall, the buried surface area suggests that due to the low solvent-accessible surface of the single phosphorylation (pTyr^{1034} or pTyr^{1035}), the unfavorable desolvation of polar groups' energy contributions also decreases.

Table 4.2: Energetic components of the total binding free energy of SOCS1 bound to the JAK1. All values are given in kcal/mol. The standard error of the mean is provided in parentheses.

component	com-2pY	com-pY1034	com-pY1035
$\Delta \boldsymbol{E_{\text{vdW}}}$	-142.0 (0.2)	-141.5 (0.2)	-140.2 (0.2)
$\Delta \boldsymbol{E_{\text{elec}}}$	-146.4 (1.4)	-17.2 (1.3)	-40.6 (1.5)
$\Delta \boldsymbol{G_{\text{pol}}}$	200.1 (1.3)	69.1 (1.2)	95.2 (1.4)
$\Delta \boldsymbol{G_{\text{np}}}$	-12.4 (0.0)	-13.1 (0.0)	-12.7 (0.0)
$\Delta \boldsymbol{G_{\text{MM}}}$	-288.4 (1.4)	-158.7 (1.3)	-180.8 (1.5)
$\Delta \boldsymbol{G_{\text{solv}}}$	187.7 (1.3)	56.0 (1.2)	82.5 (1.2)
$\Delta \boldsymbol{G_{\text{pol,elec}}}$	53.7 (1.9)	51.9 (1.8)	54.6 (2.0)
$\Delta \boldsymbol{G_{\text{bind}}}$	-100.7 (1.9)	-102.8 (1.8)	-98.3 (1.9)

$\Delta G_{\text{MM}} = \Delta E_{\text{vdW}} + \Delta E_{\text{elec}}$; $\Delta G_{\text{solv}} = \Delta G_{\text{pol}} + \Delta G_{\text{np}}$; $\Delta G_{\text{pol,elec}} = \Delta G_{\text{pol}} + \Delta E_{\text{elec}}$

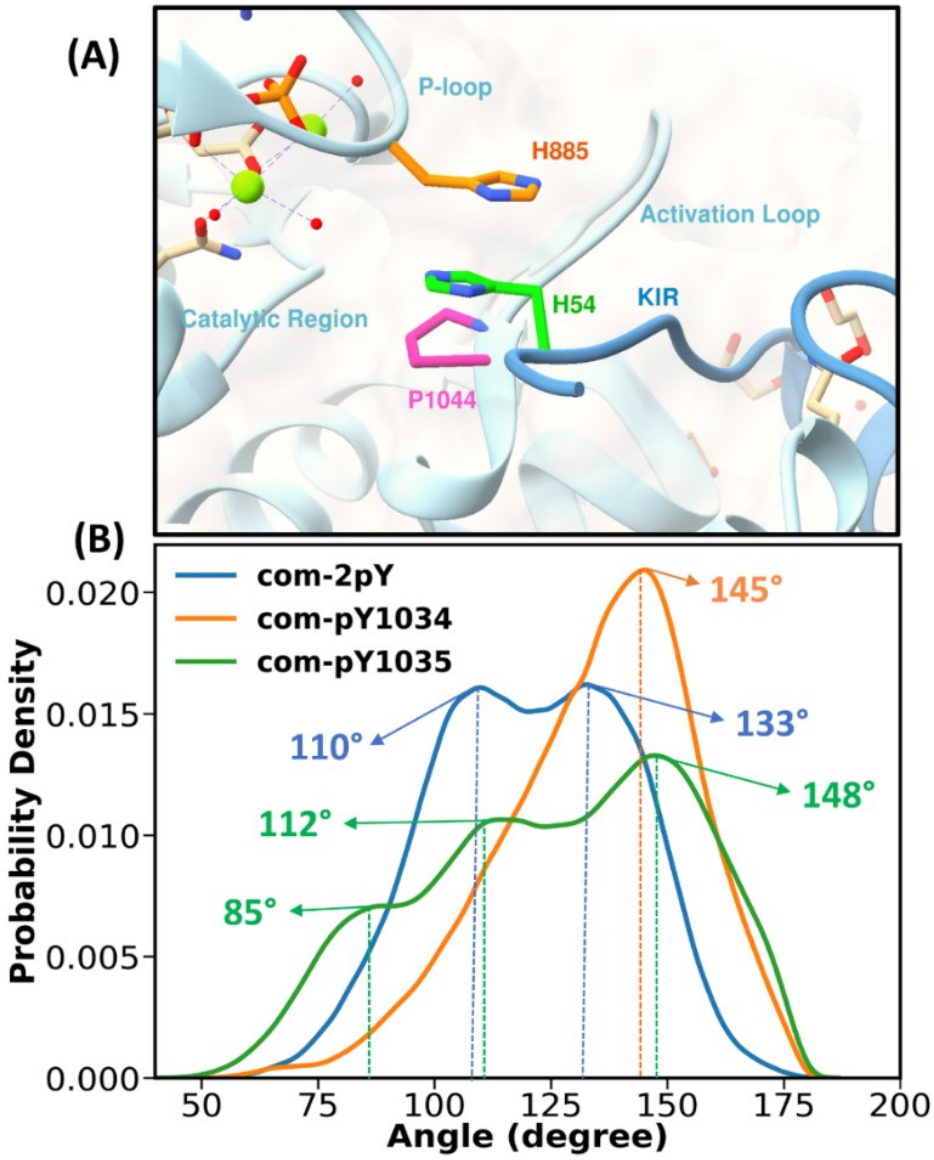


Figure 4.9: The sandwiched-like structure and **(B)** the probability density function of P-loop (His^{885}), A-loop (Pro^{1044}), and KIR (His^{54}) angle for three phosphorylated JAK1/SOCS1 complexes.

4.3.11. Identification of Key interacting residues

To identify residues from JAK1 and SOCS1 involved in the JAK1/SOCS1 complexation, the total binding free energy was further decomposed at the individual residue level by utilizing the MM/GBSA method. The resultant protein-protein interaction spectra are shown in **Figure 4.10(A-C)**, while the location of the residues is shown in **Figure 4.10(D-F).** It is evident from **Figure 4.10** that the interaction spectra of all three complexes are more or less similar. The experimental study suggested that the JAK1/SOCS1 interactions

are mainly observed at the KIR and ESS subdomain of SOCS1[309], in good agreement with our results. Moreover, we identified some hotspot residues from SOCS1 (Gly[1097], Gln[1098], Met[1099], Arg[1103]) and JAK1 (Asp[1040] to Ser[1056] (A-loop and P+1 loop)). As shown in **Figure 4.10**, residues from His[54] to Arg[59] of the KIR region primarily interact with the A-loop and P+1 loop of JAK1. Further, residues from Asp[105] to Phe[112] of the BC-loop of SOCS1 also interact with the GQM motif and αG-helix of JAK1. We also explored the energetic contribution of the A-loop alone. In the case of pY1034, the favorable contribution of the A-loop is higher than the other two phosphorylation states, which agrees with the experimental study[309]. We also observed that Gln[1098] (-7.1 kcal/mol), Arg[56] (-7.0 kcal/mol) and Phe55 (-7.1 kcal/mol) in com-2pY, Gln[1098] (-7.7 kcal/mol), Arg[1103] (-7.8 kcal/mol) and Phe[55] (-7.0 kcal/mol) in com-pY1034 and Gln[1098] (-7.5 kcal/mol) and Phe[55] (-6.9 kcal/mol) in com-pY1035 are most attractive contributing residues. Moreover, we also notice that some specific polar residues like Asp[1040] and Asp[1042] of JAK1 and Arg[56], Arg[59], Asp[105,] and Lys[109] of KIR and BC-loop of SOCS1 play crucial roles in the interactions via hydrogen bonds or salt bridge networks (as demonstrated in **Figure 4.10(A-C)** and **Appendix B, Figure B12)** and are consistent with the previous study by Wei *et al.*[310].

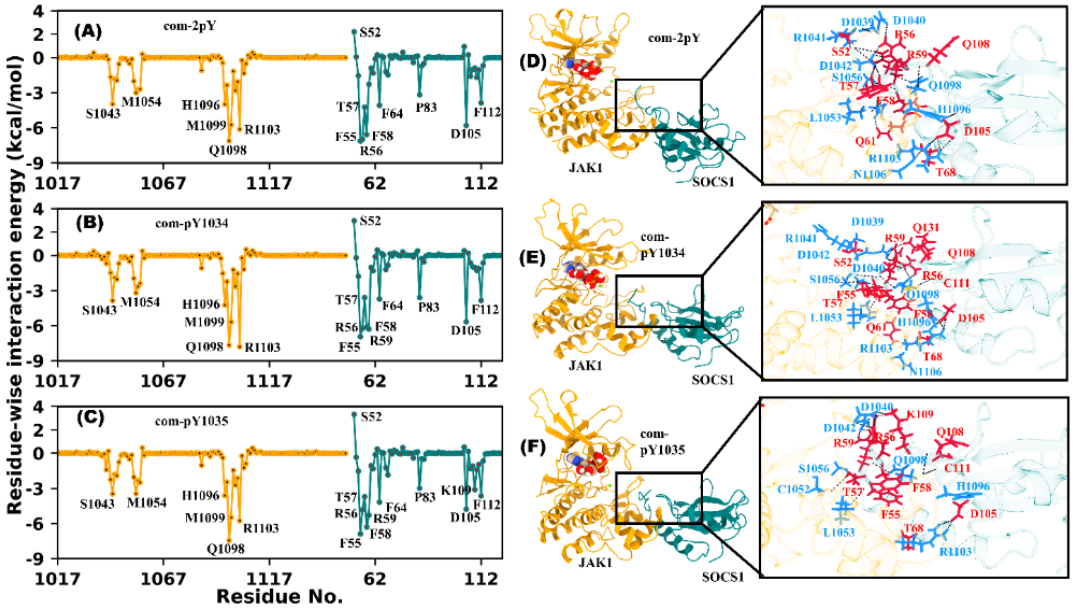


Figure 4.10: (**A-C**) The free energy contribution of each residue to the JAK1/SOCS1 binding. (**D-F**) Details of the interactions between JAK1 and SOCS1 in the final production simulation. Black dashed lines indicate hydrogen bonds.

Finally, we explored the hydrogen bond (HB) and salt bridge formations between JAK1 and SOCS1 (see **Appendix B, Table B3** and **Figure 4.10**). As indicated in **Table B3,** residues of SOCS1 strongly bond with JAK1, although their probabilities of formation of HB are different due to JAK1 phosphorylation states. The KIR's residue Phe^{58} forms a strong hydrogen bond with the hydroxyl group of Ser^{1056} with an occupancy of more than 90% for all cases, which is consistent with the experimental study[309]. In addition, Thr^{57}, Phe^{55}, Arg^{56}, Arg^{59}, Asp^{105}, $Lys^{109,}$ and Cys^{111} residues from the KIR and BC region also form hydrogen bonds with the A-loop and GQM motif of JAK1 (see **Table B3).** These hydrogen bonds play an essential role in stabilizing SOCS1. The distance distributions of all these KIR residues with JAK1 are shown in **Appendix B, Figure B12**, **B14**, and **B15**.

4.3.12. Hydrogen bonds within JAK1, salt-bridges, and secondary structure propensity of A-loop

Additionally, we calculated all hydrogen bonds within the residues of JAK1 and SOCS1. The time evolution of these total hydrogen bonds is shown in **Appendix B, Figure B16**. Hydrogen bonds analysis suggested that the pY1034 contained more bonds within the residues. Moreover, we found some new salt bridges formed due to the phosphorylation of tyrosine (Tyr1034/1035) residues, shown in the right panel of **Figure B16**. Salt bridge analysis clearly suggests that pY1034 achieved extra stability due to the strong salt bridge between $pTyr^{1034}$ and Lys^{1032}.

Further, we have also performed the time-dependent secondary structure propensity (DSSP) analysis of the A-loop segment, see **Appendix B, Figure B17**. **Figure B17(A-C)** clearly shows that the secondary structures are dynamic in the case of doubly phosphorylated tyrosine (2pY) and second phosphorylated tyrosine (pY1035). In the case of the first phosphorylated tyrosine, pY1034 is moderately high stable. Similarly, the compactness analysis also agrees with the above DSSP analysis, see **Figure B17(D). Figure B17(D)** suggested that the A-loop of pY1034 is highly compact and less flexible.

4.4. Conclusion

In this study, all-atom MD simulations of cumulative length 13 μs were conducted to investigate the effect of varying phosphorylation (single and double) on the structural dynamics of JAK1 and its binding with SOCS1. Our simulation results suggest that the

phosphorylation at Tyr^{1034} stabilizes the flexible activation segment and confers conformational rigidity by forming stable hydrogen bonds between JAK1 and SOCS1. The potential energy surface of the P-loop shows conformational variability, leading to various inactive conformations that can be targeted to design a better specific inhibitor. The structural protein network analysis suggests that the com-pY1034 system is stabilized due to higher values of network hubs than the other two complex systems, which is further confirmed by the RMSF and PCA results. The MM/PBSA binding free energy calculations predicted that the single pTyr1034 phosphorylation state shows a slightly higher affinity towards SOCS1 than the other two phosphorylation states. The polar residues, including Arg^{56}, $Arg^{59,}$ and Asp^{105} from SOCS1 and Asp^{1040} and Asp^{1042} from JAK1, contribute significantly to the binding. Moreover, KIR of SOCS1 and the GQM motif, a fragment of BC-loop, and Ser^{1043}, Val^{1045}, M^{1054}, Ser^{1056}, and His^{1096} are identified as the most hotspot residues. Our results provide a better understanding of how phosphorylation states of A-loop tyrosine residues rearrange the conformational dynamics of JAK1 and the binding mechanisms of SOCS1. These findings help us understand the structure-to-function relationship of protein and interaction mechanisms between JAK1 and SOCS1. We hope that these results will aid the development of new peptide inhibitors with higher affinity and specificity.

Chapter 5

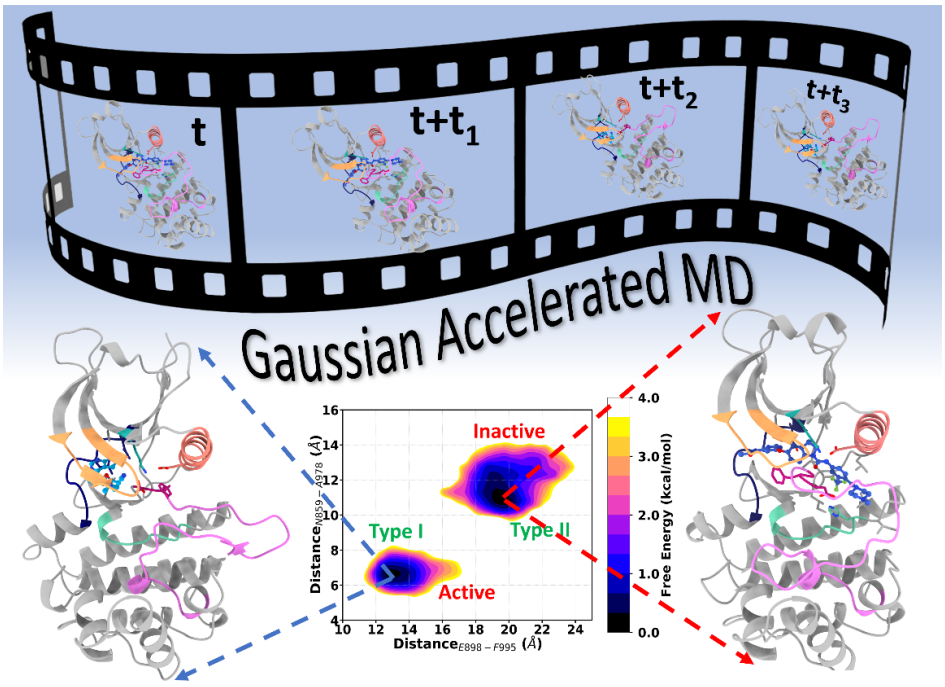
Chapter 5

Exploration of conformational dynamics of phosphorylated active and inactive Janus kinase 2 using Gaussian accelerated molecular dynamics simulations

AIM: To elucidate the conformational dynamics of phosphorylated active and inactive JAK2 induced by the different type of inhibitors and their molecular basis of inhibition.

This chapter is reproduced from our original article:

Sk, M.F. & Kar, P. (2022). "Exploration of conformational dynamics of phosphorylated active and inactive Janus kinase 2 using the Gaussian accelerated molecular dynamics simulations". *J. Chem. Inf. Model. (In Revision)*



5.1. Introduction

The Janus kinase 2 (JAK2)[341] is a cytoplasmic non-receptor[301] tyrosine kinase[289] and a member of the JAK family. JAK family[33] has three other members, JAK1, JAK3, and TYK2[291,293]; these kinases play an essential role in the growth, development, and survival of different types of cells and are master regulators of the JAK/STAT (signal transducers and activators of transcription) signaling pathway[75,342,343]. The JAK/STAT pathway is the primary signaling pathway in the immune system[15,344] and is involved in various biological cell developments[36]. Cytokines binding to its receptor activates JAK2, then activates STAT protein, promoting dimerization and translocating to the nucleus to turn on the gene expression[345]. Therefore, the dysregulation of the JAK/STAT cascade has implicated many pathological conditions, including autoimmune disorders and cancers[291–293].

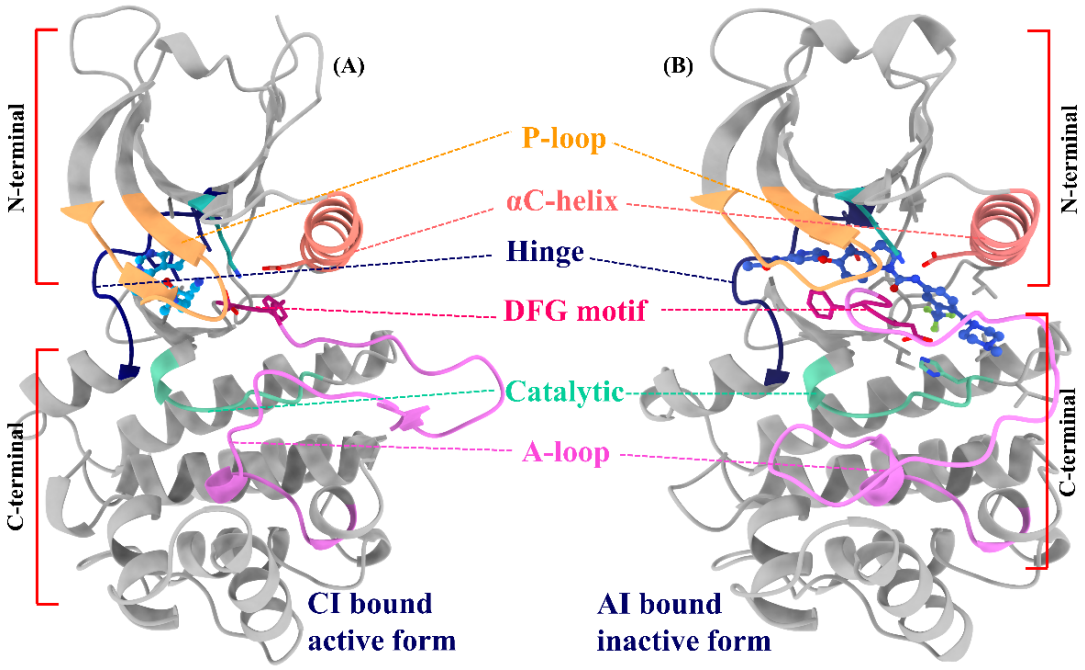


Figure 5.1: The three-dimensional structures (ribbon representation) of (**A**) CP-690550 (ATP-competitive inhibitor, CI; PDB Id; 3FUP) bound active and (**B**) BBT594 (allosteric inhibitor, AI; PDB Id: 3UGC) bound inactive JAK2. Locations of the hinge, P-loop, αC-helix, DFG-motif, catalytic loop, and A-loop are indicated by color-coded.

Like all tyrosine kinases, the kinase domain (KD) in JAK is formed by N- and C-terminals, which jointly include a deep cleft in which the ATP (Adenosine triphosphate) binds[31] (see

Figure 5.1). Then phosphotransferase reaction occurs; the ATP transfers the γ-phosphate onto the substrate proteins' tyrosine residue, which regulates the JAK/STAT signaling pathway. The small N-lobe is dominated by a five-stranded anti-parallel β-sheet (β1-β5)[52]. It also contains a vital regulatory αC-helix and ATP-phosphate binding loop (P-loop). The β3-strand typically has catalytic residue Lys^{882} (JAK2), forming a stable salt bridge with αC-helix Glu^{898} (JAK2), which corresponds to an "αC-*in*" conformation[31], otherwise "αC-*out*" conformation. The large lobe of JAK protein is dominated by mainly α-helical with six conserved segments (αD-αI)[49,52] and a vital activation loop (A-loop). Notably, the catalytic process is highly dependent on the conformation of the flexible A-loop. In JAK2, this is a sequence of 30 residues bounded by the conserved Asp^{994}-Phe^{995}-Gly^{996} (DFG-motif) and Ala^{1022}-Pro^{1023}-Glu^{1024} (APE-motif)[31]. ATP and other substrates can only bind and undergo the phosphotransferase reaction when the A-loop is fully open. Several computational and experimental studies have reported that phosphorylation at Tyr^{1007} stabilizes the A-loop[70,346] in a fully open conformation, promoting ATP binding and thus full kinase activation.

JAK2 kinase exists in an equilibrium of active and inactive conformations, in which the conformational adjustment induced by tyrosine phosphorylation of the A-loop occurs in the DFG motif[347]. Moreover, JAK2 kinase activity also depends on the conformation adopted by the DFG motif[348]. Only the "DFG-*in*" conformation in which the Asp^{994} coordinates the catalytic magnesium (Mg^{2+}) ion and allows the phosphotransferase reaction[31]. In this state, the side chain of Phe^{995} faces away from the ATP binding site, allowing a proper formation of the regulatory spine (R-spine)[49,247], whose adequate alignment plays an essential role in the kinase activity. Conversely, in the "DFG-*out"* state, flipped Asp^{994} side chain causes the loss of catalytic kinase function, and the reorientation of the Phe^{995} side chain leads to the misalignment of the R-spines[70].

Based on the binding location of the inhibitors and protein intrinsic conformations, JAK2 inhibitors can be classified into two groups: type-I or conventional ATP-competitive inhibitors (CI) bind exclusively at the ATP-binding cleft of the active state (DFG-*in*)[349,350] (see **Figure 5.1A**), and type-II or allosteric inhibitors (AI), which binds on to the JAK2 on the allosteric site adjacent to the ATP-binding pocket and inactive form of JAK2 (DFG-

out)[351,352] (see **Figure 5.1B**). Due to the high degree of sequence identity and similarity at the ATP-binding pocket among the JAK isoform, the type-I inhibitors lack specificity toward individual JAK[70]. In contrast, it may be easier to achieve kinase selectivity for the type-II inhibitors attributed to an additional and hydrophobic site with less sequence conservation relative to the ATP-binding site. Therefore, type-II inhibitors could be the better choice for specific drug designs[352]. Many X-ray crystal structures of active JAK2 bound with type-I inhibitors are available in the PDB database[353]. Still, only one inactive conformation of JAK2 bound with a type-II inhibitor, BBT594[354]. However, the crystal structures describe the static structural and binding pose of protein–inhibitor complexes, and the dynamic nature of protein remains poorly defined. What does affect the A-loop spatial conformation is still not clear? What are the possible molecular mechanisms that modulate the inactive and active conformation of JAK2 that have not been reported so far? Molecular dynamics (MD) simulations based on atomic models may help answer these questions by providing energetics, kinetics, and conformational information that is not easily accessible from experiments.

In this work, we have provided structural insights into the conformational diversity of active and inactive states and the molecular basis of type-I and type-II inhibitors binding to the JAK2 kinase using extensive microsecond long Gaussian accelerated Molecular Dynamics simulation (GaMD)[179,355]. Moreover, this study aims to elucidate the intrinsic dynamics among the native JAK2 apo form and bound with type-I and type-II inhibitors using the popular free energy surface analyses. In addition, the protein motion of apo and complex is calculated using the principal component analysis (PCA)[268], and inhibitor binding energy spectra are investigated by using the molecular mechanics Poisson-Boltzmann surface area (MM/PBSA)[224,312,356] scheme and the conformational entropic cost is estimated by normal mode analysis[230]. We believed that the obtained information could provide unique possibilities for structure-based drug discovery.

5.2. Materials, method, and protocols

5.2.1. Preparation of apo and type I/II inhibitor bound forms of JAK2

The crystal structure of the JAK2 complexed with type-I inhibitor (DFG-*in*) and type-II (DFG-*out*) was retrieved from the protein databank with PBD entry 3FUP[33] (resolution 2.40 Å) and 3UGC[354] (resolution 1.34 Å), respectively. The initial type-II inhibitor-bound complex contained two mutations (Tyr^{1007}Phe and Tyr^{1008}Phe) in the A-loop, which were reverted from Phe to Tyr and phosphorylated these residues manually. Missing residues in both complexes were modeled using the Modeller[313] plugin in the UCSF chimera[251]. Subsequently, the structures' steric clashes were cleaned by minimizing the structures using the Schrödinger maestro[357] visualizer. The Propka3.1 module[254] determined the protonation state of all ionizable residues. The histidine (His) residue was at the Nε2-protonated (Hie) τ tautomer state, and the rest of all charged residues were deprotonated. The *LEaP* module[314] of AmberTools19[118] was used to add all the missing hydrogen atoms. The main objective of this study is to explore the intrinsic conformational dynamics of JAK2 upon binding to different inhibitors (type I/II) and their native active apo conformation. The apo JAK2 structure was obtained by removing the inhibitor of the complex.

The molecular dynamics of three systems were performed using the *pmemd.cuda* module of AMBER18[118] suite. The AMBER antechamber[358] module calculated the atomic charges (AM1-BCC) of inhibitors. The protein and small molecules ligand were defined by ff14SB[123] and general Amber force field (Gaff2)[133] force fields, respectively. The Phosaa10[260] force field characterized the phosphorylated (Tyr^{1007}/Tyr^{1008}) residues. The explicit TIP3P[147] force field was used to define the solvent water molecules, and an octahedron periodic box with a distance cutoff of 10 Å was used. An appropriate number of monovalent counterions were added to the solvated systems to maintain each system's physiological salt concentration and charge neutrality. Finally, all three systems (apo, CI, and AI) were considered for subsequent molecular dynamics simulations.

5.2.2. Conventional molecular dynamics (cMD) simulations

Each system was subjected to energy minimized using the steepest descent and conjugate gradient methods. Subsequently, all complexes were gradually heated from 0 to 300 K for 50 ps at the canonical (*NVT*) ensemble with a restraint force constant of 2.0 kcal mol^{-1} $Å^{-2}$ on the solute and the temperature was controlled by a Langevin thermostat[316]. All covalent bonds involving hydrogen atoms were constrained using the SHAKE algorithm[141]. The particle-mesh Ewald summation (PME)[162] was used to compute long-range interaction with a 10 Å nonbonded cutoff. After that, systems were equilibrated in an isothermal-isobaric (*NPT*) ensemble for 20.0 ns, and the pressure was controlled using a Berendsen barostat[155] at 1 atm. A 2.0 fs time step was used for all simulations, and the coordinates were recorded every 10.0 ps for further analysis. A 2.0 ns short production NVT simulation was performed without any constrain to collect the initial potential statistics.

5.2.3. Gaussian accelerated molecular dynamics (GaMD) simulations

The GaMD[179] is an enhanced sampling simulation technique developed to smooth the surface of potential energy with a harmonic boost potential and reduce the potential barrier. The advantages of GaMD[180,355] are no need for predefined reaction coordinates to run the simulations, in contrast to other enhanced sampling schemes. The boost potential follows the Gaussian distribution, which can be properly recovered via reweighting methods[182]. Recently, the method has been widely used to investigate the conformational dynamics of proteins[359] and nucleic acids[360], ligand binding[361,362], ion channels[363], agonists, and antagonist binding in the μ-OR[364,365], bacterial effector proteins[366], and the allosteric mechanism in CRISPR-Cas9[367,368]. The mathematical formulations of GaMD were discussed by Miao et al. [179]. and Wang et al. [180]. Here, we briefly discuss the essential background of GaMD simulation.

Consider a system with N atoms at position $\boldsymbol{r} = \{\boldsymbol{r}_1,..,\boldsymbol{r}_N\}$, when a system potential, $V(\boldsymbol{r})$, is lower than threshold energy, E, then a boost potential, $\Delta V(\boldsymbol{r})$, is added and can be expressed as,

$$\Delta V(\boldsymbol{r}) = \frac{1}{2} k(E - V(\boldsymbol{r}))^2; \qquad V(\boldsymbol{r}) < E \qquad [5.1]$$

Where k is the harmonic force constant. Then the modified potential is V*(***r***) and can be written as,

$$V*(\boldsymbol{r}) = V(\boldsymbol{r}) + \frac{1}{2}k(E - V(\boldsymbol{r}))^2; \qquad V(\boldsymbol{r}) < E \qquad [5.2]$$

Otherwise, when the system potential is above the threshold energy, i.e., $V(\boldsymbol{r}) \geq E$, the boost potential is set to zero, and $V^*(\boldsymbol{r}) = V(\boldsymbol{r})$. To smoothen the potential energy surface, three enhanced sampling principles were applied to boost potential in GaMD. First, consider two arbitrary potential values, $V_1(\boldsymbol{r})$ and $V_2(\boldsymbol{r})$, found on the original energy surface; if $V_1(\boldsymbol{r}) < V_2(\boldsymbol{r})$, then ΔV should be a monotonic function and does not change the modified bias potential energy, i.e., $V^*_1(\boldsymbol{r}) < V^*_2(\boldsymbol{r})$. Second, if $V_1(\boldsymbol{r}) < V_2(\boldsymbol{r})$, the potential difference observed on the smoothed energy surface should be smaller than that of the original, i.e., $V_2^*(\boldsymbol{r}) - V^*_1(\boldsymbol{r}) < V_2(\boldsymbol{r}) - V_1(\boldsymbol{r})$. Combining these two criteria, the threshold energy needs to be set in the following range:

$$V_{max} \leq E \leq V_{min} + \frac{1}{k} \qquad [5.3]$$

Where V_{max} and V_{min} are the system's maximum and minimum potential energies. To ensure the existence of equation (3), k must satisfy the following criterion,

$$k \leq \frac{1}{V_{max} - V_{min}}; \text{ let us define, } k \equiv k_0 \frac{1}{V_{max} - V_{min}} \text{ and } 0 \leq k_0 \leq 1 \qquad [5.4]$$

Third, the standard deviation of ΔV needs to be small enough (i.e., narrow distribution) to ensure appropriate energetic reweighting: $\sigma_{\Delta V} = k\,(E - V_{avg})\,\sigma_V \leq \sigma_0$, where V_{avg} and σ_V are the average, and standard deviation of the system potential energies, and $\sigma_{\Delta V}$ is the standard deviation of ΔV with σ_0 as a user-specified upper limit (e.g., $10k_BT$) for appropriate reweighting. When E is set to the lower bound, $E = V_{max}$, k_0 can be calculated as:

$$k_0 = min\,(1.0, \frac{\sigma 0}{\sigma V}.\frac{v_{max} - V_{min}}{V_{max} - V_{avg}}) \qquad [5.5]$$

On the other side, if E is set to the upper bound, $V_{min} + \frac{1}{k}$; k_0 can be calculated as:

$$k_0 = (1.0 - \frac{\sigma 0}{\sigma V}).(\frac{v_{max} - V_{min}}{V_{max} - V_{avg}}) \qquad [5.6]$$

GaMD simulation provided the total (ΔV_p), dihedral (ΔV_D), and dual potential (ΔV_D *and* ΔV_p) boost to accelerate the MD simulations. The dual-boost GaMD generally provides higher acceleration than the other two types of simulations. So, a dual potential boost was applied to our GaMD simulations, and the system threshold energy is set as $E = V_{max}$ for all in this study. The $V_{\max}$, $V_{\min}$, V_{avg}, and σ_{v} values of initial potential parameters were taken from the 2.0 ns cMD simulation with no boost potential. Afterward, a 50.0 ns GaMD simulation was performed, in which the potential boost was updated every 200 ps to reach equilibrium values. Finally, all the systems underwent 3 × 1 μs production simulations in the NVT ensemble, and coordinates were saved every 10.0 ps, resulting in 300,000 frames being generated for further analyses.

5.2.4. Simulation trajectories analysis

All the trajectories were evaluated and analyzed with the help of the Cpptraj[369] module of AmberTools19. All the comparatives stability (protein backbone root mean square deviation, RMSD; the radius of gyration, Rg; solvent accessible surface area, SASA, and intermolecular hydrogen bond occupancy) analyses were performed using the equilibration time (~ 500 ns) with respect to their energy minimized initial conformations. We also computed the structural flexibility using root-mean-square fluctuations (RMSF) and intrinsic residual motions by cross-correlation map (DCCM). We also highlighted the principal motions of each system by reducing the dimensional spaces of the data sets from the MD trajectories by PCA[268] analysis. The free energy surface (FES)[370,371] was generated by different reaction coordinates (RC) of the system, including RMSD, Rg, distances, and principal components (PC). The FES can be described by the following equation below;

$$\Delta G \; = \; - \, k_B T \; ln(\rho *) - \kappa - (\sum_{k=1}^{\infty} \frac{(\beta)^k}{k!} \; C_k - \text{const.}) \qquad [5.7]$$

Where, k_B is the Boltzmann constant, T is the absolute temperature of the system, and $\rho *$ is the modified probability density along with a selected RC. The parameter κ is defined in such a way that the global minimum is located at a 0 kcal/mol energy state. The parameter β is the Boltzmann factor, and C_k is the expansion coefficient. Thereafter, the potential boost combined with various RC calculated from PCA and MD trajectories was applied to recover the FES by suitable reweighting methods[182], as shown in the bracket in equation

(7). A toolkit of python scripts for GaMD reweighting "PyReweighting" is developed by McCammon group[182] and distributed free of charge at http://miao.compbio.ku.edu/PyReweighting/.

The binding free energy calculations were estimated by applying a computer-friendly reliable, and widely used MM/PBSA[220] scheme utilizing the *MMPBSA.py* script of the AMBER18 suite. The details mathematical formulations were already discussed in our earlier study[325,372–374]. We have extracted 15000 (3x5000) frames from the equilibrated MD trajectory for this calculation. Besides, the MM/GBSA[143,375] (generalized Born) was also applied for residue-wise interaction estimations.

5.3. Results and discussion

Here, we try to explore the dynamic nature of JAK2 structure and the effect of inhibitors (type I/II) binding on the structure and dynamics of the protein. To fulfill the objectives, we performed MD simulations on the kinase domains (KDs) of the JAK2 protein. Three independent simulations for each system were analyzed to reduce the sampling problem and avoid false-positive conclusions. Finally, we considered the last 500 ns trajectories from each simulation for subsequently used for analyses of MD trajectories.

5.3.1. Structural stability and fluctuation analysis of active and inactive JAK2

The stability of the simulation trajectories of each system was initially assessed using their backbone atoms RMSD and the time evolution of RMSD values is shown in **Appendix C Figure C1** and the total energy (and temperature) of these systems (data not shown here). The RMSD analysis of each system revealed that all systems are stabilized well after ~300 ns, except run1 of AI. In each scenario, the independent simulations also showed a similar pattern of their protein backbone atom RMSD. The change in AI and CI's atomic positional deviation pattern is different from their native apo form (see **Figure C1**). The kinase domain (KD) backbone atom's RMSD probability density (see **Figure 5.2A**) suggested that apo JAK2 showed a lower deviation than the complex. It is also suggested that the most probable RMSD values of apo, CI, and AI bound JAK2 is ~1.6, ~1.7, and ~2.1 Å, respectively. The AI-bound JAK2 (see Figure 2A) shows a higher RMSD value probable

conformation at ~2.4 Å. These RMSD trends are also supported by previous protein kinase B (PKB/Akt)[376], Abl kinase[377], and with-no-lysine (WNK)[359] kinase bound with allosteric inhibitors studies.

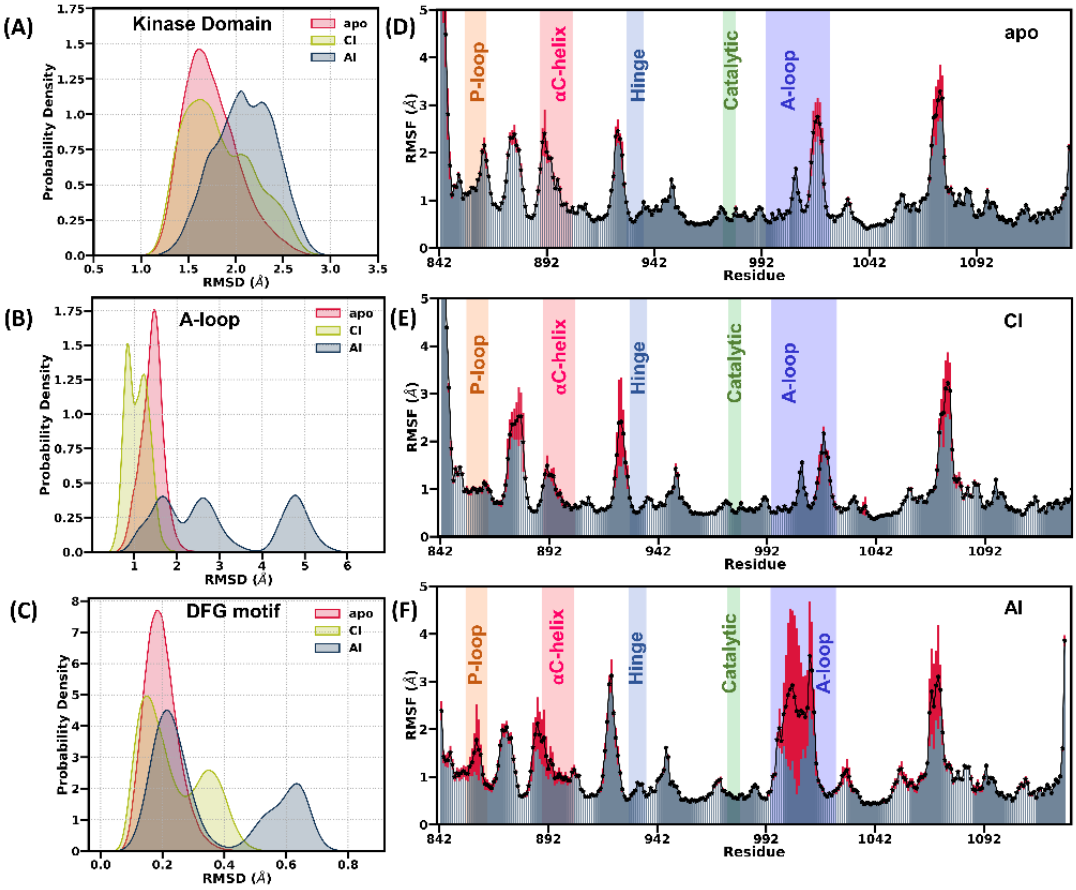


Figure 5.2: (A) The backbone atoms RMSD probability density of three systems; (B-C) the probability density distributions of the activation loop (A-loop) and DFG motif of JAK2 protein. (D-F) The backbone Cα-atoms average RMSF of three sets of MD trajectories (black) and their respective standard deviation (SD) of each fluctuation (red line).

Moreover, we also estimated the backbone atom's RMSD values of four particular mobile structural elements; the P-loop (see **Appendix C Figure C2A**), αC helix (**Figure C2B**), A-loop (**Figure 5.2B**), and DFG motif (**Figure 5.2C**). It is evident from **Figure C2A** that the P-loop of apo systems shows a single most probable RMS deviation at ~0.52 Å. After inhibitor binding, the P-loop shows two different deviation patterns, one most probable and the second comparatively less likely. The CI-bound most probable peak at ~0.44 Å and the

second higher deviated peak at ~0.63 Å (see **Figure C2A**). Similarly, AI-bound JAK2 contained a less probable, low deviation peak at ~0.42 Å and, most probable, high deviation peak at ~0.81 Å. These results suggested that the P-loop dynamics are regulated by the inhibitors[378] (type I/II) and play an essential role in the inhibitor selectivity[329,379]. The RMSD probability distribution of αC helix suggested that the apo and CI-bound JAK2 show similar RMS deviation from their energy minimized conformations (see **Figure C2B**). The AI-bound complex showed a higher variation and well-spreading RMSD than the apo and CI-bound complex. The most probable peaks are at ~0.38, ~0.37, and ~0.68 Å, respectively. These results suggested that αC helix conformational plasticity increases by allosteric inhibitor bindings, further supported by a previous kinase[359] study. Next, we checked the structural stability of the A-loop and its initial part (DFG motif) by monitoring their backbone atom's RMS deviation. The probability distribution of these RMSD is shown in **Figure 5.2(B, C).** As suggested in **Figure 5.2B**, the apo system showed a sharp peak at ~1.53 Å, indicating a well-stabilized conformation from the initial. In contrast to the apo results, inhibitors (type I/II) bound systems showed more than one peak. The CI-bound complex shows two peaks, one most probable (less deviate) at ~0.91 Å and little bit high deviate, shoulder peak at ~1.32 Å. Similarly, the AI-bound showed three equiprobable peaks ~1.71, ~2.74, and ~4.82 Å. These three peaks indicating that the three distinct conformations are possible in the systems and verified further by the free energy landscapes (FEL). In a similar fashion, the structural deviation of DFG motif also monitored by RMSD calculation, shown in Figure 2C. The RMSD probability distribution suggested that the inhibitor bound JAK2 shows some two different deviation peaks but apo constrain in single peak. Herein, we observe that CI-bound has a distinct most probable peak at ~0.18 Å with a shoulder peak at ~0.38 Å. Similarly, the AI-bound complex shows the two probable conformation and their RMSD peaks are at ~0.21 Å and ~0.62 Å. Overall, the RMSD analysis suggested that the change in the fluctuation pattern of JAK2 complexes was higher than the apo form.

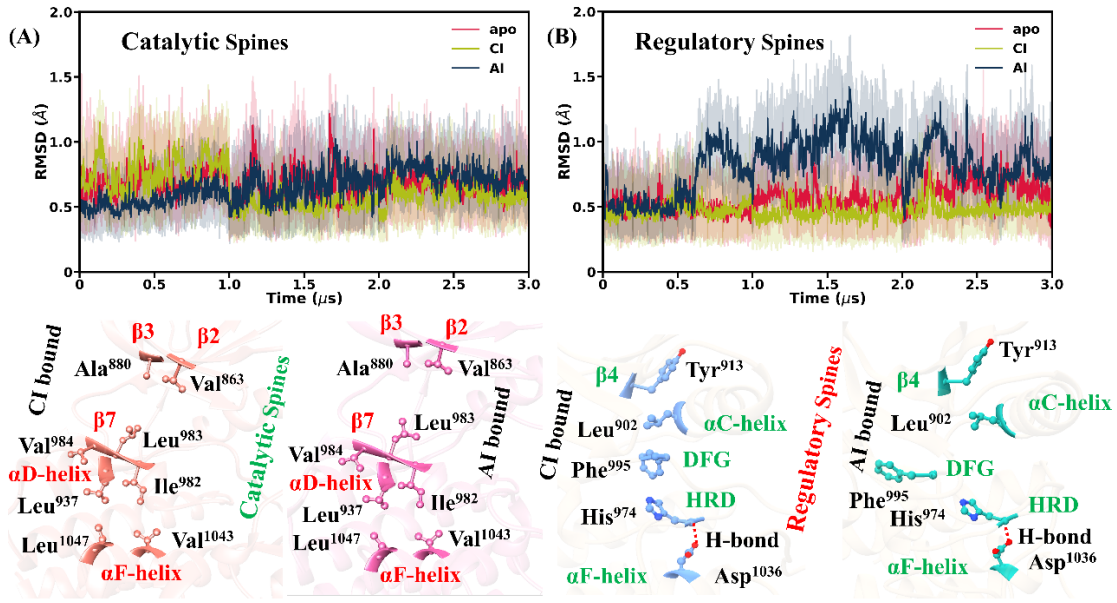


Figure 5.3: The time-evolution of backbone atoms RMSD values of hydrophobic (A) catalytic spine and (B) regulatory spines of apo, CI-bound, and AI-bound JAK2, respectively. Their corresponding structural alignment is also shown below.

To further reveal the structural movement and fluctuation around its mean position at their residual level upon inhibitor binding, we performed a root-mean-square fluctuation (RMSF)[333] analysis for the three systems, shown in **Figure 5.2(D-F).** It is evident from **Figure 5.2(D-F)** that the overall fluctuations pattern of apo and CI-bound systems looks similar. Still, the AI-bound, A-loop region is more flexible than other regions (see **Figure 5.2F**). A similar type of RMSF pattern was also seen in the case of WNK[359] kinase. This result supports higher RMSD values of AI-bound complex and suggests that the significant deviation arises from the A-loop region. Apart from that, the A-loop, the P-loop, and αC helix also show high flexibility in the case of an AI-bound complex, which agrees with the previous RMSD results in **Figure C2**. However, the catalytic loop and hinge retained their rigidity for all.

5.3.2. The Hydrophobic spines (C-spine and R-spine)

Hydrophobic spines alignment is another structural feature that is very crucial for maintaining the catalytic and regulatory activity of kinases[245]. The kinase skeletal assembly is controlled by eight catalytic (Val^{863}, Ala^{880}, Val^{984}, Leu^{983}, Ile^{982}, Leu^{937}, Leu^{1047}, and

Val[1043]) and four non-consecutive hydrophobic residues (Tyr[913], Leu[902], Phe[995], and His[974]) along with the anchored residue, Asp[1036], of the hydrophobic αF-helix, which serve as a base for hydrophobic regulatory spines. The R-spine of an active kinase is nearly aligned with linear and least solvent-exposed, while in an inactive form, this alignment is broken. As these two spines link the two lobes of KD and couple with A-loop, it is necessary to explore the structural evolution of these spines throughout the simulations. The time evolution of heavy atoms RMSD is shown in **Figure 5.3**. It is evident from **Figure 5.3A** that the catalytic spines are well aligned throughout the simulation for each case.

In contrast to C-spine, the R-spine shows a drastic change in its RMSD value in the case of the AI-bound system (see **Figure 5.3B**). This result suggests that in the AI-bound complex, R-spines are broken due to the DFG-out conformation, and in the CI-bound system, due to DFG-*in* conformation, R-spines are well aligned. Moreover, we estimated the one-dimensional potential of mean force (PMF) as a function of the RMSD of the spines (see **Appendix C Figure C3**), demonstrating the free energy profile of the stability of the spines before and after inhibitors bind. **Figure C3** suggests that the conformation of C-spines is very similar ("On" conformation in **Figure C3A**) in all three systems, but the R-spine shows a different free energy profile in the AI-bound system. The R-spine PMF of apo and CI-bound system showed similar well aligned ("On" conformation in **Figure C3B**) and AI-bound R-spine show "Off" conformation (broken conformation). A similar type of result was also obtained for the WNK[359] kinase and c-SRC[67] kinase.

5.3.3. The kinase intradomain motion

To gain deep insight into the impact of ligand binding on the independent dynamics among different intradomain of active and inactive forms, a traditional Pearson cross-correlation or dynamic cross-correlation (DCC) was used to calculate the correlation analysis, which was conducted and averaged over all MD trajectories and shown in **Appendix C Figure C4**. The DCC analysis describes the colinear correlation between the two residues Cα-atoms, reflecting whether they move in the correlated motions (DCC > 0) or the anti-correlated (DCC < 0) movements.

In **Figure C4**, the lower triangular part corresponds to apo systems, and the upper triangular part represents the CI-bound (**Figure C4A**) and AI-bound (**Figure C4B**)

systems, respectively. Overall, compared to the apo conformation, the inhibitors bound systems decreased anticorrelation movements (indicated by the blues patches in **Figure C4**). The first small, dotted box on the DCC map represents the N-lobe of JAK2, and the second rectangle dotted lines correspond to the A-loop of the C-lobe. The N-lobe containing the P-loop (residues 856-861) and αC-helix (residues 888-904), and C-lobe (residues 937-1129) shows anti-correlated movements, which is also noticed on the WNK[359], anaplastic lymphoma kinase (ALK)[380], BCR-ABL[381] and EGFR[382] kinase. In the CI-bound system, the DCC motion amplitude is the lowest (**Figure C4A**, upper triangle), reflecting that the residues interactions within the CI-bound system are the most stable, which further supports our RMSF results. Finally, we compared the concerted motions of the A-loop (**Figure C4,** dotted rectangle lines), and the results suggested that CI-bound along with apo systems showed higher correlated motions than the AI-bound system, which again showed consistent with the previous RMSF and RMSD analyses. Overall, the DCC map indicates that the binding of a type-II (allosteric) destabilized the A-loop and the allosteric binding sites (DFG-motif and αC helix), as evident from their anticorrelated movement with other domains, influencing their catalytic activity and renders the kinase deactivation.

5.3.4. Local motions and conformational free energy surfaces

To explore the large-scale collective movements in the three systems and the conformational interconversion of the protein through different low-energy states, we applied a systematic dimensionality reduction method, principal component analysis (PCA). Here we used residual Cartesian coordinates (heavy atoms) for PCA and explored the free energy surfaces (FES) of these systems using the first two principal components (PC1 and PC2) associated with significant conformational changes in the systems and find out their local low energy conformations. The FES and the porcupine diagrams of the first mode of these systems are shown in **Appendix C Figure C5**. **Figure C5** indicated that the dynamic motions of the apo and CI-bound systems are primarily restricted on its N-lobe and αG-αH loop, which is consistent with the RMSF data (**Figure 5.2**). The AI-bound systems exhibited the most prominent differences in the movement vectors of the A-loop (see **Figure C5E, F**), as we see in the three most probable low energy structures from the FES and representative structures extracted by the K-means clustering. By comparing the

FES of the three systems, we observe that the FES basin is more restricted in the CI-bound complex than the apo and AI-bound complex. This result also suggests that the CI-JAK2 interactions are more stable than the AI-JAK2 complex.

Further, the directions of motion mode of N-lobe (upward) and C-lobe (downward) in apo and CI-bound JAK2 are just opposite in directions, as well as amplitude, which signifies that the N-lobe and C-lobe are anti-correlated, which is in accordance with the DCC matrix analysis (see **Figure C4**). The AI-bound systems, the superimposed low energy conformations, and porcupines have suggested the major conformational changes are in the A-loop segment, which agrees with our previous RMSD, RMSF, and DCC map. A similar type of essential mode of domain motion was noticed in the case of Aurora-A kinase[68].

5.3.5. Conformational states of the HRD motif

In addition to the activation loop and αC-helix, the dynamics of HRD motif residue play an essential role in the kinase catalysis activity[383] and participate in the stabilization of active conformation[57]. Many kinases work as a central hub of the signaling network for kinase autophosphorylation[377]. This motif in JAK2 comprises three consecutive conserved residues (His^{974}-Arg^{975}-Asp^{976}) located at the beginning of the catalytic loop. The His^{974} is a component of R-spines, Arg^{975} stabilizes one of the phosphorylated Tyr1007 of the A-loop, and Asp976 plays a vital role in the substrate-binding, a mediating factor of phosphorylation reaction. The all-heavy atoms RMSD-based one-dimensional free energy landscape of the HRD motif is shown in **Appendix C Figure C6A** for the three respective systems. According to **Figure C6A**, the CI-bound system, HRD motif stabilized its global free energy minima around ~0.51 Å of RMSD, and less frequent local minima at ~1.48 Å, and the energy barriers between these two conformations are more than 2.0 kcal/mol. The apo JAK2 and AI-bound systems showed their stabilized global low energy conformation around ~1.0 and ~1.12 Å of RMSD. Moreover, the AI-bound state showed a local free energy basin at ~1.72 Å, but the energy barrier with the global is less than 1.0 kcal/mol and can be accessible at normal room temperature.

Subsequently, we investigated the local conformational change at their residual levels and performed a torsional angle-based FES analysis, shown in **Figure 5.4,** for Arg^{975} residue.

The time evolution of torsional angles is shown in **Figure C7**. **Figure 5.4** suggests that the energy basins of active structures (apo and CI-bound) lead drastically dissimilar. The active-like conformations (apo and CI-bound) showed their low energy basins at (175°, -75°) and (175°, 50°), see **Figure 5.4A, B**. The inactive AI-bound complex shows the energy basin at two different locations on the FES. The first energy basin in the AI-bound system is located at (-100°, -70°), and the second is at (65°, -180°). The representative structures of the basin are extracted from the MD trajectory and shown in **Figure 5.4D, E**. The Arg975 sidechain is exceptionally extended in the active form due to a higher degree of a torsional angle than the inactive structure, as suggested by their representative conformations. The closed Arg^{975} in inactive JAK2 formed a salt bridge with A-loop phosphorylated Tyr^{1007} and Asp^{1004} and stabilized the closed inactive A-loop conformation (see **Figure 5.4F**).

Previously, Oruganty[384] *et al*. experimentally and Meng[67] *et al***.** computationally studied the details structural analyses of the HRD motif and found that the backbone conformation of the HRD motif is in strained geometry in case of active-like conformations and formed a stable hydrogen bond with a non-catalytic residue. Here, we also checked their HRD motif strained geometry proposed by Oruganty et al. The backbone bond angle (N-Cα-C) of Arg975 is computed for apo, CI-bound, and AI-bound JAK2; its probability of density is shown in **Figure C6B**. The computed average bond angle is the same (~117°), which agrees with the experimental structures. The consequence of this observation is that the backbone of the HRD motif is in strained conformations and formed a strong hydrogen bond with the Asp^{1036} residue of the hydrophobic αF helix. To monitor these hydrogen bonds, we measure the distances of His^{974} and Arg^{975} backbone amide hydrogen atoms and Asp^{1036} side-chain oxygen atoms during the MD simulations. The probability density distributions of these distances are shown in **Figures C6C and D,** and their average values are reported in **Appendix C Table C1**. **Table C1** suggested that His^{974} and Arg^{975} formed strong hydrogen bonds with aspartate residue of αF helix, which stabilized the HRD motif backbone skeletal and agreed with the experimental study by Oruganty et al.

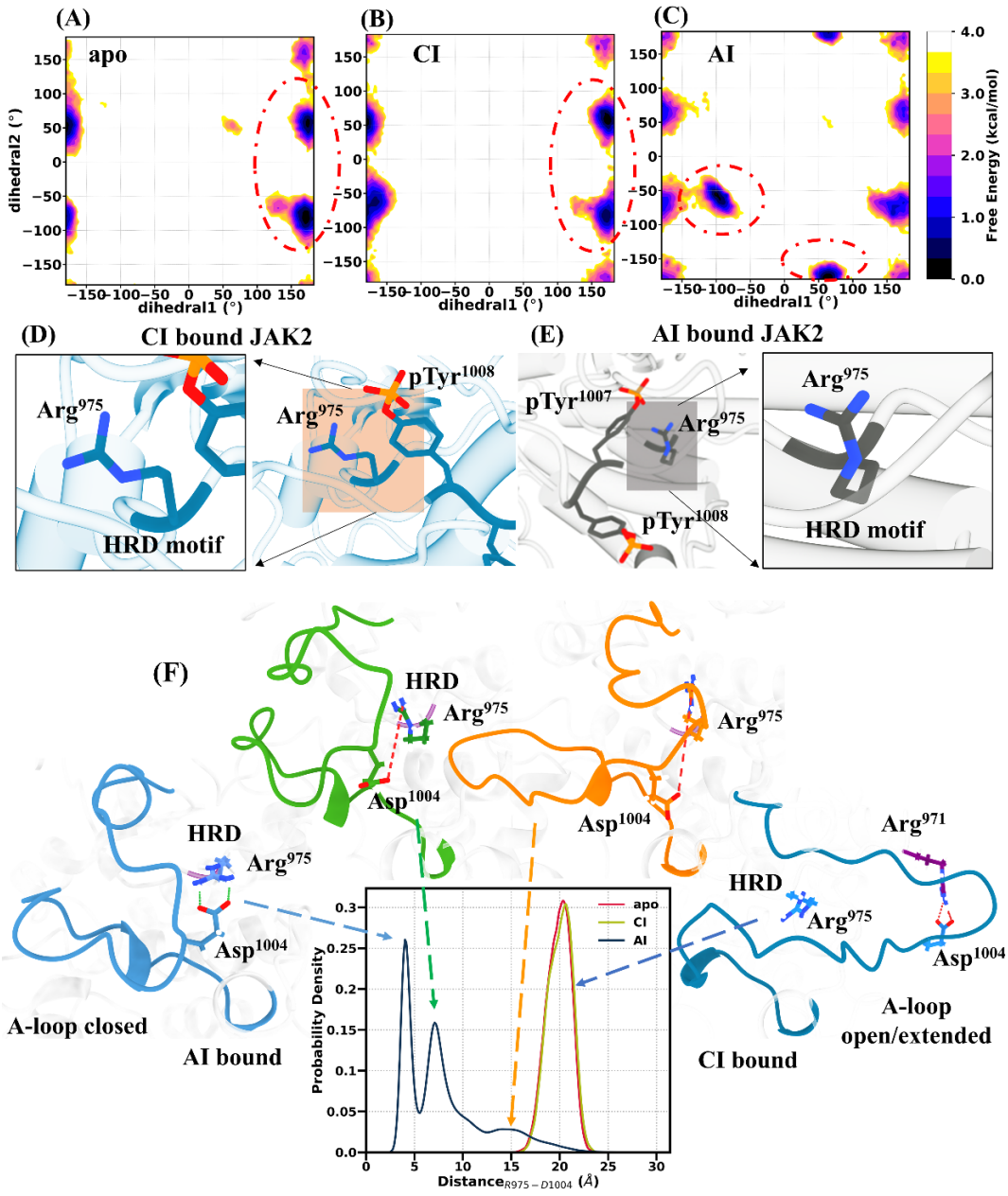


Figure 5.4: (A-E) The pseudo-dihedral-based free energy surfaces of HRD motif Arg^{975}. Here we considered the last 500 ns trajectories from every three replicas for all systems to calculate the free energy. **(F)** salt-bridge between Arg^{975} (HRD) and Asp^{1004} (A-loop) interactions in apo, CI-bound, and AI-bound JAK2.

5.3.6. Conformational diversity in αC helix and DFG motif

In protein kinases, the active and several inactive conformations are distinguished by various order parameters or the wide variety of structures and locations of the specific

residues of the kinase domain. Similarly, the residual position of Asp^{994}-Phe^{995}-Gly^{996} (DFG motif) and αC helix also controlled the active and various inactive conformations of kinases[385]. According to Tsai *et al.* [326], the free energy landscapes of the DFG motif and αC helix can be explored using two order parameters or reaction coordinates (RC); the pseudo torsion angle formed by Ile^{992}, Gly^{993}, Asp^{994}, and Phe^{995} (ζ(Gly993)) to define the DFG-*out* (ζ ≥ 120°) and DFG-*in* (ζ < 120°) states, which was first introduced by H. Möbitz[386] and the hallmark salt-bridge (Lys^{882}@NZ-Glu^{898}@CD) distance (R) monitor the αC helix movement, αC-*in* (R ≤ 4.5 Å) and αC-*out* (R > 4.5 Å) states as suggested by M. Sultan et al. [387]. The initial conformations of the DFG motif of active and inactive JAK2 are shown in **Figure 5.5(A, B)**. The conformational space sampled by the above-mentioned RC (ζ and R) is shown in **Figure 5.5C,** and their MD simulation extracted representative conformations are also shown in **Figure 5.5(D-I)**. **Figure 5.5C** indicates that the apo and CI-bound systems maintained their DFG-*in* conformations and AI-bound system in the DFG-*out* state during simulations. In contrast, the αC helix shows both active and inactive conformations for apo and AI-bound systems during the MD simulations. The CI-bound JAK2 shows quite restricted sampling space compared to the apo and AI-bound JAK2, which might help the CI-bound system to stabilize their αC helix in αC-in form. The apo system has probability to destabilized αC helix suggested by the αC-out conformation (see **Figure 5.5E**). Apart from the native active conformation, the AI-bound JAK2 shows the various inactive conformations sampled by out simulation, shown in **Figure 5.5(G-I)**. The similar types of αC-out conformations were seen in JAK1 kinase due to influence of phosphorylation of A-loop[373].

To capture the residual localization of DFG-Phe^{995} residue[385], we measured its distance from two conserved N-lobe residues: the β3-Lys^{882} and αC-Leu^{902}. The Leu^{902} (one of the R-spine residues) is the fourth residue after the conserved Glu^{898}. The distance with DFG-Phe995 signifies the DFG spatial conformations and the opening/closing of the allosteric site of JAK2. The distance between conserved Lys^{882} and Phe^{995} captures the closeness of the N-lobe and DFG motif, thus giving an estimate of the upward movement of the Phe^{995} ring and distinguishing the spatial orientations of the DFG motif during MD simulations. The distances are plotted against each other in a scatter plot and their free energy surfaces (FES) to capture the low energy conformations (see **Figure 5.6**). We have performed K-

means clustering of these data to find out low energy representative conformation shown in **Figure 5.6(B, C)**. The restricted conformational sampling spaces and a single low-energy basin of apo and CI-bound JAK2 suggested that the allosteric site is closed (no space for inhibitors), and DFG-Phe^{995} is in fully *in*-state. In construct, the conformational sampling space was broad, and a broad global low energy basin was obtained in the case of the AI-bound complex. It indicates the flexibility of the structure to many inactive-like conformations. The high distance gap between DFG-Phe^{995} and αC helix is suitable for the tight binding of an inhibitor in the allosteric binding pocket.

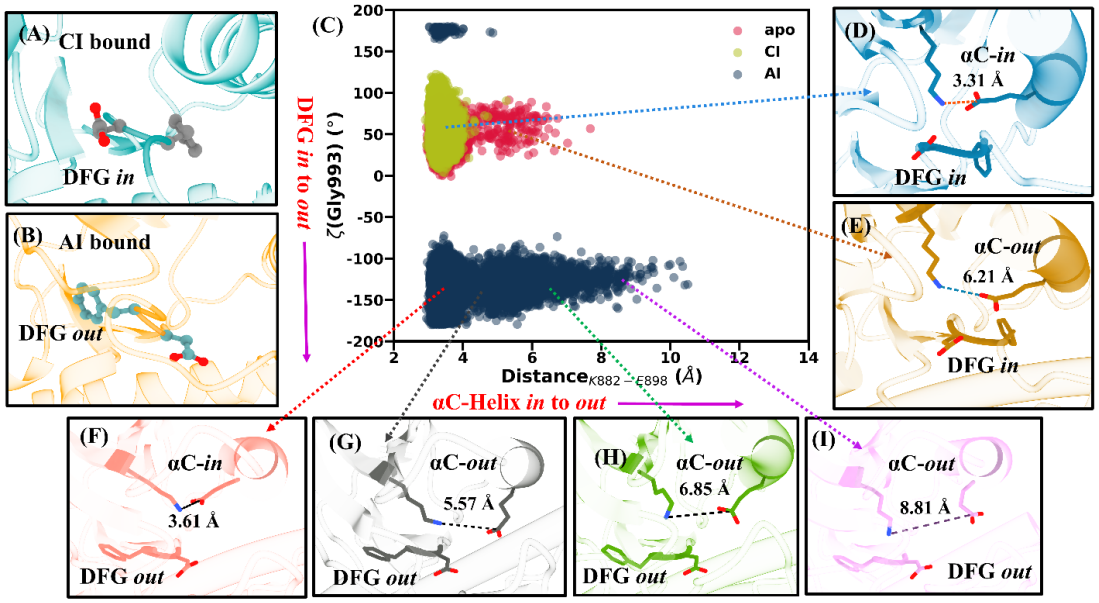


Figure 5.5: (A-B) The initial conformations of JAK2 bound with CI and AI. (C) The conformational space is sampled by three systems (red: apo; yellowgreen: CI and darkslategrey: AI) along with the pseudo-torsional angle (ζ) and distance between β3 (Lys^{882}) and αC-helix (Glu^{898}) residues. (D-I) The corresponding representative conformations of active and inactive JAK2 structures.

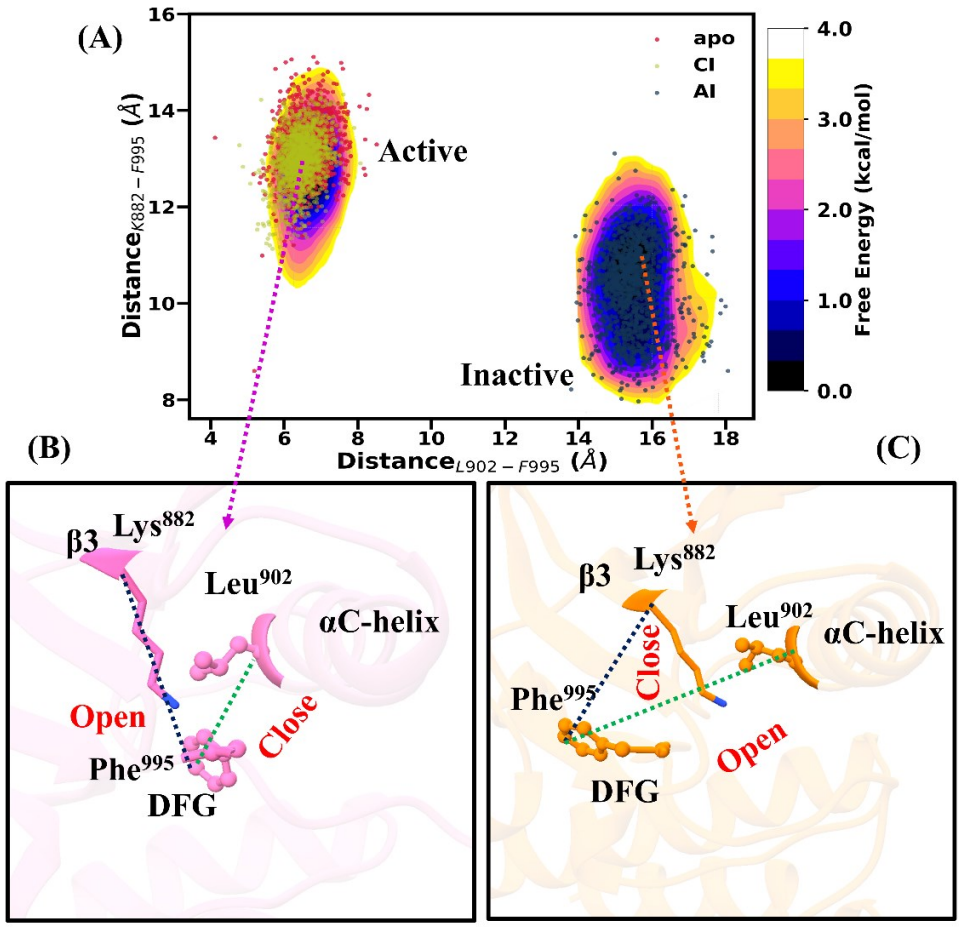


Figure 5.6: Two-dimensional free energy surfaces underlying the dynamics of critical structural motifs. The X-axis represents the allosteric pocket opening/closing by the distance between Phe^{995}@CZ (DFG motif) and Leu^{902}@CA (αC) residues. The Y-axis represents the closeness of DFG-Phe^{995} to N-lobe β-sheet strands by the distance between Phe^{995}@CZ (DFG motif) and Lys^{882}@CA (β3) residues.

5.3.7. Comparison of A-loop conformational variability

To characterize the conformational dynamics of the A-loop induced by inhibitor binding, we constructed the 2D- FES of the A-loop along with the two global reactional coordinates: the backbone atoms RMSD and the radius of gyration (ROG) of the A-loop. The RMSD indicates the overall stability of the loop, and ROG helps us to estimate the A-loop compactness (closed/extended). The FES results of these systems are shown in **Appendix C Figure C8**. The FES results suggest the apo and CI-bound systems offer a narrow single low energy basin at the upper left corner of **Figure C8(A, B)** containing well-extended

open A-loop conformation (active-like), which is consistent with our previous RMSD and DCC map analyses. The A-loop of AI-bound systems has three low-energy basins located at three distinct locations in the FES map with a minimum energy barrier of more than 2.0 kcal/mol (see **Figure C8C**). The high free energy barrier between these conformations suggests distinct probable inactive (due to closed-form) conformations adopted by AI-bound JAK2.

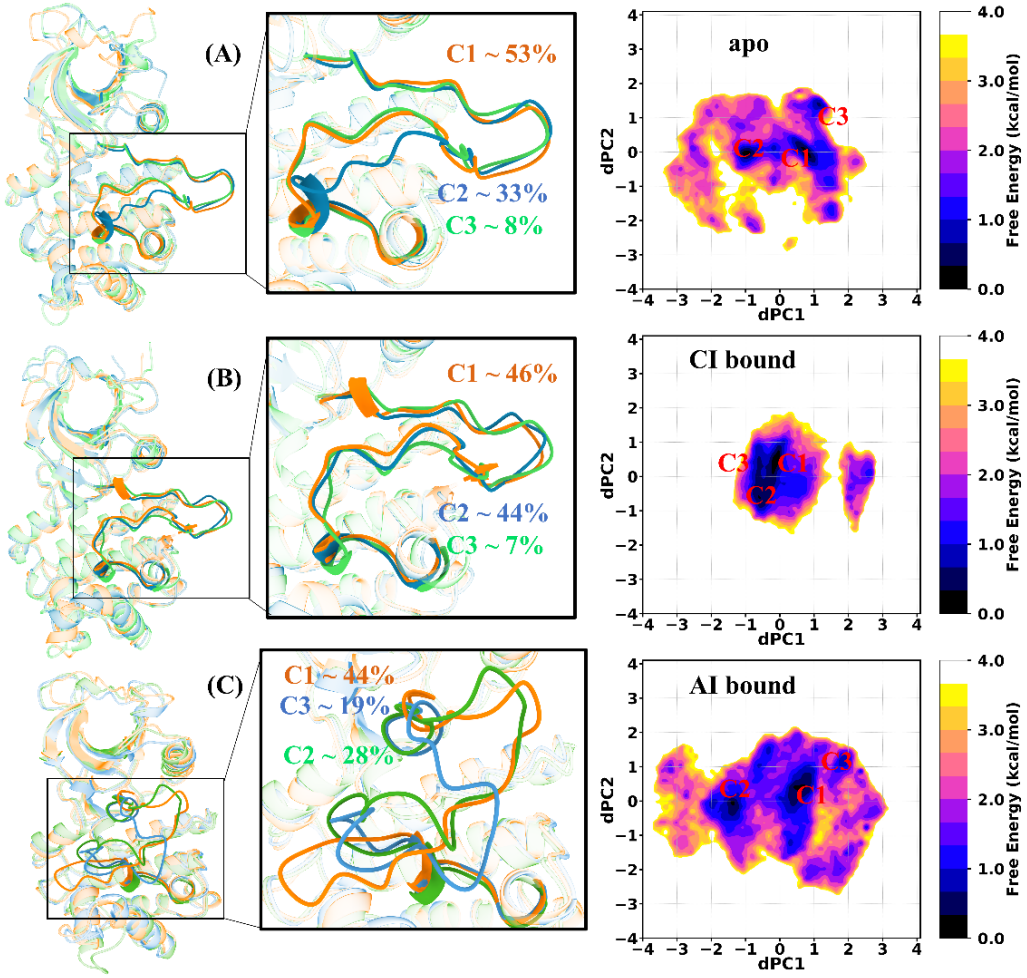


Figure 5.7: The two-dimensional free energy surface (FES) obtained from the dihedral principal component analysis (dPCA) of the A-loop for (**A**) apo and (**B, C**) inhibitors (CI/AI) bound JAK2, respectively. Each local minimum is highlighted and numbered (C1 to C3) with increasing energy. The K-means cluster analysis shows their representative structures on the middle panels.

To get more insight into the internal conformational dynamics of the A-loop, we calculated all all-possible dihedral angles of the A-loop. We performed a dihedral angle-based PCA

(dPCA) approach to reduce their motion dynamics to a lower dimension and better picture the A-loop inactive conformations. In the PCA analysis, we observed that the first principal vector (dPC) showed a higher percentage of variance. The FES of dPCA is constructed using dPC1 and dPC2 and shown in **Figure 5.7**. Furthermore, we used the K-means clustering scheme to extract the low energy conformation from the PCA data, as shown in **Figure 5.7** (left panel). **Figure 5.7A** shows three conformations of A-loop in apo systems, in which C1 shows the highest occurrence of 53%. The dPCA-based FES results suggested that the CI-bound systems showed restricted sampling space. The free energy barrier among the low energy structures is shallow and room temperature accessible (see **Figure 5.7B**). One of the A-loop Lys^{999} formed a salt-bridge interaction with DFG-Asp^{994} in an AI-bound system but not in active JAK2 (instead of Asp^{994}, it creates a salt-bridge with Glu^{1012}), as suggested by **Figure C9**.

5.3.8. Salt-bridge interactions control the conformational switch of the A-loop

The well-ordered open/extended conformation of the A-loop (see **Figure C10)** shows many significant interactions, which helps the A-loop stabilize in its open/extended form in active JAK2[388]. The base of the A-loop is stabilized by Arg^{893}, $Asp^{894,}$ and Arg^{897} residues from the αC helix and Arg^{975} from the HRD motif. The tip of the A-loop stabilized in its extended form by the interactions of Arg^{971} and Lys^{1030}. The phosphorylated $pTyr^{1007}$ is also stabilized by several lysine residues, including Lys^{1005}, $Lys^{1030,}$ and Lys^{1009} residues. The A-loop (APE motif) end is well stabilized during simulation with another salt bridge between APE-Glu^{1024} and Arg^{1113} residue (see **Figure C10**). The Tyr^{972} and Val^{1000} hydrophobic interaction also stabilized the A-loop in its extended spatial conformation in active JAK2[31].

Next, the inactive DFG-out AI-bound structure shows three independent conformations (see **Figure 5.8C**), and we checked their unique conformational residual interaction patterns (see **Appendix C Figure C11**). The conformation, C1 (probability ~44%), shows a distinct interaction pattern than the other two conformations (C2 and C3). **Figure C11A** shows that the beginning part of the A-loop stabilized in C1 by many interactions, including Asp^{1004}-Arg^{897}, Asp^{894}-$Lys^{999,}$ and Asp^{994}-Arg^{975} salt-bridges. The $pTyr^{1007}$ and $pTyr^{1008}$

area is mainly stabilized by two lysine residues, Lys^{1009} and Lys^{1011}, which formed salt bridges. The APE motif (end of A-loop) and before APE motif stabilized by Arg^{1113} and Arg^{938} residue. The second probable conformation, the C2 (probability ~28%) interactions pattern (see **Figure C11B**), suggests that the $pTyr^{1007}$ residue forms a salt bridge with Arg^{975} instead of Lys^{1009} and Lys^{1011} residue and $pTyr^{1008}$ with Lys^{1005} instead of Lys^{1009}. Finally, the conformation C3 (probability ~19%) interaction pattern (see **Figure C11C**) suggests that Lys^{999} residue form an interaction with DFG-Asp^{994} instead of αC-Asp^{894} residue. Similarly, HRD-Arg^{975} is now forming a salt bridge with $Asp^{1004,}$ and $pTyr^{1007}$ is not involved in any salt bridge interaction. Overall, these distinct interactions pattern of A-loop (inactive closed) with its nearby residues help to stabilize their particular rearrangements in three inactive-like conformations.

5.3.9. The Tyr^{1007}-Tyr^{1008} region of A-loop

As mentioned above, the A-loop of JAK2 required phosphorylation of two consecutive tyrosine, Tyr^{1007} and Tyr^{1008}, which are located at the center of the A-loop[31]. Due to their critical role in activating the JH1 domain, we monitored their dihedral angle (φ, ψ) (see **Appendix C Figure C12A, D**), and their 2D-scatter plot is shown in **Figure 5.8**, along with representative structures. The 2D conformational (φ, ψ) sampling space of $pTyr^{1007}$ (see **Figure 5.8A**) in an AI-bound system shows that the conformational flexibility but in active structure is relatively stabilized. This result is consistent with our previous free energy landscapes and salt-bridges analysis. **Figure 5.8B** suggests that pTyr1008 shows relative stability in its initial spatial conformation due to strong salt bridge interactions, including Arg^{893}, $Arg^{897,}$ and Lys^{999} in a CI-bound system (see **Figure C10**) and Lys^{1005} and Lys^{1009} in an AI-bound system (see **Figure C11**).

Moreover, to complement these results to estimate their radial distribution function (g(***r***)) of water molecules around the oxygen atoms of these phosphorylated tyrosine residues along with solvent accessibility surface area (SASA), shown in **Figure C12**. The g(r) and SASA results indicate that the $pTyr^{1007}$ residue in the CI-bound system gets more solvent-exposed than in the AI-bound system (see **Figure C12B, C**). A similar type of opposite phenomenon is noticed in the case of second phosphorylated tyrosine (pTyr1008), which

is a more exposed solvent in AI-bound cases that CI-bound (see **Figure C12E, F**). Overall, these results indicate that in active form, both the phosphorylated residues are opposite directions, and in case inactive, these are in flip their conformations.

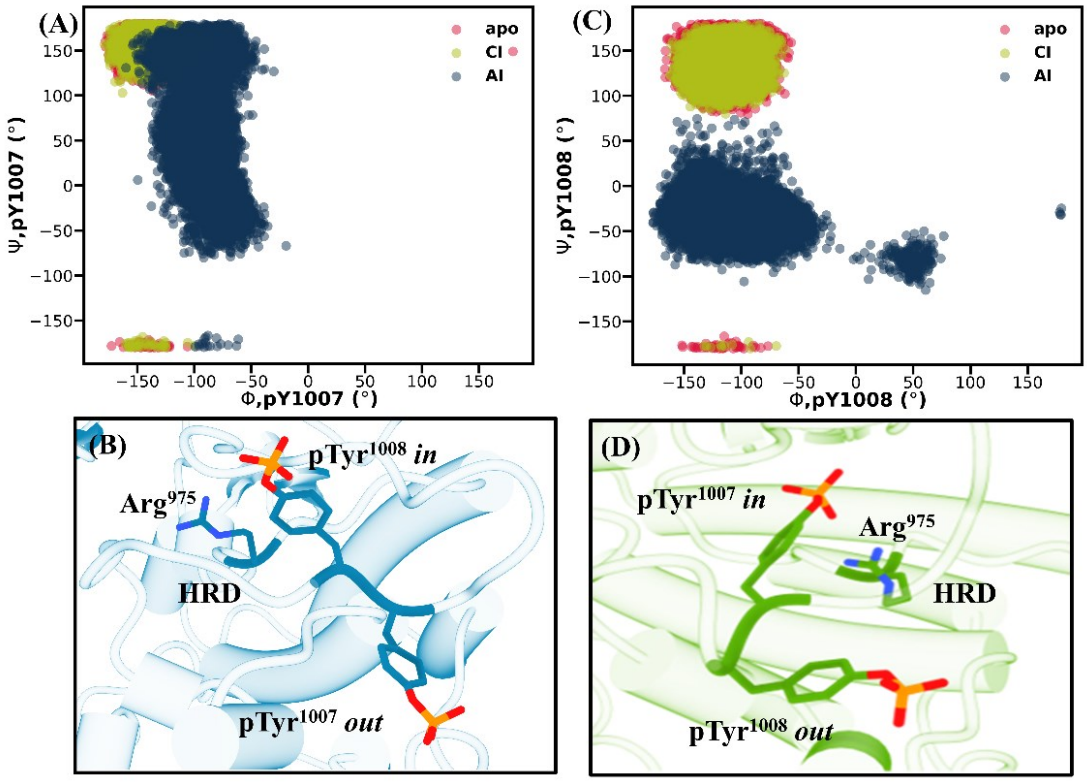


Figure 5.8: The conformational space sampled by three systems and the dihedral angles (φ/ψ) from phosphorylated Tyr^{1007} and Tyr^{1008} and their corresponding snapshots.

5.3.10. Community network analysis of kinase domain

Protein structure network (PSN)[207] approaches were used for long-space communications and to study the allostery effect on network connectivity. We used the most probable conformations (C1 conformations) from our previous PCA analysis in this calculation. The PSN depicts a structural network comprising nodes, edges, hubs, and community networks. Here, the C_α-atom of amino acid residues represents a node, and edges show non-covalent interactions among nodes with a cutoff distance of 4.5 Å between two residues. The PSN properties, such as the number of nodes, hubs, and community, were also analyzed for apo

and complex systems of JAK2. The average network properties are shown in **Table C2,** and the corresponding hubs and communities are shown in **Figure 5.9.** Interestingly, a significant difference in the average network properties was observed among the different systems. Residues making four or more edges are referred to as hubs at that particular I_{min}. It is evident from **Table C2** that the total number of linked nodes is much high for the AI-bound system, and after inhibitor bound, the number of links and nodes in the systems also increases suggested by the low value for apo (see **Appendix C Figure C13**). Similarly, the three systems' overall total hubs and community networks are significantly different. The active JAK2 conformations (apo and CI-bound) possess 27 and 32 network hubs, respectively. The high value of nodes, links, and hubs suggested that AI-bound (C1) systems are relatively more stable than apo (C1) and CI-bound (C1) systems (see **Figure C13**).

Further, we compare their community network structure map and schematic representation of their community, shown in **Figure 5.9(D-F)**. The apo structure shows that the community network is distributed throughout the N-lobe and C-lobe. **Figure 5.9(D-F)** suggests that the highest-ranked community network, C1, mainly stabilized the C-lobe of JAK2 for all cases. The second highest community, C2, is located at N-lobe for apo JAK2 but relocated to the A-loop area in case of inhibitors bound case. Let's compared the CI-bound and AI-bound community networks. It is clear from **Figure 5.9(E-F)** that their community network connectivity is disturbed due to the types of inhibitors that induce perturbation in the conformations. These comparative community network analyses provide more insight into the conformational rearrangement of JAK by inhibitors.

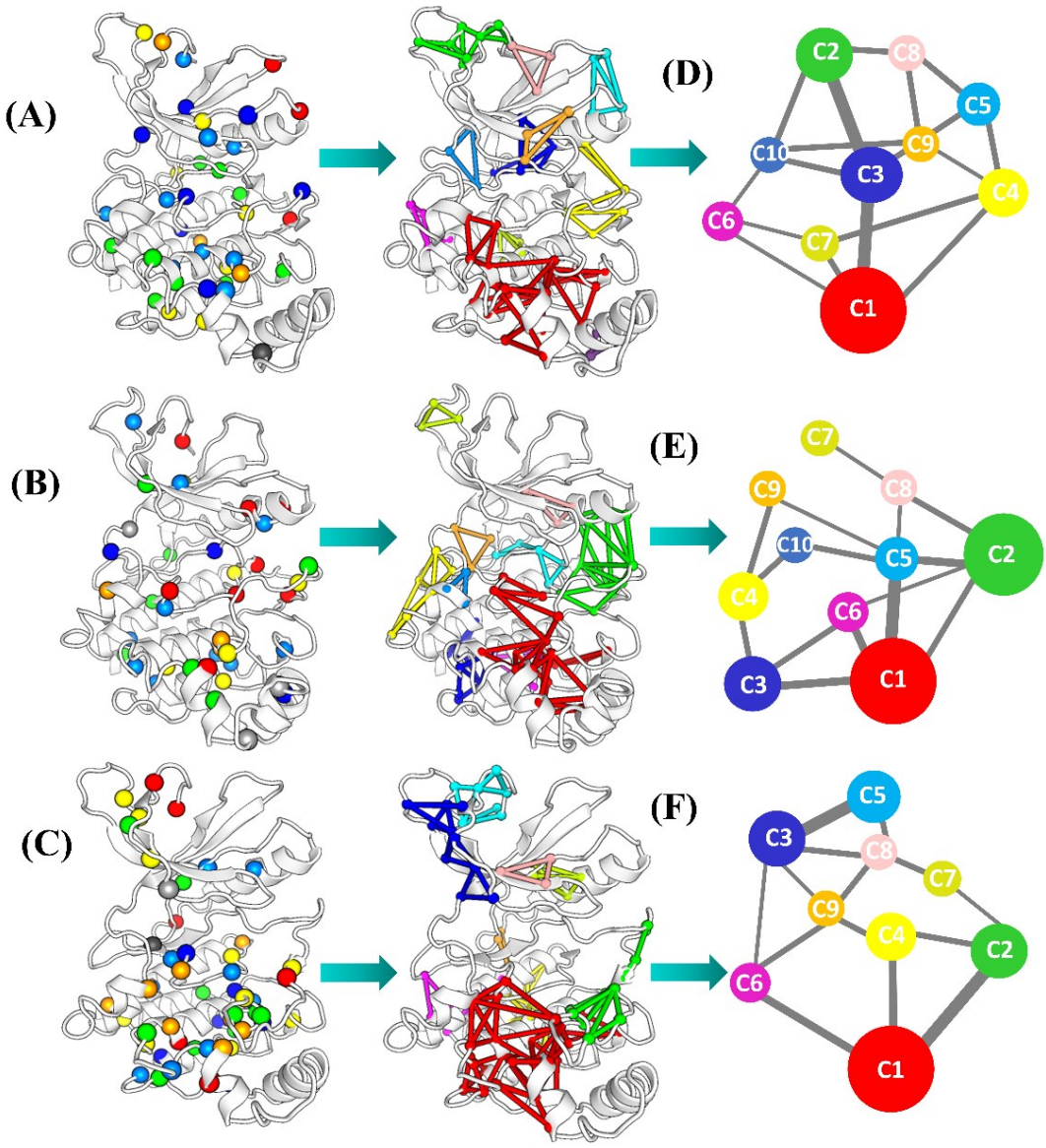


Figure 5.9: Network hubs and community of residue contact network in (**A, D**) apo; (**B, E**) CI-bound, and (**C, F**) AI-bound system, respectively. The hubs are indicated as spheres centered on the Cα-atoms on the protein. The community network describes a fundamental feature of residual contact networks. Each color represents an individual community, and lines connecting different color spheres visualize the inter-community connections, and the thickness of these lines is proportional to the corresponding edge connectivity.

5.3.11. Comparative binding modes

The binding free energy calculations, residue-wise decomposition of the total binding free energy, and the hydrogen bonds between the inhibitors and JAK2 analyses were performed. The total average binding free energy of CI-bound and AI-bound systems are listed in

Table 5.1, and their replica run calculations are listed in **Appendix C Table C3**. It is evident from **Table 5.1** that the type-II inhibitor (NVP-BBT594) shows the lowest free energy ($\Delta \boldsymbol{G_{bind}}$) or highest affinity (-16.83 ± 0.92 kcal/mol) toward JAK2 than a type-I competitive inhibitor, tofacitinib (-9.76 ± 0.83 kcal/mol). These results qualitatively agree with the experimental IC^{50} values, 21.7 nM[33] and 0.99 nM[350] for tofacitinib and NVP-BBT594, respectively. **Table 5.1** also suggests that the AI-bound systems show more attractive van der Waals energy (-71.91 kcal/mol) and total non-polar free energy ($\Delta G_{\mathrm{vdW,np}}$) as well (-78.29 kcal/mol) and supported by the previous study by Kong *et al.* [349]. Similarly, the binding dis-favorable total polar energy ($\Delta G_{\mathrm{pol,elec}}$) is high for AI-bound complexes than CI-bound complexes. The binding favor intermolecular electrostatics interaction is overcompensated by their polar desolvation free energy for both cases. Therefore, the primary interactions between these complexes are the van der Waals interactions.

Table 5.1: The energetics components of the total binding free energy for CI-bound and AI-bound JAK2 complexes in kcal/mol.

Component	CI-bound	AI-bound
$\Delta \boldsymbol{E_{vdW}}$	-45.47 ± 0.08	-71.91 ± 0.07
$\Delta \boldsymbol{E_{elec}}$	-32.68 ± 0.12	-23.78 ± 0.17
$\Delta \boldsymbol{G_{pol}}$	51.96 ± 0.13	61.17 ± 0.23
$\Delta \boldsymbol{G_{np}}$	-3.80 ± 0.0	-6.38 ± 0.0
$\Delta \boldsymbol{G_{solv}}$	48.16 ± 0.13	54.79 ± 0.23
$\Delta \boldsymbol{G_{pol,elec}}$	19.28 ± 0.17	37.39 ± 0.28
$\Delta \boldsymbol{G_{vdW,np}}$	-49.27 ± 0.08	-78.29 ± 0.07
$\Delta \boldsymbol{E_{MM}}$	-78.15 ± 0.14	-95.69 ± 0.18
$\Delta \boldsymbol{H_{bind}}$	-29.96 ± 0.09	-40.92 ± 0.20
$-\mathbf{T}\Delta \boldsymbol{S_{conf}}$	20.20 ± 0.83	24.09 ± 0.90
$\Delta \boldsymbol{G_{bind}}$	-9.76 ± 0.83	-16.83 ± 0.92
IC^{50} (nM)	21.7	0.99

To gain more insight into their interaction spectra at the amino acid level, the total binding free energy was decomposed at their residual level for both the complexes and shown in **Figure 5.10A,** and the independent replica runs decomposed energy shown in **Appendix C Figure C14**. The residues contributing lower than -1.0 kcal/mol free energy to the total

binding free energy are highlighted in **Figure 5.10A** and listed in **Appendix C Table C4**. Residues Leu[855] and Val[863] from P-loop contributed significantly to the binding of both CI[33,284] and AI[349,350] inhibitors. One β3 strand hydrophobic residue, Ala[880,] also took part in the interaction and contributed attractively, see **Figure 5.10A and Table C4**. Interestingly, the αC helix involves in the interaction of AI and JAK2 but not in CI-bound JAK2, as suggested by their interaction spectra, **Figure 5.10A** and their representative binding poses, **Figure 5.10(B-C)**. The involvement of αC helix (-2.93 kcal/mol) (residues Ile[901] and Leu[902]) increases the binding affinity of allosteric inhibitor, NVP-BBT594 than tofacitinib (-0.03 kcal/mol), see **Figure 5.10D**. Note that the hydrophilic Glu[898] residue (shown in red color in **Figure 5.10A**) from the αC helix shows an unfavorable polar interaction (0.60 kcal/mol) with the hydrophobic moiety of AI inhibitor, which is consistent with the previously reported study by Jiao Jiao Li *et al.* [350]. Interestingly, type I, tofacitinib is strongly stabilized by hinge region Glu[930] (-1.21 kcal/mol), Tyr[931] (-1.67 kcal/mol), and Leu[932] (-1.64 kcal/mol) and type II, NVP-BBT594 by residues Tyr[931] (-1.90 kcal/mol) and Leu[932] (-2.81 kcal/mol). Apart from these interactions, our calculations also reveal another hydrophobic residue from β7 strands (end of the catalytic loop), Leu[983,] actively participates in both types of inhibitors binding (see **Table C4**). The A-loop actively participates in the allosteric inhibitor binding, as suggested by **Figure 5.10D,** and the residues Gly[993] (-1.78 kcal/mol), Asp[994] (-1.16 kcal/mol), and Phe[995] (-1.95 kcal/mol) provide a favorable force to the total binding affinity. Overall, many hydrophobic interactions suggested that due to DFG-out-like inactive conformation, type II AI extended to residues Ile901 and Leu902 with strong non-polar interactions, increasing its total binding affinity. All the factors we mentioned in this study need to be considered when searching for new inhibitors for JAK2 kinase.

Apart from the hydrophobic interactions, the electrostatic interactions and hydrogen bond network between these inhibitors and JAK2 play a crucial role in enhancing the total binding affinity. These inhibitors' hydrogen bonds are summarized in Supplementary Information Table S5. The pyrrolopyrimidine core of CI (tofacitinib) exhibited very strong hydrogen bonds with the hinge region of JAK2; Glu[930] and Leu[932] formed two hydrogen bonds with an occupancy of 84.14 and 79.07 %, respectively. These results are consistent with earlier experimental[33] and theoretical[284] studies. In a similar fashion, the type-II AI

also formed three stable hydrogen bonds with DFG-*out* JAK2 *via* Asp^{994} and Leu^{932} (two hydrogen bonds), which agree with the study by Jiao Jiao Li. The DFG-Asp994 involves the highest occupied hydrogen bond (Lig@O17 ⋯ Asp^{994}@N-H) with AI. Two hydrogen bond interactions (Leu^{932}@O ⋯ Lig@N43 and Lig@N39 ⋯ Leu^{932}@N-H) formed by Leu^{932} with occupancies of 50.67 and 40.84, respectively. Therefore, these hydrogen bonds play an essential role in stabilizing the inhibitors and JAK2 interactions favorably.

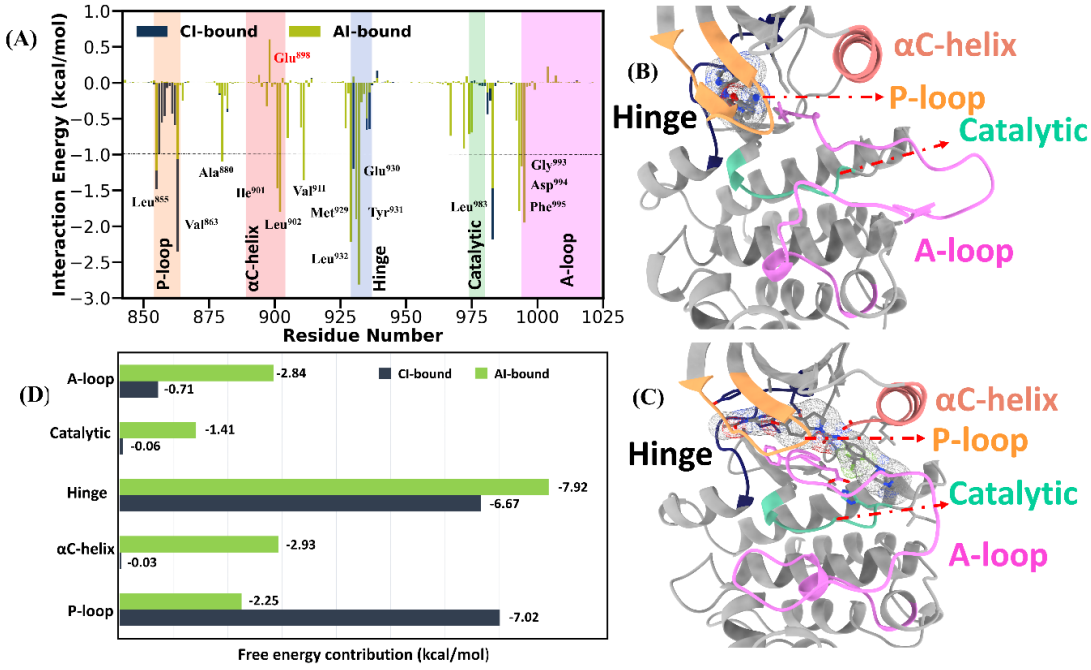


Figure 5.10: (**A**) Average per-residue energy contribution and the location of these residues in the JAK2 structure; highlighted residues are contributed lower than -1.0 kcal/mol. (**B-C**) JAK2 and essential areas are shown color-coded with CI and AI bound states. **(D)** The binding free energy contribution of some important competitive and allosteric inhibitor binding sites.

5.4. Conclusions

Herein, we performed extensive multiple GaMD simulations of cumulative 9 μs (with three different initial velocities) to elucidate the inhibitor-induced conformational dynamics of JAK2. This study compared the conformational dynamics and inhibition mechanisms of competitive inhibitors (CI) and allosteric inhibitors (AI) and compared them with their native apo conformation. In view of the comparative analysis among these systems, our results reveal that the allosteric inhibitor-induced inactive conformations of JAK2 are

unique from their typical inactive conformations. Hence these distinct inactive conformations can be exploited to design a new selective inhibitor. Our results also suggested significant conformational variations upon inhibitor binding are noticed in the A-loop and αC helix motions. Furthermore, our study reveals some unique conformational features of inactive and active JAK2, including an open catalytic cleft between the N- and C-terminal toward the N-lobe movement of the P-loop, the outward motion of αC helix, the closed inactive-like A-loop conformation, and the unique DFG motif orientation. Moreover, the outward sign of αC helix impacts the hallmark salt bridge formation of Lys^{882} and Glu^{898} in an inactive conformation. Finally, we compared their binding poses and free energy by the MM/PBSA approach. The free energy calculations suggested that the AI's binding affinity is higher than CI to JAK2 due to an increased favorable contribution from the total non-polar interactions relative to CI. We have also identified some prominent residues whose involvement in the inhibition mechanism is significant for AI inhibitors, including the residues (Ile^{901} and Leu^{902}) from the αC helix region. These individual inactive states and their binding patterns provide an essential starting point to design specific novel allosteric inhibitors in the future.

Chapter 6

Chapter 6

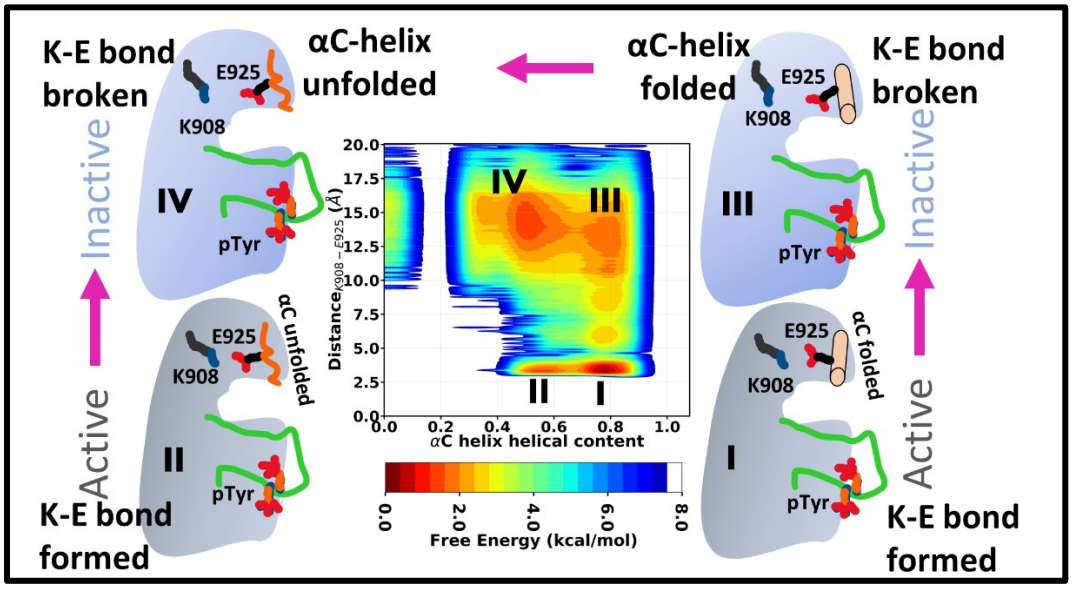
Finding Inhibitors and Deciphering Inhibitor-induced Conformational Plasticity in the Janus Kinase via Multiscale Simulations

AIM: Finding out novel JAK inhibitors from the phytochemicals databases to treat autoimmune-related disorders and inhibitor-induced inactive conformation.

This chapter is reproduced from our original article:

Sk, M.F., & Kar, P. (2022). "Finding inhibitors and deciphering inhibitor-induced conformational plasticity in the Janus kinase *via* multiscale simulations". ***SAR and QSAR in Environmental Research.***

Table of Content (TOC)



6.1. Introduction

The Janus kinase (JAK) signal transducers and activators of the transcription (STAT) pathway plays a significant role in controlling cytokine-dependent immune systems, hormones, growth factors, and hematopoiesis[389,390]. The JAKs are cytoplasmic, non-receptor tyrosine kinases and have four family members: JAK1, JAK2, JAK3, and tyrosine kinase 2 (TYK2). Cytokine binds to its respective receptor and activates JAKs, which in turn phosphorylates the tyrosine residues within the receptor's cytoplasmic domain. These phosphotyrosines then serve as docking sites for the cytoplasmic STAT. Then the STATs dimerize and move into the nucleus to regulate the transcription of specific target genes [391].

Reported genetic and biochemical studies have shown that JAK1 is involved in cytokine signaling, including the beta common (βc), gamma common (γc), and type-I type-II interferon, gp130, IL-10, and IL-6 subfamilies[389,392]. Typically, the JAK-STAT pathway transmits signals from the cell membrane to the nucleus in response to cytokines and extracellular growth factors. Dysregulated JAK-STAT functionality can lead to various immune diseases and human tumors like gliomas, whose most common site is the brain[393]. Recent studies suggested that the JAK/STAT signaling cascade is a crucial factor that promotes neuroinflammation[14,15] in neurodegenerative diseases, including multiple sclerosis[16] (MS) and Parkinson's diseases[394] (PD). Consequently, JAKs have been characterized as potential targets for developing anti-inflammatory, anti-tumor drugs[395]. Further, JAKs have also been targeted for treating hematological disorders, including myeloproliferative disorders (MPDs)[396], acute leukemia, and lymphoproliferative disorders (LPDs)[397,398]. Computer-aided drug designing tools offer initial support to wet-lab experiments, and the JAKinibs (JAK inhibitors) design is no exception in this regard[399].

Till today, there are few FDA-approved JAKinibs, including Ruxolitinib (2011)[400], Tofacitinib (2012)[401], Baricitinib (2018)[402], and Upadacitinib (2019)[403], that work as ATP-competitive inhibitors and block several downstream signaling cascades[404]. Moreover, it is noteworthy that the insufficient selectivity of these inhibitors complicates their clinical use and results in future harmful side effects. Therefore, the development of JAKinibs for inflammatory arthritis and immune-related diseases remains challenging and an open issue

due to the high structural conservation of the kinase domain (JH1). A previous study reported that JAKs isoforms have sequence identity ranges from 46-55% and sequence similarity of 60-70%[390]. Moreover, the amino acids sequence at the activation loop (A-loop), Catalytic loop, glycine-rich Phosphate binding loop (P-loop), and hinge region are conserved and have a sequence identity 65-78%[390].

Herein, we reported the structure-based virtual screening, ADMET (absorption, distribution, metabolism, excretion, and toxicity) pharmacological and molecular dynamics simulation to find the hit compound that can design potential novel inhibitors against the JAK family (see **Figure 6.1**). We screened more than 194610 natural small molecules from different natural product databases, including *AnalyticCon* discovery (https://ac-discovery.com/) (32396 compounds), InterBioScreen (https://www.ibscreen.com/) (161102 compounds), and PubChem (https://pubchem.ncbi.nlm.nih.gov/) (1113 compounds) by the structure-based high throughput virtual screening (HTVS) against JAK1. After VS, we filtered each top selected hit from their ADMET profile. Each docked complex proceeded for further binding posed and structural refinement with MD simulations. Finally, we measured the binding free energy of each compound against JAK1 using the molecular mechanics Poisson-Boltzmann surface area (MM-PBSA) scheme.

6.2. Materials and methods

6.2.1. Target protein preparation and receptor grid generation

The crystal structure of JAK1 (PDB ID 3EYG, resolution 1.90 Å)[33] was used as the target protein for our study, and both Tyr^{1034} and Tyr^{1035} were phosphorylated. Missing residues were added using the Modeller plugin of UCSF Chimera [323], and preparation before docking was performed using the protein preparation wizard of the Schrodinger suite (Version 2021-1, Schrodinger). Necessary missing hydrogen atoms were added, charges were assigned, and Het states were generated using Epik at pH 7.0 ± 2.0. The essential water molecules remained the same, and the other molecules apart from the co-crystallized ligand were deleted. Finally, the protein was minimized using the OPLS3 force field[405]. A grid was created by considering the co-crystal ligand, which was included in the active site

of the selected protein target. A receptor grid was generated around the protein (PDB-3EYG [coordinates X-11.22, Y- 13.78, Z- -15.13, 10 × 10 × 10] by choosing the inhibitory co-crystallized ligand (i.e., MI1). The centroid of the selected ligand was used to create a grid box around it and van der Waals radii of protein atoms were scaled to 1.00 Å with a partial atomic charge cut-off for polarity set to 0.25. The grid was standardized by redocking the co-crystallized ligand MI1. Further, the docking protocol was validated by superimposing the re-docked ligand with a co-crystallized ligand, and the RMSD was calculated between them (see **Appendix D, Figure D1**).

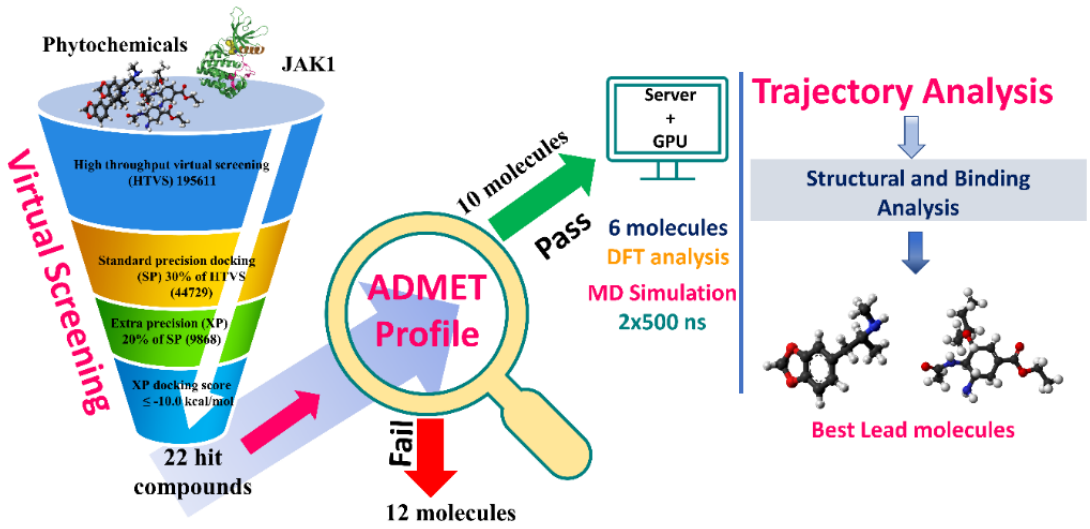


Figure 6.1: Workflow of identifying best lead molecules against JAK1 via systematic structure-based virtual screening, ADMET profiles, and molecular dynamics simulations.

6.2.2. Ligand preparation and receptor-based virtual screening

Our main goal is to find a novel JAK kinase inhibitor from the natural phytochemical database. We considered ~ 194610 natural compounds for the current study. All these compounds were subjected to ligand preparation using the LigPrep module of the Schrödinger suite. The different possible ionization states were generated using the Epik module at pH 7.0 ± 2.0. Almost 32 conformers of all the combinations of each ligand were generated while keeping other parameters as default. Finally, all ligand molecules were minimized and optimized using the OPLS3 force field.

After that, we used a stepwise QikProp and Lipinski's rule of five filtering protocols through the virtual screening workflow (VSW) module [406]. Then the following stage compounds were docked using HTVS docking. The top 30% of HTVS filtered compounds were further docked with Glide-SP (standard precession)[406]. Subsequently, the final 20% of SP filtered molecules were taken for the extra precession (XP) docking [407], and only one pose per ligand was retained. Finally, 22 hits were chosen based on their Glide-XP docking score ≥ -10.0 kcal/mol.

6.2.3. Pharmacological properties and toxicity profiles of hits compounds

The physicochemical properties were predicted by employing the QikProp module of the Schrodinger suite. The ADME profiles were predicted using reliable online servers, including SwissADME [408] and pkCSM [409], while the toxicity profiles were obtained from the ProTox-II[410] and ADMETlab 2.0 [411] web servers. After that, the top 10 selected compounds (based on ADMET profiles) were taken for short molecular dynamics simulation (100 ns) for further structure minimization and pose refinement. Finally, we decided on six compounds for long replica MD simulations based on the highest affinity from the previous stage.

6.2.4. Density function theory and receptor-ligand based e-pharmacophore analysis

To analyze the electronic effects, including structure, molecular orbital, and energy states of every atom of the selected six hits, a quantum chemistry method, Density Functional Theory (DFT), was applied using the Jaguar program. Becke's three-parameter exchange potential[412] and Lee-Yang-Parr correlation functional theory[413] (B3LYP) with the basis set of 6-13G* was used for the energy computation. The PBF (Poisson Boltzmann Finite) implicit solvation model was employed for the single-point energy calculations. The molecular orbital surfaces, atomic electrostatic potential (ESP), and molecular electrostatic potential (MESP) were used to compute the highest occupied molecular orbitals (HOMO) and lowest unoccupied molecular orbitals (LUMO) energies for all five hits. All these descriptors were calculated based on the electron density map. Furthermore, the difference in HOMO and LUMO energies was calculated to identify the electronic excitation energy required for analyzing the stability and reactivity of the receptor-ligand complex.

Further, we used the Glide energy descriptors' information of these top six lead phytochemicals for e-pharmacophore hypothesis generation. The six in-built pharmacophore features were quantified with the Glide XP's energy score, and the top-ranked features were selected to build the pharmacophore. In e-pharmacophores, hydrogen bond donors (D) and acceptors (A) were defined as vectors directed to the donor and acceptor positions at the receptor's binding site.

6.2.5. Molecular dynamics simulations and trajectory analysis

All the selected six complexes were taken for further long replica MD simulations (2x500 ns) with two different initial velocities using AMBER18 [255] under the periodic boundary condition. The proteins were described by the Amber ff14SB force field [257] and small molecules by the updated generalized Amber force field (GAFF2) [256]. The water molecules were taken as solvent at a buffer distance of 10.0 Å in all directions from the solute, and parameters were updated by the TIP3P force field [414]. An appropriate number of ions were added to neutralize the system, and the SHAKE algorithm [262] was applied to fix all covalent bond lengths, including the hydrogen atoms. The long-range electrostatic interaction was computed using the particle-mesh Ewald (PME) [263] method at a non-bonded cut-off distance of 10.0 Å. We used a 2.0 fs time step for each MD run, steepest descent (500 cycles), and conjugate gradient (500 cycles) algorithms for system energy minimization. During the simulation, the Langevin thermostat[316] and Berendsen Barostat[265] maintained the system temperature (300 K) and pressure (1 bar), respectively. Finally, we proceeded with the (2x500 ns) production simulation run at the isothermal-isobaric ensemble (NPT). The MD trajectories were collected every 5000 steps for later analysis, and overall, 100,000 snapshots were generated from two replica runs. All the post-simulation calculations were done using the *Cpptraj* module [266] of Amber18.

6.2.6. Ligand binding affinity calculation and energy decomposition

The molecular mechanics/Poisson–Boltzmann (generalized Born) surface area (MM/PBSA)[272,273,372,375,415] method was employed to estimate the protein-ligand binding free energy. Here, we used 2000 snapshots extracted from the last 400 (2X200) ns simulations with an interval of 200 ps. We have already discussed this method in our earlier protein-ligand studies [274,275,277,322,324,416–419]. Briefly, the method is described as follows:

$$\Delta G_{bind} = \Delta H - T\Delta S \qquad [6.1]$$

$$\Delta G_{bind} \approx \Delta E_{ele} + \Delta E_{vdW} + \Delta G_{pol} + \Delta G_{nonpol} - T\Delta S \quad [6.2]$$

where ΔE_{ele}, ΔE_{vdW}, ΔG_{pol}, ΔG_{nonpol}, and $-T\Delta S$ represent the electrostatic, van der Waals interactions, polar solvation, non-polar solvation free energy, and configurational entropy, respectively. The entropic contribution was calculated using normal-mode analysis. Moreover, we also computed the per-residue-based energy decomposition using the MM/GBSA scheme.

6.3. Results and Discussions

6.3.1. Systematic virtual screening of phytochemical dataset

The large phytochemical dataset was first screened with the QikProp module and Lipinski's rule of five filters, in which a total of 44729 compounds passed for the next HTVS docking module. The HTVS module passed 10000 phytochemicals for the SP docking protocol. The SP docking program filtered around 881 compounds and proceeded with the extra precision (XP) docking program for more accurate screening. Finally, we chose 22 compounds (XP-docking score ≤ -10.0 kcal/mol) from the XP docking program and listed them in **Appendix D, Table D1**. Epicatechin showed the lowest XP-docking score (~ -10.9 kcal/mol).

6.3.2. Predicting the pharmacological and toxicity properties of hit compounds

Further, to evaluate the potential of these selected 22 hits to be developed as a human drug in the future, we predicted their pharmacological profiles, including ADME using the widely used web server SwissADME, ADMETlab, and the QikProp module. All the predicted properties were listed in **Appendix D, Table D1, and Table D2**. As suggested by **Appendix D, Table D3** that all the small molecules have a molecular weight below 500 g/mol, indicating these are better absorptions than higher molecular weight[420]. All molecules were satisfied with the Lipinski rule of five except Myricetin (H-bond Donor ≤ 5). We also predicted the octanol/water partition coefficient (logP) in the acceptable ranges, suggesting that these molecules have good permeability to cross the cell membrane. Similarly, all the hits have a molecular refractive index in the excellent acceptable range.

Moreover, we calculated the aqueous solubility (logS), human colon adenocarcinoma (Caco-2) cell lines permeability[421], IC_{50} value for blockage of HERG K+ channels

(logHERG), Madin-Darby canine kidney (MDCK) cells permeability and percentage (%) of human oral absorption. Wang et.al[421] suggested that the Caco-2 cell permeability is a well-established *in-silico* method to screen oral absorption and the transport mechanisms of drugs. Only four phytochemicals, including Myricetin, Morin, Ellagic acid, and STOCK1N-37646, were not accepted due to the low value of Caco-2 cells permeability. Another four molecules, namely, STOCK1N-72541, STOCK1N-76211, STOCK1N-98539, and Quercetin, showed high IC_{50} values for HERG K^+ channels. Keeping in mind the ADME, as mentioned earlier, we found that 13 hits excellently fulfilled the minimum criteria.

Moreover, we predicted the toxicity properties of each hit using widely used online ProTox-II and ADMETlab 2.0 servers. To examine each candidate molecule's safety characteristics, we computationally predicted various toxicology properties, including hepatotoxicity, carcinogenicity, mutagenicity, cytotoxicity, LD_{50} (lethal dose 50) values, and their toxicity class, see **Appendix D, Table D4**. It is evident from **Table D4** that all molecules passed the hepatotoxicity and cytotoxicity. Among all, Myricetin and Ellagic acid showed carcinogenicity. In addition, Myricetin, Lucidin, Cordycepin, STOCK1N-31772, STOCK1N-71990, STOCK1N-72541, Quercetin, and STOCK1N-68737 displayed mutagenicity. All the phytochemicals showed LD_{50} in the acceptable range (LD50 = < 300 mg/kg) except Myricetin, STOCK1N-71990, Quercetin, and STOCK1N-68642.

Moreover, the predicted toxicity class ranges from 3 to 6, where a higher number indicates low toxicity. Among all, epicatechin was found to be the best compound belonging to the class of 6. Therefore, we extensively reviewed ADMET profiles and selected the top 10 molecules that had a better probability of crossing the biological barrier in the ADME and toxicology study. The binding interaction pattern of these ten molecules is shown in **Appendix D, Figure D2**.

6.3.3. Local structure and binding pose refinement

After the systematic virtual screening and ADMET profile prediction, we proceeded with the top 10 molecules complexed with JAK1 for further local structure and binding pose refinement using a 100 ns long molecular dynamics simulation. Based on the structural stability of the binding pocket and free energy of each ligand, we further screened six lead

molecules (namely Epicatechin, Leonurine, STOCK1N-68642, STOCK1N-82656, STOCK1N-85809, and STOCK1N-98615) out of 10 hits, which have lower energy (higher affinity) ≥ -23.0 kcal/mol (see **Appendix D, Table D5**).

6.3.4. Ligand chemical structural similarity and clustering

We performed the ligand clustering and chemical structural similarity (Tanimoto coefficient) of these selected six lead molecules with the FDA-approved JAK kinase inhibitor (Baricitinib, Itacitinib, Filgotinib, and Upadacitinib) to know the atom pair and maximum common substructure (MCS)[422] similarity algorithm among them. We used the online webserver ChemMine Tools[423] to predict the Tanimoto coefficient and ligand clustering. The Tanimoto coefficient ranges from 0 to 1, close to 1.0, representing higher similarity than lower ones. The similarity score based on the MCS Tanimoto coefficient of the best-screened compounds has been added in **Appendix D, Figure D3**. Most lead compounds show an excellent range of dissimilarity among others and FDA-approved drugs. The Tofacitinib and STOCK1N-68642 offer a moderately high degree of similarity with the Tanimoto coefficient of 0.6, but the other molecules show a lower value of the Tanimoto coefficient (0.09 to 0.31).

The clustering of these molecules was performed using the multidimensional scaling (MDS)[424] algorithm in the ChemMine interface based on their structural and physicochemical similarities. The dissimilarity analysis was performed with a similarity cut-off of 0.4. A similarity matrix calculates the distance matrices of each substructure based on the common and unique features observed among all-against-all pairs using the Tanimoto coefficient. The predicted dissimilarity matrix-like scatter plot is shown in **Appendix D, Figure D4**. It is evident from **Figure D4** that lead phytochemicals are segregated in scatter plots, and most of these compounds show high dissimilarity with the FDA-approved JAK inhibitors except STOCK1N-68642. Therefore, all the studied phytochemicals are structurally unique compared to available drugs.

6.3.5. Density functional theory (DFT) and e-pharmacophore analysis

The calculated DFT profiles of six lead phytochemicals are shown in **Table 6.1** and **Appendix D, Figure D5**. Here, we described the molecular orbital properties of the lead phytochemicals using five different molecular descriptors, including electrostatic potential

(ESP), the energy of the highest occupied molecular orbit (HOMO), the energy of the lowest unoccupied molecular orbit (LUMO), the energy gap between HOMO and LUMO orbit (HOMO-LUMO gap) and the single point Poisson-Boltzmann Finite (PBF) solvation energy. HOMO characterizes the ligand's ability to donate an electron, whereas LUMO refers to the ability of the ligand to accept an electron. Low values of LUMO indicate the higher propensity of the molecule to accept electrons, whereas higher values of HOMO dictate the increased tendency to donate electrons to unoccupied molecular orbitals. The HOMO-LUMO energy gap signifies the compound stability index and its chemical activities and bioactive properties. The molecule with the lower gap is highly active; the higher energy gap refers to the kinetic instability of the molecule.

A higher energy value of HOMO (STOCK1N-68642 = -6.37 eV, STOCK1N82656 = -6.27 eV) is associated with the better electron-donating ability and is related to ionization potential of the molecule, whereas a lower LUMO (Epicatechin = -0.26 eV, STOCK1N98615 = -0.41 eV, STOCK1N68642 = -0.94 eV) is directly related to electron affinity and its ability to accept electrons. Epicatechin (5.60 eV), STOCK1N68642 (5.43 eV), and STOCK1N98615 (5.66 eV) show similar types of HOMO-LUMO energy gaps, and lower for STOCK1N82656 (0.86 eV). The lower energy gap for STOCK1N82656 denotes better stability of the molecule with better bioactive properties. Thus, we conclude that the HOMO-LUMO energy of all the lead molecules indicates that they are capable of binding to JAK1.

Table 6.1: The six lead molecules' electrostatic potential (ESP), solvation energy, HOMO, and LUMO energy were generated by density functional theory (DFT) analysis in the Jaguar module of the Schrodinger suite.

Top Lead	ESP min (kcal/mol)	HOMO (eV)	LUMO (eV)	HOMO–LUMO gap (eV)	Solvation energy (kcal/mol)
Epicatechin	-45.16	-5.86	-0.26	5.60	-30.00
Leonurine	-57.57	-5.59	-1.35	4.24	-28.86
STOCK1N-68642	-58.32	-6.37	-0.94	5.43	-34.13
STOCK1N-82656	-49.10	-6.27	-5.41	0.86	-31.71
STOCK1N-85809	-66.85	-6.07	-1.22	4.85	-37.54
STOCK1N-98615	-117.72	-6.07	-0.41	5.66	-58.97

Furthermore, we created the receptor structure-based e-pharmacophore model using a single JAK1-phytochemical complex (shown in **Appendix D, Figure D6**). This type of receptor structural information is crucial for specific drug design[425]. **Figure D6** suggests that all six phytochemicals show a distinct pharmacophore model, which agrees with the Tanimoto coefficient (**Figure D3**) and multidimensional substructure similarity analysis (**Figure D4**). In summary, the pharmacophore models provide more profound insight into the structural features of phytochemical as an inhibitor.

6.3.6. Molecular dynamics simulation and binding conformational stability analysis

Molecular dynamics simulation is a robust and versatile approach to analyzing the internal interaction and structural stability between the small molecules and their respective target receptor protein under the specific physiological environment condition. Herein, the top six lead molecules were subjected to 2 x 500 ns MD simulations. Each simulation's stability was monitored by measuring the root mean square deviation (RMSD) of backbone atoms and plotted along the simulation time of 500 ns (see **Appendix D, Figure D7**). The RMSD results from the two independent replicas show a similar atomic deviation, except for Leonurine and STOCK1N-85809 complexes. The RMSD values of Epicatechin, STOCK1N-68642 and STOCK1N-85909 show the minimum fluctuation and are maintained throughout the simulation (run1 and run2). For Leonurine, STOCK1N-82656, and STOCK1N-85809 bound complexes, the RMSD value continuously increases for the initial 200 ns, and after that, the plateau is reached.

Further, we also calculated the RMSD distribution of the combined trajectory of two replicas, shown in **Figure 6.2A**. It is clear from **Figure 6.2A** that the backbone RMSD distribution of epicatechin, STOCK1N-68642, STOCK1N-82656, and STOCK1N-98615 show restricted fluctuations and the most probable peak is found at ~2.06, 1.92, 2.64, and 1.98 Å, respectively. The Leonurine/JAK1 complex showed two probable RMSD peaks at ~2.43 and ~3.08 Å. In the case of STOCK1N-85809/JAK1, the distribution is characterized by a sharp peak at ~1.56 Å and hump at ~2.03 Å. In our previous study[325], we reported that the backbone RMSD of JAK1 decreases after inhibitors binding. A similar trend is observed in the current study also.

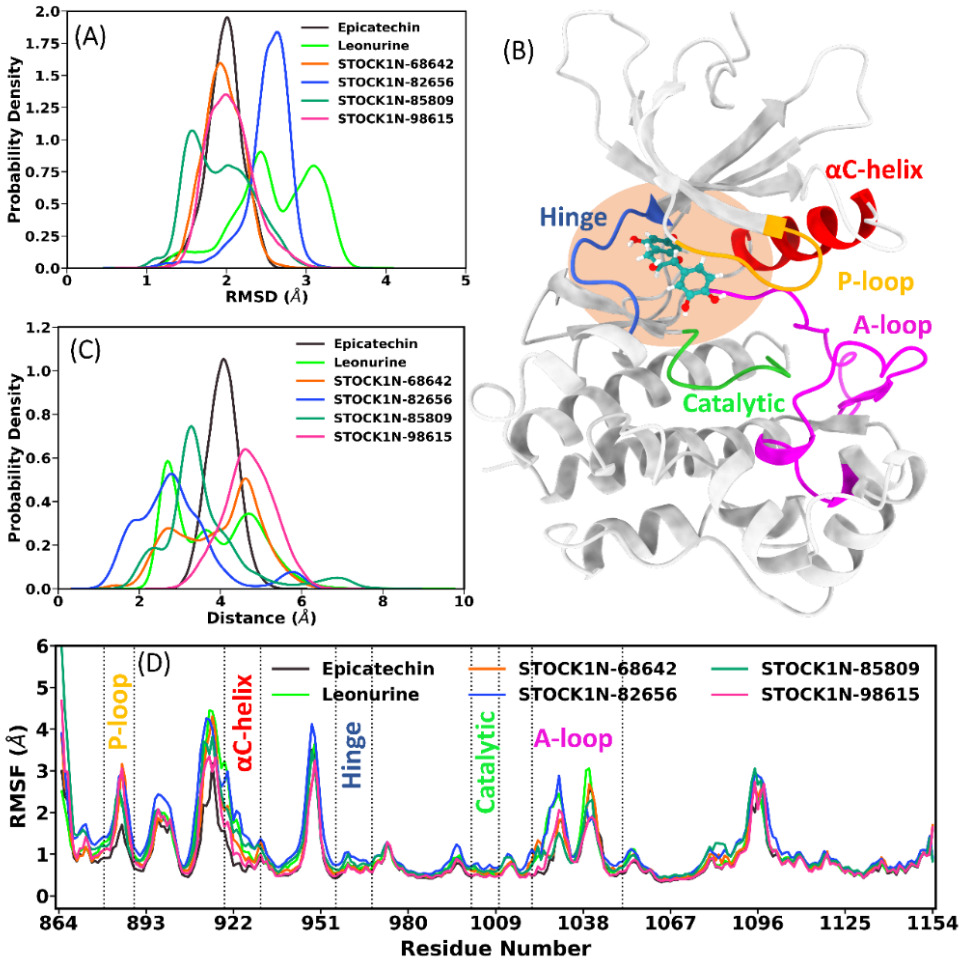


Figure 6.2: (A) The probability of density function (PDF) plots of the backbone RMSD of target receptor protein JAK1 complexed with the phytochemicals. (B) The three-dimensional crystal structure of receptor protein, JAK1, highlighted the crucial binding residues, including P-loop, Hinge, αC-helix, catalytic, and activation loop (A-loop). (C) The probability of the center of the mass distance between the ATP-binding pocket and phytochemicals during two replica MD simulations. (D) The residue-wise C_α-atom RMSF of JAK1.

The movement of each small molecule was monitored by measuring the center of mass (CoM) distance between the ATP-binding pocket residues (see **Figure 5.2B**, orange shaded area) and the heavy atoms of the small molecule. The time evolution of each distance is shown in **Appendix D (Figure D8),** and the probability density of the combined trajectory is shown in **Figure 6.2C**. **Figure 6.2C** suggests that Epicatechin and STOCK1N-98615 molecules maintain a stable interaction with a most probable peak at ~4.09 and 4.62 Å. On the other hand, the rest of the four molecules show a widely spreading distance map with

more than one probable distance peak. Interestingly, all the molecules show their most probable distance of less than 5.0 Å, suggesting that these phytochemicals are the good binding partner of JAK1. Besides the distance analysis, we also measured the heavy atoms RMSD throughout the simulation time and the probability density distribution (see **Appendix D, Figure D9A**). **Figure D9A** suggests that Epicatechin and STOCK1N-68642 have two probable RMSD peaks and probable high peaks at ~1.41 and ~1.54 Å and relatively low probable heights at ~0.26 and ~2.11 Å, respectively. The most probable peak of Leonurine, STOCK1N-82656, and STOCK1N-85809 are located at ~1.41, ~1.31, and ~2.00 Å, respectively. Their hump peaks are at ~2.14, ~2.31, and ~0.78 Å, respectively.

Additionally, we determined the conformational dynamics of JAK1, which get affected by the inhibitor's binding. We plotted the backbone-atoms RMSD probability density distribution of critical areas, namely, the Phosphate binding loop (P-loop), hinge area, catalytic loop, activation segment, DFG motif, catalytic spines, and regulatory spines (see **Figure 6.2B** and **Figure D9 B-H**). **Figure D9B** indicates that P-loop is highly dynamic and shows conformational variability after ligand binding. The structurally conserved hinge area is stable and rigid throughout the simulation, but STOCK1N-85809 shows the bimodal distribution (see **Figure D9C**). The catalytic site shows conformational variability for STOCK1N-68642, STOCK1N-85809, and STCOK1N-98615 complexes (see **Figure D9D**). The A-loop RMSD distribution suggests higher A-loop stability for the Epicatechin/JAK1 complex. Other complexes show widely spread RMSD distributions, which means that the phytochemicals binding increases A-loop flexibility (see **Figure D9E**). Except for Epicatechin and STOCK1N-98615, the DFG motif shows conformational variability (see **Figure D9F**). The catalytic spine (C-spine) is highly rigid throughout the simulation time except for the STOCK1N-85809 complex (see **Figure D9G**). The regulatory spines of Epicatechin and STOCK1N-98615 are well aligned, and other complexes show the rearrangement in alignment (see **Figure D9H**).

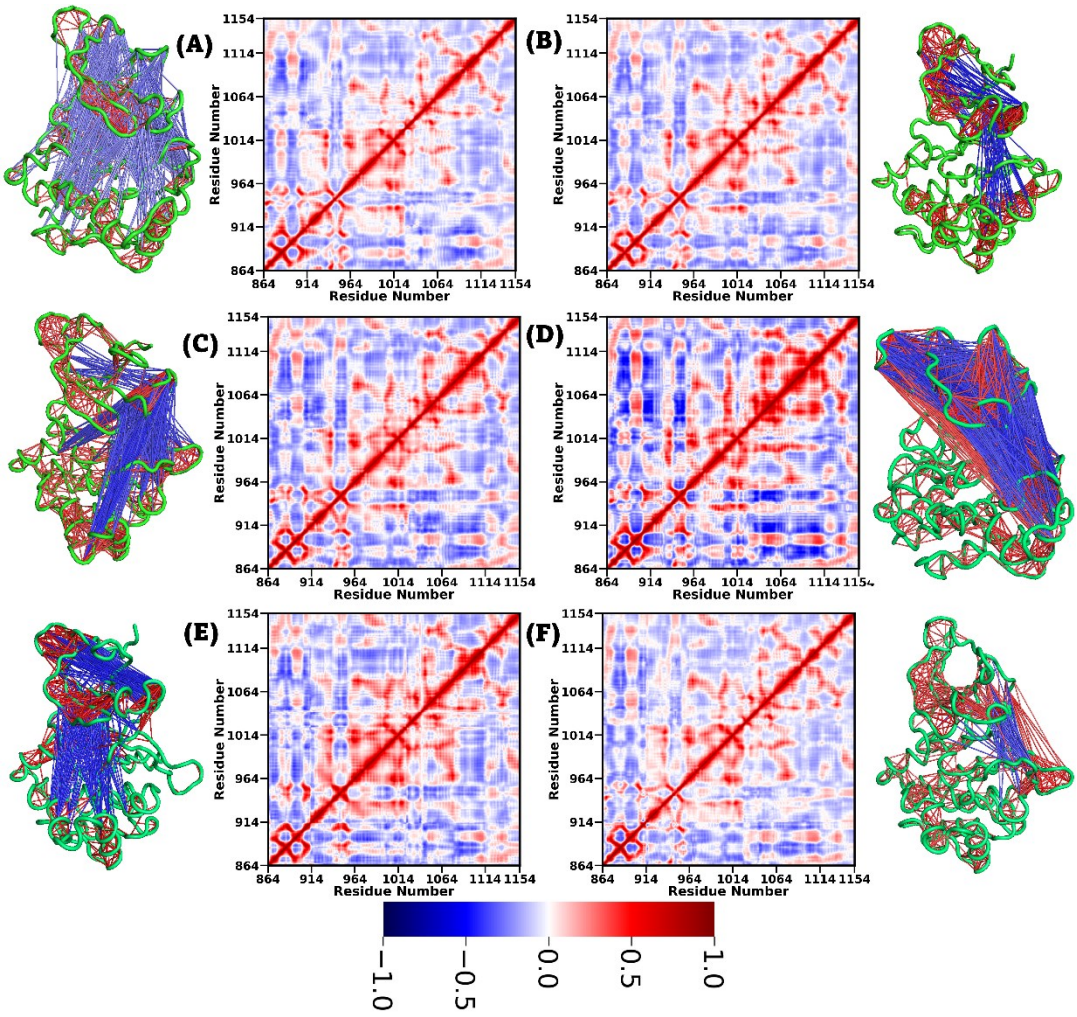


Figure 6.3: The dynamic cross-correlation map of JAK1 receptor (Cα-atoms): **(A)** JAK1/Epicatechin, **(B)** JAK1/Leonurine, **(C)** JAK1/STOCK1N-68642, **(D)** JAK1/STOCK1N-82656, **(E)** JAK1/STOCK1N-85809, and **(F)** JAK1/STOCK1N-98615. The blue (-1 to 0) and red (1 to 0) colors correspond to the anti-correlated and correlated motions of the atoms, and the white color suggests no correlation map among the residues. The red and blue lines in the 3D structures represent the strong correlation and anticorrelation motion, respectively.

Subsequently, we investigated the atomic level flexibility of backbone C_α-atoms of each complex system by calculating the root-mean-square fluctuations (RMSF), as shown in **Figure 6.2D**. It is clear from **Figure 6.2D** that all six complexes show similar RMS fluctuation patterns. The terminal residues and loop (P-loop and A-loop) areas display high RMSF values. The ligand-binding site regions, including catalytic and hinge areas, show a lower RMSF. These results are also supported by our previous JAK study[325] and a study

by K. Sanachai and co-workers[390]. Overall, STOCK1N-82656 shows relatively higher values of RMSF.

Subsequently, we measured the SASA of each complex and the probability density distribution shown in **Figure D10**. The most probable peak of Epicatechin, Leonurine, STOCK1N-68642, STOCK1N-82656, STOCK1N-85809, and STOCK1N-98615 are at ~13705, ~14106, ~13382, ~14135, ~13875, and ~13670 Å^2, respectively. Overall, these results suggest that the protein compactness increases and SASA values decrease compared to apo JAK1 protein[325].

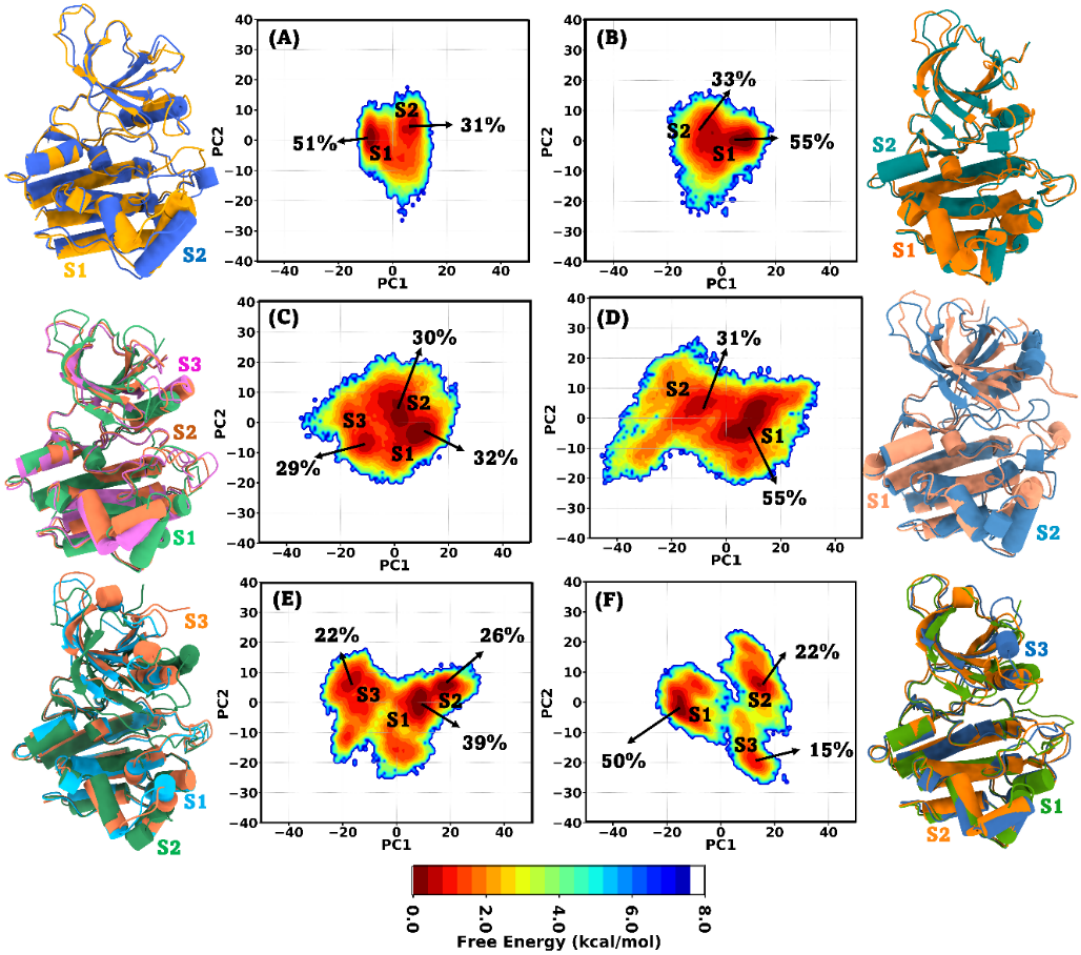


Figure 6.4: The two-dimensional free energy landscapes generated by projecting the first two principal components, PC1, and PC2, for **(A)** JAK1/Epicatechin, **(B)** JAK1/Leonurine, **(C)** JAK1/STOCK1N-68642, **(D)** JAK1/STOCK1N-82656, **(E)** JAK1/STOCK1N-85809, and **(F)** JAK1/STOCK1N-98615. The representative structures of corresponding minima compared to the X-ray structure are shown on the left panel.

6.3.7. Essential dynamics of protein and Free energy landscape

Here, we performed a dynamic cross-correlation matrix (DCCM) of residual backbone Cα-atoms to explore protein's internal collective domain motions. The structural 3D maps of these proteins were constructed by the Bio3D[426] package. These maps help us identify the group of pairs of residues contributing to a conformational change in the protein. The DCCM plot and the corresponding map on the 3D structures are shown in **Figure 6.3(A-F)**. The correlated motions are shown in red, while the anti-correlated movements are in blue. In our previous study[325], we discussed the importance of residual collective motions. We also reported that the apo system shows anticorrelation motion, and the inhibitors bound complexes show moderate correlation/anticorrelation movements. This study also found a similar trend in the DCCM of phytochemicals bound JAK1 complex. Overall, it is evident that STOCK1N-82656 exhibits a relatively high anti-correlated tendency consistent with RMSF analyses (**Figure 6.2D**).

We performed principal component analysis (PCA) on the last 500 (2 X 250) ns trajectory to further characterize the structural variation and internal protein dynamics. The corresponding eigenvalues of each eigenvector are shown in **Appendix D, Figure D11**. We extensively analyzed each principal component (PC). We found that the first 10 PC accounts for ~71.48 to ~93.75% of the overall motions, and the first two (PC1 and PC2) components capture ~54 to 86.12% of the total motions and collective localized fluctuations in the protein. Subsequently, we considered these two (PC1 and PC2) eigenvectors as reaction coordinates and generated the free energy landscapes (at 300 K) for six complexes, shown in **Figure 6.4**. It is evident from **Figure 6.4** that the binding of the inhibitor to JAK1 adopts distinct conformation and conformational space sampled by all complexes. The conformational space shows multiple dispersed basins, and each basin corresponds to a most probable minimum energy conformation. The Epicatechin, Leonurine, and STOCK1N-82656 bound complexes exhibited two probable conformations: global free energy minimum accounts for ~51, ~55, and ~55%, and local energy minimum conformations account for ~31 to 33%, respectively (see **Figure 6.4A-B, D**).

Additionally, we extracted these probable free energy minimum structures (see **Figure 6.4**), and significant changes are seen in the αC-helix and A-loop regions. In contrast, STOCK1N-68642, STOCK1N-85809, and STOCK1N-98615 show the three probable energy basins, and the primary basin accounts for ~32, ~39, and ~50% of the observed conformations, respectively (see **Figure 6.4C, E-F**). Overall, these results suggest that the different complexes visited the different conformational spaces, and the inhibitor binding induces the structural rearrangement in receptor protein.

6.3.8. The conformational landscape of αC-helix and its relationship to ligand binding

In understanding protein kinase, computational investigations are powerful complements to experiments in the recent era due to increasing advancements in computing power. Here we performed extensive atomistic molecular dynamics with a total aggregate sampling time of 1.0 μs. Previously reported studies suggested that the kinase conformational landscape can be characterized by the DFG motif and αC-helix states' plasticity. Recently, Chan and his coworkers[427] showed how the structural rearrangement controlled the neurotransmitter serotonin from the synaptic cleft by the serotonin transporter.

To explore the conformational landscape of the DFG motif and αC-helix, we considered a characteristic salt-bridge distance Lys^{908} (β3) - Glu^{925} (αC) as one of the reaction coordinates (RC1) to define the αC-in (R < 4.5 Å) and αC-out (R > 4.5 Å) states, suggested by Sultan et al.[387] and Tsai et al. [326]. We also accounted for pseudo torsional angle formed by Ile^{1019}, Gly^{1020}, Asp^{1021}, and Phe^{1022}, first proposed by Möbitz[386] to define the DFG-*in* ($\zeta < 120^{o}$) and DFG-*out* ($\zeta \geq 120^{o}$) states. The free energy landscape (FEL) as a function of R and ζ for all six complexes is shown in **Figure 6.5**. The vertical line at 4.5 Å divide the conformational landscape into two areas: αC-helix in and αC-helix out. It is evident from **Figure 6.5(A, F)** that Epicatechin and STOCK1N-98615 bound αC-helix of JAK1 visits a smaller conformational space, indicating the rigidity in term of DFG-*in* and αC-*in* conformations. In the case of rest four complexes, the αC-helix of JAK1 visits a larger conformational space, indicating its overall augmented flexibility in the protein structure, see **Figure 6.5B-E**. To examine the various low energy conformations in the FEL, we extracted the structures form the minimum energy basins (Figure 5). We rigorously

explored the each low energy conformations and found that all conformation are in DFG-in states (ζ < 120°).

Leonurine displays (**Figure 6.5B**) a global minimum structure at R around 3.31 Å (αC-in) and ζ of 60-45 (DFG-in) and an extended local minimum centered at R ~11.41 Å (αC-out) and ζ of 15-25 (DFG-in). A similar type of αC-helix transition (αC-in to αC-out) is noticed in the case of STOCK1N-68642 (**Figure 6.5C**), STOCK1N-82656 (**Figure 6.5D**), and STOCK1N-85809 (**Figure 6.5E**) bound JAK1. A structurally similar inactive state has also been observed in Abl kinase[328] and epidermal growth factor receptor kinase (EGFR)[428]. Interestingly, a shallow energy barrier between these low energy structures (αC-in and αC-out) is high enough, about 3-5 kcal/mol based on the FEL, indicating that these conformations are distinct and can be a target for allosteric inhibitors.

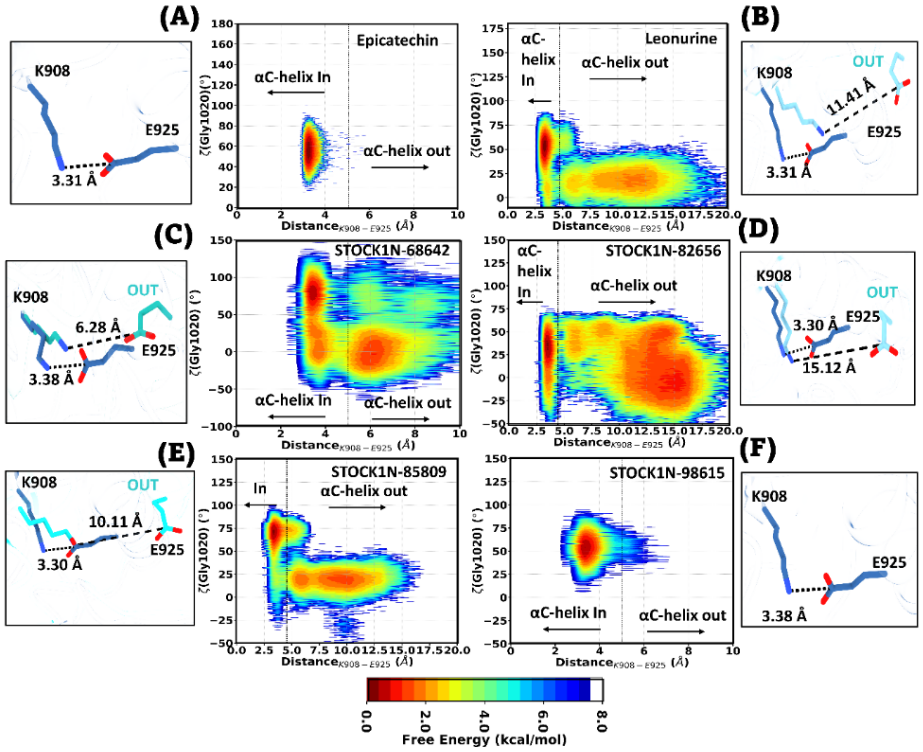


Figure 6.5: The conformational landscape of JAK1 complexed with **(A)** Epicatechin, **(B)** Leonurine, **(C)** STOCK1N-68642, **(D)** STOCK1N-82656, **(E)** STOCK1N-85809, and **(F)** STOCK1N-98615. Free energy landscape (FES) as a function of the pseudo torsion ζ(Gly1020) and distance R between Lys[908]:NZ and Glu[925]:CD. ζ(Gly1020) is defined using the CA atoms of Ile[1019], Gly[1020], Asp[1021], and Phe[1022]. Vertical lines are plotted at 4.5 Å to define cut-off values for αC-helix in/out structure. Representative snapshots represented the FES minima. Lys[908], Glu[925], Asp[1021], and Phe[1022] are shown.

Moreover, we investigated the secondary structural disorders in terms of the probability of helical content of all the residues of the αC-helix for all six complexes, see **Figure 6.6**. It is evident from **Figure 6.6** that Epicatechin, STOCK1N-68642, and STOCK1N-98615 bound states show a high degree of integration in the αC-helix. On the other hand, Leonurine and STOCK1N-85809 bound states show an intermediate degree of disorder (**Figure 6.6**) and break the conserved catalytic Lys^{908}-Glu^{925} salt bridge (**Figure 6.5B, E**). The STOCK1N-82656 bound state shows a high degree of helical disintegration, especially in the Ala^{920}-Asp^{921}-Leu^{922} and Ile^{926}-Glu^{927}-Ile^{928}-Leu^{929}-Arg^{930} of αC-helix. Overall, these results suggested that the JAK1 structure shows massive conformational plasticity in the αC-helix upon competitive inhibitor binding, causing its helix to unfold and outward movement leading to inactive conformations. These findings are also consistent with previously reported kinase studies, including human epidermal growth factor receptor (EGFR)[429] and with no-lysine kinase (WNK)[416] and eukaryotic protein kinases, Brassinosteroid-insensitive-1 (BRI1) and BRI1-associated kinase 1 (BAK1)[430].

6.3.9. Binding free energy analysis of phytochemicals against JAK1

Molecular simulations, in conjunction with free energy estimations, are a good approach to exploring the ligand-binding stability and their interaction strength. To elucidate the binding pattern of phytochemicals against JAK1, we calculated the various individual contribution of free energy, listed in **Table 6.2** and **Appendix D, Table D6**. The intermolecular electrostatic (ΔE_{elec}), van der Waals (ΔE_{vdW}) interactions, and the non-polar solvation free energy ($\Delta G_{non\text{-}polar}$) strengthen the complex formation. In contrast, the polar solvation free energy (ΔG_{polar}) and entropic penalty ($-T\Delta S_{conf}$) opposes the complexation. The average binding affinity was calculated from the replicas of six complexes (Epicatechin, Leonurine, STOCK1N-68642, STOCK1N-82656, STOCK1N-85809, and STOCK1N-98615) also reported in **Table 6.2**.

The predicted total binding affinity values of six complexes range from -13.86 ± 0.93 to -9.67 ± 1.06 kcal/mol. The STOCK1N-68642 (-13.86 ± 0.93 kcal/mol) and Epicatechin (-12.96 ± 0.94 kcal/mol) shows relatively higher binding affinity against JAK1 followed by STOCK1N-98615 (-11.72 ± 0.82 kcal/mol), STOCK1N-85809 (-11.67 ± 1.07 kcal/mol), Leonurine (-11.34 ± 0.91 kcal/mol), and STOCK1N-82656 (-9.67 ± 1.06 kcal/mol). As

shown in **Table 6.2**, the van der Waals contribution for all complexes ranges from -47.76 ± 0.11 to -33.01 ± 0.07 kcal/mol, and the favorable electrostatic interaction (ΔE_{elec}) component of binding free energy is overcompensated by its desolvation polar energy (ΔG_{pol}). For all the complexes, the non-polar solvation energy (ΔG_{np}) is more or less similar, and configurational entropic cost varies from 15.02 to 20.48 kcal/mol, consistent with our previous results[325]. The binding of phytochemicals with JAK1 is mainly driven by the van der Waals interactions[276,279,431]. The most negative electrostatic energy contribution (ΔE_{elec} = -39.04 ± 0.11 kcal/mol) and medium range of disfavourable polar group contributions (ΔG_{pol} = 41.06 ± 0.26 kcal/mol) of STOCK1N-68642 contributes for having higher binding affinity. Overall, it suggests that among the six lead molecules, STOCK1N-68642 binds strongly with the JAK1 although Epicatechin exhibits the highest XP-docking score during the virtual screening.

6.3.9. Key residues involved in the recognition process

The binding free energies were further decomposed into their residual contributions to generate an energy spectrum at residue levels, as shown in **Appendix D, D12**. Note that the phytochemicals binding pattern with JAK1 in two independent simulations are not significantly different, and the overall interaction spectra of all complexes are pretty similar. In **Figure D12**, positive and negative interaction energy destabilizes and stabilizes phytochemicals in the ATP-binding pocket. Herein, we only present the binding pocket residues, including P-loop, αC-helix, hinge, catalytic, and the initial part A-loop (near DFG motif) residues.

Table 6.2: Average energetic components of the binding free energy for phytochemicals bound to the JAK1. All values are given in kcal/mol.

	Epicatechin	**Leonurine**	**STOCK1N-68642**	**STOCK1N-82656**	**STOCK1N-85809**	**STOCK1N-98615**
ΔE_{vdW}	**-33.01 ± 0.07**	**-41.75 ± 0.11**	**-39.04 ± 0.11**	**-36.80 ± 0.11**	**-47.76 ± 0.11**	**-40.77 ± 0.11**
ΔE_{elec}	**-36.26 ± 0.26**	**-32.04 ± 0.31**	**-32.25 ± 0.25**	**-25.84 ± 0.31**	**-32.98 ± 0.29**	**-25.08 ± 0.28**
ΔG_{pol}	**41.93 ± 0.22**	**47.20 ± 0.17**	**41.06 ± 0.26**	**40.67 ± 0.25**	**53.58 ± 0.27**	**39.10 ± 0.21**
ΔG_{np}	**-3.56 ± 0**	**-4.82 ± 0**	**-3.89 ± 0**	**-4.85 ± 0**	**-4.61 ± 0**	**-4.29 ± 0**
$-T\Delta S_{\text{conf}}$	**17.93 ± 0.88**	**20.07 ± 0.84**	**20.26 ± 0.85**	**17.15 ± 0.98**	**20.09 ± 0.99**	**15.54 ± 0.73**
ΔG_{bind}	**-12.96 ± 0.94**	**-11.34 ± 0.91**	**-13.86 ± 0.93**	**-9.67 ± 1.06**	**-11.67 ± 1.07**	**-11.72 ± 0.82**

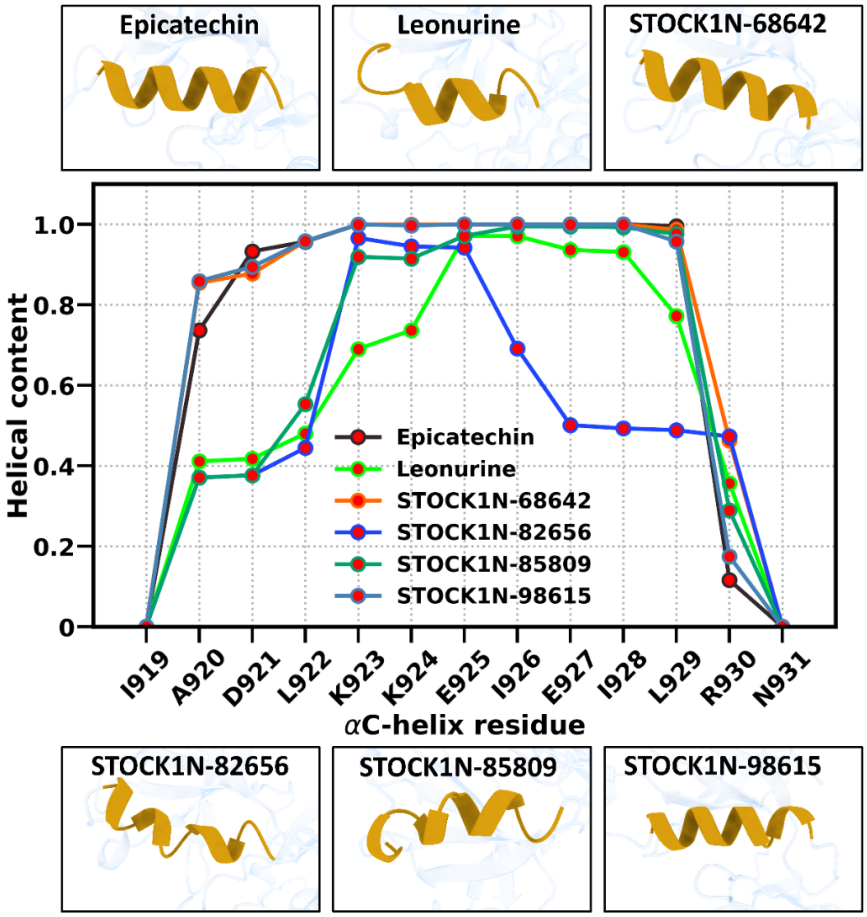


Figure 6.6: The probability of helical content of αC-helix residues calculated in the combined trajectory for all six complexes and their corresponding representative structures.

It is evident from **Figure D12** that the P-loop, hinge, and catalytic areas are mainly involved in the stabilizations of phytochemicals. The hydrophobic residues, including Leu[881], Phe[886], Val[889], Leu[959], and Leu[1010], exhibited an energetic contribution of < -1.0 kcal/mol (see **Figure D12**). The location of residues with high contributing interaction energy for all six complexes is shown in **Figure 6.7**. Interestingly, the hinge region strongly stabilizes all the phytochemicals, mainly by Glu959 and Leu959 residues. The second-generation FDA-approved drugs[325] and Tofacitinib[390] also show similar energy spectra against JAK1. It is noteworthy to mention that the hydrophobic interaction plays a crucial role in phytochemicals and JAK1 complexation, which agrees with the other JAK inhibitor[390,432,433]. Apart from the hydrophobic interactions, the hinge region also takes part in the electrostatic interactions and hydrogen bond formation with the help of Glu[957]

residue. The attractive electrostatic contributions from catalytic (HRD motif) Arg^{1007} residue also helps stabilize the phytochemicals firmly. In the case of STOCK1N-68642, the DFG motif residue Asp^{1021} contributes disfavorably to complexation and destabilizes STOCK1N-68642, which contrasts with other phytochemicals. The Asp^{1021} residue also destabilized[325] the Baricitinib, Itacitinib, and Upadacitinib complexation with JAK1.

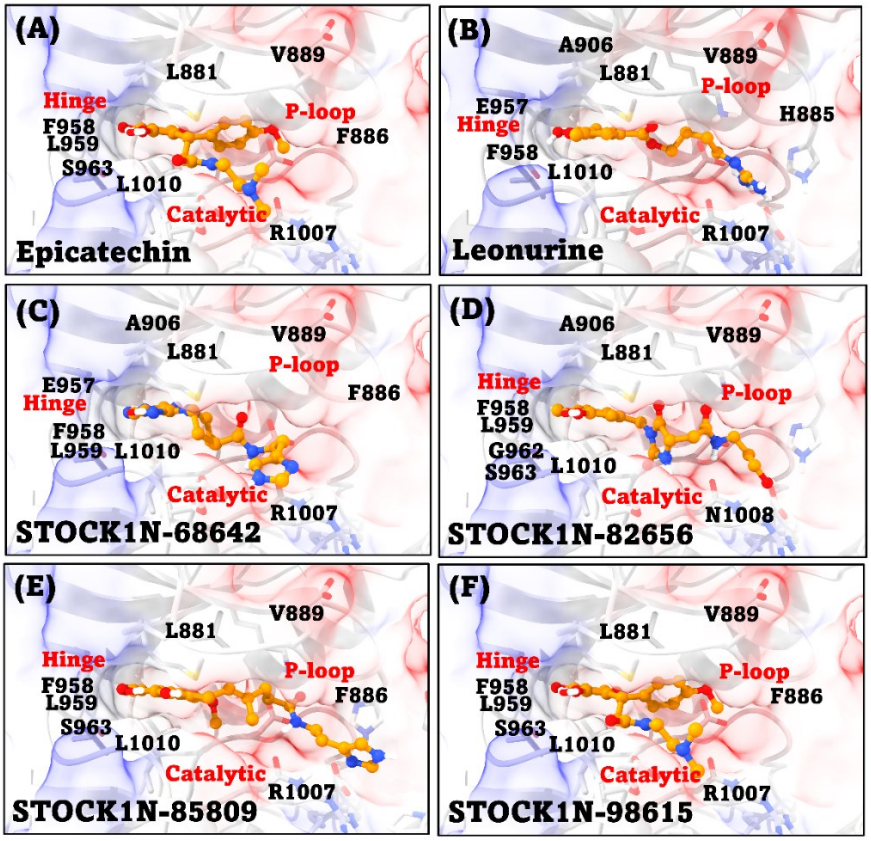


Figure 6.7: The highlighted residues are the crucial residues of JAK1 from the P-loop, hinge, and catalytic area, which take part in the phytochemical binding affinity. The shaded surface represents the electrostatic surface area of JAK near the ATP-competitive inhibitor binding pocket, and phytochemicals are described in the ball and stick model in the binding cavity of JAK1.

6.3.10. Hydrogen bonds pattern between the phytochemicals and JAK1

To further complement the energetic analysis of the phytochemicals with JAK1, hydrogen bonds (HB) formation between the phytochemical and JAK1 was explored. The number of HB for all complexes, along with the simulation time, is shown in **Appendix D (Figure D13)**. The residues involved in HB formation and their respective distance and occupancy

(run1 and run2) are enlisted in **Appendix D (Table D7)** and shown in **Figure 6.8** for all six complexes.

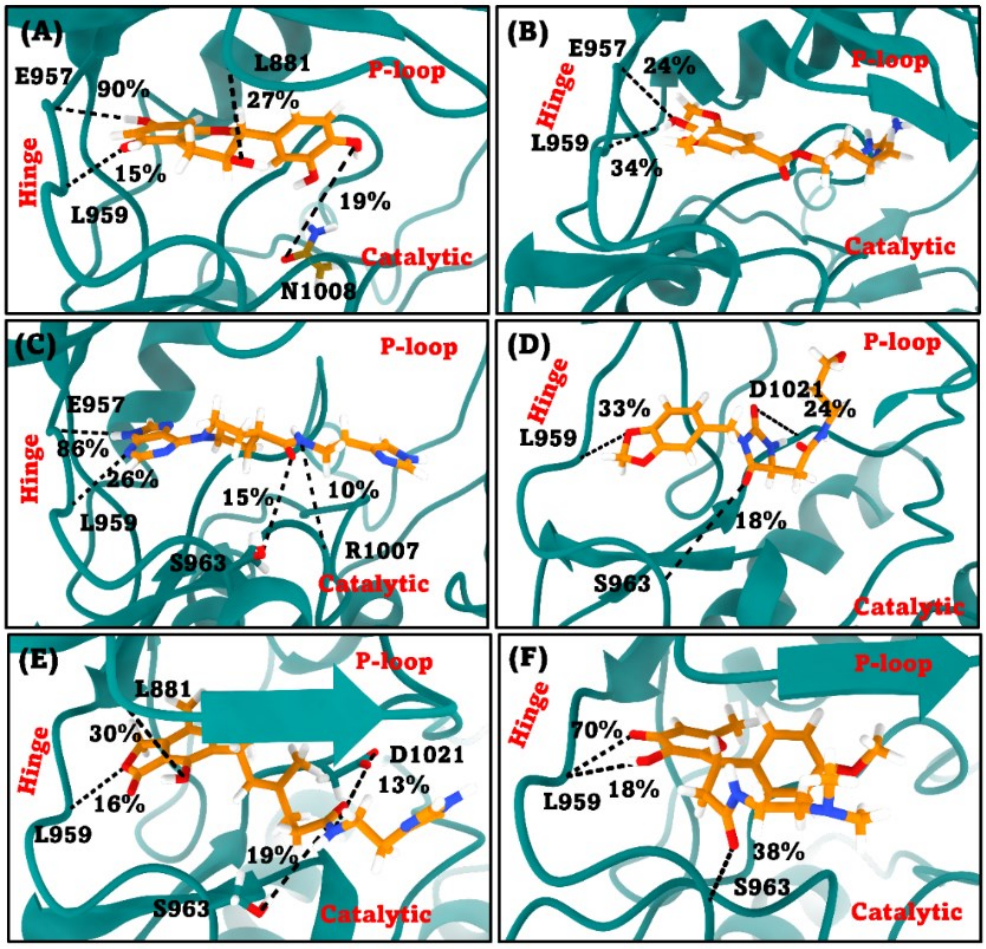


Figure 6.8: Percentage of hydrogen bond occupation with the essential residues in the hinge, P-loop, and catalytic loop region of JAKs. The data were derived from the last 300 ns of the two simulations.

It is evident from **Table 6.4** and **Figure 6.8** that the Glu957 and Leu959 residues form stable and strong hydrogen bonds with the phytochemicals and play an essential role in stabilizing the binding pocket. This finding is consistent with the previous experimental study by Williams et al. [33]. and theoretical studies[325,390]. The HB occupancy of Glu^{957} with Epicatechin (Glu^{957}@O - Lig@O2), Leonurine (Glu^{957}@O - Lig@O2), and STOCK1N-68642 (Glu^{957}@O – Lig@N1) are 90.08, 24.45, and 86.35%, respectively. Similarly, Leu959 residue HB occupancy varies between 15.14 to 70.40%. In the case of the

Epicatechin/JAK1 complex (see **Figure 6.8A**), Leu^{881} (P-loop) and Asn^{1008} (catalytic) form HBs with occupancies of 27 and 19%, respectively. The DFG motif residue Asp^{1021} forms an HB with STOCK1N-82656 (Asp^{1021}@N – Lig@O1) and STOCK1N-85809 (Asp^{1021}@OD1 – Lig@N3) with an occupancy of 24% and 13%, respectively. Interestingly, the catalytic residue Arg^{1007} stabilizes the phytochemicals by forming an HB but destabilizes the pan-JAK Tofacinib[390] and Upadacitinib[325] inhibitors. Another non-conserved residue, Ser^{963}, from the end of the hinge area takes part in the HB network phytochemicals with JAK1. Overall, these findings explain the high binding affinity of the studied phytochemical against JAK1.

6.4. Conclusions

Herein, we performed a structure-based virtual screening of natural products using JAK1 as a target to identify prospective JAK inhibitors for treating neuroinflammation and autoimmune disorders. For this purpose, we adopted a systematic molecular docking protocol using the Schrodinger suite's Virtual Screening Workflow (VSW) scheme. We obtained 22 molecules with more than -10.0 kcal/mol docking energy. Out of these 22 molecules, only ten molecules displayed acceptable pharmacological and pharmacokinetic properties. Then, we performed local structure refinement analysis for these ten complexes using 100 ns MD simulations in conjunction with the free energy calculations. Based on the structural stability and binding enthalpy, we shortlisted six compounds and extended our studies by performing 2 X 500 ns MD simulations for these six ligand-bound complexes. MD simulation results suggested that all six best-screened ligands bind strongly with JAK1 and remain stable inside the ATP-binding pocket during the entire simulation. Based on affinity comparison among these phytochemicals, we found STOCK1N-68642 (-13.86 ± 0.93 kcal/mol) and Epicatechin (-12.96 ± 0.94 kcal/mol) showed higher affinity, followed by STOCK1N-98615, STOCK1N-85809, Leonurine, and STOCK1N-82656. The per-residue-based free energy decomposition and HB analyses suggested that the interactions pattern of these phytochemicals well resembles FDA-approved pan-JAK inhibitors.

Further, we explored the structural dynamics of the JAK1 complexes. The binding pocket RMSD of all complexes was stable, and the RMSF values of those residues were very low, suggesting strong stability inside the binding pocket. The R_g and SASA values indicated that the protein compactness increases, and solvent accessibility decrease in the JAK1 complexes relative to apo-JAK1. The DCCM results showed that overall anti-correlated/correlated motion diminishes after ligand binding. The PCA-based free energy basin clearly showed a conformational rearrangement in the complexes. In addition to PCA, we also performed a 2D FEL analysis, considering hallmark salt bridge (R) and pseudo-torsional angle (ζ) as reaction coordinates. The hallmark salt bridge distance and distorted αC-helix analyses suggested conformational change from active (αC-*in*) to inactive (αC-*out*) due to ligand binding in the ATP-binding pocket of JAK1. As these characteristics arise from structural rearrangements far from the ATP-binding site, they could be used to develop a new class of allosteric inhibitors.

Chapter 7

Chapter 7

Conclusions and Future Perspective

7.1. Overall conclusions

In this doctoral dissertation, I have described the conformational dynamics of Janus kinases implicated in autoimmune disorders and inflammations. I have also discussed the development of potential inhibitors for treating autoimmune diseases. Earlier studies identify numerous molecular targets in the JAK/STAT signaling cascade that could be targeted to discover new therapeutics to control aberrant functioning. The present thesis work has provided insights into the mysterious contributions of non-covalent interactions and the conformational dynamics of JAK kinases due to phosphorylation states and ligand bindings. The outcome of this study explains the molecular targets of inflammation that are well appreciated for designing novel drugs for chronic inflammatory diseases.

Firstly, we have extensively investigated the molecular basis of competitive inhibitors binding of JAK by atomistic molecular dynamics simulations and free energy calculations. Here, we successfully docked the four next-generation JAK inhibitors against JAK1. Our rigorous MD simulation studies suggested that the calculated binding affinity decreases in the order upadacitinib > itacitinib > filgotinib > baricitinib. Due to the increased favorable intermolecular electrostatic contribution, upadacitinib is more selective to JAK1 than the other three inhibitors. We also interpreted the structural information, including the dynamics cross-correlation (DCC) and principal component analyses (PCA) showed that different inhibitor bindings significantly affect the binding site dynamics of JAK1. Furthermore, our studies indicated that the hydrophobic residues and hydrogen bonds from the hinge region (Glu^{957} and Leu^{959}) of JAK1 played an essential role in stabilizing the inhibitors.

After understanding the molecular basis of inhibitor binding mechanisms, we are interested in exploring Janus kinase's conformational plasticity by phosphorylation states and binding of small SOCS protein. We performed an aggregate of 13 μs molecular dynamics simulations on the activation loop's three different phosphorylation (double and single)

states. Results from our simulations show that the single Tyr^{1034} phosphorylation could stabilize the JAK1/SOCS1 complex as well as the flexible activation segment. The phosphate-binding loop (P-loop) shows conformational variability at dual and single phosphorylated states. Principal component analysis and protein structure network (PSN) analysis revealed that the different phosphorylation states and SOCS1 binding induce intermediate inactive conformations of JAK1, which could be a better target for future JAK1 selective drug design. PSN analysis suggests that the com-pY1034 system is stabilized due to higher values of network hubs than the other two complex systems. Moreover, the binding free energy calculations indicate that $pTyr^{1034}$ states show a higher affinity toward SOCS1 than the dual and $pTyr^{1035}$ states.

The active state of kinases resembles, while its inactive state exists in different conformations as it does not have to perform a stringent phosphorylation function. Hence, these distinct conformations can be exploited in the design of inhibitors. The complex conformations of the inhibitor-induced inactive state are an extensive time-scale process; therefore, we used the enhanced Gaussian accelerated molecular dynamic simulations (GaMD) instead of the conventional MD simulations to explore sizeable conformational space. Our results indicate significant conformational variations upon inhibitor binding in the A-loop and αC helix motions. Our studies also reveal that the DFG-*out* inactive conformation is characterized by a close A-loop rearrangement, open catalytic cleft of N and C-lobe, the outward movement of the αC helix, and open P-loop states. Moreover, the outward sign of αC helix impacts the hallmark salt bridge formation of Lys^{882} and Glu^{898} in an inactive conformation. Finally, we compared their binding poses and free energy by the MM/PBSA approach. The free energy calculations suggested that the AI's binding affinity is higher than CI against JAK2 due to an increased favorable contribution from the total non-polar interactions and the involvement of the αC helix.

Finally, we keep in mind the above information and try to find future drug candidates by the conventional systematic method of high-throughput virtual screening of an extensive library of phytochemicals, followed by molecular dynamics simulation studies in conjunction with free energy calculations and inhibitor-induced conformational dynamics. Our MD simulations show that Leonurine, STOCK1N-68642, STOCK1N-82656, and

STOCK1N-85809 bound JAK1 exhibited a smooth transition from an active conformation (αC-in) to a completely inactive conformation (αC-out). We also found that ligand binding induces disorders in the αC-helix. Subsequently, the molecular mechanics Poisson Boltzmann surface area (MM/PBSA) calculation suggested three phytochemicals, namely STOCK1N-68642, Epicatechin, and STOCK1N-98615, have higher binding affinity compared to other ligand molecules. The ligand-induced conformational plasticity revealed by our simulations differs significantly from the available crystal structures, which might help design new rational and allosteric drugs. However, further molecular and pharmacokinetics are required to test the selectivity and affinity *in-vitro* and *in-vivo*.

7.2. Outlook and future directions

In this thesis, I have discussed the conformational plasticity and molecular basis for inhibition with small molecule inhibitors and negative feedback regulator protein SOCS1 to control the dysfunctions of the JAK/STAT signaling pathway through multiscale simulation techniques. Here I have tried to elucidate the protein-ligand interaction and the role of non-covalent forces in stabilizing/destabilizing the complexes. Side by side, I also provide details information for inactive and active JAK kinase structural rearrangement for the first time. As we know, understanding protein conformational plasticity remains a challenge in drug discovery, and I believe that our results on the protein conformational variability give an insight framework to understand protein plasticity and the structure-based drug design. Here, I only considered the active kinases for drug library screening, but the future researcher can also explore more inactive states and consider JAK isoforms.

Although, if we go back to the signaling cascade of JAK/STAT, we found that we can control the aberrant function of this cascade by controlling the functional mechanisms of other members of this pathway, including cytokines receptors, STAT protein dimerization, and trans-phosphorylation of JAK. The experimental studies suggested that trans-phosphorylation plays an important role in activating JAK and the response signaling cascade. Still, the molecular basis of this transphosphorylation is not much studied theoretically. Therefore, our study would be a great starting point for further novel drug development and conformational study of JAK dimers and STAT inhibitions.

Appendix

Appendix A

Table A1: The average backbone RMSD values (in Å unit) at four different segments of the last 100 ns trajectory (four equal segments of 25 ns trajectory) along their respective standard deviation (SD) values.

Run1				
System/Time	**100-125 ns**	**125-150 ns**	**150-175 ns**	**175-200 ns**
apo	1.91 ± 0.12	1.99 ± 0.16	1.98 ± 0.14	1.99 ± 0.15
Baricitinib	1.99 ± 0.22	1.95 ± 0.11	1.94 ± 0.09	1.86 ± 0.08
Filgotinib	1.90 ± 0.18	1.90 ± 0.13	1.67 ± 0.10	1.77 ± 0.16
Itacitinib	1.56 ± 0.11	1.62 ± 0.11	1.54 ± 0.11	1.54 ± 0.09
Upadacitinib	2.18 ± 0.14	2.25 ± 0.15	2.14 ± 0.18	2.32 ± 0.15
Run2				
apo	2.31 ± 0.27	2.27 ± 01.9	2.31 ± 0.24	2.06 ± 0.26
Baricitinib	2.63 ± 0.20	2.36 ± 0.20	2.09 ± 0.19	2.17 ± 0.21
Filgotinib	2.78 ± 0.19	2.78 ± 0.18	2.65 ± 0.16	2.65 ± 0.19
Itacitinib	2.50 ± 0.16	2.34 ± 0.15	2.49 ± 0.18	2.41 ± 0.19
Upadacitinib	1.71 ± 0.14	1.78 ± 0.12	1.86 ± 0.10	1.84 ± 0.14

Table A2: The average RMSD during the last 100 ns production simulations (**Run1 & Run2**) of the JAK1-inhibitor complexes and standard errors of the mean (SEM) in the parentheses.

System/RMSD	JAK1/apo	JAK1/BTN	JAK1/FTN	JAK1/ITN	JAK1/UTN
P-loop area	1.98 (0.05)	1.97 (0.04)	2.04 (0.04)	1.37 (0.02)	2.17 (0.03)
Catalytic	0.42 (0.01)	0.40 (0.01)	0.37 (0.01)	0.38 (0.01)	0.42 (0.01)
Hinge area	0.29 (0.01)	0.38 (0.01)	0.33 (0.01)	0.40 (0.01)	0.57 (0.01)
Activation	2.14 (0.08)	2.14 (0.09)	2.57 (0.08)	1.53 (0.01)	1.50 (0.02)
DFG loop	0.17 (0.01)	0.17 (0.01)	0.21 (0.01)	0.21 (0.01)	0.20 (0.01)
C-spine	0.67 (0.02)	0.55 (0.01)	0.45 (0.01)	0.66 (0.02)	0.79 (0.01)
R-spine	0.51 (0.01)	0.64 (0.02)	0.50 (0.01)	0.55 (0.01)	0.77 (0.02)

** BTN: Baricitinib; FTN: Filgotinib; ITN: Itacitinib and UTN: Upadacitinib

Table A3: Average radius of gyration (R_g), protein solvent accessible surface area ($SASA_{protein}$), and ligand SASA ($SASA_{ligand}$) for last 100 ns trajectories **(Run1 & Run2**). Standard errors of the mean are provided in the parentheses.

System	$<R_g>$ (Å)	$<SASA_{protein}>$ (Å^2)	$<SASA_{ligand}>$ (Å^2)
JAK1/apo	19.63 (0.03)	14134.55 (86.95)	-
JAK1/BTN	19.98 (0.03)	14241.07 (73.49)	90.51 (7.28)
JAK1/FTN	19.83 (0.03)	14035.58 (77.82)	174.50 (2.75)
JAK1/ITN	19.68 (0.01)	13613.88 (62.36)	168.90 (10.82)
JAK1/UTN	19.74 (0.01)	13883.57 (42.80)	55.63 (3.78)

** BTN: Baricitinib; FTN: Filgotinib; ITN: Itacitinib and UTN: Upadacitinib

Table A4: Average network properties of Jak1 unbound and bound states with inhibitors generated through PSNs.

Parameters	Apo	Jak1/BTN	Jak1/FTN	Jak1/ITN	Jak1/UTN
No. of linked nodes	279	279	276	284	279
No. of Links	328	354	318	316	312
No. of hubs	40	48	33	29	29
No. of linked mediated hubs	149	188	132	117	120
No. of communities	12	8	7	9	13
No. of nodes involved in communities	86	111	57	47	63
No. of links involved in communities	116	162	82	62	77

Table A5: Decomposition of binding free energy into contributions from individual residues[a]

Residue	E_{vdW}	E_{elec}	G_{pol}	G_{np}	G_{side_chain}	$G_{backbone}$	G_{total}
JAK1/Baricitinib							
Leu^{1010}	-2.49	-0.10	0.29	-0.29	-2.53	-0.06	-2.59
Val^{889}	-2.15	0.28	-0.28	-0.29	-2.13	-0.31	-2.44
Leu^{959}	-1.08	-1.82	1.03	-0.04	-0.90	-1.02	-1.92
Phe^{958}	-1.15	-1.30	0.72	-0.01	-0.72	-1.02	-1.74
Leu^{881}	-1.76	0.05	0.44	-0.21	-1.35	-0.14	-1.49
Glu^{957}	-0.02	-4.87	3.47	-0.01	-0.14	-1.29	-1.43

Gly^{882}	-1.04	-0.40	0.49	-0.11	-0.28	-0.78	-1.06
JAK1/Filgotinib							
Leu^{959}	-1.09	-4.74	2.28	-0.05	-0.98	-2.62	-3.60
Leu^{881}	-2.63	0.22	0.47	-0.33	-2.24	-0.03	-2.27
Val^{889}	-1.90	-0.13	0.02	-0.25	-1.95	-0.31	-2.26
Phe^{958}	-2.03	-0.96	0.94	-0.09	-1.70	-0.45	-2.15
Leu^{1010}	-1.95	-0.12	0.35	-0.24	-1.93	-0.04	-1.97
Arg^{1007}	-1.59	-1.67	2.50	-0.27	-1.13	-0.10	-1.03
Ala^{906}	-0.90	-0.34	0.27	-0.05	-0.80	-0.23	-1.03
JAK1/Itacitinib							
Leu^{959}	-0.94	-4.19	2.41	-0.06	-1.00	-1.78	-2.78
Val^{889}	-2.24	-0.58	0.45	-0.28	-2.17	-0.48	-2.65
Leu^{1010}	-2.32	-0.02	0.14	-0.30	-2.43	-0.07	-2.50
Arg^{1007}	-2.08	-0.68	0.94	-0.35	-2.00	-0.17	-2.17
Phe^{958}	-1.25	-1.19	0.57	-0.03	-0.93	-0.97	-1.90
Leu^{881}	-1.82	-0.87	1.26	-0.25	-1.55	-0.13	-1.68
Phe^{886}	-1.54	-0.65	1.00	-0.17	-1.00	-0.36	-1.36
JAK1/Upadacitinib							
Leu^{1010}	-2.51	-0.26	0.39	-0.26	-2.42	-0.23	-2.65
Val^{889}	-2.30	0.29	-0.32	-0.26	-2.24	-0.37	-2.61
Leu^{881}	-2.54	-1.04	1.57	-0.26	-1.98	-0.30	-2.28
Phe^{958}	-1.21	-1.44	0.87	-0.02	-0.73	-1.09	-1.82
Leu^{959}	-1.03	-1.54	0.90	-0.02	-0.70	-1.00	-1.70
Gly^{882}	-0.83	-1.72	1.01	-0.09	-0.22	-1.41	-1.63
Glu^{957}	0.02	-5.22	3.72	-0.01	-0.12	-1.37	-1.49
Ala^{906}	-0.92	-0.33	0.25	-0.04	-0.77	-0.27	-1.04

[a] Energetics contributions from the van der Waals (E_{vdW}) and electrostatic interactions (E_{elec}) as well as polar (G_{pol}) and non-polar solvation energy (G_{np}) and the total contribution of given residue (G_{total}) for JAK1-inhibitor complexes are listed. G_{side_chain} and $G_{backbone}$ represent the side chain and backbone contributions. Only residues with $|\Delta G| \geq 1.0$ kcal/mol are shown. All values are given in kcal/mol. The standard error of free energy estimated as well as of individual components are not larger than 0.4 kcal/mol.

Table A6: The energy contributions of the different parts of the receptor toward the inhibitor bindings. All are in kcal/mol units.

Systems/Energy	E_{vdW}	E_{elec}	G_{pol}	G_{np}	G_{side_chain}	$G_{backbone}$	G_{total}
P-loop							
Baricitinib	-7.60	1.14	0.70	-0.93	-4.15	-2.52	-6.68
Filgotinib	-7.65	0.36	2.35	-1.03	-4.60	-1.37	-5.97
Itacitinib	-10.96	-2.85	5.58	-1.23	-6.21	-3.25	-9.46

Upadacitinib	-8.70	-1.83	3.14	-0.90	-5.05	-3.22	-8.27
Hinge region							
Baricitinib	-4.68	-7.57	5.60	-0.29	-2.99	-3.94	-6.94
Filgotinib	-9.14	-11.20	11.91	-0.89	-5.54	-3.77	-9.31
Itacitinib	-5.15	-6.22	5.10	-0.42	-3.18	-3.51	-6.69
Upadacitinib	-5.39	-7.13	5.27	-0.33	-3.03	-4.55	-7.58
Catalytic region							
Baricitinib	-4.81	-1.75	3.02	-0.53	-3.22	-0.85	-4.08
Filgotinib	-4.18	-2.05	3.30	-0.54	-3.22	-0.25	-3.47
Itacitinib	-5.68	-1.20	1.94	-0.67	-4.78	-0.83	-5.61
Upadacitinib	-4.70	-0.97	2.80	-0.42	-2.90	-0.40	-3.30
Activation region							
Baricitinib	-1.61	-4.25	6.87	-0.23	0.64	0.13	0.77
Filgotinib	-0.65	-2.21	3.23	-0.11	0.18	0.06	0.24
Itacitinib	-2.75	3.75	0.09	-0.41	0.49	0.18	0.68
Upadacitinib	-1.61	-9.12	11.44	-0.25	0.51	-0.06	0.45

Table A7: Types of interactions involved in recognition of inhibitors on JAK1

Bonds	Distance	Categories	Types
JAK1/Baricitinib			
Leu^{959}@HN - Lig@N1	2.28	Hydrogen Bond	Conventional Hydrogen Bond
Lig@HAT - Glu^{957}@O	1.88	Hydrogen Bond	Conventional Hydrogen Bond
Gly^{1020}@HA2 - Lig@NAI	3.00	Hydrogen Bond	Carbon Hydrogen Bond

Lig@H2 - Leu^{959}@O	2.47	Hydrogen Bond	Carbon Hydrogen Bond
Met^{956}@SD – Lig	4.90	Other	Pi-Sulfur
Lig – Leu^{881}	5.15	Hydrophobic	Pi-Alkyl
Lig – Ala^{906}	4.48	Hydrophobic	Pi-Alkyl
Lig - Leu^{1010}	4.12	Hydrophobic	Pi-Alkyl
Lig – Val^{889}	4.63	Hydrophobic	Pi-Alkyl
JAK1/Filgotinib			
Leu^{959}@HN – Lig@N22	2.05	Hydrogen Bond	Conventional Hydrogen Bond
Lig@H24 – Leu^{959}@O	1.77	Hydrogen Bond	Conventional Hydrogen Bond
Leu^{881} – Lig	4.97	Hydrophobic	Pi-Alkyl
Val^{889} – Lig	4.23	Hydrophobic	Pi-Alkyl
Ala^{906} – Lig	4.33	Hydrophobic	Pi-Alkyl
Leu^{959} – Lig	5.42	Hydrophobic	Pi-Alkyl
Leu^{1010} – Lig	4.07	Hydrophobic	Pi-Alkyl
JAK1/Itacitinib			
Leu^{959}@HN- Lig@N8	2.00	Hydrogen Bond	Conventional Hydrogen Bond
Lig@H22-Leu^{959}@O	1.92	Hydrogen Bond	Conventional Hydrogen Bond
Arg^{1007}@HD1-Lig@F1	2.43	Hydrogen Bond	Carbon Hydrogen Bond
Arg1007@HD2-Lig@)O1	2.62	Hydrogen Bond	Carbon Hydrogen Bond
Asn1008@HA-Lig@N5	2.62	Hydrogen Bond	Carbon Hydrogen Bond
Lig@H21-Glu957@)O	2.51	Hydrogen Bond	Carbon Hydrogen Bond
Arg1007@CZ-Lig@F3	3.61	Halogen	Halogen (Fluorine)
Leu881 - Lig	2.62	Hydrophobic	Pi-Sigma, Pi-Alkyl
Val889 – Lig	2.64	Hydrophobic	Pi-Sigma
Arg1007 - Lig	5.48	Hydrophobic	Alkyl, Pi-Alkyl
Leu1010 - Lig	5.09	Hydrophobic	Pi-Alkyl
Ala^{906} - Lig	3.63	Hydrophobic	Pi-Alkyl
Leu^{959} – Lig	4.86	Hydrophobic	Pi-Alkyl
JAK1/Upadacitinib			

Leu^{959}@HN – Lig@N5	2.17	Hydrogen Bond	Conventional Hydrogen Bond
Lig@H13 – Asp^{1021}@OD1	1.94	Hydrogen Bond	Conventional Hydrogen Bond
Lig@H18 – Glu^{957}@O	1.89	Hydrogen Bond	Conventional Hydrogen Bond
Gly^{882}@HA2 – Lig@O1	2.96	Hydrogen Bond	Carbon Hydrogen Bond
Gly^{884}@HA1 – Lig@F1	2.95	Halogen	Halogen (Fluorine)
Gly^{962}@HA2 – Lig@N3	2.81	Hydrogen Bond	Carbon Hydrogen Bond
Arg^{1007}@HD1 – Lig@F3	2.49	Hydrogen Bond	Carbon Hydrogen Bond
Lig@H3 – Asp^{1021}@OD1	2.83	Hydrogen Bond	Carbon Hydrogen Bond
Lig@H12 – Leu^{881}@O	2.76	Hydrogen Bond	Carbon Hydrogen Bond
Lig@H14 – Leu^{959}@O	2.29	Hydrogen Bond	Carbon Hydrogen Bond
Lig@H16 – Asp^{1021}@OD1	2.99	Hydrogen Bond	Carbon Hydrogen Bond
Glu^{883}@O – Lig@F1	3.00	Halogen	Halogen (Fluorine)
Leu^{881}@HB2 – Lig	2.76	Hydrophobic	Pi-Sigma
Met^{956}@SD - Lig	4.95	Other	Pi-Sulfur
Val^{889} – Lig	4.41	Hydrophobic	Alkyl
Lig@C8 – Leu^{1010}	3.94	Hydrophobic	Alkyl
Lig - Leu^{881}	4.40	Hydrophobic	Pi-Alkyl
Lig – Ala^{906}	4.32	Hydrophobic	Pi-Alkyl
Lig - Leu^{1010}	4.42	Hydrophobic	Pi-Alkyl
Lig - Val^{938}	5.44	Hydrophobic	Pi-Alkyl

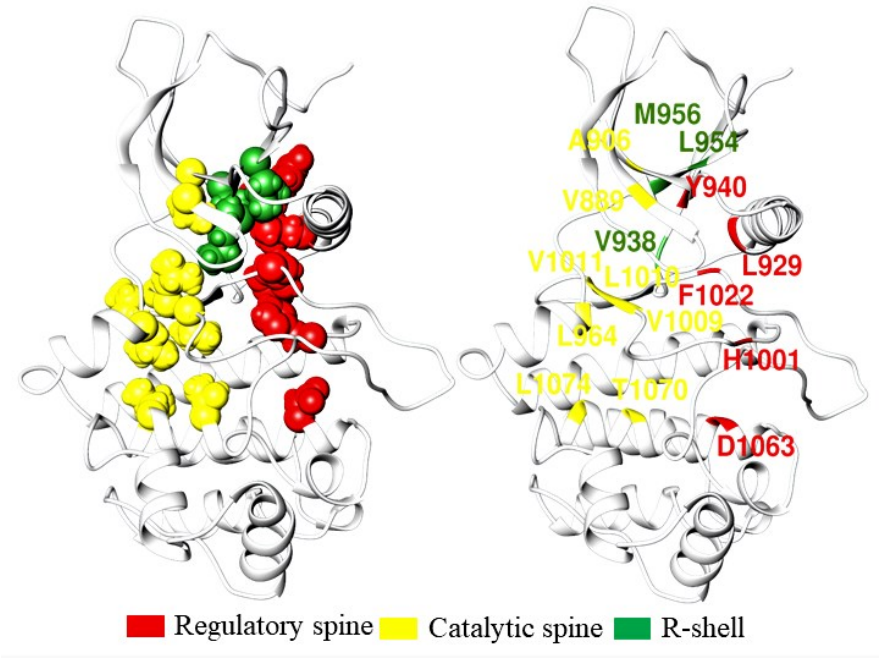


Figure A1: The position of each regulatory spine, catalytic spine, and R-shell in JAK1 structure. R-spine residues are Tyr^{940}, Leu^{929}, Phe^{1022}, His^{1001} and D^{1063}; R-shell residues are Leu^{954}, Met^{956} (gatekeeper residue) and Val^{938}; C-spine residues are Val^{889}, Ala^{906}, Leu^{964}, Val^{1009}, Leu^{1010}, Val^{1011}, Thr^{1070} and Leu^{1074}.

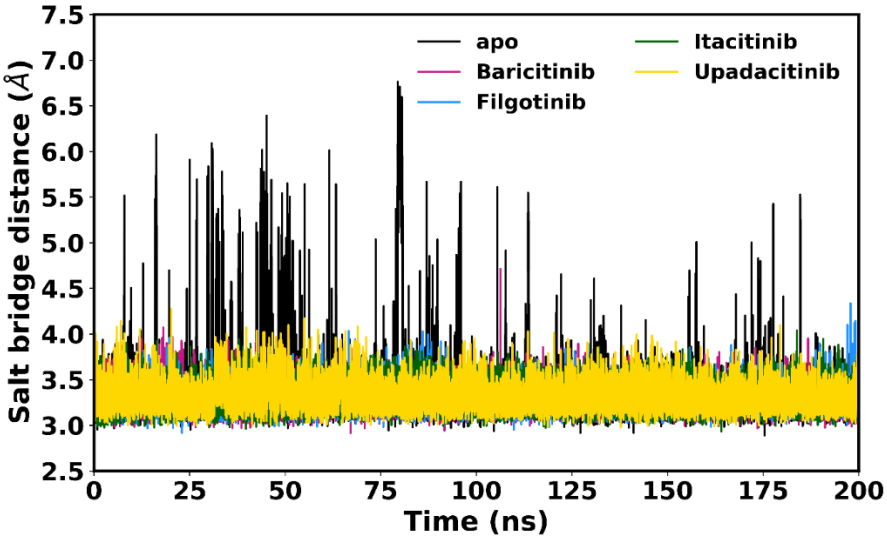


Figure A2: Evolution of salt bridge distance K908@NZ and E925@CD over simulation time.

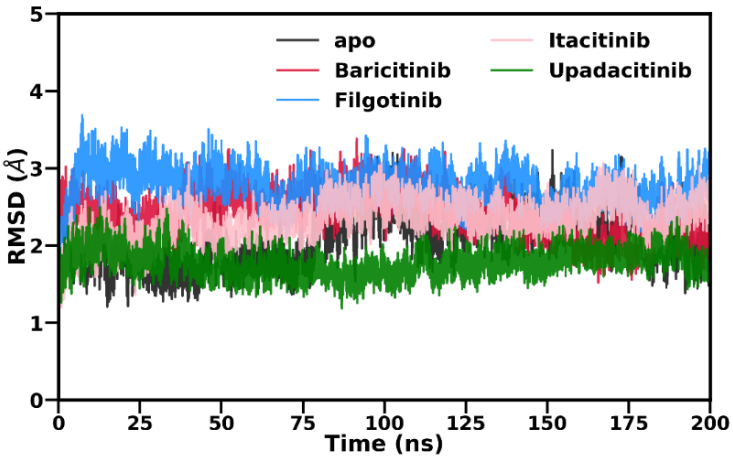


Figure A3: The time evolution of backbone RMSD of the JAK1/ligand complexes, including the apo system. The data are derived from the replica simulation (**Run2**).

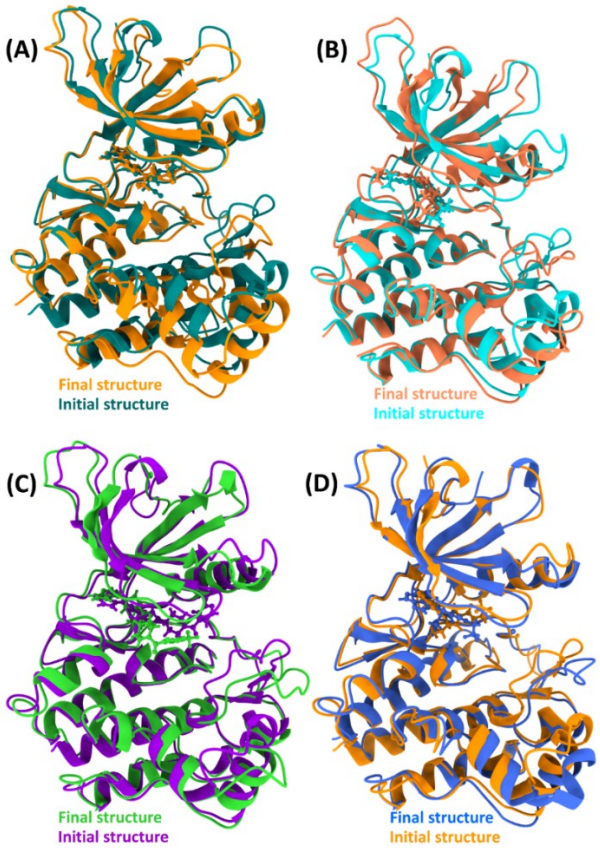


Figure A4: The superimposed structures of initial and final conformations of (A) Baricitinib, (B) Filgotinib, (C) Itacitinib, and (D) Upadacitinib complexed with JAK1.

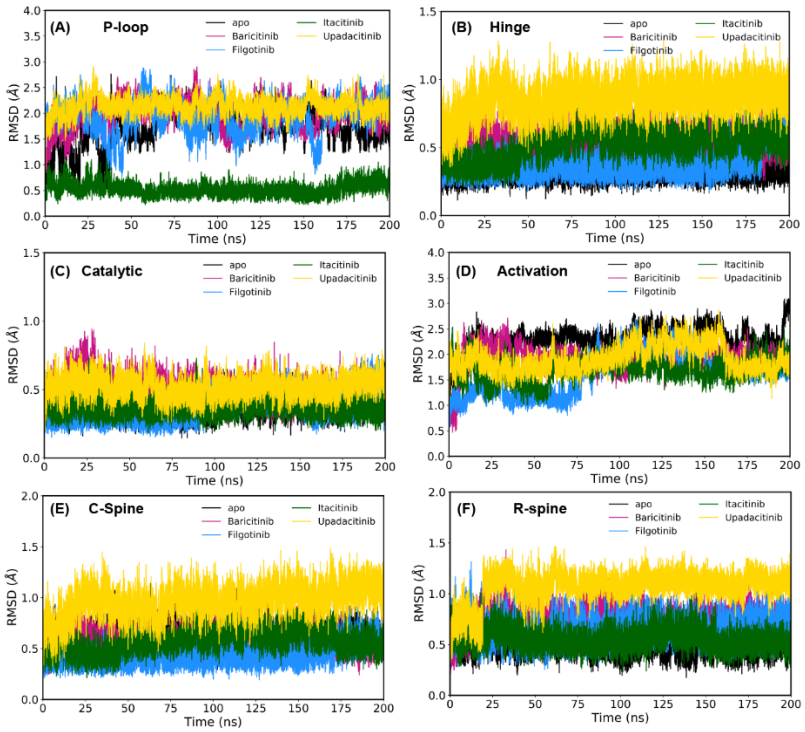


Figure A5: Time dependence root-mean-square deviations (RMSDs) of C_α atoms relative to their energy minimized and equilibrated structures of (A) P-loop, (B) hinge region, (C) catalytic region, (D) activation segment, (E) C-spine residues, (F) R-spine residues.

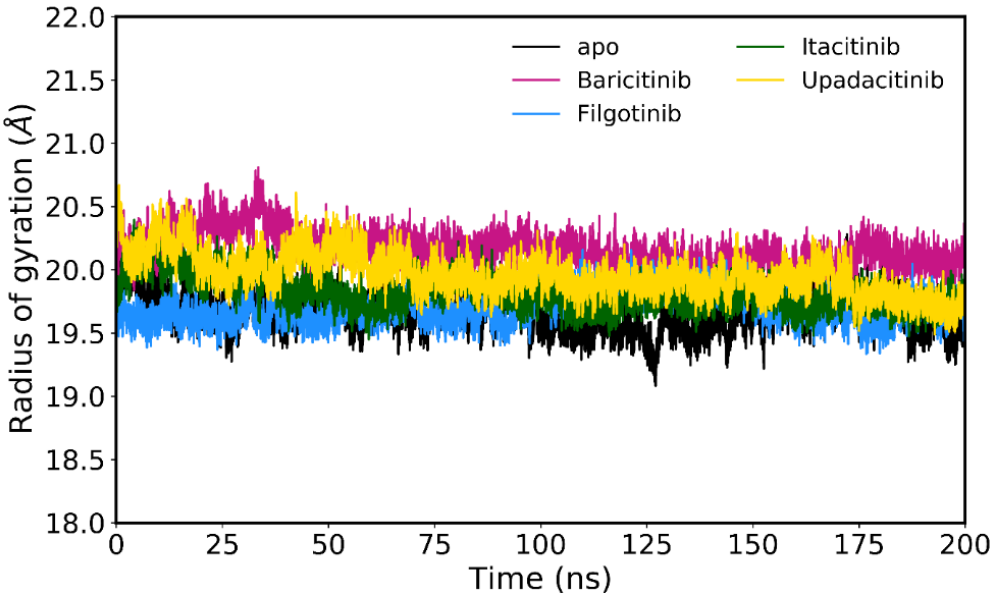


Figure A6: The time evolution of the radius of gyration of the four complexes during MD simulation.

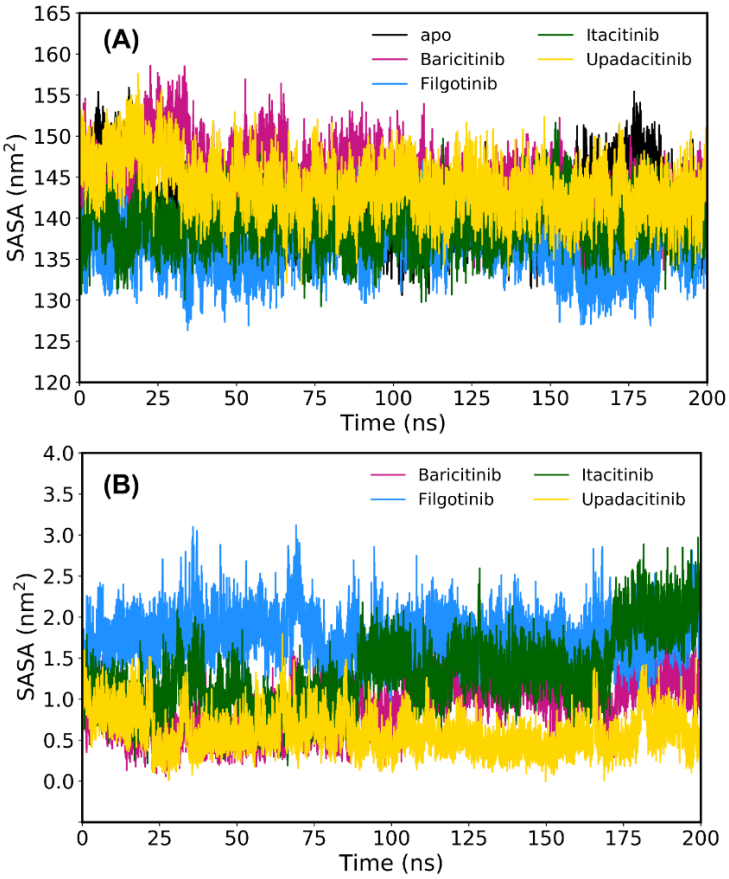


Figure A7: The time evolution of the solvent accessible surface area of (A) protein/ligand and (B) ligands.

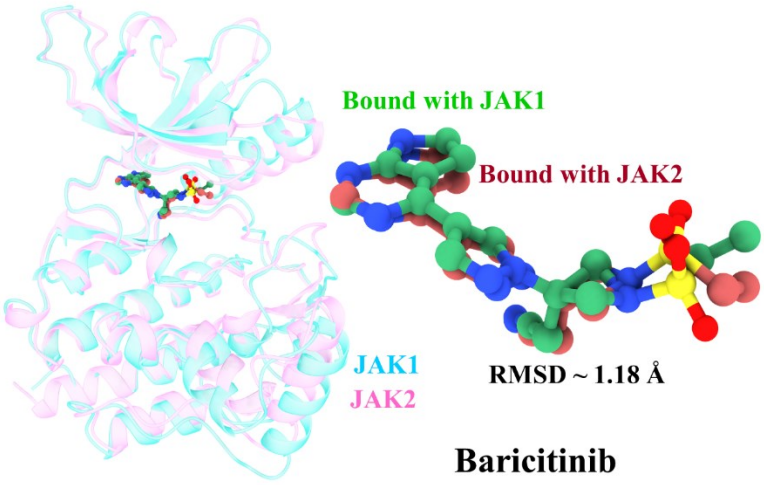


Figure A8: Superposition of JAK1 (light blue) and JAK2 (light pink, PDB:6VN8) JH1 domain complexed with Baricitinib. The RMSD of Baricitinib is 1.18 Å.

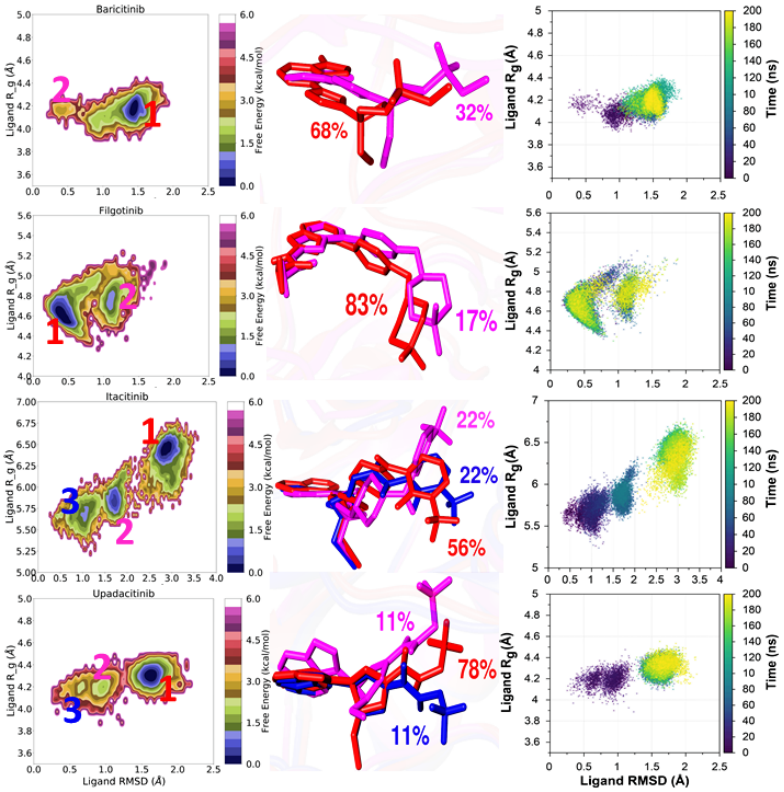


Figure A9: (Left) Two-dimensional free energy landscape of ligand movements. The ligand RMSD and radius of gyration of each ligand are taken as reaction coordinates. (Right) Ligand movements scatter plot of 20000 snapshots from MD simulation along RMSD and radius of gyration of ligand and plotted against simulation time.

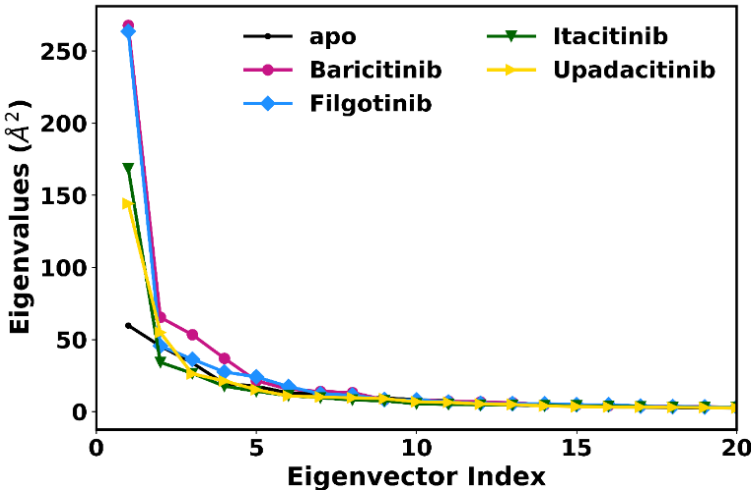


Figure A10: Eigenvalues as a function of the eigenvectors. Here we calculated first 20 eigenvectors and its corresponding eigenvalues of covariance matrices of PCA.

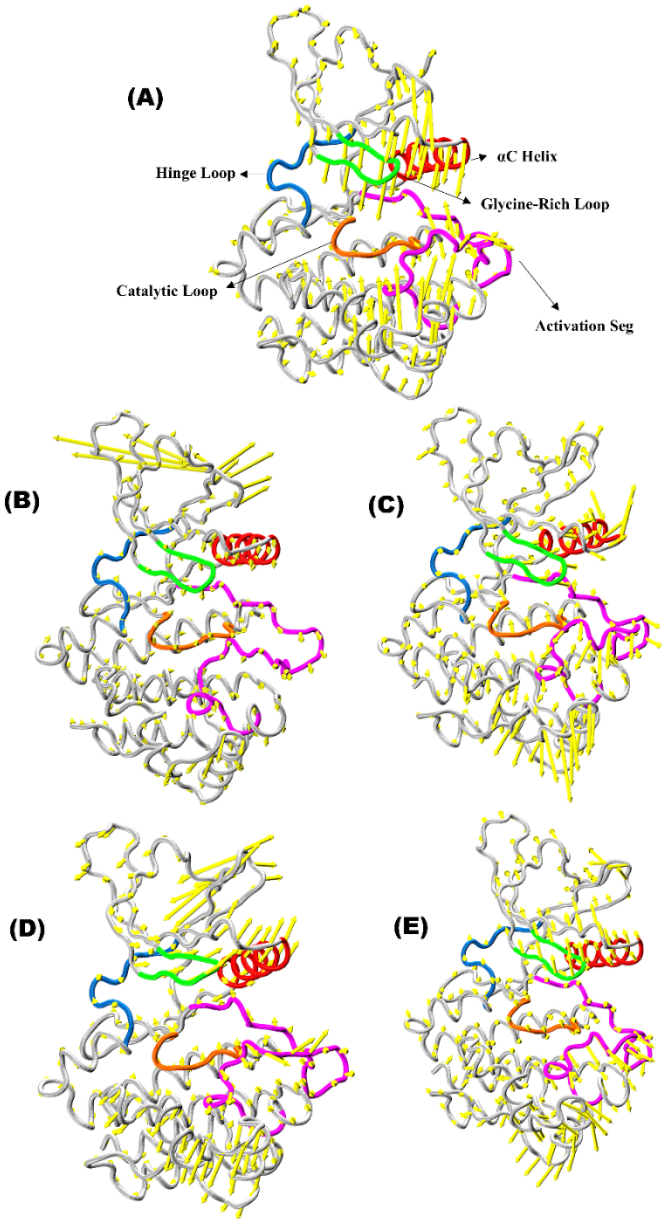


Figure A11: Porcupine plots of the first PC motion modes of (A) apo, (B) JAK1/Baricitinib, (C) JAK1/Filgotinib, (D) JAK1/Itacitinib, and (E) JAK1/Upadacitinib models. The protein is shown in tube representation; the dark red color corresponds to αC helix, light green is the P-loop region, dark orange is the catalytic loop, the blue color corresponds to the hinge region, and dark magenta represents activation segment. The prominent motions are shown in the yellow color corresponding to the direction of the eigenvector, and the magnitude of length represents the eigenvalues.

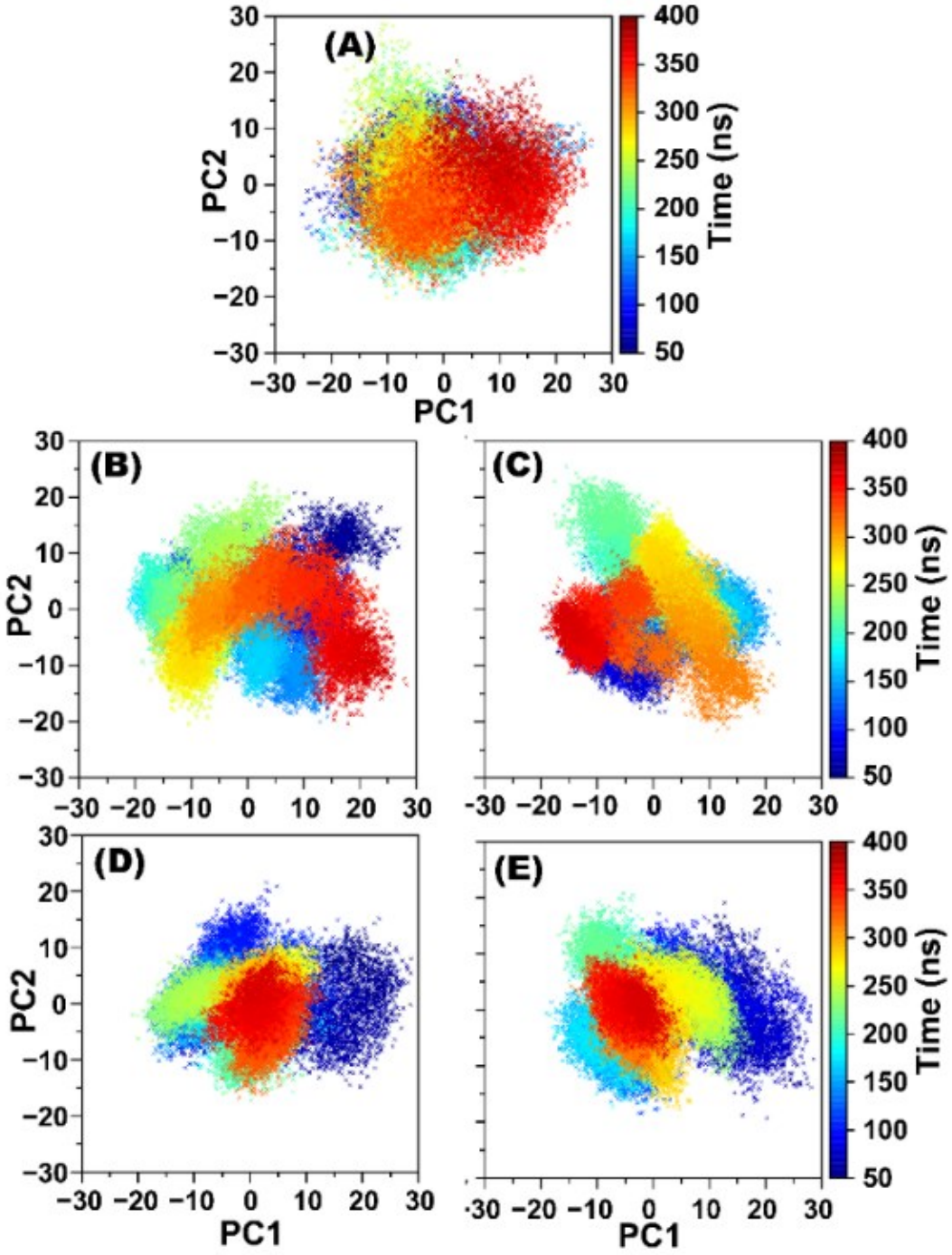


Figure A12: PCA scatter plot of last 17000 snapshots from MD simulation along with first two principal components and plotted against simulation time for (A) apo, (B) JAK1/Baricitinib, (C) JAK1/Filgotinib, (D) JAK1/Itacitinib and (E) JAK1/Upadacitinib, respectively.

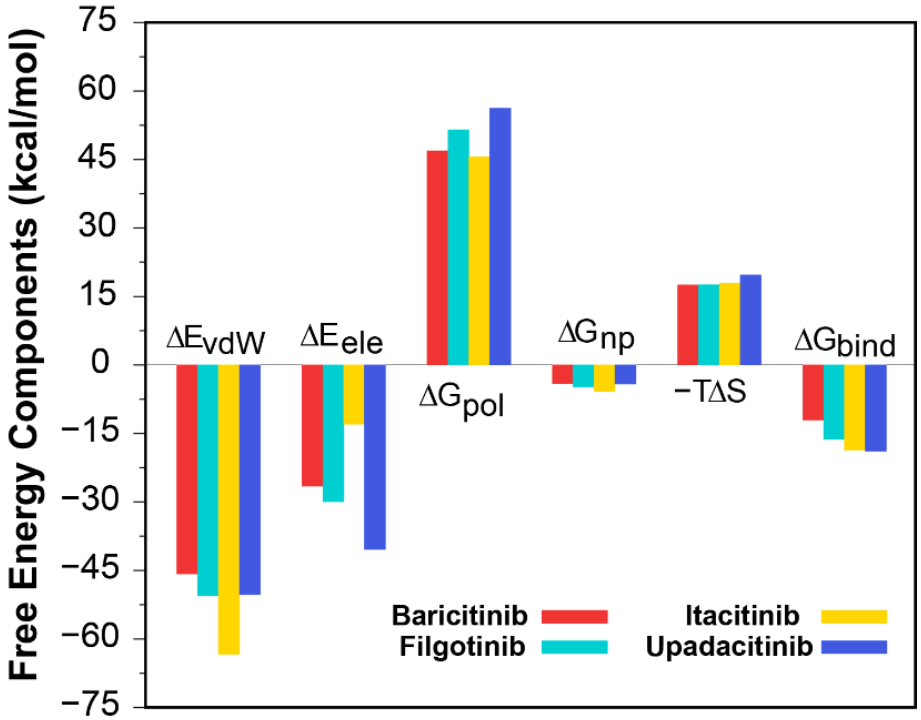


Figure A13: Energy components (kcal/mol) for the binding of inhibitors to the JAK1. $\Delta E_{vdW,}$ van der Waals energy; $\Delta E_{ele,}$ electrostatics energy in the gas phase; $\Delta G_{pol,}$ polar solvation energy; $\Delta G_{np,}$ non-polar solvation energy; $T\Delta S$, configurational entropy contribution.

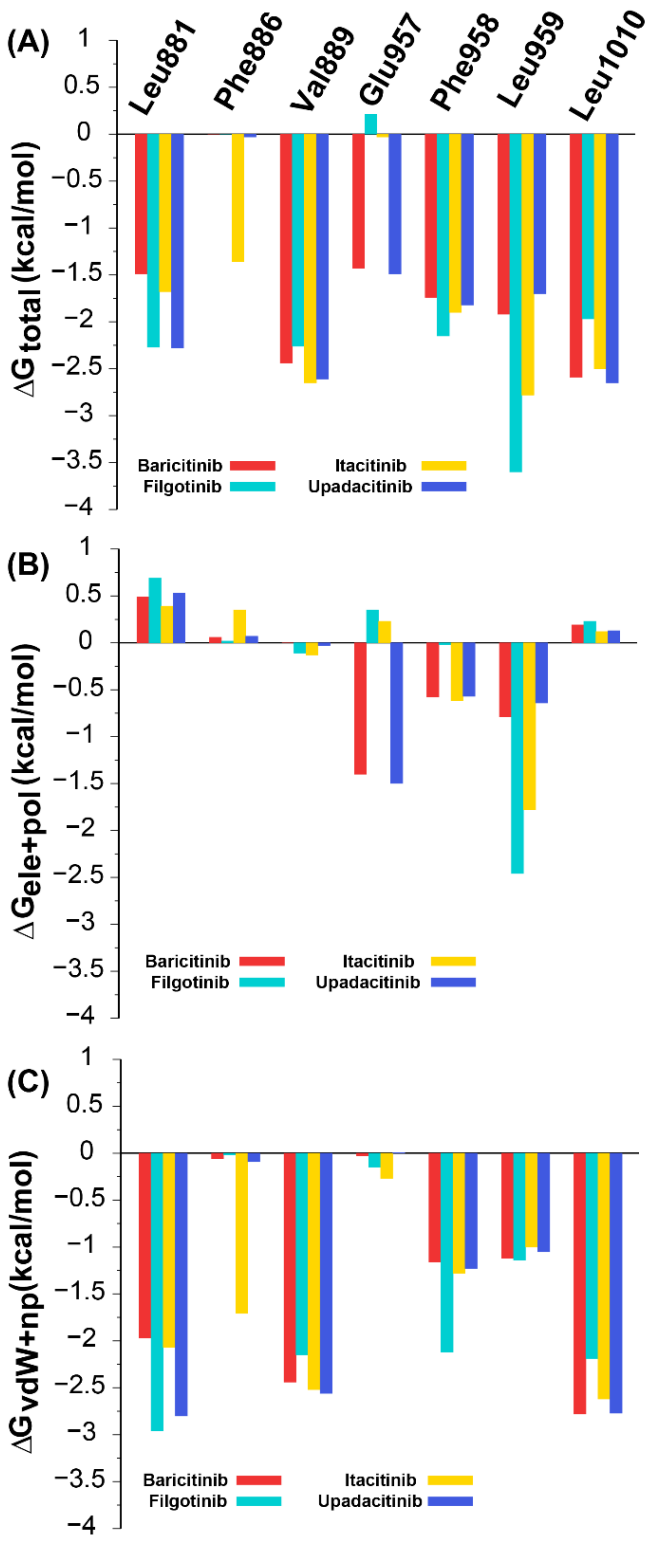


Figure A14: The key residue interactions for JAK1 kinase with four inhibitors. (A) The total binding contributions (ΔG_{Total}); (B) the total polar contributions ($\Delta G_{ele+pol}$); (C) the total non-polar contributions (ΔG_{vdW+np}).

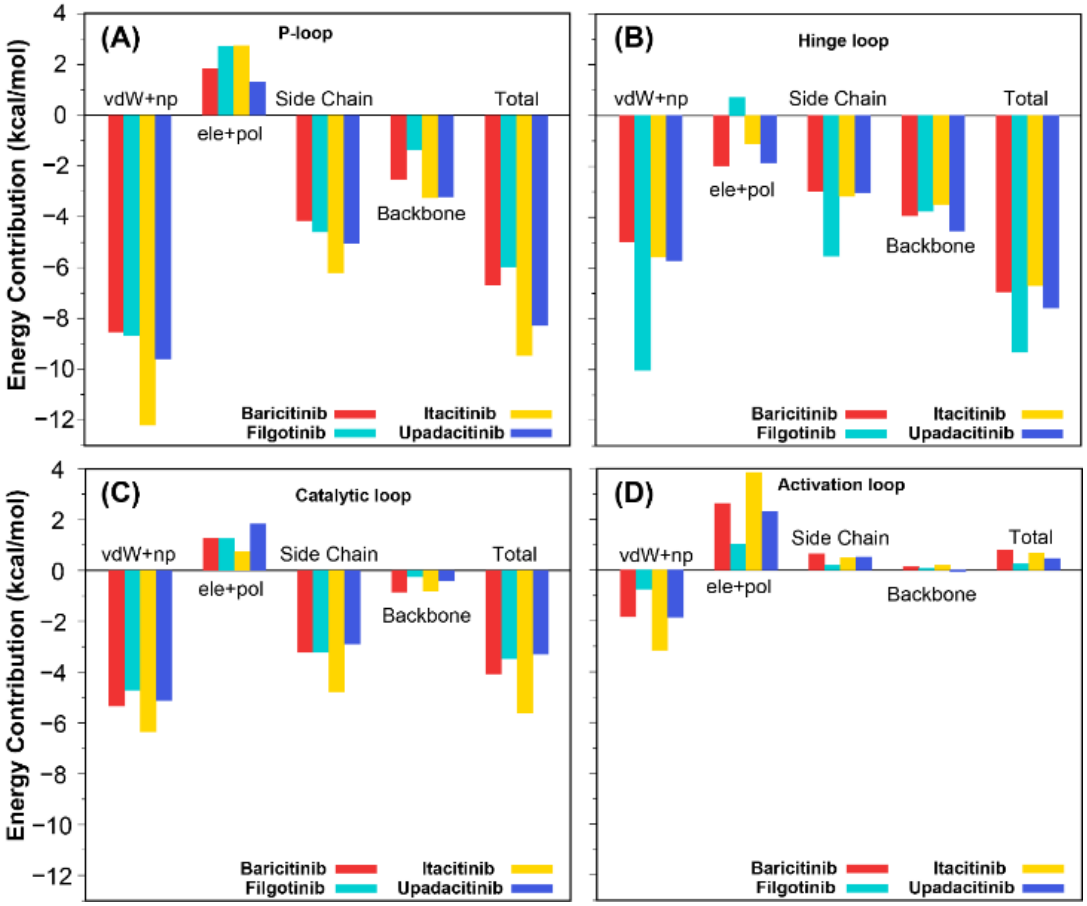


Figure A15: Energy decompositions of the different important regions of JAK1 and inhibitors complexes. (**A**) The glycine-rich, P-loop; (**B**) hinge region; (**C**) catalytic loop; (**D**) activation segment.

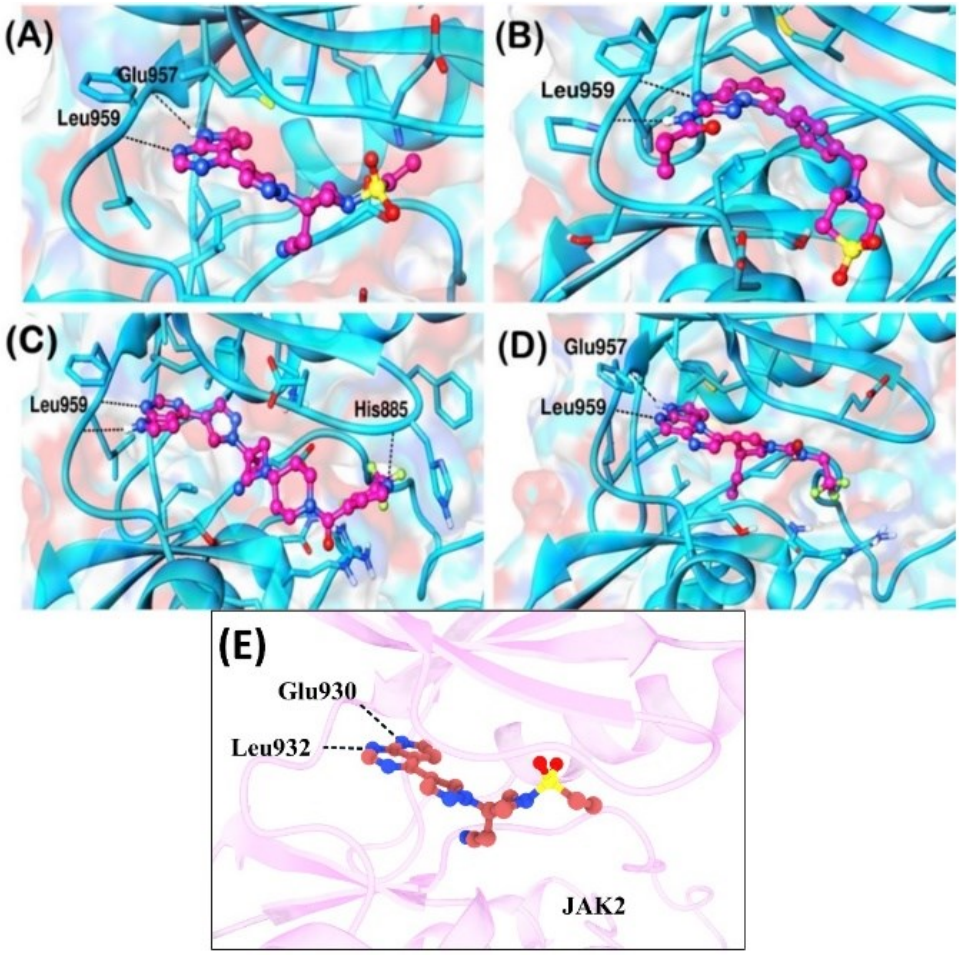


Figure A16: The final (200 ns) snapshot of each system (A) JAK1/Baricitinib, (B) JAK1/Filgotinib, (C) JAK1/Itacitinib, and (D) JAK1/Upadacitinib, the corresponding inhibitors are shown in ball and stick representations. (E) The hydrogen bond interaction pattern of JAK2 with Baricitinib. The protein is in new cartoons and with essential regions, and hydrogen-bonded residues are labeled, and bonds are shown in the black dotted line.

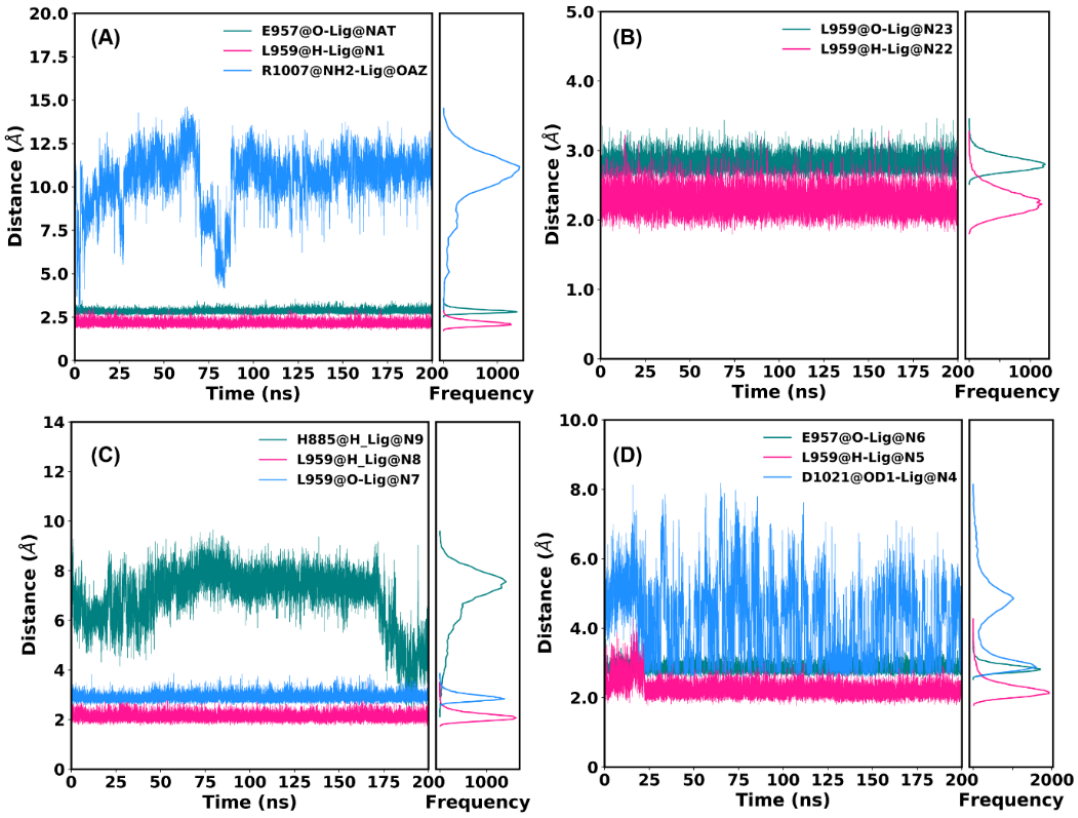


Figure A17: The hydrogen-bonded distance calculations from the simulations between the JAK1 and corresponding ligands, i.e., (A) Baricitinib, (B) Filgotinib, (C) Itacitinib, and (D) Upadacitinib, respectively.

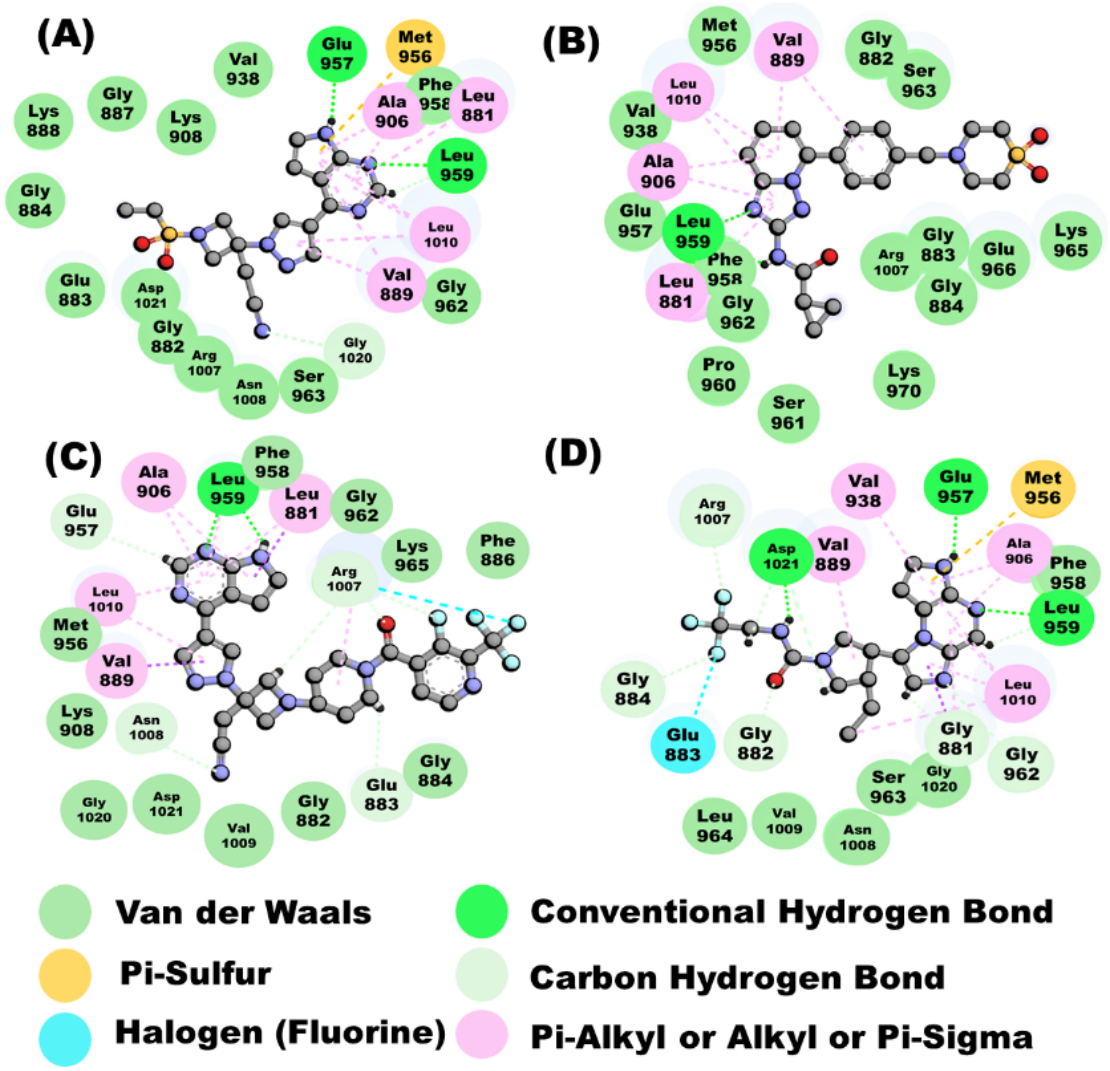


Figure A18: The inter-molecular interaction of JAK1 with (A) Baricitinib, (B) Filgotinib, (C) Itacitinib, and (D) Upadacitinib.

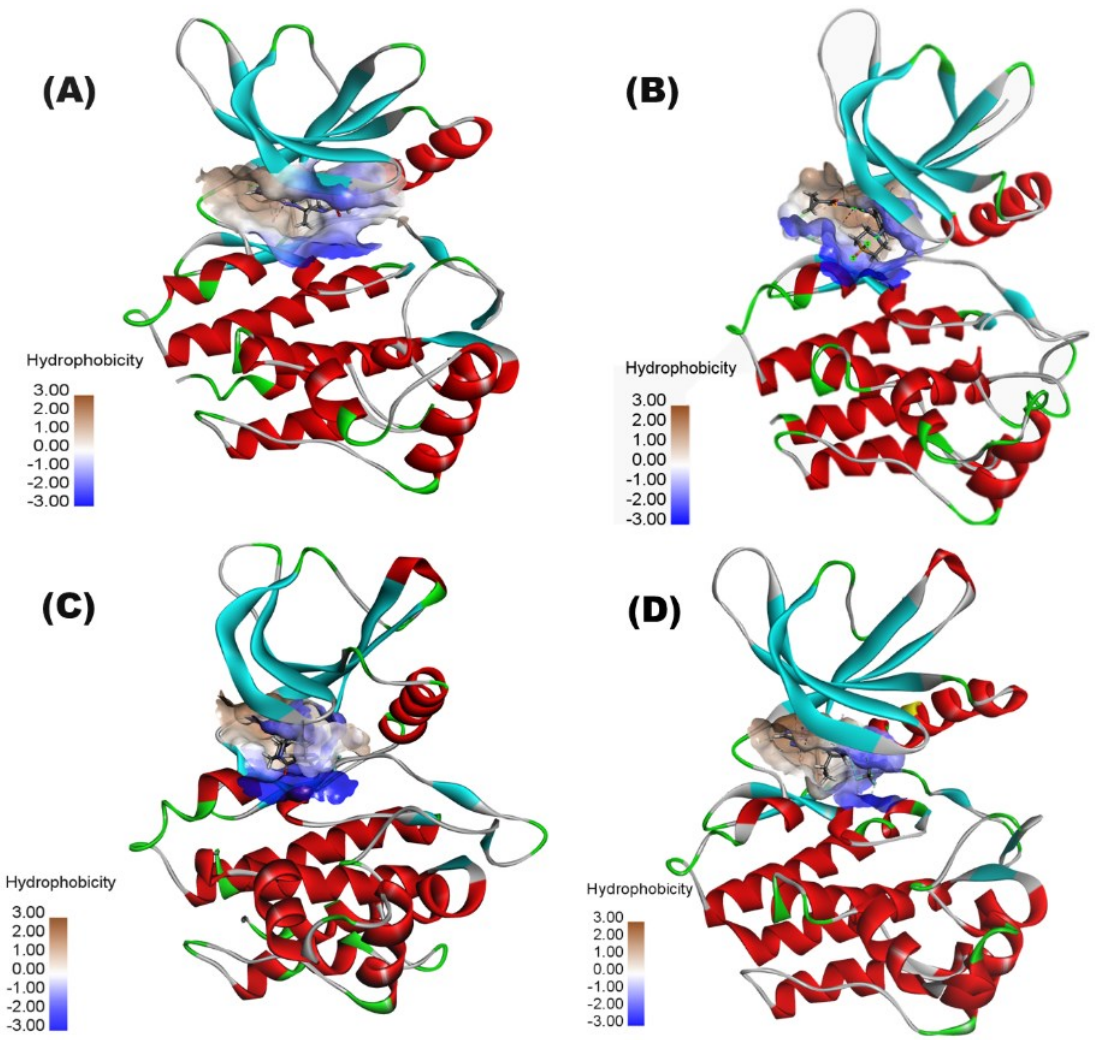


Figure A19: The hydrophobicity surface of the binding pocket of JAK1 for four different inhibitors. (A) JAK1/Baricitinib; (B) JAK1/Filgotinib; (C) JAK1/Itacitinib; and (D) JAK1/Upadacitinib.

Appendix B

Table B1: The average values of backbone RMSD, the radius of gyration (Rg), and solvent accessible surface area (SASA) for all seven systems. The statistical error was estimated based on the deviation between block averages and provided in the parentheses.

Systems		**RMSD (Å)**	**Rg (Å)**	**SASA (Å^2)**
Double phosphorylation, 2pY				
Com-2pY	**JAK1^{2pY} + SOCS1**	2.78 (0.06)	25.14 (0.05)	18981.08 (45.81)
	JAK1^{2pY}	2.25 (0.05)	19.80 (0.03)	12671.42 (33.23)
	SOCS1	1.16 (0.01)	14.13 (0.01)	6085.63 (8.00)
Apo-2pY	**JAK1^{2pY}**	2.38 (0.03)	19.60 (0.01)	13632.23 (16.22)
Single phosphorylation, pY1034				
Com-pY1034	**JAK1^{pY1034} + SOCS1**	1.93 (0.02)	24.72 (0.02)	18626.20 (22.50)
	JAK1^{pY1034}	1.80 (0.02)	19.55 (0.01)	12375.75 (15.76)
	SOCS1	1.35 (0.01)	14.15 (0.01)	6133.13 (9.84)
Apo-pY1034	**JAK1^{pY1034}**	**1.96 (0.01)**	19.67 (0.01)	13843.84 (13.90)
Single phosphorylation, pY1035				
Apo-pY1035	**JAK1^{pY1035} + SOCS1**	2.38 (0.06)	24.74 (0.02)	18701.44 (32.23)
	JAK1^{pY1035}	2.25 (0.0)	19.52 (0.02)	12470.50 (31.06)
	SOCS1	1.15 (0.01)	14.13 (0.01)	6126.24 (8.91)
Apo-pY1035	**JAK1^{pY1035}**	**3.05 (0.03)**	19.76 (0.01)	13926.92 (21.30)
SOCS1				
Apo-SOCS1	**SOCS1**	3.00 (0.04)	13.96 (0.01)	6889.55 (12.73)

Table B2: Average network properties of three phosphorylated JAK1 bound with SOCS1, generated through PSNs.

Parameters	**com-2pY**	**com-pY1034**	**com-pY1035**
No. of linked nodes	387	391	384
No. of Links	434	465	447
No. of hubs	37	54	47
No. of linked mediated hubs	157	216	191
No. of communities	16	9	13
No. of nodes involved in communities	79	115	100
No. of links involved in communities	97	163	129

Table B3: Hydrogen bond (HB) bonds formed between JAK1 and SOCS1 and the corresponding average distance and percentage of HB occupancy determined using 3 μs trajectories in the MD simulations.

Binding couples		**Distance (Å)**	**Hydrogen bonds occupancy (%)**
Acceptor (A)	Donor (D)	A-D distance	MD analysis
2pY			
Phe58@O	Ser1056@OG	2.71	94.75
Asp105@OD2	Arg1103@NH1	2.74	72.05
Ser1056@OG	**Thr57@OG1**	2.80	68.30
Phe55@O	Ser1043@N	2.84	64.31
Asp105@OD2	Arg1103@NH2	2.79	59.72
Leu1053@O	**Phe58@N**	2.88	53.37
Ser1056@O	**Phe55@N**	2.89	44.47
Arg56@O	Gln1098@NE2	2.82	40.50
Cys111@O	Gln1098@N	2.86	30.18
Asp105@OD1	Arg1103@NH2	2.84	27.76
Asp1040@OD2	**Arg59@NH2**	2.79	27.46
Asp1040@OD1	**Arg59@NH2**	2.79	24.58
Asp105@O	His1096@NE2	2.85	24.18
Gln1098@OE1	**Arg56@NH2**	2.81	22.87
pY1034			
Phe58@O	Ser1056@OG	2.73	91.32
Phe55@O	Ser1043@N	2.85	72.63
Asp105@OD2	Arg1103@NH2	2.80	67.81
Asp105@OD1	Arg1103@NH1	2.73	57.65
Leu1053@O	**Phe58@N**	2.88	57.12
Ser1056@OG	**Thr57@OG1**	2.80	53.87
Ser1043@O	**Phe55@N**	2.90	36.67
Arg56@O	Gln1098@NE2	2.84	31.80
Asp1040@OD1	**Arg59@NH2**	2.78	27.20
Asp105@O	His1096@NE2	2.86	27.06
Asp1040@OD2	**Arg59@NH2**	2.79	23.76
Asp105@OD1	Arg1103@NH2	2.78	20.98
Cys111@O	Gln1098@N	2.87	20.73
pY1035			
Phe58@O	Ser1056@OG	2.73	95.12
Asp105@OD2	Arg1103@NH1	2.75	64.30
Phe55@O	Ser1043@N	2.85	63.06

Ser1056@OG	**Thr57@OG1**	2.81	62.58
Leu1053@O	**Phe58@N**	2.88	53.79
Asp105@OD2	Arg1103@NH2	2.77	50.49
Arg56@O	Gln1098@NE2	2.84	38.45
Cys111@O	Gln1098@N	2.86	36.22
Ser1043@O	**Phe55@N**	2.90	35.76
Asp105@O	His1096@NE2	2.86	31.98
Asp105@OD1	Arg1103@NH2	2.81	30.19
Asp1040@OD1	**Arg59@NH2**	2.80	27.15
Asp1040@OD2	**Arg59@NH2**	2.80	21.30

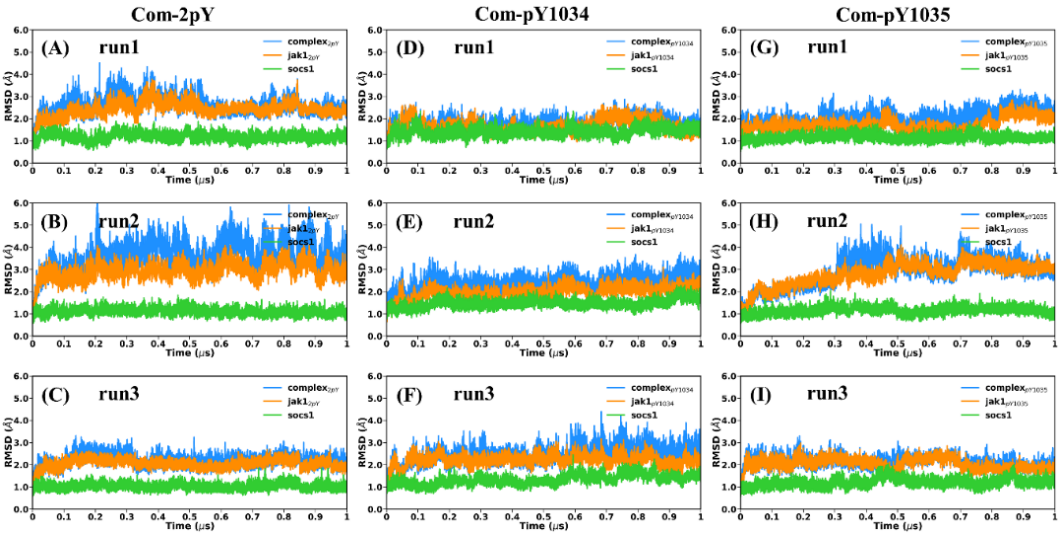


Figure B1: The time evolution of the backbone RMSD of JAK1/SOCS1 complexes with three different phosphorylated states.

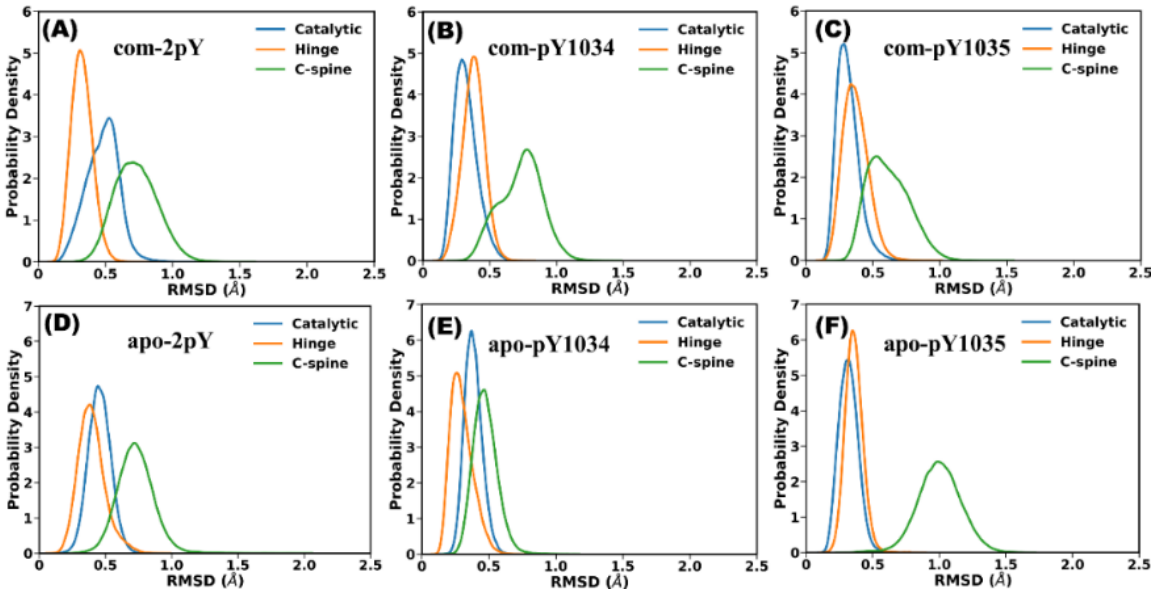


Figure B2: The probability of density function (PDF) plots based on backbone RMSD of catalytic, hinge, and catalytic spine (C-spine) of JAK1 in complexed with SOCS1 and apo conformation.

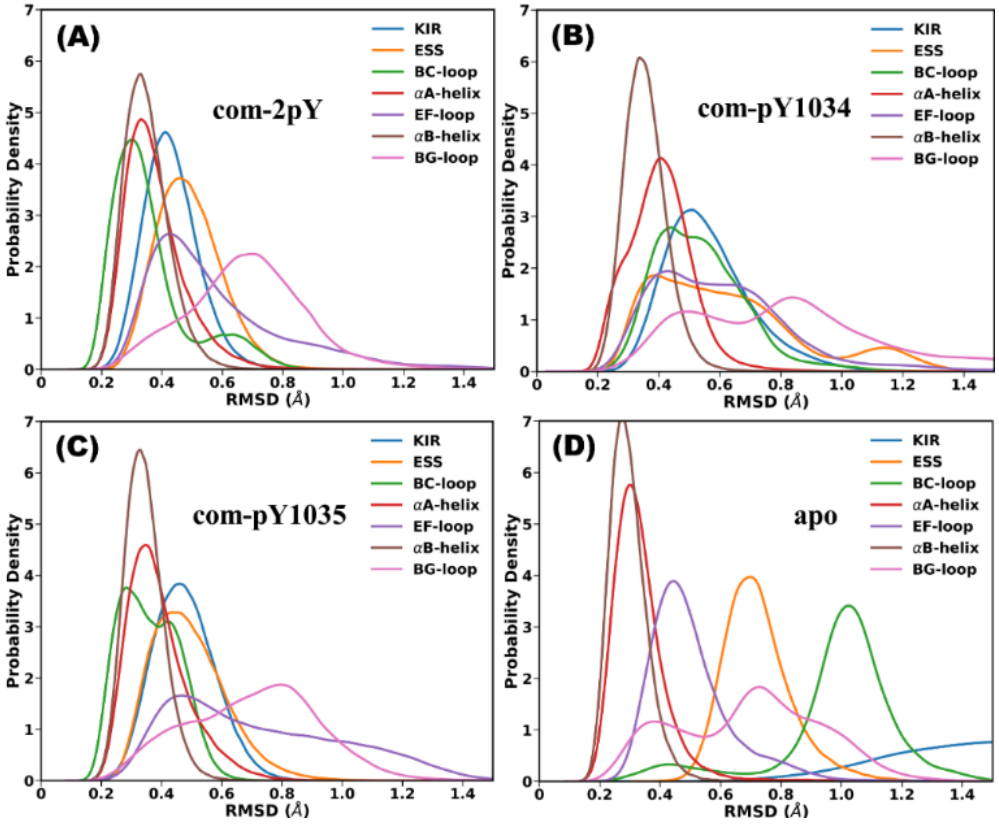


Figure B3: The probability of density function (PDF) plots based on backbone RMSD of KIR, ESS, BC-loop, αA-helix, EF-loop, αB-helix, and BG-loop SOCS1 in complexed with JAK1 and apo conformation.

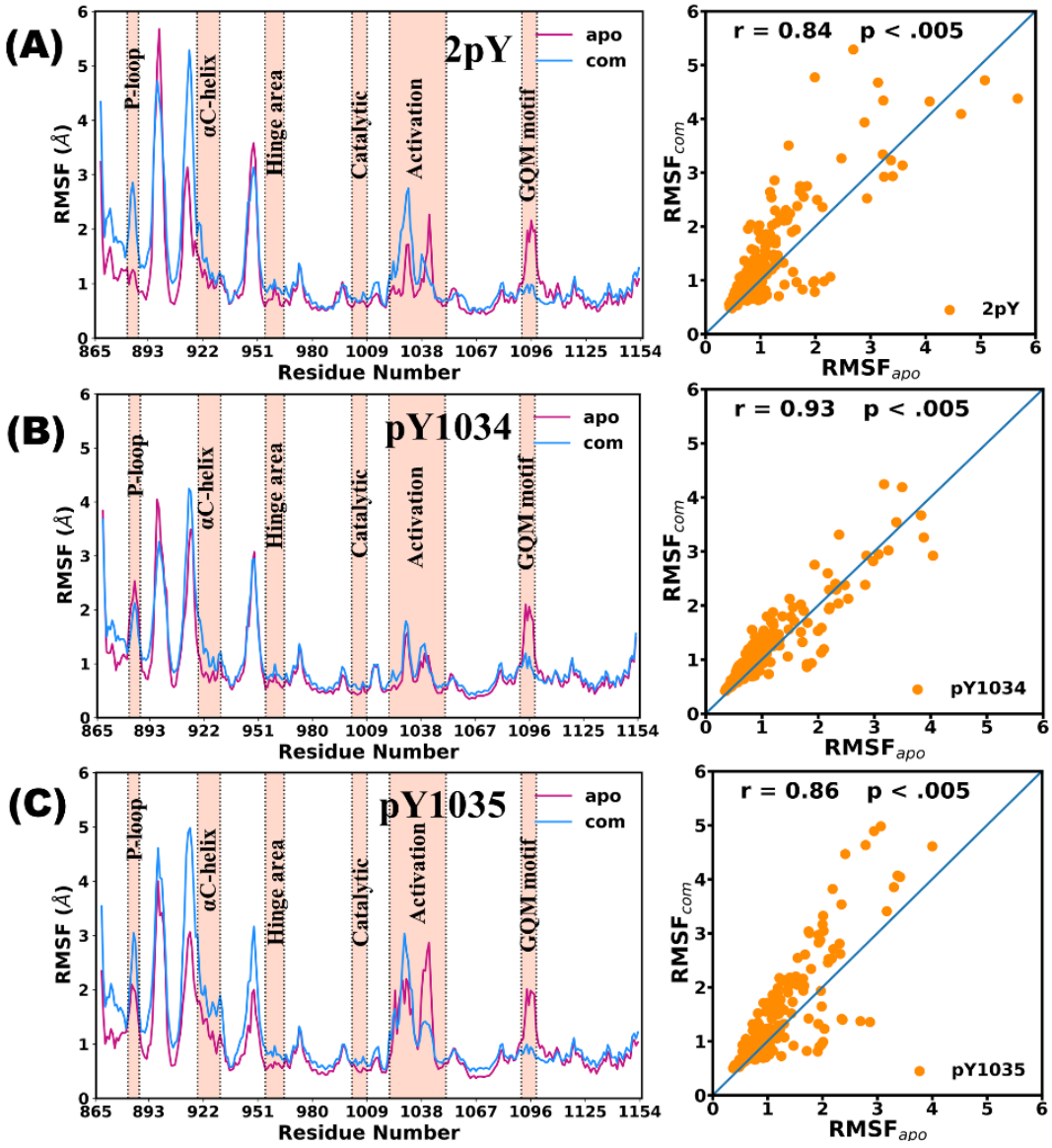


Figure B4: Residue-wise C_α RMSF of apo and complex form of JAK1, is plotted for apo (magenta) and complex (light blue). A scatter plot of C_α RMSF of residues in complex is plotted against that of the equivalent residues in apo in **(A)** the double phosphorylated A-loop, 2pY, **(B)** single phosphorylated A-loop, pY1034, and **(C)** single phosphorylated A-loop, pY1035. The blue diagonal indicates unity slope, across which the points are expected to be equally distributed if there is no systematic difference in Cα RMSF of residues in complex and apo Jak1 (Chi-square test, p-value < 0.05).

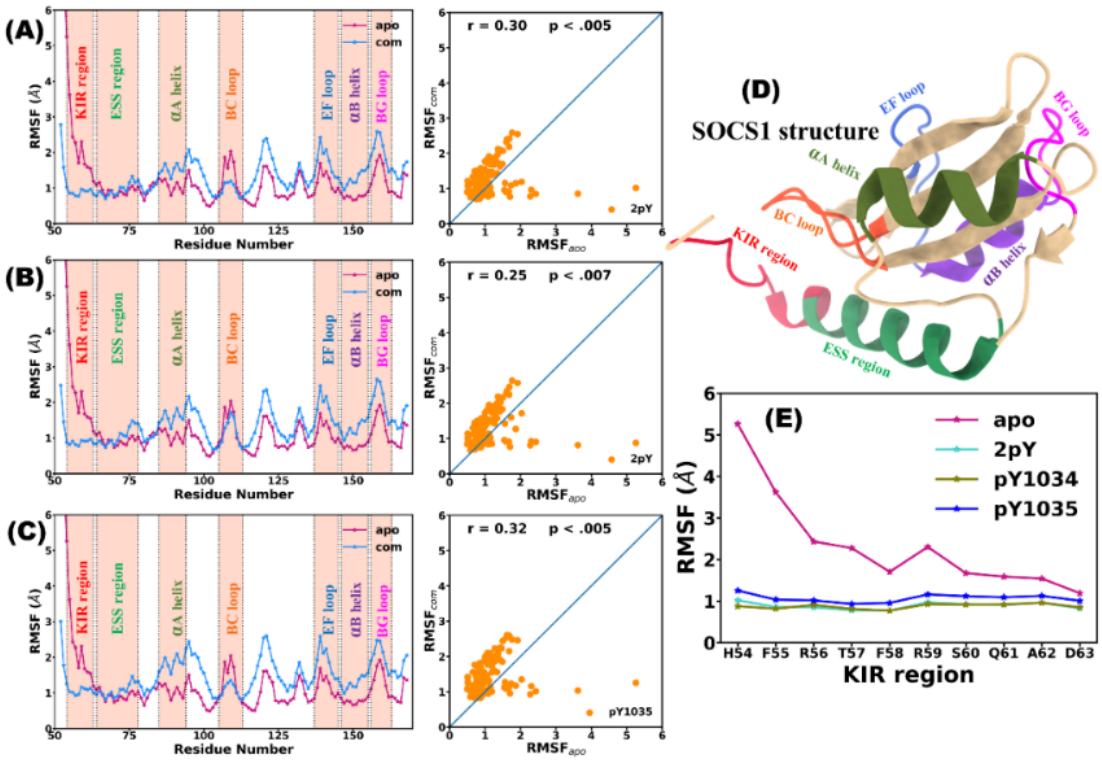


Figure B5: Residue-wise C_α RMSF of the apo and complex form of SOCS1 is plotted for apo (magenta) and complex (light blue). A scatter plot of C_α RMSF of residues in the complex is plotted against that of the equivalent residues in apo in **(A)** the double phosphorylated A-loop, 2pY, **(B)** single phosphorylated A-loop, pY1034, and **(C)** single phosphorylated A-loop, pY1035. The blue diagonal indicates a unity slope, across which the points are expected to be equally distributed if there is no systematic difference in Cα RMSF of residues in complex and apo SOCS1 (Chi-square test, p-value < 0.05). **(D)** the three-dimensional structure of SOCS1 and **(E)** RMSF values of KIR region residues.

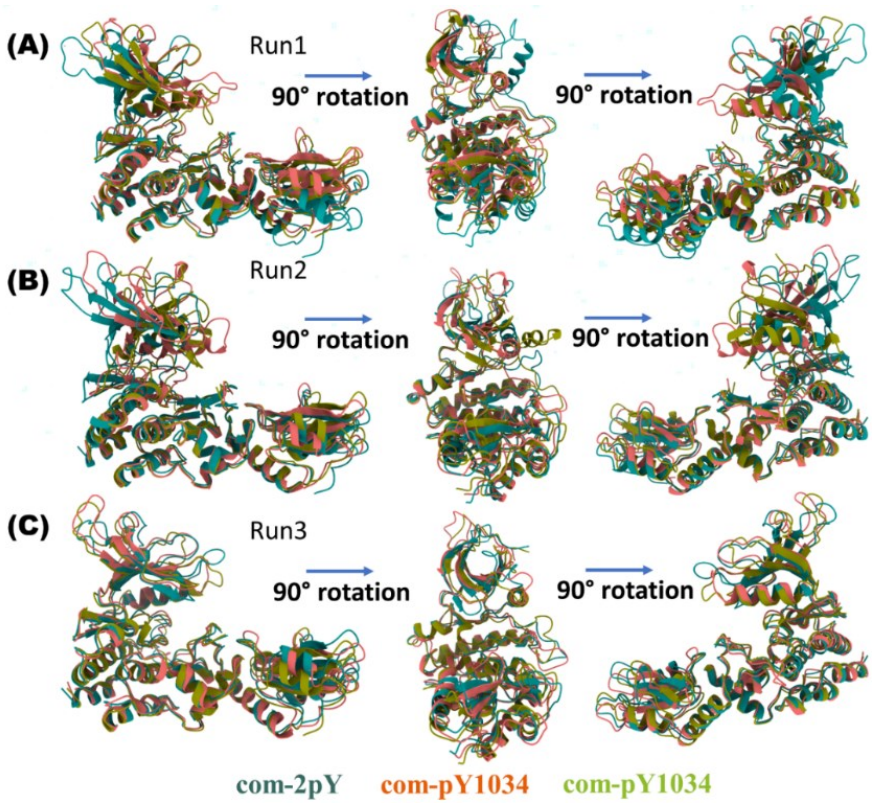


Figure B6: The superimposed final snapshots of three phosphorylated JAK1/SOCS1 complex in three replicas, **(A)** run1, **(B)** run2, and **(C)** run3, respectively.

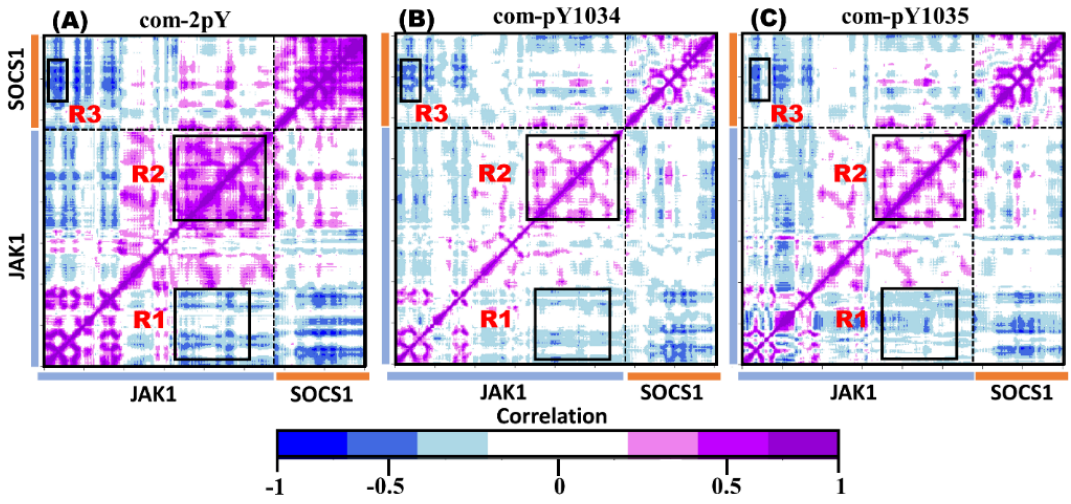


Figure B7: Dynamic cross-correlation matrices (DCCM) of the fluctuations of Cα atoms in com-2pY **(A)**, com-pY1034 **(B)** and com-pY1035 **(C)** JAK1/SOCS1 complexes. The negative, blue color bar represents the negative correlation, and the positive, dark pink color bar indicates positively correlated motions in protein

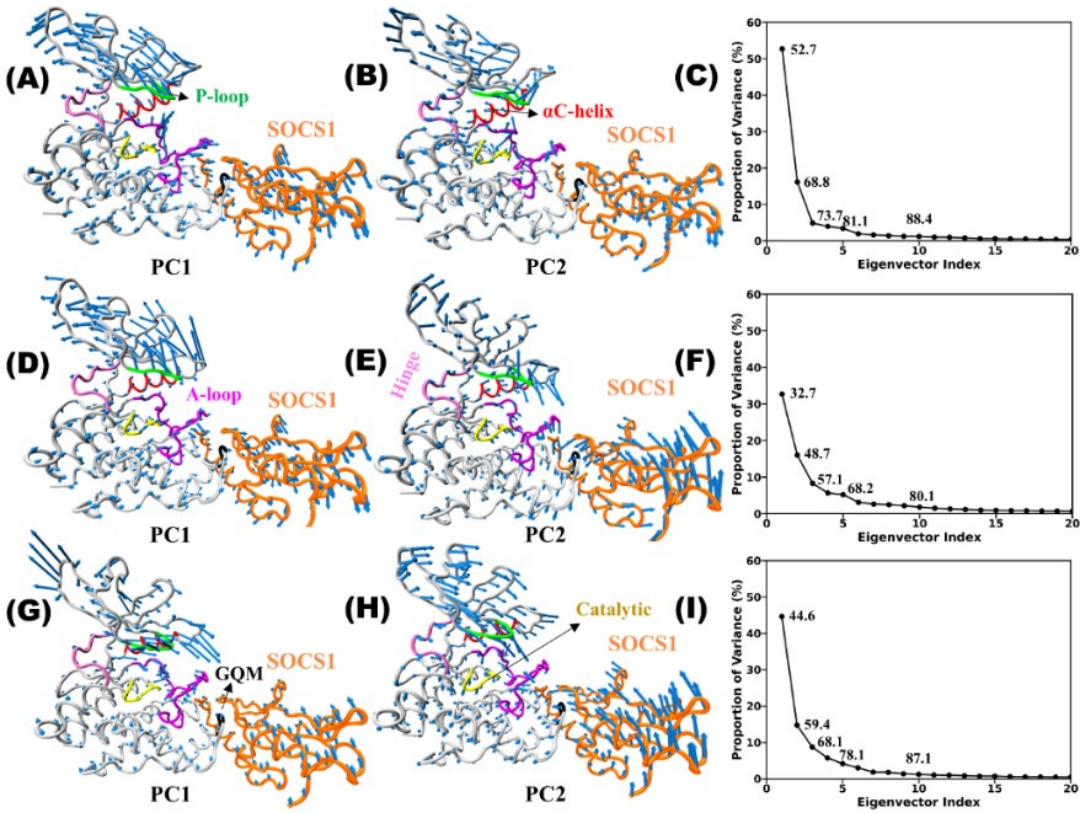


Figure B8: Porcupine illustration of the PC1 and PC2 and percentage of the total variance captured by each eigenvector, with numbers indicating the cumulative variance accounted for all preceding eigenvectors for three phosphorylated states of JAK1/SOCS1 complex; (**A - C**) for com-2pY, (**D - F**) for com-pY1034 and (**G - I**) for com-pY1035, respectively. The blue arrow represents the direction of motion; the length characterizes the amplitude of motion. The highlighted areas are: red (αC-helix), magenta (activation segment), pink (hinge loop), yellow (catalytic domain), black (GQM motif), green (P-loop), and orange (SOCS1).

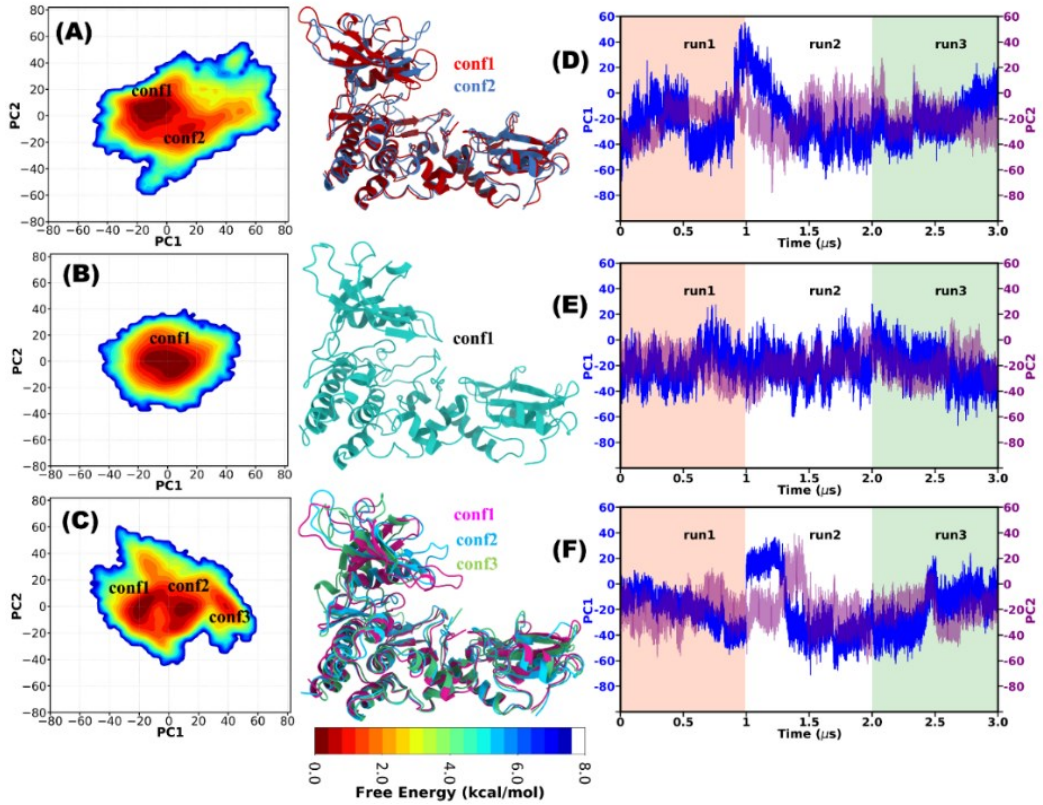


Figure B9: Two-dimensional free energy landscape from MD trajectories for **(A)** com-2pY, **(B)** com-pY1034, and **(C)** com-pY1035 JAK1/SOCS1 along with the lowest energy conformations at 300 K. The first two principal components (PC1 and PC2) were used as reaction coordinates, and free energy values are given in kcal/mol as indicated by the colour bar. The time evolution of the first two principal components for **(D)** com-2pY, **(E)** com-pY1034, and **(F)** com-pY1035 is also shown on its right panel.

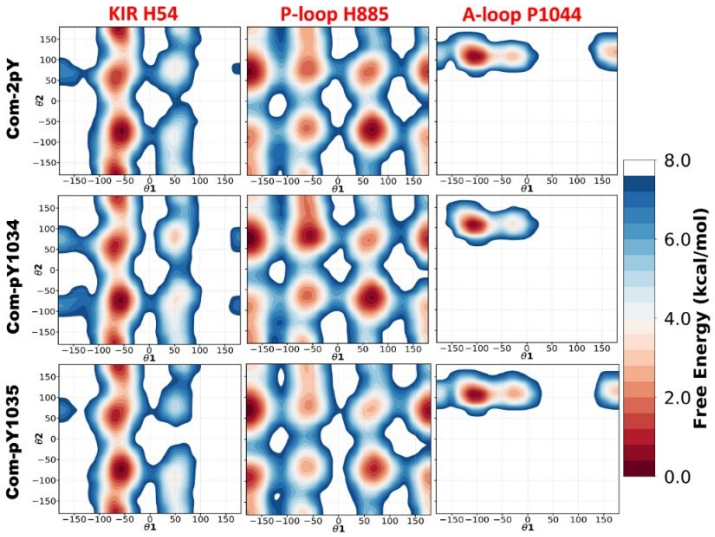


Figure B10: The 2D free energy potential surface based on pseudo-dihedral angles, θ1 and θ2. Where, θ1 and θ2 defined separately for H54 (θ1: N, C_α, C_β, C_γ; θ2: C_α, C_β, C_γ, ND1), H885 (θ1: N, C_α, C_β, C_γ; θ2: C_α, C_β, C_γ, ND1) and P1044 (θ1: O, C, C_α, N; θ2: C, C_α, N, C_γ, ND1) amino acid. Residue H54, H885, and P1044 belong to SOCS1's KIR, JAK1's P-loop, and A-loop regions, respectively.

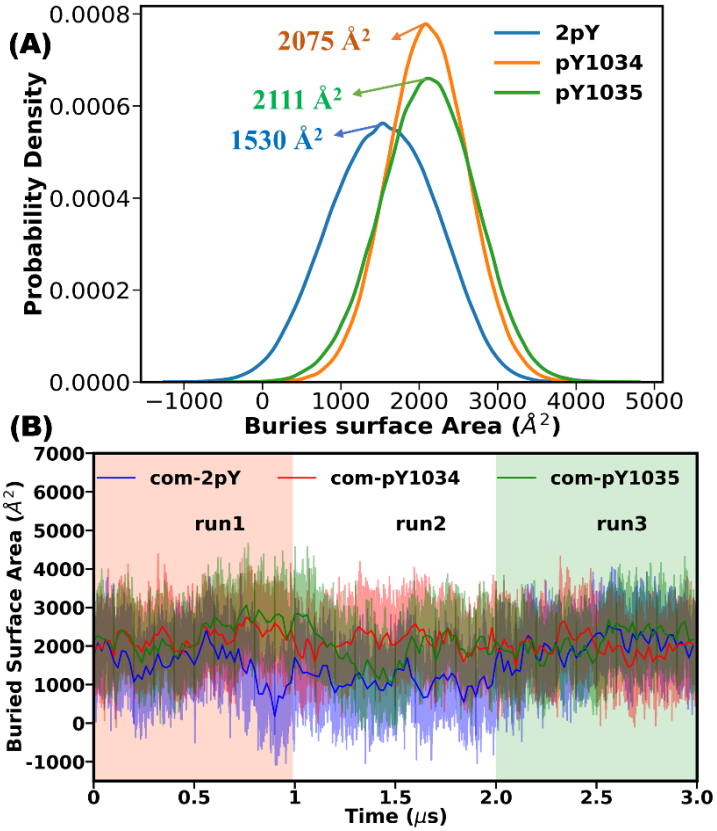


Figure B11: The probability density and time evolution of the buried surface area of JAK1/SOCS1 complexes.

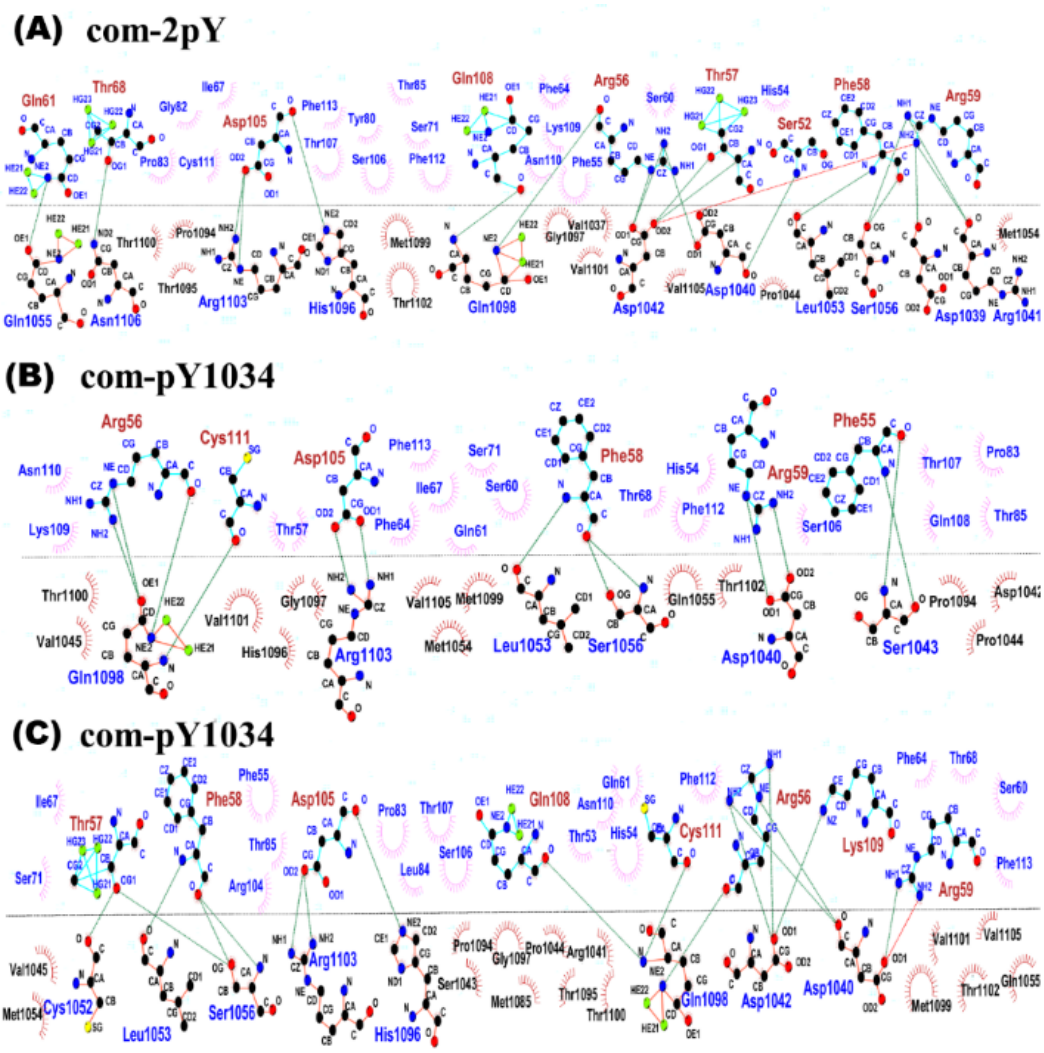


Figure B12: Protein-protein interaction diagrams for the JAK1/SOCS1 complexes. The upper panel corresponds to SOCS1, and the lower panel corresponds to JAK1 in the complex. The plot was generated with the dimplot module of LigPlot+. Hydrogen bonds are shown as a green dotted line, and hydrophobic residues are represented in red.

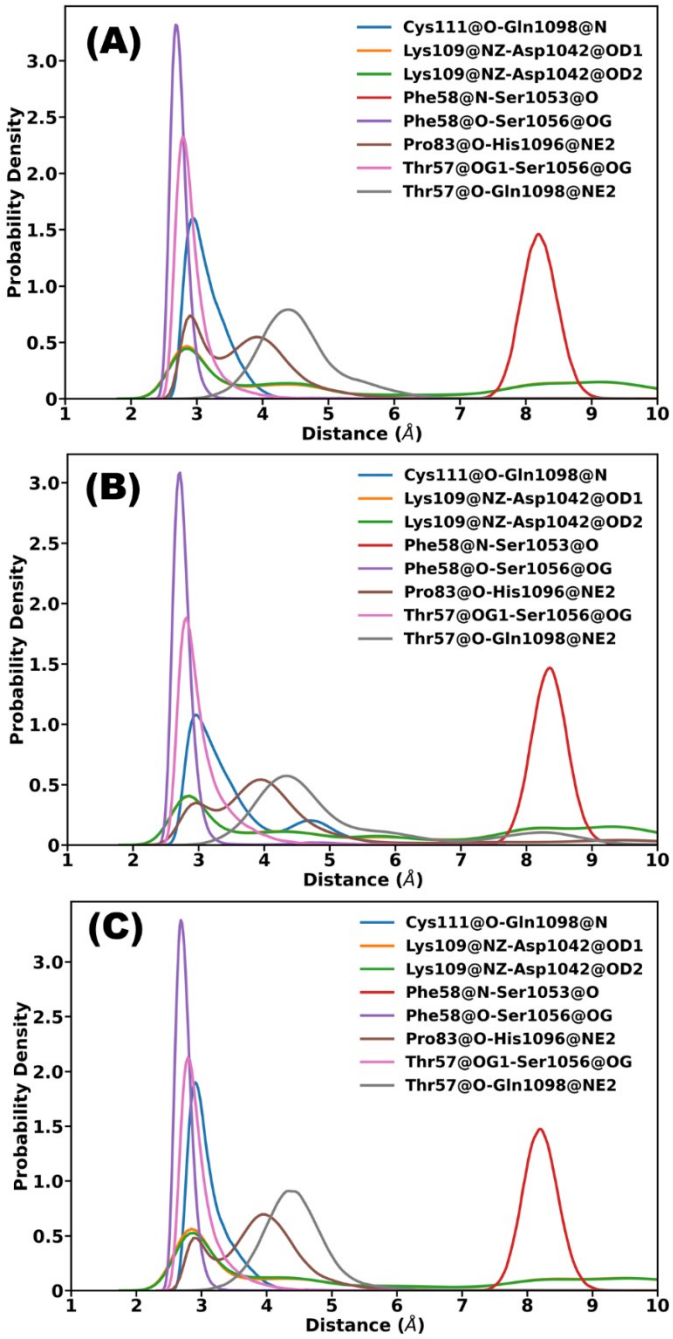


Figure B13: The probability density function of HBs distances for Cys[111], Lys[109], Phe[58], Pro[83] and Thr[57] with JAK1 obtained by analysis of combined 3 μs trajectory. (**A**) com-2pY; (**B**) com-pY1034 and (**C**) com-pY1035, respectively.

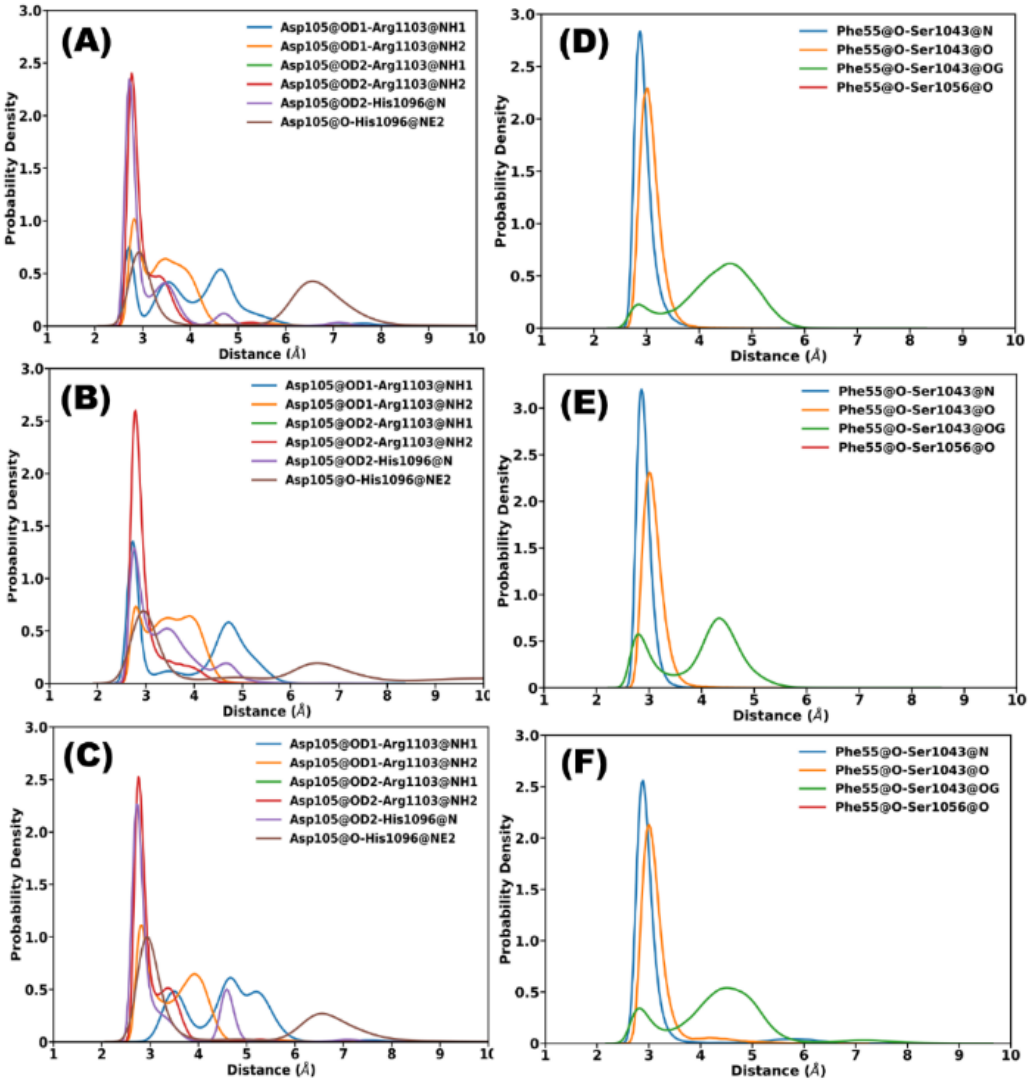


Figure B14: The probability density function of HBs distances for Asp^{105} and Phe^{55} (BC-loop and KIR of SOCS1) with JAK1 obtained by analysis of combined 3 μs trajectory. (**A & D**) com-2pY; (**B & E**) com-pY1034 and (**C & F**) com-pY1035, respectively.

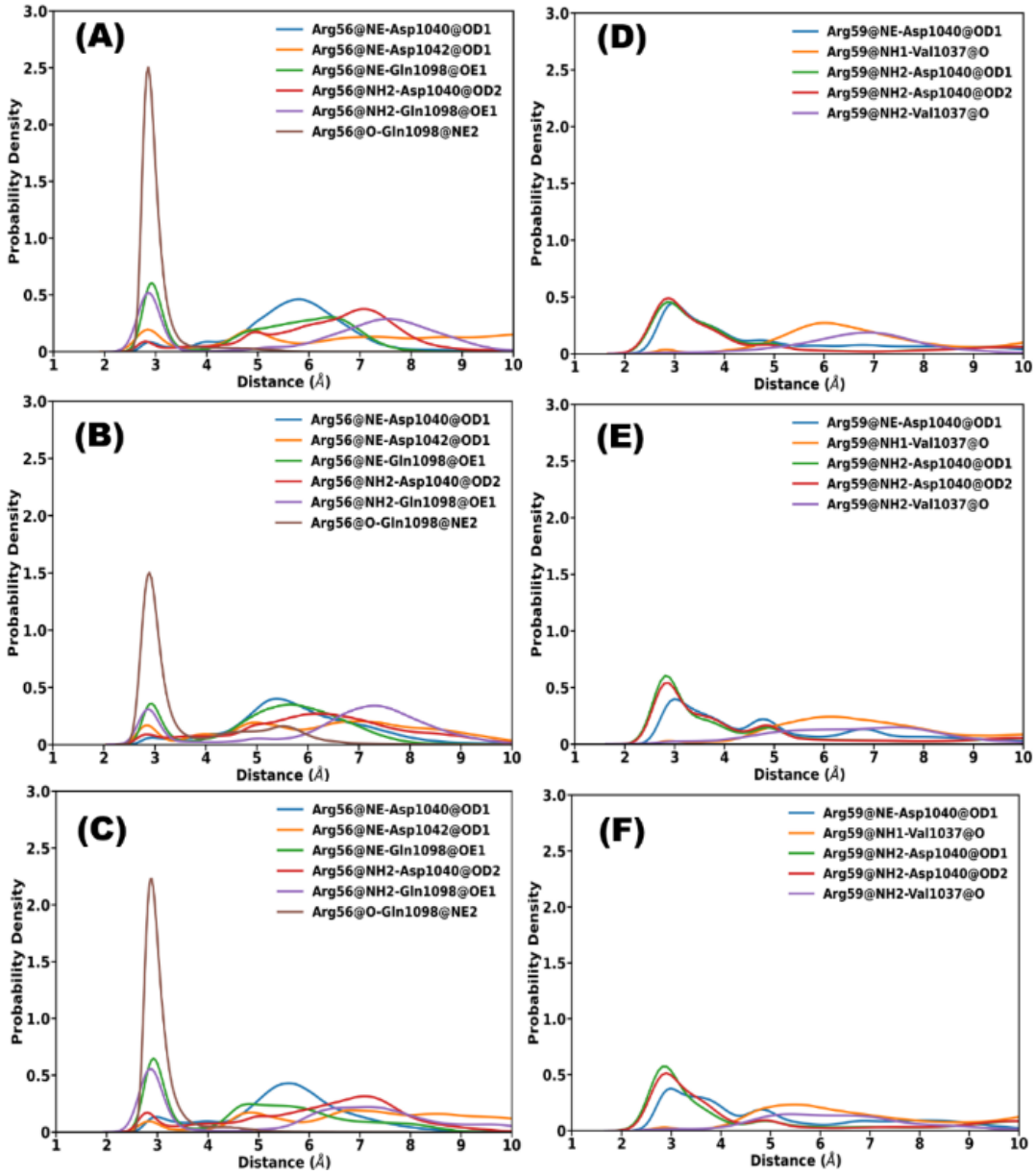


Figure B15: The probability density function of HBs distances for Arg^{56} and Arg^{59} (KIR of SOCS1) with JAK1 obtained by analysis of combined 3 μs trajectory. (**A & D**) com-2pY; (**B & E**) com-pY1034 and (**C & F**) com-pY1035, respectively.

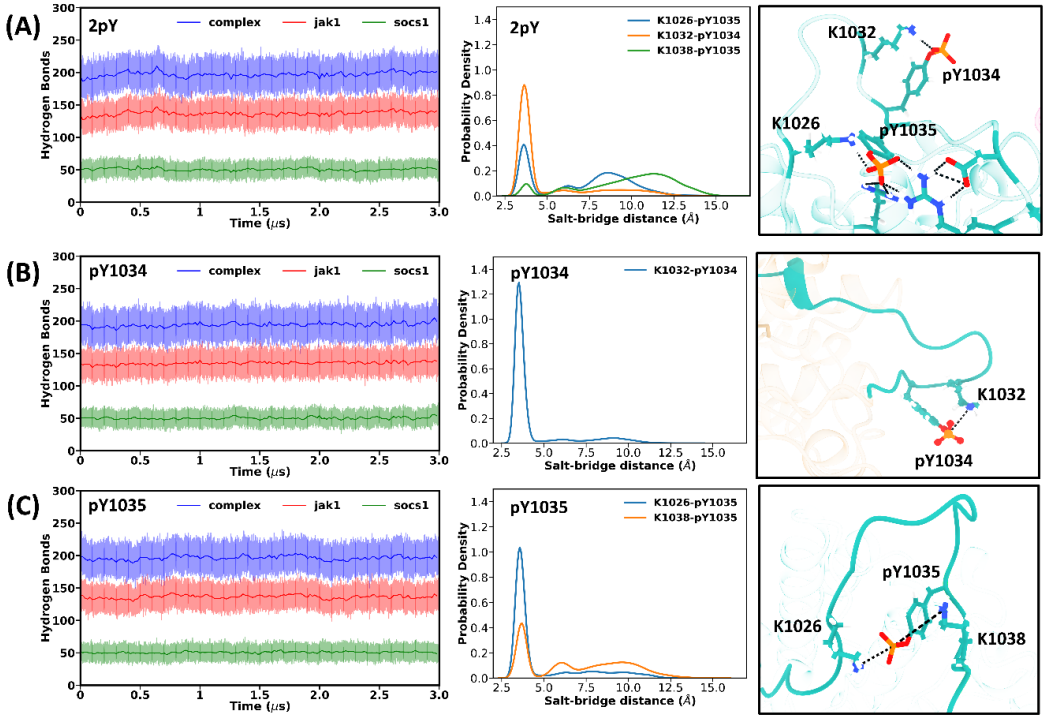


Figure B16: The time evolution of the total number of internal hydrogen bonds formed by the amino acid residues of JAK1 and SOCS1. The probability density of the salt-bridges formed by phosphorylated two tyrosine residues is shown in the right panel.

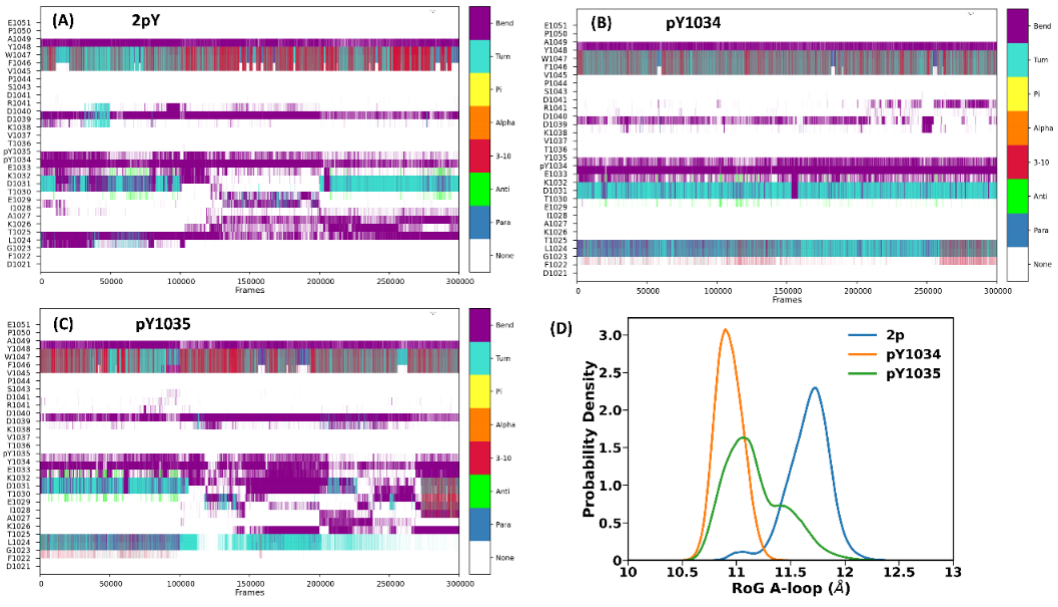


Figure B17: The time-dependent behavior of secondary structure prediction of A-loop segment of JAK1. The probability density of compactness (radius of gyration) of the A-loop segment.

Appendix C

Table C1: The mean value of the distance between donor hydrogen and acceptor.

Hydrogen bonds (HRD—αF helix)	**apo**	**CI-bound**	**AI-bound**
His974@H **---** Asp1036@OD1	1.9	1.9	1.9
Arg975@H --- Asp1036@OD2	2.0	1.9	1.9

The average value is calculated based on the MD equilibrium simulation.

Table C2: Average network properties of three systems, generated through PSNs.

Parameters	**apo**	**CI-bound**	**AI-bound**
I_{min}	3.78	3.52	3
No. of linked nodes	273	279	284
No. of Links	309	321	343
No. of hubs	27	32	46
No. of linked mediated hubs	155	165	183
No. of communities	10	10	9
No. of nodes involved in communities	71	65	86
No. of links involved in communities	97	92	122

Table C3: Energetic components of the binding free energy for phytochemicals bound to the JAK1. All values are given in kcal/mol.

	ΔE_{vdW}	ΔE_{elec}	ΔG_{pol}	ΔG_{np}	ΔG_{bind}
CI-bound					
Run1	-46.97 ± 0.05	-34.57 ± 0.07	54.52 ± 0.07	-3.79 ± 0	-30.81 ± 0.06
Run2	-45.55 ± 0.05	-33.60 ± 0.08	53.57 ± 0.08	-3.79 ± 0	-29.37 ± 0.06
Run3	-43.82 ± 0.05	-29.88 ± 0.07	47.81 ± 0.08	-3.83 ± 0	-29.72 ± 0.05
Avg.	-45.47 ± 0.08	-32.68 ± 0.12	51.96 ± 0.13	-3.80 ± 0	-29.96 ± 0.09
AI-bound					
Run1	-71.35 ± 0.04	-26.27 ± 0.12	64.70 ± 0.20	-6.41 ± 0	-39.33 ± 0.17
Run2	-70.89 ± 0.05	-23.70 ± 0.09	60.57 ± 0.09	-6.38 ± 0	-40.40 ± 0.08
Run3	-73.50 ± 0.05	-21.38 ± 0.08	58.23 ± 0.08	-6.37 ± 0	-43.02 ± 0.07
Avg.	-71.91 ± 0.07	-23.78 ± 0.17	61.17 ± 0.23	-6.38 ± 0	-40.92 ± 0.20

Table C4: Total energy decomposed into their residual levels for CI and AI bound systems*.

Residues	JAK2 region	Total energy contribution (kcal/mol)	
		CI-bound	AI-bound
Leu^{855}	**P-loop**	**-1.48**	**-1.23**
Val^{863}	**P-loop**	**-2.35**	**-1.06**
Ala^{880}	β3 strands	-1.09	-0.89
Glu^{898}	αC helix	0.03	0.60
Ile^{901}	αC helix	-0.01	-1.47
Leu^{902}	αC helix	-0.03	-1.80
Val^{911}	αC – β4 loop	-0.57	-1.35
Met^{929}	Hinge	-0.68	-2.22
Glu^{930}	Hinge	-1.21	0.08
Tyr^{931}	**Hinge**	**-1.67**	**-1.90**
Leu^{932}	**Hinge**	**-1.64**	**-2.81**
Leu^{983}	**β7 strands**	**-2.18**	**-1.47**
Gly^{993}	DFG-1 motif	-0.23	-1.78
Asp^{994}	DFG motif	-0.75	-1.16
Phe^{995}	DFG motif	-0.03	-1.95

*values lower than -1.0 kcal/mol are listed here.

Table C5: Hydrogen bond (HB) bonds formed between JAK2, and inhibitors (type-I/type-II) and the corresponding average distance and percentage of HB occupancy determined using the 3 μs trajectories in the MD simulations.

Binding couples		Distance (Å)	Hydrogen bonds occupancy (%)
Acceptor (A)	Donor (D)	A-D distance	MD analysis
CI-bound			
Glu^{930}@O	Lig@N12	2.82	84.14
Lig@N17	Leu^{932}@N	2.92	79.07
AI-bound			
Lig@O17	Asp^{994}@N	2.84	66.75
Leu^{932}@O	Lig@N43	2.85	50.67
Lig@N39	Leu^{932}@N	2.92	40.84

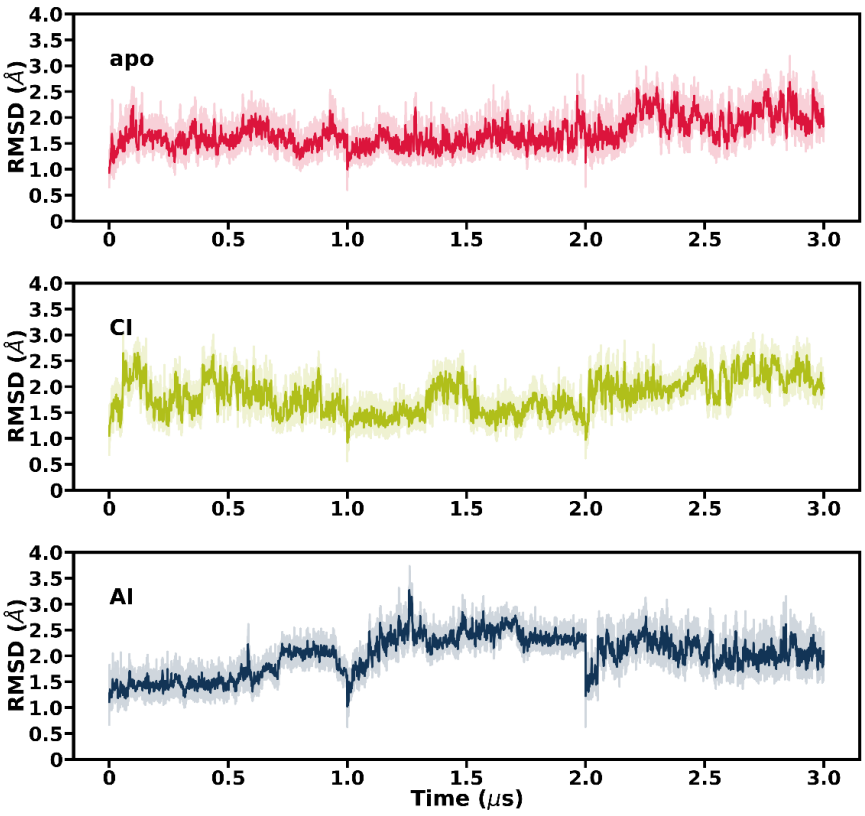


Figure C1: Backbone atoms root mean square deviation (RMSD) plot of the apo, competitive inhibitor (CI) bound, and allosteric inhibitor (AI) bound complexes. The data are derived from the three independent simulations with different initial velocities.

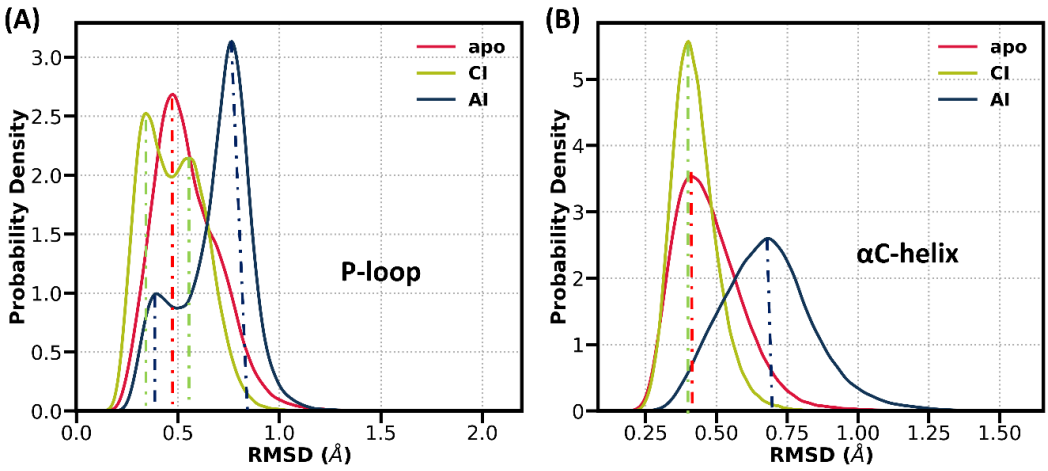


Figure C2: The probability of RMSD distributions of (A) P-loop, (B) αC helix backbone atoms of all residues. Their probable RMSD values are shown in dotted lines with their corresponding color code.

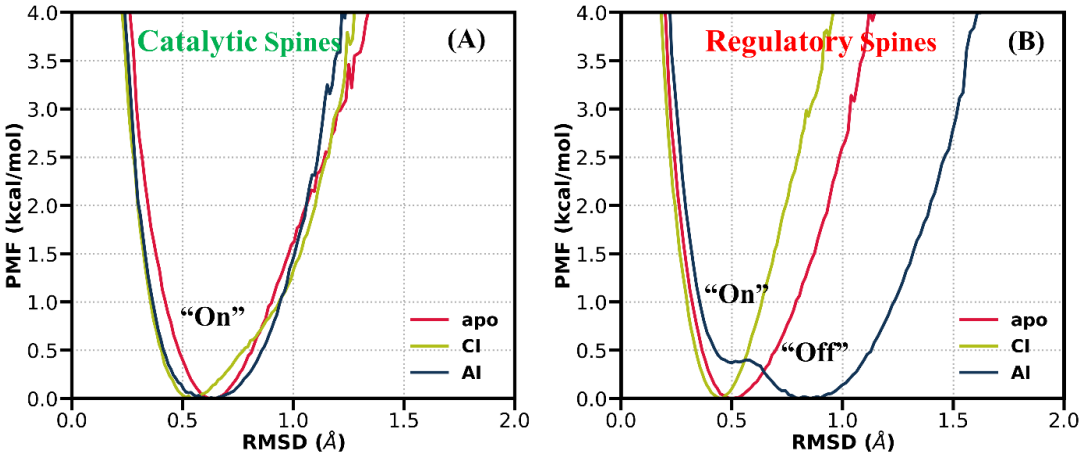


Figure C3: The one-dimensional potential of mean force (PMF) of (**A**) catalytic and (**B**) regulatory hydrophobic spines. Here the PMFs are a function of the RMSD of regulatory and catalytic spines relative to their energy-minimized structures. All heavy atoms in the spines are used in this calculation of RMSD. The "On" and "Off" states correspond to the hydrophobic spines formed and broken conformations.

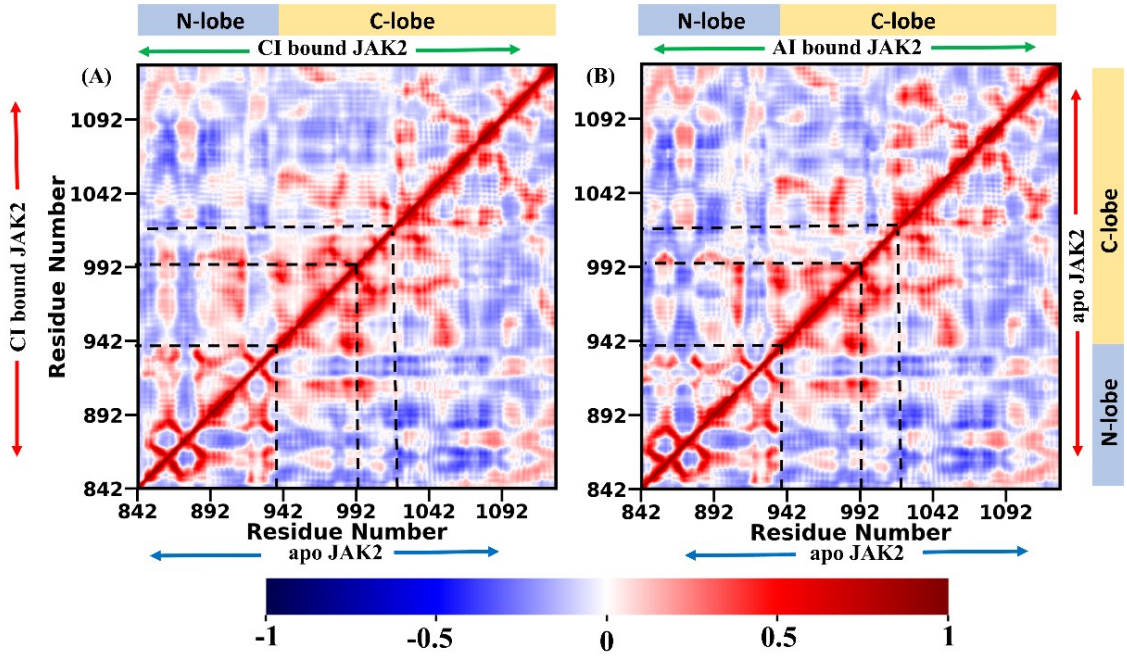


Figure C4: The dynamic cross-correlation residue motion of three systems. The upper triangle matrix corresponds to CI and AI-bound JAK2, and the lower triangle matrix corresponds to apo-JAK2. The dotted line box denotes the N-terminal residues, and the second box represents the A-loop residues.

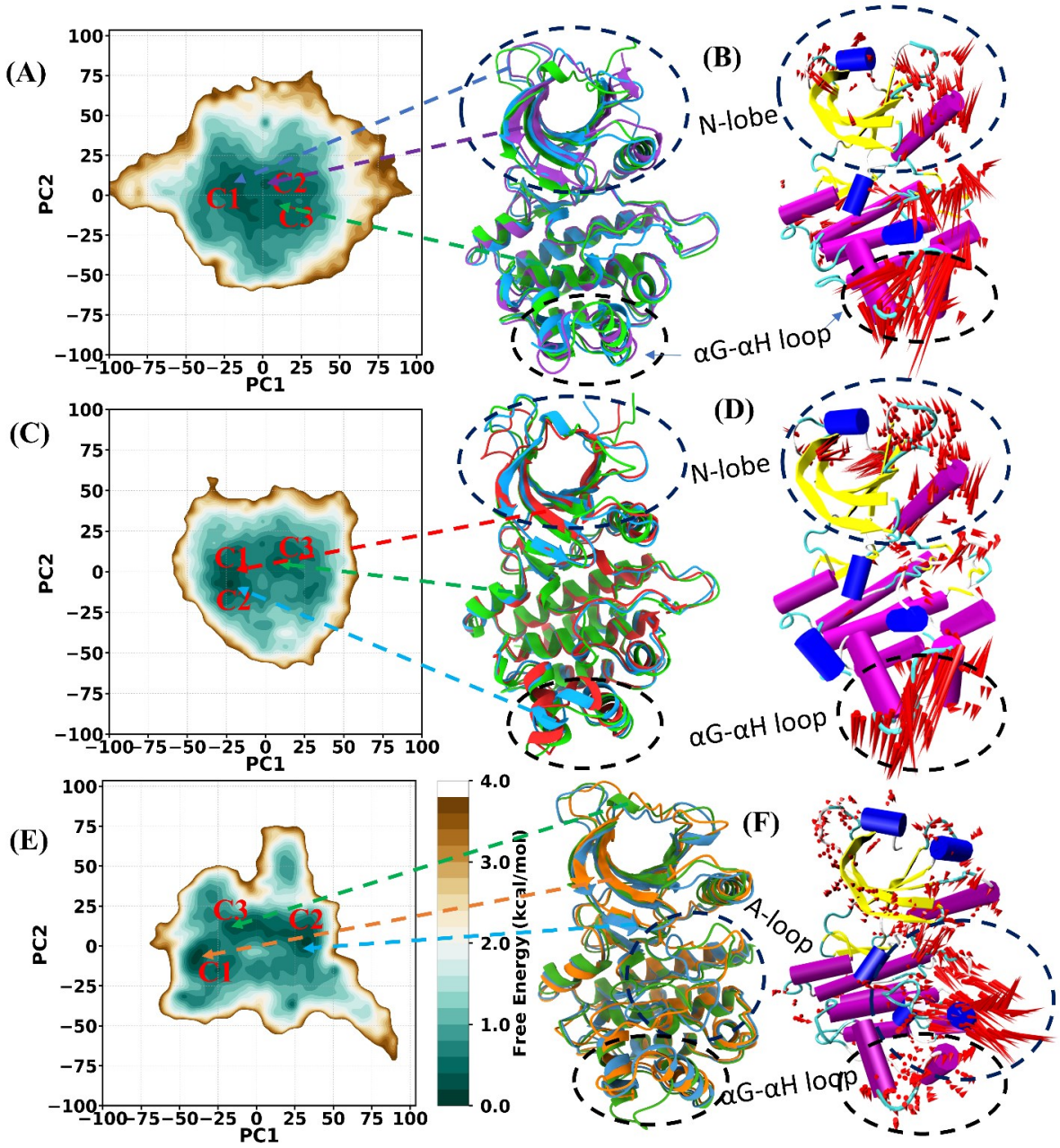


Figure C5: The two-dimensional free energy surfaces generated by projecting the first two principal components (PC1 and PC2) as a reaction coordinate during simulations, shown in left panels. The low energy conformations are highlighted (C1, C2, C3), and their superimposed representative structures are shown on middle panels by the K-means clustering analysis. The PCA porcupines of eigenvector 1 map on the 3D-structures shown in the right panels. (A-B) for apo, (C-D) CI-bound, and (E-F) AI-bound JAK2 kinase.

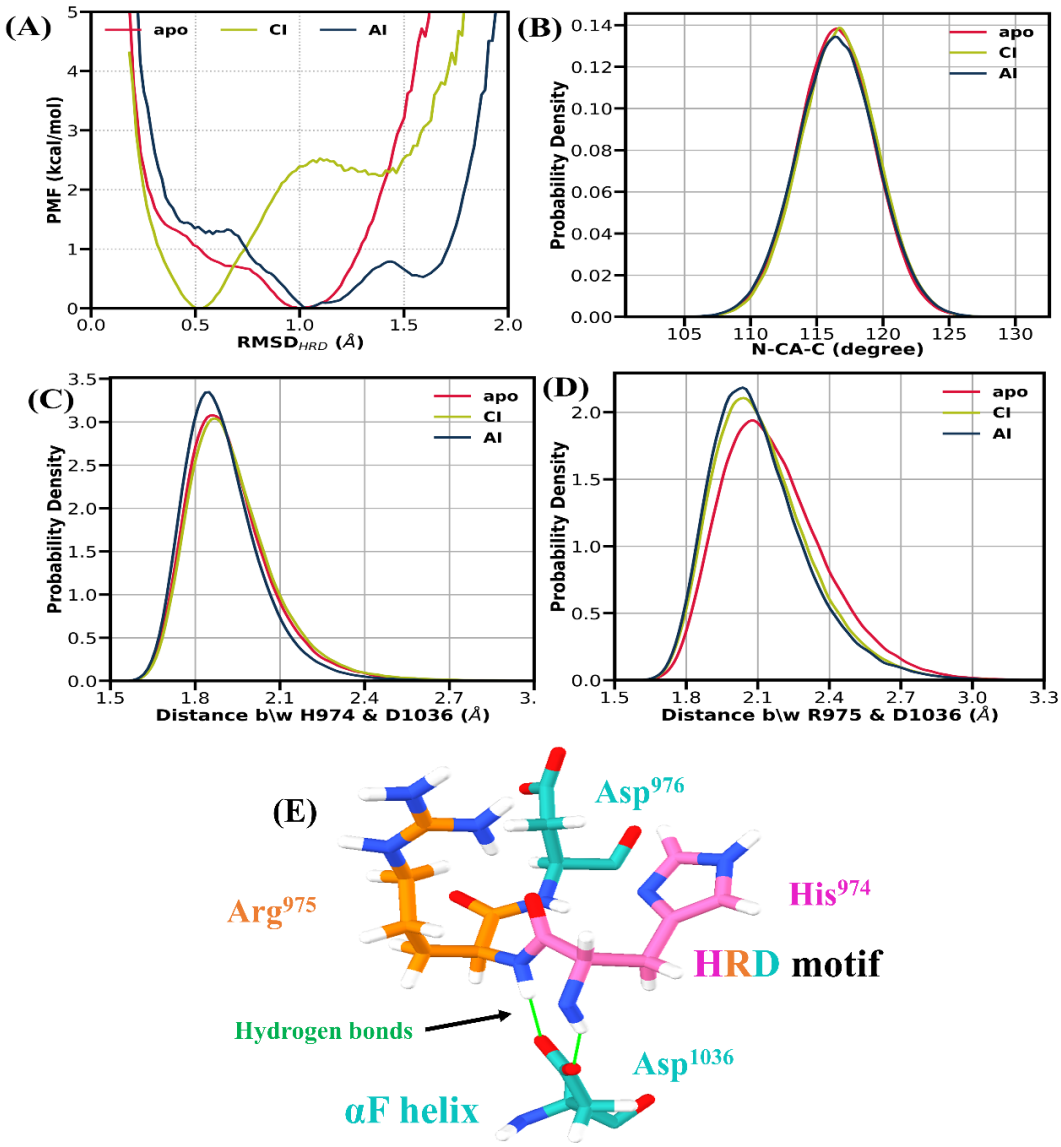


Figure C6: (**A**) The one-dimensional potential of mean force (PMF) as a function of HRD motif heavy atoms RMSD. (**B**) The probability density of the backbone bond angle (N-Cα-C) of Arg^{975}. (**C**) The probability density of the distance between His^{974} (HRD motif) amide hydrogen atom and Asp^{1036} (αF helix) side-chain oxygen atom. (**D**) Probability density of the distance between Arg^{975} amide hydrogen atom and Asp^{1036} side-chain oxygen atom. Since the two-oxygen atom in the sidechain of Asp1036 is equivalent chemically, distances from the hydrogen atom to both oxygen atoms are estimated, and the smaller one is used to make these probability density plots. **(E)** The representative structure of hydrogen bond formation.

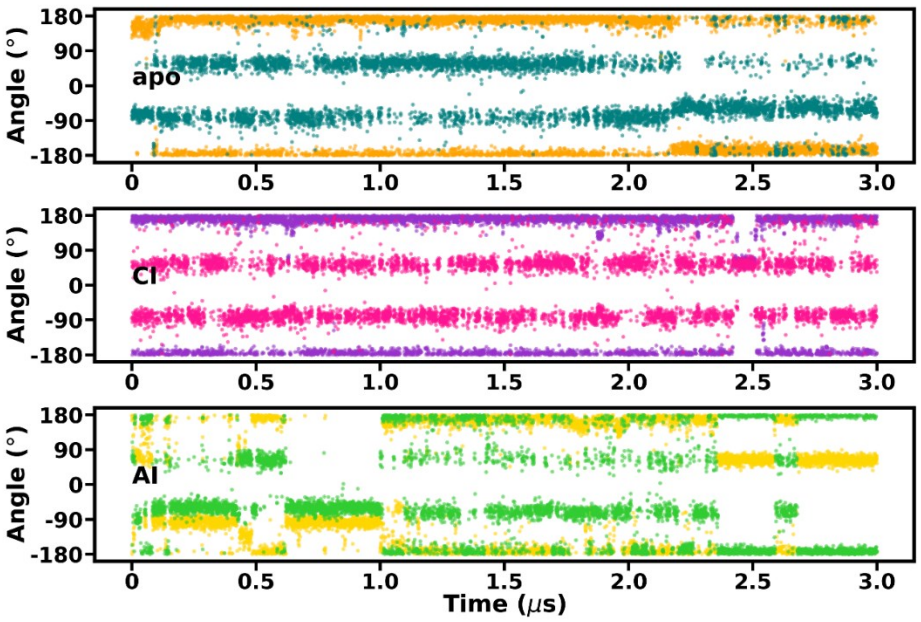


Figure C7: Time dependence of the dihedral angle of Arg^{975} in apo, CI-bound, and AI-bound JAK2 kinase.

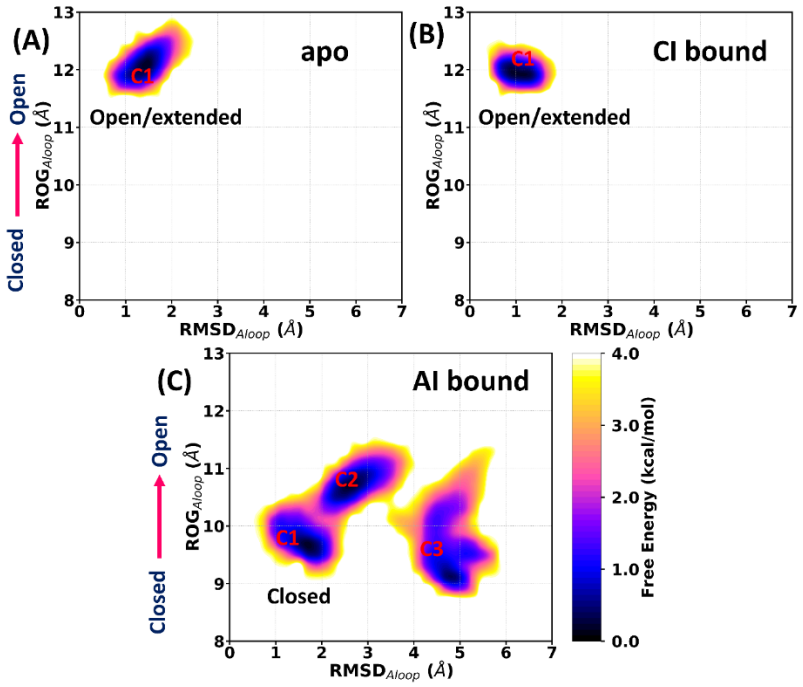


Figure C8: The two-dimensional free energy surfaces as a function of backbone atoms RMSD and radius of gyration (RoG) of A-loop segment of JAK2 in (A) apo, (B) CI-bound, and (C) AI-bound inhibitors, respectively. Local minima are highlighted in C1, C2, and C3.

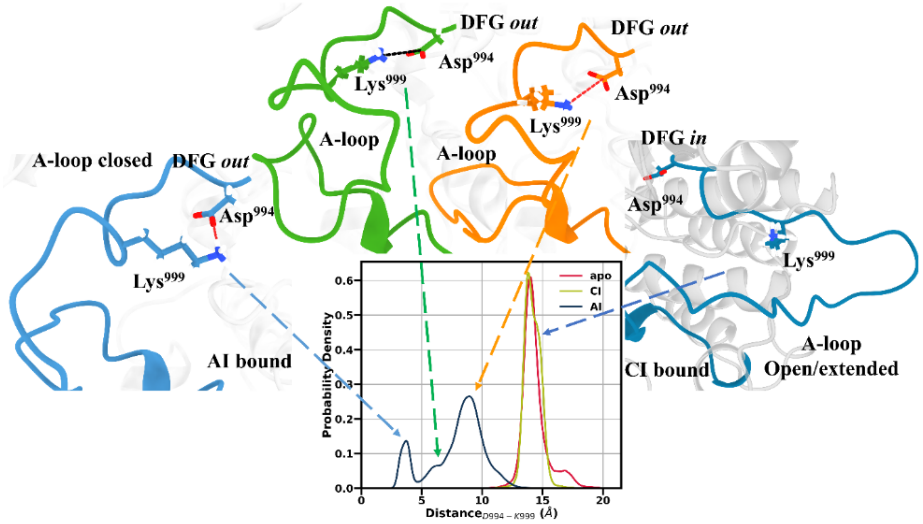


Figure C9: Interactions enhanced the particular rearrangement of the A-loop of JAK2 upon binding of type I/II inhibitors. (A) salt-bridge between Asp^{994} (DFG) and Lys^{999} (A-loop).

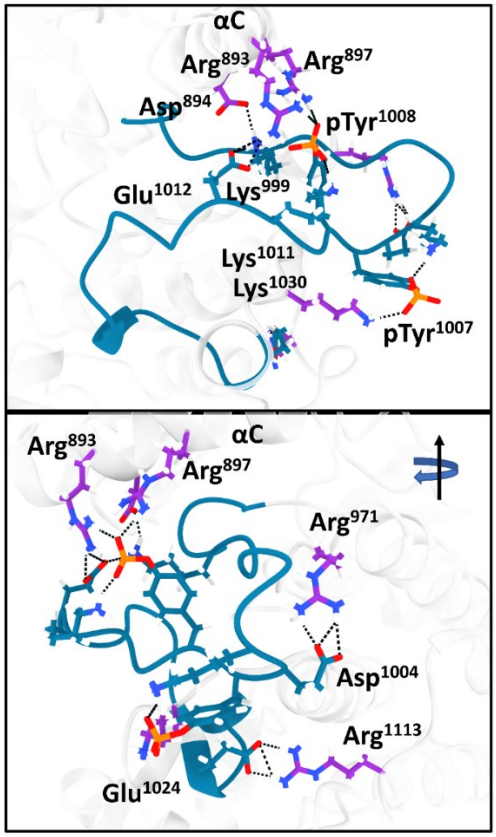


Figure C10: The interaction pattern and residues involved in the open/extended conformations of A-loop segment in active-like apo and CI-bound JAK2. The hydrogen bonds and salt bridges are shown as dotted lines.

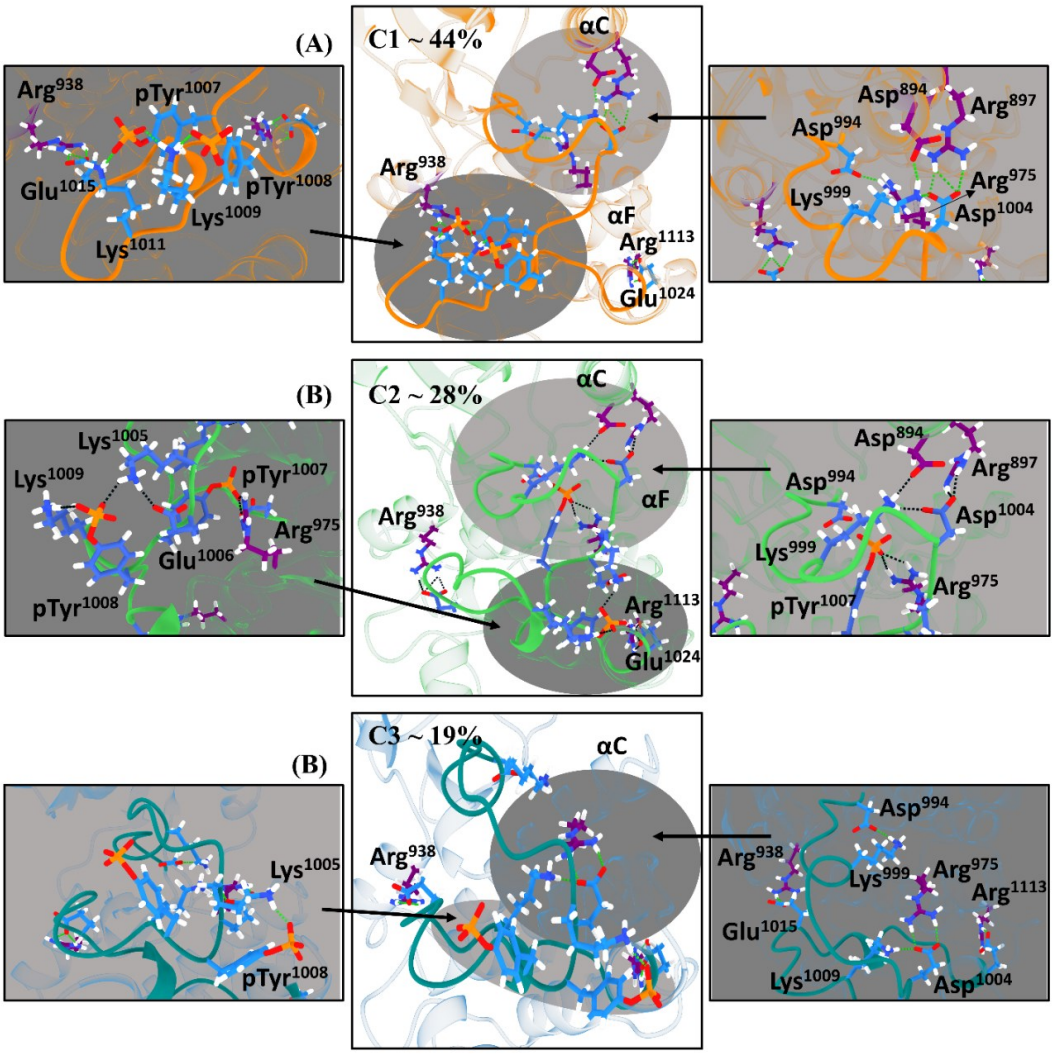


Figure C11: The residual interactions pattern of activation loop and its nearby residues to maintain the unique conformations of closed A-loop in inactive-like JAK2. These conformations are obtained from the K-means clustering analysis of dPCA eigenvectors of inactive DFG-out AI-bound JAK2 kinase during GaMD simulations. The hydrogen bonds and salt bridges are shown as dotted lines.

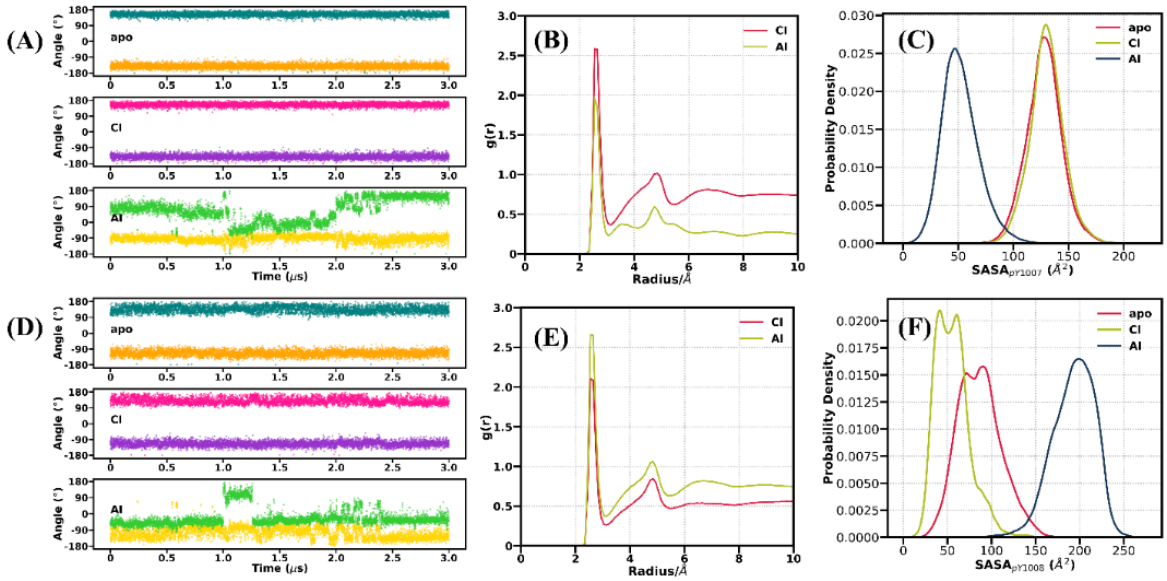


Figure C12: The time evolution of the dihedral angles of phosphorylated residues (Tyr^{1007}/Tyr^{1008}), the radial distribution functions (RDF, g(r)) of water accessibility around the residues, and their solvent accessible surface area (SASA).

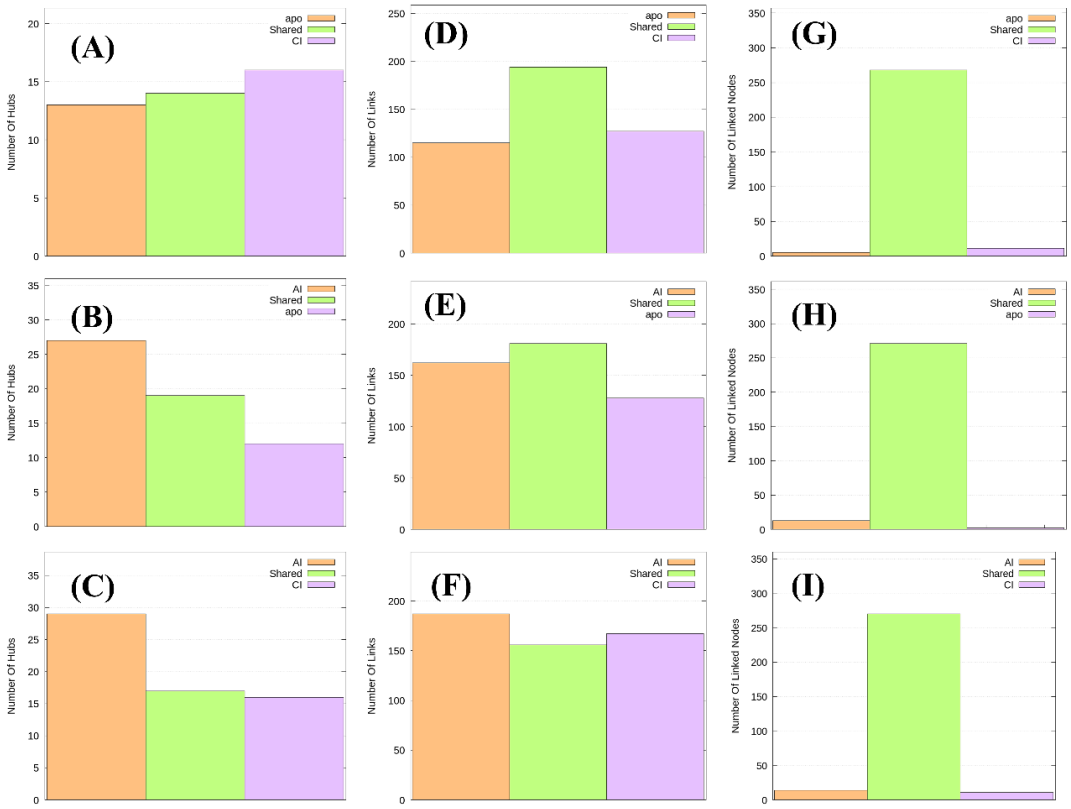


Figure C13: Network comparison of apo, CI-bound, and AI-bound JAK2. (A-C) shows the number of hubs; (D-F) the number of links in the structure; (G-I) the number of linked nodes in the systems. The first row represents the difference between the apo and CI-bound complex, the middle row for apo and AI-bound JAK2, and the below row is CI-bound and AI-bound JAK2.

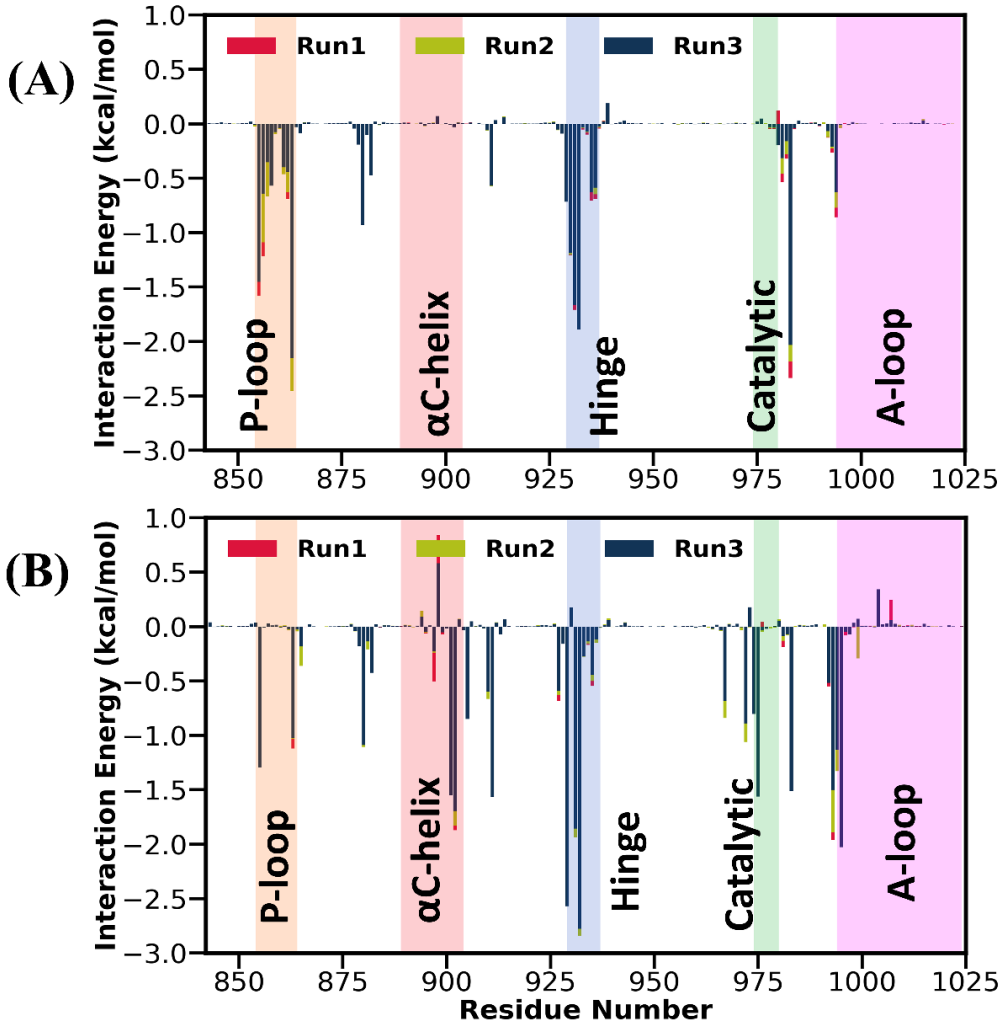
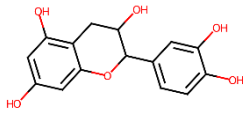
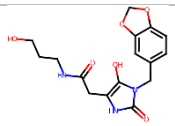
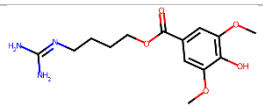
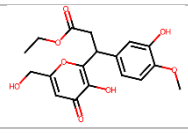
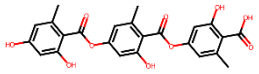
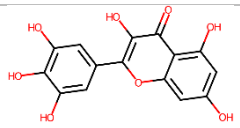
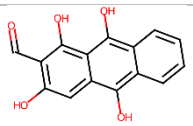
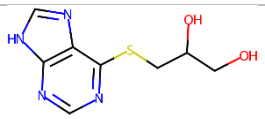
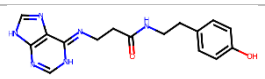
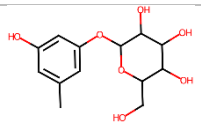
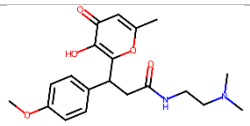
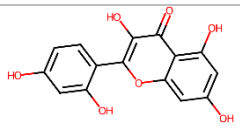
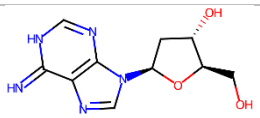
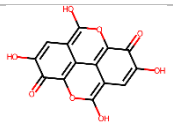
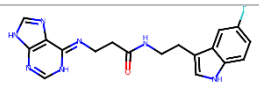
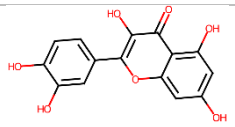
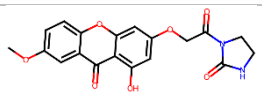
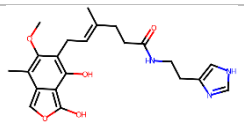
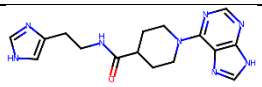
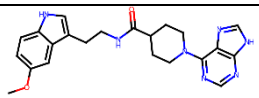
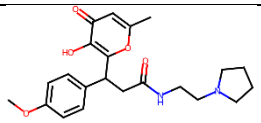
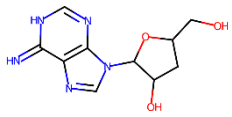


Figure C14: The per-residue binding free energy decomposition of (A) CI-bound and (B) AI-bound JAK2 systems of three independent replica runs.

Appendix D

Table D1: List of top hit compounds with their Glide XP-Docking scores in kcal/mol, obtained from the virtual screening workflow and ADMETlab 2.0.

Compounds		**Physico-chemical property**			**Glide-XP**
	Structure	**MW[a] (g/mol)**	**Vol_{vdW}[b]**	**TPSA[c]**	**Docking score**
Epicatechin		290.27	279.25	110.38	-10.9
STOCK1N-82656		349.34	329.13	125.81	-10.8
Leonurine		311.33	305.90	129.39	-10.8
STOCK1N-77498		364.35	354.64	126.43	-10.8
STOCK1N-37646		468.40	454.25	170.82	-10.7
Myricetin		318.23	291.55	151.16	**-10.7**
Lucidin		270.24	265.19	97.99	-10.7
STOCK1N-31772		226.26	199.34	94.92	-10.4
STOCK1N-71990		326.35	322.09	119.05	-10.4
Sakakin		286.28	269.91	119.61	-10.3
STOCK1N-98615		374.40	384.85	92.01	-10.2
Morin		302.23	282.77	131.36	-10.2

STOCK1N-51172		251.24	226.65	120.04	-10.2
Ellagic Acid		302.19	265.71	141.34	-10.1
STOCK1N-72541		367.16	353.76	114.61	-10.1
Quercetin		302.04	282.76	131.36	-10.1
STOCK1N-76211		384.10	362.75	118.31	-10.1
STOCK1N-85809		413.47	419.25	120.61	-10.1
STOCK1N-68642		340.38	329.37	115.48	-10.1
STOCK1N-68737		419.50	417.11	111.82	-10.1
STOCK1N-98539		400.20	410.89	92.01	-10.1
Cordycepin		251.24	226.65	120.04	-10.0

Acceptable ranges- [a]MW: Molecular Weight, (0-500) g/mol; [b]Vol_{vdW}: van der Waals volume; [c]TPSA: Topological Polar surface Area, (0-140)

Table D2: Prediction of essential ADME properties of all lead compounds obtained from QikProp.

Compound Name	logP(o/w)[a]	logS[b]	Caco-2d[c]	logHERG[d]	PMDCK[e]	%Human Oral absorption[f]
Epicatechin	0.5	-2.6	52.7	-4.8	26.7	61
Leonurine	0.9	-3.5	39.8	-5.0	25.2	61
Myricetin	-0.3	-2.6	**7**	-5.0	**2.3**	27
Lucidin	0.6	-2.3	96.5	-4.5	39.4	66
Sakakin	-0.8	-2.0	51.5	-4.4	25.3	53
Morin	0.4	-2.8	**21.7**	-5.0	**7.9**	53
Ellagic Acid	-1.3	-1.9	**7**	-3.8	**2.2**	34
Cordycepin	-1.0	-1.7	99.7	-3.8	40.9	57
STOCK1N-82656	0.62	-2.4	79.5	-3.6	57.5	65
STOCK1N-77498	1.2	-3.4	60.0	-4.9	**23.6**	66
STOCK1N-37646	3.3	-5.9	**2.11**	-4.3	**0.8**	52
STOCK1N-31772	-0.2	-1.6	153.45	-4.2	93.4	65
STOCK1N-71990	1.4	-3.1	62.5	-4.6	44.4	67
STOCK1N-98615	1.6	-2.5	86.4	-4.8	61.2	71
STOCK1N-51172	-1.0	-1.8	96.0	-3.9	39.3	57
STOCK1N-72541	2.4	-4.5	102.9	**-5.1**	131.2	77
Quercetin	0.3	-2.9	19.3	**-5.1**	**6.9**	52
STOCK1N-76211	2.3	-4.3	110.2	**-5.6**	45.6	77
STOCK1N-85909	2.4	-4.3	97.2	-4.2	67.0	76
STOCK1N-68642	0.8	-3.4	119.2	-4.3	79.4	69
STOCK1N-68737	2.5	-4.3	316.2	-4.4	198.3	86
STOCK1N-98539	2.0	-3.0	82.8	**-5.0**	60.1	73

a-Predicted octanol/water partition coefficient log P (acceptable range −2.0 to 6.5)
b—Predicted aqueous solubility log S in mol/L (acceptable range: −6.5 to 0.5)
c—Predicted Caco-2 cell permeability in nm/s (acceptable range: < 25 is poor and > 500 is great)
d—Predicted IC 50 value for blockage of HERG K+ channels (concern below −5.0)
e—Predicted apparent MDCK cell permeability in nm/s (acceptable range: < 25 is poor > 500 is great)
f—Percentage of human oral absorption (acceptable range: < 25% is poor and > 80% is high)

Table D3: The drug-likeliness properties calculated using the SwissADME webserver. Lipinski's rule of 5 helps in distinguishing between drug-like and non-drug-like molecules. It predicts a high probability of success or failure due to drug-likeness for molecules complying with two or more rules.

Compounds (CID)	Compound Name/ID	Molweight[a]	Octanol/water partition coefficient[b] (logP)	No. of H-bond donors[c]	No. of H-bond acceptors[d]	Molecular refractivity[e]
9064	**Epicatechin**	290.27	1.55	5	6	72.62
161464	**Leonurine**	311.33	0.62	5	8	81.40
5281672	**Myricetin**	318.23	1.71	**6**	8	75.71
10163	**Lucidin**	270.24	1.36	3	5	69.00
70700401	**Sakakin**	286.28	-1.12	5	7	67.24
5281670	**Morin**	302.23	2.01	5	7	74.05
5281855	**Ellagic Acid**	302.19	1.24	4	8	68.45
6303	**Cordycepin**	251.24	-0.95	4	7	61.35
75531711	**STOCK1N-82656**	349.34	0.38	3	6	92.29
71753414	**STOCK1N-77498**	364.35	1.64	3	8	91.61
135728	**STOCK1N-37646**	468.40	3.57	5	10	119.19
3579846	**STOCK1N-31772**	226.26	-0.2	3	5	55.34
42506840	**STOCK1N-71990**	326.35	1.68	4	5	89.73
135640282	**STOCK1N-98615**	374.40	2.25	2	6	102.65
13730	**STOCK1N-51172**	251.24	-0.37	3	6	61.51
42507459	**STOCK1N-72541**	367.40	2.6	4	5	99.52
5280343	**Quercetin**	302.23	1.99	5	7	78.04
52905134	**STOCK1N-76211**	384.30	1.86	2	7	106.43
91638074	**STOCK1N-85909**	413.47	3.12	3	6	111.83
16407730	**STOCK1N-68642**	340.38	1.11	3	5	95.19
16407819	**STOCK1N-68737**	419.50	2.87	3	5	121.39
135640585	**STOCK1N-98539**	400.50	2.73	2	6	114.07

[a]Molecular weight range: 150-500 g/mol
[b]Octanol/water partition coefficient range (log P) ≤ 4.15
[c]No. Of H-bond Donor ≤ 5
[d]No. of H-bond Acceptor ≤ 10
[e]Molecular refractivity range: 40-130

Table D4: Toxicity prediction of all screened molecules using ProTox-II webserver.

Compound Name	Hepatotoxicity (Probability)	Carcinogenicity (Probability)	Mutagenicity (Probability)	Cytotoxicity (Probability)	Predicted LD50 (mg/kg)	Predicted Toxicity Class
Epicatechin	0.72, Inactive	0.51, Inactive	0.55, Inactive	0.84, Inactive	10000	6
Leonurine	0.88, Inactive	0.54, Inactive	0.62, Inactive	0.67, Inactive	682	4
Myricetin	0.69, Inactive	0.68, **Active**	0.51, **Active**	0.99, Inactive	159	3
Lucidin	0.84, Inactive	0.72, Inactive	0.90, **Active**	0.86, Inactive	5000	5
Sakakin	0.92, Inactive	0.84, Inactive	0.76, Inactive	0.85, Inactive	1380	4
Morin	0.68, Inactive	0.72, Inactive	0.52, Inactive	0.98, Inactive	3919	5
Ellagic Acid	0.83, Inactive	0.59, **Active**	0.84, Inactive	0.90, Inactive	2991	4
Cordycepin	0.83, Inactive	0.68, Inactive	0.69, **Active**	0.58, Inactive	NA	NA
STOCK1N-82656	0.81, Inactive	0.61, Inactive	0.57, Inactive	0.73, Inactive	2700	5
STOCK1N-77498	0.85, Inactive	0.69, Inactive	0.78, Inactive	0.80, Inactive	2000	4
STOCK1N-37646	0.55, Inactive	0.88, Inactive	0.79, Inactive	0.86, Inactive	962	4
STOCK1N-31772	0.50, Inactive	0.65, Inactive	0.54, **Active**	0.63, Inactive	2000	4
STOCK1N-71990	0.78, Inactive	0.55, Inactive	0.55, **Active**	0.72, Inactive	175	3

STOCK1N-98615	0.87, Inactive	0.57, Inactive	0.71, Inactive	0.64, Inactive	1000	4
STOCK1N-51172	0.51, Inactive	0.69, Inactive	0.91, Inactive	0.64, Inactive	5000	5
STOCK1N-72541	0.70, Inactive	0.61, Inactive	0.54, **Active**	0.73, Inactive	500	4
Quercetin	0.69, Inactive	0.68, Inactive	0.51, **Active**	0.99, Inactive	159	3
STOCK1N-76211	0.82, Inactive	0.68, Inactive	0.59, Inactive	0.70, Inactive	1780	4
STOCK1N-85909	0.83, Inactive	0.52, Inactive	0.70, Inactive	0.68, Inactive	1000	4
STOCK1N-68642	0.78, Inactive	0.64, Inactive	0.52, Inactive	0.90, Inactive	175	3
STOCK1N-68737	0.79, Inactive	0.58, Inactive	0.55, **Active**	0.74, Inactive	369	4
STOCK1N-98539	0.89, Inactive	0.60, Inactive	0.74, Inactive	0.64, Inactive	1000	4

Table D5: Energetic components of the total binding free energy of natural compounds bound to the JAK1 for 1000 frames of 100 ns trajectory. All values are given in kcal/mol. The standard error of the mean is provided in parentheses.

Compounds	ΔE_{vdW}	ΔE_{elec}	ΔG_{pol}	ΔG_{np}	ΔG_{bind}
Epicatechin	-28.66 (0.09)	-28.62 (0.30)	37.50 (0.21)	-3.34 (0.0)	-23.14 (0.14)
STOCK-82656	-41.02 (0.10)	-22.72 (0.25)	44.66 (0.24)	-3.98 (0.0)	-23.07 (0.17)
Leonurine	-38.35 (0.11)	-22.93 (0.27)	40.33 (0.23)	-4.05 (0.0)	-25.00 (0.18)
STOCK-77498	-35.60 (0.11)	-24.87 (0.36)	46.47 (0.28)	-4.07 (0.0)	-18.08 (0.14)
Sakakin	-32.81 (0.09)	-20.87 (0.19)	43.93 (0.16)	-3.54 (0.0)	-13.30 (0.14)
STOCK-98615	-40.96 (0.11)	-22.09 (0.19)	40.74 (0.20)	-4.27 (0.0)	-26.57 (0.13)
Morin	-26.67 (0.09)	-24.95 (0.26)	41.20 (0.19)	-3.29 (0.0)	-13.71 (0.14)
STOCK-51172	-31.71 (0.08)	-26.09 (0.21)	40.09 (0.23)	-3.12 (0.0)	-20.84 (0.12)
STOCK-85809	-52.06 (0.17)	-32.65 (0.34)	60.57 (0.29)	-4.75 (0.0)	-28.90 (0.24)
STOCK-68642	-41.62 (0.12)	-25.69 (0.24)	44.57 (0.25)	-4.04 (0.0)	-26.79 (0.13)

Table D6: Energetic components of the binding free energy for phytochemicals bound to the JAK1. All values are given in kcal/mol.

	ΔE_{vdW}	ΔE_{elec}	ΔG_{pol}	ΔG_{np}	$-T\Delta S_{conf}$	ΔG_{bind}
Epicatechin						
Run1	-32.49 ± 0.05	-34.10 ± 0.16	40.32 ± 0.13	-3.56 ± 0	16.59 ± 0.71	-13.24 ± 0.61
Run2	-33.53 ± 0.06	-38.43 ± 0.20	43.55 ± 0.18	-3.55 ± 0	19.28 ± 0.67	-12.68 ± 0.72
Leonurine						
Run1	-38.27 ± 0.09	-26.27 ± 0.25	38.76 ± 0.14	-5.56 ± 0	20.89 ± 0.58	-10.45 ± 0.65
Run2	-45.24 ± 0.06	-37.81 ± 0.18	55.64 ± 0.09	-4.08 ± 0	19.25 ± 0.61	-12.24 ± 0.64
STOCK1N-68642						
Run1	-38.44 ± 0.09	-30.45 ± 0.16	37.84 ± 0.15	-3.85 ± 0	20.48 ± 0.71	-14.42 ± 0.75
Run2	-39.64 ± 0.07	-34.06 ± 0.19	44.29 ± 0.21	-3.93 ± 0	20.05 ± 0.47	-13.29 ± 0.55
STOCK1N-82656						
Run1	-41.02 ± 0.09	-22.72 ± 0.25	39.78 ± 0.20	-5.27 ± 0	18.16 ± 0.83	-11.07 ± 0.89
Run2	-32.59 ± 0.06	-28.96 ± 0.19	41.56 ± 0.16	-4.43 ± 0	16.15 ± 0.53	-8.26 ± 0.59
STOCK1N-85809						
Run1	-50.11 ± 0.06	-35.61 ± 0.20	58.73 ± 0.19	-4.69 ± 0	20.11 ± 0.68	-11.57 ± 0.73
Run2	-45.42 ± 0.10	-30.35 ± 0.21	48.45 ± 0.20	-4.53 ± 0	20.07 ± 0.73	-11.78 ± 0.79
STOCK1N-98615						
Run1	-39.20 ± 0.09	-22.14 ± 0.17	39.88 ± 0.14	-4.11 ± 0	16.07 ± 0.59	-9.50 ± 0.63
Run2	-42.33 ± 0.07	-26.03 ± 0.23	46.22 ± 0.16	-4.46 ± 0	15.02 ± 0.43	-11.58 ± 0.52

Table D7: Hydrogen bonds formed between JAK1, and top six lead molecules and the corresponding average distance and percentage determined using the simulation trajectories of run1 and run2.

Binding couples		**Molecular Dynamics**			
Acceptor	**Donor**	**Distance (Å)**	**Occupancy (%)**		
Epicatechin			**Run1**	**Run2**	**Avg.**
Glu^{957}@O	Lig@O2	2.73	94.35	85.82	90.08
Leu^{881}@O	Lig@O4	2.77	33.81	19.69	26.75
Asn^{1008}@OD1	Lig@O5	2.74	25.23	12.08	18.65
Lig@O1	Leu^{959}@N	2.93	14.52	15.76	15.14
Leonurine					
Leu^{959}@O	Lig@O2	2.76	24.42	43.58	34.00
Glu^{957}@O	Lig@O2	2.77	37.75	11.14	24.45
STOCK1N-68642					
Glu^{957}@O	Lig@N1	2.84	87.97	84.73	86.35
Lig@N6	Leu^{959}@N	2.92	17.11	34.31	25.71
Lig@O1	Ser^{963}@OG	2.80	7.11	23.31	15.21
Arg^{1007}@O	Lig@N8	2.86	11.09	10.22	10.65
STOCK1N-82656					
Lig@O4	Leu^{959}@N	2.90	35.36	31.22	33.29
Asp^{1021}@N	Lig@O1	2.87	37.48	11.33	24.40
Lig@O5	Ser^{963}@N	2.87	19.17	17.31	18.24
Pro^{960}@O	Lig@O6	2.76	15.06	13.58	14.32
STOCK1N-85809					
Leu^{881}@O	Lig@O4	2.76	35.12	25.21	30.16
Lig@O3	Ser^{963}@OG	2.70	15.31	22.32	18.81
Lig@O1	Leu^{959}@N	2.91	11.20	20.22	15.71
Asp^{1021}@OD1	Lig@N3	2.83	15.89	11.03	13.46
STOCK1N-98615					
Lig@O2	Leu^{959}@N	2.84	75.11	65.69	70.40
Lig@O3	Ser^{963}@N	2.86	45.98	29.59	37.78
Leu^{959}@O	Lig@O4	2.83	20.05	16.78	18.41

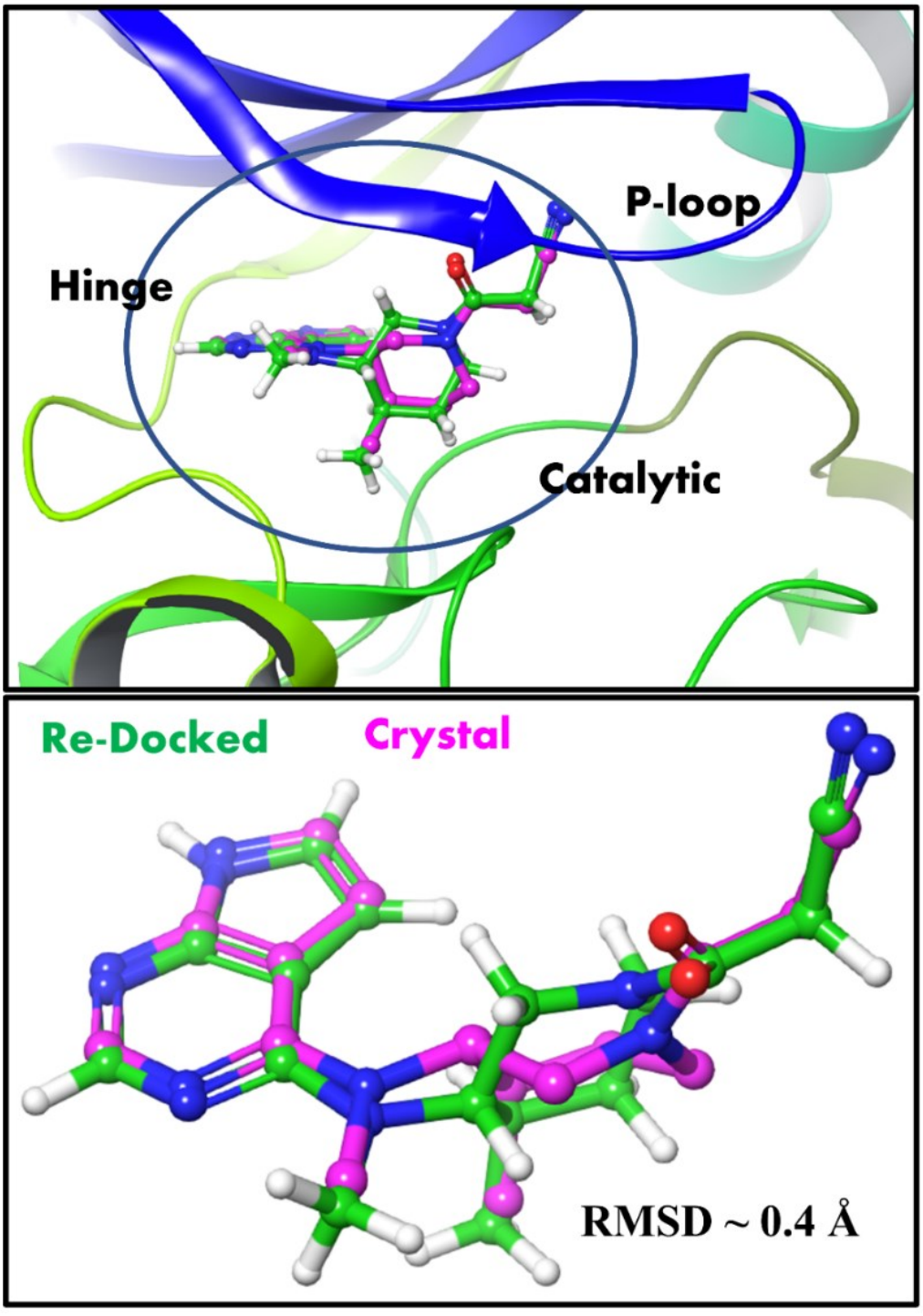


Figure D1: Superimposing the redocked ligand with co-crystallized ligand MI1. The co-crystallized ligand is drawn in magenta. The ligands as docked with Glide is in green.

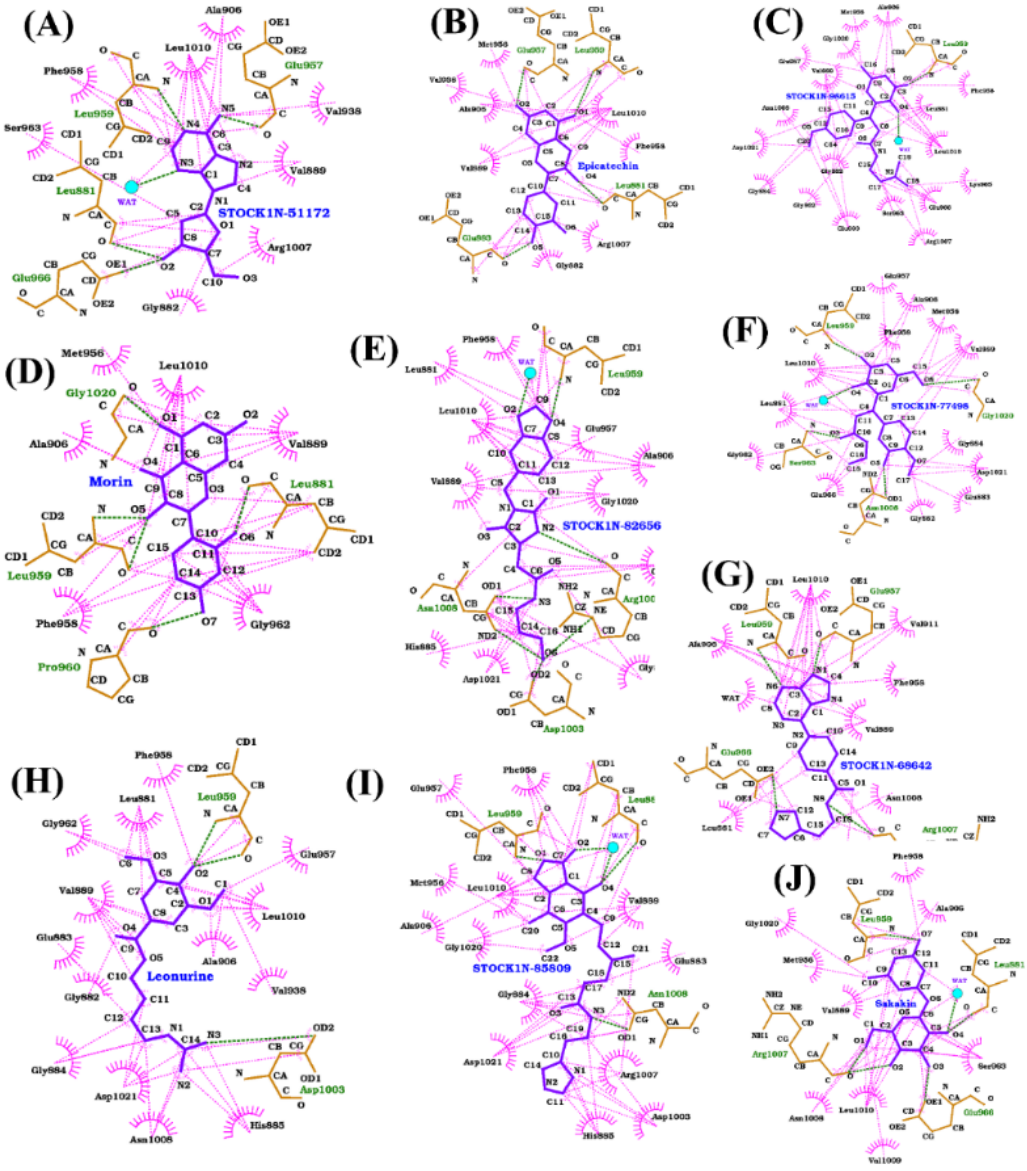


Figure D2: The interaction pattern of the top 10 hit molecules with the ATP-binding pocket of the JAK1 receptor. All plots were generated by a widely used LigPlot++ software.

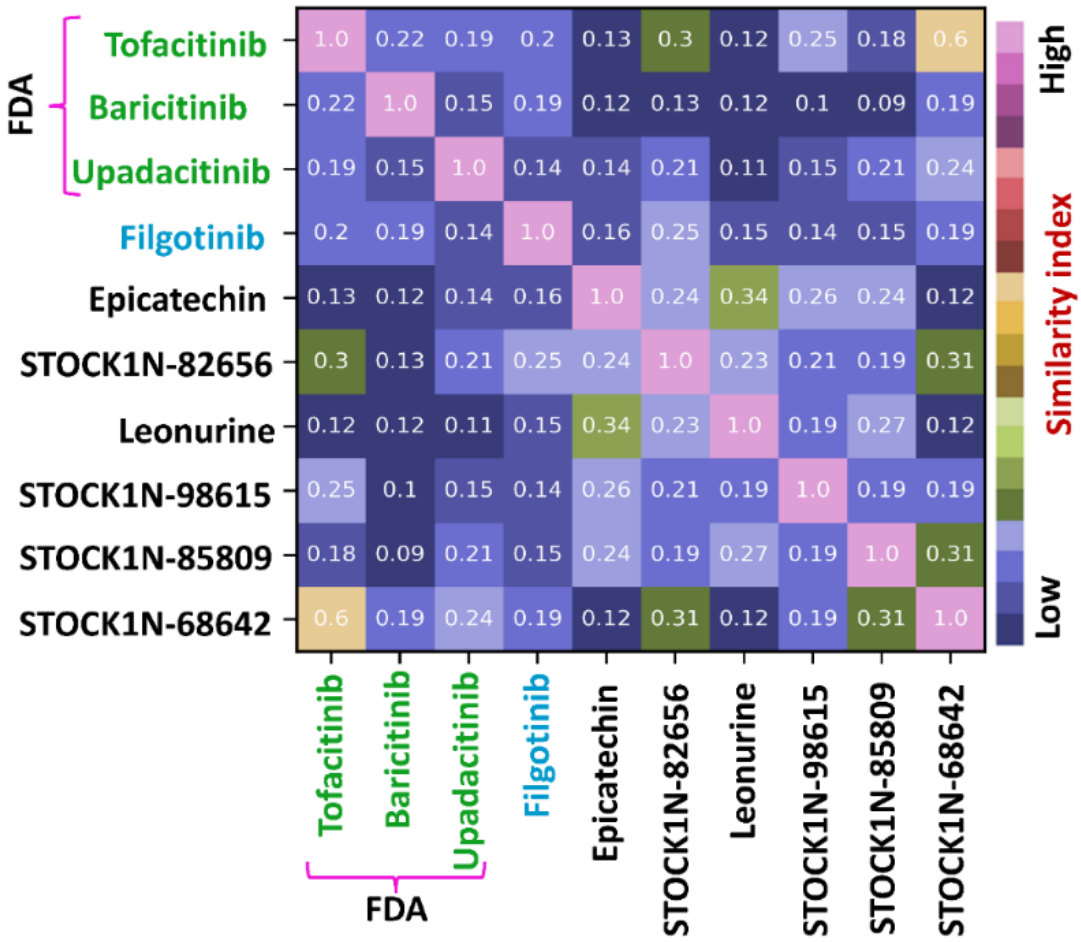


Figure D3: The similarity scores between compounds calculated in the Similarity Workbench interface of ChemMine Tools, calculating the atom pair and maximum common substructure (MCS) similarities with the Tanimoto coefficient.

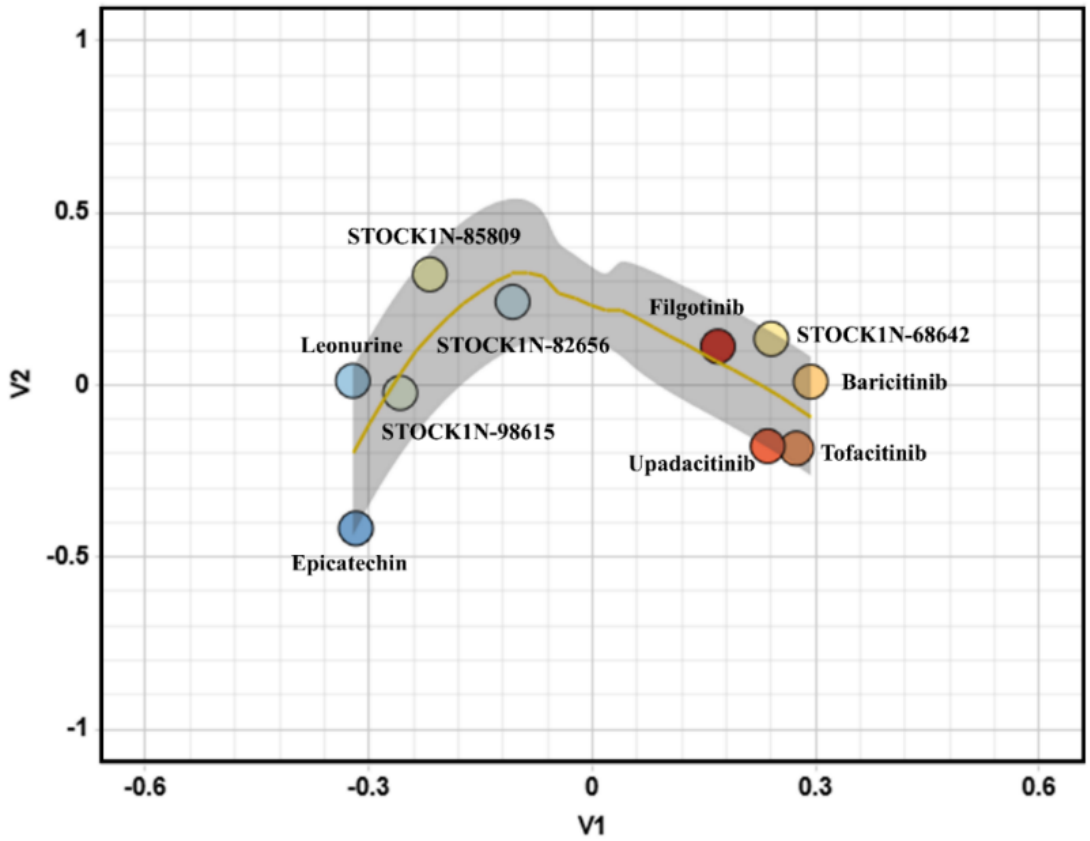


Figure D4: Multidimensional (2D) scaling of the chemical structures of the dataset reveals high structural diversity. The axes (V1 and V2) represent chemical space in two directions.

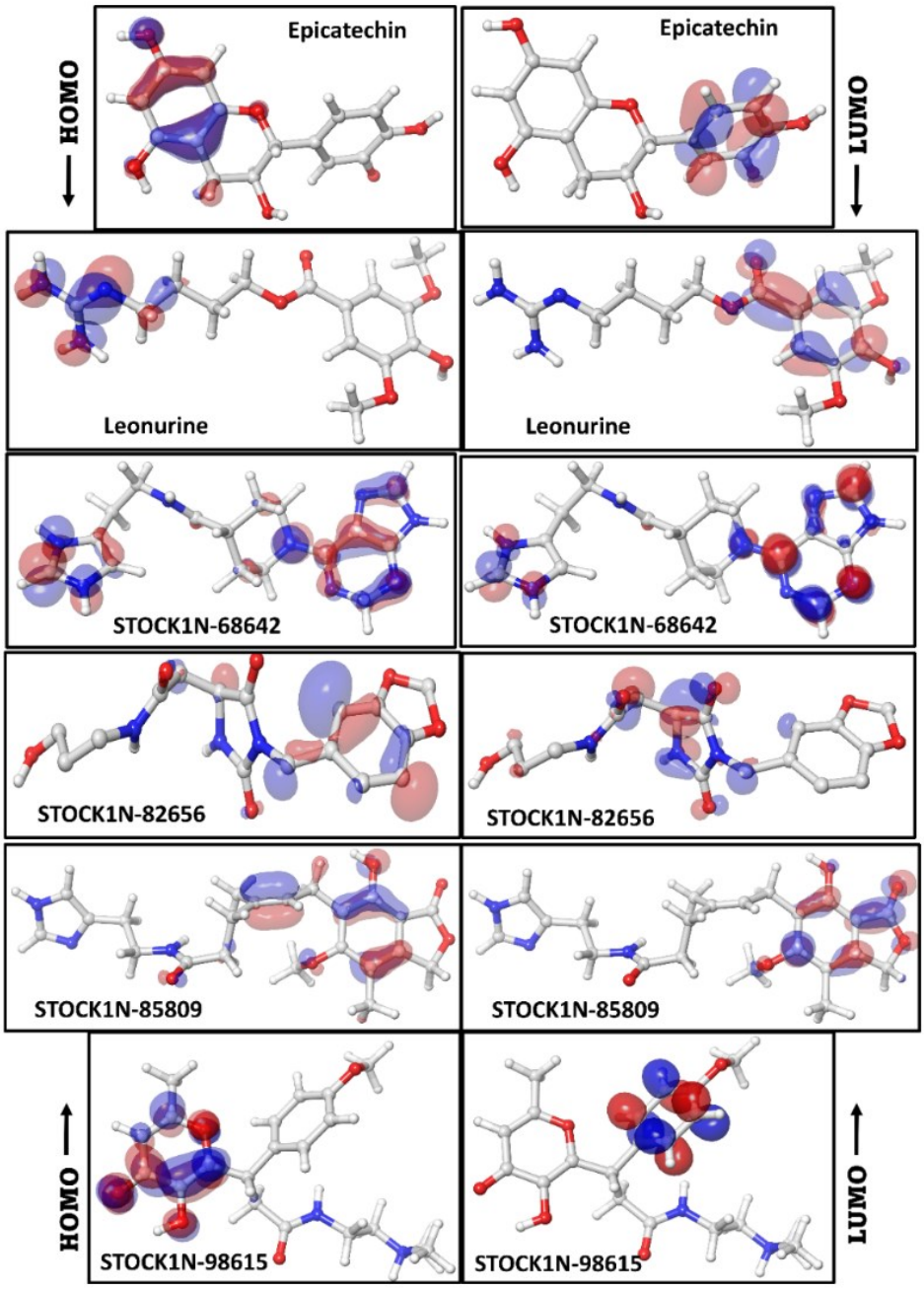


Figure D5: Frontier molecular orbitals (highest occupied molecular orbitals, HOMO and lowest unoccupied molecular orbitals, LUMO) surface of six lead phytochemical molecules. The conformers were analyzed via the Lee-Yang-Parr correlation function (B3LYP) and Beche's three-parameter exchange potential using 6-31G** basic set level. The color-coding region was red, indicating the most negative potential region, and blue indicating the most positive region of the molecules.

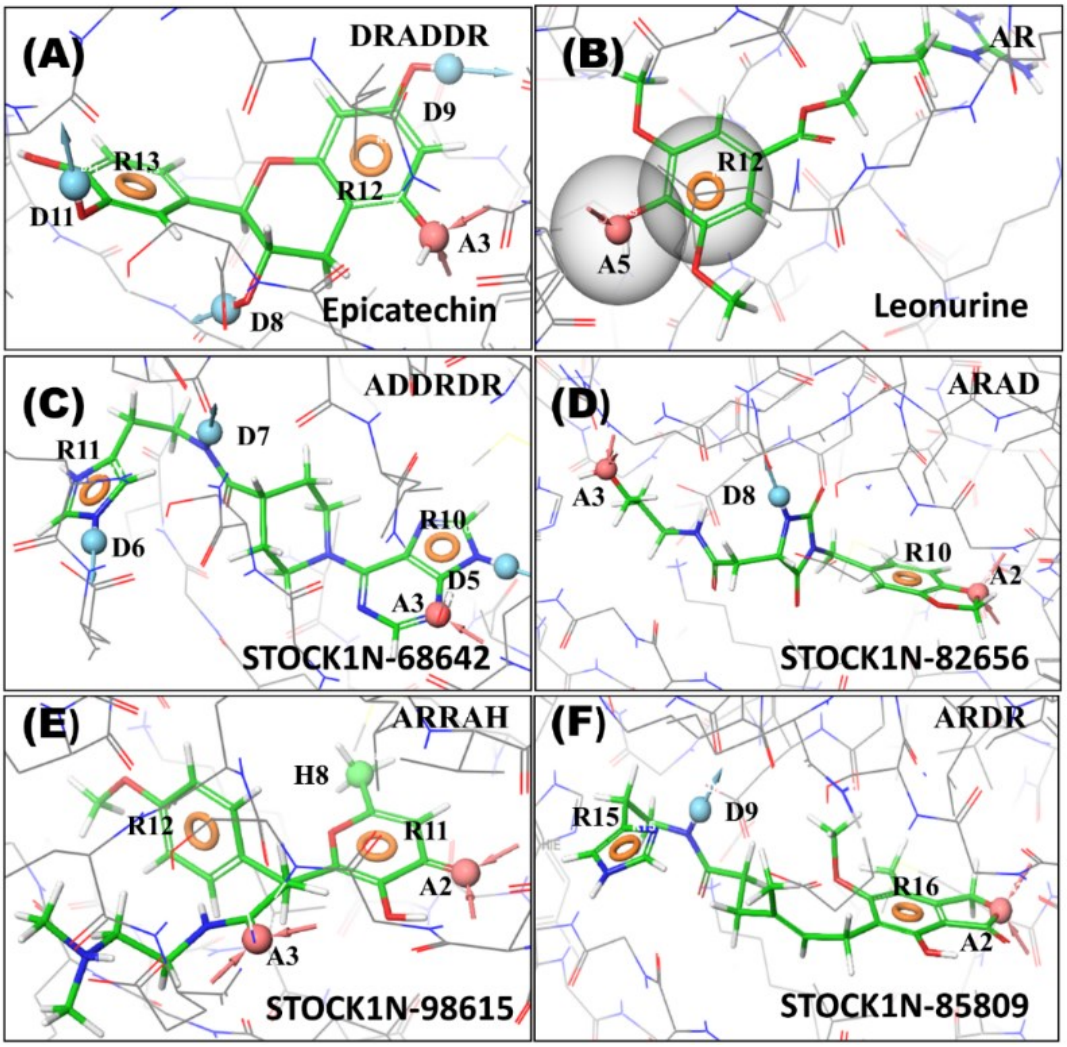


Figure D6: Structure-based e-pharmacophores generated using JAK1 complex with phytochemicals. The phytochemicals are shown in green sticks (orange ring: aromatic ring, blue ball: hydrogen bond donor, pink ball: hydrogen bond acceptor).

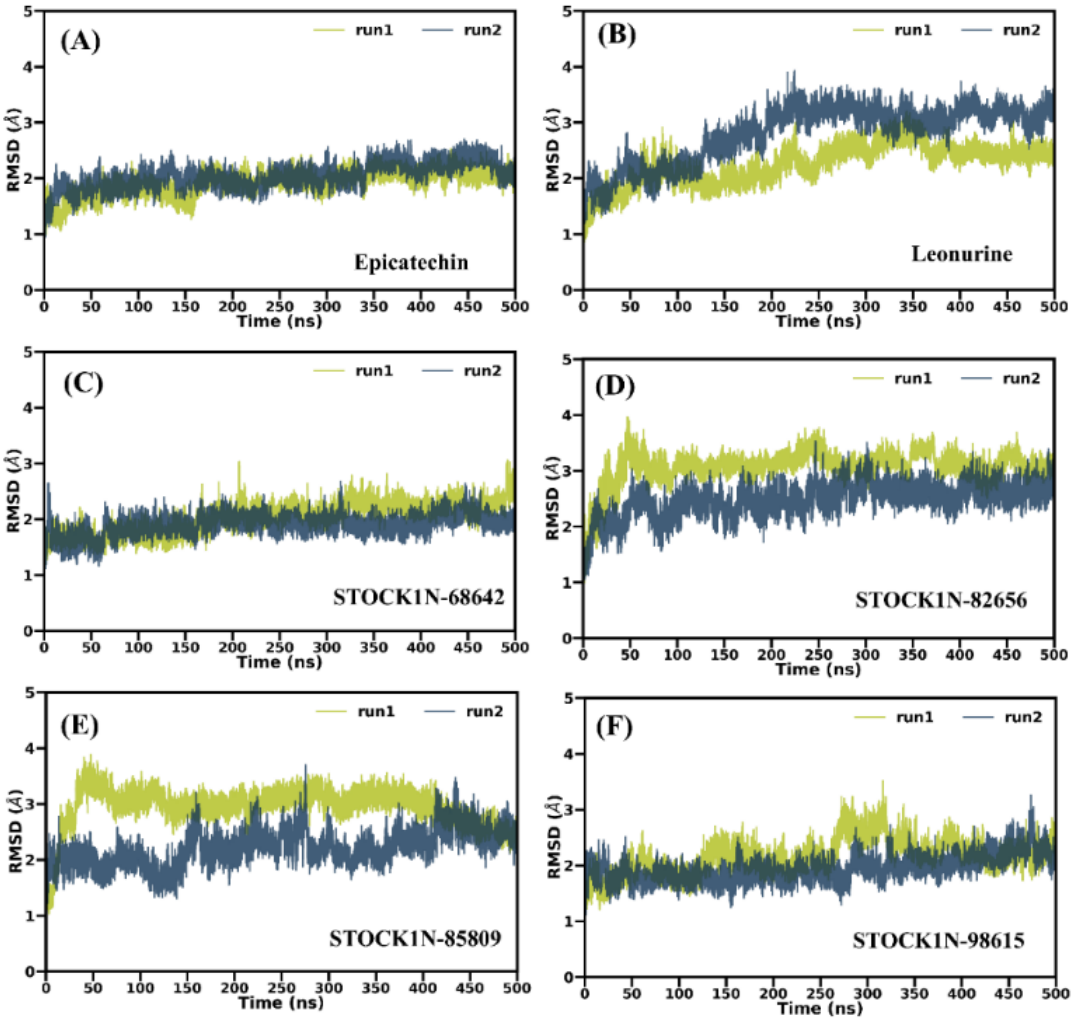


Figure D7: The time evolution of the root mean-square deviations (RMSD) of backbone atoms of each JAK1 complexed with the top six lead molecules.

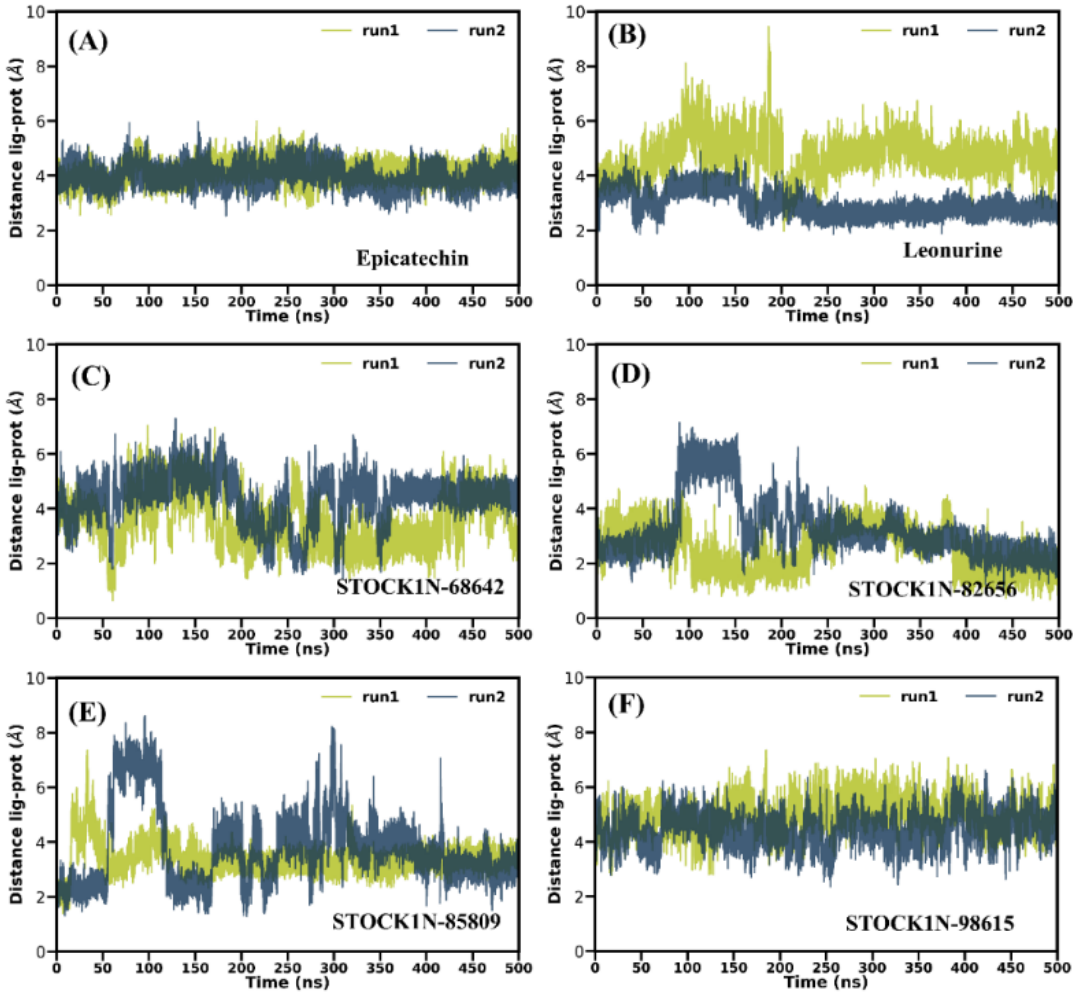


Figure D8: The ligand and protein (binding pocket residues only) center of mass distance with respect to simulation time.

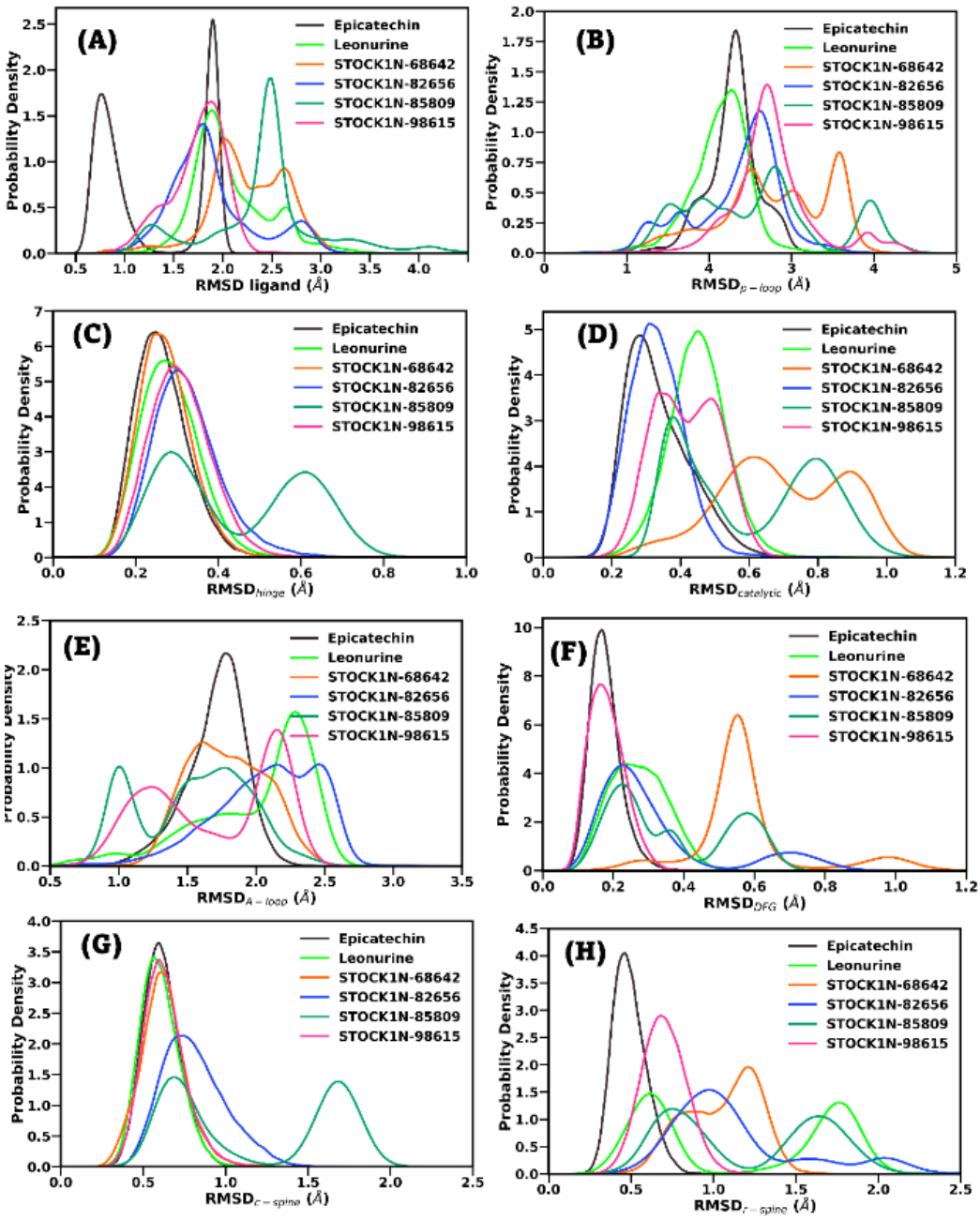


Figure D9: The RMSD distribution of (A) ligand heavy atoms, backbone atoms of (B) P-loop, (C) Hinge region, (D) Catalytic loop, (E) Activation loop, (F) DFG motif, (G) catalytic spines and (H) regulatory spines.

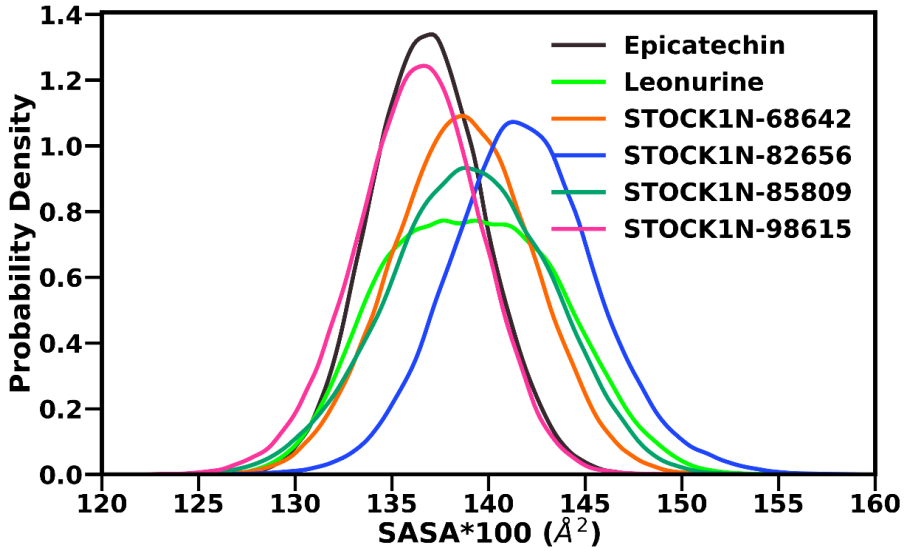


Figure D10: The solvent accessible surface area probability density distribution of six complexes during the combined two replicas.

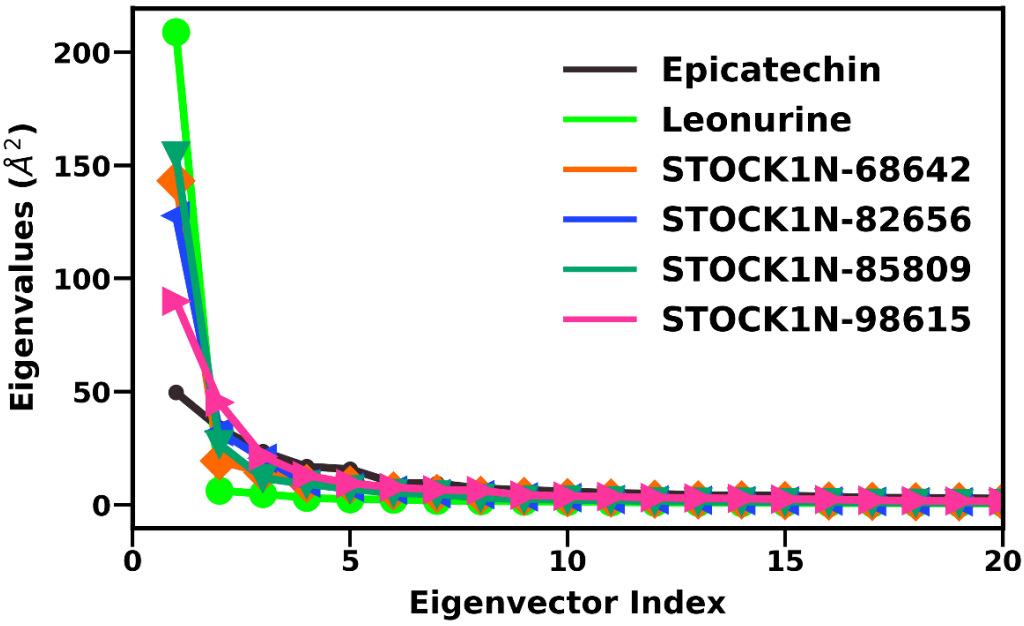


Figure D11: The eigenvalues plotted against the corresponding eigenvector indices from the Cα atoms covariance matrix constructed form the last 500 (2x250) ns trajectory of all six systems.

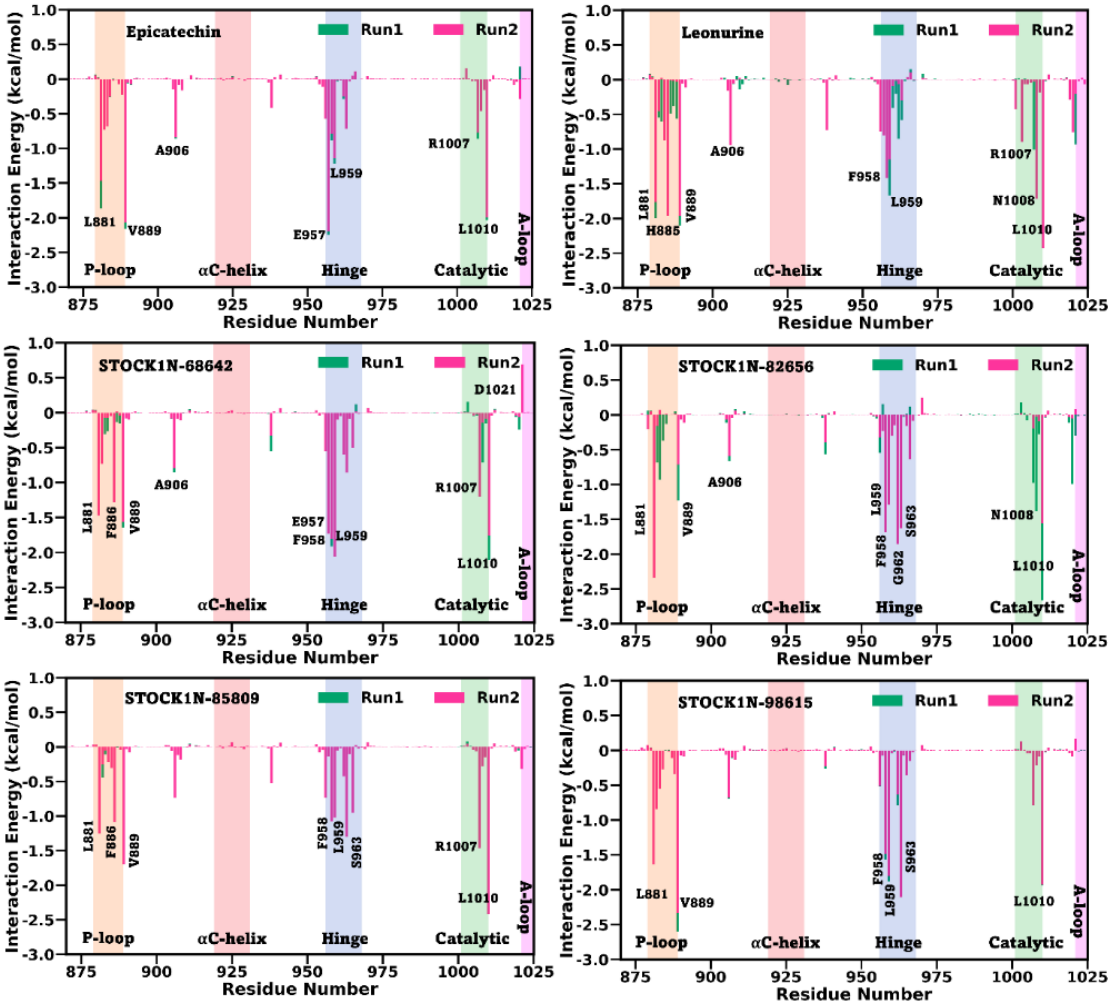


Figure D12: The per-residue energy decomposition plot of phytochemicals and JAK1 complexes. Here, we only considered the ATP-binding area of the JAK kinase and colour shaded areas represented the different important binding area.

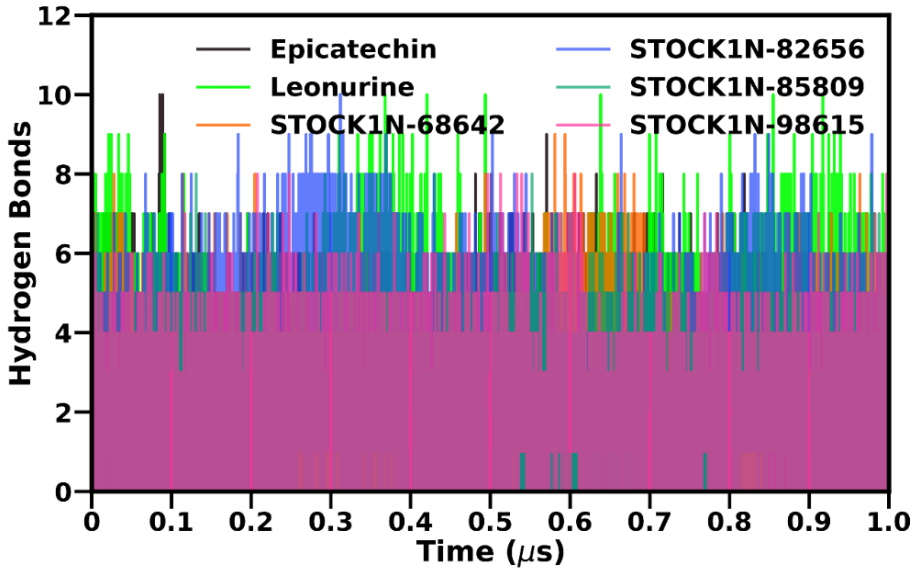


Figure D13: Number of hydrogen bonds formed between phytochemicals and JAK1 along 1.0 µs (2X500 ns) simulation.

Bibliography

Bibliography

(1) Cori, G. T.; Green, A. A. CRYSTALLINE MUSCLE PHOSPHORYLASE: II. PROSTHETIC GROUP. *Journal of Biological Chemistry* **1943**, *151* (1), 31–38. https://doi.org/10.1016/S0021-9258(18)72111-X.

(2) Krebs, E. G.; Fischer, E. H. The Phosphorylase b to a Converting Enzyme of Rabbit Skeletal Muscle. *Biochim Biophys Acta* **1956**, *20* (1), 150–157. https://doi.org/10.1016/0006-3002(56)90273-6.

(3) Wosilait, W. D.; Sutherland, E. W. The Relationship of Epinephrine and Glucagon to Liver Phosphorylase. II. Enzymatic Inactivation of Liver Phosphorylase. *J Biol Chem* **1956**, *218* (1), 469–481.

(4) Brognard, J.; Hunter, T. Protein Kinase Signaling Networks in Cancer. *Curr Opin Genet Dev* **2011**, *21* (1), 4–11. https://doi.org/10.1016/j.gde.2010.10.012.

(5) Lu, Z.; Hunter, T. Degradation of Activated Protein Kinases by Ubiquitination. *Annu Rev Biochem* **2009**, *78*, 435–475. https://doi.org/10.1146/annurev.biochem.013008.092711.

(6) Henneke, G.; Koundrioukoff, S.; Hübscher, U. Multiple Roles for Kinases in DNA Replication. *EMBO Rep* **2003**, *4* (3), 252–256. https://doi.org/10.1038/sj.embor.embor774.

(7) Depamphilis, M. L.; de Renty, C. M.; Ullah, Z.; Lee, C. Y. "The Octet": Eight Protein Kinases That Control Mammalian DNA Replication. *Front Physiol* **2012**, *3*, 368. https://doi.org/10.3389/fphys.2012.00368.

(8) Cao, M.; Meng, D.; Wang, L.; Bei, S.; Snell, W. J.; Pan, J. Activation Loop Phosphorylation of a Protein Kinase Is a Molecular Marker of Organelle Size That Dynamically Reports Flagellar Length. *Proc Natl Acad Sci U S A* **2013**, *110* (30), 12337–12342. https://doi.org/10.1073/pnas.1302364110.

(9) Mashud, R.; Nomachi, A.; Hayakawa, A.; Kubouchi, K.; Danno, S.; Hirata, T.; Matsuo, K.; Nakayama, T.; Satoh, R.; Sugiura, R.; Abe, M.; Sakimura, K.; Wakana, S.; Ohsaki, H.; Kamoshida, S.; Mukai, H. Impaired Lymphocyte Trafficking in Mice Deficient in the Kinase Activity of PKN1. *Sci Rep* **2017**, *7* (1), 7663. https://doi.org/10.1038/s41598-017-07936-9.

(10) Adams, J. A. Kinetic and Catalytic Mechanisms of Protein Kinases. *Chem Rev* **2001**, *101* (8), 2271–2290. https://doi.org/10.1021/cr000230w.

(11) Shoni, M.; Lui, K. O.; Vavvas, D. G.; Muto, M. G.; Berkowitz, R. S.; Vlahos, N.; Ng, S.-W. Protein Kinases and Associated Pathways in Pluripotent State and Lineage Differentiation. *Curr Stem Cell Res Ther* **2014**, *9* (5), 366–387. https://doi.org/10.2174/1574888x09666140616130217.

(12) He, W.-Q.; Peng, Y.-J.; Zhang, W.-C.; Lv, N.; Tang, J.; Chen, C.; Zhang, C.-H.; Gao, S.; Chen, H.-Q.; Zhi, G.; Feil, R.; Kamm, K. E.; Stull, J. T.; Gao, X.; Zhu, M.-S. Myosin Light Chain Kinase Is Central to Smooth Muscle Contraction and Required for Gastrointestinal Motility in Mice. *Gastroenterology* **2008**, *135* (2), 610–620. https://doi.org/10.1053/j.gastro.2008.05.032.

(13) Sathishkumar, K.; Yallampalli, U.; Elkins, R.; Yallampalli, C. Raf-1 Kinase Regulates Smooth Muscle Contraction in the Rat Mesenteric Arteries. *J Vasc Res* **2010**, *47* (5), 384–398. https://doi.org/10.1159/000277726.

(14) Nicolas, C. S.; Amici, M.; Bortolotto, Z. A.; Doherty, A.; Csaba, Z.; Fafouri, A.; Dournaud, P.; Gressens, P.; Collingridge, G. L.; Peineau, S. The Role of JAK-STAT Signaling within the CNS. *JAK-STAT* **2013**, *2* (1), e22925. https://doi.org/10.4161/jkst.22925.

(15) Yan, Z.; Gibson, S. A.; Buckley, J. A.; Qin, H.; Benveniste, E. N. Role of the JAK/STAT Signaling Pathway in Regulation of Innate Immunity in Neuroinflammatory Diseases. *Clinical Immunology* **2018**, *189*, 4–13. https://doi.org/10.1016/j.clim.2016.09.014.

(16) Jain, M.; Singh, M. K.; Shyam, H.; Mishra, A.; Kumar, S.; Kumar, A.; Kushwaha, J. Role of JAK/STAT in the Neuroinflammation and Its Association with Neurological Disorders. *Annals of Neurosciences* **2022**, 09727531211070532. https://doi.org/10.1177/09727531211070532.

(17) Bermingham, D. P.; Blakely, R. D. Kinase-Dependent Regulation of Monoamine Neurotransmitter Transporters. *Pharmacol Rev* **2016**, *68* (4), 888–953. https://doi.org/10.1124/pr.115.012260.

(18) Endicott, J. A.; Noble, M. E. M.; Johnson, L. N. The Structural Basis for Control of Eukaryotic Protein Kinases. *Annu Rev Biochem* **2012**, *81*, 587–613. https://doi.org/10.1146/annurev-biochem-052410-090317.

(19) Manning, G.; Whyte, D. B.; Martinez, R.; Hunter, T.; Sudarsanam, S. The Protein Kinase Complement of the Human Genome. *Science* **2002**, *298* (5600), 1912–1934. https://doi.org/10.1126/science.1075762.

(20) Lahiry, P.; Torkamani, A.; Schork, N. J.; Hegele, R. A. Kinase Mutations in Human Disease: Interpreting Genotype-Phenotype Relationships. *Nat Rev Genet* **2010**, *11* (1), 60–74. https://doi.org/10.1038/nrg2707.

(21) Blume-Jensen, P.; Hunter, T. Oncogenic Kinase Signalling. *Nature* **2001**, *411* (6835), 355–365. https://doi.org/10.1038/35077225.

(22) Roskoski, R. A Historical Overview of Protein Kinases and Their Targeted Small Molecule Inhibitors. *Pharmacological Research* **2015**, *100*, 1–23. https://doi.org/10.1016/j.phrs.2015.07.010.

(23) Cohen, P. Protein Kinases--the Major Drug Targets of the Twenty-First Century? *Nat Rev Drug Discov* **2002**, *1* (4), 309–315. https://doi.org/10.1038/nrd773.

(24) Hollingsworth, S. A.; Dror, R. O. Molecular Dynamics Simulation for All. *Neuron* **2018**, *99* (6), 1129–1143. https://doi.org/10.1016/j.neuron.2018.08.011.

(25) Karplus, M.; McCammon, J. A. Molecular Dynamics Simulations of Biomolecules. *Nat. Struct Biol.* **2002**, *9* (9), 646–652. https://doi.org/10.1038/nsb0902-646.

(26) Walsh, C. T.; Garneau-Tsodikova, S.; Gatto, G. J. Protein Posttranslational Modifications: The Chemistry of Proteome Diversifications. *Angew Chem Int Ed Engl* **2005**, *44* (45), 7342–7372. https://doi.org/10.1002/anie.200501023.

(27) Cohen, P. The Role of Protein Phosphorylation in Human Health and Disease. The Sir Hans Krebs Medal Lecture. *Eur J Biochem* **2001**, *268* (19), 5001–5010. https://doi.org/10.1046/j.0014-2956.2001.02473.x.

(28) Liu, Q.; Sabnis, Y.; Zhao, Z.; Zhang, T.; Buhrlage, S. J.; Jones, L. H.; Gray, N. S. Developing Irreversible Inhibitors of the Protein Kinase Cysteinome. *Chem Biol* **2013**, *20* (2), 146–159. https://doi.org/10.1016/j.chembiol.2012.12.006.

(29) Kannan, N.; Taylor, S. S.; Zhai, Y.; Venter, J. C.; Manning, G. Structural and Functional Diversity of the Microbial Kinome. *PLoS Biol* **2007**, *5* (3), e17. https://doi.org/10.1371/journal.pbio.0050017.

(30) Ubersax, J. A.; Ferrell, J. E. Mechanisms of Specificity in Protein Phosphorylation. *Nat Rev Mol Cell Biol* **2007**, *8* (7), 530–541. https://doi.org/10.1038/nrm2203.

(31) Roskoski, R. Janus Kinase (JAK) Inhibitors in the Treatment of Inflammatory and Neoplastic Diseases. *Pharmacol Res* **2016**, *111*, 784–803. https://doi.org/10.1016/j.phrs.2016.07.038.

(32) Kennelly, P. J. Protein Kinases and Protein Phosphatases in Prokaryotes: A Genomic Perspective. *FEMS Microbiol Lett* **2002**, *206* (1), 1–8. https://doi.org/10.1111/j.1574-6968.2002.tb10978.x.

(33) Williams, N. K.; Bamert, R. S.; Patel, O.; Wang, C.; Walden, P. M.; Wilks, A. F.; Fantino, E.; Rossjohn, J.; Lucet, I. S. Dissecting Specificity in the Janus Kinases: The Structures of JAK-Specific Inhibitors Complexed to the JAK1 and JAK2 Protein Tyrosine Kinase Domains. *Journal of Molecular Biology* **2009**, *387* (1), 219–232. https://doi.org/10.1016/j.jmb.2009.01.041.

(34) Ghoreschi, K.; Laurence, A.; O'Shea, J. J. Janus Kinases in Immune Cell Signaling. *Immunological Reviews* **2009**, *228* (1), 273–287. https://doi.org/10.1111/j.1600-065X.2008.00754.x.

(35) Roskoski, R. Classification of Small Molecule Protein Kinase Inhibitors Based upon the Structures of Their Drug-Enzyme Complexes. *Pharmacol Res* **2016**, *103*, 26–48. https://doi.org/10.1016/j.phrs.2015.10.021.

(36) O'Shea, J. J.; Schwartz, D. M.; Villarino, A. V.; Gadina, M.; McInnes, I. B.; Laurence, A. The JAK-STAT Pathway: Impact on Human Disease and Therapeutic Intervention. *Annu Rev Med* **2015**, *66*, 311–328. https://doi.org/10.1146/annurev-med-051113-024537.

(37) Stark, G. R.; Darnell, J. E. The JAK-STAT Pathway at Twenty. *Immunity* **2012**, *36* (4), 503–514. https://doi.org/10.1016/j.immuni.2012.03.013.

(38) Morris, R.; Kershaw, N. J.; Babon, J. J. The Molecular Details of Cytokine Signaling via the JAK/STAT Pathway. *Protein Sci* **2018**, *27* (12), 1984–2009. https://doi.org/10.1002/pro.3519.

(39) Pérez-Jeldres, T.; Tyler, C. J.; Boyer, J. D.; Karuppuchamy, T.; Yarur, A.; Giles, D. A.; Yeasmin, S.; Lundborg, L.; Sandborn, W. J.; Patel, D. R.; Rivera-Nieves, J. Targeting Cytokine Signaling and Lymphocyte Traffic via Small Molecules in Inflammatory Bowel Disease: JAK Inhibitors and S1PR Agonists. *Frontiers in Pharmacology* **2019**, *10*.

(40) 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. https://doi.org/10.1016/j.intimp.2020.106210.

(41) Yoshimura, A.; Yasukawa, H. JAK's SOCS: A Mechanism of Inhibition. *Immunity* **2012**, *36* (2), 157–159. https://doi.org/10.1016/j.immuni.2012.01.010.

(42) Yoshimura, A.; Ohkubo, T.; Kiguchi, T.; Jenkins, N. A.; Gilbert, D. J.; Copeland, N. G.; Hara, T.; Miyajima, A. A Novel Cytokine-Inducible Gene CIS Encodes an SH2-Containing Protein That Binds to Tyrosine-Phosphorylated Interleukin 3 and

Erythropoietin Receptors. *EMBO J* **1995**, *14* (12), 2816–2826. https://doi.org/10.1002/j.1460-2075.1995.tb07281.x.

(43) Endo, T. A.; Masuhara, M.; Yokouchi, M.; Suzuki, R.; Sakamoto, H.; Mitsui, K.; Matsumoto, A.; Tanimura, S.; Ohtsubo, M.; Misawa, H.; Miyazaki, T.; Leonor, N.; Taniguchi, T.; Fujita, T.; Kanakura, Y.; Komiya, S.; Yoshimura, A. A New Protein Containing an SH2 Domain That Inhibits JAK Kinases. *Nature* **1997**, *387* (6636), 921–924. https://doi.org/10.1038/43213.

(44) Wilks, A. F. Two Putative Protein-Tyrosine Kinases Identified by Application of the Polymerase Chain Reaction. *Proc Natl Acad Sci U S A* **1989**, *86* (5), 1603–1607. https://doi.org/10.1073/pnas.86.5.1603.

(45) Wilks, A. F.; Harpur, A. G.; Kurban, R. R.; Ralph, S. J.; Zürcher, G.; Ziemiecki, A. Two Novel Protein-Tyrosine Kinases, Each with a Second Phosphotransferase-Related Catalytic Domain, Define a New Class of Protein Kinase. *Mol Cell Biol* **1991**, *11* (4), 2057–2065. https://doi.org/10.1128/mcb.11.4.2057-2065.1991.

(46) Krolewski, J. J.; Lee, R.; Eddy, R.; Shows, T. B.; Dalla-Favera, R. Identification and Chromosomal Mapping of New Human Tyrosine Kinase Genes. *Oncogene* **1990**, *5* (3), 277–282.

(47) Glassman, C. R.; Tsutsumi, N.; Saxton, R. A.; Lupardus, P. J.; Jude, K. M.; Garcia, K. C. Structure of a Janus Kinase Cytokine Receptor Complex Reveals the Basis for Dimeric Activation. *Science* **2022**, *376* (6589), 163–169. https://doi.org/10.1126/science.abn8933.

(48) Thompson, J. D.; Higgins, D. G.; Gibson, T. J. CLUSTAL W: Improving the Sensitivity of Progressive Multiple Sequence Alignment through Sequence Weighting, Position-Specific Gap Penalties and Weight Matrix Choice. *Nucleic Acids Res* **1994**, *22* (22), 4673–4680. https://doi.org/10.1093/nar/22.22.4673.

(49) Meharena, H. S.; Chang, P.; Keshwani, M. M.; Oruganty, K.; Nene, A. K.; Kannan, N.; Taylor, S. S.; Kornev, A. P. Deciphering the Structural Basis of Eukaryotic Protein Kinase Regulation. *PLoS Biol* **2013**, *11* (10), e1001680. https://doi.org/10.1371/journal.pbio.1001680.

(50) Johnson, D. A.; Akamine, P.; Radzio-Andzelm, E.; Madhusudan, M.; Taylor, S. S. Dynamics of CAMP-Dependent Protein Kinase. *Chem Rev* **2001**, *101* (8), 2243–2270. https://doi.org/10.1021/cr000226k.

(51) 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.

(52) Taylor, S. S.; Keshwani, M. M.; Steichen, J. M.; Kornev, A. P. Evolution of the Eukaryotic Protein Kinases as Dynamic Molecular Switches. *Philos Trans R Soc Lond B Biol Sci* **2012**, *367* (1602), 2517–2528. https://doi.org/10.1098/rstb.2012.0054.

(53) Shaw, A. S.; Kornev, A. P.; Hu, J.; Ahuja, L. G.; Taylor, S. S. Kinases and Pseudokinases: Lessons from RAF. *Mol Cell Biol* **2014**, *34* (9), 1538–1546. https://doi.org/10.1128/MCB.00057-14.

(54) Yamaguchi, H.; Hendrickson, W. A. Structural Basis for Activation of Human Lymphocyte Kinase Lck upon Tyrosine Phosphorylation. *Nature* **1996**, *384* (6608), 484–489. https://doi.org/10.1038/384484a0.

(55) Jeffrey, P. D.; Russo, A. A.; Polyak, K.; Gibbs, E.; Hurwitz, J.; Massagué, J.; Pavletich, N. P. Mechanism of CDK Activation Revealed by the Structure of a

CyclinA-CDK2 Complex. *Nature* **1995**, *376* (6538), 313–320. https://doi.org/10.1038/376313a0.

(56) Canagarajah, B. J.; Khokhlatchev, A.; Cobb, M. H.; Goldsmith, E. J. Activation Mechanism of the MAP Kinase ERK2 by Dual Phosphorylation. *Cell* **1997**, *90* (5), 859–869. https://doi.org/10.1016/s0092-8674(00)80351-7.

(57) Nolen, B.; Taylor, S.; Ghosh, G. Regulation of Protein Kinases; Controlling Activity through Activation Segment Conformation. *Mol Cell* **2004**, *15* (5), 661–675. https://doi.org/10.1016/j.molcel.2004.08.024.

(58) Gotoh, N.; Tojo, A.; Hino, M.; Yazaki, Y.; Shibuya, M. A Highly Conserved Tyrosine Residue at Codon 845 within the Kinase Domain Is Not Required for the Transforming Activity of Human Epidermal Growth Factor Receptor. *Biochem Biophys Res Commun* **1992**, *186* (2), 768–774. https://doi.org/10.1016/0006-291x(92)90812-y.

(59) Tice, D. A.; Biscardi, J. S.; Nickles, A. L.; Parsons, S. J. Mechanism of Biological Synergy between Cellular Src and Epidermal Growth Factor Receptor. *Proc Natl Acad Sci U S A* **1999**, *96* (4), 1415–1420. https://doi.org/10.1073/pnas.96.4.1415.

(60) Tournier, C.; Dong, C.; Turner, T. K.; Jones, S. N.; Flavell, R. A.; Davis, R. J. MKK7 Is an Essential Component of the JNK Signal Transduction Pathway Activated by Proinflammatory Cytokines. *Genes Dev* **2001**, *15* (11), 1419–1426. https://doi.org/10.1101/gad.888501.

(61) Xie, X.; Gu, Y.; Fox, T.; Coll, J. T.; Fleming, M. A.; Markland, W.; Caron, P. R.; Wilson, K. P.; Su, M. S. Crystal Structure of JNK3: A Kinase Implicated in Neuronal Apoptosis. *Structure* **1998**, *6* (8), 983–991. https://doi.org/10.1016/s0969-2126(98)00100-2.

(62) Mishra, P.; Günther, S. New Insights into the Structural Dynamics of the Kinase JNK3. *Sci Rep* **2018**, *8* (1), 9435. https://doi.org/10.1038/s41598-018-27867-3.

(63) Berteotti, A.; Cavalli, A.; Branduardi, D.; Gervasio, F. L.; Recanatini, M.; Parrinello, M. Protein Conformational Transitions: The Closure Mechanism of a Kinase Explored by Atomistic Simulations. *J Am Chem Soc* **2009**, *131* (1), 244–250. https://doi.org/10.1021/ja806846q.

(64) Gan, W.; Yang, S.; Roux, B. Atomistic View of the Conformational Activation of Src Kinase Using the String Method with Swarms-of-Trajectories. *Biophys J* **2009**, *97* (4), L8–L10. https://doi.org/10.1016/j.bpj.2009.06.016.

(65) Paul, F.; Meng, Y.; Roux, B. Identification of Druggable Kinase Target Conformations Using Markov Model Metastable States Analysis of Apo-Abl. *J. Chem. Theory Comput.* **2020**, *16* (3), 1896–1912. https://doi.org/10.1021/acs.jctc.9b01158.

(66) Shan, Y.; Seeliger, M. A.; Eastwood, M. P.; Frank, F.; Xu, H.; Jensen, M. Ø.; Dror, R. O.; Kuriyan, J.; Shaw, D. E. A Conserved Protonation-Dependent Switch Controls Drug Binding in the Abl Kinase. *Proc Natl Acad Sci U S A* **2009**, *106* (1), 139–144. https://doi.org/10.1073/pnas.0811223106.

(67) Meng, Yilin; Roux, Benoit. Locking the Active Conformation of C-Src Kinase through the Phosphorylation of the Activation Loop. *Journal of Molecular Biology* **2014**, *426* (2), 423–435. https://doi.org/10.1016/j.jmb.2013.10.001.

(68) Musavizadeh, Z.; Grottesi, A.; Guarguaglini, G.; Paiardini, A. Phosphorylation, Mg-ADP, and Inhibitors Differentially Shape the Conformational Dynamics of the

A-Loop of Aurora-A. *Biomolecules* **2021**, *11* (4), 567. https://doi.org/10.3390/biom11040567.

(69) Gosu, V.; Choi, S. Structural Dynamic Analysis of Apo and ATP-Bound IRAK4 Kinase. *Sci Rep* **2014**, *4*, 5748. https://doi.org/10.1038/srep05748.

(70) Feng, J.; Witthuhn, B. A.; Matsuda, T.; Kohlhuber, F.; Kerr, I. M.; Ihle, J. N. Activation of Jak2 Catalytic Activity Requires Phosphorylation of Y1007 in the Kinase Activation Loop. *Mol Cell Biol* **1997**, *17* (5), 2497–2501. https://doi.org/10.1128/MCB.17.5.2497.

(71) Zhou, Y. J.; Hanson, E. P.; Chen, Y. Q.; Magnuson, K.; Chen, M.; Swann, P. G.; Wange, R. L.; Changelian, P. S.; O'Shea, J. J. Distinct Tyrosine Phosphorylation Sites in JAK3 Kinase Domain Positively and Negatively Regulate Its Enzymatic Activity. *Proc Natl Acad Sci U S A* **1997**, *94* (25), 13850–13855. https://doi.org/10.1073/pnas.94.25.13850.

(72) Gauzzi, M. C.; Velazquez, L.; McKendry, R.; Mogensen, K. E.; Fellous, M.; Pellegrini, S. Interferon-Alpha-Dependent Activation of Tyk2 Requires Phosphorylation of Positive Regulatory Tyrosines by Another Kinase. *J Biol Chem* **1996**, *271* (34), 20494–20500. https://doi.org/10.1074/jbc.271.34.20494.

(73) Banerjee, S.; Biehl, A.; Gadina, M.; Hasni, S.; Schwartz, D. M. JAK–STAT Signaling as a Target for Inflammatory and Autoimmune Diseases: Current and Future Prospects. *Drugs* **2017**, *77* (5), 521–546. https://doi.org/10.1007/s40265-017-0701-9.

(74) Aaronson, D. S.; Horvath, C. M. A Road Map for Those Who Don't Know JAK-STAT. *Science* **2002**, *296* (5573), 1653–1655. https://doi.org/10.1126/science.1071545.

(75) O'Shea, J. J.; Plenge, R. JAK and STAT Signaling Molecules in Immunoregulation and Immune-Mediated Disease. *Immunity* **2012**, *36* (4), 542–550. https://doi.org/10.1016/j.immuni.2012.03.014.

(76) Ivashkiv, L. B.; Donlin, L. T. Regulation of Type I Interferon Responses. *Nat Rev Immunol* **2014**, *14* (1), 36–49. https://doi.org/10.1038/nri3581.

(77) Villarino, A. V.; Kanno, Y.; O'Shea, J. J. Mechanisms and Consequences of Jak-STAT Signaling in the Immune System. *Nat Immunol* **2017**, *18* (4), 374–384. https://doi.org/10.1038/ni.3691.

(78) Hu, X.; Li, J.; Fu, M.; Zhao, X.; Wang, W. The JAK/STAT Signaling Pathway: From Bench to Clinic. *Sig Transduct Target Ther* **2021**, *6* (1), 1–33. https://doi.org/10.1038/s41392-021-00791-1.

(79) T. Virtanen, A.; Haikarainen, T.; Raivola, J.; Silvennoinen, O. Selective JAKinibs: Prospects in Inflammatory and Autoimmune Diseases. *BioDrugs* **2019**, *33* (1), 15–32. https://doi.org/10.1007/s40259-019-00333-w.

(80) Murray, P. J. The JAK-STAT Signaling Pathway: Input and Output Integration. *The Journal of Immunology* **2007**, *178* (5), 2623–2629. https://doi.org/10.4049/jimmunol.178.5.2623.

(81) Alamanos, Y.; Voulgari, P. V.; Drosos, A. A. Incidence and Prevalence of Rheumatoid Arthritis, Based on the 1987 American College of Rheumatology Criteria: A Systematic Review. *Seminars in Arthritis and Rheumatism* **2006**, *36* (3), 182–188. https://doi.org/10.1016/j.semarthrit.2006.08.006.

(82) Chopra, A.; Abdel-Nasser, A. Epidemiology of Rheumatic Musculoskeletal Disorders in the Developing World. *Best Practice & Research Clinical Rheumatology* **2008**, *22* (4), 583–604. https://doi.org/10.1016/j.berh.2008.07.001.
(83) Rendon, A.; Schäkel, K. Psoriasis Pathogenesis and Treatment. *Int J Mol Sci* **2019**, *20* (6), E1475. https://doi.org/10.3390/ijms20061475.
(84) Crispino, N.; Ciccia, F. JAK/STAT Pathway and Nociceptive Cytokine Signalling in Rheumatoid Arthritis and Psoriatic Arthritis. *Clin Exp Rheumatol* **2021**, *39* (3), 668–675.
(85) Luo, Y.; Alexander, M.; Gadina, M.; O'Shea, J. J.; Meylan, F.; Schwartz, D. M. JAK-STAT Signaling in Human Disease: From Genetic Syndromes to Clinical Inhibition. *J Allergy Clin Immunol* **2021**, *148* (4), 911–925. https://doi.org/10.1016/j.jaci.2021.08.004.
(86) Bryan, J. C.; Verstovsek, S. Overcoming Treatment Challenges in Myelofibrosis and Polycythemia Vera: The Role of Ruxolitinib. *Cancer Chemother Pharmacol* **2016**, *77* (6), 1125–1142. https://doi.org/10.1007/s00280-016-3012-z.
(87) Shuai, K.; Ziemiecki, A.; Wilks, A. F.; Harpur, A. G.; Sadowski, H. B.; Gilman, M. Z.; Darnell, J. E. Polypeptide Signalling to the Nucleus through Tyrosine Phosphorylation of Jak and Stat Proteins. *Nature* **1993**, *366* (6455), 580–583. https://doi.org/10.1038/366580a0.
(88) Pfeifer, A. C.; Timmer, J.; Klingmüller, U. Systems Biology of JAK/STAT Signalling. *Essays Biochem* **2008**, *45*, 109–120. https://doi.org/10.1042/BSE0450109.
(89) Seif, F.; Khoshmirsafa, M.; Aazami, H.; Mohsenzadegan, M.; Sedighi, G.; Bahar, M. The Role of JAK-STAT Signaling Pathway and Its Regulators in the Fate of T Helper Cells. *Cell Commun Signal* **2017**, *15* (1), 23. https://doi.org/10.1186/s12964-017-0177-y.
(90) Krebs, D. L.; Uren, R. T.; Metcalf, D.; Rakar, S.; Zhang, J.-G.; Starr, R.; De Souza, D. P.; Hanzinikolas, K.; Eyles, J.; Connolly, L. M.; Simpson, R. J.; Nicola, N. A.; Nicholson, S. E.; Baca, M.; Hilton, D. J.; Alexander, W. S. SOCS-6 Binds to Insulin Receptor Substrate 4, and Mice Lacking the SOCS-6 Gene Exhibit Mild Growth Retardation. *Mol Cell Biol* **2002**, *22* (13), 4567–4578. https://doi.org/10.1128/MCB.22.13.4567-4578.2002.
(91) Trengove, M. C.; Ward, A. C. SOCS Proteins in Development and Disease. *Am J Clin Exp Immunol* **2013**, *2* (1), 1–29.
(92) Vuong, B. Q.; Arenzana, T. L.; Showalter, B. M.; Losman, J.; Chen, X. P.; Mostecki, J.; Banks, A. S.; Limnander, A.; Fernandez, N.; Rothman, P. B. SOCS-1 Localizes to the Microtubule Organizing Complex-Associated 20S Proteasome. *Mol Cell Biol* **2004**, *24* (20), 9092–9101. https://doi.org/10.1128/MCB.24.20.9092-9101.2004.
(93) Chikuma, S.; Kanamori, M.; Mise-Omata, S.; Yoshimura, A. Suppressors of Cytokine Signaling: Potential Immune Checkpoint Molecules for Cancer Immunotherapy. *Cancer Sci* **2017**, *108* (4), 574–580. https://doi.org/10.1111/cas.13194.
(94) Fragoulis, G. E.; McInnes, I. B.; Siebert, S. JAK-Inhibitors. New Players in the Field of Immune-Mediated Diseases, beyond Rheumatoid Arthritis. *Rheumatology* **2019**, *58* (Supplement_1), i43–i54. https://doi.org/10.1093/rheumatology/key276.

(95) Sperti, M.; Malavolta, M.; Ciniero, G.; Borrelli, S.; Cavaglià, M.; Muscat, S.; Tuszynski, J. A.; Afeltra, A.; Margiotta, D. P. E.; Navarini, L. JAK Inhibitors in Immune-Mediated Rheumatic Diseases: From a Molecular Perspective to Clinical Studies. *Journal of Molecular Graphics and Modelling* **2021**, *104*, 107789. https://doi.org/10.1016/j.jmgm.2020.107789.
(96) Garber, K. Pfizer's First-in-Class JAK Inhibitor Pricey for Rheumatoid Arthritis Market. *Nature Biotechnology* **2013**, *31* (1), 3–4. https://doi.org/10.1038/nbt0113-3.
(97) Mayence, A.; Vanden Eynde, J. J. Baricitinib: A 2018 Novel FDA-Approved Small Molecule Inhibiting Janus Kinases. *Pharmaceuticals* **2019**, *12* (1), 37. https://doi.org/10.3390/ph12010037.
(98) Duggan, S.; Keam, S. J. Upadacitinib: First Approval. *Drugs* **2019**, *79* (16), 1819–1828. https://doi.org/10.1007/s40265-019-01211-z.
(99) Taylor, P. C.; Abdul Azeez, M.; Kiriakidis, S. Filgotinib for the Treatment of Rheumatoid Arthritis. *Expert Opinion on Investigational Drugs* **2017**, *26* (10), 1181–1187. https://doi.org/10.1080/13543784.2017.1372422.
(100) Markwick, P. R. L.; Malliavin, T.; Nilges, M. Structural Biology by NMR: Structure, Dynamics, and Interactions. *PLoS Comput Biol* **2008**, *4* (9), e1000168. https://doi.org/10.1371/journal.pcbi.1000168.
(101) Shi, Y. A Glimpse of Structural Biology through X-Ray Crystallography. *Cell* **2014**, *159* (5), 995–1014. https://doi.org/10.1016/j.cell.2014.10.051.
(102) Renaud, J.-P.; Chari, A.; Ciferri, C.; Liu, W.; Rémigy, H.-W.; Stark, H.; Wiesmann, C. Cryo-EM in Drug Discovery: Achievements, Limitations and Prospects. *Nat Rev Drug Discov* **2018**, *17* (7), 471–492. https://doi.org/10.1038/nrd.2018.77.
(103) Lazim, R.; Suh, D.; Choi, S. Advances in Molecular Dynamics Simulations and Enhanced Sampling Methods for the Study of Protein Systems. *IJMS* **2020**, *21* (17), 6339. https://doi.org/10.3390/ijms21176339.
(104) Feig, M.; Sugita, Y. Whole-Cell Models and Simulations in Molecular Detail. *Annu. Rev. Cell Dev. Biol.* **2019**, *35* (1), 191–211. https://doi.org/10.1146/annurev-cellbio-100617-062542.
(105) Hohenberg, P.; Kohn, W. Inhomogeneous Electron Gas. *Phys. Rev.* **1964**, *136* (3B), B864–B871. https://doi.org/10.1103/PhysRev.136.B864.
(106) Kmiecik, S.; Gront, D.; Kolinski, M.; Wieteska, L.; Dawid, A. E.; Kolinski, A. Coarse-Grained Protein Models and Their Applications. *Chem. Rev.* **2016**, *116* (14), 7898–7936. https://doi.org/10.1021/acs.chemrev.6b00163.
(107) Aminpour, M.; Montemagno, C.; Tuszynski, J. A. An Overview of Molecular Modeling for Drug Discovery with Specific Illustrative Examples of Applications. *Molecules* **2019**, *24* (9), 1693. https://doi.org/10.3390/molecules24091693.
(108) Leach, A. R. *Molecular Modelling: Principles and Applications*, Addison Wesley Longman Limited:Essex.; Pearson Education Limited: England, 2001.
(109) Schrödinger, E. Quantisierung als Eigenwertproblem. *Ann. Phys.* **1926**, *384* (4), 361–376. https://doi.org/10.1002/andp.19263840404.
(110) Hartree, D. R. The Wave Mechanics of an Atom with a Non-Coulomb Central Field. Part II. Some Results and Discussion. *Math. Proc. Camb. Phil. Soc.* **1928**, *24* (1), 111–132. https://doi.org/10.1017/S0305004100011920.

(111) Cremer, D. Møller-Plesset Perturbation Theory: From Small Molecule Methods to Methods for Thousands of Atoms: Møller-Plesset Perturbation Theory. *WIREs Comput Mol Sci* **2011**, *1* (4), 509–530. https://doi.org/10.1002/wcms.58.
(112) Bartlett, R. J.; Musiał, M. Coupled-Cluster Theory in Quantum Chemistry. *Rev. Mod. Phys.* **2007**, *79* (1), 291–352. https://doi.org/10.1103/RevModPhys.79.291.
(113) Born, M.; Oppenheimer, R. Zur Quantentheorie der Molekeln. *Ann. Phys.* **1927**, *389* (20), 457–484. https://doi.org/10.1002/andp.19273892002.
(114) Gerber, R. B.; Shemesh, D.; Varner, M. E.; Kalinowski, J.; Hirshberg, B. Ab Initio and Semi-Empirical Molecular Dynamics Simulations of Chemical Reactions in Isolated Molecules and in Clusters. *Phys. Chem. Chem. Phys.* **2014**, *16* (21), 9760–9775. https://doi.org/10.1039/C3CP55239J.
(115) Warshel, A.; Levitt, M. Theoretical Studies of Enzymic Reactions: Dielectric, Electrostatic and Steric Stabilization of the Carbonium Ion in the Reaction of Lysozyme. *Journal of Molecular Biology* **1976**, *103* (2), 227–249. https://doi.org/10.1016/0022-2836(76)90311-9.
(116) Nishizawa, H.; Nishimura, Y.; Kobayashi, M.; Irle, S.; Nakai, H. Three Pillars for Achieving Quantum Mechanical Molecular Dynamics Simulations of Huge Systems: Divide-and-Conquer, Density-Functional Tight-Binding, and Massively Parallel Computation. *J. Comput. Chem.* **2016**, *37* (21), 1983–1992. https://doi.org/10.1002/jcc.24419.
(117) Case, D. A.; Cheatham, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. The Amber Biomolecular Simulation Programs. *J Comput Chem* **2005**, *26* (16), 1668–1688. https://doi.org/10.1002/jcc.20290.
(118) Case, D. A.; Ben-Shalom, I. Y.; Brozell, S. R.; Cerutti, D. S.; Cheatham, III, T. E.; Cruzeiro, V. W. D.; Darden, T. A.; Duke, R. E.; Ghoreishi, D.; Gilson, M. K.; Gohlke, H.; Goetz, A. W.; Greene, D.; Harris, R.; Homeyer, N.; Huang, Y.; Izadi, S.; Kovalenko, A.; Kurtzman, T.; Lee, T. S.; LeGrand, S.; Li, P.; Lin, C.; Liu, J.; Luchko, T.; Luo, R.; Mermelstein, D. J.; Merz, K. M.; Miao, Y.; Monard, G.; Nguyen, C.; Nguyen, H.; Omelyan, I.; Onufriev, A.; Pan, F.; Qi, R.; Roe, D. R.; Roitberg, A.; Sagui, C.; Schott-Verdugo, S.; Shen, J.; Simmerling, S. L.; Smith, J.; Salomon-Ferrer, R.; Swails, J.; Walker, R. C.; Wang, J.; Wei, H.; Wolf, R. M.; Wu, X.; Xiao, L.; York, D. M.; Kollman, P. A. AMBER 2018. *University of California, San Francisco* **2018**.
(119) Brooks, B. R.; Brooks, C. L.; Mackerell, A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M. CHARMM: The Biomolecular Simulation Program. *J. Comput. Chem.* **2009**, *30* (10), 1545–1614. https://doi.org/10.1002/jcc.21287.
(120) Jorgensen, W. L.; Tirado-Rives, J. The OPLS [Optimized Potentials for Liquid Simulations] Potential Functions for Proteins, Energy Minimizations for Crystals of Cyclic Peptides and Crambin. *J. Am. Chem. Soc.* **1988**, *110* (6), 1657–1666. https://doi.org/10.1021/ja00214a001.

(121) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High Performance Molecular Simulations through Multi-Level Parallelism from Laptops to Supercomputers. *SoftwareX* **2015**, *1–2*, 19–25. https://doi.org/10.1016/j.softx.2015.06.001.

(122) Tian, C.; Kasavajhala, K.; Belfon, K. A. A.; Raguette, L.; Huang, H.; Migues, A. N.; Bickel, J.; Wang, Y.; Pincay, J.; Wu, Q.; Simmerling, C. Ff19SB: Amino-Acid-Specific Protein Backbone Parameters Trained against Quantum Mechanics Energy Surfaces in Solution. *J. Chem. Theory Comput.* **2020**, *16* (1), 528–552. https://doi.org/10.1021/acs.jctc.9b00591.

(123) 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. *J. Chem. Theory Comput.* **2015**, *11* (8), 3696–3713. https://doi.org/10.1021/acs.jctc.5b00255.

(124) 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* (3), 712–725. https://doi.org/10.1002/prot.21123.

(125) Harder, E.; Damm, W.; Maple, J.; Wu, C.; Reboul, M.; Xiang, J. Y.; Wang, L.; Lupyan, D.; Dahlgren, M. K.; Knight, J. L.; Kaus, J. W.; Cerutti, D. S.; Krilov, G.; Jorgensen, W. L.; Abel, R.; Friesner, R. A. OPLS3: A Force Field Providing Broad Coverage of Drug-like Small Molecules and Proteins. *J. Chem. Theory Comput.* **2016**, *12* (1), 281–296. https://doi.org/10.1021/acs.jctc.5b00864.

(126) Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, M.; de Groot, B. L.; Grubmüller, H.; MacKerell, A. D. 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.

(127) Zgarbová, M.; Šponer, J.; Otyepka, M.; Cheatham, T. E.; Galindo-Murillo, R.; Jurečka, P. Refinement of the Sugar–Phosphate Backbone Torsion Beta for AMBER Force Fields Improves the Description of Z- and B-DNA. *J. Chem. Theory Comput.* **2015**, *11* (12), 5723–5736. https://doi.org/10.1021/acs.jctc.5b00716.

(128) Zgarbová, M.; Šponer, J.; Jurečka, P. Z-DNA as a Touchstone for Additive Empirical Force Fields and a Refinement of the Alpha/Gamma DNA Torsions for AMBER. *J. Chem. Theory Comput.* **2021**, *17* (10), 6292–6301. https://doi.org/10.1021/acs.jctc.1c00697.

(129) Hart, K.; Foloppe, N.; Baker, C. M.; Denning, E. J.; Nilsson, L.; MacKerell, A. D. Optimization of the CHARMM Additive Force Field for DNA: Improved Treatment of the BI/BII Conformational Equilibrium. *J. Chem. Theory Comput.* **2012**, *8* (1), 348–362. https://doi.org/10.1021/ct200723y.

(130) Denning, E. J.; Priyakumar, U. D.; Nilsson, L.; Mackerell, A. D. Impact of 2′-Hydroxyl Sampling on the Conformational Properties of RNA: Update of the CHARMM All-Atom Additive Force Field for RNA. *J. Comput. Chem.* **2011**, *32* (9), 1929–1943. https://doi.org/10.1002/jcc.21777.

(131) Kirschner, K. N.; Yongye, A. B.; Tschampel, S. M.; González-Outeiriño, J.; Daniels, C. R.; Foley, B. L.; Woods, R. J. GLYCAM06: A Generalizable

Biomolecular Force Field. Carbohydrates. *J Comput Chem* **2008**, *29* (4), 622–655. https://doi.org/10.1002/jcc.20820.

(132) Klauda, J. B.; Venable, R. M.; Freites, J. A.; O'Connor, J. W.; Tobias, D. J.; Mondragon-Ramirez, C.; Vorobyov, I.; MacKerell, A. D.; Pastor, R. W. Update of the CHARMM All-Atom Additive Force Field for Lipids: Validation on Six Lipid Types. *J. Phys. Chem. B* **2010**, *114* (23), 7830–7843. https://doi.org/10.1021/jp101759q.

(133) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and Testing of a General Amber Force Field. *J Comput Chem* **2004**, *25* (9), 1157–1174. https://doi.org/10.1002/jcc.20035.

(134) Dodda, L. S.; Vilseck, J. Z.; Tirado-Rives, J.; Jorgensen, W. L. 1.14*CM1A-LBCC: Localized Bond-Charge Corrected CM1A Charges for Condensed-Phase Simulations. *J. Phys. Chem. B* **2017**, *121* (15), 3864–3870. https://doi.org/10.1021/acs.jpcb.7b00272.

(135) Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; Mackerell, A. D. CHARMM General Force Field: A Force Field for Drug-like Molecules Compatible with the CHARMM All-Atom Additive Biological Force Fields. *J. Comput. Chem.* **2009**, NA-NA. https://doi.org/10.1002/jcc.21367.

(136) Ponder, J. W.; Wu, C.; Ren, P.; Pande, V. S.; Chodera, J. D.; Schnieders, M. J.; Haque, I.; Mobley, D. L.; Lambrecht, D. S.; DiStasio, R. A.; Head-Gordon, M.; Clark, G. N. I.; Johnson, M. E.; Head-Gordon, T. Current Status of the AMOEBA Polarizable Force Field. *J. Phys. Chem. B* **2010**, *114* (8), 2549–2564. https://doi.org/10.1021/jp910674d.

(137) Li, H.; Chowdhary, J.; Huang, L.; He, X.; MacKerell, A. D.; Roux, B. Drude Polarizable Force Field for Molecular Dynamics Simulations of Saturated and Unsaturated Zwitterionic Lipids. *J. Chem. Theory Comput.* **2017**, *13* (9), 4535–4552. https://doi.org/10.1021/acs.jctc.7b00262.

(138) Grubmüller, H.; Heller, H.; Windemuth, A.; Schulten, K. Generalized Verlet Algorithm for Efficient Molecular Dynamics Simulations with Long-Range Interactions. *Molecular Simulation* **1991**, *6* (1–3), 121–142. https://doi.org/10.1080/08927029108022142.

(139) Van Gunsteren, W. F.; Berendsen, H. J. C. A Leap-Frog Algorithm for Stochastic Dynamics. *Molecular Simulation* **1988**, *1* (3), 173–185. https://doi.org/10.1080/08927028808080941.

(140) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. LINCS: A Linear Constraint Solver for Molecular Simulations. *J. Comput. Chem.* **1997**, *18* (12), 1463–1472. https://doi.org/10.1002/(SICI)1096-987X(199709)18:12<1463::AID-JCC4>3.0.CO;2-H.

(141) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of n-Alkanes. *Journal of Computational Physics* **1977**, *23* (3), 327–341. https://doi.org/10.1016/0021-9991(77)90098-5.

(142) Onufriev, A. V.; Case, D. A. Generalized Born Implicit Solvent Models for Biomolecules. *Annu. Rev. Biophys.* **2019**, *48* (1), 275–296. https://doi.org/10.1146/annurev-biophys-052118-115325.

(143) Onufriev, A.; Bashford, D.; Case, D. A. Exploring Protein Native States and Large-Scale Conformational Changes with a Modified Generalized Born Model. *Proteins* **2004**, *55* (2), 383–394. https://doi.org/10.1002/prot.20033.
(144) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. Semianalytical Treatment of Solvation for Molecular Mechanics and Dynamics. *J. Am. Chem. Soc.* **1990**, *112* (16), 6127–6129. https://doi.org/10.1021/ja00172a038.
(145) Onufriev, A. V.; Izadi, S. Water Models for Biomolecular Simulations. *WIREs Comput Mol Sci* **2018**, *8* (2). https://doi.org/10.1002/wcms.1347.
(146) Kadaoluwa Pathirannahalage, S. P.; Meftahi, N.; Elbourne, A.; Weiss, A. C. G.; McConville, C. F.; Padua, A.; Winkler, D. A.; Costa Gomes, M.; Greaves, T. L.; Le, T. C.; Besford, Q. A.; Christofferson, A. J. Systematic Comparison of the Structural and Dynamic Properties of Commonly Used Water Models for Molecular Dynamics Simulations. *J. Chem. Inf. Model.* **2021**, *61* (9), 4521–4536. https://doi.org/10.1021/acs.jcim.1c00794.
(147) Price, D. J.; Brooks, C. L. A Modified TIP3P Water Potential for Simulation with Ewald Summation. *The Journal of Chemical Physics* **2004**, *121* (20), 10096–10103. https://doi.org/10.1063/1.1808117.
(148) Jorgensen, W. L. Quantum and Statistical Mechanical Studies of Liquids. 10. Transferable Intermolecular Potential Functions for Water, Alcohols, and Ethers. Application to Liquid Water. *J. Am. Chem. Soc.* **1981**, *103* (2), 335–340. https://doi.org/10.1021/ja00392a016.
(149) Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. The Missing Term in Effective Pair Potentials. *J. Phys. Chem.* **1987**, *91* (24), 6269–6271. https://doi.org/10.1021/j100308a038.
(150) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Hermans, J. Interaction Models for Water in Relation to Protein Hydration. In *Intermolecular Forces*; Pullman, B., Ed.; The Jerusalem Symposia on Quantum Chemistry and Biochemistry; Springer Netherlands: Dordrecht, 1981; Vol. 14, pp 331–342. https://doi.org/10.1007/978-94-015-7658-1_21.
(151) Jorgensen, W. L.; Madura, J. D. Temperature and Size Dependence for Monte Carlo Simulations of TIP4P Water. *Molecular Physics* **1985**, *56* (6), 1381–1392. https://doi.org/10.1080/00268978500103111.
(152) Khalak, Y.; Baumeier, B.; Karttunen, M. Improved General-Purpose Five-Point Model for Water: TIP5P/2018. *J. Chem. Phys.* **2018**, *149* (22), 224507. https://doi.org/10.1063/1.5070137.
(153) Mahoney, M. W.; Jorgensen, W. L. A Five-Site Model for Liquid Water and the Reproduction of the Density Anomaly by Rigid, Nonpolarizable Potential Functions. *The Journal of Chemical Physics* **2000**, *112* (20), 8910–8922. https://doi.org/10.1063/1.481505.
(154) Nada, H.; van der Eerden, J. P. J. M. An Intermolecular Potential Model for the Simulation of Ice and Water near the Melting Point: A Six-Site Model of H[Sub 2]O. *J. Chem. Phys.* **2003**, *118* (16), 7401. https://doi.org/10.1063/1.1562610.
(155) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular Dynamics with Coupling to an External Bath. *The Journal of Chemical Physics* **1984**, *81* (8), 3684–3690. https://doi.org/10.1063/1.448118.

(156) Andersen, H. C. Molecular Dynamics Simulations at Constant Pressure and/or Temperature. *The Journal of Chemical Physics* **1980**, *72* (4), 2384–2393. https://doi.org/10.1063/1.439486.
(157) Schneider, T.; Stoll, E. Molecular-Dynamics Study of a Three-Dimensional One-Component Model for Distortive Phase Transitions. *Phys. Rev. B* **1978**, *17* (3), 1302–1322. https://doi.org/10.1103/PhysRevB.17.1302.
(158) Hünenberger, P. H. Thermostat Algorithms for Molecular Dynamics Simulations. In *Advanced Computer Simulation*; Dr. Holm, C., Prof. Dr. Kremer, K., Eds.; Abe, A., Joanny, J.-F., Albertsson, A.-C., Duncan, R., Kausch, H.-H., Kobayashi, S., Dušek, K., Lee, K.-S., de Jeu, W. H., Leibler, L., Nuyken, O., Long, T. E., Terentjev, E. M., Voit, B., Manners, I., Wegner, G., Möller, M., Series Eds.; Advances in Polymer Science; Springer Berlin Heidelberg: Berlin, Heidelberg, 2005; Vol. 173, pp 105–149. https://doi.org/10.1007/b99427.
(159) Parrinello, M.; Rahman, A. Polymorphic Transitions in Single Crystals: A New Molecular Dynamics Method. *Journal of Applied Physics* **1981**, *52* (12), 7182–7190. https://doi.org/10.1063/1.328693.
(160) Martyna, G. J.; Tobias, D. J.; Klein, M. L. Constant Pressure Molecular Dynamics Algorithms. *The Journal of Chemical Physics* **1994**, *101* (5), 4177–4189. https://doi.org/10.1063/1.467468.
(161) Martyna, G. J.; Tuckerman, M. E.; Tobias, D. J.; Klein, M. L. Explicit Reversible Integrators for Extended Systems Dynamics. *Molecular Physics* **1996**, *87* (5), 1117–1157. https://doi.org/10.1080/00268979600100761.
(162) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An $N \cdot \log(N)$ Method for Ewald Sums in Large Systems. *The Journal of Chemical Physics* **1993**, *98* (12), 10089–10092. https://doi.org/10.1063/1.464397.
(163) Lane, T. J.; Shukla, D.; Beauchamp, K. A.; Pande, V. S. To Milliseconds and beyond: Challenges in the Simulation of Protein Folding. *Current Opinion in Structural Biology* **2013**, *23* (1), 58–65. https://doi.org/10.1016/j.sbi.2012.11.002.
(164) Englander, S. W.; Mayne, L. The Nature of Protein Folding Pathways. *Proc Natl Acad Sci U S A* **2014**, *111* (45), 15873–15880. https://doi.org/10.1073/pnas.1411798111.
(165) Onuchic, J. N.; Luthey-Schulten, Z.; Wolynes, P. G. Theory of Protein Folding: The Energy Landscape Perspective. *Annu Rev Phys Chem* **1997**, *48*, 545–600. https://doi.org/10.1146/annurev.physchem.48.1.545.
(166) Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. THE Weighted Histogram Analysis Method for Free-Energy Calculations on Biomolecules. I. The Method. *J. Comput. Chem.* **1992**, *13* (8), 1011–1021. https://doi.org/10.1002/jcc.540130812.
(167) Torrie, G. M.; Valleau, J. P. Nonphysical Sampling Distributions in Monte Carlo Free-Energy Estimation: Umbrella Sampling. *Journal of Computational Physics* **1977**, *23* (2), 187–199. https://doi.org/10.1016/0021-9991(77)90121-8.
(168) Bešker, N.; Gervasio, F. L. Using Metadynamics and Path Collective Variables to Study Ligand Binding and Induced Conformational Transitions. *Methods Mol Biol* **2012**, *819*, 501–513. https://doi.org/10.1007/978-1-61779-465-0_29.
(169) Laio, A.; Gervasio, F. L. Metadynamics: A Method to Simulate Rare Events and Reconstruct the Free Energy in Biophysics, Chemistry and Material Science. *Rep.*

Prog. Phys. **2008**, *71* (12), 126601. https://doi.org/10.1088/0034-4885/71/12/126601.
(170) Darve, E.; Wilson, M. A.; Pohorille, A. Calculating Free Energies Using a Scaled-Force Molecular Dynamics Algorithm. *Molecular Simulation* **2002**, *28* (1–2), 113–144. https://doi.org/10.1080/08927020211975.
(171) Darve, E.; Rodríguez-Gómez, D.; Pohorille, A. Adaptive Biasing Force Method for Scalar and Vector Free Energy Calculations. *The Journal of Chemical Physics* **2008**, *128* (14), 144120. https://doi.org/10.1063/1.2829861.
(172) Israelewitz, B.; Baudry, J.; Gullingsrud, J.; Kosztin, D.; Schulten, K. Steered Molecular Dynamics Investigations of Protein Function. *J Mol Graph Model* **2001**, *19* (1), 13–25. https://doi.org/10.1016/s1093-3263(00)00133-9.
(173) Sugita, Y.; Okamoto, Y. Replica-Exchange Molecular Dynamics Method for Protein Folding. *Chemical Physics Letters* **1999**, *314* (1–2), 141–151. https://doi.org/10.1016/S0009-2614(99)01123-9.
(174) Okamoto, Y. Generalized-Ensemble Algorithms: Enhanced Sampling Techniques for Monte Carlo and Molecular Dynamics Simulations. *J Mol Graph Model* **2004**, *22* (5), 425–439. https://doi.org/10.1016/j.jmgm.2003.12.009.
(175) Bouvier, B.; Grubmüller, H. A Molecular Dynamics Study of Slow Base Flipping in DNA Using Conformational Flooding. *Biophys J* **2007**, *93* (3), 770–786. https://doi.org/10.1529/biophysj.106.091751.
(176) Grubmüller, H. Predicting Slow Structural Transitions in Macromolecular Systems: Conformational Flooding. *Phys. Rev. E* **1995**, *52* (3), 2893–2906. https://doi.org/10.1103/PhysRevE.52.2893.
(177) Hansmann, U. H. E. Parallel Tempering Algorithm for Conformational Studies of Biological Molecules. *Chemical Physics Letters* **1997**, *281* (1–3), 140–150. https://doi.org/10.1016/S0009-2614(97)01198-6.
(178) Hamelberg, D.; Mongan, J.; McCammon, J. A. Accelerated Molecular Dynamics: A Promising and Efficient Simulation Method for Biomolecules. *The Journal of Chemical Physics* **2004**, *120* (24), 11919–11929. https://doi.org/10.1063/1.1755656.
(179) Miao, Y.; Feher, V. A.; McCammon, J. A. Gaussian Accelerated Molecular Dynamics: Unconstrained Enhanced Sampling and Free Energy Calculation. *J. Chem. Theory Comput.* **2015**, *11* (8), 3584–3595. https://doi.org/10.1021/acs.jctc.5b00436.
(180) Wang, J.; Arantes, P. R.; Bhattarai, A.; Hsu, R. V.; Pawnikar, S.; Huang, Y. M.; Palermo, G.; Miao, Y. Gaussian Accelerated Molecular Dynamics: Principles and Applications. *WIREs Comput Mol Sci* **2021**, *11* (5). https://doi.org/10.1002/wcms.1521.
(181) Miao, Y.; Bhattarai, A.; Wang, J. Ligand Gaussian Accelerated Molecular Dynamics (LiGaMD): Characterization of Ligand Binding Thermodynamics and Kinetics. *J. Chem. Theory Comput.* **2020**, *16* (9), 5526–5547. https://doi.org/10.1021/acs.jctc.0c00395.
(182) Miao, Y.; Sinko, W.; Pierce, L.; Bucher, D.; Walker, R. C.; McCammon, J. A. Improved Reweighting of Accelerated Molecular Dynamics Simulations for Free Energy Calculation. *J. Chem. Theory Comput.* **2014**, *10* (7), 2677–2689. https://doi.org/10.1021/ct500090q.

(183) Oshima, H.; Re, S.; Sugita, Y. Replica-Exchange Umbrella Sampling Combined with Gaussian Accelerated Molecular Dynamics for Free-Energy Calculation of Biomolecules. *J. Chem. Theory Comput.* **2019**, *15* (10), 5199–5208. https://doi.org/10.1021/acs.jctc.9b00761.

(184) Wang, J.; Miao, Y. Peptide Gaussian Accelerated Molecular Dynamics (Pep-GaMD): Enhanced Sampling and Free Energy and Kinetics Calculations of Peptide Binding. *J Chem Phys* **2020**, *153* (15), 154109. https://doi.org/10.1063/5.0021399.

(185) Sousa, S. F.; Fernandes, P. A.; Ramos, M. J. Protein-Ligand Docking: Current Status and Future Challenges. *Proteins* **2006**, *65* (1), 15–26. https://doi.org/10.1002/prot.21082.

(186) Sousa, S. F.; Ribeiro, A. J. M.; Coimbra, J. T. S.; Neves, R. P. P.; Martins, S. A.; Moorthy, N. S. H. N.; Fernandes, P. A.; Ramos, M. J. Protein-Ligand Docking in the New Millennium--a Retrospective of 10 Years in the Field. *Curr Med Chem* **2013**, *20* (18), 2296–2314. https://doi.org/10.2174/0929867311320180002.

(187) Halgren, T. A.; Murphy, R. B.; Friesner, R. A.; Beard, H. S.; Frye, L. L.; Pollard, W. T.; Banks, J. L. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 2. Enrichment Factors in Database Screening. *J. Med. Chem.* **2004**, *47* (7), 1750–1759. https://doi.org/10.1021/jm030644s.

(188) McGann, M. R.; Almond, H. R.; Nicholls, A.; Grant, J. A.; Brown, F. K. Gaussian Docking Functions. *Biopolymers* **2003**, *68* (1), 76–90. https://doi.org/10.1002/bip.10207.

(189) T. J. Ewing; Kunts, I. D.,. Critical Evaluation of Search Algorithms for Automated Molecular Docking and Database Screening. *J. Comput. Chem.* **1997**, *18* (9), 1175–1189.

(190) Böhm, H. J. The Computer Program LUDI: A New Method for the de Novo Design of Enzyme Inhibitors. *J Comput Aided Mol Des* **1992**, *6* (1), 61–78. https://doi.org/10.1007/BF00124387.

(191) Rarey, M.; Kramer, B.; Lengauer, T.; Klebe, G. A Fast Flexible Docking Method Using an Incremental Construction Algorithm. *J Mol Biol* **1996**, *261* (3), 470–489. https://doi.org/10.1006/jmbi.1996.0477.

(192) Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W. E.; Belew, R. K.; Olson, A. J. Automated Docking Using a Lamarckian Genetic Algorithm and an Empirical Binding Free Energy Function. *J. Comput. Chem.* **1998**, *19* (14), 1639–1662. https://doi.org/10.1002/(SICI)1096-987X(19981115)19:14<1639::AID-JCC10>3.0.CO;2-B.

(193) Goodsell, D. S.; Morris, G. M.; Olson, A. J. Automated Docking of Flexible Ligands: Applications of AutoDock. *J Mol Recognit* **1996**, *9* (1), 1–5. https://doi.org/10.1002/(sici)1099-1352(199601)9:1<1::aid-jmr241>3.0.co;2-6.

(194) Jones, G.; Willett, P.; Glen, R. C. Molecular Recognition of Receptor Sites Using a Genetic Algorithm with a Description of Desolvation. *J Mol Biol* **1995**, *245* (1), 43–53. https://doi.org/10.1016/s0022-2836(95)80037-9.

(195) Eldridge, M. D.; Murray, C. W.; Auton, T. R.; Paolini, G. V.; Mee, R. P. Empirical Scoring Functions: I. The Development of a Fast Empirical Scoring Function to Estimate the Binding Affinity of Ligands in Receptor Complexes. *Journal of Computer-Aided Molecular Design* **1997**, *11* (5), 425–445. https://doi.org/10.1023/A:1007996124545.

(196) Tao, P.; Lai, L. Protein Ligand Docking Based on Empirical Method for Binding Affinity Estimation. *Journal of Computer-Aided Molecular Design* **2001**, *15* (5), 429–446. https://doi.org/10.1023/A:1011188704521.
(197) Wang, R.; Lai, L.; Wang, S. Further Development and Validation of Empirical Scoring Functions for Structure-Based Binding Affinity Prediction. *Journal of Computer-Aided Molecular Design* **2002**, *16* (1), 11–26. https://doi.org/10.1023/A:1016357811882.
(198) Gohlke, H.; Hendlich, M.; Klebe, G. Knowledge-Based Scoring Function to Predict Protein-Ligand Interactions. *Journal of Molecular Biology* **2000**, *295* (2), 337–356. https://doi.org/10.1006/jmbi.1999.3371.
(199) DeWitte, R. S.; Shakhnovich, E. I. SMoG: De Novo Design Method Based on Simple, Fast, and Accurate Free Energy Estimates. 1. Methodology and Supporting Evidence. *J. Am. Chem. Soc.* **1996**, *118* (47), 11733–11744. https://doi.org/10.1021/ja960751u.
(200) Bahar, I.; Atilgan, A. R.; Demirel, M. C.; Erman, B. Vibrational Dynamics of Folded Proteins: Significance of Slow and Fast Motions in Relation to Function and Stability. *Phys. Rev. Lett.* **1998**, *80* (12), 2733–2736. https://doi.org/10.1103/PhysRevLett.80.2733.
(201) Skjaerven, L.; Martinez, A.; Reuter, N. Principal Component and Normal Mode Analysis of Proteins; a Quantitative Comparison Using the GroEL Subunit: Comparison of PCA and NMA for Proteins. *Proteins* **2011**, *79* (1), 232–243. https://doi.org/10.1002/prot.22875.
(202) Mu, Y.; Nguyen, P. H.; Stock, G. Energy Landscape of a Small Peptide Revealed by Dihedral Angle Principal Component Analysis. *Proteins* **2005**, *58* (1), 45–52. https://doi.org/10.1002/prot.20310.
(203) Yang, Y.; Pan, L.; Lightstone, F. C.; Merz, K. M. The Role of Molecular Dynamics Potential of Mean Force Calculations in the Investigation of Enzyme Catalysis. In *Methods in Enzymology*; Elsevier, 2016; Vol. 577, pp 1–29. https://doi.org/10.1016/bs.mie.2016.05.040.
(204) Chong, S.-H.; Im, H.; Ham, S. Explicit Characterization of the Free Energy Landscape of PKID–KIX Coupled Folding and Binding. *ACS Cent. Sci.* **2019**, *5* (8), 1342–1351. https://doi.org/10.1021/acscentsci.9b00200.
(205) Wales, D. J.; Bogdan, T. V. Potential Energy and Free Energy Landscapes. *J. Phys. Chem. B* **2006**, *110* (42), 20765–20776. https://doi.org/10.1021/jp0680544.
(206) Pietrucci, F. Strategies for the Exploration of Free Energy Landscapes: Unity in Diversity and Challenges Ahead. *Reviews in Physics* **2017**, *2*, 32–45. https://doi.org/10.1016/j.revip.2017.05.001.
(207) Felline, A.; Seeber, M.; Fanelli, F. WebPSN v2.0: A Webserver to Infer Fingerprints of Structural Communication in Biomacromolecules. *Nucleic Acids Research* **2020**, *48* (W1), W94–W103. https://doi.org/10.1093/nar/gkaa397.
(208) Chakrabarty, B.; Naganathan, V.; Garg, K.; Agarwal, Y.; Parekh, N. NAPS Update: Network Analysis of Molecular Dynamics Data and Protein–Nucleic Acid Complexes. *Nucleic Acids Research* **2019**, *47* (W1), W462–W470. https://doi.org/10.1093/nar/gkz399.
(209) Clark, A. J.; Gindin, T.; Zhang, B.; Wang, L.; Abel, R.; Murret, C. S.; Xu, F.; Bao, A.; Lu, N. J.; Zhou, T.; Kwong, P. D.; Shapiro, L.; Honig, B.; Friesner, R. A. Free

Energy Perturbation Calculation of Relative Binding Free Energy between Broadly Neutralizing Antibodies and the Gp120 Glycoprotein of HIV-1. *Journal of Molecular Biology* **2017**, *429* (7), 930–947. https://doi.org/10.1016/j.jmb.2016.11.021.

(210) Reddy, M. R.; Erion, M. D. Calculation of Relative Binding Free Energy Differences for Fructose 1,6-Bisphosphatase Inhibitors Using the Thermodynamic Cycle Perturbation Approach. *J. Am. Chem. Soc.* **2001**, *123* (26), 6246–6252. https://doi.org/10.1021/ja0103288.

(211) Hirono, S.; Kollman, P. A. Calculation of the Relative Binding Free Energy of 2'GMP and 2'AMP to Ribonuclease T1 Using Molecular Dynamics/Free Energy Perturbation Approaches. *J Mol Biol* **1990**, *212* (1), 197–209. https://doi.org/10.1016/0022-2836(90)90315-D.

(212) Woo, H.-J.; Roux, B. Calculation of Absolute Protein–Ligand Binding Free Energy from Computer Simulations. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102* (19), 6825–6830. https://doi.org/10.1073/pnas.0409005102.

(213) Mobley, D. L.; Graves, A. P.; Chodera, J. D.; McReynolds, A. C.; Shoichet, B. K.; Dill, K. A. Predicting Absolute Ligand Binding Free Energies to a Simple Model Site. *Journal of Molecular Biology* **2007**, *371* (4), 1118–1134. https://doi.org/10.1016/j.jmb.2007.06.002.

(214) Zwanzig, R. W. High-Temperature Equation of State by a Perturbation Method. I. Nonpolar Gases. *The Journal of Chemical Physics* **1954**, *22* (8), 1420–1426. https://doi.org/10.1063/1.1740409.

(215) Kollman, Peter. Free Energy Calculations: Applications to Chemical and Biochemical Phenomena. *Chem. Rev.* **1993**, *93* (7), 2395–2417. https://doi.org/10.1021/cr00023a004.

(216) Straatsma, T. P.; Berendsen, H. J. C. Free Energy of Ionic Hydration: Analysis of a Thermodynamic Integration Technique to Evaluate Free Energy Differences by Molecular Dynamics Simulations. *The Journal of Chemical Physics* **1988**, *89* (9), 5876–5886. https://doi.org/10.1063/1.455539.

(217) Straatsma, T. P.; Berendsen, H. J. C.; Postma, J. P. M. Free Energy of Hydrophobic Hydration: A Molecular Dynamics Study of Noble Gases in Water. *The Journal of Chemical Physics* **1986**, *85* (11), 6720–6727. https://doi.org/10.1063/1.451846.

(218) Straatsma, T. P.; McCammon, J. A. Multiconfiguration Thermodynamic Integration. *The Journal of Chemical Physics* **1991**, *95* (2), 1175–1188. https://doi.org/10.1063/1.461148.

(219) Wang, C.; Greene, D.; Xiao, L.; Qi, R.; Luo, R. Recent Developments and Applications of the MMPBSA Method. *Front. Mol. Biosci.* **2018**, *4*, 87. https://doi.org/10.3389/fmolb.2017.00087.

(220) Wang, C.; Nguyen, P. H.; Pham, K.; Huynh, D.; Le, T.-B. N.; Wang, H.; Ren, P.; Luo, R. Calculating Protein-Ligand Binding Affinities with MMPBSA: Method and Error Analysis. *J Comput Chem* **2016**, *37* (27), 2436–2446. https://doi.org/10.1002/jcc.24467.

(221) Martins, S. A.; Perez, M. A. S.; Moreira, I. S.; Sousa, S. F.; Ramos, M. J.; Fernandes, P. A. Computational Alanine Scanning Mutagenesis: MM-PBSA vs TI. *J. Chem. Theory Comput.* **2013**, *9* (3), 1311–1319. https://doi.org/10.1021/ct4000372.

(222) Gouda, H.; Kuntz, I. D.; Case, D. A.; Kollman, P. A. Free Energy Calculations for Theophylline Binding to an RNA Aptamer: Comparison of MM-PBSA and Thermodynamic Integration Methods. *Biopolymers* **2003**, *68* (1), 16–34. https://doi.org/10.1002/bip.10270.
(223) Srinivasan, J.; Cheatham, T. E.; Cieplak, P.; Kollman, P. A.; Case, D. A. Continuum Solvent Studies of the Stability of DNA, RNA, and Phosphoramidate−DNA Helices. *J. Am. Chem. Soc.* **1998**, *120* (37), 9401–9409. https://doi.org/10.1021/ja981844+.
(224) Kollman, P. A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W.; Donini, O.; Cieplak, P.; Srinivasan, J.; Case, D. A.; Cheatham, T. E. Calculating Structures and Free Energies of Complex Molecules: Combining Molecular Mechanics and Continuum Models. *Acc. Chem. Res.* **2000**, *33* (12), 889–897. https://doi.org/10.1021/ar000033j.
(225) Weiser, J.; Shenkin, P. S.; Still, W. C. Fast, Approximate Algorithm for Detection of Solvent-Inaccessible Atoms. *Journal of Computational Chemistry* **1999**, *20* (6), 586–596.
(226) Fogolari, F.; Brigo, A.; Molinari, H. The Poisson-Boltzmann Equation for Biomolecular Electrostatics: A Tool for Structural Biology. *J. Mol. Recognit.* **2002**, *15* (6), 377–392. https://doi.org/10.1002/jmr.577.
(227) Mongan, J.; Simmerling, C.; McCammon, J. A.; Case, D. A.; Onufriev, A. Generalized Born Model with a Simple, Robust Molecular Volume Correction. *J Chem Theory Comput* **2007**, *3* (1), 156–169. https://doi.org/10.1021/ct600085e.
(228) Onufriev, A.; Bashford, D.; Case, D. A. Modification of the Generalized Born Model Suitable for Macromolecules. *J. Phys. Chem. B* **2000**, *104* (15), 3712–3720. https://doi.org/10.1021/jp994072s.
(229) Kassem, S.; Ahmed, M.; El-Sheikh, S.; Barakat, K. H. Entropy in Bimolecular Simulations: A Comprehensive Review of Atomic Fluctuations-Based Methods. *J Mol Graph Model* **2015**, *62*, 105–117. https://doi.org/10.1016/j.jmgm.2015.09.010.
(230) Xu, B.; Shen, H.; Zhu, X.; Li, G. Fast and Accurate Computation Schemes for Evaluating Vibrational Entropy of Proteins. *J. Comput. Chem.* **2011**, *32* (15), 3188–3193. https://doi.org/10.1002/jcc.21900.
(231) Karplus, M.; Kushick, J. N. Method for Estimating the Configurational Entropy of Macromolecules. *Macromolecules* **1981**, *14* (2), 325–332. https://doi.org/10.1021/ma50003a019.
(232) Carlsson, J.; Åqvist, J. Calculations of Solute and Solvent Entropies from Molecular Dynamics Simulations. *Phys. Chem. Chem. Phys.* **2006**, *8* (46), 5385–5395. https://doi.org/10.1039/B608486A.
(233) Ihle, J. N. Cytokine Receptor Signalling. *Nature* **1995**, *377* (6550), 591–594. https://doi.org/10.1038/377591a0.
(234) Burja, B.; Mertelj, T.; Frank-Bertoncelj, M. Hi-JAKi-Ng Synovial Fibroblasts in Inflammatory Arthritis With JAK Inhibitors. *Frontiers in Medicine* **2020**, *7*.
(235) Genovese, M. C.; McKay, J. D.; Nasonov, E. L.; Mysler, E. F.; da Silva, N. A.; Alecock, E.; Woodworth, T.; Gomez-Reino, J. J. Interleukin-6 Receptor Inhibition with Tocilizumab Reduces Disease Activity in Rheumatoid Arthritis with Inadequate Response to Disease-Modifying Antirheumatic Drugs: The

Tocilizumab in Combination with Traditional Disease-Modifying Antirheumatic Drug Therapy Study. *Arthritis & Rheumatism* **2008**, *58* (10), 2968–2980. https://doi.org/10.1002/art.23940.

(236) Geiler, J.; Buch, M.; McDermott, M. F. Anti-TNF Treatment in Rheumatoid Arthritis. *Current Pharmaceutical Design 17* (29), 3141–3154.

(237) Kremer, J. M.; Westhovens, R.; Leon, M.; Di Giorgio, E.; Alten, R.; Steinfeld, S.; Russell, A.; Dougados, M.; Emery, P.; Nuamah, I. F.; Williams, G. R.; Becker, J.-C.; Hagerty, D. T.; Moreland, L. W. Treatment of Rheumatoid Arthritis by Selective Inhibition of T-Cell Activation with Fusion Protein CTLA4Ig. *New England Journal of Medicine* **2003**, *349* (20), 1907–1915. https://doi.org/10.1056/NEJMoa035075.

(238) Parmentier, J. M.; Voss, J.; Graff, C.; Schwartz, A.; Argiriadi, M.; Friedman, M.; Camp, H. S.; Padley, R. J.; George, J. S.; Hyland, D.; Rosebraugh, M.; Wishart, N.; Olson, L.; Long, A. J. In Vitro and in Vivo Characterization of the JAK1 Selectivity of Upadacitinib (ABT-494). *BMC Rheumatology* **2018**, *2* (1), 23. https://doi.org/10.1186/s41927-018-0031-x.

(239) Tanaka, Y. Current Concepts in the Management of Rheumatoid Arthritis. *Korean J Intern Med* **2016**, *31* (2), 210–218. https://doi.org/10.3904/kjim.2015.137.

(240) Schwartz, D. M.; Bonelli, M.; Gadina, M.; O'Shea, J. J. Type I/II Cytokines, JAKs, and New Strategies for Treating Autoimmune Diseases. *Nat Rev Rheumatol* **2016**, *12* (1), 25–36. https://doi.org/10.1038/nrrheum.2015.167.

(241) McInnes, I. B.; Buckley, C. D.; Isaacs, J. D. Cytokines in Rheumatoid Arthritis — Shaping the Immunological Landscape. *Nat Rev Rheumatol* **2016**, *12* (1), 63–68. https://doi.org/10.1038/nrrheum.2015.171.

(242) Leonard, W. J.; O'Shea, J. J. JAKS AND STATS: Biological Implications. *Annual Review of Immunology* **1998**, *16* (1), 293–322. https://doi.org/10.1146/annurev.immunol.16.1.293.

(243) McInnes, I. B.; Byers, N. L.; Higgs, R. E.; Lee, J.; Macias, W. L.; Na, S.; Ortmann, R. A.; Rocha, G.; Rooney, T. P.; Wehrman, T.; Zhang, X.; Zuckerman, S. H.; Taylor, P. C. Comparison of Baricitinib, Upadacitinib, and Tofacitinib Mediated Regulation of Cytokine Signaling in Human Leukocyte Subpopulations. *Arthritis Research & Therapy* **2019**, *21* (1), 183. https://doi.org/10.1186/s13075-019-1964-1.

(244) Gao, H.; Mei, S.; Zhao, J.; Zheng, K.; Liao, S. Study on the Binding Mode of a Pyrrolotriazin Derivative with JAK2 by Docking and MD Simulation. *Molecular Simulation* **2019**, *45* (3), 230–238. https://doi.org/10.1080/08927022.2018.1557330.

(245) Taylor, S. S.; Kornev, A. P. Protein Kinases: Evolution of Dynamic Regulatory Proteins. *Trends in Biochemical Sciences* **2011**, *36* (2), 65–77. https://doi.org/10.1016/j.tibs.2010.09.006.

(246) Kornev, A. P.; Haste, N. M.; Taylor, S. S.; Eyck, L. F. T. Surface Comparison of Active and Inactive Protein Kinases Identifies a Conserved Activation Mechanism. *PNAS* **2006**, *103* (47), 17783–17788. https://doi.org/10.1073/pnas.0607656103.

(247) Kornev, A. P.; Taylor, S. S.; Eyck, L. F. T. A Helix Scaffold for the Assembly of Active Protein Kinases. *PNAS* **2008**, *105* (38), 14377–14382. https://doi.org/10.1073/pnas.0807988105.

(248) McInnes, I. B.; Higgs, R.; Lee, J.; Macias, W. L.; Na, S.; Ortmann, R. A.; Rocha, G.; Wehrman, T.; Zhang, X.; Zuckerman, S. H.; Taylor, P. C.; Perrier, C. P059 Ex Vivo Comparison of Baricitinib, Upadacitinib, Filgotinib and Tofacitinib for Cytokine Signalling in Human Leukocyte Subpopulations. *Annals of the Rheumatic Diseases* **2018**, *77* (Suppl 1), A38–A38. https://doi.org/10.1136/annrheumdis-2018-EWRR2018.78.

(249) Gadina, M.; Le, M. T.; Schwartz, D. M.; Silvennoinen, O.; Nakayamada, S.; Yamaoka, K.; O'Shea, J. J. Janus Kinases to Jakinibs: From Basic Insights to Clinical Practice. *Rheumatology* **2019**, *58* (Supplement_1), i4–i16. https://doi.org/10.1093/rheumatology/key432.

(250) Hamaguchi, H.; Amano, Y.; Moritomo, A.; Shirakami, S.; Nakajima, Y.; Nakai, K.; Nomura, N.; Ito, M.; Higashi, Y.; Inoue, T. Discovery and Structural Characterization of Peficitinib (ASP015K) as a Novel and Potent JAK Inhibitor. *Bioorganic & Medicinal Chemistry* **2018**, *26* (18), 4971–4983. https://doi.org/10.1016/j.bmc.2018.08.005.

(251) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF Chimera-A Visualization System for Exploratory Research and Analysis. *J. Comput. Chem.* **2004**, *25* (13), 1605–1612. https://doi.org/10.1002/jcc.20084.

(252) Webb, B.; Sali, A. Comparative Protein Structure Modeling Using MODELLER. *Current Protocols in Bioinformatics* **2016**, *54* (1). https://doi.org/10.1002/cpbi.3.

(253) Kim, S.; Chen, J.; Cheng, T.; Gindulyte, A.; He, J.; He, S.; Li, Q.; Shoemaker, B. A.; Thiessen, P. A.; Yu, B.; Zaslavsky, L.; Zhang, J.; Bolton, E. E. PubChem 2019 Update: Improved Access to Chemical Data. *Nucleic Acids Res* **2019**, *47* (D1), D1102–D1109. https://doi.org/10.1093/nar/gky1033.

(254) Olsson, M. H. M.; Søndergaard, C. R.; Rostkowski, M.; Jensen, J. H. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical PKa Predictions. *J Chem Theory Comput* **2011**, *7* (2), 525–537. https://doi.org/10.1021/ct100578z.

(255) Case, D. A.; Ben-Shalom, I. Y.; Brozell, S. R.; Cerutti, D. S.; Cheatham, III, T. E.; Cruzeiro, V. W. D.; Darden, T. A.; Duke, R. E.; Ghoreishi, D.; Gilson, M. K.; Gohlke, H.; Goetz, A. W.; Greene, D.; Harris, R.; Homeyer, N.; Huang, Y.; Izadi, S.; Kovalenko, A.; Kurtzman, T.; Lee, T. S.; LeGrand, S.; Li, P.; Lin, C.; Liu, J.; Luchko, T.; Luo, R.; Mermelstein, D. J.; Merz, K. M.; Miao, Y.; Monard, G.; Nguyen, C.; Nguyen, H.; Omelyan, I.; Onufriev, A.; Pan, F.; Qi, R.; Roe, D. R.; Roitberg, A.; Sagui, C.; Schott-Verdugo, S.; Shen, J.; Simmerling, S. L.; Smith, J.; Salomon-Ferrer, R.; Swails, J.; Walker, R. C.; Wang, J.; Wei, H.; Wolf, R. M.; Wu, X.; Xiao, L.; York, D. M.; Kollman, P. A. AMBER 2018. *University of California, San Francisco* **2018**.

(256) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and Testing of a General Amber Force Field. *J Comput Chem* **2004**, *25* (9), 1157–1174. https://doi.org/10.1002/jcc.20035.

(257) 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. *J. Chem. Theory Comput.* **2015**, *11* (8), 3696–3713. https://doi.org/10.1021/acs.jctc.5b00255.

(258) Jakalian, A.; Jack, D. B.; Bayly, C. I. Fast, Efficient Generation of High-Quality Atomic Charges. AM1-BCC Model: II. Parameterization and Validation. *J. Comput. Chem.* **2002**, *23* (16), 1623–1641. https://doi.org/10.1002/jcc.10128.

(259) Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. Automatic Atom Type and Bond Type Perception in Molecular Mechanical Calculations. *Journal of Molecular Graphics and Modelling* **2006**, *25* (2), 247–260. https://doi.org/10.1016/j.jmgm.2005.12.005.

(260) Homeyer, N.; Horn, A. H. C.; Lanig, H.; Sticht, H. AMBER Force-Field Parameters for Phosphorylated Amino Acids in Different Protonation States: Phosphoserine, Phosphothreonine, Phosphotyrosine, and Phosphohistidine. *J Mol Model* **2006**, *12* (3), 281–289. https://doi.org/10.1007/s00894-005-0028-4.

(261) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *The Journal of Chemical Physics* **1983**, *79* (2), 926–935. https://doi.org/10.1063/1.445869.

(262) Kräutler, V.; Gunsteren, W. F. van; Hünenberger, P. H. A Fast SHAKE Algorithm to Solve Distance Constraint Equations for Small Molecules in Molecular Dynamics Simulations. *Journal of Computational Chemistry* **2001**, *22* (5), 501–508. https://doi.org/10.1002/1096-987X(20010415)22:5<501::AID-JCC1021>3.0.CO;2-V.

(263) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An $N \cdot \log(N)$ Method for Ewald Sums in Large Systems. *The Journal of Chemical Physics* **1993**, *98* (12), 10089–10092. https://doi.org/10.1063/1.464397.

(264) Pastor, R. W.; Brooks, B. R.; Szabo, A. An Analysis of the Accuracy of Langevin and Molecular Dynamics Algorithms. *Molecular Physics* **1988**, *65* (6), 1409–1419. https://doi.org/10.1080/00268978800101881.

(265) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular Dynamics with Coupling to an External Bath. *The Journal of Chemical Physics* **1984**, *81* (8), 3684–3690. https://doi.org/10.1063/1.448118.

(266) Roe, D. R.; Cheatham, T. E. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J. Chem. Theory Comput.* **2013**, *9* (7), 3084–3095. https://doi.org/10.1021/ct400341p.

(267) Hünenberger, P. H.; Mark, A. E.; van Gunsteren, W. F. Fluctuation and Cross-Correlation Analysis of Protein Motions Observed in Nanosecond Molecular Dynamics Simulations. *Journal of Molecular Biology* **1995**, *252* (4), 492–503. https://doi.org/10.1006/jmbi.1995.0514.

(268) Ichiye, T.; Karplus, M. Collective Motions in Proteins: A Covariance Analysis of Atomic Fluctuations in Molecular Dynamics and Normal Mode Simulations. *Proteins* **1991**, *11* (3), 205–217. https://doi.org/10.1002/prot.340110305.

(269) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput.* **2008**, *4* (3), 435–447. https://doi.org/10.1021/ct700301q.

(270) Sun, W. The Relationship Between Low-Frequency Motions and Community Structure of Residue Network in Protein Molecules. *Journal of Computational Biology* **2018**, *25* (1), 103–113. https://doi.org/10.1089/cmb.2017.0171.

(271) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. *Journal of Molecular Graphics* **1996**, *14* (1), 33–38. https://doi.org/10.1016/0263-7855(96)00018-5.
(272) Miller, B. R.; McGee, T. D.; Swails, J. M.; Homeyer, N.; Gohlke, H.; Roitberg, A. E. MMPBSA.Py: An Efficient Program for End-State Free Energy Calculations. *J. Chem. Theory Comput.* **2012**, *8* (9), 3314–3321. https://doi.org/10.1021/ct300418h.
(273) Kollman, P. A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W.; Donini, O.; Cieplak, P.; Srinivasan, J.; Case, D. A.; Cheatham, T. E. Calculating Structures and Free Energies of Complex Molecules: Combining Molecular Mechanics and Continuum Models. *Acc. Chem. Res.* **2000**, *33* (12), 889–897. https://doi.org/10.1021/ar000033j.
(274) Singh, S.; Sk, M. F.; Sonawane, A.; Kar, P.; Sadhukhan, S. Plant-Derived Natural Polyphenols as Potential Antiviral Drugs against SARS-CoV-2 via RNA-Dependent RNA Polymerase (RdRp) Inhibition: An in-Silico Analysis. *J Biomol Struct Dyn* **2020**, 1–16. https://doi.org/10.1080/07391102.2020.1796810.
(275) Sk, M. F.; Jonniya, N. A.; Roy, R.; Poddar, S.; Kar, P. Computational Investigation of Structural Dynamics of SARS-CoV-2 Methyltransferase-Stimulatory Factor Heterodimer Nsp16/Nsp10 Bound to the Cofactor SAM. *Front. Mol. Biosci.* **2020**, *7*, 590165. https://doi.org/10.3389/fmolb.2020.590165.
(276) Sk, M. F.; Roy, R.; Jonniya, N. A.; Poddar, S.; Kar, P. Elucidating Biophysical Basis of Binding of Inhibitors to SARS-CoV-2 Main Protease by Using Molecular Dynamics Simulations and Free Energy Calculations. *Journal of Biomolecular Structure and Dynamics* **2021**, *39* (10), 3649–3661. https://doi.org/10.1080/07391102.2020.1768149.
(277) Jonniya, N. A.; Sk, M. F.; Kar, P. Investigating Phosphorylation-Induced Conformational Changes in WNK1 Kinase by Molecular Dynamics Simulations. *ACS Omega* **2019**, *4* (17), 17404–17416. https://doi.org/10.1021/acsomega.9b02187.
(278) Jonniya, N. A.; Kar, P. Investigating Specificity of the Anti-Hypertensive Inhibitor WNK463 against With-No-Lysine Kinase Family Isoforms via Multiscale Simulations. *Journal of Biomolecular Structure and Dynamics* **2020**, *38* (5), 1306–1321. https://doi.org/10.1080/07391102.2019.1602079.
(279) Sk, M. F.; Roy, R.; Kar, P. Exploring the Potency of Currently Used Drugs against HIV-1 Protease of Subtype D Variant by Using Multiscale Simulations. *Journal of Biomolecular Structure and Dynamics* **2021**, *39* (3), 988–1003. https://doi.org/10.1080/07391102.2020.1724196.
(280) Sk, M. F.; Jonniya, N. A.; Kar, P. Exploring the Energetic Basis of Binding of Currently Used Drugs against HIV-1 Subtype CRF01_AE Protease via Molecular Dynamics Simulations. *Journal of Biomolecular Structure and Dynamics* **2020**, *0* (0), 1–18. https://doi.org/10.1080/07391102.2020.1794965.
(281) Jonniya, N. A.; Sk, M. F.; Kar, P. A Comparative Study of Structural and Conformational Properties of WNK Kinase Isoforms Bound to an Inhibitor: Insights from Molecular Dynamic Simulations. *Journal of Biomolecular Structure and Dynamics* **2020**, *0* (0), 1–16. https://doi.org/10.1080/07391102.2020.1827035.

(282) Hartigan, J. A.; Wong, M. A. Algorithm AS 136: A K-Means Clustering Algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics)* **1979**, *28* (1), 100–108. https://doi.org/10.2307/2346830.
(283) Chang, C. A.; Chen, W.; Gilson, M. K. Ligand Configurational Entropy and Protein Binding. *PNAS* **2007**, *104* (5), 1534–1539. https://doi.org/10.1073/pnas.0610494104.
(284) Sanachai, K.; Mahalapbutr, P.; Choowongkomon, K.; Poo-arporn, R. P.; Wolschann, P.; Rungrotmongkol, T. Insights into the Binding Recognition and Susceptibility of Tofacitinib toward Janus Kinases. *ACS Omega* **2020**, *5* (1), 369–377. https://doi.org/10.1021/acsomega.9b02800.
(285) Chen, D.; Oezguen, N.; Urvil, P.; Ferguson, C.; Dann, S. M.; Savidge, T. C. Regulation of Protein-Ligand Binding Affinity by Hydrogen Bond Pairing. *Sci. Adv.* **2016**, *2* (3), e1501240. https://doi.org/10.1126/sciadv.1501240.
(286) Huang, Y. M.; Kang, M.; Chang, C. A. Switches of Hydrogen Bonds during Ligand–Protein Association Processes Determine Binding Kinetics. *Journal of Molecular Recognition* **2014**, *27* (9), 537–548. https://doi.org/10.1002/jmr.2377.
(287) Davis, R. R.; Li, B.; Yun, S. Y.; Chan, A.; Nareddy, P.; Gunawan, S.; Ayaz, M.; Lawrence, H. R.; Reuther, G. W.; Lawrence, N. J.; Schönbrunn, E. Structural Insights into JAK2 Inhibition by Ruxolitinib, Fedratinib, and Derivatives Thereof. *J. Med. Chem.* **2021**, *64* (4), 2228–2241. https://doi.org/10.1021/acs.jmedchem.0c01952.
(288) Mahalapbutr, P.; Kongtaworn, N.; Rungrotmongkol, T. Structural Insight into the Recognition of S-Adenosyl-L-Homocysteine and Sinefungin in SARS-CoV-2 Nsp16/Nsp10 RNA Cap 2′-O-Methyltransferase. *Computational and Structural Biotechnology Journal* **2020**, *18*, 2757–2765. https://doi.org/10.1016/j.csbj.2020.09.032.
(289) O'Shea, J. J.; Gadina, M.; Schreiber, R. D. Cytokine Signaling in 2002: New Surprises in the Jak/Stat Pathway. *Cell* **2002**, *109 Suppl*, S121-131. https://doi.org/10.1016/s0092-8674(02)00701-8.
(290) Campbell, I. L. Cytokine-Mediated Inflammation, Tumorigenesis, and Disease-Associated JAK/STAT/SOCS Signaling Circuits in the CNS. *Brain Res Brain Res Rev* **2005**, *48* (2), 166–177. https://doi.org/10.1016/j.brainresrev.2004.12.006.
(291) Banfield, C.; Scaramozza, M.; Zhang, W.; Kieras, E.; Page, K. M.; Fensome, A.; Vincent, M.; Dowty, M. E.; Goteti, K.; Winkle, P. J.; Peeva, E. The Safety, Tolerability, Pharmacokinetics, and Pharmacodynamics of a TYK2/JAK1 Inhibitor (PF-06700841) in Healthy Subjects and Patients With Plaque Psoriasis. *J Clin Pharmacol* **2018**, *58* (4), 434–447. https://doi.org/10.1002/jcph.1046.
(292) Farmer, L. J.; Ledeboer, M. W.; Hoock, T.; Arnost, M. J.; Bethiel, R. S.; Bennani, Y. L.; Black, J. J.; Brummel, C. L.; Chakilam, A.; Dorsch, W. A.; Fan, B.; Cochran, J. E.; Halas, S.; Harrington, E. M.; Hogan, J. K.; Howe, D.; Huang, H.; Jacobs, D. H.; Laitinen, L. M.; Liao, S.; Mahajan, S.; Marone, V.; Martinez-Botella, G.; McCarthy, P.; Messersmith, D.; Namchuk, M.; Oh, L.; Penney, M. S.; Pierce, A. C.; Raybuck, S. A.; Rugg, A.; Salituro, F. G.; Saxena, K.; Shannon, D.; Shlyakter, D.; Swenson, L.; Tian, S.-K.; Town, C.; Wang, J.; Wang, T.; Wannamaker, M. W.; Winquist, R. J.; Zuccola, H. J. Discovery of VX-509 (Decernotinib): A Potent and Selective Janus Kinase 3 Inhibitor for the Treatment of Autoimmune Diseases. *J*

Med Chem **2015**, *58* (18), 7195–7216. https://doi.org/10.1021/acs.jmedchem.5b00301.

(293) Welsch, K.; Holstein, J.; Laurence, A.; Ghoreschi, K. Targeting JAK/STAT Signalling in Inflammatory Skin Diseases with Small Molecule Inhibitors. *Eur J Immunol* **2017**, *47* (7), 1096–1107. https://doi.org/10.1002/eji.201646680.

(294) Hilton, D. J.; Richardson, R. T.; Alexander, W. S.; Viney, E. M.; Willson, T. A.; Sprigg, N. S.; Starr, R.; Nicholson, S. E.; Metcalf, D.; Nicola, N. A. Twenty Proteins Containing a C-Terminal SOCS Box Form Five Structural Classes. *Proceedings of the National Academy of Sciences* **1998**, *95* (1), 114–119. https://doi.org/10.1073/pnas.95.1.114.

(295) Naka, T.; Narazaki, M.; Hirata, M.; Matsumoto, T.; Minamoto, S.; Aono, A.; Nishimoto, N.; Kajita, T.; Taga, T.; Yoshizaki, K.; Akira, S.; Kishimoto, T. Structure and Function of a New STAT-Induced STAT Inhibitor. *Nature* **1997**, *387* (6636), 924–929. https://doi.org/10.1038/43219.

(296) Starr, R.; Willson, T. A.; Viney, E. M.; Murray, L. J.; Rayner, J. R.; Jenkins, B. J.; Gonda, T. J.; Alexander, W. S.; Metcalf, D.; Nicola, N. A.; Hilton, D. J. A Family of Cytokine-Inducible Inhibitors of Signalling. *Nature* **1997**, *387* (6636), 917–921. https://doi.org/10.1038/43206.

(297) Liau, N. P. D.; Babon, J. J. Expression and Purification of JAK1 and SOCS1 for Structural and Biochemical Studies. *Methods Mol Biol* **2018**, *1725*, 267–280. https://doi.org/10.1007/978-1-4939-7568-6_21.

(298) Liang, Y.-B.; Tang, H.; Chen, Z.-B.; Zeng, L.-J.; Wu, J.-G.; Yang, W.; Li, Z.-Y.; Ma, Z.-F. Downregulated SOCS1 Expression Activates the JAK1/STAT1 Pathway and Promotes Polarization of Macrophages into M1 Type. *Mol Med Rep* **2017**, *16* (5), 6405–6411. https://doi.org/10.3892/mmr.2017.7384.

(299) Stec, W. J.; Zeidler, M. P. Drosophila SOCS Proteins. *Journal of Signal Transduction* **2011**, *2011*, 1–8. https://doi.org/10.1155/2011/894510.

(300) Linossi, E. M.; Babon, J. J.; Hilton, D. J.; Nicholson, S. E. Suppression of Cytokine Signaling: The SOCS Perspective. *Cytokine Growth Factor Rev* **2013**, *24* (3), 241–248. https://doi.org/10.1016/j.cytogfr.2013.03.005.

(301) Sasaki, A.; Yasukawa, H.; Suzuki, A.; Kamizono, S.; Syoda, T.; Kinjyo, I.; Sasaki, M.; Johnston, J. A.; Yoshimura, A. Cytokine-Inducible SH2 Protein-3 (CIS3/SOCS3) Inhibits Janus Tyrosine Kinase by Binding through the N-Terminal Kinase Inhibitory Region as Well as SH2 Domain. *Genes Cells* **1999**, *4* (6), 339–351. https://doi.org/10.1046/j.1365-2443.1999.00263.x.

(302) Babon, J. J.; Kershaw, N. J.; Murphy, J. M.; Varghese, L. N.; Laktyushin, A.; Young, S. N.; Lucet, I. S.; Norton, R. S.; Nicola, N. A. Suppression of Cytokine Signaling by SOCS3: Characterization of the Mode of Inhibition and the Basis of Its Specificity. *Immunity* **2012**, *36* (2), 239–250. https://doi.org/10.1016/j.immuni.2011.12.015.

(303) Kershaw, N. J.; Murphy, J. M.; Liau, N. P. D.; Varghese, L. N.; Laktyushin, A.; Whitlock, E. L.; Lucet, I. S.; Nicola, N. A.; Babon, J. J. SOCS3 Binds Specific Receptor-JAK Complexes to Control Cytokine Signaling by Direct Kinase Inhibition. *Nat Struct Mol Biol* **2013**, *20* (4), 469–476. https://doi.org/10.1038/nsmb.2519.

(304) Starr, R.; Metcalf, D.; Elefanty, A. G.; Brysha, M.; Willson, T. A.; Nicola, N. A.; Hilton, D. J.; Alexander, W. S. Liver Degeneration and Lymphoid Deficiencies in Mice Lacking Suppressor of Cytokine Signaling-1. *Proc Natl Acad Sci U S A* **1998**, *95* (24), 14395–14399. https://doi.org/10.1073/pnas.95.24.14395.

(305) Vogelstein, B.; Papadopoulos, N.; Velculescu, V. E.; Zhou, S.; Diaz, L. A.; Kinzler, K. W. Cancer Genome Landscapes. *Science* **2013**, *339* (6127), 1546–1558. https://doi.org/10.1126/science.1235122.

(306) Yoshikawa, H.; Matsubara, K.; Qian, G. S.; Jackson, P.; Groopman, J. D.; Manning, J. E.; Harris, C. C.; Herman, J. G. SOCS-1, a Negative Regulator of the JAK/STAT Pathway, Is Silenced by Methylation in Human Hepatocellular Carcinoma and Shows Growth-Suppression Activity. *Nat Genet* **2001**, *28* (1), 29–35. https://doi.org/10.1038/ng0501-29.

(307) Liu, S.; Ren, S.; Howell, P.; Fodstad, O.; Riker, A. I. Identification of Novel Epigenetically Modified Genes in Human Melanoma via Promoter Methylation Gene Profiling. *Pigment Cell Melanoma Res* **2008**, *21* (5), 545–558. https://doi.org/10.1111/j.1755-148X.2008.00484.x.

(308) Oshimo, Y.; Kuraoka, K.; Nakayama, H.; Kitadai, Y.; Yoshida, K.; Chayama, K.; Yasui, W. Epigenetic Inactivation of SOCS-1 by CpG Island Hypermethylation in Human Gastric Carcinoma. *Int J Cancer* **2004**, *112* (6), 1003–1009. https://doi.org/10.1002/ijc.20521.

(309) Liau, N. P. D.; Laktyushin, A.; Lucet, I. S.; Murphy, J. M.; Yao, S.; Whitlock, E.; Callaghan, K.; Nicola, N. A.; Kershaw, N. J.; Babon, J. J. The Molecular Basis of JAK/STAT Inhibition by SOCS1. *Nat Commun* **2018**, *9* (1), 1558. https://doi.org/10.1038/s41467-018-04013-1.

(310) Wei, Y.; Zhang, Z.; She, N.; Chen, X.; Zhao, Y.; Zhang, J. Atomistic Insight into the Inhibition Mechanisms of Suppressors of Cytokine Signaling on Janus Kinase. *Phys. Chem. Chem. Phys.* **2019**, *21* (24), 12905–12915. https://doi.org/10.1039/C9CP02257K.

(311) Knapp, B.; Ospina, L.; Deane, C. M. Avoiding False Positive Conclusions in Molecular Simulation: The Importance of Replicas. *J. Chem. Theory Comput.* **2018**, *14* (12), 6127–6138. https://doi.org/10.1021/acs.jctc.8b00391.

(312) Rastelli, G.; Del Rio, A.; Degliesposti, G.; Sgobba, M. Fast and Accurate Predictions of Binding Free Energies Using MM-PBSA and MM-GBSA. *J Comput Chem* **2010**, *31* (4), 797–810. https://doi.org/10.1002/jcc.21372.

(313) 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.

(314) Salomon-Ferrer, R.; Case, D. A.; Walker, R. C. An Overview of the Amber Biomolecular Simulation Package: Amber Biomolecular Simulation Package. *WIREs Comput Mol Sci* **2013**, *3* (2), 198–210. https://doi.org/10.1002/wcms.1121.

(315) Salomon-Ferrer, R.; Götz, A. W.; Poole, D.; Le Grand, S.; Walker, R. C. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald. *J. Chem. Theory Comput.* **2013**, *9* (9), 3878–3888. https://doi.org/10.1021/ct400314y.

(316) Loncharich, R. J.; Brooks, B. R.; Pastor, R. W. Langevin Dynamics of Peptides: The Frictional Dependence of Isomerization Rates of N-Acetylalanyl-N′-

Methylamide. *Biopolymers* **1992**, *32* (5), 523–535. https://doi.org/10.1002/bip.360320508.
(317) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of n-Alkanes. *Journal of Computational Physics* **1977**, *23* (3), 327–341. https://doi.org/10.1016/0021-9991(77)90098-5.
(318) Jeffrey, G. A. *An Introduction to Hydrogen Bonding*; Topics in physical chemistry; Oxford University Press: New York Oxford, 1997.
(319) Hu, G.; Ma, A.; Dou, X.; Zhao, L.; Wang, J. Computational Studies of a Mechanism for Binding and Drug Resistance in the Wild Type and Four Mutations of HIV-1 Protease with a GRL-0519 Inhibitor. *IJMS* **2016**, *17* (6), 819. https://doi.org/10.3390/ijms17060819.
(320) Shi, S.; Zhang, S.; Zhang, Q. Insight into Binding Mechanisms of Inhibitors MKP56, MKP73, MKP86, and MKP97 to HIV-1 Protease by Using Molecular Dynamics Simulation. *J Biomol Struct Dyn* **2018**, *36* (4), 981–992. https://doi.org/10.1080/07391102.2017.1305296.
(321) Sheahan, T. P.; Sims, A. C.; Leist, S. R.; Schäfer, A.; Won, J.; Brown, A. J.; Montgomery, S. A.; Hogg, A.; Babusis, D.; Clarke, M. O.; Spahn, J. E.; Bauer, L.; Sellers, S.; Porter, D.; Feng, J. Y.; Cihlar, T.; Jordan, R.; Denison, M. R.; Baric, R. S. Comparative Therapeutic Efficacy of Remdesivir and Combination Lopinavir, Ritonavir, and Interferon Beta against MERS-CoV. *Nat Commun* **2020**, *11* (1), 222. https://doi.org/10.1038/s41467-019-13940-6.
(322) Jakhmola, S.; Jonniya, N. A.; Sk, M. F.; Rani, A.; Kar, P.; Jha, H. C. Identification of Potential Inhibitors against Epstein–Barr Virus Nuclear Antigen 1 (EBNA1): An Insight from Docking and Molecular Dynamic Simulations. *ACS Chem. Neurosci.* **2021**. https://doi.org/10.1021/acschemneuro.1c00350.
(323) Goddard, T. D.; Huang, C. C.; Meng, E. C.; Pettersen, E. F.; Couch, G. S.; Morris, J. H.; Ferrin, T. E. UCSF ChimeraX: Meeting Modern Challenges in Visualization and Analysis: UCSF ChimeraX Visualization System. *Protein Science* **2018**, *27* (1), 14–25. https://doi.org/10.1002/pro.3235.
(324) Roy, R.; Jonniya, N. A.; Poddar, S.; Sk, M. F.; Kar, P. Unraveling the Molecular Mechanism of Recognition of Human Interferon-Stimulated Gene Product 15 by Coronavirus Papain-Like Proteases: A Multiscale Simulation Study. *J. Chem. Inf. Model.* **2021**, *61* (12), 6038–6052. https://doi.org/10.1021/acs.jcim.1c00918.
(325) Sk, M. F.; Jonniya, N. A.; Roy, R.; Kar, P. Unraveling the Molecular Mechanism of Recognition of Selected Next-Generation Antirheumatoid Arthritis Inhibitors by Janus Kinase 1. *ACS Omega* **2022**, *7* (7), 6195–6209. https://doi.org/10.1021/acsomega.1c06715.
(326) Tsai, C.-C.; Yue, Z.; Shen, J. How Electrostatic Coupling Enables Conformational Plasticity in a Tyrosine Kinase. *J. Am. Chem. Soc.* **2019**, *141* (38), 15092–15101. https://doi.org/10.1021/jacs.9b06064.
(327) Sun, H.; Li, Y.; Tian, S.; Wang, J.; Hou, T. P-Loop Conformation Governed Crizotinib Resistance in G2032R-Mutated ROS1 Tyrosine Kinase: Clues from Free Energy Landscape. *PLoS Comput Biol* **2014**, *10* (7), e1003729. https://doi.org/10.1371/journal.pcbi.1003729.

(328) Meng, Y.; Gao, C.; Clawson, D. K.; Atwell, S.; Russell, M.; Vieth, M.; Roux, B. Predicting the Conformational Variability of Abl Tyrosine Kinase Using Molecular Dynamics Simulations and Markov State Models. *J Chem Theory Comput* **2018**, *14* (5), 2721–2732. https://doi.org/10.1021/acs.jctc.7b01170.

(329) Guimarães, C. R. W.; Rai, B. K.; Munchhof, M. J.; Liu, S.; Wang, J.; Bhattacharya, S. K.; Buckbinder, L. Understanding the Impact of the P-Loop Conformation on Kinase Selectivity. *J. Chem. Inf. Model.* **2011**, *51* (6), 1199–1204. https://doi.org/10.1021/ci200153c.

(330) Xie, T.; Saleh, T.; Rossi, P.; Kalodimos, C. G. Conformational States Dynamically Populated by a Kinase Determine Its Function. *Science* **2020**, *370* (6513), eabc2754. https://doi.org/10.1126/science.abc2754.

(331) Collie, G. W.; Michaelides, I. N.; Embrey, K.; Stubbs, C. J.; Börjesson, U.; Dale, I. L.; Snijder, A.; Barlind, L.; Song, K.; Khurana, P.; Phillips, C.; Storer, R. I. Structural Basis for Targeting the Folded P-Loop Conformation of c-MET. *ACS Med. Chem. Lett.* **2021**, *12* (1), 162–167. https://doi.org/10.1021/acsmedchemlett.0c00392.

(332) Seeliger, M. A.; Ranjitkar, P.; Kasap, C.; Shan, Y.; Shaw, D. E.; Shah, N. P.; Kuriyan, J.; Maly, D. J. Equally Potent Inhibition of C-Src and Abl by Compounds That Recognize Inactive Kinase Conformations. *Cancer Res* **2009**, *69* (6), 2384–2392. https://doi.org/10.1158/0008-5472.CAN-08-3953.

(333) Kalaivani, R.; Narwani, T. J.; de Brevern, A. G.; Srinivasan, N. Long-Range Molecular Dynamics Show That Inactive Forms of Protein Kinase A Are More Dynamic than Active Forms. *Protein Sci* **2019**, *28* (3), 543–560. https://doi.org/10.1002/pro.3556.

(334) Xu, B.; Min, X.; Stippec, S.; Lee, B.-H.; Goldsmith, E. J.; Cobb, M. H. Regulation of WNK1 by an Autoinhibitory Domain and Autophosphorylation. *J Biol Chem* **2002**, *277* (50), 48456–48462. https://doi.org/10.1074/jbc.M207917200.

(335) Shukla, D.; Meng, Y.; Roux, B.; Pande, V. S. Activation Pathway of Src Kinase Reveals Intermediate States as Targets for Drug Design. *Nat Commun* **2014**, *5* (1), 3397. https://doi.org/10.1038/ncomms4397.

(336) Mehta, I. D.; Beck, B. W. Modeling Proteins as Residue Interaction Networks to Understand Structure-Function Relationship. *Biophysical Journal* **2014**, *106* (2), 50a. https://doi.org/10.1016/j.bpj.2013.11.358.

(337) Di Paola, L.; De Ruvo, M.; Paci, P.; Santoni, D.; Giuliani, A. Protein Contact Networks: An Emerging Paradigm in Chemistry. *Chem. Rev.* **2013**, *113* (3), 1598–1613. https://doi.org/10.1021/cr3002356.

(338) Tse, A.; Verkhivker, G. M. Molecular Dynamics Simulations and Structural Network Analysis of C-Abl and c-Src Kinase Core Proteins: Capturing Allosteric Mechanisms and Communication Pathways from Residue Centrality. *J. Chem. Inf. Model.* **2015**, *55* (8), 1645–1662. https://doi.org/10.1021/acs.jcim.5b00240.

(339) Hubbard, S. R.; Wei, L.; Ellis, L.; Hendrickson, W. A. Crystal Structure of the Tyrosine Kinase Domain of the Human Insulin Receptor. *Nature* **1994**, *372* (6508), 746–754. https://doi.org/10.1038/372746a0.

(340) Marsh, J. A. Buried and Accessible Surface Area Control Intrinsic Protein Flexibility. *J Mol Biol* **2013**, *425* (17), 3250–3263. https://doi.org/10.1016/j.jmb.2013.06.019.

(341) Harpur, A. G.; Andres, A. C.; Ziemiecki, A.; Aston, R. R.; Wilks, A. F. JAK2, a Third Member of the JAK Family of Protein Tyrosine Kinases. *Oncogene* **1992**, *7* (7), 1347–1353.
(342) Rawlings, J. S.; Rosler, K. M.; Harrison, D. A. The JAK/STAT Signaling Pathway. *J Cell Sci* **2004**, *117* (Pt 8), 1281–1283. https://doi.org/10.1242/jcs.00963.
(343) Schindler, C.; Levy, D. E.; Decker, T. JAK-STAT Signaling: From Interferons to Cytokines. *J Biol Chem* **2007**, *282* (28), 20059–20063. https://doi.org/10.1074/jbc.R700016200.
(344) Shuai, K.; Liu, B. Regulation of JAK-STAT Signalling in the Immune System. *Nat Rev Immunol* **2003**, *3* (11), 900–911. https://doi.org/10.1038/nri1226.
(345) Babon, J. J.; Lucet, I. S.; Murphy, J. M.; Nicola, N. A.; Varghese, L. N. The Molecular Regulation of Janus Kinase (JAK) Activation. *Biochem J* **2014**, *462* (1), 1–13. https://doi.org/10.1042/BJ20140712.
(346) Chatti, K.; Farrar, W. L.; Duhé, R. J. Tyrosine Phosphorylation of the Janus Kinase 2 Activation Loop Is Essential for a High-Activity Catalytic State but Dispensable for a Basal Catalytic State. *Biochemistry* **2004**, *43* (14), 4272–4283. https://doi.org/10.1021/bi036109b.
(347) Ung, P. M.-U.; Schlessinger, A. DFGmodel: Predicting Protein Kinase Structures in Inactive States for Structure-Based Discovery of Type-II Inhibitors. *ACS Chem Biol* **2015**, *10* (1), 269–278. https://doi.org/10.1021/cb500696t.
(348) McInnes, C.; Mezna, M.; Kontopidis, G. Catch the Kinase Conformer. *Chem Biol* **2006**, *13* (7), 693–694. https://doi.org/10.1016/j.chembiol.2006.07.001.
(349) Kong, X.; Sun, H.; Pan, P.; Li, D.; Zhu, F.; Chang, S.; Xu, L.; Li, Y.; Hou, T. How Does the L884P Mutation Confer Resistance to Type-II Inhibitors of JAK2 Kinase: A Comprehensive Molecular Modeling Study. *Sci Rep* **2017**, *7* (1), 9088. https://doi.org/10.1038/s41598-017-09586-3.
(350) Li, J. J.; Tu, J.; Cheng, P.; Zhai, H. L.; Zhang, X. Y. Insights into DFG-in and DFG-out JAK2 Binding Modes for a Rational Strategy of Type II Inhibitors Combined Computational Study. *RSC Adv.* **2016**, *6* (51), 45540–45552. https://doi.org/10.1039/C6RA06266K.
(351) Zhao, Z.; Wu, H.; Wang, L.; Liu, Y.; Knapp, S.; Liu, Q.; Gray, N. S. Exploration of Type II Binding Mode: A Privileged Approach for Kinase Inhibitor Focused Drug Discovery? *ACS Chem Biol* **2014**, *9* (6), 1230–1241. https://doi.org/10.1021/cb500129t.
(352) Liu, Y.; Gray, N. S. Rational Design of Inhibitors That Bind to Inactive Kinase Conformations. *Nat Chem Biol* **2006**, *2* (7), 358–364. https://doi.org/10.1038/nchembio799.
(353) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res* **2000**, *28* (1), 235–242. https://doi.org/10.1093/nar/28.1.235.
(354) Andraos, R.; Qian, Z.; Bonenfant, D.; Rubert, J.; Vangrevelinghe, E.; Scheufler, C.; Marque, F.; Régnier, C. H.; De Pover, A.; Ryckelynck, H.; Bhagwat, N.; Koppikar, P.; Goel, A.; Wyder, L.; Tavares, G.; Baffert, F.; Pissot-Soldermann, C.; Manley, P. W.; Gaul, C.; Voshol, H.; Levine, R. L.; Sellers, W. R.; Hofmann, F.; Radimerski, T. Modulation of Activation-Loop Phosphorylation by JAK Inhibitors

Is Binding Mode Dependent. *Cancer Discov* **2012**, *2* (6), 512–523. https://doi.org/10.1158/2159-8290.CD-11-0324.

(355) Miao, Y.; McCammon, J. A. Gaussian Accelerated Molecular Dynamics: Theory, Implementation, and Applications. *Annu Rep Comput Chem* **2017**, *13*, 231–278. https://doi.org/10.1016/bs.arcc.2017.06.005.

(356) Miller, B. R.; McGee, T. D.; Swails, J. M.; Homeyer, N.; Gohlke, H.; Roitberg, A. E. MMPBSA.Py: An Efficient Program for End-State Free Energy Calculations. *J. Chem. Theory Comput.* **2012**, *8* (9), 3314–3321. https://doi.org/10.1021/ct300418h.

(357) Maestro. Schrödinger Release 2022-1: *Schrödinger* **2021**, *LLC, New York* (NY).

(358) Junmei Wang, Wei Wang, Peter A. Kollman and David A. Case. Antechamber: An Accessory Software Package for Molecular Mechanical Calculations. **2001**, *222*, U403.

(359) Amarnath Jonniya, N.; Fulbabu Sk, M.; Kar, P. Characterizing an Allosteric Inhibitor-Induced Inactive State in with-No-Lysine Kinase 1 Using Gaussian Accelerated Molecular Dynamics Simulations. *Physical Chemistry Chemical Physics* **2021**, *23* (12), 7343–7358. https://doi.org/10.1039/D0CP05733A.

(360) Roy, R.; Mishra, A.; Poddar, S.; Nayak, D.; Kar, P. Investigating the Mechanism of Recognition and Structural Dynamics of Nucleoprotein-RNA Complex from *Peste Des Petits Ruminants Virus* via Gaussian Accelerated Molecular Dynamics Simulations. *Journal of Biomolecular Structure and Dynamics* **2020**, 1–14. https://doi.org/10.1080/07391102.2020.1838327.

(361) Bhattarai, A.; Miao, Y. Gaussian Accelerated Molecular Dynamics for Elucidation of Drug Pathways. *Expert Opin Drug Discov* **2018**, *13* (11), 1055–1065. https://doi.org/10.1080/17460441.2018.1538207.

(362) Miao, Y.; Caliman, A. D.; McCammon, J. A. Allosteric Effects of Sodium Ion Binding on Activation of the M3 Muscarinic G-Protein-Coupled Receptor. *Biophys J* **2015**, *108* (7), 1796–1806. https://doi.org/10.1016/j.bpj.2015.03.003.

(363) Zhang, J.; Wang, N.; Miao, Y.; Hauser, F.; McCammon, J. A.; Rappel, W.-J.; Schroeder, J. I. Identification of SLAC1 Anion Channel Residues Required for CO2/Bicarbonate Sensing and Regulation of Stomatal Movements. *Proceedings of the National Academy of Sciences* **2018**, *115* (44), 11129–11137. https://doi.org/10.1073/pnas.1807624115.

(364) Liao, J.-M.; Wang, Y.-T. In Silico Studies of Conformational Dynamics of Mu Opioid Receptor Performed Using Gaussian Accelerated Molecular Dynamics. *Journal of Biomolecular Structure and Dynamics* **2019**, *37* (1), 166–177. https://doi.org/10.1080/07391102.2017.1422025.

(365) Wang, Y.-T.; Chan, Y.-H. Understanding the Molecular Basis of Agonist/Antagonist Mechanism of Human Mu Opioid Receptor through Gaussian Accelerated Molecular Dynamics Method. *Sci Rep* **2017**, *7* (1), 7828. https://doi.org/10.1038/s41598-017-08224-2.

(366) Park, J. B.; Kim, Y. H.; Yoo, Y.; Kim, J.; Jun, S.-H.; Cho, J. W.; El Qaidi, S.; Walpole, S.; Monaco, S.; García-García, A. A.; Wu, M.; Hays, M. P.; Hurtado-Guerrero, R.; Angulo, J.; Hardwidge, P. R.; Shin, J.-S.; Cho, H.-S. Structural Basis for Arginine Glycosylation of Host Substrates by Bacterial Effector Proteins. *Nat Commun* **2018**, *9* (1), 4283. https://doi.org/10.1038/s41467-018-06680-6.

(367) East, K. W.; Newton, J. C.; Morzan, U. N.; Narkhede, Y. B.; Acharya, A.; Skeens, E.; Jogl, G.; Batista, V. S.; Palermo, G.; Lisi, G. P. Allosteric Motions of the CRISPR-Cas9 HNH Nuclease Probed by NMR and Molecular Dynamics. *J Am Chem Soc* **2020**, *142* (3), 1348–1358. https://doi.org/10.1021/jacs.9b10521.

(368) Ricci, C. G.; Chen, J. S.; Miao, Y.; Jinek, M.; Doudna, J. A.; McCammon, J. A.; Palermo, G. Deciphering Off-Target Effects in CRISPR-Cas9 through Accelerated Molecular Dynamics. *ACS Cent Sci* **2019**, *5* (4), 651–662. https://doi.org/10.1021/acscentsci.9b00020.

(369) Roe, D. R.; Cheatham, T. E. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J. Chem. Theory Comput.* **2013**, *9* (7), 3084–3095. https://doi.org/10.1021/ct400341p.

(370) Frauenfelder, H.; Sligar, S. G.; Wolynes, P. G. The Energy Landscapes and Motions of Proteins. *Science* **1991**, *254* (5038), 1598–1603. https://doi.org/10.1126/science.1749933.

(371) Tsai, C.-J.; Ma, B.; Nussinov, R. Folding and Binding Cascades: Shifts in Energy Landscapes. *Proceedings of the National Academy of Sciences* **1999**, *96* (18), 9970–9972. https://doi.org/10.1073/pnas.96.18.9970.

(372) Indari, O.; Sk, M. F.; Jakhmola, S.; Jonniya, N. A.; Jha, H. C.; Kar, P. Decoding the Host–Parasite Protein Interactions Involved in Cerebral Malaria Through Glares of Molecular Dynamics Simulations. *J. Phys. Chem. B* **2022**, *126* (2), 387–402. https://doi.org/10.1021/acs.jpcb.1c07850.

(373) Sk, M. F.; Jonniya, N. A.; Roy, R.; Kar, P. Phosphorylation-Induced Conformational Dynamics and Inhibition of Janus Kinase 1 by Suppressors of Cytokine Signaling 1. *J. Phys. Chem. B* **2022**, *126* (17), 3224–3239. https://doi.org/10.1021/acs.jpcb.1c10733.

(374) Sk, M. F.; Jonniya, N. A.; Roy, R.; Poddar, S.; Kar, P. Computational Investigation of Structural Dynamics of SARS-CoV-2 Methyltransferase-Stimulatory Factor Heterodimer Nsp16/Nsp10 Bound to the Cofactor SAM. *Front. Mol. Biosci.* **2020**, *0*. https://doi.org/10.3389/fmolb.2020.590165.

(375) Gohlke, H.; Kiel, C.; Case, D. A. Insights into Protein–Protein Binding by Binding Free Energy Calculation and Free Energy Decomposition for the Ras–Raf and Ras–RalGDS Complexes. *Journal of Molecular Biology* **2003**, *330* (4), 891–913. https://doi.org/10.1016/S0022-2836(03)00610-7.

(376) Chen, S.-F.; Cao, Y.; Han, S.; Chen, J.-Z. Insight into the Structural Mechanism for PKBα Allosteric Inhibition by Molecular Dynamics Simulations and Free Energy Calculations. *J Mol Graph Model* **2014**, *48*, 36–46. https://doi.org/10.1016/j.jmgm.2013.12.002.

(377) La Sala, G.; Riccardi, L.; Gaspari, R.; Cavalli, A.; Hantschel, O.; De Vivo, M. HRD Motif as the Central Hub of the Signaling Network for Activation Loop Autophosphorylation in Abl Kinase. *J. Chem. Theory Comput.* **2016**, *12* (11), 5563–5574. https://doi.org/10.1021/acs.jctc.6b00600.

(378) Liu, C.; Li, Z.; Liu, Z.; Yang, S.; Wang, Q.; Chai, Z. Understanding the P-Loop Conformation in the Determination of Inhibitor Selectivity Toward the Hepatocellular Carcinoma-Associated Dark Kinase STK17B. *Frontiers in Molecular Biosciences* **2022**, *9*.

(379) Gohda, K.; Hakoshima, T. A Molecular Mechanism of P-Loop Pliability of Rho-Kinase Investigated by Molecular Dynamic Simulation. *J Comput Aided Mol Des* **2008**, *22* (11), 789–797. https://doi.org/10.1007/s10822-008-9214-7.

(380) Liang, S.; Wang, Q.; Qi, X.; Liu, Y.; Li, G.; Lu, S.; Mou, L.; Chen, X. Deciphering the Mechanism of Gilteritinib Overcoming Lorlatinib Resistance to the Double Mutant I1171N/F1174I in Anaplastic Lymphoma Kinase. *Front. Cell Dev. Biol.* **2021**, *9*, 808864. https://doi.org/10.3389/fcell.2021.808864.

(381) Zhang, H.; Zhu, M.; Li, M.; Ni, D.; Wang, Y.; Deng, L.; Du, K.; Lu, S.; Shi, H.; Cai, C. Mechanistic Insights Into Co-Administration of Allosteric and Orthosteric Drugs to Overcome Drug-Resistance in T315I BCR-ABL1. *Front. Pharmacol.* **2022**, *13*, 862504. https://doi.org/10.3389/fphar.2022.862504.

(382) Qiu, Y.; Yin, X.; Li, X.; Wang, Y.; Fu, Q.; Huang, R.; Lu, S. Untangling Dual-Targeting Therapeutic Mechanism of Epidermal Growth Factor Receptor (EGFR) Based on Reversed Allosteric Communication. *Pharmaceutics* **2021**, *13* (5), 747. https://doi.org/10.3390/pharmaceutics13050747.

(383) Vijayan, R. S. K.; He, Peng, P.; Modi, V.; Duong-Ly, K. C.; Ma, H.; Peterson, J. R.; Dunbrack, R. L.; Levy, R. M. Conformational Analysis of the DFG-Out Kinase Motif and Biochemical Profiling of Structurally Validated Type II Inhibitors. *J. Med. Chem.* **2015**, *58* (1), 466–479. https://doi.org/10.1021/jm501603h.

(384) Oruganty, K.; Talathi, N. S.; Wood, Z. A.; Kannan, N. Identification of a Hidden Strain Switch Provides Clues to an Ancient Structural Mechanism in Protein Kinases. *Proceedings of the National Academy of Sciences* **2013**, *110* (3), 924–929. https://doi.org/10.1073/pnas.1207104110.

(385) Modi, V.; Dunbrack, R. L. Defining a New Nomenclature for the Structures of Active and Inactive Kinases. *Proceedings of the National Academy of Sciences* **2019**, *116* (14), 6818–6827. https://doi.org/10.1073/pnas.1814279116.

(386) Möbitz, H. The ABC of Protein Kinase Conformations. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* **2015**, *1854* (10, Part B), 1555–1566. https://doi.org/10.1016/j.bbapap.2015.03.009.

(387) Sultan, M. M.; Kiss, G.; Pande, V. S. Towards Simple Kinetic Models of Functional Dynamics for a Kinase Subfamily. *Nature Chem* **2018**, *10* (9), 903–909. https://doi.org/10.1038/s41557-018-0077-9.

(388) Lucet, I. S.; Fantino, E.; Styles, M.; Bamert, R.; Patel, O.; Broughton, S. E.; Walter, M.; Burns, C. J.; Treutlein, H.; Wilks, A. F.; Rossjohn, J. The Structural Basis of Janus Kinase 2 Inhibition by a Potent and Specific Pan-Janus Kinase Inhibitor. *Blood* **2006**, *107* (1), 176–183. https://doi.org/10.1182/blood-2005-06-2413.

(389) Schindler, C.; Levy, D. E.; Decker, T. JAK-STAT Signaling: From Interferons to Cytokines*. *Journal of Biological Chemistry* **2007**, *282* (28), 20059–20063. https://doi.org/10.1074/jbc.R700016200.

(390) Sanachai, K.; Mahalapbutr, P.; Choowongkomon, K.; Poo-arporn, R. P.; Wolschann, P.; Rungrotmongkol, T. Insights into the Binding Recognition and Susceptibility of Tofacitinib toward Janus Kinases. *ACS Omega* **2020**, *5* (1), 369–377. https://doi.org/10.1021/acsomega.9b02800.

(391) Levy, D. E.; Darnell, J. E. STATs: Transcriptional Control and Biological Impact. *Nat Rev Mol Cell Biol* **2002**, *3* (9), 651–662. https://doi.org/10.1038/nrm909.

(392) Song, L.; Schindler, C. Chapter 249 - JAK-STAT Signaling. In *Handbook of Cell Signaling (Second Edition)*; Bradshaw, R. A., Dennis, E. A., Eds.; Academic Press: San Diego, 2010; pp 2041–2048. https://doi.org/10.1016/B978-0-12-374145-5.00249-7.

(393) Tu, Y.; Zhong, Y.; Fu, J.; Cao, Y.; Fu, G.; Tian, X.; Wang, B. Activation of JAK/STAT Signal Pathway Predicts Poor Prognosis of Patients with Gliomas. *Med Oncol* **2011**, *28* (1), 15–23. https://doi.org/10.1007/s12032-010-9435-1.

(394) Lashgari, N.-A.; Roudsari, N. M.; Momtaz, S.; Sathyapalan, T.; Abdolghaffari, A. H.; Sahebkar, A. The Involvement of JAK/STAT Signaling Pathway in the Treatment of Parkinson's Disease. *Journal of Neuroimmunology* **2021**, *361*, 577758. https://doi.org/10.1016/j.jneuroim.2021.577758.

(395) Lesgidou, N.; Eliopoulos, E.; Goulielmos, G. N.; Vlassi, M. Insights on the Alteration of Functionality of a Tyrosine Kinase 2 Variant: A Molecular Dynamics Study. *Bioinformatics* **2018**, *34* (17), i781–i786. https://doi.org/10.1093/bioinformatics/bty556.

(396) Sato, T.; Toki, T.; Kanezaki, R.; Xu, G.; Terui, K.; Kanegane, H.; Miura, M.; Adachi, S.; Migita, M.; Morinaga, S.; Nakano, T.; Endo, M.; Kojima, S.; Kiyoi, H.; Mano, H.; Ito, E. Functional Analysis of JAK3 Mutations in Transient Myeloproliferative Disorder and Acute Megakaryoblastic Leukaemia Accompanying Down Syndrome. *Br J Haematol* **2008**, *141* (5), 681–688. https://doi.org/10.1111/j.1365-2141.2008.07081.x.

(397) Ghoreschi, K.; Laurence, A.; O'Shea, J. J. Janus Kinases in Immune Cell Signaling. *Immunological Reviews* **2009**, *228* (1), 273–287. https://doi.org/10.1111/j.1600-065X.2008.00754.x.

(398) Mullighan, C. G.; Zhang, J.; Harvey, R. C.; Collins-Underwood, J. R.; Schulman, B. A.; Phillips, L. A.; Tasian, S. K.; Loh, M. L.; Su, X.; Liu, W.; Devidas, M.; Atlas, S. R.; Chen, I.-M.; Clifford, R. J.; Gerhard, D. S.; Carroll, W. L.; Reaman, G. H.; Smith, M.; Downing, J. R.; Hunger, S. P.; Willman, C. L. JAK Mutations in High-Risk Childhood Acute Lymphoblastic Leukemia. *Proc Natl Acad Sci U S A* **2009**, *106* (23), 9414–9418. https://doi.org/10.1073/pnas.0811761106.

(399) T. Virtanen, A.; Haikarainen, T.; Raivola, J.; Silvennoinen, O. Selective JAKinibs: Prospects in Inflammatory and Autoimmune Diseases. *BioDrugs* **2019**, *33* (1), 15–32. https://doi.org/10.1007/s40259-019-00333-w.

(400) Harrison, C.; Kiladjian, J.-J.; Al-Ali, H. K.; Gisslinger, H.; Waltzman, R.; Stalbovskaya, V.; McQuitty, M.; Hunter, D. S.; Levy, R.; Knoops, L.; Cervantes, F.; Vannucchi, A. M.; Barbui, T.; Barosi, G. JAK Inhibition with Ruxolitinib versus Best Available Therapy for Myelofibrosis. *N Engl J Med* **2012**, *366* (9), 787–798. https://doi.org/10.1056/NEJMoa1110556.

(401) Ghoreschi, K.; Jesson, M. I.; Li, X.; Lee, J. L.; Ghosh, S.; Alsup, J. W.; Warner, J. D.; Tanaka, M.; Steward-Tharp, S. M.; Gadina, M.; Thomas, C. J.; Minnerly, J. C.; Storer, C. E.; LaBranche, T. P.; Radi, Z. A.; Dowty, M. E.; Head, R. D.; Meyer, D. M.; Kishore, N.; O'Shea, J. J. Modulation of Innate and Adaptive Immune Responses by Tofacitinib (CP-690,550). *J Immunol* **2011**, *186* (7), 4234–4243. https://doi.org/10.4049/jimmunol.1003668.

(402) Mogul, A.; Corsi, K.; McAuliffe, L. Baricitinib: The Second FDA-Approved JAK Inhibitor for the Treatment of Rheumatoid Arthritis. *Ann Pharmacother* **2019**, *53* (9), 947–953. https://doi.org/10.1177/1060028019839650.

(403) Parmentier, J. M.; Voss, J.; Graff, C.; Schwartz, A.; Argiriadi, M.; Friedman, M.; Camp, H. S.; Padley, R. J.; George, J. S.; Hyland, D.; Rosebraugh, M.; Wishart, N.; Olson, L.; Long, A. J. In Vitro and in Vivo Characterization of the JAK1 Selectivity of Upadacitinib (ABT-494). *BMC Rheumatology* **2018**, *2* (1), 23. https://doi.org/10.1186/s41927-018-0031-x.

(404) Wu, P.; Nielsen, T. E.; Clausen, M. H. FDA-Approved Small-Molecule Kinase Inhibitors. *Trends in Pharmacological Sciences* **2015**, *36* (7), 422–439. https://doi.org/10.1016/j.tips.2015.04.005.

(405) Harder, E.; Damm, W.; Maple, J.; Wu, C.; Reboul, M.; Xiang, J. Y.; Wang, L.; Lupyan, D.; Dahlgren, M. K.; Knight, J. L.; Kaus, J. W.; Cerutti, D. S.; Krilov, G.; Jorgensen, W. L.; Abel, R.; Friesner, R. A. OPLS3: A Force Field Providing Broad Coverage of Drug-like Small Molecules and Proteins. *J. Chem. Theory Comput.* **2016**, *12* (1), 281–296. https://doi.org/10.1021/acs.jctc.5b00864.

(406) Halgren, T. A.; Murphy, R. B.; Friesner, R. A.; Beard, H. S.; Frye, L. L.; Pollard, W. T.; Banks, J. L. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 2. Enrichment Factors in Database Screening. *J. Med. Chem.* **2004**, *47* (7), 1750–1759. https://doi.org/10.1021/jm030644s.

(407) Friesner, R. A.; Murphy, R. B.; Repasky, M. P.; Frye, L. L.; Greenwood, J. R.; Halgren, T. A.; Sanschagrin, P. C.; Mainz, D. T. Extra Precision Glide: Docking and Scoring Incorporating a Model of Hydrophobic Enclosure for Protein−Ligand Complexes. *J. Med. Chem.* **2006**, *49* (21), 6177–6196. https://doi.org/10.1021/jm051256o.

(408) Daina, A.; Michielin, O.; Zoete, V. SwissADME: A Free Web Tool to Evaluate Pharmacokinetics, Drug-Likeness and Medicinal Chemistry Friendliness of Small Molecules. *Sci Rep* **2017**, *7* (1), 42717. https://doi.org/10.1038/srep42717.

(409) Pires, D. E. V.; Blundell, T. L.; Ascher, D. B. PkCSM: Predicting Small-Molecule Pharmacokinetic and Toxicity Properties Using Graph-Based Signatures. *J. Med. Chem.* **2015**, *58* (9), 4066–4072. https://doi.org/10.1021/acs.jmedchem.5b00104.

(410) Banerjee, P.; Eckert, A. O.; Schrey, A. K.; Preissner, R. ProTox-II: A Webserver for the Prediction of Toxicity of Chemicals. *Nucleic Acids Research* **2018**, *46* (W1), W257–W263. https://doi.org/10.1093/nar/gky318.

(411) Xiong, G.; Wu, Z.; Yi, J.; Fu, L.; Yang, Z.; Hsieh, C.; Yin, M.; Zeng, X.; Wu, C.; Lu, A.; Chen, X.; Hou, T.; Cao, D. ADMETlab 2.0: An Integrated Online Platform for Accurate and Comprehensive Predictions of ADMET Properties. *Nucleic Acids Research* **2021**, *49* (W1), W5–W14. https://doi.org/10.1093/nar/gkab255.

(412) Becke, A. D. Density-Functional Exchange-Energy Approximation with Correct Asymptotic Behavior. *Phys. Rev. A* **1988**, *38* (6), 3098–3100. https://doi.org/10.1103/PhysRevA.38.3098.

(413) Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti Correlation-Energy Formula into a Functional of the Electron Density. *Phys. Rev. B* **1988**, *37* (2), 785–789. https://doi.org/10.1103/PhysRevB.37.785.

(414) Price, D. J.; Brooks, C. L. A Modified TIP3P Water Potential for Simulation with Ewald Summation. *The Journal of Chemical Physics* **2004**, *121* (20), 10096–10103. https://doi.org/10.1063/1.1808117.
(415) Jayaram, B.; Liu, Y.; Beveridge, D. L. A Modification of the Generalized Born Theory for Improved Estimates of Solvation Energies and PK Shifts. *J. Chem. Phys.* **1998**, *109* (4), 1465–1471. https://doi.org/10.1063/1.476697.
(416) Amarnath Jonniya, N.; Fulbabu Sk, M.; Kar, P. Characterizing an Allosteric Inhibitor-Induced Inactive State in with-No-Lysine Kinase 1 Using Gaussian Accelerated Molecular Dynamics Simulations. *Physical Chemistry Chemical Physics* **2021**, *23* (12), 7343–7358. https://doi.org/10.1039/D0CP05733A.
(417) Jonniya, N. A.; Sk, M. F.; Kar, P. Elucidating Specificity of an Allosteric Inhibitor WNK476 among With-No-Lysine Kinase Isoforms Using Molecular Dynamic Simulations. *Chemical Biology & Drug Design* **2021**, *98* (3), 405–420. https://doi.org/10.1111/cbdd.13863.
(418) Masand, V. H.; Sk, M. F.; Kar, P.; Rastija, V.; Zaki, M. E. A. Identification of Food Compounds as Inhibitors of SARS-CoV-2 Main Protease Using Molecular Docking and Molecular Dynamics Simulations. *Chemometrics and Intelligent Laboratory Systems* **2021**, *217*, 104394. https://doi.org/10.1016/j.chemolab.2021.104394.
(419) Sk, M. F.; Haridev, S.; Roy, R.; Kar, P. Investigating Potency of TMC-126 against Wild-Type and Mutant Variants of HIV-1 Protease: A Molecular Dynamics and Free Energy Study. *SAR and QSAR in Environmental Research* **2021**, *32* (11), 941–962. https://doi.org/10.1080/1062936X.2021.1999318.
(420) Srimai, V.; Ramesh, M.; Satya Parameshwar, K.; Parthasarathy, T. Computer-Aided Design of Selective Cytochrome P450 Inhibitors and Docking Studies of Alkyl Resorcinol Derivatives. *Med Chem Res* **2013**, *22* (11), 5314–5323. https://doi.org/10.1007/s00044-013-0532-5.
(421) Wang, Z.; Hop, C. E.; Leung, K. H.; Pang, J. Determination of in Vitro Permeability of Drug Candidates through a Caco-2 Cell Monolayer by Liquid Chromatography/Tandem Mass Spectrometry. *J Mass Spectrom* **2000**, *35* (1), 71–76. https://doi.org/10.1002/(SICI)1096-9888(200001)35:1<71::AID-JMS915>3.0.CO;2-5.
(422) Cao, Y.; Jiang, T.; Girke, T. A Maximum Common Substructure-Based Algorithm for Searching and Predicting Drug-like Compounds. *Bioinformatics* **2008**, *24* (13), i366–i374. https://doi.org/10.1093/bioinformatics/btn186.
(423) Backman, T. W. H.; Cao, Y.; Girke, T. ChemMine Tools: An Online Service for Analyzing and Clustering Small Molecules. *Nucleic Acids Research* **2011**, *39* (suppl), W486–W491. https://doi.org/10.1093/nar/gkr320.
(424) Cao, Y.; Charisi, A.; Cheng, L.-C.; Jiang, T.; Girke, T. ChemmineR: A Compound Mining Framework for R. *Bioinformatics* **2008**, *24* (15), 1733–1734. https://doi.org/10.1093/bioinformatics/btn307.
(425) Tan, S.; Gong, X.; Liu, H.; Yao, X. Virtual Screening and Biological Activity Evaluation of New Potent Inhibitors Targeting LRRK2 Kinase Domain. *ACS Chem. Neurosci.* **2021**, *12* (17), 3214–3224. https://doi.org/10.1021/acschemneuro.1c00399.

(426) Grant, B. J.; Rodrigues, A. P. C.; ElSawy, K. M.; McCammon, J. A.; Caves, L. S. D. Bio3d: An R Package for the Comparative Analysis of Protein Structures. *Bioinformatics* **2006**, *22* (21), 2695–2696. https://doi.org/10.1093/bioinformatics/btl461.

(427) Chan, M. C.; Procko, E.; Shukla, D. Structural Rearrangement of the Serotonin Transporter Intracellular Gate Induced by Thr276 Phosphorylation. *ACS Chem. Neurosci.* **2022**. https://doi.org/10.1021/acschemneuro.1c00714.

(428) Jura, N.; Zhang, X.; Endres, N. F.; Seeliger, M. A.; Schindler, T.; Kuriyan, J. Catalytic Control in the EGF Receptor and Its Connection to General Kinase Regulatory Mechanisms. *Molecular Cell* **2011**, *42* (1), 9–22. https://doi.org/10.1016/j.molcel.2011.03.004.

(429) Shan, Y.; Arkhipov, A.; Kim, E. T.; Pan, A. C.; Shaw, D. E. Transitions to Catalytically Inactive Conformations in EGFR Kinase. *Proc. Natl. Acad. Sci. U.S.A.* **2013**, *110* (18), 7270–7275. https://doi.org/10.1073/pnas.1220843110.

(430) Moffett, A. S.; Bender, K. W.; Huber, S. C.; Shukla, D. Molecular Dynamics Simulations Reveal the Conformational Dynamics of Arabidopsis Thaliana BRI1 and BAK1 Receptor-like Kinases. *Journal of Biological Chemistry* **2017**, *292* (30), 12643–12652. https://doi.org/10.1074/jbc.M117.792762.

(431) Jonniya, N. A.; Sk, M. F.; Roy, R.; Kar, P. Discovery of Potential Competitive Inhibitors against With-No-Lysine Kinase 1 for Treating Hypertension by Virtual Screening, Inverse Pharmacophore-Based Lead Optimization, and Molecular Dynamics Simulations. *SAR and QSAR in Environmental Research* **2022**, *33* (2), 1–25. https://doi.org/10.1080/1062936X.2021.2023218.

(432) Wang, J. L.; Cheng, L. P.; Wang, T. C.; Deng, W.; Wu, F. H. Molecular Modeling Study of CP-690550 Derivatives as JAK3 Kinase Inhibitors through Combined 3D-QSAR, Molecular Docking, and Dynamics Simulation Techniques. *Journal of Molecular Graphics and Modelling* **2017**, *72*, 178–186. https://doi.org/10.1016/j.jmgm.2016.12.020.

(433) Itteboina, R.; Ballu, S.; Sivan, S. K.; Manga, V. Molecular Docking, 3D QSAR and Dynamics Simulation Studies of Imidazo-Pyrrolopyridines as Janus Kinase 1 (JAK 1) Inhibitors. *Computational Biology and Chemistry* **2016**, *64*, 33–46. https://doi.org/10.1016/j.compbiolchem.2016.04.009.