Published in final edited form as:

Phys Chem Chem Phys. 2019 January 30; 21(5): 2463–2471. doi:10.1039/c8cp05644g.

Computational Studies of Protein-Protein Dissociation by Statistical Potential and Coarse-grained Simulations: a Case Study on Interactions between Colicin E9 Endonuclease and Immunity Proteins

Zhaoqian Su¹ and Yinghao Wu^{1,*}

¹Department of Systems and Computational Biology, Albert Einstein College of Medicine, 1300 Morris Park Avenue, Bronx, NY, 10461

Abstract

Proteins carry out their diverse functions in cells by forming interactions with each other. The dynamics of these interactions are quantified by the measurement of association and dissociation rate constants. Relative to the efforts made to model the association of biomolecules, little has been studied to understand the principles of protein complex dissociation. Using the interaction between colicin E9 endonucleases and immunity proteins as a test system, here we develop a coarse-grained simulation method to explore the dissociation mechanisms of protein complexes. The interactions between proteins in the complex are described by the knowledge-based potential that was constructed by the statistics from available protein complexes in the structural database. Our study provides the supportive evidences to the dual recognition mechanism for the specificity of binding between E9 DNase and immunity proteins, in which the conserved residues of helix III of Im2 and Im9 proteins act as the anchor for binding, while the sequence variations in helix II make positive or negative contributions to specificity. Beyond that, we further suggest that this binding specificity is rooted in the process of complex dissociation instead of association. While we increased the flexibility of protein complexes, we further found that they are less prone to dissociation, suggesting that conformational fluctuations of protein complexes play important functional roles in regulating their binding and dissociation. Our studies therefore bring new insights to the molecule mechanisms of protein-protein interactions, while the method can serve as a new addition to a suite of existing computational tools for the simulations of protein complexes.

Graphical Abstract

*Corresponding authors: Yinghao Wu, Phone: (718) 678-1232, Fax: (718) 678-1018, yinghao.wu@einstein.yu.edu.

Declaration of interests

Competing financial interests: The authors declare no conflict of financial interests.

A coarse-grained simulation method and a knowledge-based potential were developed to explore the dissociation mechanisms of protein complexes.

1. Introduction

Proteins form an elaborate network through their interactions in cells [1, 2]. The functions of most cellular processes are determined by the dynamics of this protein-protein interaction (PPI) network [3]. For a specific pair of PPI in the network, the dynamics of its interaction is modulated by the association rate (k_{on}) and dissociation rate (k_{off}) constants [4]. Therefore, the cellular functions of protein interactions are closely correlated to these two parameters. However, the general mechanism underlying the association and dissociation of a protein complex and the determinants of their rate constants are still not fully understood [5–7]. Different from association rates that are concentration dependent and limited by the molecular diffusion in cells, dissociation rates are entirely counted on the short-range interactions of residues at the interfaces of protein complexes. Moreover, many disease-associated mutations that occur at the protein-protein binding interfaces have shown minimal impacts on the association rates of their interactions, but cause significant changes to the dissociation rates [8]. This highlights the importance of understanding dissociation of protein complexes and the implications in their biological functions.

Relative to the time-consuming and labor-intensive lab techniques which experimentally measure rate constants of protein binding [9–12], computational modeling approaches show a unique advantage that is able to reveal the atomic details of binding processes. Unfortunately, most of these computational studies focus on the simulations of biomolecular association [6, 13–19]. In comparison, much less effort has been made to decipher the dissociation mechanism of protein complexes. Effect of mutations on changes in rates of protein-protein dissociation has been assessed by characterizing the energetic features of hotspot residues at binding interfaces [20]. In addition to the research on bioinformatics, dissociation processes were also estimated by molecular dynamic (MD) simulations [21, 22]. However, because the capacity of current MD simulations is below the normal time-scale for dissociation of most native protein complexes, their applications have been limited to protein-protein interactions with very weak binding affinity ($K_d \sim 10^{-4}$ M).

In this work, we applied a coarse-grained (CG) simulation method to study the dissociation mechanisms between proteins. The CG model of a protein is represented by the Ca atoms and sidechain centers of mass. The parameters of interactions between proteins are described by the statistical potential that was derived from a non-redundant structural library of protein

complexes. A Langevin dynamic simulation algorithm is further used to guide the process of dissociation starting from the bound structure of native protein complex. We tested our method to the complexes of colicin E9 endonucleases (DNase) with immunity (Im) proteins [23]. Interestingly, it shows that dissociation between E9 DNase and its cognate ligand protein Im9 is five orders of magnitude slower than its non-cognate ligand Im2 [24], although Im9 and Im2 share high sequence identity and structural similarity (Figure 1a). Our results show that the difference of dissociation between these two complexes and the effects of mutations on change of their dissociation rates can be reproduced by our simulations. While we increased the flexibility of protein complexes, we further found that they are less prone to dissociation, suggesting the functional role of conformational fluctuations in regulating protein unbinding. Taken together, we demonstrate that our simulation method can be a useful tool to unravel molecular mechanisms of protein-protein dissociation.

2. Results

We first applied all atom molecular dynamic (MD) simulations to study the interactions between protein E9 DNase and its two ligands: Im9 and Im2. Specifically, 100 nanosecond-long simulation trajectories were carried out for both complexes using GROMACS. Detailed simulation setups are described in the methods. We plotted the backbone root-mean-square-differences (RMSD) as a function of simulation time for both systems, as shown in Figure 1b. We also calculated the root-mean-square-fluctuations (RMSF) of each residue in two complexes, as shown in Figure 1c. The figures indicate that both RMSD and RMSF of complex E9 DNase-Im9 are lower than the complex E9 DNase-Im2. The MD simulations thus suggest that the complex E9 DNase-Im9 is more stable, which is able to give indirect explanation about why its dissociation is slower. However, given the small amplitude of conformational dynamics within the timescale of all atom MD simulations, we still cannot observe the initiation of protein complex dissociation.

In order to simulate how protein complexes dissociate and understand the dissociation mechanisms, the coarse-grained model developed in this article was first used to study the diner complex between colicin E9 DNase and its cognate immunity protein Im9. Specifically, the native structure of the complex with PDB id 1EMV was used as the initial conformation (Figure 2a) [25]. Based on the crystal structure, we also plotted the contact map for residues that form inter-molecular interactions in the native complex, as shown in Figure 3a. The y-axis in the figure indicates the residue index of E9 DNase, while the x-axis indicates the residue index of its ligand. The contact map shows that binding of the native complex is largely maintained by the residues in the second and the third helices (helix II and helix III) of immunity protein Im9. Following the initial conformation, 10 independent trajectories of Langevin dynamic simulation were carried out on the coarse-grained model of the complex. Each trajectory contains 10^6 simulation steps. The process of dissociation between E9 DNase and Im9 is guided by the interactions at their binding interfaces which parameters were derived from the knowledge-based potential.

We counted the percentage of native contacts (PNC) during dissociation along all simulation trajectories. A native contact was defined as an intermolecular interaction between a residue from E9 DNase and the other residue from Im9 which was observed in the native structure

of the complex. The percentage of native contacts was derived by calculating the ratio between the numbers of native contacts left during dissociation versus the original number before simulation. Figure 2c gives the changes of calculated PNC along one of the simulation trajectories. The initial value of PNC is 1 in the native structure. The figure shows that the PNC dropped after simulation started, suggesting that residues at the binding interface of the complex lost their contacts. The final value at the end of the trajectory is 0.7, indicating that around one third of the native contacts have been broken during simulation. The final structural conformation of the complex at the end of the trajectory is shown in Figure 2b. The energy at binding interface of the complex that was calculated by equation (2) was also plotted along the trajectory (Figure 2d). The figure shows that the interface energy increased after the simulation, indicating that the complex became less stable. Moreover, the fastest increases of interface energy are at the first 2×10^5 steps of the simulation. This suggests that the dissociation is initiated by breaking the intermolecular interactions that make the most significant contributions to the binding of the complex. In order to give more quantitative analysis to the detailed mechanism of how complex dissociates, we plotted the contact map for the final conformation in Figure 3b. By comparing with the contact map of the native structure (Figure 3a), we found that most interactions between residues in E9 DNase and residues in helix III of Im9 are maintained. In contrast, most interactions between residues in E9 DNase and residues in helix II of Im9 are lost. This result indicates that the dissociation of E9 DNase from Im9 is initiated and thus mainly regulated by helix II. On the other hand, more stable interactions are formed for residues from helix III in Im9.

In order to show which specific residues in helix II and helix III of Im9 remain contacts with E9 DNase and which residues lose contacts along simulations, the binding interface of native complex is zoomed in and the sidechains of residues that form atomic interactions are highlighted (Figure 4a). As shown in the figure, the interactions between E9 DNase and Im9 helix III are maintained through the two tyrosine residues Y54 and Y55. They form hydrophobic interactions with non-polar residues F86 and P88 in E9 DNase. These interactions mostly remained after simulations (Figure 3b). In contrast, the contacts between E9 DNase and Im9 helix II are maintained by two groups of interactions: one is between the electrostatic interaction between E41 in helix II of Im9 and K89 in E9 DNase; the other group is between the hydrophobic residues V34 and V37 in helix II of Im9 and F86 in E9 DNase. These interactions are mostly broken after simulations (Figure 3b). Previous experiments demonstrated that binding between E9 DNase and Im9 is largely driven by entropy [26], consistent with the nonpolar nature of the interface of this complex.

In addition to the cognate ligand protein Im9, E9 DNase can also bind to a non-cognate ligand, immunity protein Im2. Both cognate and non-cognate ligands share 68% sequence identity. However, experimental measurement showed that dissociation between E9 DNase Im2 ($7.3 \times 10^{-1} \text{s}^{-1}$) is five orders of magnitude faster than Im9 ($2.4 \times 10^{-6} \text{s}^{-1}$) [27]. To understand the difference of dissociation between protein Im9 and Im2, and further reveal the binding mechanism between E9 DNase and its different ligands, we also applied the coarse-grained simulation to the protein complex between E9 DNase and immunity protein Im2. Specifically, the native structure of the complex is taken from the PDB id 2WPT [28] and was used as the initial conformation of simulations. The contact map of the

intermolecular interactions in the native complex is plotted in Figure 3c. Similar patterns are found in the contact map of Figure 3a, indicating that both complexes of Im9 and Im2 are formed through similarly binding interfaces. We also carried out 10 independent trajectories of Langevin dynamic simulation for the dissociation of E9 DNase and Im2, in which each trajectory contains 10^6 simulation steps. The final calculated PNC for all trajectories are plotted as histogram in Figure 5a, in decreasing order. Comparing with the final PNC of simulations between E9 DNase and Im9, which is shown in decreasing order as the stripe bars in the figure, lower values of final PNC (grey bars in Figure 5a) are found for the simulations between E2 DNase and Im2. This is consistent with the experiments that dissociation between E9 DNase and Im9 is more difficult than Im2. In order to further assess the statistical significance of obtained differences in final PNCs from 10 trajectories of simulations between two systems, a student's t-test was performed to the two datasets. In detail, the average PNC of E9 DNase-Im9 complex simulations is 0.58 and the standard deviation is 0.103, while the average PNC of E9 DNase-Im2 complex simulations is 0.48 and the standard deviation is 0.066. The null hypothesis that no difference exists between two sets was tested at a 95% confidence interval. Consequently, the calculated t-score equals 2.6 and the corresponding P-value is 0.0287. Therefore, the small P-value for the t-test suggests that we can reject the null hypothesis and accept the alternative hypothesis, i.e., the differences of final PNCs between simulations of E9 DNase-Im9 complex and E9 DNase-Im2 complex are significant.

Moreover, we plotted the contact map of the final conformation in Figure 3d. Relative to the native structure of the complex we found that almost all contacts between residues in E9 DNase and residues in helix II of Im2 are lost after simulation. On the other hand, interactions of residues from helix III are largely remained. Therefore, after comparing the simulations of the complex between E9 DNase and Im2 with the complex between E9 DNase and Im9, the similarity and difference in the dissociation pathways of these two systems can be manifested. The interactions between E9 DNase and the residues in helix III of both Im2 and Im9 proteins are stable during the initial stage of dissociation. On the contrary, residues in helix II show higher propensity to be broken from the binding interfaces. Furthermore, fewer residues in helix II of protein Im2 remain interacting with E9 DNase than residues in helix II of protein Im9, which explains the lower values of final PNC in the simulations of E9 DNase-Im2 system. In other words, protein Im2 is easier to dissociate from E9 DNase than Im9, and helices II in both proteins are most responsible to the difference. Previous study on the system proposed a dual recognition mechanism for this binding specificity [28], in which the residues of helix III of Im2 and Im9 proteins act as the anchor for binding, while the residues of helix II make positive or negative contributions to specificity. It is also worth mentioning that the sequence of residues from helix III are more conserved, especially the two tyrosine residues Y54 and Y55, as well as the aspartic acid (D51) which are presented at the binding interfaces between E9 DNase and two immunity proteins. We speculate that these residues could serve as conserved hotspot that dominates the association process of protein complexes, considering that measured values of k_{on} in both systems are not too much different. In comparison, the residues in helix II of immunity protein are much more variable. Especially, the hydrophobic residue V34 in the helix of Im9 is replaced by asparagine, which is a polar residue (Figure 4b). Since residue V34 is

involved in the nonpolar interface of E9 DNase-Im9 complex, this sequence variation can destabilize the binding. We thus deduce that sequence variations in helix II gives the result that non-cognate ligand protein Im2 dissociation much faster than the cognate protein Im9 and modulate the binding selectivity. Therefore, our study provides the supportive evidences to the dual recognition mechanism for the specificity of binding between E9 DNase and immunity proteins. Moreover, we suggest that this binding specificity is rooted in the process of complex dissociation instead of association.

The E colicin DNases are a group of bacterial toxins including E2, E7 and E8 in addition to E9. Specific for each colicin, producing strains co-synthesize an immunity protein (Im2, Im7, Im8, and Im9) to neutralize the catalytic activity of the toxin. They are two families of highly conserved proteins. Previous experiments showed that all 16 possible DNase-Im protein combinations can form highly specific complexes [26]. The binding of cognate complexes is 10^6 - 10^{10} -fold tighter than non-cognate complexes. It was proposed that specificity across all DNase-Im interactions is governed by the same “dual recognition” mechanism [29], in which conserved helices III anchor the binding, while helices II stabilize cognate complexes and likely destabilize non-cognate complexes. Moreover, previous experiments also showed that both cognate and non-cognate complexes bind with similar association rate constants [30]. This provides the supportive evidence to our simulation results that the distinction between cognate and non-cognate binding arises from the dissociation kinetics. Therefore, although the simulations in current study are only carried out for E9 DNase, it can bring insights to the kinetic mechanism of dissociation for the entire group of DNase-Im protein interactions.

After we tested our method on wild-type complexes, we also carried out the simulations of dissociation to the systems in which residues at the binding interface between E9 DNase and Im9 are mutated. Specifically, two systems are studied. Both systems contain double mutants. In the first system, one of the mutations is from E9 DNase. This mutation replaces residue F86, which is located on the binding interface of E9 DNase, with alanine. The other mutation in the first system is from protein Im9, in which residue L33 is changed into alanine. L33 is located in helix II of the protein. The mutant F86A is also included in the second system. Differently, the other mutation replaces Y54 from Im9 with alanine. Y54 is one of the hotspot residues located in helix II of the protein. Experimental data show that the values of k_{off} become much larger for both mutants ($1.5 \times 10^{-1} \text{s}^{-1}$ and $3.2 \times 10^0 \text{s}^{-1}$) [31], indicating that mutations of these residues accelerate dissociation of the complex.

In order to perform simulations on these systems, the structures of two complexes with mutated residues were constructed as described in the methods. Following the same procedure, 10 independent trajectories of Langevin dynamic simulation were carried out to test the dissociation of mutated complexes in both systems, while each trajectory contains 10^6 simulation steps. We monitored the PNC along all the simulation trajectories. Their final values are plotted in decreasing order as histograms in Figure 5b. The white bars in the figure correspond to the system of F86A, L33A mutants, while the black bars correspond to the system of F86A, Y54A mutants. In comparison, the final values of PNC for the simulations of wild-type complex are plotted in decreasing order as stripe bars. The figure shows that much less native interactions between E9 DNase and Im9 are left after

simulations in both mutant systems than those in the wild-type, suggesting that the structures of mutant complexes are much less stable. This result indicates that mutations of two residues at binding interfaces can strongly facilitate the dissociation of the complex, which is consistent with the previous experiments. Based on our tests on the wild-type complexes between E9 DNase and its different ligand protein Im2 and Im9, as well as various mutations which disturb the native binding interfaces, we demonstrate that our coarse-grained simulation method is sensitive enough, and thus efficient to qualitatively capture the mechanistic details of protein complex dissociation. Student's t-tests were carried out to assess the statistical significance in obtained differences of final PNCs from 10 trajectories of simulations between wild-type and two mutated complex systems. The average PNC of complex with F86A, L33A mutants is 0.334 and the standard deviation is 0.032, while the average PNC of complex with F86A, Y54A mutants is 0.42 and the standard deviation is 0.084. As a result, the calculated t-score between wild-type and double mutants F86A, L33A is 7.27 and the corresponding P-value is 0.0001. Similarly, the calculated t-score between wild-type and double mutants F86A, Y54A is 4.0 and the corresponding P-value is 0.0031. The small p-values from the t-tests demonstrate that the differences of final PNCs between simulations of wild-type and mutated complexes are significant.

The entropic effect due to conformational fluctuations in a protein complex plays an important role to regulating its stability [32]. In our coarse-grained simulation, the intramolecular degrees of freedom for each protein are maintained by the elastic potential which constrains both the local conformation of backbone and the long-range interactions between residue sidechains. In attempt to evaluate how conformational fluctuations affect dissociation of a protein complex, we reduced the strength of force constants to increase the flexibility of proteins in a complex. Specifically, the force constants of bond angles and bond dihedrals between consecutive C α atoms along the backbone, as well as the force constants of elastic potential that constrains the long-range interactions between sidechains of non-bonded residues were reduced. Practically, these parameters were lowered to one third of their original values. As a result, conformational fluctuations can be enhanced when the complex becomes more flexible. We applied the new parameters to simulate dissociation of the wild-type complex between E9 DNase and protein Im9 with 10 independent trajectories. We calculated the root-mean-square fluctuations (RMSF) of residues in E9 DNase by averaging over 10^6 simulation steps of all trajectories (Figure 6a). We compared the fluctuations of the new system with the original simulations in which the complex is less flexible. The black curve in the figure shows the RMSF of each residue in the new system, while the red curve represents the original system. The fluctuation profiles of two systems show similar distributions along the residue index. This indicates that the new system still follows the dynamic properties of the original system despite a big change of force constants. Moreover, the overall higher RMSF in the new simulations confirm that the protein undergoes larger conformational changes by given the weaker force constraints.

In addition to the fluctuations, we further compare the PNC between the new and the original systems. The grey bars in Figure 6b show the PNC that correspond to the new system with higher fluctuations, while the stripe bars stand for the PNC of the original system with lower fluctuations. The figure suggests that although the protein complex undergoes higher level of conformational variations, much less native interactions are broken

during the simulations. Student's t-tests were carried out to assess the statistical significance of obtained differences in final PNCs between 10 trajectories of simulations with high and low structural flexibility. Comparing to the original level of flexibility in which the average PNC is 0.58 and the standard deviation is 0.103, the average PNC of 0.69 and the standard deviation of 0.06 were derived for simulations with higher flexibility. Consequently, the calculated t-score equals 2.93 and the corresponding P-value is 0.0168. The small P-value suggests that the differences between simulations with high and low conformational fluctuations are statistically significant.

The result from Figure 6b indicates that high structural flexibility of proteins doesn't facilitate dissociation. In contrast, it prevents these proteins from leaving each other. In principle, this is due to the fact that higher conformational entropy in a more flexible protein complex makes positive contribution to the stability of the system. In recent works, it has been shown that the mutations of residues that are not located at interfaces of a protein complex still are able to change its binding affinity [33]. Since these residues are not directly involved in the inter-molecular interactions, their functional roles in regulating the stability of protein complexes are not fully understood. Our simulations might provide a potential mechanism of how non-interface residues correlate to protein-protein interactions. We speculate that mutations of these non-interface residues change the flexibility of the entire protein through the variation of intramolecular interactions. The change of molecular flexibility in turn affects protein binding, as we illustrated above. Our hypothesis can be testified in the future, after we collect more quantitative evidence on the relation between mutations of non-interface residues and changes of protein conformational flexibility.

3. Concluding discussions

Proteins fulfill their versatile functions by forming complexes together [13, 33–37]. For an example, bacteria that produce colicins are protected against their cytotoxic activity through the high specificity interactions with small immunity proteins to inactivate the endonuclease [25]. The dynamics of these interactions are quantified by the measurement of association and dissociation rate constants. However, most relevant experimental methods are both time-consuming and labor-expensive, which offers a large opportunity to computational simulations that can also be used to study the detailed mechanism of protein interactions. Unfortunately, relative to the efforts made to simulate the association of biomolecules, there are few reports specifically focusing on the modeling of protein complex dissociation. To tackle this problem, protein complex formed between the E9 DNase domains of bacterial colicins and their immunity proteins are used as a test system. Our previously developed coarse-grained (CG) model [38] was extended to simulate the dissociation of the complex. In detail, two upgrades have been made. Firstly, instead of using a C α -based model in the previous method, each residue is now represented by two points: one is the C α atom and the other is the sidechain center of mass. Secondly, the interactions between proteins in the complex are modeled by a new knowledge-based potential. As described in the Methods, the potential was constructed by the statistics from available protein complexes in the structural database. Given the new representation and potential, the process of dissociation is guided by Langevin dynamic simulation algorithm. In the future, the method can also be upgraded by using a multiscale strategy to improve the accuracy of simulation results. For instance, a

hybrid molecular mechanics/coarse-grained simulations (MM/CG) approach [39, 40] can be used to study the dissociation of protein complexes. In this MM/CG approach, the binding site can be described by an atomistic potential whereas the rest of the protein is treated with a CG potential.

Consistent with the experiments, we found that dissociation between E9 DNase and immunity protein Im9 is more difficult than the dissociation of immunity protein Im2. Moreover, our study provides the supportive evidences to the dual recognition mechanism for the specificity of binding between E9 DNase and immunity proteins, in which the conserved residues of helix III of Im2 and Im9 proteins act as the anchor for binding, while the sequence variations in helix II make positive or negative contributions to specificity. Beyond that, we further suggest that this binding specificity is rooted in the process of complex dissociation instead of association. Additionally, in order to explore determinants for protein dissociation, we changed the energy parameters in simulations or tried different binding interfaces by using artificially generated structural model. We demonstrated that that higher conformational entropy of more flexible proteins negatively regulates the dissociation of the complex. In summary, our studies bring new insights to the molecule mechanisms of protein-protein interactions, while the method can serve as a new addition to a suite of existing computational tools for the simulations of protein complexes.

4. Methods

Representation of the coarse-grained model

The initial conformation of the complex formed between colicin E9 DNase and its cognate immunity protein Im9 was adopted from the crystal structure with PDB id 1EMV. Similarly, the initial conformation of the complex formed between colicin E9 DNase and its non-cognate immunity protein Im2 was adopted from the crystal structure with PDB id 2WPT. Consequently, the atomic structure of protein complexes was reduced to the following simplified model in simulations. Each residue is coarse-grained into two sites: one is the position of its Ca atom, while the other indicates the position of its sidechain. This is represented by the center of mass for all atoms in sidechain of the residue.

In comparison to the wild-type, the structures of mutant protein complexes with mutated residues were constructed by first keeping the backbone of the protein complex as its native conformation and replacing the sidechain conformations using SCWRL4 [41]. The constructed sidechain rotamers were then coarse-grained by their center of mass.

Overall simulation algorithm

Following our previous study [41], Brownian Dynamics (BD) algorithm was used to update the Cartesian coordinates and velocities for all particles in each time step of simulations. Specifically, the equation of motion for the particle i , which could be either a Ca atom or a sidechain center of mass, in protein complex can be written as:

$$\gamma_i M_i \frac{d\vec{r}_i}{dt} = -\nabla_i E(\vec{r}) + \sqrt{2\gamma_i M_i k_B T} \vec{R}_i(t). \quad (1)$$

In equation (1), γ and M are the friction coefficient and mass of each particle. The first term on the right side is the total forces applied to the particle, in which ∇_i is the gradient operator applied to it, and E is the potential calculated on the protein complex under current conformation. The second term describes the stochastic force from surrounding environments, in which k_B is the Boltzmann constant, T is the temperature of the system, and $R(t)$ is a delta-correlated stationary Gaussian process with zero-mean.

In our simulations, the friction coefficient and average mass of each particle equals 0.1ps^{-1} and 100g/mol , respectively. The detailed form of potential energies will be described in the next part. Forces are calculated in the unit of pico-newton (pN). All simulations were run with a time step of 0.01 ns and a heat bath at 300 K temperature. Finally, to calibrate the computational performance, we compared our coarse-grained model with the all-atom MD simulations. Using the E9 DNase-Im9 complex as a test system, the coarse-grained simulation takes approximately one CPU hours to generate a trajectory of 100 ns on a regular computer processor. In comparison, it takes approximately 3000 CPU hours of the same computer processor to generate a $100\text{ nanosecond-long}$ trajectory using GROMACS all atom MD simulation on the same system. Therefore, our CG simulations are much faster than the traditional all-atom MD simulations. In practical, the multi-node parallelization-scheme (MPI) was implemented for GROMACS simulation on 20 processors in the high-performance computing center of our institute. The detailed protocol of GROMACS simulation is specified below.

Potential for inter-molecular interactions

The short-range interaction and intra-molecular long-range interactions are based on the traditional Go-potential [42], which have been described in our previous study. They are used to constrain the tertiary structure of each protein monomers in the complex around their native conformation, but still allow a small degree of flexibility. Different from the previous study, the inter-molecular interactions in the complex have the following form.

$$E_{\text{inter}}(\vec{r}) = \sum_{ij \in \text{inter}} |u_{SC}(i, j)| \left[\left(\frac{r_{ij}^0}{r_{ij}} \right)^{12} - f(u_{SC}(i, j)) \left(\frac{r_{ij}^0}{r_{ij}} \right)^6 \right] \quad (2)$$

The summation in equation (2) is taken over all residue pairs at the binding interface of the protein complex, one from the residue of E9 DNase and the other from the residue of Im2 or Im9 proteins. A pair of residues in the summation is recognized if the inter-molecular distance between any atoms of their side-chains is less than 5.5 Angstrom . The distance between sidechain centers of mass for these residues is r_{ij} , while r_{ij}^0 is their distance in the native structure. In turn, the strength of the interaction between residue types i and j is determined by the statistics from available protein complexes in the structural database. Specifically, the sidechain-based binding energy $u_{sc}(i, j)$ can be written as follow.

$$u_{sc}(i, j) = -kT \ln \frac{N_{obs}(i, j)}{\chi_i \chi_j N_{obs}} \quad (3)$$

In equation (3), $N_{obs}(i, j)$ is the observed number of residue pairs i and j at binding interfaces; N_{obs} is the total residue pairs at binding interface; and χ_i is the mole fraction of residue type i at binding interfaces [43]. Two residues are considered in contact if a pair of any atoms belonging to the sidechains of these residues is closer than the same cut-off value (5.5 Angstrom). The values of parameters in equation (3) were derived by counting the corresponding observed numbers in a large-scale structural library of protein complexes that collected from the 3did database [44]. The 3did database selected inter-domain interactions in all protein complexes for which high-resolution three-dimensional structures are available. The database contains a large group of items defined as interacting domain pairs (IDP). Each IDP could be homodimer, heterodimer, or inter-domain interaction within a single subunit. Information about Pfam index is given for both domains of an IDP [45]. Each IDP further includes different number of instances. The specific instances are called 3D items in which information about PDB index, chain id and residue range are provided for both interacting protein domains. In order to reduce sequence redundancy in our structural library of protein complexes, we selected only one representative 3D item from each IDP in 3did database. This leads to the final library consisting of 4960 entries of protein-protein interactions from either homodimers or heterodimers. Detailed values of all derived energy parameters are plotted as two-dimensional contour in Figure 7. The figure shows that the first-principle physical and chemical characteristics of molecular recognition such as electrostatic interactions and hydrophobic effect can be captured by the potential. Finally, the function $f(x)$ equals 1 when x is smaller than 0, and -1 when x is larger than 0. This function assures that the negative energy parameters derive from the statistics are attractive in the dynmsic simulations, while the positive parameters are repulsive.

Protocol of all atom molecular dynamic simulations

Both MD simulations of DNase E9-Im9 and DNase E9-Im2 complexes are carried out using GROMACS with the CHARMM36 force-field and the TIP3P water model. The initial conformation of the DNase E9-Im2 complex was adopted from the crystal structure with PDB id 2WPT. Similarly, the initial structure of DNase E9-Im9 was constructed by the PDB entry 1EMV. The size of simulation boxes in both systems was approximately $85 \times 85 \times 85 \text{ \AA}^3$. Both simulation boxes were solvated with 19174 and 19079 water molecules, respectively. The net charge of the simulation box was neutralized by adding three Na^+ ions. Both systems were relaxed for 4 ns at 310 K and 1 atm to remove unrealistic contacts, and they are analyzed along 100 ns trajectories. A uniform integration step of 2 fs was used for all types of interactions, throughout all simulations. A cutoff of 12 $\text{\AA}$ was used for van der Waals interactions, and electrostatic interactions were calculated with the particle mesh technique for Ewald summations, also with a cutoff of 12 $\text{\AA}$. Temperature (310 K) and pressure (1 bar) are controlled using the v-rescale thermostat and the Parrinello-Rahman barostat, respectively.

Acknowledgement

This work was supported by the National Institutes of Health under Grant Numbers R01GM120238 and R01GM122804. The work is also partially supported by a start-up grant from Albert Einstein College of Medicine. Computational support was provided by Albert Einstein College of Medicine High Performance Computing Center.

References

1. MacPherson RE, et al., Skeletal muscle PLIN proteins, ATGL and CGI-58, interactions at rest and following stimulated contraction. *Am J Physiol Regul Integr Comp Physiol*, 2013 304(8): p. R644–50. [PubMed: 23408028]
2. Pawson T and Nash P, Protein-protein interactions define specificity in signal transduction. *Genes Dev*, 2000 14(9): p. 1027–47. [PubMed: 10809663]
3. Barabasi AL and Oltvai ZN, Network biology: understanding the cell's functional organization. *Nat Rev Genet*, 2004 5(2): p. 101–13. [PubMed: 14735121]
4. Kastiris PL and Bonvin AM, On the binding affinity of macromolecular interactions: daring to ask why proteins interact. *J R Soc Interface*, 2013 10(79): p. 20120835. [PubMed: 23235262]
5. Ma B, et al., Folding funnels and binding mechanisms. *Protein Eng*, 1999 12(9): p. 713–20. [PubMed: 10506280]
6. Csermely P, Palotai R, and Nussinov R, Induced fit, conformational selection and independent dynamic segments: an extended view of binding events. *Trends Biochem Sci*, 2010 35(10): p. 539–46. [PubMed: 20541943]
7. Bui JM and McCammon JA, Protein complex formation by acetylcholinesterase and the neurotoxin fasciculin-2 appears to involve an induced-fit mechanism. *Proc Natl Acad Sci U S A*, 2006 103(42): p. 15451–6. [PubMed: 17021015]
8. David A, et al., Protein-protein interaction sites are hot spots for disease-associated nonsynonymous SNPs. *Hum Mutat*, 2012 33(2): p. 359–63. [PubMed: 22072597]
9. Ghirlardo R, The analysis of macromolecular interactions by sedimentation equilibrium. *Methods*, 2011 54(1): p. 145–56. [PubMed: 21167941]
10. Daghestani HN and Day BW, Theory and Applications of Surface Plasmon Resonance, Resonant Mirror, Resonant Waveguide Grating, and Dual Polarization Interferometry Biosensors. *Sensors*, 2010 10(11): p. 9630–9646. [PubMed: 22163431]
11. Velazquez-Campoy A, Leavitt SA, and Freire E, Characterization of protein-protein interactions by isothermal titration calorimetry. *Methods Mol Biol*, 2015 1278: p. 183–204. [PubMed: 25859950]
12. Pierce MM, Raman CS, and Nall BT, Isothermal titration calorimetry of protein-protein interactions. *Methods-a Companion to Methods in Enzymology*, 1999 19(2): p. 213–221.
13. Schreiber G, Haran G, and Zhou HX, Fundamental aspects of protein-protein association kinetics. *Chem Rev*, 2009 109(3): p. 839–60. [PubMed: 19196002]
14. Xie ZR, Chen J, and Wu Y, Predicting Protein-protein Association Rates using Coarse-grained Simulation and Machine Learning. *Sci Rep*, 2017 7: p. 46622. [PubMed: 28418043]
15. Wang B, et al., Integrating Structural Information to Study the Dynamics of Protein-Protein Interactions in Cells. *Structure*, 2018.
16. Alsallaq R and Zhou HX, Prediction of protein-protein association rates from a transition-state theory. *Structure*, 2007 15(2): p. 215–24. [PubMed: 17292839]
17. Alsallaq R and Zhou HX, Energy landscape and transition state of protein-protein association. *Biophys J*, 2007 92(5): p. 1486–502. [PubMed: 17142273]
18. Alsallaq R and Zhou HX, Electrostatic rate enhancement and transient complex of protein-protein association. *Proteins-Structure Function and Bioinformatics*, 2008 71(1): p. 320–335.
19. Qin S, Pang X, and Zhou HX, Automated prediction of protein association rate constants. *Structure*, 2011 19(12): p. 1744–51. [PubMed: 22153497]
20. Agius R, et al., Characterizing changes in the rate of protein-protein dissociation upon interface mutation using hotspot energy and organization. *PLoS Comput Biol*, 2013 9(9): p. e1003216. [PubMed: 24039569]

21. Zhang L, Borthakur S, and Buck M, Dissociation of a Dynamic Protein Complex Studied by All-Atom Molecular Simulations. *Biophys J*, 2016 110(4): p. 877–86. [PubMed: 26910424]
22. Liu FF, et al., Molecular basis for the dissociation dynamics of protein A-immunoglobulin G1 complex. *PLoS One*, 2013 8(6): p. e66935. [PubMed: 23776704]
23. Kleanthous C and Walker D, Immunity proteins: enzyme inhibitors that avoid the active site. *Trends Biochem Sci*, 2001 26(10): p. 624–31. [PubMed: 11590016]
24. Wallis R, et al., Protein-protein interactions in colicin E9 DNase-immunity protein complexes. 2. Cognate and noncognate interactions that span the millimolar to femtomolar affinity range. *Biochemistry*, 1995 34(42): p. 13751–9. [PubMed: 7577967]
25. Kuhlmann UC, et al., Specificity in protein-protein interactions: the structural basis for dual recognition in endonuclease colicin-immunity protein complexes. *J Mol Biol*, 2000 301(5): p. 1163–78. [PubMed: 10966813]
26. Keeble AH, et al., Calorimetric dissection of colicin DNase-immunity protein complex specificity. *Biochemistry*, 2006 45(10): p. 3243–54. [PubMed: 16519519]
27. Li W, et al., Dual recognition and the role of specificity-determining residues in colicin E9 DNase-immunity protein interactions. *Biochemistry*, 1998 37(34): p. 11771–9. [PubMed: 9718299]
28. Meenan NA, et al., The structural and energetic basis for high selectivity in a high-affinity protein-protein interaction. *Proc Natl Acad Sci U S A*, 2010 107(22): p. 10080–5. [PubMed: 20479265]
29. Keeble AH and Kleanthous C, The kinetic basis for dual recognition in colicin endonuclease-immunity protein complexes. *J Mol Biol*, 2005 352(3): p. 656–71. [PubMed: 16109424]
30. Li W, et al., Highly discriminating protein-protein interaction specificities in the context of a conserved binding energy hotspot. *J Mol Biol*, 2004 337(3): p. 743–59. [PubMed: 15019791]
31. Keeble AH, et al., Experimental and computational analyses of the energetic basis for dual recognition of immunity proteins by colicin endonucleases. *J Mol Biol*, 2008 379(4): p. 745–59. [PubMed: 18471830]
32. Grunberg R, Nilges M, and Leckner J, Flexibility and conformational entropy in protein-protein binding. *Structure*, 2006 14(4): p. 683–93. [PubMed: 16615910]
33. Kastitis PL, et al., Proteins feel more than they see: fine-tuning of binding affinity by properties of the non-interacting surface. *J Mol Biol*, 2014 426(14): p. 2632–52. [PubMed: 24768922]
34. Plewczynski D and Ginalski K, The interactome: predicting the protein-protein interactions in cells. *Cell Mol Biol Lett*, 2009 14(1): p. 1–22. [PubMed: 18839074]
35. Janin J and Chothia C, The structure of protein-protein recognition sites. *J Biol Chem*, 1990 265(27): p. 16027–30. [PubMed: 2204619]
36. Xenarios I, et al., DIP: the database of interacting proteins. *Nucleic Acids Res*, 2000 28(1): p. 289–91. [PubMed: 10592249]
37. Zhou HX and Bates PA, Modeling protein association mechanisms and kinetics. *Curr Opin Struct Biol*, 2013 23(6): p. 887–93. [PubMed: 23850142]
38. Chen J, Xie ZR, and Wu Y, Study of protein structural deformations under external mechanical perturbations by a coarse-grained simulation method. *Biomech Model Mechanobiol*, 2016 15(2): p. 317–29. [PubMed: 26049804]
39. Leguebe M, et al., Hybrid molecular mechanics/coarse-grained simulations for structural prediction of G-protein coupled receptor/ligand complexes. *PLoS One*, 2012 7(10): p. e47332. [PubMed: 23094046]
40. Neri M, et al., Coarse-grained model of proteins incorporating atomistic detail of the active site. *Physical Review Letters*, 2005 95(21): p. 4.
41. Krivov GG, Shapovalov MV, and Dunbrack RL Jr., Improved prediction of protein side-chain conformations with SCWRL4. *Proteins*, 2009 77(4): p. 778–95. [PubMed: 19603484]
42. Ravikumar KM, Huang W, and Yang S, Coarse-grained simulations of protein-protein association: an energy landscape perspective. *Biophys J*, 2012 103(4): p. 837–45. [PubMed: 22947945]
43. Zhou HY and Zhou YQ, Distance-scaled, finite ideal-gas reference state improves structure-derived potentials of mean force for structure selection and stability prediction. *Protein Science*, 2002 11(11): p. 2714–2726. [PubMed: 12381853]

44. Mosca R, et al., 3did: a catalog of domain-based interactions of known three-dimensional structure. *Nucleic Acids Res*, 2014 42(Database issue): p. D374–9. [PubMed: 24081580]
45. Finn RD, et al., iPfam: a database of protein family and domain interactions found in the Protein Data Bank. *Nucleic Acids Res*, 2014 42(Database issue): p. D364–73. [PubMed: 24297255]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 1:

All atom molecular dynamic (MD) simulations were carried out on two systems: the protein complex between E9 DNase and Im9 versus the protein complex between E9 DNase and Im2. Both dimers were superimposed in (a). The chains of E9 DNase and Im9 in one dimer are shown in yellow and red, while the chains of E9 DNase and Im2 in the other dimer are shown in blue and green, respectively. Along the MD simulations, the backbone RMSD are plotted in (b) as a function of time for both complexes. Furthermore, the RMSF of each residue in two complexes are calculated in (c). In both (b) and (c), the black curves represent the complex between E9 DNase and Im2, while the red curves represent the complex between E9 DNase and Im9.

Figure 2:

The native structure of the complex between E9 DNase and immunity proteins Im9 is plotted in (a). The backbone of E9 DNase is shown in red, while the backbone of Im9 is shown in green. The side-chain centers of mass in the complex are represented as cyan sticks. Using the same representation, the final structural conformation of the complex at the end of one simulation trajectory is shown in (b). The changes of calculated percentage of native contacts and interface energy along the same simulation trajectory are plotted in (c) and (d), respectively. The figure shows that as the energy at binding interface of the complex increases, around one third of the native contacts have been broken by the end of the simulation trajectory.

Figure 3:

The contact map for residues that form inter-molecular interactions in the native complex between E9 DNase and Im9 protein is plotted in (a). The helix II and helix III regions in the immunity protein are shown by blue and yellow shaded areas, respectively. In contrast, the inter-molecular contact map of E9 DNase/Im9 complex after simulation is plotted in (b). Similarly, the inter-molecular contact map of native E9 DNase/Im2 complex is plotted in (c), while the inter-molecular contact map of E9 DNase/Im2 complex after simulation is plotted in (d).

Figure 4:

The sidechain contacts at the binding interface between the native structure of E9 DNase and protein Im9 are shown in (a), while the similar sidechain contacts at the binding interface between E9 DNase and Im2 are shown in (b). In both plots, the E9 DNase is shown in yellow on the left, while the immunity protein ligands are shown in light blue on the right. Residues that form native interactions are highlighted in both complexes. The residues in helix II of immunity proteins are more variable than the residues in helix III. For instance, the hydrophobic residue V34 in the helix II of Im9 is replaced by asparagine in Im2.

Figure 5:

The comparison of final calculated PNC between E9 DNase/Im9 and E9 DNase/Im2 complexes are plotted in (a) for all 10 trajectories. Similarly, the comparison of final calculated PNC between wild-type E9 DNase/Im9 complex and two mutated systems are plotted in (b) for all 10 trajectories. The values of PNC in both plots are ranked in decreasing order.

Figure 6:

The root-mean-square fluctuations for all residues in E9 DNase are plotted in (a). The results with reduced strength of force constants are shown by the red curve, while the results of the original system in which the proteins are less flexible are shown by the black curve. Moreover, the comparison of final calculated PNC between these two systems are plotted in (b) for all 10 trajectories. The values of PNC are ranked in decreasing order.

Figure 7:

The interactions between proteins in the complex are described by the knowledge-based potential. The pair-wise potential was constructed by collecting the corresponding information of binding in a large-scale structural library that consists of 4960 entries of protein-protein interactions from either homodimers or heterodimers. Detailed values of these pair-wise energy parameters are plotted as two-dimensional contour, indexed by types of amino acids. The regions that correspond to the first-principle physical and chemical characteristics of molecular recognition are also indicated in the figure.