12-2021

Deciphering Chemomechanical Couplings in Proteins Using Microsecond-level Molecular Dynamics Simulations

Vivek Govind Kumar
University of Arkansas, Fayetteville

Deciphering Chemomechanical Couplings in Proteins Using
Microsecond-level Molecular Dynamics Simulations

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy in Cell and Molecular Biology

by

Vivek Govind Kumar
McGill University
Bachelor of Science in Biochemistry, 2013
University of Bradford
Master of Science in Cancer Pharmacology, 2015

December 2021
University of Arkansas

This dissertation is approved for recommendation to the Graduate Council.

_____
Mahmoud Moradi, Ph.D.
Dissertation Director

_____
Chenguang Fan, Ph.D.
Committee Member

_____
Shilpa Iyer, Ph.D.
Committee Member

_____
Suresh Thallapuranam, Ph.D.
Committee Member

ABSTRACT

All-atom molecular dynamics (MD) simulations combine the high temporal resolution of experimental methods like smFRET and spatial resolution of methods like x-ray crystallography, to provide a detailed dynamic picture of biomolecular processes. Here, microsecond-level atomistic MD simulations have been used to characterize chemomechanical couplings in human fibroblast growth factor 1 (hFGF1) and the spike proteins of SARS CoV-1 and SARS-CoV-2. hFGF1 is a globular signaling protein that is involved in several physiological processes ranging from cell proliferation to wound healing. Experimental studies have previously described the low proteolytic and thermal stability of hFGF1, in addition to the stabilizing role of heparin. Here, a conformational change in the hFGF1 heparin-binding pocket that occurs only when heparin is absent, is described for the first time. Comparisons with experimental data indicate that this conformational transition is implicated in the low thermal stability of hFGF1. Unique electrostatic interactions that contribute to heparin-mediated stabilization are also described. This work also describes a novel binding affinity estimation approach involving restrained umbrella sampling simulations. The absolute binding affinity for the hFGF1-heparin interaction determined using this approach is in very good agreement with data from isothermal titration calorimetry (ITC) experiments. This binding affinity study revealed that restraining ligand orientation is essential for effective sampling along a protein-ligand distance collective variable. The differential dynamic behavior of the SARS-CoV-1 and CoV-2 spike proteins is also described in this work. Spike protein activation is the first step in the “effective binding” process leading to interaction with the human ACE2 receptor. This study shows that the active form of the CoV-1 spike protein is less stable than that of the CoV-2 spike protein and that the energy

barriers associated with activation and inactivation are higher in CoV-2. A “pseudo-inactive” state of the CoV-1 spike protein is described for the first time, wherein the N-terminal domain (NTD) interacts with the receptor-binding domain (RBD). This highlights the potential role of the NTD in spike protein inactivation. The relatively slower kinetics of spike protein activation and inactivation in CoV-2 indicate that it might spend more time bound to the ACE2 receptor than CoV-1, which in turn might provide an explanation for the higher transmissibility of CoV-2.

ACKNOWLEDGEMENTS

I would first like to thank my supervisor, Dr. Mahmoud Moradi, for his invaluable support over the course of my PhD program. I am extremely grateful for his insightful guidance on the overall direction of my research and the formulation of my research projects. I would also like to thank my committee member and research collaborator, Dr. Suresh Kumar Thallapuranam, for his excellent advice and guidance on the hFGF1 projects. In addition, I extend my thanks to my committee members, Dr. Chenguang Fan and Dr. Shilpa Iyer, for their help and support during my program.

I thank my colleagues from the Biomolecular Simulations Group (Dylan Ogden, Adithya Polasa, Ugochi Isu, James Losey, Mortaza Molayousefi, Matthew Brownd and Shadi Badiee) for their friendship, advice and support over the past 4 years. I also thank Shilpi Agrawal for providing the experimental data for the hFGF1 projects.

I thank Dr. Ragupathy Kannan for pointing me towards the University of Arkansas and for his friendship and support during my studies. Finally, I would like to thank my parents and my sister for their unconditional love and support, without which I would have given up a long time ago.

TABLE OF CONTENTS

LIST OF PUBLISHED/SUBMITTED PAPERS

1. Chapter 1 – **Govind Kumar, V.**; Agrawal, S.; Kumar, T. K. S.; Moradi, M. Mechanistic Picture for Monomeric Human Fibroblast Growth Factor 1 Stabilization by Heparin Binding. *J. Phys. Chem. B* **2021**. https://pubs.acs.org/doi/10.1021/acs.jpcb.1c07772.

2. Chapter 2 - **Govind Kumar, V**.; Agrawal, S.; Suresh Kumar, T. K.; Moradi, M. Binding Affinity Estimation From Restrained Umbrella Sampling Simulations. *bioRxiv* **2021**, 2021.10.28.466324. https://doi.org/10.1101/2021.10.28.466324.

3. Chapter 3 – **Govind Kumar, V.**; Ogden, D. S.; Isu, U. H.; Polasa, A.; Losey, J.; Moradi, M. Differential Dynamic Behavior of Prefusion Spike Proteins of SARS Coronaviruses 1 and 2. *bioRxiv* **2020,** 2020.12.25.424008. https://doi.org/10.1101/2020.12.25.424008. Submitted to Journal of Biological Chemistry.

INTRODUCTION

Due to advances in force field development, high-throughput modeling techniques and supercomputing power, atomistic molecular dynamics (MD) simulation has become a standard technique for investigating the behavior of proteins in explicit aqueous and membrane environments. Experimental methods such as cryogenic electron microscopy (cryo-EM) and X-ray crystallography essentially capture static images of specific conformational states which are not necessarily functionally relevant[1-3]. The effectiveness of all-atom MD arises from the fact that it is currently the only method capable of providing a dynamic picture of biomolecular processes based on a combination of high spatio-temporal resolution and atomic-level detail[3,4]. Most MD-based studies make the assumption that local conformational changes observed in short, nanosecond-level simulations can be used to accurately describe functionally relevant large-scale conformational transitions occurring on much longer timescales[5,6]. A recent study has shown that longer simulations at the microsecond-level are the minimum requirement for an accurate and effective characterization of both large-scale and local conformational transitions[7].

This work focuses on the use of microsecond-level MD simulations to investigate the conformational transitions associated with chemomechanical couplings in proteins, namely human fibroblast growth factor 1 (hFGF1) and the spike proteins of SARS CoV-1 and SARS-CoV-2. hFGF1 is a small globular signaling protein that is involved in numerous physiological processes such as cell proliferation, wound healing, tumor growth and neurogenesis[8,9]. The interaction of hFGF1 with heparin, its glycosaminoglycan (GAG) binding partner, is of particular interest. While studies have shown that heparin binding does not cause any global conformational transitions in hFGF1, it is thought to stabilize the protein and impart protection

against proteolysis[10-13]. Microsecond-level unbiased MD simulations described in this work reveal that the heparin-binding pocket of hFGF1 undergoes a conformational change when heparin is absent. Based on comparisons with data from equilibrium unfolding experiments, it is proposed that this conformational transition is implicated in the enhanced thermal instability[14,15] displayed by unbound hFGF1. This study also describes the intermolecular and intramolecular electrostatic interactions that contribute to the conformational dynamics of the hFGF1-heparin complex.

Binding affinity calculations are an important aspect of any experimental or computational study aiming to characterize protein-ligand interactions. Several computational approaches have been developed to calculate binding affinities for biomolecular[16-20]. Extensive sampling of ligand translational and rotational movements with respect to the protein, in addition to ligand conformational dynamics, is a critical factor in the accurate quantification of the entropic reduction caused by ligand binding[21-23]. A major disadvantage of popular methods like Molecular Mechanics/Poisson-Boltzmann-Surface Area (MM-PBSA) is that they do not treat entropic contributions to the binding free energy rigorously[21,24]. In addition, adequate sampling of the overall conformational landscape of a system using traditional unrestrained MD simulations is essentially impossible due to the timescales involved[25,26]. Computational approaches that involve the calculation of a potential of mean force (PMF) along biased, discretized reaction coordinates also tend towards inadequate sampling of factors like ligand orientation[16,27]. Woo and Roux have developed an approach that revolves around the use well-defined geometrical restraints on the orientation and conformation of both protein and ligand[16,27]. This method improves PMF convergence via a reduction in the conformational entropy of the system[16,27]. In this work, a purely physics-based approach involving restrained umbrella

sampling simulations, which is somewhat similar to the strategy devised by Woo and Roux[16,27], is described. The absolute binding affinity for the heparin-hGF1 interaction estimated via this approach is in very good agreement with the binding affinity determined from isothermal titration calorimetry (ITC) experiments.

As discussed in this work, hFGF1 undergoes a conformational transition and becomes unstable in the absence of its binding partner (heparin hexasaccharide). In this case, the presence of a binding partner is necessary to prevent a conformational change. At the other end of the spectrum, the CoV-1 and CoV-2 spike proteins must undergo major conformational transitions prior to interaction with the human ACE2 receptor. Several studies have shown that CoV-2 has a higher transmissibility than the very similar CoV-1[28-34], but the mechanistic basis for this differential transmissibility remains largely unexplored. The spike protein plays a critically important role in the infection process of these two viruses, wherein the receptor-binding domain (RBD) transitions from an inactive "down" position to an active "up" position prior to binding the human ACE2 receptor[28,33,35-41]. Several studies have focused on the endpoint of this transition (i.e. RBD-ACE2 binding) while mostly ignoring the mechanistic aspects that drive it. This work investigates the potential contributions of this large-scale conformational transition to the differential transmissibility and infectivity of CoV-1 and CoV-2. Extensive microsecond-level unbiased MD simulations described in this work show that the active CoV-2 spike protein is more stable than the active CoV-1 spike protein. In these simulations, the RBD of the active CoV-1 spike protein spontaneously transitions to a "pseudo-inactive" state where it interacts with the N-terminal domain (NTD). This conformation has not been observed in any cryo-EM or X-ray crystallography studies thus far. Steered MD (SMD) simulations[42] and nonequilibrium work calculations[43] have also been used to study conformational dynamics inaccessible to the shorter

timescales of the unbiased simulations. These simulations strongly indicate that the energy barriers for the confirmational transitions involved in activation or inactivation are lower for the CoV-1 spike protein and that the CoV-2 spike protein undergoes conformational changes relatively slowly. The slower kinetics of the CoV-2 spike protein provides an explanation for why it might remain bound to ACE2 longer than the CoV-1 spike protein. This in turn might potentially be connected to the higher transmissibility of CoV-2.

REFERENCES

1. Punjani, A.; Fleet, D. J. 3D Variability Analysis: Resolving Continuous Flexibility and Discrete Heterogeneity from Single Particle Cryo-EM. *bioRxiv* **2021**.https://doi.org/10.1101/2020.04.08.032466.

2. Torrens-Fontanals, M.; Stepniewski, T. M.; Aranda-García, D.; Morales-Pastor, A.; Medel-Lacruz, B.; Selent, J. How Do Molecular Dynamics Data Complement Static Structural Data of GPCRs. *Int. J. Mol. Sci*. **2020**, *21*, 5933.

3. Srivastava, A.; Nagai, T.; Srivastava, A.; Miyashita, O.; Tama, F. Role of Computational Methods in Going beyond X-Ray Crystallography to Explore Protein Structure and Dynamics. *Int. J. Mol. Sci.* **2018**, *19*, 3401.

4. Dror, R. O.; Jensen, M. Ø.; Borhani, D. W.; Shaw, D. E. Exploring Atomic Resolution Physiology on a Femtosecond to Millisecond Timescale Using Molecular Dynamics Simulations. *J. Gen. Physiol.* **2010**, *135*, 555-562.

5. Aduri, N. G.; Prabhala, B. K.; Ernst, H. A.; Jørgensen, F. S.; Olsen, L.; Mirza, O. Salt Bridge Swapping in the EXXERFXYY Motif of Proton-Coupled Oligopeptide Transporters. *J. Biol. Chem.* **2015**, *290*, 29931– 29940.

6. Doki, S.; Kato, H.E.; Solcan, N.; Iwaki, M.; Koyama, M.; Hattori, M.; Iwase, N.; Tsukazaki, T.; Sugita, Y.; Kandori, H. et al. Structural Basis for Dynamic Mechanism of

Proton-Coupled Symport by the Peptide Transporter POT. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 11343– 11348.

7. Immadisetty, K.; Hettige, J.; Moradi, M. What Can and Cannot Be Learned from Molecular Dynamics Simulations of Bacterial Proton-Coupled Oligopeptide Transporter GkPOT? *J. Phys. Chem. B* **2017**, *121*, 3644-3656.

8. Ornitz, D. M.; Herr, A. B.; Nilsson, M.; Westman, J.; Svahn, C. M.; Waksman, G. FGF Binding and FGF Receptor Activation by Synthetic Heparan-Derived Di- and Trisaccharides. *Science* **1995**, *268*, 432-436.

9. Culajay, J. F.; Blaber, S. I.; Khurana, A.; Blaber, M. Thermodynamic Characterization of Mutants of Human Fibroblast Growth Factor 1 with an Increased Physiological Half-Life. *Biochemistry* **2000**, *39*, 7153-7158.

10. DiGabriele, A. D.; Lax, I.; Chen, D. I.; Svahn, C. M.; Jaye, M.; Schlessinger, J.; Hendrickson, W. A. Structure of a Heparin-Linked Biologically Active Dimer of Fibroblast Growth Factor. *Nature* **1998**, *393*, 812-817.

11. Zhu, X. T.; Hsu, B. T.; Rees, D. C. Structural Studies of The Binding of The Antiulcer Drug Sucrose Octasulfate to Acidic Fibroblast Growth-Factor. *Structure* **1993***, 1,* 27-34.

12. Carter, E. P.; Fearon, A. E.; Grose, R. P. Careless Talk Costs Lives: Fibroblast Growth Factor Receptor Signalling and the Consequences of Pathway Malfunction. *Trends Cell Biol*. **2015,** *25*, 221-233.

13. Goetz, R.; Mohammadi, M. Exploring Mechanisms of FGF Signalling through the Lens of Structural Biology. *Nat. Rev. Mol. Cell Biol*. **2013**, *14*, 166-180.

14. Copeland, R. A.; Ji, H.; Halfpenny, A. J.; Williams, R. W.; Thompson, K. C.; Herber, W. K.; Thomas, K. A.; Bruner, M. W.; Ryan, J. A.; Marquis-Omer, D. et al. The Structure of Human Acidic Fibroblast Growth Factor and Its Interaction with Heparin. *Arch. Biochem. Biophys.* **1991**, *289*, 53-61.

15. Blaber, M.; Adamek, D. H.; Popovic, A.; Blaber, S. I. Biophysical and Structural Analysis of Human Acidic Fibroblast Growth Factor. *Tech. Protein Chem.* **1997**, *8*, 745-753.

16. 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*, 6825-6830.

17. Ytreberg, F.M.; Zuckerman, D.M. Simple estimation of absolute free energies for biomolecules. *J. Chem. Phys.* **2006**, *124*, 104105.

18. Wang, J.; Deng, Y.; Roux, B. Absolute binding free energy calculations using molecular dynamics simulations with restraining potentials. *Biophys. J.* **2006**, *91*, 2798–2814.

19. Doudou, S.; Burton, N.A.; Henchman, R.H. Standard free energy of binding from a one-dimensional potential of mean force. *J. Chem. Theory Comput.* **2009**, *5*, 909–918.

20. Deng, Y.; Roux, B. Computations of standard binding free energies with molecular dynamics simulations. *J. Phys. Chem. B.* **2009**, *113*, 2234–2246.

21. Fu, H.; Gumbart, J. C.; Chen, H.; Shao, X.; Cai, W.; Chipot, C. BFEE: A User-Friendly Graphical Interface Facilitating Absolute Binding Free-Energy Calculations. *J. Chem. Inf. Model*. **2018**, *58*, 556–560.

22. Chipot, C. Frontiers in free-energy calculations of biological systems. In *Wiley Interdisciplinary Reviews: Computational Molecular Science*, **2014**, *4*, 71– 89.

23. Chodera, J. D.; Mobley, D. L. Entropy-enthalpy compensation: Role and ramifications in biomolecular ligand recognition and design. *Annu. Rev. Biophys*. **2013**, *42*, 121– 142.

24. Wang, C. H.; Nguyen, P. H.; Pham, K.; Huynh, D.; Le, T. B. N.; Wang, H. L. et al. Calculating protein-ligand binding affinities with MMPBSA: method and error analysis. *J. Comput. Chem.* **2016**, *37*, 2436–2446.

25. van Gunsteren, W.F.; Mark, A.E. Validation of molecular dynamics simulation. *J. Chem. Phys*. **1998**, *108*, 6109–6116.

26. van Gunsteren, W.F.; Dolenc, J.; Mark, A.E. Molecular simulation as an aid to experimentalists. *Curr. Opin. Struct. Biol.* **2008**, *18*, 149–153.

27. Gumbart, J.C.; Roux, B.; Chipot, C. Standard binding free energies from computer simulations: What is the best strategy? *J.* Chem. *Theory Comput.* **2013**, *9***,** 794–802.

28. Walls, A. C.; Park, Y. J.; Tortorici, M. A.; Wall, A.; McGuire, A. T.; Veesler, D. Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell* **2020**, *181*, 281 – 292.e6.

29. Wu, K.; Peng, G.; Wilken, M.; Geraghty, R. J.; Li, F. Mechanisms of Host Receptor Adaptation by Severe Acute Respiratory Syndrome Coronavirus. *J. Biol. Chem.* **2012**, *287*, 8904–8911.

30. Petersen, E.; Koopmans, M.; Go, U.; Hamer, D. H.; Petrosillo, N.; Castelli, F.; Storgaard, M.; Al Khalili, S.; Simonsen, L. Comparing SARS-CoV-2 with SARS-CoV and Influenza Pandemics. *Lancet Infect. Dis*. 2020, *20*, e238–e244.

31. Ferretti, L.; Wymant, C.; Kendall, M.; Zhao, L.; Nurtay, A.; Abeler-Dörner, L.; Parker, M.; Bonsall, D.; Fraser, C. Quantifying SARS-CoV-2 Transmission Suggests Epidemic Control with Digital Contact Tracing. *Science* **2020**, *368*, eabb6936.

32. Zhou, P.; Yang, X. Lou; Wang, X. G.; Hu, B.; Zhang, L.; Zhang, W.; Si, H. R.; Zhu, Y.; Li, B.; Huang, C. L. et al. A Pneumonia Outbreak Associated with a New Coronavirus of Probable Bat Origin. *Nature* **2020**, *579*, 270–273.

33. Tortorici, M. A.; Veesler, D. Structural Insights into Coronavirus Entry. In *Adv. Virus Res.* 2019, *105*, 93–116.

34. Walls, A. C.; Tortorici, M. A.; Snijder, J.; Xiong, X.; Bosch, B. J.; Rey, F. A.; Veesler, D. Tectonic Conformational Changes of a Coronavirus Spike Glycoprotein Promote Membrane Fusion. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114*, 11157–11162.

35. Walls, A. C.; Xiong, X.; Park, Y. J.; Tortorici, M. A.; Snijder, J.; Quispe, J.; Cameroni, E.; Gopal, R.; Dai, M.; Lanzavecchia, A.; Zambon, M.; Rey, F. A.; Corti, D.; Veesler, D. Unexpected Receptor Functional Mimicry Elucidates Activation of Coronavirus Fusion. *Cell* **2019**, *176*, 1026–1039.e15.

36. Ou, X.; Liu, Y.; Lei, X.; Li, P.; Mi, D.; Ren, L.; Guo, L.; Guo, R.; Chen, T.; Hu, J.; Xiang, Z.; Mu, Z.; Chen, X.; Chen, J.; Hu, K.; Jin, Q.; Wang, J.; Qian, Z. Characterization of Spike Glycoprotein of SARS-CoV-2 on Virus Entry and Its Immune Cross-Reactivity with SARS-CoV. *Nat. Commun.* **2020**, *11*, 1620.

37. Tai, W.; He, L.; Zhang, X.; Pu, J.; Voronin, D.; Jiang, S.; Zhou, Y.; Du, L. Characterization of the Receptor-Binding Domain (RBD) of 2019 Novel Coronavirus: Implication for Development of RBD Protein as a Viral Attachment Inhibitor and Vaccine. *Cell. Mol. Immunol.* **2020**, *17*, 613–620.

38. Wrapp, D.; Wang, N.; Corbett, K. S.; Goldsmith, J. A.; Hsieh, C. L.; Abiona, O.; Graham, B. S.; McLellan, J. S. Cryo-EM Structure of the 2019-NCoV Spike in the Prefusion Conformation. *Science* **2020**, *367*, 1260–1263.

39. Yuan, Y.; Cao, D.; Zhang, Y.; Ma, J.; Qi, J.; Wang, Q.; Lu, G.; Wu, Y.; Yan, J.; Shi, Y.; Zhang, X.; Gao, G. F. Cryo-EM Structures of MERS-CoV and SARS-CoV Spike Glycoproteins Reveal the Dynamic Receptor Binding Domains. *Nat. Commun.* **2017**, *8*, 15092.

40. Zhou, T.; Tsybovsky, Y.; Gorman, J.; Rapp, M.; Cerutti, G.; Chuang, G. Y.; Katsamba, P. S.; Sampson, J. M.; Schön, A.; Bimela, J. et al. Cryo-EM Structures of SARS-CoV-2 Spike without and with ACE2 Reveal a PH-Dependent Switch to Mediate Endosomal Positioning of Receptor-Binding Domains. *Cell Host Microbe* **2020**, *28*, 867–879.e5.

41. Cai, Y.; Zhang, J.; Xiao, T.; Peng, H.; Sterling, S. M.; Walsh, R. M.; Rawson, S.; Rits-Volloch, S.; Chen, B. Distinct Conformational States of SARS-CoV-2 Spike Protein. *Science* **2020**, *369*, 1586–1592.

42. Schlitter, J.; Engels, M.; Krüger, P. Targeted Molecular Dynamics: A New Approach for Searching Pathways of Conformational Transitions. *J. Mol. Graph.* **1994**, *12*, 84–89.

43. Jarzynski, C. Nonequilibrium equality for free energy differences. *Phys. Rev. Lett*. **1997**, *78*, 2690–2693.

Mechanistic Picture for Monomeric Human Fibroblast Growth Factor 1 Stabilization by Heparin Binding

Vivek Govind Kumar[1], Shilpi Agrawal[1], Thallapuranam Krishnaswamy Suresh Kumar[1], and Mahmoud Moradi[1]

[1] Department of Chemistry and Biochemistry, University of Arkansas, Fayetteville, AR 72701, United States of America

ABSTRACT

Human fibroblast growth factor (FGF) 1 or hFGF1 is a member of the FGF family that is involved in various vital processes such as cell proliferation, cell differentiation, angiogenesis and wound healing. hFGF1, which is associated with low stability *in vivo,* is known to be stabilized by binding heparin sulfate, a glycosaminoglycan that aids the protein in the activation of its cell surface receptor. The poor thermal and proteolytic stability of hFGF1 and the stabilizing role of heparin have long been observed experimentally; however, the mechanistic details of these phenomena are not well understood. Here, we have used microsecond-level equilibrium molecular dynamics (MD) simulations to quantitatively characterize the structural dynamics of monomeric hFGF1 in the presence and absence of heparin hexasaccharide. We have observed a conformational change in the heparin-binding pocket of hFGF1 that occurs only in the absence of heparin. Several intramolecular interactions were also identified within the heparin-binding pocket, that form only when hFGF1 interacts with heparin. The loss of both intermolecular and intramolecular interactions in the absence of heparin plausibly leads to the observed conformational change. This conformational transition results in increased flexibility of the heparin-binding pocket and provides an explanation for the susceptibility of *apo* hFGF1 to

proteolytic degradation and thermal instability. This study provides a glimpse into mechanistic details of heparin-mediated stabilization of hFGF1 and encourages the use of microsecond-level MD in studying the effect of binding on protein structure and dynamics. In addition, the observed differential behavior of hFGF1 in the absence and presence of heparin provides an example, where microsecond-level all-atom MD simulations are necessary to see functionally relevant biomolecular phenomena that otherwise will not be observed on sub-microsecond timescales.

INTRODUCTION

Thanks to the ever-increasing power of computers, improved force fields, and high-throughput modeling, all-atom MD is now routinely used to simulate proteins in simplified but explicit aqueous/membrane environments. MD simulations combine the high spatial resolution of experimental methods such as X-ray crystallography with the high temporal resolution of experimental methods such as single-molecule FRET spectroscopy[1,2]. However, many MD studies implicitly assume that local conformational transitions observed in short, nanosecond-level simulations can be used to describe global protein conformational transitions that typically occur on microsecond or millisecond time scales[3,4]. We have recently demonstrated that longer microsecond-level simulations are essential for a more precise statistical characterization of both local and global conformational transitions[5]. Here, we use microsecond-level unbiased MD simulations to investigate the conformational and structural dynamics of monomeric hFGF1[6] and the chemo-mechanical coupling between hFGF1 and heparin, its glycosaminoglycan (GAG) binding partner.

Fibroblast growth factors (FGFs) are signaling proteins that are involved in an extensive variety of physiological processes[7,8,9]. The biological activity of FGFs is regulated through interactions with linear anionic polysaccharides called glycosaminoglycans (GAGs), which facilitate binding to specific receptors on the cell surface (FGFRs)[10,11,12,13,14,15,16,17]. Human acidic fibroblast growth factor (hFGF1) is an important signaling molecule expressed in embryonic and adult tissues for angiogenesis, cell proliferation and differentiation, tumor growth, neurogenesis and wound healing[10,18]. Glycosaminoglycans (GAGs) consist of a class of negatively charged and large linear polysaccharides formed of repeating disaccharide units in which a uronic acid (either glucuronic acid or iduronic acid) moiety is combined with an amino sugar (either *N-acetyl*-D-glucosamine or *N-acetyl*-D-galactosamine)[19,20]. Heparin is a GAG made up of 2-O-sulfated iduronic acid and 6-O-sulfated, N-sulfated glucosamine (IdoA(2S)-GlcNS(6S)), connected by α-(1→4) glycosidic linkages[21]. The anionic nature of GAGs leads to electrostatic interactions with positively charged (Lysine/Arginine-rich) regions of their target proteins[19,20]. The hFGF1-heparin complex is the most broadly studied protein-GAG complex[22,23].

The interaction of hFGF1 with specific heparin sulfate proteoglycans may be influenced by the flexibility of the heparin-binding pocket[24]. In addition to the structural features of hFGF1, GAG sulfation patterns also determine the functionality and specificity of protein-GAG interactions[25,26]. hFGF1 is known to selectively recognize the GlcNS-IdoA2S-GlcNS sulfation motif[27]. DiGabriele *et al*.[28] crystallized a dimeric hFGF1-heparin sandwich complex (PDB entry: 2AXM) and showed that heparin binding does not result in any global conformational changes within hFGF1[28,29]. Solution nuclear magnetic resonance (NMR) and experimental binding studies suggest that a monomeric hFGF1-heparin complex is also fully functional[23,30]. *Apo* hFGF1 shows relatively low thermal stability and is known to be susceptible to thermal

degradation[31,32]. Binding to heparin sulfate proteoglycans is thought to protect hFGF1 against proteolytic degradation[33,34].

Our microsecond-level all-atom equilibrium MD simulations reveal that a conformational change occurs in the heparin-binding pocket of hFGF1 in the absence of heparin. We postulate that this conformational change is responsible for the susceptibility of unbound hFGF1 to thermal instability, as seen in equilibrium unfolding experiments. We have also studied the intermolecular interactions of the hFGF1-heparin complex and the intramolecular interactions that are unique to heparin-bound hFGF1 in order to obtain a clearer picture of the heparin-mediated stabilization.

METHODS

Equilibrium unfolding of hFGF1 with heparin hexasaccharide

The temperature-based denaturation experiment was performed using the JASCO-1500 Circular dichroism spectrophotometer cohered with fluorescence detector. hFGF1 was diluted with 10 mM phosphate buffer containing 100 mM NaCl at pH 7.2, to get a concentration of 33μM. The experiment was performed with and without heparin. For the measurements with heparin, protein to heparin ratio of 1:10 was used. The fluorescence spectra were collected in 5 °C intervals from 25 °C to 90 °C. The fraction of denatured protein (F_d) at each temperature was determined as $F_d = (Y - Y_N)/(Y_D - Y_N)$; where, Y, Y_N, and Y_D are the fluorescence signals of the 305/350 nm fluorescence ratio at the native state (25 °C), each consecutive temperature, and the denatured state (90 °C) respectively. The data set was fit using MS Excel. T_m, the temperature at

which 50 % of the protein molecules exist in the denatured state(s), was calculated from the fraction denatured protein population versus temperature graph.

All-atom equilibrium MD simulations

We have used all-atom equilibrium MD simulations to characterize the conformational dynamics of hFGF1 with and without heparin hexasaccharide. Our simulations were based on the x-ray crystal structures of the unbound hFGF1 monomer (PDB: 1RG8, resolution: 1.1 angstroms)[6] and the dimeric complex with a heparin hexasaccharide (PDB:2AXM, resolution: 3.0 angstroms)[28]. We built three different models – monomeric *apo* hFGF1 from 1RG8; monomeric heparin-bound hFGF1 (1RG8) using the heparin hexasaccharide from the dimeric complex (2AXM) (Model 1) and monomeric heparin-bound hFGF1 from the dimeric complex (2AXM) (Model 2). Residues 12-137 in the PDB files correspond to residues 26-151 in the experimental sequence. The experiments were performed using a truncated version of hFGF1 (residues 13-154) which did not contain the unstructured 12 amino acid N-terminal segment. The unstructured N-terminal segment is not known to be involved in receptor activation or heparin binding. The heparin hexasaccharide consists of N, O6 disulfo-glucosamine and 2-O-sulfo-alpha-L-idopyranuronic acid repeats[28].

MD simulations were performed using the NAMD 2.13[35] simulation package with the CHARMM36 all-atom additive force field[36]. The input files for energy minimization and production were generated using CHARMM-GUI[37,38]. For heparin-bound Model 1, the heparin hexasaccharide segment from 2AXM was added to the 1RG8 structure using psfgen. The models were then solvated in a box of TIP3P waters and 0.15 M NaCl. The heparin-bound systems had approximately 23000 atoms while the *apo* system had 27,000 atoms.

Initially, we energy-minimized each system for 10,000 steps using the conjugate gradient algorithm[39]. Subsequently, we relaxed the systems using restrained MD simulations in a stepwise manner (for a total of ~1 ns) using the standard CHARMM-GUI protocol[37]. The initial relaxation was performed in an NVT ensemble while all production runs were performed in an NPT ensemble. Simulations were carried out using a 2-fs time step at 300 K using a Langevin integrator with a damping coefficient of $\gamma = 0.5\ ps^{-1}$. The pressure was maintained at 1 atm using the Nosé−Hoover Langevin piston method[39,40]. The smoothed cutoff distance for non- bonded interactions was set to 10−12 Å and long-range electrostatic interactions were computed with the particle mesh Ewald (PME) method[41]. The initial production run for each model lasted 15 nanoseconds, in which the conformations were collected every 2 ps. After each model was equilibrated for 15 ns, the production runs were extended on the supercomputer Anton 2 (Pittsburgh Supercomputing Center) for 4.8 μs each, with a timestep of 2.5 fs. Conformations were collected every 240 picoseconds.

VMD[42] was used to analyze the simulation trajectories. The RMSD Trajectory tool[42] was used to calculate the RMSD and C_α atoms were considered for these calculations. For internal RMSD, the region of interest was aligned against its own initial configuration and RMSD was calculated with respect to this configuration. RMSF of individual residues was calculated using the C_α atoms by aligning the trajectory against the crystal structure. The HBond[42] and Salt Bridge[42] plugins were used to generate the data for hydrogen bonding and salt-bridge analysis respectively. For all interactions of interest, the number of frames with 1 or more hydrogen bonds was counted to get the occupancy percentage. An occupancy cutoff of 50%, a donor-acceptor distance cutoff of 4 Å and an angle cutoff of 35° were used to define hydrogen

bond/salt bridge interactions. The salt bridge plugin[42] was used to calculate the distance between the two salt bridge residues over the course of the simulation, which is the distance between the oxygen atom of the participating acidic residue and the nitrogen atom of the basic residue. The Timeline plugin[42] was used to analyze protein secondary structure. An internal measurement method in VMD was used to count the number of water molecules within 3 Å of the heparin-binding pocket[42].

RESULTS AND DISCUSSION

The putative role of heparin is to prevent the degradation of hFGF1. However, the specifics of this heparin-mediated stabilization are still unclear. To address this issue, we have used microsecond-level all-atom MD to compare and characterize the *apo* and heparin-bound forms of hFGF1.

We have performed three unbiased all-atom MD simulations of monomeric hFGF1, each for 4.8 μs. One *apo* and two heparin-bound models were simulated in the presence of explicit water. The *apo* model and one of the heparin-bound models (Model 1) are based on the crystal structure of monomeric *apo* hFGF1 (PDB entry: 1RG8)[6]. In order to examine the reproducibility of our results, we have also made a second heparin-bound model of hFGF1. The second heparin-bound model is extracted as a monomeric model from the crystal structure of dimeric heparin-bound hFGF1 (PDB entry: 2AXM)[28] (Supplemental Figure S1). Both heparin-bound models use a heparin hexasaccharide. The heparin-bound models are quite similar and involve an hFGF1 monomer bound to heparin hexasaccharide. As the only difference between the *apo* and heparin-bound models is the presence or absence of heparin, we can make meaningful comparisons between all three sets of simulations (i.e., two *holo* and one *apo* models).

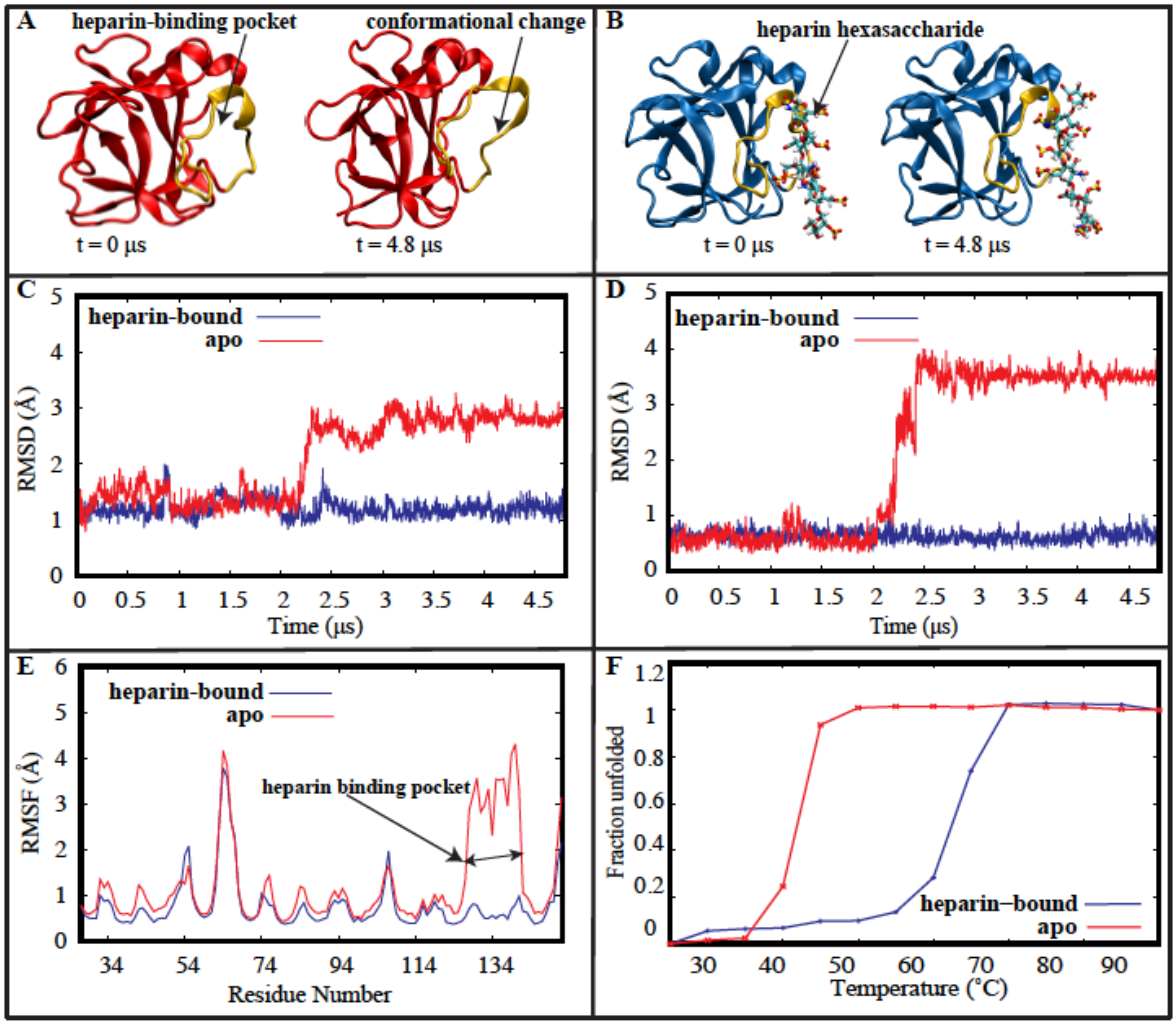


Figure 1. Conformational change in the heparin-binding pocket of *apo* hFGF1. (A,B) Cartoon representation of *apo* (red) and heparin-bound (blue) hFGF1 at the beginning and end of the 4.8-μs simulations. The heparin-binding pocket (gold) moves away from the beta-trefoil core of the *apo* protein. (C,D) RMSD time series for the *apo* (red) and heparin-bound (blue) models of hFGF1 protein (C) and its heparin binding pocket (D). (E) RMSF estimations for the *apo* (red) and heparin-bound (blue) models of hFGF1. (F) Thermal denaturation data for hFGF1 in the absence (red) and presence (blue) of heparin. The presence of heparin causes the T_m value to increase by around 20°C, indicating that heparin stabilizes the protein.

A conformational change occurs in the heparin-binding pocket of the apo model

The most noticeable observation in our simulations is that the heparin-binding pocket (residues 126-142) of the *apo* model becomes elongated and extends further outward and away from the core beta-trefoil structure after approximately 2 μs (Figure 1A). This conformational change is not observed in either of the two heparin-bound models (Figure 1B, Supplemental Figure S2A). Comparing the internal root mean square deviation (RMSD) of the hFGF1 monomer from each system reveals that the presence of heparin hexasaccharide stabilizes the protein and prevents this conformational change from occurring (Figure 1C, Supplemental Figures S2A-B). All 3 models initially have internal RMSD values of approximately 1 Å from their initial conformations, indicating little flexibility at least within the first 2 μs of simulations. Both heparin-bound models settle down into a stable conformation within 0.5 μs (RMSD=1.5 to 2 Å approx.) (Figure 1C, Supplemental Figures S2A-B). On the other hand, the *apo* model clearly undergoes a conformational change after 2 μs (RMSD=3 Å approx.) (Figure 1C, 1A). This new conformation then remains stable for the remainder of the simulation (around 2.8 μs) (Figure 1C, 1A).

A comparison of the internal RMSD of the heparin-binding pocket reveals that this region plays a key role in the differential behavior of the *apo* (RMSD≈4 Å) and heparin-bound models (RMSD≈0.5 Å) (Figure 1D, Supplemental Figures S2C-D). This indicates that the absence of interactions with heparin leads to the decreased stability of the *apo* model. These results are also supported by the root mean square fluctuation (RMSF) data for each model, which was calculated for the C_α atoms of all protein residues (Figure 1E, Supplemental Figure S2E-F). All three models show similar trends in the fluctuations for different regions, with the

exception of the heparin-binding pocket. As expected, the heparin-binding pocket is much more flexible in the *apo* model than in the heparin-bound models.

Thermal denaturation experiments were performed on monomeric hFGF1, in the absence and presence of heparin hexasaccharide, to further validate our computational results. The T_m value for the *apo* experimental model was approximately 42°C while the T_m value for the heparin-bound experimental model was approximately 62.5°C (Figure 1F). The presence of heparin thus increases the T_m value by around 20°C, indicating that heparin stabilizes the protein. These observations are in qualitative agreement with the computational RMSD/RMSF data.

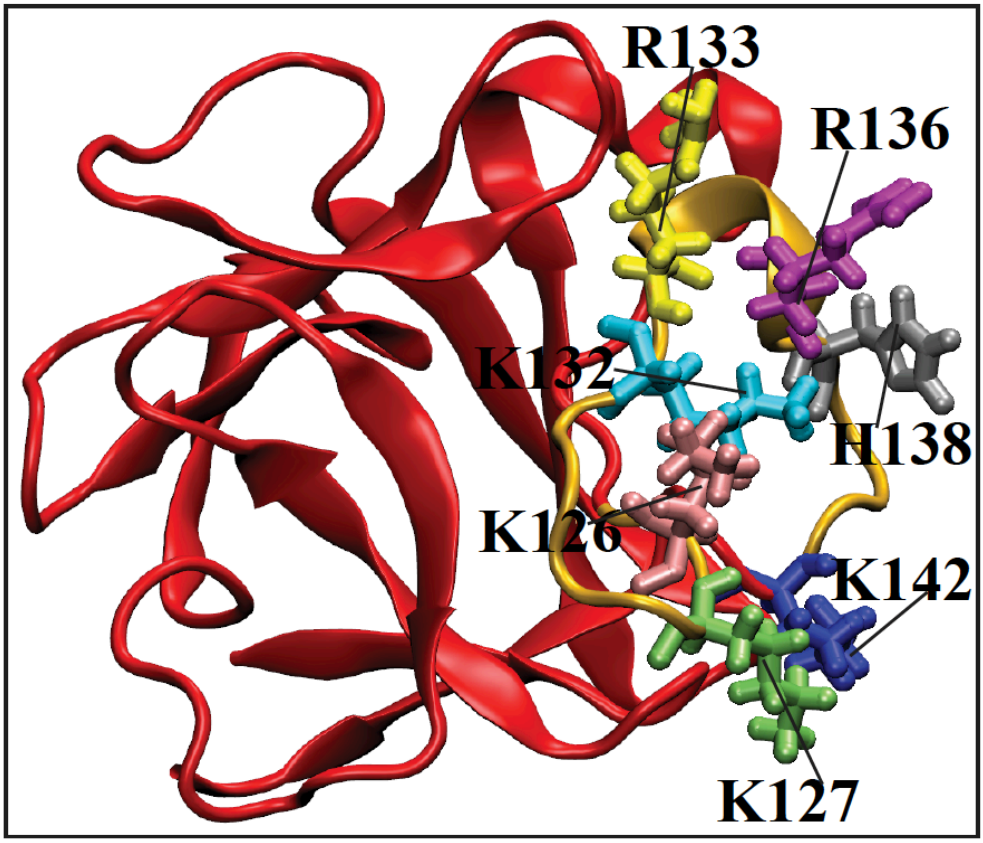
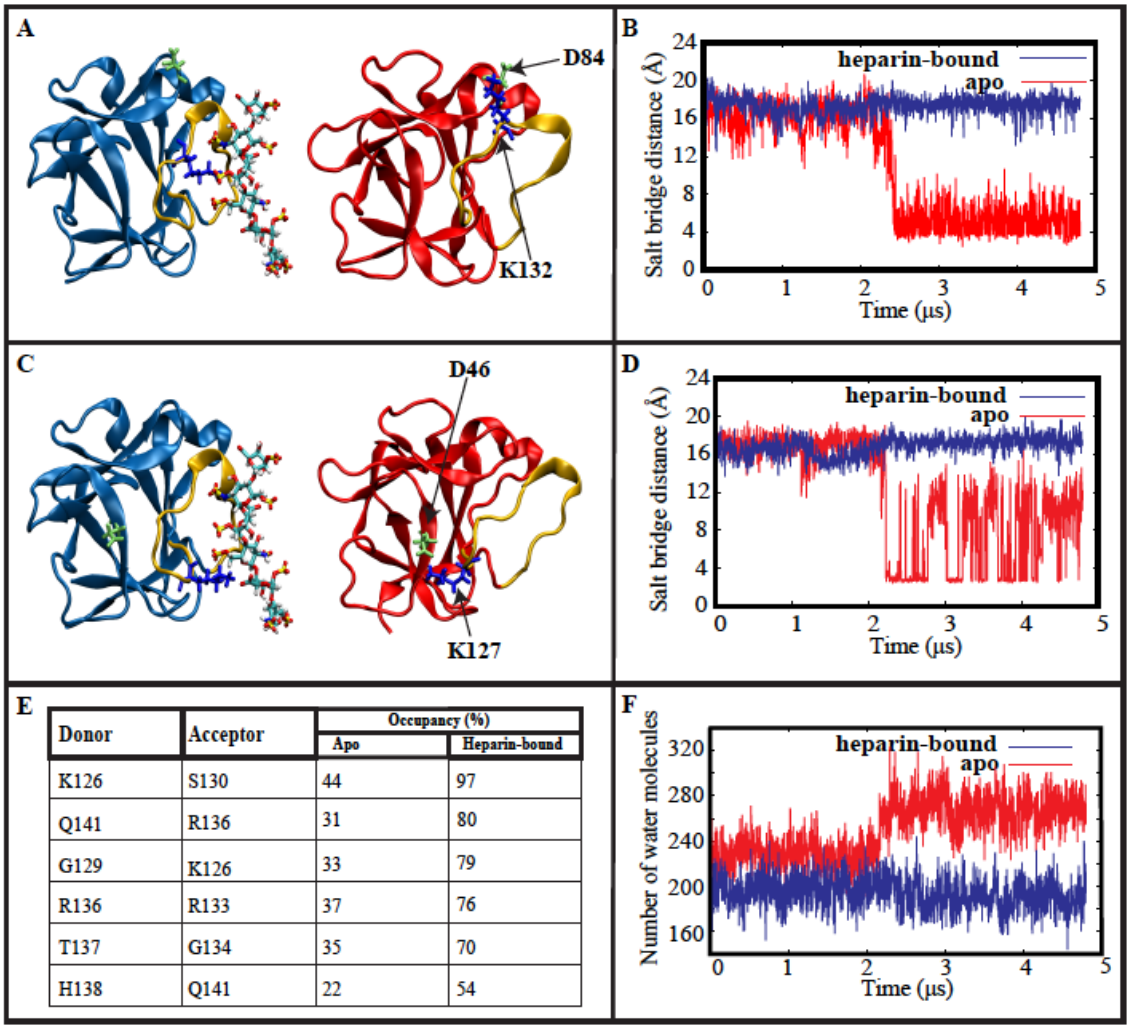


Figure 2. Cartoon representation of *apo* hFGF1. The positively charged residues of the heparin-binding pocket (gold) are shown using stick representation. These residues are involved in both intramolecular and intermolecular electrostatic interactions.



Donor	Acceptor	Occupancy (%)	
		Apo	Heparin-bound
K126	S130	44	97
Q141	R136	31	80
G129	K126	33	79
R136	R133	37	76
T137	G134	35	70
H138	Q141	22	54

Figure 3. Unique salt-bridge interactions facilitate the conformational change in the *apo* model. (A) K132 (blue) of the heparin-binding pocket (gold) forms a salt-bridge with D84 of the beta-trefoil core in the *apo* model (red). This interaction does not form in the heparin-bound protein (blue). (B) Time series of the D84-K132 donor-acceptor salt bridge distances in the presence (blue) and absence (red) of heparin. (C) K127 (blue) of the heparin-binding pocket (gold) forms a weak salt-bridge with D46 of the beta-trefoil core in the *apo* model (red). This interaction does not form in the heparin-bound Model 1 (blue). (D) Time series of the D46-K127 donor-acceptor salt bridge distances in the presence (blue) and absence (red) of heparin. (E) Table of intramolecular interactions unique to the heparin-binding pocket of heparin-bound hFGF1. (F) Time series of water molecule count within 3 Å of the heparin-binding pocket.

Unique salt-bridge interactions facilitate the conformational change in the apo model

The conformational change that occurs in the *apo* model is localized in the heparin-binding pocket. Electrostatic interactions between positively charged residues in the heparin-binding pocket (Figure 2) and negatively charged residues in the beta-trefoil core help stabilize the new conformation. We have identified two salt bridge interactions that are unique to the *apo* model. They do not form in the two heparin-bound models (Figure 3A-D, Supplemental Figure S3A-D). D84 of the beta-trefoil core interacts with K132 of the heparin-binding pocket (Figure 2, Figure 3A-B), while D46 of the beta-trefoil core interacts with K127 of the heparin-binding pocket (Figure 2, Figure 3C-D). The destabilization of the heparin-binding pocket is accompanied with the formation of a weak salt bridge between D46 and K127 (Figure 3C-D), followed by the formation of a stronger salt bridge between D84 and K132 when the heparin-binding pocket becomes elongated and is extended outward (Figure 3A-B) and away from the beta-trefoil core. Both K127 and K132 are known to interact with negatively charged heparin residues[30]. Interactions with negatively charged residues of the beta-trefoil core possibly compensate for the absence of interactions with heparin. Together, these salt bridges play a key role in stabilizing the new conformation of the heparin-binding pocket for almost 2.8 μs.

Hydration analysis of the heparin-binding pocket (quantified by the number of water molecules within 3 Å of this domain) provides additional evidence for a conformational change within the heparin-binding pocket. Around 200 water molecules are present throughout both heparin-bound trajectories (Figure 3F, Supplemental Figure S3E-F). During the *apo* simulation, however, the count of water molecules increases to 280 after 2 μs (Figure 3F), thus coinciding with the observed conformational change.

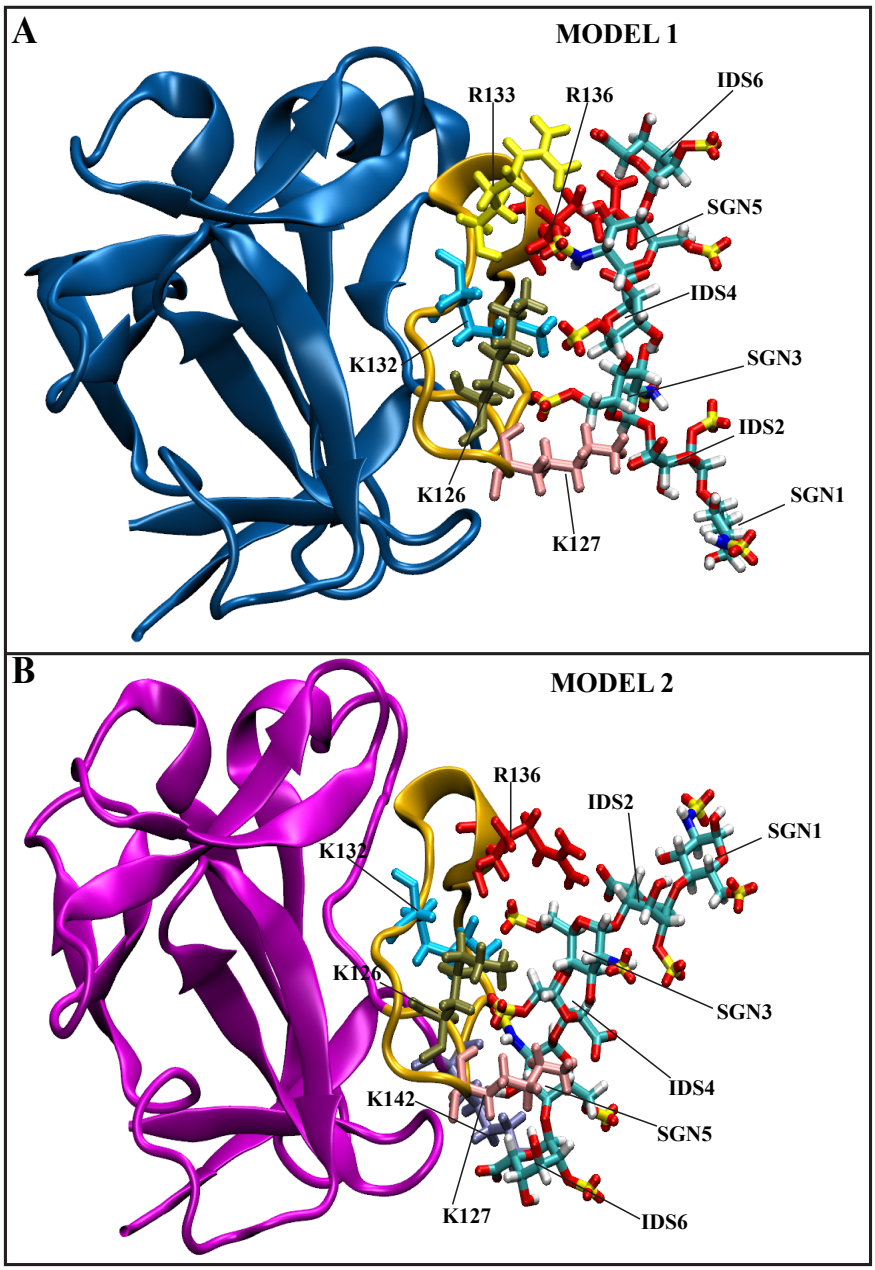


Figure 4. Cartoon representation of the final frames of the two heparin-bound trajectories. (A) Model 1 (blue) – heparin hexasaccharide from PDB entry 2AXM with monomeric hFGF1 from PDB entry 1RG8. (B) Model 2 (magenta) – one monomer and heparin hexasaccharide from 2AXM (dimeric). 6 residues in the heparin-binding pocket (R136, K132, K126, K127, R133, K142) were found to interact with heparin hexasaccharide.

Intramolecular interactions within the heparin-binding pocket help stabilize heparin-bound hFGF1

Thus far, we have shown that monomeric hFGF1 is destabilized in the absence of heparin and that a conformational change occurs within the heparin-binding pocket. This conformational change does not occur in the heparin-bound models, which are considerably more stable than the *apo* model. We have identified several unique intramolecular interactions within the heparin-binding pocket that contribute to the increased stability of the heparin-bound hFGF1 models (Figure 3E, Supplemental Figure S4). Hydrogen bond occupancies are quite similar in both Model 1 and Model 2 (Figure 3E, Supplemental Figure S4). While these interactions are also present briefly in the *apo* model, none of them meet the occupancy criteria that would allow them to be described as hydrogen bonds (Figure 3E). We propose that these intramolecular interactions within the protein may form as a consequence of intermolecular interactions between positively charged residues of the heparin-binding pocket (Figure 2) and negatively charged residues of heparin hexasaccharide. The strength of these intramolecular interactions (occupancies between 54-97% in Model 1 and 64-94% in Model 2) might thus be a factor that prevents the conformational change observed in the *apo* model from occurring in the heparin-bound models. Among intramolecular hydrogen bonds observed in the *apo* model, only one (L145-K142) involves the heparin-binding pocket (occupancy: 84%). All of the interactions observed in the *apo* model—including salt bridges D46-R38 (90%), D53-R38 (87%), and E67-K114 (60%)—also occur in the heparin-bound models with similar occupancies.

Secondary structure analysis reveals that parts of the heparin-binding pocket of the *apo* model become unstructured and unravel into random coils when the conformational

change occurs (Supplemental Figure S5A-B). This change in the secondary structure is then maintained for the remaining 2.8 μs of the *apo* trajectory. This change in the secondary structure is not observed in the heparin-bound models (Supplemental Figure S5C-D). The lack of strong intramolecular interactions in the heparin-binding pocket of the *apo* model (Figure 3E) could thus account for the observed changes in the secondary structure.

Our findings are further validated by internal RMSD analysis of the heparin-binding pocket of the *apo* (RMSD of ~4 Å) and heparin-bound (RMSD of ~0.5 Å) models (Figure 1D, Supplemental Figures S2C-D). This analysis demonstrates that the heparin-binding pockets of *apo* and heparin-bound hFGF1 have different internal conformations. Therefore, these observations confirm the role of heparin-derived intramolecular interactions in maintaining and promoting the structured nature of the heparin-binding pocket.

Characterization of intermolecular interactions that contribute to the stabilizing effects of heparin

The heparin hexasaccharide in Model 1 fluctuates considerably before it eventually undergoes a 180° rotation to settle down into a more stable conformation (Supplemental Figure S6A). This transition occurs at the 1.25 μs mark and continues until the 2 μs mark (Supplemental Figure S6A). On the other hand, the heparin molecule in Model 2 does not undergo any major positional changes and attains a stable conformation very quickly (Supplemental Figure S6B). As a result of the differences in behavior and position of the heparin hexasaccharide in each model, slightly different intermolecular hydrogen bond interactions occur in each model in terms of both occupancy as well as the residues involved (Figure 4, Table 1). Six residues in the heparin-binding pocket (R136, K132, K126, K127, R133, K142) were found to be involved in

these interactions (Figure 4A-B). With the exception of R133, intermolecular hydrogen bonds involving these residues are present in the dimeric crystal structure (PDB entry: 2AXM)[28]. R133 was found to interact with heparin only in Model 1 (Figure 4A), while K142 was found to interact with heparin only in Model 2 (Figure 4B). Intermolecular interactions involving N32, N128 and Q141 are also present in the dimeric crystal structure[28] but these residues only interact briefly (hydrogen bond occupancies < 35%) with heparin in our simulation trajectories.

Occupancies are fairly similar for interactions involving R136 and K126 in both models, while they are somewhat different for interactions involving residues K132 and K127. R133 and K142 only interact with heparin hexasaccharide in Models 1 and 2, respectively. See Table 1 and Fig. S7 for more details. As discussed previously, we have also identified six major intramolecular interactions within the heparin-binding pocket that are unique to heparin-bound hFGF1 (Figure 3E, Supplemental Figure S4). The presence of heparin ostensibly leads to the formation of these intramolecular hydrogen bonds, which consequently contribute to the stabilization of heparin-bound hFGF1. This is consistent with the thermal denaturation experiments described above, where the T_m value increases by around 20°C upon heparin binding (Fig. 1F), indicating an increase in the strength of protein intramolecular interactions upon heparin binding.

Table 1. Characterization of the hFGF1-heparin intermolecular interactions in the heparin binding pocket. Table of intermolecular hydrogen-bonding interactions observed in the last microsecond of both heparin-bound trajectories. R133 interacts with heparin only in Model 1, while K142 interacts with heparin only in Model 2.

Donor	Acceptor	
	Model 1	Model 2
R136	IDS4 (78%)	IDS2 (91%)
K126	SGN5 (68%) IDS4 (51%)	SGN3 (75%) IDS4 (72%)
K132	SGN3 (53%)	SGN3 (66%) IDS4 (52%)
K127	SGN3 (54%) IDS2 (61%)	IDS4 (75%)
R133	IDS6 (57%) SGN5 (64%)	none
K142	none	SGN5 (85%) IDS6 (57%)

Conclusions

In this study, we used microsecond-level MD simulations to compare the behavior of hFGF1 in the absence and presence of heparin hexasaccharide at the molecular level. These simulations reveal a significant conformational difference within the heparin-binding pocket in the absence and presence of the ligand. We conclude that the conformational change observed in

the heparin-binding pocket of the hFGF1 model in the absence of the heparin is directly linked to the thermal instability displayed by unbound monomeric hFGF1 experimentally. In addition to the intermolecular interactions between hFGF1 and heparin hexasaccharide, we have identified several intramolecular interactions within the heparin-binding pocket that are unique to the heparin-bound models. Thermal denaturation experiments have revealed that the T_m value for hFGF1 increases by approximately 20°C when bound to heparin. This suggests that the intramolecular interactions play a key role in stabilizing monomeric hFGF1. Further experimental and computational research is needed to elucidate the functional relevance of these specific intramolecular interactions.

SUPPORTING INFORMATION

Supporting Figures S1-S7 provide the results of additional analyses. Cartoon representation of dimeric hFGF1 with heparin; Stability of heparin-bound hFGF1 – RMSD/RMSF; Salt-bridge analysis for heparin-bound models; Intramolecular interactions in the heparin-binding pocket of heparin-bound Model 2; Secondary structure analysis related to conformational change in *apo* model; Behavior of heparin in both heparin-bound models; Time series of hFGF1-heparin interactions.

ACKNOWLEDGEMENTS

This research is supported by National Science Foundation grant CHE 1945465 and OAC 1940188 and the Arkansas Biosciences Institute. Anton 2 computer time was provided by the Pittsburgh Supercomputing Center (PSC) through Grant R01GM116961 from the National

Institutes of Health. The Anton 2 machine at PSC was generously made available by D.E. Shaw Research. This research is also part of the Blue Waters sustained-petascale computing project, which is supported by the National Science Foundation (awards OCI-0725070 and ACI-1238993) and the state of Illinois. This work also used the Extreme Science and Engineering Discovery Environment (allocation MCB150129), which is supported by National Science Foundation grant number ACI-1548562. This research is also supported by the Arkansas High Performance Computing Center, which is funded through multiple National Science Foundation grants and the Arkansas Economic Development Commission. This work is also supported by the Department of Energy (DE-FG02-01ER15161), the National Institutes of Health/National Cancer Institute (NIH/NCI) (1 RO1 CA 172631) and the NIH through the COBRE program (P30 GM103450).

AUTHOR CONTRIBUTIONS

M.M. and T.K.S.K. designed the research. V.G.K performed the simulations and analyzed the simulation data. S.A. performed the experiments and analyzed the experimental data. V.G.K, M.M., T.K.S.K., S.A. wrote the manuscript.

COMPETING INTERESTS

The authors declare no competing interests.

AVAILABILITY OF DATA

All data will be shared upon request to corresponding author.

REFERENCES

1. Kim, J. Y.; Kim, C.; Lee, N. K. Real-Time Submillisecond Single-Molecule FRET Dynamics of Freely Diffusing Molecules with Liposome Tethering. *Nat. Commun.* **2015**, *6*, 6992.

2. Srivastava, A.; Nagai, T.; Srivastava, A.; Miyashita, O.; Tama, F. Role of Computational Methods in Going beyond X-Ray Crystallography to Explore Protein Structure and Dynamics. *Int. J. Mol. Sci.* **2018**, *19*, 3401.

3. Aduri, N. G.; Prabhala, B. K.; Ernst, H. A.; Jørgensen, F. S.; Olsen, L.; Mirza, O. Salt Bridge Swapping in the EXXERFXYY Motif of Proton-Coupled Oligopeptide Transporters. *J. Biol. Chem.* **2015**, *290*, 29931– 29940.

4. Doki, S.; Kato, H.E.; Solcan, N.; Iwaki, M.; Koyama, M.; Hattori, M.; Iwase, N.; Tsukazaki, T.; Sugita, Y.; Kandori, H. et al. Structural Basis for Dynamic Mechanism of Proton-Coupled Symport by the Peptide Transporter POT. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 11343– 11348.

5. Immadisetty, K.; Hettige, J.; Moradi, M. What Can and Cannot Be Learned from Molecular Dynamics Simulations of Bacterial Proton-Coupled Oligopeptide Transporter GkPOT? *J. Phys. Chem. B* **2017**, *121*, 3644-3656.

6. Bernett, M. J.; Somasundaram, T.; Blaber, M. An Atomic Resolution Structure for Human Fibroblast Growth Factor 1. *Proteins* **2004**, *57*, 626-634.

7. Eswarakumar, V. P.; Lax, I.; Schlessinger, J. Cellular Signaling by Fibroblast Growth Factor Receptors. *Cytokine Growth Factor Rev.* **2005**, *16*, 139-149.

8. Beenken, A.; Mohammadi, M. The FGF Family: Biology, Pathophysiology and Therapy. *Nat. Rev. Drug Discovery* **2009**, *8*, 235-253.

9. Kuro-o, M. Endocrine FGFs and Klothos: Emerging Concepts. *Trends Endocrinol. Metab.* **2008**, *19*, 239-245.

10. Ornitz, D. M.; Herr, A. B.; Nilsson, M.; Westman, J.; Svahn, C. M.; Waksman, G. FGF Binding and FGF Receptor Activation by Synthetic Heparan-Derived Di- and Trisaccharides. *Science* **1995**, *268*, 432-436.

11. Mohammadi, M.; Olsen, S. K.; Ibrahimi, O. A. Structural Basis for Fibroblast Growth Factor Receptor Activation. *Cytokine Growth Factor Rev.* **2005**, *16*, 107-137.

12. Johnson, D. E.; Williams, L. T. Structural and Functional Diversity in the FGF Receptor Multigene Family. *Adv. Cancer Res.* **1993***, 60*, 1-41.

13. Lin, X.; Buff, E. M.; Perrimon, N.; Michelson, A. M. Heparan Sulfate Proteoglycans Are Essential for FGF Receptor Signaling during Drosophila Embryonic Development. *Development* **1999**, *126*, 3715-3723.

14. Yayon, A.; Klagsbrun, M.; Esko, J. D.; Leder, P.; Ornitz, D. M. Cell Surface, Heparin-like Molecules Are Required for Binding of Basic Fibroblast Growth Factor to Its High Affinity Receptor. *Cell* **1991**, *64*, 841-848.

15. Rapraeger, A. C.; Krufka, A.; Olwin, B. B. Requirement of Heparan Sulfate for BFGF-Mediated Fibroblast Growth and Myoblast Differentiation. *Science* **1991**, *252*, 1705-1708.

16. Taipale, J.; Keski-Oja, J. Growth Factors in the Extracellular Matrix. *FASEB J.* **1997**, *11*, 51-59.

17. McKeehan, W. L.; Wang, F.; Kan, M. The Heparan Sulfate-Fibroblast Growth Factor Family: Diversity of Structure and Function. *Prog. Nucleic Acid Res. Mol. Biol.* **1998**, *59*, 135-176.

18. Culajay, J. F.; Blaber, S. I.; Khurana, A.; Blaber, M. Thermodynamic Characterization of Mutants of Human Fibroblast Growth Factor 1 with an Increased Physiological Half-Life. *Biochemistry* **2000**, *39*, 7153-7158.

19. Agostino, M.; Gandhi, N. S.; Mancera, R. L. Development and Application of Site Mapping Methods for the Design of Glycosaminoglycans. *Glycobiology* **2014**, *24*, 840-851.

20. Imberty, A.; Lortat-Jacob, H.; Pérez, S. Structural View of Glycosaminoglycan-Protein Interactions. *Carbohydr. Res*. **2007**, *342*, 430-439.

21. ten Dam, G. B.; Van De Westerlo, E. M. A.; Smetsers, T. F. C. M.; Willemse, M.; Van Muijen, G. N. P.; Merry, C. L. R.; Gallagher, J. T.; Kim, Y. S.; Van Kuppevelt, T. H. Detection of 2-O-Suifated Iduronate and N-Acetylglucosamine Units in Heparan Sulfate by an Antibody Selected against Acharan Sulfate (IdoA2S-GlcNAc)N. *J. Biol. Chem.* **2004**, *279*, 38346-38352.

22. Babik, S.; Samsonov, S. A.; Pisabarro, M. T. Computational Drill down on FGF1-Heparin Interactions through Methodological Evaluation. *Glycoconj. J.* **2017***, 34*, 427-440.

23. Muñoz-García, J. C.; García-Jiménez, M. J.; Carrero, P.; Canales, A.; Jiménez-Barbero, J.; Martín-Lomas, M.; Imberty, A.; de Paz, J. L.; Angulo, J.; Lortat-Jacob, H. et al. Importance of the Polarity of the Glycosaminoglycan Chain on the Interaction with FGF-1. *Glycobiology* **2014**, *24* 1004-1009.

24. Raman, R.; Venkataraman, G.; Ernst, S.; Sasisekharan, V.; Sasisekharan, R. Structural Specificity of Heparin Binding in the Fibroblast Growth Factor Family of Proteins. *Proc. Natl. Acad. Sci. U. S. A.* **2003***, 100*, 2357-2362.

25. Habuchi, H.; Habuchi, O.; Kimata, K. Sulfation Pattern in Glycosaminoglycan: Does It Have a Code? *Glycoconj. J.* **2004**, *21*, 47-52.

26. Gama, C. I.; Tully, S. E.; Sotogaku, N.; Clark, P. M.; Rawat, M.; Vaidehi, N.; Goddard, W. A.; Nishi, A.; Hsieh-Wilson, L. C. Sulfation Patterns of Glycosaminoglycans Encode Molecular Recognition and Activity. *Nat. Chem. Biol.* **2006**, *2*, 467-473.

27. Saxena, K.; Schieborr, U.; Anderka, O.; Duchardt-Ferner, E.; Elshorst, B.; Gande, S. L.; Janzon, J.; Kudlinzki, D.; Sreeramulu, S.; Dreyer, M. K. et al. Influence of Heparin Mimetics on Assembly of the FGF·FGFR4 Signaling Complex. *J. Biol. Chem.* **2010**, *285*, 26628-26640.

28. DiGabriele, A. D.; Lax, I.; Chen, D. I.; Svahn, C. M.; Jaye, M.; Schlessinger, J.; Hendrickson, W. A. Structure of a Heparin-Linked Biologically Active Dimer of Fibroblast Growth Factor. *Nature* **1998**, *393*, 812-817.

29. Zhu, X. T.; Hsu, B. T.; Rees, D. C. Structural Studies of The Binding of The Antiulcer Drug Sucrose Octasulfate to Acidic Fibroblast Growth-Factor. *Structure* **1993***, 1,* 27-34.

30. Canales, A.; Lozano, R.; López-Méndez, B.; Angulo, J.; Ojeda, R.; Nieto, P. M.; Martín-Lomas, M.; Giménez-Gallego, G.; Jiménez-Barbero, J. Solution NMR Structure of a Human FGF-1 Monomer, Activated by a Hexasaccharide Heparin-Analogue. *FEBS J.* **2006**, *273*, 4716-4727.

31. Copeland, R. A.; Ji, H.; Halfpenny, A. J.; Williams, R. W.; Thompson, K. C.; Herber, W. K.; Thomas, K. A.; Bruner, M. W.; Ryan, J. A.; Marquis-Omer, D. et al. The Structure of Human Acidic Fibroblast Growth Factor and Its Interaction with Heparin. *Arch. Biochem. Biophys.* **1991**, *289*, 53-61.

32. Blaber, M.; Adamek, D. H.; Popovic, A.; Blaber, S. I. Biophysical and Structural Analysis of Human Acidic Fibroblast Growth Factor. *Tech. Protein Chem.* **1997**, *8*, 745-753.

33. Carter, E. P.; Fearon, A. E.; Grose, R. P. Careless Talk Costs Lives: Fibroblast Growth Factor Receptor Signalling and the Consequences of Pathway Malfunction. *Trends Cell Biol*. **2015,** *25*, 221-233.

34. Goetz, R.; Mohammadi, M. Exploring Mechanisms of FGF Signalling through the Lens of Structural Biology. *Nat. Rev. Mol. Cell Biol*. **2013**, *14*, 166-180.

35. Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K. Scalable Molecular Dynamics with NAMD. *J. Comput. Chem*. **2005**, *26*, 1781-1802.

36. Best, R. B.; Zhu, X.; Shim, J.; Lopes, P. E. M.; Mittal, J.; Feig, M.; MacKerell, A. D. Optimization of the Additive CHARMM All-Atom Protein Force Field Targeting Improved Sampling of the Backbone φ, ψ and Side-Chain X1 and X2 Dihedral Angles. *J. Chem. Theory Comput.* **2012**, *8*, 3257-3273.

37. Jo, S.; Kim, T.; Im, W. Automated Builder and Database of Protein/Membrane Complexes for Molecular Dynamics Simulations. *PLoS One* **2007**, 2, e880.

38. Lee, J.; Cheng, X.; Swails, J. M.; Yeom, M. S.; Eastman, P. K.; Lemkul, J. A.; Wei, S.; Buckner, J.; Jeong, J. C.; Qi, Y. et al. CHARMM-GUI Input Generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM Simulations Using the CHARMM36 Additive Force Field. *J. Chem. Theory Comput.* **2016**, *12*, 405-413.

39. *Large Sparse Sets of Linear Equations*. Reid, J.K., Ed.; Academic Press: London and New York, 1971.

40. Martyna, G. J.; Tobias, D. J.; Klein, M. L. Constant Pressure Molecular Dynamics Algorithms. *J. Chem. Phys.* **1994**, *101*, 4177-4189.

41. Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An N·log(N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98*, 10089-10092.

42. Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. *J. Mol. Graph.* **1996**, *14*, 33-38.

Binding Affinity Estimation From Restrained Umbrella Sampling Simulations

Vivek Govind Kumar[1], Shilpi Agrawal[1], Thallapuranam Krishnaswamy Suresh Kumar[1], and Mahmoud Moradi[1]

[1]Department of Chemistry and Biochemistry, University of Arkansas, Fayetteville, AR 72701

ABSTRACT

The protein-ligand binding affinity quantifies the binding strength between a protein and its ligand. Computer modeling and simulations can be used to estimate the binding affinity or binding free energy using data- or physics-driven methods or a combination thereof. Here, we discuss a purely physics-based sampling approach based on biased molecular dynamics (MD) simulations, which in spirit is similar to the stratification strategy suggested previously by Woo and Roux. The proposed methodology uses umbrella sampling (US) simulations with additional restraints based on collective variables such as the orientation of the ligand. The novel extension of this strategy presented here uses a simplified and more general scheme that can be easily tailored for any system of interest. We estimate the binding affinity of human fibroblast growth factor 1 (hFGF1) to heparin hexasaccharide based on the available crystal structure of the complex as the initial model and four different variations of the proposed method to compare against the experimentally determined binding affinity obtained from isothermal calorimetry (ITC) experiments. Our results indicate that enhanced sampling methods that sample along the ligand-protein distance without restraining other degrees of freedom do not perform as well as those with additional restraint. In particular, restraining the orientation of the ligands plays a crucial role in reaching a reasonable estimate for binding affinity. The general framework presented here provides a flexible scheme for designing practical binding free energy estimation methods.

INTRODUCTION

Accurate quantification of absolute binding affinities remains a problem of major importance in computational biophysics[1,2,3,4]. In principle, accurate binding free energy calculations should be the cornerstone of any study investigating protein-ligand interactions. However, the high computational costs that typically accompany such calculations necessitate the improvement of the computational methods traditionally used to investigate complex biomolecular interactions[3,4,5]. Experimentally determined binding affinities are commonly used as benchmarks to judge the accuracy of various computational binding affinity estimation methods[5,6,7]. Several experimental techniques can be used to study protein-ligand binding equilibria[5,8]. For instance, isothermal titration calorimetry (ITC) can detect the interaction of binding partners based on changes in solution heat capacity and binding partner concentration[8,9,10]. Other methods such as fluorescence spectroscopy rely on changes in fluorescence intensity upon ligand binding[8,11,12]. Surface plasmon resonance (SPR) can be used to calculate binding affinities based on changes in refractive index that occur when an immobilized binding partner interacts with a free binding partner[8,13,14]. Studies have found that experimental binding affinities can vary depending on the experimental method used[5,6,15]. Therefore, a thorough understanding of the experimental conditions used to generate reference data is essential when comparing computationally determined binding affinities with experimental values.

Several computational methods at varying levels of rigor and complexity have been used to determine binding affinities for biomolecular interactions[3,16-28]. Knowledge-based statistical potentials and force field scoring potentials are typically used to rank docked protein-ligand or protein-protein complexes but can also be used for binding affinity prediction[29,30,31].

A major disadvantage of these methods is that they do not treat the entropic effects rigorously, which effectively decreases the accuracy of such binding affinity predictions[5,32]. This is also the case for methods like Molecular Mechanics/Poisson-Boltzmann-Surface Area (MM-PBSA) and Molecular Mechanics/Generalized Born-Surface Area (MM-GBSA), which combine sampling of conformations from explicit solvent molecular dynamics (MD) simulations with free energy estimation based on implicit continuum solvent models[33,34,35]. Adequate sampling of both ligand conformational dynamics as well as ligand roto-translational movements with respect to the protein is essential for accurately quantifying the entropic reduction arising from the binding event[35,36,37]. MM-PBSA/GBSA methods typically neglect the contribution of these entropic terms to the binding free energy[34,35].

One of the best-known binding free energy estimation methods is alchemical free energy perturbation (FEP), where scaling of non-bonded interactions enables reversal decoupling of the ligand from its environment in the bound state as well as the unbound state[38,39,40,41]. Most entropic and enthalpic contributors to changes in binding affinity are typically considered during FEP simulations, thus avoiding the approximations used by methods like MM-PBSA/GBSA[5,42]. A disadvantage of FEP is the fact that ligands tend to move away from the binding site during the decoupling process, which results in poorly defined target states of the FEP calculation being used as starting states for the re-coupling process[43]. Using receptor-ligand restraints to resolve this issue[17,40,44,45] introduces some ambiguity to the way a standard state is defined, with a level of correlation between the size of the simulation cell and the standard state[46]. This can be corrected via the use of appropriate geometrical restraints[16,47,23].

Unrestrained long timescale MD simulations should theoretically allow for the investigation and accurate quantification of protein-ligand or protein-protein binding events[48,49].

While microsecond-level MD simulations provide a more accurate description of protein conformational dynamics as compared to shorter simulations[50], efficient sampling of the conformational landscape remains a major issue and requires access to timescales beyond the capabilities of current MD simulations[51,52]. Several methods have been developed to tackle the sampling problem. Markov state models allow the sampling and characterization of native as well as alternative binding states[53,54,55]. Similarly, weighted ensemble (WE) simulations sample the conformational landscape along one or more discretized reaction coordinates based on the assignment of a statistical weight to each simulation[56,57]. More traditionally, umbrella sampling along such reaction coordinates can be used to guide the binding or unbinding of a ligand, after which algorithms like the weighted histogram analysis method (WHAM) can be used to calculate a unidimensional potential of mean force (PMF) which quantifies ligand binding and unbinding along a reaction coordinate[58,59]. Better convergence of the calculated free energy profiles can be achieved by the exchange of conformations between successive umbrella-sampling windows as in the bias-exchange umbrella sampling (BEUS)[60,61,62]. Other methods based on similar principles include umbrella integration[63], well-tempered metadynamics[64], adaptive biasing force (ABF) simulations[65] and variations of these techniques.

Incomplete sampling of important degrees of freedom, such as orientation of the ligand with respect to the protein, remains a major disadvantage of unidimensional PMF-based methods[3,4]. To resolve this problem, Woo and Roux[3] have devised a method wherein explicitly defined geometrical restraints on the orientation and conformation of the binding partners are used to reduce the conformational entropy of the biomolecular system being studied[3,4]. This results in improved convergence of the PMF calculation[3,4]. The introduction of a restraining potential based on the root-mean-square deviation (RMSD) of the ligand relative to its average

bound conformation, reduces the flexibility of the ligand and the number of conformations that need to be sampled[3,4]. This method avoids the need to decouple the ligand from its surrounding environment as required by alchemical FEP[3,4,38-41]. Recent studies have described applications and extensions of the methodology proposed by Woo and Roux[4,66].

Here, we describe a purely physics-based enhanced sampling method based on biased MD simulations, which is similar in principle to the stratification strategy proposed by Woo and Roux[3,4]. Although we use the US method as our enhanced sampling technique, the methodology is generalizable to other techniques as long as they can be combined with additional restraints. A major difference between our method and that of Woo and Roux[3,4] is the use of the unidimensional orientation angle of the ligand with respect to the protein as a collective variable for restraining, as opposed to the use of three Euler angles. The formalism has also some other major differences that are discussed in more detail below. We have used use this methodology to calculate the binding affinity for the interaction of human fibroblast growth factor 1 (hFGF1) with heparin hexasaccharide, its glycosaminoglycan (GAG) binding partner. hFGF1 is an important signaling protein that is implicated in physiological processes such as cell proliferation and differentiation, neurogenesis, wound healing, tumor growth and angiogenesis[67,68,69,70,71]. GAGs are linear anionic polysaccharides that interact with positively charged regions of FGF binding partners to regulate their biological activity[70,72-80]. The hFGF1-heparin complex is the most well-known and broadly characterized protein-GAG complex[81,82]. Heparin binding is thought to stabilize hFGF1 and impart protection against proteolytic degradation. In this study, we show that the absolute binding affinity for the hFGF1-heparin interaction calculated using our novel approach, is in good agreement with binding affinity data from ITC experiments.

THEORETICAL FOUNDATION

Binding affinity is often quantified using the equilibrium dissociation constant (K_d), defined as:

$$K_d = [P][L]/[P{:}L] \qquad (1)$$

where $[P]$, $[L]$, and $[P{:}L]$ are the concentrations of protein, ligand, and the protein-ligand complex, respectively. Computationally, the absolute binding free energy ($\Delta G°$), which is the standard molar free energy of binding, is more convenient to calculate. The dissociation constant and the absolute binding free energy are related via

$$\Delta G° = RT \ln \frac{K_d}{1M} \qquad (2)$$

where R is the gas constant and T is the temperature. Various strategies have been used to estimate $\Delta G°$, some of which were briefly discussed above. The methodology proposed here has a significant resemblance to the stratification strategy of Woo and Roux[3,4]. However, the two methods have major differences as will be discussed later.

Absolute binding free energy or $\Delta G°$ is the free energy change associated with moving the ligand from the bulk to the binding pocket. Within the formalism presented in this work, $\Delta G°$ is determined from the grid PMF $G(\boldsymbol{x})$, where $\boldsymbol{x}$ is the position of the ligand mass center from the center of the binding pocket, $G(\boldsymbol{x})$ is the potential of mean force (PMF) associated with the ligand position $\boldsymbol{x}$. In practice, we need to bin the 3D space and define the PMF at every bin or grid point as:

$$G(\boldsymbol{x}) = -RT \ln p(\boldsymbol{x}) \qquad (3)$$

where $p(\boldsymbol{x})$ is the probability of finding the ligand at bin $\boldsymbol{x}$.

We define $\Delta G(\boldsymbol{x}) = G(\boldsymbol{x}) - G(\boldsymbol{0})$, where $\boldsymbol{x} = \boldsymbol{0}$ (i.e., the center of the binding pocket) is defined as the grid point associated with the lowest grid PMF. One can show:

$$\Delta G^{\circ} = -RT \ln \frac{\int_{pocket} e^{-\frac{G(\boldsymbol{x})}{RT}} dV}{\int_{bulk} e^{-\frac{G(\boldsymbol{x})}{RT}} dV} = -RT \ln \frac{\int_{pocket} e^{-\frac{\Delta G(\boldsymbol{x})}{RT}} dV}{\int_{bulk} e^{-\frac{\Delta G(\boldsymbol{x})}{RT}} dV} \qquad (4)$$

in which the pocket refers to all $\boldsymbol{x}$ where the ligand is considered bound (i.e., the binding pocket) and bulk refers to all $\boldsymbol{x}$ where the ligand is not interacting with the protein. Since $\Delta G(\boldsymbol{x})$ is the same everywhere in the bulk, we can simplify Relation (4) as follows:

$$\Delta G^{\circ} = -RT \ln \frac{V_P}{e^{-\frac{\Delta G(\boldsymbol{x}_B)}{RT}} V_B} = -\Delta G(\boldsymbol{x}_B) - RT \ln \frac{V_P}{V_B} \qquad (5)$$

where V_B is the bulk volume per protein associated with the standard concentration (i.e., 1 M), $\boldsymbol{x}_B$ is any grid point in the bulk, and V_P is the binding pocket volume defined as:

$$V_P = \int_{pocket} e^{-\frac{\Delta G(\boldsymbol{x})}{RT}} dV \qquad (6)$$

Defining ΔG_V as the contribution of the difference between the volume of the binding pocket and the bulk to the binding free energy:

$$\Delta G_V = -RT \ln \frac{V_P}{V_B} \qquad (7)$$

Combining (5) and (7), we have:

$$\Delta G^{\circ} = -\Delta G(\boldsymbol{x}_B) + \Delta G_V \qquad (8)$$

We can find the bulk volume (V_B) associated with the standard concentration for a single protein approximately as:

$$V_B = \frac{\frac{1}{N_A} \text{ mol}}{1\ M} = \frac{1}{N_A}\ L \approx 1661\ \text{Å}^3 \qquad (9)$$

where N_A is the Avogadro's constant. We can now rewrite ΔG_V as:

$$\Delta G_V = -RT \ln \frac{V_P}{V_B} = -RT \ln \frac{V_P}{\text{Å}^3} + RT \ln \frac{V_B}{\text{Å}^3} = \Delta G_P - \Delta G_B \qquad (10)$$

in which ΔG_B is the bulk volume contribution and ΔG_P is the binding pocket contribution:

$$\begin{cases} \Delta G_B = -RT \ln \frac{V_B}{\text{Å}^3} \approx -7.42RT \\ \Delta G_P = -RT \ln \frac{V_P}{\text{Å}^3} = -RT \ln \int_{pocket} e^{-\frac{\Delta G(\boldsymbol{x})}{RT}} \frac{dV}{\text{Å}^3} \end{cases} \quad (11)$$

Determining both $\Delta G(\boldsymbol{x}_B)$ and ΔG_P requires finding the grid PMF $\Delta G(\boldsymbol{x})$. $\Delta G(\boldsymbol{x}_B)$ is the PMF difference between the binding pocket center and the bulk and ΔG_P also requires an estimate for $\Delta G(\boldsymbol{x})$ within the binding pocket. We therefore do not need to find $\Delta G(\boldsymbol{x})$ for all $\boldsymbol{x}$ if we have a good estimate for $\Delta G(\boldsymbol{x})$ within the binding pocket and in the bulk. Ideally, $\Delta G(\boldsymbol{x})$ for these points can be determined by pulling the ligand out of the binding pocket towards the bulk and using an enhanced sampling technique such as US to sample the space of a collective variable such as d, i.e., the distance between the mass centers of the ligand and protein. $\Delta G(\boldsymbol{x})$ can be estimated for all sampled grid points $\boldsymbol{x}$ using this distance-based US simulation. Note that the collective variable used for biasing would be d, while the collective variable used for the PMF calculations would be the 3D position vector of the mass center of ligand with respect to protein's binding pocket center. One may estimate the grid PMF from the distance-based US simulations using a non-parametric reweighting algorithm as discussed in the Methods section. $\Delta G(\boldsymbol{x})$ can also be used to estimate ΔG_P as defined in Relation (11). There is often no need to strictly define the binding pocket since only low $\Delta G(\boldsymbol{x})$ values have nonnegligible contribution to V_P and thus even if we include all sampled grid points, only those close to the binding pocket center have nonnegligible contributions.

A practical issue with determining $\Delta G(\boldsymbol{x}_B)$ is the convergence. The key obstacles for the sampling that slow down the convergence are the orientation of the ligand, and the conformational changes of the ligand and protein. Using an approach similar in spirit to the

previously proposed stratification strategy[3,4,35], we can circumvent extensive sampling of these degrees of freedom. Let us first focus on the orientation of the ligand (Ω). We can restrain Ω during the distance-based US simulations using a biasing potential ($\frac{1}{2}k\Omega^2$) and later correct the free energy difference based on the PMF associated with the Ω, which is different in the bulk ($F(\boldsymbol{x}_{\mathrm{B}},\Omega)$) and in the binding pocket ($F(\mathbf{0},\Omega)$). More generally, for any grid point $\boldsymbol{x}$, we may determine $\Delta G(\boldsymbol{x})$ based on the PMF associated with the Ω at $\boldsymbol{x}$ ($F(\boldsymbol{x},\Omega)$) and $\mathbf{0}$ ($F(\mathbf{0},\Omega)$):

$$e^{-\frac{\Delta G(\boldsymbol{x})}{RT}} = \frac{\int_0^\pi e^{-\frac{F(\boldsymbol{x},\Omega)}{RT}}\,d\Omega}{\int_0^\pi e^{-\frac{F(\mathbf{0},\Omega)}{RT}}\,d\Omega} \qquad (12)$$

Note that $F(\boldsymbol{x},\Omega)$ is the PMF associated with $\boldsymbol{x}$ and Ω, defined such that:

$$G(\boldsymbol{x}) = c - RT\ln\int_0^\pi e^{-\frac{F(\boldsymbol{x},\Omega)}{RT}}\,d\Omega \qquad (13)$$

where c is an arbitrary constant. We therefore have:

$$e^{-\frac{\Delta G(\boldsymbol{x})}{RT}} = \frac{\int_0^\pi e^{-\frac{F(\boldsymbol{x},\Omega)}{RT}}\,d\Omega}{\int_0^\pi e^{-\frac{F(\boldsymbol{x},\Omega)+\frac{1}{2}k\Omega^2}{RT}}\,d\Omega} \times \frac{\int_0^\pi e^{-\frac{F(\mathbf{0},\Omega)+\frac{1}{2}k\Omega^2}{RT}}\,d\Omega}{\int_0^\pi e^{-\frac{F(\mathbf{0},\Omega)}{RT}}\,d\Omega} \times \frac{\int_0^\pi e^{-\frac{F(\boldsymbol{x},\Omega)+\frac{1}{2}k\Omega^2}{RT}}\,d\Omega}{\int_0^\pi e^{-\frac{F(\mathbf{0},\Omega)+\frac{1}{2}k\Omega^2}{RT}}\,d\Omega} \qquad (14)$$

We now define $G_\Omega(\boldsymbol{x})$ as the grid PMF of the restrained system (by Ω):

$$G_\Omega(\boldsymbol{x}) = -RT\ln\int_0^\pi e^{-\frac{F(\boldsymbol{x},\Omega)+\frac{1}{2}k\Omega^2}{RT}}\,d\Omega \qquad (15)$$

We also define $U_\Omega(\boldsymbol{x})$ as the average biasing potential at grid point $\boldsymbol{x}$:

$$U_\Omega(\boldsymbol{x}) = -RT\ln\langle e^{-\frac{\frac{1}{2}k\Omega^2}{RT}}\rangle_{\boldsymbol{x}} = -RT\ln\frac{\int_0^\pi e^{-\frac{F(\mathbf{x},\Omega)+\frac{1}{2}k\Omega^2}{RT}}\,d\Omega}{\int_0^\pi e^{-\frac{F(\mathbf{x},\Omega)}{RT}}\,d\Omega} \qquad (16)$$

Now we have from Relations (14), (15), and (16):

$$\Delta G(\boldsymbol{x}) = \Delta G_\Omega(\boldsymbol{x}) - \Delta U_\Omega(\boldsymbol{x}) \qquad (17)$$

where the free energy of grid point $\boldsymbol{x}$ from the center $\mathbf{0}$ ($\Delta G(\boldsymbol{x})$) is calculated based on its equivalent free energy ($\Delta G_{\Omega}(\boldsymbol{x})$) in a system biased by a harmonic restraint on Ω and a correction term $\Delta U_{\Omega}(\boldsymbol{x})$. For $\boldsymbol{x} = \boldsymbol{x}_B$:

$$\Delta U_{\Omega}(\boldsymbol{x}_B) = -RT \ln \frac{\langle e^{-\frac{\frac{1}{2}k\Omega^2}{RT}} \rangle_{bulk}}{\langle e^{-\frac{\frac{1}{2}k\Omega^2}{RT}} \rangle_{pocket}} \qquad (18)$$

To determine the above ensemble averages, we need to determine the PMF along Ω for the bound and unbound ligand and calculate the ensemble averages analytically using Relation (16). $\Delta G_{\Omega}(\boldsymbol{x}_B)$ can be determined from PMF calculations, where the distance between the protein and ligand is varied and the orientation of the ligand is restrained (distance-based BEUS with restrained orientation). We note that:

$$V_P = \int_{pocket} e^{-\frac{\Delta G(\boldsymbol{x})}{RT}} dV = \int_{pocket} e^{-\frac{\Delta G_{\Omega}(\boldsymbol{x}) - \Delta U_{\Omega}(\boldsymbol{x})}{RT}} dV \approx \int_{pocket} e^{-\frac{\Delta G_{\Omega}(\boldsymbol{x})}{RT}} dV \quad (19)$$

where we assume $\Delta U_{\Omega}(\boldsymbol{x})$ is negligible for $\boldsymbol{x}$ within the binding pocket. In other words, $\langle e^{-\frac{\frac{1}{2}k\Omega^2}{RT}} \rangle_{\boldsymbol{x}} \approx \langle e^{-\frac{\frac{1}{2}k\Omega^2}{RT}} \rangle_{\mathbf{0}}$ for $\boldsymbol{x}$ close to $\mathbf{0}$.

In brief, if we choose to restrain the orientation, our absolute binding free energy estimate includes the following terms (using Relations (8) and (17)):

$$\Delta G^{\circ} = -\Delta G_{\Omega}(\boldsymbol{x}_B) + \Delta \mathrm{U}_{\Omega}(\boldsymbol{x}_B) + \Delta G_V \qquad (20)$$

$F(\boldsymbol{x}_B, \Omega)$ can be calculated numerically from orientation angle distribution of a free ligand: $\mathrm{F}(\boldsymbol{x}_B, \Omega) = -RT \ln p(\Omega)$, where $p(\Omega)$ is determined from the distribution of Euler angles ($p(\phi, \theta, \psi) = \frac{1}{8\pi^2} \sin\theta$, where $0 \leq \phi, \psi \leq 2\pi$ and $0 \leq \theta \leq \pi$) given that:

$$\cos\frac{\Omega}{2} = \cos\frac{\phi}{2}\cos\frac{\theta}{2}\cos\frac{\psi}{2} + \sin\frac{\phi}{2}\sin\frac{\theta}{2}\sin\frac{\psi}{2} \qquad (21)$$

$\langle e^{-\frac{\frac{1}{2}k\Omega^2}{RT}}\rangle_{bulk}$ can then be calculated using numerically calculated $F(\boldsymbol{x}_B, \Omega)$ and k as used in the simulations using Relation (16). $F(\mathbf{0}, \Omega)$ can be determined approximately using orientation-based US simulations of bound ligand. F$(\mathbf{0}, \Omega)$ can then be used to estimate $\langle e^{-\frac{\frac{1}{2}k\Omega^2}{RT}}\rangle_{pocket}$ using Relation (16).

The above strategy can be extended to other degrees of freedom for which unbiased sampling may hinder the convergence. Most notably, the internal conformational changes of the ligand and that of the protein may also play a crucial role in slowing down the convergence. In the following, we show how one can restrain not only the orientation of the ligand but also the root-mean-square-deviation (RMSD) of the ligand (denoted here by r) in distance-based US simulations (along d) to speed up convergence. In this case, the grid PMF difference $\Delta G(\boldsymbol{x})$ is calculated based on $\Delta G_{\Omega,r}(\boldsymbol{x})$, the grid PMF of a system whose Ω and r are both restrained:

$$e^{-\frac{\Delta G(\boldsymbol{x})}{RT}} = \frac{\int_0^\infty \int_0^\pi e^{-\frac{G(\boldsymbol{x},\Omega,\mathrm{r})}{RT}}\, d\Omega\, dr}{\int_0^\infty \int_0^\pi e^{-\frac{G(\mathbf{0},\Omega,\mathrm{r})}{RT}}\, d\Omega\, dr} \qquad (22)$$

Using a similar strategy as in Relation (14), we have:

$$e^{-\frac{\Delta G(\boldsymbol{x})}{RT}} = \frac{\int_0^{\infty}\int_0^{\pi} e^{-\frac{F(\boldsymbol{x},\Omega,\mathrm{r})}{RT}}\, d\Omega\, dr}{\int_0^{\infty}\int_0^{\pi} e^{-\frac{F(\mathbf{x},\Omega,\mathrm{r})+\frac{1}{2}k'\mathrm{r}^2}{RT}}\, d\Omega\, dr} \times \frac{\int_0^{\infty}\int_0^{\pi} e^{-\frac{F(\mathbf{x},\Omega,\mathrm{r})+\frac{1}{2}k'\mathrm{r}^2}{RT}}\, d\Omega\, dr}{\int_0^{\infty}\int_0^{\pi} e^{-\frac{F(\mathbf{x},\Omega,\mathrm{r})+\frac{1}{2}k'\mathrm{r}^2+\frac{1}{2}k\Omega^2}{RT}}\, d\Omega\, dr}$$

$$\times \frac{\int_0^{\infty}\int_0^{\pi} e^{-\frac{F(\mathbf{0},\Omega,\mathrm{r})+\frac{1}{2}k'\mathrm{r}^2+\frac{1}{2}k\Omega^2}{RT}}\, d\Omega\, dr}{\int_0^{\infty}\int_0^{\pi} e^{-\beta(F(\mathbf{0},\Omega,\mathrm{r})+\frac{1}{2}k'\mathrm{r}^2)}\, d\Omega\, dr} \times \frac{\int_0^{\infty}\int_0^{\pi} e^{-\frac{F(\mathbf{0},\Omega,\mathrm{r})+\frac{1}{2}k'\mathrm{r}^2}{RT}}\, d\Omega\, dr}{\int_0^{\infty}\int_0^{\pi} e^{-\beta F(\mathbf{0},\Omega,\mathrm{r})}\, d\Omega\, dr}$$

$$\times \frac{\int_0^{\infty}\int_0^{\pi} e^{-\frac{F(\mathbf{x},\Omega,\mathrm{r})+\frac{1}{2}k'\mathrm{r}^2+\frac{1}{2}k\Omega^2}{RT}}\, d\Omega\, dr}{\int_0^{\infty}\int_0^{\pi} e^{-\frac{F(\mathbf{0},\Omega,\mathrm{r})+\frac{1}{2}k'\mathrm{r}^2+\frac{1}{2}k\Omega^2}{RT}}\, d\Omega\, dr} \qquad (23)$$

which results in:

$$e^{-\frac{\Delta G(\boldsymbol{x})}{RT}} = \frac{\langle e^{-\frac{\frac{1}{2}k'\mathrm{r}^2}{RT}}\rangle_{\mathbf{0}}}{\langle e^{-\frac{\frac{1}{2}k'\mathrm{r}^2}{RT}}\rangle_{\boldsymbol{x}}} \times \frac{\langle e^{-\frac{\frac{1}{2}k\Omega^2}{RT}}\rangle_{\mathbf{0}}^{r}}{\langle e^{-\frac{\frac{1}{2}k\Omega^2}{RT}}\rangle_{\boldsymbol{x}}^{r}} \times \frac{e^{-\beta G_{\Omega,\mathrm{r}}(\boldsymbol{x})}}{e^{-\beta G_{\Omega,\mathrm{r}}(\mathbf{0})}} \qquad (24)$$

Here we have defined $G_{\Omega,\mathrm{r}}(\boldsymbol{x})$ as:

$$G_{\Omega,\mathrm{r}}(\boldsymbol{x}) = -RT\ln\int_0^{\infty}\int_0^{\pi} e^{-\frac{F(\mathbf{x},\Omega,\mathrm{r})+\frac{1}{2}k'\mathrm{r}^2+\frac{1}{2}k\Omega^2}{RT}}\, d\Omega\, dr \qquad (25)$$

We also define $U_r(\boldsymbol{x})$ similar to $U_\Omega(\boldsymbol{x})$ in Relation (15) except for using r instead of Ω. $U_\Omega^r(\boldsymbol{x})$ is also defined similar to $U_\Omega(\boldsymbol{x})$ except for the additional restraint on r:

$$U_\Omega^r(\boldsymbol{x}) = -RT\ln\langle e^{-\frac{\frac{1}{2}k\Omega^2}{RT}}\rangle_{\boldsymbol{x}}^{r} = -RT\ln\frac{\int_0^{\infty}\int_0^{\pi} e^{-\frac{F(\mathbf{x},\Omega,\mathrm{r})+\frac{1}{2}k'\mathrm{r}^2+\frac{1}{2}k\Omega^2}{RT}}\, d\Omega\, dr}{\int_0^{\infty}\int_0^{\pi} e^{-\beta\left(F(\mathbf{x},\Omega,\mathrm{r})+\frac{1}{2}k'\mathrm{r}^2\right)}\, d\Omega\, dr} \qquad (26)$$

Finally, we have:

$$\Delta G(\boldsymbol{x}) = \Delta G_{\Omega,\mathrm{r}}(\boldsymbol{x}) - \Delta U_{\mathrm{r}}(\boldsymbol{x}) - \Delta U_\Omega^r(\boldsymbol{x}) \qquad (27)$$

In brief, if we choose to restrain both the orientation and RMSD, our absolute binding free energy estimate includes the following terms:

$$\Delta G^\circ = -\Delta G_{\Omega,\mathrm{r}}(\boldsymbol{x}_B) + \Delta \mathrm{U}_\mathrm{r}(\boldsymbol{x}_B) + \Delta \mathrm{U}^\mathrm{r}_\Omega(\boldsymbol{x}_B) + \Delta G_V \qquad (28)$$

Here we are using an approximation similar to that in Relation (19):

$$V_P \approx \int_{pocket} e^{-\frac{\Delta G_{\Omega,\mathrm{r}}(\boldsymbol{x})}{RT}} dV \qquad (29)$$

Using Relations (20) and (28), we can generalize the stratification strategy to include three restraints on arbitrary collective variables α, β, and γ:

$$\Delta G^\circ = -\Delta G_{\alpha,\beta,\gamma}(\boldsymbol{x}_B) + \Delta \mathrm{U}_\gamma(\boldsymbol{x}_B) + \Delta \mathrm{U}^\gamma_\beta(\boldsymbol{x}_B) + \Delta \mathrm{U}^{\beta,\gamma}_\alpha(\boldsymbol{x}_B) + \Delta G_V \qquad (30)$$

where:

$$\Delta G_V \approx -RT \ln \int_{pocket} e^{-\frac{\Delta G_{\alpha,\beta,\gamma}(\boldsymbol{x})}{RT}} \frac{dV}{\mathrm{\AA}^3} - \Delta G_B \qquad (31)$$

METHODS

Isothermal titration calorimetry of hFGF-1 with heparin hexasaccharide

Isothermal titration calorimetry (ITC) data was obtained using MicroCal iTC 200 (Malvern Inc.). The change in heat during the biomolecular interaction was measured by titrating the heparin (loaded in the syringe) to the hFGF1 solution in the calorimetric cell. Both the protein and the heparin samples were made in the buffer containing 10 mM phosphate buffer with 100 mM NaCl at pH 7.2 and were degassed prior to loading. The protein to heparin ratio was maintained at 1:10 with the protein concentration being 100 μM and the heparin concentration being 1mM. A total of 30 injections were conducted with a constant temperature of 25 °C and stirring speed of 300 rpm. One set of sites binding model was used for the ITC binding curve[85]. The standard binding free energy ΔG° was determined from dissociation constant via Relation (2) at $T = 25℃$.

All-atom MD simulations

Our simulations were based on the x-ray crystal structure of the dimeric complex with a heparin hexasaccharide (PDB:2AXM, resolution: 3.0 angstroms)[86]. One of the hFGF1 protomers was removed leaving one protein and one ligand in the model of the holo protein. The model for the apo protein was based on the x-ray crystal structure of unbound monomeric hFGF1 (PDB: 1RG8, resolution: 1.1 angstroms)[87]. The simulations (residues 12-137 in the PDB file correspond to residues 26-151 in the experimental sequence) and experiments were performed using a truncated version of hFGF1 (residues 13-154) which did not contain the unstructured 12 amino acid N-terminal segment. The heparin hexasaccharide consists of N, O6 disulfo-glucosamine and 2-O-sulfo-alpha-L-idopyranuronic acid repeats[86]. The models were solvated in a box of TIP3P waters and 0.15 M NaCl. MD simulations were performed using the NAMD 2.13[88] simulation package with the CHARMM36m all-atom additive force field[89] Initially, we energy-minimized the systems for 10,000 steps using the conjugate gradient algorithm[90]. Subsequently, we relaxed the systems using restrained MD simulations in a stepwise manner (for a total of ~1 ns) using the standard CHARMM-GUI protocol[91,92]. The initial relaxations were performed in an NVT ensemble while the production runs were performed in an NPT ensemble. Simulations were carried out using a 2-fs time step at 300 K using a Langevin integrator with a damping coefficient of $\gamma = 0.5\ ps^{-1}$. The pressure was maintained at 1 atm using the Nosé−Hoover Langevin piston method[90,93]. The smoothed cutoff distance for non-bonded interactions was set to 10−12 Å and long-range electrostatic interactions were computed with the particle mesh Ewald (PME) method[94]. The initial runs lasted 15 nanoseconds, followed by the productions run on the supercomputer Anton 2 (Pittsburgh Supercomputing Center) for 4.8 μs, with a timestep of 2.5 fs. These equilibrium simulations have previously been described in a related study[95]. We

used these equilibrium simulations to construct the PMF in terms of the RMSD of the protein (r_P) both for the apo[95] and holo proteins (for bulk and binding pocket, respectively). We also used the holo protein simulations[95] to construct the PMF in terms of the RMSD of the ligand (r_L) in the binding pocket.

MD simulations of free heparin hexasaccharide

The heparin hexasaccharide[86] was simulated in a rectangular water box without the protein. The system was set up as described previously. The final conformation after relaxation was then used as the starting conformation for 10 production runs for 40 ns each. The total simulation time was around 400 ns. We used these unbiased simulations instead of US simulations to construct the PMF of free heparin in the bulk in terms of ligand RMSD (r_L).

Steered Molecular Dynamics (SMD) simulations

The final conformation of the hFGF1-heparin equilibrium simulation[95] was used to generate starting conformations for the non-equilibrium pulling simulations. Two collective variables[96] were used for SMD simulations[97]: (1) distance between the heavy-atom center of mass of heparin and that of the protein (d) and (2) the orientation angle of heparin with respect to the protein (Ω). Two independent sets of simulations were performed. The distance-based SMD simulation was run for 9.5 ns, while the orientation based SMD simulation was run for 8 ns. The distance-based SMD simulation was used to pull the heparin away from the protein by approximately 30 Å (10→40 Å) with a force constant of 100 kcal/(mol.Å^2). The orientation angle was also restrained

in these simulations with a force constant of 0.5 kcal/(mol.$degree^2$) to stay close to its initial orientation in the bound state. The orientation-based SMD simulation was used to rotate the bound heparin locally with respect to the protein (0°→73°) with a force constant of 100 kcal/(mol.$degree^2$).

Bias Exchange Umbrella Sampling (BEUS) simulations

Bias exchange umbrella sampling[62,98,99] (BEUS), which is a variation of the US simulation method, was performed to estimate grid PMF. Four independent sets of distance (d) based BEUS simulations were performed, with no restraints, restraint on Ω, restraint on r_L and r_P, and restraints on Ω, r_L, and r_P. Two sets of BEUS simulations were also performed using the Ω collective variable, one with and one without a restraint on r_L and r_P. Selected SMD conformations were assigned to individual BEUS windows with equal spacing in each one of these BEUS simulations. The distance-based BEUS simulation ran for 10 ns with 31 replicas/windows and the orientation-based simulation ran for 10 ns with 30 replicas/windows. The force constant used for ligand-protein distance (d) in distance-based BEUS was 2 kcal/(mol.Å^2) while the orientation was restrained as in SMD simulations using a force constant of 0.5 kcal/(mol.$degree^2$). For orientation-based BEUS simulations, the force constant for the ligand orientation angle (as in SMD simulations) was set to 0.5 kcal/(mol.$degree^2$). The force constant used for r_L and r_P was 1 kcal/(mol.Å^2).

Free energy calculations using non-parametric reweighting

Once the BEUS simulations described above were converged, a non-parametric reweighting method[98,100], which is somewhat similar to the multi-state Bennett acceptance ratio method[101], was used to construct PMF. In this method[98], each sampled configuration will be assigned a weight, which can be used to construct the PMF in terms of a desired collective variable. Suppose that a system is biased (for instance, within a BEUS scheme) using N different biasing potentials $U_i(\boldsymbol{r})$, where $i = 1, \ldots, N$, and $\boldsymbol{r}$ represents all atomic coordinates. Typically, $U_i(\boldsymbol{r})$ is a harmonic potential defined in terms of a collective variable with varying centers for different i. Assuming an equal number of sampled configurations from each of the N generated trajectories, we can combine them in a single set of samples $\{\boldsymbol{r}_k\}$ (irrespective of which bias was used to generate each sample $\boldsymbol{r}_k$) and determine the weight of each sample as:

$$w_k = c/\sum_i e^{-\beta(U_i(\boldsymbol{r}_k)-F_i)}$$

where c is the normalization constant such that $\sum_k w_k = 1$ and both $\{w_k\}$ and $\{F_i\}$ are determined iteratively using the above equation and the following:

$$e^{-\beta F_i} = \sum_k w_k e^{-\beta U_i(\boldsymbol{r}_k)}$$

Converged w_k values can be used to construct any ensemble averages including any PMF (e.g., $G(\boldsymbol{\zeta})$) not only in terms of the collective variable used for biasing but also any other collective variables that are sufficiently sampled. One may use a weighted histogram method to construct the PMF as follows:

$$G(\boldsymbol{\zeta}_i) = -RT \ln \sum_k w_k \, \delta(\boldsymbol{\zeta}(\boldsymbol{r}_k) - \boldsymbol{\zeta}_i),$$

$$\delta(\boldsymbol{\zeta}(\boldsymbol{r}_k) - \boldsymbol{\zeta}_i) = \begin{cases} 1, & |\boldsymbol{\zeta}(\boldsymbol{r}_k) - \boldsymbol{\zeta}_i| < \left|\boldsymbol{\zeta}(\boldsymbol{r}_k) - \boldsymbol{\zeta}_j\right| \; for \; j \neq i \\ 0, & otherwise \end{cases}$$

RESULTS AND DISCUSSION

We have calculated the absolute binding free energy for the interaction of hFGF1 with heparin hexasaccharide using four variations of the stratification scheme described above, based on a combination of SMD and BEUS simulations. The details of the methodology are discussed in the Methods section. Four different methods are used with varying effectiveness in estimating the absolute binding free energy. These methods include (1) the traditional distance-based BEUS simulations that do not employ any additional restraining, (2) distance-based BEUS simulations employing a restraint on the orientation of the ligand (Ω) defined based on the orientation quaternion, (3) distance-based BEUS simulations employing a restraint on the RMSD of both ligand and protein ($\mathrm{r_L}, r_P$), (4) distance-based BEUS simulations employing a restraint on the RMSD of both ligand and protein as well as the orientation of the ligand ($\Omega, \mathrm{r_L}, r_P$). In each case, appropriate correction terms are calculated as discussed in the Theoretical Foundation section above and shown in Table 1.

Table 1: Summary of free energy calculation results.

	No restraints	Ω restraint	r_L, r_P restraint	Ω, r_L, r_P restraint
Grid PMF difference (kcal/mol)	$\Delta G(\boldsymbol{x}_B) = -19.7 \pm 1.1^{*}$	$\Delta G_{\Omega}(\boldsymbol{x}_B) = -13.2 \pm 0.3$	$\Delta G_{r_L, r_P}(\boldsymbol{x}_B) = -17.7 \pm 1.0$	$\Delta G_{\Omega, r_L, r_P}(\boldsymbol{x}_B) = -17.0 \pm 0.5$
Orientation correction (kcal/mol)	N/A	$\Delta U_{\Omega}(\boldsymbol{x}_B) = 4.4 \pm 0.3$	N/A	$\Delta U_{\Omega}^{r_L, r_P}(\boldsymbol{x}_B) = 4.4 \pm 0.3$
Ligand RMSD correction (kcal/mol)	N/A	N/A	$\Delta U_{\mathrm{r_L}}(\boldsymbol{x}_B) = 0.5 \pm 0.1$	$\Delta U_{\mathrm{r_L}}(\boldsymbol{x}_B) = 0.5 \pm 0.1$
Protein RMSD correction (kcal/mol)	N/A	N/A	$\Delta U_{\mathrm{r_P}}^{r_L}(\boldsymbol{x}_B) = 0.8 \pm 0.1$	$\Delta U_{\mathrm{r_P}}^{r_L}(\boldsymbol{x}_B) = 0.8 \pm 0.1$
ΔG_V (kcal/mol)	3.7 ± 0.2	2.5 ± 0.2	2.3 ± 0.2	2.7 ± 0.2
$\Delta G°$ (kcal/mol)	-16.0 ± 1.2	-6.3 ± 0.5	-14.1 ± 1.0	-8.5 ± 0.7
K_d $(\mu M)^{**}$	$O(10^{-6})$	25	$O(10^{-5})$	0.6
K_d range $(\mu M)^{***}$	$10^{-7} - 10^{-5}$	$11 - 58$	$10^{-6} - 10^{-4}$	$0.2 - 2.0$

* All error estimates are based on one standard deviation (s.d.). ** K_d values are determined directly from mean $\Delta G°$ values using Relation (2). *** K_d range is determined from the lower and upper limits of $\Delta G°$ values (mean $\pm$ s.d.) using Relation (2).

The most successful method is expected to be the one employing restraints on $\Omega, \mathrm{r_L}, r_P$. The largest contributor to the free energy is the difference between the grid PMF associated with the heparin hexasaccharide at a grid point at the center of the binding pocket and at any grid point in the bulk, which is -17.0 ± 0.5 kcal/mol (see Figure 1A and Table 1). We denote the PMF of the ligand at a given position $\boldsymbol{x}$ (with respect to the center of the heparin binding pocket) as the grid PMF, since the PMF is estimated at different grid points in this approach.

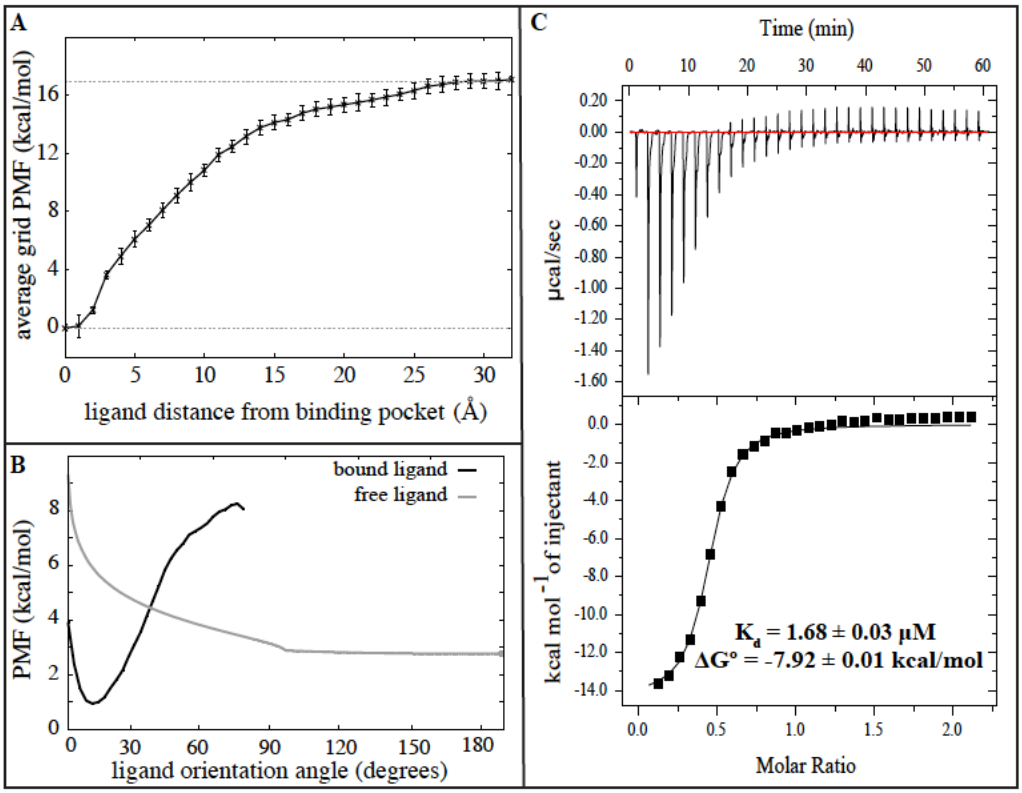


Figure 1. Computational (A-B) and experimental (C) heparin-hFGF1 binding free energy measurements. (A) Average grid PMF in terms of $|\boldsymbol{x}|$,where $\boldsymbol{x}$ is the 3D position vector of the ligand with respect to the center of binding pocket determined from distance-based BEUS simulations with Ω, r_L, r_P restraints. The x axis represents $|\boldsymbol{x}|$ and the y axis represents $\Delta G_{\Omega,r_L,r_P}(|\boldsymbol{x}|)$, which is an average over all $\Delta G_{\Omega,r_L,r_P}(\boldsymbol{x})$ with the same $|\boldsymbol{x}|$, i.e., the ligand distance from the center of binding pocket. The error bar represents the standard deviation obtained from all values of $\Delta G_{\Omega,r_L,r_P}(\boldsymbol{x})$ at various grid points $\boldsymbol{x}$ with the same $|\boldsymbol{x}|$. The dashed line represents the value associated with $\Delta G_{\Omega,r_L,r_P}(|\boldsymbol{x}|)$ at $|\boldsymbol{x}| = 30$ Å. (B) The PMF associated with the ligand orientation angle (Ω) for the bound heparin (i.e., $\boldsymbol{x} \approx \boldsymbol{0}$, ligand in the binding pocket) and free heparin (i.e., $\boldsymbol{x} \approx \boldsymbol{x}_B$, ligand in the bulk). (C) Isothermogram representing the titration of hFGF1 with heparin hexasaccharide. The inset is the experimentally estimated dissociation constant and its associated binding free energy.

The PMF calculations above are based on the BEUS simulations along the protein-ligand distance; however, the orientation and RMSD of the ligand and the RMSD of the protein are restrained to speed up convergence. To account for the orientation bias, a correction term needs to be applied, which is calculated from the PMF associated with the ligand orientation angle at the bulk and binding pocket (Figure 1B). The orientation bias is estimated to be 4.4 ± 0.3

kcal/mol (Table 1). Similarly, a correction term is calculated based on the PMF of the ligand RMSD and that of the protein (Figure 2). These correction terms are estimated to be 0.5 ± 0.1 and 0.8 ± 0.1 kcal/mol, for the ligand and protein, respectively.

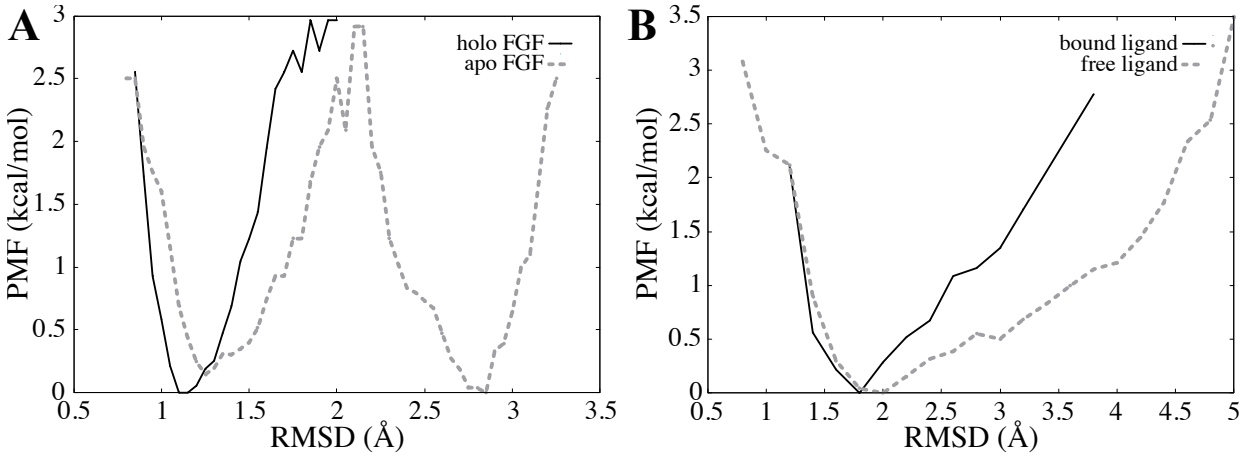


Figure 2. PMF in terms of internal conformational fluctuations of the protein and ligand. (A) PMF associated with the internal RMSD of heparin-bound (solid line) and apo (dashed line) hFGF1, obtained from the equilibrium simulations. (D) PMF associated with the internal RMSD of FGF1-bound (solid line) and free (dashed line) heparin hexasaccharide, obtained from equilibrium simulations.

Finally, another term is needed to account for the difference in the volume accessible to the ligand in the binding pocket and in the bulk (volume contribution). Figure 3 shows that ΔG_P (or V_P) for the distance-based BEUS simulations with no restraint as determined from 20 lowest free energy grid points is almost equal to that obtained from all visited grid points inside or outside the binding pocket. For the distance-based BEUS simulations with $\Omega, \mathrm{r_L}, r_P$ restraints, this term is estimated to be 2.7 ± 0.2 kcal/mol, which results in an absolute binding free energy of -8.5 ± 0.7 kcal/mol. Based on our error analysis, K_d values calculated from the absolute binding free energy were found to be in the micromolar range with an average value of 0.6 μM (using the mean $\Delta G°$ estimate) and ranging from 0.2 to 2.0 μM (based on the lower and upper bounds of free energy estimates). These are in very good agreement with the K_d value obtained from ITC experiments

that is 1.68 μM (Figure 1C). The free energy calculated from the experimental K_d (-7.91 kcal/mol) is also in good agreement with the computationally calculated binding free energy (Figure 1C and Table 1).

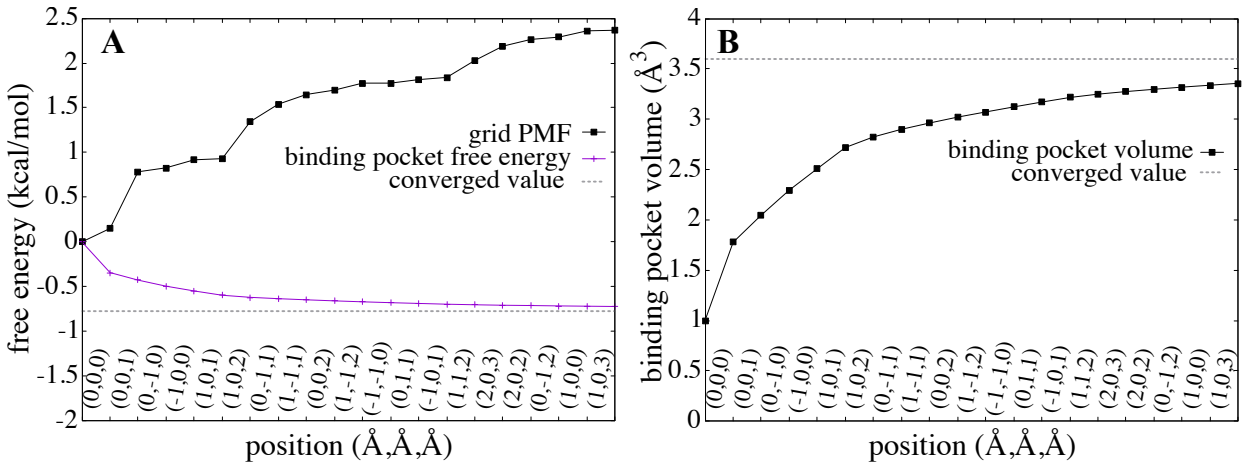


Figure 3. Estimating binding pocket volume (V_P) and its contribution to absolute binding free energy (ΔG_P). (A) Grid PMF ($\Delta G(\boldsymbol{x})$) associated with grid points with the 20 lowest PMF values (black) along with estimated ΔG_P based on the first 20 grid points (shown in an accumulative manner in magenta). The distance-based BEUS simulations with no restraints are used for these calculations. Dashed line shows the estimated ΔG_P based on all visited grid points inside or outside the binding pocket. The x axis shows the position vector of these 20 grid points. (B) Binding pocket volume (V_P) calculated from 20 lowest grid PMF values (similar to A). Dashed line shows the V_P estimated from all visited grid points inside or outside the binding pocket. See the Methods section for more details.

The quantitative agreement between the computational and experimental binding affinity estimates is a great indicator of the accuracy of our absolute binding free energy calculation method. However, if proper restraining is not used as in the distance-based BEUS simulations with no restraints or only RMSD restraints, the binding affinity estimates would be off by several orders of magnitude. The simulations that only restrain the orientation of the ligand are interestingly quite successful as well, being off only by one order of magnitude in terms of binding affinity, which is generally considered a good estimate. This provides some evidence that the orientation of the ligand is perhaps the degree of freedom with the most significant

contribution to the absolute binding free energy besides the ligand-protein distance. We note that the average grid PMF profiles along the ligand-protein distance for the four different methods used here (as shown in Figure 4), confirm the differential behavior of these methods; however, it is important to note that the correction terms should ideally eliminate these differences. This is seen to some extent when comparing the two methods involving orientation restraints that happen to estimate binding affinities that are reasonably close to the experimentally determined value.

Recent computational studies have used the MM-GBSA method to calculate the binding free energy of the hFGF1-heparin interaction, with values ranging from -84.9 kcal/mol to -106.1 kcal/mol[102]. The results obtained from the MM-GBSA approach are very different from our own results, which is to be expected given that MM-GBSA ignores various contributors to the free energy[33,34,35]. Studies have shown that the binding affinity and free energy results derived from computational methods can be compared to experimental binding affinities obtained from ITC experiments[9,10]. However, for a reliable computational free energy estimate, employing purely physics-based free energy calculation methods such as those employed here has proven to be difficult. Here we showed that using a careful strategy that considers all relevant free energy terms and ensures the use of powerful enhanced sampling techniques, could result in good quantitative agreements between the computational and experimental binding affinity estimates.

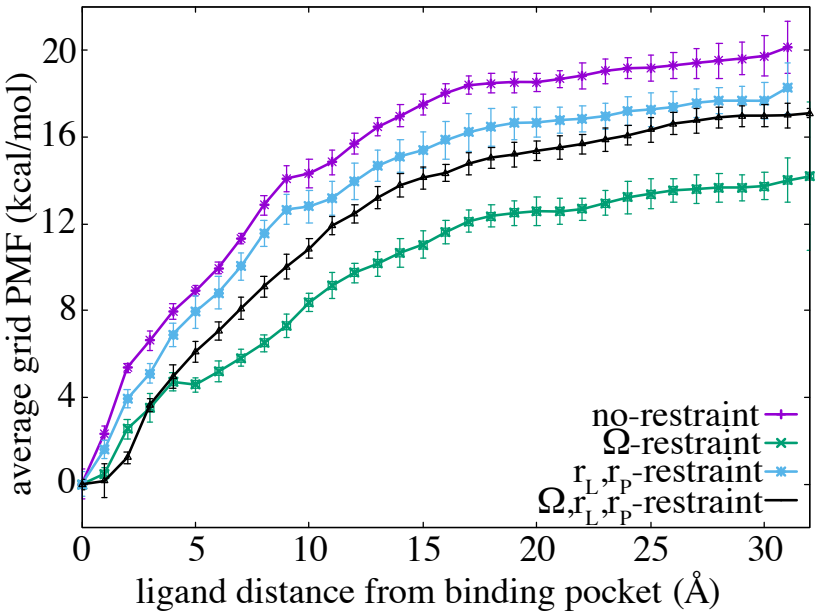


Figure 4. Average grid PMF based on four alternative distance-based BEUS simulations. Average grid PMF in terms of $|\boldsymbol{x}|$,where $\boldsymbol{x}$ is the 3D position vector of the ligand with respect to the center of binding pocket. The x axis represents $|\boldsymbol{x}|$ and the y axis represents $\Delta F(|\boldsymbol{x}|)$, which is an average over all $\Delta F(\boldsymbol{x})$ with the same $|\boldsymbol{x}|$, i.e., the ligand distance from the center of binding pocket. The error bar represents the standard deviation obtained from all values of $\Delta F(\boldsymbol{x})$ at various grid points $\boldsymbol{x}$ with the same $|\boldsymbol{x}|$. The dashed line represents the value associated with $\Delta F(|\boldsymbol{x}|)$ at $|\boldsymbol{x}| = 30$ Å. The inset summarizes different free energy terms involved in the calculation of the absolute binding free energy and the dissociation constant.

The formalism presented in this work has significant similarities to the method previously proposed by Woo and Roux[3], and later implemented[4,66]. However, there are major differences that make the current method more practical. The grid PMF and its various estimates provide a simple conceptual framework to understand how restraining can be accounted for with appropriate correction terms. The average grid PMF in terms of the ligand-protein distance provides an alternative to the PMF in terms of d as is often constructed. Relation (30) is a general scheme that can be easily adapted to any number of restraints. The orientation angle of the ligand with respect to the protein as determined using the orientation quaternion formalism, provides a simple way of determining the absolute binding free energy with a feasible

computational cost. Among the four different sets of restraints, the two involving orientation restraints predict binding free energies similar to that determined experimentally.

ACKNOWLEDGEMENTS

This research is supported by National Science Foundation grant CHE 1945465 and OAC 1940188 and the Arkansas Biosciences Institute. Anton 2 computer time was provided by the Pittsburgh Supercomputing Center (PSC) through Grant R01GM116961 from the National Institutes of Health. The Anton 2 machine at PSC was generously made available by D.E. Shaw Research. This research is also part of the Blue Waters sustained-petascale computing project, which is supported by the National Science Foundation (awards OCI-0725070 and ACI-1238993) and the state of Illinois. This work also used the Extreme Science and Engineering Discovery Environment (allocation MCB150129), which is supported by National Science Foundation grant number ACI-1548562. This research is also supported by the Arkansas High Performance Computing Center, which is funded through multiple National Science Foundation grants and the Arkansas Economic Development Commission. This work is also supported by the Department of Energy (DE-FG02-01ER15161), the National Institutes of Health/National Cancer Institute (NIH/NCI) (1 RO1 CA 172631) and the NIH through the COBRE program (P30 GM103450).

AUTHOR CONTRIBUTIONS

M.M. and T.K.S.K. designed the research. V.G.K performed the simulations and analyzed the simulation data. S.A. performed the experiments and analyzed the experimental data. V.G.K, M.M., T.K.S.K., S.A. wrote the manuscript.

COMPETING INTERESTS

The authors declare no competing interests.

AVAILABILITY OF DATA

All data will be shared upon request to corresponding author.

REFERENCES

1. Mobley, D.L.; Gilson, M.K. Predicting binding free energies: frontiers and benchmarks. *Annu. Rev. Biophys*. **2017**, *46*, 531-58.

2. Wan, S.; Bhati, A.P.; Zasada, S.J.; Coveney, P.V. Rapid, accurate, precise and reproducible ligand–protein binding free energy prediction. *Interface Focus*. **2020**, *10*, 20200007.

3. 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*, 6825-6830.

4. Gumbart, J.C.; Roux, B.; Chipot, C. Standard binding free energies from computer simulations: What is the best strategy? *J. Chem. Theory Comput*. **2013**, *9*, 794–802.

5. Siebenmorgen, T.; Zacharias, M. Computational prediction of protein–protein binding affinities. *WIREs Comput. Mol. Sci*. **2020**, *10*, e1448.

6. Kastritis, P.L., Bonvin, A.M. Are scoring functions in protein-protein docking ready to predict interactomes? Clues from a novel binding affinity benchmark. *J. Proteome Res.* **2010**, *9*, 2216–2225.

7. Gromiha, M.M.; Yugandhar, K.; Jemimah, S. Protein–protein interactions: Scoring schemes and binding affinity. *Curr. Opin. Struct. Biol*. **2017**, *44*, 31–38.

8. Du ,X.; Li, Y.; Xia, Y.L.; Ai, S.M.; Liang, J.; Sang, P.; Ji, X.L.; Liu, S.Q. Insights into Protein-Ligand Interactions: Mechanisms, Models, and Methods. *Int. J. Mol. Sci.* **2016**, *17*, 144.

9. Fenley, A. T.; Henriksen, N. M.; Muddana, H. S.; Gilson, M. K. Bridging Calorimetry and Simulation through Precise Calculations of Cucurbituril-Guest Binding Enthalpies. *J. Chem. Theory Comput.* **2014**, *10*, 4069–4078.

10. Talhout, R.; Villa, A.; Mark, A. E.; Engberts, J. B. F. N. Understanding binding affinity: A combined isothermal titration calorimetry/molecular dynamics study of the binding of

a series of hydrophobically modified benzamidinium chloride inhibitors to trypsin. *J. Am. Chem. Soc*. **2003**, *125*, 10570-10579.

11. Weiss, S. Measuring conformational dynamics of biomolecules by single molecule fluorescence spectroscopy. *Nat. Struct. Biol.* **2000**, *7*, 724–729.

12. Rossi, A.; Taylor, C. Analysis of protein-ligand interactions by fluorescence polarization. *Nat. Protoc.* **2011**, *6*, 365–387.

13. Jönsson, U.; Fägerstam, L.; Ivarsson, B.; Johnsson, B.; Karlsson, R.; Lundh, K.; Löfås, S.; Persson, B.; Roos, H.; Rönnberg, I. et al. Real-time biospecific interaction analysis using surface plasmon resonance and a sensor chip technology. *Biotechniques.* **1991**, *11*, 620–627.

14. Torreri, P.; Ceccarini, M.; Macioce, P.; Petrucci, T.C. Biomolecular interactions by Surface Plasmon Resonance technology. *Ann. Ist. Super Sanita.* **2005**, *41*, 437–441.

15. Kastritis, P.L.; Moal, I.H.; Hwang, H. et al. A structure-based benchmark for protein–protein binding affinity. *Protein Sci.* **2011**, *20*, 482–491.

16. Dixit, S.B.; Chipot, C. Can absolute free energies of association be estimated from molecular mechanical simulations ? The biotin–streptavidin system revisited. *J. Phys. Chem. A.* **2001**, *105*, 9795–9799.

17. Boresch, S.; Tettinger, F.; Leitgeb, M.; Karplus, M. Absolute binding free energies: A quantitative approach to their calculation. *J. Phys. Chem. B.* **2003**, *107*, 9535–9551.

18. Huang, D.; Caflisch, A. Efficient evaluation of binding free energy using continuum electrostatics solvation. *J. Med. Chem.* **2004**, *47*, 5791–5797.

19. Rodinger, T.; Howell, P.L.; Pomès, R. Absolute free energy calculations by thermodynamic integration in four spatial dimensions. *J. Chem. Phys.* **2005**, *123*, 34104.

20. Ytreberg, F.M.; Zuckerman, D.M. Simple estimation of absolute free energies for biomolecules. *J. Chem. Phys.* **2006**, *124*, 104105.

21. Wang, J.; Deng, Y.; Roux, B. Absolute binding free energy calculations using molecular dynamics simulations with restraining potentials. *Biophys. J.* **2006**, *91*, 2798–2814.

22. Rodinger, T.; Howell, P.L.; Pomès, R. Calculation of absolute protein-ligand binding free energy using distributed replica sampling. *J. Chem. Phys.* **2008**, *129*, 155102.

23. Deng, Y.; Roux, B. Computations of standard binding free energies with molecular dynamics simulations. *J. Phys. Chem. B.* **2009**, *113*, 2234–2246.

24. Boyce, S.E.; Mobley, D.L.; Rocklin, G.J.; Graves, A.P.; Dill, K.A.; Shoichet, B.K. Predicting ligand binding affinity with alchemical free energy methods in a polar model binding site. *J. Mol. Biol.* **2009**, *394*, 747–763.

25. Doudou, S.; Burton, N.A.; Henchman, R.H. Standard free energy of binding from a one-dimensional potential of mean force. *J. Chem. Theory Comput.* **2009**, *5*, 909–918.

26. Jiang, W.; Roux, B. Free energy perturbation Hamiltonian replica-exchange molecular dynamics (FEP/H-REMD) for absolute ligand binding free energy calculations. *J. Chem. Theory Comput.* **2010**, *6*, 2559–2565.

27. Singh, N.; Warshel, A. Absolute binding free energy calculations: On the accuracy of computational scoring of protein-ligand interactions. *Proteins.* **2010**, *78*, 1705–1723.

28. General, I.J.; Dragomirova, R.; Meirovitch, H. Absolute free energy of binding of avidin/biotin, revisited. *J. Phys. Chem. B.* **2012**, *116*, 6628–6636.

29. Zhang, C.; Liu, S.; Zhu, Q.; Zhou, Y. A knowledge-based energy function for protein−ligand, protein−protein, and protein−DNA complexes. *J. Med. Chem*. **2005**, *48*, 2325–2335.

30. Chéron, J.-B.; Zacharias, M.; Antonczak, S.; Fiorucci, S. Update of the ATTRACT force field for the prediction of protein–protein binding affinity. *J. Comput. Chem*. **2017**, *38*, 1887–1890.

31. Lensink, M.F.; Wodak, S.J. Docking and scoring protein interactions: CAPRI 2009. *Proteins Struct. Funct. Bioinf*. **2010**, *78*, 3073–3084.

32. Böhm, H.-J. Prediction of Non-bonded Interactions in Drug Design. *In Protein-Ligand Interactions*. Mannhold, R.; Kubinyi, H.; Folkers, G.; Böhm, H.-J.; Schneider, G., Eds.; 2003.

33. 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*, 9401– 9409.

34. Wang, C. H.; Nguyen, P. H.; Pham, K.; Huynh, D.; Le, T. B. N.; Wang, H. L. et al. Calculating protein-ligand binding affinities with MMPBSA: method and error analysis. *J. Comput. Chem.* **2016**, *37*, 2436–2446.

35. Fu, H.; Gumbart, J. C.; Chen, H.; Shao, X.; Cai, W.; Chipot, C. BFEE: A User-Friendly Graphical Interface Facilitating Absolute Binding Free-Energy Calculations. *J. Chem. Inf. Model*. **2018**, *58*, 556–560.

36. Chipot, C. Frontiers in free-energy calculations of biological systems. In *Wiley Interdisciplinary Reviews: Computational Molecular Science*, **2014**, *4*, 71– 89.

37. Chodera, J. D.; Mobley, D. L. Entropy-enthalpy compensation: Role and ramifications in biomolecular ligand recognition and design. *Annu. Rev. Biophys*. **2013**, *42*, 121– 142.

38. *Free energy calculations. Theory and applications in chemistry and biology.* Chipot, C.; Pohorille, A., Eds; Springer Verlag: 2007.

39. Kollman, P.A. Free energy calculations: Applications to chemical and biochemical phenomena. *Chem. Rev.* **1993**, *93*, 2395–2417.

40. Gilson, M.K.; Given, J.A.; Bush, B.L.; McCammon, J.A. The statistical–thermodynamic basis for computation of binding affinities: A critical review. *Biophys. J.* **1997**, *72*, 1047–1069.

41. Hermans, J.; Wang, L. Inclusion of loss of translational and rotational freedom in theoretical estimates of free energies of binding. Application to a complex of benzene and mutant T4 lysozyme. *J. Am. Chem. Soc.* **1997**, *119*, 2707–2714.

42. Fratev, F.; Sirimulla, S. An Improved Free Energy Perturbation FEP+ Sampling Protocol for Flexible Ligand-Binding Domains. *Sci. Rep.* **2019***, 9***,** 16829.

43. Jorgensen, W.L. Free-energy calculations: A breakthrough for modeling organic chemistry in solutions. *Acc. Chem. Res.* **1989**, *22*, 184–189.

44. Hermans, J.; Shankar, S. The free energy of xenon binding to myoglobin from molecular dynamics simulation. *Isr. J. Chem.* **1986**, *27*, 225–227.

45. Roux, B.; Nina, M.; Pomès, R.; Smith, J.C. Thermodynamic stability of water molecules in the bacteriorhodopsin proton channel: A molecular dynamics free energy perturbation study. *Biophys. J.* **1996**, *71*, 670–681.

46. Fujitani, H.; Tanida, Y.; Ito, M.; Jayachandran, G.; Snow, C.D.; Shirts, M.R.; Sorin, E.J.; Pande, V.S. Direct calculation of the binding free energies of FKBP ligands. *J. Chem. Phys.* **2005**, *123*, 084108.

47. Deng, Y.; Roux, B. Calculation of standard binding free energies: Aromatic molecules in the T4 lysozyme L99A mutant. *J. Chem. Theory Comput.* **2006**, *2*, 1255–1273.

48. Karplus, M.; McCammon, J.A. Molecular dynamics simulations of biomolecules. *Nat. Struct. Biol*. **2002**, *9*, 646–652.

49. Dror, R.O.; Dirks, R.M.; Grossman, J.P.; Xu, H.; Shaw, D.E. Biomolecular simulation: A computational microscope for molecular biology. *Annu. Rev. Biophys*. **2012**, *41*, 429–452.

50. Immadisetty, K.; Hettige, J.; Moradi, M. What Can and Cannot Be Learned from Molecular Dynamics Simulations of Bacterial Proton-Coupled Oligopeptide Transporter GkPOT? *J. Phys. Chem. B* **2017**, *121*, 3644-3656.

51. van Gunsteren, W.F.; Mark, A.E. Validation of molecular dynamics simulation. *J. Chem. Phys.* **1998**, *108*, 6109–6116.

52. van Gunsteren, W.F.; Dolenc, J.; Mark, A.E. Molecular simulation as an aid to experimentalists. *Curr. Opin. Struct. Biol.* **2008**, *18*, 149–153.

53. Plattner, N.; Doerr, S.; De Fabritiis, G.; Noé, F. Complete protein–protein association kinetics in atomic detail revealed by molecular dynamics simulations and Markov modelling. Nat. Chem. 2017, 9, 1005–1011.

54. Thayer, K. M.; Lakhani, B.; Beveridge, D. L. Molecular Dynamics–Markov State Model of Protein Ligand Binding and Allostery in CRIB-PDZ: Conformational Selection and Induced Fit. *J. Phys. Chem. B* **2017**, *121*, 5509–5514.

55. Chodera, J.D.; Noé, F. Markov state models of biomolecular conformational dynamics. *Curr. Opin. Struct. Biol*. **2014**, *25*, 135–144.

56. Zuckerman, D.M.; Chong, L.T. Weighted ensemble simulation: Review of methodology, applications, and software. *Annu. Rev. Biophys*. **2017**, *46*, 43–57.

57. Zwier, M.C.; Adelman, J.L.; Kaus, J.W. et al. WESTPA: An interoperable, highly scalable software package for weighted ensemble simulation and analysis. *J. Chem. Theory Comput.* **2015**, *11*, 800–809.

58. Kumar, S.; Rosenberg, J.M.; Bouzida, D.; Swendsen, R.H.; Kollman, P.A. The weighted histogram analysis method for free-energy calculations on biomolecules. *J. Comput. Chem.* **1992**, *13*, 1011–1021.

59. Souaille, M.; Roux, B. Extension to the weighted histogram analysis method: Combining umbrella sampling with free energy calculations. *Comput. Phys. Commun*. **2001**, *135*, 40–57.

60. Luitz, M.; Bomblies, R.; Ostermeir, K.; Zacharias, M. Exploring biomolecular dynamics and interactions using advanced sampling methods. *J. Phys. Condens. Matter*. **2015**, *27*, 323101.

61. Kokubo, H.; Tanaka, T.; Okamoto, Y. Ab initio prediction of protein–ligand binding structures by replica-exchange umbrella sampling simulations. *J. Comput. Chem*. **2011**, *32*, 2810–2821.

62. Moradi, M.; Tajkhorshid, E. Mechanistic picture for conformational transition of a membrane transporter at atomic resolution. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110,* 18916–18921.

63. Kästner, J.; Thiel, W. Bridging the gap between thermodynamic integration and umbrella sampling provides a novel analysis method: "Umbrella integration.". *J. Chem. Phys.* **2005**, *123*, 144104.

64. Barducci, A.; Bussi, G.; Parrinello, M. Well-Tempered Metadynamics: A Smoothly Converging and Tunable Free-Energy Method. *Phys. Rev. Lett*. **2008**, *100*, 020603.

65. Comer, J.; Gumbart, J.C.; Hénin, J.; Lelièvre, T.; Pohorille, A.; Chipot, C. The adaptive biasing force method: Everything you always wanted to know but were afraid to ask. *J. Phys. Chem. B*. **2015**, *119*, 1129–1151.

66. Gumbart, J.C.; Roux, B.; Chipot, C. Efficient determination of protein–protein standard binding free energies from first principles. *J. Chem. Theory Comput*. **2013**, *9*, 3789–3798.

67. Eswarakumar, V. P.; Lax, I.; Schlessinger, J. Cellular Signaling by Fibroblast Growth Factor Receptors. *Cytokine Growth Factor Rev.* **2005**, *16*, 139-149.

68. Beenken, A.; Mohammadi, M. The FGF Family: Biology, Pathophysiology and Therapy. *Nat. Rev. Drug Discovery* **2009**, *8*, 235-253.

69. Kuro-o, M. Endocrine FGFs and Klothos: Emerging Concepts. *Trends Endocrinol. Metab.* **2008**, *19*, 239-245.

70. Ornitz, D. M.; Herr, A. B.; Nilsson, M.; Westman, J.; Svahn, C. M.; Waksman, G. FGF Binding and FGF Receptor Activation by Synthetic Heparan-Derived Di- and Trisaccharides. *Science* **1995**, *268*, 432-436.

71. Culajay, J. F.; Blaber, S. I.; Khurana, A.; Blaber, M. Thermodynamic Characterization of Mutants of Human Fibroblast Growth Factor 1 with an Increased Physiological Half-Life. *Biochemistry* **2000**, *39*, 7153-7158.

72. Mohammadi, M.; Olsen, S. K.; Ibrahimi, O. A. Structural Basis for Fibroblast Growth Factor Receptor Activation. *Cytokine Growth Factor Rev.* **2005**, *16*, 107-137.

73. Johnson, D. E.; Williams, L. T. Structural and Functional Diversity in the FGF Receptor Multigene Family. *Adv. Cancer Res.* **1993***, 60*, 1-41.

74. Lin, X.; Buff, E. M.; Perrimon, N.; Michelson, A. M. Heparan Sulfate Proteoglycans Are Essential for FGF Receptor Signaling during Drosophila Embryonic Development. *Development* **1999**, *126*, 3715-3723.

75. Yayon, A.; Klagsbrun, M.; Esko, J. D.; Leder, P.; Ornitz, D. M. Cell Surface, Heparin-like Molecules Are Required for Binding of Basic Fibroblast Growth Factor to Its High Affinity Receptor. *Cell* **1991**, *64*, 841-848.

76. Rapraeger, A. C.; Krufka, A.; Olwin, B. B. Requirement of Heparan Sulfate for BFGF-Mediated Fibroblast Growth and Myoblast Differentiation. *Science* **1991**, *252*, 1705-1708.

77. Taipale, J.; Keski-Oja, J. Growth Factors in the Extracellular Matrix. *FASEB J.* **1997**, *11*, 51-59.

78. McKeehan, W. L.; Wang, F.; Kan, M. The Heparan Sulfate-Fibroblast Growth Factor Family: Diversity of Structure and Function. *Prog. Nucleic Acid Res. Mol. Biol.* **1998**, *59*, 135-176.

79. Agostino, M.; Gandhi, N. S.; Mancera, R. L. Development and Application of Site Mapping Methods for the Design of Glycosaminoglycans. *Glycobiology* **2014**, *24*, 840-851.

80. Imberty, A.; Lortat-Jacob, H.; Pérez, S. Structural View of Glycosaminoglycan-Protein Interactions. *Carbohydr. Re*s. **2007**, *342*, 430-439.

81. Babik, S.; Samsonov, S. A.; Pisabarro, M. T. Computational Drill down on FGF1-Heparin Interactions through Methodological Evaluation. *Glycoconj. J.* **2017***, 34*, 427-440.

82. Muñoz-García, J. C.; García-Jiménez, M. J.; Carrero, P.; Canales, A.; Jiménez-Barbero, J.; Martín-Lomas, M.; Imberty, A.; de Paz, J. L.; Angulo, J.; Lortat-Jacob, H. et al. Importance of the Polarity of the Glycosaminoglycan Chain on the Interaction with FGF-1. *Glycobiology* **2014**, *24* 1004-1009.

83. Carter, E. P.; Fearon, A. E.; Grose, R. P. Careless Talk Costs Lives: Fibroblast Growth Factor Receptor Signalling and the Consequences of Pathway Malfunction. *Trends Cell Biol*. **2015,** *25*, 221-233.

84. Goetz, R.; Mohammadi, M. Exploring Mechanisms of FGF Signalling through the Lens of Structural Biology. *Nat. Rev. Mol. Cell Biol*. **2013**, *14*, 166-180.

85. Le, V. H.; Buscaglia, R.; Chaires, J. B.; Lewis, E. A. Modeling complex equilibria in isothermal titration calorimetry experiments: thermodynamic parameters estimation for a three-binding-site model. *Anal. Biochem.* **2013**, *434*, 233–241.

86. DiGabriele, A. D.; Lax, I.; Chen, D. I.; Svahn, C. M.; Jaye, M.; Schlessinger, J.; Hendrickson, W. A. Structure of a Heparin-Linked Biologically Active Dimer of Fibroblast Growth Factor. *Nature* **1998**, *393*, 812-817.

87. Bernett, M. J.; Somasundaram, T.; Blaber, M. An Atomic Resolution Structure for Human Fibroblast Growth Factor 1. *Proteins* **2004**, *57*, 626-634.

88. Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K. Scalable Molecular Dynamics with NAMD. *J. Comput. Chem*. **2005**, *26*, 1781-1802.

89. Huang, J.; Rauscher, S.; Nawrocki, G. *et al.* CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nat. Methods.* **2017**, *14*, 71–73.

90. *Large Sparse Sets of Linear Equations*. Reid, J.K., Ed.; Academic Press: London and New York, 1971.

91. Jo, S.; Kim, T.; Im, W. Automated Builder and Database of Protein/Membrane Complexes for Molecular Dynamics Simulations. *PLoS One* **2007**, 2, e880.

92. Lee, J.; Cheng, X.; Swails, J. M.; Yeom, M. S.; Eastman, P. K.; Lemkul, J. A.; Wei, S.; Buckner, J.; Jeong, J. C.; Qi, Y. et al. CHARMM-GUI Input Generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM Simulations Using the CHARMM36 Additive Force Field. *J. Chem. Theory Comput.* **2016**, *12*, 405-413.

93. Martyna, G. J.; Tobias, D. J.; Klein, M. L. Constant Pressure Molecular Dynamics Algorithms. *J. Chem. Phys.* **1994**, *101*, 4177-4189.

94. Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An N·log(N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98*, 10089-10092.

95. Kumar, V. G.; Agrawal, S.; Kumar, T. K. S.; Moradi, M. Mechanistic Picture for Monomeric Human Fibroblast Growth Factor 1 Stabilization by Heparin Binding (preprint). *bioRxiv*. **2020**.

96. Fiorin, G.; Klein, M. L.; Henin, J. Using collective variables to drive molecular dynamics simulations. *Mol. Phys.* **2013***, 111*, 3345-3362.

97. Izrailev, S.; Stepaniants, S.; Balsera, M.; Oono, Y.; Schulten, K. Molecular dynamics study of unbinding of the avidin-biotin complex. *Biophys. J.* **1997***, 72*, 1568-1581.

98. Moradi, M.; Tajkhorshid, E. Computational recipe for efficient description of large-scale conformational changes in biomolecular systems. *J. Chem. Theory Comput.* **2014**, *10,* 2866–2880.

99. Moradi, M.; Enkavi, G.; Tajkhorshid, E. Atomic-level characterization of transport cycle thermo-dynamics in the glycerol-3-phosphate:phosphate transporter. *Nat. Commun.* **2015**, *6*, 8393.

100. Bartels, C. Analyzing biased Monte Carlo and molecular dynamics simulations. *Chem. Phys. Lett.* **2000***, 331*, 446-454.

101. Shirts, M. R.; Chodera, J. D. Statistically optimal analysis of samples from multiple equilibrium states. *J. Chem. Phys.* **2008***, 129*, 124105.

102. Bojarski, K. K.; Sieradzan, A. K.; Samsonov, S. A. Molecular dynamics insights into protein-glycosaminoglycan systems from microsecond-scale simulations. *Biopolymers.* **2019***, 110*, e23252.

Prefusion Spike Protein Conformational Changes Are Slower in SARS-CoV-2 Relative to SARS-CoV-1

Vivek Govind Kumar[1], Dylan S Ogden, Dylan S Ogden,[1] Ugochi H Isu,[1] Adithya Polasa,[1] James Losey,[1] and Mahmoud Moradi[1]

[1]Department of Chemistry and Biochemistry, University of Arkansas, Fayetteville, AR 72701

ABSTRACT

Within the last two decades, severe acute respiratory syndrome (SARS) coronaviruses 1 and 2 (SARS- CoV-1 and SARS-CoV-2) have caused two major outbreaks. For reasons yet to be fully understood the COVID-19 outbreak caused by SARS-CoV-2 has been significantly more widespread than the 2003 SARS epidemic caused by SARS-CoV-1, despite striking similarities between the two viruses. The spike protein,which binds to the host cell angiotensin converting enzyme 2 (ACE2) in both SARS-CoV-1 and 2, has been implied to be a potential source of their differential transmissibility. However, the mechanistic details of prefusion spike protein binding to ACE2 remain elusive at the molecular level. Here, we have used an extensive set of equilibrium and nonequilibrium microsecond-level all-atom molecular dynamics (MD) simulations of SARS-CoV-1 and 2 prefusion spike proteins to determine their differential dynamic behavior. Our results indicate that the active form of the SARS-CoV-2 spike protein is more stable than that of SARS-CoV-1 and the energy barrier associated with the activation is higher in SARS-CoV-2. Our results suggest that not only the receptor binding domain (RBD) but also other domains such as the N-terminal domain (NTD), could play a crucial role in the differential binding behavior of SARS-CoV-1 and 2 spike proteins.

INTRODUCTION

Within the last two decades, SARS CoV-1 [1–3] and 2 [4–8] (CoV-1 and CoV-2, respectively) have caused SARS epidemic and coronavirus disease 2019 (COVID-19) pandemic, respectively. Various studies have shown that CoV-2 is more easily transmissible between humans in comparison to CoV-1 [9–12]. However, given the striking similarity of the two viruses, the molecular-level explanation of their differential transmissibility is largely missing. The two viruses share several highly conserved structural and functional features [4, 13, 14]. The spike protein plays a crucial role in the infection process [13, 15–17] and has been the primary target of various candidate drugs and vaccines [18–26].

CoV-1 and CoV-2 spike proteins have a high sequence identity of approximately 79% [4] and the RBDs of both proteins interact with the human ACE2 receptor [9, 16, 17, 27–31]. Studies have shown that several regions of the CoV-2 spike protein are susceptible to mutations, with the RBD being particularly vulnerable in this regard [32–35]. It is possible that therapeutic agents targeting only the RBD-ACE2 interaction might eventually be rendered ineffective due to the appearance of emerging variants. Therefore, diversifying the hot spots of the protein being targeted by therapeutics and vaccines is essential in increasing their long-term efficacy. The current study provides a rational framework for such directions by systematically studying the differential behavior of the CoV-1 and CoV-2 spike proteins, highlighting significant regions of the protein that are involved in the activation process, i.e., a large-scale conformational change in the prefusion spike protein, which occurs prior to ACE2 binding.

Recently, several cryogenic electron microscopy (cryo-EM) and computational studies have shed light on the differential receptor binding behavior of the CoV-1 and CoV-2 spike proteins [9, 17, 27, 36, 37]. The RBD of the spike protein undergoes a large-scale conformational

transition from an inactive “down” position to an active “up” position in order to access the ACE2 receptors on the host-cell surface [9, 17, 27, 38–40]. Experimental studies investigating the binding affinity of the spike protein RBD for the ACE2-peptidase domain (PD) have produced varying results. Using surface plasmon resonance (SPR) and flow cytometry techniques, respectively, Wrapp et al. [27] and Tai et. al. [17] have reported that the CoV-2 RBD has a higher binding affinity for ACE2-PD than the CoV-1 RBD. For instance, the SPR-based assay shows that the dissociation constant of the CoV-2 spike protein ($K_d \approx 14.7\ nM$) is 10-20 times lower than that of the CoV- 1 spike protein [27, 41]. In a different study, biolayer interferometry has shown that the CoV-2 dissociation constant ($K_d \approx 1.2\ nM$) is only 4 times lower than that of CoV-1, indicating that the binding affinities are generally comparable [9]. Such quantitative inconsistencies emphasize the need to improve our understanding of the mechanistic aspects of the RBD-ACE2 interaction. A disadvantage of experimental techniques like SPR and biolayer interferometry is that they require the protein to be immobilized prior to measuring the binding affinity [42, 43]. This introduces a level of bias into these experimental assays, particularly if the binding behavior of a protein is conformation-dependent, as is the case for the coronavirus spike proteins. One may argue that some studies have neglected the fact that the binding process involves not only the RBD- ACE2 interaction but also the spike protein activation, a large-scale conformational change with a potentially significant contribution to the differential binding behavior of SARS-CoV-1 and 2. Therefore, to gain a clearer understanding of the enhanced infectivity of SARS-CoV-2, “effective binding” involving both the RBD-ACE2 interaction and the spike protein activation/inactivation process needs to be investigated. Here, we focus on the latter, which has received less attention in the literature.

Cryo-EM studies have successfully resolved structures of both spike proteins in the inactive state, active unbound state, and active ACE2-bound state [9, 27, 31, 38, 44]. However, cryo-EM and X-ray crystallography studies essentially capture static pictures of specific protein conformations [45–47]. In addition, given the substantial differences in the experimental and physiological conditions, it is not clear whether all relevant conformational states are captured using these techniques. For instance, a recent single-molecule fluorescence resonance energy transfer (smFRET) study has captured an alternative inactive conformation for the CoV-2 spike protein [48] that is not consistent with those obtained from cryo-EM. It is thus important to investigate the differential conformational landscapes of the CoV-1 and CoV-2 spike proteins in terms of both important functional states and their dynamics. For this purpose, we use an extensive set of microsecond-level unbiased and biased MD simulations. Here, we make certain assumptions to be able to make progress towards deciphering the differential behavior of the two spike proteins, such as relying on cryo-EM structures as our initial models, excluding the unresolved transmembrane domain of the spike protein, and excluding the glycan chains in the simulations. However, we treat the spike proteins of both viruses similarly so that a reliable comparison can be made.

Allowing for the fact that this study has certain limitations as discussed previously, our extensive all-atom equilibrium MD simulations show that the active CoV-2 spike protein is potentially more stable than the active CoV-1 spike protein. We also report that the RBD of the active CoV-1 spike protein can undergo a spontaneous conformational transition to a pseudo-inactive state characterized by the interaction of the NTD and RBD, a state not observed in any of the previous experimentally reported structures of CoV-1 or CoV-2 spike protein. This observation is broadly in line with the recent smFRET experimental results indicating the

presence of alternative inactive spike protein conformations [48]. More specifically, electrostatic interaction analyses reveal that unique salt-bridge interactions between the NTD and RBD of the CoV-1 spike protein, are involved in the major conformational transition observed in our simulations. No large-scale conformational changes occur in any of the active CoV-2 spike protein simulations or any of the inactive CoV-1 or CoV-2 spike protein simulations within the timescale of our unbiased MD simulations (5 μs).

In order to investigate the longer timescale conformational dynamics inaccessible to unbiased simulations [49], we have also employed extensive steered MD (SMD) simulations [50] along with nonequilibrium work calculations [51] to make a semi-quantitative comparison between the two proteins [52, 53]. The SMD simulations shed light on the energetics of the conformational change associated with the activation and inactivation processes. The results obtained from these enhanced simulations strongly suggest that the energy barriers for such conformational transitions are significantly lower for the CoV-1 spike protein and that conformational changes occur more slowly for the CoV-2 spike protein. This provides an explanation for the conformational plasticity displayed by the active CoV-1 spike protein in our simulations as well as the relative conformational stability of the active CoV-2 spike protein. The results from our equilibrium and nonequilibrium simulations thus provide a self-constituent picture of the long timescale conformational dynamics of the CoV-1 and CoV-2 spike proteins. We note that our results are not conclusive with regards to the thermodynamics of activation and inactivation. Instead, they provide a semi-quantitative picture of the kinetics. The propensity of the active CoV-2 spike protein to maintain the "up" RBD conformation for a longer period of time as compared to CoV-1 might explain why the CoV-2 has a better chance of remaining

bound to ACE2 long enough to allow for the next step in the viral entry process, which in turn could potentially be linked to the CoV-2's comparatively high human-to-human transmissibility.

RESULTS

We have performed 5-μs-long unbiased all-atom MD simulations of both inactive and active CoV-1 and CoV-2 spike proteins in explicit water. The active CoV-1 and CoV- simulations were repeated additionally twice for another 5 μs each (see Supporting Information - MD Simulation details). We have also performed 80 independent nonequilibrium SMD simulations of the CoV-1 and 2 spike proteins, each for 100 ns, to compare the activation and inactivation of CoV-1 and CoV-2 spike proteins that are otherwise generally inaccessible to unbiased MD. We have thus generated 40 μs of equilibrium and 8 μs of nonequilibrium simulation trajectories in aggregate.

Within the timescale of our unbiased equilibrium simulations (i.e., 5 μs), the inactive forms of both CoV-1 and CoV-2 spike proteins do not undergo any major conformational transitions, with the RBDs remaining in the "down" position (Fig. 1A) [9, 38]. On the other hand, a spontaneous large-scale conformational change occurs in the active CoV-1 spike protein simulation (Fig. 1B), with the RBD moving from an active "up" position to a pseudo-inactive "down" conformation that is different from the inactive conformation in the cryo-EM structure [38]. This spontaneous conformational transition appears to occur due to interactions between the NTD and RBD of the CoV-1 spike protein (Fig. 1B). Unlike the active CoV-1, the active CoV-2 spike protein does not undergo any large-scale conformational transitions and remains in the active state within the 5-μs simulations (Fig. 1B). Movie S1 in Supporting Information demonstrates the differential behavior of CoV-1 and CoV-2 clearly.

To examine the reproducibility of the above observations, the active CoV-1 and -2 simulations were repeated twice (see Supporting Information - MD Simulation details). Consistent with Set 1, the active CoV-2 simulations do not show any significant conformational change in Sets 2 and 3. The active CoV-1 simulations, on the other hand, undergo some significant conformational change in Set 2 and Set 3; although these conformational changes are not the same in the three different repeats. The dramatic change from the "up" to "down" (or pseudo-inactive) conformation of the CoV-1 spike protein is only observed in Set 1; however, all three sets show some significant conformational changes that are not observed in any of the CoV-2 simulations. Root mean square deviation (RMSD) (Fig. S1) and root mean square fluctuation (RMSF) (Fig. S2) analyses demonstrate the relative stability of the active CoV-2 as compared to the active CoV-1 spike protein. A comparison of individual protomer RMSDs from all 3 repeats of the active CoV-1 and CoV-2 spike protein trajectories, clearly shows that the active CoV-1 spike protein is less stable overall as compared to the active CoV-2 (Fig. S1). Similarly, RMSF analysis indicates that the RBD and NTD regions of the active CoV-1 spike protein fluctuate more than the corresponding regions of the active CoV-2 (Fig. S2).

In order to quantify the spontaneous conformational transition that occurs in the active CoV-1 spike protein, we measured the center-of-mass distance between the receptor-binding motif (RBM) of protomer A and the S2 trimer of the spike protein (Fig. 1C). The RBM-S2 distance remains stable for both inactive states at ~85 Å over 5 μs. For both the CoV-1 and CoV-2 active states, the RBM-S2 distance is initially ~100 Å but decreases to ~85 Å for CoV-1 after 2 μs (Fig. 1C). This analysis clearly demonstrates that the final conformation adopted by the RBD of the active CoV-1 spike is similar to the inactive state RBD conformations of both CoV-1 and CoV-2, in terms of the RBM-S2 trimer distance (Figure 1C). On the other hand, the

RBM-S2 trimer distance for the active CoV-2 spike protein remains relatively unchanged over 5 μs (Fig. 1C), consistent with the molecular images shown in Figs. 1A-B. Similarly, the angle between the RBM of protomer A and the S2 trimer remains relatively unchanged for the CoV-2 active state, while the CoV-1 active simulation shows a behavior during the last 3 μs that is similar to that of the inactive states of CoV-1 and CoV-2 (Fig. 1D).

The RBD-NTD contact analysis also demonstrates the RBD-NTD association in the so-called pseudo-inactive conformation observed in our CoV-1 simulations. We specifically calculated the minimum distance between the RBD and NTD of protomer A for each system (Fig. 1E). While the RBM-S2 distance and angle calculations indicate that the behavior of the CoV-1 active state eventually resembles that of both inactive systems (Fig. 1C-D), the NTD-RBD distance calculation showcases the unique behavior of the pseudo-inactive CoV-1 spike protein. The NTD-RBD distance of the active protomer in CoV-1 fluctuates considerably over the first 2 μs of the trajectory, after which it decreases sharply to settle down around 2 Å (Figure 1E). This clearly demonstrates that the RBD of the pseudo-inactive CoV-1 spike protein, that results from the inactivation of the active CoV-1 spike, is in close proximity to the NTD as also shown in the cartoon representations (Fig. 1B). This is not observed during any of the simulations of active CoV-2 spike protein or either of the inactive spike proteins (Fig. 1A-B, 1E), thus indicating that the pseudo-inactive conformation adopted by the initially active CoV-1 spike protein is unique.

The RBM hydration analysis provides more evidence that the pseudo-inactive CoV-1 is truly inactive since its exposure to water (as a proxy to ACE2 accessibility) is quite similar to that of inactive CoV-1 and 2 states. This is quantified using the estimated probability distribution for the number of water molecules near the RBM during the last 500 ns of simulations (Fig. 1F).

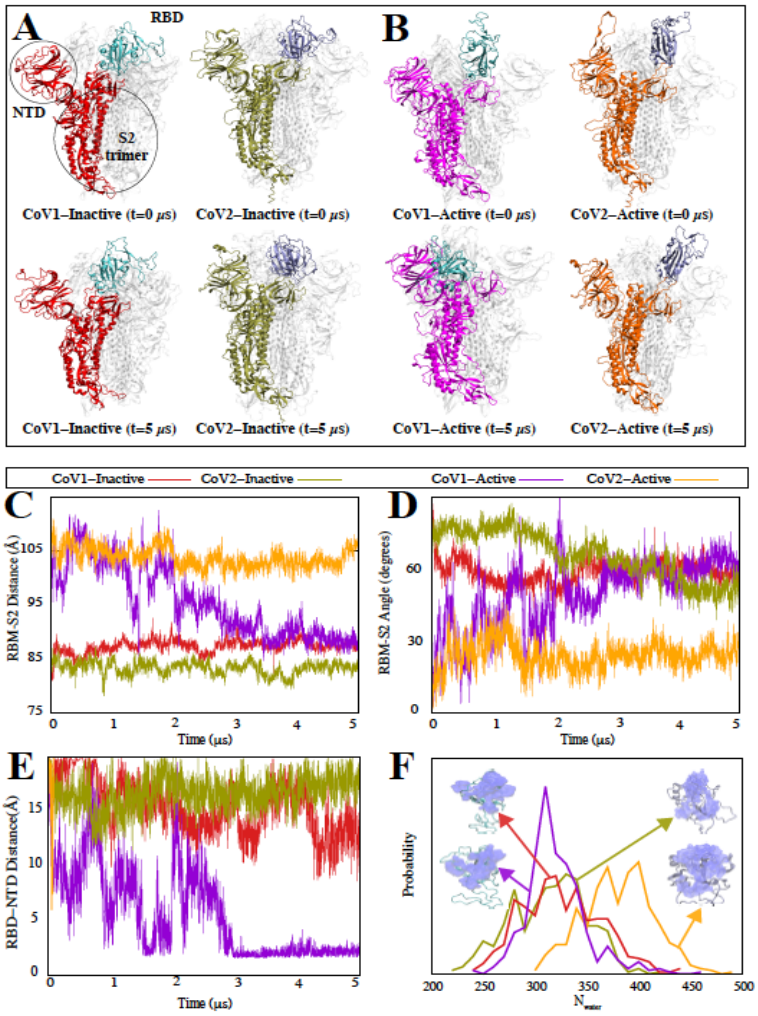


FIG. 1. Unbiased simulations of the CoV-1 and CoV-2 spike proteins show a differential dynamic behavior. (A-B) The initial and final MD snapshots of CoV-1 and CoV-2 spike proteins starting from both inactive and active states. Protomer A in each protein is colored and protomers B and C are shown in white. The RBD of the colored protomer has a distinctive color from the rest of the protomer. Based on multiple repeats of these simulations, we have observed that the active form of the CoV-2 spike protein is consistently more stable than the active CoV-1 spikeprotein. The active CoV-1 spike protein transitions spontaneously to a pseudo-inactive conformation. **(C)** The center-of-mass distance between the S2 trimer of the spike protein and the RBM of protomer A shown as a function of time. **(D)** The angle between the S2 trimer of the spike protein and RBM of protomer A shown as a function of time. **(E)** Minimum distance between the NTD and RBD of protomer A as a function of time for CoV-1 and CoV-2 spike proteins in both active and inactive state simulations. **(F)** Probability density map of water within 5 Å of the RBM for the final 500 ns of simulation. In panels C-F, the same color code is used to represent CoV-1-inactive (blue), CoV-1-active (magenta), CoV-2-inactive (red) and CoV-2-active (orange).

The water molecule count for the pseudo-inactive state (here, represented by the last 500 ns of the simulation starting with the CoV-1 active state) is considerably lower than that of the CoV-2 active state and is comparable to the counts for the CoV-1/2 inactive states, further confirming that the active CoV-1 spike protein undergoes a large-scale conformational transition and becomes inactive (Fig. 1F).

While the measures discussed above provide clear evidence that the CoV-2 spike protein behaves more as a stable structure in its active state as compared to CoV-1 spike protein, more insight can be gained from more systematic analysis techniques such as principal component analysis (PCA) [54] and dynamic network analysis (DNA) [55]. For instance, considering the (PC1,PC2) space shows that the region sampled by the active protomer of the CoV-1 spike protein is considerably larger than the region sampled by the corresponding protomers of the CoV-2 spike protein (Fig. S3). Interestingly, the PCA analysis reveals that the most pronounced conformational change (i.e., PC1) is related to the motion of the RBD towards the NTD in the CoV-1 spike protein (Figure S3). For more PCA based analysis, see Supporting Discussion and Figures S3-S5 in Supporting Information. Similarly, the DNA analysis provides more details on the differential behavior of the spike proteins of CoV-1 and CoV-2. For instance, CoV-1 protomer A (i.e., the active protomer) shows several high inter-domain correlations (indicating concerted motions), while these correlations are missing in the same protomer of CoV-2 (Fig. S6). Similar trends were observed in all three Sets of CoV-1 and CoV-2 active state simulations (Figs. S7-S10). Inter-protomer correlations also highlight the differential behavior of the active CoV-1 and CoV-2 spike proteins (Fig. S11). For more DNA based analysis, see Supporting Discussion and Figures S6-S11 in Supporting Information.

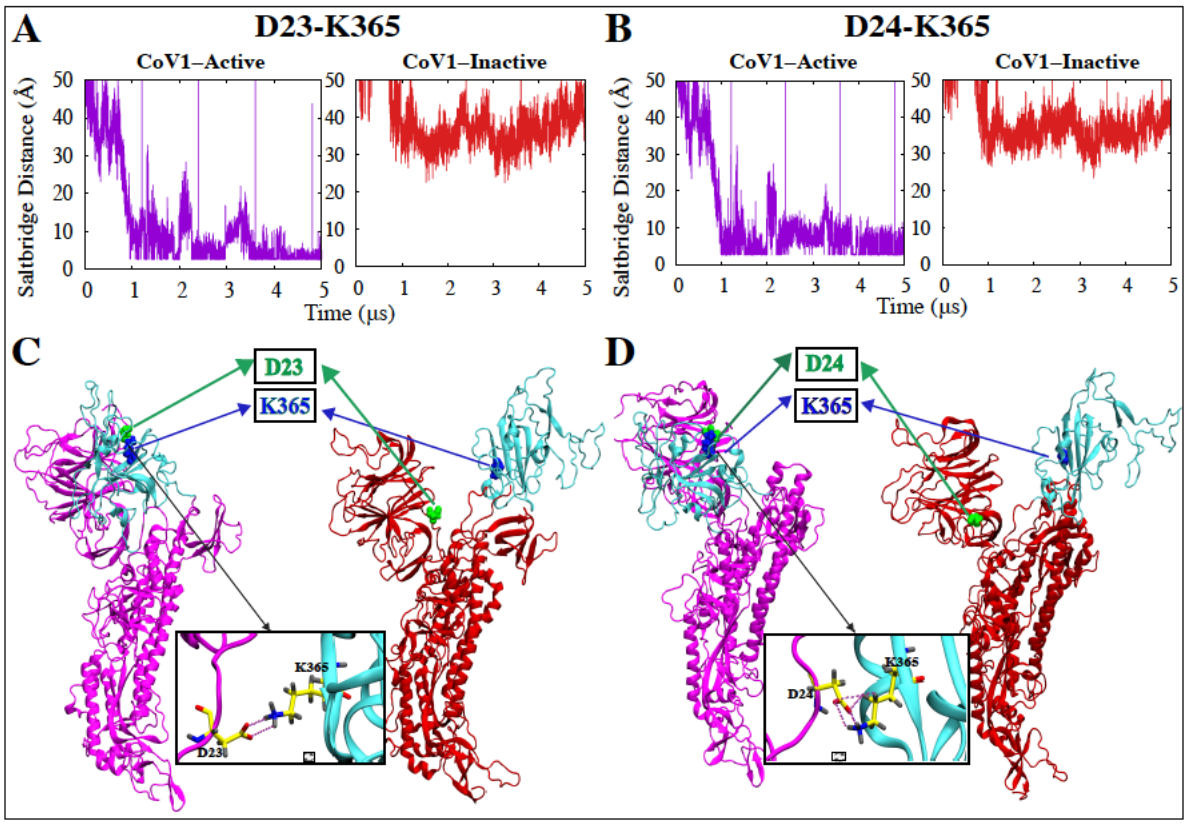


FIG. 2. Unique salt-bridge interactions between the RBD and NTD of the active CoV-1 spike protomer facilitate the transition to a pseudo-inactive conformation. (A-B) Time series of D23/24-K365 (A/B) salt-bridge distances in CoV-1 spike protein simulations and (C-D) visual representations of salt-bridge formation in the initially active CoV-1 protomer A. D23 and D24 (green) of the NTD form a salt-bridge with K365 (blue) of the RBD only in the pseudo-inactive state of CoV-1. D23 and D24 are not present in the CoV-2 spike protein.

Our extensive electrostatic interaction analysis reveals that the driving force behind the unique conformational transition observed in the initially active CoV-1 spike protein simulation (Fig. 1) is at least partly a set of salt-bridge interactions that are unique to CoV-1. Residues D23 and D24 in the NTD interact with K365 in the RBD, forming stable salt bridges in the active CoV-1 spike protein but not in the inactive state (Fig. 2). These fairly stable salt-bridges form around the 1 μs mark (Fig. 2A,B), prior to the final movement of the RBD towards the NTD (Fig. 1E). Residues D23 and D24 are not conserved in the SARS-CoV-2 spike protein. Differential behavior is also observed for two sets of residues that are conserved in both CoV-1 and CoV-2 spike proteins (Fig.S12). R328 and D578 form a stable salt bridge in both active and

inactive CoV-2 spike proteins while R315 and D564 do not form a salt-bridge in the CoV-1 spike proteins (Fig. S12A). Similarly, R273 and D290 form a stable salt bridge in both active and inactive CoV-2 spike proteins while K258 and D277 do not form a salt-bridge in the CoV-1 spike proteins (Fig. S12B). Additionally, a conserved pair of residues form an intra-RBD hydrogen bond in the active/inactive CoV-2 spike protein (Y396-E516) and the inactive CoV-1 spike protein (Y383-E502), but not in the active CoV-1 spike protein (Y383-E502) (Fig. S13). These electrostatic interactions thus potentially contribute to the relative stability of the active SARS-CoV-2 spike protein.

SMD simulations were performed to semi-quantitatively characterize the energetics of the activation-inactivation process for the CoV-1 and 2 spike proteins. To induce the activation or inactivation of individual protomers, we used the Cα RMSD of each protomer with respect to a target structure (the inactive state for the inactivation process and the active state for the activation process). 10 sets of 100 ns SMD simulations were performed for each system. The conformational transition of an inactive RBD to the active "up" position was accompanied by a decrease in the RBM-S2 angle and an increase in the RBM-S2 distance, as expected (Figure 3A-B). Similarly, the inactivation of an active protomer was characterized by an increase in the RBM-S2 angle and a decrease in the RBM-S2 distance, as expected (Fig. 3A-B).

Without performing strict free-energy calculations, we have used nonequilibrium work measurements to compare the energetics of the CoV-1/CoV-2 spike protein activation-inactivation process in a semi-quantitative manner. We have previously used similar methods to investigate conformational transitions of other biomolecular systems [52, 53, 56, 57]. The accumulated nonequilibrium work measured during the inactivation of an initially active CoV-2 protomer or the activation of an initially inactive CoV-2 protomer, is significantly larger than the work measured during the inactivation or activation of the corresponding CoV-1 protomer (Fig. 3C-D). Similarly, the change in the associated Jarzynski average is also much higher for the CoV-

2 protomers (Fig. 3C-D, inset). We note that the Jarzynski average would only quantify the true free energy if converged, which requires many more repeats. However, here we are only interested inrelative behavior of the CoV-1 and CoV-2 in a qualitative or semi-quantitative manner rather than accurately calculating any free energies [52, 56, 57]. These results suggest that the CoV-2 spike protein has slower kinetics, due to higher barriers, in both directions. In other words, the conformational changes associated with activation or inactivation of the spike protein proceeds more slowly in CoV-2 relative to CoV-1. This is in good agreement with our observations on the relative conformational stability of the active CoV-2 spike protein from the unbiased simulations. The difference in the kinetics explains why we have been able to observe large-scale conformational changes in some of the SARS-CoV-1 spike protein simulations but not in any of the SARS-CoV-2 spike protein simulations. It is also important to note that the work analysis here does not provide much information on the thermodynamics. To be able to make statements about thermodynamics,we need to perform very accurate free energy calculations.

Our SMD simulations show that it is relatively difficult for the CoV-2 spike protein to undergo a large-scale conformational transition between active and inactive states, when compared to the CoV-1 spike protein. Although these SMD simulations were run using the full trimers, they involved only a single protomer (protomer A) in the biasing schemes while the other two protomers were not biased. These simulations were also repeated with all 3 protomers being biased (Fig. S14), which verified the large difference between the CoV-1 and CoV-2 kinetics. Our results indicate that the energy barriers associated with conformational changes that are required for activation and inactivation are larger in the CoV-2 spike protein as compared to CoV-1.

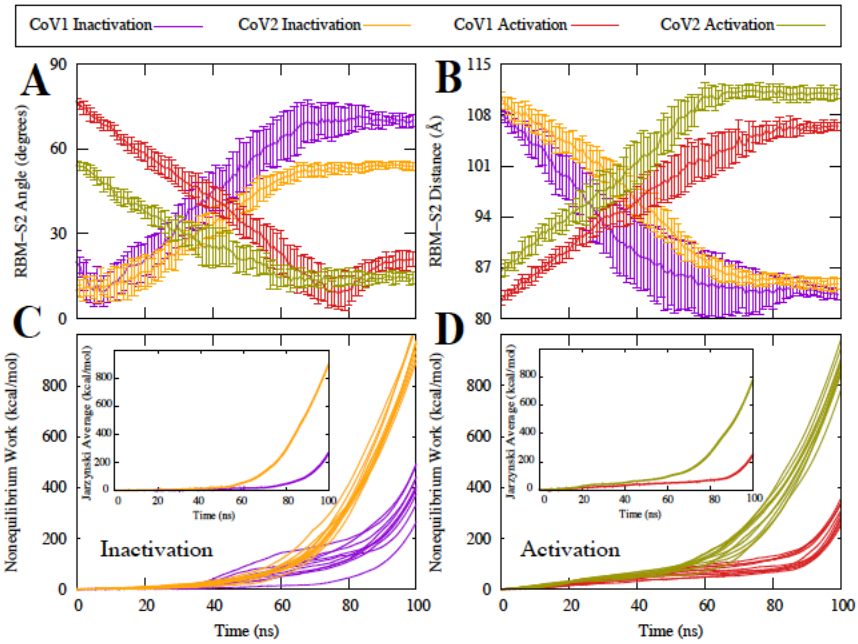


FIG. 3. SMD simulations show that the CoV-2 spike protein has higher energy barriers between active and inactive states as compared to the CoV-1 spike protein. (A) RBM-S2 angle between the beta sheet region of the RBM and the alpha helical region of S2, shown as a function of time during SMD simulations. Protomer activation is characterized by a decrease in the RBM-S2 angle. **(B)** RBM-S2 COM Distance between the beta sheet regionof the RBM and the alpha helical region of S2, as shown as a function of time during SMD simulations. Protomer activation is characterized by an increase in the RBM-S2 distance.**(C,D)** Accumulated non-equilibrium work as a function of time during SMD simulations for individual simulations. Inset: The Jarzynski average over 10 individual work profiles shown in Panels C and D.

DISCUSSION

Using microsecond-level equilibrium and nonequilibrium MD simulations, we have demonstrated that the active CoV-1 and CoV-2 spike proteins exhibit differential dynamic behavior. The active CoV-2 spike protein remains relatively stable over 5 μs, whereas the active CoV-1 spike protein undergoes conformational changes and adopts, at least in one simulation, a pseudo-inactive conformation that is distinct from the well-characterized inactive "RBD-down" conformation [38]. Our observation of a pseudo-inactive state of the CoV-1 spike protein essentially agrees with the results of an experimental smFRET study that describes alternative inactive states of the CoV-2 spike protein [48]. While this pseudo-inactive conformation is not

observed in our CoV-2 spike protein simulations, it is certainly plausible that the CoV-2 spike protein samples alternative conformational states during the spike protein activation process that is dependent on the experimental/physiological conditions. In general, the key conclusion from the observation of this pseudo-inactive state is that the published cryo-EM structures that are produced under non-physiological conditions do not necessarily represent all relevant conformational states of the spike protein.

While our unbiased simulations provide some insight into the spike protein inactivation process, SMD simulations can access longer timescale conformational dynamics which allows for a more detailed characterization of both activation and inactivation. An investigation of the energetics of the activation-inactivation process using SMD simulations revealed that relative to CoV-1, it is difficult for the CoV-2 spike protein to undergo a major conformational transition from the active state to the inactive state or vice-versa. Nonequilibrium work measurements indicate that large-scale conformational transitions occur relatively slowly in the CoV-2 spike protein, which complements our observations on the relative conformational stability of the active CoV-2 spike protein from the equilibrium simulations, explaining the spontaneous conformational transition observed in the initially active CoV-1 equilibrium trajectory. The results from our equilibrium and nonequilibrium simulations are thus very consistent and provide extensive insights into the long-term dynamics of the CoV-1 and CoV-2 spike proteins. A recent computational study has shown that the RBD of the CoV-2 spike protein has greater mechanical stability than the RBD of the CoV-1 spike protein [58], which agrees with our observations on the conformational stability of the active CoV-2 spike protein.

Several cryo-EM studies have reported differing results on the propensity of the CoV-1 and CoV-2 spike proteins to adopt certain conformations (eg. 1 RBD "up" or 3 RBDs "down"). For instance, Kirchdoerfer et al. state that the single RBD "up" conformation is highly favored by the CoV-1 spike protein, with 58% of particles belonging to this population [41]. They did not

observe the 3 RBDs "down" conformation [41]. On the other hand, Yuan et al. and Gui et al. report that particles in the 3 RBDs "down" conformation make up approximately 56% and 27% of the population respectively [38, 59]. Similarly, for the CoV-2 spike protein, Walls et al. observe an approximately even split between the 1 RBD "up" and 3 RBDs "down" conformations while Wrapp et al. only observe the 1 RBD "up" conformation [9, 27]. In our study, we do not make any claims about the predominance or relative stability of these conformations for the CoV-1 or CoV-2 spike protein. Instead, we focus exclusively on the differential dynamic behavior of the CoV-1 active and CoV-2 active spike proteins. Our study provides new insights into the kinetics, and not the thermodynamics, of the CoV-1 and CoV-2 spike protein activation process.

Using surface plasmon resonance and protein pull-down assays, Shang et al. have shown that the CoV-2 spike RBD has significantly higher ACE2 binding affinity than the CoV-1 spike RBD [60]. However, their results also indicate that the ACE2 binding affinity of the entire CoV-2 spike protein is similar to or lower than that of the CoV-1 spike protein [60]. To explain this "paradox", the authors hypothesize that although the CoV-2 has a higher-affinity RBD as compared to CoV-1, the CoV-1 favors the up state of the RBD more than the CoV-2 and thus has a higher accessibility to ACE2. Since we do not make any claims regarding the thermodynamics (i.e., up vs down stability), we can neither rule out nor provide evidence for this hypothesis based on our simulations. However, given the fact that the spike-ACE2 binding is only the first step in a cascade of events that result in S1-S2 cleavage and membrane fusion, it is important for the spike-ACE2 association to last long enough so the rest of the process is triggered. Therefore, kinetics is perhaps as important if not more important than thermodynamics here. Our hypothesis based on the slower kinetics is that once the SARS-CoV-2 spike protein is activated, not only is it ready to bind to ACE2 but it also favors staying bound for a long enough time such that a cascade of events necessary for membrane fusion can occur.

Unlike X-ray crystallography and cryo-EM, MD simulations facilitate the elucidation of detailed hypotheses on the dynamic behavior of proteins and other biomolecules [46, 47]. However, each computational or experimental technique has its own assumptions and limitations. Here, for instance, we chose to work with the non-glycosylated spike proteins of CoV-1 and 2 to avoid complications when making comparisons. A recent study has shown that glycosylation of the spike proteins might play an important role in the conformational dynamics of the RBD [61, 62]. At this stage, we have not simulated the glycosylated spike proteins due to the difficulty of modeling the correct glycan chains. It would be quite difficult to determine whether conformational changes occur as a result of the intrinsic protein dynamics or the differential glycosylation patterns of the CoV-1 and CoV-2 spike proteins imposed by our modeling. However, we use the non-glycosylated form of the spike protein for both CoV-1 and 2, which makes the comparison justifiable.

Investigation of the "effective binding" process involving both receptor interaction and spike protein activation will provide deeper insights into the enhanced infectivity of SARS-CoV-2. Several studies have investigated RBD-ACE2 binding for both SARS CoV-1 [63–69] and SARS-CoV-2 [9, 17, 27, 36, 37], while ignoring the conformational dynamics of spike protein activation and inactivation. We propose that the “effective binding” process is different in the CoV-1 and CoV-2 spike proteins, not only because of the variability of the RBD but also due to the contribution of other regions, particularly the NTD, as seen in the CoV-1 pseudo-inactive state, where the NTD interacts with RBD and therefore could block the ACE2 binding to RBD. This is in qualitative agreement with the results of recent experimental and clinical studies which highlight the importance of the spike protein NTD in the SARS-CoV-2 infection process [70–75].

Several circulating SARS-CoV-2 variants with mutations or deletions in the NTD show greatly reduced recognition by NTD-specific neutralizing monoclonal antibodies [70–75]. This

strongly suggests that the NTD is under selective pressure from the host humoral immune response [70–75]. Based on the observation of a previously unknown pseudo-inactive conformational state for the spike protein, we hypothesize that the RBD-NTD interaction could play a crucial role in the inactivation of the spike protein and that mutations in the spike NTD could potentially have an effect on the transmissibility of the coronavirus. More generally, our simulations suggest that the differential conformational dynamics associated with inactivation and activation of the coronavirus spike protein might contribute to the increased transmissibility of SARS-CoV-2 as compared to SARS-CoV-1 and some variants of SARS-CoV-2 as compared to some other variants.

Several experiments could be performed in order to test the hypotheses presented in our computational study. For instance, the importance of residues D23 and D24 from the CoV-1 spike NTD could be investigated via site-directed mutagenesis. This might provide some additional insights on the conformational dynamics of the CoV-1 spike protein. Similarly, the conserved residue pairs that exhibit differential behavior in terms of salt-bridge interactions could be mutated in both spike proteins. Additionally, smFRET experiments could be used to investigate a potential RBD-NTD interaction by measuring the distance between fluorophores attached to each domain. Disulfide cross-linking experiments could also be used to investigate residues in the NTD and RBD that potentially interact with each other.

As discussed previously, our study primarily sheds light on the conformational dynamics of the SARS-CoV-1 and SARS-CoV-2 spike proteins. While differences in the dynamic behavior of these spike proteins almost certainly contribute to differences in transmissibility and infectivity, factors such as spike protein glycosylation and the behavior of other viral proteins also need to be considered in order to provide a more complete hypothesis. Additional experimental and computational studies are thus needed to fully investigate the differential infectivity and transmissibility of SARS-CoV-1 and SARS-CoV-2.

METHODS

Our simulations were based on cryo-EM structures of the SARS-CoV-2 spike protein in the active (PDB entry:6VYB) [9] and inactive (PDB entry:6VXX) [9] states and the SARS-CoV-1 spike protein in the active (PDB entry:5X5B) [38] and inactive (PDB entry:5X58) [38] states. The protein was solvated in a box of TIP3P waters with 0.15 M NaCl and was simulated using CHARMM36m all-atom additive force field [76]. For details of our simulation and analysis methods see Supporting Information.

DATA AVAILABILITY

Simulation and analysis scripts are available at https://github.com/bslgroup/Spike_Protein.

SUPPORTING INFORMATION

This article contains Supporting Information.

ACKNOWLEDGEMENTS

This research is supported by National Science Foundation grant CHE 1945465. Simulations in this study have been performed primarily using Anton 2, Frontera, and Longhorn. We acknowledge COVID-19 HPC Consortium for providing access to these resources. This research is part of the Frontera computing project at the Texas Advanced Computing Center, made possible by National Science Foundation award OAC-1818253. Anton 2 computer time was provided by the Pittsburgh Supercomputing Center (PSC) through Grant R01GM116961 from the National Institutes of Health. The Anton 2 machine at PSC was generously made available by D.E. Shaw Research. This research is also supported by the Arkansas High Performance Computing Center which is funded through multiple National Science Foundation grants and the Arkansas Economic Development Commission.

REFERENCES

[1] Cherry, J. D. The Chronology of the 2002-2003 SARS Mini Pandemic. *Paediatr. Respir. Rev.* **2004**, *5*, 262–269.

[2] Drosten, C.; Günther, S.; Preiser, W.; van der Werf, S.; Brodt, H.-R.; Becker, S.; Rabenau, H.; Panning, M.; Kolesnikova, L.; Fouchier, R. A. M. et al. Identification of a Novel Coronavirus in Patients with Severe Acute Respiratory Syndrome. *N. Engl. J. Med.* **2003**, *348*, 1967–1976.

[3] Ksiazek, T. G.; Erdman, D.; Goldsmith, C. S.; Zaki, S. R.; Peret, T.; Emery, S.; Tong, S.; Urbani, C.; Comer, J. A.; Lim, W. et al. A Novel Coronavirus Associated with Severe Acute Respiratory Syndrome. *N. Engl. J. Med.* **2003**, *348*, 1953–1966.

[4] Zhou, P.; Yang, X. Lou; Wang, X. G.; Hu, B.; Zhang, L.; Zhang, W.; Si, H. R.; Zhu, Y.; Li, B.; Huang, C. L. et al. A Pneumonia Outbreak Associated with a New Coronavirus of Probable Bat Origin. *Nature* **2020**, *579*, 270–273.

[5] Lu, R.; Zhao, X.; Li, J.; Niu, P.; Yang, B.; Wu, H.; Wang, W.; Song, H.; Huang, B.; Zhu, N. et al. Genomic Characterisation and Epidemiology of 2019 Novel Coronavirus: Implications for Virus Origins and Receptor Binding. *Lancet* **2020**, *395*, 565-574.

[6] Chan, J. F. W.; Yuan, S.; Kok, K. H.; To, K. K. W.; Chu, H.; Yang, J.; Xing, F.; Liu, J.; Yip, C. C. Y.; Poon, R. W. S. et al. A Familial Cluster of Pneumonia Associated with the 2019 Novel Coronavirus Indicating Person-to-Person Transmission: A Study of a Family Cluster. *Lancet* **2020**, *395*, 514–523.

[7] Huang, C.; Wang, Y.; Li, X.; Ren, L.; Zhao, J.; Hu, Y.; Zhang, L.; Fan, G.; Xu, J.; Gu, X. et al. Clinical Features of Patients Infected with 2019 Novel Coronavirus in Wuhan, China. *Lancet* **2020**, *395*, 497–506.

[8] Zhu, N.; Zhang, D.; Wang, W.; Li, X.; Yang, B.; Song, J.; Zhao, X.; Huang, B.; Shi, W.; Lu, R. et al. A Novel Coronavirus from Patients with Pneumonia in China, 2019. *N. Engl. J. Med.* **2020**, *382*, 727–733.

[9] Walls, A. C.; Park, Y. J.; Tortorici, M. A.; Wall, A.; McGuire, A. T.; Veesler, D. Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell* **2020**, *181*, 281 – 292.e6.

[10] Wu, K.; Peng, G.; Wilken, M.; Geraghty, R. J.; Li, F. Mechanisms of Host Receptor Adaptation by Severe Acute Respiratory Syndrome Coronavirus. *J. Biol. Chem.* **2012**, *287*, 8904–8911.

[11] Petersen, E.; Koopmans, M.; Go, U.; Hamer, D. H.; Petrosillo, N.; Castelli, F.; Storgaard, M.; Al Khalili, S.; Simonsen, L. Comparing SARS-CoV-2 with SARS-CoV and Influenza Pandemics. *Lancet Infect. Dis.* **2020**, *20*, e238–e244.

[12] Ferretti, L.; Wymant, C.; Kendall, M.; Zhao, L.; Nurtay, A.; Abeler-Dörner, L.; Parker, M.; Bonsall, D.; Fraser, C. Quantifying SARS-CoV-2 Transmission Suggests Epidemic Control with Digital Contact Tracing. *Science* **2020,** *368*, eabb6936.

[13] Tortorici, M. A.; Veesler, D. Structural Insights into Coronavirus Entry. In *Advances in Virus Research* **2019**, *105*, 93–116. Academic Press Inc.

[14] Walls, A. C.; Tortorici, M. A.; Snijder, J.; Xiong, X.; Bosch, B. J.; Rey, F. A.; Veesler, D. Tectonic Conformational Changes of a Coronavirus Spike Glycoprotein Promote Membrane Fusion. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114*, 11157–11162.

[15] Walls, A. C.; Xiong, X.; Park, Y. J.; Tortorici, M. A.; Snijder, J.; Quispe, J.; Cameroni, E.; Gopal, R.; Dai, M.; Lanzavecchia, A. et al. Unexpected Receptor Functional Mimicry Elucidates Activation of Coronavirus Fusion. *Cell* **2019**, *176*, 1026–1039.e15.

[16] Ou, X.; Liu, Y.; Lei, X.; Li, P.; Mi, D.; Ren, L.; Guo, L.; Guo, R.; Chen, T.; Hu, J. et al. Characterization of Spike Glycoprotein of SARS-CoV-2 on Virus Entry and Its Immune Cross-Reactivity with SARS-CoV. *Nat. Commun.* **2020**, *11*, 1620.

[17] Tai, W.; He, L.; Zhang, X.; Pu, J.; Voronin, D.; Jiang, S.; Zhou, Y.; Du, L. Characterization of the Receptor-Binding Domain (RBD) of 2019 Novel Coronavirus: Implication for Development of RBD Protein as a Viral Attachment Inhibitor and Vaccine. *Cell. Mol. Immunol.* **2020**, *17*, 613–620.

[18] Guo, Y.; Kawaguchi, A.; Takeshita, M.; Sekiya, T.; Hirohama, M.; Yamashita, A.; Siomi, H.; Murano, K. Potent Mouse Monoclonal Antibodies That Block SARS-CoV-2 Infection. *J. Biol. Chem.* **2021**, *296*, 100346.

[19] Baisa, G.; Rancour, D.; Mansfield, K.; Burns, M.; Martin, L.; Cunha, D.; Fischer, J.; Muecksch, F.; Hatziioannou, T.; Bieniasz, P. D. et al. A Recombinant Protein SARS-CoV-2 Candidate Vaccine Elicits High-Titer Neutralizing Antibodies in Macaques. *Res. Sq.* **2021**. https://doi.org/10.21203/rs.3.rs-137857/v1.

[20] Akhter, J.; Quéromès, G.; Pillai, K.; Kepenekian, V.; Badar, S.; Mekkawy, A. H.; Frobert, E.; Valle, S. J.; Morris, D. L. The Combination of Bromelain and Acetylcysteine (Bromac) Synergistically Inactivates Sars-Cov-2. *Viruses* **2021**, *13*, 425.

[21] Lv, Z.; Deng, Y. Q.; Ye, Q.; Cao, L.; Sun, C. Y.; Fan, C.; Huang, W.; Sun, S.; Sun, Y.; Zhu, L. et al. Structural Basis for Neutralization of SARS-CoV-2 and SARS-CoV by a Potent Therapeutic Antibody. *Science* **2020**, *369*, 1505–1509.

[22] Liu, Y. M.; Shahed-Al-Mahmud, M.; Chen, X.; Chen, T. H.; Liao, K. S.; Lo, J. M.; Wu, Y. M.; Ho, M. C.; Wu, C. Y.; Wong, C. H. et al. A Carbohydrate-Binding Protein from the Edible Lablab Beans Effectively Blocks the Infections of Influenza Viruses and SARS-CoV-2. *Cell Rep.* **2020**, *32*, 108016.

[23] Bangaru, S.; Ozorowski, G.; Turner, H. L.; Antanasijevic, A.; Huang, D.; Wang, X.; Torres, J. L.; Diedrich, J. K.; Tian, J. H.; Portnoff, A. D. et al. Structural Analysis of Full-Length SARS-CoV-2 Spike Protein from an Advanced Vaccine Candidate. *Science* **2020**, *370*, 1089–1094.

[24] Polack, F. P.; Thomas, S. J.; Kitchin, N.; Absalon, J.; Gurtman, A.; Lockhart, S.; Perez, J. L.; Pérez Marc, G.; Moreira, E. D.; Zerbini, C. et al. Safety and Efficacy of the BNT162b2 MRNA Covid-19 Vaccine. *N. Engl. J. Med.* **2020**, *383*, 2603–2615.

[25] Baden, L. R.; El Sahly, H. M.; Essink, B.; Kotloff, K.; Frey, S.; Novak, R.; Diemert, D.; Spector, S. A.; Rouphael, N.; Creech, C. B. et al. Efficacy and Safety of the MRNA-1273 SARS-CoV-2 Vaccine. *N. Engl. J. Med.* **2021**, *384*, 403–416.

[26] Zhao, J.; Zhao, S.; Ou, J.; Zhang, J.; Lan, W.; Guan, W.; Wu, X.; Yan, Y.; Zhao, W.; Wu, J. et al. COVID-19: Coronavirus Vaccine Development Updates. *Front. Immunol*. **2020**, *11*, 602256.

[27] Wrapp, D.; Wang, N.; Corbett, K. S.; Goldsmith, J. A.; Hsieh, C. L.; Abiona, O.; Graham, B. S.; McLellan, J. S. Cryo-EM Structure of the 2019-NCoV Spike in the Prefusion Conformation. *Science* **2020**, *367*, 1260–1263.

[28] Li, F.; Li, W.; Farzan, M.; Harrison, S. C. Structural Biology: Structure of SARS Coronavirus Spike Receptor-Binding Domain Complexed with Receptor. *Science* **2005**, *309*, 1864–1868.

[29] Ge, X. Y.; Li, J. L.; Yang, X. Lou; Chmura, A. A.; Zhu, G.; Epstein, J. H.; Mazet, J. K.; Hu, B.; Zhang, W.; Peng, C. et al. Isolation and Characterization of a Bat SARS-like Coronavirus That Uses the ACE2 Receptor. *Nature* **2013**, *503*, 535–538.

[30] Li, W.; Moore, M. J.; Vasllieva, N.; Sui, J.; Wong, S. K.; Berne, M. A.; Somasundaran, M.; Sullivan, J. L.; Luzuriaga, K.; Greeneugh, T. C. et al. Angiotensin-Converting Enzyme 2 Is a Functional Receptor for the SARS Coronavirus. *Nature* **2003**, *426*, 450–454.

[31] Song, W.; Gui, M.; Wang, X.; Xiang, Y. Cryo-EM Structure of the SARS Coronavirus Spike Glycoprotein in Complex with Its Host Cell Receptor ACE2. *PLoS Pathog.* **2018**, *14*, e1007236.

[32] Koyama, T.; Platt, D.; Parida, L. Variant Analysis of SARS-Cov-2 Genomes. *Bull. World Health Organ.* **2020**, *98*, 495–504.

[33] Isabel, S.; Graña-Miraglia, L.; Gutierrez, J. M.; Bundalovic-Torma, C.; Groves, H. E.; Isabel, M. R.; Eshaghi, A. R.; Patel, S. N.; Gubbay, J. B.; Poutanen, T. et al. Evolutionary and Structural Analyses of SARS-CoV-2 D614G Spike Protein Mutation Now Documented Worldwide. *Sci. Rep.* **2020**, *10*, 14031.

[34] Kaushal, N.; Gupta, Y.; Goyal, M.; Khaiboullina, S. F.; Baranwal, M.; Verma, S. C. Mutational Frequencies of SARS-CoV-2 Genome during the Beginning Months of the Outbreak in USA. *Pathogens* **2020**, *9*, 1–16.

[35] van Dorp, L.; Acman, M.; Richard, D.; Shaw, L. P.; Ford, C. E.; Ormond, L.; Owen, C. J.; Pang, J.; Tan, C. C. S.; Boshier, F. A. T. et al. Emergence of Genomic Diversity and Recurrent Mutations in SARS-CoV-2. *Infect. Genet. Evol.* **2020**, *83*, 104351.

[36] Spinello, A.; Saltalamacchia, A.; Magistrato, A. Is the Rigidity of SARS-CoV-2 Spike Receptor-Binding Motif the Hallmark for Its Enhanced Infectivity? Insights from All-Atom Simulations. *J. Phys. Chem. Lett.* **2020**, *11*, 4785–4790.

[37] Cao, W.; Dong, C.; Kim, S.; Hou, D.; Tai, W.; Du, L.; Im, W.; Zhang, X. F. Biomechanical Characterization of SARS-CoV-2 Spike RBD and Human ACE2 Protein-Protein Interaction. *Biophys. J.* **2021**, *120*, 1011-1019.

[38] Yuan, Y.; Cao, D.; Zhang, Y.; Ma, J.; Qi, J.; Wang, Q.; Lu, G.; Wu, Y.; Yan, J.; Shi, Y. et al. Cryo-EM Structures of MERS-CoV and SARS-CoV Spike Glycoproteins Reveal the Dynamic Receptor Binding Domains. *Nat. Commun.* **2017**, *8*, 15092.

[39] Zhou, T.; Tsybovsky, Y.; Gorman, J.; Rapp, M.; Cerutti, G.; Chuang, G. Y.; Katsamba, P. S.; Sampson, J. M.; Schön, A.; Bimela, J. et al. Cryo-EM Structures of SARS-CoV-2 Spike without and with ACE2 Reveal a PH-Dependent Switch to Mediate Endosomal Positioning of Receptor-Binding Domains. *Cell Host Microbe* **2020**, *28*, 867–879.e5.

[40] Cai, Y.; Zhang, J.; Xiao, T.; Peng, H.; Sterling, S. M.; Walsh, R. M.; Rawson, S.; Rits-Volloch, S.; Chen, B. Distinct Conformational States of SARS-CoV-2 Spike Protein. *Science* **2020,** *369*, 1586–1592.

[41] Kirchdoerfer, R. N.; Wang, N.; Pallesen, J.; Wrapp, D.; Turner, H. L.; Cottrell, C. A.; Corbett, K. S.; Graham, B. S.; McLellan, J. S.; Ward, A. B. Stabilized Coronavirus Spikes Are Resistant to Conformational Changes Induced by Receptor Recognition or Proteolysis. *Sci. Rep.* **2018**, *8*, 15701.

[42] Patching, S. G. Surface Plasmon Resonance Spectroscopy for Characterisation of Membrane Protein-Ligand Interactions and Its Potential for Drug Discovery. *Biochim. Biophys. Acta Biomembr.* **2014**, *1838*, 43–55.

[43] Concepcion, J.; Witte, K.; Wartchow, C.; Choo, S.; Yao, D.; Persson, H.; Wei, J.; Li, P.; Heidecker, B.; Ma, W. et al. Label-Free Detection of Biomolecular Interactions Using BioLayer Interferometry for Kinetic Characterization. *Comb. Chem. High Throughput Screen.* **2009**, *12*, 791–800.

[44] Xu, C.; Wang, Y.; Liu, C.; Zhang, C.; Han, W.; Hong, X.; Wang, Y.; Hong, Q.; Wang, S.; Zhao, Q. et al. Conformational Dynamics of SARS-CoV-2 Trimeric Spike Glycoprotein in Complex with Receptor ACE2 Revealed by Cryo-EM. *Sci. Adv.* **2021**, *7*, eabe5575.

[45] Punjani, A.; Fleet, D. J. 3D Variability Analysis: Resolving Continuous Flexibility and Discrete Heterogeneity from Single Particle Cryo-EM. *J. Struct. Biol.* **2021**, *213*, 107702.

[46] Torrens-Fontanals, M.; Stepniewski, T. M.; Aranda-García, D.; Morales-Pastor, A.; Medel-Lacruz, B.; Selent, J. How Do Molecular Dynamics Data Complement Static Structural Data of GPCRs. *Int. J. Mol. Sci*. **2020**, *21*, 5933.

[47] Srivastava, A.; Nagai, T.; Srivastava, A.; Miyashita, O.; Tama, F. Role of Computational Methods in Going beyond X-Ray Crystallography to Explore Protein Structure and Dynamics. *Int. J. Mol. Sci*. **2018**, *19*, 3401.

[48] Lu, M.; Uchil, P. D.; Li, W.; Zheng, D.; Terry, D. S.; Gorman, J.; Shi, W.; Zhang, B.; Zhou, T.; Ding, S. et al. Real-Time Conformational Dynamics of SARS-CoV-2 Spikes on Virus Particles. *Cell Host Microbe* **2020**, *28*, 880–891.e8.

[49] Immadisetty, K.; Hettige, J.; Moradi, M. What Can and Cannot Be Learned from Molecular Dynamics Simulations of Bacterial Proton-Coupled Oligopeptide Transporter GkPOT? *J. Phys. Chem. B* **2017**, *121*, 3644–3656.

[50] Schlitter, J.; Engels, M.; Krüger, P. Targeted Molecular Dynamics: A New Approach for Searching Pathways of Conformational Transitions. *J. Mol. Graph.* **1994**, *12*, 84–89.

[51] Jarzynski, C. Nonequilibrium equality for free energy differences. *Phys. Rev. Lett*. **1997**, *78*, 2690–2693.

[52] Moradi, M.; Sagui, C.; Roland, C. Calculating relative transition rates with driven nonequilibrium simulations. *Chem. Phys. Lett.* **2011**, *518*, 109–113.

[53] Moradi, M.; Tajkhorshid, E. Mechanistic Picture for Conformational Transition of a Membrane Transporter at Atomic Resolution. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 18916–18921.

[54] Amadei, A.; Linssen, A. B. M.; Berendsen, H. J. C. Essential Dynamics of Proteins. *Proteins Struct. Funct. Bioinforma.* **1993**, *17*, 412–425.

[55] Sethi, A.; Eargle, J.; Black, A. A.; Luthey-Schulten, Z. Dynamical Networks in TRNA: Protein Complexes. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 6620–6625.

[56] Moradi, M.; Sagui, C.; Roland, C. Investigating Rare Events with Nonequilibrium Work Measurements. I. Nonequilibrium Transition Path Probabilities. *J. Chem. Phys.* **2014**, *140*, 034114.

[57] Moradi, M.; Sagui, C.; Roland, C. Investigating Rare Events with Nonequilibrium Work Measurements. II. Transition and Reaction Rates. *J. Chem. Phys.* **2014**, *140*, 034115.

[58] Moreira, R. A.; Chwastyk, M.; Baker, J. L.; Guzman, H. V.; Poma, A. B. Quantitative Determination of Mechanical Stability in the Novel Coronavirus Spike Protein. *Nanoscale* **2020**, *12*, 16409–16413.

[59] Gui, M.; Song, W.; Zhou, H.; Xu, J.; Chen, S.; Xiang, Y.; Wang, X. Cryo-electron microscopy structures of the SARS-CoV spike glycoprotein reveal a prerequisite conformational state for receptor binding. *Cell Res*. **2017**, *27*, 119-129.

[60] Shang, J.; Wan, Y.; Luo, C.; Ye, G.; Geng, Q.; Auerbach, A.; Li, F. Cell Entry Mechanisms of SARS-CoV-2. *Proc. Natl. Acad. Sci. U. S. A.* **2020,** *117*, 11727–11734.

[61] Casalino, L.; Gaieb, Z.; Goldsmith, J. A.; Hjorth, C. K.; Dommer, A. C.; Harbison, A. M.; Fogarty, C. A.; Barros, E. P.; Taylor, B. C.; Mclellan, J. S. et al. Beyond Shielding: The Roles of Glycans in the SARS-CoV-2 Spike Protein. *ACS Cent. Sci.* **2020**, *6*, 1722–1734.

[62] Sztain, T.; Ahn, S. H.; Bogetti, A. T.; Casalino, L.; Goldsmith, J. A.; Seitz, E.; McCool, R. S.; Kearns, F. L.; Acosta-Reyes, F.; Maji, S. et al. A Glycan Gate Controls Opening of the SARS-CoV-2 Spike Protein. *Nat. Chem.* **2021,** *13*, 963–968.

[63] Guan, Y.; Zheng, B. J.; He, Y. Q.; Liu, X. L.; Zhuang, Z. X.; Cheung, C. L.; Luo, S. W.; Li, P. H.; Zhang, L. J.; Guan, Y. J. et al. Isolation and Characterization of Viruses Related to the SARS Coronavirus from Animals in Southern China. *Science* **2003**, *302*, 276–278.

[64] Li, W.; Greenough, T. C.; Moore, M. J.; Vasilieva, N.; Somasundaran, M.; Sullivan, J. L.; Farzan, M.; Choe, H. Efficient Replication of Severe Acute Respiratory Syndrome Coronavirus in Mouse Cells Is Limited by Murine Angiotensin-Converting Enzyme 2. *J. Virol.* **2004**, *78*, 11429–11433.

[65] Li, W.; Zhang, C.; Sui, J.; Kuhn, J. H.; Moore, M. J.; Luo, S.; Wong, S. K.; Huang, I. C.; Xu, K.; Vasilieva, N. et al. Receptor and Viral Determinants of SARS-Coronavirus Adaptation to Human ACE2. *EMBO J.* **2005**, *24*, 1634–1643.

[66] Wan, Y.; Shang, J.; Graham, R.; Baric, R. S.; Li, F. Receptor Recognition by the Novel Coronavirus from Wuhan: An Analysis Based on Decade-Long Structural Studies of SARS Coronavirus. *J. Virol.* **2020,** *94*, e00127–20.

[67] He, J. F.; Peng, G. W.; Min, J.; Yu, D. W.; Liang, W. J.; Zhang, S. Y.; Xu, R. H.; Zheng, H. Y.; Wu, X. W.; Xu, J. et al. Molecular Evolution of the SARS Coronavirus, during the Course of the SARS Epidemic in China. *Science* **2004**, *303*, 1666–1669.

[68] Kan, B.; Wang, M.; Jing, H.; Xu, H.; Jiang, X.; Yan, M.; Liang, W.; Zheng, H.; Wan, K.; Liu, Q. et al. Molecular Evolution Analysis and Geographic Investigation of Severe Acute Respiratory Syndrome Coronavirus-Like Virus in Palm Civets at an Animal Market and on Farms. *J. Virol.* **2005**, *79*, 11892–11900.

[69] Sui, J.; Li, W.; Murakami, A.; Tamin, A.; Matthews, L. J.; Wong, S. K.; Moore, M. J.; Tallarico, A. S. C.; Olurinde, M.; Choe, H. et al. Potent Neutralization of Severe Acute Respiratory Syndrome (SARS) Coronavirus by a Human MAb to S1 Protein That Blocks Receptor Association. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 2536–2541.

[70] McCallum, M.; De Marco, A.; Lempp, F. A.; Tortorici, M. A.; Pinto, D.; Walls, A. C.; Beltramello, M.; Chen, A.; Liu, Z.; Zatta, F. et al. N-Terminal Domain Antigenic Mapping Reveals a Site of Vulnerability for SARS-CoV-2. *Cell* **2021**, *184*, 2332-2347.e16.

[71] McCarthy, K. R.; Rennick, L. J.; Nambulli, S.; Robinson-McCarthy, L. R.; Bain, W. G.; Haidar, G.; Paul Duprex, W. Recurrent Deletions in the SARS-CoV-2 Spike Glycoprotein Drive Antibody Escape. *Science* **2021**, *371*, 1139–1142.

[72] Choi, B.; Choudhary, M. C.; Regan, J.; Sparks, J. A.; Padera, R. F.; Qiu, X.; Solomon, I. H.; Kuo, H.-H.; Boucau, J.; Bowman, K. et al. Persistence and Evolution of SARS-CoV-2 in an Immunocompromised Host. *N. Engl. J. Med.* **2020**, *383*, 2291–2293.

[73] Avanzato, V. A.; Matson, M. J.; Seifert, S. N.; Pryce, R.; Williamson, B. N. Journal Pre-Proof Case Study: Prolonged Infectious SARS-CoV-2 Shedding from an Asymptomatic Immunocompromised Cancer Patient. *Cell* **2020**, *183*, 1901–1912.e9.

[74] Weisblum, Y.; Schmidt, F.; Zhang, F.; DaSilva, J.; Poston, D.; Lorenzi, J. C. C.; Muecksch, F.; Rutkowska, M.; Hoffmann, H. H.; Michailidis, E. et al. Escape from Neutralizing Antibodies by SARS-CoV-2 Spike Protein Variants. *eLife* **2020**, *9*, e61312.

[75] Andreano, E.; Piccini, G.; Licastro, D.; Casalino, L.; Johnson, N. V; Paciello, I.; Dal Monego, S.; Pantano, E.; Manganaro, N.; Manenti, A. et al. SARS-CoV-2 Escape in Vitro from a Highly Neutralizing COVID-19 Convalescent Plasma. *bioRxiv* **2020**. https://doi.org/10.1101/2020.12.28.424451.

[76] 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* **2016**, *14*, 71–73.

CONCLUSION

Microsecond-level MD simulations have successfully been used to investigate chemomechanical coupling in two very different proteins – hFGF1 and the CoV-1/CoV-2 spike protein. Using a combination of microsecond-level unbiased MD simulations, SMD simulations, restrained umbrella sampling simulations and a state-of-the-art binding affinity estimation approach, the conformational dynamics of the heparin-hFGF1 complex has been characterized in detail. For the first time, a destabilizing conformational transition was observed in the hFGF1 heparin-binding pocket in the absence of heparin. This provides an explanation for the experimentally observed thermal instability of hFGF1. Unique intramolecular interactions occurring within the heparin-binding pocket in the presence of heparin potentially play an important role in stabilizing hFGF1. More research needs to be carried out to understand the functional relevance of these interactions. The computationally determined binding affinity for the heparin-hFGF1 interaction is in very good agreement with the binding affinity obtained from ITC experiments. The results of the binding affinity study indicate that sampling along a protein-ligand distance without restraints is less effective than sampling with restraints. Specifically, the results show that restraining ligand orientation is key to arriving at a reasonable computational binding affinity estimate.

Extensive equilibrium and nonequilibrium simulations were also used to investigate the conformational dynamics of spike protein activation, which is an important part of the "effective binding" process that leads to interaction with the human ACE2 receptor. The observation of a previously unknown "pseudo-inactive" state of the CoV-1 spike protein, where the NTD interacts with the RBD, suggests that the NTD could play an important role in spike inactivation by blocking the RBD-ACE2 interaction. Recent experimental and clinical studies have shown that CoV-2 variants with mutations in the NTD experience reduced recognition by NTD-specific monoclonal antibodies, indicating that the NTD is under selective pressure from the host immune

response[1-6]. This suggests that mutations in the NTD could potentially be linked to differential transmissibility. This study also shows that the kinetics of the activation or inactivation process are much slower for the CoV-2 spike protein compared to the CoV-1 spike protein, suggesting that CoV-2 may potentially be more transmissible than CoV-1 as a consequence of spending more time bound to the ACE2 receptor. Additional experimental and computational studies are needed to further investigate these hypotheses.

REFERENCES

1) McCallum, M.; De Marco, A.; Lempp, F. A.; Tortorici, M. A.; Pinto, D.; Walls, A. C.; Beltramello, M.; Chen, A.; Liu, Z.; Zatta, F. et al. N-Terminal Domain Antigenic Mapping Reveals a Site of Vulnerability for SARS-CoV-2. *Cell* **2021**, *184*, 2332-2347.e16.

2) McCarthy, K. R.; Rennick, L. J.; Nambulli, S.; Robinson-McCarthy, L. R.; Bain, W. G.; Haidar, G.; Paul Duprex, W. Recurrent Deletions in the SARS-CoV-2 Spike Glycoprotein Drive Antibody Escape. *Science* **2021**, *371*, 1139–1142.

3) Choi, B.; Choudhary, M. C.; Regan, J.; Sparks, J. A.; Padera, R. F.; Qiu, X.; Solomon, I. H.; Kuo, H.-H.; Boucau, J.; Bowman, K. et al. Persistence and Evolution of SARS-CoV-2 in an Immunocompromised Host. *N. Engl. J. Med.* **2020**, *383*, 2291–2293.

4) Avanzato, V. A.; Matson, M. J.; Seifert, S. N.; Pryce, R.; Williamson, B. N. Journal Pre-Proof Case Study: Prolonged Infectious SARS-CoV-2 Shedding from an Asymptomatic Immunocompromised Cancer Patient. *Cell* **2020**, *183*, 1901–1912.e9.

5) Weisblum, Y.; Schmidt, F.; Zhang, F.; DaSilva, J.; Poston, D.; Lorenzi, J. C. C.; Muecksch, F.; Rutkowska, M.; Hoffmann, H. H.; Michailidis, E. et al. Escape from Neutralizing Antibodies by SARS-CoV-2 Spike Protein Variants. *eLife* **2020**, *9*, e61312.

6) Andreano, E.; Piccini, G.; Licastro, D.; Casalino, L.; Johnson, N. V; Paciello, I.; Dal Monego, S.; Pantano, E.; Manganaro, N.; Manenti, A. et al. SARS-CoV-2 Escape in Vitro from a Highly Neutralizing COVID-19 Convalescent Plasma. *bioRxiv* **2020**. https://doi.org/10.1101/2020.12.28.424451.

APPENDIX

Supporting Information - Mechanistic Picture for Monomeric Human Fibroblast Growth Factor 1 Stabilization by Heparin Binding

Vivek Govind Kumar[1], Shilpi Agrawal[1], Thallapuranam Krishnaswamy Suresh Kumar[1], and Mahmoud Moradi[1]

[1]Department of Chemistry and Biochemistry, University of Arkansas, Fayetteville, AR 72701

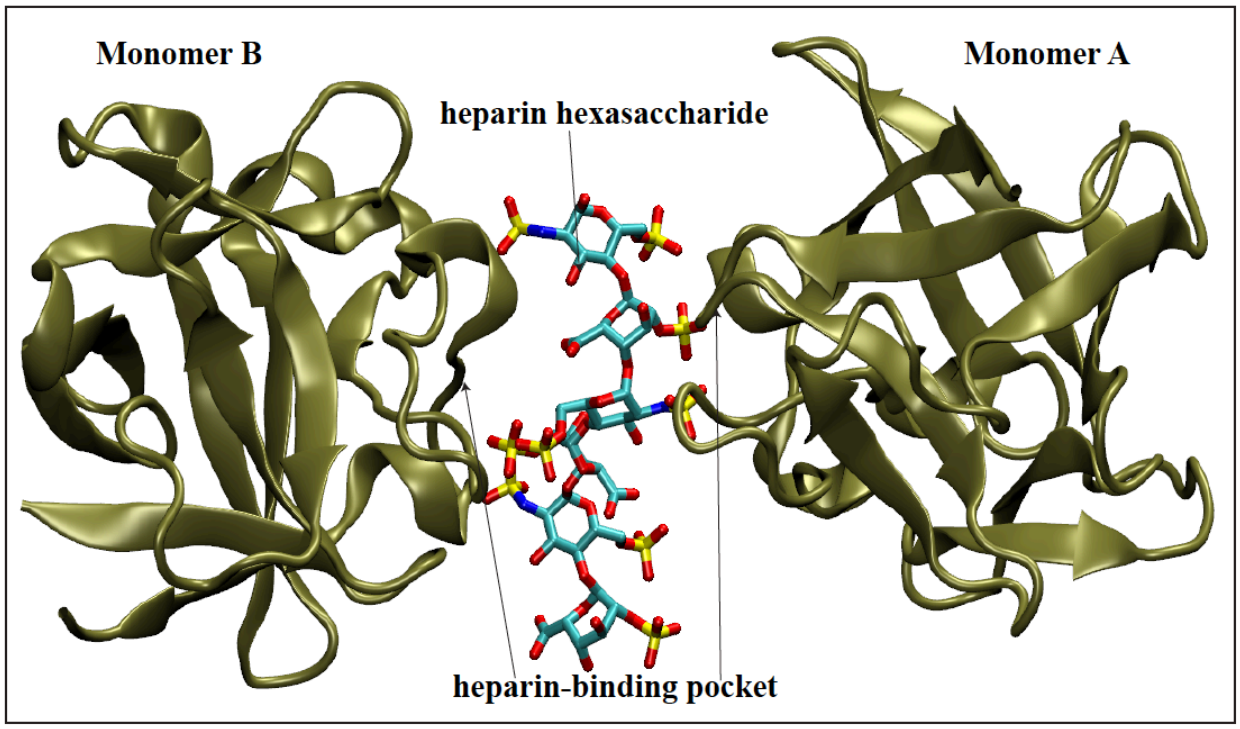


Figure S1. Cartoon representation of the dimeric hFGF1 X-ray crystal structure with heparin hexasaccharide (PDB entry 2AXM) (related to Figures 1-4 and Table 1).

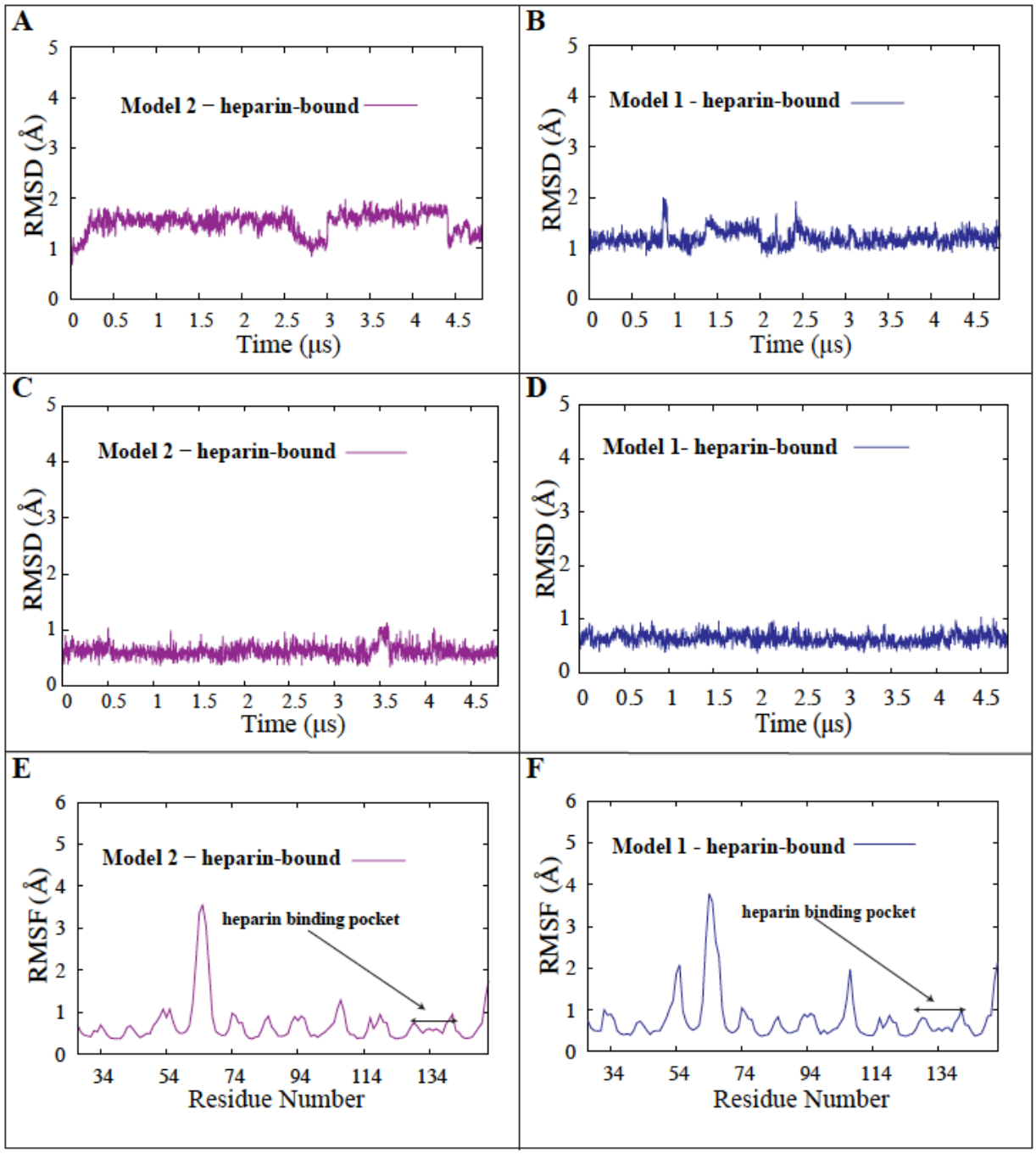


Figure S2. Stability of heparin-bound hFGF1 assessed through RMSD and RMSF calculations (related to Figure 1). (A) Internal RMSD time series for heparin-bound hFGF1 (Model2). (B) Internal RMSD time series for heparin-bound hFGF1 (Model1). (C) Internal RMSD time series for the heparin-binding pocket of heparin-bound hFGF1 (Model2). (D) Internal RMSD time series for the heparin-binding pocket of heparin-bound hFGF1 (Model1). (E) RMSF estimation for heparin-bound hFGF1 (Model2). (F) RMSF estimation for heparin-bound hFGF1 (Model1)

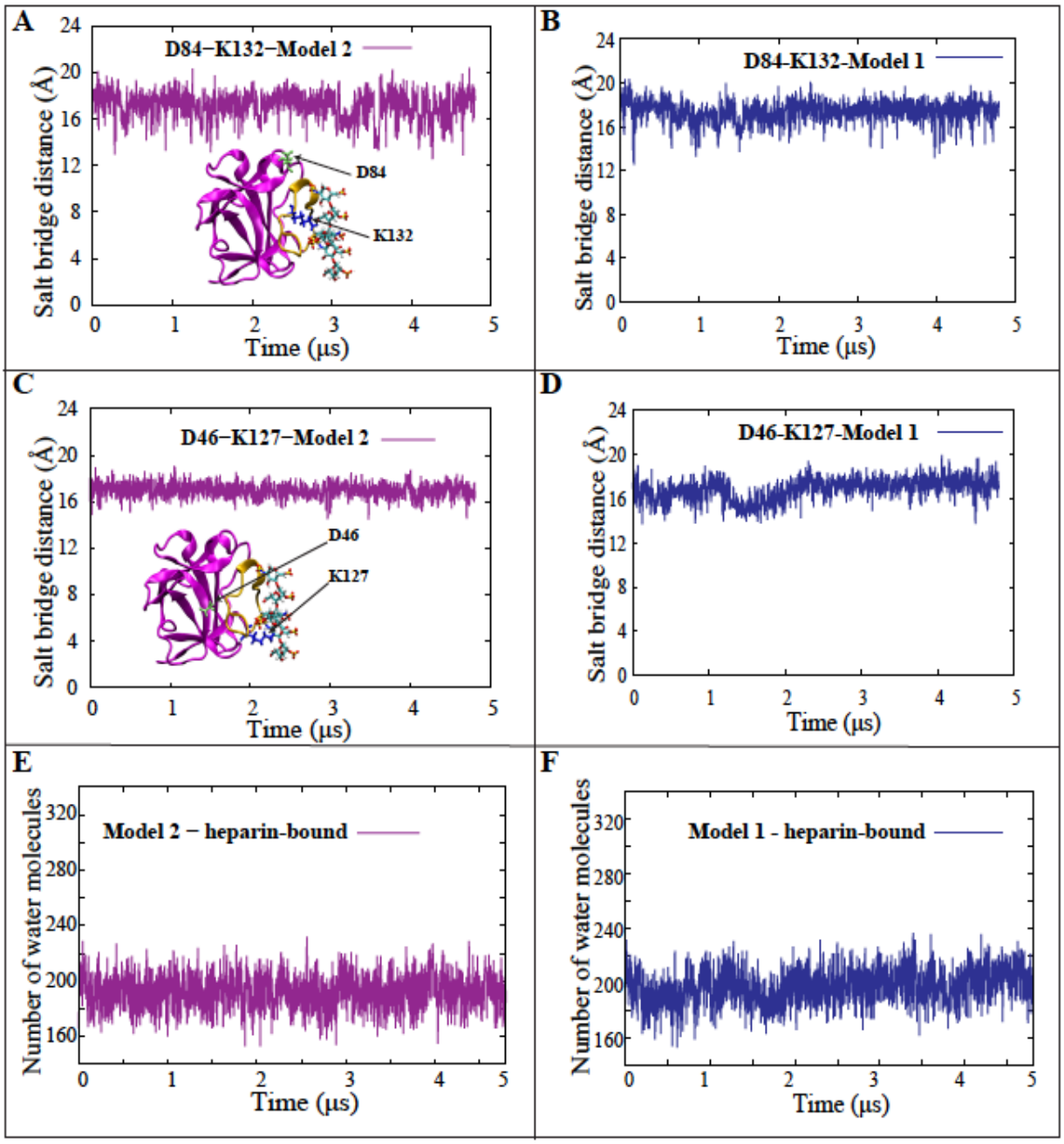


Figure S3. Salt bridges associated with the conformational change in the apo model do not form in the heparin-bound models (related to Figure 3). (A) Time series and cartoon representation of the D84-K132 donor-acceptor salt bridge distance for heparin-bound Model 2. (B) Time series of the D84-K132 donor-acceptor salt bridge distance for heparin-bound Model 1. (C) Time series and cartoon representation of the D46-K127 donor-acceptor salt bridge distance for heparin-bound Model 2. (D) Time series of the D46-K127 donor-acceptor salt bridge distance for heparin-bound Model 1. (E) Time series of water molecule count within 3 Å of the heparin-binding pocket for heparin-bound Model 2. (F) Time series of water molecule count within 3 Å of the heparin-binding pocket for heparin-bound Model 2.

Figure S4. Table of intramolecular interactions unique to the heparin-binding pocket of heparin-bound hFGF1 (Model 2) (Related to Figure 3E). Hydrogen-bonding occupancies are similar in heparin-bound Model 1.

Donor	Acceptor	Occupancy (%)
K126	S130	94
Q141	R136	82
G129	K126	83
R136	R133	74
T137	G134	69
H138	Q141	64

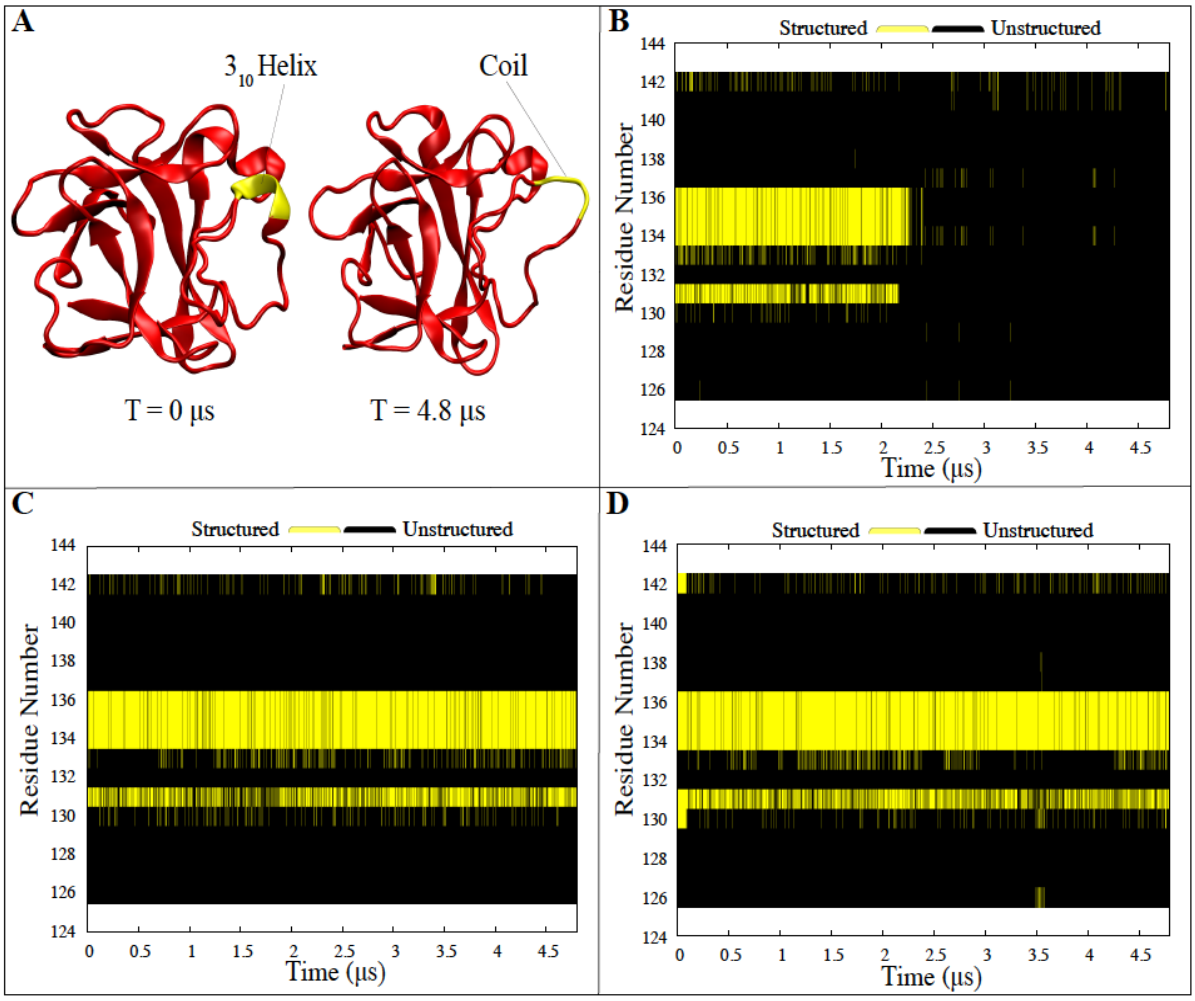


Figure S5. The conformational change causes secondary structural changes in the apo model (related to Figure 1 and Figure 3). (A) Cartoon representation of the secondary structural change that occurs in the heparin-binding pocket of the apo model due to the conformational change. (B) Secondary structure of the heparin-binding pocket of apo hFGF1 as a function of simulation time. Parts of the heparin-binding pocket become unstructured after 2 microseconds. (C) Secondary structure of the heparin-binding pocket of heparin-bound hFGF1 (Model 1) as a function of simulation time. (D) Secondary structure of the heparin-binding pocket of heparin-bound hFGF1 (Model 2) as a function of simulation time. The heparin-binding pocket remains structured in both heparin-bound trajectories.

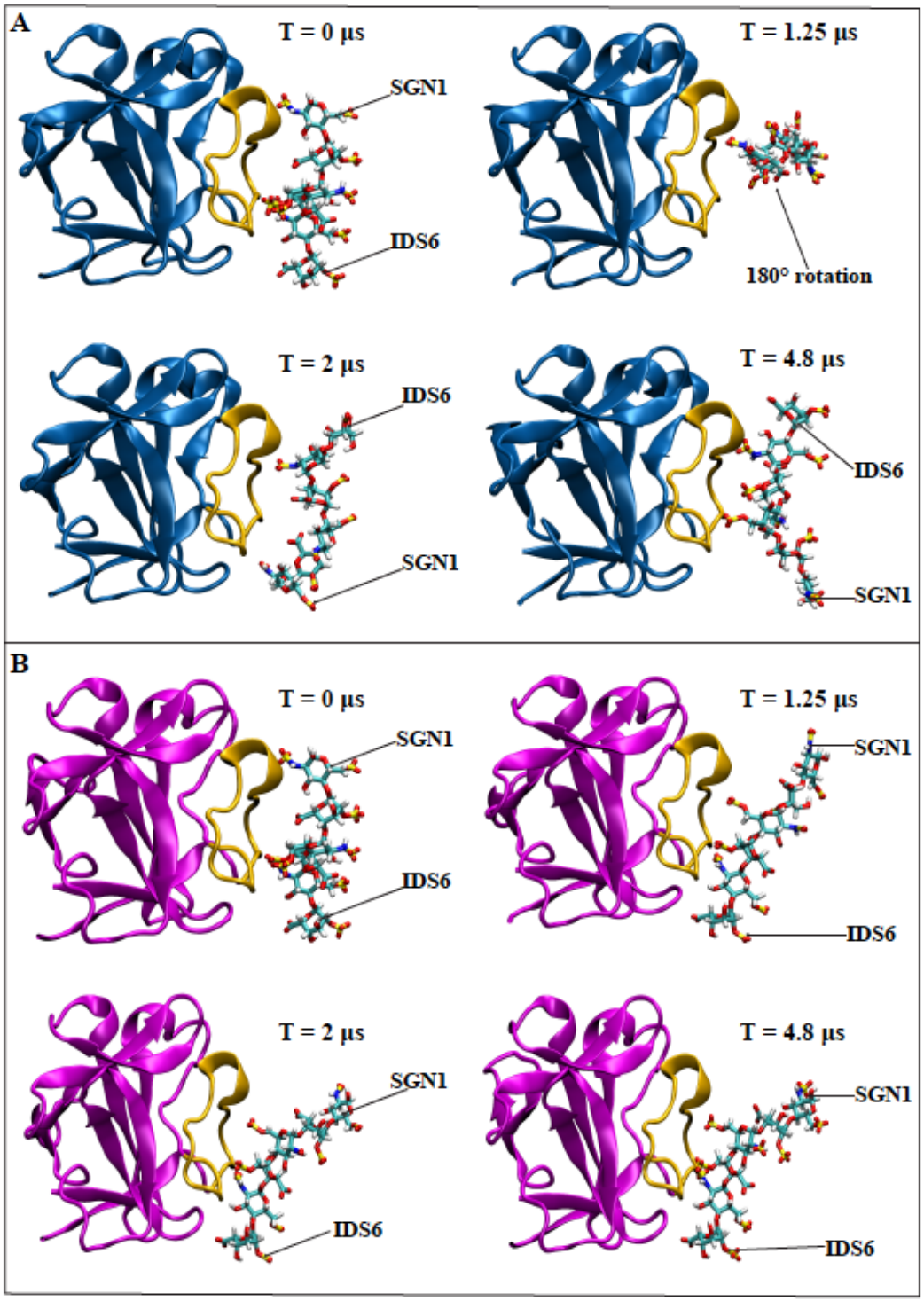


Figure S6. Behavior of heparin hexasaccharide in the heparin-bound trajectories (related to Figure 4 and Table 1). (A) The heparin hexasaccharide in Model 1 (blue) fluctuates considerably before undergoing a 180° rotation. It settles into a more stable conformation after 2 µs. (B) The heparin hexasaccharide in Model 2 does not undergo any major positional changes. Due to the differences in behavior of heparin in each model, slightly different intermolecular interactions occur in terms of both occupancy as well as the residues involved.

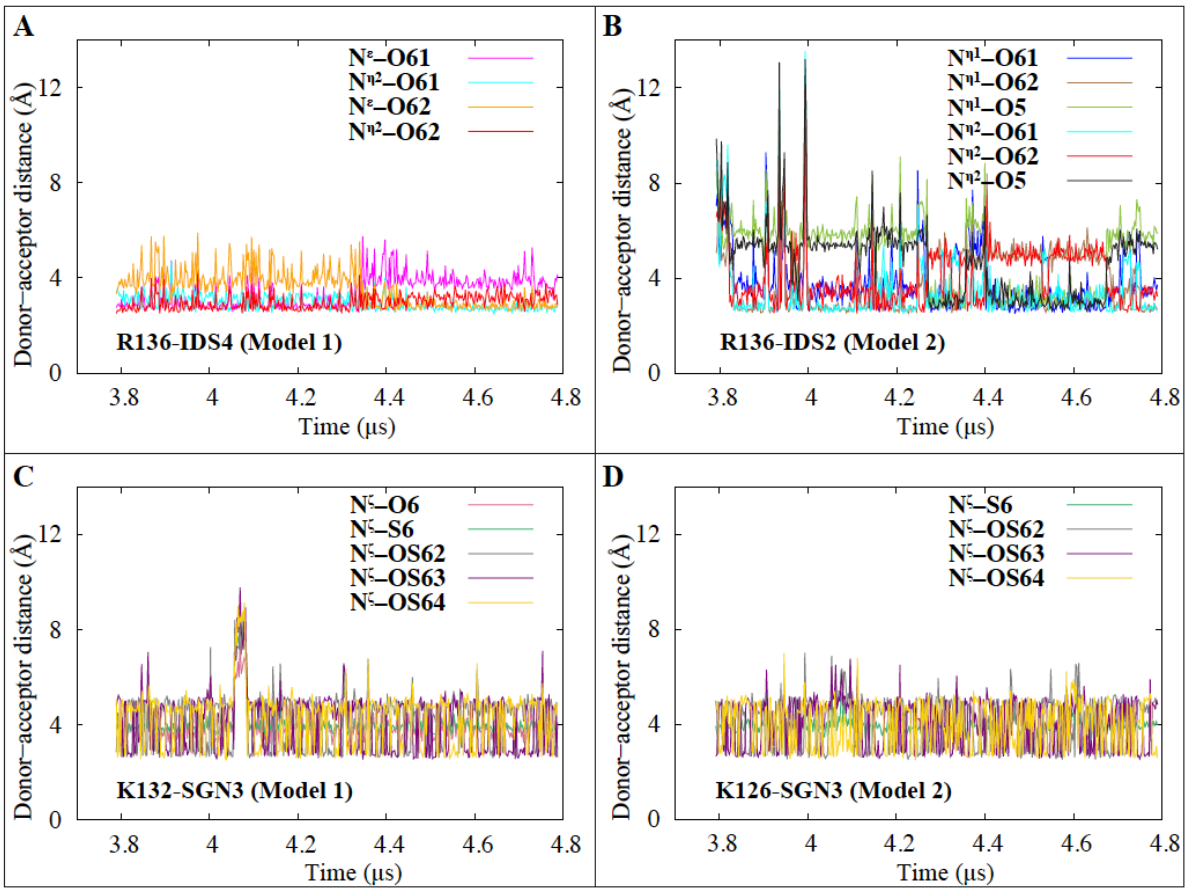


Figure S7. Time series of hFGF1-heparin intermolecular interactions (related to Figure 4 and Table 1). (A) Time series of hydrogen-bonding interactions between R136 and IDS4 (Model 1). (B) Time series of hydrogen-bonding interactions between R136 and IDS2 (Model 2). (C) Time series of hydrogen-bonding interactions between K132 and SGN3 (Model 1). (D) Time series of hydrogen-bonding interactions between K126 and SGN3 (Model 2)

Supporting Information - Prefusion Spike Protein Conformational Changes Are Slower in SARS-CoV-2 Relative to SARS-CoV-1

Vivek Govind Kumar[1], Dylan S Ogden,[1] Ugochi H Isu,[1] Adithya Polasa,[1] James Losey,[1] and Mahmoud Moradi[1]

[1]Department of Chemistry and Biochemistry, University of Arkansas, Fayetteville, AR 72701

SIMULATION AND ANALYSIS DETAILS

MD Simulation Details

We have used all-atom equilibrium and nonequilibrium MD simulations to characterize the conformational dynamics of the spike protein from SARS-CoV-2 and SARS-CoV-1. Our simulations were based on cryo-EM structures of the SARS-CoV-2 spike protein in the active (PDB entry:6VYB)[1] and inactive (PDB entry:6VXX)[1] states and the SARS-CoV-1 spike protein in the active (PDB entry:5X5B)[2] and inactive (PDB entry:5X58)[2] states. Missing residues for all 4 models were generated using Modeller.[3] 10,000 Monte Carlo iterations were used to generate the initial models for the equilibrium simulations.[3] CHARMM-GUI[4,5] was then used to build the simulation systems. Engineered residues were mutated back to the wildtype and disulfide bonds were added to each model based on the information provided in the respective PDB files.[1,2] The protein was solvated in a box of TIP3P waters, and 0.15 M NaCl (in addition to the counterions used to neutralize the protein) using CHARMM-GUI.[4,5] All simulations were performed using the NAMD 2.13[6] simulation package with the CHARMM36m all-atom additive force field[7]. The input files for energy minimization and production were generated using CHARMM-GUI.[4,5] Initially, we energy-minimized each

system for 10,000 steps using the conjugate gradient algorithm.[8] Then, we relaxed the systems using restrained MD simulations in a stepwise manner using the standard CHARMM-GUI protocol[4,5] ("relaxation step"). In the next step, backbone and sidechain restraints were used for 10 ns with a force constant of 1 kcal/mol.Å^2 and 0.5 kcal/mol.Å^2, respectively ("restraining step"). The systems were then equilibrated with no bias for another 10 ns ("equilibration step"). The initial relaxation was performed in an NVT ensemble while the rest of the simulations were performed in an NPT ensemble. Simulations were carried out using a 2-fs time step at 310 K using a Langevin integrator with a damping coefficient of γ = 0.5 ps^{-}1. The pressure was maintained at 1 atm using the Nose-Hoover Langevin piston method.[8,9] The smoothed cutoff distance for non-bonded interactions was set at 10 to 12 Å and long-range electrostatic interactions were computed with the particle mesh Ewald (PME) method.[10]

These initial simulations were executed on TACC Longhorn. The production run for each model was then extended to 5 μs on Anton2,[11] with a timestep of 2.5 fs. Conformations were collected every 240 picoseconds. Initial processing of the Anton2 simulation trajectories was carried out on Kollman. [11] Two additional 5 μs simulations were performed for both the CoV-2 and CoV-1 active models on Anton2 (referred to as Set 2 and Set 3 in the manuscript). As stated previously, cryo-EM structures (PDB entries:6VYB,5X5B)[1,2] were used as starting conformations for each model. In order to generate initial conformations for Set 2, the original production run ("equilibration step" described previously) for each model was extended by 0.5 ns on TACC Longhorn. The production runs were then extended again by 0.5 ns to generate the initial conformations for Set 3. 40 μs of simulation data was generated in aggregate – 15 μs each for the active Cov-1/Cov-2 spike proteins and 5 μs each for the inactive spike proteins.

RBM-S2 Distance and Angle

To quantify the RBM-S2 distance, we defined centers of mass based on residues that form a beta-sheet in the RBM region of each RBD (CoV-1: RBM residues 439 to 441, 479 to 481; CoV-2: RBM residues 452 to 454, 492 to 494) and residues that encompass the S2 trimer (CoV-1: S2 residues 672 to 1104; CoV-2: S2 residues 690 to 1147). We then measured the vector distance between the two centers of mass and used the vector magnitude to quantify the overall distance.

For the RBM-S2 angle, we chose residues at the top and bottom of the straightest region of the S2 Trimer (alpha-helical regions in CoV-1: residues 970 and 1016; CoV-2: residues 914 and 987). Similarly, we also chose residues from the beta-sheet region of the RBM and one at the bottom of the RBD (CoV-1: residues 348 and 478; CoV-2: residues 391 and 493). We then defined a vector direction using the vector subtraction of the two chosen residues in the S2 region and the residues of the RBD region, which were defined as v_1 and v_2. The vector angle between the RBD and S2 was then calculated with the following equation: $\arccos(\frac{v_1 \cdot v_2}{|v_1||v_2|})$. The computed angle was subtracted from 180°. An angle above $\approx 60^\circ$ would indicate an RBD in the inactive conformation with respect to S2, and 0-40° would indicate an RBD in the active conformation.

NTD-RBD Distance

To characterize conformational changes in the active and inactive states of both CoV-1 and CoV-2 spike proteins, we calculated the minimum distance between every residue of the receptor-binding domain (RBD) and the N-terminal domain (NTD). We measured the

distance between each residue pair in these regions (maximum distance cutoff was 20 Å) as a function of time. The domains were defined as follows: CoV-2 RBD (residues 330 to 515); CoV-2 NTD (residues 60 to 270); CoV-1 RBD (residues 330 to 550); CoV-1 NTD (residues 35 to 255).

RBM Hydration Analysis

The amount of solvent around the receptor-binding motif (RBM) was quantified using a VMD[12] script. We calculated the number of water molecules within 5 Å of the RBM for every frame of the last 500 ns of each trajectory and also plotted probability density maps for each water count.

Principal Component Analysis (PCA)

PCA[13] performed with ProDy[14] was used to quantify the persistent conformational changes and relative motions of the active and inactive states. Only the position of the C-α atoms of the spike protein was considered when building the covariance matrix of atomic positions, in order to focus on the large conformational changes and ignore side chain fluctuations. Each trajectory was aligned with the positions from the cryo-EM structure before analysis to remove translational motion of the protein from the variance calculations.

The CoV-1/CoV-2 active state (Set 1) and CoV-1/CoV-2 inactive state trajectories were stripped down to trajectories of the individual protomers from each simulation. The individual protomers were then analyzed together to compare and quantify the relative motions of the active and inactive states. Through eigenvalue decomposition, the top twenty principal components (PCs) were calculated for each protomer. The top two PCs for each protomer have been plotted to identify the major motions of the protein.

Dynamic Network Analysis (DNA)

DNA[15] of the correlated motions of the protein provided further quantitative information on the concerted motions of the C-α atoms of the protein. MD-TASK, [16] a software suite of MD analysis tools, was used to calculate the correlation coefficient for the motion of each C-α atom relative to the other C-α atoms. A correlation matrix M was generated for each of the three protomers in all the simulated trajectories. Additionally, a correlation matrix for the entire trimer was calculated for each simulation to explore correlations between structures of different protomers. A step size of four frames was used during the correlation calculations to reduce the processing times, given the large number of residues.

To quantify the differences in correlation between a protomer and some reference, a difference matrix, Δ was calculated,

$$\Delta = |M_i - M_{Ref.}|, \quad (1)$$

where M_i is the correlation matrix of interest, and M_{Ref} is the correlation matrix of a reference conformation. In this work, the difference between a protomer in an active conformationand an inactive conformation was of interest. For this reason, the protomers in the active simulations were compared with Protomer C in the inactive simulation, which displayed relatively little motion.

Interaction Analysis

To identify interactions that contribute to the stability of the Cov-2 spike protein or play key roles in the CoV-1 active conformational transition, we performed salt-bridge and hydrogen-bond analysis for all SARS-CoV-2 and SARS-CoV-1 systems. Salt bridges were identified using the VMD Timeline plugin[12] at a cutoff distance of 4.0 Å. The salt-bridge cutoff

distance is defined as the distance between the oxygen atom of the participating acidic residue and the nitrogen atom of the basic residue. The VMD HBond plugin[12] was used for hydrogen bond analysis. The donor-acceptor distance and angle cutoffs used were 3.5 Å and 30 degrees respectively. We report salt-bridge and hydrogen-bond interactions that illustrate the differential behavior of the SARS-CoV-2 and CoV-1 spike proteins.

Steered Molecular Dynamics (SMD) analysis

To induce activation/inactivation of a protomer initially in the inactive/active conformation, we defined collective variables based on the Cα RMSD of each protomer in the CoV-1 and CoV-2 systems. Reference coordinates were taken from the corresponding active/inactive structure for both CoV-1 and CoV-2 protomers. The atoms chosen were based on the total number of modeled residues in the CoV-2 structures. Structural analysis of CoV-1 and CoV-2 was employed to ensure that equivalent Cα atoms were steered in all simulation sets. 1037 atoms were steered for any given protomer and the following residue range was used: 27 to 239, 244 to 315, 322 to 662, 673 to 809, and 831 to 1104. These atoms span the entire protomer, starting from the NTD and ending approximately at the C-terminus of the S2 region. A force constant of 250 $kcal/mol/Å^2$ was used for SMD simulations involving a single protomer and a force constant of 750 $kcal/mol/Å^2$ was used for SMD simulations involving all three protomers. The systems used for each simulation were taken from the outcome of the "equilibration step" as explained above. Utilizing the multi-copy capabilities of NAMD, we performed 10 sets of 100 ns RMSD steering for each system – 8 μs of simulation time in aggregate.

For all SMD time series analyses, each data point was averaged for the 10 sets and standard deviation was calculated. Each analysis was plotted with 100 points and error bars

were derived from the standard deviation. The RBM-S2 distance and angle calculations were performed as described previously. Using the Jarzynski relation[17] we calculate the Jarzynski average at time t during the activation or inactivation process as:

$$-k_B T \ln \sum_{i=1}^{N} \exp\left(\frac{-W_i(t)}{k_B T}\right)/N$$

where k_B and T are the Boltzmann constant and the temperature, respectively and $W_i(t)$ is the work accumulated from the beginning of the SMD simulation i up to time t. The above average would converge to the free energy for large number of trajectories ($N \rightarrow \infty$). For $N = 10$, the above average simply provides a semi-quantitative measure for relative energetic comparisons.[18–21]

Supporting Discussion : Principal Component Analysis and Dynamic Network Analysis

We performed principal component analysis (PCA) to validate our claim that the active form of the CoV-2 spike protein is more stable than the active CoV-1 spike protein and to provide insight into the mechanistic aspects of the spike protein activation-inactivation process. When the individual protomer trajectories (see Methods section) from the CoV-1/CoV-2 active (Set 1) and inactive simulations are projected onto the space of their first two principal components (PC1 and PC2), it clearly demonstrates that the CoV-1 active protomer A samples a much larger region in the PC1 space than CoV-2 active protomer A (Figures S3A, S3C). This is further evidence of the relative stability of the active CoV-2 spike protein in comparison to the active CoV-1 spike protein.

A visual representation of PC1 for all protomers from the CoV-1 spike protein simulations shows that the RBD undergoes the most pronounced motions directed inward towards the NTD (Figure S3B). On the other hand, a visual representation of PC1 for the

CoV-2 spike protein shows that the RBD and NTD tend to move away from each other slightly and that the fluctuations are significantly smaller than in the CoV-1 spike protein (Figure S3D). The most pronounced collective motion in each system (PC1) describes the distinct motions associated with the RBD, that play key roles in the inactivation of the active CoV-1 spike protein and maintenance of the active conformation of the CoV-2 spike protein (Figure 1). This highlights the differential dynamic behavior of the active CoV-1 spike protein.

PC2 describes the relative motions of the NTD and RBD, showing that the NTD motion is more pronounced in CoV-1 (Figure S4). The motions associated with PC2 are roughly the opposite of those associated with PC1 in terms of direction. PC2 also shows that the CoV-1 spike protein has more regions outside the NTD and RBD that show high variance (Figure S4). Similar trends are observed in Sets 2 and 3 of the active state simulations (Figure S5). While different protomers are involved, the active CoV-1 spike protein still undergoes more pronounced motions in both PC1 and PC2 compared to the active CoV-2 spike protein (Figure S5). These observations are in agreement with our claim that the active CoV-2 spike protein is relatively stable and that the active CoV-1 spike protein transitions spontaneously to a pseudo-inactive conformation.

The inferences drawn from PCA are also supported by dynamic network analysis (DNA). Differential behavior of the active CoV-1 and CoV-2 spike proteins manifests in the correlation of motions between the various domains in individual protomers. In Figure S6A, correlation heat maps of active CoV-1 protomer A (Set 1) and inactive CoV-1 protomer C are presented, along with the difference between the active state and the reference structure (inactive protomer C). The heat map for active Cov-1 protomer A shows regions of high correlation and anticorrelation between several domains of the protomer. The NTD correlates strongly with itself while anticorrelating with the RBD and parts of the S2 region. The

reference protomer, inactive CoV-1 protomer C, shows a general reduction in correlation across all regions (Figure S6A). The NTD does correlate with itself, but not as strongly as in the active CoV-1 protomer A.

Similarly, the NTD-RBD anticorrelations were reduced. The Δ matrix of differences between active CoV-1 protomer A and inactive protomer C identified the regions where the correlations were most different. Correlations between S1-C and the NTD/RBD changed significantly, as did correlations between the RBD and S2 region (Figure S6A).

The correlations and anti-correlations observed for active CoV-2 protomer A (Set 1) were not as strong as those observed for active CoV-1 protomer A (Figure S6B). Similar to CoV-1, anti-correlation occurs between the NTD and RBD but is not as pronounced. Very low correlation was observed between the NTD and S1-C/S2 regions, also differentiating CoV-2 from CoV-1. The active CoV-2 protomer A is closer to the stable inactive CoV-2 protomer C, as shown in the Δ matrix (Figure S6B). DNA correlation heat maps for all protomers in Set 1 of the CoV-1/CoV-2 active state simulations are shown in Figures S7 and S8 respectively. Similar trends were observed in Set 2 and Set 3 of the CoV-1 and CoV-2 active state simulations (Figure S9-S10). These observations thus provide further evidence of the relative stability of the active CoV-2 spike protein.

The concerted movements of each protomer relative to the rest of the trimer also highlight the differences between the active CoV-1 and CoV-2 spike proteins. Heat maps showing correlations between NTD regions of different protomers are presented in Figure S11A. Stronger correlations and anticorrelations occurred in Sets 2 and 3 of the active CoV-1 simulations (Figure S11A). Set 2 showed moderately strong anticorrelations between NTDs A-C and NTDs B-C. Stronger anti-correlations between NTDs A-B and NTDs B-C occurred in Set 3, with moderate correlations between NTDs A-C. The active CoV-2 simulations showed similar correlations across all three simulation sets, with slightly

increased values in Set 3 (Figure S11A). These observations are consistent with a more stable conformation for the active CoV-2 spike protein.

Figure S11B shows a similar trend with correlations between the NTD and RBD regions of different protomers. Sets 2 and 3 of the active CoV-1 spike protein trajectories showed stronger correlations between the NTD and RBD regions than the corresponding CoV-2 trajectories (Figure S11B). In particular, RBD C of Sets 2 and 3 had strong correlations or anticorrelations with the NTDs of all protomers (Figure S11B). The CoV-2 simulations displayed lower correlations for all the NTD-RBD combinations, with similar results for both active state and inactive state trajectories (Figure S11B). This recapitulates our other observations of greater conformational stability of the active CoV-2 spike protein relative to the active CoV-1 spike protein (Figures 1, S3, S6).

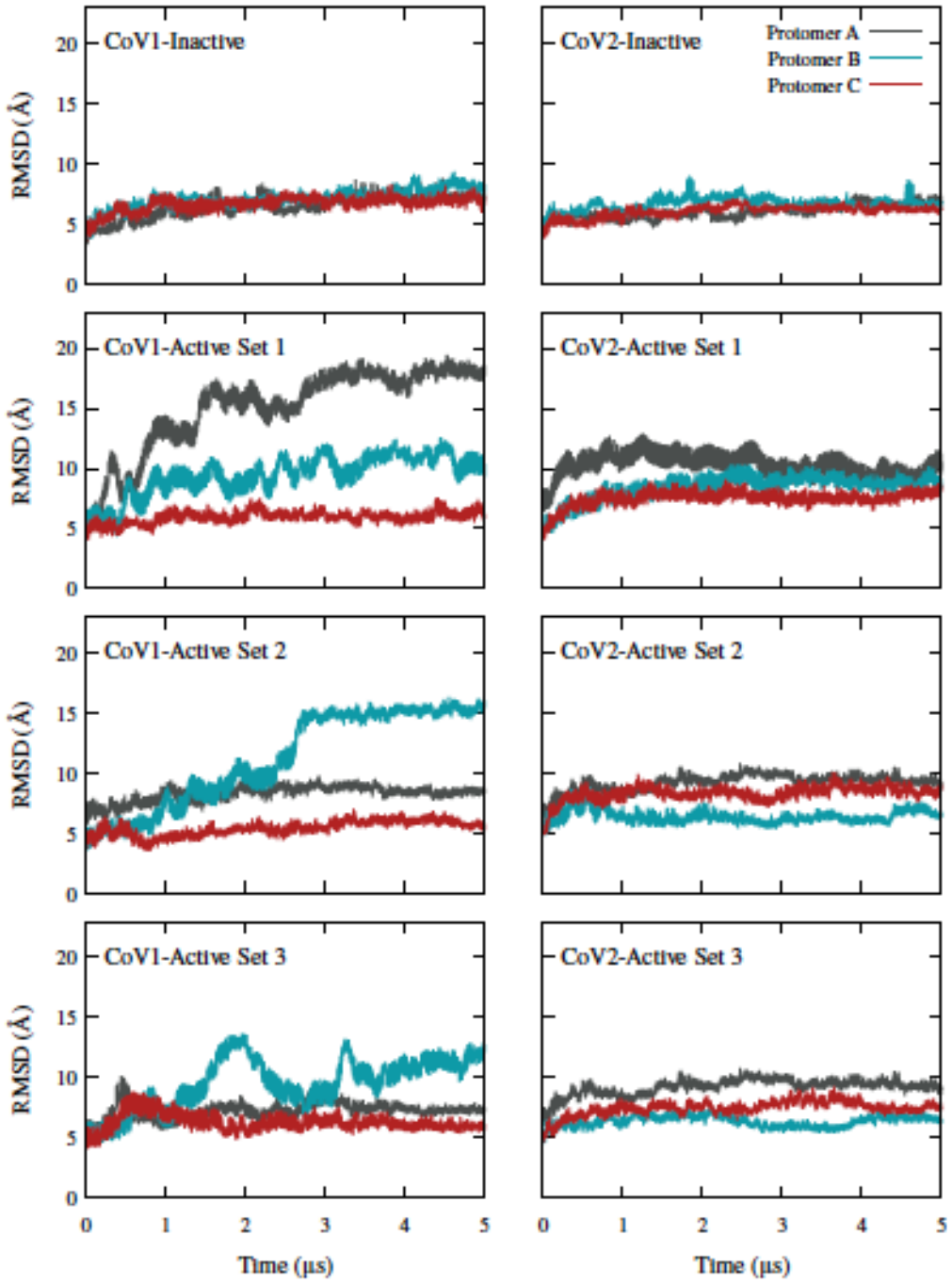


Figure S1: **C-α RMSD for individual protomers.** The C-α RMSD calculated for each protomer relative to the initial cryo-EM structure over the 5 μs simulation is plotted for the inactive spike simulationsand three sets of active spike simulations. Protomer A is colored dark grey, protomer B is colored light blue, and protomer C is colored dark red. The active CoV-2 spike protein is more stable overall than the active CoV-1 spike protein.

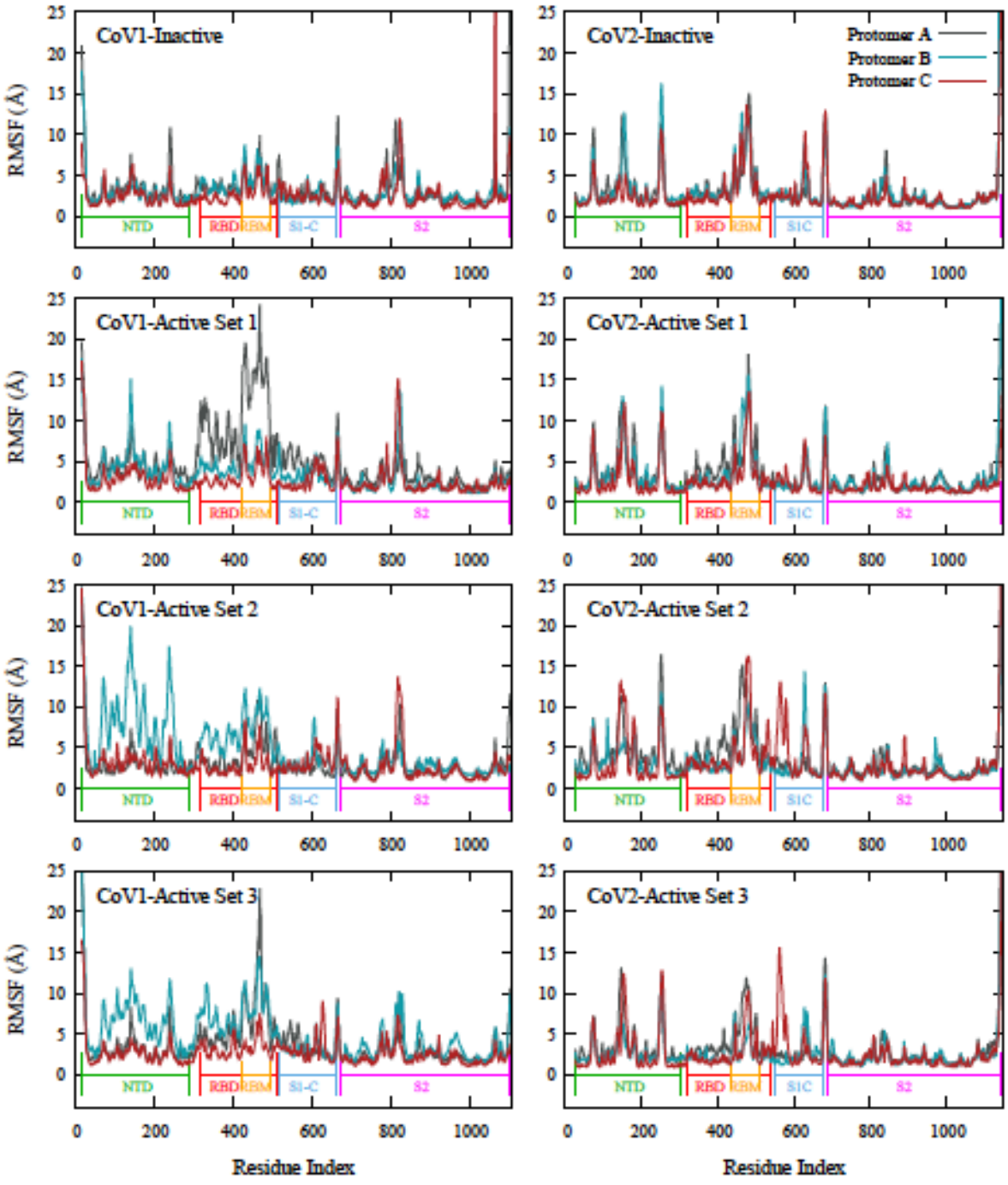


Figure S2: **C-α RMSF for individual protomers.** The C-α RMSF for each protomer relative to the initial cryo-EM structure position was calculated for the inactive spike simulations and three sets of activespike simulations. Protomer A is colored dark grey, protomer B is colored light blue, and protomer C is colored dark red. The NTD and RBD of the active CoV-1 spike are more flexible than the corresponding regions of the active CoV-2 spike.

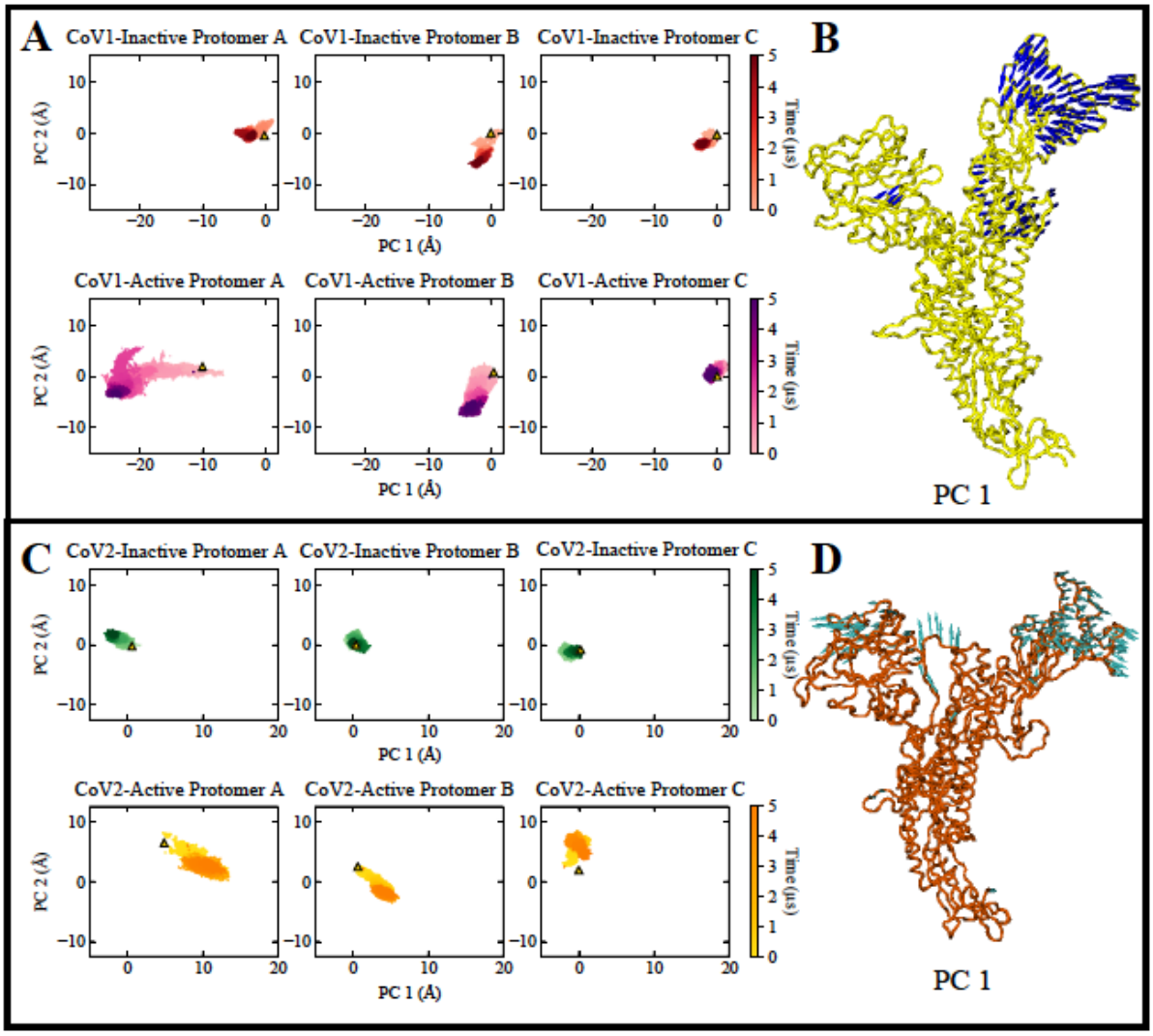


Figure S3: **Principal component analysis demonstrates that the active CoV-2 spike protein is more stable than the active Cov-1 spike protein. (A)** Scatter plot of PC1 and PC2 for each protomer in the active and inactive CoV-1 simulations. Protomers from inactive state simulations are colored red while protomers from active state simulations are colored magenta. Lighter/darker colors represent earlier/later stages in the simulation. **(B)** Visual representation of PC1 with the blue arrows at each C-α atom indicating direction and magnitude of variance. The RBD of the CoV-1 spike protein shows pronounced motions in the direction of the NTD. **(C)** Scatter plot of PC1 and PC2 for each protomer in the inactive and active CoV-2 simulations. Protomers from inactive state simulations are colored green while protomers from active state simulations are colored yellow. The active CoV-2 spike protein is relatively stable and samples significantly fewer conformations in the PC1 space in comparison to the active Cov-1 spike protein. **(D)** Visual representation of PC1 with the cyan arrows at each C-α atom indicating direction and magnitude of variance. The NTD and RBD of the CoV-2 spike protein show slight movement away from each other.

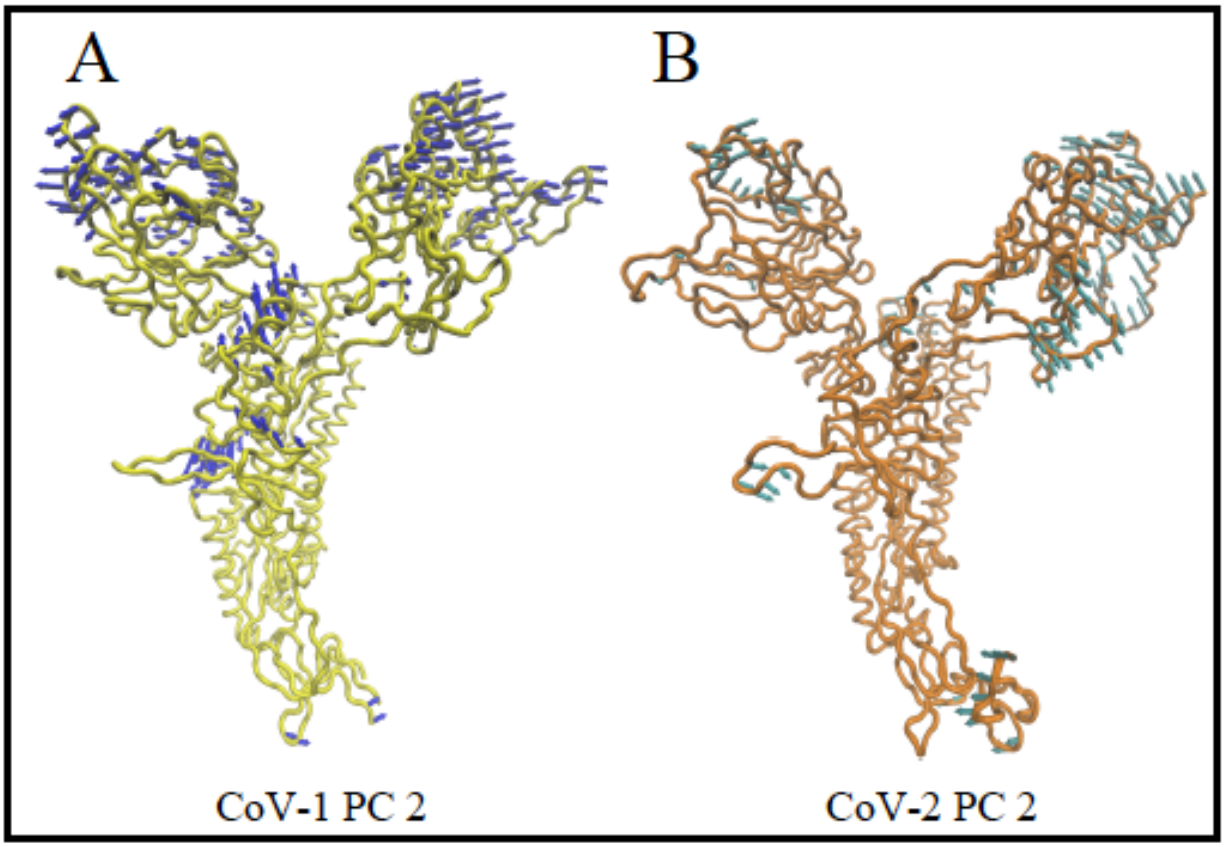


Figure S4: **Visual representation of PC2 for all protomers in the inactive and active (Set 1) spike simulations for CoV-1 and CoV-2. (A)** Visual representation of PC2 for all CoV-1 protomers with the blue arrows at each C-α atom indicating direction and magnitude of variance. **(B)** Visual representationof PC2 for all CoV-2 protomers with the blue arrows at each C-α atom indicating direction and magnitude of variance. The NTD motions contribute more to the conformations sampled in the PC2 space than the PC1 space. These NTD motions are more pronounced in the CoV-1 spike, which also has more regions outsidethe NTD/RBD that show high variance.

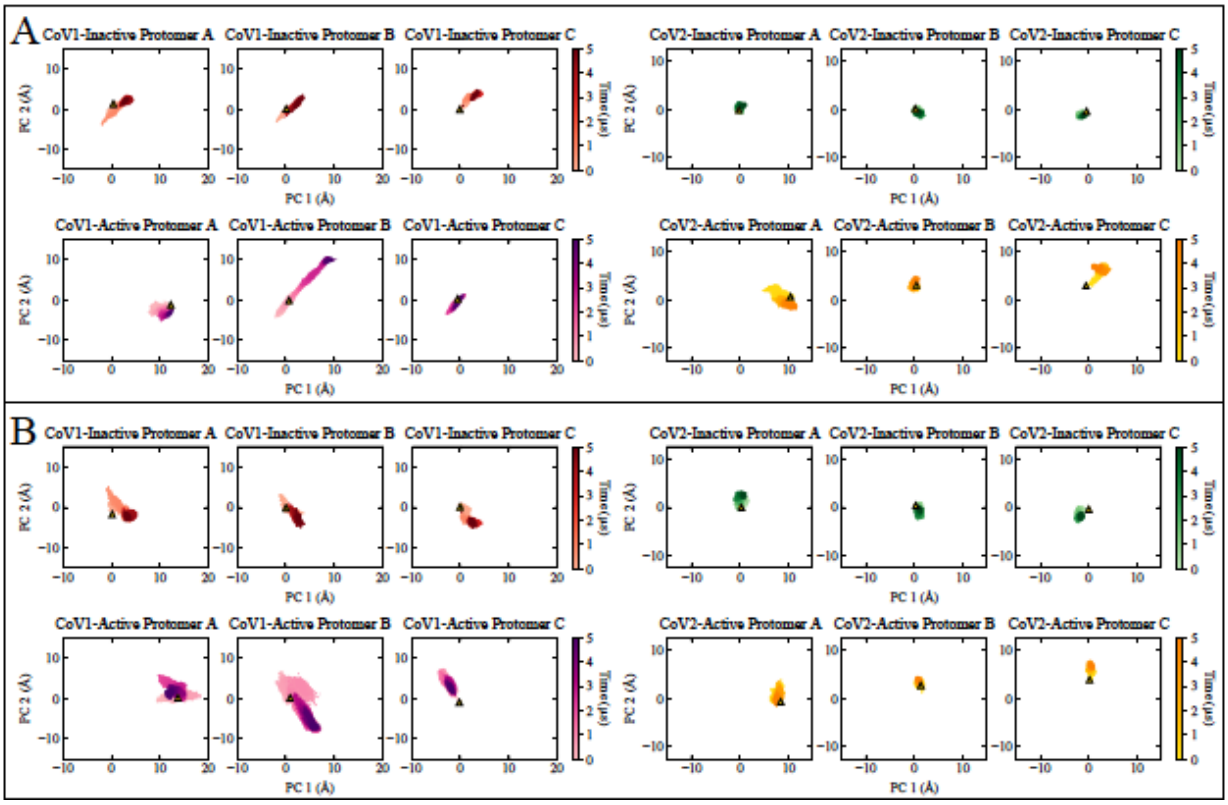


Figure S5: **PCA of all protomers in the inactive and active (Sets 2 and 3) simulations for CoV-1 and CoV-2. (A)** Scatter plot of PC1 and PC2 for Set 2 of CoV-1 and CoV-2 active and inactive spike simulations. **(B)** Scatter plot of PC1 and PC2 for Set 3 of CoV-1 and CoV-2 active and inactive spike simulations. The coloring is the same as seen in Figure 3 with darker shades representing frames towards the end of the simulations. The active CoV-2 spike clearly samples fewer conformations in both PC1 and PC2 spaces than the active CoV-1 spike.

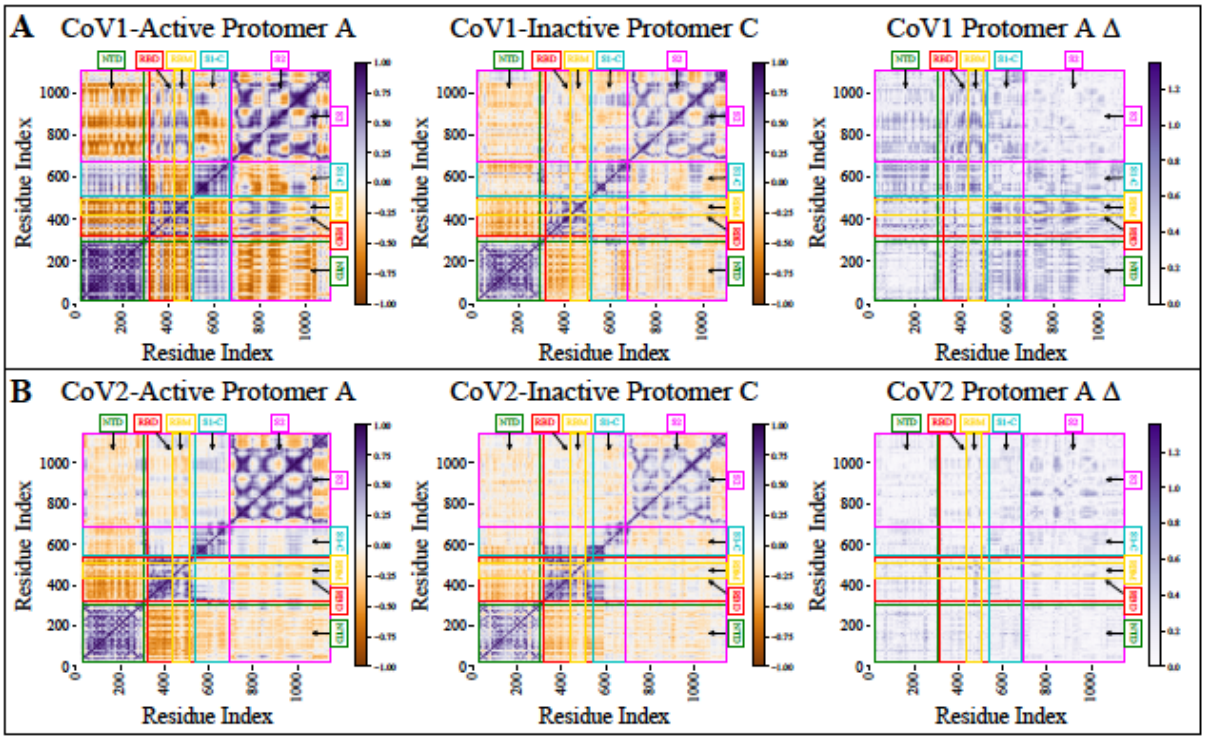


Figure S6: **Dynamic network analysis shows that intra-protomer correlations and anticorrelations are relatively strong in the active CoV-1 spike protein simulations. (A)** DNA heat maps showing the correlation of motions for the active CoV-1 protomer A, inactive protomer C (reference), andthe difference matrix. **(B)** DNA heat maps showing the correlation of motions for the active CoV-2 protomer A, inactive protomer C (reference), and the difference matrix. Correlations are shown in purple and anti-correlations are shown in orange, with the darker colors indicating greater correlation/anti-correlation.Colored labels for the NTD (green), RBD (red), RBM (yellow), S1-C (cyan), and S2 (magenta) regions are positioned over the appropriate residues. The delta matrix identifies differences in protomer correlation between the active and reference inactive protomer. A theoretical maximum for Δ is 2, but the observed maximum was less than 1.3. Differences in correlation are shown as a purple gradient with darker purple indicating larger difference.

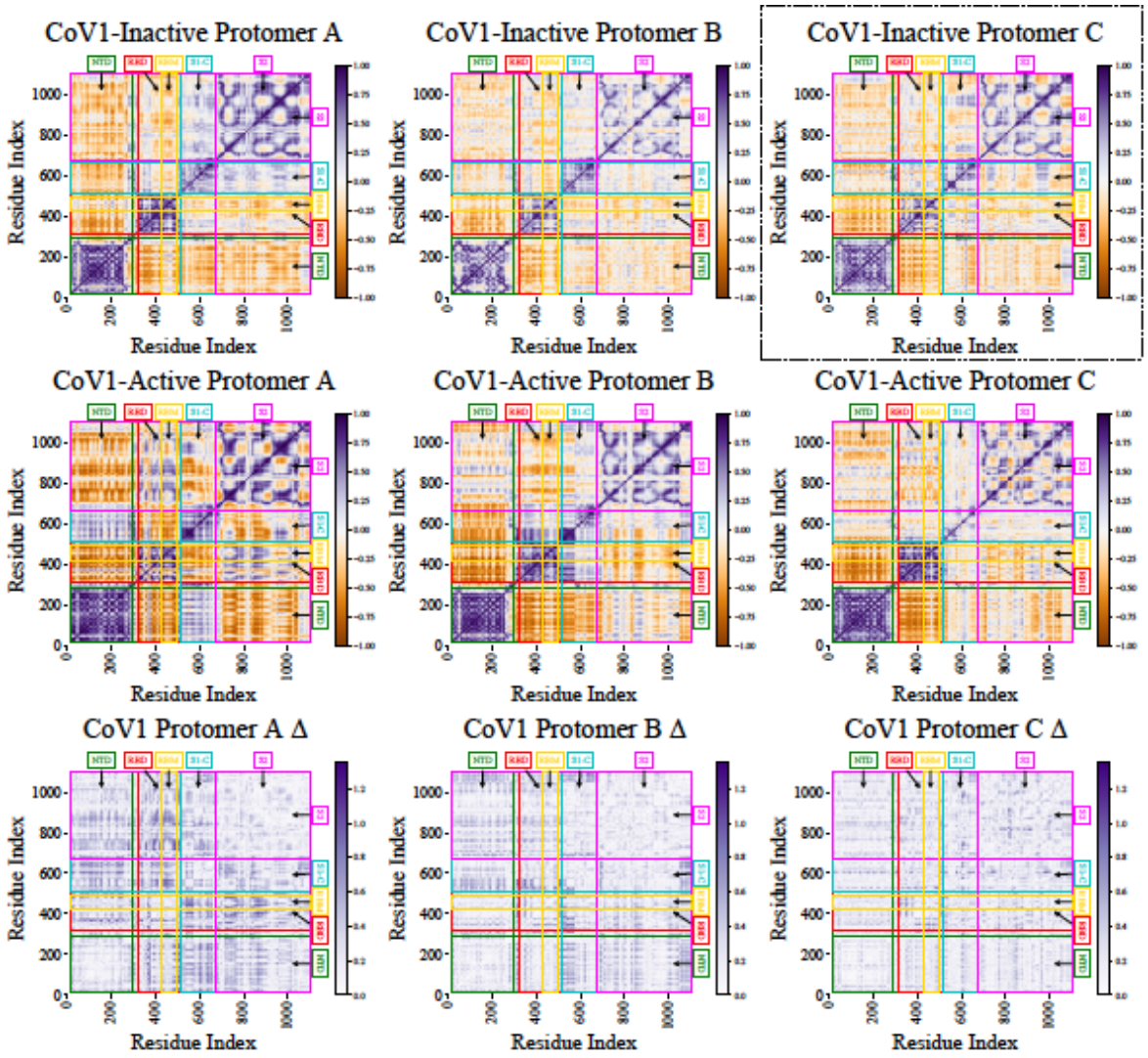


Figure S7: DNA correlation heat maps and Δ matrix for all protomers from the CoV-1 inactive and CoV-1 active (Set 1) spike simulations. DNA heat maps showing the correlation of motions for the CoV-1 inactive protomers (first row), the CoV-1 active protomers from Set 1 (second row) and the difference matrices. The inactive protomer C correlation matrix, indicated by the dotted box, is the reference used for calculating the Δ matrix.

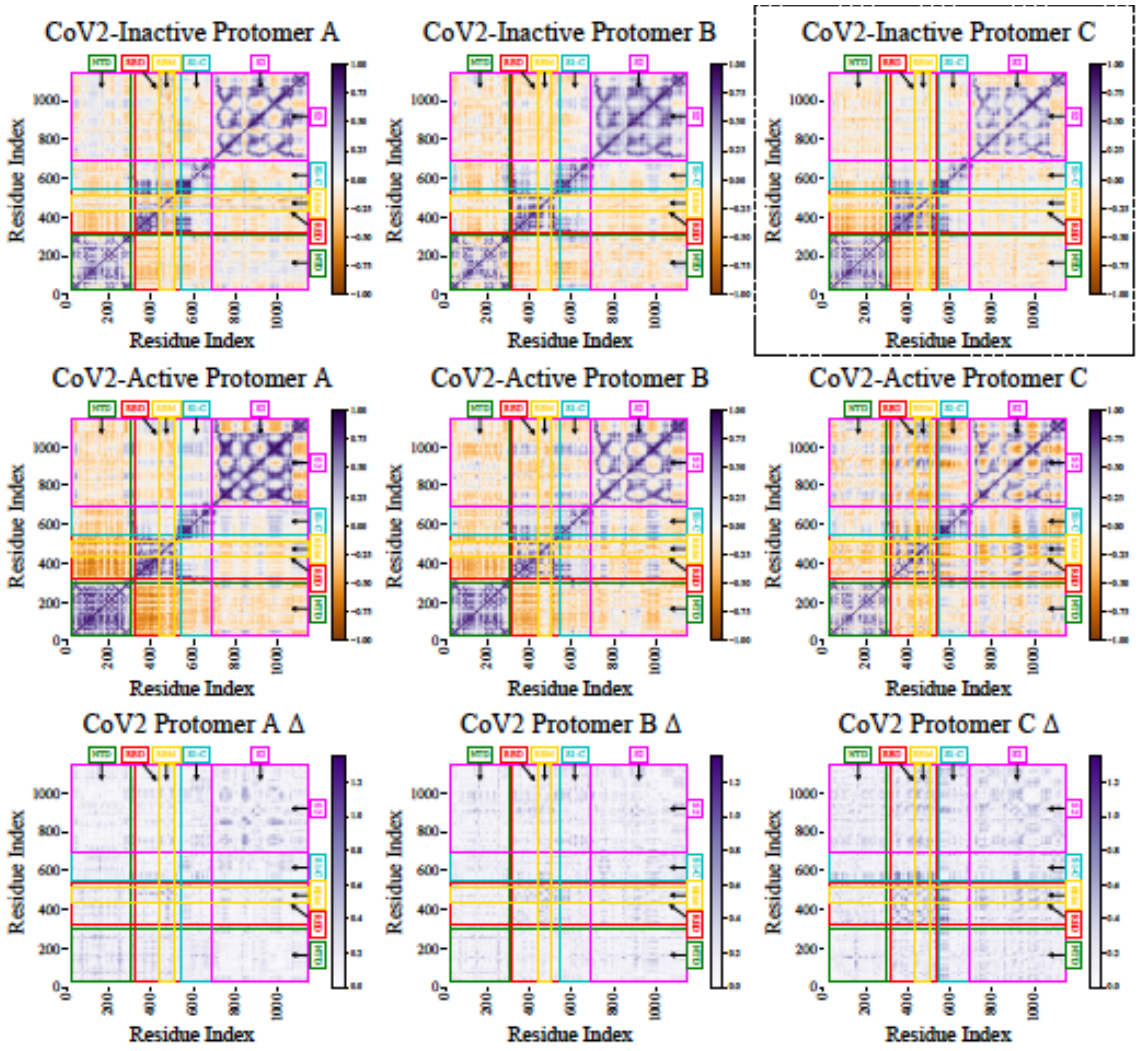


Figure S8: DNA correlation heat maps and Δ matrix for all protomers from the CoV-2 inactive and CoV-2 active (Set 1) spike simulations. DNA heat maps showing the correlation of motions for the CoV-2 inactive protomers (first row), the CoV-2 active protomers from Set 1 (second row) and the difference matrices. The inactive protomer C correlation matrix, indicated by the dotted box, is the reference used for calculating the Δ matrix.

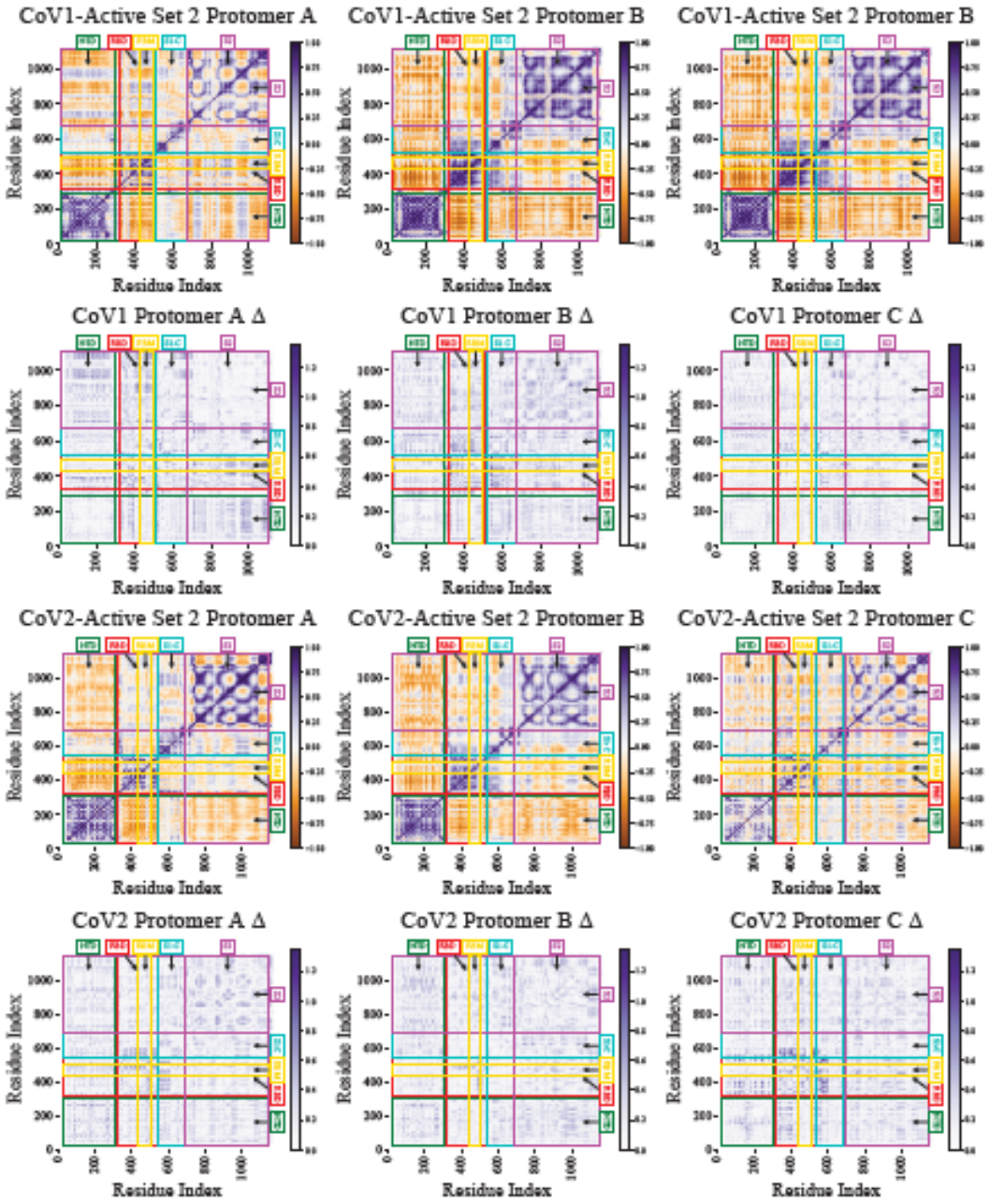


Figure S9: DNA correlation heat maps and Δ matrix for all protomers from Set 2 of the CoV-1 and CoV-2 active spike simulations. (A) DNA heat maps showing the correlation of motions for the CoV-1 active (Set 2) protomers (first row) and the difference matrices (second row). The reference matrix from Figure S7 was used for Δ matrix calculations. **(B)** DNA heat maps showing the correlation of motions for the CoV-2 active (Set 2) protomers (third row) and the difference matrices (fourth row). The reference matrix from Figure S8 was used for Δ matrix calculations.

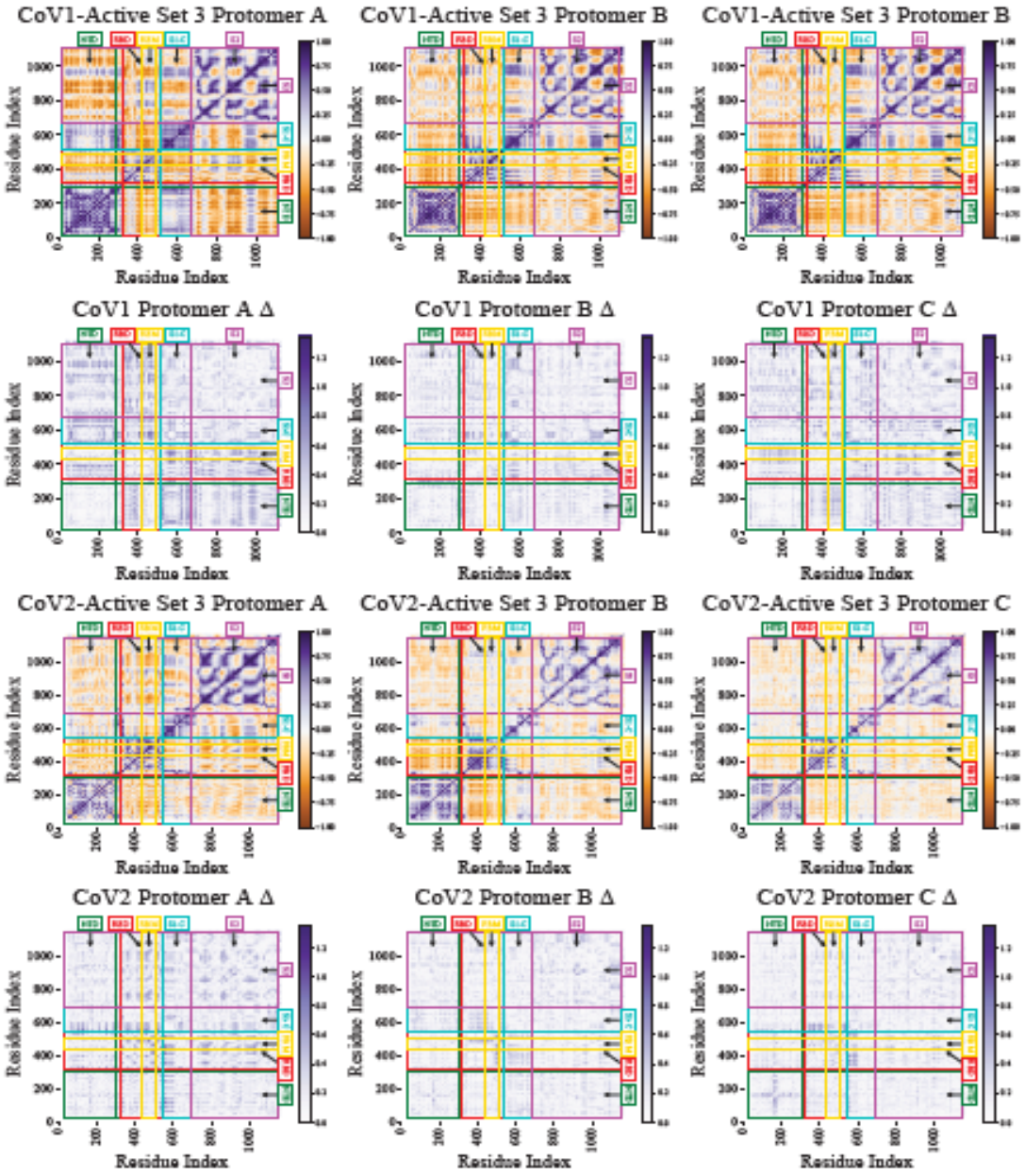


Figure S10: DNA correlation heat maps and Δ matrix for all protomers from Set 3 of the CoV-1 and CoV-2 active spike simulations. (A) DNA heat maps showing the correlation of motions for the CoV-1 active (Set 3) protomers (first row) and the difference matrices (second row). The reference matrix from Figure S7 was used for Δ matrix calculations. **(B)** DNA heat maps showing the correlation of motions for the CoV-2 active (Set 3) protomers (third row) and the difference matrices (fourth row). The reference matrix from Figure S8 was used for Δ matrix calculations.

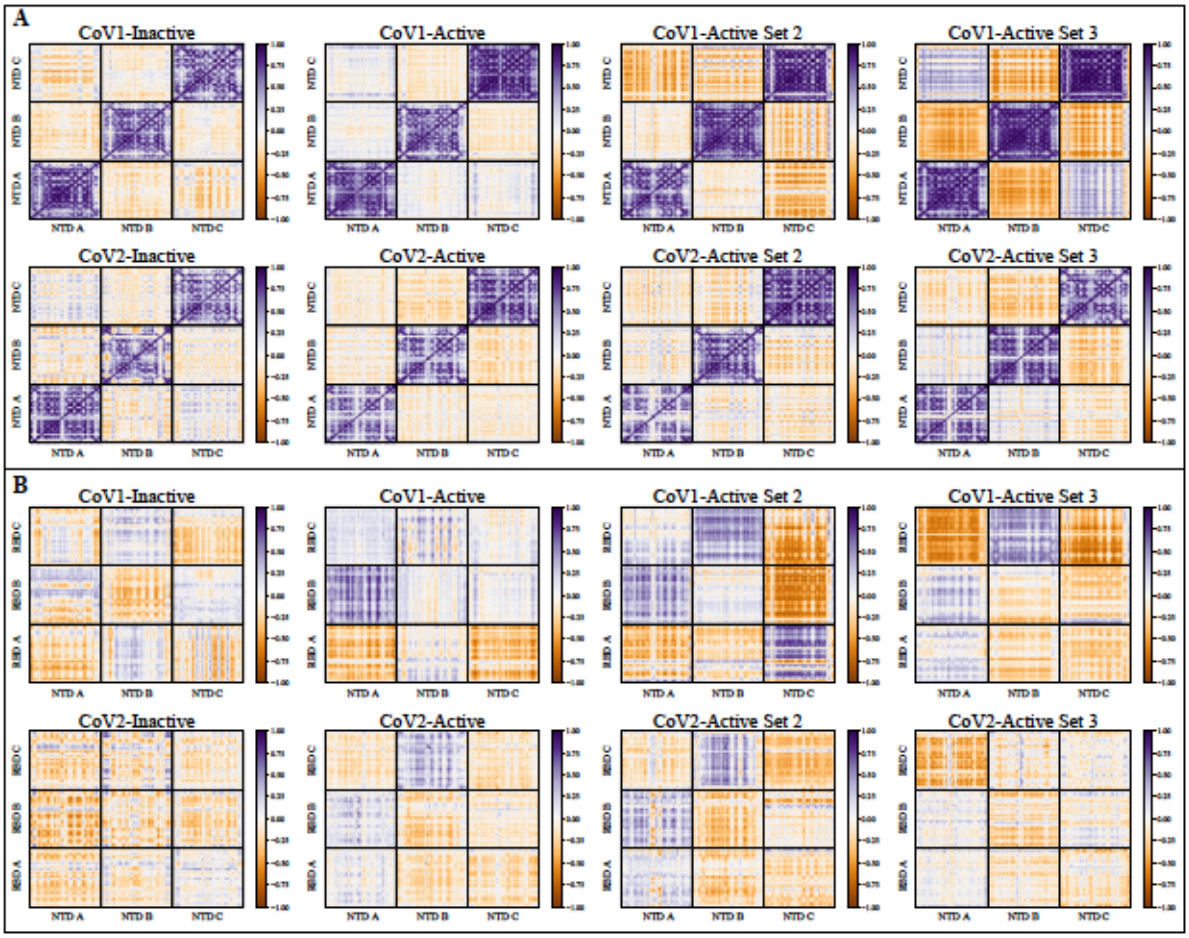


Figure S11: Dynamic network analysis shows that inter-protomer correlations and anticorrelations are relatively strong in the active CoV-1 spike protein simulations. (A) DNA heat maps showing the correlation of motion between the NTD regions of different protomers. **(B)** DNA heat maps showing the correlation of motion between the NTD and RBD regions of different protomers. Correlationsare shown in purple and anti-correlations are shown in orange, with the darker colors indicating greater correlation/anti-correlation.

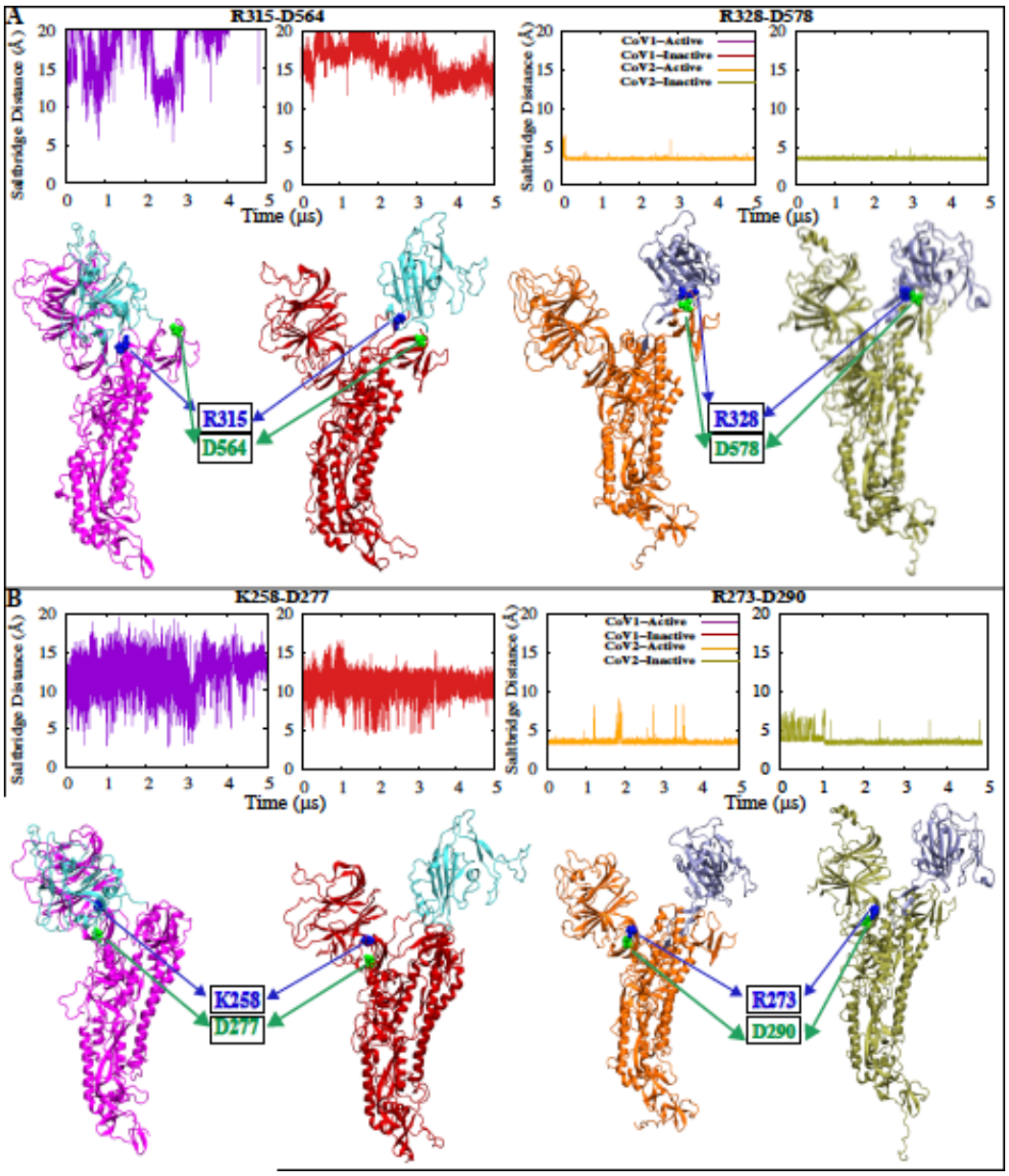
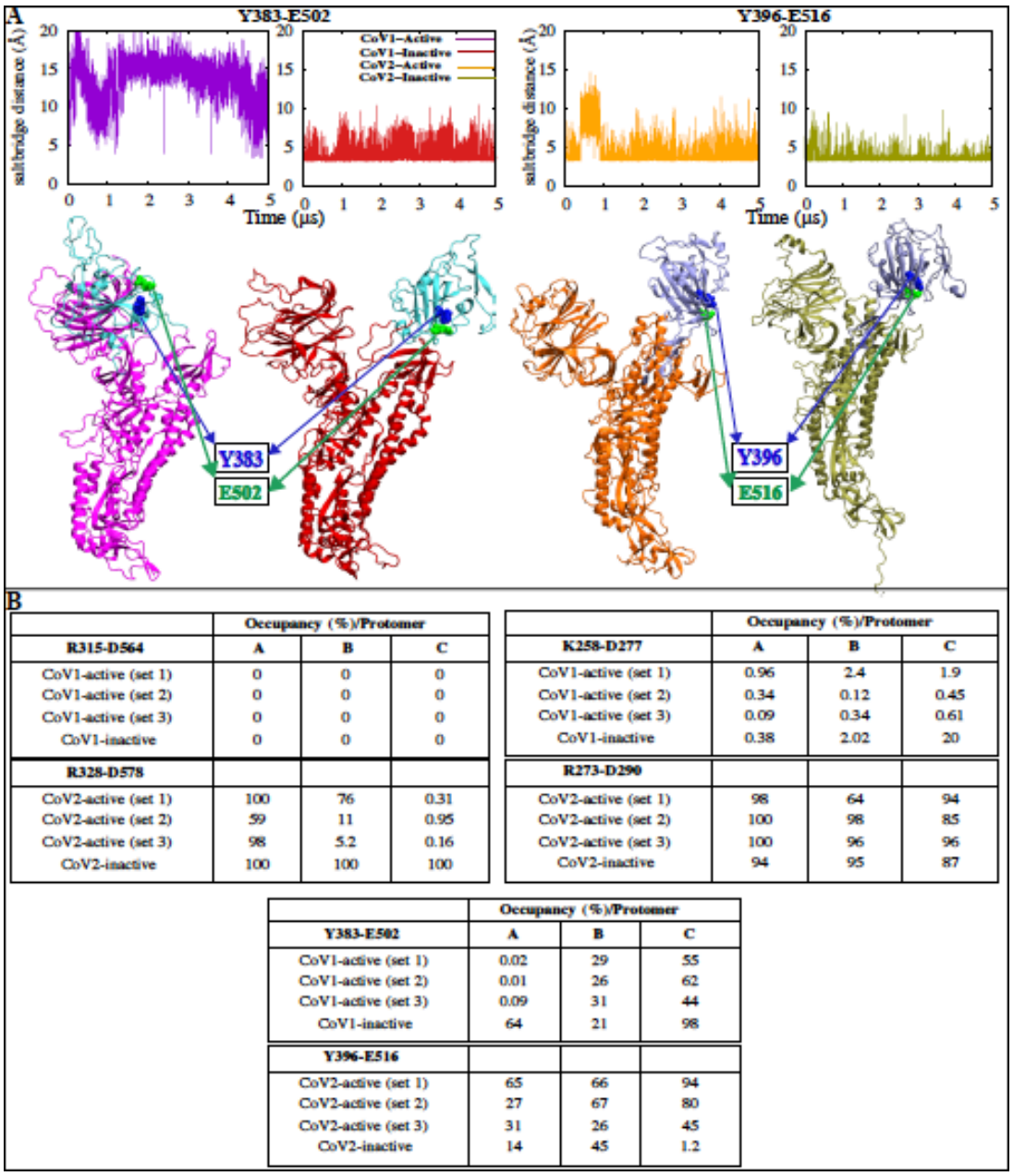


Figure S12: Conserved residues show distinct differential behavior in the CoV-1 and CoV-2 spike proteins. Time series and visual representation of the minimum salt-bridge distance for **(A)** R315/328 (blue) - D564/578 (green) and **(B)** K258/R273 (blue) - D277/290 (green), shows that salt-bridges are formed in the CoV-2 spike protein but are absent in the CoV-1 spike protein. These salt-bridges potentially contribute to the higher relative stability of the CoV-2 spike protein. CoV-1 inactive is colored red, CoV-1 active is colored magenta, CoV-2 inactive is colored olive-green and CoV-2 active is colored orange.



B

	Occupancy (%)/Protomer		
R315-D564	**A**	**B**	**C**
CoV1-active (set 1)	0	0	0
CoV1-active (set 2)	0	0	0
CoV1-active (set 3)	0	0	0
CoV1-inactive	0	0	0
R328-D578			
CoV2-active (set 1)	100	76	0.31
CoV2-active (set 2)	59	11	0.95
CoV2-active (set 3)	98	5.2	0.16
CoV2-inactive	100	100	100

	Occupancy (%)/Protomer		
K258-D277	**A**	**B**	**C**
CoV1-active (set 1)	0.96	2.4	1.9
CoV1-active (set 2)	0.34	0.12	0.45
CoV1-active (set 3)	0.09	0.34	0.61
CoV1-inactive	0.38	2.02	20
R273-D290			
CoV2-active (set 1)	98	64	94
CoV2-active (set 2)	100	98	85
CoV2-active (set 3)	100	96	96
CoV2-inactive	94	95	87

	Occupancy (%)/Protomer		
Y383-E502	**A**	**B**	**C**
CoV1-active (set 1)	0.02	29	55
CoV1-active (set 2)	0.01	26	62
CoV1-active (set 3)	0.09	31	44
CoV1-inactive	64	21	98
Y396-E516			
CoV2-active (set 1)	65	66	94
CoV2-active (set 2)	27	67	80
CoV2-active (set 3)	31	26	45
CoV2-inactive	14	45	1.2

Figure S13: **Hydrogen bond analysis for a conserved residue pair within the RBD. (A)** Time series and visual representation of the minimum H-bond donor-acceptor distance between Y383/396 (blue)and E502/516 (green), in the CoV-1 and CoV-2 spike respectively. CoV-1 inactive is colored red, CoV-1 active is colored magenta, CoV-2 inactive is colored olive green and CoV-2 active is colored orange. Table **(B)** shows the occupancy (%) of the salt-bridge and hydrogen-bond interactions between conserved residue pairs, for all protomers from all simulation sets.

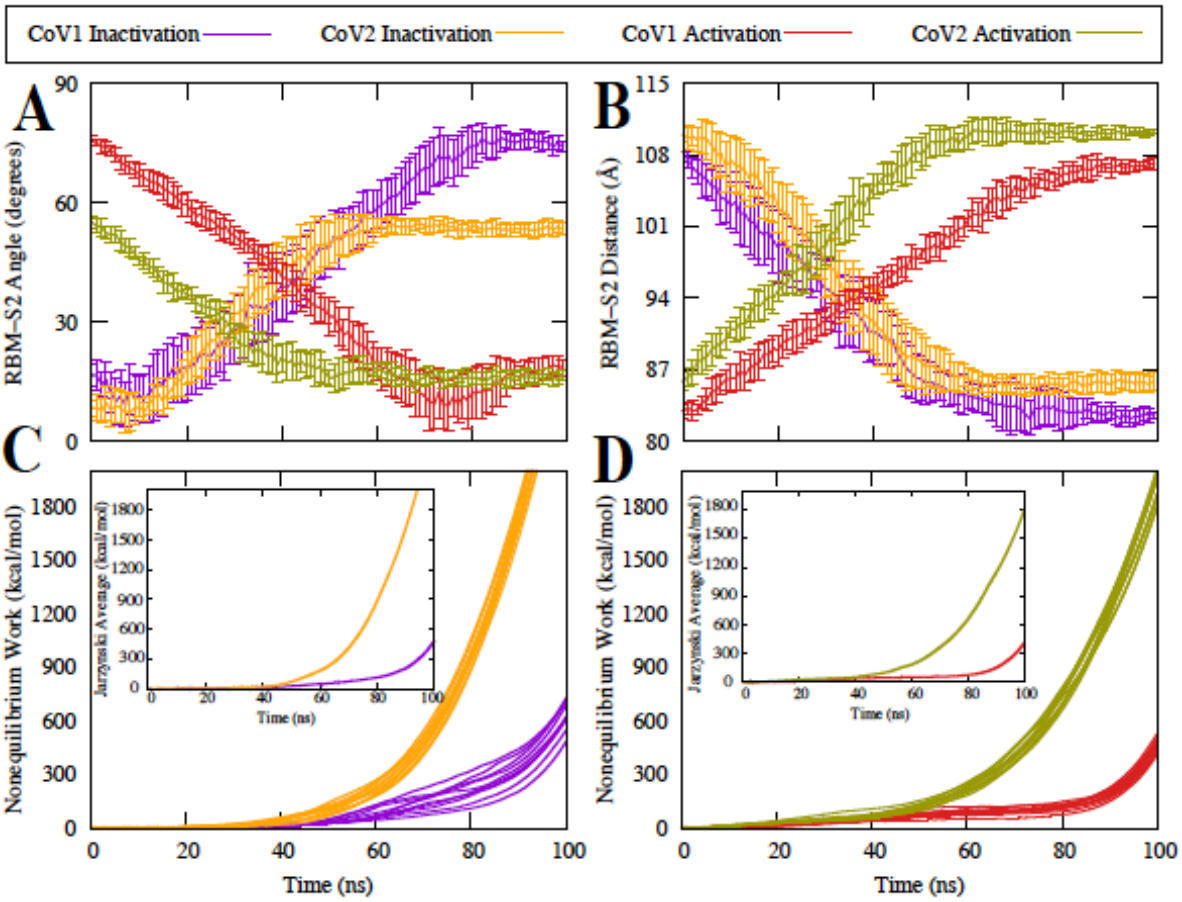


Figure S14: **Three protomer SMD simulations. (A)** RBM-S2 Angle between the beta sheet region of the RBM and the alpha helical region of S2, shown as a function of time. Protomer activation is characterized by a decrease in the RBM-S2 angle. **(B)** RBM-S2 COM Distance between the beta sheet region of the RBM and the alpha helical region of S2, shown as shown as a function of time. Protomer activation is characterized by an increase in the RBM-S2 distance.**(C,D)** Accumulated non-equilibrium work as a function of simulation time. Inset: Jarzynski average of non-equilibrium work.

LIST OF OTHER PUBLISHED PAPERS

1. Baucom, D. R.; Furr, M.; **Govind Kumar, V**.; Okoto, P.; Losey, J. L.; Henry, R. L.; Moradi, M.; Suresh Kumar, T. K.; Heyes, C. D. Transient Local Secondary Structure in the Intrinsically Disordered C-Term of the Albino3 Insertase. *Biophys. J.* **2021**, *120*, 4992-5004.

2. Agrawal, S., **Govind Kumar, V.**, Gundampati, R.K. et al. Characterization of the structural forces governing the reversibility of the thermal unfolding of the human acidic fibroblast growth factor. *Sci. Rep.* **2021**, *11*, 15579.

3. Harkey, T., **Govind Kumar, V**., Hettige, J. et al. The Role of a Crystallographically Unresolved Cytoplasmic Loop in Stabilizing the Bacterial Membrane Insertase YidC2. *Sci. Rep.* **2019**, *9*, 14451.

4. Faouri, R.A., Krueger, E., **Govind Kumar, V.** et al. An Effective Electric Dipole Model for Voltage-induced Gating Mechanism of Lysenin. *Sci. Rep.* **2019**, *9*, 11440.

5. Kiaei, M., Balasubramaniam, M., **Govind Kumar, V.** et al. ALS-causing mutations in profilin-1 alter its conformational dynamics: A computational approach to explain propensity for aggregation. *Sci. Rep.* **2018**, *8*, 13102.