PROTEINS
STRUCTURE ■ FUNCTION ■ BIOINFORMATICS

Multi-scale characterization of the energy landscape of proteins with application to the C3d/Efb-C complex

Nurit Haspel,[1] Brian V. Geisbrecht,[2] John Lambris,[3] and Lydia Kavraki[1]*

[1] Department of Computer Science, Rice University, Houston, Texas 77005
[2] Division of Cell Biology and Biophysics, School of Biological Sciences, University of Missouri, Kansas City, Missouri
[3] Department of Pathology and Laboratory Medicine, University of Pennsylvania, Philadelphia, Pennsylvania

ABSTRACT

We present a novel multi-level methodology to explore and characterize the low energy landscape and the thermodynamics of proteins. Traditional conformational search methods typically explore only a small portion of the conformational space of proteins and are hard to apply to large proteins due to the large amount of calculations required. In our multi-scale approach, we first provide an initial characterization of the equilibrium state ensemble of a protein using an efficient computational conformational sampling method. We then enrich the obtained ensemble by performing short Molecular Dynamics (MD) simulations on selected conformations from the ensembles as starting points. To facilitate the analysis of the results, we project the resulting conformations on a low-dimensional landscape to efficiently focus on important interactions and examine low energy regions. This methodology provides a more extensive sampling of the low energy landscape than an MD simulation starting from a single crystal structure as it explores multiple trajectories of the protein. This enables us to obtain a broader view of the dynamics of proteins and it can help in understanding complex binding, improving docking results and more. In this work, we apply the methodology to provide an extensive characterization of the bound complexes of the C3d fragment of human Complement component C3 and one of its powerful bacterial inhibitors, the inhibitory domain of Staphylococcus aureus extra-cellular fibrinogen-binding domain (Efb-C) and two of its mutants. We characterize several important interactions along the binding interface and define low free energy regions in the three complexes.

Proteins 2010; 00:000–000.
© 2009 Wiley-Liss, Inc.

Key words: protein dynamics; MD simulations; multi-scale methods; Efb-C/C3d; immune system inhibition.

INTRODUCTION

Proteins are flexible molecules[1–3] that can assume various conformations via structural changes that range from small-scale movements to large domain motions. From a computational point of view, the prediction and modeling of protein flexibility is a challenging problem due to the large number of calculations required. Conformational modeling methods such as Molecular Dynamics (MD)[4] and Monte Carlo[5] can sample atomic level dynamic processes, yet their usefulness is limited for several reasons. First, they require large computational resources, but only allow for modeling of interactions that take place on very small time scales (e.g., several hundreds of nanoseconds). This is a significant shortcoming, since most physiological processes occur within milliseconds or seconds. Second, because the computational resources needed to conduct these studies scale up quickly with the size of the protein, carrying out a computational simulation of large proteins or protein complexes is especially time consuming. Finally, an MD simulation outputs a single trajectory that depends on the accuracy of the starting model, which is usually a single crystal structure that represents the average conformation at equilibrium state. As a result, a single MD simulation may only cover a small portion of the conformational space of a protein. To extend the capability of MD, several approaches have been proposed. Computational multi-scale methods have been shown to work well in characterizing the large-scale molecular motions that underlie the function of many proteins.[6] These methods use local perturbations,[7] coarse grained modeling,[8] elastic networks,[9] and other approaches to sample larger portion of protein dynamics than MD. The subject is further reviewed elsewhere.[6]

Earlier work in our research group includes the Fragment/Protein Ensemble Method FEM/PEM.[10,11] This method uses an efficient sampling algorithm inspired by robotics to characterize

Additional Supporting Information may be found in the online version of this article.
Grant sponsor: NIH; Grant numbers: 5R01GM078988; 5R01AI071028; Grant sponsor: National Science Foundation; Grant number: CNS-0421109. Teragrid allocation grants MCB080051 and MCB080118.
*Correspondence to: Lydia Kavraki, Department of Computer Science, Rice University, Houston, TX 77005. E-mail: kavraki@rice.edu.
Received 23 June 2009; Revised 24 August 2009; Accepted 8 September 2009
Published online 28 September 2009 in Wiley InterScience (www.interscience.wiley.com).
DOI: 10.1002/prot.22624

the low energy equilibrium state ensemble for a given protein under physiological conditions. Later, using only sequence data and a novel multi-scale approach, our group produced conformational ensembles of single, relatively small proteins that have multiple functional states such as Calbindin, Calmodulin, and Adenylate Kinase.[12]

Inspired by the success of that study, in this article, we take a different approach. We consider protein complexes where the size of the complex makes the application of the earlier methods computationally infeasible. We develop a method to provide a multi-scale characterization of the local conformational flexibility and thermodynamics of proteins by combining an efficient conformational modeling algorithm, MD simulations, and a non-linear dimensionality reduction technique.[13] Our work seeks to further enrich the set of available approaches for the exploration of protein dynamics. We first generate an initial conformational ensemble for the tested proteins using the fragment ensemble method (FEM).[10,11] We then enrich the initial ensemble by conducting short (2 ns) MD simulations using each of the initial ensemble conformations as a distinct starting point. We show that this can provide a more extensive sampling of the conformational space and produce multiple trajectories rather than a single one, as is typically the case for a long MD simulation starting from a single structure. Using thousands of processors we are able to cover a larger portion of the conformational space in an equivalent amount of time, whereas a single MD simulation generally scales up to several hundreds of processors.[13,14] To facilitate data analysis, the resulting multiple trajectories are then embedded on a low-dimensional landscape using a non-linear dimensionality reduction technique called SciMap.[15] This method provides a small number of collective coordinates that best characterizes the variance in the data. These global coordinates are used as reaction coordinates for free energy calculations and enable the analysis and visualization of datasets consisting of a large number of individual observations. Non-linear dimensionality reduction techniques were shown to provide a better characterization of complex, non-linear events such as protein folding and binding than linear methods such as principal component analysis (PCA).[16]

As shown by our results, the proposed methodology is suitable for extensive characterization of local conformational changes in proteins and complexes, as it uses an adaptation of two local conformational search methods, FEM and MD. Caution should be taken when applying this method to complexes where there are known significant or allosteric changes to the proteins upon complex binding. In these cases, the method is not directly applicable. The ensemble generation should use multiple distinct conformations of the involved molecules and docking methods may be applied to aid in assembling the complex following the initial ensemble generation. Modeling of large-scale conformational transitions is beyond the scope of this article.

We applied our method to the complex formed between the C3d domain of human Complement Component C3 and one of its bacterial inhibitors, the *Staphylococcus aureus* (*S. aureus*) extra-cellular fibrinogen-binding protein (Efb-C). Previous studies[17] identified Arg-131 and Asn-138 as two residues on Efb-C that create a number of discrete contacts with C3d, and thus, play an important role in formation and maintainence of the Efb-C/C3d complex. Simultaneous mutation of both residues to either alanine (RA/NA) or glutamic acid (RE/NE) resulted in a complete loss of both C3d binding and complement inhibition, whereas the single mutants, R131A and N138A, formed stable complexes that still retained some function. Previously,[18] we used a combination of crystallography, isothermal titration calorimetry (ITC), surface plasmon resonance, and MD simulations to characterize the thermodynamics, kinetics, and energetics of the complex and the two single mutants, N138A and R131A. We found that while the mutations had little effect on the structure of the complex, they had a significant adverse effect on the binding energy and the kinetics of the complex. We further characterized several potential, though previously unidentified interactions along the Efb-C/C3d binding interface that appear to contribute to the intricate network of salt bridges and hydrogen bonds that anchor Efb-C to C3d and that support its potent complement inhibitory properties. In this work, using our extended analysis of the complexes described above, we characterize several distinct low free energy states for each of these three complexes. Using our previous knowledge about the binding interface between C3d and Efb-C, we analyze the low energy states, with the goal of describing correlations between low energy regions and specific interactions. We find that the low free energy regions correspond to a large number of native contacts between C3d and Efb-C. In addition to Arg-131 and Asn-138, we find that both the N- and C-terminal portions of Efb-C and several other residues located on helices $\alpha 2$ and $\alpha 3$ play a major role in the binding of C3d in both the wildtype and mutant complexes. The findings reported here provide further insight into the contribution of individual residues of Efb-C in disrupting C3 function.

METHODS

Generation of initial equilibrium state ensembles for C3d and Efb-C

Sampling of protein conformational space was done using the FEM.[10,11] This algorithm models flexible regions in proteins and produces an ensemble of conformations representing the near-equilibrium conformational preference of the input protein. It uses cyclic coordinate descent (CCD)[19] and subsequently minimizes the generated conformations using the AMBER program[4,20] for a

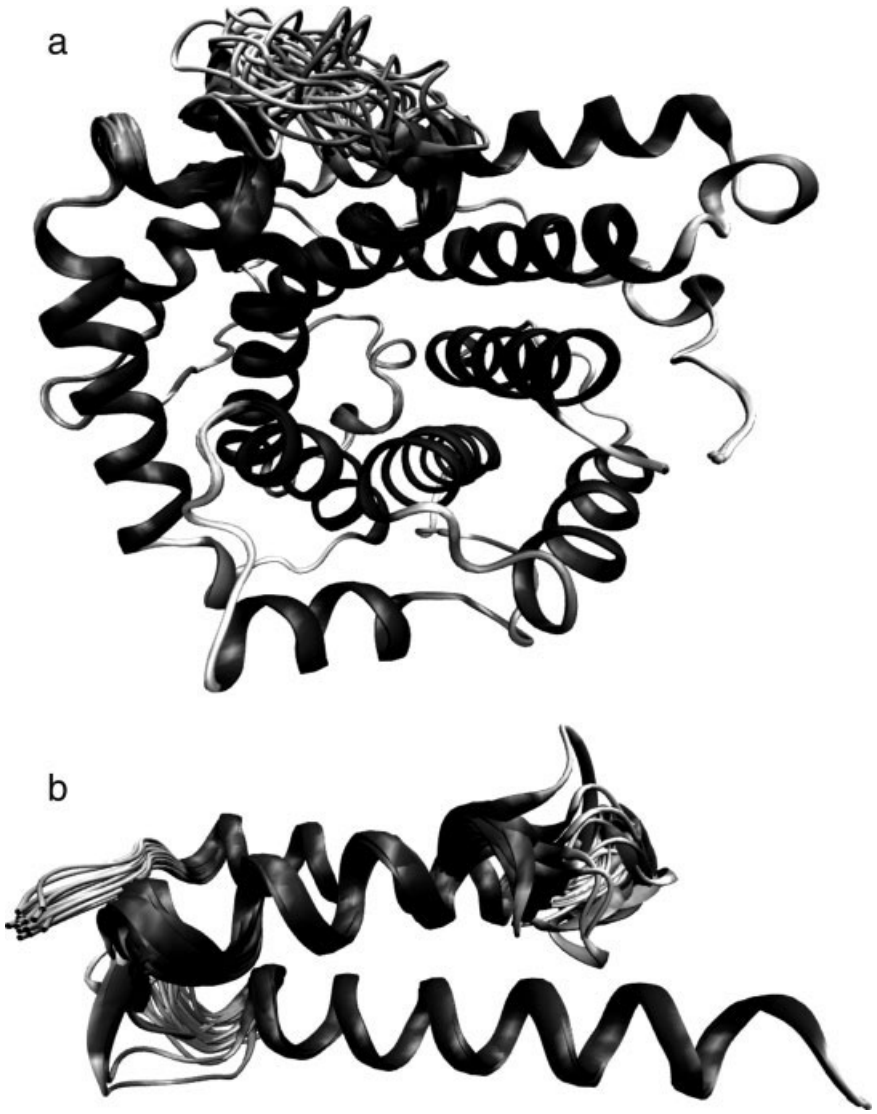

Figure 1

The initial low energy ensemble generated for the C3d molecule (**a**) and Efb-C molecule (**b**) of the wildtype complex. The modeling of C3d was done only on the three binding loops. The lowest energy conformations of this ensemble were subject to further enrichment by multiple MD simulations. Helices are depicted in dark gray and loops in light gray.

duration of 1500 steps to allow relaxation of the structures without causing large structural changes. The resulting minimized conformations are weighted according to their Boltzmann probability and only those within 10 kcal/mol of the native structure are retained.

The initial structures for C3d and Efb-C were taken from PDB accession code 2GOX for the wildtype complex and the structures for the N138A and R131A were taken from PDB accession codes 3D5S and 3D5R, respectively.[18] To generate the ensemble to the C3d/Efb-C complex, we generated the conformational ensembles for each of C3d and Efb-C separately using FEM. We then combinatorially assembled each conformation generated for C3d with each conformation generated for Efb-C to produce an ensemble for the entire C3d/Efb-C complex. The process was repeated similarly for both the N138A and R131A mutants. An initial set of 10,000 conformations was generated for each of the three C3d structures and each of the wildtype and mutant Efb-C. The ensembles were generated for all of Efb-C and residues 1029–1050, 1089–1099, and 1157–1166 of C3d, which comprise the three inter-helix loops that provide the binding interface for Efb-C. Following energy minimization, a total of 90, 78, and 84 starting structures constituted the initial ensembles for the wildtype, R131A, and N138A complexes, respectively. Figure 1 shows an example of the ensembles generated for the wildtype complex. No docking was needed for the assembly since the input confor-

mations were obtained directly from the refined coordinates of the relevant co-crystal structures.[18] FEM was applied to only limited regions of C3d and did not cause drastic structural changes (see Fig. 1). Each assembled complex was subject to energy minimization prior to the MD simulation to resolve clashes that may have occurred as a result of the assembly.

Molecular dynamics setup

Each conformation generated as described in the previous section was the starting point of an MD simulation. Prior to this, missing side chains from the PDB files, as well as hydrogen atoms, were added using the AMBER 9 leap utility.[21] The simulations were conducted using the AMBER ff03 force field.[22] Implicit solvent was used via the General Born solvation method.[23] Each conformation was subject to 2500 steepest descent minimization steps[24] followed by 2500 conjugate gradient minimization steps.[25] This minimization was required to resolve initial clashes between the C3d and Efb-C structures, which were generated separately as described earlier, and to allow relaxation of the structure. Minimization was followed by 20 ps of gradual heating to 300 K and equilibration. The MD simulations were performed for 2 ns for each equilibrated conformation using a constant temperature of 300 K. The SHAKE algorithm[26] was used to restrain the length of bonds involving hydrogen. This allowed an integration time step of 2 fs.

Binding energy calculation

The binding energy of the complex was estimated using the mm-pbsa/gbsa approach implemented as a part of the AMBER package.[27] The average complex binding energy is obtained using the following formula:

$$G_{\text{binding}} = G_{\text{total;complex}} - G_{\text{total;C3d}} - G_{\text{total;Efb-C}}$$

where:

$$G_{\text{total}} = E_{\text{MM;gas}} + G_{\text{solv}}$$

$E_{\text{MM;gas}}$ is gas phase molecular mechanics energy and G_{solv} is the change in the free energy upon solvation. The solvation free energy is the sum of two terms, the electrostatic contribution and the non-polar contribution:

$$G_{\text{solv}} = G_{\text{el}} + G_{\text{np}}$$

The electrostatic contribution was estimated using the General Born method.[23,28,29] The non-polar term (G_{np}) was calculated from the solvent accessible surface area (SASA), which is estimated using the LPCO method.[30] γ was set to 0.0072 kcal/(mol*A^2) and β was set to 0 kcal/mol in the following equation:

$$G_{\text{np}} = \gamma^{*}\text{SASA} + \beta$$

In the mm-pbsa/gbsa approach, the binding energy is calculated on an ensemble of uncorrelated snapshots extracted from the equilibrated simulation. A snapshot was taken every 5 ps. The first 500 ps of the simulations were excluded from the binding and free energy calculations to ensure that the simulation has equilibrated. The binding energy was calculated using the single trajectory approach. The C3d and Efb-C molecules were not simulated separately, but rather the snapshot structures for the energy calculations of the C3d-Efb-C complex and separated C3d and Efb-C were taken from the MD trajectory of the C3d/Efb-C complex.[18]

Dimensionality reduction and free energy calculation

The trajectories of all MD simulations for each complex were projected onto a low-dimensional subspace using SciMAP,[15] a non-linear dimensionality reduction technique based on the Isomap algorithm.[31] The values of the reaction coordinates are scaled and are specific to the input set of coordinates thus the actual value of the reaction coordinates cannot be compared across two separate sets of conformations.

The low-dimensional representations of the trajectories were used as reaction coordinates for potential of mean force (PMF) calculations using the Weighted Histogram Analysis Method (WHAM),[32] with the binding energy mentioned above as the potential energy. WHAM outputs non-negative values since the PMF values are calculated relative to a minimum reference value, hence, the positive free energy values in Figure 3.

RESULTS AND DISCUSSION

The low energy landscape of the wildtype and mutants C3d/Efb-C

Ninety initial conformations were used as starting points for MD simulations of the wildtype, 78 conformations were generated for the R131A mutant and 84 conformations were generated for the N138A mutants. Overall 180 ns, 156 ns, and 168 ns were simulated for the wildtype, the R131A and the N138A mutants, respectively. Each set of trajectories produced 27,000, 23,400, and 25,200 conformations, respectively. To test whether the simulations had reached equilibrium, which is necessary for free energy calculations, we measured the evolution of the potential energy as a function of the simulation time for the last 1.5 ns of the simulations (as mentioned above, the first 0.5 ns was excluded from the calculation to assure convergence). Figure 2 shows an example of such a calculation for one conformation of each of the wildtype and mutant complexes. As seen, the potential energy changes very little and remains within a

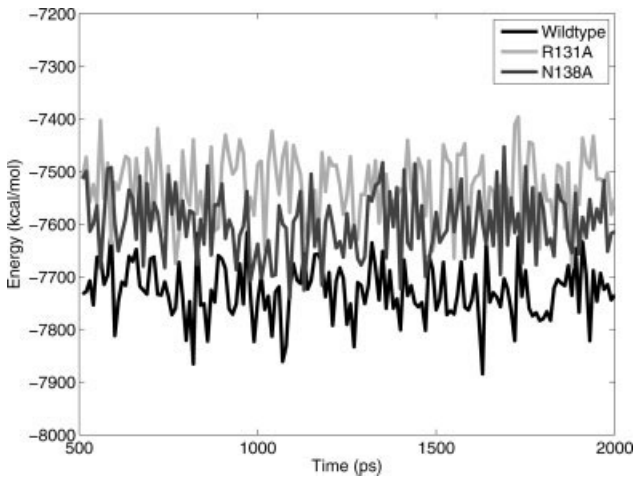

Figure 2

The evolution of the potential energy during the simulation time for an example of a wildtype, R131A mutant and N138A mutant conformation. The values are shown for the last 1.5 ns of the MD simulation. These conformations represent the minimum free energy values shown in Figures 4–6.

range of 250 kcal/mol. The root mean square fluctuation (RMSF) of the potential energy was ∼56–58 kcal/mol. Other conformations display a similar pattern (results not shown). The conformations were the input for binding energy calculations and for low-dimensional embedding using SciMap (see Methods section for details about the MD simulation setup and the criteria for selecting conformations for free energy calculations). In each case, low-dimensional embedding produced three reaction coordinates which captured ∼75–80% of the variance in the data. Figure 3 shows the energy surfaces of the three complexes along the first two reaction coordinates, which encompass ∼60% of the variance in the data in each case. The data points are colored by energy values on the red (high energy) to blue (low energy) scale. The potential energy maps along the first and second reaction coordinates are shown in Figure 3(a,c,e) for the wildtype, R131A mutant and N138A mutant, respectively. Generally, the reaction coordinates do not necessarily carry a biological meaning. However, in this case, the values of the first reaction coordinate were well-correlated with the values of the potential energy for each conformation (correlation coefficients of 0.42, −0.51, and −0.47, respectively). We also found by visual inspection that in all three cases, the first reaction coordinate was also correlated to the position of the 1089–1099 loop of C3d of the given conformation with respect to Efb-C. This accounts for a large number of native contacts formed between this loop and helices α2 and 3 of Efb-C. Figure 3(b,d,f) show the free energy surfaces of the complexes as a function of the first and the second reaction coordinates of the wildtype and the R131A and N138A mutants, respectively. In the following sections, we present an in-depth analysis of the three free energy surface map.

The low energy landscape of the wildtype C3d/Efb-C

Figure 3(b) shows a clear free energy minimum at the top left part of the figure, as well as several other local minima. The low free energy conformations are all associated with particularly strong interactions between the 1089–1099 loop in the C3d and helix α2–3 of Efb-C. Figure 4(a) shows an example of a conformation corresponding to the low energy minimum on Figure 3(b). Maps for the pairwise interactions between C3d and Efb-C are shown in Figure 4(b,c,d). For clarity, only interactions lower than −1.5 kcal/mol are displayed. Notable interactions include the 1089–1099 loop with helices α2 and 3, a set of interactions between the N-terminal domain of Efb-C and the 1057–1066 loop of C3d, and a strong interaction between the C-terminal residue of Efb-C and the 1029–1050 loop of C3d. Strong interactions in the N-terminal domain of Efb-C include the following: Glu-1159, Glu-1160/Lys-110, Glu-1159/Lys-107, and multiple hydrogen bonds involving Gln1161 with the N-terminal domain of Efb-C and Ser-101 with the 1157–1176 loop of C3d. Interactions along the 1089–1098 loop include Lys-145/Asp-1096, Lys-148/Ile-1095-Ser-1097, as well as multiple hydrogen bonds formed by Asn-138 and by Ser-1097. The strong involvement of the N-terminal Lysine residues of Efb-C correlates well with our previous studies[17] and further emphasizes the importance of the salt bridges involving these residues to the binding interaction. In addition, these findings reaffirm the role Asn-138 plays in complex binding, which was detected to a lesser extent in our previous study.[17] The 1029–1050 loop is anchored to Efb-C by a network of salt bridges between Arg-131/Asp-1029-Glu-1030 and between Arg-165/Glu-1032,Glu-1035 as well as Lys-160/Glu-1047 and Asp-156/Lys-1050. This loop showed very little mobility in our analysis, which further reaffirms the role played by Arg-131 and Arg-165 in maintaining the stability of the binding interface.

The low energy surfaces of the R131A mutant

The free energy map of the R131A mutant is shown in Figure 3(d). A broad low energy region can be seen in the map. Two distinct minima can be viewed within this region. An examination of the low energy regions shows that as is the case with the wildtype, they too are associated with a large number of pairwise contacts involving the N-terminal region and helices α2–3 in Efb-C. Figure 5(a) shows an example of a conformation associated with the low minimum A. Plots displaying pairwise interactions are shown in Figure 5(b,c,d), in a similar fashion to the wildtype above. The absence of Arg-131 is partly compensated for by the hydrogen bond formed by the

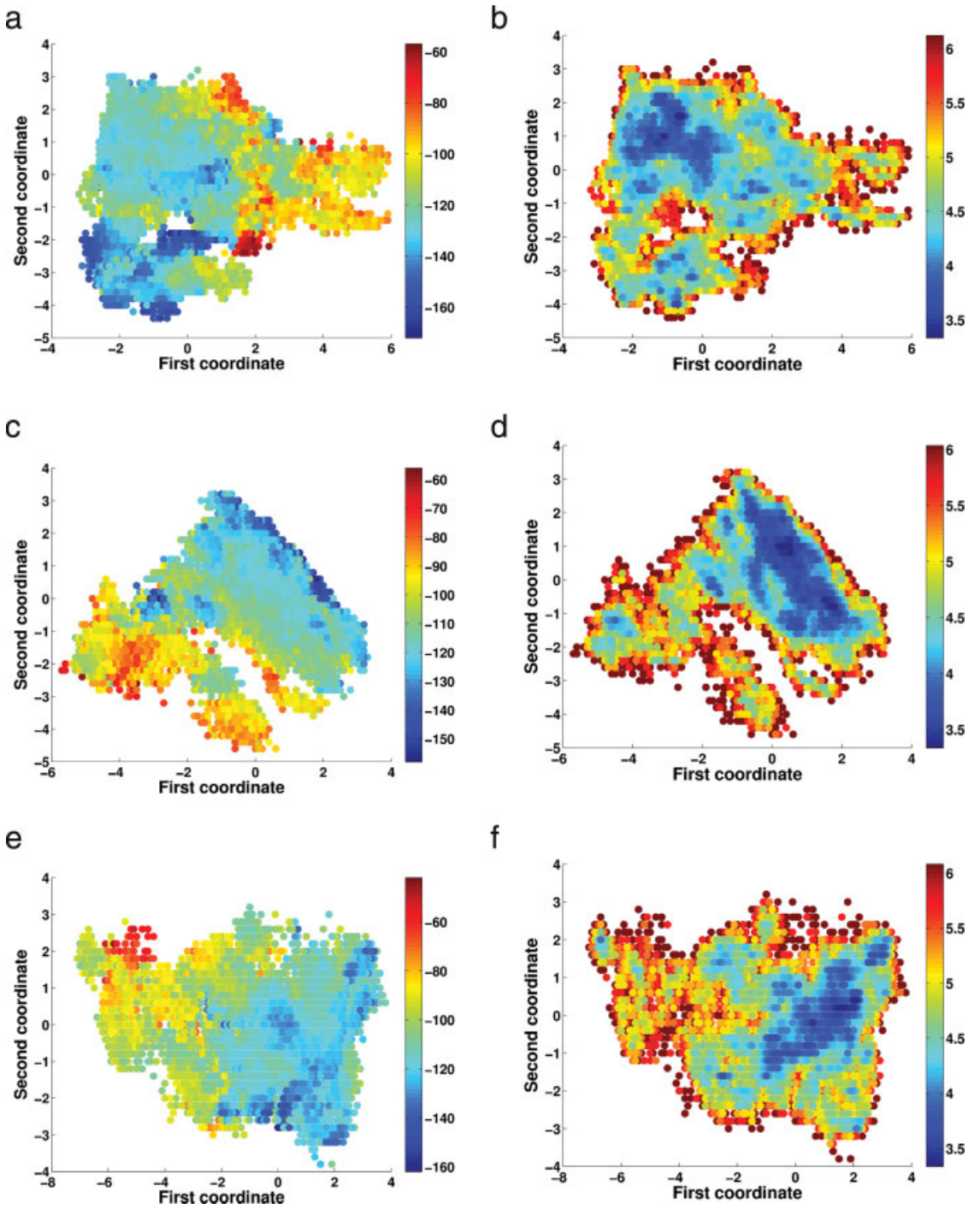

Figure 3

A two-dimensional representation of the potential energy (**a,c,e**) and free energy (**b,d,f**) of the ensemble generated from short MD simulations for the C3d/Efb-C complex. (a) and (b) show the wildtype complex, (c) and (d) show the R131A mutant, and (e) and (f) show the N138A mutant. The points are colored according to their free energy: Shades of blue indicate lower energy and shades of red indicate higher energy. Notice the different energy value ranges. The energy is measured in kcal/mol. The free energy values in (b,d,f) are relative to a minimum free energy conformation, as output by WHAM,[29] and therefore, are non-negative. Refer to the Methods section for an explanation.

backbone of Ala-132 and the side-chain of Asn-1091 in C3d; In addition, there is a large number of contacts formed by nearby residues with residues 1028–1042 of C3d, which help in anchoring Efb-C to the 1029–1042 loop of C3d. A bifurcated salt bridge involving Arg-165/Glu-1032,Glu-1035 is also observed as it appears in the

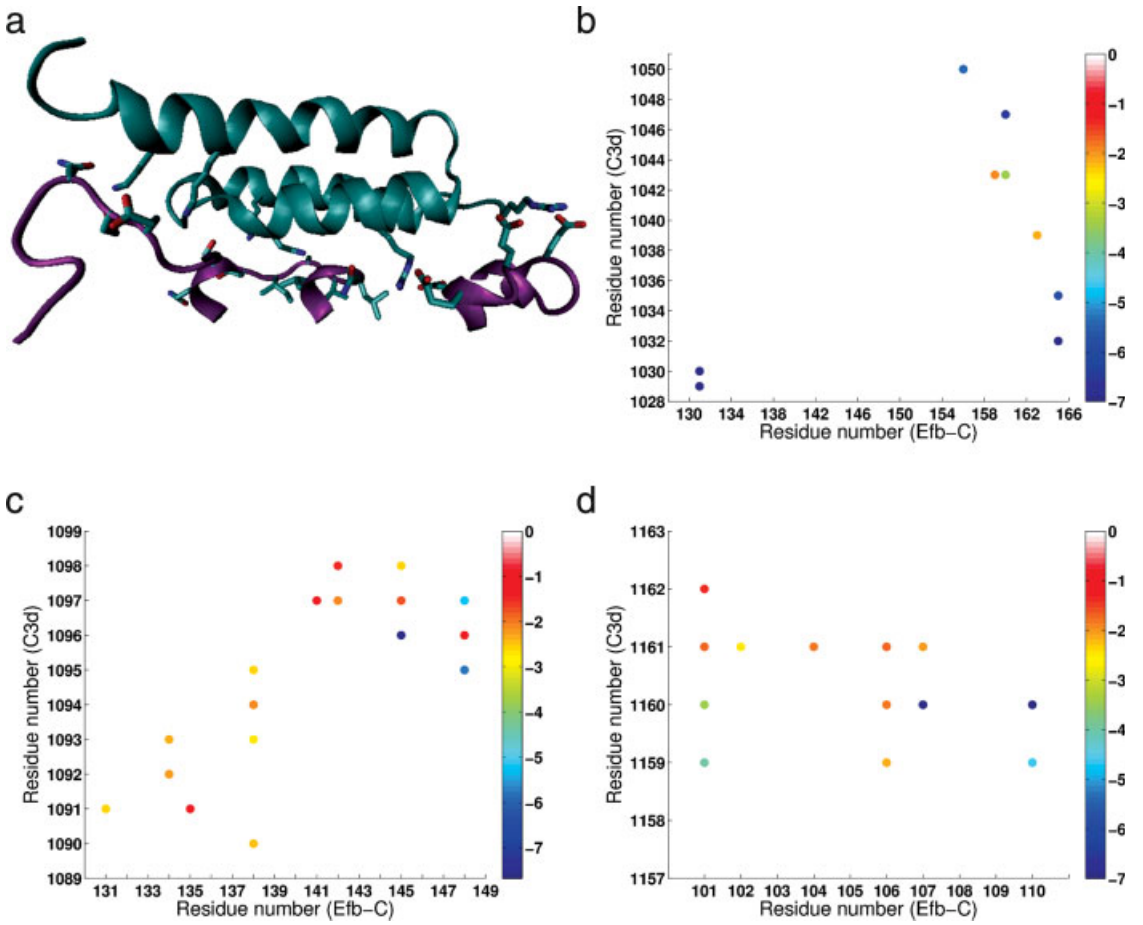

Figure 4

Representative structure from the minimum free energy conformations of the wildtype C3d/Efb-C complex. The free energy landscape of the complex can be found in Figure 2(b). (**a**) Structure of Efb-C (green) and the conformation of the binding loops from C3d (purple). (**b–d**) Energetic contributions of each pairwise interaction between Efb-C and C3d residues along the 1029–1050 loop (b), 1089–1099 loop (c), and 1157–1167 loop (d).

wildtype complex. Most of the other interactions in that area are hydrogen bonds which are weaker than the salt bridges formed by Arg-131. Together, these observations better explain the lower-affinity complex formed by R131A mutant when compared to wildtype. Other notable interactions in the N-terminal domain of Efb-C involve Glu-1159/Lys-110, Gln-1061/Lys-110, and Glu-1159/Lys-107. Notable interactions along the 1089–1098 loop include Lys-145/Asp-1096, Lys-148/Ile-1095-Ser-1097 as well as multiple hydrogen bonds formed by Asn-138 and by other residues in its vicinity. This is similar to the picture seen in the wildtype complex, although the specific residues involve in the interactions may be different.

The low energy surfaces of the N138A mutant

The free energy map of the N138A mutant is shown in Figure 3(f). A low energy region can be seen in the right side of the free energy map. An examination of the low energy region shows that, as is the case with the wildtype complex and the R131A mutant, it is associated with a large number of pairwise contacts involving helices α2–3 and the N-terminal region in Efb-C. However, since Asn-138 is missing, the number of contacts involving helix α2 is significantly smaller than in the case of the previous two examples. Figure 6(a) shows an example of a conformation associated with the low energy minimum. Plots showing interactions involving the two are shown in Figure 6(b,c,d). The strongest interaction between the N-terminal domain of Efb-C and C3d is Glu-1160/Lys-110. Arg-131 is associated with two strong salt bridges with Asp-1029 and Glu-1030 and participates in a hydrogen bond with Asn-1091. Other strong interactions are found between Arg-165/Glu1032,Glu-1035, and Lys-148/Asp-1096. As in the previous examples, this set of interactions anchors Efb-C to the binding interface of C3d. The absence of Asn-138, which makes a large number of indi-

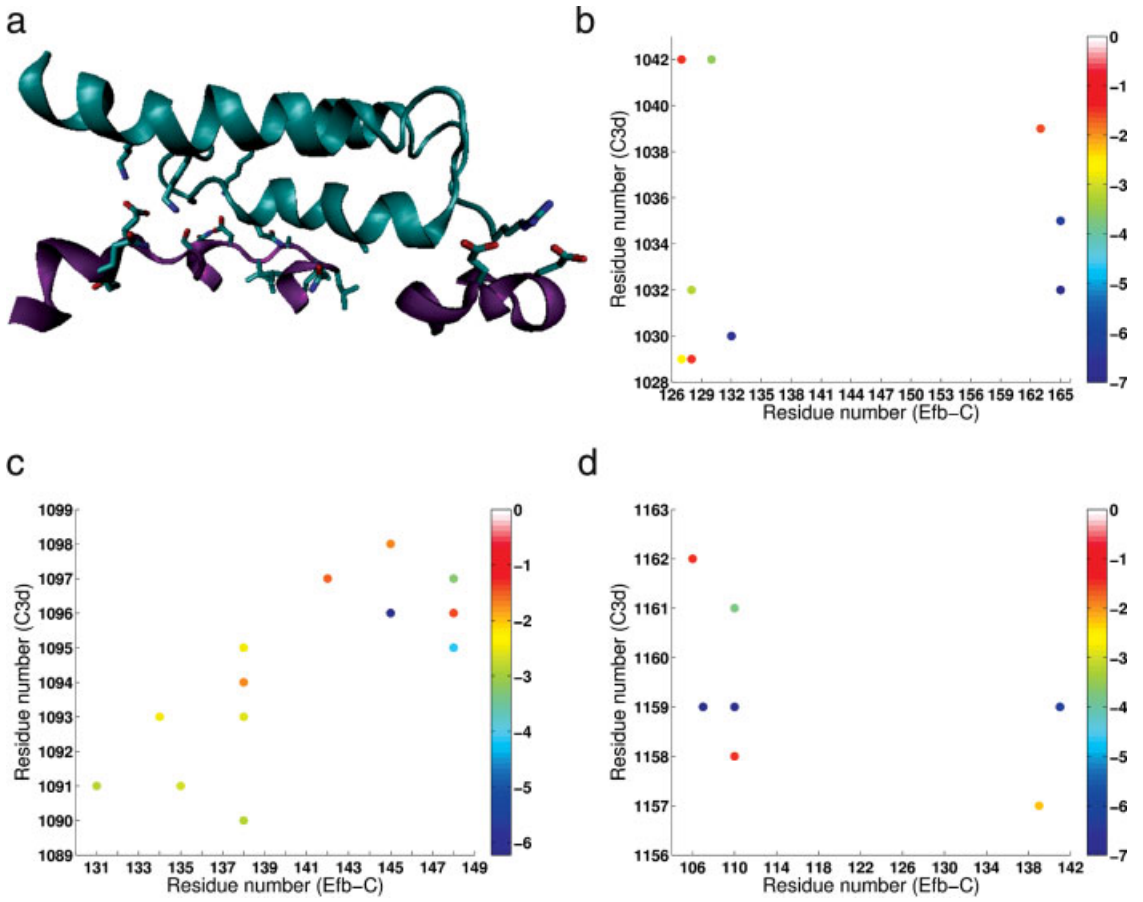

Figure 5

Representative structure from the minimum free energy conformations of the R131A mutant complex. The free energy landscape of the complex can be found in Figure 2(d). (**a**) Structure of Efb-C (green) and the conformation of the binding loops from C3d (purple). (**b–d**) Energetic contributions of each pairwise interaction between Efb-C and C3d residues along the 1029–1050 loop (b), 1089–1099 loop (c), and 1157–1167 loop (d).

vidual contacts with C3d can explain the previously known fact that N138A forms a weaker complex than the wildtype.

Conserved interactions

All the low energy regions described above exhibit a variety of pairwise contacts between Efb-C and C3d. However, several interactions are conserved throughout all of the low energy conformations we have associated with low free energy. Foremost, those interactions are the salt bridges formed by Arg-131 and the 1029–1042 loop and the multiple contacts formed by Asn-138 and the 1089–1099 loop, whose absences have previously been shown to compromise Efb-C binding to C3, and, therefore, its ability to inhibit C3 function.[17,18,33] Other than these previously known interactions, all the conformations exhibited a network of salt bridges and hydrogen bonds between the lysine-rich N -terminal part of Efb-C and the 1057–1067 loop of C3d, which involved mostly Lys-107 and Lys-110 with Glu-1159/1160. Other interactions are Arg-165/Glu-1032,Glu-1035, Lys-148/Ile1095-Ser1097, and Arg-131/Asn-1091. The latter is a hydrogen bond between the backbone of Arg-131 and the side-chain of Asn-1091; it is worth noting that this bond is also found in the R131A mutant. The results are summarized in Table I. Our results further emphasize the role played by Arg-131 and Asn-138, as well as the lysine-rich region found near the N-terminus of Efb-C in promoting C3d binding.

Performance analysis

To quantitatively analyze the amount of coverage obtained by our method when compared to a single MD simulation, we present the following calculation: A 1 ns of MD simulation as performed in our previous work[18] and in this work requires ~6 h using 64 CPUs on a Cray XT3 machine. A single 20 ns MD simulation would therefore require a total of 120 h. Using the method

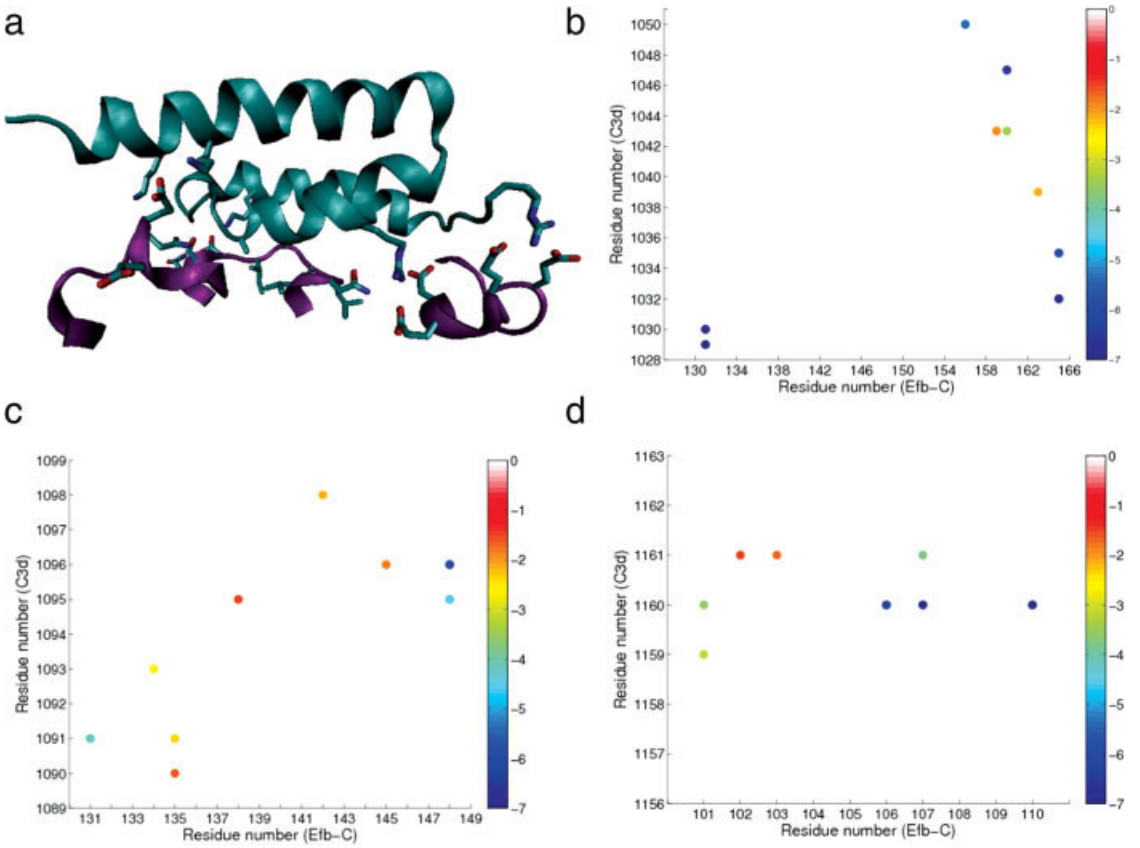

Figure 6

Representative structure from the minimum free energy conformations of the N138A mutant complex. The free energy landscape of the complex can be found in Figure 2(f). (**a**) Structure of Efb-C (green) and the conformation of the binding loops from C3d (purple). (**b–d**) Energetic contributions of each pairwise interaction between Efb-C and C3d residues along the 1029–1050 loop (b), 1089–1099 loop (c), and 1157–1167 loop (d).

outlined in this article, we can utilize 640 processors to invoke 10 simultaneous 2 ns MD simulations in 12 h. Using all 640 processors in a single MD simulation trying to get the same effect will not result in linear speedup due to the nature of the MPI protocol used by standard MD packages.[20,34] This way we achieve a practical linear speedup as a function of the number of processors, at the cost of losing time-related information which can be achieved by a single, long MD simulation. A significant gain in performance is still obvious after taking into account the time spent on generating the initial conformations using FEM. FEM is embarrassingly parallel since the loops are generated independently of one another.[11] When the generation of 10,000 conformations is divided into 200 processors, each producing 50 conformations, each processor requires ~1.5 h on a Cray XD1 machine

Table I

Conserved Interactions and their Values in the Low Energy Conformations of C3d/Efb-C and the Two Mutants

Residues involved (Efb-C/C3d)	Interaction energy (kcal/mol), wildtype	Interaction energy (kcal/mol), R131A mutant	Interaction energy (kcal/mol), N138A mutant
Lys-106-107, Lys-110/Glu-1159, Glu1160	−26.3	−17.53	−25.25
Arg-131/Asn-1091	−2.59	−2.89	−4.22
Lys-148/Ile-1095-Ser-1097	−12.45	−8.86	−10.62
Arg-165/Glu-1032, Glu-1035	−13.36	−15.96	−18.54
Arg-131/Asp-1029, Glu-1030	−16.81	N/A	−12.79
Asn-138/Val-1090, Ile-1093-1095	−9.8	−9.74	N/A

The energy values may be the sum of several individual interactions.

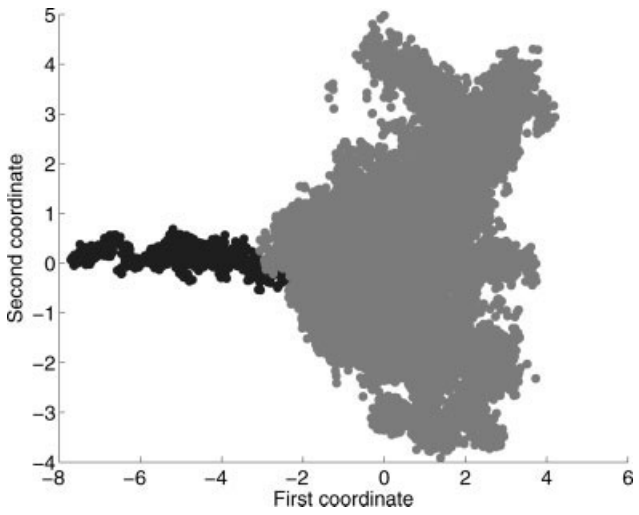

Figure 7

The conformational landscape of the wildtype C3d/Efb-C complex. The landscape represents both the trajectories obtained by a 20 ns MD simulation in our previous work[17] (dark shade) and the analysis performed in this study (light shade).

for C3d and 15 min for Efb-C. The energy minimization of the 10,000 conformations can also be done in parallel. Minimizing a single C3d conformation requires ∼45 min on a single processor using a Cray XT3 machine, and minimizing a single Efb-C conformation requires less than 5 min. Therefore, utilizing 600 processors, the entire minimization of 10,000 C3d conformations can be completed in ∼12 h and 10,000 conformations of Efb-C can be minimized within 1.5 h. The filtering and selection of low energy conformations takes several minutes in total. Overall, FEM requires approximately the same amount of time as the MD simulations. The rate limiting step is the minimization of C3d. This can be accelerated by either using more processors or a faster minimization protocol, since at this stage we are not interested in the global minimum conformations but rather in the lowest energy conformations with respect to the generated ensemble.

Figure 7 Shows the conformational landscape of the wildtype complex as obtained by our method (light shade) and using a single MD simulation as in our previous work[17] (dark shade). As seen, the single 20 ns MD simulation extends a single trajectory that has a small overlap with the rest of the conformational space captured by our method, which provides a broader coverage of other conformational preferences of the complex.

CONCLUSIONS

We have presented an efficient computational framework for the analysis of the low energy landscape of proteins that combines MD simulations and a robotics-inspired geometric conformational sampling algorithm, FEM. MD, although being a powerful tool for exploring the energetics and dynamics of proteins, can be limited to exploring local dynamics due to its large computational requirements. FEM enables us to capture larger scale motions, but is limited in its ability to explore large proteins. The combined approach presented in this article allows us to overcome the limitations of each method and provides an extensive sampling of the energy surfaces of the proteins in question. To provide a more complete picture, we used a non-linear dimensionality reduction technique that captures the main variability of the data and facilitates both the visualization and analysis of the protein low energy landscape. The proposed methodology is suitable for the extensive characterization of proteins and complexes which undergo small and medium scale conformational changes.

The study presented in this article enabled us to identify several distinct low energy conformations of the C3d/Efb-C complex and its two mutants, R131A and N138A, and identify interactions detected by experimental methods such as ITC, SPR, and mass spectrometry in addition to other previously uncharacterized interactions. The method presented here is a first step towards obtaining a complete picture of the conformational motions of large proteins in near-equilibrium conditions. We are currently working on developing methods that will allow us to scale up to even larger complexes and larger conformational motions.

ACKNOWLEDGMENT

The authors thank Dr. Amarda Shehu for providing help using the FEM software. This research was supported in part by the National Science Foundation through TeraGrid resources[35] provided by PSC and NICS.

REFERENCES

1. Dodson G, Verma C. Protein flexibility: its role in structure and mechanism revealed by molecular simulations. Cell Mol Life Sci 2006;63:207–219.
2. Jones S, Thornton J. Principles of protein-protein interactions. Proc Natl Acad Sci USA 1996;93:13–20.
3. Teague S. Implications of protein flexibility for drug discovery. Nat Rev Drug Discov 2003;2:527–541.
4. Case DA, Cheatham ITE, Darden T, Gohlke H, Luo R, Merz JKM, Onufriev A, Simmerling C, Wang B, Woods R. The AMBER biomolecular simulation programs. J Comput Chem 2005;26:1668–1688.
5. Kirkpatrick S, Gelatt CDJ, Vecchi MP. Optimization by simulated annealing. Science 1983;220:671–680.
6. Sherwood P, Brooks RB, Sansom M. Multiscale methods for macromolecular simulations. Curr Opin Struct Biol 2008;18:630–640.
7. Ho B, Agard D. Probing the flexibility of large conformational changes in protein structures through local perturbations. PLoS Comput Biol 2009;5:1–13.
8. Thorpe I, Zhou J, Voth G. Peptide folding using multi-scale coarse-grained models. J Phys Chem B 2008;112:13079–13090.

9. Lyman E, Pfaendtner J, Voth G. Systematic multiscale parameterization of heterogeneous elastic network models of proteins. Biophys J 2008;95:4183–4192.
10. Shehu A, Kavraki LE, Clementi C. On the characterization of protein native state ensembles. Biophys J 2006;92:1503–1511.
11. Shehu A, Kavraki LE, Clementi C. Modeling protein conformational ensembles: from missing loops to equilibrium fluctuations. Proteins: Struct Funct Bioinf 2006;65:164–179.
12. Shehu A, Clementi C, Kavraki LE. Multiscale characterization of protein conformational ensembles. Proteins: Struct Funct Bioinf 2009;76:837–851.
13. Kale L, Skeel R, Bhandarkar M, Brunner R, Gursoy A, Krawetz N, Phillips J, Shinozaki A, Varadarajan K, Schulten K. NAMD2: greater scalability for parallel molecular dynamics. J Comput Phys 1999; 151:283–312.
14. Phillips J, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kale LKS. Scalable molecular dynamics with NAMD. J Comput Chem 2005;26:1781–1802.
15. Das P, Moll M, Stamati H, Kavraki LE, Clementi C. Low dimensional, free energy landscapes of protein folding reactions by nonlinear dimensionality reduction. Proc Natl Acad Sci USA 2006;103: 9885–9890.
16. Lee J, Verleysen M. Non-linear dimensionality reduction. New-York: Springer; 2007.
17. Hammel M, Sfyroera G, Ricklin D, Magotti P, Lambris JD, Geisbrecht BV. A structural basis for complement inhibition by Staphylococcus aureus. Nat Immunol 2007;8:430–437.
18. Haspel N, Ricklin D, Geisbrecht BV, Kavraki LE, Lambris JD. Electrostatic contributions drive the interaction between staphylococcus aureus protein Efb-C and its complement target C3d. Prot Sci 2008;17:1894–1906.
19. Canutescu AA, Dunbrack RLJ. Cyclic coordinate descent: a robotics algorithm for protein loop closure. Prot Sci 2003;12:963–972.
20. Case DA, Darden TA, Cheatham ITE, Simmerling CL, Wang J, Duke RE, Luo R, Merz KM, Pearlman DA, Crowley M, Walker RC, Zhang W, Wang B, Hayik S, Roitberg A, Seabra G, Wong KF, Paesani F, Wu X, Brozell S, Tsui V, Gohlke H, Yang L, Tan C, Mongan J, Hornak V, Cui G, Beroza P, Mathews DH, Schafmeister C, Ross WS, Kollman PA. AMBER 9. San Francisco: University of California; 2006.
21. Pearlman DA, Case DA, Caldwell JW, Ross WS, Cheatham ITE, DeBolt S, Ferguson D, Seibel G, Kollman P. AMBER, a package of computer programs for applying molecular meechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. Comput Phys Commun 1995;91:1–41.
22. Duan Y, Wu C, Chowdhury S, Lee MC, Xiong G, Zhang W, Yang R, Cieplak P, Luo R, Lee T. A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. J Comput Chem 2003;24:1999–2012.
23. Srinivasan J, Trevathan M, Beroza P, Case D. Application of a pairwise generalized Born model to proteins and nucleic acids: inclusion of salt effects. Theor Chem Acc 1999;101:426–434.
24. Arfken G. The method of steepest descents. Mathematical methods for physicists, 3rd ed. Orlando, FL: Academic Press; 1985. pp 428–436.
25. Hestenes MR, Stiefel H. Methods of conjugate gradients for solving linear system. J Res Nat Bur Stand 1952;49:409–436.
26. Miyamoto S, Kollman PA. SETTLE: an analytical version of the SHAKE and RATTLE algorithm for rigid water models. J Comput Chem 1992;13:952–962.
27. Kollman PA, Massova I, Reyes C, Kuhn B, Huo S, Chuong L, Lee M, Lee T, Duan Y, Wang W, Donini O, Cieplak P, Srinivasan J, Case DA, Cheatham TE, III. Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. Acc Chem Res 2000;33:889–897.
28. Tsui V, Case DA. Theory and applications of the generalized Born solvation model in macromolecular simulations. Biopolymers (Nucl Acid Sci) 2001;56:275–291.
29. Onufriev A, Bashford D, Case DA. Exploring protein native states and large-scale conformational changes with a modified generalized born model. Proteins: Struct Funct Bioinf 2004;55:383–394.
30. Weiser J, Shenkin PS, Still WC. Approximate atomic surfaces from linear combinations of pairwise overlaps(LCPO). J Comput Chem 1999;20:217–230.
31. Tenenbaum J, de Silva V, Langford J. A global geometric framework for nonlinear dimensionality reduction. Science 2000;290:2319–2323.
32. Kumar S, Rosenberg J, Bouzida D, Swendsen R, Kollman P. The weighted histogram analysis method for free-energy calculations on biomolecules. J Comput Chem 2004;13:1011–1021.
33. Ricklin D, Ricklin-Lichtsteiner S, Markiewski M, Geisbrecht B, Lambris J. Cutting edge: staphylococcus aureus extracellular fibrinogen -binding protein inhibits the interaction of c3d with complement receptor 2. J Immunol 2008;181:7463–7467.
34. Alam SR, Agarwal PK, Hampton SS, Ong H, Vetter JS. Impact of multicore on large scale molecular dynamics simulations. IEEE International Workshop on High Performance Computing in Biology (HiCOMB), in conjunction with IPDPS. 2008.
35. Catlett C. Teragrid: analysis of organization, system architecture, and middleware enabling new types of applications. Advances in parallel computing series. Amsterdam: IOS Press. 2007.