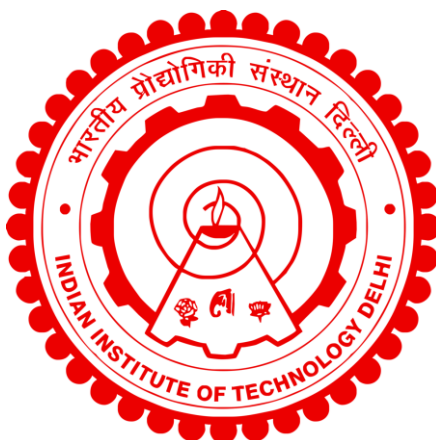


COMPUTATIONAL CHARACTERIZATION OF COOPERATIVE DYNAMICS AND LARGE-SCALE CONFORMATIONAL TRANSITIONS IN BIOMOLECULES AND THEIR ASSEMBLIES

RAJAT PUNIA



**DEPARTMENT OF CHEMICAL ENGINEERING
INDIAN INSTITUTE OF TECHNOLOGY DELHI
APRIL 2025**

© Indian Institute of Technology Delhi (IITD), New Delhi, 2025

Computational Characterization of Cooperative Dynamics and Large-Scale Conformational Transitions in Biomolecules and their Assemblies

by

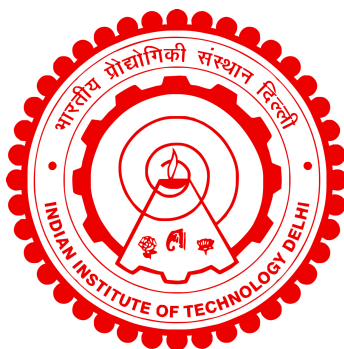
Rajat Punia

Department of Chemical Engineering

submitted in fulfillment of the requirements of the degree of

Doctor of Philosophy

to the



Indian Institute of Technology Delhi

April 29, 2025

*Dedicated to my wife Shreya for her unwavering support
throughout this academic journey*

Certificate

This is to certify that the thesis entitled “**Computational Characterization of Cooperative Dynamics and Large-Scale Conformational Transitions in Biomolecules and their Assemblies**”, submitted by **Rajat Punia** to the Indian Institute of Technology Delhi, for the award of the degree of **Doctor of Philosophy**, is a record of the original, bona fide research work carried out by him under my supervision and guidance. The thesis has reached the standards fulfilling the requirements of the regulations related to the award of the degree.

The results contained in this thesis have not been submitted in part or in full to any other University or Institute for the award of any degree or diploma to the best of our knowledge.

Date:

Prof. Gaurav Goel

Department of Chemical Engineering,
Indian Institute of Technology Delhi
New Delhi, India

Acknowledgements

First and foremost, I would like to express my heartfelt gratitude to my supervisor, Prof. G. Goel, for granting me the opportunity to join his research group. His profound intelligence, meticulous attention to detail, and unwavering pursuit of excellence have been both inspiring and instrumental throughout my research journey. Prof. Goel's exceptional communication skills fostered an environment of open dialogue, allowing for enriching technical discussions, constructive critiques, and insightful perspectives during our weekly meetings. His continuous support and patience provided me with the freedom to explore and develop on multiple levels. I am particularly grateful for his extensive support, not only pertaining to this research project but also concerning my future endeavors. Without his professional guidance, deep knowledge, and invaluable suggestions, this thesis would not have reached its current form.

I extend my deep gratitude to my research committee members, Prof. M. Banerjee, Prof. A. S. Rathore, Prof. S. Gupta, and Prof. S. Pattnaik, for their valuable suggestions, time, and thoughtful corrections throughout the duration of my Ph.D. I would also like to thank the faculty members and office staff of the Department for their cordiality and assistance whenever needed. Special thanks to Krishna Kumar for his help with lab organization. I am truly grateful to the HPC facility at IIT Delhi for providing the computational resources essential to my research.

Additionally, I extend my sincere gratitude to my research collaborators, Prof. M. Banerjee, Prof. R. Kulsheshta, and their research group. Their collaboration, creative discussions, and unwavering support have significantly enriched my research journey. I am deeply thankful to my lab members, Akshay, Ankit, Rit Pratik, and Julie, for their invaluable help, active discussions, and for being great friends throughout this process. My heartfelt thanks also go to Pratyasha, Abhilasha, Shivam, and Vasu, with whom I had the pleasure of engaging in stimulating scientific discussions and whose assistance has been greatly appreciated.

Lastly, I owe my deepest gratitude to my family, whose unwavering support, boundless patience, and belief in me have been the foundation of my perseverance.

Rajat Punia

Abstract

Cooperative conformational transitions in biomolecules are crucial for several biological processes, including ATP catalysis, membrane transport, viral infectivity, and oxygen transport by hemoglobin. These transitions often involve collective, large-scale structural rearrangements integral to biomolecular function and regulation. Characterizing these transitions poses significant challenges due to the vast timescales involved, spanning milliseconds to seconds, which exceed the capabilities of traditional molecular dynamics (MD) simulations limited to the nanosecond to microsecond range. This thesis introduces ‘linear response driven molecular dynamics,’ a novel computational approach designed to efficiently and accurately characterize these transitions by leveraging the intrinsic dynamics of biomolecules. Our approach fits within a broader class of methods that use coarse-grained dynamics in the subspace of internal degrees of freedom, with normal mode analysis (NMA) being one of the most prominent techniques in this category.

While such methods have shown utility in applications like flexible docking, structural refinement to fit cryo-EM density maps, and predicting conformational pathways from apo to holo states, they have notable limitations. First, many existing methods rely heavily on experimental data to guide conformational sampling. In the absence of this data, they face computational challenges, where sampling time scales exponentially with the number of degrees of freedom (or modes) considered. Additionally, these methods often involve fixed displacements along selected modes without considering the thermodynamic feasibility of the resulting conformational change pathway or the metastability of intermediate states. Most of these approaches are limited in their applicability to small, globular proteins; the computational complexity becomes prohibitive for larger protein assemblies, such as viral capsids, necessitating significant methodological developments for broader applications. While existing methods account for the elastic properties of selected modes, they often overlook frictional damping, which is particularly limiting for biomolecules like lipid bilayers, where frictional effects play a crucial role in governing conformational transitions.

To address these limitations, we developed a multi-step, linear response-driven molecular dynamics simulation protocol that samples the transition pathways and

intermediate states associated with large-scale conformational changes. This approach employs the computationally efficient elastic network model (ENM) to represent the intrinsic vibrations of biomolecules and linear response theory (LRT) to identify a small, relevant subset of modes. Instead of fixed conformational displacements, structural changes along these selected modes are dynamically accelerated using excited normal modes molecular dynamics (MDeNM) simulations. This protocol includes a systematic approach for selecting excitation velocity vectors and incorporates frictional damping where necessary, yielding a computationally efficient and physically accurate method to explore complex conformational transitions in biomolecular systems.

The first objective of this thesis addresses the challenge of delineating protein conformational transition pathways upon ligand binding. Employing an elastic network model description and LRT, we identified slow normal modes of protein vibrations that correspond with conformational shifts from the apo to holo states. Conformational along these modes were dynamically updated using MDeNM simulations, providing a series of intermediate states. Herein, the MDeNM excitation velocity was optimized on-the-fly on the basis of its coupling to protein dynamics and ligand–protein interactions. Thus, a determined set of structures was validated against crystallographic and simulation data on four protein-ligand systems, namely, adenylate kinase–di(adenosine-5')pentaphosphate, ribose binding protein– β -d-ribose, DNA β -glucosyltransferase–uridine-diphosphate, and G-protein α subunit–guanosine-5'-triphosphate, which present important differences in protein conformational heterogeneity, ligand binding mechanism, viz. induced-fit or conformational selection, extent, and nonlinearity in protein conformational changes upon ligand binding, and presence of allosteric effects.

In the second objective, we explored the native-state dynamics of the Flock House virus (FHV) capsid and its role in the release mechanism of cell-penetrating γ -peptides, a crucial process for viral infectivity. Through MD simulations and NMA, we found that γ -peptide release is intrinsically linked to the capsid's vibrational

dynamics. MDeNM simulations revealed large-scale structural rearrangements, enabling a detailed mapping of the free energy landscape and elucidating the molecular-level mechanisms of peptide release. We analyzed the intrinsic motions of a pentameric sub-structure of the FHV capsid, accounting for the influence of the surrounding capsid on the substructure’s dynamics. We observed that a single normal mode of native state dynamics was dominantly associated with the initial direction of γ -peptide release. Sampling along this mode using MDeNM simulations showed a large-scale structural rearrangement in the capsid and an orientation flip of the γ_A peptide during release. These findings were used to identify relevant collective variables (CVs) and to map the free energy landscape using bias-exchange metadynamics, thereby elucidating the molecular-level mechanism of peptide release.

In third objective, we investigated lipid membrane remodeling, which is vital for numerous cellular processes. Slow dynamics and the lack of knowledge of reaction coordinates (RCs) for biasing methods pose a challenge for all-atom simulations to study conformational changes in lipid bilayers. We characterized bilayer vibrational dynamics by mapping system dynamics onto Langevin dynamics in normal mode space, derived from an ENM of the lipid-water Hamiltonian. All-atom MD simulations were used to determine model parameters, and Langevin dynamics predictions for bilayer structural, mechanical, and dynamic properties are validated against MD simulations and experiments. This method elucidated generic reaction coordinates (RCs) for pore formation in DMPC and DPPC bilayers. Enhanced sampling along these RCs, using path-metadynamics and umbrella sampling, revealed the thermodynamics of pore formation and highlighted the roles of lipid solvation and packing.

The importance of such precise characterizations of protein conformational transitions in drug discovery is exemplified in the fourth objective, which focused on the investigation of differential transitions in the hypoxia-inducible factor 2 (HIF-2) upon agonist and antagonist binding. HIF-2 is critical for cellular adaptation to low oxygen conditions and, both activation and inhibition of HIF-2 activity have significant therapeutic relevance, with potential applications in treating cancer, renal anemia, and acute kidney injury. We showed that multiple HIF-2 structures with distinct binding pocket conformations were required for docking-based ligand screening to effectively capture the plasticity of the binding domain. Using all-atom MD simulations of HIF-2 with various ligands and decoys, we identified unique

structural, dynamic, and thermodynamic changes associated with different ligand categories. We analyzed the allosteric mechanism of ligand-induced local structural and dynamic changes, as well as their global effects and corresponding time scales. Our findings demonstrate that distinct ligand-induced conformational changes can effectively distinguish agonists from antagonists, aiding in the identification of potential active compounds and minimizing false positives in virtual screening.

सार

बायोमोलेक्यूल्स में सहकारी संरचनात्मक परिवर्तन एटीपी उत्प्रेरण, झिल्ली परिवहन, वायरल संक्रमणशीलता, और हीमोग्लोबिन द्वारा ऑक्सीजन परिवहन जैसी मौलिक जैविक प्रक्रियाओं के लिए महत्वपूर्ण हैं। ये परिवर्तन अक्सर सामूहिक, बड़े पैमाने पर संरचनात्मक पुनर्व्यवस्थाओं को शामिल करते हैं जो बायोमोलेक्यूल्स के कार्य और विनियमन के लिए अनिवार्य हैं। इन परिवर्तनों का वर्णन करना महत्वपूर्ण चुनौतियाँ प्रस्तुत करता है क्योंकि ये मिलीसेकंड से सेकंड तक फैले विस्तृत समय सीमाओं में होते हैं, जो पारंपरिक मॉलिक्यूलर डायनेमिक्स (MD) सिमुलेशन की क्षमताओं से परे हैं, जो नैनोसेकंड से माइक्रोसेकंड तक सीमित हैं। यह शोध प्रबंध 'लिनियर रिस्पॉन्स ड्रिवन मॉलिक्यूलर डायनेमिक्स' नामक एक नवीन कम्प्यूटेशनल दृष्टिकोण प्रस्तुत करता है, जो बायोमोलेक्यूल्स की आंतरिक डायनेमिक्स का लाभ उठाकर इन परिवर्तनों को प्रभावी और सटीक रूप से वर्णित करने के लिए डिज़ाइन किया गया है। हमारा दृष्टिकोण उन विधियों के व्यापक वर्ग में फिट होता है जो आंतरिक स्वतंत्रता डिग्री के उपक्षेत्र में कोर्स-ग्रेन्ड डायनेमिक्स का उपयोग करती हैं, जिसमें नॉर्मल मोड एनालिसिस (NMA) इस श्रेणी में सबसे प्रमुख तकनीकों में से एक है।

लांकि ऐसी विधियों ने फ्लेक्सिबल डॉकिंग, क्रायो-ईएम घनत्व मानचित्रों में फिट होने के लिए संरचनात्मक परिशोधन, और एपो से होलो अवस्थाओं तक संरचनात्मक मार्गों की भविष्यवाणी जैसे अनुप्रयोगों में उपयोगिता दिखाई है, लेकिन उनकी उल्लेखनीय सीमाएँ हैं। पहला, कई मौजूदा विधियाँ संरचनात्मक सैम्पलिंग को मार्गदर्शन देने के लिए प्रयोगात्मक डेटा पर अत्यधिक निर्भर हैं। इस डेटा की अनुपस्थिति में, वे कम्प्यूटेशनल चुनौतियों का सामना करती हैं जो विचार किए गए स्वतंत्रता डिग्री (या मोड्स) की संख्या के साथ घातीय रूप से बढ़ती हैं। इसके अलावा, ये विधियाँ अक्सर चयनित मोड्स के साथ निश्चित विस्थापन शामिल करती हैं, बिना परिणामी संरचनात्मक परिवर्तनों की उष्मागतिकीय संभवता या मध्यवर्ती अवस्थाओं की अर्धस्थायित्व को ध्यान में लिए। इनमें से अधिकांश दृष्टिकोणों की प्रयोज्यता छोटे, गोलाकार प्रोटीनों तक सीमित है; बड़ी प्रोटीन असेंबलीज जैसे वायरल कैप्सिड्स के लिए कम्प्यूटेशनल जटिलता अत्यधिक हो जाती है, जो व्यापक अनुप्रयोगों के लिए महत्वपूर्ण पद्धति विकास की आवश्यकता उत्पन्न करती है। जबकि मौजूदा विधियाँ चयनित मोड्स के लोचदार गुणों को ध्यान में रखती हैं, वे अक्सर घर्षणीय डैम्पिंग को नजरअंदाज कर देती हैं, जो लिपिड बाइलेयर्स जैसे बायोमोलेक्यूल्स के लिए विशेष रूप से सीमित है, जहाँ घर्षणीय प्रभाव संरचनात्मक संक्रमणों को नियंत्रित करने में महत्वपूर्ण भूमिका निभाते हैं।

इन सीमाओं को दूर करने के लिए, हमने एक बहु-चरणीय, लिनियर रिस्पॉन्स-चालित कम्प्यूटेशनल प्रोटोकॉल विकसित किया है जो बड़े पैमाने पर संरचनात्मक परिवर्तनों से जुड़े संक्रमण मार्गों और मध्यवर्ती अवस्थाओं का नमूना लेता है। यह दृष्टिकोण बायोमोलेक्यूल्स की आंतरिक कंपन को दर्शाने के लिए कम्प्यूटेशनल रूप से कुशल इलास्टिक नेटवर्क मॉडल (ENM) का उपयोग करता है और मोड्स के एक छोटे, प्रासंगिक उपसमूह की पहचान करने के लिए लिनियर रिस्पॉन्स थ्योरी (LRT) को अपनाता है। निश्चित संरचनात्मक विस्थापनों के बजाय, चयनित मोड्स के साथ संरचनात्मक परिवर्तन उत्तेजित नॉर्मल मोड्स

मॉलिक्यूलर डायनेमिक्स (MDeNM) सिमुलेशनों का उपयोग करके गतिशील रूप से तीव्र किए जाते हैं। यह प्रोटोकॉल उत्तेजना वेग वेक्टरों के चयन के लिए एक प्रणालीगत दृष्टिकोण शामिल करता है और जहां आवश्यक हो वहां घर्षणीय डैम्पिंग को सम्मिलित करता है, जिससे बायोमोलेक्यूलर सिस्टम में जटिल संरचनात्मक संक्रमणों का पता लगाने के लिए एक कम्प्यूटेशनल रूप से कुशल और भौतिक रूप से सटीक विधि प्राप्त होती है।

इस शोध प्रबंध का पहला उद्देश्य लिगेंड बाइंडिंग के समय प्रोटीन संरचनात्मक संक्रमण मार्गों को परिभाषित करने की चुनौती को संबोधित करता है। इलास्टिक नेटवर्क मॉडल विवरण और LRT का उपयोग करके, हम प्रोटीन कंपन के धीमे नॉर्मल मोड्स की पहचान करते हैं जो एपो से होली अवस्थाओं में संरचनात्मक परिवर्तनों के अनुरूप होते हैं। इन मोड्स के साथ संरचनाओं को MDeNM सिमुलेशनों का उपयोग करके गतिशील रूप से अपडेट किया गया, जिससे मध्यवर्ती अवस्थाओं की एक श्रृंखला प्राप्त हुई। यहां, MDeNM उत्तेजना वेग को प्रोटीन डायनेमिक्स और लिगेंड-प्रोटीन इंटरैक्शन के साथ उसके कपलिंग के आधार पर ऑन-द-फ्लाई ऑप्टिमाइज़ किया गया। इस प्रकार, संरचनाओं के एक निर्धारित सेट को चार प्रोटीन-लिगेंड सिस्टमों पर क्रिस्टलोग्राफिक और सिमुलेशन डेटा के विरुद्ध सत्यापित किया गया, अर्थात् एडेनिलेट किनेज-डाइ(एडेनोसिन-5)पेंटाफॉस्फेट, राइबोज बाइंडिंग प्रोटीन- β -D-रिबोपाइरानोज़, DNA β -ग्लूकोसिलट्रांसफेरेज़-यूरिडीन-डाइफॉस्फेट, और G-प्रोटीन α सबयूनिट-गुआनोसिन-5-ट्राइफॉस्फेट, जो प्रोटीन संरचनात्मक विषमता, लिगेंड बाइंडिंग तंत्र जैसे प्रेरित-फिट या संरचनात्मक चयन, सीमा और लिगेंड बाइंडिंग पर प्रोटीन संरचनात्मक परिवर्तनों में गैर-रेखीयता, और एलोस्टेरिक प्रभावों की उपस्थिति में महत्वपूर्ण अंतर प्रस्तुत करते हैं।

दूसरे उद्देश्य में, हम Flock House वायरस (FHV) कैप्सिड की मूल-राज्य डायनेमिक्स और सेल-प्रवेश करने वाले γ -पेप्टाइड्स के रिलीज़ तंत्र में उसकी भूमिका का अध्ययन करते हैं, जो वायरल संक्रमणशीलता के लिए एक महत्वपूर्ण प्रक्रिया है। MD सिमुलेशन्स और NMA के माध्यम से, हमने पाया कि γ -पेप्टाइड रिलीज़ कैप्सिड की कंपन डायनेमिक्स से आंतरिक रूप से जुड़ी हुई है। MDeNM सिमुलेशन्स ने बड़े पैमाने पर संरचनात्मक पुनर्व्यवस्थाओं को उजागर किया, जिससे मुक्त ऊर्जा परिदृश्य का विस्तृत मानचित्रण संभव हुआ और पेप्टाइड रिलीज़ के आणविक-स्तर के तंत्रों को स्पष्ट किया गया। MD सिमुलेशन्स और NMA का उपयोग करते हुए, हमने FHV कैप्सिड की एक पेंटामेरिक उप-संरचना की आंतरिक गतियों का विश्लेषण किया, जिसमें उप-संरचना की डायनेमिक्स पर आसपास के कैप्सिड के प्रभाव को ध्यान में रखा गया। हमने देखा कि मूल-राज्य डायनेमिक्स के एक एकल नॉर्मल मोड का γ -पेप्टाइड्स की प्रारंभिक रिलीज़ दिशा के साथ प्रमुख संबंध था। इस मोड के साथ MDeNM सिमुलेशन्स द्वारा सैम्पलिंग करने पर कैप्सिड में बड़े पैमाने पर संरचनात्मक पुनर्व्यवस्था और रिलीज़ के दौरान γ_A पेप्टाइड्स के अभिविन्यास में उलटफेर का संकेत मिला। इन निष्कर्षों का उपयोग प्रासंगिक सामूहिक चर की पहचान करने और बायस-एक्सचेंज मेटाडायनेमिक्स का उपयोग करके मुक्त ऊर्जा परिदृश्य को मैप करने के लिए किया गया, जिससे पेप्टाइड रिलीज़ के आणविक-स्तर के तंत्र को स्पष्ट किया जा सका।

तीसरे उद्देश्य में, हम लिपिड मेम्ब्रेन रिमॉडलिंग की जांच करते हैं, जो कई कोशिकीय प्रक्रियाओं के लिए महत्वपूर्ण है। धीमी डायनेमिक्स और बायसिंग विधियों के लिए प्रतिक्रिया समन्वयकों (RCs) के ज्ञान की कमी, लिपिड बाइलेयर्स में संरचनात्मक परिवर्तनों का अध्ययन करने के लिए ऑल-एटम सिमुलेशन्स के लिए एक चुनौती प्रस्तुत करती है। लिपिड-पानी हैमिल्टोनियन के एक ENM से व्युत्पन्न नॉर्मल मोड स्पेस में सिस्टम डायनेमिक्स को लैंग्विन डायनेमिक्स पर मैप करके, हमने बाइलेयर के कंपन डायनेमिक्स का वर्णन किया। मॉडल पैरामीटर्स निर्धारित करने के लिए ऑल-एटम MD सिमुलेशन्स का उपयोग किया जाता है, और बाइलेयर की संरचनात्मक, यांत्रिक और डायनेमिक गुणों के लिए लैंग्विन डायनेमिक्स की भविष्यवाणियों को MD सिमुलेशन्स और प्रयोगों के विरुद्ध सत्यापित किया जाता है। इस विधि ने DMPC और DPPC बाइलेयर्स में पोर्स निर्माण के लिए सामान्य RCs को स्पष्ट किया। पाथ-मेटाडायनेमिक्स और अम्ब्रेला सैम्पलिंग का उपयोग करके इन RCs के साथ संवर्धित सैम्पलिंग ने पोर्स निर्माण की उष्मागतिकी को उजागर किया और लिपिड सॉल्वेशन और पैकिंग की भूमिकाओं को रेखांकित किया।

प्रोटीन के संरचनात्मक संक्रमणों की सटीक पहचान का दवा अनुसंधान में महत्व, हमारे चौथे उद्देश्य में स्पष्ट होता है, जिसमें हाइपोक्सिया-इंड्यूसिबल फैक्टर 2 (HIF-2) के एगोनिस्ट और एंटागोनिस्ट बाइंडिंग पर अलग-अलग संक्रमणों की जांच पर ध्यान केंद्रित किया गया। HIF-2, जो कोशिकाओं को कम ऑक्सीजन परिस्थितियों में अनुकूलन के लिए महत्वपूर्ण है, का सक्रियण और अवरोधन दोनों ही कैंसर, गुर्दे की एनीमिया और तीव्र गुर्दे की चोट जैसे रोगों के उपचार में महत्वपूर्ण चिकित्सीय उपयोगिता रखते हैं। हमने दिखाया कि डॉकिंग-आधारित लिगेंड स्क्रीनिंग के लिए HIF-2 की कई संरचनाएँ, जिनमें बाइंडिंग पॉकेट के विभिन्न रूप होते हैं, आवश्यक हैं ताकि बाइंडिंग डोमेन की प्लास्टिसिटी को प्रभावी ढंग से समझा जा सके। HIF-2 के विभिन्न लिगेण्ड्स और डिक्लोइस के साथ ऑल-एटम MD सिमुलेशन का उपयोग करते हुए, हमने अलग-अलग प्रकार के लिगेण्ड्स से जुड़े अद्वितीय संरचनात्मक, गतिशील और ऊष्मागतिकीय परिवर्तनों की पहचान की। हमने लिगेण्ड-प्रेरित स्थानीय संरचनात्मक और गतिशील परिवर्तनों की एलोस्टेरिक प्रक्रिया का विश्लेषण किया, साथ ही उनके वैश्विक प्रभाव और संबंधित समय-सीमाओं का अध्ययन किया। हमारे निष्कर्षों ने यह दर्शाया कि विशिष्ट लिगेण्ड-प्रेरित संरचनात्मक परिवर्तन एगोनिस्ट और एंटागोनिस्ट को प्रभावी रूप से अलग कर सकते हैं, जिससे सक्रिय यौगिकों की पहचान में सहायता मिलती है और वर्चुअल स्क्रीनिंग में गलत सकारात्मक परिणामों को कम किया जा सकता है।

Contents

Certificate

Acknowledgements

Abstract

Contents

List of Figures

List of Tables

Abbreviations

1	Introduction	1
2	Literature Review	11
2.1	Computation of protein structure transition pathway upon ligand binding	15
2.2	Mechanism of membrane-lytic peptides release in Flock House Virus .	19
2.3	Molecular mechanism of slow and cooperative structural transitions in lipid bilayers	22
2.4	Mechanisms and therapeutic potential of differential HIF-2 activity modulation	26
3	Research Methodology	29
3.1	Molecular dynamics simulations	29
3.1.1	Protein conformational transitions on ligand binding: Protein-ligand complex MD simulations	31
3.1.2	MD simulations of Flock House Virus (FHV) capsid	32

3.1.3	MD simulations of lipid bilayers	32
3.1.4	MD simulations of HIF-2/ARNT complex	33
3.2	Normal mode analysis	34
3.2.1	Normal modes calculations for protein-ligand systems	35
3.2.2	Normal mode analysis of FHV capsid dynamics	36
3.2.3	Normal mode analysis of lipid-bilayers	37
3.3	Excited normal modes molecular dynamics	38
3.4	Linear response direction of protein conformational change	39
3.4.1	Calculation of ligand-induced force	40
3.4.2	Multi-step MDeNM simulations	41
3.4.3	Structure similarity and structure selection metrics	42
3.5	Langevin model of lipid bilayer dynamics	43
3.5.1	Estimation of bilayer properties from uL-NMS	44
3.5.2	Pore formation pathway from MDeNM	45
3.6	Free energy calculations	46
3.6.1	Protein apo to holo conformational transitions	47
3.6.2	Transient pore formation in lipid bilayers	48
3.6.3	Peptides release in FHV capsid	49
4	Computation of Protein Conformational Transition Pathway on Ligand Binding by Linear Response-Driven Molecular Dynamics	51
4.1	Theory	52
4.1.1	Nonlinear conformational change	53
4.1.2	Direction of protein conformational change	54
4.1.3	Excited normal modes molecular dynamics (MDeNM)	55
4.2	Results	58
4.2.1	Intrinsic motions triggered by ligand binding	59
4.2.2	Holo-state and intermediates on apo to holo pathway	64
4.2.2.1	Single- and multi-step MDeNM	64
4.2.2.2	MFEP from path metadynamics	71
4.2.2.3	Apo to holo transition pathway of adenylate kinase	73
4.2.2.4	Apo-holo transition pathway of ribose binding protein	76
4.2.3	Multi-step MDeNM application to system with large apo-state conformational heterogeneity	78
4.3	Discussion	82
5	Mechanism of Release of Membrane-lytic Peptides in Flock House Virus is Intrinsic to its Conformational Dynamics	85
5.1	Results	86
5.1.1	Normal mode analysis of peptides release in Flock House Virus capsid	86

5.1.2	Excited normal modes molecular dynamics simulations of peptides release	90
5.1.3	Free energy calculations of peptide release using bias-exchange metadynamics	93
6	Free Energy Surface and Molecular Mechanism of Slow Structural Transitions in Lipid Bilayers	99
6.1	Results	100
6.1.1	Langevin model for a lipid bilayer in water	100
6.1.2	Quantitative assessment of the Langevin model	101
6.1.3	ENM for a protein embedded in a lipid bilayer	106
6.1.4	Pore formation in tensionless lipid bilayers	109
6.2	Discussion	115
7	Key Descriptors Reveal the Differential Mechanism of HIF-2 Activity Modulators	123
7.1	Significance of accounting for HIF-2 PAS-B domain conformational plasticity	124
7.2	Differential effect of agonists and antagonists on HIF-2/ARNT structure and dynamics	128
7.3	Large-scale virtual screening of small-molecules	133
8	Scope for Future Work	137
A	Supplementary Information: Computation of Protein Conformational Transition Pathway on Ligand Binding by Linear Response-Driven Molecular Dynamics	197
A.1	Theory: Direction of protein conformational change on ligand binding	197
A.1.1	Weak perturbation limit	200
A.1.2	Strong perturbation, Gaussian fluctuations limit	201
A.2	Supplementary Figures	203
A.3	Supplementary Tables	216
A.4	Supplementary Methods	219
A.4.1	Protein-ligand complex preparation	219
A.4.2	Normal mode analysis	220
A.4.3	Calculation of linear-response direction, enhanced-sampling, and selection of output structures	222
A.4.4	Collective variables	225
A.4.5	Path collective variables and metadynamics simulations	226
A.4.6	Principal component analysis and clustering	227
A.4.7	Structure displacement along normal modes	228

A.4.8	Adenylate Kinase LID-NMP domains interface area	228
A.4.9	MM/PBSA protein-ligand binding energy analysis	229
B	Supplementary Information: Mechanism of release of membrane-lytic peptides in Flock House Virus is intrinsic to its conformational dynamics	231
B.1	Vibrational subsystem analysis of FHV capsid pentameric unit dynamics	231
B.2	Supplementary Methods	234
B.2.1	Molecular dynamics simulations	234
B.2.2	Determination of ENM parameters	235
B.2.3	Collective Variables	236
B.2.4	Bias-Exchange Metadynamics Simulations	237
B.3	Supplementary Figures	238
C	Supplementary Information: Free energy surface and molecular mechanism of slow structural transitions in lipid bilayers	241
C.1	Methods	241
C.1.1	Simulation systems, protocols and parameters	241
C.1.2	Langevin model for a lipid bilayer system in water environment	243
C.1.2.1	Determination of Hessian matrix	244
C.1.2.2	Determination of Friction matrix	249
C.1.3	Undulations structure factor: Static and relaxation	250
C.1.3.1	MD	251
C.1.3.2	uL-NMS	252
C.1.4	Dynamic structure factor	255
C.1.4.1	MD	256
C.1.4.2	uL-NMS	256
C.1.5	Excited normal modes molecular dynamics	259
C.1.6	Path CVs and free energy calculations	260
C.1.7	Lipid bilayer order parameters	264
C.1.7.1	Bilayer thickness profile	264
C.1.7.2	Lipid orientation profile	264
C.1.7.3	Coordination of lipid polar and non-polar atoms	265
C.1.7.4	Acyl chain order parameter	266
C.1.7.5	Formation and rupture of water channel	267
C.1.8	Miscellaneous	267
C.1.8.1	Determination of lipid-bilayer ENM parameters	267
C.1.8.2	Exploring static structure factor for fixing ENM parameters	269
C.1.8.3	Previously defined pore-formation reaction coordinates	270

C.1.8.4	MM/PBSA energy analysis	270
C.2	Supplementary Figures	272
C.3	Supplementary Tables	272
D	Supplementary Information: Key descriptors reveal the differential mechanism of HIF-2 activity modulators	285
D.1	Supplementary Methods	285
D.1.1	Ligand Force Field Parameterization	285
D.1.2	Amide Hydrogen Protection Factors	287
D.2	Supplementary Figures	288
D.3	Supplementary Tables	291
	References	293

List of Figures

4.1	Estimation of ANM parameters and identification of normal modes triggered by ligand binding for adenylate kinase. (A) Pearson correlation (ρ) between residue mean squared fluctuations predicted by ANM (with different R_c values) and estimated from a 1 ns MD simulation is on left axis ($\bullet$). Lower bound at $\rho = 0.95\rho_{max}$ is shown ($\cdots$). Variation of Shannon entropy measure of first M dominant modes, η (equation 4.10), with R_c is on right axis ($\circ$). (B) Mean squared fluctuations of AdK residues in apo-state obtained from 1 ns MD simulation ($---$) and estimated by ANM at $R_c = 10 \text{ \AA}$ ($---$). (C) Overlap (equation 4.4) of each of the first 20 normal modes with $\mathbf{d}_{LRT}$. (D) Cumulative overlap of top M normal modes (equation 4.11) with $\mathbf{d}_{LRT}$, K_M ($\bullet$); reference lines at $K_M = 0.9, 0.95$, and 0.99 ($\cdots$) are also shown. (E) Direction of residue displacement along Mode 1, Mode 2, and Mode 3 are shown by blue arrows on the protein (in cartoon representation) - ligand (in bond stick) complex.	61
4.2	Structural change on one step excitation along ($\mathbf{d}_{LRT}$) for AdK–Ap5A. (A) Displacement along ($\mathbf{d}_{LRT}$), (D_P), in a 100 ps MDeNM simulation at various excitation temperatures, ΔT , ($\bullet$) along with its three-point average ($---$). Regions with clusters of similar D_P values (plateau regions) are shown as separate color bands. (B) C_α -RMSD and (C) average GDT (global distance test) score, GDT_{avg} , (average calculated using three cutoff values: $2 \text{ \AA}$, $4 \text{ \AA}$, and $8 \text{ \AA}$) between the experimental holo structure and the final structure of MDeNM simulation. Structure with highest D_P in each plateau region is highlighted ($\odot$) in all panels.	65

- 4.3 **Holo-state structure prediction from multi-step MDeNM.** (A, B and C) Cartoon diagram of initial docked (left) and MDeNM predicted (right) holo structure (green) superimposed on the corresponding crystal structure (cyan) for AdK–Ap5A[1], RBP–RIP[2], and BGT–UDP[3], respectively. (D, E and F) Mean square displacement of predicted (—●—) and experimental (---○---) holo-structure from the initial docked structure (I_1), D_m , along first 10 normal modes of I_1 for AdK, RBP, and BGT, respectively (calculated using equation A.36). (G, H and I) Trajectory of predicted apo $\rightarrow$ holo transition pathway in the subspace constituted by top three normal modes of I_1 (in order of decreasing D_m) for AdK, RBP, and BGT, respectively. 67
- 4.4 **Minimum free-energy path (MFEP) from path metadynamics.** (A, B, and C) Projection of MFEP obtained in the space of two path collective variables (Supplementary Figure A.8), $s(\mathbf{R})$ and $z(\mathbf{R})$, along the path $s(\mathbf{R})$ variable for AdK–Ap5A, RBP–RIP, and BGT–UDP, respectively. The apo-state lies at $s(\mathbf{R}) = 0$ (labeled I_1) and the holo-state at $s(\mathbf{R}) = 1$. The metastable states along this 1D projection are labeled as M_i and the intermediates from multi-step MDeNM are labeled as E_i . 71
- 4.5 **Apo $\rightarrow$ holo transition pathway of AdK.** Projection of transition intermediates from the multi-step MDeNM protocol (●) on the $(\theta_{LID}, \theta_{NMP})$ space. The lowest free energy path and intermediates (---○---) obtained using long-timescale MD and bias-exchange metadynamics are also shown. Projection of lowest free energy path, obtained in path CVs space using well-tempered metadynamics (shown in Supplementary Figure A.8(A)), on the $(\theta_{LID}, \theta_{NMP})$ space (—); error bars shows the region of uncertainty. Surface diagrams of apo- (I_0), initial docked state (refined in short MD, I_1), transition intermediates (E_1 to E_4) and the holo-state (E_5) obtained from multi-step MDeNM are shown with colour scheme: green (CORE domain), blue (LID domain), red (NMP domain) and yellow (Ap5A). 75

- 4.6 **Apo $\rightarrow$ holo transition pathway of RBP.** (A) Surface diagrams of apo- (I_0), the initial docked state (refined in short MD, I_1), transition intermediates (E_1, E_2) and the holo-state (E_3) obtained from multi-step MDeNM: N-terminal domain (orange), C-terminal domain (green), hinge region (white), ligand (RIP, yellow), and RBP residues that affect ribose (RIP) transport but not binding, viz. 11, 12, 44, 45, 52, 67, 70, 72, 73, 134, 165, and 166 (blue)[4]. (B) C_α -RMSD of predicted intermediates calculated with respect to RBP crystallographic structures with PDB IDs 1BA2_A, 1BA2_B, 1URP, 2GX6, and 2DRI. (C) Projection of the transition intermediates predicted by our protocol ($\bullet$) along with the crystallographic structures ($\circ$) and metastable states of RBP obtained using umbrella sampling simulation[5] (stars) on the (θ, ϕ) -space. Projection of lowest free energy path, obtained in path CVs space using well-tempered metadynamics (shown in Supplementary Figure A.8(B)), on the (θ, ϕ) -space (---); error bars shows the region of uncertainty. 78
- 4.7 **Apo $\rightarrow$ holo transition for the G_α -GTP system.** (A) Cartoon diagram of nine rigid-body docked structures (outer periphery) and an overlay of multi-step MDeNM predicted holo structures (pink; all except C_6 are transparent) with the experimental holo-state (blue). The apo state population of nine cluster centers is also listed. (B) Projection of the transition path to the holo state from each of the nine rigid-body docked complexes ($\text{---}\bullet$) in the space of top 3 as well as top 2 PCs. Crystallographic holo-state structure is also shown. . . . 80
- 5.1 **Normal Mode Analysis of γ Peptides Release from FHV capsid** (A) Structure of the FHV capsid (PDB ID: 4FTB). The red dashed lines indicate the boundaries of an icosahedral asymmetric unit (iASU). The β_A , β_B , and β_C subunits within the iASU are highlighted in white, gray, and blue, respectively. The 5-fold axis of symmetry is also clearly marked. (B–C) Pentameric substructure (5iA) of the FHV capsid utilized for simulations, consisting of 5 iASUs. (B) External view showing the arrangement of iASUs. (C) Internal view displaying γ_A , γ_B , and γ_C peptides, colored red, orange, and green, respectively. (D) C_α RMSD of the 5iA with (blue solid line) and without (black solid line) harmonic restraints. (E) RMSF of C_α atoms 5iA residues from 40 ns all-atom MD simulation (red dashed line) and estimated from ENM (black solid line). (F) Change in order parameter ξ for fixed displacement of 2 Å along normal modes 88

5.2	Excited Normal Mode MD Simulations of γ_A Peptides Release. (A–C) Surface diagrams of 5iA configurations at the end of 100 ps MDeNM simulations, shown at excitation temperatures of $\Delta T = 0$ K (A), $\Delta T = 35$ K (B), and $\Delta T = 70$ K (C). The β_A , β_B , and β_C subunits are depicted with white, gray, and blue transparent surfaces, respectively, while γ_A peptides are highlighted in dark red. (D) Progress variable ξ measurements at the end of MDeNM simulations conducted at varying ΔT from 0 K to 75 K, in 5 K increments; inset shows the time evolution of ξ for ΔT ranging from 0 K to 70 K in 10 K increments. (E–G) Values of order parameters, d_{mode} (E), d_{base} (F), and ϕ (G) at the end of MDeNM simulations.	91
5.3	Minimum free energy pathway of γ-peptides release. (A–C) 2-dimensional projections of the free energy landscape (ΔG), which is function of 3 OPs: d_{mode} , d_{base} and ϕ . (D) Visualization of ΔG across the full 3-dimensional parameter space. The red-dashed line delineates the minimum free energy pathway (MFEP). Minima along the MFEP are labeled (a)–(d) and corresponding structural configurations are illustrated.	95
5.4	Collectivity of γ-Peptides Release. (A) Displacement of individual β subunits of 5iA along mode 3, calculated at each intermediate along MFEP, as detailed in Figure 5.3. (B–C) Distance from the base, defined by the COM of the β_B and β_C subunits (B), and orientation angle ϕ (C), calculated for each γ_A peptide for different MFEP intermediates.	96
6.1	Validation of uL-NMS model of a DMPC lipid bilayer. (A) Static undulations spectra, $I_u(q)$ (Eq. C.11): MD ($\circ$), uL-NMS ($\diamond$, Eq. C.23). (B) Decay rate for undulation auto-correlation, $\tau_u(q)$: MD simulations (this work ($\circ$), Brandt et al.[6] ($\blacktriangleright$)), uL-NMS ($\diamond$, Eq. C.24), Nuclear Spin Echo (NSE) experiments (Nagao et al.[7] ($\blacktriangleright$), Yi et al.[8] ($\blacktriangleright$)). (C) Dynamic structure factor, $S(q = 0.70 \text{ nm}^{-1}, \omega)$ (Eq. C.29): MD (---), uL-NMS (---); inset: magnified view of $S(q, \omega)$ in low ω region. (D) Rayleigh widths, Γ_h (Eq. C.39), calculated from the structure factor data (points) and the quadratic fits (line), $\Gamma_h(q) = D_T q^2$, are shown for MD ($\circ$, ---), uL-NMS ($\diamond$, ---), NSE experiments (Tarek et al.[9] ($\bullet$, ---), Chen et al.[10] ($\blacksquare$, ---)). A 1 μs AA MD simulation of a 1024 DMPC bilayer is used for calculation of all properties reported here. Error bars in the MD data correspond to the standard error for I_u and 95% confidence interval for fitted parameters τ and Γ_h . Curve fittings were performed over the average data distribution from the MD trajectory. Error bars correspond to the standard deviation in the uL-NMS data, where the uL-NMS estimates were obtained from 10 independent bilayer starting conformations selected every 100 ns from a 1 μs trajectory. . .	103

- 6.2 **Integration of lipid-bilayer ENM with protein ENM for γ -secretase embedded in DPPC bilayer.** (A) A representative configuration of γ -secretase (cartoon) in DPPC bilayer (ball-stick). The protein complex comprises four subunits, viz. extra-cellular domain nicastrin (NCT) that also comprises a membrane-embedded 15 residue helical unit, and transmembrane proteins presenilin-1 (PS1), anterior pharynx defective 1 (APH-1), and presenilin enhancer 2 (PEN-2). (B) Pearson correlation between C_α -RMSF of membrane-bound γ -secretase calculated from a 200 ns MD simulation and from ENM-PL at various $\gamma_{\text{PL}}/\gamma_{\text{LL}}$: κ_{mem} for membrane-embedded residues ($\triangleleft$), κ_{EC} for extracellular-domain residues ($\circ$), and κ_0 for the whole complex ($\blacksquare$). (C) RMSF of γ -secretase residues obtained from the MD simulation ($---$), ENM-P ($\cdots$), and ENM-PL ($—$). 107
- 6.3 **Direction of the local minimum free energy pathway for pore formation from uL-NMS.** (A–C) DMPC lipid bilayer configurations corresponding to $\xi = 0, 0.4$, and 0.75 , respectively. (D) Contribution of 50 slowest normal modes (α_i) to $\hat{\mathbf{d}}_{\text{pore}}$ for a flat bilayer. (E) Projection of $\hat{\mathbf{d}}_{\text{pore}}$ in the bilayer plane, representing lateral displacement of lipid beads (arrows) and predicted density change (contour lines) from the continuity equation, $\partial\rho/\partial t = \nabla \cdot (\rho\mathbf{u})$ 111
- 6.4 **Order parameter evolution and the potential of mean force (PMF) for prepore opening and closing in a DMPC bilayer.** Variation of (A) lipid density order parameter ξ , (B) water density order parameter ξ_{MF} , (C) PCV $z(\mathbf{R})$, and (D) the pore formation PMF ΔG_{pore} in umbrella sampling windows restrained along $s(\mathbf{R})$, with starting configurations taken from either the pore-opening ($\blacksquare$ or $—$) or the pore-closing path ($\blacklozenge$ or $—$). The last 30 ns trajectory of each umbrella window was used for analysis. One standard-deviation from mean is represented by vertical bars in (A)–(C) (estimated from respective temporal profiles) and by shaded region in (D) (estimated from WHAM as per Eq. C.49). One-dimensional projection of the pore-formation free energy along $s(\mathbf{R})$, obtained from wt-metad on PCVs $s(\mathbf{R})$ and $z(\mathbf{R})$, is also shown in (D) ($=$). MD snapshots of the representative conformation on pore-opening (E–G) and pore-closing (H–J) pathways: (E, H) the flat-bilayer ($s(\mathbf{R}) : 0.04, 0.08$), (F, I) the transition-state ($s(\mathbf{R}) : 0.63, 0.64$), and (G, J) a stable prepore ($s(\mathbf{R}) : 0.82, 0.82$). 112

- 6.5 Mechanistic characterization of pore formation in a DMPC bilayer.** (A) Two-dimensional projection of $\hat{\mathbf{d}}_{\text{pore}}$ in bilayer plane (Eq. 3.25), (B) gradient of membrane thickness (t_{XY} , Appendix C Sec. C.1.7.1) along $\hat{\mathbf{d}}_{\text{pore}}$, (C) Membrane thickness profile ($\langle t_{XY} \rangle$, Appendix C Sec. C.1.7.1), and (D) membrane normal projection along a vector from membrane surface to pore-centre ($\langle \hat{\mathbf{n}}_{XY} \cdot \hat{\mathbf{r}}_{XY} \rangle$, Appendix C Sec. C.1.7.2). In (A)–(D), data for MDeNM intermediates I_0 (flat bilayer, $s(\mathbf{R}) = 0.09$) is on left and I_4 (last intermediate before water channel formation, $s(\mathbf{R}) = 0.50$) is on right. The values reported are averages over a 1 ns trajectory from the stable portion of corresponding MDeNM run. (E–H) Variation of order parameters along the PCV $s(\mathbf{R})$: (E) $\langle \hat{\mathbf{n}}_{XY} \cdot \hat{\mathbf{r}}_{XY} \rangle_{cyl}$ (Appendix C Sec. C.1.7.2), (F) $\langle t_{XY} \rangle_{cyl}$ (Appendix C Sec. C.1.7.1), (G) fractional contribution of lipid polar atom (LP) - water coordination to total LP coordination, $\langle f_{LP-water} \rangle_{cyl}$ (Appendix C Sec. C.1.7.3), and (H) number of water in coordination only with lipid tail atoms, N_w^{tail} (Appendix C Sec. C.1.7.3). Averages calculated both from 1 ns MD trajectory of MDeNM intermediates (—■—) and from last 30 ns umbrella sampling simulation in given window (●). The vertical bars corresponds to one standard deviation. 116
- 6.6 Transition state of DMPC bilayer pore formation.** (A) Distribution of PCV $s(\mathbf{R})$ and N_w^{tail} (Appendix C Sec. C.1.7.3) calculated from the last 30 ns of the umbrella window simulation trajectory corresponding to the transition-state (●: $f_w = 1$ and ○: $f_w < 1$); inset: evolution of f_w (Appendix C Sec. C.1.7.5) during last 30 ns of the umbrella sampling simulation trajectory for TS window (—●—) and the consecutive preceding (—▲—) and succeeding (—◆—) windows. (B–J) Multiple configurations are mapped across mean (μ) and standard deviation (σ) values of $s(\mathbf{R})$ and N_w^{tail} distribution: (B–D) configurations are shown at $\{s(\mathbf{R}), N_w^{\text{tail}}\} = \{0.626, 2\}$ (B), $\{0.627, 6\}$ (C), $\{0.627, 9\}$ (D), $\{0.631, 1\}$ (E), $\{0.630, 6\}$ (F), $\{0.632, 9\}$ (G), $\{0.636, 1\}$ (H), $\{0.637, 5\}$ (I) and $\{0.638, 9\}$ (J). The lipid headgroup atoms N (blue) and P (orange) are represented as spheres, while the remaining atoms of the lipid molecules are displayed as transparent sticks. Only water oxygen atoms in coordination with lipids are shown: represented as spheres (red); oxygen atoms belonging to hydrophobic water molecules are enlarged. 118

7.1	A) Cartoon representation of the HIF-2 α /ARNT complex, with HIF-2 α shown in green and ARNT in yellow. (B) 2D histogram of the HIF-2 PAS-B domain cavity volume and PAS-B C α RMSD relative to the experimental apo-state, computed from 1.5 μ s long MD simulations in apo-state. (C) ROC curve assessing enrichment over decoys in docking calculations using Group B conformers (apo-state cluster centres plus holo-state structures); inset: ROC curve using only apo-state cluster centers. (D) Histogram of $\Delta\Delta G$ values for agonists, antagonists, and decoys.	125
7.2	(A) Distribution of ϕ_{Y281} and ϕ_{M252} angles in long MD simulations of the apo-state (blue contours), PT2385-bound (red contours), and M1001-bound (green contours) states, compared to docking followed by short 1 ns MD simulations (markers). (B) Histogram of DBD C α RMSD relative to the DNA-bound conformation of the DBD. (C) Conformational variance of the HIF-2 α /ARNT complex, computed from a series of 50 ns windows within a total of 500 ns long simulations. (D) Change in log(protection factors) relative to the apo-state for DBD domain residues.	130
7.3	(A) Actual Vina docking score versus D-MPNN predicted Vina score. (B) Root mean squared error (RMSE) and Pearson correlation between actual and D-MPNN predicted Vina docking scores for models trained with RDKit features (solid line) and without RDKit features (dashed line). (C) Distribution of Vina scores for 10 million ligands selected randomly and by the D-MPNN model. (D) Recovery rate of the top 0.1 percent ligands within the top p percent ligands identified by the D-MPNN model.	134
A.1	Multi-step MDeNM protocol for prediction of ligand-induced protein structure transitions	204
A.2	Estimation of ANM parameters and identification of normal modes triggered by ligand binding for ribose binding protein. (A) Pearson correlation (ρ) between residue mean squared fluctuations predicted by ANM (with different R_c values) and estimated from a 1 ns MD simulation is on left axis ($\bullet$). Lower bound at $\rho = 0.95\rho_{max}$ is shown ($---$). Variation of Shannon entropy measure of first M dominant modes, η , with R_c is on right axis ($\circ$). (B) Mean squared fluctuations of RBP residues in apo-state obtained from 1 ns MD simulation ($---$) and estimated by ANM at $R_c = 10 \text{ \AA}$ ($---$). (C) Overlap of each of the first 20 normal modes with $\mathbf{d}_{LRT}$. (D) Cumulative overlap of top M normal modes with $\mathbf{d}_{LRT}$, K_M ($\bullet$); reference lines at $K_M = 0.9, 0.95$, and 0.99 ($---$) are also shown. (E) Direction of residue displacement along Mode 1, Mode 2, and Mode 3 are shown by blue arrows on the protein (in cartoon representation) - ligand (spheres) complex.	205

A.3	Estimation of ANM parameters and identification of normal modes triggered by ligand binding for β-glucosyltransferase. (A) Pearson correlation (ρ) between residue mean squared fluctuations predicted by ANM (with different R_c values) and estimated from a 1 ns MD simulation is on left axis ($\bullet$). Lower bound at $\rho = 0.95\rho_{max}$ is shown ($\cdots$). Variation of Shannon entropy measure of first M dominant modes, η , with R_c is on right axis ($\circ$). (B) Mean squared fluctuations of BGT residues in apo-state obtained from 1 ns MD simulation ($---$) and estimated by ANM at $R_c = 8 \text{ \AA}$ ($---$). (C) Overlap of each of the first 20 normal modes with $\mathbf{d}_{LRT}$. (D) Cumulative overlap of top M normal modes with $\mathbf{d}_{LRT}$, K_M ($\bullet$); reference lines at $K_M = 0.9, 0.95$, and 0.99 ($---$) are also shown. (E) Direction of residue displacement along Mode 1 and Mode 2 are shown by blue arrows on the protein (in cartoon representation) - ligand (spheres) complex.	206
A.4	Excited normal modes molecular dynamics (MDeNM). (A) Variation of temperature of the AdK in water system ($\triangle$, $\circ$, $\square$) and RMSD of protein from initial structure ($\blacktriangle$, $\bullet$, $\blacksquare$) in a MDeNM excitation of 5 ps along Mode 1 at excitation temperature $\Delta T = 10 \text{ K}$ ($\square$, $\blacksquare$), 30 K ($\circ$, $\bullet$) and 50 K ($\triangle$, $\blacktriangle$). Displacement along the excited direction (D_P) in a 100 ps MDeNM excitation of AdK along Mode 1 ($---$), Mode 5 ($---$) and Mode 9 ($---$) (B) at fixed excitation temperature $\Delta T = 25 \text{ K}$ and (C) at scaled excitation temperature (for i^{th} mode, $\Delta T_i \sim \omega_i^2$): $\Delta T_1 = 8.5 \text{ K}$, $\Delta T_5 = 21 \text{ K}$ and $\Delta T_9 = 30 \text{ K}$. Maximum displacement along the excited direction (D_P^{max}) are marked for each case with dashed lines.	207
A.5	Structural change on one step excitation along ($\mathbf{d}_{LRT}$) for RBP-RIP system. (A) Displacement along ($\mathbf{d}_{LRT}$), (D_P), in a 100 ps MDeNM simulation at various values of excitation temperature, ΔT , ($\bullet$) along with its three-point average ($---$). Regions with clusters of almost similar D_P values (plateau regions) are shown as separate color bands. (B) C_α -RMSD and (C) average GDT (global distance test) score, GDT_{avg} , (average calculated using three cutoff values: $2 \text{ \AA}$, $4 \text{ \AA}$, and $8 \text{ \AA}$) between the experimental holo structure and the final structure of MDeNM simulation. Structure with highest D_P in each plateau region at increasingly higher ΔT is highlighted ($\odot$) in all panels.	208

- A.6 **Structural change on one step excitation along ($\mathbf{d}_{LRT}$) for BGT–UDP system.** (A) Displacement along ($\mathbf{d}_{LRT}$), (D_P), in a 100 ps MDeNM simulation at various values of excitation temperature, ΔT , ($\bullet$) along with its three-point average (---). Regions with clusters of almost similar D_P values (plateau regions) are shown as separate color bands. (B) C_α -RMSD and (C) average GDT (global distance test) score, GDT_{avg} , (average calculated using three cutoff values: 2 Å, 4 Å, and 8 Å) between the experimental holo structure and the final structure of MDeNM simulation. Structure with highest D_P in each plateau region at increasingly higher ΔT is highlighted ($\odot$) in all panels. 209
- A.7 **Structural change on multi-step excitation along ($\mathbf{d}_{LRT}$) for AdK–Ap5A system.** (A to H) Displacement along ($\mathbf{d}_{LRT}$), (D_P), in a 100 ps MDeNM simulation at various values of excitation temperature, ΔT , ($\bullet$) along with its three-point average (---) for first to eighth step of excitation simulations respectively. First plateau regions (clusters of almost similar D_P values) are shown as separate color bands and structure with highest D_P in the first plateau region at increasingly higher ΔT is highlighted ($\odot$) in all panels. 210
- A.8 **Free energy surface in the space of path collective variables** (A, B and C) Free energy (unit: kcal mol⁻¹) surface (contours) of AdK–Ap5A, RBP–RIP and BGT–UDP systems in the space of path collective variables (PCVs), $s(\mathbf{R})$ (equation A.32) and $z(\mathbf{R})$ (equation A.33) obtained from 800 ns, 500 ns and 500 ns long well-tempered metadynamics simulations respectively. Minimum free energy pathway from metastable state at one end ($s = 0$) to metastable state at the other end ($s = 1$) obtained using nudged elastic band method (dashed line). Projection of structures in PCVs space obtained after every 200 ps interval of 10 ns long unbiased MD simulations started from intermediates obtained using our protocol are also shown (scattered data). (D, E and F) Evolution of PCV $s(\mathbf{R})$ with time during 10 ns long unbiased MD simulations started from intermediates obtained using our protocol for AdK–Ap5A, RBP–RIP and BGT–UDP systems. The $s(\mathbf{R})$ value corresponding to the fictitious intermediates (one each for RBP–RIP and BGT–UDP systems (F_1); none for AdK–Ap5A system) used in path-metadynamics simulations (dashed line). (G, H and I) RMSD-matrix of structures used to define initial apo to holo transition pathway. 211

A.9	Minimum free-energy pathway and comparison of intermediates predicted using our multi-step MDeNM based method. (A,B and C) Free energy curve (FEC) along the minimum free energy pathway (see Supplementary Figure A.8 (A,B and C)) for AdK-Ap5A, RBP-RIP and BGT-UDP systems. Metastable states (M_i) identified from FEC are along the apo to holo transition pathway are marked. C_α RMSD (D, E and F) and all-atom RMSD (G, H and I) comparison of identified metastable states with the intermediates predicted using our multi-step MDeNM based protocol for AdK-Ap5A, RBP-RIP and BGT-UDP systems.	212
A.10	Normal modes triggered by ligand in the intermediate state E_3 of AdK-Ap5A system. Direction of residue displacement along Mode 1 and 2 shown by blue arrows on protein (in cartoon representation) - ligand (in bond stick) complex.	213
A.11	Selection of principal components (PCs) for structural clustering of G-protein α subunit. (A) Ras domain (red) and α -H domain(blue) of G-protein α subunit (B) Protein C_α RMSD change corresponding to displacement of 1 standard deviation along a PC from the mean state. (C) Cumulative population distribution along top 5 PCs in 10 independent 100 ns (total 1 μ s) long unbiased MD simulations started from 10 different apo-state NMR conformations (PDB: 5JS7)[11]. (D) Direction of residue displacement along top 3 PCs shown by blue arrows on protein average structure (in green cartoon representation). (E) Evolution of distance between centre of mass of Ras domain and α -H domain (d_{COM}) in ten 100 ns long MD simulations. Values of d_{COM} corresponding to crystal structures 1CIP and 3SN6 are also shown (dashed lines).	214
A.12	Analysis of apo-state cluster centres of G-protein α subunit (A and B) C_α -RMSD matrix and TM-score matrix of a set of 9 cluster centres identified using robust density-based clustering method[12] and 10 apo-state NMR structures (PDB: 5JS7)[11] (C and D) Pair-wise C_α -RMSD and TM-score of a set of 9 cluster centres.	215
A.13	Protein-ligand interface area (A) during the 1 ns unrestrained NPT equilibration simulation of AdK-Ap5A (—), RBP-RIP (—) and BGT-UDP (—) protein-ligand complexes. Interface area (A) is calculated using equation: $A = (SASA_{protein} + SASA_{ligand} - SASA_{complex})/2$, where $SASA$ is solvent-accessible surface area calculated with solvent probe radius of 0.14 nm.	216

B.1	(A) Visualization of 5iA residues colored by the maximum strength of the harmonic spring among all residue-residue harmonic restraints involving each residue, with a color scale ranging from blue ($0 \text{ kJ mol}^{-1} \text{ nm}^{-2}$) through green to red ($1000 \text{ kJ mol}^{-1} \text{ nm}^{-2}$). (B) Scatter plot illustrating the relationship between the strength of inferred harmonic restraints ($\Gamma_{\text{extra},ij}$) and the distance between residue pairs (d_{ij}). (C) Pearson correlation (ρ) between root mean squared fluctuations (RMSF) of 5iA residues predicted by ENM with different cutoff radii (R_c) and those estimated from a 40 ns MD simulation.	239
B.2	(A) Blue arrows shows the direction of residue displacement for motion along Mode 1 (A), Mode 2 (B), and Mode 3 (C) on the 5iA structure (surface representation). (D) A diagram featuring two diametrically opposite 5iA units in cartoon representation to clarify the definition of the progress variable $\xi = R_{\gamma_A} - R_{\gamma_B}$; γ_A peptides are marked in red, while the rest of the structure is shown in green. (E) Change in the progress variable ξ , separately defined for analyzing the release of γ_A , γ_B , and γ_C peptides, calculated for a fixed displacement of $2 \text{ \AA}$ along the normal modes.	240
C.1	(A) All-atomic configuration of DMPC lipid-bilayer (yellow) in water environment (red), and (B) the corresponding coarse-grained configuration.	243
C.2	An illustrative diagram of 2-dimensional simulation box (shaded blue) and its 8 periodic images in direct contact (shaded white), along with the transformation relationships to generate these images from the coordinates of simulation box.	246
C.3	DMPC bilayer ENM parameters and friction coefficients along normal modes. (A) Pearson correlation (κ) between RMSF of lipid beads, averaged over all lipid molecules, calculated using MD simulations (Eq. C.52) and the ENM (Eq. C.53), the latter at several R_c and γ_{SE}/γ_{SS} , with $\gamma_{EE}/\gamma_{SS} = 10^{-5}$. (B) Variation of κ with γ_{EE}/γ_{SS} , with $\gamma_{SE}/\gamma_{SS} (= 1)$ and $R_c = 10 \text{ \AA}$ ($\bullet$) or $R_c = 14 \text{ \AA}$ ($\square$). (C) Static undulation spectra calculated using MD ($\circ$) and ENM for $\gamma_{SS}[\text{J mol}^{-1} \text{ nm}^{-2}] = 50$ ($\square$), 125 (∇) and 500 ($\bullet$), with $\gamma_{SE}/\gamma_{SS} = 1$, $\gamma_{EE}/\gamma_{SS} = 10^{-5}$, and $R_c = 14 \text{ \AA}$. (D) Mass-weighted friction coefficients ($\bullet$) along normal modes (Γ_m for mode m with frequency ω_m ; Eq. C.9). 95% confidence interval is shown as grey-shaded region. AA MD simulation of a 256 DMPC bilayer is used for calculation of all properties and ENM parameterization.	248

C.4	Undulatory normal modes of lipid-bilayer. Height displacement profile of DMPC lipid bilayer (1024 lipids) corresponding to selected undulatory normal modes of wave-vector: (A and B) $q = 0.35 \text{ nm}^{-1}$, (C) $q = 0.49 \text{ nm}^{-1}$ and (D) $q = 0.70 \text{ nm}^{-1}$. Height displacement profile is calculated over a square grid and height displacement at a grid-point is calculated as average displacement along bilayer normal, of all the atoms belonging to that grid-point.	252
C.5	ENM parametrization using undulations structure factor. (A) Relative root mean squared logarithmic error (RRMSLE) between MD-derived and ENM-derived undulations spectra of DMPC bilayer (Eq. C.55), for a range of R_c and γ_{ss} values. (B) Variation of optimal value of γ_{ss} (corresponding to minimum RRMSLE in Fig. C.5(A)) for different values of ENM-parameter R_c (-○-).	272
C.6	ENM parametrization using static structure factor (A-F) Static structure factor, $S(q)$, calculated using MD (—) and ENM (—) for different values of γ_{ss} (shown in plot window, units= $\text{J mol}^{-1} \text{ nm}^{-2}$), for fixed value of $R_c = 14 \text{ \AA}$ and $\gamma_{SE}/\gamma_{ss} = 1$. (G) Relative root mean squared error (RRMSE) between MD-derived and ENM-derived $S(q)$ (Eq. C.56), for a whole q-range (◊-) and for low-q only (-○-).	272
C.7	Normal modes of DMPC lipid-bilayer. (A) Normalized root mean squared fluctuations (nRMSF) of lipid beads in DMPC bilayer (defined in Eq. C.54), averaged over all lipid molecules, calculated from MD simulation (○) and from two different ENM models with $R_c = 10 \text{ \AA}$ (□) and $R_c = 14 \text{ \AA}$ (△). Inset shows the numbering of lipid beads in the coarse-grained description of DMPC lipid molecule. (B) Free-energy profile along selected undulatory normal modes ($q_1 = 0.35 \text{ nm}^{-1}$ and $q_2 = 0.70 \text{ nm}^{-1}$), calculated using ENM and estimated from statistics of a $1 \mu\text{s}$ long unbiased MD simulation. Error bars are shown as standard deviation of free energy profiles estimated along: modes 1–4 (corresponds to undulatory motion of wave-vector q_1), and modes 7–9 (corresponds to undulatory motion of wave-vector q_2). (C) Mode frequencies of slowest 20 normal modes, obtained using two different ENM models with $R_c = 10 \text{ \AA}$ (□) and $R_c = 14 \text{ \AA}$ (△). (D) DMPC lipid-bilayer (comprising of 1024 lipid molecules) motion corresponding to a set of selected modes. (E) Contribution of wave-vectors (q) to selected normal modes of DMPC bilayer (1024 lipids), calculated from the fourier transform of bilayer undulations profile corresponding to the mode displacement vector.	273

- C.8 Friction coefficients along normal modes.** (A) Histogram of mass-weighted friction coefficients of lipid beads. (B) Distribution of friction coefficient of each type of lipid bead. The inset shows the coarse-grained configuration of a DMPC lipid molecule, with bead indices on top of each bead. (C) Density profile of lipid and water molecules along the bilayer normal, where $z = 0$ corresponds to the bilayer centre. (D) Mass-weighted friction-coefficients sub-matrix for top 100 normal modes. 274
- C.9 Determination of bilayer material properties** (A) Static undulations spectra $I_u(q)$ (Eq. C.11) of DMPC bilayer (1024 lipids) estimated from all-atom MD and uL-NMS. Solid lines shows the best-fit of $I_u(q)$ by Eq. C.25, used to estimate the membrane bending (k_c) and tilt (k_θ) moduli. (B) Decay rate for undulation auto-correlation, $\tau_u(q)$ from all-atom MD and uL-NMS. Solid lines shows the fit of the data to equation $\tau_u(q) \sim q^{-b}$ to estimate the scaling coefficient for $\tau_u(q)$ dependence on wave-vector q . (C) Dynamic structure factor, $S(q = 0.70 \text{ nm}^{-1}, \omega)$ (Eq. C.29) from all-atom MD. Solid line shows the fit of $S(q, \omega)$ to the Rayleigh-Brillouin triplet model (Eq. C.39), used to estimate the width of the Brillouin and Rayleigh peaks. . . . 274
- C.10 Dynamic and intermediate structure factor.** (A–D) Dynamic structure factor of DMPC lipid-bilayer calculated from ENM (—) and from MD simulations (—) for wavevectors: (A) $q = 1 \text{ nm}^{-1}$, (B) $q = 2 \text{ nm}^{-1}$, (C) $q = 8 \text{ nm}^{-1}$ and (D) $q = 16 \text{ nm}^{-1}$. (E) Comparison of intermediate structure factors calculated from ENM (solid lines) and MD simulations (dashed lines) for different wavevectors described in figure legend. (F) Comparison of intermediate structure factors calculated from ENM (solid lines), scaled to match the MD simulations data (dashed lines) at time 1 ps to compare the long-time relaxations. 275
- C.11 Determination of DPPC lipid-bilayer ENM parameters** (A) Pearson correlation κ between root mean squared fluctuations (RMSF) of DPPC lipid beads, averaged over all lipid molecules, calculated from MD simulations (Eq. C.52) and from ENM (Eq. C.53) for a varying range of γ_{SE}/γ_{SS} values, with fixed values of $R_c = 14 \text{ \AA}$ and $\gamma_{EE}/\gamma_{SS} = 10^{-5}$. A similar variation of κ with γ_{SE}/γ_{SS} is obtained for other R_c values in the range $10 \text{ \AA}$ to $18 \text{ \AA}$ (data not shown), with highest correlation obtained for $R_c = 14 \text{ \AA}$. (B) Relative root mean squared logarithmic error (RRMSLE) between MD-derived and ENM-derived undulations spectra of DMPC bilayer (Eq. C.55), to determine optimal γ_{SS} value. The optimal set of ENM parameters thus obtained for DPPC bilayer is: $R_c = 14 \text{ \AA}$, $\gamma_{SS} = 0.9 \text{ kJ mol}^{-1} \text{ nm}^{-2}$, $\gamma_{SE}/\gamma_{SS} = 1.1$, and $\gamma_{EE}/\gamma_{SS} = 10^{-5}$ 275

- C.12 Elastic network model of membrane protein γ -secretase.** (A) Correlation (calculated as vector dot product) between the set of slowest 10 normal modes of γ -secretase calculated using ENM without including lipid-membrane environment i.e. protein ENM only (ENM-P) and calculated using ENM including lipid-membrane environment (ENM-PL) (B-C) Protein conformational variance in 200 ns all-atomic MD simulation ($\bullet$), calculated as $C\alpha$ mean-squared displacement, along slowest 100 modes calculated from (B) ENM-P ($\bullet$) and (C) ENM-PL ($\bullet$). (D) Cross-correlations map of γ -secretase residue fluctuations [$C_{ij} = \langle \Delta \mathbf{R}_i \cdot \Delta \mathbf{R}_j \rangle / (\langle \Delta \mathbf{R}_i^2 \rangle \langle \Delta \mathbf{R}_j^2 \rangle)^{1/2}$], calculated from ENM-P, ENM-PL and 200 ns MD simulation. 276
- C.13 Excited normal modes molecular dynamics (MDeNM).** (A) Progress along the DMPC bilayer pore formation order parameter, ξ , in 1 ns long MDeNM simulations at different excitation temperatures ΔT (B) Progress along ξ in multiple 500 ps MDeNM simulations, with random initial velocities but fixed excitation temperature ($\Delta T = 50$ K), all started from an initial flat bilayer configuration (C) Mean progress along ξ at the end of 500 ps MDeNM simulations at varying ΔT values (0 K to 100 K), with and without inclusion of friction effects in MDeNM excitation velocity determination. For each ΔT , mean progress is calculated from ten independent MDeNM simulations, and the standard deviation is shown as the shaded region. . 276
- C.14 Model application to obtain pore formation pathway in phospholipid bilayers.** (A) Progress along the pore-formation order parameter ξ in multi-step MDeNM simulations of DMPC ($\boxplus$) and DPPC ($\boxplus$) bilayers. (B) Free energy surface contours of DMPC bilayer in the space of PCVs $s(\mathbf{R})$ (Eq. C.44) and $z(\mathbf{R})$ (Eq. C.45) obtained from 1 μ s long well-tempered metadynamics (wt-metad) simulation and minimum free energy path (---) of pore formation obtained using nudged elastic band method. MD snapshots of a single member from the ensemble of flat-bilayer (left), transition-state (middle) and stable pore (right) are shown. (C) Pore formation free energy (ΔG_{pore}) profile as function of $s(\mathbf{R})$ for DMPC (—) and DPPC (—) bilayer, obtained from umbrella sampling. Shaded region shows the uncertainty in estimates of ΔG_{pore} from WHAM (Eq. C.49). 1-dimensional projection of DMPC free energy, obtained from wt-metad, along $s(\mathbf{R})$ is also shown for comparison (=). 277

- C.15 **PMF profile and hysteresis along the pore reaction coordinate in DPPC bilayer.** (A-B) Variation of order parameter ξ defined by Tolpekina et al. [13] and order parameter ξ_{MF} (number density of water molecule within a fixed membrane-spanning cylinder) defined by Mirjalili and Feig [14], in umbrella sampling windows restrained along $s(\mathbf{R})$, with starting configurations taken from both forward path (—■) and backward path (—●). (C) Variation of $z(\mathbf{R})$ in umbrella sampling windows with starting configurations taken from backward path (—○). (D) Pore formation free energy (ΔG_{pore}) profile along $s(\mathbf{R})$ obtained from umbrella sampling, using starting configurations from forward (—■) and backward (—●) paths. Shaded region shows the uncertainty in estimates of ΔG_{pore} from WHAM (Eq. C.49). 278
- C.16 **Lipid ordering in flat DMPC and DPPC bilayers** (A) Atomic numbering of sn1 and sn2 chains of DMPC and DPPC lipid molecules (B) Static undulations spectra of DMPC and DPPC lipid bilayers, calculated from 100 ns all-atomic MD simulations. (C) Order parameter S_{CH} (Eq. C.51) calculated from 100 ns MD simulation of flat DMPC and DPPC bilayer. 279
- C.17 **Structural and energetic analysis of pore formation** (A) Thermodynamic cycle used to estimate the free energy of pore formation ΔG_{pore} using MM/PBSA method. ΔG_{pore}^0 is the pore formation energy in the gas-phase, and ΔG_{solv}^{flat} and ΔG_{solv}^{pore} are the solvation free energy of lipid bilayer in flat and pore state respectively. (B) Mean lipid chain order (S_{CH}) dependence with lipid lateral distance from the pore centre. S_{CH} is calculated by averaging S_{CH} of Carbon atoms 4–10 of sn1 and sn2 chains (corresponds to the plateau region in Fig. C.16(C)). Further calculation details can be found in Ref. C.1.7.4. Dashed lines shows the S_{CH} values in the flat bilayer state. 280
- C.18 **Pore metastability in DMPC and DPPC lipid-bilayers.** (A–B) Path progress variable $s(\mathbf{R})$ in five independent MD simulations, with random initial velocities, started from a pore-state of (A) DMPC and (B) DPPC bilayer. For both bilayers, the last frame of the last umbrella window is taken as the the initial configuration in these simulations. (C–D) 10 ns moving average of change in the number of lipid-water (—), water-water (—) and total (—) hydrogen bonds, relative to the corresponding average value in flat-bilayer state, in the five independent MD simulations started from a pore-state of (C) DMPC and (D) DPPC bilayer. 281

C.19	Characterization of initial stages of pore formation in a DMPC bilayer. (A–E) Two-dimensional projection of $\hat{\mathbf{d}}_{\text{pore}}$, and (F–J) gradient of membrane thickness profile (Sec. C.1.7.1) along $\hat{\mathbf{d}}_{\text{pore}}$, for MDeNM intermediates I_0 , I_1 , I_2 , I_3 and I_4 respectively. MDeNM intermediate I_0 corresponds to flat-bilayer, and I_4 is the last MDeNM intermediate obtained before water channel formation. (K–O) Membrane thickness profile ($\langle t_{XY} \rangle$, Sec. C.1.7.1), and (P–T) profiles of membrane normal projection along a vector from membrane surface to pore-centre ($\langle \hat{\mathbf{n}}_{XY} \cdot \hat{\mathbf{r}}_{XY} \rangle$, Sec. C.1.7.2), averaged over 1 ns simulation trajectory for intermediates I_0 to I_4 respectively. (U) Dot product between two-dimensional projection of $\hat{\mathbf{d}}_{\text{pore}}$ calculated for intermediates I_0 to I_4 . (V) Variation of order parameters $\langle \hat{\mathbf{n}}_{XY} \cdot \hat{\mathbf{r}}_{XY} \rangle_{\text{cyl}}$ (Sec. C.1.7.2; $\bullet$) and $f_{\text{LP-water}}$ (Sec. C.1.7.3; $\diamond$) with the mean bilayer thickness within the pore-spanning cylindrical region $\langle t_{XY} \rangle_{\text{cyl}}$ (Sec. C.1.7.1). (W) Variation of average coordination number ($\langle n_{\text{LP-X}} \rangle_{\text{cyl}}$; details in Sec. C.1.7.3) of a lipid head-group polar atom (within the pore-spanning cylindrical region) with water oxygens ($\blacktriangle$, X = water), with all inter-molecular lipid polar atoms ($\blacktriangledown$, X = lipid), and all inter-molecular polar atoms ($\bullet$, X = {lipid, water}) with $s(\mathbf{R})$	282
D.1	Ligand RMSD relative to the initial state in MD simulations performed with charges determined using CGenFF, DFT, and QM/MM for (A) M1001 and (B) PT2385.	286
D.2	Kinetic model for hydrogen/deuterium exchange. Conformational shifts expose internal amide groups (blue) (closed state) to the solvent (open state), allowing the exchange of amide hydrogens (white) with deuterium (yellow) at an intrinsic rate constant k_{int} . Figure adapted from Ref. [15].	287
D.3	(A) Distribution of HIF-2 PAS-B cavity volume and $\text{C}\alpha$ RMSD in a 1.5 μs apo-state simulation trajectory. The volume and RMSD of cluster centers obtained from pocket-shape based clustering of the trajectory (circles) and experimental holo-state crystal structures (stars) are marked. (B–C) ROC curves obtained from ranking ligands and decoys based on increasing $\Delta\Delta G$ for antagonists (B) and decreasing $\Delta\Delta G$ for agonists (C).	288

- D.4 (A) Definition of θ_{M252} that captures the displacement of HIF-2 residue M252 from inside the PAS-B pocket towards the ARNT subunit upon antagonist binding, weakening heterodimerization. It is the angle made by the lines joining atom $C\alpha$ and $C\epsilon$ of M252 and atom $C\alpha$ and the pocket center. The pocket center is the geometric center of $C\alpha$ positions of HIF-2 residues: 246, 252, 277, 278, 280, 281, 289, 292, 293, 296, 302, 303, 304, 309, 321, and 322. (B) Definition of ϕ_{Y281} , that captures the displacement of HIF-2 residue Y281 from inside the PAS-B pocket towards the ARNT subunit upon agonist binding, where it forms a hydrogen bond with ARNT residue Y456, strengthening heterodimerization. It is defined as the dihedral angle by atoms N, $C\alpha$, $C\beta$, and $C\gamma$ of HIF-2 residue Y281. (C-D) Distribution of hydrogen-acceptor distance d_{HA} (C) and donor-hydrogen-acceptor angle θ_{DHA} (D) to assess hydrogen bond formation between HIF-2 residue Y281 and ARNT residue Y456. (E) Fraction of trajectories (out of 50) started from the same initial configuration (ligand docked to apo-state structure) in which residue M252 is displaced outwards upon antagonist binding. (F) Fraction of trajectories (out of 50) in which residue Y281 is displaced outwards upon agonist binding. 289
- D.5 (A) Residue $C\alpha$ root mean squared fluctuations (RMSF) profile of the HIF-2/ARNT complex in the apo-state. The mean RMSF value of each residue is calculated from three independent 500 ns MD trajectories. Residue indices 1-365 correspond to ARNT residues 100-464, and residue indices 366-698 correspond to HIF-2 residues 28-360. (B) Changes in the average RMSF profile upon ligand binding. (C) Histogram of MM/PBSA interaction energy between HIF-2 and ARNT for apo, PT2385-bound, and M1001-bound states. 290
- D.6 (A) Crystal structure of the HIF-2/ARNT complex with hypoxia response element (HRE) DNA, obtained from PDB accession code 4ZPK [16]. (B) Mean $C\alpha$ RMSD of DNA-binding bHLH domains relative to the 4ZPK structure, computed from three independent replicas of 500 ns MD trajectories for apo, PT2385-bound (holo_antag), and M1001-bound (holo_ag) states. The shaded region represents standard deviations. (C) Distribution of $C\alpha$ RMSD of DNA-binding bHLH domains relative to the 4ZPK structure for apo, active ligand-bound, and decoy-bound states. 290

List of Tables

4.1	Comparison of single-step and multi-step protocols. Two structural metrics are reported for protein structures obtained for the three protein–ligand systems: C_α -RMSD and GDT_{avg} with respect to the experimental holo-structure. Both single and multi-step approaches were initiated from the protein-ligand complex obtained from a short, unbiased MD simulation, I_1 . The change in protein-ligand binding free energy during structural evolution is reported as $\Delta\Delta G$, calculated with respect to I_1 (see equation 3.21). The structure with largest (most negative) $\Delta\Delta G$ is taken as the predicted holo structure (values in bold).	66
6.1	Bilayer properties from uL-NMS are compared with the MD simulation and the experimental data. The bending (k_c) and tilt moduli (k_θ) are estimated from $I_u(q)$ (Eq. C.25), and thermal diffusivity (D_T) from $\Gamma_h(q)$ (Fig. 6.1(D)). Errors ($\pm$) are reported as the 95% confidence interval for the fitted parameters k_c and k_θ	104
6.2	Dominant modes contributing to $\hat{\mathbf{d}}_{pore}$ (Eq. 3.25) in the flat-bilayer state of DMPC bilayer. Mode indices (m) are listed in decreasing order of their contribution (α_m^2) to the $\hat{\mathbf{d}}_{pore}$ norm. $\sum_i (U_{m,3i})^2$ quantifies the proportion of the vibrational motion in mode m that occurs along the z-axis.	110
A.1	Experimental and LRT-predicted (calculated using equation A.16) magnitude of protein structural change upon ligand binding, measured as protein C_α RMSD, for the three protein-ligand systems i.e. AdK–Ap5A, RBP–RIP and BGT–UDP.	216
A.2	Pearson correlation between D_P (structure displacement along $\mathbf{d}_{LRT}$) and structure improvement (measured as decrease in RMSD with experimental holo structure) for the three protein-ligand systems in state I_1 perturbed along $\mathbf{d}_{LRT}$ using MDeNM-based velocity excitation protocol	217

A.3	Comparison of predicted holo-states of G-protein α subunit with experimental holo-state	
	Holo-states of G-protein α subunit are predicted using multi-step MDeNM approach, where excitation simulations are started from 9 distinct conformations ($C_1, C_2, \dots, C_9$), obtained by ligand-docking on 9 apo-state cluster centres followed by short MD relaxation of protein-ligand complex. One energetic metric ($\Delta G_{\text{bind}}^{\text{mm/pbsa}}$ [kJ mol^{-1}]: protein-ligand mm/pbsa binding energy averaged over 50 ps trajectory) and eight structural metrics (protein-RMSD [$\text{\AA}$]: C α RMSD of protein structure w.r.t. experimental holo-state (PDB: 1CIP); protein TM-score: TM-score of protein w.r.t. experimental holo-state; ligand-RMSD [$\text{\AA}$]: all heavy-atom RMSD of ligand in protein-ligand complex w.r.t. ligand in experimental holo-state, calculated after optimal alignment of protein structures; $f_{\text{R}}^{\text{contact}}$: fraction of protein residues in contact with the ligand out of total 18 residues in contact in experimental holo-state; domain-RMSD (Ras domain [$\text{\AA}$], α -H domain [$\text{\AA}$]): domain-wise RMSD w.r.t. respective domains in the experimental holo-state; d_{COM} [$\text{\AA}$]: distance between centre of mass of Ras domain and α -H domain; d_{PC} [nm]: distance of protein conformation from holo-state in the sub-space of top 3 principal components; RMSD^{avg} [$\text{\AA}$]: mean protein-RMSD w.r.t initial nine apo-state cluster centres) are reported for protein-ligand complexes: (before) before starting MDeNM excitation simulations, (after) predicted as holo-state using our multi-step MDeNM approach, and (10 ns) obtained after 10 ns long unbiased MD relaxation of predicted holo-state.	218
C.1	Energy decomposition analysis of pore formation in phospholipid bilayers.	
	$\Delta\langle E_{\text{MM}} \rangle$ is the change in average potential energy on pore formation in DMPC and DPPC lipid bilayer comprising 256 lipid molecules. It can be decomposed as contributions from (a) change in average bonded energy including bonds, angles and dihedrals energy ($\Delta\langle E_{\text{bonded}} \rangle$), (b) change in average Lennard-Jones potential energy of intramolecular 1–4 pairs ($\Delta\langle E_{\text{LJ-14}} \rangle$), (c) change in average Lennard-Jones potential energy of rest of the atomic pairs ($\Delta\langle E_{\text{LJ}} \rangle$), (d) change in average Coulomb potential energy of intramolecular 1–4 pairs ($\Delta\langle E_{\text{Coulomb-14}} \rangle$), (e) change in average short-range and long-range Coulomb potential energy excluding intramolecular 1–4 pairs ($\Delta\langle E_{\text{Coulomb-short}} \rangle$ and $\Delta\langle E_{\text{Coulomb-long}} \rangle$ respectively). Further, $\Delta\langle E_{\text{LJ-14}} \rangle$ and $\Delta\langle E_{\text{Coulomb-short}} \rangle$ are decomposed into contributions from lipid-lipid (LL), lipid-solvent (LS) and solvent-solvent (SS) interactions.	283

D.1	Ligands known to experimentally bind the HIF-2 α PAS-B domain. M1001 and M1002 are weak agonists, while the remaining ligands are antagonists.	291
-----	---	-----

Abbreviations

AA	All Atomic
ANM	Anisotropic Network Model
ARNT	Aryl hydrocarbon Receptor Nuclear Translocator
CG	Coarse Grained
CV	Collective Variable
D-MPNN	Direct Message Passing Neural Network
ENM	Elastic Network Model
FEL	Free Energy Landscape
FHV	Flock House Virus
HIF	Hypoxia Inducible Factor
LRT	Linear Response Theory
MD	Molecular Dynamics
MDeNM	Molecular Dynamics with excited Normal Modes
MFEP	Minimum Free Energy Pathway
NMA	Normal Mode Analysis
PCV	Path Collective Variable
RC	Reaction Coordinate
RMSD	Root Mean Squared Deviation
RMSF	Root Mean Squared Fluctuations
uL-NMS	uncoupled Langevin equations in the Normal Modes Space