2017

Investigating medicinally important portein-protein and protein-ligand interactions : a computational approach

Cooper Ashley Taylor
University of Richmond

Investigating Medicinally Important Protein-Protein and Protein-Ligand Interactions: A Computational Approach

Cooper Ashley Taylor

Approved by:

Dr. Kelling Donald
Associate Professor of Chemistry
School of Arts and Sciences
University of Richmond

Approved by:

Dr. Carol Parish
Advisor
School of Arts and Sciences
University of Richmond

Investigating Medicinally Important Protein-Protein and Protein-Ligand Interactions: A Computational Approach

By
Cooper Ashley Taylor

Honor's Thesis
Submitted to the
Department of Chemistry
University of Richmond
Richmond, VA

April 20, 2017
Advisor: Dr. Carol A. Parish

Table of Contents

Acknowledgements

I would like to acknowledge and thank the following people who have supported me during the completion of not only the included research, but also throughout my Bachelor's degree.

Firstly, I would like to thank Dr. Carol Parish, to whom I am forever indebted. Carol's guidance has truly shaped who I am as a scientist and as a researcher. The lessons Carol has taught me will forever guide my thought process when approaching problems, both scientific and layman.

I would also like to thank Dr. Billy Miller III, my mentor and friend. Without his instruction, advice, and counsel I would not have been able to accomplish the feats I have up to this point in my scientific career. Bill's encouragement and support of my endeavors have been a true driving force behind the development of my scientific curiosity and passion.

Lastly, Dr. Kelling Donald and Dr. Michael Leopold have also contributed greatly to my academic advancement and personal growth, and I am truly grateful for their direction over the course of my undergraduate studies.

Abstract

Molecular dynamics (MD) simulations and computational chemistry allow for an atomistic understanding of protein-protein and protein-ligand binding motifs. Through the use of MD, medicinally relevant complexes can be examined in detail unattainable by experimental methods. Within this work, systems pertinent to both Alzheimer's Disease and HIV-1 are probed and thoroughly sampled to help elucidate potential therapeutic pathways. We used molecular dynamics and free energy estimations to gauge the affinity for the binary and ternary complexes of KLC1, APP and JIP1, three proteins all believed to be involved in the propagation of Alzheimer's Disease. Two areas of thought exist suggesting that APP is either transported in a binary KLC1:APP complex, or is assisted by a third protein, JIP1, in a ternary KLC1:JIP1:APP complex. We find that all binary and ternary complexes (KLC1:APP, KLC1:JIP1, APP:JIP1, and KLC1:JIP1:APP) contain conformations with favorable binding free energies, and that the ternary KLC1:JIP1:APP complex shows signs of being thermodynamically more favorable than the binary KLC1:APP complex. With regards to the HIV-1 studies, a pyrazolo-piperdine ligand was recently synthesized and the corresponding biological data showed good binding to both CCR5 and CXCR4, receptors involved in the HIV-1 lifecycle, thus effectively preventing HIV-1 entry. After extensive sampling, we find that π-stacking interactions between the ligand and receptor residues, as well as electrostatic interactions involving the protonated piperidine nitrogen are the driving forces behind the ligand-protein binding. We also propose and computationally verify a new, synthetically-accessible derivative designed to increase the electrostatic interactions without compromising the π-stacking features.

Introduction

A common method to gain an atomistic understanding of protein-protein and protein-ligand binding is using computational chemistry, and more specifically, molecular dynamics (MD) simulations.[1-3] MD is based on Newton's second law (Equation 1). Newton's equation of motion can also be recast as the negative gradient of the potential energy (Equation 2). Given that acceleration is the second derivative of displacement, it is possible to combine Equations 1 and 2 to reveal the relationship between the change in potential energy and the movement of the system particles as function of time (Equation 3).

Equation 1. Newton's second law.

$$F_i = m_i a_i$$

Equation 2. Newton's second law in terms of potential energy.

$$F_i = -\nabla_i \mathrm{V}$$

Equation 3. Development of potential energy service as particle moves over time.

$$-\frac{dV}{dx_i} = m_i \frac{d^2 x_i}{dt^2}$$

This relationship depicts how Newtonian physics can be utilized to sample a molecular potential energy surface. Simulations are initiated by the assignment of random velocities to each atom as determined by the Boltzmann distribution at the specified temperature. It is at this point that Newtonian physics defines the path on which the various components of the system travel, allowing for the atoms to transfer momentum and energy to one another. By integrating Eq 3, and using classical force fields (*vide infra*) to define the potential energy at each point in time, the various positions, velocities, and accelerations of each atom can be determined. This trajectory can

be used to further examine the state of the system at various time stamps, as well as determine the average values of properties such as the free energy of the system.

Classical force fields are an essential component of MD simulations. Classical molecular force fields utilize covalent and noncovalent energy terms to calculate a system's potential energy (Equation 4). Each energy term has a functional form by which it is calculated, but these calculations may vary depending on the force field used. Each force field assigns the energy terms parameters to most accurately derive the potential energy surface. The chosen parameters tend to be derived from experimental results, as well as from high level quantum mechanical calculation on model compounds. The energy terms, in conjunction with the geometric coordinates, partial charges, atomic masses, atomic radii, and defined parameters form the overall molecular force field.

Equation 4. Potential energy of a classical force field.

$$E_{potential} = E_{bond} + E_{angle} + E_{dihedral} + E_{electrostatic} + E_{vdW}$$

The binding free energies are estimated from the resulting MD trajectories using the MM-PBSA and MM-GBSA approximations.[4] In this study the binding free energy between the dual-tropic inhibitors **3** and **3.1**, and the CCR5 and CXCR4 receptors, was determined using MM-GBSA. We also used this approach to determine the per residue free energy decomposition of key amino acids within the CCR5 and CXCR4 receptors. The MM-G(P)BSA approach is based upon the thermodynamic cycle shown in Equation 5.[5] This equation demonstrates that the overall protein-ligand binding free energy is the difference between the free energy of the fully solvated complex and the fully solvated, individual, protein and ligand.

Equation 5. Calculation of the binding free energy of a protein-ligand complex.

$$\Delta G_{binding,solvated} = \Delta G_{complex,solvated} - \left(\Delta G_{receptor,solvated} + \Delta G_{ligand,solvated}\right)$$

As seen in Equation 6, the free energy of each right-hand term of Eq 5 can be described as consisting of 1.) a gas-phase energy component, which can be approximated by force-field based molecular mechanics calculations (E_{gas}), 2.) the free energy of solute-solvent and solvent-solvent interactions as approximated by an implicit solvent model ($\Delta G_{solvation}$), and 3.) the entropic contribution of the solute (S_{solute}) scaled by the temperature of interest.[5] However, it is important to note that the entropic contribution of the solute is computationally demanding and is often not included in the determination of the binding free energies.

Equation 6. Calculation of the free energy terms of a protein-ligand complex.

$$\Delta G_{solvated} = E_{gas} + \Delta G_{solvation} - TS_{solute}$$

The calculated free energies from Eq 6 can be utilized to further examine the specific contributions of individual (or pairs of) residues within a system, resulting in the per-residue free energy decomposition equation (Equation 7).[6] This approach allows for the examination of specific contributions of each residue by breaking down the overall free energy contribution into the internal, van der Waals, electrostatic, polar solvation, and non-polar solvation energy terms. The ΔG_{decomp} only pertains to the contribution of individual residues to the overall free energy of the system, while Eq 5 describes the free energy of the solvated system as a whole.

Equation 7. Calculation of the per-residue free energy decomposition.

$$\Delta G_{decomposition} = E_{internal} + E_{vdW} + E_{electrostatic} + E_{polar\ solv} + E_{non-polar\ solv}$$

These theories and calculations allow for a thorough investigation of the potential of a drug or inhibitor to bind to a target receptor or enzyme.[1-3,6] The binding free energy of a system provides a broad characterization of a system, while the energy values of the various components of each specific residue as provided by the decomposition calculation provides reasoning for the binding free energy results.

In the following studies, we aimed to 1) evaluate the binding dynamics between the binary complexes of three proteins believed to be involved in Alzheimer's disease, 2) examine the binding interactions of potential ternary complexes of those same such proteins, and 3) investigate the protein-ligand interactions between a previously reported dual-tropic inhibitor of CCR5 and CXCR4, the two main receptors responsible for HIV-1 infection of healthy T-helper cells, as well as a derivative thereof. We believe a more thorough comprehension of the atomistic interactions and dynamics of these protein-protein and protein-ligand complexes will provide fundamental information with which to elucidate future therapeutic pathways.

A Molecular Dynamics Study of the Binary Complexes of APP, JIP1, and the Cargo Binding Domain of KLC

Cooper A. Taylor[1], Bill R. Miller III[2], Soleil S. Shah[1], and Carol A. Parish[1]

[1]*Department of Chemistry, Gottwald Center for the Sciences, University of Richmond, 28 Westhampton Way, Richmond, VA 23173*

[2]*Department of Chemistry, Truman State University, 100 E. Normal Ave, Kirksville, MO 63501*

Mutations in the Amyloid Precursor Protein (APP) are responsible for the formation of Amyloid-β peptides. These peptides play a role in Alzheimer's and other dementia-related diseases. The cargo binding domain of the kinesin-1 light chain motor protein (KLC1) may be responsible for transporting APP either directly or via interaction with C-jun N-terminal kinase-interacting protein 1 (JIP1). However, to date there has been no direct experimental or computational assessment of such binding at the atomistic level. We used molecular dynamics and free energy estimations to gauge the affinity for the binary complexes of KLC1, APP and JIP1. We find that all binary complexes (KLC1:APP, KLC1:JIP1 and APP:JIP1) contain conformations with favorable binding free energies. For KLC1:APP the inclusion of approximate entropies reduces the favorability. This is likely due to the flexibility of the 42-residue APP protein. In all cases we analyze atomistic/residue driving forces for favorable interactions.

Introduction

Alzheimer's disease (AD) is the leading cause of dementia in the elderly. Early onset familial AD has been shown to be caused by mutations in three different gene products; Presenilin 1 (PS1), Presenilin 2 (PS2), and Amyloid Precursor Protein (APP). Significant research is underway to understand the role these proteins play in all dementia-like pathologies.[7-11] APP in particular has been the subject of much study as proteolysis releases amyloid-β peptides.[12] It is the accumulation of these peptides and subsequent senile plaque formation that is considered a causative agent in AD.[13-16]

Kinesin-1 is a tetrameric protein composed of two heavy chains (KHC) and two light chains (KLC), which make up its three domains: the globular motor head, which contains the

ATPase and is responsible for microtubule binding activity; the dimeric, coiled-coil stalk connecting region; and the cargo-binding tail.[17] The cargo-binding region is compromised of the C-terminal 150 amino acids of the heavy chains plus the light chains. The light chains have been shown to consist of highly helical TPR units that may function as a cargo-binding domain via a protein-protein interaction motif with APP.[18-20] Each TPR is a 34 amino acid sequence that adopts a helix-turn-helix structure and occurs in tandem arrays, typically 3-16 repeats in length.[21] Each repeat motif forms two antiparallel α-helices that stack together in a parallel arrangement to produce an extended molecule with an overall right-handed super-helical tertiary structure. This spiral staircase-like tertiary fold presents an inner amphipathic concave surface to which cargo has been suggested to bind.[22-24] When at least 3 TPRs are present, a groove is formed which is ideally suited for binding to the α-helix of a target protein such as APP.[25]

Alzheimer's-related proteins such as APP have been shown to reside within the axon and undergo fast, plus-end directed (anterograde) axonal transport.[26-28] Models have been proposed that suggest APP is transported via the light chain units of kinesin-1 (Figure 1).[27,29-32] Co-immunoprecipitation experiments, along with *in vitro* binding measurements, have demonstrated that the C-terminal end of APP displays direct nanomolar binding (15-20 nM) to the tetratricopeptide repeat (TPR) region of the kinesin-1 light chain (KLC1) with a stoichiometry of two APP units per KLC1.[33] We hypothesize that a mis-regulation of the cargo binding domain in motor proteins may disrupt the normal degradation processes leading to an accumulation of Aβ and subsequent formation of plaques. In light of recent findings that amyloid-β may play a protective role in healthy individuals an atomistic examination of this hypothesis is necessary.[11]

APP binding to KLC1 may be facilitated by the interaction of the c-Jun N-terminal kinase (JNK)-interacting protein 1 (JIP1), as seen in Figure 1B.[34] The 707 residue JIP1 protein is localized

in the neuronal region of the brain, suggesting a critical role in cell communication.[35] Barr et al. have shown that an 11 residue JIP1 peptide fragment (residues 553-563) functions similarly to the full protein in various known binding interactions.[36] Two experimental structures containing ten of the eleven residues of the JIP1 peptide fragment are available (PDBID 2H96 and 1UHK).[37,38] The missing residue is Arg553 and this is the first residue of the fragment. Experimentally, Arg556, Pro557, Leu560 and Leu562 have been shown to be the most critical residues for the retention of function. Additionally, it was shown that removing either the C-terminus or N-terminus of the 11 residue peptide had no impact on function.[36] Since Arg553 is the N-terminus of the 11-mer, and its absence showed no major impact on binding affinity, it was not added to the 10 residue crystal structure.

JIP1 can also interact directly with KLC1.[39] A previous study suggested that JIP1 binds to a "polar patch" region located in the center of the concave side of KLC1.[40] JIP1 and APP may also interact directly. Fluorescence resonance energy transfer (FRET) was used to probe direct interaction of JIP1 with the intracellular domain of APP.[41] In addition, an experimental study of neurons on JIP1 knock-out mice showed a 21% decrease in frequency and a nearly 1.0 μm/s decrease in velocity of APP transport (from 2.7 to 1.83 μm/s).[42] Further information describing the various models of APP transport and aggregation can be found in a 2015 review article by van der Kant and Goldstein.[43]

There is significant evidence that KLC1:APP axonal transport dysfunction plays a critical role in AD progression.[29,32,43] However, this hypothesis and the corresponding model shown in Figure 1 is not without some controversy – results published in 2005 refute the existence of any specific association between KLC1 and APP *in vivo*.[44] However, later work in 2005 by Stokin et al. affirmed that APP undergoes fast anterograde axonal transport and that APP does interact with

KLC1, but suggested the hydrophobic nature of the TPR motifs prevents cargo-binding specificity.[29] While the Goldstein group has demonstrated tight binding between mouse KLC1 and hAPP *in vivo*, to the best of our knowledge, no one has yet performed direct binding experiments using purified hKLC1 and hAPP compounds, *in vivo* or computationally.

In this study, we aimed to evaluate the potential for the formation of binary complexes between KLC1, APP and JIP1. Towards that end we 1) evaluate the propensity for APP to bind directly to the TPR region of human KLC1, 2) determine the efficacy of JIP1 binding to the TPR region of human KLC1, and 3) assess the binding of JIP1 to APP. We believe that a better understanding of the structure and dynamics of these binary protein-complexes will provide fundamental information with which to better assess the role of KLC1 and JIP1 in APP processing for Alzheimer's disease and other dementia-related diseases as well as to cast light on the fundamental binding behaviors of a motor protein cargo-binding domain.

Methods

Structure Retrieval

Structures of KLC1 and JIP1 were obtained from the Protein Databank.[45] The original file (PDB 3NF1) of KLC1 consisted of 304 amino acids. We removed the last 11 N-terminal residues so that the KLC consisted of only six α-helices.[46] These 11 amino acids represented an artificially engineered histidine tail and were removed so the hKLC1 model would more accurately resemble the *in vivo* structure. The structure for JIP1 was derived from PDB 2H96.[47] The PDB file contained a dimer of JIP1 as well as the mitogen-activated protein kinase 8. We removed the kinase, as well as one of the two JIP1 structures (chain G), leaving the 10-residue monomer. The solution-phase NMR structure of the C-terminal 42 residues of APP was obtained through personal

communications with Dr. Charles Sanders at Vanderbilt University.[48] Schrodinger's Protein Preparation workflow was used to add missing hydrogen atoms to each 3D structure as well as to ensure physiologically correct protonation states.[49]

Analyzing Binding Sites in KLC1

The SITEMAP package from Schrodinger was used to identify and score putative binding sites in KLC1 based upon shape complementarity and solvent accessibility. SITEMAP identifies binding sites based upon an analysis of hydrophobic surface area, surface concavity, as well as location of hydrogen-bond donor and acceptor regions.[50,51] Three sites were identified within the cargo binding domain (Figure 2). Site 1 occurs within the extended loop region between TPR 5 and TPR 6. The lack of definitive secondary structure in the loop region makes this a difficult target to characterize. Experimental evidence suggests that JIP1 interacts with the TPR regions of KLC1 [39] so we are focusing largely on binding sites within the TPR regions and not on the extended loop region of Site 1. Although preliminary results are reported below, we will reserve a more detailed conformational characterization of the loop regions to a future study. Site 2 spanned TPR 1 and TPR 2 in a perpendicular orientation. Site 3 lies parallel to TPR4 and corresponds to the "polar patch" region first discussed by Zhu et al.[40]

Preparing Binary Complexes using Docking Tools

Two different docking algorithms were used to prepare the binary complexes. This increases the likelihood of obtaining good coverage of conformational space and generating a diverse set of docked poses.[52-56] GRAMM-X, a public protein docking software developed by the Vakser Lab at the University of Kansas, uses a rigid-body Fast Fourier Transform algorithm and smoothed potentials to generate the desired number of docked and scored poses.[55,56] We chose to use GRAMM-X in this study because it was specifically designed for protein-protein docking and

has shown good ability to predict protein-protein interactions in the CAPRI (Critical Assessment of PRediction of Interactions) competition.[55,56] We also used Schrodinger's Glide program to flexibly dock JIP1 into KLC1 and to generate docked poses of JIP1 with APP.[52-54] All Glide docking calculations used the Standard Precision mode. The Glide aglorithm provides a systematic search of flexible protein and ligand space followed by energetic refinement and pose enrichment using Monte Carlo sampling.[57] Glide has been shown to produce ligand poses in good agreement with experimental data.[58]

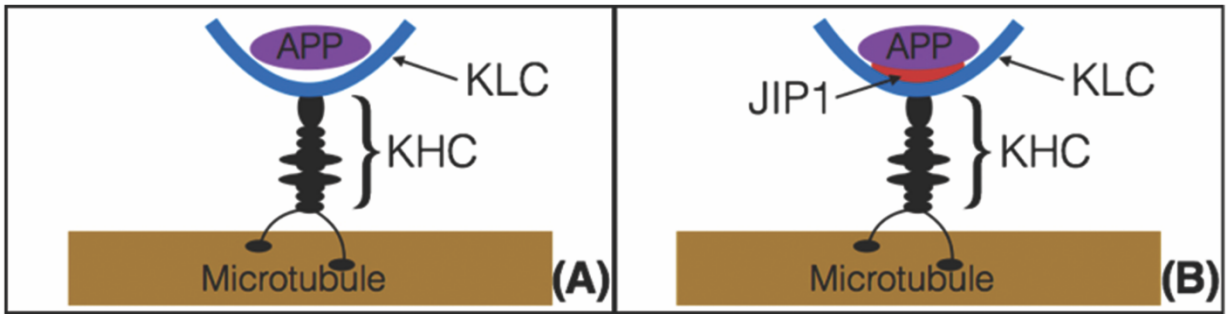

Figure 1. Two proposed models for the transportation of APP across microtubules either by (A) a binary KLC1:APP binding motif or by (B) a ternary KLC1:JIP1:APP binding motif. These models are not drawn to scale with respect to the size of the proteins and the positioning of the cargoes within KLC1.

KLC1:APP

GRAMM-X was used to produce 10 models of varying KLC1:APP starting conformations.[55,56] Of the 10 complexes that were produced, 8 positioned APP within the cargo binding domain, while 2 placed the APP outside of the cup of the KLC-1. GRAMM-X ranks each model based upon energetic criteria including intermolecular interaction energy, volume, contact energy and binding free energy estimation. The top four poses in which APP was positioned inside of the groove of KLC1 (2 each within Site 2 and 3) were subjected to subsequent MD. Glide was not used to generate any KLC:APP conformations due to the program's maximum atom limit that prevented either molecule from being treated as a ligand.

JIP1:APP

GRAMM-X and Glide were used to dock JIP1 into APP.[52-56] The Glide receptor grid was centered on APP as APP is the larger of the two proteins. A 30x40x40Å boundary box that included the entire polypeptide chain was used. Glide produced 19 conformations while GRAMM-X produced 20 conformations, resulting in a total of 39 binary APP:JIP1 complexes. All of these poses were used to initiate molecular dynamics simulations due to the increased conformational flexibility expected for this binary complex.

KLC1:JIP1 Site 1

To generate preliminary information regarding the importance of binding to the flexible loop region, we also performed MD on three highly ranked GRAMM-X poses that positioned JIP1 in Site 1 (Figure 2). These three models will be referred to as "KLC1:JIP1 Site 1."

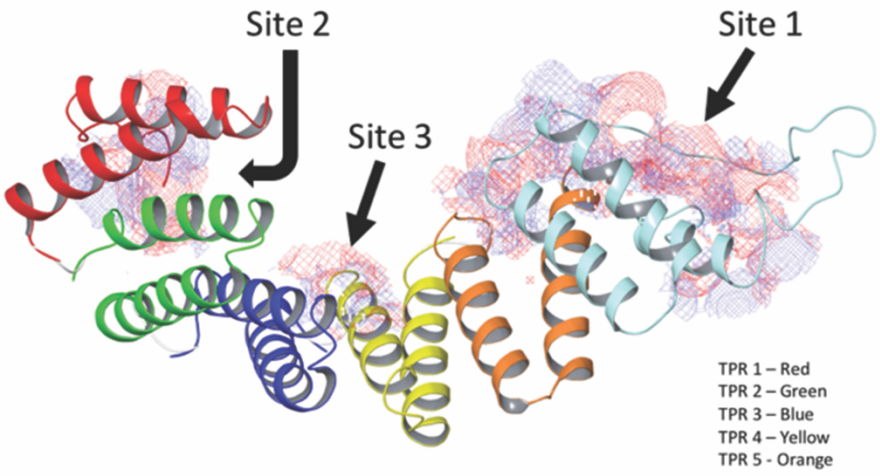

Figure 2. KLC1 active sites were produced using SITEMAP. Site 1 exists within the extended loop region of the KLC1, Site 2 is perpendicularly oriented across TPR 1 and TPR 2, and Site 3 lies within the "polar patch" region, mostly parallel to TPR 4.

KLC1:JIP1 Site 2

A Glide receptor grid of 40x40x40Å was centered on the cargo-binding domain of KLC1.[42] Experimental evidence suggests that JIP1 acts as a mediator between APP and KLC1 and binds inside the cargo-binding domain.[42] Therefore, we chose the 10 highest ranked Glide poses that

positioned JIP1 within the KLC1 cargo binding domain for subsequent molecular dynamics simulation. Since the JIP1 structure within these models tended to migrate towards Site 2 of the KLC1 during simulations, as discussed below in the Results, these models will be referred to as "KLC1:JIP1 Site 2."

KLC1:JIP1 Site 3

To assess the importance of the polar patch (binding Site 3; Figure 2) we used GRAMM-X with KLC1:JIP1 and retained the top three poses in which JIP1 was positioned within the polar patch region.[40,52-56] We also used Glide with a 26x32x26Å grid centered on the central four TPR units of the cargo-binding region of KLC1 (Figure S1). This produced four highly ranked poses in which JIP1 was located within Site 3 (Figure 2). All seven of these poses (GRAMM-X plus Glide results) were used to initiate MD simulations.

Zhu et al. suggested an asparagine to serine point mutation within the polar patch region of Site 3 (N157S) results in a drop in binding affinity, so GRAMM-X and Glide were also utilized to dock JIP1 to this mutated form of KLC1.[40] GRAMM-X produced 20 unique models; 3 of which positioned JIP1 within the mutated polar patch region. The receptor grid used for Glide docking focused the search on Site 3 (as described above for WT; Figure S1), producing 11 poses which positioned JIP1 within the mutated polar patch region. These models (both WT and mutant) will be referred to as "KLC1:JIP1 *Site 3*."

Molecular Dynamics Analysis of Binary Complexes KLC1:APP, KLC1:JIP1, and APP:JIP1

All simulations were performed under identical constraints. The Amber *ff12SB* force field was applied to all simulations.[59] AmberTools' *tleap* was used to neutralize each system with Na^+/Cl^- ions and solvate a truncated octahedron periodic box with TIP3P water molecules.[60,61]

There was initially at least 12.0 Å of solvent between the solute and edge of the unit cell. The GPU-accelerated *pmemd* code of Amber 14 was used to perform all steps of MD for each simulation.[61,62] The initial structures were minimized, heated, equilibrated, and then run under unrestrained MD. The minimization process consisted of seven stages, each comprising a maximum of 5,000 steps. The first 1,000 steps were of steepest descent minimization, and the remaining 4,000 steps were of conjugate gradient minimization. The first of the seven stages was given a restraining force of 10.0 kcal/mol/Å^2 on the heavy atoms of the solute and this was methodically lowered to 0.0 kcal/mol/Å^2 by stage seven. After minimization, each structure was heated from 10 K to 300 K with a restraining force of 10.0 kcal/mol/Å^2 on the solute. Equilibration consisted of lowering the restraining force every 500 ps from 10.0 kcal/mol/Å^2 to 0.0 kcal/mol/Å^2 over seven stages. After the kinetic energy equilibrated between the solvent and structure with no restraining force, unrestrained MD at constant pressure (1 atm) and temperature (300 K) was commenced. The SHAKE algorithm was used to restrain all covalent bonds to hydrogen atoms.[63] This improved the computational efficiency of the simulation and allowed the use of a 2-fs time-step in the molecular dynamics simulation.

Unrestrained MD was used to explore the conformational flexibility of all docked poses described above. Details are shown in Table 1. APP is larger (42 residues) and has more degrees of freedom relative to JIP1 (10 residues), so the KLC1:APP simulations were run longer than either the KLC:JIP1 or APP:JIP1 simulations to better sample the more complex potential energy surfaces. In the case of KLC:JIP1 Site 2, preliminary MM-GBSA binding energy analysis was used to focus the search to three (down from ten total) of the tightest binding conformations. The sequence of each protein being studied can be found in Figure S3.

Table 1. Simulation time for all docked poses used as initial structures for molecular dynamics.

	# of GRAMM-X poses	MD simulation time (ns)	# of Glide poses	MD simulation time (ns)
KLC1:APP	4	350		
KLC1:JIP1 *Site 1*	3	75		
KLC1:JIP1 *Site 2*			6	25
			1	50
			3	125
KLC1:JIP1 *Site 3* (WT)	3	25	4	25
KLC1:JIP1 *Site 3* (mutant)	3	25	11	25
APP:JIP1	20	25	19	25

MD Analysis

The calculation of root-mean-square deviations (RMSD), distance measurements, RMS residue fluctuations (RMSFs), average structures, and clustering was performed using *cpptraj* commands included in AmberTools 14.[61,64] Binding free energies, pairwise per residue free energy decomposition values (ΔG_{bind}), and entropies were calculated using MMPBSA.py.[5] We utilized the MM-GBSA approximation of binding free energies as implemented by Onefriev et al.[65] Normal mode analysis was used to calculate the entropic contribution to the binding free energy.[5] Binding free energies, per residue free energy decomposition values, and alanine scanning were calculated every 0.1 ns for the binary KLC1:APP simulations. Entropy was calculated every 35 ns for the binary KLC1:APP simulations.[66] When entropy was considered in the overall binding free energy values, the relative binding affinities were scaled to experimentally realistic values. Normal mode entropy calculations have been found to improve the correlation and ranking of binding affinities as well.[67] Entropies were computed only for the KLC1:APP simulations. APP is roughly

four times the size of JIP1. Because of this difference in size, we expect entropy to play a lesser role in the KLC1:JIP1 complex. As a test of this hypothesis, we ran entropy calculations on one trajectory initiated from KLC1:JIP1 structure 4 and obtained an entropy ($T\Delta S$) value that was roughly half of the entropy for KLC1:APP. Additionally, the decreased flexibility of JIP1 should produce less variability in the computed entropy values between various simulations leading to the costly inclusion of entropy amounting to no more than a simple scaling factor. Structures and trajectories were visualized through VMD and PyMOL.[68,69]

Results and Discussion

The purpose of this study was to characterize the binding of APP and JIP1, separately, to the cargo-binding domain of KLC1, as well as the modes of interaction between JIP1 and APP. To investigate these potential interactions, a total of 3.85 μs of unrestrained MD was performed on various binary complexes of KLC1:APP, KLC1:JIP1, and APP:JIP1 (Table 1). To the best of our knowledge, experimental structures of these binary complexes are unavailable, so two docking methods as well as multiple randomized initial starting velocities were utilized to effectively sample the conformational space across all three binary complexes.[52-56] In this way, we are utilizing efficient docking approaches (GRAMM-X and Glide) to sample globally the surface of each binary complex while relying on MD to sample the local environment of each starting structure.

In lieu of experimental structures, we have evaluated the structural fidelity of our initial complexes by analyzing the total number of hydrogen-bonds and salt bridges as well as the total contact surface area and amount of buried hydrophobic surface area (Table S2). This allows a comparison between docked poses as well as an overall assessment of the reasonableness of our

initial structures. For KLC1:APP, our initial structures contain between 3 and 9 hydrogen bonds and 1 or 2 salt bridges. The contact surface area between KLC1 and APP ranges from 954 to 1580 $Å^2$. This represents 5–8% of the total surface area available to the complex and is the range of contact areas for proteins that are known to bind favorably.[70] We see a range of 0–8 intermolecular hydrogen bonds and no salt bridges between KLC1 and JIP1. The total contact area is relatively large with values ranging from 1481–2437 $Å^2$ or 9–14% of the total complex surface area. For our 39 structures of APP:JIP1, we observe a range of 1-11 hydrogen bonds and 1-2 salt bridges. The total contact area ranges between 393 and 790 $Å^2$ or 6–13% of the total.

KLC1:APP Binary Complex

Docking Followed by MD Simulations Failed to Produce a Stable APP:APP:KLC1 Complex

Experimentally it has been suggested that APP binds to KLC1 with a 2:1 stoichiometry. [33] To explore this further, GRAMM-X was used to generate 10 initial structures by docking a second molecule of APP into the lowest energy conformation of KLC1:APP (KLC1:APP Structure 4 at 205.1 ns, ΔG_{bind} = -108 kcal/mol). Each model was subjected to 25 ns of MD and in all cases at least one molecule of APP dissociated from the cargo-binding domain before the end of the 25 ns simulations.

MM-GBSA Binding Energies Suggest Favorable Interaction; However, Inclusion of Entropy Negates Such Effects

We performed 350 ns of MD simulation on four KLC1:APP binary poses obtained from GRAMM-X docking (Figure S4). The final conformation of each 350 ns simulation can be found in Figure S5. We utilized MM-GBSA (Molecular Mechanics-Generalize Born Surface Area)

calculations to estimate the binary binding affinities and obtained values that ranged from -22 to -59 kcal/mol. When entropic effects were included explicitly, the binding affinities are no longer negative in value suggesting a lack of favorable interaction between KLC1 and APP (Table 2). This lack of stability for the KLC1:APP binary complex is in agreement with a previous report by Chiba et al.[42] However, the retention of APP within the KLC1 cargo-binding domain throughout each simulation indicates favorable interactions that may suggest a kinetic barrier to dissociation. It should be noted that the MM-GBSA and MM-GBSA/entropy calculations are approximate. Neither computational approach provides absolute free energies but rather relative free energies that can be used to compare with similar calculations on other complexes.

Table 2. The MM-GBSA and MM-GBSA/entropy binding energy of KLC1:APP. MM-GBSA and MM-GBSA/entropy values were calculated for every 0.1 and 35 ns, respectively, for the full 350 ns trajectory of each simulation.

Simulation	Average ΔG MM-GBSA (kcal/mol)	Standard Error of the Mean MM-GBSA (kcal/mol)	Sampled TΔS (kcal/mol)	Adjusted ΔG (kcal/mol)
KLC1:APP Pose 1	-22.5	0.2	-67.3	+44.8
KLC1:APP Pose 2	-42.9	0.2	-52.4	+9.5
KLC1:APP Pose 3	-52.9	0.3	-77.7	+24.8
KLC1:APP Pose 4	-59.2	0.3	-59.3	+0.1

Figure 3 shows the average structure obtained from each 350 ns simulation on KLC1:APP. We began our simulations with a conformation of APP that was partially folded into an N-terminal α-helix with a relatively disordered C-terminal tail. We initiated all simulations with a KLC1 conformation organized into well-defined tetratricopeptide repeat units that formed a super-helical groove. These structural motifs persisted in APP and KLC1 throughout all of the simulations; however, the relative orientation of APP to KLC1 changed. These intermolecular variations play a large role in the overall energetics of each complex.

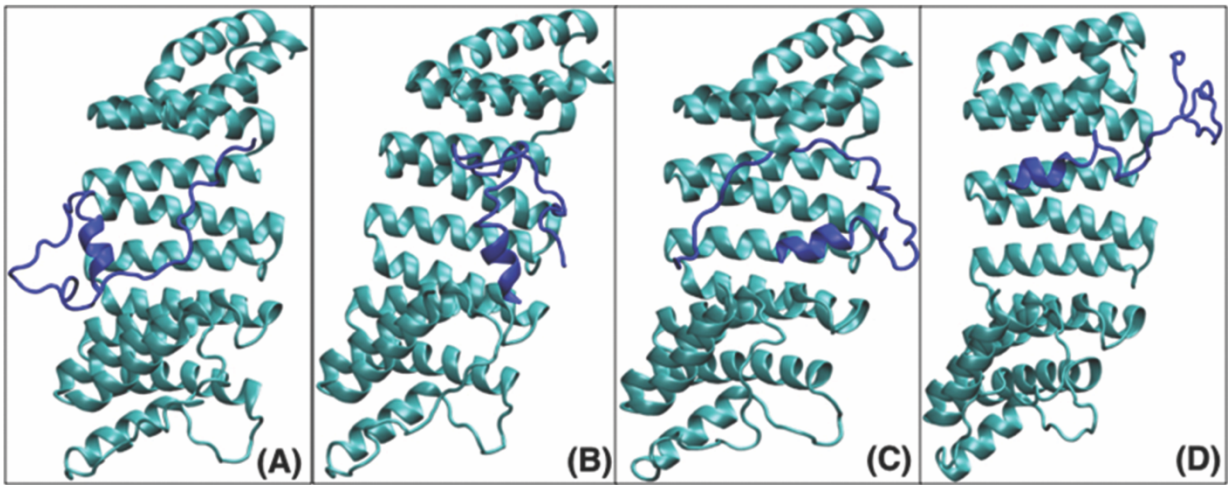

Figure 3. The average structure of KLC1:APP after 350 ns initiated from four different GRAMM-X poses. (a) KLC1:APP Pose 1 (b) KLC1:APP Pose 2 (c) KLC1:APP Pose 3 (d) KLC1:APP Pose 4. KLC1 is shown in teal ribbons and APP is displayed as blue ribbons.

As shown in Figure 3A, in the average structure of KLC1:APP Pose 1, the α-helix of APP is perpendicular to the central KLC1 TPR units. The APP C-terminal tail protrudes into solvent and then folds back into the cargo-binding domain. In the average structure for KLC1:APP Pose 2, the APP α-helix was also perpendicular to the central TPR units of the KLC1, however the C-terminal tail did not wrap back into the center area of the KLC1 cargo-binding domain but rather formed many points of interaction with Site 2. Atomistic explanations of these higher energy

complexes, in respect to non-entropic MM-GBSA average ΔG values, can be found in the Supporting Information.

Both KLC1:APP Pose 3 and KLC1:APP Pose 4 oriented the APP α-helix parallel to the central TPR units of the KLC1. The C-terminal tail of KLC1:APP Pose 3 folded back into the cargo-binding domain, while that of KLC1:APP Pose 4 mostly remained exposed to solvent. For KLC1:APP Pose 3, one of the few consistent interactions of the C-terminal tail was the oxygen atom on the backbone of Ala13 on the APP with Asn73 and Arg99 of KLC1. These interactions, occurring within Site 2 of the KLC1, began at 5 ns and were maintained for the rest of the simulation. It is important to note that the APP α-helix seemed to be drawn back inside of the cargo-binding domain of the KLC1 through hydrogen bonding interactions of the Lys196 on the KLC1 and the Asn42 on the APP. These key interactions helped stabilize the bulk of the APP within the cargo-binding domain, but the α-helix as well as the end of the C-terminal tail of the APP were highly dynamic relative to the other three simulations, which likely accounts for the high entropic contribution to its binding energy.

For KLC1:APP Pose 4, the very tight binding suggested by the MM-GBSA energies and relatively smaller TΔS estimates could be attributed to the APP α-helix moving quickly into Site 2 of KLC1. The Phe36 in the APP α-helix created a non-polar bonding network with Ile35, Ile74, and Leu77 around 15 ns, and these interactions persisted as the simulation progressed. This interaction was so strong due to the Leu35 on KLC1 folding over the APP's Phe36, in a quasi "knob-hole interaction" between the Phe36 and the other three specified residues on the KLC1 (Figure 4).[71] KLC1:APP 4 had the strongest relative MM-GBSA binding energy; however, the addition of entropy contributions seems to mostly negate such favorability. This might suggest that the binary KLC1:APP complex may not persist under standard experimental conditions. However,

because the MM-GBSA values are an estimation and provide only relative energetic comparisons, this conjecture will need to be validated with experimental data.

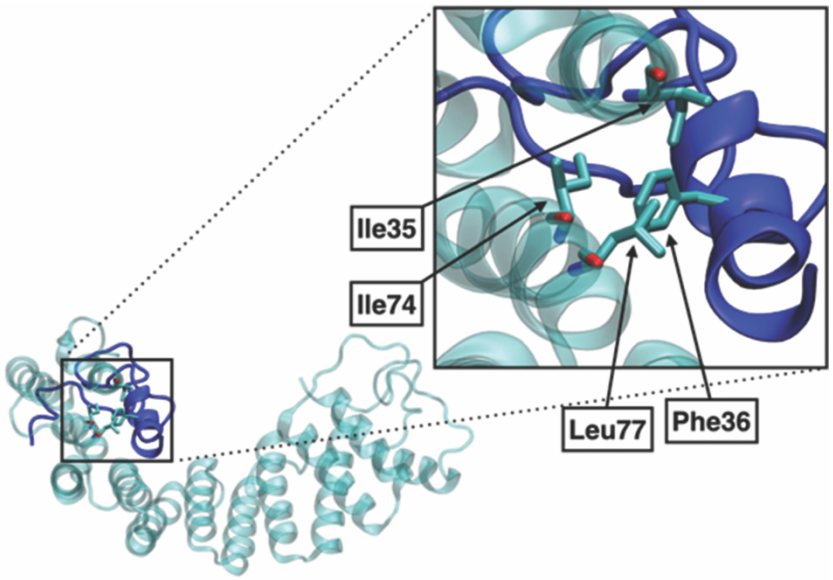

Figure 4. The final conformation of KLC1:APP Pose 4 simulation is shown here, highlighting the quasi knob-hole interaction of Phe36 on the APP with Ile35, Ile74, and Leu77 on the KLC1. Ile35 can be seen folding over top of Phe36, trapping the APP within Site 2, accounting for its stability.

KLC1:JIP1 Binary Complex

KLC Site 1 Exhibits Most Favorable MM-GBSA Binding Energies

MD simulations initiated from poses that oriented JIP1 in the KLC Site 1 binding pocket confirm the strength of Site 1 as predicted by Sitemap.[50,51] The average MM-GBSA binding energies for Site 1 (Table 3) are generally more favorable than those found for Site 2 and Site 3 (Tables 4, 6-7). Simulations begun from Site 1 Structure 1 and 3 showed JIP1 burying into the loop of Site 1, yielding a much more favorable MM-GBSA binding energy. However, simulations initiated from Site 1 Structure 2, produced trajectories in which JIP1 displayed highly dynamic behavior within the loop region as a whole. When the binding free energies were compared to the distance between the center of mass of JIP1 and that of Site 1 over each 75 ns trajectory, it was clear that the closer the center of masses were, the stronger the binding (Table 3 and Figure S6).

The simulation with the most favorable binding free energy, KLC1:JIP1 Site 1 Structure 1, had the shortest average distance between the center of masses (5.71 Å); while the simulation with the least favorable binding free energy, KLC1:JIP1 Site 1 Structure 2, had the longest average distance (13.05 Å). The starting and final conformations of each KLC1:JIP1 Site 1 structure can be found in Figures S7-S8. Further characterization of the binding affinity of this loop region as well as its effect on the binding of APP to KLC1's cargo-binding domain has been reserved for a future study due to the well-known conformational flexibility and disorder typically associated with loop domains.

Table 3. The average MM-GBSA binding energies for each KLC1:JIP1 Site 1 model. All simulations were run for 75 ns. Structures 1 and 3 have relatively strong binding affinities as a result of the short distance between the center of mass of the JIP1 and that of Site 1, while Structure 2's energy was diminished by its highly dynamic behavior.

Model	Avg. ΔG_{bind} (kcal/mol)	Average KLC1:JIP center of mass distance (Å)
KLC1:JIP1 Site 1 Structure 1	-62.9	5.71
KLC1:JIP1 Site 1 Structure 2	-21.3	13.05
KLC1:JIP1 Site 1 Structure 3	-38.3	6.49

Key Interactions in KLC Site 2 Result in Favorable Binding Energies

The average binding energies for the four most favorable KLC1:JIP1 Site 2 structures (KLC1:JIP1 Site 2 Structures 2, 4, 6 and 8) are shown in Table 4. After the completion of an initial 25 ns of unrestrained MD on the ten KLC1:JIP1 Site 2 Glide-generated poses, Structures 2, 4, 6, and 8 showed the strongest average binding affinities and decreasing binding energies with respect to time (negative slopes; Figure S9). Results from these four simulations were chosen for further molecular dynamics simulation because of the negative binding affinities and the dynamical behavior of the resulting trajectories gave evidence of successfully sampling the potential energy

surface (Figure S10). After an additional 50 ns of simulation time, the simulation initiated from Structure 6 was discontinued due to a loss of favorable binding interactions (Figure S9C).

Table 4. The RMSD between KLC1:JIP1 Site 2 Structures 2, 4, 6, and 8 were calculated against a reference structure (Structure 4). The average ΔG_{bind} was determined for the KLC1:JIP1 complex over the entirety of each simulation.

Complex	Time (ns)	RMSD against Structure 4	Average ΔG_{bind} (kcal/mol)	Final Distance between Center of Masses of JIP1 and Site 2 (Å)
KLC1:JIP1 Site 2 Structure 2	125	1.184 Å	-15.1	21.4
KLC1:JIP1 Site 2 Structure 4	125	0.000 Å	-35.3	17.0
KLC1:JIP1 Site 2 Structure 6	50	1.304 Å	-13.9	16.5
KLC1:JIP1 Site 2 Structure 8	125	2.184 Å	-32.3	16.7

From these results, it was determined that JIP1 favored a perpendicular orientation relative to KLC1 TPRs 1 and 2, as the four most stable simulations displayed a slight shift in JIP1 as it moved from its starting conformation to lie closely inside binding Site 2 of KLC1 (Figure 5). Several intermolecular interactions appear to drive this structural shift. It is also important to note that all four models show the JIP1 located on the "topside" of KLC1, which is where cargo is known to bind to KLC1.[22-24]

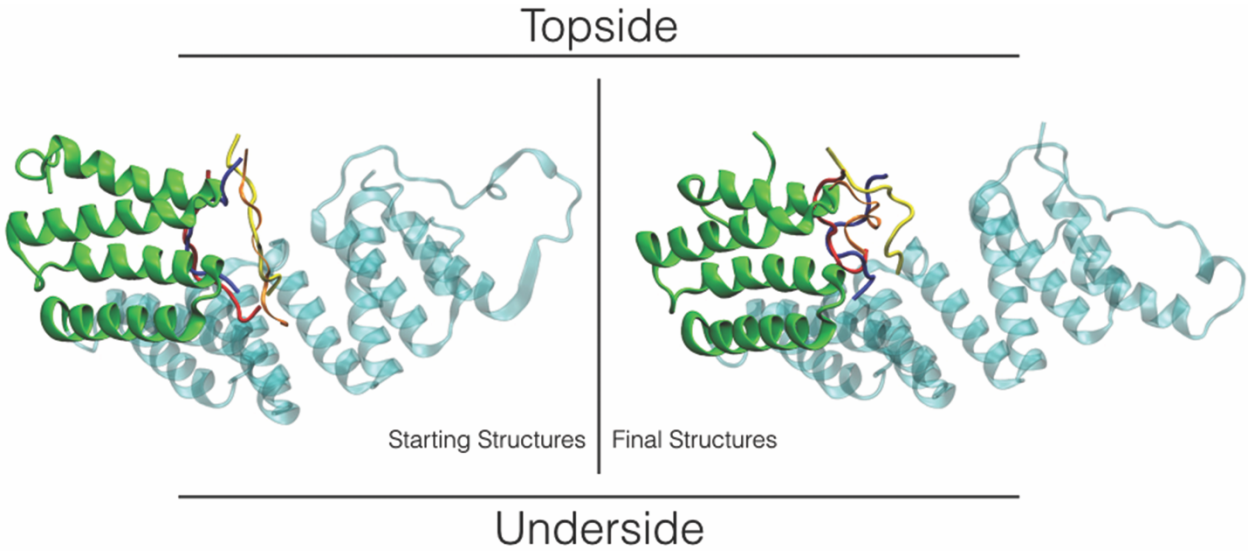

Figure 5. The starting and final conformations of Site 2 KLC:JIP1 Structures 2, 4, 6, and 8 (for clarity, only the KLC1 from Structure 4 is shown). TPRs 1 and 2 defining Site 2 are highlighted in green. Yellow = Structure 2, Orange = Structure 4, Red = Structure 6, Blue = Structure 8. The topside and underside are defined here because the binding sites of KLC1 are solely on the topside of the cargo-binding domain, and exclude the underside of those same residues.

Hydrogen bond networks were very important in the stabilization of KLC1:JIP1 Site 2 Structures 4 and 8. Generally, the fluctuations in binding affinity were inversely proportional to the amount of hydrogen bonding between the KLC1 and JIP1 (Figure S11). For Structure 4, the complex experienced a large drop in binding energy (~ -10 to -40 kcal/mol) when the side chain nitrogen atoms of JIP1 Arg556 formed a hydrogen bonding interaction with the side chain oxygen atoms of KLC1 Glu44 (ΔG_{decomp} = -16.2 kcal/mol) and the backbone oxygen atom of Gly41 (ΔG_{decomp} = -5.8 kcal/mol) (Figure 6). (All pairwise per residue free energy decomposition values discussed here are available in the supplemental information - Figure S12). Interactions between Arg556 on JIP1 and glutamates on KLC1 are seen to be favored throughout the other simulations as well, which corresponds to experimental results.[38] Arg556 on the JIP-1 was found to experimentally favor hydrogen bonding with a glutamate in an inhibited kinase (JNK-1), which was essential for the overall binding of the two proteins. The other important stabilizing

interactions with KLC occurred at the C-terminal of the JIP1 peptide. The backbone oxygen of Leu562 created a strong hydrogen bonding interaction with KLC1 Arg80 (ΔG_{decomp} = -3.2 kcal/mol), pulling the JIP1 phenylalanine terminus into Site 2. This interaction then allowed for the terminal phenylalanine backbone (Phe563) to alternate hydrogen bonds with KLC1 Lys123 (ΔG_{decomp} = -4.1 kcal/mol) and Arg80 (ΔG_{decomp} = -7.9 kcal/mol). Although Lys123 is not within Site 2 itself, it helped to stabilize the phenylalanine tail of JIP1, stabilizing the rest of the peptide within Site 2. The hydrogen bonding interactions described above for Structure 4 were very similar to those observed in the trajectory initiated from Structure 8, but each was slightly weaker in Structure 8, resulting in the overall ΔG_{bind} to be lower compared to the Structure 4 trajectory (Table 4). A description and schematic of the dominant interactions of the trajectory initiated from Structure 8 can be found in the Supplemental Information (Figure S13).

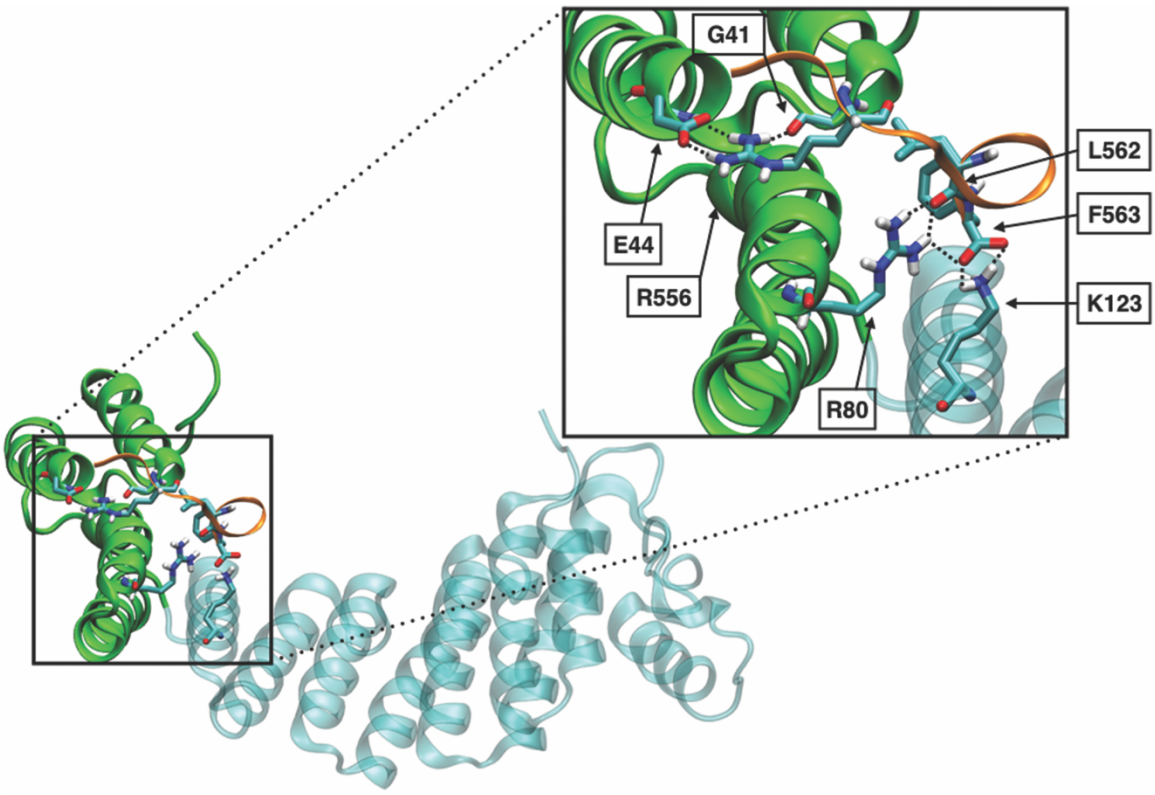

Figure 6. The dominant hydrogen bond interactions in KLC1:JIP1 Site 2 Structure 4 as determined by pairwise decomposition energies. The conformation imaged here is at 93.2 ns due to its having

the strongest binding affinity (-56.8 kcal/mol). These binding motifs were found in 5-40% of the simulation. Hydrogen bonds are represented by dotted lines between the hydrogen bond pairs. Site 2, composed of TPRs 1 and 2, is highlighted in green.

An analysis of the trajectories indicates that the conformational space of Site 2 was well sampled. This is best illustrated by RMSD values relative to the average structure from KLC1:JIP1 Site 2 Structure 4 (Table 4) and center of mass distances between the JIP1 and Site 2 (Table 4 and Figure 7). Structure 4 was chosen as the reference structure as it has the lowest average binding energy. All other pairwise RMSD values of the four averaged structures can be found in the Supplemental Information (Table S1). From this analysis, it can be seen that the JIP1 peptide sampled various locations both inside and outside of Site 2. While the amount of movement varied, it is clear that the final JIP1 conformations from each trajectory were roughly the same distance away from the center of mass of Site 2 (Table 4). Each trajectory sampled the entire site from TPR 1 to TPR 4, before stabilizing along TPR 1 and 2 through the interactions described previously. The respective average ΔG_{bind} values from each of the 4 simulations reveals that the binding strength of KLC Site 2 ranges from -15 to -35 kcal/mol (Table 4).

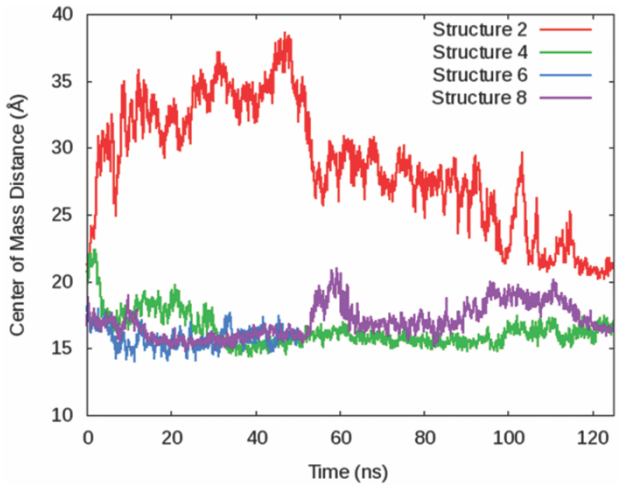

Figure 7. The distance between the center of masses of JIP1 and Site 2 for each KLC1:JIP1 Site 2 Structures 2, 4, 6, and 8 calculated against time. Each structure's final conformation was in close proximity with respect to the distance between the center of mass of JIP1 and that of Site 2.

To estimate the significance of entropy in the KLC1:JIP1 binary complex, a sample entropy computation was performed on Structure 4 using normal mode analysis. For this system, the entropy term (TΔS) was determined to be approximately -36 kcal/mol; this is half of the entropy value determined for the KLC1:APP complex. If we treat this KLC:JIP1 entropy value as a general correction [67] to the MM-GBSA energy, than 2 of the 4 trajectories described in Table 4 yield unfavorable ΔG_{bind} values while the other 2 trajectories result in energetically neutral ΔG_{bind}. MM-GBSA and entropy-corrected MM-GBSA values are estimations and provide information that is useful on a relative scale. If we compare the average entropy-adjusted ΔG_{bind} of KLC1:APP and KLC1:JIP1 binary complexes we obtain values of +19.8 and +11.9 kcal/mol, respectively. These results suggest a preferential binding of KLC1 to JIP1, relative to binding with APP.

KLC1 Polar Patch is not the Most Favorable Binding Site

A study by Zhu et al. suggested the binding of JIP1 to KLC1 occurred in the "polar patch" region of KLC. This region corresponds to Site 3 in our SITEMAP results (Figure 2) and is comprised of highly polar residues within the cargo-binding domain.[40] Previously, Zhu et al. suggested that the KLC-TPR4 point mutation N157S reduced significantly the binding affinity of JIP1 to KLC1. These experimental results indicated that the polar patch region was an important component of the binding motif between JIP1 and KLC1.[50,51] We ran MD initiated from JIP1 structures docked into both WT and N157S mutant KLC1 Site 3 in order to provide an atomistic understanding of the KLC1:JIP1 binding. Our results indicate that while the "polar patch" of KLC1 showed potential for binding JIP1, it was not as favorable as Site 2. For the WT models, we ran seven simulations for 25 ns each. In three of the seven trajectories JIP1 remained within the polar patch region throughout the simulation. However, in four of the seven trajectories JIP1 exited the polar patch. The average binding energies from the three simulations that remained in the polar

patch region (Table 5) suggest that WT KLC1 displays stronger average binding than the mutated KLC1, agreeing with the results obtained by Zhu et al.[40]

Table 5. The average binding energy of binary KLC1:JIP1 and mutated KLC1:JIP1 simulations in which JIP1 remained in, as well as exited, the polar patch region. This is the average ΔG_{bind} of all simulations of the same complex combined.

Complex	Average Binding Energy (kcal/mol)	Number of Simulations
KLC1:JIP1 Site 3 (WT) Remained in Polar Patch	-19.5903	3
KLC1:JIP1 Site 3 (WT) Exited Polar Patch	-12.0033	4
KLC1:JIP1 Site 3 (mutant) Remained in Polar Patch	-14.5491	5
KLC1:JIP1 Site 3 (mutant) Exited Polar Patch	-14.0993	9

Our results are in agreement with Zhu et al. that the polar patch is a favorable binding site as there is a nearly 8 kcal/mol difference in binding energy between the WT simulations that remained in the polar patch and those that dissociated to the area around it (Table 5).[40] However, the inconsistency of JIP1 to remain in the polar patch may indicate that this site is not as important as previously thought. This idea needs further study using significantly more starting structures initiated from within the polar patch and with longer MD trajectories obtained with methods that are more able to sample conformational interconversions (REMD, simulated annealing, etc.) The polar patch is clearly a valid option for the binding of JIP1 to KLC1, but in this study it is not found to be the strongest candidate.

Four of the ten trajectories initiated from poses docked in Site 2 did sample Site 3; however, in each case the JIP1 residence time in Site 3 was relatively short ranging from 27 – 57% of the overall simulation time. (Table 6 and Figure S15). The average binding free energy for structures within and outside of the polar patch was determined using a 20 Å cutoff distance between the center of mass of the polar patch and JIP1. Of the four simulations, two produced results suggesting that JIP1 binding within and outside of the polar patch is energy equivalent. These were initiated from KLC1:JIP1 Site 2 Structure 3 and Structure 6. Our longest simulation, initiated from Structure 2, suggests that JIP1 binds ~7 kcal/mol more favorably within the polar patch. Interestingly, during the 46.4 ns that JIP1 was within 20 Å of the polar patch it curved around the edge of the KLC1 cargo-binding domain and interacted with the underside of the polar patch.[68] Zhu et al.'s results suggested that only the topside of the polar patch was able to interact with JIP1.[40] To better understand the relative lack of JIP1 affinity for the polar patch region, further discussion on the trajectories started from Structures 3 and 5 is provided in the Supporting Information. The starting and final structures of each KLC1:JIP1 Site 3 pose that remained within the polar patch can be found in Figures S16-S17.

Table 6. Comparison of average binding free energy for when the center of mass of JIP1 was inclusively within and outside of 20 Å of the center of mass of the polar patch region of KLC1. Entropy was not considered in these calculations.

Simulation	Total Time (ns)	Time < 20 Å (ns)	Avg. ΔG < 20 Å (kcal/mol)	Time 20+ Å (ns)	Avg. ΔG 20+ Å (kcal/mol)
KLC1:JIP1 Site 2 Structure 2	125	46.4	-19.9*	78.6	-12.3
KLC1:JIP1 Site 2 Structure 3	25	6.7	-16.5	18.3	-18.0

KLC1:JIP1 Site 2 Structure 5	25	14.2	-9.8	10.8	-18.9
KLC1:JIP1 Site 2 Structure 6	50	28.3	-14.6	21.7	-13.0

* Binding free energy is not characteristic of being solely within the polar patch region. See above text for further explanation.

On average, JIP1 binding to Site 2 is approximately 4.6 kcal/mol more favorable than to Site 3. (average ΔG_{bind} from Table 4 -24.2 kcal/mol; average ΔG_{bind} from Table 5 -19.6 kcal/mol). Our results are in agreement with Zhu et al. in that we find the polar patch to be a relatively favorable binding pocket. We also find KLC1:JIP1 binding to be less favorable when Asn157 is mutated to a serine. However, our MM-GBSA estimation of binding free energies suggest Site 2 is a more favorable binding pocket within the cargo-binding domain of KLC1. Though favorable, the polar patch region may be too small to maintain JIP1 binding.

APP:JIP1 Complex

Two Conformations Dominate over 39 Separate 25 ns Simulations (975 ns total).

Docking methods were used to generate 39 unique APP:JIP1 starting structures (20 from GRAMM-X and 19 from Glide) and these were subjected to 25 ns of unrestrained MD.[52-56] It was important to sample as many binary APP:JIP1 conformations as possible because the binding of APP to KLC1 is thought to be facilitated by JIP1 and there is less experimental information available for either APP or JIP1 relative to KLC1.[34,39,41,42] In all 39 simulations, JIP1 and APP persisted in maintaining a bound binary complex. The starting and final structures of the four lowest energy APP:JIP1 simulations can be found in Figures S18-S19. Cluster analysis (using *cpptraj*) over 975 ns of simulation data revealed three dominant conformations (Figure 8).[61] We used the hierarchical agglomerative clustering method available in the *cpptraj* module of the

AMBER software suite. We chose this method as it has been shown to be particularly useful when the cluster count is not known in advance.[72] Molecular structures were grouped into clusters based on an atomic coordinate RMSD distance metric. Structures within clusters had smaller atomic RMSD values relative to RMSD differences between clusters. *Cpptraj* was a very useful tool for this analysis as it provided frame by frame cluster summaries as well as a representative conformation for each cluster found.

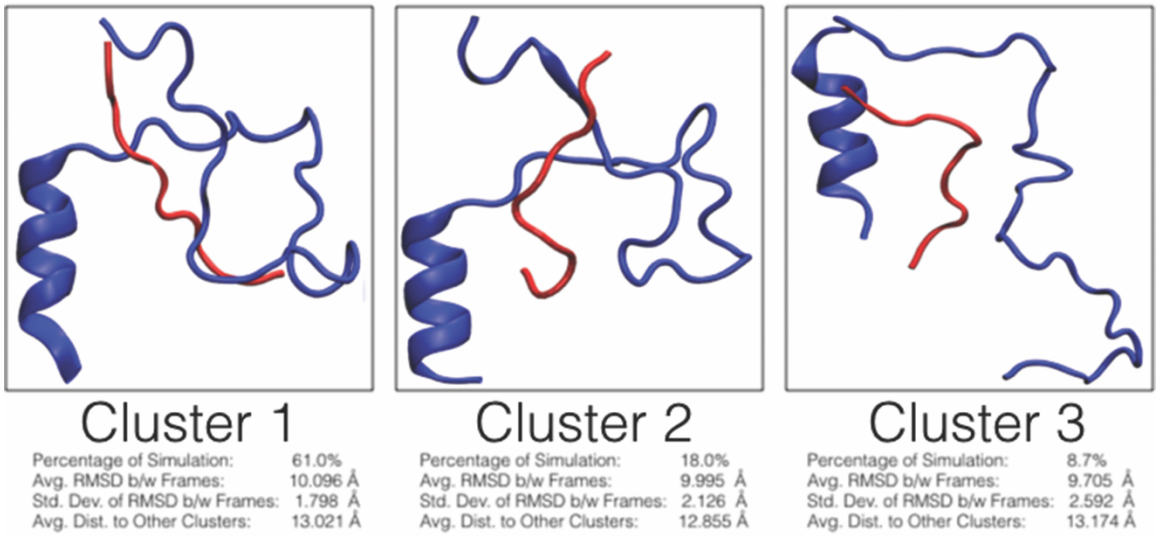

Figure 8. Cluster analysis of all 39 binary APP:JIP1 simulations performed using *cpptraj* from the AmberTools 14 package.

The majority of the 39 simulations showed little to no interconversion between clusters, suggesting simulation convergence for each trajectory (Figures S20-S21). We were not seeking to sample conformational interconversions; only to fully minimize and sample well the local environment of each starting structure. In this way, we were relying on the docking approach to sample globally and the MD simulation to sample locally. Only one simulation (GRAMM-X Pose 20) transitioned between four clusters. Two simulations (Glide Pose 9 and GRAMM-X Pose 13) converted between three clusters, while the remaining 36 simulations either remained within a single cluster for the entirety of the simulation or made a single cluster change. This suggests that

the majority of the trajectories on which our MM-GBSA data depends were converged with respect to time. The most common conversion was from Cluster 1 to Cluster 3, but this transition only occurred in 4 simulations indicating no particular favorability in cluster interconversion.

As seen in Figure 8, the APP alpha-helix comprising residues 31-42 is present in all clusters. JIP1 and the C-terminal tail of JIP1 remain relatively disordered. Cluster 1 occurred 61.0% of the simulation time. This cluster contained structures in which JIP1 is "threaded" through the APP, between the alpha helix and the C-terminal tail. By threading through APP, JIP1 does not interact with solvent and remains centrally located within the space between the alpha helix and C-terminal tail of APP. It is not clear whether this binding motif adopted by the functional 10-residue fragment would be possible for the full 707-residue JIP1 protein.[45] The structures in Cluster 1 are also oriented such that APP appears to block any subsequent binding of JIP1 to other macromolecules, such as KLC1.

Cluster 2 occurs 18.0 % of the simulation time. The structures in this cluster do not display the threading or blocking behavior of the structures from Cluster 1. Figure 8 shows the JIP1 peptide between the alpha helix and the C-terminal tail of the APP. Instead of tucking back through this gap, like in Cluster 1, the JIP1 is solvent exposed. This positioning exposes the JIP1 in a manner that would allow for another potential binding partner. Cluster 3 occurs 8.1% of the simulation time. The JIP1 within the Cluster 3 conformations is located once again between the alpha helix and C-terminal tail of the APP. The C-terminal tail of APP in Cluster 3 is more extended than in Clusters 1 and 2. This extension could allow for the APP to bind over a larger surface. This orientiation also allows for the JIP1 to potentially be exposed to a second binding partner, similar to the JIP1 within the Cluster 2 conformations. Though the conformations within Cluster 1 would

allow for a tighter binding between APP and JIP1, the orientation of the complexes within Clusters 2 and 3 would provide an opening to which KLC1 could bind.

Potential Ternary KLC1:JIP1:APP Complex

Preliminary testing of a ternary KLC1:JIP1:APP complex was performed by docking APP to the binary KLC1:JIP1 lowest Site 2 binding free energy conformation (ΔGbind = -60.9 kcal/mol) using GRAMM-X to yield three starting conformations.[55,56] We used GRAMM-X as Glide was not able to accommodate the total number of atoms in the ternary structure. Three different initial structures were simulated in duplicate for one μs under the same conditions as each of the binary complexes described above. MM-GBSA binding computations were performed every 0.8 ns for each simulation (Table 7). These preliminary results on the ternary complex yielded relatively stronger binding affinities than the binary KLC1:APP complex (Table 2), suggesting that JIP1 could play a role in the energetic stabilization of KLC1:APP binding. However, further conformational sampling is required before this model can be viewed with confidence. Further studies will include the docking of lower energy APP:JIP1 binary complexes from each of the conformational clusters outlined within this work to the KLC1 crystal structure.

Table 7. The MM-GBSA binding energy of the ternary KLC1:JIP1-APP complexes. MM-GBSA 0.8 ns for the full 1000 ns trajectory of each simulation.

Simulation	Time (ns)	Average ΔG (kcal/mol)	Std Error of the Mean (kcal/mol)
KLC1:JIP1:APP 1 Seed 1	1000	-66.8	0.5
KLC1:JIP1:APP 2 Seed 1	1000	-115.8	0.9
KLC1:JIP1:APP 3 Seed 1	1000	-69.0	0.6
KLC1:JIP1:APP 1 Seed 2	1000	-76.7	0.4

KLC1:JIP1:APP 2 Seed 2	1000	-82.4	0.5
KLC1:JIP1:APP 3 Seed 2	1000	-87.0	0.4

Conclusion

Understanding the atomistic molecular dynamics of the APP, KLC1 and JIP1 proteins will lead to a better fundamental understanding of their intermolecular interactions. Such studies could have significance for dementia-related diseases by determining likely conformations that may lead to the progression of amyloid aggregation, as well as finding plausible mutation sites for the disruption of these processes. Binary complexes of KLC1, JIP1, and APP persist over long time scale molecular dynamics investigation and in many cases produce favorable MM-GBSA binding free energies. Site 2 on KLC1 binds most favorably to JIP1, while the polar patch region of Site 3 bound JIP1 somewhat less favorably (Figure 2). Preliminary results suggest that KLC Site 1 is also a favorable binding site but further characterization of this highly dynamic loop is needed. The MM-GBSA binding analysis of KLC1:APP complexes suggest favorable interactions, however when normal mode entropic effects are included the free energies are less favorable. (Table 2). These results suggest possible modes of binding between binary KLC1:APP and KLC1:JIP1 complexes but are not able to rule out the possibility that APP does not bind either directly or indirectly to KLC1 due to blocking of APP by other biologically relevant proteins. This is Cluster 1 of the APP:JIP1 complex occurred most often, but Clusters 2 and 3 could be a more practical conformation when discussing the linking of KLC1 to APP through the mediation of JIP1. The orientation of JIP1 relative to APP in Clusters 2 and 3 create an opening to which KLC1 could reasonably bind, while avoiding threading and blocking behavior. Further studies are needed to examine the difference between the binary KLC1:APP complex and a ternary KLC1:JIP1:APP

complex, as well as the effect the conformational difference of Cluster 1 and Clusters 2 and 3 has on the binding of the ternary complex.

A Molecular Dynamics Study of the Ternary Complexes of APP, JIP1, and the Cargo Binding Domain of KLC

Cooper A. Taylor[1], Bill R. Miller III[2], and Carol A. Parish[1]

[1]*Department of Chemistry, Gottwald Center for the Sciences, University of Richmond, 28 Westhampton Way, Richmond, VA 23173*

[2]*Department of Chemistry, Truman State University, 100 E. Normal Ave, Kirksville, MO 63501*

The misregulation of the cargo binding domain of kinesin-1 light chain (KLC1) has been linked to various neurodegenerative diseases, most notably Alzheimer's, as KLC1 may play a role in amyloid precursor protein (APP) transport and subsequent degradation. Recent literature suggests a third protein, C-Jun-amino-terminal kinase-interacting protein 1 (JIP1), acts as a mediator protein for binding of KLC to APP. Computational molecular dynamics (MD) was conducted on this ternary complex using the Amber MD package to understand the atomistic interactions between the structures. The online GRAMM-X rigid docking server was used to dock binary APP:JIP1 complexes to KLC1 as well as APP to the binary KLC1:JIP1 complex. All binary complexes were derived from previous studies. The simulations were used to examine low energy conformations of the bound complex, since this conformation has not been resolved experimentally. Binding free energy calculations suggest the ternary complex is more stable than the binary complex, and hydrogen bond networks between the three proteins prove essential to successful binding motifs.

Introduction

Alzheimer's disease (AD) is the leading cause of dementia in the elderly. Early onset familial AD has been shown to be caused by mutations in three different gene products, with Amyloid Precursor Protein (APP) being the focus of many studies.[7-11] This focus is due to the fact that proteolysis releases amyloid-β peptides.[12] The accumulation of these peptides along with plaque formation is considered the driving force of AD.[13-16]

It has been shown that Alzheimer's-related proteins such as APP undergo fast, plus-end directed (anterograde) axonal transport as they reside within the axon.[26-28] The light chains of kinesin-1 (KLC1) have been shown to be a potential mode of transport for APP.[27,29-32] This

transportation could also require the facilitation of a third protein, c-Jun N-terminal kinase (JNK)-interacting protein 1 (JIP1).[34] The 707 residue JIP1 protein is localized in the neuronal region of the brain, suggesting a critical role in cell communication, a role which could be extrapolated to being involved in the transport of APP within the brain.[35]

In a recent study, we had provided preliminary results of the ternary KLC1:JIP1:APP complex.[73] These preliminary results on the ternary complex yielded relatively stronger binding affinities than the binary KLC1:APP complex, suggesting that JIP1 could play a role in the energetic stabilization of KLC1:APP binding. However, these results could not be viewed with confidence without further sampling of the potential energy surface for this full ternary complex as well as a more holistic understanding of the witnessed energies. Thus, further studies including the docking of low energy APP:JIP1 binary complexes from the defined conformational clusters found in the previous binary studies to the KLC1 crystal structure have been performed, and will be discussed within this work.

In this study, we aimed to evaluate the potential for the formation of ternary complexes between KLC1, APP and JIP1. Towards that end, we 1) evaluate the effect of docking order of the three proteins, 2) determine the influence of JIP1 orientation in respect to KLC1 and APP, and 3) assess the various binding motifs witnessed amongst the ternary complexes. We believe that a more foundational insight of the dynamics of these ternary protein-complexes will elucidate the roles KLC1 and JIP1 play in APP processing for Alzheimer's disease and other dementia-related diseases.

Methods

Structure Retrieval

Structures of KLC1 and JIP1 were obtained from the Protein Databank.[45] The crystalized KLC1 (PDB 3NF1) consisted of 304 amino acids, but the final 11 N-terminal amino acids were removed so the KLC1 consisted of only six α-helices.[46] This was done to more accurately resemble the *in vivo* structure of hKLC1 since these 11 residues were an artificially engineered histidine tail. The JIP1 structure was derived from PDB 2H96, which contained a dimer of JIP1 along with mitogen-activated protein kinase 8.[47] The 10-residue monomer was procured through the removal of the kinase and one of the two JIP1 structures (chain G). The C-terminal 42 residues of APP solution-phase NMR structure was obtained through personal communications with Dr. Charles Sanders at Vanderbilt University.[48] The complexes of KLC1:JIP1 and APP:JIP1 discussed within were obtained from our previous studies of the binary complexes of KLC1:APP, KLC1:JIP1, and APP:JIP1.[73]

Docking of KLC1:JIP1:APP and KLC1:JIP1:APP

Previous work has shown three major binding sites within the KLC1 protein.[73] After extensive sampling through the unrestrained molecular dynamics (MD) of several docked positions of JIP1 within each binding site, it was found that Site 2, which exists perpendicularly across TPRs 1-2 of the KLC1 cargo-binding domain, yielded the most consistently strong binding of JIP1 in a physiologically relevant manner. Each binary KLC1:JIP1 Site 2 simulation that underwent extended unrestrained MD in our previous study was analyzed and the coordinates of the conformation corresponding to the lowest binding free energy (ΔG_{bind} = -60.9 kcal/mol) were used for the ternary docking of KLC1:JIP1:APP, Set 1 of the ternary complexes. The c-terminal 42 amino acid APP peptide was docked to this binary (KLC1:JIP1) structure using Vakser Lab's GRAMM-X Protein-Protein Docking Web Server, producing 10 ternary models of varying starting

conformations.[55,56] Glide was not able to dock APP to the binary KLC1:JIP1 complex due to its large atom count, prompting the use of GRAMM-X.[53,57,74]

An additional type of ternary complexes, KLC1:JIP1:APP, was generated by docking two different binary APP:JIP1 complexes into KLC1 with GRAMM-X, creating Sets 2 and 3 of the ternary systems (Figure 9).[55,56] Set 2 of the ternary complexes utilized the lowest binding free of all binary APP:JIP1 complexes previously analyzed.[73] The coordinates were taken from APP:JIP1 Glide Pose 8 (ΔG_{bind} = -97.4 kcal/mol). The JIP1 was located mostly within the APP, thus prompting the use of a second APP:JIP1 binary complex, the coordinates of which were taken from APP:JIP1 GRAMM-X Pose 4 (ΔG_{bind} = -59.9 kcal/mol), in which the JIP1 was externally positioned in relation to the APP. GRAMM-X generated 10 unique models for each of Sets 2 and 3 of the ternary KLC1:JIP1:APP complexes.[55,56]

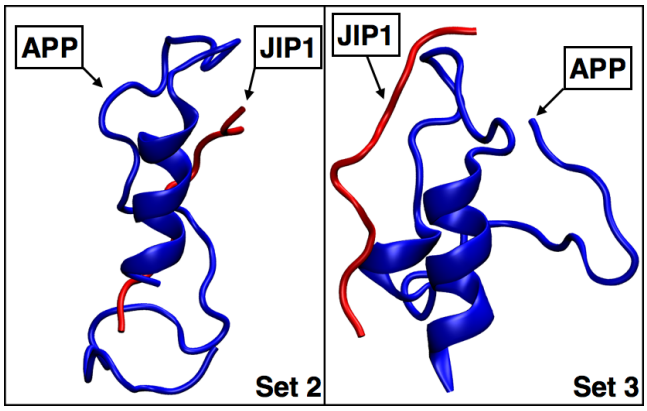

Figure 9. APP:JIP1 binary complexes used to generate Sets 2 and 3 of the KLC1:JIP1:APP ternary complexes.

Molecular Dynamics of Ternary Complexes KLC1:JIP1:APP and KLC1:JIP1:APP

All ternary simulations were performed under identical constraints. The *ff14SB* force field was applied to all simulations.[75] Each system was neutralized with Cl^- ions and solvated within a truncated octahedron periodic box of TIP3P water molecules by AmberTools' *tleap*.[60,61] At least 12.0 Å of solvent between the solute and edge of the unit cell was initially required. All steps of

MD were performed using the GPU-accelerated *pmemd* code of Amber 14.[61,76] Each system was minimized, heated, equilibrated, and then run under unrestrained MD. The minimization process consisted of seven stages, each made up of 5,000 steps. The first 1,000 steps were of steepest descent minimization, and the remaining 4,000 steps were of conjugate gradient minimization. The first of the seven stages was given a restraining force of 10.0 kcal/mol/Å^2 on the heavy atoms of the solute and was methodically lowered to 0.0 kcal/mol/Å^2 by stage seven. After minimization, each structure was heated from 10 K to 300 K with a restraining force of 10.0 kcal/mol/Å^2 on the solute. Equilibration consisted of lowering the restraining force every 500 ps from 10.0 kcal/mol/Å^2 to 0.0 kcal/mol/Å^2 over seven stages. After the kinetic energy equilibrated between the solvent and structure with no restraining force, unrestrained MD at constant pressure (1 atm) and temperature (300 K) was conducted on the system. Unrestrained MD was performed in 25 ns intervals with the complexes' coordinates, trajectories, and energies saved every 100 ps. The SHAKE algorithm was used to restrain all covalent bonds to hydrogen atoms.[63] This improved the computational efficiency of the simulation and allowed the use of a 2-fs time-step in the molecular dynamics simulation.

After 25 ns of unrestrained MD was conducted on the ten Set 1 KLC1:JIP1:APP ternary complex models, only two models were kept for further examination. The other eight models were ruled invalid due to unrealistic or implausible interactions between atoms and amino acids of the various proteins. Of these eight invalid simulations, five had the APP tangled in the edges of KLC1, two had APP leave the cargo-binding domain of KLC1 altogether, and one had the JIP1 slide out of the cargo-binding domain of KLC1. After 75 more ns of unrestrained MD was carried out on the two valid structures, the four frames with the lowest free binding energy from across both ternary simulations were determined and their coordinates were saved as PDB files (ΔG_{bind} =

-96.4 kcal/mol, ΔG_{bind} = -95.5 kcal/mol, ΔG_{bind} = -94.5 kcal/mol, and ΔG_{bind} = -93.9 kcal/mol). Each of these four conformations were used as initial coordinates for four new simulations run for 100 ns each before the conformation with the third most negative initial conformation (ΔG_{bind} = -94.5 kcal/mol) was discontinued due to separation between the APP and the binary structure of JIP1 and KLC1. The remaining three simulations were run until they each totaled one μs of elapsed simulation time. Duplicates of each structure were also run with different starting velocity vectors for a μs, each, to gain a wider variety of sampling. These simulations will be referred to as Set 1 of the ternary KLC:JIP1-APP simulations, with the syntax "Set 1 KLC1:JIP1:APP Simulation # Seed #" (i.e. Set 1 KLC1:JIP1:APP 1 Seed 1).

All ten poses of Set 2 of the ternary complexes, in which the lowest binding energy coordinates of the binary APP:JIP1 models was docked to KLC1, were subjected to an initial 25 ns of simulation. After 25 ns, four simulations (models 6-9) were discontinued either due to the binary complex disconnecting completely from the KLC1 cargo-binding domain or due to invalid interactions like those found in the discontinued simulations of Set 1, such as the entanglement of the APP within the extended loop region of the KLC1. The remaining 6 simulations were extended for an additional 50 ns, totaling 75 ns for each. These simulations will be referred to as Set 2 of the ternary KLC1:JIP1:APP simulations, with the syntax "Set 2 KLC1:JIP1:APP Simulation #" (i.e. Set 2 KLC1:JIP1:APP 1).

Set 3 of the ternary complexes consisted of the 10 docked poses of the low binding energy coordinates of an APP:JIP1 binary model in which the JIP1 was located externally with respect to the APP protein. All 10 models were subjected to 100 ns of unrestrained MD, and no models experienced any disconnect between the three proteins. These simulations will be referred to as Set 3 of the ternary KLC1:JIP1:APP simulations, with the syntax akin to that of Set 2.

MD Analysis

The calculation of RMSDs, distance measurements, RMSFs, and average structures was performed using *cpptraj* commands included in AmberTools 14.[61] Binding free energies, per residue free energy decomposition values, alanine scanning, and entropies were calculated using MMPBSA.py.[5] Binding free energies, per residue free energy decomposition values, and alanine scanning were calculated every 0.8 ns for the Set 1 KLC1:JIP1:APP simulations, while binding free energies were calculated every 0.1 ns for Sets 2 and 3 of the KLC1:JIP1:APP simulations. The computation of the binding free energies were performed two ways, one of which considered the KLC1:JIP1 binary complex of each ternary model to be the "receptor" with the APP being the "ligand" and the other in which just the KLC1 was considered the "receptor" while the binary complex of APP:JIP1 was considered the "ligand". Normal mode analysis was used when calculating entropy values. Entropy was calculated every 50 ns for the Set 1 KLC1:JIP1:APP simulations. Entropies were excluded from all binding free energy calculations, unless otherwise stated. When entropy was considered in the overall binding free energy values, the relative binding affinity between each simulation was not significantly affected, but was simply scaled to experimentally realistic values. Structures and trajectories were visualized through VMD and PyMOL.[69,77]

Results and Discussion

Set 1 KLC1:JIP1:APP simulations most representative

Rigid-body docking methods along with extensive sampling of binary complexes were utilized to generate three sets of ternary KLC1:JIP1:APP models, which were then each subjected to unrestrained molecular dynamics.[55,56,73] Set 1 consisted of models generated by docking APP

into the lowest binding free energy complex of a previously reported binary KLC1:JIP1 complex.[73] Set 2 was comprised of models in which the lowest binding free energy conformation of the binary APP:JIP1 complex was docked into KLC1, while Set 3 was produced when another low binding free energy APP:JIP1 conformation in which the JIP1 was located externally with respect to the APP as opposed to internally like in Set 2 was docked to KLC1.

After each set of ternary complexes underwent various amounts of unrestrained molecular dynamics, the binding free energies of each model was computed through two MM-GBSA calculations. One calculation considered the binary complex of KLC1:JIP1 to be the "receptor" while the APP was the "ligand", while the other calculation utilized just the KLC1 as the "receptor" while the binary APP:JIP1 complex served as the "ligand" (Table 8).[5] By utilizing both methods of calculation, it is possible to determine the influence the KLC1 and JIP1, both together and individually, contribute to the overall binding of the ternary complex.

Table 8. The average of the two MM-GBSA binding energy values (one with the binary KLC1:JIP1 as the "receptor" and one with KLC1 as the "receptor") for each ternary KLC1:JIP1:APP model. MM-GBSA values were calculated for every 0.8 ns for Set 1 models and every 0.1 ns for models from Sets 2 and 3. The difference between the two values is also reported. The order of each model is provided as well, in which an order of KLC1:JIP1:APP means that the docked complex yielded a conformation in which the JIP1 was located between the KLC1 and APP proteins.

Simulation	Time (ns)	Avg Avg ΔG_{bind} (kcal/mol)	$\Delta G_{KLC:JIP1}$-ΔG_{KLC}	Order
Set 1 KLC1:JIP1:APP 1 Seed 1	1000	-64.1	-5.4	KLC1:JIP1:APP
Set 1 KLC1:JIP1:APP 2 Seed 1	1000	-113.6	-4.5	KLC1:JIP1:APP
Set 1 KLC1:JIP1:APP 3 Seed 1	1000	-81.5	24.9	KLC1:JIP1:APP
Set 1 KLC1:JIP1:APP 1 Seed 2	1000	-86.0	18.6	KLC1:JIP1:APP
Set 1 KLC1:JIP1:APP 2 Seed 2	1000	-92.9	21.0	KLC1:JIP1:APP
Set 1 KLC1:JIP1:APP 3 Seed 2	1000	-94.9	15.7	KLC1:JIP1:APP
Set 2 KLC1:JIP1:APP 1	75	-56.1	-81.8	KLC:APP-JIP1

Set 2 KLC1:JIP1:APP 2	75	-63.7	-76.6	KLC:APP-JIP1
Set 2 KLC1:JIP1:APP 3	75	-65.9	-66.9	KLC:APP-JIP1
Set 2 KLC1:JIP1:APP 4	75	-47.6	-52.9	KLC:JIP1:APP
Set 2 KLC1:JIP1:APP 5	75	-44.2	-64.7	KLC:APP:JIP1
Set 2 KLC1:JIP1:APP 6	25	-32.4	-72.2	KLC:APP:JIP1
Set 2 KLC1:JIP1:APP 7	25	-41.0	-74.7	KLC:APP:JIP1
Set 2 KLC1:JIP1:APP 8	25	-32.1	-64.1	KLC:APP:JIP1
Set 2 KLC1:JIP1:APP 9	25	-64.0	-48.4	KLC:APP:JIP1
Set 2 KLC1:JIP1:APP 10	75	-60.3	-90.8	KLC:JIP1:APP
Set 3 KLC1:JIP1:APP 1	100	-41.5	-36.2	KLC:APP:JIP1
Set 3 KLC1:JIP1:APP 2	100	-25.1	-41.9	KLC:APP:JIP1
Set 3 KLC1:JIP1:APP 3	100	-78.5	-20.4	KLC:APP:JIP1
Set 3 KLC1:JIP1:APP 4	100	-37.2	-34.1	KLC:APP:JIP1
Set 3 KLC1:JIP1:APP 5	100	-32.9	-30.6	KLC:JIP1:APP
Set 3 KLC1:JIP1:APP 6	100	-40.1	-52.0	KLC:JIP1:APP
Set 3 KLC1:JIP1:APP 7	100	-46.2	-42.6	KLC:APP:JIP1
Set 3 KLC1:JIP1:APP 8	100	-31.0	-33.4	KLC:APP:JIP1
Set 3 KLC1:JIP1:APP 9	100	-26.3	-19.2	KLC:APP:JIP1
Set 3 KLC1:JIP1:APP 10	100	-39.6	-32.9	KLC:JIP1:APP

Physically, and in accordance to the literature, it would make sense for the KLC1 to be the most essential component for binding, with assistance being provided by the JIP1, and not vice-versa. Therefore, it would be desired that the difference between the two MM-GBSA values ($\Delta G_{KLC:JIP1}$-ΔG_{KLC}) would be positive, or at least near zero. Furthermore, the optimal average of the two values would be largely negative. After observing these two sets of optimal conditions, Set 1 of the ternary complexes mostly fulfills both. The first two simulations of Set 1 yielded negative differences between the two MM-GBSA values, but the difference is minimal, especially with respect to those witnessed among Sets 2 and 3.

The values observed in Sets 2 and 3 suggest that the "binding" of the ternary complex was ultimately driven by the interaction between APP and JIP1, and not so much the interaction between KLC1 and the binary complex of APP and JIP1. Physically, this would suggest that KLC1 is not the main component of APP binding, even though KLC1 is the transporter protein. The reason for this disparity is most likely due to the JIP1 not being located between the KLC1 and APP proteins as noted by the "Order" column of Table 8 for many models within Sets 2 and 3. Even those models that did have JIP1 exist between the two other proteins, the criteria as outlined in which a relatively large magnitude negative value of the average of the two MM-GBSA values and the small or positive difference between the values was not met. Due to these reasons, it was determined that Set 1 was the most representative of the KLC1:JIP1:APP ternary complex, and will thus far be used to describe the atomistic characteristics of the system.

Set 1 KLC1:JIP1:APP – relatively strong binding affinity

The three differing starting complexes of the Set 1 KLC1:JIP1:APP complex were simulated in duplicate for one μs. These results are displayed in Table 9. The ternary complex yielded much stronger binding affinities than the binary KLC1:APP complex which yielded MM-GBSA values ranging from -22.5 kcal/mol to -59.2 kcal/mol when computed in a similar fashion, suggesting that JIP1 plays a major role in stabilizing the binding of APP to KLC1.[73] Four of the six ternary simulations yielded negative adjusted binding free energy values when entropy was included. The adjusted binding energies of the ternary complex were still much stronger relative to the adjusted binding free energies of the binary KLC1:APP complexes (ranging from +0.1 kcal/mol to +44.8 kcal/mol), further supporting the theory that JIP1 acts as a link between APP and KLC1.

Table 9. The MM-GBSA binding free energy, with KLC1:JIP1 set as the "receptor", along with normal mode entropy correction for each Set 1 KLC1:JIP1:APP model subjected to long timescale

unrestrained MD. The MM-GBSA binding free energy values were computed every 0.8 ns, while the normal mode analysis was performed every 50 ns. Each simulation was run for 1.0 μs.

Simulation	Average ΔG_{bind} (kcal/mol)	Std Err (kcal/mol)	Sampled TΔS (kcal/mol)	Adjusted ΔG (kcal/mol)
Set 1 KLC1:JIP1:APP 1 Seed 1	-66.8	0.5	-74.7	+7.9
Set 1 KLC1:JIP1:APP 2 Seed 1	-115.8	0.9	-79.0	-36.8
Set 1 KLC1:JIP1:APP 3 Seed 1	-69.0	0.6	-69.9	+0.9
Set 1 KLC1:JIP1:APP 1 Seed 2	-76.7	0.4	-65.4	-11.3
Set 1 KLC1:JIP1:APP 2 Seed 2	-82.4	0.5	-70.3	-12.1
Set 1 KLC1:JIP1:APP 3 Seed 2	-87.0	0.4	-69.1	-17.9

The average structures of each of the six Set 1 KLC1:JIP1:APP complexes show a range of conformations, even between two separate initial seeds corresponding to the same initial model (Figure 10). For Set 1 KLC1:JIP1:APP 1, the first seed yielded the weakest of all six average binding affinities, which is likely caused by the JIP1 being curled away from the KLC1, as well as the C-terminal tail of the APP bending up and away from the cargo-binding domain. When a second set of starting velocity vectors was used for this model, though, it produced one of the stronger binding affinities, which could be accounted for by the JIP1 being positioned in an extended configuration perpendicular to the TPR units of the KLC1, as well as the C-terminal of the APP wrapping back towards the JIP1, increasing intermolecular interactions.

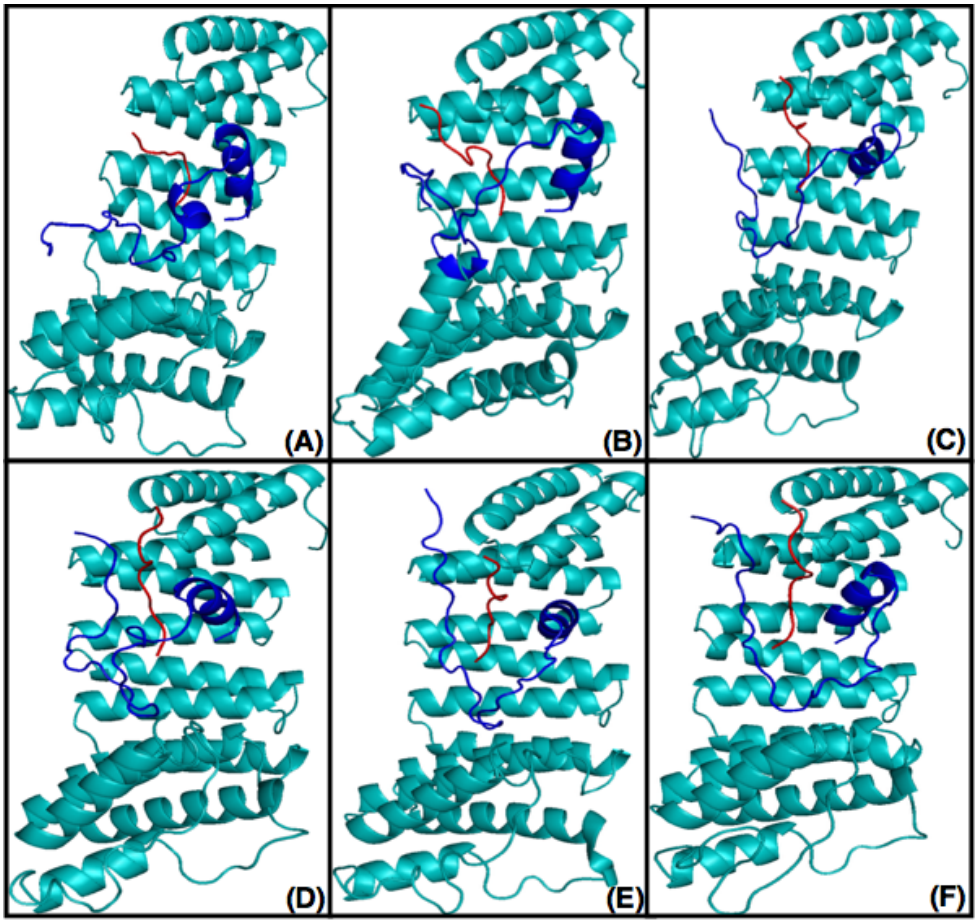

Figure 10. The average structure of each ternary KLC1:JIP1:APP complex was calculated. (a) Set 1 KLC1:JIP1:APP 1 Seed 1 (b) Set 1 KLC1:JIP1:APP 2 Seed 1 (c) Set 1 KLC1:JIP1:APP 3 Seed 1 (d) Set 1 KLC1:JIP1:APP 1 Seed 2 (e) Set 1 KLC1:JIP1:APP 2 Seed 2 (f) Set 1 KLC1:JIP1:APP 3 Seed 2. Teal = KLC1, Red = JIP1, Blue = APP

Set 1 KLC1:JIP1:APP 2 produced the first and third strongest average binding affinities amongst all the ternary simulations for its first and second seed, respectively. The first seed of this model showed the C-terminal tail of the APP once again wrapping back towards the JIP1 just as it had done in the second seed of the first ternary model, but it also has the JIP1 not as extended as it was in the second seed of the first ternary model. With the JIP1 in this more condensed form, it appeared to be exposed for further interaction with more of the KLC1 and other parts of the APP, not just the C-terminal tail. The α-helix of the APP in this first seed also seemed to be positioned ideally to maximize intermolecular interactions, as it was not perpendicular to KLC1's TPR units, rather positioned flatly across them. The second seed of this model had the α-helix of APP, on

average, positioned perpendicularly so to protrude outwards from the TPR units, as well as a protracted C-terminal that stretched past the nearly flat JIP1. This model, although each piece of each protein was very elongated and spread out, was very ordered in the sense that many of the amino acids in the JIP1 were aligned with those in the APP and KLC1 that allowed for strong interactions between the three molecules.

The third Set 1 ternary model, Set 1 KLC1:JIP1:APP 3, produced the second weakest and second strongest adjusted average binding affinities between its first and second seeds, respectively, due to two noticeable differences in their average conformations. In the average structure of the first seed, the C-terminal of the APP is wrapped back towards the JIP1, but we observe a major gap (~ 12 Å) between the two, suggesting minimal interaction throughout the simulation. The α-helix of the APP was also extended perpendicularly upward off the TPR unit like the second seed of the second model. JIP1 is also extended, but at a location of minimal interaction with APP. The second seed, on the other hand, has a much tighter configuration of the extended C-terminal of APP and JIP1. The α-helix of APP also is positioned deeper within KLC1 cargo-binding domain. The second seed shows a clear picture of how the JIP1 can be positioned between the APP and KLC1 to act as the linking protein as suggested by Chiba et al.[42] This pattern of KLC1 and APP clamping down on the JIP1 is seen in all four of the strong and reasonable configurations.

Hydrogen bond Analysis Reveals Lower ΔG Patterns

H-bond analysis was performed to determine the importance of polar interactions within the ternary complex. The amount of hydrogen bonds was calculated between each binary pairing of complex proteins (i.e. KLC1:JIP1, KLC1:APP, APP:KLC1, APP:JIP1, JIP1:KLC1, JIP1:APP) and then these results were normalized by dividing the number of hydrogen bonds by the total

amount of residues within the binary pairing. These graphs are available in the Supplemental Information (Figure S22). It was determined that the overall ternary complex proved more stable when polar interactions were maximized among the pairings of KLC1:APP, APP:JIP1, and JIP1:APP. The other three pairings showed no significant pattern of hydrogen bonding. This suggests that JIP1, half of which is polar amino acids, forms a hydrogen-bonding network between the KLC1 and APP.

The total number of normalized hydrogen bonds between the three pairs of KLC1 and APP, KLC1 and JIP1, and JIP1 and APP of the weakest (Set 1 KLC1:JIP1:APP 1 Seed 1 - Figure 11A) and strongest (Set 1 KLC1:JIP1:APP 2 Seed 1 - Figure 11B) models (Model 1 and Model 2, respectively) reveal two major differences that suggest why there is such a disparity between their binding affinities. First, there is little fluctuation in the overall hydrogen bonding between KLC1 and APP throughout the simulation for Model 1 (between 0.006 and 0.06 normalized hydrogen bonds at each frame, with a slope of -1.5e-05), while there is a smooth upward trend in total hydrogen bonding between the KLC1 and APP in the simulation for Model 2 (between 0.01 and 0.09 normalized hydrogen bonds at each frame, with a slope of 2.8e-0.5). This suggests that hydrogen bonding between KLC1 and APP plays a role in stabilizing the overall complex. Second, the fluctuation in total hydrogen bonding between JIP1 and APP in Model 1 is highly sporadic (between 0.0 and 0.23 hydrogen bonds at each frame, with a slope of 2.2e-06) while that of Model 2 experiences a steady increase throughout the simulation (between 0.0 and 0.2 normalized hydrogen bonds at each frame, with a slope of 5.7e-05). When compared to the binding energy against time for each respective model, there are major drops in binding energy during periods of increasing hydrogen bonds between both KLC1 and APP and JIP1 and APP. The role of hydrogen bonds between KLC1 and JIP1 appears minimal in this context due to the normalization process.

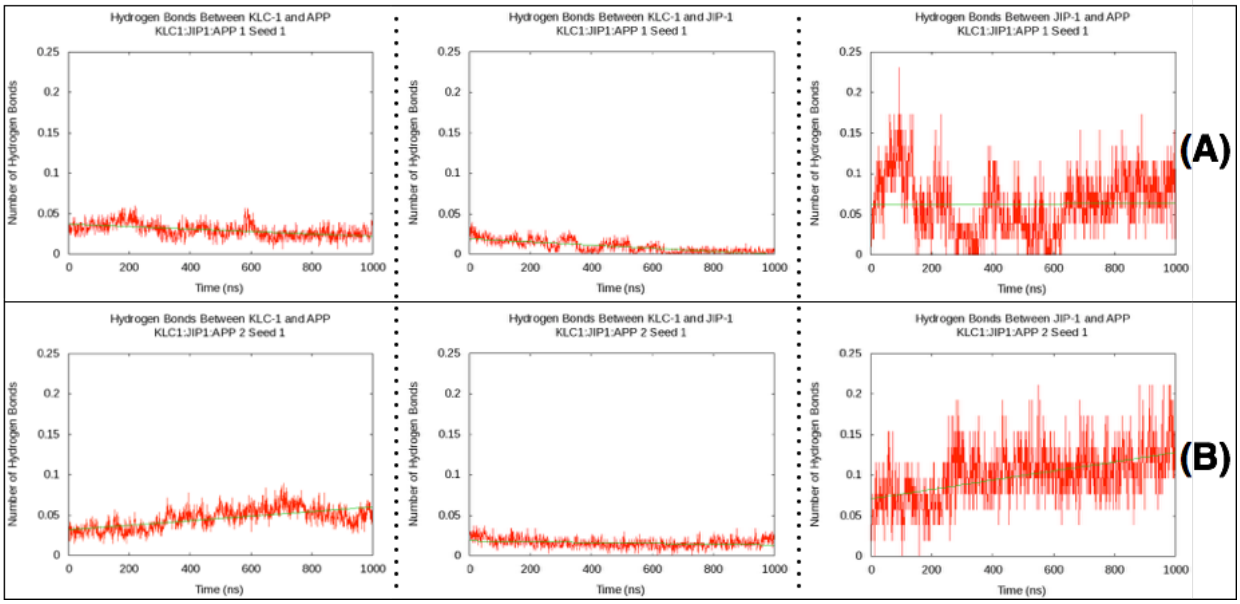

Figure 11. The total, normalized hydrogen bonds against time for the following protein pairs were calculated for (A) Set 1 KLC1:JIP1:APP 1 Seed 1 (weakest average binding affinity) and (B) Set 1 KLC1:JIP1:APP 2 Seed 1 (strongest average binding affinity) and are displayed in a left to right fashion: KLC1 and APP, KLC1 and JIP1, and JIP1 and APP.

Important Residues in Stabilizing KLC1:JIP1:APP

Alanine Scanning of Residues with Significant Decomposition Energy Values

The analysis of individual per residue free energy decomposition graphs of each of the 345 residues involved in the ternary KLC1:JIP1:APP complex revealed six residues as consistently having a major effect on the binding free energy. An example of this selection process is outlined in the Supplemental Information (Figure S23). Not all models revealed all six of these amino acids to be important, but after comparison of the residues described herein for each model, these six were seen to be the most commonly involved in stabilizing the binding within the ternary complex. Three of the six amino acids of interest are part of the central region of JIP1 (Arg556, Pro557, Leu560), while the other three are from the APP (Ile3, His4, His20). The relative configurations of all six residues can be seen in Figure 12A. Alanine scanning was performed to individually mutate these six residues and to calculate the difference in average binding energy that the

mutation caused. From these six residues, His20 from the APP showed a major contribution to binding energy, between -4.9 kcal/mol and -9.2 kcal/mol, once mutated to an alanine in all six simulations. The largest change in binding energy was -10.7 kcal/mol when Arg556 of the JIP1 was mutated to an alanine in KLC1:JIP1:APP 2 Seed 1. The relative configuration of His20 on the APP and Arg556 on the JIP1 can be seen in Figure 12B. There were no trends observed in the weakening of binding affinity when mutating these residues to an alanine.

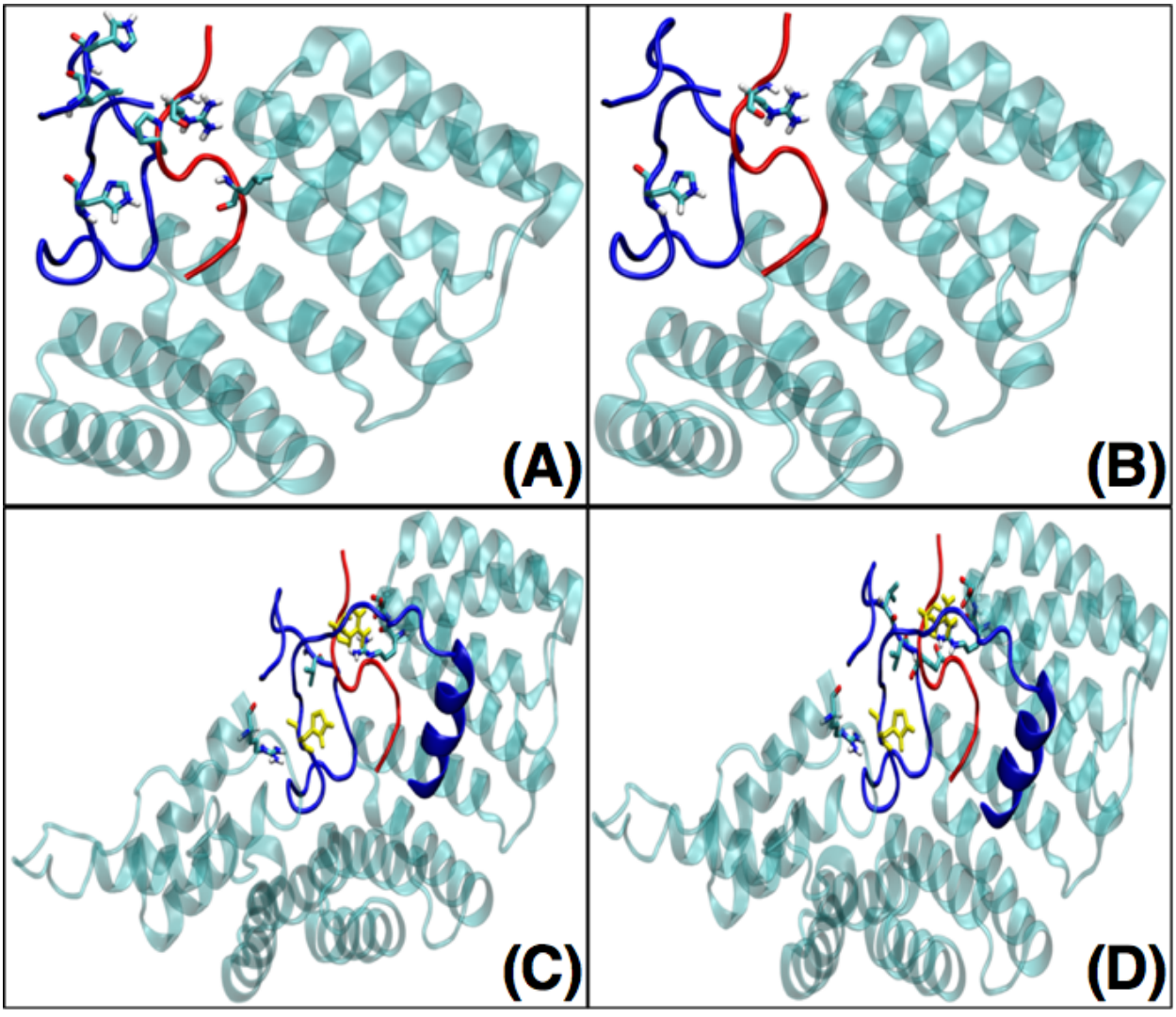

Figure 12. Six residues in the ternary KLC1:JIP1:APP complex were determined to be major contributors to the overall binding free energy after analyzing each residue's individual per residue free energy decomposition graph. (A) The six residues are Arg556, Pro557, and Leu560 on the JIP1 and Ile3, His4, and His20 on the APP. (B) After alanine scanning was performed for each of

these six amino acids, Arg556 on the JIP1 and His20 on the APP showed a major increase in negative contribution to the overall binding affinity. Only partial forms of the APP and KLC1 are shown in (A) and (B). (C) All residues within 2.5 Å of Arg556 and His20 (highlighted in yellow) are shown. (D) All residues within 3 Å of Arg556 and His20 (highlighted in yellow) are shown. KLC1 = cyan, JIP1 = red, APP = blue.

Pairwise Decomposition Values of Important Residues

Patterns that suggest why this change occurred were revealed when pairwise per residue free energy decomposition was calculated for these two amino acids for each of the six ternary simulations. When Arg556 of the JIP1 had maximized polar interactions with both KLC1 and APP, a strong hydrogen bond network was formed to increase binding affinity. For instance, all models with negative average $\Delta G_{adjusted}$ had negative decomposition energy between Arg556 on the JIP1 and at least one residue on the KLC1 as well as at least one residue on the APP. Neither of the two models with a positive average $\Delta G_{adjusted}$ met this criteria, and only had significant interaction with at least one residue on the APP, in regards to the Arg556. Significant in this case means less than -1.0 kcal/mol. For instance, Set 1 KLC1:JIP1:APP 1 Seed 1 ($\Delta G_{adjusted}$ = +7.9 kcal/mol) had a pairwise free energy decomposition value of -5.7 kcal/mol between Arg556 on the JIP1 and Asp11 on the APP, and -1.1 kcal/mol between Arg556 on the JIP1 and Glu18 on the APP. However, there were no significant interactions between the Arg556 with any KLC1 residues. The decomposition energy result between Arg556 on JIP1 and Asp11 on APP is due to the side chains of the two residues hydrogen bonding for 34.10% of the simulation, as well as the backbones of the two residues hydrogen bonding for 22.85% of the simulation. The side chains of Arg556 on JIP1 and Glu18 on APP hydrogen bond for 4.97% of the simulation.

KLC1:JIP1:APP 2 Seed 1 ($\Delta G_{adjusted}$ = -36.8 kcal/mol) met both requirements of the Arg556 hydrogen bonding network. The Arg556 had a pairwise per residue free energy decomposition of -1.2 kcal/mol with Asp80 on the KLC1, as well as -1.2 kcal/mol with Val7 and -13.9 kcal/mol with Glu9 on the APP. The -1.2 kcal/mol decomposition free energy between

Arg556 and Asp80 is derived from the hydrogen bonding between the two residues' side chains 75.37% of the simulation. Glu9 on the APP had the strongest interaction with Arg556 on JIP1 since the side chains of the two residues hydrogen bond for 83.61% of the simulation. The Val10 on the KLC1 had insignificant hydrogen bonding between its backbone atoms and the side chain of Arg556 (0.51% of the simulation) and did not play a large role in the hydrogen bond network. Interactions between Arg556 and glutamates on either the KLC1 or APP are seen to be favored throughout the other simulations as well, which corresponds to experimental results.[38] Arg556 on the JIP1 was found to experimentally favor hydrogen bonding with a glutamate in an inhibited kinase (JNK-1), which was essential for the overall binding of the two proteins. JIP1 repeats this pattern theoretically in the ternary KLC1:JIP1:APP complexes, since the binding affinity of APP to the binary KLC1:JIP1 complex strengthens once Arg556 on the JIP1 hydrogen bonds to a glutamate either on KLC1 or APP.

His20 on the APP did not take part in a three-way hydrogen bonding network like Arg556 on the JIP1, but there was a clear requirement that this residue had to meet to strengthen the ternary complex's binding affinity. As shown in Figure 13, His20 is located near the midpoint of the tail region of APP. His20 can be considered as a junction point in the tail region of APP based upon the interactions it makes with residues in the KLC1 determines the conformation of the rest of the APP tail. Set 1 KLC1:JIP1:APP 1 Seed 1 ($\Delta G_{adjusted}$ = +7.9 kcal/mol) and Set 1 KLC1:JIP1:APP 2 Seed 1 ($\Delta G_{adjusted}$ = -36.8 kcal/mol) will be used once again as examples to describe this function, but first it is important to note that all six ternary KLC1:JIP1:APP simulations have a core set of residues that interact, to some degree, with His20 on the APP: Asp273, Ser274, Pro275, Thr278, and Thr279. KLC1:JIP1:APP 1 Seed 1 only has significant interactions with this core group of residues (Figure 13A), with pairwise per residue free energy decomposition values of -2.9

kcal/mol, -1.1 kcal/mol, -2.1 kcal/mol, -2.1 kcal/mol, and -1.4 kcal/mol, respectively. Since His20 on the APP only significantly interacts with these core residues on the KLC1, the rest of the APP tail is free to be oriented away from the JIP1 since the core residues are located below and across from the His20. These residues simply hold the His20 in place, but has no effect on the orientation of the rest of the tail.

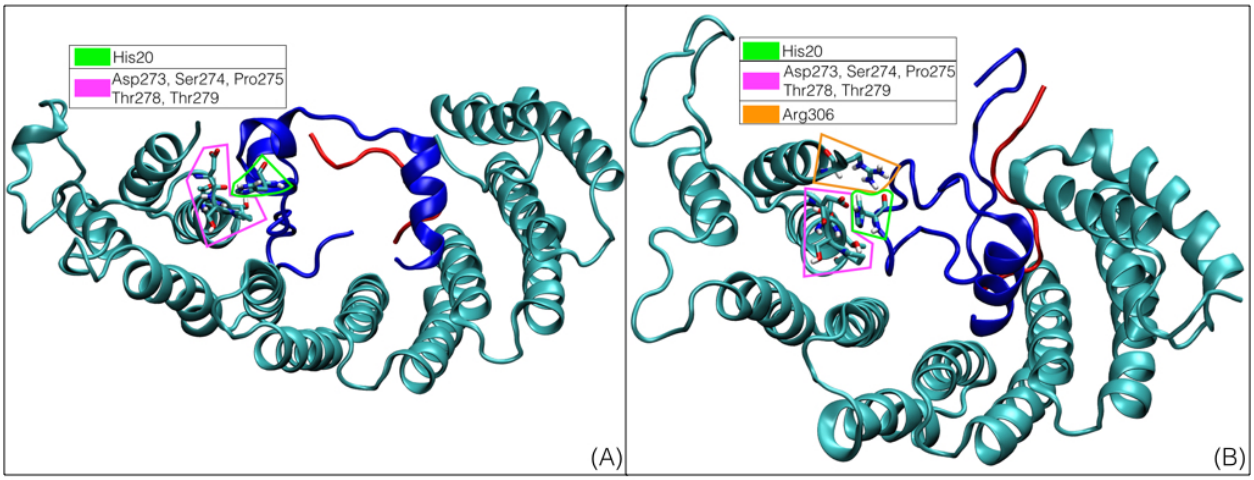

Figure 13. Asp273, Ser274, Pro275, Thr278, and Thr279 act as core residues on KLC1 that interact with His20 on the APP. (A) The His20 of the APP in Set 1 KLC1:JIP1:APP 1 Seed 1 only had signficant interactions with the core KLC1 residues. (B) The His20 of the APP in Set 1 KLC1:JIP1:APP 2 Seed 1 only had significant interactions with Asp273, Thr278, and Arg306 on the KLC1. The Arg306 on the KLC1 forced the tail of the APP to wrap back towards the JIP1 molecule due to its conformation being above that of the His20 on the APP.

Set 1 KLC1:JIP1:APP 2 Seed 1 interacts with both the core group of KLC1 residues as well as a residue located on the TPR unit located above the core group (Arg306). The His20 significantly interacts with Asp273, Thr278, and Arg306 with pairwise per residue free energy decomposition values of -1.5 kcal/mol, -1.3 kcal/mol, and -3.0 kcal/mol, respectively. Figure 13B shows how Arg306 is located above His20 on the APP, forcing the rest of the APP tail to flip back towards the JIP1. By acting as a junction point, His20 on the APP causes the APP tail to wrap back towards the JIP1, allowing for the Arg556 on the JIP1 to initiate its hydrogen bond network between the tail residues of APP as well as KLC1. The four ternary KLC1:JIP1:APP simulations

with negative average $\Delta G_{adjusted}$ values had some interaction between His20 on the APP and Arg306 on the KLC1, forcing the tail turn around.

Conclusion

An examination of the atomistic molecular dynamics of the KLC1, JIP1, and APP proteins in complex with one another can provide significant for determining likely pathways that could ultimately lead to the aggregation of amyloid-beta peptides within dementia-related diseases, as well as lead to the discovery of potential sites of mutation to prevent such processes. Ternary complexes of KLC1, JIP1, and APP persist over long time scale molecular dynamics investigation and were seen to produce favorable MM-GBSA binding free energies, both when the binary KLC1:JIP1 complex was utilized as the receptor, and when just KLC1 was characterized as the receptor. However, it was determined that of the three sets of ternary models produced, Set 1 was the most representative due to its producing of large negative average values of the two MM-GBSA produced binding free energies, as well as its positive or minimally negative difference between them (i.e. $\Delta G_{KLC:JIP1}$-ΔG_{KLC}). Adjusted binding free energy values of Set 1 of the ternary complex yielded much more energetically favorable values than those witnessed in the binary KLC1:APP models previously reported, supporting the preliminary results stated in our earlier study and suggesting that the ternary complex is the more physically accessible mode of transportation of APP by KLC1. After hydrogen bond analysis of the Set 1 complexes, it was determined that stability was optimized when JIP1 served as a hydrogen bond bridge between the KLC1 and the APP. It is important to note that the half of the JIP1 peptide studied is made up of polar amino acids. The key residue that served as this hydrogen bond bridge between KLC1 and APP was Arg556 of the JIP1, which is consistent with previous studies.[38,73] His20 on the APP

acted as a junction point, in which it allowed for the Arg556 on JIP1 to serve as the hydrogen bond bridge between the tail of APP and KLC1 since it caused the APP tail to fold back towards and align with the JIP1 when interacting with Arg306 of the KLC1. Further studies are needed to determine potential mutations that would strengthen the key interactions of His20 on the APP, Arg306 on KLC1, and Arg556 on JIP1, thus potentially preventing the propagation of amyloid-beta aggregation.

Design and Computational Support for the Binding Stability of a New CCR5/CXCR4 Dual Tropic Inhibitor

Cooper A. Taylor[1], Bill R. Miller III[2], and Carol A. Parish[1]

[1]*Department of Chemistry, Gottwald Center for the Sciences, University of Richmond, 28 Westhampton Way, Richmond, VA 23173*

[2]*Department of Chemistry, Truman State University, 100 E. Normal Ave, Kirksville, MO 63501*

The human immunodeficiency virus (HIV) infects healthy human cells by binding to the glycoprotein cluster of differentiation 4 receptors on the surface of helper T-cells, along with either of two chemokine receptors, C-C chemokine receptor type 5 (CCR5) or C-X-C chemokine receptor type 4 (CXCR4). Recently, a pyrazolo-piperdine ligand was synthesized and the corresponding biological data showed good binding to both chemokine receptors, effectively blocking HIV-1 entry. Here, we exhaustively assess the atomistic binding interactions of this compound with both CCR5 and CXCR4, and we find that binding is driven by π–stacking interactions between aromatic rings on the ligand and receptor residues, as well as electrostatic interactions involving the protonated piperidine nitrogen. However, these favorable binding interactions were partially offset by unfavorable desolvation of active site glutamates and aspartates, prompting our proposal of a new, synthetically-accessible derivative designed to increase the electrostatic interactions without compromising the π–stacking features.

Introduction

Over 35 million people worldwide live with the human immunodeficiency virus (HIV), a lentivirus that ultimately leads to the onset of acquired immunodeficiency syndrome (AIDS). Recently, a dual-tropic inhibitor has been reported that could prevent HIV-1 from attacking healthy T-cells.[78] The earliest known receptor to play a role in the HIV infection of healthy cells was the glycoprotein cluster of differentiation 4 (CD4) on the surface of helper T-cells. Designing drugs to inhibit CD4 receptors is a difficult task because helper T-cells are a vital part of the human immune defense. Through various experiments[79-82], CD4 helper T-cells have been shown to prevent immune system failures as well as autoimmune system diseases. HIV targets the CD4

receptor to gain entry into human T-cells. Several early studies[83-86] showed the selective reduction of CD4 helper T-cells by HIV. For instance, Gottlieb et al.[83] used monoclonal antibody analysis to show that HIV-afflicted individuals displayed a major decrease in T-helper cells along with a concomitant increase in T-suppressor cells. Klatzmann et al.[84] found that antibody inhibition of CD4 receptors prevented HIV infection of the cell. Another experimental study in 1993[82] utilized PCR to confirm the dependency on CD4 receptors for HIV1 infection of healthy human cells. However, it was also found that HIV inhibition of helper T-cells was not solely CD4-dependent, but was codependent on CD4 and chemokine receptors.[87] Two chemokine receptors, C-C chemokine receptor type 5 (CCR5) and C-X-C chemokine receptor type 4 (CXCR4), were determined to be necessary for the full infection of human helper T-cells by HIV. Experimentally it has been shown, in the presence of HIV-1, that co-expression of CD4 with CCR5 or CXCR4 correlated with an increase in helper T-cells, suggesting a codependence on CD4 and the chemokine receptors.[88,89]

Chemokines are generally responsible for pro-inflammatory behaviors of leukocyte cells. CCR5 is a β-chemokine that tends to exist on monocytes, lymphocytes, basophils, and eosinophils, while CXCR4 is an α-chemokine that is primarily found on neutrophils.[90,91] Furthermore, it has been found that the expression of CCR5 and CXCR4 receptors is not uniform and can show large variance between T-cell lines and subsets.[92] It is for this reason that HIV treatment methods have required a medley of drugs, specific to the level of expression of each HIV target receptor in an infected individual.

The discovery of CD4/CCR5 and CD4/CXCR4 codependence of HIV inhibition opened new pathways for therapeutic targets.[87-89] To date, the most successful form of treatment for HIV and AIDS is highly active antiretroviral therapy (HAART), which utilizes a combination of several

drugs to ultimately slow the advancement to AIDS through various inhibition of nucleoside/non-nucleoside reverse transcriptase, entry and fusion, integrase, and protease.[93] The utilization of several single-target inhibitors can lead to unintended side effects caused by drug-drug interactions, prompting the need for a single inhibitor capable of multi-site inhibition.[78,94-96]

HIV infects healthy human helper T-cells by binding to CCR5 and/or CXCR4, thereby prompting the development of dual CCR5/CXCR4 inhibitors.[78,97,98] Most recently, one of the more effective CCR5 and CXCR4 dual-inhibitor drugs was reported by Liotta and coworkers.[78] Compound **3** (Figure 14) was shown to be an effective drug with IC_{50} values of 36 μM (CCR5) and 52 μM (CXCR4); however, the mechanism for binding was not known.

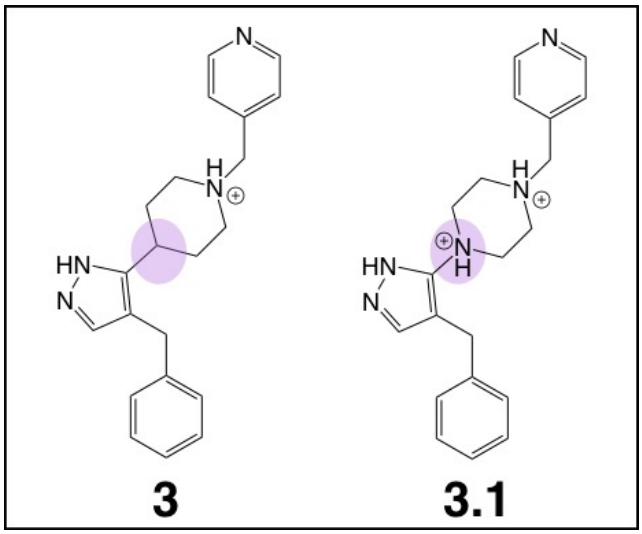

Figure 14. Compound **3** was originally synthesized by the Liotta group, while **3.1** was conceived in this study, replacing the γ-carbon of the piperidine ring of **3** with a second protonated nitrogen.

In this study, we aimed to 1) evaluate the binding dynamics between **3** and CCR5 and CXCR4, 2) determine the atomistic interactions of **3** with each receptor, and 3) design a new, synthetically-accessible derivative of **3** that would increase the inhibition of CCR5 and CXCR4. We believe that a better understanding of the atomistic details and molecular dynamics of these protein-ligand complexes will provide fundamental information with which to better assess potential drugs for dual chemokine receptor ligation.

Methods

Structure Retrieval

Structures of the CCR5 chemokine receptor, and the CXCR4 chemokine receptor in complex with Compound **3**, were obtained through personal communications with Dr. Dennis Liotta at Emory University. The initial docking of these complexes has been described previously.[78] The CCR5 receptor was taken from PDB 4MBS.[99] Two CXCR4 crystal structures were published in 2010, one inhibited by IT1t (PDB ID: 3ODU) and the other by CVX15 (PDB ID: 3OE0).[100] The sequences in these two structures are very similar but not identical. CXCR4:IT1t has a T240P mutation relative to CXCR4:CVX15, and a fragment ~20 residues long beginning after residue 303 that is missing in CXCR4:CVX15 (Figure 15). Novel CXCR4 inhibitors have previously been shown to have higher energy docked poses when docked into CXCR4:IT1t than when docked into CXCR4:CVX15, suggesting that both crystal structures may need to be used to ensure full characterization of binding.[101] Cox et al. attributed the discrepancy in docked poses to differences in binding pocket topology caused by crystallization with the different ligands. The binding pocket in CXCR4:CVX15 is larger than in CXCR4:IT1t as the CVX15 ligand binds to major and minor pockets in CXCR4 whereas IT1t binds only to the minor pocket.[101] Although the original ligands (IT1t and CVX15) were removed from the CXCR4 receptors, the CXCR4 from PDB 3ODU will be referred to as CXCR4:IT1t and the CXCR4 from PDB 3OE0 will be referred to as CXCR4:CVX15. In addition to initiating studies from the previously published Liotta docked complexes, we also performed independent Glide docking to generate five more initial structures (described below) used in this study, each sufficiently different from the Liotta docked complexes, as determined by RMSD and visual inspection.[52,53]

```
                10         20         30         40         50         60
       CCR5 ---------- --------PC QKINVKQIAA RLLPPLYSLV FIFGFVGNML VILILINCKR
CXCR4:CVX15 ---------- ---------- ------PCFR EENANFNKIF LPTIYSIIFL TGIVGNGLVI
 CXCR4:IT1t ---------- ---------- ------PCFR EENANFNKIF LPTIYSIIFL TGIVGNGLVI
                70         80         90        100        110        120
       CCR5 LKSMTDIYLL NLAISDLFFL LTVPFWAHYA AAQWDFGNTM CQLLTGLYFI GFFSGIFFII
CXCR4:CVX15 LVMGYQKKLR SMTDKYRLHL SVADLLFVIT LPFWAVDAVA NWYFGNFLCK AVHVIYTVNL
 CXCR4:IT1t LVMGYQKKLR SMTDKYRLHL SVADLLFVIT LPFWAVDAVA NWYFGNFLCK AVHVIYTVNL
               130        140        150        160        170        180
       CCR5 LLTIDRYLAV VHAVFALKAR TVTFGVVTSV ITWVVAVFAS LPGIIFTRSQ KEGLHYTCSS
CXCR4:CVX15 YSSVWILAFI SLDRYLAIVH ATNSQRPRKL LAEKVVYVGV WIPALLLTIP DFIFANVSEA
 CXCR4:IT1t YSSVWILAFI SLDRYLAIVH ATNSQRPRKL LAEKVVYVGV WIPALLLTIP DFIFANVSEA
               190        200        210        220        230        240
       CCR5 HFPYSQYQFW KNFQTLKIVI LGLVLPLLVM VICYSGILKT LLRCRNEKKR HRAVRLIFTI
CXCR4:CVX15 DDRYICDRFY PNDLWVVVFQ FQHIMVGLIL PGIVILSCYC IIISKLSHSK GHQKRKALKP
 CXCR4:IT1t DDRYICDRFY PNDLWVVVFQ FQHIMVGLIL PGIVILSCYC IIISKLSHSK GHQKRKALKT
               250        260        270        280        290        300
       CCR5 MIVYFLFWAP YNIVLLLNTF QEFFGLNNCS SSNRLDQAMQ VTETLGMTHC CINPIIYAFV
CXCR4:CVX15 TVILILAFFA CWLPYYIGIS IDSFILLEII KQGCEFENTV HKWISITEAL AFFHCCLNPI
 CXCR4:IT1t TVILILAFFA CWLPYYIGIS IDSFILLEII KQGCEFENTV HKWISITEAL AFFHCCLNPI
               310        320        330        340        350        360
       CCR5 GEKFRNYLLV FFQ------- ---------- ---------- ---------- ----------
CXCR4:CVX15 LYA------- ---------- ---------- ---------- ---------- ----------
 CXCR4:IT1t LYAFLGAKFK TSAQHALTSG RPLEVLFQ-- ---------- ---------- ----------
               900        910        920        930        940        950
       CCR5 ---------- ---------- ---------- ---------- ---------- ----------
CXCR4:CVX15 ---------- ---------- ---------- ---------- ---------- ----------
 CXCR4:IT1t ---------G S--------- ---------- ---------- ---------- ----------
              1000       1010       1020       1030       1040       1050
       CCR5 ---------- ---------- ---------- ---------- ---------- ----------
CXCR4:CVX15 --------SG SNIFEMLRID EGLRLKIYKD TEGYYTIGIG HLLTKSPSLN AAKSELDKAI
 CXCR4:IT1t ---------- -NIFEMLRID EGLRLKIYKD TEGYYTIGIG HLLTKSPSLN AAKSELDKAI
              1060       1070       1080       1090       1100       1110
       CCR5 ---------- ---------- ---------- ---------- ---------- ----------
CXCR4:CVX15 GRNTNGVITK DEAEKLFNQD VDAAVRGILR NAKLKPVYDS LDAVRRAALI NMVFQMGETG
 CXCR4:IT1t GRNTNGVITK DEAEKLFNQD VDAAVRGILR NAKLKPVYDS LDAVRRAALI NMVFQMGETG
              1120       1130       1140       1150       1160       1170
       CCR5 ---------- ---------- ---------- ---------- ---------- ----------
CXCR4:CVX15 VAGFTNSLRM LQQKRWDEAA VNLAKSRWYN QTPNRAKRVI TTFRTGTWDA YG--------
 CXCR4:IT1t VAGFTNSLRM LQQKRWDEAA VNLAKSRWYN QTPNRAKRVI TTFRTGTWDA Y---------
              1180       1190       1200       1210       1220       1230
       CCR5 ---------- ---------- ---------- ---------- ---------- ----------
CXCR4:CVX15 ---------- ---------- ---------- ---------- ---------- ----------
 CXCR4:IT1t ---------- ---------- ---------G S--------- ---------- ----------
```

Figure 15. The sequences of each protein utilized for molecular dynamics simulations. Residues that were added during refinement are bolded and underlined. The active site residues are highlighted in red, all negatively charged residues within the active site are highlighted in blue, and the negatively charged residues that produce a positive per residue free energy decomposition value (*vida infra*) are colored green. The active site residues are defined as those residues located within the 10 x 10 x 10 Å Glide docking grid centered on the binding site within the extracellular domain of each receptor. Since the binding sites determined by SITEMAP were different between the two CXCR4 experimental structures, the docking grid contained different sets of residues, leading to different residues highlighted for CXCR4:CVX15 and CXCR4:IT1t.

Structure Preparation

Each complex required structural refinement. The CCR5 crystal structure was missing a loop between residues 241 and 245. The Schrodinger Prime package was utilized to reconstruct the missing loop with the appropriate residues as found within the wild-type FASTA sequence (Cys242, Arg243, and Asn244 were added).[102-104] These three residues are located at the bottom of the CCR5 crystal structure, roughly 40 Å away from the ligand in each of the docked poses, and have negligible effect, if any, on the calculated dynamics and free energies (Figure S24). The 4MBS crystal structure was of a mutated form (C58Y, G163N, Y251D), so the residues were changed to match the wild-type sequence.

The 3OE0 CXCR4:CVX15 structure was missing three loops that were also constructed using Prime.[102-104] According to the FASTA sequence, the missing residues were 67-70 (Lys, Lys, Leu, Arg), 999-1001 (Ser, Gly, Ser), 229-230 (Ser, Lys) and 1162 (Gly).[100] Each of these residues were at least 36 Å away from the ligand, also having negligible to no effect on the computed binding free energies and dynamics of the complex (Figure S24). FASTA sequences of the CCR5 and CXCR4 models used in this study are shown in Figure 2.

Potential Binding Sites in CCR4, CXCR4:CVX15, and CXCR4:IT1t

The SITEMAP package from Schrodinger was used to identify and score putative binding sites based upon shape complementarity and solvent accessibility.[50,105] Identification of binding sites through SITEMAP is based upon the analysis of hydrophobic surface area, surface concavity, and the location of hydrogen-bond donor and acceptor regions.[50,105] Four unique binding sites were identified for each of the three crystal structures, but in each case only one of the four binding sites was located within the extracellular region of the proteins as described by Liotta and coworkers.[78] For CCR5 and CXCR4:IT1t, the top ranked site was located in the extracellular region, while for

CXCR4:CVX15 the second ranked site was in this region (Figure S25).

Glide Docking of Compound 3 into CCR5, CXCR4:CVX15, and CXCR4:IT1t

Schrodinger's Glide program was used to flexibly dock Compound 3 into each of the refined protein structures.[52,53] We utilized a 10 x 10 x 10 Å Glide grid and standard precision (SP) to dock Compound 3 into the binding pocket of each protein.[50,78,105] This was the same approach taken by Cox et al.[52,53,78] The Glide grid was centered on Site 1 for CCR5 and CXCR4:IT1t, and Site 2 for CXCR4:CVX15 (Figure S26).

Five unique poses were generated for each crystal structure. These structures are labeled as "Receptor Structure 2-6" (i.e. CCR5 Structure 2). The RMSD of the ligand within each complex was calculated against the ligand in the respective Cox et al. docked structure ("Receptor Structure 1"). The RMSD comparison was made using the RMSD Trajectory Tool in VMD after superimposing each complex and aligning the backbone residues (Table 10).[68] The variability in RMSD values show the spatial sampling of the ligand within the active site for each complex. Larger values represent a flipped orientation of **3** about the long molecular axis (Figure S27).

Table 10. The RMSD of **3** in each of the Receptor Structures compared to **3** from the Structure 1.

	RMSD (Å)		
Simulation Initiated from Structure:	CCR5	CXCR4:IT1t	CXCR4:CVX15
1	0.000	0.000	0.000
2	7.271	2.089	1.945
3	9.090	7.128	5.054
4	7.509	7.592	3.804
5	7.347	3.967	2.962
6	5.167	4.870	4.869

Preparation and Glide Docking of Compound 3.1

A derivative of **3**, Compound **3.1**, was designed by replacing the γ-carbon of the piperidine ring with a second protonated nitrogen using Gaussview.[106] This compound was then subjected to the same procedure as **3** when docking to CCR5 and CXCR4:IT1t. **3.1** was only docked to CXCR4:IT1t and not CXCR4:CVX15 since no refinements to the CXCR4 protein were necessary. Also, even though there were small differences in the sequence and structure of CXCR4:IT1t and CXCR4:CVX15, our binding analyses suggested that both crystal structures resulted in comparable binding affinities when in complexation with **3** and so we expect the same result for **3.1.**

Molecular Dynamics of the Binding of Compounds 3 and 3.1

All simulations were performed under identical constraints and utilized the *ff14SB* force field.[59] Each system was neutralized with Cl^- ions and solvated with a truncated octahedron periodic box with TIP3P water molecules using AmberTools' *tleap*.[60,61] Initially, a 12.0 Å buffer of solvent between the solute and the edge of the unit cell was required. All steps of MD were performed using the GPU-accelerated *pmemd* code of Amber 14.[61,62] The initial structures were minimized, heated, equilibrated, and then run under unrestrained MD. A seven stage minimization process was performed with each stage consisting of 5,000 steps. The first 1,000 steps were of steepest descent minimization before conjugate gradient minimization was executed on the remaining 4,000 steps. A restraining force of 10.0 kcal/mol/Å^2 on the heavy atoms of the solute was assigned in the first stage, which was then methodically lowered to 0.0 kcal/mol/Å^2 by stage seven. Each structure was then heated from 10 K to 300 K with a restraining force of 10.0 kcal/mol/Å^2 on the solute. After heating, equilibration consisted of lowering the restraining force from 10.0 kcal/mol/Å^2 to 0.0 kcal/mol/Å^2 systematically over seven stages, with each stage lasting

500 ps. Once kinetic equilibrium was reached with no restraining force necessary between the solvent and complex, unrestrained MD at constant pressure (1 atm) and temperature (300 K) was performed. The SHAKE algorithm was used to restrain all covalent bonds to hydrogen atoms, improving computational efficiency as well as allowing the use of a 2-fs time-step in the molecular dynamics simulation.[63]

Unrestrained MD was used to explore the conformational flexibility of all docked poses. Five different, randomly selected seeds were run for 250 ns on Structure 1 of each receptor. A seed represents the set of initial velocities assigned to each atom within the solvated system at the beginning of the simulation. Root-mean-square deviation analyses of MD trajectories indicated minimal movement of the ligand within the active site. Because of this, and because the goal of our simulations was to determine binding free energies and not extensive conformational sampling, we ran subsequent structures for 50 ns. Recently, studies on protein-ligand complexes have used MD simulation times of 50 ns or less and observed reliable binding free energy results.[1-3] Structures 2-6 were studied with a single seed and subjected to 50 ns of unrestrained MD. Simulations initiated from previously published protein-ligand structures were run longer (250ns) than the Glide docked structures (50 ns) we are reporting in this study. We used the longer trajectories to ensure that our methods were correctly describing the previously reported experimental work.[78] Each **3.1** docked complex was used to initiate 250 ns of unrestrained MD.

Analysis

Root mean square deviations (RMSD) and center of mass distances were calculated using *cpptraj* included in AmberTools 14.[61] Binding free energies, per residue free energy decomposition values, and pairwise per residue free energy decomposition values were calculated using MMPBSA.py.[5] Binding free energies and per residue free energy decomposition values

were calculated every 0.1 ns for each simulation using MM-GBSA computations.[4] Normal mode analysis was used to calculate entropy values every 50 ns for a representative simulation of each Structure 1 complex (one each for CCR5, CXCR4:IT1t, and CXCR4:CVX15), as well as every 10 ns for each receptor using one of the five Structures 2-6 structures. Entropy calculations were performed on trajectories having the lowest ΔG_{bind} energies before entropy corrections. When entropy was considered in the overall binding free energy values, the relative binding affinities are suggested to be scaled to more experimentally realistic values.[67] Normal mode entropy calculations have also been found to improve the correlation and ranking of binding affinities.[67] Structures and trajectories were visualized using VMD.[68]

Results and Discussion

The purpose of this study was to obtain an atomistic understanding of the binding of a previously reported dual-inhibitor with the CCR5 and CXCR4 HIV receptors.[78] Using this information, we designed and computationally tested a synthetically accessible derivative of the original compound, that shows enhanced binding to these receptors. Our ability to compute binding free energies for the original compound **3**, in agreement with previously published experimental results, support the viability of the results for the proposed derivative **3.1** in the given receptors.[78] Six unique starting poses of **3** within CCR5, CXCR4:IT1t, and CXCR4:CVX15 were simulated for a combined 1.5 μs for each of the three receptors, as outlined in the Methods section. A total of 1.0 μs of unrestrained MD was also performed on two distinct poses of **3.1** in each of the CCR5 and CXCR4:IT1t receptors. These two poses were selected from among 5 poses obtained via SP Glide docking[52,53] because they sample the flipped orientation observed in some of the **3** and **3.1** docked pose (i.e. where the molecule reorients along the long intramolecular axis).

Stability of Compound 3 Bound to HIV Receptors

MM-GBSA Calculations Reveal Correlation with Experimental Results

In order to assess the reliability of our simulations, MM-GBSA binding affinities were determined for **3** with CCR5, CXCR4:IT1t, and CXCR4:CVX15 (Table 11)[4] for direct comparison with experimental binding data. Experimentally, it has been shown that **3** inhibits CCR5 (IC_{50} = 36 μM) more favorably than CXCR4 (IC_{50} = 52 μM).[78] As seen in Table 11, our computational results are in general agreement with this, with **3** binding to CCR5 with an average ΔG_{bind} of -42.9 kcal/mol across all ten simulations and an average ΔG_{bind} of -31.7 kcal/mol across all twenty CXCR4 simulations. This MM-GBSA approximation does not provide absolute free energies but rather relative free energies that can be used for comparison between similar complexes.[5] When comparing the relative experimental and computational binding free energies there is remarkable agreement, i.e. experimentally we see a CCR5/CXCR4 ratio of 52/36 or 1.44 whereas computationally we see a ratio of -42.9/-31.7 or 1.35. It is important to note that the ΔG_{bind} values derived from the MM-GBSA calculations do not consider entropic effects. The standard error of the mean ranged between 0.1 and 0.3 kcal/mol for each MM-GBSA computation. All graphs of ΔG_{bind} as a function of time for each complex with **3** are available in the Supporting Information (Figs S30-S32).

Table 11. The average binding affinity, ΔG_{bind} of **3** derived from MM-GBSA computations. Normal mode entropy analysis was performed on the simulations shown in bold.

	Avg. ΔG_{bind} (kcal/mol)		
Simulation Initiated from Structure:	CCR5	CXCR4:IT1t	CXCR4:CVX15
1 Seed 1	-46.9	-30.3	-29.8
1 Seed 2	**-47.8**	-34.4	**-34.2**
1 Seed 3	-40.0	-32.5	-34.1
1 Seed 4	-42.3	-26.8	-29.0

1 Seed 5	-42.4	**-44.2**	-27.0
2	-40.8	-21.4	**-40.8**
3	**-45.3**	**-36.1**	-29.1
4	-35.4	-26.1	-31.1
5	-43.4	-32.8	-38.0
6	-44.5	-28.5	-27.9

When normal mode analysis was utilized to perform entropy calculations on the trajectories with the most favorable average binding free energies (shown bolded in Table 11), it was found to retain the ranking of binding affinities, as seen in Table 12. The inclusion of entropy has previously shown to improve MM-GBSA results.[67] After the entropy was incorporated into the MM-GBSA derived binding affinities, CXCR4 remained less strongly inhibited by **3** than CCR5, in agreement with the experimental results of Liotta and coworkers.[78] Due to the high computational demand of normal mode entropy calculations, the six values found in Table 12 will be representative of the alignment of the relative binding results between experiment and theory. As the binding is discussed throughout the remainder of this study, we will be using the relative, non-entropic ΔG_{bind} values. This is possible since normal mode entropy calculation has been found to scale the binding affinities to experimentally realistic results as well as to refine the ranking of binding affinities across different systems, but it is not known to dramatically alter the ranking within the same receptor-ligand complex.[67] This allows for the non-entropic ΔG_{bind} values to be used as an accurate description of the relative binding between similar receptors.

Table 12. The effect of the inclusion of entropy on binding affinity rankings for select CCR5 and CXCR4 complexes with **3**. These simulations were selected for normal mode entropy analysis as they had the most favorable binding free energies prior to the inclusion of entropic effects.

Simulation Initiated from Structure:	Average ΔG (kcal/mol)	Standard Error of the Mean (kcal/mol)	Sampled TΔS (kcal/mol)	Adjusted ΔG (kcal/mol)
CCR5 1 Seed 2	-47.8	0.1	-25.8	-22.0
CCR5 3	-45.3	0.2	-24.3	-21.0

CXCR4:IT1t 1 Seed 5	-44.2	0.1	-24.1	-20.1
CXCR4:IT1t 3	-36.1	0.3	-30.3	-5.8
CXCR4:CVX15 1 Seed 2	-34.2	0.1	-25.0	-9.2
CXCR4:CVX15 2	-40.8	0.2	-24.3	-16.5

Ligand RMSD Values Reveal Minimal Movement from Docked Position

3 was structurally stable in the active site of CCR5 and CXCR4 during all 4.5 μs of unrestrained MD, as seen by ligand RMSD values of less than 4 Å relative to the equilibrated structures. All RMSD values as a function of time can be found in Figure S28.

Pairwise free energy decomposition calculations reveal common interaction patterns

MM-GBSA analysis was performed on all 30 complexes. The good agreement between our computational results and the experimental binding affinities supports the underlying atomistic picture that emerges from the MD trajectories regarding ligand-receptor interactions. In addition to determining MM-GBSA free energies, we also calculated the pairwise free energy decomposition every 0.1 ns between **3** and every residue within the respective receptors.[5] Residues with a pairwise free energy decomposition value more negative than -1.0 kcal/mol were considered to be significant. Although there was some variability as to which residues met the significance criteria across all 30 simulations, there were only a few residues that consistently showed an important contribution to the stabilization of the ligand within the binding site. Based on this atomistic analysis, the stability of **3** within CCR5 and CXCR4 can be attributed to two major interactions: π–stacking and hydrogen bonding.

*Pi-stacking helps stabilize Compound **3** within CCR5*

Pi-stacking interactions occur when two aromatic systems are oriented in a parallel displaced or T-shaped arrangement to one another. This has been previously shown to be a highly stabilizing interaction.[107] π–stacking drives binding of **3** to the receptors due to the presence of three aromatic rings present in the ligand: benzene, pyridine, and pyrazole. In our simulations, the benzene and pyridine rings at either end of the molecule were more flexible due to the single bonds connecting each aromatic ring to the rest of the compound. The dynamics of the pyrazole ring, on the other hand, was more restricted due to its location in the center of the molecule. Three CCR5 residues were consistently found to have some degree of π–stacking with the aromatic rings on **3**: Trp86, Tyr108, and Phe109. The relative influence of these three residues on the overall complex was determined by their average per residue free energy decomposition values for each simulation (Table 13).

Table 13. The average pairwise per residue free energy decomposition values (kcal/mol) of Trp86, Tyr108, and Phe109 on CCR5 with **3**. The moiety on **3** that interacts with each residue via π–stacking is shown. For calibration, pairwise decomposition values between CCR5 and **3** for all interactions (not just π–stacking) ranged from ~ -15 - 0 kcal/mol. The average binding energy for each CCR5 complex is also shown here.

Simulation Initiated from Structure:	Trp86	Tyr108	Phe109	Avg. ΔG_{bind}
CCR5 1 Seed 1	-7.4 (pd)	-7.0 (b)	-3.1 (b)	-46.9
CCR5 1 Seed 2	-6.3 (pd)	-3.8 (pd)	-2.2 (b)	-47.8
CCR5 1 Seed 3	-5.3 (pd)	-5.1 (pd)	-2.5 (b)	-40.0
CCR5 1 Seed 4	-6.0 (pd)	-7.4 (b)	-2.8 (b)	-42.3
CCR5 1 Seed 5	-6.1 (pd)	-6.4 (b)	-3.3 (b)	-42.4
CCR5 2	-4.4 (pa)	-3.4 (pd)	-4.9 (pd)	-40.8
CCR5 3	-8.2 (pd)	-4.2 (pd)	-1.1 (pd)	-45.3
CCR5 4	-6.6 (pd)	-4.2 (pd)	-1.5 (pd)	-35.4
CCR5 5	-4.3 (pa)	-4.2 (pd)	-4.3 (pd)	-43.4
CCR5 6	-7.6 (pd)	-4.8 (pd)	-2.5 (pa)	-44.5

Abbreviations: pd = pyridine, pa = pyrazole, b = benzene.

In our simulations, the CCR5 residue that interacts with **3** most strongly via π–stacking was found to be Trp86, especially via the interaction with the pyridine ring of **3**. With pairwise per residue free energy decomposition values for this π–stacking interaction ranging from -5.3 kcal/mol to -8.2 kcal/mol, the π–stacking interaction between Trp86 and the pyridine ring dominates the majority of the other π–stacking interactions within the CCR5 complex. In all simulations, the π–stacking motif with Trp86 always occurred in the parallel-displaced conformation (Figure 16).

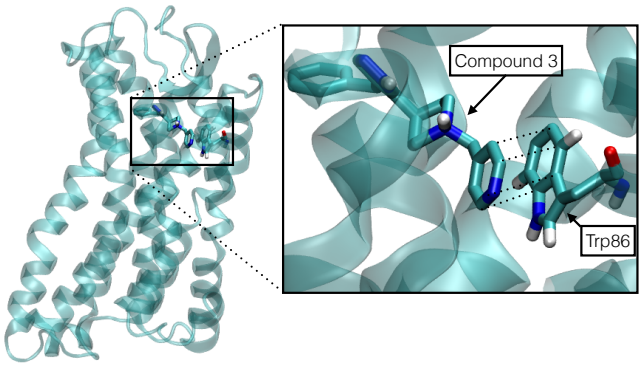

Figure 16. Parallel displaced π–stacking interaction between Trp86 and the pyridine ring of **3**. This image is at 156.2 ns, with a ΔG_{bind} of -63.4 kcal/mol, the lowest binding free energy observed in CCR5 Structure 1 Seed 2.

The other notably strong interaction was the T-shaped π–stacking of Tyr108 with the benzene ring of **3**, with values ranging from -6.4 kcal/mol to -7.0 kcal/mol. This occurred in three of the ten simulations, but shows the importance of the variability of potential π–stacking partners in **3**.

Influence of π–stacking is consistent within CXCR4 models

The π–stacking interactions of the pyridine ring in **3** have the greatest impact on the overall binding to CCR5. However, it is the benzene ring of **3** that contributes the most favorable interaction when binding to CXCR4. Tables S3 and S4 show average pairwise per residue free

energy decomposition values of all π–stacking interactions between CXCR4 residues and the various components of **3**.

For CXCR4:CVX15, the most consistent residue that π–stacked with **3** was His177, which was involved in a π–stacking motif in 9/10 of the simulations. For 6/9 simulations, His177 π–stacked with the benzene ring, while the remaining 3/9 simulations showed π–stacking with the pyrazole ring, which yielded stronger pairwise decomposition values than the benzene π–stacking motif (Table S3). In CXCR4:IT1t, the most significant π–stacking interaction occurred between either of the three aromatic rings of **3** with Trp68. π–stacking via aromatic rings of **3** was observed in all ten CXCR4 simulations, with values ranging from -1.6 (pyridine) to -11.8 (pyrazole) kcal/mol. With three rings available for π–stacking, the structure of **3** is well tuned for favorable interactions with CCR5 and CXCR4.

Unfavorable Polar Solvation Energies Poorly Countered by Hydrogen Bonding

Individual per residue free energy decomposition values were calculated on all 30 trajectories, and each complex revealed a specific pattern of distinct residue interactions that produced positive decomposition values, suggesting a significant source of complex instability. For instance, in all ten CCR5 simulations Glu283 produced significantly positive (greater than 0.5 kcal/mol) per residue free energy decomposition values (Table 14).

Table 14. The per residue free energy decomposition values (kcal/mol) of the residues that produced a significantly positive increase binding free energy in the CCR5 complex with **3**, along with their respective electrostatic and polar solvation energy terms.

Simulation Initiated from Structure:	Residue	Per Residue Free Energy Decomposition	Electrostatic	Polar Solvation
CCR5 1 Seed 1	Glu283	+2.1	-44.5	+48.7
CCR5 1 Seed 2	Glu283	+1.4	-49.0	+52.1
CCR5 1 Seed 3	Glu283	+2.4	-49.0	+52.9

CCR5 1 Seed 4	Glu283	+1.9	-38.6	+42.9
CCR5 1 Seed 5	Glu283	+2.6	-42.3	+47.1
CCR5 2	Glu283	+3.2	-49.4	+54.6
CCR5 3	Glu283	+1.7	-38.0	+40.9
CCR5 4	Glu283	+2.9	-38.1	+42.9
CCR5 5	Glu283	+2.8	-48.4	+52.7
CCR5 6	Glu283	+2.8	-47.7	+52.9

A close examination of the decomposition values reveals that the electrostatic and polar solvation terms dominate the contributions to the binding energies. The unfavorable polar solvation term was consistently of greater magnitude than the favorable electrostatic term, leading to overall positive residue free energy decomposition values. The large polar solvation energy term is due to the desolvation of the negatively charged Glu283 that occurs upon binding to **3**. The electrostatic terms, however, are also large due to Glu238 interacting with **3**. The non-protonated oxygen atoms of Glu283 hydrogen bond with the protonated nitrogen of the piperidine ring of **3**, thus causing the electrostatic energy term to be negative (Figure 17).

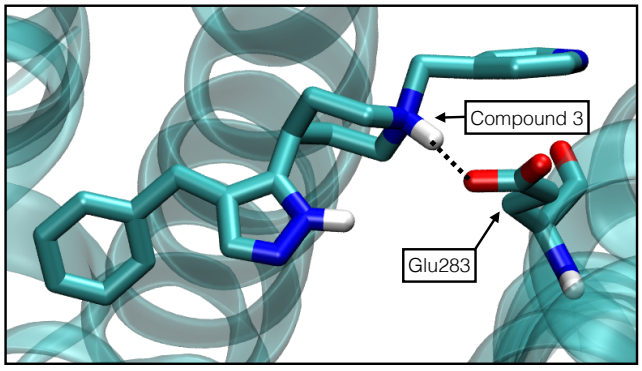

Figure 17. Electrostatic interaction between Glu283 and the protonated nitrogen of the piperidine ring of **3**. This image is the same CCR5 Structure 1 Seed 2 trajectory snapshot as in Figure 16.

The influence of this interaction is further highlighted by tracking the per residue free energy decomposition value of Glu283 against the binding affinity of the overall complex in a

representative CCR5 trajectory (CCR 5 Structure 1 Seed 2), as seen in Figure S29. All major dips and spikes in binding affinity are directly correlated to the free energy decomposition of Glu283 in the CCR5 complexes, which is positive for the majority of the simulation.

The same pattern is observed in the CXCR4 complexes, not only with glutamate, but also with aspartate, the only other negatively charged amino acid at pH 7. Again, the large unfavorable polar solvation energy terms are caused by **3** blocking out a sufficient amount of solvent that could normally stabilize these negatively charged residues. Only one simulation, CXCR4:CVX15 Structure 1 Seed 3, had no significantly positive per residue free energy decomposition peaks, from among 20 different CXCR4 simulations (Table S5). A visualization of the trajectory associated with this exception shows that **3** does not shield any of the active site aspartates or glutamates from interacting with solvent as it did in the other 19 CXCR4 simulations. This is because, in this one simulation, strong π–bonding between the benzene ring of **3** with both Phe173 ($\Delta G_{pairwise}$ = -1.4 kcal/mol) and His177 ($\Delta G_{pairwise}$ = -3.1 kcal/mol) (Table S3) positioned **3** such that the negatively charged residues remained solvent exposed. Only two other CXCR4 simulations (CXCR4:CVX15 Structure 1 Seed 5 and CXCR4:CVX15 Structure 2) showed a π–stacking motif in which the benzene of **3** π–stacked with Phe173 and His177, but these pairwise interactions were not as strong, allowing for **3** to lie closer to Glu288. This reduced the solvent exposure of this negatively charged residue, leading to a significantly positive per residue free energy decomposition value (Table S5). The CXCR4:CVX15 models either showed one or more of the followings residues having a positive decomposition value caused by the polar solvation energy being larger than the electrostatic interactions: Asp171, Asp262, and/or Glu288. The following are the residues in CXCR4:IT1t that showed positive decomposition values for the same reason: Asp97, Asp187, Glu288. These residues, along with Glu283 of CCR5, are colored green in Figure 15. All other

active site glutamates and aspartates are colored blue in Figure 15 because they result in negative per-residue decomposition free energies. The per residue free energy decomposition values, well as electrostatic and polar solvation terms for CXCR4:CVX15 and CXCR4:IT1t are provided in Tables S5 and S6, respectively.

A close examination of the CXCR4:CVX15 and CXCR4:IT1t results suggests similar binding behavior on the part of each experimental structure, in spite of small differences in the number of residues. In part, this is because differences in sequence occur far from the extracellular active site. It has previously been suggested that CXCR4:CVX15, not CXCR4:IT1t, was able to dock small molecules such as **3** via strong interactions with Trp94, Asp97, Asp171, and Glu288.[78,101] We find that these interactions occur regardless of which CXCV4 experimental structure is utilized. The most common interaction observed was between the protonated piperidine nitrogen on **3** and the Glu288 of CXCR4 (Tables S5 and S6).

Maximizing electrostatic interactions with Compound 3.1

Based on this atomistic picture of **3** binding to the CCR5 and CXCR4 chemokine receptors, we postulated that binding could be enhanced by structural modifications that would increase electrostatic interactions between **3** and the receptor. We believed that enhanced electrostatic interactions would counter the unfavorable polar solvation terms that occur when the binding of **3** blocks the interaction of negatively charged glutamates and aspartates with solvent. Keeping synthetic accessibility in mind, these design characteristics were achieved by adding a protonated nitrogen at the γ-carbon of the piperidine ring, forming a piperazine ring (Figure 14). To test this hypothesis, we employed an MM-GBSA computational estimation of binding free energies. Given the good agreement between the computational and experimental binding energies for **3** and the

structural similarities between **3** and **3.1**, we expect the computational results to be predictive of the binding interactions of **3.1** with CCR5 and CXCR4.[78]

Two unique poses of Compound **3.1** were simulated within CCR5, as well as CXCR4:IT1t, for 250 ns each. Two poses were utilized for each receptor to allow for flipped orientations of **3.1** within the active site and to facilitate conformational sampling. The conformational space of the ligand was determined through the flexible docking of Glide producing the two conformations studied for each, along with the extended unrestrained MD.[52,53] This space is well defined as supported by the energetic convergence of the binding free energies computed through MM-GBSA methods.[4] When average ΔG_{bind} was computed for each simulation, it was found that **3.1** yielded relatively more favorable binding free energies with CCR5 than any of the simulations of **3** with CCR5. Simulations of CXCR4:IT1t with **3.1** consistently produced results in the upper range of binding affinities observed with **3** and CXCR4:IT1t. These MM-GBSA based ΔG_{bind} estimations showed **3.1** to bind about 9 and 7 kcal/mol more strongly than **3** with CCR5 and CXCR4, respectively, when the average of all ΔG_{bind} values across the respective simulations were taken into account (Table 15). This is a relative binding increase of about 20% within both receptors.

Table 15. The average binding affinity, ΔG_{bind}, of **3.1** in each of the receptor poses.

CCR5	Avg. ΔG_{bind} (kcal/mol)	Std. Error (kcal/mol)	CXCR4	Avg. ΔG_{bind} (kcal/mol)	Std. Error (kcal/mol)
Pose 1	-50.0	0.1	Pose 1	-44.0	0.1
Pose 2	-53.0	0.1	Pose 2	-32.5	0.1

A close examination of the resulting trajectories shows that **3.1** does indeed have stronger and more favorable electrostatic interactions with CCR5; and, this results in improved binding free energies. As theorized, the electrostatic enhancement offsets the unfavorable polar solvation

energy term of Glu283 when a ligand is present in the active site. No significantly positive per residue free energy decomposition energies (values greater than +0.5 kcal/mol) were observed in either of the CCR5 models (Table 16). Simulations initiated from Pose 1 of CXCR4 with **3.1** produced only one significantly positive decomposition energy with residues within the active site (1.4 kcal/mol with Asp97), while Pose 2 produced two significantly positive decomposition energies at Asp187 (1.7 kcal/mol) and Glu288 (4.3 kcal/mol). A close analysis of the simulation trajectories suggests that **3.1** can consistently stabilize one of the three negatively charged amino acids within the CXCR4 active site as seen by interaction with Glu288 in Pose 1 and Asp97 in Pose 2. Asp187 in Pose 2 also experienced a favorable ΔG due to its high level of solvent exposure.

Table 16. The per residue free energy decomposition values (kcal/mol) from simulations of **3.1** with CCR5/CXCR4. Shown are residues that produced a significantly positive increase in the binding free energy, along with the respective electrostatic and polar solvation energy terms.

Simulation Initiated from Structure:	Residue	Per Residue Free Energy Decomposition	Electrostatic	Polar Solvation
CCR5 Pose 1	Glu283	-1.4	-94.6	+94.6
CCR5 Pose 2	Glu283	-5.0	-94.6	+91.8
CXCR4 Pose 1	Asp 97	+1.4	-55.1	+57.7
	Asp187	-0.2	-32.4	+32.8
	Glu288	+0.5	-84.5	+85.6
CXCR4 Pose 2	Asp97	+0.4	-30.1	+31.0
	Asp187	+1.7	-53.1	+55.7
	Glu288	+4.3	-76.0	+81.8

The binding free energies suggest that each simulation has reached a sampling convergence for the various poses of **3.1** in each receptor (Figure 18). Although the average ΔG_{bind} of each

CXCR4 pose suggests a binding affinity akin to **3**, it is important to note the significant sampling of binding free energies between -50 and -60 kcal/mol, suggesting the potential for **3.1** to experimentally bind more tightly with CXCR4 than **3**. The computed relative energies for both CXCR4 and CCR5 suggest that the change from a piperidine to a piperazole ring in **3** could provide a significant enhancement to the overall binding affinity.

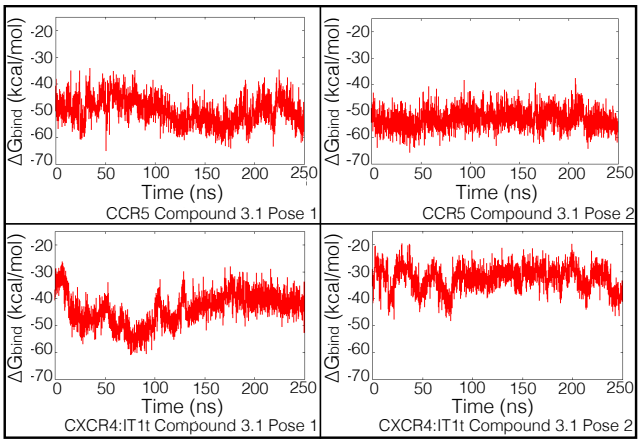

Figure 18. The binding free energy of each **3.1** inhibited complex throughout the entirety of the 250 ns simulation.

Conclusion

An atomistic understanding of the inhibition of CCR5 and CXCR4 by the Liotta et al. synthesized **3** has been determined. Based on both experimental and computational results, **3** shows higher affinity for CCR5 than CXCR4. **3** remained within the active site for all 30 molecular dynamics simulations, and the trajectories from those simulations produced MM-GBSA binding energies in good agreement with experiment. The binding of **3** could be attributed to two main factors: three aromatic rings capable of π–stacking, and the presence of a positively charged hydrogen bond donor that could assist in stabilizing the negatively charged glutamates and aspartates within the active site. However, the presence of **3** within the active site reduces the solvent exposure of those negatively charged residues, and this is a destabilizing effect. Our results

suggest that changing the protonated piperidine ring to a doubly protonated piperazine ring in **3** could enhance the electrostatic interactions between the ligand and the negatively charged residues, thus counteracting the destabilization effects of the polar solvation energy term, while retaining the potential π–stacking motifs. This minor structural change yielded a roughly 20% relative increase in MM-GBSA binding affinities within both CCR5 and CXCR4.

References

(1) Miller, B. R.; Roitberg, A. E. *J Mol Graph Model* **2013**, *45*, 10.1016/j.jmgm.2013.08.009.

(2) Chong Teoh, T.; Heidelberg, T.; Rizman-Idid, M. *DARU Journal of Pharmaceutical Sciences* **2011**, *19*, 469.

(3) Kheirabadi, M.; Maleki, J.; Soufian, S.; Hosseini, S. *Molecular Biology Research Communications* **2016**, *5*, 19.

(4) Massova, I.; Kollman, P. A. *Perspectives in Drug Discovery and Design* **2000**, *18*, 113.

(5) Miller, B. R.; McGee, D. T.; Swails, J. M.; Homeyer, N.; Gohlke, H.; Roitberg, A. E. *Journal of Chemical Theory and Computation* **2012**, *8*, 3314.

(6) Gohlke, H.; Kiel, C.; Case, D. A. *Journal of Molecular Biology* **2003**, *330*, 891.

(7) Sherrington, R.; Rogaev, E. I.; Liang, Y.; Rogaeva, E. A.; Levesque, G.; Ikeda, M.; Chi, H.; Lin, C.; Li, G.; Holman, K.; Tsuda, T.; Mar, L.; Foncin, J. F.; Bruni, A. C.; Montesi, M. P.; Sorbi, S.; Rainero, I.; Pinessi, L.; Nee, L.; Chumakov, I.; Pollen, D.; Brookes, A.; Sanseau, P.; Polinsky, R. J.; Wasco, W.; Da Silva, H. A.; Haines, J. L.; Perkicak-Vance, M. A.; Tanzi, R. E.; Roses, A. D.; Fraser, P. E.; Rommens, J. M.; St George-Hyslop, P. H. *Nature* **1995**, *375*, 754.

(8) Goate, A.; Chartier-Harlin, M. C.; Mullan, M.; Brown, J.; Crawford, F.; Fidani, L.; Giuffra, L.; Haynes, A.; Irving, N.; James, L. *Nature* **1991**, *349*, 704.

(9) Levy-Lahad, E.; Wasco, W.; Poorkaj, P.; Romano, D. M.; Oshima, J.; Pettingell, W. H.; Yu, C. E.; Jondro, P. D.; Schmidt, S. D.; Wang, K. *Science* **1995**, *269*.

(10) Stokin, G. B.; Almenar-Queralt, A.; Gunawardena, S.; Rodrigues, E. M.; Falzone, T.; Kim, J.; Lillo, C.; Mount, S. L.; Roberts, E. A.; McGowan, E.; Williams, D. S.; Goldstein, L. S. *Human Molecular Genetics* **2008**, *17*, 3474.

(11) Kumar, D. K. V.; Choi, S. H.; Washicosky, K. J.; Eimer, W. A.; Tucker, S.; Ghofrani, J.; Lefkowitz, A.; McColl, G.; Goldstein, L. E.; Tanzi, R. E.; Moir, R. D. *Science Translational Medicine* **2016**, *8*.

(12) Stokin, G. B.; Goldstein, L. S. B. *Journal of Physiology - Paris* **2006**, *99*, 193.

(13) Stokin, G. B.; Goldstein, L. S. B. *Annual Review Biochemistry* **2006**, *75*, 607.

(14) O'Brien, R. J.; Wong, P. C. *Annual Review Neurscience* **2011**, *34*, 185.

(15) Tanzi, R. E.; L., B. *Cell* **2005**, *120*, 545.

(16) Zhang, Y.-w.; Thompson, R.; Zhang, H.; Xu, H. *Molecular Brain* **2011**, *4*, 1.

(17) Gauger, A. K.; Goldstein, L. S. B. *J. Biol. Chem.* **1993**, *268*, 13657.

(18) Fisher, S.; Weck, M.; Landers, J.; Emrich, J.; Middleton, S.; Cox, J.; Gentile, L.; Parish, C. *Journal of Structural Biology* **2012**, *177*, 602.

(19) Lamb, J. R.; Michaud, W. A.; Sikorski, R. S.; Hieter, P. A. *EMBO J.* **1994**, *13*, 4321.

(20) Gindhart, J. G., Jr.; Goldstein, L. S. B. *Trends Biochem. Sci.* **1996**, *21*, 52.

(21) Sikorski, R. S.; Boguski, M. S.; Goebl, M.; Hieter, P. *Cell (Cambridge, Mass.)* **1990**, *60*, 307.

(22) Main, E.; Xiong, Y.; Cocco, M. J.; D'Andrea, L.; Regan, L. *Structure* **2003**, *11*, 497.

(23) Main, E. R. G.; Jackson, S. E.; Regan, L. *Current Opinion in Structural Biology* **2003**, *13*, 482.

(24) Scheufler, C.; Brinker, A.; Bourenkov, G.; Pegoraro, S.; Moroder, L.; Bartunik, H.; Hartl, F. U.; Moarefi, I. *Cell (Cambridge, Mass.)* **2000**, *101*, 199.
(25) Jarymowycz, V. A.; Cortajarena, A. L.; Regan, L.; Stone, M. J. *J. Biomol. NMR* **2008**, *41*, 169.
(26) Kaether, C.; Skehel, P.; Dotti, C. G. *Molecular Biology of the Cell* **2000**, *11*, 1213.
(27) Kamal, A.; Almenar-Queralt, A.; LeBlanc, J. F.; Roberts, E. A. *Nature* **2001**, *414*, 643.
(28) Sheng, J. G.; Price, D. L.; Koliatsos, V. E. *Experimental neurology* **2003**, *184*, 1053.
(29) Stokin, G. B.; Lillo, C.; Falzone, T. L.; Brusch, R. G.; Rockenstein, E.; Mount, S. L.; Raman, R.; Davies, P.; Masliah, E.; Williams, D. S.; Goldstein, L. S. *Science* **2005**, *307*, 1282.
(30) Amaratunga, A.; Leeman, S. E.; Kosik, K. S.; Fine, R. E. *Journal of neurochemistry* **1995**, *64*, 2374.
(31) Akhmanova, A.; Hammer, J. A. *Current Opinion in Cell Biology* **2010**, *22*, 479.
(32) Weaver, C.; Leidel, C.; Szpankowski, L.; Farley, N. M.; Shubeita, G. T.; Goldstein, L. S. *Traffic* **2013**, *14*, 295.
(33) Kamal, A.; Stokin, G. B.; Yang, Z.; Xia, C.-H.; Goldstein, L. S. B. *Neuron* **2000**, *28*, 449.
(34) Matsuda, S.; Matsuda, Y.; D'Adamio, L. *The Journal of Biological Chemistry* **2003**, *278*, 38601.
(35) Pellet, J.-B.; Haefliger, J.-A.; Staple, J.; Widmann, C.; Welker, E.; Hirling, H.; Bonny, C.; Nicod, P.; Catsicas, S.; Waeber, G.; Riederer, B. *European Journal of Neuroscience* **2000**, *12*, 621.
(36) Barr, R. K.; Kendrick, T. S.; Bogoyevitch, M. A. *Journal of Biological Chemistry* **2002**, *277*, 10987.
(37) Zhao, H.; Serby, M. D.; Xin, Z.; Szczepankiewicz, B. G.; Liu, M.; Kosogof, C.; Liu, B.; Nelson, L. T. J.; Johnson, E. F.; Wang, S.; Pederson, T.; Gum, R. J.; Clampit, J. E.; Haasch, D. L.; Abad-Zapatero, C.; Fry, E. H.; Rondinone, C.; Trevillyan, J. M.; Sham, H. L.; Liu, G. *J. Med. Chem.* **2006**, *49*, 4455.
(38) Heo, Y.-S.; Kim, S.-K.; Seo, C.; Kim, Y.; Sung, B.-J.; Lee, H.; Lee, J.; Park, S.-Y.; Kim, J.; Hwang, K.; Hyun, Y.-L.; Jeon, Y.; Ro, S.; Cho, J.; Lee, T.; Yang, C.-H. *The EMBO Journal* **2004**, *23*, 2185.
(39) Verhey, K.; Meyer, D.; Deehan, R.; Blenis, J.; Schnapp, B.; Rapoport, T.; Margolis, B.; Verhey, K. *The Journal of Cell Biology* **2001**, *152*, 959.
(40) Zhu, H.; Lee, H.; Tong, Y.; Hong, B.-S.; Kim, K.-P.; Shen, Y.; Lim, K.; Mackenzie, F.; Tempel, W.; Park, H.-W. *PLoS ONE* **2012**, *7*.
(41) Scheinfeld, M. H.; Roncarati, R.; Vito, P.; Lopez, P. A.; Abdallah, M.; D'Adamio, L. *Journal of Biological Chemistry* **2002**, *277*, 3767.
(42) Chiba, K.; Araseki, M.; Nozawa, K.; Furukori, K.; Araki, Y.; Matsushima, T.; Nakaya, T.; Hata, S.; Saito, Y.; Uchida, S.; Okada, Y.; Nairn, A. C.; Davis, R. J.; Yamamoto, T.; Kinjo, M.; Taru, H.; Suzuki, T. *Molecular Biology of the Cell* **2014**, *25*, 3569.
(43) van der Kant, R.; Goldstein, L. S. *Developmental Cell* **2015**, *32*, 502.
(44) Lazarov, O.; Morfini, G. A.; Lee, E. B.; Farah, M. H.; Szodorai, A.; DeBoer, S. R.; Koliatsos, V. E.; Kins, S.; Lee, V. M.-Y.; Wong, P. C.; Price, D. L.; Brady, S. T.; Sisodia, S. S. *The Journal of Neuroscience* **2005**, *25*, 2386.
(45) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. *Nucleic Acids Research* **2000**, *28*, 235.
(46) Tong, Y.; Tempel, W.; Shen, L.; Shen, Y.; Nedyalkova, L.; Arrowsmith, C. H.; Edwards, A. M.; Bountra, C.; Weigelt, J.; Bochkarev, A.; Park, H. *To be published* **2010**.

(47) Zhao, H.; Serby, M. D.; Xin, Z.; Szczepankiewicz, B. G.; Liu, M.; Kosogof, C.; Liu, B.; Nelson, L. T.; Johnson, E. F.; Wang, S.; Pederson, T.; Gum, R. J.; Clampit, J. E.; Haasch, D. L.; Abad-Zapatero, C.; Fry, E. H.; Rondinone, C.; Trevillyan, J. M.; Sham, H. L.; Liu, G. *Journal of medicinal chemistry* **2006**, *49*, 4455.
(48) Barrett, P. J.; Song, Y.; Horn, V. W. D.; Hustedt, E. J.; Schafer, J. M.; Hadziselimovic, A.; Beel, A. J.; Sanders, C. R. *Science* **2012**, *336*, 1168.
(49) Sastry, M. G.; Adzhigirey, M.; Day, T.; Annabhimoju, R.; Sherman, W. *Journal of Computer-Aided Molecular Design* **2013**, *27*, 221.
(50) Halgren, T. *Chemical Biology & Drug Design* **2007**, *69*, 146.
(51) Halgren, T. A. *Journal of Chemical Information and Modeling* **2009**, *49*, 377.
(52) Friesner, R. A.; Banks, J. L.; Murphy, R. B. *Journal of Medicinal Chemistry* **2004**, 1739.
(53) Halgren, T. A.; Murphy, R. B.; Friesner, R. A. *Journal of Medicinal Chemistry* **2004**, 1750.
(54) Friesner, R. A.; Murphy, R. B.; Repasky, M. P. *Journal of Medicinal Chemistry* **2006**, 6177.
(55) Tovchigrechko, A.; Vakser, I. A. *Proteins: Structure, Function and Bioinformatics* **2005**, *60*, 296.
(56) Tovchigrechko, A.; Vakser, I. A. *Nucleic Acids Research* **2006**, *34*, W310.
(57) Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S. *Journal of Medicinal Chemistry* **2004**, *47*, 1739.
(58) Cross, J. B.; Thompson, D. C.; Rai, B. K.; Baber, J. C.; Fan, K. Y.; Hu, Y.; Humblet, C. *Journal of Chemical Information and Modeling* **2009**, *49*, 1455.
(59) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B. *Proteins: Structure, Function, and Bioinformatics* **2006**, 712.
(60) Jorgensen, W. L.; Madura, J. D. *Journal of the American Chemical Society* **1983**, *105*, 1407.
(61) Case, D. A.; Babin, V.; Berryman, J. T.; Betz, R. M.; Cai, Q.; Cerutti, D. S.; Cheatham, I. T. E.; Darden, T. A.; Duke, R. E.; Gohlke, H.; Goetz, A. W.; Gusarov, S.; Homeyer, N.; Janowski, P.; Kaus, J.; Kolossváry, I.; Kovalenko, A.; Lee, T. S.; LeGrand, S.; Luchko, T.; Luo, R.; Madej, B.; Merz, K. M.; Paesani, F.; Roe, D. R.; Roitberg, A.; Sagui, C.; Salomon=Ferrer, R.; Seabra, G.; Simmerling, C. L.; Smith, W.; Swails, J.; Walker, R. C.; Wang, J.; Wolf, R. M.; Wu, X.; Kollman, P. A. *AMBER 14* **2014**.
(62) Salomon-Ferrer, R.; Götz, A. W.; Poole, D.; Grand, S.; Walker, R. C. *Journal of Chemical Theory and Computation* **2013**, 3878.
(63) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *Journal of Computational Physics* **1977**, *23*, 327.
(64) Roe, D. R.; Cheatham, T. E. *Journal of Chemical Theory and Computation* **2013**, *9*, 3084.
(65) Onufriev, A.; Bashford, D.; Case, D. A. *Proteins* **2004**, *55*, 383.
(66) McQuarrie, D. A. *Statistical Mechanics*; 2 ed.; University Science Books, 2000.
(67) Hou, T.; Wang, J.; Li, Y.; Wang, W. *Journal of Chemical Information Model* **2010**, 69.
(68) Humphrey, W.; Dalke, A.; Schulten, K. *Journal of Molecular Graphics* **1996**, *14*, 33.
(69) Schrodinger, LLC 2015.
(70) Chen, J.; Sawyer, N.; Regan, L. *Protein Science : A Publication of the Protein Society* **2013**, *22*, 510.

(71) Hirano, T.; Kinoshita, N.; Morikawa, K.; Yanagida, M. *Cell (Cambridge, Mass.)* **1990**, *60*, 319.
(72) Shao, J.; Tanner, S. W.; Thompson, N.; Cheatham, T. E. *Journal of Chemical Theory and Computation* **2007**, *3*, 2312.
(73) Taylor, C. A.; Miller, B. R.; Shah, S. S.; Parish, C. A. *Proteins: Structure, Function, and Bioinformatics* **2016**, *85*, 221.
(74) Friesner, R. A. *Annual Reviews in Physical Chemistry* **1991**, *42*, 341.
(75) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. *Proteins* **2006**, *65*, 712.
(76) Salomon-Ferrer, R.; Götz, A. W.; Poole, D.; Le Grand, S.; Walker, R. C. *Journal of Chemical Theory and Computation* **2013**, *9*, 3878.
(77) Humphrey, W.; Dalke, A.; Schulten, K. *J Mol Graphics* **1996**, *14*, 33.
(78) Cox, B. D.; Prosser, A. R.; Sun, Y.; Li, Z.; Lee, S.; Huang, M. B.; Bond, V. C.; Snyder, J. P.; Krystal, M.; Wilson, L. J.; Liotta, D. C. *ACS Med Chem Lett* **2015**, *6*, 753.
(79) Linterman, M. A.; Rigby, R. J.; Wong, R. K.; Yu, D.; Brink, R.; Cannons, J. L.; Schwartzberg, P. L.; Cook, M. C.; Walters, G. D.; Vinuesa, C. G. *The Journal of experimental medicine* **2009**, *206*, 561.
(80) King, C.; Tangye, S. G.; Mackay, C. R. *Annual review of immunology* **2008**, *26*, 741.
(81) Simpson, N.; Gatenby, P. A.; Wilson, A.; Malik, S.; Fulcher, D. A.; Tangye, S. G.; Manku, H.; Vyse, T. J.; Roncador, G.; Huttley, G. A.; Goodnow, C. C.; Vinuesa, C. G.; Cook, M. C. *Arthritis and rheumatism* **2010**, *62*, 234.
(82) Collin, M. H., G.; Montaner, L.; Gordon, S. *Research in Virology* **1993**, *144*, 13.
(83) Gottlieb , M. S.; Schroff , R.; Schanker , H. M.; Weisman , J. D.; Fan , P. T.; Wolf , R. A.; Saxon , A. *New England Journal of Medicine* **1981**, *305*, 1425.
(84) Klatzmann, D.; Champagne, E.; Chamaret, S.; Gruest, J.; Guetard, D.; Hercend, T.; Gluckman, J.-C.; Montagnier, L. *Nature* **1984**, *312*, 767.
(85) Dalgleish, A. G.; Beverley, P. C.; Clapham, P. R.; Crawford, D. H.; Greaves, M. F.; Weiss, R. A. *Nature* **1984**, *312*, 763.
(86) McDougal, J. S.; Cort, S. P.; Kennedy, M. S.; Cabridilla, C. D.; Feorino, P. M.; Francis, D. P.; Hicks, D.; Kalyanaraman, V. S.; Martin, L. S. *Journal of Immunological Methods* **1985**, *76*, 171.
(87) Maddon, P. J.; Dalgleish, A. G.; McDougal, J. S.; Clapham, P. R.; Weiss, R. A.; Axel, R. *Cell* **1986**, *47*, 333.
(88) Choe, H.; Farzan, M.; Sun, Y.; Sullivan, N.; Rollins, B.; Ponath, P. D.; Wu, L.; Mackay, C. R.; LaRosa, G.; Newman, W.; Gerard, N.; Gerard, C.; Sodroski, J. *Cell* **1996**, *85*, 1135.
(89) Delezay, O.; Koch, N.; Yahi, N.; Hammache, D.; Tourres, C.; Tamalet, C.; Fantini, J. *Aids* **1997**, *11*, 1311.
(90) Baggiolini, M.; Dewald, B.; Moser, B. *Advances in Immunology* **1993**, *55*, 97.
(91) Schall, T. J.; Bacon, K. B. *Current opinion in immunology* **1994**, *6*, 865.
(92) Lee, B.; Sharron, M.; Montaner, L. J.; Weissman, D.; Doms, R. W. *Proceedings of the National Academy of Sciences* **1999**, *96*, 5215.
(93) Detels, R.; Munoz, A.; McFarlane, G.; Kingsley, L. A.; Margolick, J. B.; Giorgi, J.; Schrager, L. K.; Phair, J. P. *Jama* **1998**, *280*, 1497.
(94) Herschhorn, A.; Gu, C.; Espy, N.; Richard, J.; Finzi, A.; Sodroski, J. G. *Nat Chem Biol* **2014**, *10*, 845.

(95) Watson Buckheit, K.; Yang, L.; Buckheit, R. W. *Antimicrob Agents Chemother* **2011**, *55*, 5243.
(96) Jenwitheesuk, E.; Horst, J. A.; Rivas, K. L.; Van Voorhis, W. C.; Samudrala, R. *Trends in pharmacological sciences* **2008**, *29*, 62.
(97) Horuk, R. *Expert reviews in molecular medicine* **2009**, *11*, e1.
(98) Princen, K.; Hatse, S.; Vermeire, K.; Aquaro, S.; De Clercq, E.; Gerlach, L. O.; Rosenkilde, M.; Schwartz, T. W.; Skerlj, R.; Bridger, G.; Schols, D. *J Virol* **2004**, *78*, 12996.
(99) Tan, Q.; Zhu, Y.; Li, J.; Chen, Z.; Han, G. W.; Kufareva, I.; Li, T.; Ma, L.; Fenalti, G.; Li, J.; Zhang, W.; Xie, X.; Yang, H.; Jiang, H.; Cherezov, V.; Liu, H.; Stevens, R. C.; Zhao, Q.; Wu, B. *Science (New York, N.Y.)* **2013**, *341*, 10.1126/science.1241475.
(100) Wu, B.; Chien, E. Y. T.; Mol, C. D.; Fenalti, G.; Liu, W.; Katritch, V.; Abagyan, R.; Brooun, A.; Wells, P.; Bi, F. C.; Hamel, D. J.; Kuhn, P.; Handel, T. M.; Cherezov, V.; Stevens, R. C. *Science (New York, N.Y.)* **2010**, *330*, 1066.
(101) Cox, B. D.; Prosser, A. R.; Katzman, B. M.; Alcaraz, A. A.; Liotta, D. C.; Wilson, L. J.; Snyder, J. P. *ChemBioChem* **2014**, *15*, 1614.
(102) In *Schrödinger Release 2016-1*; 4.3 ed.; Schrödinger, LLC: New York, NY, 2016.
(103) Jacobson, M. P. F., R.A.; Xiang, Z.; Honig, B. *Journal of Molecular Biology* **2002**, *320*, 597.
(104) Jacobson, M. P. P., D. L.; Rapp, C. S.; Day, T. J. F.; Honig, B.; Shaw, D. E.; Friesner, R. A. *Proteins: Structure, Function and Bioinformatics* **2004**, *55*, 351.
(105) Halgren, T. A. *Journal of Chemical Information and Modeling* **2009**, *49*, 377.
(106) Dennington, R. K., Todd; Millam, John. ; 5.0.9 ed.; Semichem Inc.: Shawnee Mission, KS, 2009.
(107) Sinnokrot, M. O.; Valeev, E. F.; Sherrill, C. D. *Journal of the American Chemical Society* **2002**, *124*, 10887.

List of Tables

unrestrained MD. The MM-GBSA binding free energy values were computed every 0.8 ns, while the normal mode analysis was performed every 50 ns. Each simulation was run for 1.0 μs.

Table 10. The RMSD of **3** in each of the Receptor Structures compared to **3** from the Structure 1.

Table 11. The average binding affinity, ΔG_{bind} of **3** derived from MM-GBSA computations. Normal mode entropy analysis was performed on the simulations shown in bold.

Table 12. The effect of the inclusion of entropy on binding affinity rankings for select CCR5 and CXCR4 complexes with **3**. These simulations were selected for normal mode entropy analysis as they had the most favorable binding free energies prior to the inclusion of entropic effects.

Table 13. The average pairwise per residue free energy decomposition values (kcal/mol) of Trp86, Tyr108, and Phe109 on CCR5 with **3**. The moiety on **3** that interacts with each residue via π-stacking is shown. For calibration, pairwise decomposition values between CCR5 and **3** for all interactions (not just π-stacking) ranged from ~ -15 - 0 kcal/mol. The average binding energy for each CCR5 complex is also shown here.

Table 14. The per residue free energy decomposition values (kcal/mol) of the residues that produced a significantly positive increase binding free energy in the CCR5 complex with **3**, along with their respective electrostatic and polar solvation energy terms.

Table 15. The average binding affinity, ΔG_{bind}, of **3.1** in each of the receptor poses.

Table 16. The per residue free energy decomposition values (kcal/mol) from simulations of **3.1** with CCR5/CXCR4. Shown are residues that produced a significantly positive increase in the binding free energy, along with the respective electrostatic and polar solvation energy terms.

List of Figures

Figure 8. Cluster analysis of all 39 binary APP:JIP1 simulations performed using *cpptraj* from the AmberTools 14 package.

Figure 9. APP:JIP1 binary complexes used to generate Sets 2 and 3 of the KLC1:JIP1:APP ternary complexes.

Figure 10. The average structure of each ternary KLC1:JIP1:APP complex was calculated. (a) Set 1 KLC1:JIP1:APP 1 Seed 1 (b) Set 1 KLC1:JIP1:APP 2 Seed 1 (c) Set 1 KLC1:JIP1:APP 3 Seed 1 (d) Set 1 KLC1:JIP1:APP 1 Seed 2 (e) Set 1 KLC1:JIP1:APP 2 Seed 2 (f) Set 1 KLC1:JIP1:APP 3 Seed 2. Teal = KLC1, Red = JIP1, Blue = APP

Figure 11. The total, normalized hydrogen bonds against time for the following protein pairs were calculated for (A) Set 1 KLC1:JIP1:APP 1 Seed 1 (weakest average binding affinity) and (B) Set 1 KLC1:JIP1:APP 2 Seed 1 (strongest average binding affinity) and are displayed in a left to right fashion: KLC1 and APP, KLC1 and JIP1, and JIP1 and APP.

Figure 12. Six residues in the ternary KLC1:JIP1:APP complex were determined to be major contributors to the overall binding free energy after analyzing each residue's individual per residue free energy decomposition graph. (A) The six residues are Arg556, Pro557, and Leu560 on the JIP1 and Ile3, His4, and His20 on the APP. (B) After alanine scanning was performed for each of these six amino acids, Arg556 on the JIP1 and His20 on the APP showed a major increase in negative contribution to the overall binding affinity. Only partial forms of the APP and KLC1 are shown in (A) and (B). (C) All residues within 2.5 Å of Arg556 and His20 (highlighted in yellow) are shown. (D) All residues within 3 Å of Arg556 and His20 (highlighted in yellow) are shown. KLC1 = cyan, JIP1 = red, APP = blue.

Figure 13. Asp273, Ser274, Pro275, Thr278, and Thr279 act as core residues on KLC1 that interact with His20 on the APP. (A) The His20 of the APP in Set 1 KLC1:JIP1:APP 1 Seed 1 only had signficant interactions with the core KLC1 residues. (B) The His20 of the APP in Set 1 KLC1:JIP1:APP 2 Seed 1 only had significant interactions with Asp273, Thr278, and Arg306 on the KLC1. The Arg306 on the KLC1 forced the tail of the APP to wrap back towards the JIP1 molecule due to its conformation being above that of the His20 on the APP.

Figure 14. Compound **3** was originally synthesized by the Liotta group, while **3.1** was conceived in this study, replacing the γ-carbon of the piperidine ring of **3** with a second protonated nitrogen.

Figure 15. The sequences of each protein utilized for molecular dynamics simulations. Residues that were added during refinement are bolded and underlined. The active site residues are

highlighted in red, all negatively charged residues within the active site are highlighted in blue, and the negatively charged residues that produce a positive per residue free energy decomposition value (*vida infra*) are colored green. The active site residues are defined as those residues located within the 10 x 10 x 10 Å Glide docking grid centered on the binding site within the extracellular domain of each receptor. Since the binding sites determined by SITEMAP were different between the two CXCR4 experimental structures, the docking grid contained different sets of residues, leading to different residues highlighted for CXCR4:CVX15 and CXCR4:IT1t.

Figure 16. Parallel displaced π-stacking interaction between Trp86 and the pyridine ring of **3**. This image is at 156.2 ns, with a ΔG_{bind} of -63.4 kcal/mol, the lowest binding free energy observed in CCR5 Structure 1 Seed 2.

Figure 17. Electrostatic interaction between Glu283 and the protonated nitrogen of the piperidine ring of **3**. This image is the same CCR5 Structure 1 Seed 2 trajectory snapshot as in Figure 16.

Figure 18. The binding free energy of each **3.1** inhibited complex throughout the entirety of the 250 ns simulation.

Supplemental Information

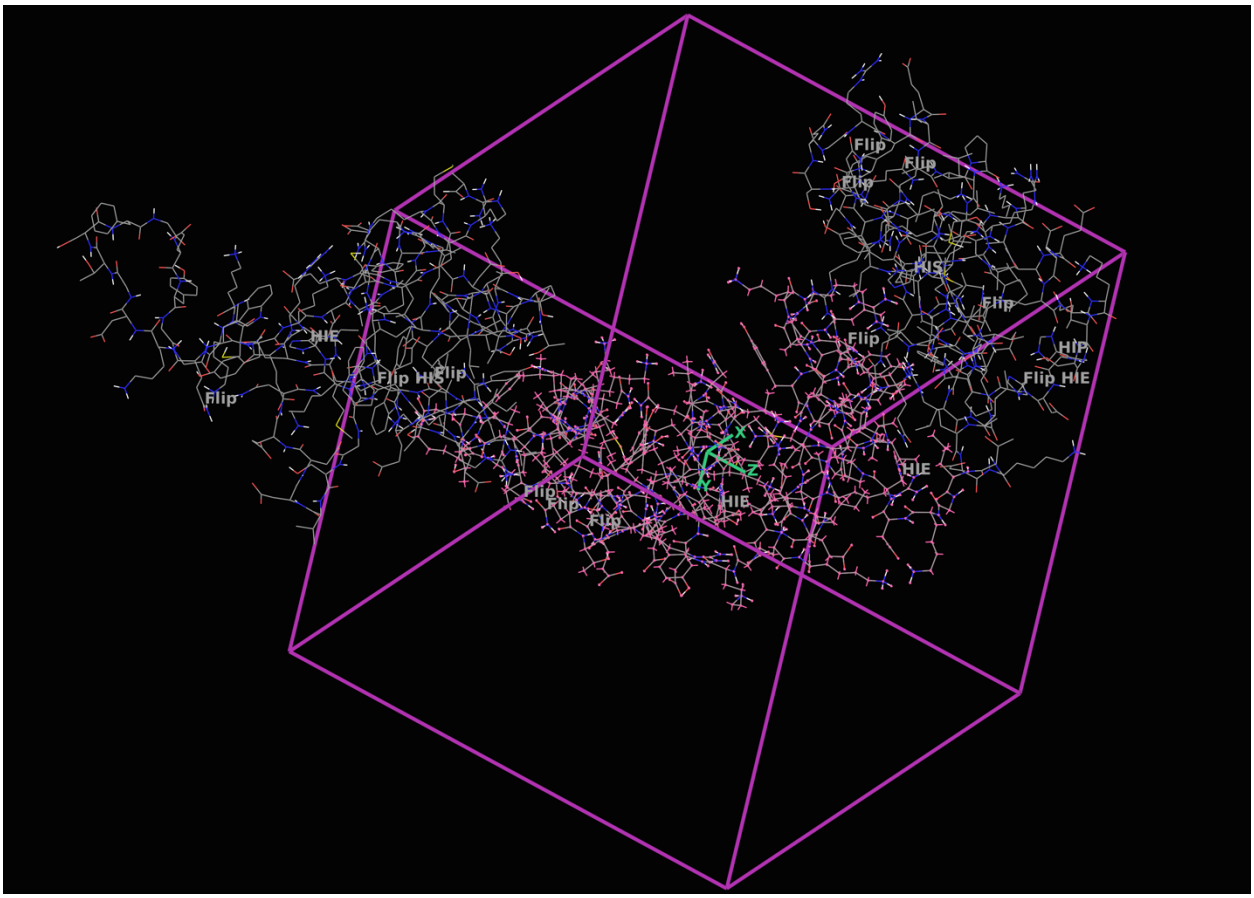

Fig S1. The grid box used when docking JIP-1 to polar patch of KLC-1 using the Glide program.

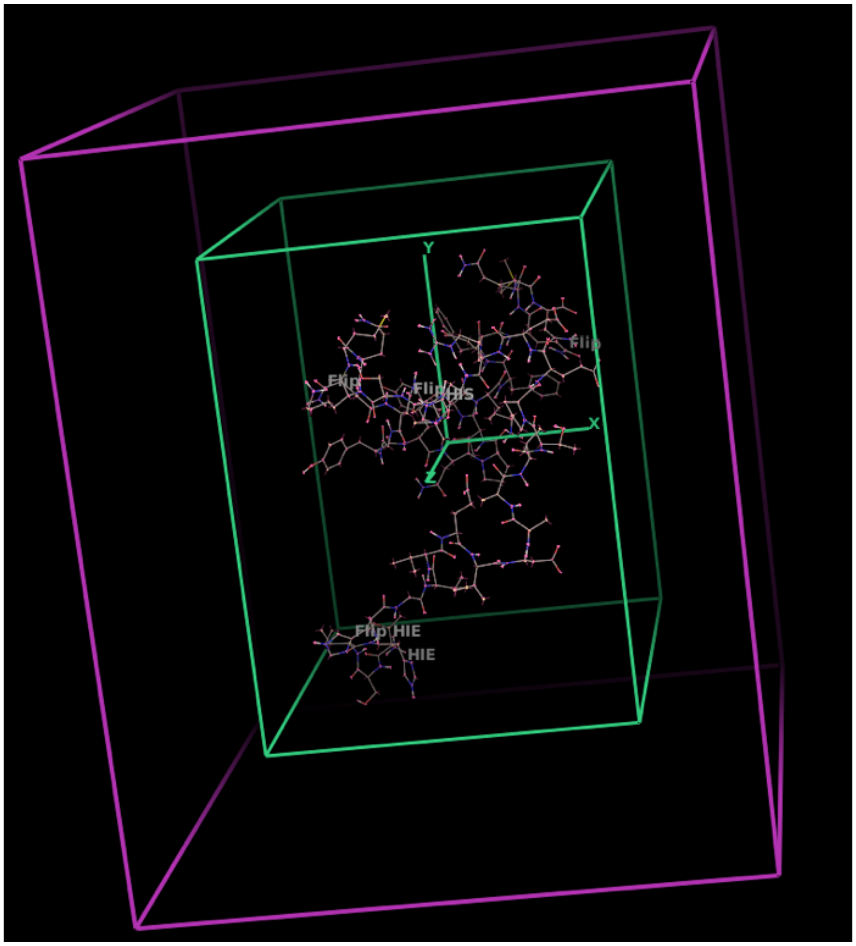

Fig S2. The 30x40x40 grid box used when docking JIP-1 to APP using the Glide program.

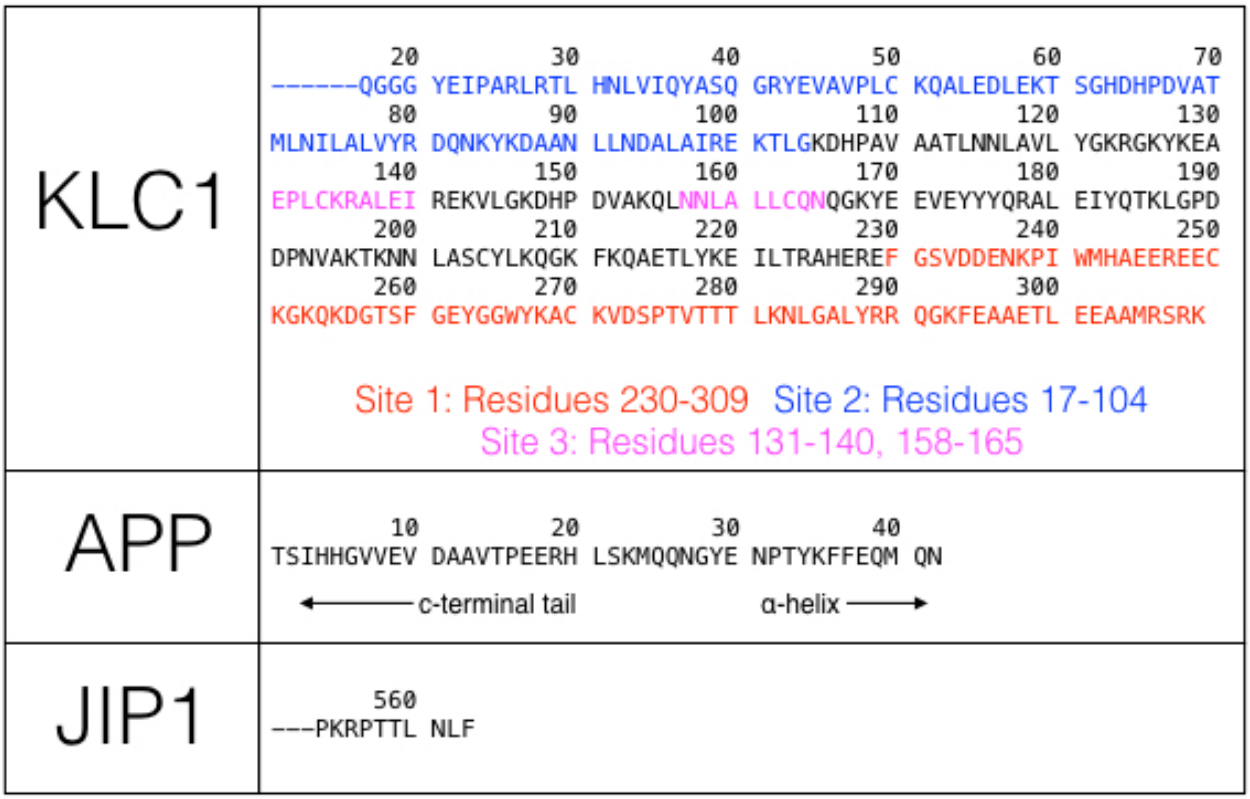

Fig S3. The FASTA sequence for each protein (KLC1, APP, and JIP1) is given here. The KLC1 sequence is highlighted in various sections to denote that those specified residues make up one of the calculated binding sites. The general alpha helix and c-terminal tail regions are denoted underneath the APP sequence. The numbering of each residue is in accordance to the respective protein's PDB numbering (excluding the APP).

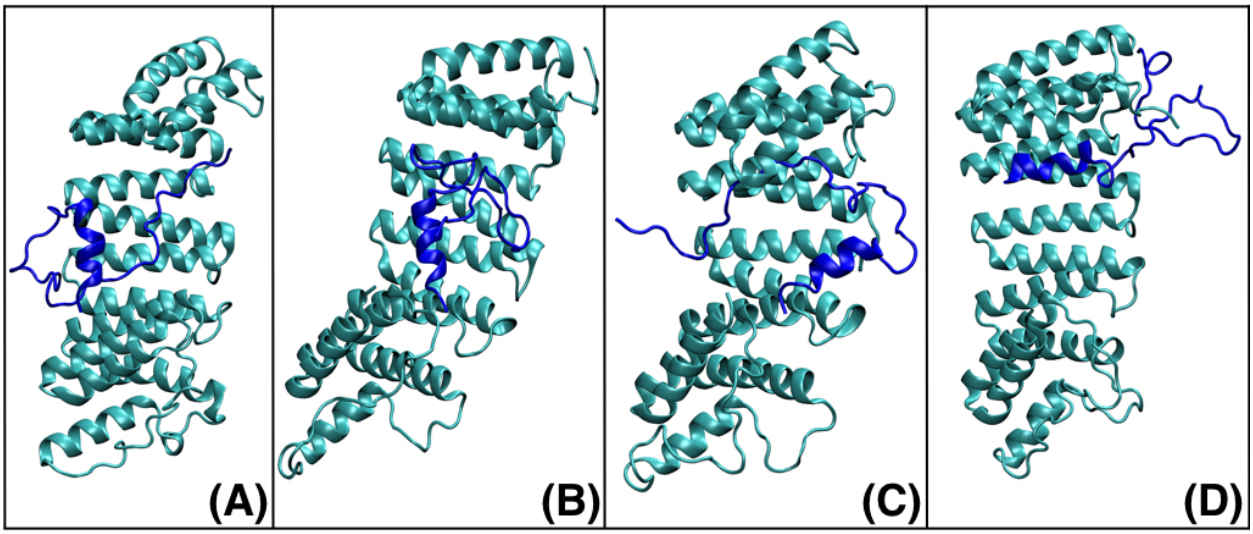

Fig S4. The starting structures of each KLC1:APP Pose. Pose 1 (A), Pose 2 (B), Pose 3 (C), and Pose 4 (D) were all docked using GRAMM-X.

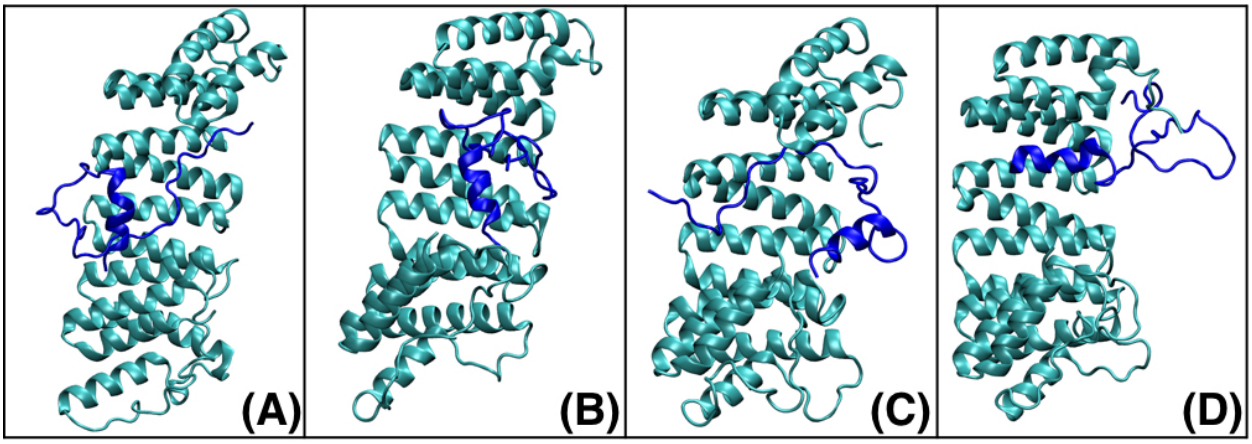

Fig S5. The final conformations of each KLC1:APP Pose. Pose 1 (A), Pose 2 (B), Pose 3 (C), and Pose 4 (D) were all docked using GRAMM-X.

Observations corroborated by both visual inspection and free energy decomposition results showed non-polar interactions between Ile3 on APP and Ile35, Ala38, and Ile74 on KLC1 Site 2 helped stabilize the APP tail for roughly the first 150 ns of the simulation. As these nonpolar interactions weakened (around 130 ns), two oxygen atoms on APP Asn42 flipped towards Arg290 on the KLC1, generating intermolecular interactions that partially stabilize the APP α-helix. KLC1:APP Pose 1 showed partial stabilization of the APP throughout the simulation, but there was no prolonged period in the trajectory in which both the α-helix and tail of APP had stabilizing interactions with KLC1. The relatively more favorable binding energy for Pose 2 is due in large part to the interactions between Glu17 on APP and KLC1 Site 2 residues Asn116, Arg99, and Asn73 that persisted throughout the simulation. The APP α-helix was also stabilized by residue Asn42 interacting with Arg228 on the KLC1. This stabilizing interaction developed around 150 ns, and persisted for the remainder of the simulation.

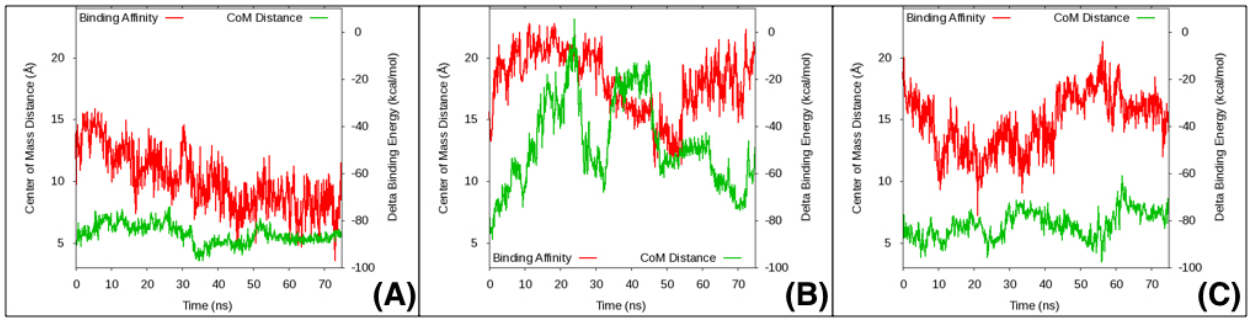

Fig S6. The distance between the center of mass of each the JIP-1 and the elongated loop region (Site 1) of the KLC1 was calculated for each of the Site 1 Structures. It can be seen that the center of mass of the JIP1 in Structures 1 (A) and 2 (C) is much closer to that of Site 1, helping to lower the overall binding affinity. The center of mass of the JIP1 in Structure 2 (B) fluctuates heavily in relation to that of Site 1, and the binding affinity is diminished because of said fluctuation.

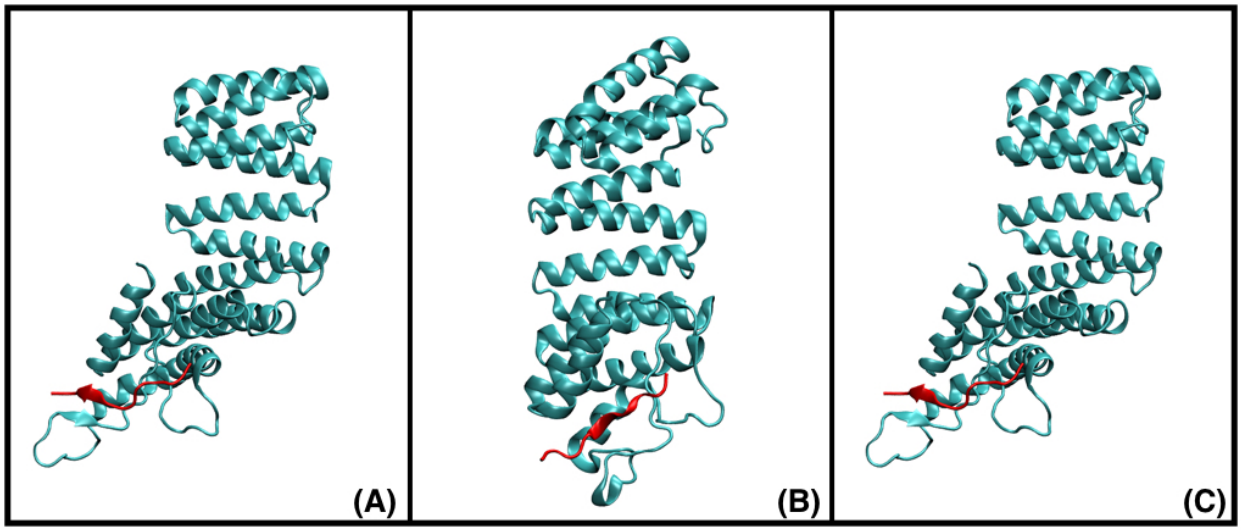

Fig S7. The starting structures of each KLC1:JIP1 Site 1 pose. Pose 1 (A), Pose 2 (B), and Pose 3 (C) were all docked using GRAMM-X.

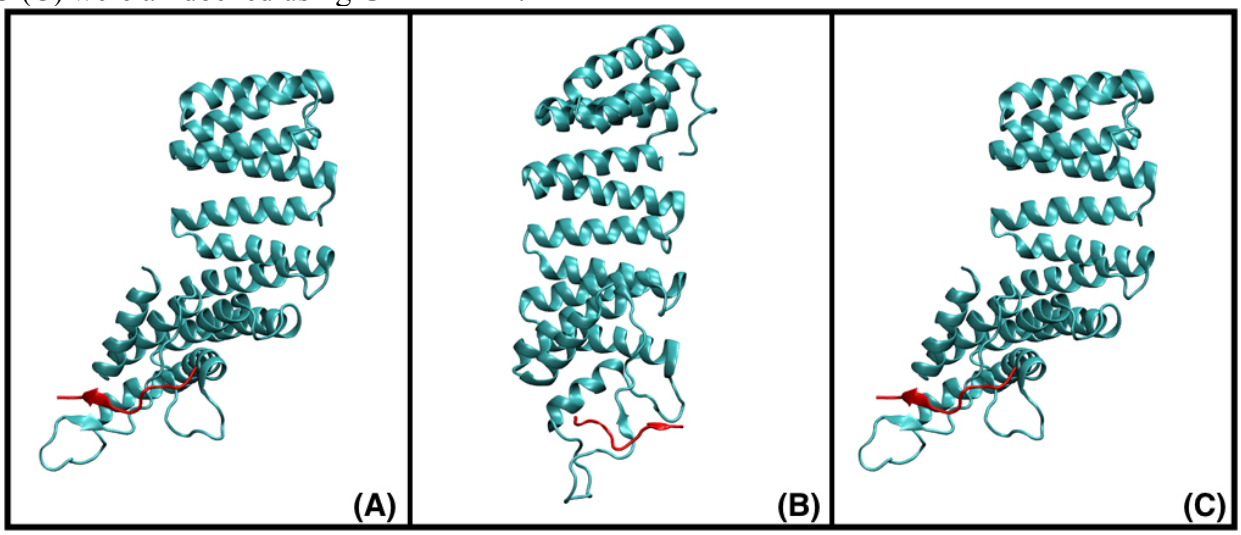

Fig S8. The final structures of each KLC1:JIP1 Site 1 pose. Pose 1 (A), Pose 2 (B), and Pose 3 (C) were all docked using GRAMM-X.

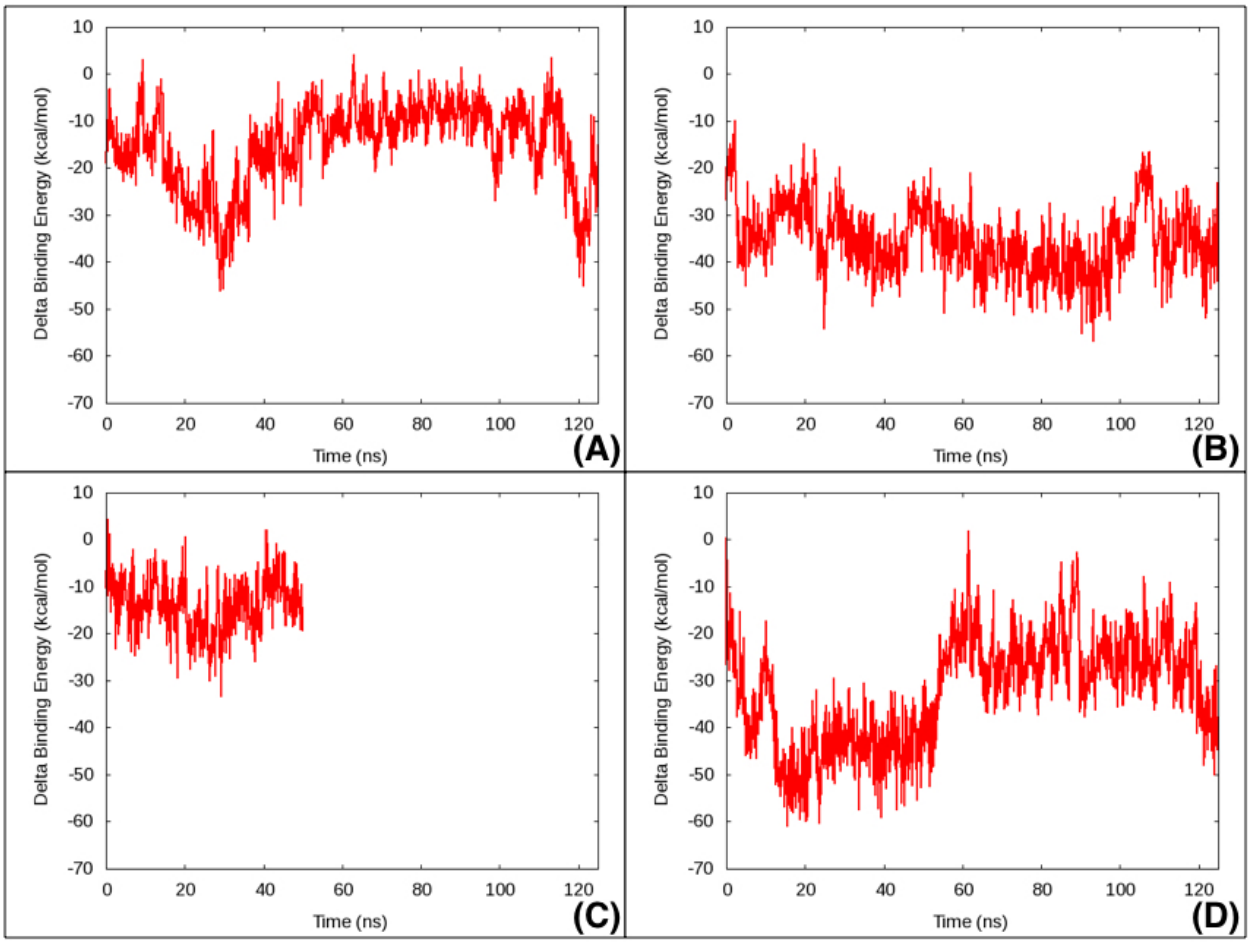

Fig S9. The ΔG_{bind} against time for KLC1:JIP1 *Site 2* Structure 2 (A), 4 (B), 6 (C), and 8 (D). ΔG_{bind} was calculated every 0.1 ns. Since Structure 6 showed no major negative peaks in binding affinity, the simulation was not continued after 50 ns of unrestrained MD.

Significant KLC1:JIP1 Site 2 interactions within higher energy complexes

The trajectory initiated from KLC1:JIP1 Site 2 Structure 2 reveals the JIP1 to be very dynamic for the first ~70 ns of the simulation. The relatively low binding affinity obtained over the course of this simulation can be attributed to this dynamic behavior dominating the majority of the trajectory. However, once JIP1 Asn561 created a hydrogen bonding network between its side chain oxygen and the two side chain nitrogen atoms of KLC1 Arg124, the complex stabilizes as JIP1 was pulled into Site 2 (Fig S11). This interaction oriented JIP1 for further hydrogen bonding between the JIP1 Asn561 backbone oxygen atom and the side chain nitrogen of Lys123 (ΔG_{decomp} = -3.6 kcal/mol). On the N-terminal of the JIP1 peptide, the nitrogen atom within the side chain of Phe554 was then able to create an alternating hydrogen bond network between the side chain of Glu44 (ΔG_{decomp} = -1.0 kcal/mol) and backbone of Gly41 (ΔG_{decomp} = -0.4 kcal/mol) on the KLC1. These are the same KLC1 residues that stabilized this end of JIP1 through hydrogen bonding in Structure 4, except the JIP1 utilized its arginine in Structure 4 instead of the terminal proline residue, resulting in a difference in ΔG_{decomp} of -15.2 kcal/mol for the Glu44 interaction and -5.4 kcal/mol for the Gly41 interaction. The weak binding affinity of KLC1:JIP1 Site 2 Structure 6 can be attributed to the failure to maintain any strong hydrogen bond networks. The strongest binding experienced by this system (-25.9 kcal/mol) occurred when there were a

few hydrogen bonds between the JIP1 and KLC1. In summary, the strongest binding affinities for the Site 2 structures were obtained when both ends of the JIP1 were involved in hydrogen bonding, as well as at least one central residue of the JIP1 peptide.

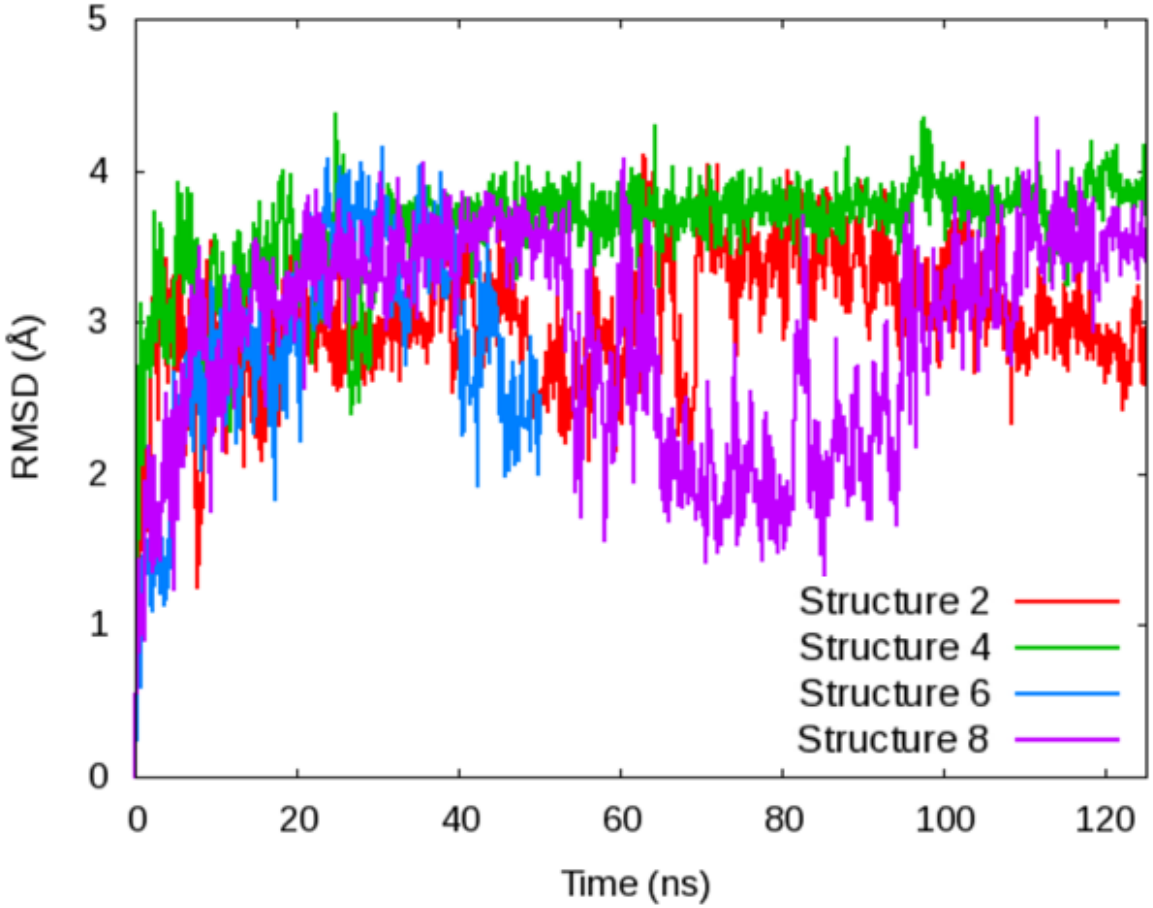

Fig S10. The RMSD (against the respective starting structure) of the JIP1 peptide for KLC1:JIP1 *Site 2* Structures 2, 4, 6, and 8. The fluctuation of more than 2.0 Å shows the JIP-1's sampling of the potential energy surface instead of remaining in a potential energy well.

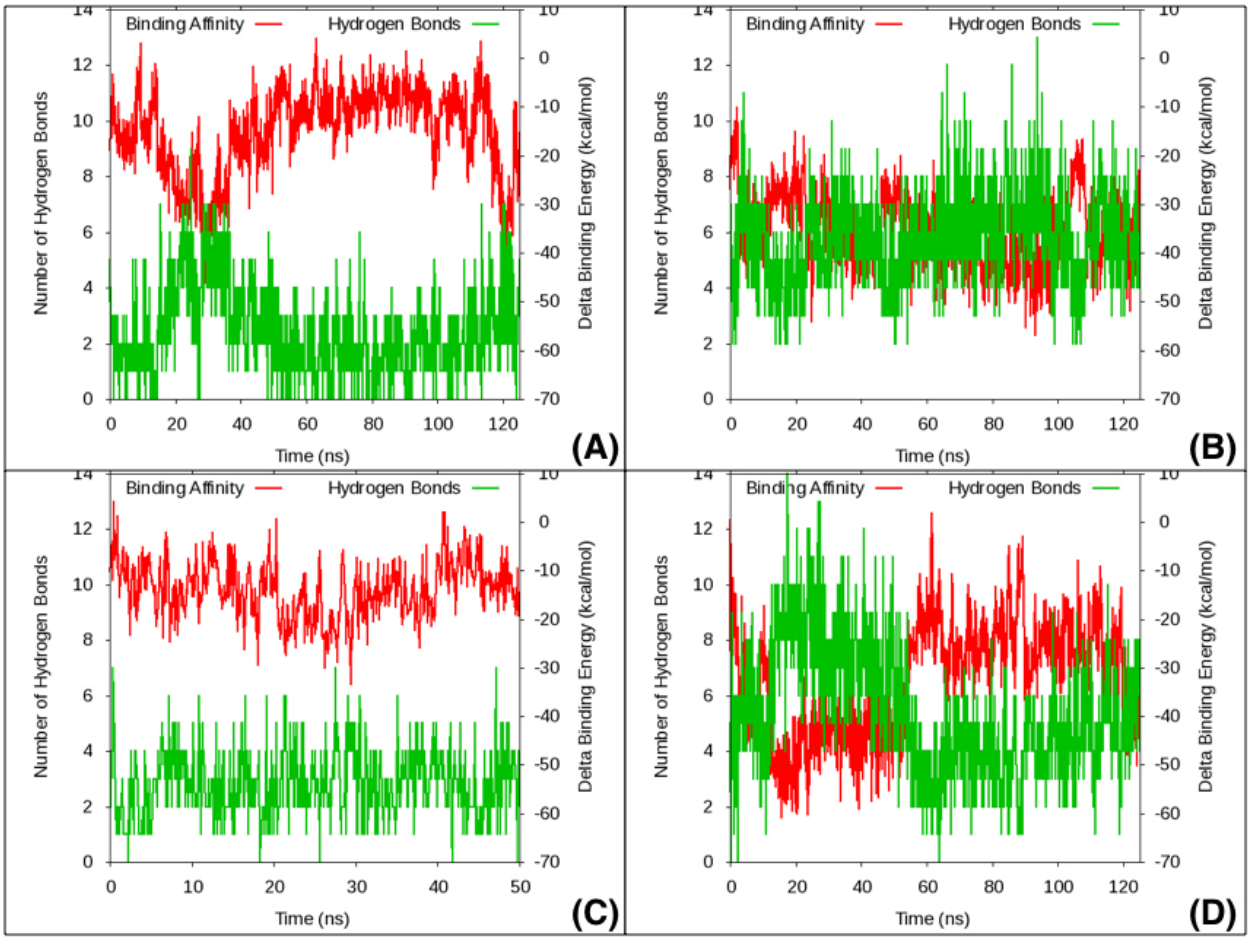

Fig S11. The ΔG_{bind} against time as well as the number of hydrogen bonds being made between the KLC1 and JIP1 for KLC1:JIP1 *Site 2* Structure 2 (A), 4 (B), 6 (C), and 8 (D). Negative spikes in binding affinity are witnessed each time the number of hydrogen bonds increase, while the binding affinity is weakened as number of hydrogen bonds lessens.

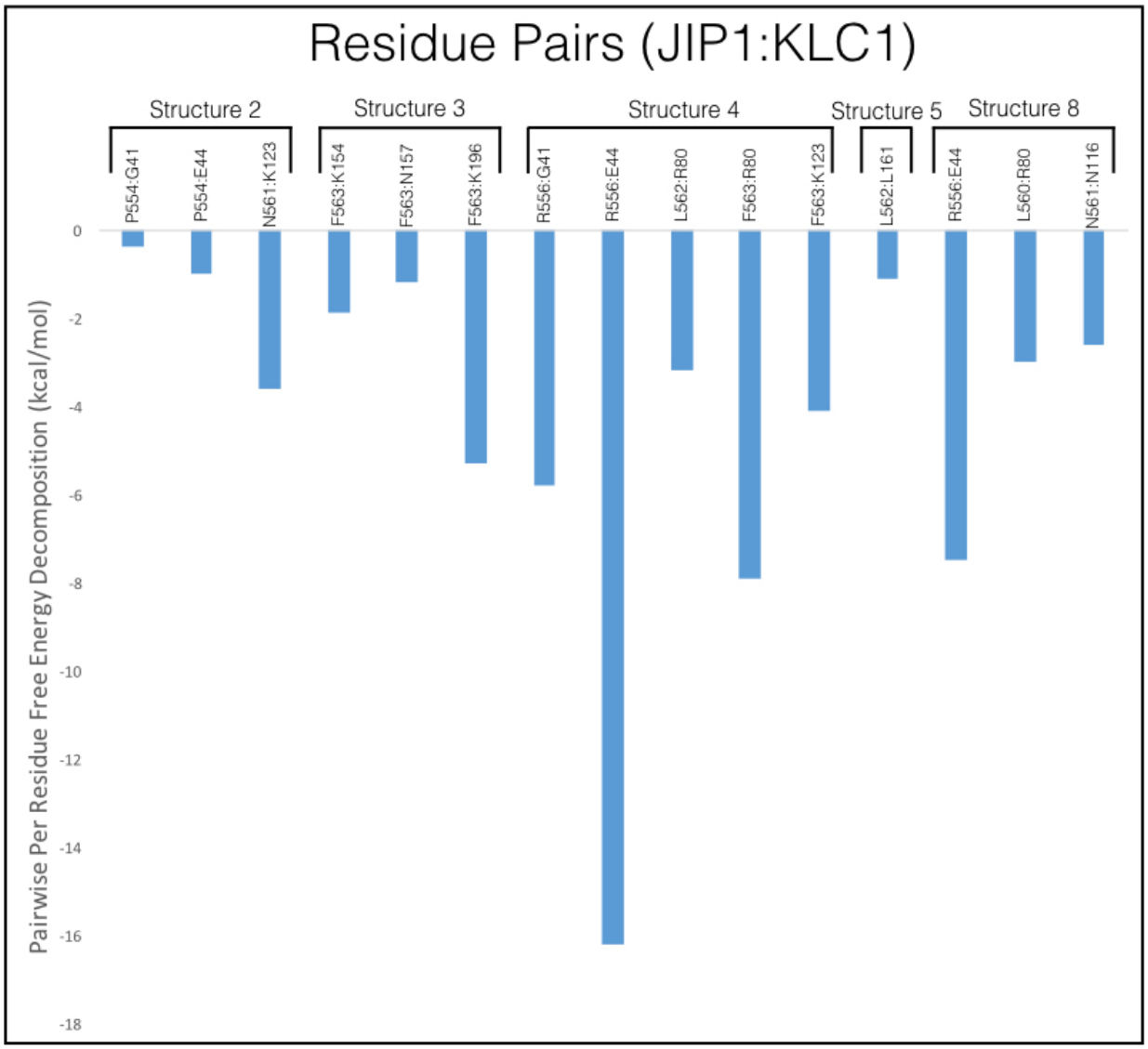

Fig S12. All important pairwise per residue free energy decomposition interactions between the KLC1:JIP1 structures as described in the main text. The syntax used for the pairs of residues is JIP1:KLC1 (i.e. P554:G41 means it is the pairwise ΔG_{decomp} between P554 on JIP1 and G41 on KLC1).

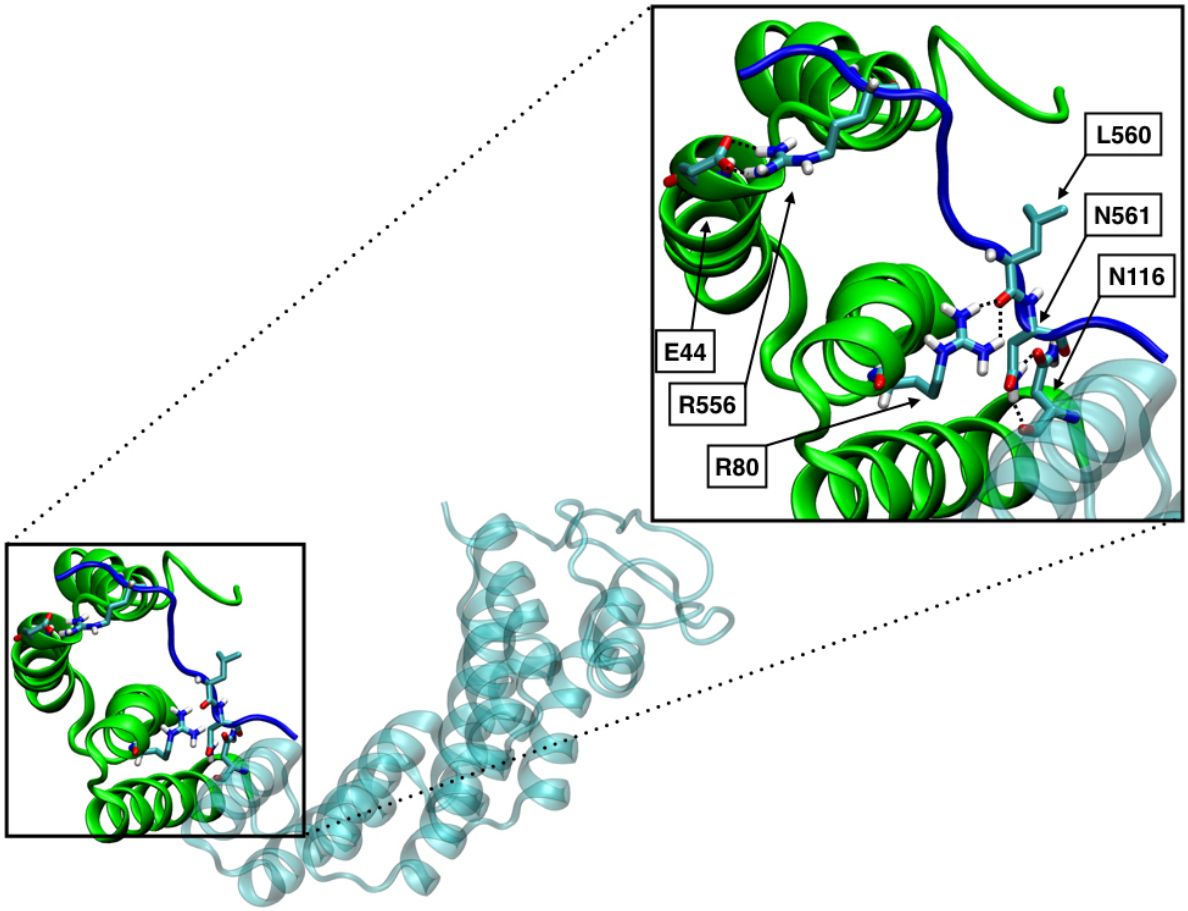

Fig S13. The dominant hydrogen bond interactions in KLC1:JIP1 *Site 2* Structure 8 as determined by pairwise decomposition energies. The conformation imaged here is at 15.6 ns due to its having the strongest binding affinity (-60.9 kcal/mol). This depiction is representative of the simulation since these binding motifs were found in 5-30% of the simulation. Hydrogen bonds are represented by dotted lines between the hydrogen bond pairs. Site 2, composed of TPRs 1 and 2, is highlighted in green.

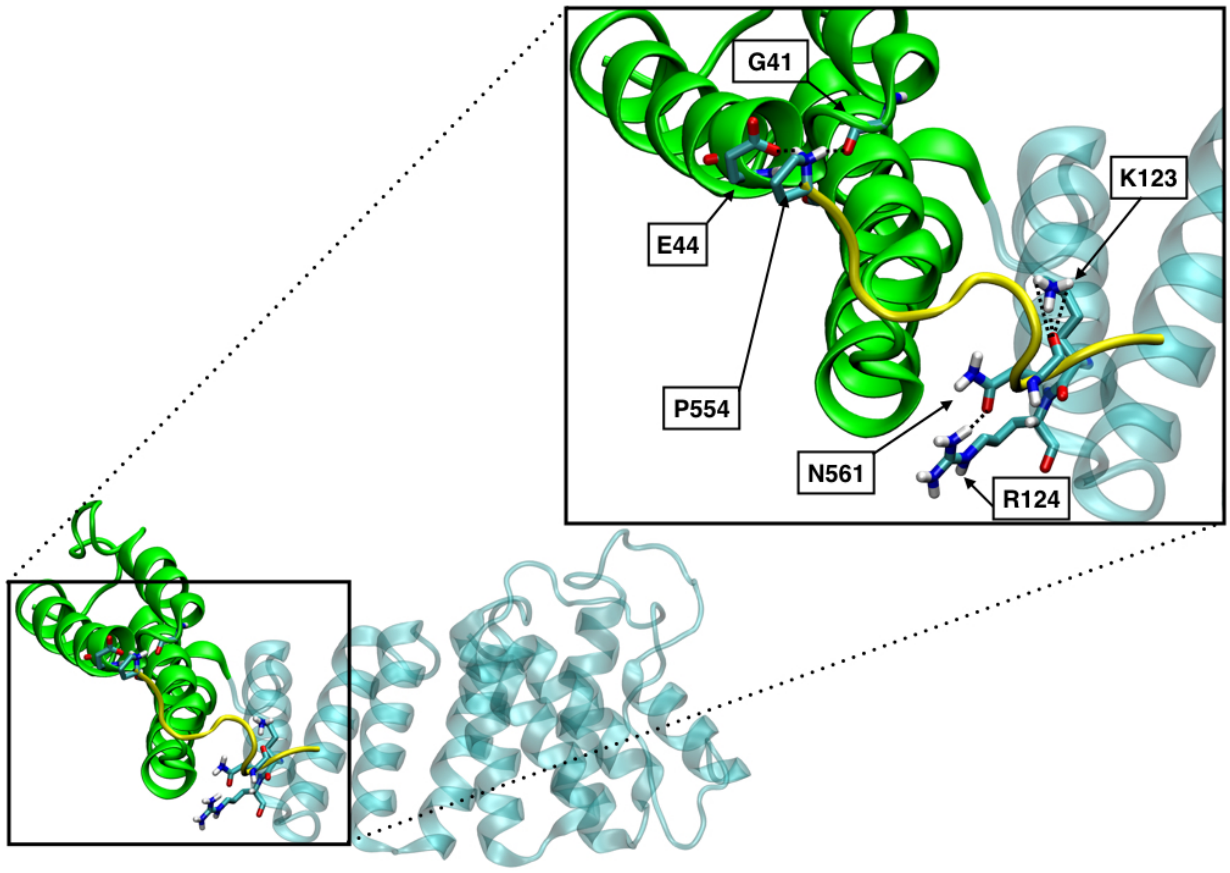

Fig S14. The dominant hydrogen bond interactions in KLC1:JIP1 *Site 2* Structure 2 as determined by pairwise decomposition energies. The conformation imaged here is at 121.3 ns due to its having the strongest binding affinity (-45.1 kcal/mol) within Site 2. This depiction is representative of the simulation since these binding motifs were found in 3-15% of the simulation. Hydrogen bonds are represented by dotted lines between the hydrogen bond pairs. Site 2, composed of TPRs 1 and 2, is highlighted in green.

Binary KLC1:JIP1 polar patch not stable

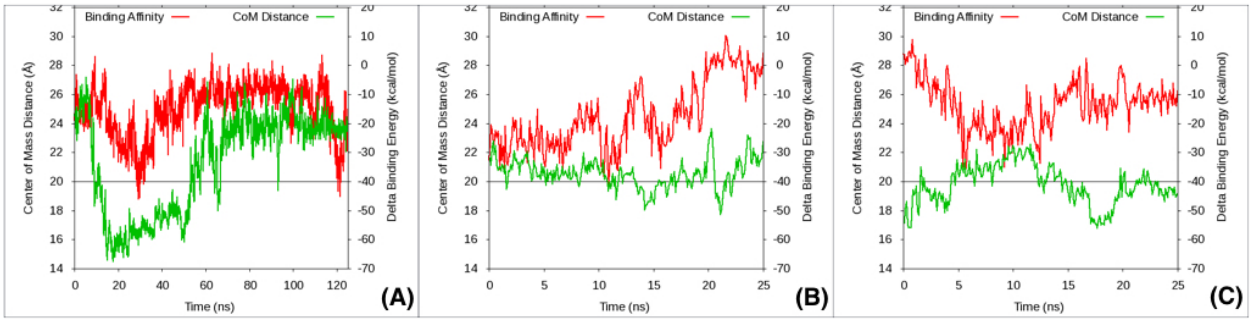

Fig S15. The graphs show the distance between the JIP1 and KLC1 "polar patch's" center of mass against time, as well as the binding energy against time. The black horizontal line is representative of the 20 Å cutoff of what is considered to be within the "polar patch." (a) In KLC1:JIP1 *Site 2* Structure 2, the center of mass of JIP1 is seen to drop below 20 Å away from the center of mass of the "polar patch" for a large part of the simulation, but the JIP1 had

wrapped around to the backside of this area, which is not within the defined constraints of the "polar patch" region (b) KLC1:JIP1 *Site 2* Structure 3 (c) KLC1:JIP1 *Site 2* Structure 5.

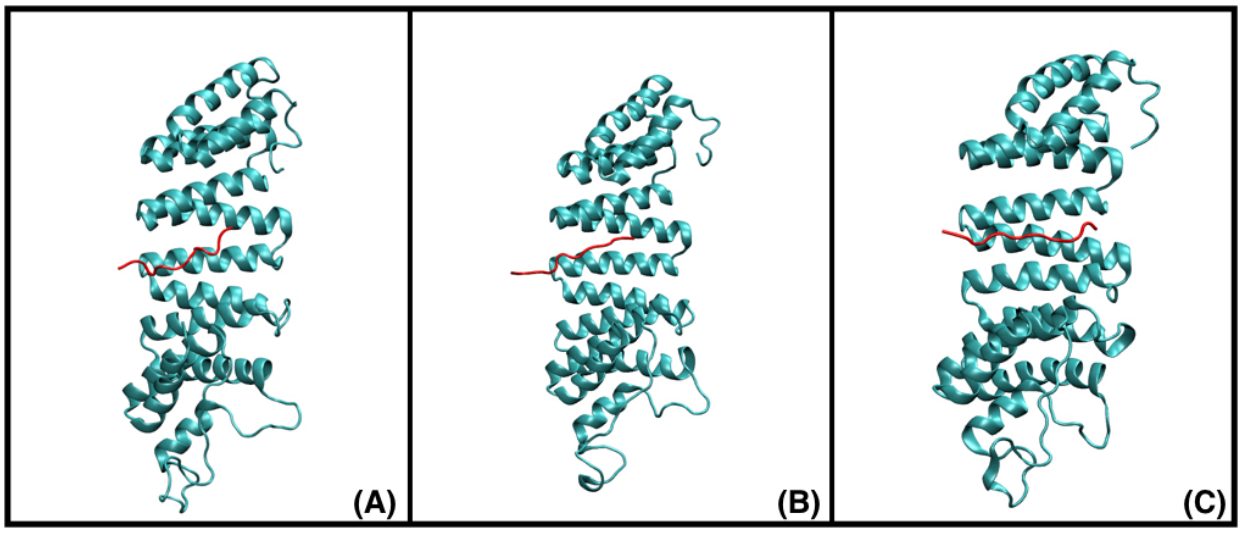

Fig S16. The starting structures of each KLC1:JIP1 Site 3 pose that remained within the polar patch. Glide Pose 3 (A), GRAMM-X Pose 1 (B), and Glide Pose 1 (C) are shown in strongest to weakest binding affinity order and were the only three Site 3 simulations in which the JIP1 remained in the polar patch.

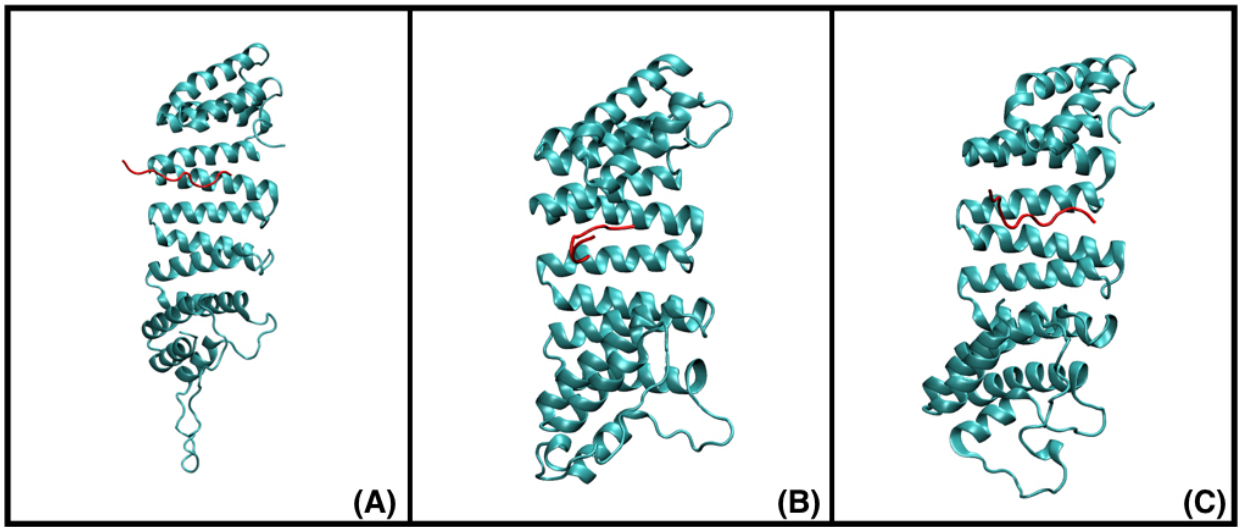

Fig S17. The final structures of each KLC1:JIP1 Site 3 pose that remained within the polar patch. Glide Pose 3 (A), GRAMM-X Pose 1 (B), and Glide Pose 1 (C) are shown in strongest to weakest binding affinity order and were the only three Site 3 simulations in which the JIP1 remained in the polar patch.

To better understand the relative lack of JIP1 affinity for the polar patch region, we investigated further the trajectories started from Structures 3 and 5. The Structure 3 trajectory interacted with the polar patch region for less than 10 ns (~27% of the simulation), and the only interaction was a weak hydrogen bond between the backbone of the terminal Phe563 on JIP1 and Asn157 on TPR 4 of the KLC1 (ΔG_{decomp} = -1.2 kcal/mol). This hydrogen bond network was

quickly broken because Lys154 (ΔG_{decomp} = -1.9 kcal/mol) and Lys196 (ΔG_{decomp} = -5.3 kcal/mol) on TPR 4 and 5, respectively, formed a new hydrogen bond network with the backbone of Phe563, displacing Asn157 as JIP1 migrated away from the polar patch region. For the trajectory initiated from Structure 5, the only significantly favorable interaction with Site 3 occurred early in the simulation as the starting structure contained non-polar interactions between Leu562 on JIP1 and Leu161 on KLC1 (ΔG_{decomp} = -1.1 kcal/mol). This interaction was quickly extinguished due to the steric clash of the Asn561 side chain on JIP1 with the side chain of Lys154 on KLC1, which forced JIP1 out of Site 3. As seen in Fig S12, JIP1 dipped back within the 20 Å cutoff distance between 15 and 20 ns, but this was once again an instance of the weak non-polar interactions that were lost once the steric clash returned.

JIP1 converged to similar conformation

Table S1. The RMSD was calculated for all pairs of the average structures of each of the four valid KLC1:JIP1.

Structure	Glide Docked 2	Glide Docked 4	Glide Docked 6	Glide Docked 8
Glide Docked 2	----------	1.184 Å	0.923 Å	1.950 Å
Glide Docked 4	1.184 Å	----------	1.304 Å	2.184 Å
Glide Docked 6	0.923 Å	1.304 Å	----------	2.567 Å
Glide Docked 8	1.950 Å	2.184 Å	2.567 Å	----------

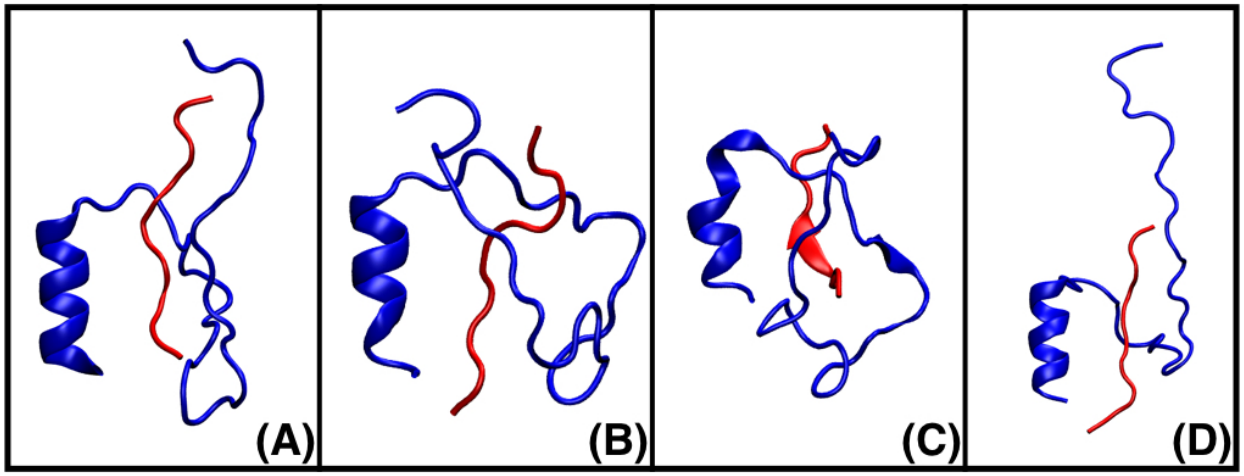

Fig S18. The starting structures of the four lowest energy APP:JIP1 simulations. GRAMM-X Pose 15 (A), Glide Pose 8 (B), Glide Pose 7 (C), and Glide Pose 12 are shown in strongest to weakest binding affinity order.

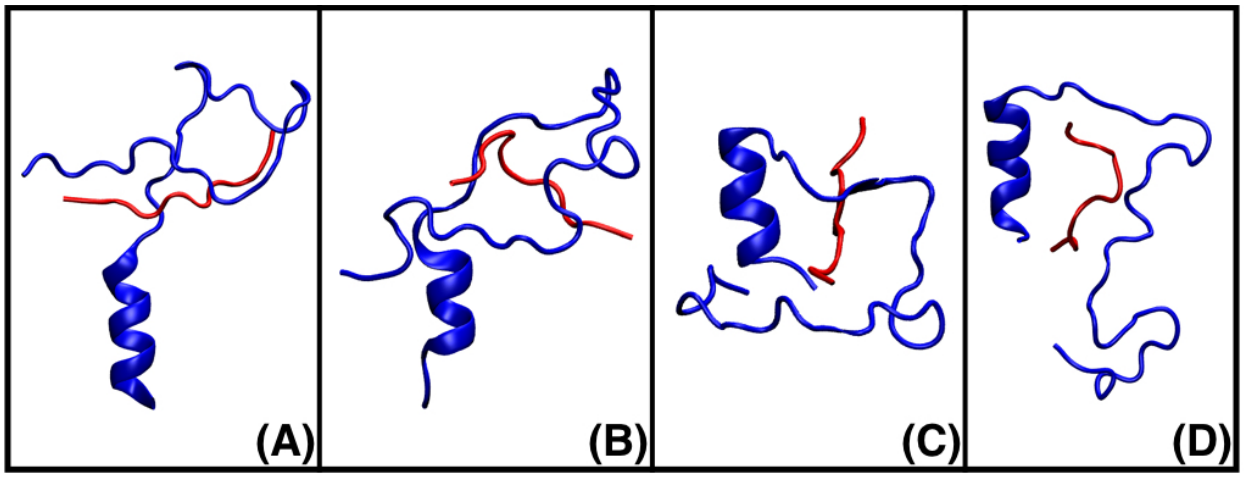

Fig S19. The final structures of the four lowest energy APP:JIP1 simulations. GRAMM-X Pose 15 (A), Glide Pose 8 (B), Glide Pose 7 (C), and Glide Pose 12 are shown in strongest to weakest binding affinity order.

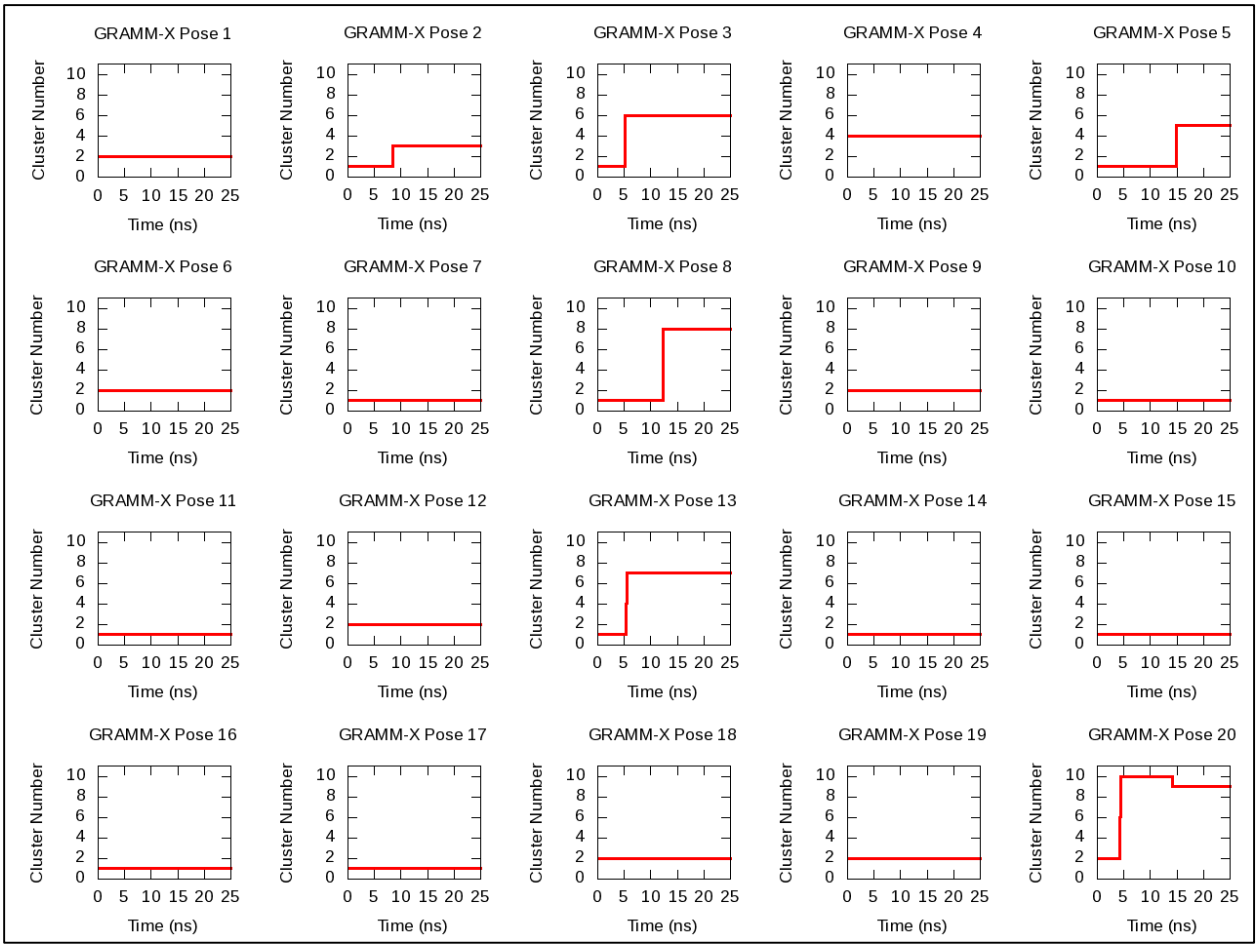

Fig S20. The graphs show the Cluster ID of each APP:JIP1 complex docked by GRAMM-X as a function of time, suggesting convergence for the majority of the simulations.

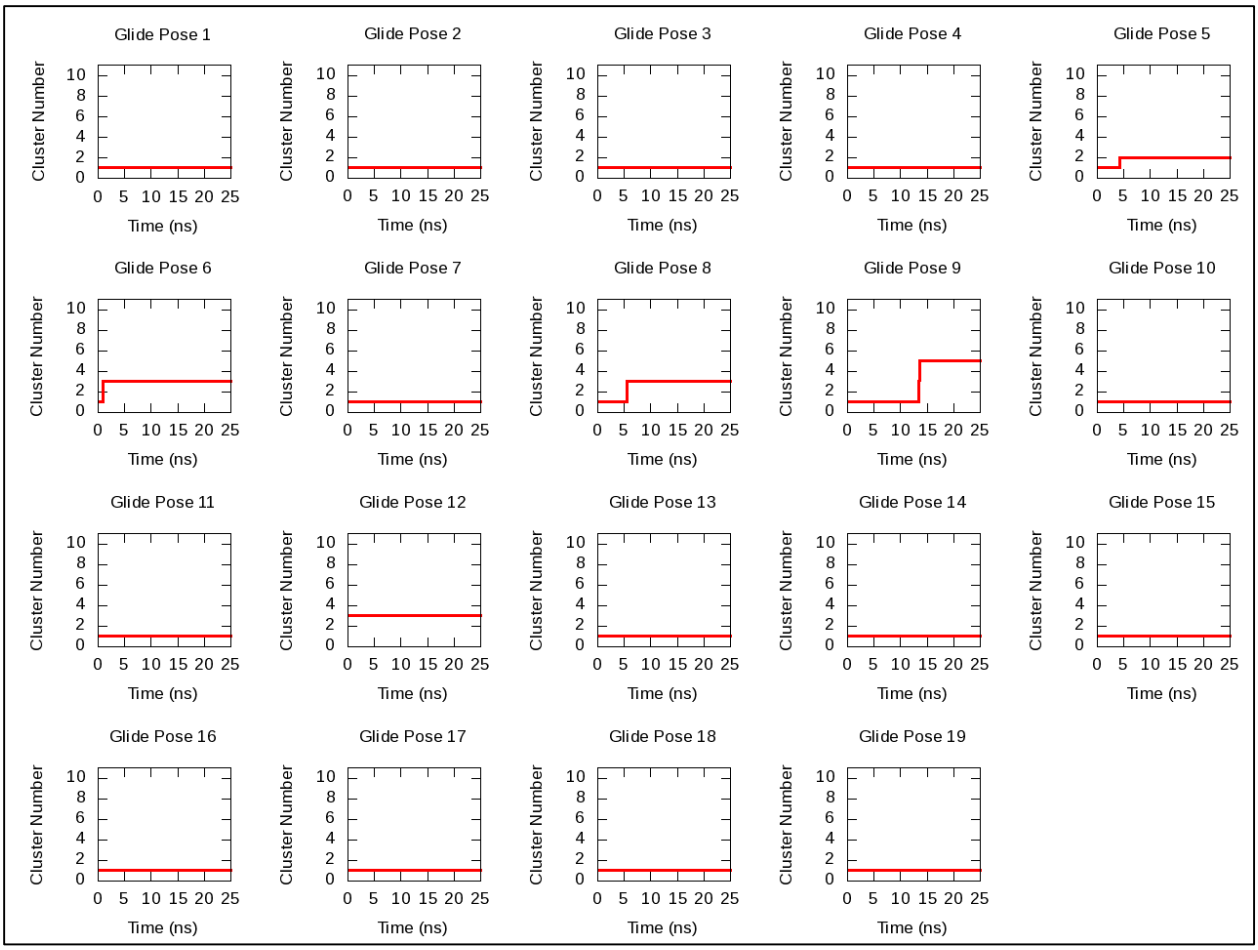

Fig S21. The graphs show the Cluster ID of each APP:JIP1 complex docked by Glide as a function of time, suggesting convergence for the majority of the simulations.

Table S2. Further atomistic details of each starting conformation for each simulation that was ran. This table includes information regarding the total number of intermolecular hydrogen bonds, salt bridges, the total buried hydrophobic surface area, as well as the total contact surface area between the two proteins in each complex.

Complex	Simulation	Number of Intermolecular H-bonds	Number of Intermolecular Salt Bridges	Buried Hydrophobic Surface Area Å^2	Total Contact Surface Area Å^2
KLC:APP	1	3	1	1238.71	1271.69
	2	5	0	1064.58	954.38
	3	9	1	1496.90	1580.05
	4	3	2	1263.96	1321.56
KLC1:JIP1 Site 1 GRAMM-X	1	3	0	445.68	2005.08
	2	2	0	355.92	1925.29
	3	2	0	354.74	1978.18
KLC1:JIP1 Site 2 Glide	1	6	0	234.00	1751.42
	2	5	0	379.35	1858.00
	3	8	0	324.23	1938.41
	4	6	0	234.00	1751.42
	5	1	0	514.30	1817.16
	6	2	0	504.73	1811.24
	7	4	0	206.08	1706.10
	8	5	0	313.26	1719.16
	9	1	0	395.76	1719.35
	10	2	0	400.82	1743.64
KLC1:JIP1 Site 3 (WT) glide	1	0	0	433.99	1963.37
	2	4	0	268.68	1788.13
	3	4	0	506.95	1826.00
	4	1	0	690.46	1923.84
KLC1:JIP1 Site 3 (WT) GRAMM-X	1	3	0	676.33	2237.73
	2	0	0	787.78	2173.50
	3	0	0	363.93	2121.67
KLC1:JIP1 Site 3 (mut) glide	1	3	0	363.93	1941.69
	2	6	0	564.20	1745.18

	3	2	0	368.98	1863.54
	4	4	0	429.05	1742.39
	5	1	0	345.21	1788.41
	6	1	0	149.79	1580.88
	7	5	0	220.87	1799.32
	8	1	0	285.82	1735.78
	9	4	0	149.23	1583.09
	10	4	0	209.93	1839.79
	11	2	0	124.73	1481.95
KLC:JIP1 Site 3 (mut) GRAMM-X	1	2	0	792.63	2436.87
	2	3	0	731.64	2118.19
	3	3	0	347.93	2082.22
APP:JIP1 Glide	1	3	0	376.98	658.64
	2	5	2	477.07	627.71
	3	6	1	448.17	638.79
	4	1	0	386.22	713.88
	5	4	1	373.70	582.71
	6	6	0	405.80	689.16
	7	4	0	385.10	659.85
	8	6	0	483.51	641.69
	9	3	0	435.49	611.15
	10	5	0	391.94	601.50
	11	8	0	365.98	681.12
	12	6	1	420.63	790.40
	13	4	0	411.94	715.05
	14	6	1	355.33	531.32
	15	5	0	439.10	505.83
	16	7	0	361.87	572.19
	17	6	1	310.55	548.23
	18	11	0	330.77	499.49
	19	8	0	430.03	608.40
APP:JIP1 GRAMM-X	1	1	0	458.38	477.97
	2	2	0	365.02	517.16
	3	6	0	235.86	393.86
	4	0	0	390.45	473.45
	5	7	1	522.59	658.76
	6	6	0	593.89	610.29
	7	5	0	521.18	484.64
	8	2	0	556.50	424.68

	9	1	0	459.23	430.73
	10	1	0	471.84	482.01
	11	1	0	389.53	501.70
	12	3	0	560.42	569.81
	13	3	0	382.64	431.44
	14	6	0	251.32	444.31
	15	8	0	550.70	623.33
	16	3	0	409.45	484.71
	17	3	0	390.38	543.25
	18	4	0	533.62	547.79
	19	3	0	413.52	653.85
	20	6	0	412.22	546.84

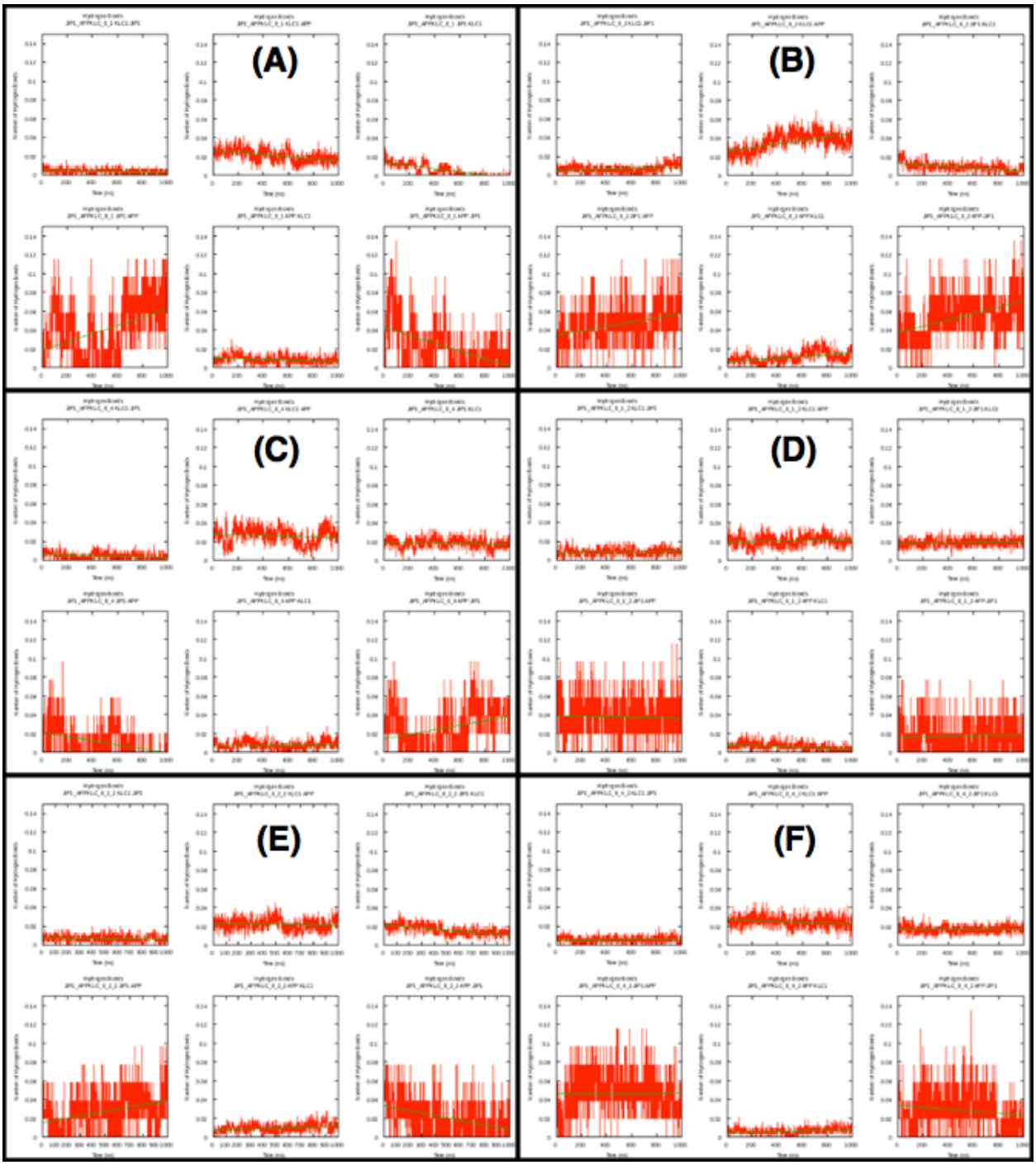

Figure S22. Normalized hydrogen-bonding analysis was performed on each pair of proteins in the fashion of "donor":"receiver" for each Set 1 KLC1:JIP1:APP complex. Each subfigure goes in the following order (from left to right, upper row to lower row): KLC1:JIP1, KLC1:APP, JIP1:KLC1, JIP:APP, APP:KLC1, APP:JIP1. (a) KLC1:JIP1:APP 1 Seed 1 (b) KLC1:JIP1:APP 2 Seed 1 (c) KLC1:JIP1:APP 4 Seed 1 (d) KLC1:JIP1:APP 1 Seed 2 (e) KLC1:JIP1:APP 2 Seed 2 (f) KLC1:JIP1:APP 4 Seed 2.

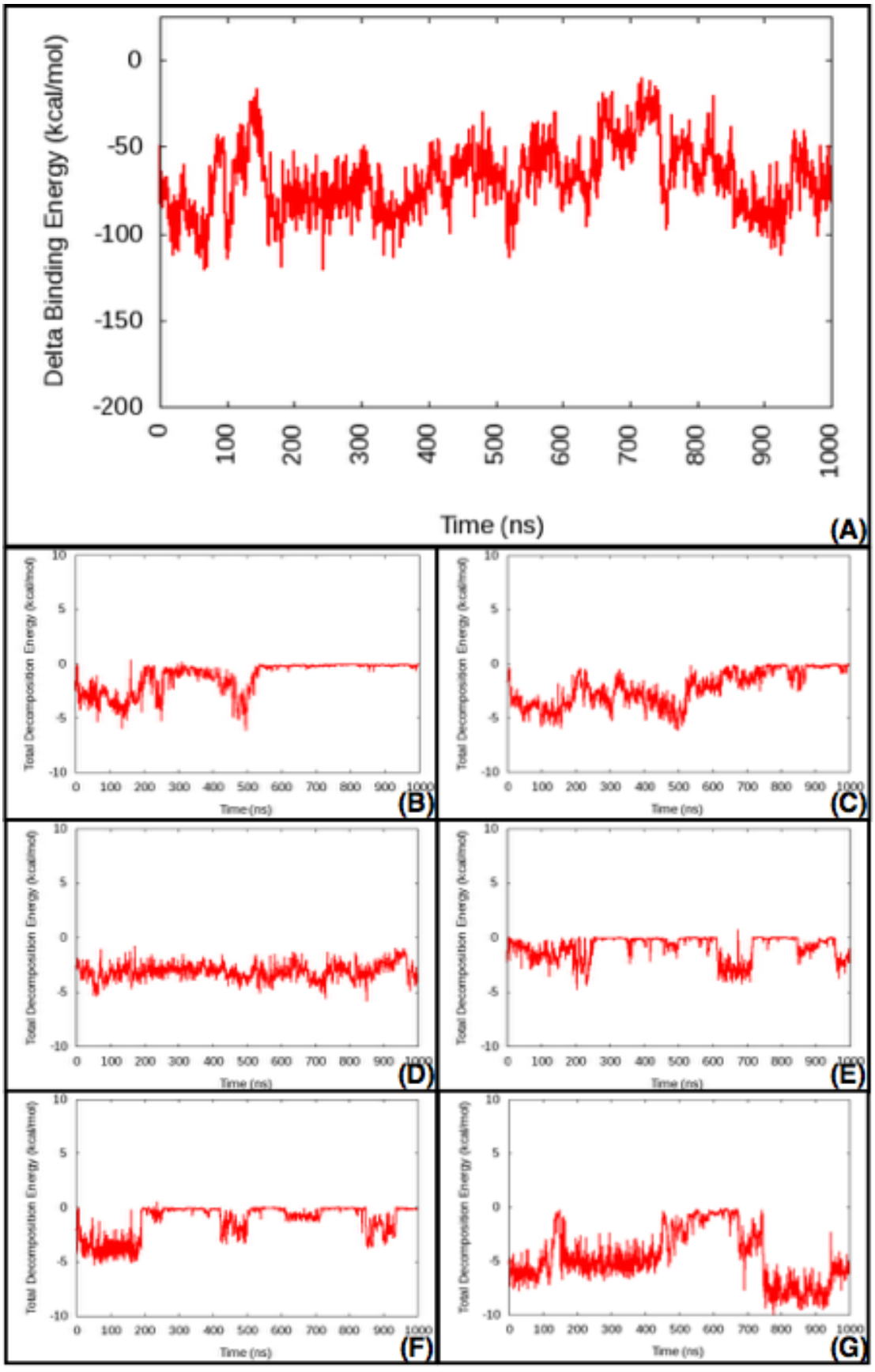
Delta Binding Energy (kcal/mol)
0
-50
-100
-150
-200
0
100
200
300
400
500
600
700
800
900
1000
Time (ns)
(A)
Total Decomposition Energy (kcal/mol)
10
5
0
-5
-10
Time (ns)
(B)
(C)
(D)
(E)
(F)
(G)

Figure S23. The per residue free energy decomposition (PRFED) graphs for each of the 345 residues in each of the six the ternary KLC1:JIP1:APP complexes were compared to their respective delta binding energy against time graph to determine level of importance. If the PRFED graph followed either a similar or close to opposite pattern as the binding affinity graph, then it was flagged as playing a large role in the determination of the binding energy. (a) The binding energy against time graph of KLC1:JIP1:APP 4 Seed 1 shows a distinct peak between 100 and 200 ns, as well as a distinct decrease in binding energy between and 750 900 ns. (b) Arg-556 in this model shows a downward spike in decomposition energy between 100 and 200 ns, so it was flagged for being important since it is opposite of what is seen in the binding energy against time graph. (c) Pro-557 shows a downward spike between 100 and 200 ns as well as an upward slope in decomposition energy in decomposition energy, both opposite of that of the binding energy graph, so it was also flagged. (d) Leu-560 failed to show any signs of importance in this model so it was not flagged. (e) Although Ile-3 shows some activity in its decomposition energy between 100 and 200 ns, it is not distinct enough to be flagged, and there is no distinct feature of the 750 to 900 ns range either. (f) Although there was a valley in the 100 to 200 ns range of the His-4, there was no major change during that time, as well as no discernable pattern in the 750 to 900 ns range, so this residue was not flagged for importance in this model. (g) His-20 was flagged for this model due to its distinct spike in decomposition energy between 100 and 200 ns, as well as its major decline in decomposition energy between 750 and 900 ns.

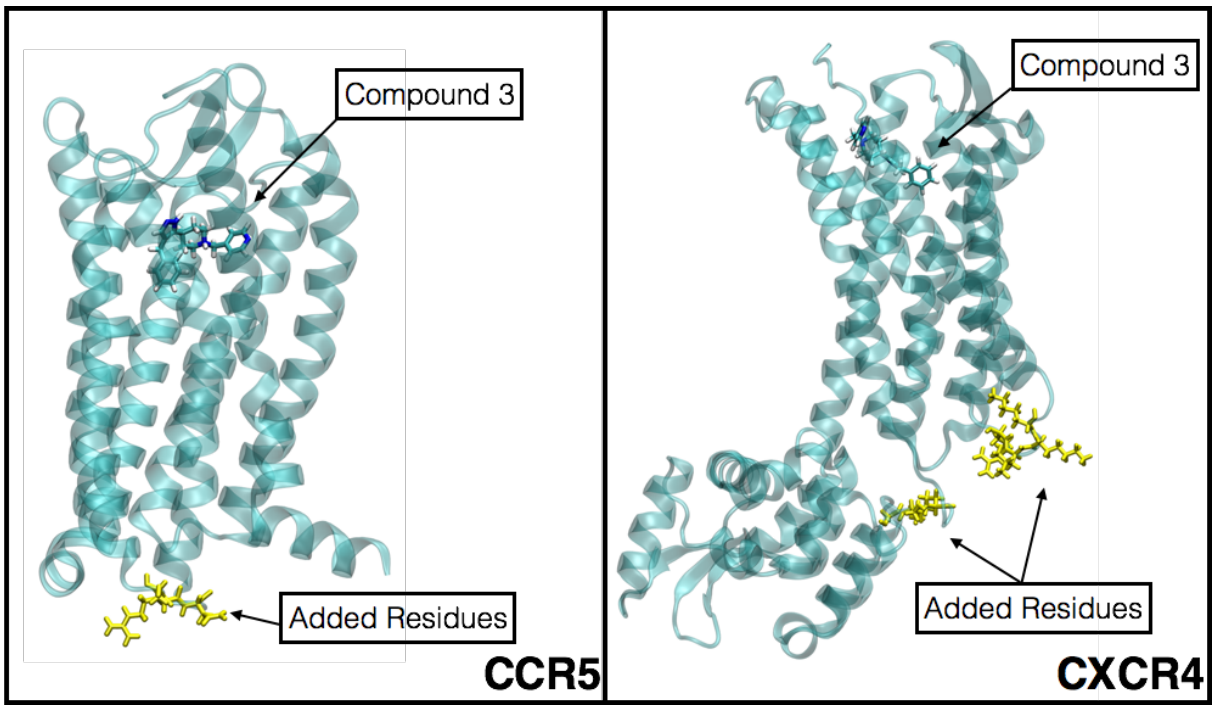

Figure S24. The crystal structures of CCR5 (PDB 4MBS) and CXCR4 (PDB 3OE0) were missing portions of the respective proteins, so Maestro was utilized to add in the missing residues. The missing residues are highlighted in yellow, and are located roughly 40 Å from 3 in CCR5 and at least 36 Å away from 3 in CXCR4, playing a negligible, to no role in the computed binding dynamics and free energies of these complexes. The added residues to CCR5 were Cys242, Arg243, and Asn244. The added residues to CXCR4 were Lys67, Lys68, Leu69, Arg70, Ser229, Lys230, Ser999, Gly1000, Ser10001, and Gly1162.

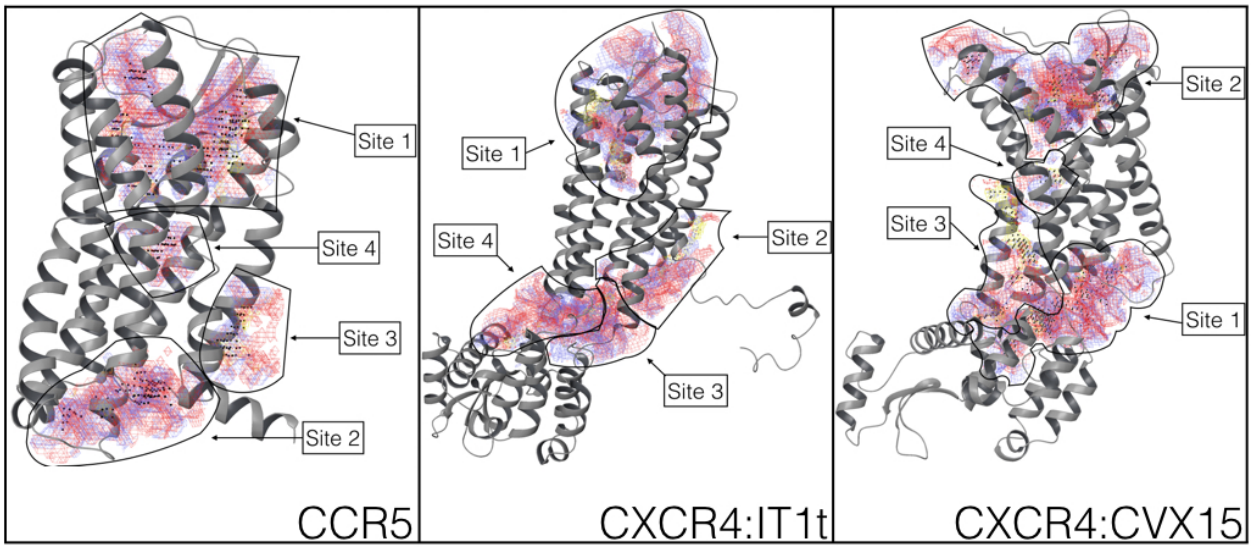

Figure S25. The binding sites as determined by the SITEMAP package from Schrodinger are depicted on their respective proteins. The extracellular region had a binding pocket in all three proteins, labeled as Site 1 for CCR5 and CXCR4:IT1t, but ranked as Site 2 for CXCR4:CVX15. The red representative of hydrogen bond acceptors, blue is hydrogen bond donors, and yellow is hydrophobic.

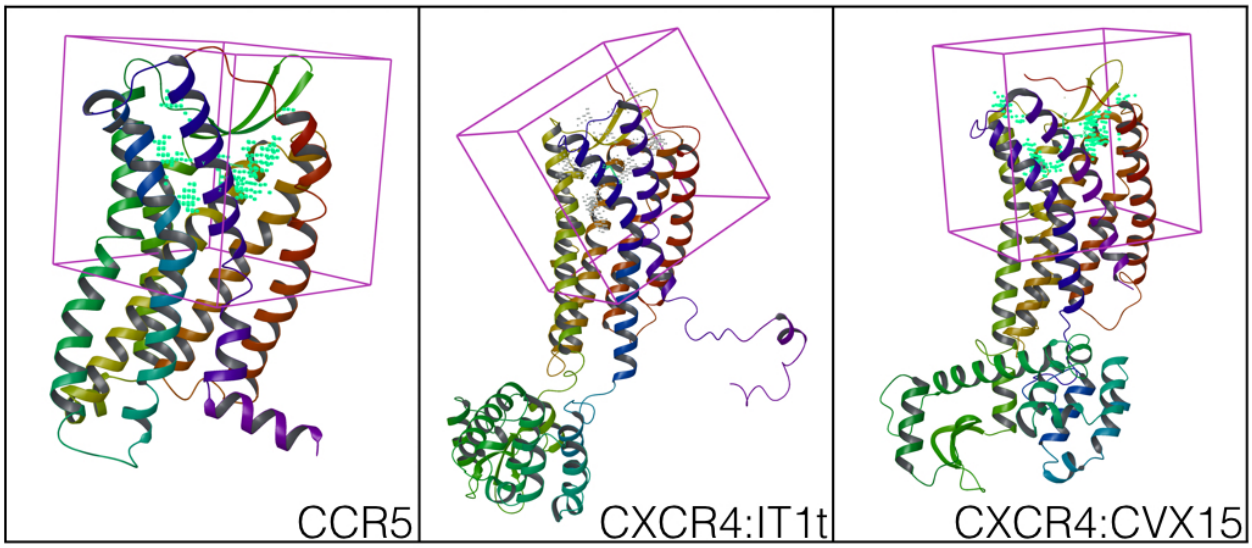

Figure S26. The Glide grids (10 x 10 x 10 Å dimensions) were centered on Site 1 for CCR5 and CXCR4:IT1t, and Site 2 for CXCR4:CVX15

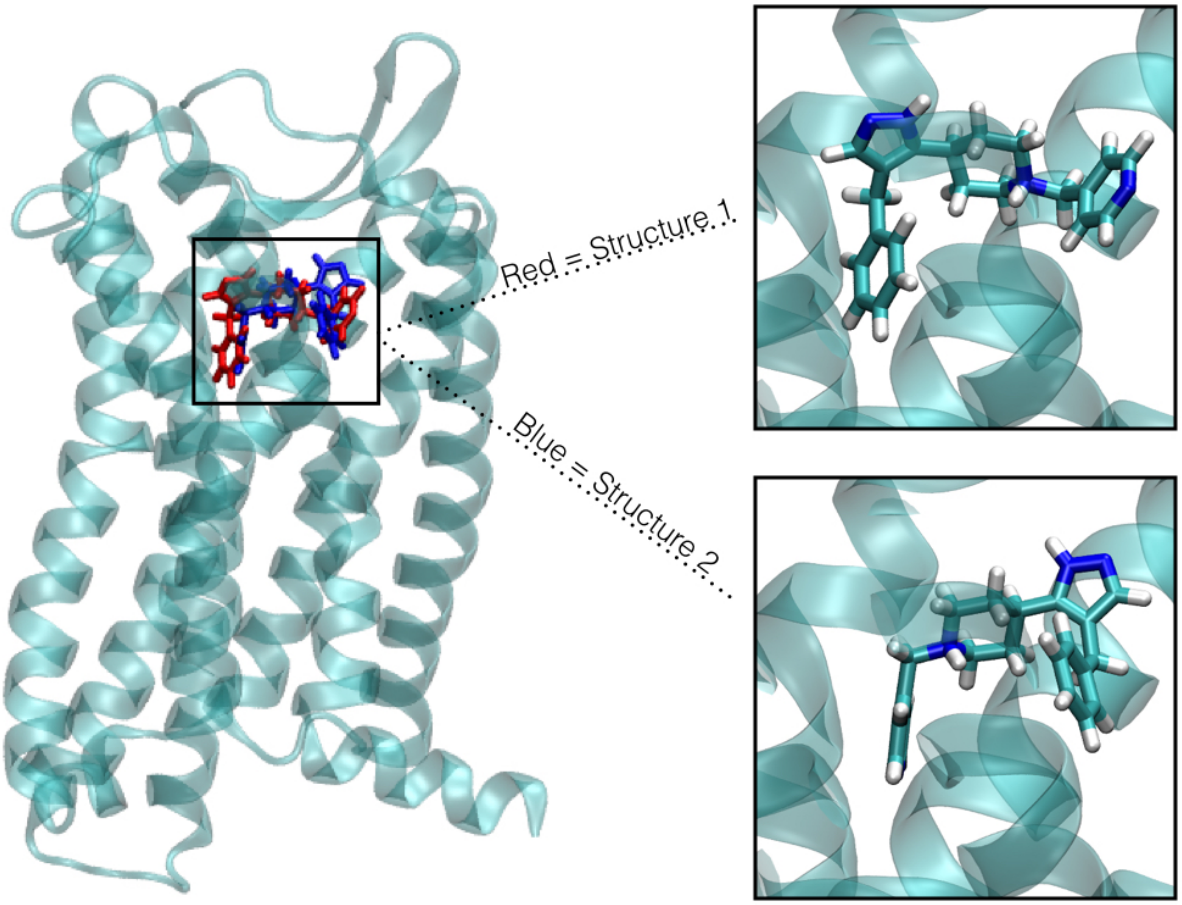

Figure S27. The flipped orientation of **3** within CCR5 Structure 1 (red ligand, top right box) and Structure 2 (blue ligand, bottom right box). It can be seen that the benzene ring of Structure 1 is "flipped" in Structure 2, located on the left from the imaged orientation in Structure 1, while being located on the right in the imaged orientation in Structure 2.

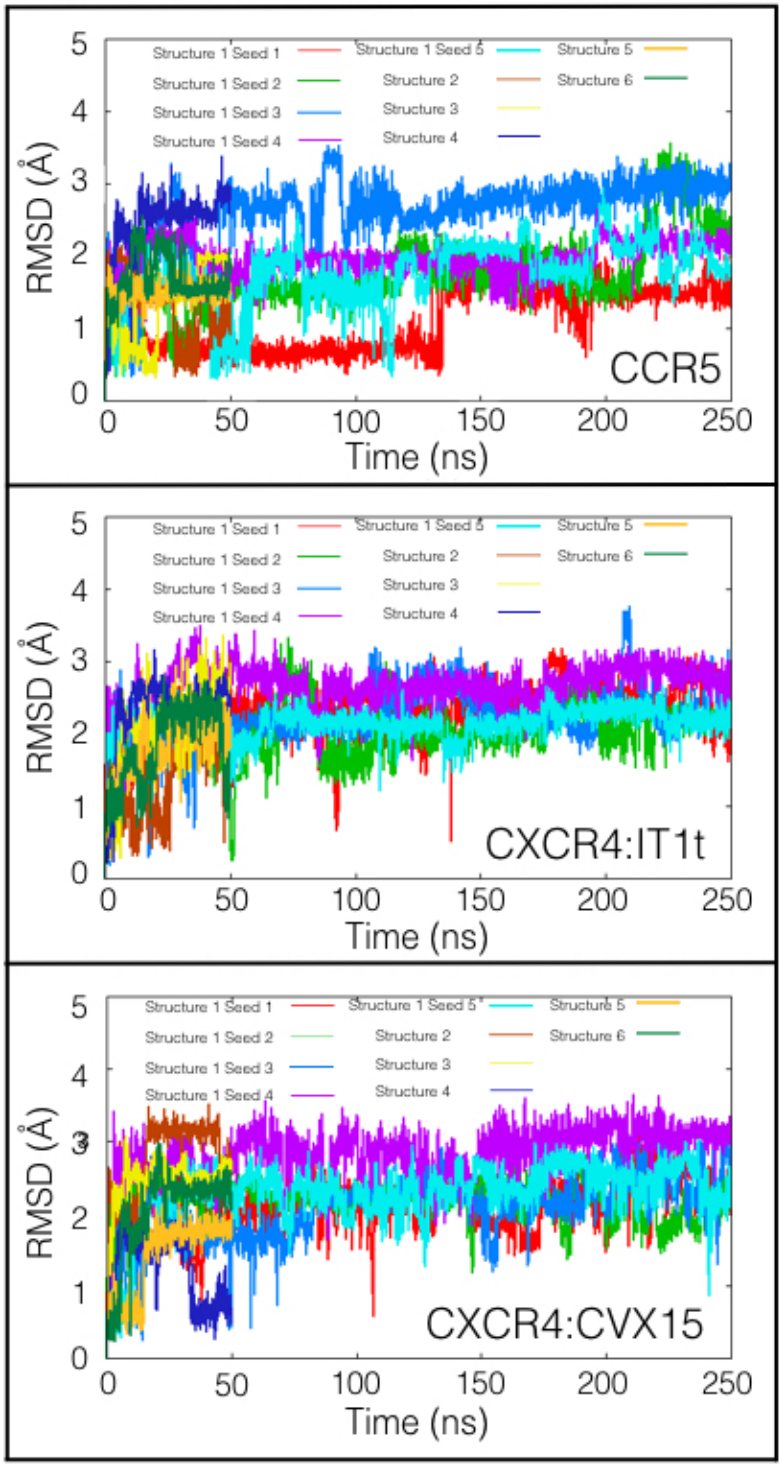

Figure S28. The RMSD values of **3**, calculated against the starting structure of the ligand, as a function of time within each receptor:ligand complex.

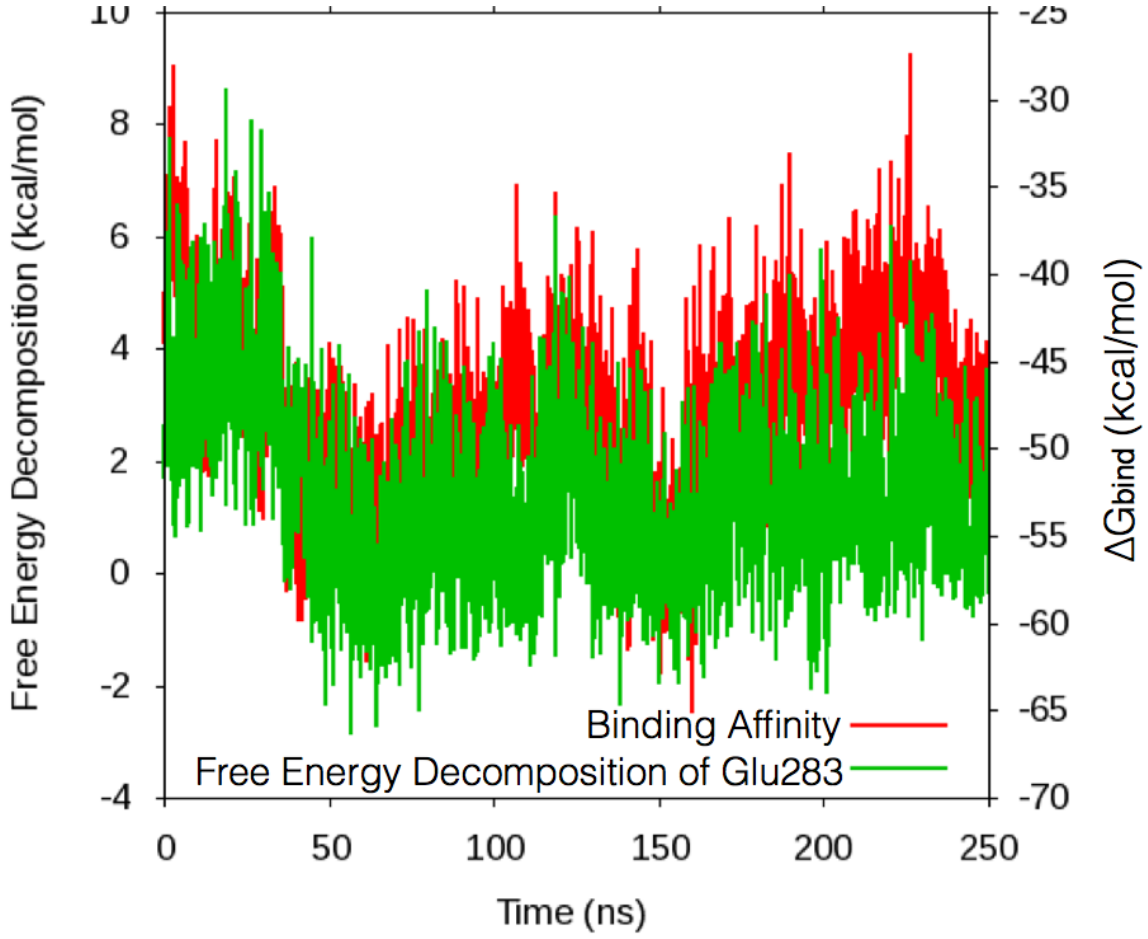

Figure S29. The per residue free energy decomposition of Glu283 and the binding affinity of 3 within the CCR5 Structure 1 Seed 2 complex during the 250 ns simulation.

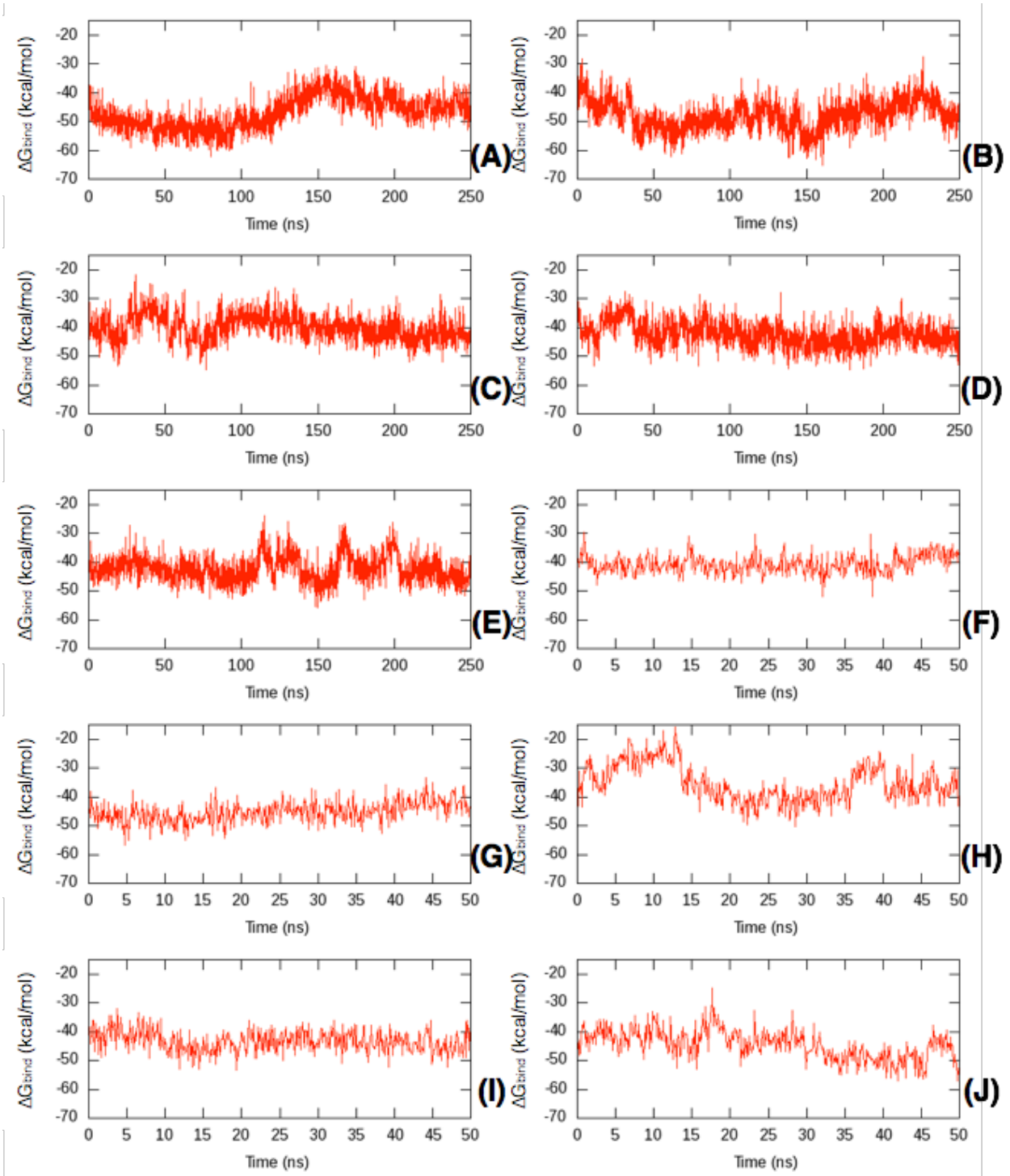

Figure S30. The ΔG_{bind} values of each **3** inhibited CCR5 simulation calculated every 0.1 ns. (A) CCR5 Structure 1 Seed 1, (B) CCR5 Structure 1 Seed 2, (C) CCR5 Structure 1 Seed 3, (D) CCR5 Structure 1 Seed 4, (E) CCR5 Structure 1 Seed 5, (F) CCR5 Structure 2, (G) CCR5 Structure 3, (H) CCR5 Structure 4, (I) CCR5 Structure 5, (J) CCR5 Structure 6.

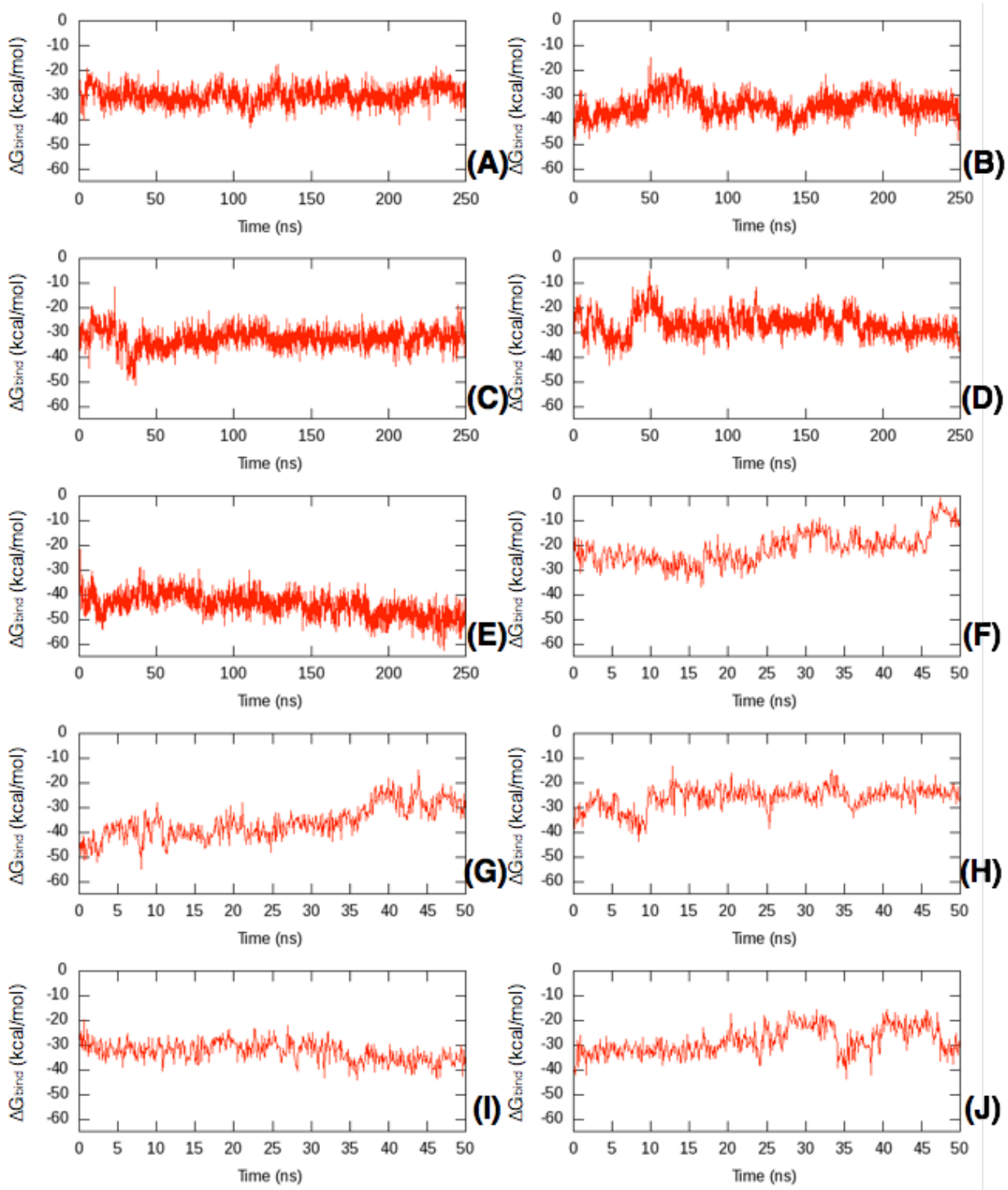

Figure S31. The ΔG_{bind} values of each **3** inhibited CXCR4:IT1t simulation calculated every 0.1 ns. (A) CXCR4:IT1t Structure 1 Seed 1, (B) CXCR4:IT1t Structure 1 Seed 2, (C) CXCR4:IT1t Structure 1 Seed 3, (D) CXCR4:IT1t Structure 1 Seed 4, (E) CXCR4:IT1t Structure 1 Seed 5, (F) CXCR4:IT1t Structure 2, (G) CXCR4:IT1t Structure 3, (H) CXCR4:IT1t Structure 4, (I) CXCR4:IT1t Structure 5, (J) CXCR4:IT1t Structure 6.

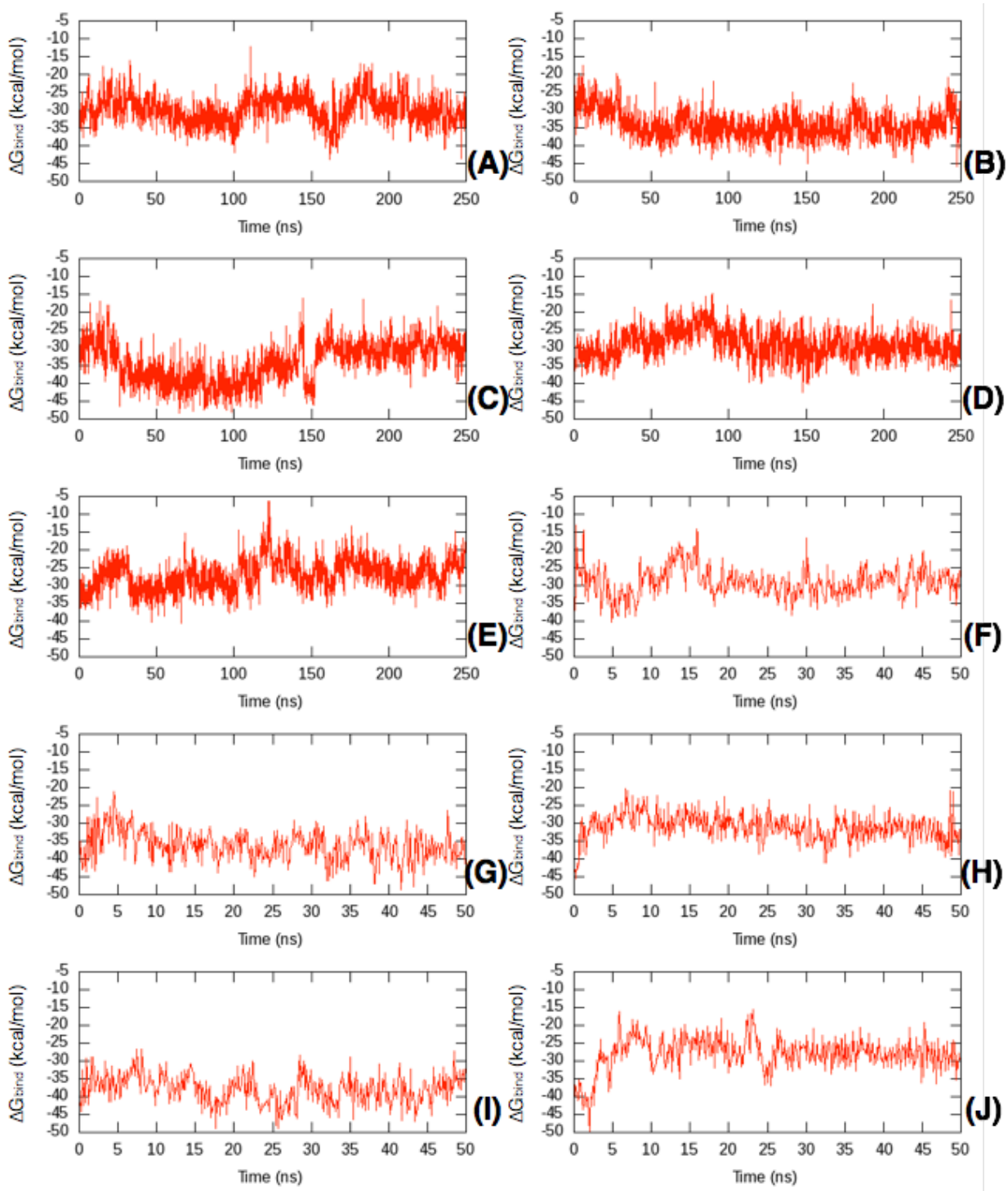

Figure S32. The ΔG_{bind} values of each **3** inhibited CXCR4:CVX15 simulation calculated every 0.1 ns. (A) CXCR4: CVX15 Structure 1 Seed 1, (B) CXCR4:CVX15 Structure 1 Seed 2, (C) CXCR4:CVX15 Structure 1 Seed 3, (D) CXCR4:CVX15 Structure 1 Seed 4, (E) CXCR4:CVX15 Structure 1 Seed 5, (F) CXCR4:CVX15 Structure 2, (G) CXCR4:CVX15 Structure 3, (H) CXCR4:CVX15 Structure 4, (I) CXCR4:CVX15 Structure 5, (J) CXCR4:CVX15 Structure 6.

Table S3. The average pairwise per residue free energy decomposition values of Tyr90, Phe173, and His177 of CXCR4:CVX15 with respect to Compound 3. The component of Compound 3 that the stated residue shows signs of pi-stacking with is listed. The shorthand of "pd" stands for interaction with the pyridine ring, "pa" the pyrazole ring, and "b" the benzene ring. The average binding energy for each CXCR4:CVX15 complex is also shown here.

Simulation	Tyr90 (kcal/mol)	Phe173 (kcal/mol)	His177 (kcal/mol)	Avg. ΔG_{bind} (kcal/mol)
CXCR4:CVX15 Structure 1 Seed 1	-2.0 (b)	-2.2 (b)	-2.7 (b)	-29.8
CXCR4:CVX15 Structure 1 Seed 2	----	-3.1 (b)	-3.9 (pa)	-34.2
CXCR4:CVX15 Structure 1 Seed 3	----	-1.4 (b)	-3.1 (b)	-34.1
CXCR4:CVX15 Structure 1 Seed 4	----	-2.4 (b)	-3.1 (pa)	-29.0
CXCR4:CVX15 Structure 1 Seed 5	----	-1.6 (b)	-1.4 (b)	-27.0
CXCR4:CVX15 Structure 2	----	-1.8 (b)	-1.3 (b)	-40.8
CXCR4:CVX15 Structure 3	-4.6 (pa)	----	-3.1 (b)	-29.1
CXCR4:CVX15 Structure 4	-2.6 (b)	-1.2 (b)	-1.3 (b)	-31.1
CXCR4:CVX15 Structure 5	-4.1 (b)	----	----	-38.0
CXCR4:CVX15 Structure 6	-2.3 (pa)	-2.0 (b)	-2.8 (pa)	-27.9

Table S4. The average pairwise per residue free energy decomposition values of Trp68 and His87 of CXCR4:IT1t with respect to Compound 3. The component of Compound 3 that the stated residue shows signs of pi-stacking with is listed. The shorthand of "pd" stands for interaction with the pyridine ring, "pa" the pyrazole ring, and "b" the benzene ring. The average binding energy for each CXCR4:IT1t complex is also shown here.

Simulation	Trp68 (kcal/mol)	His87 (kcal/mol)	Avg. ΔG_{bind} (kcal/mol)
CXCR4:IT1t Structure 1 Seed 1	-6.1 (pd)	-1.4 (b)	-29.8
CXCR4: IT1t Structure 1 Seed 2	-4.8 (b)	-2.4 (pa)	-34.2
CXCR4: IT1t Structure 1 Seed 3	-3.7 (pa)	-4.0 (pa)	-34.1
CXCR4: IT1t Structure 1 Seed 4	-4.5 (pa)	-3.9 (b)	-29.0
CXCR4: IT1t Structure 1 Seed 5	-11.8 (pa)	-3.3 (pa)	-27.0
CXCR4: IT1t Structure 2	-3.7 (b)	-2.9 (b)	-40.8
CXCR4: IT1t Structure 3	-5.9 (pd)	-1.6 (pd)	-29.1
CXCR4: IT1t Structure 4	-1.6 (pd)	-1.7 (pd)	-31.1
CXCR4: IT1t Structure 5	-6.0 (pd)	-3.1 (pa)	-38.0
CXCR4: IT1t Structure 6	-5.5 (pd)	-1.3 (pd)	-27.9

Table S5. The per residue free energy decomposition values (kcal/mol) of the residues that produced a significantly positive spike in the CXCR4:CVX15 complexes, along with their respective electrostatic and polar solvation energy terms.

Simulation	Residue	Per Residue Free Energy Decomposition	Electrostatic	Polar Solvation
CXCR4:CVX15 Structure 1 Seed 1	Asp262	+0.8	-27.6	+29.0
	Glu288	+1.1	-22.3	+23.8
CXCR4:CVX15 Structure 1 Seed 2	Asp262	+1.1	-27.1	+28.8
	Glu288	+0.9	-23.8	+25.1
CXCR4:CVX15 Structure 1 Seed 3	N/A	N/A	N/A	N/A
CXCR4:CVX15 Structure 1 Seed 4	Asp171	+0.7	-14.2	+15.1
	Asp262	+0.7	-29.4	+30.7
CXCR4:CVX15 Structure 1 Seed 5	Glu288	+1.7	-26.6	+28.9
CXCR4:CVX15 Structure 2	Glu288	+0.5	-19.1	+19.8
CXCR4:CVX15 Structure 3	Asp171	+0.7	-14.0	+15.1
	Glu288	+3.4	-24.8	+39.5
CXCR4:CVX15 Structure 4	Glu288	+4.9	-47.3	+54.2
CXCR4:CVX15 Structure 5	Glu288	+3.3	-46.8	+51.8
CXCR4:CVX15 Structure 6	Asp262	+0.8	-25.5	+26.9
	Glu288	+1.2	-25.6	+27.2

Table S6. The per residue free energy decomposition values (kcal/mol) of the residues that produced a significantly positive spike in the CXCR4:IT1t complexes, along with their respective electrostatic and polar solvation energy terms.

Simulation	Residue	Per Residue Free Energy Decomposition	Electrostatic	Polar Solvation
CXCR4:IT1t Structure 1 Seed 1	Asp97	+0.9	-23.8	+25.7
	Glu288	+5.1	-40.8	+47.1
CXCR4:IT1t Structure 1 Seed 2	Asp97	+1.2	-25.4	+27.4
	Glu288	+0.8	-27.1	+29.1
CXCR4:IT1t Structure 1 Seed 3	Asp97	+1.2	-19.8	+21.6
	Glu288	+2.6	-40.9	+44.4
CXCR4:IT1t Structure 1 Seed 4	Asp97	+0.7	-21.8	+23.3
	Glu288	+1.9	-30.3	+33.6
CXCR4:IT1t Structure 1 Seed 5	Asp97	+1.4	-43.4	+46.5
	Glu288	+0.6	-19.3	+20.1
CXCR4:IT1t Structure 2	Asp97	+1.0	-17.3	+19.0
	Asp187	+1.6	-27.1	+29.6
	Glu288	+1.5	-31.8	+34.2
CXCR4:IT1t Structure 3	Glu288	+0.9	-20.8	+22.0
CXCR4:IT1t Structure 4	Asp187	+2.7	-42.0	+46.1
	Glu288	+3.5	-33.8	+38.5
CXCR4:IT1t Structure 5	Asp97	+0.8	-23.5	+25.1
	Glu288	+2.8	-38.6	+42.6
CXCR4:IT1t Structure 6	Asp97	+2.8	-29.2	+33.6
	Glu288	+2.6	-42.9	+46.7