Parameter Optimization for Interaction between C-Terminal Domains of HIV-1 Capsid Protein

Hao Sha and Fangqiang Zhu[*]

Department of Physics, Indiana University - Purdue University Indianapolis, 402 N Blackford, LD 154, Indianapolis, Indiana 46202, United States

ABSTRACT

HIV-1 capsid proteins (CAs) assemble into a capsid that encloses the viral RNA. The binding between a pair of C-terminal domains (CTDs) constitutes a major interface in both the CA dimers and the large CA assemblies. Here we attempt to use a general residue-level coarse-grained model to describe the interaction between two isolated CTDs in Monte Carlo simulations. With the standard parameters that depend only on the residue types, the model predicts a much weaker binding in comparison to the experiments. Detailed analysis reveals that some Lennard−Jones parameters are not compatible with the experimental CTD dimer structure, thus resulting in an unfavorable interaction energy. To improve the model for the CTD binding, we introduce ad hoc modifications to a small number of Lennard−Jones parameters for some specific pairs of residues at the binding interface. Through a series of extensive Monte Carlo simulations, we identify the optimal parameters for the CTD-CTD interactions. With the refined model parameters, both the

binding affinity (with a dissociation constant of 13 ± 2 μM) and the binding mode are in good agreement with the experimental data. This study demonstrates that the general interaction model based on the Lennard−Jones potential, with some modest adjustment of the parameters for key residues, could correctly reproduce the reversible protein binding, thus potentially applicable for simulating the thermodynamics of the CA assemblies.

INTRODUCTION

The HIV-1 genome is enclosed by a protein shell termed capsid,[1] which is formed by ~1500 capsid proteins (CAs). Each CA contains one N-terminal domain (NTD) and one C-terminal domain (CTD), connected by a flexible linker.[2-5] CAs predominantly dimerize in either crystal or solution,[4, 6, 7] primarily through the CTD-CTD contacts by the parallel packing of the H9 helices.[2, 4, 7-11] Sitting in the center of H9, residues W184 and M185 are considered essential in forming the dimer, and mutations in these locations abolish dimerization.[4] Sedimentation measurement[4] showed that the dissociation constants for the dimers of the full-length CAs and of the isolated CTDs are ~18 μM and ~10 μM, respectively. An NMR experiment[2] reported a similar dissociation constant (~9.8 μM) for the isolated CTDs and further suggested that the interface in the full-length CA dimer is essentially identical to that in the CTD dimer.

In the assemblies, CAs form hexamers and pentamers through the NTD-NTD and NTD-CTD contacts.[12-18] Adjacent hexamers and pentamers in the assembly are connected to each other through the CTD-CTD interfaces similar to that in the CA dimers.[14] In vivo, the HIV-1 capsids are primarily in a conical shape consisting of ~250 hexamers and 12 pentamers,[19, 20] whereas in vitro, the CA assemblies exhibit various morphologies such as tubes, cones, and spheres.[14, 21-23] The flexibility of the binding interfaces may be responsible for the curvatures and the polymorphism of the capsid.[17] The assembly of the capsid in vivo is prevailingly considered a de novo process following the complete disassembly of the immature capsid.[24-26] It is either driven by concentration or guided by ribonucleoprotein complex (RNP), although there is no consensus as to whether the assembly starts from the narrow end[25] or from the broad end.[24] Alternatively, a recent cryo-electron microscopy (cryoEM) experiment observed multiple immature capsids coexisting in a single viral envelope, implying that the mature capsids may gradually grow from the immature ones.[27]

Monitoring the entire assembly process remains challenging given that the process is too transient to capture by experiments. Furthermore, as the association/dissociation between multiple proteins can take milliseconds or longer, typical all-atom simulations[17] are not sufficiently long to observe spontaneous binding/unbinding transitions of the CAs. To fill the gap, coarse-grained (CG) models[28-37] representing the protein at various levels of resolution have been developed. For instance, some models have a helix-level resolution with each α-helix of the CA represented by a cylinder[28, 34, 35] or a few spheres,[30] while other models have a residue-level resolution with each residue represented by a bead.[29, 32, 36] The number of CG sites with attractive interactions also varies, ranging from a few per monomer[28-31, 33-35] to one per residue.[32, 36] Most of these CG models[28-31, 33-35] follow an ad hoc approach, employing experimental CA-complexes as the targets. For example, the attractive CG sites can be placed at the centroids of the residues that are essential[28, 34, 35] or conserved[30] in the target assembly. The binding strengths between the attractive CG sites are chosen to either realize[28, 34, 35] or stabilize[30] the target CA lattices seen in the capsid. Remarkably, a recent ad hoc model,[29] with each CA dimer represented by an elastic network and with four attractive CG sites per monomer, is able to capture the nucleation and growth of the CA lattices. Most of the CG models are designed to represent a stable CA assembly, rather than to reproduce the reversible binding and the associated thermodynamics. In particular, the CTD dimer is commonly modeled as a single rigid domain without the possibility of unbinding, whereas the interfacial plasticity is likely responsible for the continuous curvature of the capsid.[17, 33]

Alternative to the ad hoc approach, a general CG model[32] was established by Kim and Hummer. The model has a residue-level granularity with one CG site per residue to quantify the inter-protein interactions. The Kim-Hummer model has two major differences from the ad hoc models. First, the standard parameters in the model are transferrable and do not include any specific information

from the known bound structures. Second, the model was designed to reproduce the reversible binding in simulations and could be used to calculate the binding affinity. Notably, the model has been verified to reproduce the experimental binding affinities for various weakly associated ($K_d >$ 1 μM) protein complexes.[32] In a recent study,[36] we adopted this model to evaluate the binding thermodynamics for a CA dimer in solution and obtained a dissociation constant of ~25 μM at 300 K, in reasonable agreement with the experimental values (~18 μM at 20°C by sedimentation,[4] or ~40 μM at 25°C by NMR[2]). However, detailed analysis revealed that the binding modes emerging from the simulations were rather diverse and mostly formed through the NTD-NTD and NTD-CTD contacts, thus contradicting with the experimental observation that CAs dimerize predominantly through the CTD-CTD contacts.[4, 6, 7]

The discrepancy above could be partly attributed to the adopted CA structure in our simulations.[36] The CAs in the crystal structure[16] (PDB ID: 3H47) were in the hexameric form, with the NTD-NTD interface enhanced by mutations A14C and E45C and the CTD-CTD interface abolished by mutations W184A and M185A. In addition, the segment (residue 176-187) that contains the major CTD-CTD contacts was missing due to the poor electron density.[16] Whereas the relevant residues were mutated back and the missing fragment was reconstructed in the structure used for the simulations, the adopted CA conformation might still have some artificial bias for the binding involving the NTDs over the CTD-CTD binding.[36]

Simulations for the binding between isolated CTDs (with the NTDs removed) could offer a more conclusive and convincing validation of the model parameters for the CTD-CTD interface. Importantly, NMR structures for the dimer of isolated CTDs[12] are available, and their binding affinity has also been measured by sedimentation[4] and NMR spectroscopy.[2] Simulations of isolated CTDs thus will not suffer the complications in our previous study[36] such as the imperfect

protein structure and the potentially competitive binding from the NTDs. Furthermore, experimental binding data could unambiguously validate our simulation results. In this study, we therefore focus on the interaction between two isolated CTDs and evaluate whether the Kim-Hummer model could reproduce their correct binding affinity and binding interface. We also present our effort to improve the model for the CTD binding by adjusting relevant parameters. Because the binding interface in the dimer of isolated CTDs is essentially identical to those in the dimer of full-length CAs[2] and in the CA assemblies, our effort here is an important step toward developing a realistic structural model for simulating large CA assemblies.

METHODS

CTD Structure. The coarse-grained protein structure in our simulations was taken from an NMR CTD dimer structure[12] (PDB ID: 2KOD) which contains the intact CTD-CTD interface similar to those in the mature CA lattice. Although the NMR structures include residues 144 to 231 of the HIV-1 CA, we only incorporated residues 144 to 219 in our model because the C-terminal tail (residues 220 to 231) is disordered and highly dynamic. Among the 30 frames in the PDB file,[12] we selected the monomer with the smallest RMSD compared to the CTDs in a tubular assembly,[18] and used the selected monomer (chain A of the seventh frame) as the CTD structure in our simulations. Our system consists of two identical copies of the CTD monomer above, each taken as a rigid body with fixed geometry.

  Energy Functions. We adopted the energy functions developed by Kim and Hummer.[32] Because the CTDs are treated as rigid bodies, the total energy U of our system is exclusively contributed by the non-bonded terms for the interactions between the two CTDs:

$$U = \sum_{ij} f_i f_j \left[u_{ij}^{LJ}(r_{ij}) + u_{ij}^{EL}(r_{ij}) \right], \tag{1}$$

in which $u_{ij}^{LJ}(r_{ij})$ and $u_{ij}^{EL}(r_{ij})$ are the Lennard–Jones (LJ) and the electrostatic energies, respectively, between residues i and j. The factor f_i is determined by the relative solvent-accessible surface area (s_i) for residue i as $f_i(s_i) = \tanh[5\tan(\pi s_i/2)]$, with s_i obtained from the GETAREA server.[38]

The LJ potential in Eq. 1 is in the form of [32]

$$u_{ij}^{LJ}(r_{ij}) = \begin{cases} 4\varepsilon_{ij}\left[(\sigma_{ij}/r_{ij})^{12} - (\sigma_{ij}/r_{ij})^6\right] + 2\varepsilon_{ij} & \text{if } \varepsilon_{ij} > 0 \text{ and } r_{ij} < 2^{\frac{1}{6}}\sigma_{ij} \\ -4\varepsilon_{ij}\left[(\sigma_{ij}/r_{ij})^{12} - (\sigma_{ij}/r_{ij})^6\right] & \text{otherwise} \end{cases}, \quad (2)$$

where ε_{ij} and σ_{ij} are the interaction strength and the characteristic distance, respectively. In particular, $\varepsilon_{ij} = \lambda(e_{ij} - e_0)$, in which $\lambda = 0.192$, $e_0 = $ -1.85 $k_B T$ (with k_B the Boltzmann constant and T the temperature),[32] and e_{ij} is the Miyazawa-Jernigan contact potential[39] based on the residue types. The standard σ_{ij} value is equal to $(\sigma_i + \sigma_j)/2$, with σ_i and σ_j determined from the van der Waals diameters of residues i and j, respectively.[32] The LJ potential with the standard parameters ε_{ij} and σ_{ij} thus represents the effective interaction between the two types of residues on average.

The electrostatic potential in Eq. 1 is in the Debye-Hückel form:[32]

$$u_{ij}^{EL}(r_{ij}) = \frac{q_i q_j}{4\pi\varepsilon r_{ij}} e^{-\frac{r_{ij}}{\zeta}}, \quad (3)$$

where q_i is the net charge of residue i, which is assigned +1 for Arg and Lys, -1 for Asp and Glu, +0.5 for His, and 0 for all other residues. The Debye screening length, ζ, is set to 10 Å, corresponding to a salt concentration of ~100 mM. The ε value corresponds to a dielectric constant of 80 for water.

We note that the LJ parameters in the Kim-Hummer model inherited the knowledge-based Miyazawa-Jernigan contact potential[39] derived from a large number of experimental protein-

complex structures, and were then calibrated[32] to fit the experimental binding affinities. Therefore, although without explicit terms for solvation energy etc., the model was indeed designed to describe the effective protein-protein interactions under physiological conditions.

Simulation Details. In our MC simulations, the CTD monomers were treated as rigid bodies. At each MC step, a random choice is made with equal probabilities for a translation or a rotation, on a randomly chosen monomer. A rigid-body translation moves the monomer in the x, y and z directions by random displacements in the range of [−0.25 Å, 0.25 Å]. A rigid-body rotation of the monomer is around a randomly selected axis through its center of mass by a random angle within 0.2 rad. The energy of the new configuration was evaluated as described earlier. All simulations were performed under the periodic boundary condition. In equations 1-3, the r_{ij} was defined as the shortest distance between residues i and j among all periodic images. The Metropolis criterion was then employed to determine whether to accept or reject the trial move. All simulations were performed at a constant temperature (300 K) for 10^9 steps. Each simulation took ~14 days to finish on a single processor in an AMD Opteron (16-core Abu Dhabi x86_64) node.

Binding Affinities. Similar to our previous study,[36] we define a pairwise contact strength between residues i and j based on their distance r_{ij}. The contact strength is 1 if $r_{ij} \leq 9$ Å, 0 if $r_{ij} \geq 12$ Å, or a value in between determined by integrating a truncated Gaussian function if 9 Å $< r_{ij} < 12$ Å. The two monomers are considered in the bound state if the total contact strength over all the residue pairs is larger than 1, or otherwise in the unbound state. The binding probability (P_b) is then given by the proportion of the bound state in the simulation trajectory, and its statistical error can be estimated from the time series using the blocking-average method.[40] The binding probability is related to the dissociation constant K_d by[36]

$$P_b = \frac{2/V}{2/V + K_d}, \tag{4}$$

in which V is the volume of the cubic unit cell in the periodic system. If simulations at multiple volumes are carried out, we estimate K_d by a nonlinear fit according to Eq. 4. Alternatively, K_d could also be obtained by a linear fit for the ratio between the probabilities of the unbound and bound states.[36]

RESULTS

We first examined the binding thermodynamics for a dimer of isolated CTDs using the standard Kim-Hummer model. Similar to our previous study[36] of the full-length CA, we performed seven MC simulations for a pair of CTDs (Table 1, Group 1) at various volumes. The dissociation constant (K_d) calculated from these simulations (see Methods) turned out to be larger than the experimental values (Table 2) by almost 300 fold. Therefore, these calculations clearly confirmed that the Kim-Hummer model in its original form underestimates the binding between the CTDs.

To identify the underlying reasons for the weak binding of the CTDs in the Kim-Hummer model, we analyzed the interaction energy when the two CTDs are in the desired binding pose as described by the experimental dimer structure. As mentioned in Methods, the total energy is the sum of the interaction energies between each pair of residues from different CTDs. Furthermore, the major component of the contacting energy between two residues comes from the LJ potential (Eq. 2) with the parameters σ_{ij} and ε_{ij} representing the range and the strength of the interaction, respectively. A close examination of these pairwise LJ terms revealed two types of problems that contributed to the unfavorable interaction in the target bound structure. First, for some residue pairs (in particular, M144-P207 and Y145-P207, see Fig. 1), the distance in the desired binding mode is small in comparison to the standard σ_{ij} value (Eq. 2) in the Kim-Hummer model, thus rendering a highly repulsive LJ energy (Fig. 2, red curves). Second, based on the structures[12] and

the mutation experiments,[4, 41] some residue pairs (in particular, W184-M185 and S178-E180, see Fig. 1) are known to form critical contact at the CTD-CTD binding interface. However, the interaction energies for these residue pairs are weak (Fig. 2, red curves) in the target binding mode because the corresponding ε_{ij} parameters are not sufficiently negative. The two problems above give rise to an unfavorable interaction energy for the bound CTD structure, thus making it unstable and resulting in the underestimation of the binding affinity.

Tuning Model Parameters. Given the problems identified above, we seek to modify relevant model parameters to reproduce the correct binding between the CTDs. As mentioned in Methods, each pair of residues in the two proteins contributes an LJ term to the total interaction energy. Therefore, a large number of such LJ parameters could potentially affect the binding, and many different sets of the parameters could possibly yield the same overall binding affinity for the protein complex. Our strategy in this study is to modify as few parameters as possible, thus keeping the vast majority of the pairwise parameters the same as in the standard Kim-Hummer model. As described in the following, our optimization is only focused on the parameters identified earlier that clearly exhibit problems according to the known dimer structures and experiments.

In the original Kim-Hummer model, the standard LJ parameters are solely determined by the residue types. Here we introduced ad hoc modifications on the LJ parameters for the specific residue pairs shown in Fig. 1 and Table 3. In particular, for M144-P207 and Y145-P207, we decreased their σ_{ij} values such that the distance in the target binding mode corresponds to the minimum of the LJ potential (Fig. 2, blue curves), while keeping the ε_{ij} parameters unchanged. The modifications completely eliminated the repulsion between these residues in the target dimer structure. For W184-M185 and S178-E180, we introduced similar modifications to the σ_{ij} parameters, but also optimized the ε_{ij} parameters to provide more favorable interactions for the

residue pairs. According to Eq. 2, a positive ε_{ij} parameter would define a purely repulsive potential, whereas a more negative ε_{ij} corresponds to a more attractive LJ energy. Therefore, by adding a negative shift $\Delta\varepsilon$ to the ε_{ij} parameters, we can increase the attraction between these residues, thus resulting in a higher binding affinity for the CTD dimer. By testing different magnitudes of $\Delta\varepsilon$, we can identify the optimal value that gives rise to the closest agreement with the experimental binding affinity, as described below.

We first scanned a broad range of $\Delta\varepsilon$, from 0 to -8 $k_B T$. For each value of $\Delta\varepsilon$, we shifted the ε_{ij} parameters for both W184-M185 and S178-E180 by the given amount, and performed an MC simulation using the modified LJ parameters. All simulations (with different LJ parameters) were run on a same system with a periodic length of 692.6 Å, corresponding to a volume at which the equilibrium binding probability (P_b) would be equal to 1/2 for the target binding affinity $K_d = 10$ μM (sedimentation measurement[4]). The results of these simulations (Table 1, Group 2 and Fig. 3) indeed showed that the binding affinity can be sensitively controlled by the magnitude of modification ($\Delta\varepsilon$), as a difference of a few $k_B T$'s in these ε_{ij} parameters could change the K_d by orders of magnitude. When $\Delta\varepsilon$ is -5 $k_B T$ or more negative, the binding becomes so strong that the two CTDs remained bound in the entire simulation without any unbinding transition. The simulations also indicated that the target binding affinity ($K_d = 10$ μM) can be achieved with the $\Delta\varepsilon$ value between -3 $k_B T$ and -4 $k_B T$. Accordingly, we performed another group of simulations, using the same protocols above, to scan this range more closely. These simulations (Table 1, Group 3 and Fig. 3) further narrowed down the optimal $\Delta\varepsilon$ value to the interval between -3.6 $k_B T$ and -3.9 $k_B T$.

To further examine the identified range of $\Delta\varepsilon$ above, we performed more simulations at multiple volumes. For each set of LJ parameters (determined by $\Delta\varepsilon$), we ran nine simulations with the

periodic length ranging from 160 Å to 800 Å with an increment of 80 Å (Table 1, Group 4), to determine a more accurate K_d (Fig. 4). Among all the tested $\Delta\varepsilon$ (Table 1, Group 4), it appears that $\Delta\varepsilon$ = -3.8 k_BT corresponds to a K_d value (13 ± 2 μM) in the closest agreement with the experimental results (see Table 2).[2, 4] Therefore, we have obtained the optimal LJ parameters for the CTD binding, with the ε_{ij} values for the residue pairs W184-M185 and S178-E180 shifted by -3.8 k_BT in comparison to the standard Kim-Hummer model (Table 3).

Binding Modes. As mentioned above, using the optimized model parameters, our simulations successfully reproduced the correct binding affinity for the CTD dimer. We also analyzed the binding mode in these simulations by calculating the RMSD for each frame in the trajectories with respect to the target experimental CTD dimer structure (2KOD).[12] Figure 5A shows the scatter plot for the RMSDs and the energies for all the bound structures sampled in our simulations. As expected, most bound states have energies significantly lower than zero, thus with favorable interaction between the two monomers. Moreover, the vast majority of the bound structures also have RMSDs (compared to the target dimer structure) lower than 4 Å, as can be seen more clearly from the cumulative distribution in Fig. 5B. In the meantime, most bound structures, including the one with the lowest energy, still have RMSDs around 3 Å with respect to the target. Indeed, the target dimer structure is not exactly at the energy minimum, thus indicating that some relaxation of that structure could further lower the energy in our model. Nonetheless, as shown in Fig. 6, at such an RMSD our bound structure matches reasonably well with the experimental CTD dimer. If one of the monomers is aligned, the other monomer would undergo a translation of 4.6 Å and a rotation of 28° to exactly superimpose with the target. Furthermore, in our binding mode the relative orientation of the two H9 helices is very similar to their parallel packing (Fig. 6) revealed by multiple experiments.[2, 4, 7-11] Overall, we conclude that there is only one major binding mode in

our simulations of the CTDs, which reasonably reproduces the CTD-CTD binding interface in the experimental structures.

Potential effect of the C-terminal tails. In this study, the CTDs are modeled as rigid domains, thus ignoring the internal conformational flexibility. This appears to be an acceptable approximation, given that the major parts of the CTD conformation are largely similar in both the unbound and the assembly forms of the CA structures.[1, 2, 12, 16] However, the C-terminal tail (residues 220 to 231) in the CTD is disordered and dynamic, and was not included in our model here. In principle, flexible tails could potentially interfere with the protein binding. To examine such effect, we performed all-atom molecular dynamics (MD) simulations of the CTD to evaluate the accessible conformations of its C-terminal tail, as described below.

Our all-atom system contained an entire CTD monomer (including the tail) in explicit solvent (15,706 TIP3P water molecules[42]). The simulation was conducted using the NAMD2 program,[43] with the CHARMM force field (version 31),[44-46] under the periodic boundary conditions at 300 K and 1 atm, and with full electrostatics calculated using the Particle Mesh Ewald method.[47, 48] We ran the simulation for a total of 300 ns, and obtained snapshots from the trajectory as representatives for the CTD monomer conformation in equilibrium. We then examined whether the C-terminal tail would interfere with the binding when a dimer is assembled from two snapshots of the monomer. It turns out that among all the virtual dimers formed by the obtained monomer conformations, the tail in one protein never has any contact with the other protein. This is not surprising, as a visual inspection also reveals that the two tails are on the opposite ends of the experimental CTD dimer structure[12] and are located far from the binding interface. Hence, each tail is highly unlikely to approach the binding interface or the other protein. We therefore conclude that the accessible conformational space of the C-terminal tail is essentially the same regardless of

whether being in a monomer or a dimer. Consequently, removing the tail should not have a large effect on the binding affinity.

DISCUSSION

In this study, we first showed that the Kim-Hummer model[32] in its original form is not sufficiently accurate to describe the binding between two CTDs[12] of the HIV-1 CA. The weak binding primarily arises from some LJ parameters that do not match the actual interactions between the residues at the desired binding interface. This is not surprising, given that the standard LJ parameters in the Kim-Hummer model only represent the interactions in the average cases. For example, the σ_{ij} parameters in the model were determined by the standard diameters of each residue type. However, the optimal C_α-C_α distance for two residues may vary significantly depending on the relative positions of their side chains. Therefore, the standard LJ parameters may not always match the actual distance between the residues at a binding interface with their side chains in highly specific positions. In such cases some ad hoc adjustments for the LJ parameters could significantly improve the accuracy of the Kim-Hummer model for a specific binding interface. As demonstrated here, after some modest modifications on the LJ parameters for four pairs of residues (while leaving all other residues still described by the standard parameters), our optimized model resulted in a dissociation constant of 13 ± 2 μM for the CTD dimer, in good agreement with the experimental measurement[2, 4] (Table 2). Moreover, the single major binding mode emerging from our simulations also reasonably resembles the experimental dimer structure (Fig. 6).

There are a variety of approaches, such as the elastic network models[49] and the Gō models,[50] to describe the protein conformation at the CG level. Moreover, some CG force field

parameterization schemes, such as the Boltzmann inversion method[51] and the force matching method,[52, 53] can be potentially applied to model the interactions between protein residues. In this study, we adopted a relatively simple strategy: given that the Kim-Hummer model performs reasonably well for the general protein-protein binding,[32] we introduced modifications on a few residues to better represent the desired specific binding. Our modified model thus combines the specificity of the CA protein with the unbiased nonspecific interactions described by the Kim-Hummer model. Our effort here is a first step toward establishing a structure-based CA model that could reproduce the proper CA assembly and in the meantime correctly represent the thermodynamics of each binding interface.

This study demonstrated that with careful calibration of the parameters, it should be feasible to use the LJ 12-6 potential to represent the residue-level interactions between protein domains. Importantly, models based on such interaction energies could quantitatively describe the reversible binding and the associated thermodynamics, as well as the correct binding mode. In addition to the CTD-CTD interface studied here, an HIV-1 CA assembly is also stabilized by several other binding interfaces. Similar optimization for the LJ parameters may reproduce the correct binding at those interfaces as well, thus eventually leading to a realistic structural model suitable for simulating the thermodynamics of the CA assemblies.

AUTHOR INFORMATION

Corresponding Author

*E-mail: fzhu0@iupui.edu.

REFERENCES

1.   Chen, B., HIV Capsid Assembly, Mechanism, and Structure. *Biochemistry* **2016**, 55, 2539-52.

2.   Deshmukh, L.; Schwieters, C. D.; Grishaev, A.; Ghirlando, R.; Baber, J. L.; Clore, G. M., Structure and Dynamics of Full-Length HIV-1 Capsid Protein in Solution. *J Am Chem Soc* **2013**, 135, 16133-47.

3.   Gamble, T. R.; Vajdos, F. F.; Yoo, S.; Worthylake, D. K.; Houseweart, M.; Sundquist, W. I.; Hill, C. P., Crystal Structure of Human Cyclophilin A Bound to the Amino-Terminal Domain of HIV-1 Capsid. *Cell* **1996**, 87, 1285-94.

4.   Gamble, T. R.; Yoo, S.; Vajdos, F. F.; von Schwedler, U. K.; Worthylake, D. K.; Wang, H.; McCutcheon, J. P.; Sundquist, W. I.; Hill, C. P., Structure of the Carboxyl-Terminal Dimerization Domain of the HIV-1 Capsid Protein. *Science* **1997**, 278, 849-53.

5.   Gres, A. T.; Kirby, K. A.; KewalRamani, V. N.; Tanner, J. J.; Pornillos, O.; Sarafianos, S. G., STRUCTURAL VIROLOGY. X-ray Crystal Structures of Native HIV-1 Capsid Protein Reveal Conformational Variability. *Science* **2015**, 349, 99-103.

6.   Momany, C.; Kovari, L. C.; Prongay, A. J.; Keller, W.; Gitti, R. K.; Lee, B. M.; Gorbalenya, A. E.; Tong, L.; McClure, J.; Ehrlich, L. S.; Summers, M. F.; Carter, C.; Rossmann, M. G., Crystal Structure of Dimeric HIV-1 Capsid Protein. *Nat. Struct. Biol.* **1996**, 3, 763-70.

7.   Worthylake, D. K.; Wang, H.; Yoo, S.; Sundquist, W. I.; Hill, C. P., Structures of the HIV-1 Capsid Protein Dimerization Domain at 2.6 A Resolution. *Acta Crystallogr D Biol Crystallogr* **1999**, 55, 85-92.

8.   Berthet-Colominas, C.; Monaco, S.; Novelli, A.; Sibai, G.; Mallet, F.; Cusack, S., Head-to-Tail Dimers and Interdomain Flexibility Revealed by the Crystal Structure of HIV-1 Capsid Protein (P24) Complexed with a Monoclonal Antibody Fab. *EMBO J* **1999**, 18, 1124-36.

9.   Du, S.; Betts, L.; Yang, R.; Shi, H.; Concel, J.; Ahn, J.; Aiken, C.; Zhang, P.; Yeh, J. I., Structure of the HIV-1 Full-Length Capsid Protein in a Conformationally Trapped Unassembled State Induced by Small-Molecule Binding. *J Mol Biol* **2011**, 406, 371-86.

10.   Rose, S.; Hensley, P.; O'Shannessy, D. J.; Culp, J.; Debouck, C.; Chaiken, I., Characterization of HIV-1 P24 Self-Association Using Analytical Affinity Chromatography. *Proteins* **1992**, 13, 112-9.

11.   Ternois, F.; Sticht, J.; Duquerroy, S.; Krausslich, H. G.; Rey, F. A., The HIV-1 Capsid Protein C-Terminal Domain in Complex with a Virus Assembly Inhibitor. *Nat Struct Mol Biol* **2005**, 12, 678-82.

12.   Byeon, I. J.; Meng, X.; Jung, J.; Zhao, G.; Yang, R.; Ahn, J.; Shi, J.; Concel, J.; Aiken, C.; Zhang, P.; Gronenborn, A. M., Structural Convergence between Cryo-EM and NMR Reveals Intersubunit Interactions Critical for HIV-1 Capsid Function. *Cell* **2009**, 139, 780-90.

13.   Ganser-Pornillos, B. K.; Cheng, A.; Yeager, M., Structure of Full-Length HIV-1 CA: a Model for the Mature Capsid Lattice. *Cell* **2007**, 131, 70-9.

14.   Li, S.; Hill, C. P.; Sundquist, W. I.; Finch, J. T., Image Reconstructions of Helical Assemblies of the HIV-1 CA Protein. *Nature* **2000**, 407, 409-13.

15.   Pornillos, O.; Ganser-Pornillos, B. K.; Banumathi, S.; Hua, Y.; Yeager, M., Disulfide Bond Stabilization of the Hexameric Capsomer of Human Immunodeficiency Virus. *J Mol Biol* **2010**, 401, 985-95.

16.   Pornillos, O.; Ganser-Pornillos, B. K.; Kelly, B. N.; Hua, Y.; Whitby, F. G.; Stout, C. D.; Sundquist, W. I.; Hill, C. P.; Yeager, M., X-ray Structures of the Hexameric Building Block of the HIV Capsid. *Cell* **2009**, 137, 1282-92.

17.   Pornillos, O.; Ganser-Pornillos, B. K.; Yeager, M., Atomic-Level Modelling of the HIV Capsid. *Nature* **2011**, 469, 424-7.

18.   Zhao, G.; Perilla, J. R.; Yufenyuy, E. L.; Meng, X.; Chen, B.; Ning, J.; Ahn, J.; Gronenborn, A. M.; Schulten, K.; Aiken, C.; Zhang, P., Mature HIV-1 Capsid Structure by Cryo-Electron Microscopy and All-Atom Molecular Dynamics. *Nature* **2013**, 497, 643-6.

19.   Briggs, J. A.; Wilk, T.; Welker, R.; Krausslich, H. G.; Fuller, S. D., Structural Organization of Authentic, Mature HIV-1 Virions and Cores. *EMBO J* **2003**, 22, 1707-15.

20.   Ganser, B. K.; Li, S.; Klishko, V. Y.; Finch, J. T.; Sundquist, W. I., Assembly and Analysis of Conical Models for the HIV-1 Core. *Science* **1999**, 283, 80-3.

21.   Ehrlich, L. S.; Agresta, B. E.; Carter, C. A., Assembly of Recombinant Human Immunodeficiency Virus Type 1 Capsid Protein in Vitro. *J Virol* **1992**, 66, 4874-83.

22.   Ehrlich, L. S.; Liu, T.; Scarlata, S.; Chu, B.; Carter, C. A., HIV-1 Capsid Protein Forms Spherical (Immature-Like) and Tubular (Mature-Like) Particles in Vitro: Structure Switching by Ph-Induced Conformational Changes. *Biophys J* **2001**, 81, 586-94.

23.   Ganser-Pornillos, B. K.; von Schwedler, U. K.; Stray, K. M.; Aiken, C.; Sundquist, W. I., Assembly Properties of the Human Immunodeficiency Virus Type 1 CA Protein. *J Virol* **2004**, 78, 2545-52.

24.   Benjamin, J.; Ganser-Pornillos, B. K.; Tivol, W. F.; Sundquist, W. I.; Jensen, G. J., Three-Dimensional Structure of HIV-1 Virus-Like Particles by Electron Cryotomography. *J Mol Biol* **2005**, 346, 577-88.

25.   Briggs, J. A.; Grunewald, K.; Glass, B.; Forster, F.; Krausslich, H. G.; Fuller, S. D., The Mechanism of HIV-1 Core Assembly: Insights from Three-Dimensional Reconstructions of Authentic Virions. *Structure* **2006**, 14, 15-20.

26.   Yu, Z.; Dobro, M. J.; Woodward, C. L.; Levandovsky, A.; Danielson, C. M.; Sandrin, V.; Shi, J.; Aiken, C.; Zandi, R.; Hope, T. J.; Jensen, G. J., Unclosed HIV-1 Capsids Suggest a Curled Sheet Model of Assembly. *J Mol Biol* **2013**, 425, 112-23.

27.   Frank, G. A.; Narayan, K.; Bess, J. W., Jr.; Del Prete, G. Q.; Wu, X.; Moran, A.; Hartnell, L. M.; Earl, L. A.; Lifson, J. D.; Subramaniam, S., Maturation of the HIV-1 Core by a Non-Diffusional Phase Transition. *Nat Commun* **2015**, 6, 5854.

28.   Chen, B.; Tycko, R., Simulated Self-Assembly of the HIV-1 Capsid: Protein Shape and Native Contacts are Sufficient for Two-Dimensional Lattice Formation. *Biophys J* **2011**, 100, 3035-44.

29.   Grime, J. M.; Dama, J. F.; Ganser-Pornillos, B. K.; Woodward, C. L.; Jensen, G. J.; Yeager, M.; Voth, G. A., Coarse-Grained Simulation Reveals Key Features of HIV-1 Capsid Self-Assembly. *Nat Commun* **2016**, 7, 11568.

30.   Grime, J. M.; Voth, G. A., Early Stages of the HIV-1 Capsid Protein Lattice Formation. *Biophys J* **2012**, 103, 1774-83.

31.   Hicks, S. D.; Henley, C. L., Coarse-Grained Protein-Protein Stiffnesses and Dynamics from All-Atom Simulations. *Phys Rev E Stat Nonlin Soft Matter Phys* **2010**, 81, 030903.

32.   Kim, Y. C.; Hummer, G., Coarse-Grained Models for Simulations of Multiprotein Complexes: Application to Ubiquitin Binding. *J Mol Biol* **2008**, 375, 1416-33.

33.   Krishna, V.; Ayton, G. S.; Voth, G. A., Role of Protein Interactions in Defining HIV-1 Viral Capsid Shape and Stability: a Coarse-Grained Analysis. *Biophys J* **2010**, 98, 18-26.

34.   Qiao, X.; Jeon, J.; Weber, J.; Zhu, F.; Chen, B., Construction of a Novel Coarse Grain Model for Simulations of HIV Capsid Assembly to Capture the Backbone Structure and Inter-Domain Motions in Solution. *Data Brief* **2015**, 5, 506-12.

35.   Qiao, X.; Jeon, J.; Weber, J.; Zhu, F.; Chen, B., Mechanism of Polymorphism and Curvature of HIV Capsid Assemblies Probed by 3D Simulations with a Novel Coarse Grain Model. *Biochim Biophys Acta* **2015**, 1850, 2353-67.

36.   Zhu, F.; Chen, B., Monte Carlo Simulations of HIV Capsid Protein Homodimer. *J Chem Inf Model* **2015**, 55, 1361-8.

37.   Noid, W. G., Perspective: Coarse-Grained Models for Biomolecular Systems. *J. Chem. Phys.* **2013**, 139, 090901.

38.   Fraczkiewicz, R.; Braun, W., Exact and Efficient Analytical Calculation of the Accessible Surface Areas and Their Gradients for Macromolecules. *J Comput Chem* **1998**, 19.

39.   Miyazawa, S.; Jernigan, R. L., Residue-Residue Potentials with a Favorable Contact Pair Term and an Unfavorable High Packing Density Term, for Simulation and Threading. *J Mol Biol* **1996**, 256, 623-44.

40.   Flyvbjerg, H.; Petersen, H. G., Error Estimates on Average of Correlated Data. *J Chem Phys* **1989**, 91, 461--466.

41.   del Alamo, M.; Neira, J. L.; Mateu, M. G., Thermodynamic Dissection of a Low Affinity Protein-Protein Interface Involved in Human Immunodeficiency Virus Assembly. *J Biol Chem* **2003**, 278, 27923-9.

42.   Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L., Comparison of Simple Potential Functions for Simulating Liquid Water. *J Chem Phys* **1983**, 79, 926--935.

43.   Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K., Scalable Molecular Dynamics with NAMD. *J Comput Chem* **2005**, 26, 1781-802.

44.   Best, R. B.; Zhu, X.; Shim, J.; Lopes, P. E.; Mittal, J.; Feig, M.; Mackerell, A. D., Jr., Optimization of the Additive CHARMM All-Atom Protein Force Field Targeting Improved Sampling of the Backbone Phi, Psi and Side-Chain Chi(1) and Chi(2) Dihedral Angles. *J Chem Theory Comput* **2012**, 8, 3257-3273.

45.   MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M., All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J Phys Chem B* **1998**, 102, 3586-616.

46.   MacKerell, A. D., Jr.; Feig, M.; Brooks, C. L., 3rd, Improved Treatment of the Protein Backbone in Empirical Force Fields. *J Am Chem Soc* **2004**, 126, 698-9.

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

48.   Essman, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G., A Smooth Particle Mesh Ewald Method. *J Chem Phys* **1995**, 103, 8577--8593.

49.   Tirion, M. M., Large Amplitude Elastic Motions in Proteins from a Single-Parameter, Atomic Analysis. *Phys Rev Lett* **1996**, 77, 1905-1908.

50.	Taketomi, H.; Ueda, Y.; Go, N., Studies on Protein Folding, Unfolding and Fluctuations by Computer Simulation. I. The Effect of Specific Amino Acid Sequence Represented by Specific Inter-Unit Interactions. *Int J Pept Protein Res* **1975**, 7, 445-59.

51.	Reith, D.; Putz, M.; Muller-Plathe, F., Deriving Effective Mesoscale Potentials from Atomistic Simulations. *J Comput Chem* **2003**, 24, 1624-36.

52.	Ercolessi, F.; Adams, J. B., Interatomic Potentials from First-Principles Calculations: The Force-Matching Method. *Europhys Lett* **1994**, 26, 583-8.

53.	Izvekov, S.; Parrinello, M.; Burnham, C. J.; Voth, G. A., Effective Force Fields for Condensed Phase Systems from Ab Initio Molecular Dynamics Simulation: a New Method for Force-Matching. *J Chem Phys* **2004**, 120, 10896-913.

Table 1. The MC simulations performed in this study[a]

	σ	$\Delta\varepsilon$ ($k_B T$)	Periodic cell length (Å)	Dissociation constant K_d (μM)
Group 1	Unmodified	0	160, 240, 320, 400, 480, 560, 640	2667 ± 68
Group 2	Modified	-0	692.6	2331 +1672/-689
		-1		1413 +1403/-472
		-2		485 +562/-172
		-3		72 +52/-23
		-4		7 +7/-4
		-5		0
		-6		0
		-7		0
		-8		0
Group 3	Modified	-3.1	692.6	31 +14/-8
		-3.2		51 +35/-16
		-3.3		19 +10/-6
		-3.4		40 +22/-12
		-3.5		69 +384/-36
		-3.6		10 +9/-5
		-3.7		10 +5/-3
		-3.8		12 +7/-4
		-3.9		9 +6/-4
Group 4	Modified	-3.6	160, 240, 320, 400, 480, 560, 640, 720, 800	16 ± 3
		-3.7		21 ± 5
		-3.8		13 ± 2
		-3.9		6 ± 2

[a]The second column specifies whether the σ_{ij} parameters for the four pairs of residues (M144-P207, Y145-P207, W184-M185 and S178-E180) are modified against the standard Kim-Hummer model (see Table 3 for the specific values). The third column specifies the shifts to the ε_{ij} parameters for the residue pairs W184-M185 and S178-E180. The fourth column specifies the periodic length of the simulation system, which determines the volume of the cubic unit cell. Multiple values indicate that multiple simulations at different volumes were performed for each parameter set. The last column provides the estimated dissociation constant from the simulation(s).

Table 2. A summary of the experimental and computational dissociation constants for the CTD dimer[a]

	Sedimentation measurement	NMR measurement	Kim–Hummer model	Optimized model
K_d (µM)	10 ± 3	9.8 ± 0.6	2667 ± 68	13 ± 2

[a]The sedimentation and NMR results are taken from Ref. 4 and Ref. 2, respectively. The values for the standard Kim-Hummer model and our optimized model are calculated from two groups of simulations (Table 1, Group 1 and Group 4 with $\Delta\varepsilon = $ -3.8 $k_B T$), respectively.

Table 3. The modified σ_{ij} and ε_{ij} values in our optimized model[a]

	M144-P207	Y145-P207	W184-M185	S178-E180
σ_{ij} (Å)	5.9 / 4.9	6.1 / 4.5	6.5 / 9.4	5.6 / 7.3
ε_{ij} ($k_B T$)			-0.7 / -4.5	0.1 / -3.7

[a]In each entry, the first value is the standard parameter in the Kim-Hummer model, and the second value is adopted in our optimized model.

FIGURE LEGENDS

Figure 1. The experimental CTD dimer structure,[12] with the side chains of some residues at the binding interface shown in licorice. The distances (in Å) between the C_α atoms of four pairs of residues (M144-P207, Y145-M207, W184-M185, and S178-E180) are indicated. The LJ parameters for these residue pairs are modified in our optimized model.

Figure 2. The LJ potential as a function of the distance (r) for the four specified pairs of residues. The red curves represent the LJ energy with the parameters in the standard Kim-Hummer model, and the blue curves represent the LJ energy in our optimized model (with the parameters given in Table 3). The vertical dashed line in each panel represents the corresponding distance in the target experimental CTD dimer structure.[12]

Figure 3. The dissociation constant K_d from simulations (Table 1, Groups 2 and 3) with different ε_{ij} parameters for residue pairs W184-M185 and S178-E180. The $\Delta\varepsilon$ indicates the shift of these ε_{ij} parameters with respect to the standard values. The dashed line indicates the experimental value ($K_d = 10\ \mu M$).[4]

Figure 4. The binding probabilities (P_b) in the simulations (Table 1, Group 4) with different volumes. The solid curves represent the nonlinear fit (Eq. 4) of P_b as a function of the effective concentration ($2/V$ multiplied by Avogadro's number). The best-fit K_d values are provided in Table 1.

Figure 5. A comparison of the bound states in the simulations with the experimental dimer structure.[12] (A) A scatter plot for the RMSD with respect to the experimental dimer structure and the energy for each bound structure in the simulations with the optimized parameters. The dashed

line indicates the average energy (which is close to zero, as expected) of the unbound states. (B) The cumulative distribution of the RMSD.

Figure 6. A superimposition for the dimer structure (blue) of the lowest energy in the simulations onto the experimental dimer structure[12] (red, with the side chains of W184 and M185 shown).
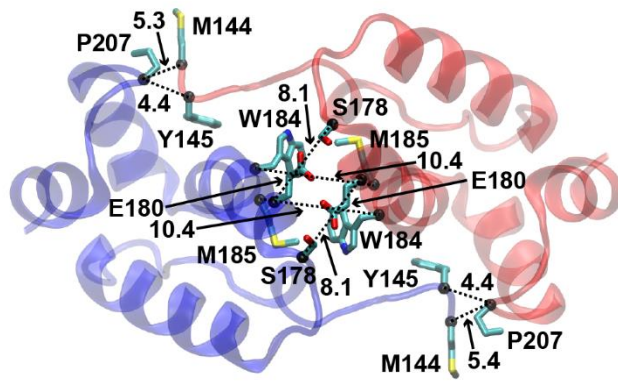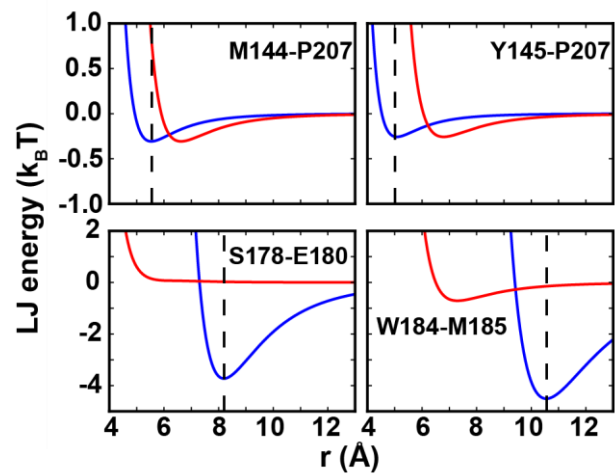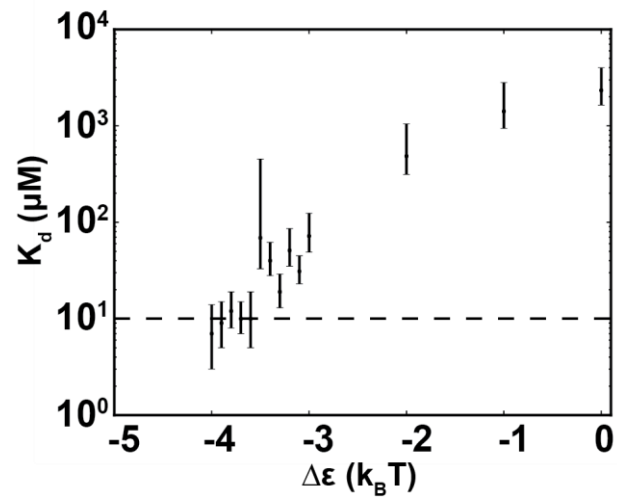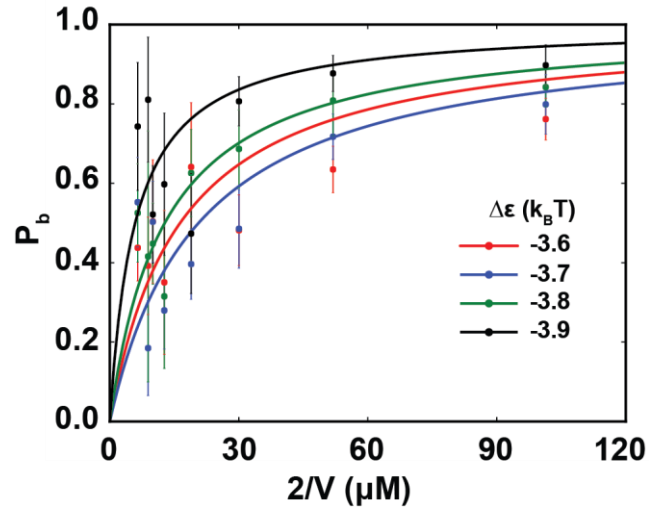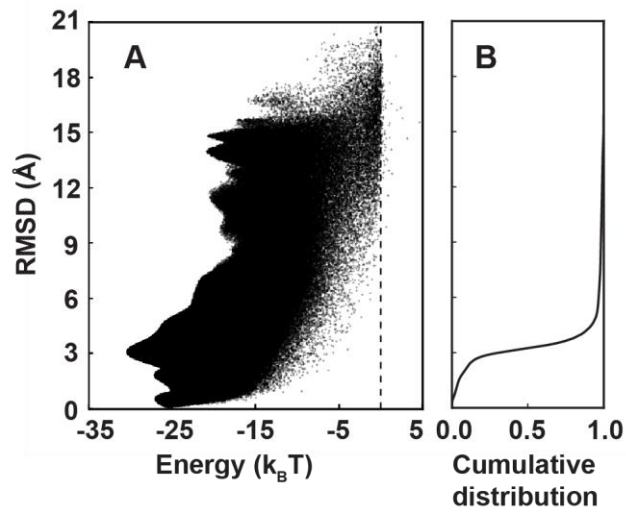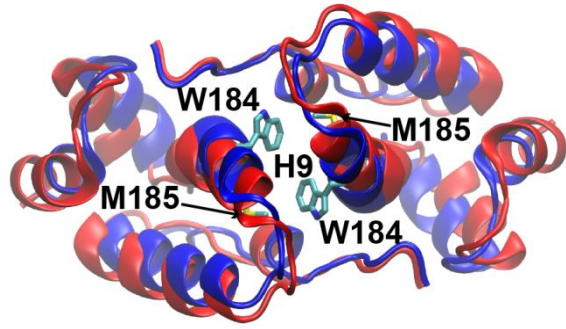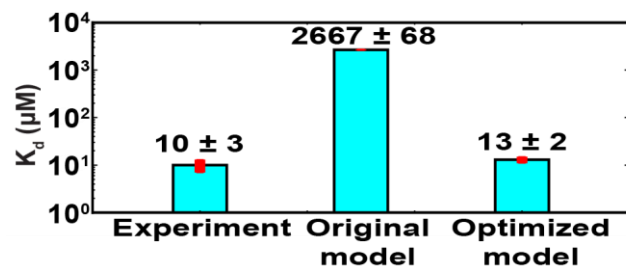
Figure 1

Figure 2

Figure 3

Figure 4

Figure 5

Figure 6

For Table of Content Graphic