

AD _____

Award Number: W81XWH-11-1-0311

TITLE: Multi-Domain Assembly of Nuclear Estrogen Receptors: Structural Insights into ER-Positive Breast Cancer Therapeutics

PRINCIPAL INVESTIGATOR: Geoffrey Greene

CONTRACTING ORGANIZATION: The University of Chicago, Chicago, IL 60637

REPORT DATE: April 2013

TYPE OF REPORT: Final

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release;
Distribution Unlimited

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision unless so designated by other documentation.

REPORT DOCUMENTATION PAGE				Form Approved OMB No. 0704-0188	
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.					
1. REPORT DATE (DD-MM-YYYY) April 2013		2. REPORT TYPE Final		3. DATES COVERED (From - To) 01 April 2011- 31 March 2013	
4. TITLE AND SUBTITLE Multi-Domain Assembly of Nuclear Estrogen Receptors: Structural Insights into ER-Positive Breast Cancer Therapeutics				5a. CONTRACT NUMBER	
				5b. GRANT NUMBER W81XWH-11-1-0311	
				5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S) Geoffrey Greene				5d. PROJECT NUMBER	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) The University of Chicago, 5801 S. Ellis Ave Chicago, IL 60637				8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Army Medical Research Fort Detrick, Maryland 21702-5012				10. SPONSOR/MONITOR'S ACRONYM(S)	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release; distribution unlimited					
13. SUPPLEMENTARY NOTES					
14. ABSTRACT The major goal of this project is to determine and compare the overall dynamic architectures of both ERs bound to different known DNA response elements (EREs). This study combines small-angle X-ray solution scattering (SAXS) data and computational modeling to address the multimeric assembly of ER DBD-hinge-LBD complexes. To date, we have successfully modeled ERβ CDE bound to its cognate DNA binding site and generated homogenous ER DBD-hinge-LDB protein samples (ERβ CDEF) needed for SAXS studies. We have also collected SAXS data on ERβ CDEF, but not ERβ CDE. We will continue to generate and compare the ERβ□CDE and CDEF fragments for small angle x-ray scattering (SAXS) analyses of ER multidomain fragments ± EREs and various peptides to obtain solution structure information. Having worked out the conditions for the expression and purification of ERβ CDEF, we do not expect difficulties in the expression and purification of ERβ CDE domain. By integrating SAXS analyses and computational biology, we expect to better understand ER multi-domain assembly mechanisms. It is anticipated that this information will help explain how domain interfaces in ER/ERE complexes modulate ligand-dependent transcriptional activity in response to various SERMs, providing novel structural insights that will facilitate improved targeting of ER-positive breast cancers, especially those that are refractory to current adjuvant therapies.					
15. SUBJECT TERMS none provided					
16. SECURITY CLASSIFICATION OF: U			17. LIMITATION OF ABSTRACT UU	18. NUMBER OF PAGES 7	19a. NAME OF RESPONSIBLE PERSON USAMRMC
a. REPORT U	b. ABSTRACT U	c. THIS PAGE U			19b. TELEPHONE NUMBER (include area code)

Table of Contents

Introduction	1
Body	1
Key Research Accomplishments	3
Reportable Outcomes.....	5
Conclusion	5
References.....	5
Appendices.....	6
Supporting Data.....	6

Introduction

This study combines complementary computational modeling and SAXS expertise (S.Y.) with expertise on SERM-mediated ER LBD structure and signaling (G.G.). It is anticipated that our studies on the DBD-hinge-LDB assembly in ER α and ER β will determine the extent to which inter-domain interfaces control ligand-dependent transcriptional activity in response to various known agonists, antagonists, or novel SERMs, providing structural insights that can be exploited for novel targeting of ER-positive breast cancers.

For clarity and completeness, I will first list the specific aims of this DoD grant and the summary of the studies and results within the first 12 months, followed by the accomplishment results within the second 12 months.

Body

A. Specific Aims

Aim 1. Determine structural models of assembly states of the DBD-hinge-LBD complexes in ER α and ER β using computational modeling.

1a. Develop an all-residue coarse-grained (ARCG) model for molecular dynamics simulations. Test the ARCG model on known ER LBD homodimers¹ first and then apply it to ER α and ER β DBD-hinge-LBD complexes.

1b. Design a hierarchical structural clustering algorithm and apply it to structural modes generated from ARCG simulations in Aim 1a. Determine a small number (~10) of assembly states using a two-step procedure that takes into account both differences in detailed residue-residue distances and overall macromolecular scattering.

1c. Compute theoretical SAXS profiles for the resulting assembly states obtained in Aim 1b by using a high-throughput, scattering computing program of “Fast-SAXS”. Compare scattering differences and examine theoretical DBD-LBD interaction mechanisms of assembly states.

Aim 2. Detect ER shape changes in response to small molecules using SAXS experiments.

2a. Express ER α and ER β DBD-hinge-LBD plasmid constructs in bacteria and purify them to a concentration of approximately 1 mg/ml. Determine the sample size distribution using dynamic light scattering. Synthesize and prepare various ligands and different ERE oligomers as agents to elicit ER conformational changes.

2b. Collect SAXS data for ERs at the APS and NSLS. SAXS experiments will be conducted with ER fragments in the presence of several ligands, including 4-hydroxytamoxifen (OHT), 17 β -estradiol (E2), raloxifene, and various known ERE oligomers. Examine differences in experimental SAXS patterns and map out the global shape changes upon the binding of ligands with diverse behaviors.

Aim 3. Determine DBD-LBD interactions and quantify molecular mechanisms of ER multi-domain assembly.

3a. Use BSS-SAXS as a structural technique²⁻³ to determine ER multi-domain assemblies. Combine structural models of assembly states obtained from computational modeling in Aim 1 and experimental SAXS data obtained in Aim 2 to quantify the population state of each multi-domain assembly state using a Monte Carlo procedure. Determine DBD and LBD interaction mechanisms under various conditions.

3b. Use hydrogen/deuterium (H/D) exchange analysis to define specific local molecular details of DBD-LBD interfaces and identify structural features of hydrophobic forces and hydrogen-bonding at DBD-LBD interfaces. Resolve the multimeric structural assembly mechanisms in ER α and ER β in response to various signals. Make specific theoretical predictions of sites at the interface(s) that stabilize specific assembly states for functional studies. Use site-directed mutagenesis to confirm predictions. Make predictions of sites at the interface(s) that can be cross-linked to stabilize specific assembly states for future crystal studies to capture such structural snapshots of energetically favorable assembly states.

B. Studies and Results in the first 12 months

The results described here include progress since this sub-award was made in April 2011. We have made progress with the simulations of ER β DBD-hinge-LBD (also known as CDE) in **Aim 1a**, structural clustering analyses of simulation trajectories in **Aim 1b**, and SAXS computing in **Aim 1c**. In addition, we have collected SAXS of ER β CDEF in complex with ERE oligomers and ligands, as proposed in **Aim 2**.

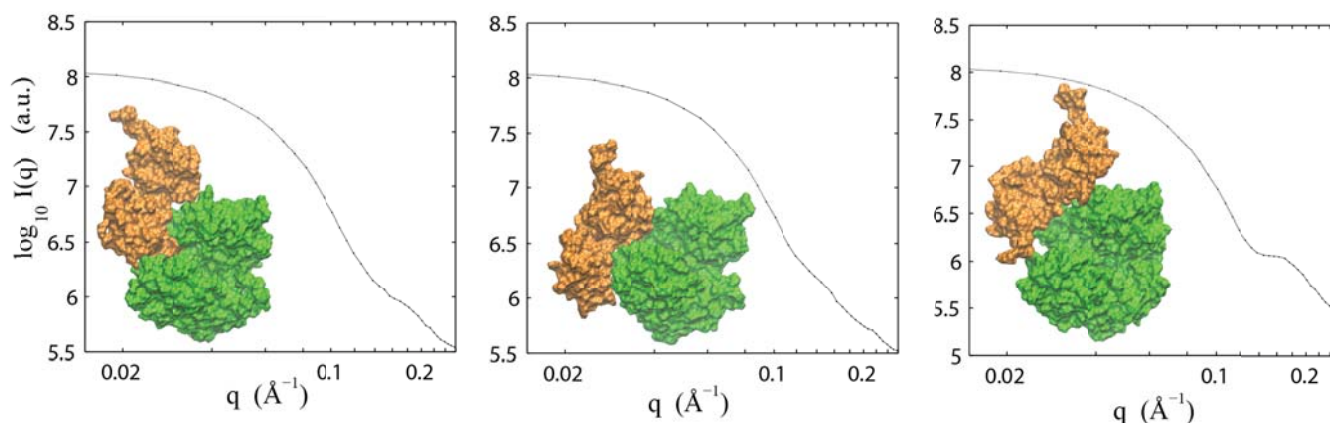


Figure 1. Three stable conformations have been identified from molecular dynamics simulations of ERbeta CDE. Their theoretical SAXS curves are also computed using in-house Fast-SAXS software (Ref. 4).

To demonstrate the results from molecular dynamics (MD) simulations (**Aim 1**), **Figure 1** shows three energetically stable conformations observed for the ER β DBD-Hinge-LBD (CDE) fragment. This is achieved by 100 separated MD simulation trajectories, and each has 10 microseconds, resulting in a total of 1 millisecond. All the resulting MD configurations are clustered into a small number of conformational clusters using a two-step procedure. For the SAXS computing, our in-house Fast-SAXS package was used for all the three conformations. The results in **Figure 1** show that each SAXS profile $I(q)$ is different from the others, suggesting experimental SAXS data can be a very powerful tool for

determining their population fractions. A previously published SAXS-based shape reconstruction method will be used once the SAXS data of this CDE complex (Ref. 2).

For **Aim 2**, the expression and purification the CDE and CDEF domains of ER β have now been largely optimized and we are able to consistently purify up to 10 mg of protein at a concentration of 2 to 3 mg/ml (**Figure 2**). Higher concentrations have been more difficult to obtain due to protein precipitation. However, these concentrations are sufficient for SAXS analyses. We have also found conditions under which the purification can be performed without ligand. Therefore the same protein preparation can be used for SAXS measurement with or without the addition of different ligand such as estradiol and 4-hydroxytamoxifen (OHT) as well as ERE oligomers of different lengths and sequence, thus allowing direct comparison between the same SAXS samples. Initial experiments were carried out with the CDEF fragment of ER β , which contains an additional 25 amino acids that are thought to be at least partially unstructured. It will be important to compare these two ER β fragments by SAXS because no existing structural data exists for the F domain of either ER subtype.

For **AIM 2b**, purified ER β CDEF fragment has been analyzed by SAXS in the presence and absence of a known oligomeric ER DNA binding site (ERE). **Figure 3** shows the SAXS data for this polypeptide in the presence of ERE, demonstrating that we have been able to obtain high-quality SAXS data for this functional complex. Shown is the plot of $I(q)$ vs q (in a logarithmic scale), where q is the amplitude of X-ray momentum transfer during the scattering. A triplicate of 20 μ L of sample was continuously flowed through a 1 mm diameter capillary and each was exposed for 30 sec. The data were recorded under the condition of the X-ray energy of 14 keV and the temperature of 10 degree. This result lays the foundation for further studies of the receptor in the presence of different ligands and different specific DNA sequences that are associated with naturally occurring ER target genes.

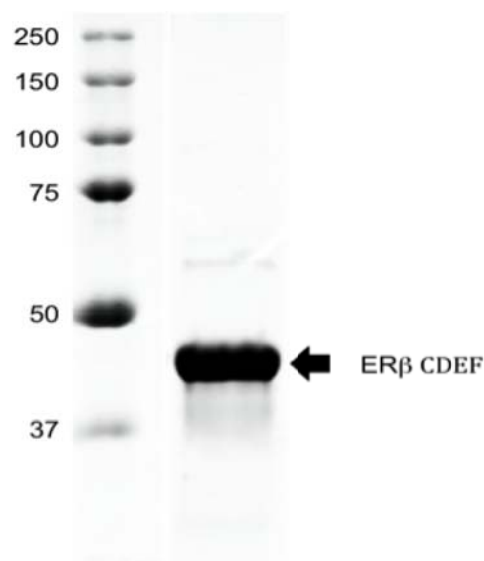


Figure 2. ER β CDEF protein is visualized by Coomassie blue staining on an SDS-PAGE gel after purification

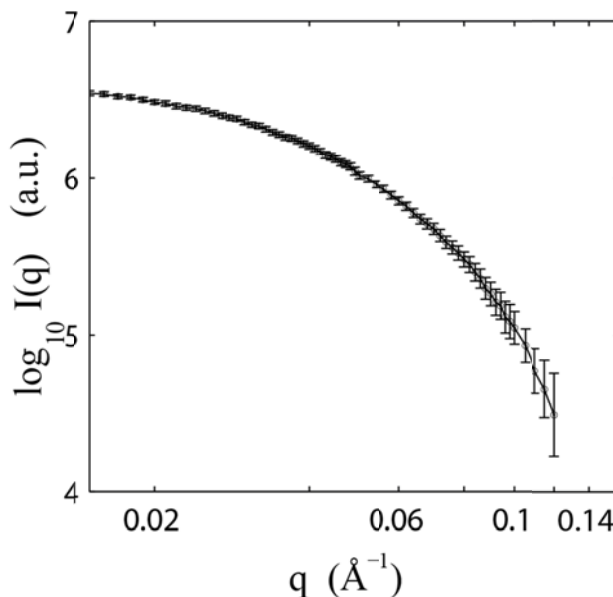


Figure 3. Experimental SAXS of ERbeta CDEF in the presence of DNA. Data were collected at the NSLS in Oct. 2011.

Key Research Accomplishments

- Successful modeling of ER β CDE bound to DNA
- Successful purification of ER β CDEF
- Successful collection of SAXS data of apo ER β CDEF in the presence of DNA at the NSLS

C. Studies and Results in the second 12 months

The results described here include progress since April 2012. First, we have developed a theoretical simulation pipeline to study the general protein-protein association as proposed in **Aim 1a**. This work has been published (5) and highlighted on the cover of *Biophysical Journal* (see Figure 4). Specifically, a coarse-grained (CG) model was implemented and its applications were demonstrated via molecular dynamics simulations for several protein complexes. Second, an enhanced search method was utilized to efficiently sample a broad range of protein conformations. Third, multiple conformations were identified and clustered from simulation data and further projected on a three-dimensional globe specifying protein orientations and interacting energies. Results from several complexes indicate that the crystal-like conformation is favorable on the energy landscape even if the landscape is relatively rugged with meta-stable conformations. A closer examination on molecular forces shows that the formation of associated protein complexes can be primarily electrostatics-driven, hydrophobics-driven, or a combination of both in stabilizing specific binding interfaces. Taken together, these results suggest that the CG simulations and analyses provide a new tool-set to study protein-protein association occurring in functional biomolecular complexes. The energy globe analysis (Figure 4) is capable of identifying distinct putative conformations that can serve as a basis set to analyze SAXS data. The application of this PPR-CGMD simulation pipeline has been tested on five different protein-protein complexes before it is applied to ER α simulations.



Figure 4. Illustrative energy landscape on a globe based on coarse-grained simulations. The globe provides an overview of the landscape by specifying the relative orientations and interaction energies of the two proteins: one is at the center of the globe (in yellow) and the other is on the surface. The globe is colored and plotted with contour lines based on the interaction energies of a large set of simulated orientations. Three distinct orientations are illustrated for the second protein in red, green, and orange, respectively, each located in an energy basin on the surface. This energy globe is intuitive in outlining energetically favorable complex-forming conformations from a landscape perspective. Taken from the August, 2012 issue of the *Biophysical Journal* cover.

Second, we have performed extensive CG simulations on ER α DBD-hinge-LDB (also known as CDE; see Figure 5) in **Aim 1a**. Based on the simulations data, a structure clustering analysis is performed on all the "free" simulation trajectories in **Aim 1b**, where molecular dynamics (MD) simulations are implemented and enhanced sampling is achieved via the *push-pull-release* (PPR) strategy in the search for different LBD-DBD orientations. Specifically, this search method is designed to facilitate domain-domain dissociation and re-association, which may be difficult for brute-force simulations once associated. This is achieved via a harmonic potential is imposed on the center-of-mass distance between LBDs and DBDs to pull them apart when they are close and to push them together when far away. Then, this bias is removed so LBDs and DBDs are free to interact when they are close enough. This strategy helps the simulations to avoid the trapping due to local stable complex-forming conformations, thus enabling the sampling of different ER conformations.



Figure 5. Schematic diagram of ER α functional domains of a ligand-binding domain (LBD) and a DNA-binding domain (DBD). LBD and DBD dimers are displayed in green and blue, respectively, and their connecting hinge loops are marked by gray dots. The DBD-specific DNA duplex is shown as ribbon in white. The activation helix H12 is highlighted in red. Taken from the Huang *et al*, 2013.

Key results from the simulations are the identification of multiple energetically stable ER α conformations on the landscape (see Figure 6). An important finding is that estradiol-bound LBDs utilize the well-described activation helix H12 to pack and stabilize LBD-DBD interactions. Our results suggest that the estradiol-bound LBDs can serve as a scaffold to

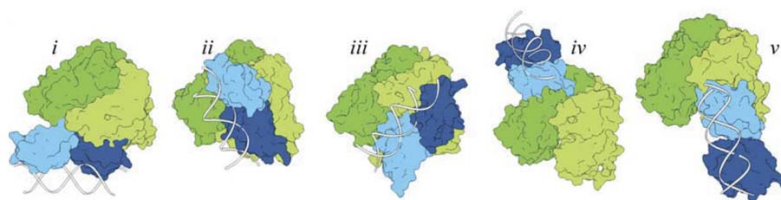


Figure 6. Possible ER α structure models identified from the simulations. Taken from the Huang *et al*, 2013.

position and stabilize the DBD-DNA complex, consistent with

experimental observations of enhanced DNA binding with the LBD. Final assessment using atomic-level simulations shows that these CG-predicted models are significantly stable within a 15-ns simulation window and specific pairs of lysine residues in close proximity at the domain interfaces. Together, these simulation results provide a molecular view of the role of ER α domain interactions in response to hormone binding. Reports on detailed findings have been published recently (see Ref. 6).

As the proposed models have very distinct global shapes, we propose later to use a synchrotron-based technique, small-angle solution X-ray scattering (SAXS), to distinguish the correct model. For a comparison with experimental data, we will compute SAXS profiles for these structure models, using the new SAXS computing algorithm that the co-PI has recently developed, termed *Fast-SAXS-pro* (see Figure 7 and Ref. 7). This method is a natural extension based on our previously developed methods for protein and RNA scattering (3,4). These calculated SAXS profiles will be used for the matching with measured experimental data to characterize the best-fit model. This new SAXS computing algorithm has been applied to predicted ER α conformations as proposed in **Aim 1c**, which will enable the data interoperation of measured experimental SAXS data.

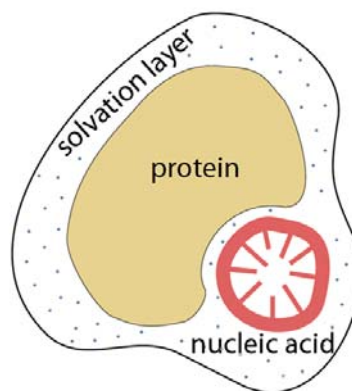


Figure 7. Schematic for a new SAXS computing algorithm for protein-DNA complexes. Modified from Ravikumar et al., (2013).

Finally, as proposed in **Aim 2**, we have been improving the sample quality of the ER β CDEF in complex with ERE oligomers and ligands. Since April 2012, the protein purity has been significantly improved, in which no sign of protein aggregation occurs. This result suggests that the ER β CDEF can form a stable complex with estradiol and the ERE oligomers. One issue still to be resolved is that we sometimes observe the presence of two protein bands on SDS-PAGE gels (data not shown). This observation suggests that ER β CDEF may be prone to protease digestion during or preceding purification, or during handling after purification. Because this is a sporadic problem, future efforts will include the use of a chromatography-coupled SAXS data setup, which will ensure the sample homogeneity critical for SAXS data collection.

Reportable Outcomes

Two papers have resulted from this funded work. One (5) has been published and the other (6) has been accepted for publication. Both are included in the appendix.

Conclusion

An important goal of this study is to generate homogenous ER DBD-hinge-LDB (also known as CDE) protein samples needed for SAXS studies. Up until now, we have collected SAXS data of CDEF samples, but not that of CDE. We will continue to generate and compare the ER β CDE and CDEF fragments for small angle x-ray scattering (SAXS) analyses of ER multidomain fragments $\pm$ EREs and various peptides to obtain solution structure information. Having worked out the conditions for the expression and purification of ER β CDEF, we do not expect difficulties in the expression and purification of ER β CDE domain. Future work will also be focused on optimizing the purification of ER α CDE. ER α protein purification has been more challenging because of the longer hinge region that separate the DBD and LBD that is more prone to proteases digestion.

References

- 1- Shiao AK et al, Cell. 1998 Dec 23; 95(7): 927-37.
- 2- Yang S. et al, Proc Natl Acad Sci U S A. 2010 Sep 7; 107(36): 15757-62.

- 3- Yang S. et al, J Phys Chem B. 2010 Aug 12; 114(31): 10039-48.
- 4- Yang S. et al, Biophys J. 2009 June 3; 96(11): 4449–4463.
- 5- Ravikumar K.M, Huang W., Yang S. Biophys J. 2012;103(4):837-45.
- 6- Huang W., Greene, G.L., Ravikumar K.M, Yang S. Proteins. 2013; *In press*.
- 7- Ravikumar K.M, Huang W., Yang S. J Chem Phys. 2013;138(2):024112.

Appendices

Pdf files are included for the following publications:

1. Ravikumar K.M, Huang W., Yang S. Biophys J. 2012;103(4):837-45.
2. Huang W., Greene, G.L., Ravikumar K.M, Yang S. Proteins. 2013; *In press*.

Supporting Data

Included in Body

List of Personnel receiving pay from the research effort

Geoffrey L. Greene
Muriel Laine

Cross-talk between the ligand- and DNA-binding domains of estrogen receptor

Short title: ER α domain cross-talk

Key words: estrogen receptor, coarse-grained model, protein-protein interaction, energy landscape, molecular dynamics

Wei Huang,¹ Geoffrey L. Greene,² Krishnakumar Mayuram Ravikumar,¹ Sichun Yang^{1*}

1. Center for Proteomics and Department of Pharmacology, Case Western Reserve University, Cleveland, OH 44106-4988
2. Ben May Department for Cancer Research, University of Chicago, Chicago, IL 60637

*Correspondence author contact information:

Sichun Yang, Ph.D.

Case Western Reserve University

2109 Adelbert Rd., BRB 929

Cleveland, OH 44106-4988

Phone: 216.368.5793

Fax: 216.368.6846

E-mail: sichun.yang@case.edu

Abstract

Estrogen receptor alpha (ER α) is a hormone-responsive transcription factor that contains several discrete functional domains including a ligand-binding domain (LBD) and a DNA-binding domain (DBD). Despite a wealth of knowledge about the behaviors of individual domains, the molecular mechanisms of cross-talk between LBD and DBD during signal transduction from hormone to DNA-binding of ER α remain elusive. Here, we apply a multi-scale approach combining coarse-grained (CG) and atomistically-detailed simulations to characterize this cross-talk mechanism via an investigation of the ER α conformational landscape. First, a CG model of ER α is built based on crystal structures of individual LBDs and DBDs, with more emphasis on their inter-domain interactions. Second, molecular dynamics (MD) simulations are implemented and enhanced sampling is achieved via the *push-pull-release* (PPR) strategy in the search for different LBD-DBD orientations. Third, multiple energetically stable ER α conformations are identified on the landscape. A key finding is that estradiol-bound LBDs utilize the well-described activation helix H12 to pack and stabilize LBD-DBD interactions. Our results suggest that the estradiol-bound LBDs can serve as a scaffold to position and stabilize the DBD-DNA complex, consistent with experimental observations of enhanced DNA binding with the LBD. Final assessment using atomic-level simulations shows that these CG-predicted models are significantly stable within a 15-ns simulation window and specific pairs of lysine residues in close proximity at the domain interfaces could serve as candidate sites for chemical cross-linking studies. Together, these simulation results provide a molecular view of the role of ER α domain interactions in response to hormone binding.

Introduction

Estrogen receptor alpha (ER α), a member of the nuclear hormone receptor (NHR) family, plays a key role in eukaryotic transcriptional regulation.^{1, 2, 3, 4} Once activated by hormones such as estradiol, ER α can bind specific DNA sequences as a homodimer and modulate the expression of genes involved in the development, homeostasis and metabolism.^{5, 6, 7, 8, 9} Because ER α is a key target for therapeutic intervention^{10, 11, 12}, it is vital to obtain a better molecular understanding of ER α action that results from hormone signaling.

ER α is a multi-domain protein fold containing a DNA-binding domain (DBD) and a ligand binding domain (LBD)^{2, 13}. Their individual crystal structures have been determined (as illustrated in Fig. 1). These two domains are connected by a 50-residue flexible hinge region that allows these domains to move freely in three dimensions. It has been observed that the DBD alone can bind a double-strand DNA^{14, 15, 16}, and that this DNA binding of the receptor is tighter in the presence of LBD when bound to estradiol⁶. At a structural level, this estradiol binding is also coupled with a major rearrangement of the critical activation helix H12 at the C-terminus, which is positioned on the LBD surface and covers the estradiol binding pocket^{17, 18, 19, 20, 21}.

The mechanisms by which ER α DBDs and LBDs interact with each other when ER α is bound to DNA are still unclear. For several other members of the NHR family, it has been demonstrated that the LBD-DBD interface plays an allosteric role to mediate the signal transduction between ligand binding and DNA association^{22, 23, 24}. For example, a single point mutation at the interface of the PPAR/RXR complex affects its DNA binding properties to a PPAR response element²². More recently, mutagenesis studies in the androgen receptor (AR) show that several mutations in the AR LBD can decrease DNA binding affinity, and conversely, mutations in the DBD also affect ligand binding²⁴. These observations strongly suggest the internal communication between LBDs and DBDs is required to facilitate *in vitro* DNA binding. For ER α , how these domains interact and communicate remains largely unknown, in part due to

the lack of structural information of a full LBD-DBD/DNA complex and the lack of knowledge about the LBD-DBD interactions at a molecular level.

In this study, we used a computational approach of coarse-grained and subsequent all-atom molecular dynamics (MD) simulations to explore the physical interactions between the ER α DBD and LBD domains from an energy landscape perspective. Specifically, a coarse-grained (CG) model and an efficient sampling method were implemented to simulate the ER α conformational landscape. The projection of simulation data onto a three-dimensional globe allows one to picture the energy landscape of ER α multi-domain interactions. Multiple ER α conformational states were identified from the landscape, highlighting different modes of LBD-DBD interactions. Final structural refinement via all-atom simulations starting from CG-predicted models provides detailed structural features at the domain interfaces that could be experimentally tested. Overall, these results provide new speculation into ER α domain cross-talk mechanisms responsible for the allosteric control of receptor binding to DNA.

Methods

Details of the CG model

A coarse-grained (CG) model was implemented where each amino acid of ER α is represented by a single bead positioned at its C α atom and each nucleotide of DNA is represented by its O5' atom. The CG energy function for ER α (E₁₈₁-P₅₅₂) is defined for each component as follows. The energy functions for the estradiol-bound LBD dimer (S₃₀₉-P₅₅₂) and the DBD dimer (E₁₈₁-K₂₅₂, in complex with a 17-bp double-stranded DNA) were built based on their crystal structures (PDB entries 1QKU and 1HCQ)^{15, 20}, using the widely used G $\ddot{o}$ -type potential. The hinge region connecting DBD and LBD (G₂₅₃-L₃₀₈) was built using loop modeling method in MODELLER²⁵. Bonded interactions, including bond, angle and dihedral terms, were modeled using typical energy functions as done previously (see details in Refs.^{26, 27}). Non-bonded

attractive interactions between native contact pairs between LBDs and DBDs were modeled using Lennard-Jones (LJ) interactions, $E_{\text{LJ}} = \sum_{i,j} \epsilon_o [5(\sigma_{ij}^o/r_{ij})^{12} - 6(\sigma_{ij}^o/r_{ij})^{10}]$, where r_{ij} is the residue-residue distance, and σ_{ij}^o is the corresponding distance in the reference structure. The contact pairs within each domain were defined using the CSU software²⁸. Native-like DBD-DNA interactions were modeled using $\epsilon_o = 5$ kcal/mol whenever residue-nucleotide pairs were within a distance of 15 Å.

The interactions between LBDs and DBDs are specified by the energy function $E_{\text{LBD-DBD}}$, which includes electrostatic (E_{elec}) and hydrophobic (E_{H}) components,

$$E_{\text{LBD-DBD}} = E_{\text{elec}} + E_{\text{H}}. \quad (1)$$

Here, $E_{\text{elec}} = \sum_{i,j} q_i q_j / (4\pi\epsilon_o D_{\text{eff}} r_{ij})$ was used where q_i is the charge of residue i and ϵ_o is the vacuum electric permittivity. An effective dielectric coefficient $D_{\text{eff}} = D_s \exp(r_{ij}/\xi)$ is applied to reflect the shielding effect between two residues separated by a distance of r_{ij} . $D_s = 10$ was used to describe the local dielectric environment when two domains are forming a close interface and $\xi = 8.2$ Å to mimic the screening effect at a 150-mM salt concentration. At pH 7, residue charges $q_i = +e$ for Lys and Arg, $-e$ for Asp and Glu, and $+0.5e$ for His (e is the elementary charge) were used^{26, 29}. Hydrophobic interactions (E_{H}) are either attractive (LJ-type) ($\epsilon_{ij} < 0$) or purely repulsive ($\epsilon_{ij} \geq 0$) where $\epsilon_{ij} = \alpha(e_{ij}^{\text{MJ}} + \beta)$ between residues i, j and e_{ij}^{MJ} is the Miyazawa-Jernigan (MJ) statistical energy.³⁰ The values of $\alpha = 0.4$ and $\beta = 1.3$ were used to balance attractive and repulsive interactions.²⁶ We used $E_{\text{H}}(i, j) = |\epsilon_{ij}| [5(\sigma_{ij}/r_{ij})^{12} - 6(\sigma_{ij}/r_{ij})^{10}]$ if $\epsilon_{ij} < 0$, and $E_{\text{H}}(i, j) = \epsilon_{ij} [5(\sigma_{ij}/r_{ij})^{12} (1 - \exp(-(r_{ij} - \sigma_{ij})/d)^2)]$ if $\epsilon_{ij} \geq 0$, where $d = 3.8$ Å. The distance σ_{ij} is defined by

$$\sigma_{ij} = \gamma(r_i + r_j) \quad (2)$$

where the value of $\gamma = 0.625$ was chosen²⁶ and r_i is the van der Waals radius of residue i ²⁹. The non-native interactions of DNA with LBDs and the hinge region were mostly repulsive with ϵ_{ij}

= 1 kcal/mol.

Details of the PPR sampling strategy

The CG model was implemented via Langevin molecular dynamics (MD) simulations in a modified version of the CHARMM program.³¹ Simulations were performed at 300 K with a friction coefficient of 50 ps⁻¹.³² A simulation time step of 10 fs was used and simulated coordinates were saved every 100 ps. A sampling strategy of *push-pull-release* (PPR) was implemented via a biasing potential E_{PPR} ²⁶,

$$E_{\text{PPR}} = \begin{cases} 0, & \text{when } R_t \leq R_c \text{ and } r_{\min} < r_o, \\ k(R - R_t)^2, & \text{otherwise.} \end{cases} \quad (3)$$

where R is the instantaneous center-of-mass distance between two selected domains, R_t is its corresponding target distance, and $r_{\min}$ is the closest residue-residue distance between two domains. The two domains move between $R_{\min} \leq R_t \leq R_{\max}$, where the values of $R_{\min} = 0$ Å and $R_{\max} = 100$ Å were used. This E_{PPR} potential is applied to each PPR cycle with a simulation time of L and repeated for N times, where each cycle includes three steps: (1) *push* the two domains away from each other when they are close, (2) *pull* the domains closer until they are separated by a threshold distance R_c , and (3) *release* them to interact freely by removing the biasing potential. For ER α simulations, the parameters $N = 5$, $L = 25$ ns, $k = 100$ kcal/(mol·Å²), $r_o = 7.6$ Å, and $R_c = 45$ Å were used. Four sets of PPR-assisted simulations were performed, each accounting for a different biasing pair between two LBDs and two DBDs. Within each set, a total of 145 simulation runs were carried out simultaneously, each with a different initial configuration of LBD-DBD relative orientation.²⁶ A simulation time of 145 μ s was achieved in total.

Details of orientation-based clustering analysis

A structural clustering analysis is applied mainly based on relative orientations between domains. First, structural alignment is performed on LBDs only. Second, resultant configurations of DBDs are used to calculate pair-wise RMSD values. Finally, these RMSD values are used as an input for a hierarchical clustering analysis as implemented in `MATLAB`. This clustering analysis was applied to selected simulation snapshots if they have the lowest $E_{\text{LBD-DBD}}$ snapshot from the last 5-ns segment in each PPR cycle, where $E_{\text{LBD-DBD}} < 0$ kcal/mol, and the RMSD values of individual DBD and LBD monomers are within 1.5 Å and 3 Å to their corresponding crystal structures, respectively. In addition, mirror conformations by exchanging the coordinates of two protein chains were generated due to ER α homodimeric symmetry. Thus, a total number of 9780 configurations were used for the clustering analysis.

Details of all-atom simulations

All-atom MD simulations were performed starting with CG-predicted conformations. Initial ER α models were built based on crystal structures of individual domains. First, the CG models were used as the templates to build atomically-detailed structures, where crystal structures of the LBDs (PDB entry 1QKU²⁰) and DBDs (PDB entry 1HCQ¹⁵) were aligned on the top of C α atoms of CG models. Second, the restraints of C α atoms within domains towards CG models were achieved via targeted MD simulations.³³ An implicit solvent GB model was used to reduce the computational cost³⁴ and a total of 40-ns simulation time was performed to reach a C α -RMSD value of 0.5 Å (relative to the CG models of ER α domains). Finally, standard all-atom simulations without any bias were performed for model refinement using the program NAMD³⁵, where the force fields of AMBER99SB³⁶ and AMBER99bsc0^{37, 38} were used. Parameters for the estradiol were generated using GAFF³⁹ and ANTECHAMBER^{40, 41, 42}. All parameters were prepared using the AMBER LEaP module⁴³. The system was placed in a rectangular water

box with more than 12 Å padding and 16 charge-neutralizing sodium ions around the solute. Langevin simulations with a total of 15 ns were performed for each state at a pressure of 1 atm and a temperature of 300 K.

Results and Discussion

A cartoon representation of the ER α multi-domain architecture is shown in Figure 1, illustrating ER α domains for both ligand and DNA binding (LBD and DBD). Here, a recently developed coarse-grained MD simulation pipeline was applied to simulate ER α dynamics.²⁶ Specifically, the search for different LBD-DBD assemblies was enhanced via a sampling strategy of *push-pull-release* (PPR). After grouping similar configurations into a conformational state via a hierarchical clustering scheme, ER α simulation data were projected onto a three-dimensional globe specifying the ER α inter-domain orientation, and the interacting energy of each state was used to visualize the LBD-DBD energy landscape and identify favorable conformations accessible on the landscape. Finally, the structural features of stable ER α conformational states were illustrated to highlight their inter-domain interactions.

CG simulations with the PPR sampling strategy

The CG model implemented in ER α simulations has two main components: intra- and inter-domain interactions. For each component of the LBD-DBD complex, a Gō-like model was used to model the dynamics within both the LBD dimer and the DBD dimer by taking advantage of the availability of their crystal structures.^{15, 20} This treatment allowed each dimer to structurally fluctuate around its native conformation, as well as maintain substantive flexibility as required for conformational deformation in an encounter complex. For the inter-domain LBD-DBD interactions, a residue-specific energy function of $E_{\text{LBD-DBD}}$ (Eq. 1) was implemented by accounting for explicit electrostatic and hydrophobic interactions (see details in Methods). Thus,

a CG model of ER α was built based on crystal structures as well as physical principles for the purpose of MD simulations.

The search for ER α conformations was enhanced via the sampling strategy of *push-pull-release* (PPR). This PPR approach was adopted to facilitate the dissociation and re-association between LBDs and DBDs (see Ref. ²⁶ and Methods). This sampling enhancement was achieved by applying a biasing potential to the center-of-mass distance of the two domains (Eq. 3), which can push the domains together to associate, pull them away to disassociate, and release them to relax by removing the biasing potential when the domains are encountered. It should be noted that during the pulling and pushing, two domains were allowed to rotate freely, thus helping achieve broader sampling in their relative orientations.

To illustrate the sampling coverage achieved by PPR, a three-dimensional sphere was used to specify LBD-DBD orientations. The LBDs were first used for the alignment of each simulation snapshot and placed in the center of this globe. The center-of-mass of DBDs was then projected onto the surface of the globe [Fig. 2(A)]. Compared to the simulations without using PPR, Figure 2(B) shows that within a same simulation window of 125 ns, the sampling efficiency was considerably enhanced by PPR with a much larger surface area spanned on the globe. The sampling of these PPR-assisted simulations was further accelerated by launching a set of 145 parallel CGMD runs each with a different initial configuration, as illustrated in Figure S1(A). In addition, four sets of such simulations were performed, each with a different biasing potential between a pair of monomeric LBD and DBD (see Methods). As a result, a total of 145 μ s simulation time was achieved and an accumulated 61.625- μ s from unbiased PPR segments was used for the rest of the analyses.

Figure 2(C) illustrates the sampled area of the globe covered by simulation snapshots, where each snapshot is represented by a blue dot. Clearly, most of the surface area was sampled by PPR-assisted simulation trajectories, suggesting that sufficient sampling of various LBD-DBD orientations was obtained. We note that both front and back views of the globe were based

on the placement of LBDs in the center [Fig. S1(B)]; an alternative placement of DBDs in the center is also shown in Figure S1(C), where more than half of the sphere is well covered (and the other half is not covered due to the blockage of its DNA binding). Together, these results suggest that this PPR strategy substantially assisted the CGMD simulations with a rather comprehensive sampling coverage in the $ER\alpha$ domain orientational space.

Energy landscape on a globe: clustering and projection

The visualization of $ER\alpha$ conformational landscape was analyzed by projecting simulation data onto a three-dimensional globe after clustering.²⁶ First, a hierarchical clustering algorithm was implemented to organize the large amount of simulation data, which was based on pair-wise RMSDs of the DBDs after LBDs were aligned (see Methods). We also note that only the last 5-ns segment within the *release* part of each PPR cycle was used for the analysis because a period of equilibration time before this segment was observed to be helpful for the relaxation of the configurations coming out of PPR pushing segments, as shown in Figure S2. As a result, a total of 4890 structures were selected from the entire CGMD simulations, each representing the lowest $E_{LBD-DBD}$ configuration from its corresponding PPR cycle. Figure 3(A) shows a plot of $E_{LBD-DBD}$ versus radius of gyration (R_g) for these configurations. Clearly, a broad range of configurations were observed in the simulations, ranging in $E_{LBD-DBD}$ from 0 kcal/mol to -54 kcal/mol and R_g from 28 Å to 39 Å.

Second, the energetics of $ER\alpha$ conformational landscape was achieved by the projection of resulting clusters onto an energy globe. Specifically, this globe was colored based on the lowest $E_{LBD-DBD}$ value within each cluster. Figure 3(B) illustrates the projected energy globe, characterizing the $ER\alpha$ overall energy landscape. This characterization was achieved with a clustering analysis procedure with a total number of clusters $N = 300$. Additional analyses using $N = 500$ and $N = 1000$ yielded a similar overall picture of the energy-colored globe [Fig. S3(A)]. Thus, the globe resulting from $N = 300$ clusters was used to complete the analysis. The lowest energy

configurations in the top 25% and 30% of all unbiased simulations were also projected onto the energy globe [Fig. S3(B)]. These results show that each state remains at a similar energy basin when different pools of configurations are used in the projection, suggesting that these CG-identified states are capable of representing and illustrating the simulated landscape. Overall, compared to traditional two-dimensional plots as shown in Figure 3(A), this energy globe analysis offers an advantage for the navigation of a complex conformational space and for the identification of stable ER α conformations.

Identification of multiple ER α conformations

The energy-landscape-on-a-globe analysis readily revealed that multiple ER α conformational states are physically accessible via CGMD simulations. Figure 3(B) outlines five energetically favorable ER α states (marked by *i-v*), which can be easily identified from this colored globe. Clearly, these states differ not only in their LBD-DBD interacting energies, but also in their domain interfaces. Based on their surface views [Fig. 3(C)] and structural ensembles (Fig. S4), both states *iv* and *v* are in a more extended conformation with larger R_g values than states *i*, *ii*, and *iii*. Within these two states, only one monomer, either LBD or DBD, is involved in LBD-DBD interactions, while both DBDs make contact with both LBDs in three other states. Notably, such a large difference in overall size, e.g., about 3 Å between state *iii* and state *iv/v*, could be experimentally characterized by small-angle X-ray scattering (SAXS).

Among these observed states, different LBD-DBD binding interfaces are formed. For example, state *iv* is located at the upper region of the globe, which has a distinct LBD binding surface compared to the rest of four states [Figs. 3(C) and S5(A)]. It should be noted that this interface is observed to be partially involved in the interactions of ER α LBD with a peptide antagonist [Fig. S5(B)].^{44, 45} A closer look at the LBD surface shows that a similar region is shared and utilized for the formation of close LBD-DBD interfaces at states *i*, *ii*, *iii*, and *v* [Fig. S6(A)], where mostly one LBD monomer participates in the interfacial interactions. From the LBD

side, this interface mainly consists of two distant sequence segments: helices H3-H5 and helix H12, marked in blue and magenta, respectively [Fig. 4(A)]. A portion of these LBD surfaces also form the binding site for the interactions with a co-activator peptide¹⁹ [Fig. S6(B)] as well as for interactions with 4-hydroxytamoxifen observed in ER β ⁴⁶.

One striking feature is that H12 is crucial in mediating LBD-DBD interactions in several conformational states. Figure 4(B) shows that H12 forms a large number of contacts with DBD at states *i*, *ii*, *iii*, and *v*, and most of them contribute to the predominant hydrophobic inter-domain interactions (Fig. S7). This involvement of H12 was further tested by mutating several charged and hydrophobic residues in helix H12 (Fig. S8). The destabilization of domain interactions is clearly reflected by the loss of interacting energies, suggesting that H12 plays a structural role in gluing and stabilizing inter-domain interactions. Furthermore, compared with previous crystallographic studies on LBDs alone, in which this H12 helix is structurally re-orientated upon estradiol binding.^{17, 19, 20} the H12 involved in domain interface maintains this estradiol-bound orientation in states *i*, *ii* and *iii*. In fact, this particular H12 adopts a similar estradiol-bound conformation, in which its native contacts with LBD were sacrificed, throughout all of these low $E_{\text{LBD-DBD}}$ states, suggesting that its interactions with DBD can stabilize receptor LBD-DBD conformations. Given the previous observation that the LBD, in the presence of hormone, promotes or stabilizes tighter DNA binding,⁶ our results strongly suggest that H12 plays a key role in receptor stabilization associated with tight DNA binding. This observation is also consistent with the fact that deletion of H12 can eliminate the ability of ER α to bind DNA.⁶ Together, these findings indicate that H12 is a key structural modulator in hormone-activated ER α allostery; it mediates internal communication by bridging LBD and DBD association, through which the re-orientation of H12, triggered by hormone binding, results in tighter DNA binding.

Another LBD surface involved in LBD-DBD interfaces consists of helices H3, H4 and H5 [labeled in blue in Fig. 4(A)]. Figure 4(C) shows that these helices make a large number of

contacts with DBD in states *i, ii, iii, v*, while almost no contact occurs in state *iv*. More details, including residue locations and total contacts, on these interfacial sites are also illustrated in Figures S5 and S6. Overall, this H3-H5 surface, together with H12, accounts for the stabilization of dominant LBD-DBD conformations.

It should be noted that one exception to this communication is state *iv*, in which H12 is not involved in the formation of its LBD-DBD interface. Figure 4(B) shows that nearly no contact is formed between H12 and DBDs at state *iv*. More strikingly, this state has a similar domain orientation regarding LBD interface-forming surfaces as observed in a new crystal structure of the nuclear receptor homodimer HNF4 α (PDB entry 4IQR)⁴⁷ (Fig. S9), where the top regions of LBD surfaces, instead of the region of H12 and H3-H5, are involved in the interface formation in both conformations. While there is a difference in their corresponding DNA segments (an inverted repeat for ER α and a direct repeat for HNF4 α), this result highlights that the simulations are capable of capturing the large-scale domain organizations.

The DBD surfaces involved in forming interfaces vary from state to state. Figure 4(D) highlights those interacting residues on DBD surfaces that form multiple contacts with its LBD partner. Among all of these states, the surface itself mainly consists of a loop connecting the first two helices as well as a large portion of the third helix. The key difference, among states *i, ii, iii*, and *v*, is that different sets of DBD residues are utilized for the interfacial packing, resulting in distinct LBD-DBD orientations. This large-scale difference of domain orientations could be examined by small-angle scattering measurements, which was not attempted here but will be reported in future communications. In addition, both DBD monomers interact with LBDs in states *i – iv* [Fig. 4(D)].

This observation that both DBD monomers can interact with LBDs suggests a possible role of LBD in receptor stabilization. It has been suggested that the interactions between two DBD monomers are relatively weak.⁴⁸ Here, based on the simulation results, their interactions with LBDs can be rather strong with a gain of energy, $E_{\text{LBD-DBD}} > 30$ kcal/mol (Fig. 3). This

stabilization via LBD interactions is also consistent with the observation that the binding of receptor to DNA is tighter in the presence of hormone-bound LBD compared to DBD alone.⁶ These results suggest that hormone-bound LBD dimers can serve as a scaffold to position the DBD dimer, resulting in the stabilization of the DBD-DNA complex and an increase in DNA binding.⁴⁸ This notion is further supported by the observed increase of DNA binding affinity when two DBD monomers are connected by an antibody or a linker.⁴⁸ Thus, it is likely that the presence of LBDs stabilizes receptor binding to DNA, similar to the antibody or the linker, in which the LBD scaffold is essential for the correct positioning of the DBD dimer.

Refinement by all-atom simulations

To examine the stability and detailed features of ER α structures, all-atom MD simulations were performed starting from the CG-predicted states (see details in Methods). Figure 5(A) shows that all five states maintain a RMSD value of 3 Å (relative to their corresponding starting configurations), suggesting that each state is significantly stable within its 15-ns simulation window. This is further illustrated by the superposition of the CG and all-atom models [Fig. 5(B)]. While large displacement (up to 3 Å) can occur at the loops or domain-connecting hinge loops, little deviation is observed at the domain interfaces or the core regions of each domain. Thus, these all-atom simulations supports the overall assessment of predicted conformational states by CG simulations.

Close proximity of lysine residues at the DBD-LBD interfaces are observed via detailed analyses of all-atom simulation data. Figure 5(B) shows that several lysine residues are within a distance of 10 Å (e.g., K₂₄₄-K₅₂₀ at state *i*, K₂₄₄-K₃₆₂ at state *ii*, K₂₃₁-K₃₀₃ or K₂₃₁-K₃₆₂ at state *iii*, and K₂₁₃-K₄₉₂ at state *iv*). **This observation indicates that these lysine residues could serve as candidate sites of chemical cross-linking for experimental validation. It should be noted that the close proximity of lysine residues is observed under a rather low salt condition [Fig. S10(A)]. At higher salt concentrations (e.g. under a physiological condition), negative ions**

could further aid favorable interactions between positively-charged lysine residues. In addition, alternative residues, such as tyrosine and cysteine, could be used in cross-linking experiments based on their pair-wise distances (shown in Fig. S10).

Conclusion

A generalized PPR-CGMD simulation pipeline has been applied to the investigation of internal interactions between the ligand- and DNA-binding domains of ER α . Projection of simulation data onto a three-dimensional energy globe reveals multiple LBD-DBD orientations that are energetically stable, instead of a singular stable conformation. This conformational multiplicity is mainly due to the long LBD-DBD hinge, whose intrinsic flexibility allows each domain to move freely in three dimensions. Given the increased demands for structural knowledge about multi-domain proteins^{49, 50, 51}, it appears that PPR-CGMD simulations can be broadly applicable to the exploration of domain interactions, e.g., within a multi-domain complex.

Structural refinement is achieved by employing atomic-level simulations starting from CG-predicted structural models. This multi-scale approach has provided a stability check for these CG models. It also reveals structural features, especially at the domain interfaces, which could be used for validation designs using SAXS and/or cross-linking measurements.

There are two major findings revealed from these simulations. First, among identified ER α conformations, the activation helix H12 is critical in mediating the cross-talk between LBD and DBD, where four out of five stable states utilize H12 as part of the domain interface. This result suggests that H12 is not only responsive upon ligand binding, as observed in crystallographic studies of LBD alone^{17, 19, 20}, but also critical for receptor-DNA binding. Second, the stabilization of LBD-DBD interaction via multiple LBD surfaces also suggests a new role of LBD as a scaffold for the stabilization of the DBD-DNA complex, as observed in *in vitro* studies where an increase of DNA binding is observed due to the presence of LBD. Together with previous

observations, our findings suggest a molecular model for this critical hormone signaling, where the LBD, when bound to estradiol, stabilizes the DBD dimer for tight DNA binding via LBD surfaces including H12.

Finally, it is worth noting that the focus here was to identify receptor conformations that are energetically stable. This focus was inspired by recent observations that formation of close LBD-DBD interface in other NHR family members^{22, 24}, as a result of the formation of compact NHR conformations. However, it should also be noted that an elongated/extended receptor conformation has been observed for NHR heterodimers in solution or at cryogenic temperatures.^{52, 53} The possibility of a similar extended conformation for ER α will be explored in future studies.

Acknowledgments

This work was supported in part by the Cleveland Foundation and the Department of Defense Breast Cancer Research Program (W81XWH-11-1033). Computational support was provided by the CWRU High Performance Computing Cluster.

References

1. Krust, A., Green, S., Argos, P., Kumar, V., Walter, P., Bornert, J., and Chambon, P. The chicken oestrogen receptor sequence: homology with v-erbA and the human estrogen and glucocorticoid receptors. *EMBO J* 5:891–897, 1986.
2. Kumar, V., Green, S., Staub, A., and Chambon, P. Localisation of the oestradiol-binding and putative DNA-binding domains of the human oestrogen receptor. *EMBO J* 5(9):2231, 1986.
3. Kumar, V., Green, S., Stack, G., Berry, M., Jin, J. R., and Chambon, P. Functional domains of the human estrogen receptor. *Cell* 51(6):941–951, 1987.
4. Evans, R. M. The steroid and thyroid hormone receptor superfamily. *Science* 240(4854):889–895, 1988.
5. Notides, A., Lerner, N., and Hamilton, D. Positive cooperativity of the estrogen receptor. *Proc Natl Acad Sci USA* 78(8):4926–4930, 1981.
6. Kumar, V. and Chambon, P. The estrogen receptor binds tightly to its responsive element as a ligand-induced homodimer. *Cell* 55(1):145 – 156, 1988.
7. Wang, H., Peters, G., Zeng, X., Tang, M., Ip, W., and Khan, S. Yeast two-hybrid system demonstrates that estrogen receptor dimerization is ligand-dependent in vivo. *J Bio Chem* 270(40):23322–23329, 1995.
8. Nilsson, S., Mkel, S., Treuter, E., Tujague, M., Thomsen, J., Andersson, G., Enmark, E., Pettersson, K., Warner, M., and Gustafsson, J. A. Mechanisms of estrogen action. *Physiol Rev* 81(4):1535–1565, 2001.
9. Dahlman-Wright, K., Cavailles, V., Fuqua, S. A., Jordan, V. C., Katzenellenbogen, J. A., Korach, K. S., Maggi, A., Muramatsu, M., Parker, M. G., and Gustafsson, J. A. Interna-

- tional union of pharmacology. LXIV. Estrogen receptors. *Pharmacol Rev* 58(4):773–781, 2006.
10. Jordan, V. C. Tamoxifen: a most unlikely pioneering medicine. *Nat Rev Drug Discov* 2(3):205–213, 2003.
 11. Ettinger, B., Black, D. M., Mitlak, B. H., Knickerbocker, R. K., Nickelsen, T., Genant, H. K., Christiansen, C., Delmas, P. D., Zanchetta, J. R., Stakkestad, J., Glücker, C. C., Krueger, K., Cohen, F. J., Eckert, S., Ensrud, K. E., Avioli, L. V., Lips, P., and Cummings, S. R. Reduction of vertebral fracture risk in postmenopausal women with osteoporosis treated with raloxifene: results from a 3-year randomized clinical trial. multiple outcomes of raloxifene evaluation (more) investigators. *JAMA* 282(7):637–645, 1999.
 12. Osborne, C. K., Zhao, H., and Fuqua, S. A. Selective estrogen receptor modulators: structure, function, and clinical use. *J Clin Oncol* 18(17):3172–3186, 2000.
 13. Tora, L., White, J., Brou, C., Tasset, D., Webster, N., Scheer, E., and Chambon, P. The human estrogen receptor has two independent nonacidic transcriptional activation functions. *Cell* 59(3):477–487, 1989.
 14. Schwabe, J. W., Chapman, L., Finch, J. T., Rhodes, D., and Neuhaus, D. DNA recognition by the oestrogen receptor: from solution to the crystal. *Structure* 1(3):187–204, 1993.
 15. Schwabe, J. W., Chapman, L., Finch, J. T., and Rhodes, D. The crystal structure of the estrogen receptor DNA-binding domain bound to DNA: how receptors discriminate between their response elements. *Cell* 75(3):567–578, 1993.
 16. Schwabe, J. W., Chapman, L., and Rhodes, D. The oestrogen receptor recognizes an imperfectly palindromic response element through an alternative side-chain conformation. *Structure* 3(2):201–213, 1995.

17. Brzozowski, A. M., Pike, A. C., Dauter, Z., Hubbard, R. E., Bonn, T., Engström, O., Ohman, L., Greene, G. L., Gustafsson, J. A., and Carlquist, M. Molecular basis of agonism and antagonism in the oestrogen receptor. *Nature* 389(6652):753–758, 1997.
18. Tanenbaum, D. M., Wang, Y., Williams, S. P., and Sigler, P. B. Crystallographic comparison of the estrogen and progesterone receptor's ligand binding domains. *Proc Natl Acad Sci USA* 95(11):5998–6003, 1998.
19. Shiau, A. K., Barstad, D., Loria, P. M., Cheng, L., Kushner, P. J., Agard, D. A., and Greene, G. L. The structural basis of estrogen receptor/coactivator recognition and the antagonism of this interaction by tamoxifen. *Cell* 95(7):927–937, 1998.
20. Gangloff, M., Ruff, M., Eiler, S., Duclaud, S., Wurtz, J. M., and Moras, D. Crystal structure of a mutant hER α ligand-binding domain reveals key structural features for the mechanism of partial agonism. *J Biol Chem* 276(18):15059–15065, 2001.
21. Shiau, A. K., Barstad, D., Radek, J. T., Meyers, M. J., Nettles, K. W., Katzenellenbogen, B. S., Katzenellenbogen, J. A., Agard, D. A., and Greene, G. L. Structural characterization of a subtype-selective ligand reveals a novel mode of estrogen receptor antagonism. *Nat Struct Biol* 9(5):359–364, 2002.
22. Chandra, V., Huang, P., Hamuro, Y., Raghuram, S., Wang, Y., Burris, T. P., and Rastinejad, F. Structure of the intact PPAR γ -RXR α nuclear receptor complex on DNA. *Nature* 350–356, 2008.
23. Zhang, J., Chalmers, M. J., Stayrook, K. R., Burris, L. L., Wang, Y., Busby, S. A., Pascal, B. D., Garcia-Ordóñez, R. D., Bruning, J. B., Istrate, M. A., Kojetin, D. J., Dodge, J. A., Burris, T. P., and Griffin, P. R. DNA binding alters coactivator interaction surfaces of the intact VDR-RXR complex. *Nat Struct Mol Biol* 18(5):556–563, 2011.

24. Helsen, C., Dubois, V., Verfaillie, A., Young, J., Trekels, M., Vancraenenbroeck, R., De Maeyer, M., and Claessens, F. Evidence for DNA-binding domain–ligand-binding domain communications in the androgen receptor. *Mol Cell Biol* 32(15):3033–3043, 2012.
25. Fiser, A., Do, R., and Šali, A. Modeling of loops in protein structures. *Protein Sci* 9(9):1753–1773, 2000.
26. Ravikumar, K., Huang, W., and Yang, S. Coarse-grained simulations of protein-protein association: An energy landscape perspective. *Biophys J* 103(4):837 – 845, 2012.
27. Yang, S., Onuchic, J. N., and Levine, H. Effective stochastic dynamics on a protein folding energy landscape. *J Chem Phys* 125(5):054910, 2006.
28. Sobolev, V., Sorokine, A., Prilusky, J., Abola, E., and Edelman, M. Automated analysis of interatomic contacts in proteins. *Bioinformatics* 15:327–332, 1999.
29. Kim, Y. C. and Hummer, G. Coarse-grained models for simulations of multiprotein complexes: Application to ubiquitin binding. *J Mol Biol* 375(5):1416 – 1433, 2008.
30. Miyazawa, S. and Jernigan, R. L. Residue-residue potentials with a favorable contact pair term and an unfavorable high packing density term, for simulation and threading. *J Mol Biol* 256:623–644, 1996.
31. Brooks, B. R., Brooks, 3rd, C., Mackerell, Jr, A., Nilsson, L., Petrella, R. J., Roux, B., Won, Y., Archontis, G., Bartels, C., Boresch, S., Caffisch, A., Caves, L., Cui, Q., Dinner, A. R., Feig, M., Fischer, S., Gao, J., Hodoscek, M., Im, W., Kuczera, K., Lazaridis, T., Ma, J., Ovchinnikov, V., Paci, E., Pastor, R. W., Post, C. B., Pu, J. Z., Schaefer, M., Tidor, B., Venable, R. M., Woodcock, H. L., Wu, X., Yang, W., York, D. M., and Karplus, M. CHARMM: the biomolecular simulation program. *J Comput Chem* 30(10):1545–1614, 2009.

32. Yang, S. and Roux, B. Src kinase conformational activation: Thermodynamics, pathways, and mechanisms. *PLoS Comput. Biol.* 4(3):e1000047, 2008.
33. Schlitter, J., Engels, M., and Krger, P. Targeted molecular dynamics: A new approach for searching pathways of conformational transitions. *Journal of Molecular Graphics* 12(2):84 – 89, 1994.
34. Tsui, V. and Case, D. A. Theory and applications of the generalized born solvation model in macromolecular simulations. *Biopolymers* 56(4):275–291, 2000.
35. Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R. D., Kale, L., and Schulten, K. Scalable molecular dynamics with namd. *Journal of computational chemistry* 26(16):1781–1802, 2005.
36. Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A., and Simmerling, C. Comparison of multiple amber force fields and development of improved protein backbone parameters. *Proteins: Structure, Function, and Bioinformatics* 65(3):712–725, 2006.
37. Pérez, A., Marchán, I., Svozil, D., Sponer, J., Cheatham III, T. E., Laughton, C. A., and Orozco, M. Refinement of the AMBER force field for nucleic acids: Improving the description of α/γ conformers. *Biophysical journal* 92(11):3817–3829, 2007.
38. Aqvist, J. Ion-water interaction potentials derived from free energy perturbation simulations. *Journal of Physical Chemistry* 94(21):8021–8024, 1990.
39. Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A., and Case, D. A. Development and testing of a general amber force field. *Journal of computational chemistry* 25(9):1157–1174, 2004.

40. Wang, J., Wang, W., Kollman, P. A., and Case, D. A. Automatic atom type and bond type perception in molecular mechanical calculations. *Journal of molecular graphics and modelling* 25(2):247–260, 2006.
41. Jakalian, A., Bush, B. L., Jack, D. B., and Bayly, C. I. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: I. method. *Journal of Computational Chemistry* 21(2):132–146, 2000.
42. Jakalian, A., Jack, D. B., and Bayly, C. I. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. parameterization and validation. *Journal of computational chemistry* 23(16):1623–1641, 2002.
43. Case, D. A., Cheatham, T. E., Darden, T., Gohlke, H., Luo, R., Merz, K. M., Onufriev, A., Simmerling, C., Wang, B., and Woods, R. J. The Amber biomolecular simulation programs. *Journal of computational chemistry* 26(16):1668–1688, 2005.
44. Paige, L. A., Christensen, D. J., Grn, H., Norris, J. D., Gottlin, E. B., Padilla, K. M., Chang, C. Y., Ballas, L. M., Hamilton, P. T., McDonnell, D. P., and Fowlkes, D. M. Estrogen receptor (ER) modulators each induce distinct conformational changes in er alpha and er beta. *Proc Natl Acad Sci USA* 96(7):3999–4004, 1999.
45. Kong, E. H., Heldring, N., Gustafsson, J. A., Treuter, E., Hubbard, R. E., and Pike, A. C. W. Delineation of a unique protein-protein interaction site on the surface of the estrogen receptor. *Proc Natl Acad Sci USA* 102(10):3593–3598, 2005.
46. Wang, Y., Chirgadze, N., Briggs, S., Khan, S., Jensen, E., and Burris, T. A second binding site for hydroxytamoxifen within the coactivator-binding groove of estrogen receptor β . *Proc Natl Acad Sci USA* 103(26):9908–9911, 2006.

47. Chandra, V., Huang, P., Potluri, N., Wu, D., Kim, Y., and Rastinejad, F. Multidomain integration in the structure of the HNF-4 α nuclear receptor complex. *Nature* 495(7441):394–398, March, 2013.
48. Kuntz, M. and Shapiro, D. Dimerizing the estrogen receptor DNA binding domain enhances binding to estrogen response elements. *J Bio Chem* 272(44):27949–27956, 1997.
49. Levitt, M. Nature of the protein universe. *Proc Natl Acad Sci USA* 106(27):11079–11084, July, 2009.
50. Zheng, W., Schafer, N. P., Davtyan, A., Papoian, G. A., and Wolynes, P. G. Predictive energy landscapes for protein-protein association. *Proc Natl Acad Sci USA* 109(47):19244–19249, 2012.
51. Tompa, P. On the supertertiary structure of proteins. *Nat Chem Biol* 8(7):597– 600, 2012.
52. Rochel, N., Ciesielski, F., Godet, J., , Moman, E., Roessle, M., Peluso-Iltis, C., Moulin, M., Haertlein, M., Callow, P., Mely, Y., Svergun, D. I., and Moras, D. Common architecture of nuclear receptor heterodimers on DNA direct repeat elements with different spacings. *Nat Struct Mol Biol* 18(5):564 – 570, 2011.
53. Orlov, I., Rochel, N., Moras, D., and Klaholz, B. P. Structure of the full human RXR/VDR nuclear receptor heterodimer complex with its DR3 target DNA. *EMBO J* 31(2):291–300, 2012.
54. Pettersen, E., Goddard, T., Huang, C., Couch, G., Greenblatt, D., Meng, E., and Ferrin, T. UCSF Chimera visualization system for exploratory research and analysis. *J Comput Chem* 25(13):1605–1612, 2004.

Figure Legends

Figure 1: Schematic diagram of ER α functional domains of a ligand-binding domain (LBD) and a DNA-binding domain (DBD). LBD and DBD dimers are displayed in green and blue, respectively, and their connecting hinge loops are marked by gray dots. The DBD-specific DNA duplex is shown as ribbon in white. The activation helix H12 is highlighted in red. Part of this figure was rendered using UCSF Chimera.⁵⁴

Figure 2: Sampling coverage using *pull-push-release* (PPR) in CGMD simulations. (A) Illustration of the three-dimensional globe specifying LBD-DBD orientations, where LBDs are in the center (in green) and DBDs are on the surface (in blue). (B) Sampling coverage without and with using PPR. The globe is projected with a single simulation trajectory without (left) and with (right) the assistance of PPR. In both simulations, a same initial configuration and a simulation length of 125 ns were used. (C) Total sampling coverage from all trajectories projected on the globe where each simulation snapshot is represented by a blue dot. Shown is a front view and alternative views are also available in Figure S1. The projected snapshots were taken only from free and released portions of PPR simulation cycles with a energy cutoff of $E_{\text{LBD-DBD}} < 0$ kcal/mol and an ensemble of 10 structures from each cycle.

Figure 3: Energy landscape of ER α multi-domain interactions. (A) A plot of LBD-DBD interacting energy $E_{\text{LBD-DBD}}$ vs ER α -DNA radius of gyration R_g . A total of 4890 snapshots were used for the plot. (B) An energy globe for the projection of simulation data and the identification of energetically stable ER α conformations. The globe is colored based on the energies of $E_{\text{LBD-DBD}}$, as indicated by the color bar on the right. From this energy plot, there are five conformational states *i*, *ii*, *iii*, *iv*, *v* that are identified as energetically stable and ranked in the order of their $E_{\text{LBD-DBD}}$ values. Their $E_{\text{LBD-DBD}}$ and R_g values are -51.6 ± 1.3 kcal/mol and 30.40 ± 0.17 Å for state *i*, -44.2 ± 1.6 kcal/mol and 29.63 ± 0.10 Å for state *ii*, -44.4 ± 1.1 kcal/mol and 28.60 ± 0.08 Å for state *iii*, -39.8 ± 1.9 kcal/mol and 31.58 ± 0.07 Å for state *iv*, and -37.2 ± 1.3 kcal/mol and 31.72 ± 0.33 Å for state *v*. A total of 9780 snapshots were used for the projection, which is doubled due to the homodimeric symmetry. (C) Surface view of a representative structure in each ER α conformational state. For clear visualization, their hinge loops are not shown in these surface views. An ensemble of structures for each state are also shown in Figure S4.

Figure 4: ER α LBD-DBD interfaces. (A) Two LBD segments: helices H3, H4, and H5 (residue 350–385) and helix H12 (residue 536–544), highlighted in blue and magenta, respectively. (B) The interactions of helix H12 with LBDs and DBDs. The calculations of H12-LBD contacts was based on the estradiol-bound LBD crystal structure (PDB entry 1QKU)²⁰, and the calculations of H12-DBD contacts was based on the criterion that the instantaneous distance is within a factor of 1.2 of the corresponding distance for each pair of amino acids as defined in Eq. (2). An ensemble of 10 highest ranked configurations based on $E_{\text{LBD-DBD}}$ were used to calculate the standard deviation (marked by error bars) for each number of contacts. (C) The interactions of helices H3, H4, and H5 with LBDs and DBDs. (D) DBD surface involved in LBD-DBD interactions. Colored are the DBD surface residues according to the number of contacts for each residue (marked by the color bar below). Results for the LBD surfaces are also shown in Figure S5(A) and S6(A).

Figure 5: All-atom simulations of five CG-predicted ER α states. (A) Simulation trajectories of five conformational states $i-v$. RMSDs were calculated based on C α atoms of the DBD-LBD complex (excluding hinge loop and DNA) with reference to the starting structure [shown in Fig. 3(C)]. (B) Comparison of all-atom and CG models. In each panel, a CG model is superimposed with the C α traces of its all-atom model taken from the last frame of the 15-ns simulation trajectory (left) and the displacement between two models is also colored based on their C α distances (right). In addition, lysine residues as candidate sites for chemical cross-linking at the domain interfaces are highlighted in magenta.

Supporting Figure Legends

Figure S1: Sampling coverage achieved by PPR-CGMD simulations. (A) A total of 145 initial configurations used in the simulations, each represented by a blue dot on the globe. Sampling coverage projected into (B) a globe placing LBDs in the center and DBDs on the surface and (C) a globe placing DBDs in the center and LBDs on the surface.

Figure S2: Illustration of simulation data used for clustering analysis. (A) A plot of RMSD vs simulation time for the *release* unbiased portions of all the PPR cycles. The last frame in each cycle was used as a reference for RMSD calculations. (B) A histogram plot of resulting RMSDs from the last 5-ns segments (also highlighted by a dashed box in (A)).

Figure S3: (A) Clustering analysis using different numbers of clusters ($N = 300, 500, 1000$). (B) Projection of the lowest-energy configurations in the top 25% (left) and top 30% of all simulation data.

Figure S4: An ensemble of structures representing each $ER\alpha$ state. A set of 10 configurations was selected from each PPR cycle with the lowest $E_{LBD-DBD}$ values within each state. LBDs are colored in green, DBDs are in blue and the DNA duplex is in white.

Figure S5: LBD surface involved in LBD-DBD interactions. Colored are the LBD surface residues involved in the interactions (A) with its DBD partners at state *iv* and (B) with a peptide antagonist as revealed from a crystal structure (PDB entry 2BJ4).⁴⁵ Each residue on the LBD surface is colored according to the color bar, based on the number of contacts it can make with its DBD or peptide partners, where a contact was considered formed if the instantaneous distance is within a factor of 1.2 of the corresponding distance for each pair of amino acids as defined in Eq. (2).

Figure S6: Comparison of LBD surfaces involved in LBD-DBD interactions. Colored are the LBD surface residues involved in the interactions (A) with its DBD partners at states *i, ii, iii, v* and (B) with the GRIP-1 peptide from a co-activator as observed in a crystal structure (PDB ID entry 3ERD).¹⁹ Each residue on the LBD surface is colored based on the number of contacts it can make with its DBD or peptide partners. A similar contact definition was used as in Figure S5.

Figure S7: Decomposition of the interaction energy into contribution from electrostatic (white bars) and hydrophobic (black bars) interactions for states $i-v$. The height of the bar is the average calculated from the ensemble of each state, and the error bar is the standard deviation.

Figure S8: Mutations in the helix H12 destabilize ER α domain interactions. (A) Shown are several charged and hydrophobic residues in the helix H12. Mutations of a charged residue E546A (B) and hydrophobic residues (C) result in the increase of domain interacting energies. The globe was colored based on the energy increase upon mutation and the energies were evaluated after 1,000 steps of energy minimization for each structure with mutant sequences.

Figure S9: Comparison of state iv (left) with the crystal structure of HNF4 α (right; PDB entry 4IQR)⁴⁷.

Figure S10: Pair-wise distance distributions $g(r)$ of cross-linker residues between DBDs and LBDs in all five states. Shown are the distributions for lysine residues between DBDs and LBDs (A), cysteine residues in the DBD and cysteine in the LBD (B), lysine residues in the DBD and tyrosine residues in the LBD (C) and tyrosine residues in the DBD and lysine residues in the LBD (D). The histogram $g(r)$ was calculated based on a total of 1,500 snapshots from the 15-ns MD simulations.

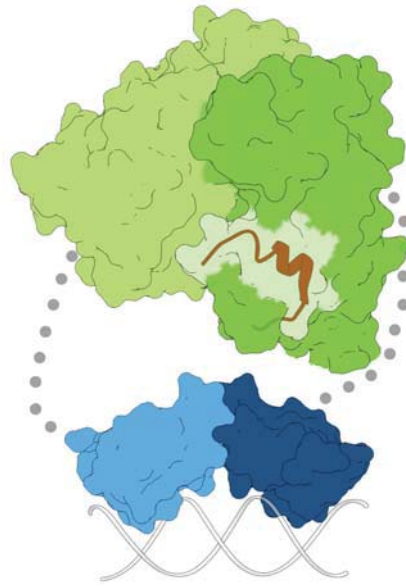


FIGURE 1

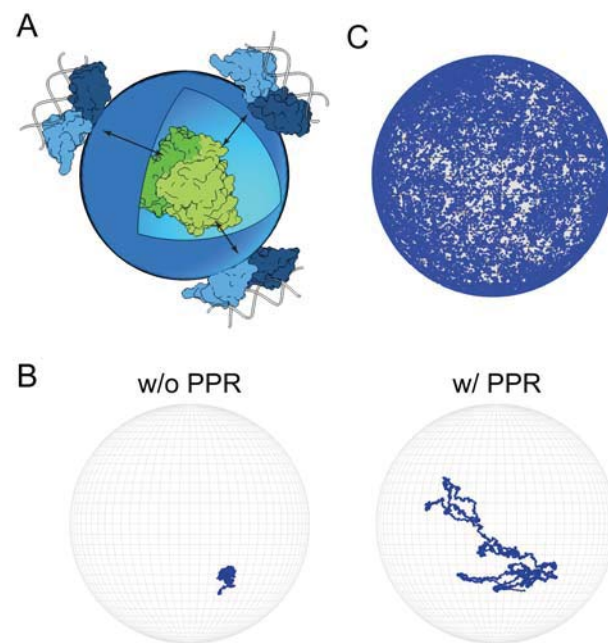


FIGURE 2

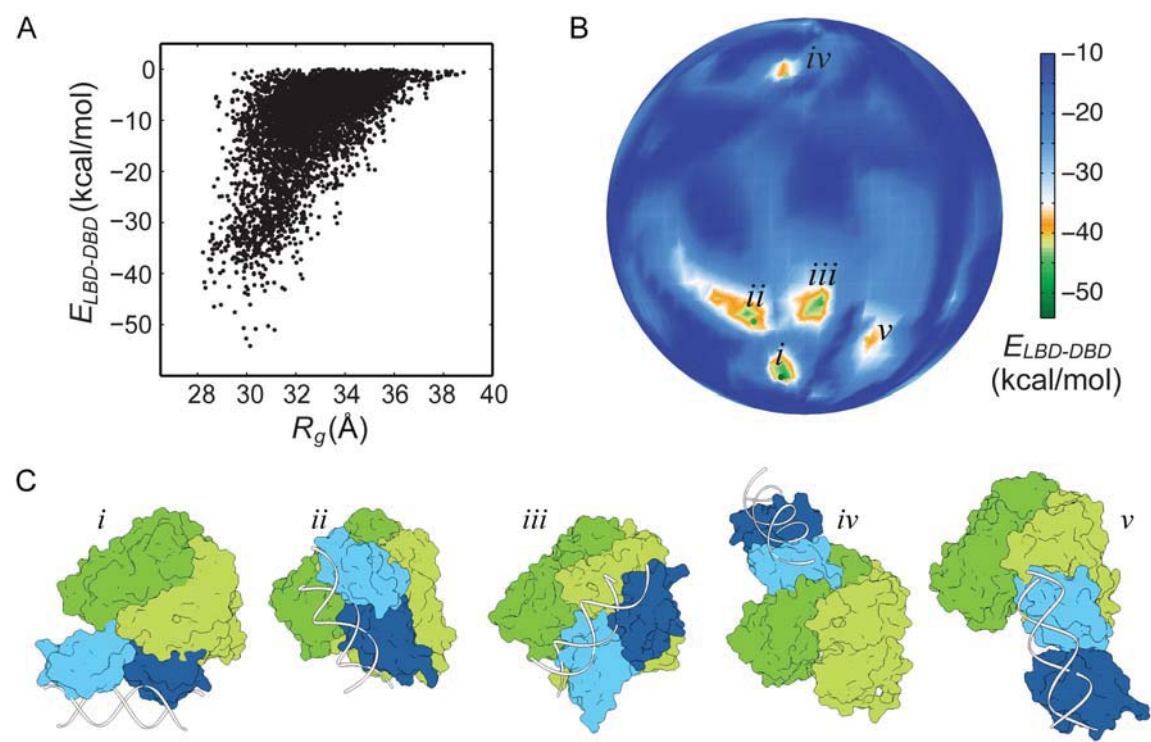


FIGURE 3

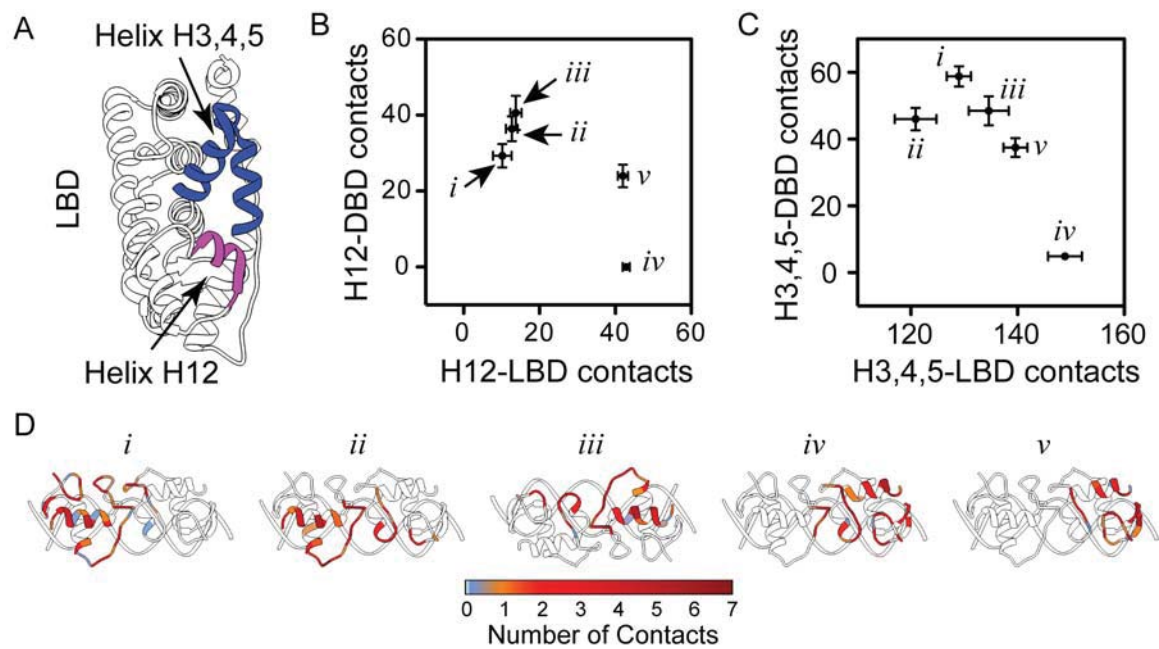


FIGURE 4

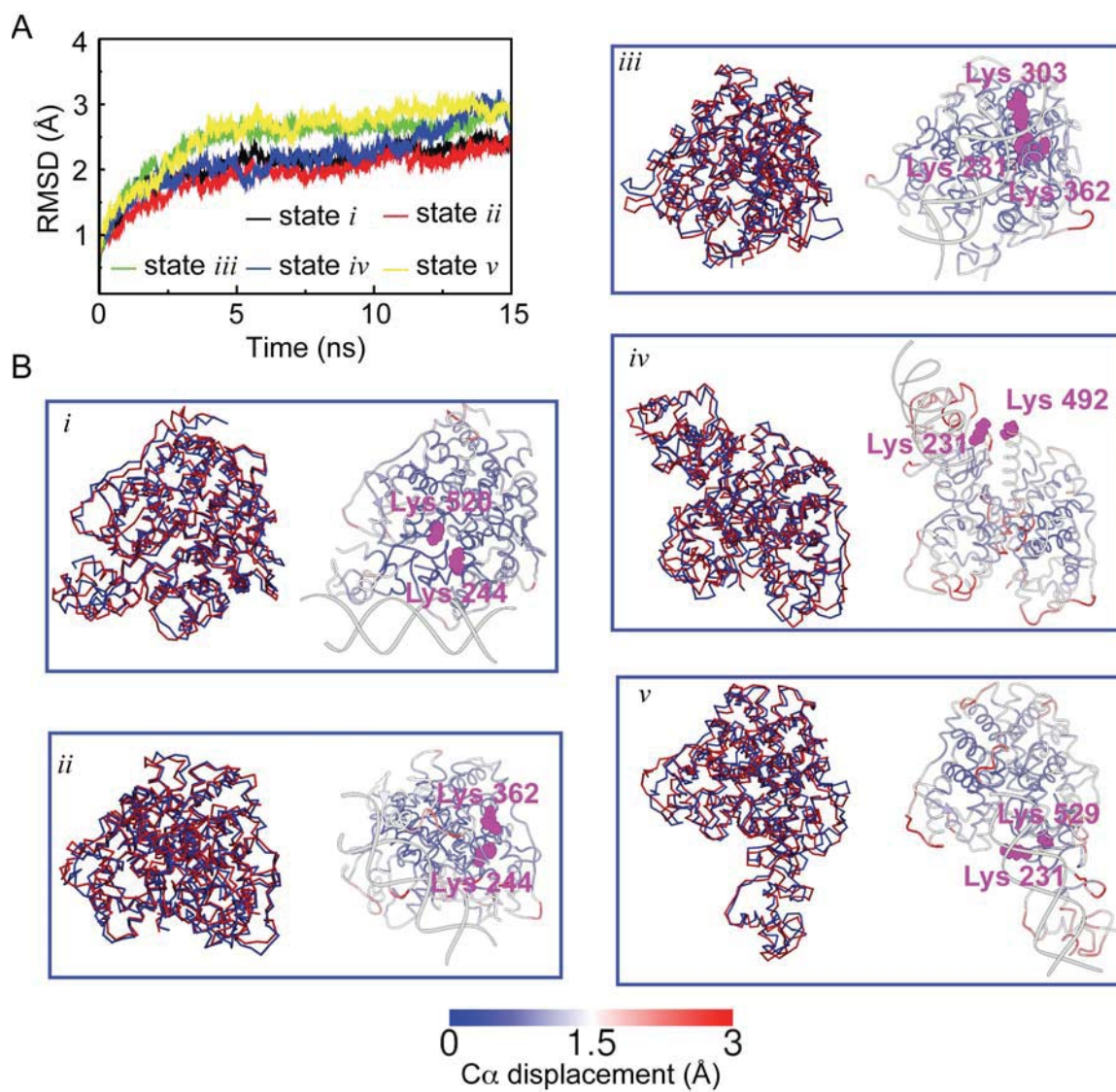


FIGURE 5

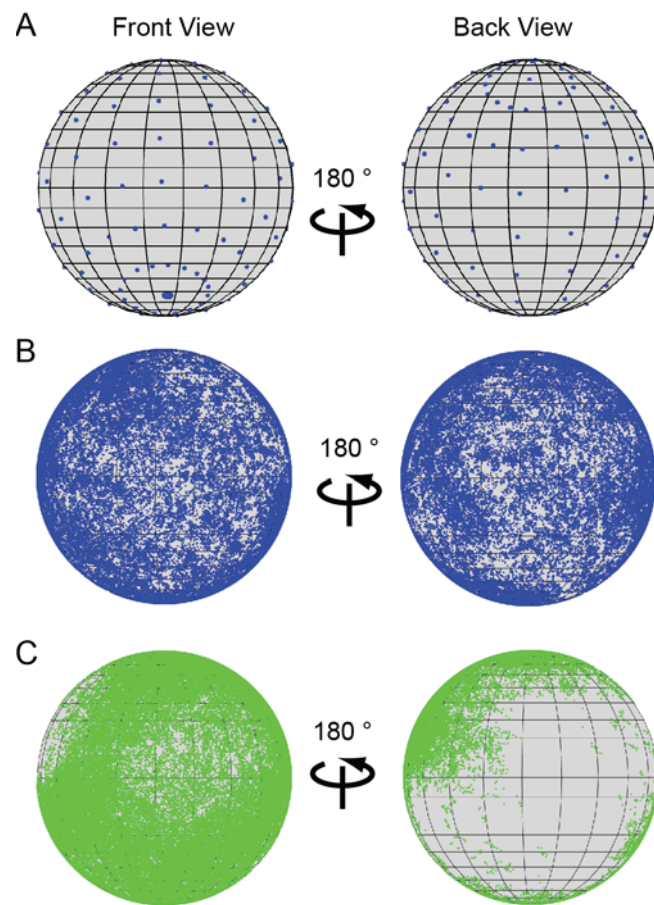


FIGURE S1

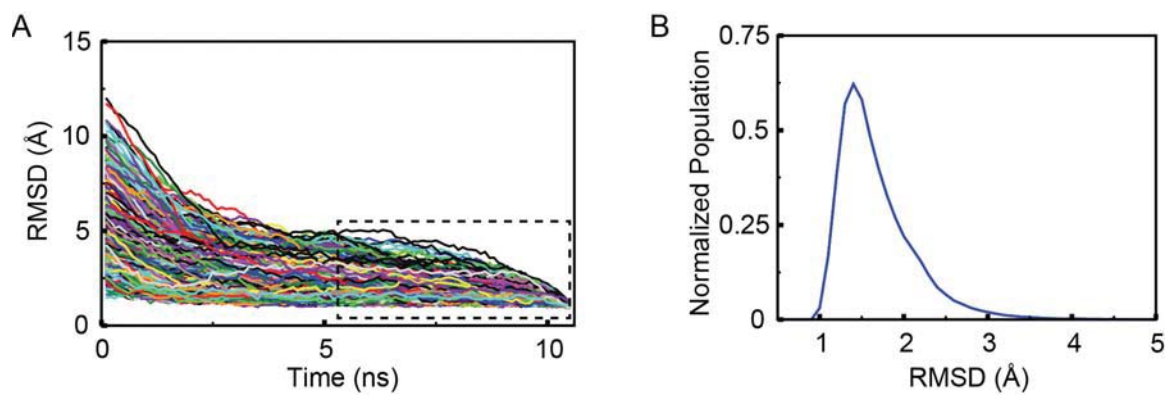


FIGURE S2

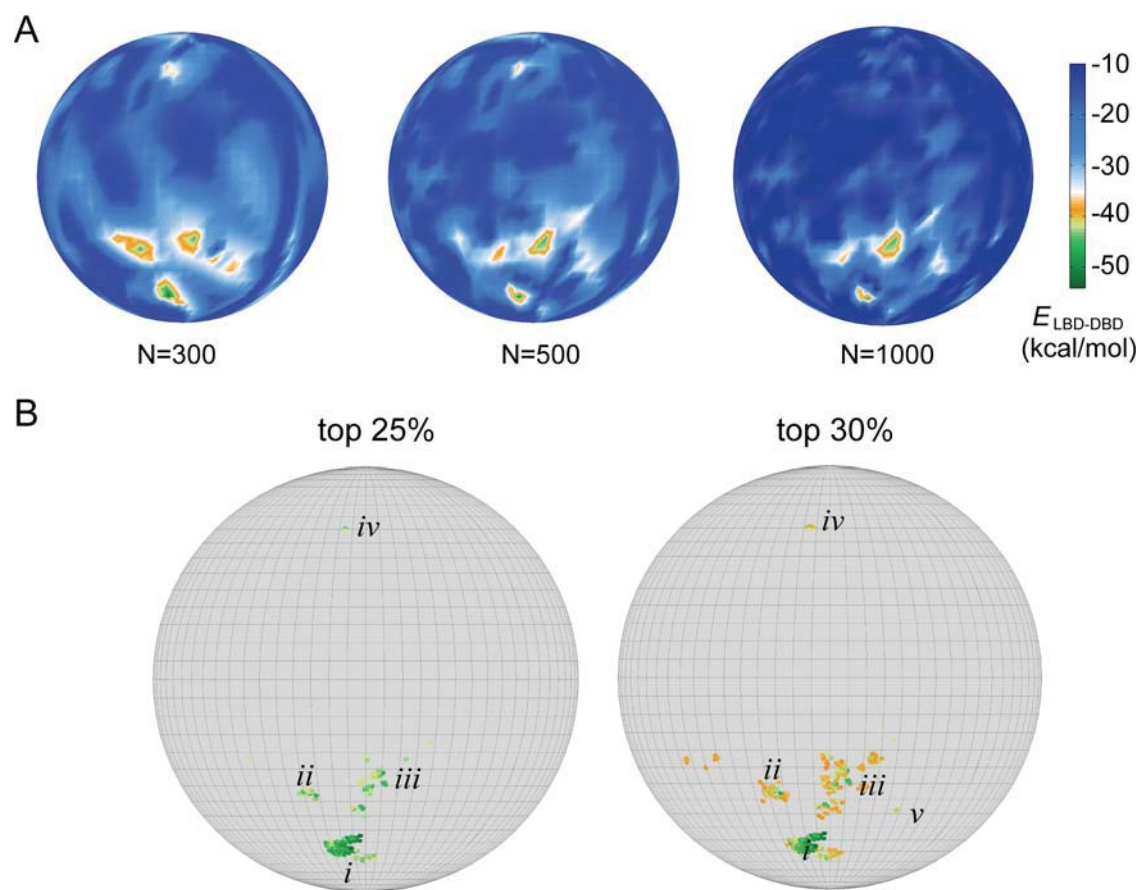


FIGURE S3

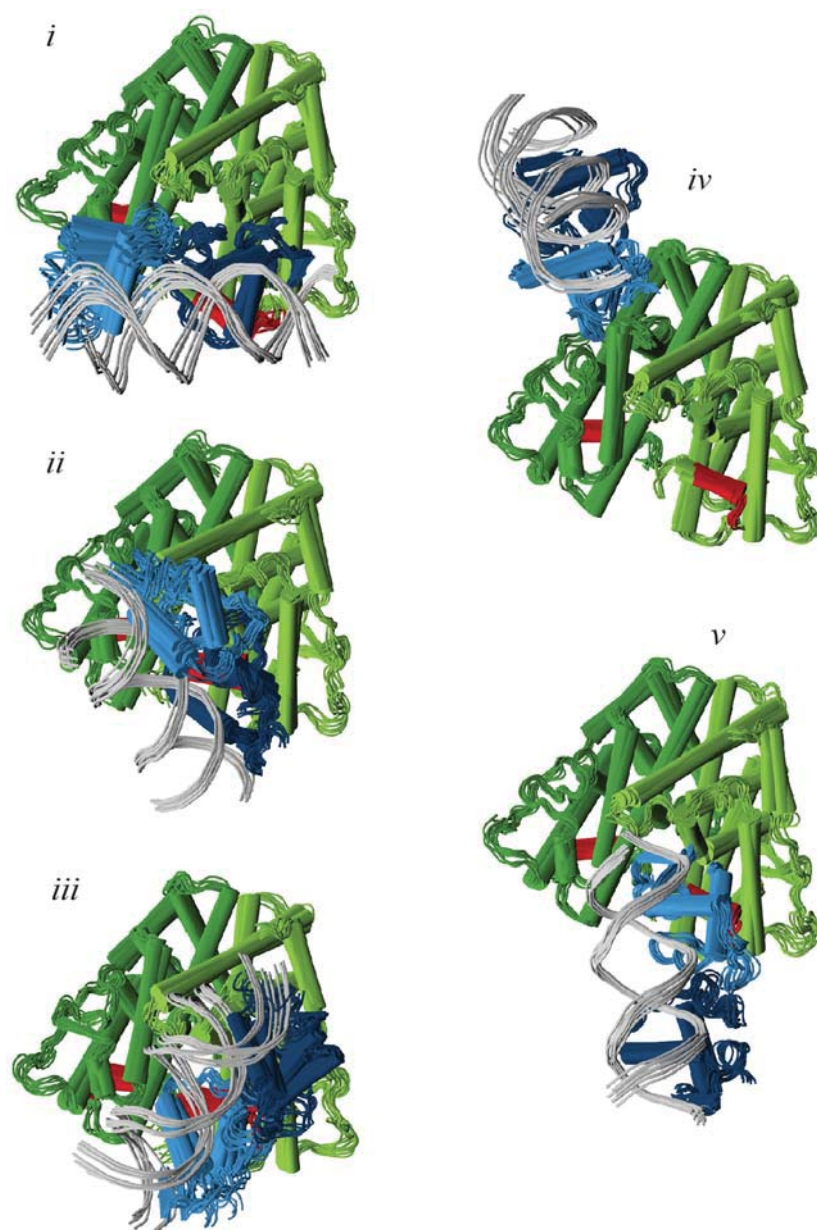


FIGURE S4

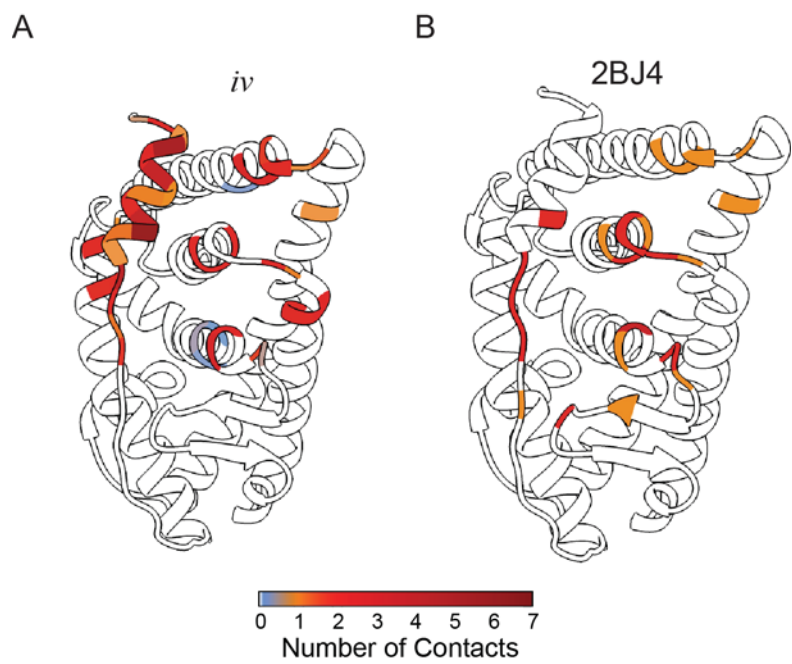


FIGURE S5

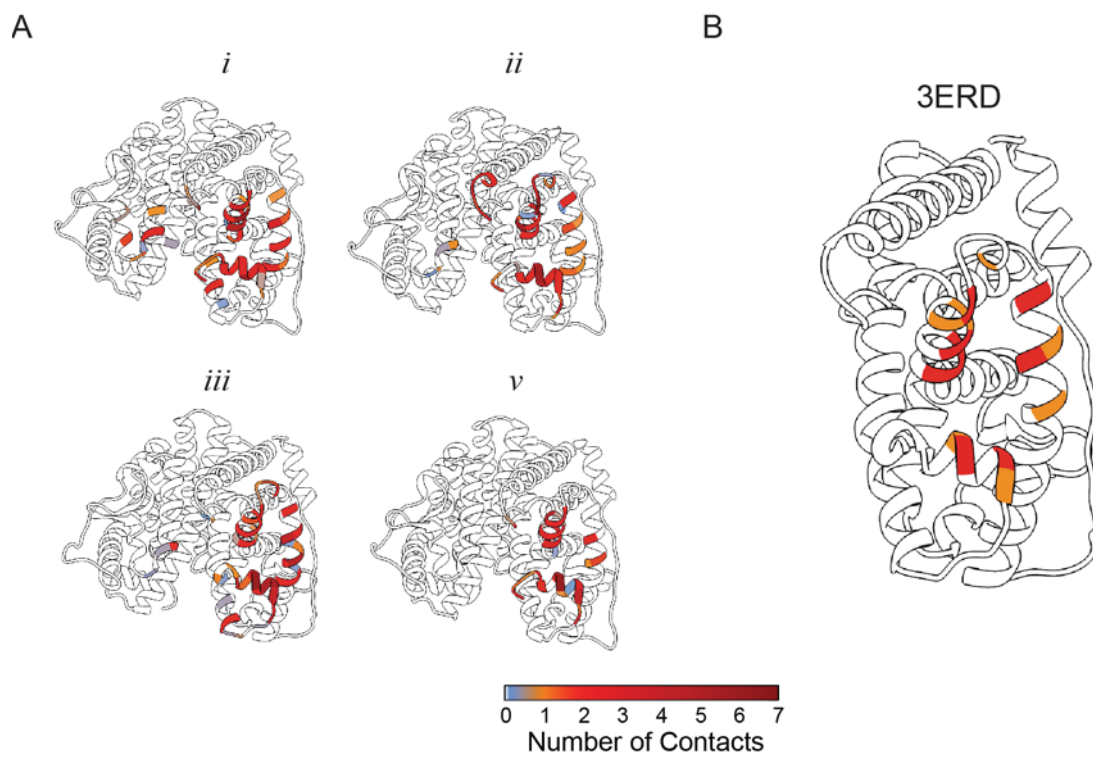


FIGURE S6

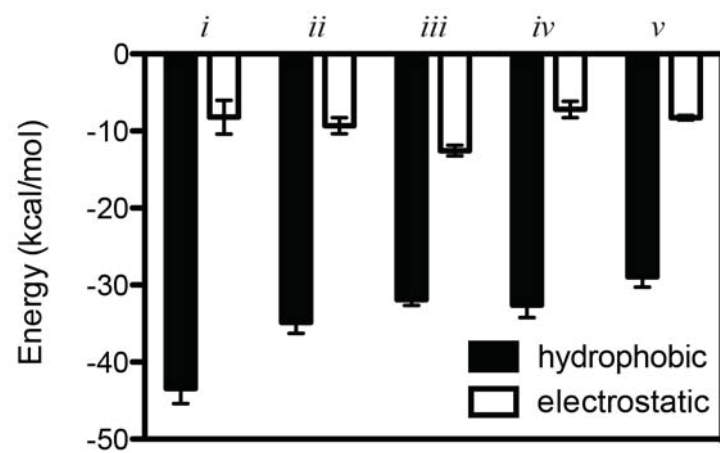


FIGURE S7

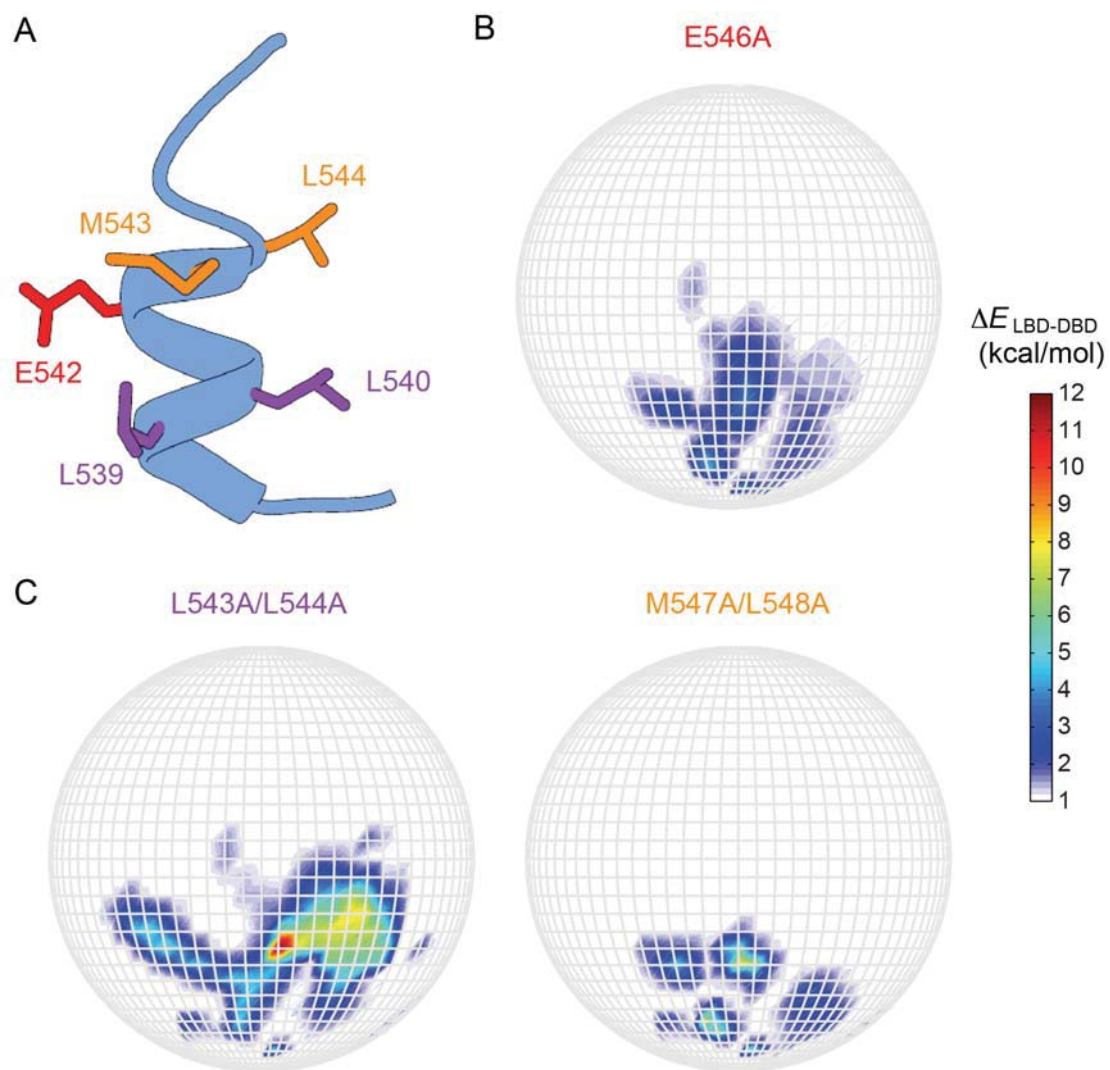


FIGURE S8

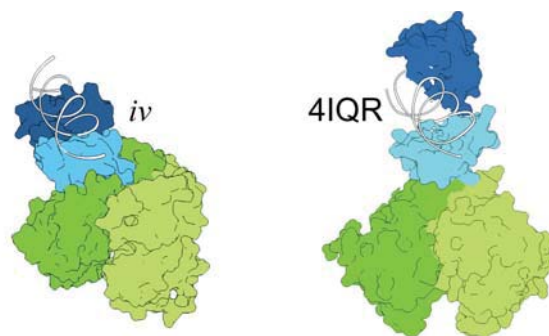


FIGURE S9

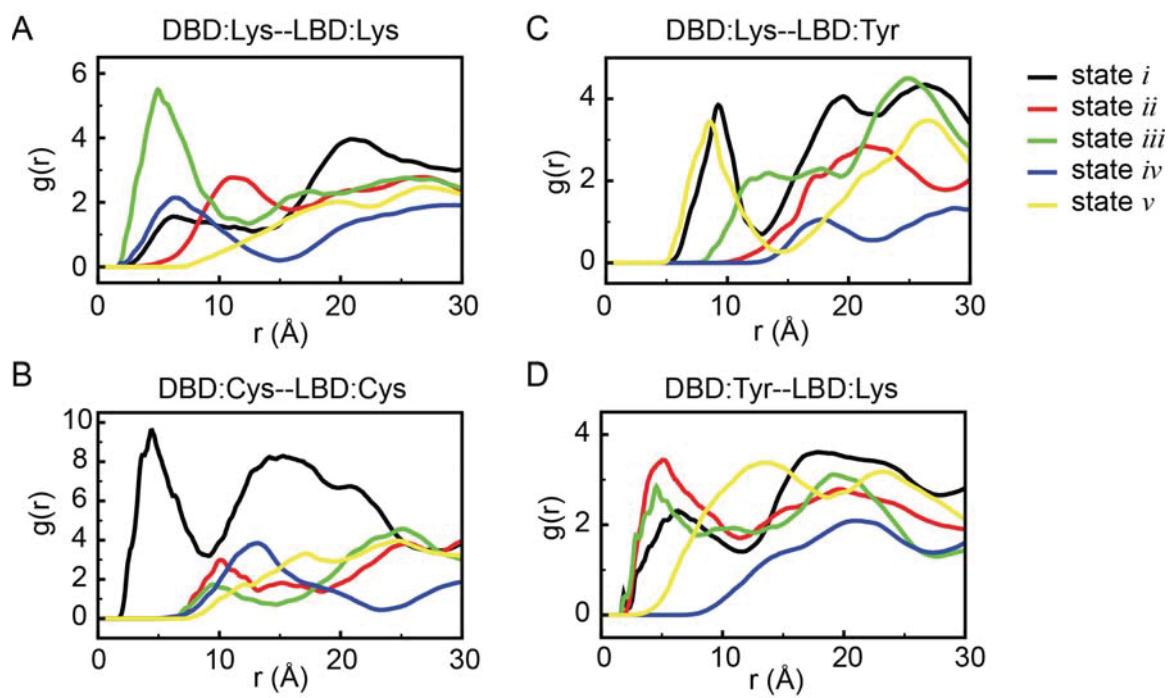


FIGURE S10

Coarse-Grained Simulations of Protein-Protein Association: An Energy Landscape Perspective

Krishnakumar M. Ravikumar, Wei Huang, and Sichun Yang*

Center for Proteomics and Department of Pharmacology, Case Western Reserve University, Cleveland, Ohio

ABSTRACT Understanding protein-protein association is crucial in revealing the molecular basis of many biological processes. Here, we describe a theoretical simulation pipeline to study protein-protein association from an energy landscape perspective. First, a coarse-grained model is implemented and its applications are demonstrated via molecular dynamics simulations for several protein complexes. Second, an enhanced search method is used to efficiently sample a broad range of protein conformations. Third, multiple conformations are identified and clustered from simulation data and further projected on a three-dimensional globe specifying protein orientations and interacting energies. Results from several complexes indicate that the crystal-like conformation is favorable on the energy landscape even if the landscape is relatively rugged with metastable conformations. A closer examination on molecular forces shows that the formation of associated protein complexes can be primarily electrostatics-driven, hydrophobics-driven, or a combination of both in stabilizing specific binding interfaces. Taken together, these results suggest that the coarse-grained simulations and analyses provide an alternative toolset to study protein-protein association occurring in functional biomolecular complexes.

INTRODUCTION

How proteins interact and associate into large functional complexes is one of the key aspects of many biological processes. The use of computational methods to understand molecular details of such protein-protein interactions has provided a powerful alternative to experimental structural characterization, especially for those who form transient but critical metastable conformational states (1,2). Picturing the landscape of protein-protein association is of importance in uncovering the hidden areas of a high-dimensional configurational space, as well as identifying new targets, e.g., using these metastable conformers, for therapeutic designs.

Two prevailing views of protein-protein association are the mechanisms of lock-and-key and induced-fit (3,4). In the former scenario, proteins are treated as rigid bodies, whereas protein flexibility due to the intrinsic dynamics is taken into account in the latter (and its generalizations). The general docking approach, driven by the lock-and-key mechanism, significantly simplifies the search in the conformational space occurring in protein-protein association (5–7). The search is typically based on atomistic representations, but has also been successfully simplified by coarse-grained (CG) models (8,9) that can accelerate energy calculations. However, the rigid body treatment cannot meaningfully account for the intrinsic protein flexibility. Although this problem can be alleviated to some extent by an after search relaxation, flexibility is inherently required for biomolecules to function, as recognized by the induced-fit mechanism. Such flexibility can be achieved computationally using a wide range of methods including molecular

dynamics (MD) simulations (10). Although the restriction to a short timescale is a bottleneck for brute-force atomistic simulations, MD simulations combined with efficient search methods have provided fruitful insights into protein folding and dynamics (10–16).

Here, we explore this general concept of the induced-fit mechanism by employing a combination of MD simulations and a simplified CG protein model with an emphasis on the energy landscape aspects of protein-protein association. This method adopts a widely used structure-based approach to model individual protein components (17–22). It also incorporates the recent implementation introduced by Kim and Hummer (8) accounting for nonnative protein-protein interactions, which would otherwise be lacking without knowing the structure of the entire complex. Furthermore, to avoid trapping due to local stable complex-forming conformations, an efficient search method is introduced with a focus on facilitating protein dissociation and reassociation.

The outline of this work is as follows. First, details of the CGMD method, with a straightforward push-pull-release (PPR) sampling strategy, are described and tested on several model systems. To organize the large amount of simulation data, a structure clustering scheme is applied. The resulting conformations are then projected on a three-dimensional energy globe for visualizing the energetics of relevant stable conformations. Finally, molecular forces stabilizing each identified conformation are briefly described.

MODELS AND METHODS

Details of the CG model

We used a CG model where each amino acid is represented by a single bead positioned at its C α atom. The CG energy function for two interacting proteins/domains (marked as 1 and 2) is formulated as

Submitted May 15, 2012, and accepted for publication July 12, 2012.

*Correspondence: sichun.yang@case.edu

Editor: Alexandre Bonvin.

© 2012 by the Biophysical Society
0006-3495/12/08/0837/9 \$2.00

<http://dx.doi.org/10.1016/j.bpj.2012.07.013>

follows: $E = E_1 + E_2 + E_{12}$. E_1 and E_2 are the energy functions for each protein, similar to the structure-based Gō-type potential (23–28), whereas E_{12} is for the interactions between proteins. In a nutshell, this CG model can be viewed as two Gō-like proteins interacting with each other in a non-Gō-like fashion.

Following Yang et al. (29), E_1 and E_2 were modeled on the basis of their corresponding crystal structures. These energy functions include the interactions for bond (E_{bond}), angle (E_{angle}), dihedral (E_{dih}), and native-like contacts modeled by Lennard-Jones (LJ)-type (E_{LJ}) potentials. Specifically, $E_1 = E_{\text{bond}} + E_{\text{angle}} + E_{\text{dih}} + E_{\text{LJ}}$, where $E_{\text{bond}} = \sum_{\text{bonds}} k_b (r - r_o)^2$, $E_{\text{angle}} = \sum_{\text{angles}} k_a (\theta - \theta_o)^2$, $E_{\text{dih}} = \sum_{\text{dihedrals}} k_\phi^{(n)} [1 + \cos(n(\phi - \phi_o))]$. r , θ , and ϕ are the instantaneous bond distances, angles, and dihedral angles, respectively; r_o , θ_o , and ϕ_o are the corresponding values in the reference structure. We note that the concept of dihedral angle among four residues was used mainly for the convenience of modeling. Force constants $k_b = 100$ kcal/(mol·Å²), $k_a = 20$ kcal/(mol·rad²), $k_\phi^{(1)} = 1.0$ kcal/(mol·rad²), and $k_\phi^{(3)} = 0.5$ kcal/(mol·rad²) were used. The LJ-type interactions for native contacts were used between residues i and j ($\geq i + 4$), $E_{\text{LJ}} = \sum_{i,j} \epsilon_o [5(\sigma_{ij}^o/r_{ij})^{12} - 6(\sigma_{ij}^o/r_{ij})^{10}]$, where $\epsilon_o = 1$ kcal/mol, r_{ij} is the residue-residue distance, and σ_{ij}^o is the corresponding distance in the reference structure. The definition of a native contact-forming pair was based on atomically detailed calculations using the CSU software (30).

The energy function E_{12} is designed for nonnative interactions between two proteins, which were extended from the Kim-Hummer model (8). It includes the electrostatic (E_{elec}) and hydrophobic (E_{H}) components,

$$E_{12} = E_{\text{elec}} + E_{\text{H}}. \quad (1)$$

We used $E_{\text{elec}} = \sum_{i,j} q_i q_j / (4\pi\epsilon_o D_{\text{eff}} r_{ij})$ where q_i is the charge of residue i and ϵ_o is the vacuum electric permittivity. An effective dielectric coefficient $D_{\text{eff}} = D_s \exp(r_{ij}/\xi)$ is applied to reflect the shielding effect between two residues separated by a distance of r_{ij} , where $D_s = 10$ was used to describe the local dielectric environment when two proteins are forming an interface, and $\xi = 8.2$ Å to mimic the screening effect at ~150 mM salt concentration. At pH 7, residue charges $q_i = +e$ for Lys and Arg, $-e$ for Asp and Glu, and $+0.5e$ for His (e is the elementary charge) were used (8). Hydrophobic interactions (E_{H}) are either attractive (LJ-type) ($\epsilon_{ij} < 0$) or purely repulsive ($\epsilon_{ij} \geq 0$) where

$$\epsilon_{ij} = \alpha (e_{ij}^{\text{MJ}} + \beta). \quad (2)$$

$e_{ij}^{\text{MJ}} (< 0)$ is the Miyazawa-Jernigan (MJ) statistical energy between residues i, j (31). β (in unit of $k_B T$) is used as an offset parameter to balance attractive and repulsive interactions, and α to scale E_{H} relative to E_{elec} (8). We used $E_{\text{H}}(i, j) = |\epsilon_{ij}| [5(\sigma_{ij}/r_{ij})^{12} - 6(\sigma_{ij}/r_{ij})^{10}]$ if $\epsilon_{ij} < 0$, and $E_{\text{H}}(i, j) = \epsilon_{ij} [5(\sigma_{ij}/r_{ij})^{12} (1 - \exp(-(r_{ij} - \sigma_{ij})/d)^2)]$ if $\epsilon_{ij} \geq 0$, where $d = 3.8$ Å. A scaling factor of γ is introduced for σ_{ij} as follows,

$$\sigma_{ij} = \gamma (r_i + r_j), \quad (3)$$

where r_i is the van der Waals radius of residue i as used in (8). Note that in this parameter set, both e_{ij}^{MJ} and σ_{ij} can vary between different pairs of residues reflecting the nature of sequence dependency. Finally, E_{12} is accounted only for surface residues with solvent accessible surface area > 10 Å², which was calculated via atomically detailed model structures of individual proteins using a probe size of 1.4 Å.

Simulation and sampling

The CG model was implemented using Langevin MD simulations in a modified version of CHARMM (32). Simulations were performed at 300 K with a friction coefficient of 50 ps⁻¹ (33). A simulation time step of 0.01 ps was used and coordinates were saved every 100 ps.

A PPR sampling strategy, illustrated in Fig. 1, was implemented using a biasing potential E_{PPR} (Eq. 4). The PPR sampling repeats a cycle including the following three parts: i), pull the two proteins away from each other when they are close, ii), push them closer when they are separated by more than a threshold distance R_c , and iii), release them to interact freely by removing the biasing potential. We used the following E_{PPR} ,

$$E_{\text{PPR}} = \begin{cases} 0, & \text{when } R_t \leq R_c \text{ and } r_{\min} < r_o, \\ k(R - R_t)^2, & \text{otherwise} \end{cases}, \quad (4)$$

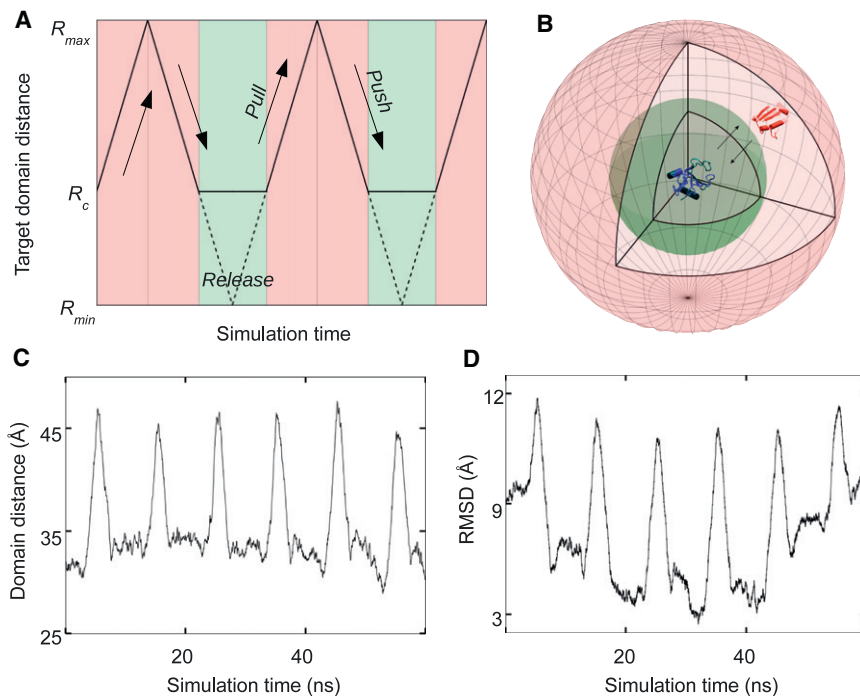


FIGURE 1 PPR sampling scheme. (A) Three parts of a PPR cycle where the pull and push portions are colored in pink and the release in green. The target trajectory (R_t) is shown in solid line ($R_{\min} \leq R_t \leq R_{\max}$). (B) A three-dimensional illustration of the PPR scheme; one protein (in blue) is positioned at the origin. The inner sphere (in green), with a radius of $R = R_c$, is the region where the biasing potential is turned off (see Eq. 4); the outer sphere has the radius of $R = R_{\max}$. R is the center-to-center domain distance between the two proteins. (C) Plot of R versus time and (D) RMSD versus time from a typical PPR simulation trajectory for the barnase/barstar complex, where $R_{\max} = 50$ Å, $R_c = 25$ Å, $R_{\min} = 0$ Å were used.

where R is the instantaneous center-to-center distance, and R_t is the target center-to-center distance (marked by *solid lines* in Fig. 1 A). r_{min} is the closest residue-residue distance between the two proteins. The value of $r_o = 7.6$ Å and a spring constant of $k = 100$ kcal/(mol·Å²) were used. A simulation length of 10 ns was used for each PPR cycle.

To further improve search efficiency, a total of 100 independent MD runs were launched simultaneously, each with an initial configuration generated by translation-and-rotation of the two proteins (see the [Supporting Material](#)). The total simulation time was 10 μs, which resulted in 10⁵ configurations. Configurations from the unbiased release parts (*highlighted in green* in Fig. 1, A and B), were used for data analysis.

Structure clustering analysis

Configurations from the simulated trajectories were grouped following a two-step clustering procedure. In the first orientational clustering step, the entire complex was aligned based only on the crystal structure of one protein. Four out-of-plane residues were then picked from the other protein and their Cartesian coordinates were used in a standard K-means clustering algorithm in MATLAB (The MathWorks, Natick, MA). From the resulting N_c clusters ($2000 \leq N_c \leq 3000$), N_c representative configurations with the lowest E_{12} within each cluster were chosen and used in the next clustering step. In the second root mean-square deviation (RMSD) clustering step, these N_c configurations from the first step were further clustered into N_f final clusters using a pairwise RMSD-based protocol (34) with a RMSD cutoff of 5 Å (for the entire complex). Similar to the first clustering step, the lowest E_{12} configuration within each cluster was selected to represent the N_f clusters. To focus on the identification of energetically stable conformations, this two-step hierarchical clustering was performed only on those configurations with $E_{12} < 0$.

RESULTS AND DISCUSSION

Here, we first describe the CG simulations with the selection of model parameters and test them on several well-characterized complexes. To accelerate the simulations, an efficient search method is introduced and compared with

brute-force simulations. This CG method is finally applied to characterize the energy landscape of several protein complexes: CCP/cc, E9/Im9, E7/Im7, RXR ligand-binding domain (LBD) dimer, and barnase/barstar. Their resultant energy landscapes are further characterized and organized according to the forces that energetically stabilize their identified favorable conformations.

The CG model

To reduce the degrees of freedom in atomistic simulations and overcome the timescale limitation, a CG approach was used in our studies of protein-protein interactions. The CG model was built on the basis of available crystal structures of individual proteins (i.e., Gō-like models; see [Models and Methods](#)). The nonnative-like interactions between proteins were effectively accounted for and optimized. Here, the first optimization is about two CG parameters (α and β in Eq. 2) used to balance the competition between hydrophobic and electrostatic interactions. To achieve this goal, brute-force CGMD simulations (without a biasing potential) were carried out on two protein complexes whose crystal structures are available. One is the barnase/barstar complex of bacterial ribonuclease and its inhibitor (PDB entry 1BRS) (35), and the other is E9/Im9 (PDB entry 1EMV), an immunity protein complex (36). To examine their energetic stability, a range of CG parameters ($\alpha = 0.2, 0.4, 0.6$ and $\beta = 0.8, 1.3, 1.9$) were used for comparison.

Fig. 2 shows the two-dimensional histogram plots of the center-to-center distance between two proteins (R) versus RMSD of the entire complex (with respect to the crystal structure). A total of nine sets of CG simulations, each

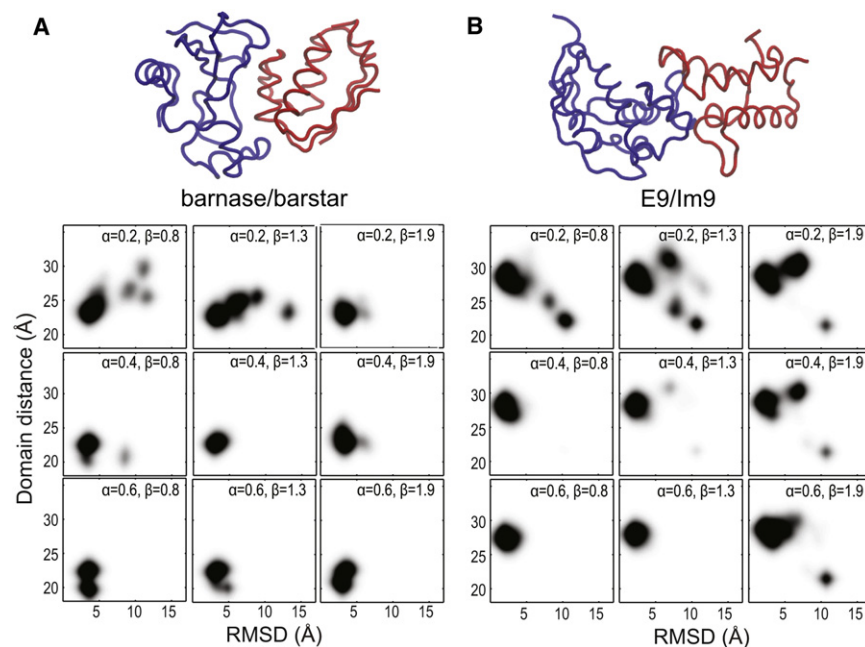


FIGURE 2 Selection of CG model parameters. (A) Histogram plots of R (domain distance) versus RMSD for barnase/barstar (PDB entry 1BRS (35)) and (B) for E9/Im9 (PDB entry 1EMV (36)). A range of α and β (Eq. 2) are used in each simulation set. For each set, 10 independent simulation runs, each lasting 100 ns, were carried out starting from their corresponding crystal conformations shown above, where $R = 23.2$ Å and 27.4 Å, respectively. The parameter $\gamma = 0.625$ (Eq. 3) is used throughout this work unless specified.

with a distinct set of α and β , were performed starting from the same crystal configuration. Comparison shows that the complex remains stable for the set of $\alpha = 0.4$ and $\beta = 1.3$; any deviation tends to result in a destabilization of the crystal conformation and the complex starts to dissociate. This set of parameters were also tested on three other complexes shown in Fig. 3, where each complex stays within a reasonable RMSD range to its crystal conformation within a simulation window of 100 ns. Additional energy calculations on protein-protein interaction, averaged over the 100-ns CGMD simulations with $\alpha = 0.4$ and $\beta = 1.3$, show that the values of E_{12} are -14.7 and -20.1 kcal/mol for barnase/barstar and E9/Im9, respectively. These results are consistent with their measured binding enthalpies of -13.9 and -19.1 kcal/mol (37,38) (Fig. S1 in the Supporting Material). Taken together, these results suggest that this CG energy function, even though highly simplified, can provide a rather detailed energy evaluation on protein-protein interactions.

Another feature of the CGMD simulations is the introduction of a scaling factor γ (Eq. 3) to account for realistic pairwise residue distances. This consideration is in part based on the observation that a typical hydrophobic pair of Leu-Ile and a typical charged pair of Asp-Asp has an optimal distance around 7.8 Å and 7 Å, respectively (39); these are substantively lower than their values of 12.4 Å and 11.2 Å, based on their van der Waals radii (8). To account for such a difference, the value of $\gamma = 0.625$ was selected for rescaling (Eq. 3). To illustrate and compare the difference, two sets of simulations, one with $\gamma = 0.625$ and the other with $\gamma = 1.0$, were performed for barnase/barstar. Fig. S2 shows that, within the same length of simulation time, the complex drifts away from its crystal conformation during the simulations with $\gamma = 1.0$, but is retained during the ones with $\gamma = 0.625$. This is systemically observed in the simulations of other complexes used in this work (data not shown). Thus, this CG parameter $\gamma = 0.625$, together with $\alpha = 0.4$ and $\beta = 1.3$, is used for the rest of the CGMD simulations.

Efficient search method

As demonstrated previously, it is difficult for brute-force CGMD simulations to observe protein dissociation events once two proteins are associated. One goal here is to search for multiple available conformations, which would require a more complete search in the configurational space. In fact, several advanced sampling techniques have been developed in the past to address this quest (11–14,16,40–45). In a similar spirit, a PPR sampling strategy is implemented here to accelerate sampling different interactions. Specifically, a biasing potential (see Eq. 4) is first applied to pull and push the two proteins to facilitate protein dissociation and reassociation, respectively; this bias is then removed and the proteins are released to interact freely when they are close enough. We repeated this PPR cycle to achieve sufficient sampling (Fig. 1).

Fig. 1, C and D, illustrates a typical PPR trajectory from the simulations of barnase/barstar. It shows that barnase and barstar dissociate and reassociate as seen in the center-to-center distance and RMSD (with respect to the crystal structure) during the push and pull parts of the PPR cycle (pink regions in Fig. 1 A). Once the two proteins are close enough, the associated complex is further relaxed by free MD simulations without any bias imposed during the release portion. We note that simulation data only from these free release portions were used for the rest of the analysis.

It is observed that this PPR scheme significantly increases the search efficiency for different protein-protein interacting conformations. Taking CCP/cc for example, with the help from PPR, a larger RMSD (with respect to the starting structure) range (up to 20 Å) and a much broader configuration space is sampled (Fig. S3); in contrast, the complex remains in a crystal-like conformation without using PPR (Fig. 3 A), and is confined near its starting point in these brute-force simulations (Fig. S3). This increased sampling efficiency is also systematically observed in other systems used in this work. To further enhance sampling, a set of

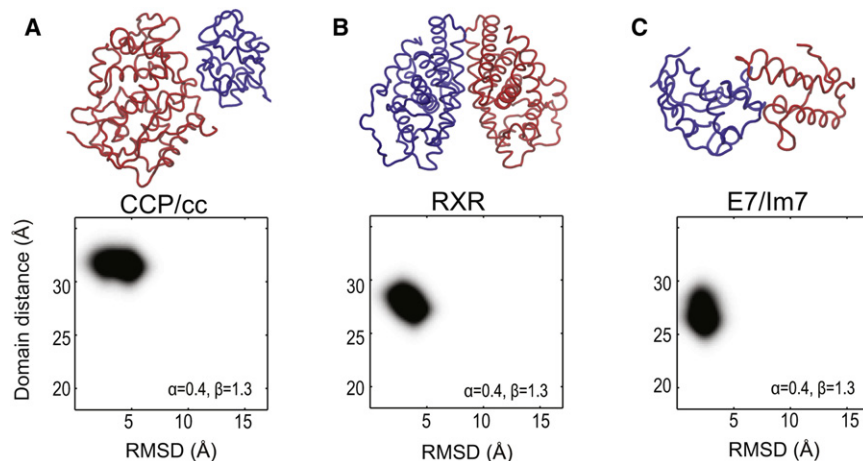


FIGURE 3 CG model parameters are tested on three protein complexes. Two-dimensional histogram plots of R (domain distance) versus RMSD from the simulations with $\alpha = 0.4$ and $\beta = 1.3$ $k_B T$ are shown. Crystal conformations are intact for (A) CCP/cc (PDB entry 2PCC (46)), (B) RXR LBD (PDB entry 1MZN (64)), and (C) E7/Im7 (PDB entry 7CEI (52)) complexes. The simulations lasted 100 ns for each protein complex. This set of parameters $\alpha = 0.4$ and $\beta = 1.3$ $k_B T$ is used throughout this work unless specified.

100 independent MD runs were launched simultaneously each starting with a random orientation between the two proteins in the complex. As shown below, a wide range of conformations are sampled for the protein complexes we examined. Thus, an efficient search method via the PPR strategy is in place to enhance the sampling of protein-protein association.

Protein interacting landscape pictured by an energy globe

To organize the large amount of simulation data, structurally similar conformations were clustered using a two-step structure clustering: the first step is based on their relative orientation and the second is based on pairwise RMSD (see Methods). Typically, clustering is performed only using pairwise RMSD (34); this would be computationally expensive because configurations generated from CGMD simulations are on the order of 10^5 . This two-step procedure overcomes this hindrance by grouping the number of configurations to the order of 10^3 clusters after the first step and finally to N_f clusters in the order of 10^2 .

To assist the navigation of the conformational diversity, we projected these N_f configurations onto a unit sphere or globe representing relative orientation between complex-forming proteins. As illustrated in Fig. 1 B, one can imagine that one protein is inside the globe, whereas the other protein takes different orientations on the surface; the globe was colored according to the interacting energy E_{12} (Eq. 1). We found that this energy-mapped globe is a useful tool to identify energetically favorable conformations with different protein positioning on the landscape. It also serves to access the sampling quality achieved by PPR-assisted CGMD simulations by examining the coverage on the globe surface. Discussed below are five protein complexes

(CCP/cc, E9/Im9, E7/Im7, RXR LBD dimer, and barnase/barstar) studied in this work; we organized them according to the decomposition of the interacting energy into electrostatic and hydrophobic components.

CCP/cc: electrostatics-driven association

To demonstrate the application of CGMD simulations, the CCP/cc complex was first examined; a similar procedure was followed for other protein complexes. CCP/cc is a complex formed between cytochrome *c* peroxidase and cytochrome *c*, whose crystal structure is shown in Fig. 3 A (46). Calculated from the PPR-assisted CGMD simulation data, Fig. 4 A shows a plot of E_{12} versus RMSD (with respect to the crystal structure). It clearly shows that a wide RMSD range is sampled for CCP/cc. A close examination also shows that the crystal-like configurations have a lower E_{12} , indicating that the CG energy function captures the molecular forces stabilizing the crystal-like conformation. It is worth noting that such a correlation between E_{12} and RMSD, where the low E_{12} conformations are funneled into low RMSD regions in the context of protein-protein interactions, somewhat resembles the funnel-like shape recognized in protein folding (47–49).

The resultant $N_f = 191$ conformations after a two-step clustering were projected into an energy globe shown in Fig. 4 B. It is fairly easy to locate conformational states that are energetically stable or metastable. Marked by arrows on Fig. 4 B are four identified lowest E_{12} conformations, each of which consists of an ensemble of five configurations. Among them, conformation (a) has a very similar CCP/cc binding interfaces to the crystal configuration (within 3 Å of RMSD), suggesting that the crystal structure is favored in the CG energy evaluation. Other alternative conformations (b–d) are metastable in E_{12} , which differ in either binding

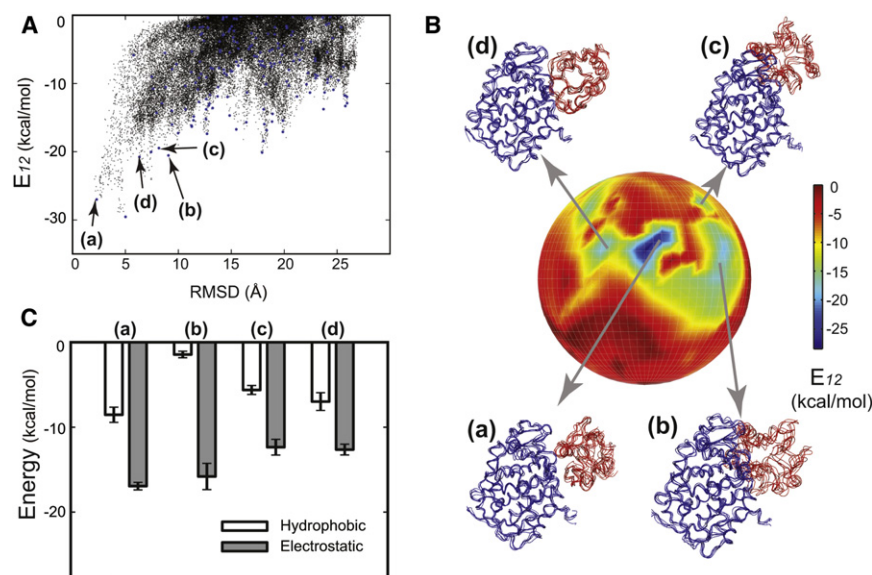


FIGURE 4 Energy landscape of the CCP/cc complex. (A) A plot of E_{12} versus RMSD. A total of 44,451 configurations from CGMD simulations are shown in black dots, where $R_{max} = 90$ Å, $R_c = 40$ Å, and $R_{min} = 0$ Å were used in the PPR scheme. (B) Front view of the energy globe colored by E_{12} . A total of $N_f = 191$ clusters were obtained (also shown in blue dots in A) after the two-step clustering. Four representative conformations (a–d) are shown on the globe, where each conformation is represented by an ensemble of five lowest energy configurations. Conformation (a) with the lowest E_{12} resembles the crystal structure (46). Note that conformations with lower E_{12} are not observed in the back of this globe. (C) The decomposition of hydrophobic (white bars) and electrostatic (gray bars) energies from E_{12} for each conformation (a–d). Their averages (and standard deviations) were calculated from the ensemble of five configurations shown in B.

interfaces or relative orientations. For example, cytochrome *c* (in *red*) is rotated in conformation (*c*) away from the crystal-like conformation (*a*). Similar to what is observed in protein folding, these metastable conformations may serve as important intermediate states right before CCP/cc forms the crystal-like conformation, either thermodynamically or kinetically.

To evaluate the molecular forces driving toward the stable conformation, electrostatic and hydrophobic energies were examined separately according to Eq. 1. Fig. 4 C shows both components in these conformations (*a–d*) where electrostatics dominate over hydrophobic contributions in the total energy E_{12} . This is consistent with that observed in the structural analysis of the CCP/cc complex (46). After all, electrostatic interactions are known to play an important role in protein-protein interactions (50,51). In addition, the crystal-like conformation (*a*) has stronger hydrophobic interactions compared to others (*b–d*), thus suggesting a possible role of hydrophobic interactions in the specificity of its crystal-like binding interface.

This phenomenon of electrostatics-driven protein association is also observed in two other complexes: E9/Im9 and E7/Im7, two immunity proteins in the form of specific bacterial toxin/inhibitor complexes (36,52). These two complexes have similar crystal structures (shown in Figs. 2 B and 3 C) with a sequence identity of 50%. Their PPR-CGMD simulation results exhibit a funnel-shaped plot of E_{12} versus RMSD (Fig. S4 A and Fig. S5 A), similar to what is seen in CCP/cc (Fig. 4 A). Each crystal-like conformation has the lowest E_{12} on the energy globe for both protein complexes (Fig. S4 B and Fig. S5 B), respectively. Four conformations (*a–d*) with low E_{12} identified from the energy globe have similar protein binding interfaces on the immunity protein (in *blue*), although their internal orientations are different. For their energy decomposition

(Fig. S4 C and Fig. S5 C), it appears that E7/Im7 has much higher electrostatic energy than E9/Im9. This difference is mainly due to more charged residues involved at the E7/Im7 interface. It was also observed from mutation studies that a tyrosine (Tyr-54 of Im9) residue at the hydrophobic core of the binding interface is important for the stability of the E9/Im9 complex (36), which might contribute to the less dominant hydrophobic interactions.

RXR LBD dimer: hydrophobics-driven association

A different molecular driving force is observed in the formation of the LBD dimer of a nuclear receptor RXR. Fig. 5 A shows the plot of E_{12} versus RMSD (with respect to the crystal conformation) for this dimer, which is more rugged compared to the CCP/cc complex in Fig. 4 A. The plot is similar to the CCP/cc complex where the crystal-like conformation is favored, although conformation (*d*) has comparable E_{12} with the crystal-like conformation (*a*) (Fig. 5, A and B). In addition, these two conformations (*a*) and (*d*) are close on the energy globe, and differ only by 7 Å in RMSD (Fig. 5 B), suggesting that conformation (*d*) may serve as an intermediate to the formation of the crystal-like conformation. It is also clear that hydrophobic interactions generally dominate the RXR dimer interface, especially in the crystal-like conformation (*a*) (Fig. 5 C). Among all the four conformations, one exception is conformation (*c*) where the second LBD (in *red*) rotates away and binds at distant sites. A close examination on LBD binding interfaces in the crystal structure also shows that parts of the dimeric interface is weakly attractive or even repulsive locally in electrostatics (data not shown), thus suggesting that hydrophobic interactions are the major molecular driving forces in the dimerization.

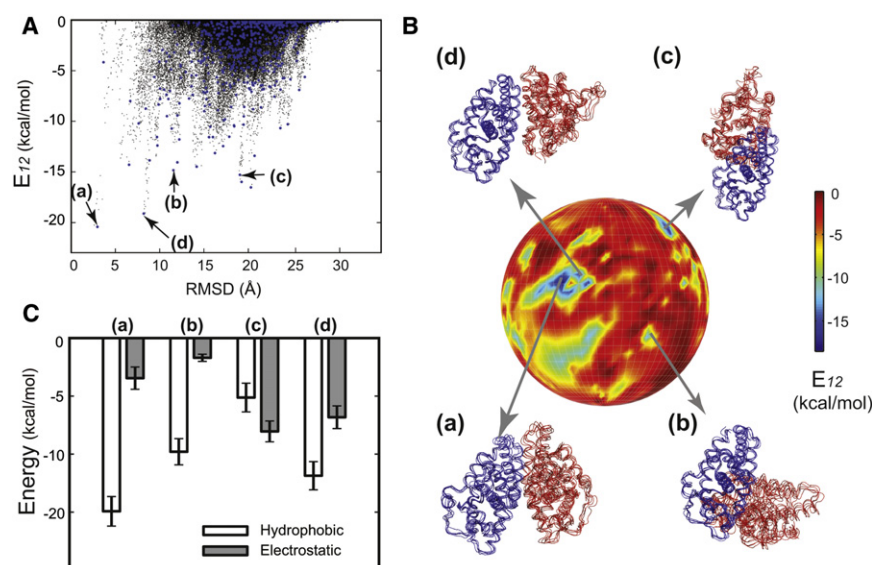


FIGURE 5 Energy landscape of the RXR LBD dimer. (A) A plot of E_{12} versus RMSD. A total of 58,480 configurations from CGMD simulations are shown in black dots, where $R_{max} = 100$ Å, $R_c = 45$ Å, and $R_{min} = 0$ Å were used in the PPR scheme. (B) Front view of the energy globe colored by E_{12} . We used a total of $N_f = 589$ clusters (also shown in blue dots in A) after the two-step clustering. Four representative conformations (*a–d*) are shown on the globe, where conformation (*a*) resembles the crystal structure (64). (C) The decomposition of hydrophobic (white bars) and electrostatic (gray bars) energies from E_{12} for each conformation (*a–d*).

Barnase/barstar: interplay between hydrophobic and electrostatic interactions

The notion of a stable crystal-like conformation is challenged by simulation results of the barnase/barstar complex. Although the crystal conformation displayed stability in brute-force simulations (Fig. 2 A), PPR-based CGMD simulations show that multiple alternative conformations are energetically stable (Fig. 6 A). Fig. 6 B shows four conformations with low E_{12} (*a–d*) on a multibasin energy globe. In particular, conformation (*b*) has a similar E_{12} with the crystal-like conformation (*a*), but barstar (in red) binds at a distant site from the C-shaped binding groove of barnase (in blue). Furthermore, energy decomposition shows stronger hydrophobic interactions in conformations (*a*) and (*d*), whereas electrostatics is stronger in conformations (*b*) and (*c*) (Fig. 6 C). This suggests that an interplay between both electrostatic and hydrophobic interactions is in place among these conformations, in accord with experimental observations on a high degree of both shape and charge complementarity (35).

These simulation results may provide a possible explanation for previous mutation studies. Structural analyses based on the crystal structure show that a set of charged residues (Lys-27, Arg-87, and His-102) in barnase interact with barstar (35), as in the conformation (*a*). In contrast, a different set of charged residues (Arg-59, Glu-60, Lys-62, Lys-66, and Arg-69) are involved in conformation (*b*), contributing to an increased electrostatic energy. One difference is that Arg-59 is at the core of a network of interactions at the binding interface in conformation (*b*), whereas it is on the edge or far away from the core binding interfaces in the crystal-like conformation (*a*). Because of such a critical role of Arg-59 in the conformation (*b*), one would imagine that any disruption might affect its complex association, either thermodynamically or kinetically. Indeed, a

significant change, >400 times in dissociation rate, has been observed in a point mutation of Arg-59 to Ala (53). Additional mutations Asn-58 and Glu-60 at the interface also show substantial change in the rate. This suggests that conformation (*b*), predicted from CGMD simulations, provides a structural basis for the observed large rate change upon mutation. We also note that the barnase surface at the conformation (*b*) is slightly deformed, suggesting that induced-fit helps achieve a better charge complementarity. Taken together, these results suggest that an energy landscape view of protein-protein interactions makes the identification of alternative conformations in barnase/barstar possible, further providing a sound structural basis for mutagenesis.

CONCLUDING REMARKS

We have established a theoretical pipeline to navigate the energy landscape of protein-protein association via PPR-CGMD simulations. The simulations naturally permit and account for the flexibility of protein domains in the realization of induced-fit mechanisms. The use of a PPR sampling scheme enables an exhaustive search to uncover hidden areas of the conformational space. An energy globe is further introduced to navigate the energy landscape of a wide range of resultant conformations. This globe also allows accessing the sampling quality determined by the extent to which the globe is covered by the simulation trajectories. Among four (out of five) protein complexes we examined, their crystal-like conformations are favorable on the energy landscape, suggesting that the CG model captures the basic features of molecular forces driving protein-protein association. One exception is barnase/barstar, where apart from the crystal-like conformation, alternative conformations are also energetically favored.

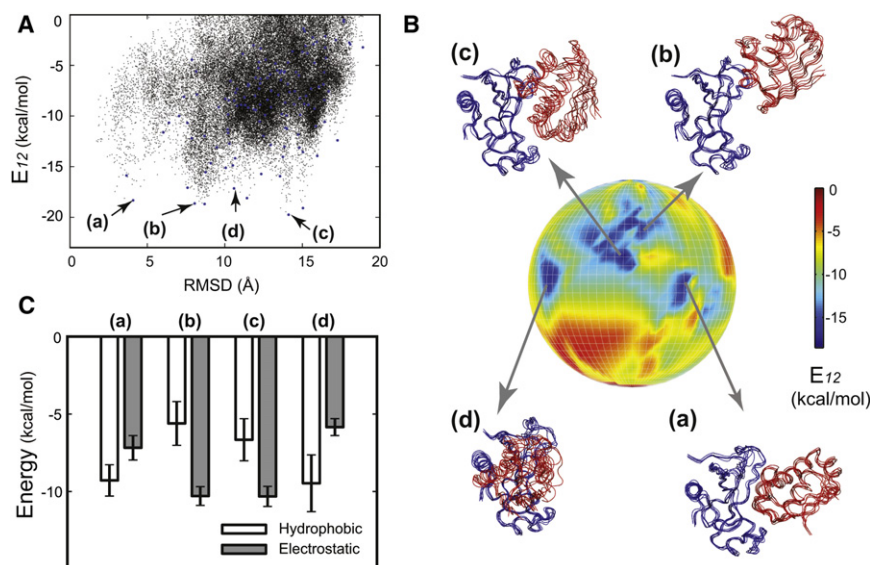


FIGURE 6 Energy landscape of the barnase/barstar complex. (A) A plot of E_{12} versus RMSD. A total of 43,554 configurations from CGMD simulations are shown in black dots, where $R_{max} = 50$ Å, $R_c = 25$ Å, and $R_{min} = 0$ Å were used in the PPR scheme. (B) Front view of the energy globe colored by E_{12} . We used a total of $N_f = 120$ clusters (also shown in blue dots in A) after the two-step clustering. Four representative conformations (*a–d*) are shown on the globe, where conformation (*a*) with the lowest E_{12} resembles the crystal structure (35). (C) The decomposition of hydrophobic (white bars) and electrostatic (gray bars) energies from E_{12} for each conformation (*a–d*).

The ability of a simple CG model to identify relevant conformations could be due to the relatively smooth binding free energy landscape for functional proteins (54–57). Of course, the inherent simplifications in such a simple model cannot characterize atomically detailed interactions (23,58–61); in that case, the CG-identified conformations can be relaxed and used as a starting point for atomistic simulations. In addition, the current CG model has captured the physical basis of protein-protein association, but it may fail to produce meaningful results on protein-ligand interactions where details can matter. We also note that folding and unfolding, which can be coupled with the protein association process (54–56,62), are not studied here. Furthermore, our focus is mainly on protein association into compact conformations that are energetically favorable; the formation of extended and entropically favorable conformations, or the kinetic process of association itself, is not fully examined here, but will be illustrated in future communications.

Finally, we wish to emphasize that this PPR-CGMD simulation pipeline can be readily applied to those protein complexes whose crystal structures are unknown, especially since considerable knowledge about individual protein subunits has been made available after decades of efforts (63). It is thus anticipated that this pipeline is positioned to serve as an alternative approach to study protein-protein interactions on a wide range of protein complexes.

SUPPORTING MATERIAL

Five figures, a table, and reference (65) are available at [http://www.biophysj.org/biophysj/supplemental/S0006-3495\(12\)00787-4](http://www.biophysj.org/biophysj/supplemental/S0006-3495(12)00787-4).

We thank Marc Parisien for critical reading of this manuscript.

This work was supported in part by the Cleveland Foundation, the American Cancer Society (ACS IRG-91-022-15), and the Department of Defense Breast Cancer Research Program (W81XWH-11-1033). Computational support was provided by the Case Western Reserve University High Performance Computing Cluster.

REFERENCES

- Elcock, A., D. Sept, and J. McCammon. 2001. Computer simulation of protein-protein interactions. *J. Phys. Chem. B.* 105:1504–1518.
- Perkins, J. R., I. Diboun, ..., C. Orengo. 2010. Transient protein-protein interactions: structural, functional, and network properties. *Structure.* 18:1233–1243.
- Koshland, Jr., D. E. 1958. Application of a theory of enzyme specificity to protein synthesis. *Proc. Natl. Acad. Sci. USA.* 44:98–104.
- Csermely, P., R. Palotai, and R. Nussinov. 2010. Induced fit, conformational selection and independent dynamic segments: an extended view of binding events. *Trends Biochem. Sci.* 35:539–546.
- Ritchie, D. W. 2008. Recent progress and future directions in protein-protein docking. *Curr. Protein Pept. Sci.* 9:1–15.
- Aloy, P., B. Böttcher, ..., R. B. Russell. 2004. Structure-based assembly of protein complexes in yeast. *Science.* 303:2026–2029.
- Parisien, M., K. F. Freed, and T. R. Sosnick. 2012. On docking, scoring and assessing protein-DNA complexes in a rigid-body framework. *PLoS ONE.* 7:e32647.
- Kim, Y. C., and G. Hummer. 2008. Coarse-grained models for simulations of multiprotein complexes: application to ubiquitin binding. *J. Mol. Biol.* 375:1416–1433.
- Solernou, A., and J. Fernandez-Recio. 2011. pyDockCG: new coarse-grained potential for protein-protein docking. *J. Phys. Chem. B.* 115:6032–6039.
- Karplus, M., and J. Kuriyan. 2005. Molecular dynamics and protein function. *Proc. Natl. Acad. Sci. USA.* 102:6679–6685.
- Hamelberg, D., T. Shen, and J. Andrew McCammon. 2005. Relating kinetic rates and local energetic roughness by accelerated molecular-dynamics simulations. *J. Chem. Phys.* 122:241103.
- Voter, A. F. 1997. Hyperdynamics: accelerated molecular dynamics of infrequent events. *Phys. Rev. Lett.* 78:3908–3911.
- Jensen, M. Ø., S. Park, ..., K. Schulten. 2002. Energetics of glycerol conduction through aquaglyceroporin GlpF. *Proc. Natl. Acad. Sci. USA.* 99:6731–6736.
- Grubmüller, H., B. Heymann, and P. Tavan. 1996. Ligand binding: molecular mechanics calculation of the streptavidin-biotin rupture force. *Science.* 271:997–999.
- Yang, S., H. Levine, and J. N. Onuchic. 2005. Protein oligomerization through domain swapping: role of inter-molecular interactions and protein concentration. *J. Mol. Biol.* 352:202–211.
- Isralewitz, B., M. Gao, and K. Schulten. 2001. Steered molecular dynamics and mechanical functions of proteins. *Curr. Opin. Struct. Biol.* 11:224–230.
- Nymeyer, H., A. E. García, and J. N. Onuchic. 1998. Folding funnels and frustration in off-lattice minimalist protein landscapes. *Proc. Natl. Acad. Sci. USA.* 95:5921–5928.
- Shea, J. E., J. N. Onuchic, and C. L. Brooks, 3rd. 1999. Exploring the origins of topological frustration: design of a minimally frustrated model of fragment B of protein A. *Proc. Natl. Acad. Sci. USA.* 96:12512–12517.
- Clementi, C., H. Nymeyer, and J. N. Onuchic. 2000. Topological and energetic factors: what determines the structural details of the transition state ensemble and “en-route” intermediates for protein folding? An investigation for small globular proteins. *J. Mol. Biol.* 298:937–953.
- Koga, N., and S. Takada. 2001. Roles of native topology and chain-length scaling in protein folding: a simulation study with a Go-like model. *J. Mol. Biol.* 313:171–180.
- Karanicolas, J., and C. L. Brooks, 3rd. 2002. The origins of asymmetry in the folding transition states of protein L and protein G. *Protein Sci.* 11:2351–2361.
- Elcock, A. H. 2006. Molecular simulations of cotranslational protein folding: fragment stabilities, folding cooperativity, and trapping in the ribosome. *PLoS Comput. Biol.* 2:e98.
- Takada, S. 2012. Coarse-grained molecular simulations of large biomolecules. *Curr. Opin. Struct. Biol.* 22:130–137.
- Onuchic, J. N., and P. G. Wolynes. 2004. Theory of protein folding. *Curr. Opin. Struct. Biol.* 14:70–75.
- Flores, S. C., J. Bernauer, ..., X. Huang. 2012. Multiscale modeling of macromolecular biosystems. *Brief Bioinform.* 13:395–405.
- Mills, B. M., and L. T. Chong. 2011. Molecular simulations of mutually exclusive folding in a two-domain protein switch. *Biophys. J.* 100:756–764.
- Graham, T. G. W., and R. B. Best. 2011. Force-induced change in protein unfolding mechanism: discrete or continuous switch? *J. Phys. Chem. B.* 115:1546–1561.
- Daily, M. D., G. N. Phillips, Jr., and Q. Cui. 2010. Many local motions cooperate to produce the adenylate kinase conformational transition. *J. Mol. Biol.* 400:618–631.
- Yang, S., J. N. Onuchic, and H. Levine. 2006. Effective stochastic dynamics on a protein folding energy landscape. *J. Chem. Phys.* 125:054910.
- Sobolev, V., A. Sorokine, ..., M. Edelman. 1999. Automated analysis of interatomic contacts in proteins. *Bioinformatics.* 15:327–332.

31. Miyazawa, S., and R. L. Jernigan. 1996. Residue-residue potentials with a favorable contact pair term and an unfavorable high packing density term, for simulation and threading. *J. Mol. Biol.* 256:623–644.
32. Brooks, B., R. Bruccoleri, ..., M. Karplus. 1983. CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* 4:187–217.
33. Yang, S., and B. Roux. 2008. Src kinase conformational activation: thermodynamics, pathways, and mechanisms. *PLoS Comput. Biol.* 4: e1000047.
34. Karpen, M. E., D. J. Tobias, and C. L. Brooks, 3rd. 1993. Statistical clustering techniques for the analysis of long molecular dynamics trajectories: analysis of 2.2-ns trajectories of YPGDV. *Biochemistry.* 32:412–420.
35. Buckle, A. M., G. Schreiber, and A. R. Fersht. 1994. Protein-protein recognition: crystal structural analysis of a barnase-barstar complex at 2.0-Å resolution. *Biochemistry.* 33:8878–8889.
36. Kühlmann, U. C., A. J. Pommer, ..., C. Kleanthous. 2000. Specificity in protein-protein interactions: the structural basis for dual recognition in endonuclease colicin-immunity protein complexes. *J. Mol. Biol.* 301: 1163–1178.
37. Stites, W. E. 1997. Protein-protein interactions: interface structure, binding thermodynamics, and mutational analysis. *Chem. Rev.* 97: 1233–1250.
38. Keeble, A. H., N. Kirkpatrick, ..., C. Kleanthous. 2006. Calorimetric dissection of colicin DNase—immunity protein complex specificity. *Biochemistry.* 45:3243–3254.
39. Simons, K. T., C. Kooperberg, ..., D. Baker. 1997. Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and Bayesian scoring functions. *J. Mol. Biol.* 268:209–225.
40. Sugita, Y., and Y. Okamoto. 1999. Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.* 314:141–151.
41. Sanbonmatsu, K. Y., and A. E. García. 2002. Structure of Met-enkephalin in explicit aqueous solution using replica exchange molecular dynamics. *Proteins.* 46:225–234.
42. Zhou, R., B. J. Berne, and R. Germain. 2001. The free energy landscape for beta hairpin folding in explicit water. *Proc. Natl. Acad. Sci. USA.* 98:14931–14936.
43. Okur, A., D. R. Roe, ..., C. Simmerling. 2007. Improving convergence of replica-exchange simulations through coupling to a high-temperature structure reservoir. *J. Chem. Theory Comput.* 3:557–568.
44. Weinan, E., W. Ren, and E. Vanden-Eijnden. 2002. String method for the study of rare events. *Phys. Rev. B.* 66:052301.
45. Maragliano, L., A. Fischer, ..., G. Ciccotti. 2006. String method in collective variables: minimum free energy paths and isocommittor surfaces. *J. Chem. Phys.* 125:24106.
46. Pelletier, H., and J. Kraut. 1992. Crystal structure of a complex between electron transfer partners, cytochrome *c* peroxidase and cytochrome *c*. *Science.* 258:1748–1755.
47. Leopold, P. E., M. Montal, and J. N. Onuchic. 1992. Protein folding funnels: kinetic pathways through compact conformational space. *Proc. Natl. Acad. Sci. USA.* 89:8721–8725.
48. Wolynes, P. G., J. N. Onuchic, and D. Thirumalai. 1995. Navigating the folding routes. *Science.* 267:1619–1620.
49. Dill, K. A., and H. S. Chan. 1997. From Levinthal to pathways to funnels. *Nat. Struct. Biol.* 4:10–19.
50. Sheinerman, F. B., R. Norel, and B. Honig. 2000. Electrostatic aspects of protein-protein interactions. *Curr. Opin. Struct. Biol.* 10:153–159.
51. Baker, N., and J. McCammon. 2005. *Electrostatic Interactions in Structural Bioinformatics*, Vol. 44. John Wiley & Sons, Hoboken, NJ.
52. Ko, T. P., C. C. Liao, ..., H. S. Yuan. 1999. The crystal structure of the DNase domain of colicin E7 in complex with its inhibitor Im7 protein. *Structure.* 7:91–102.
53. Schreiber, G., and A. R. Fersht. 1993. Interaction of barnase with its polypeptide inhibitor barstar studied by protein engineering. *Biochemistry.* 32:5145–5150.
54. Shoemaker, B. A., J. J. Portman, and P. G. Wolynes. 2000. Speeding molecular recognition by using the folding funnel: the fly-casting mechanism. *Proc. Natl. Acad. Sci. USA.* 97:8868–8873.
55. Andrec, M., A. K. Felts, ..., R. M. Levy. 2005. Protein folding pathways from replica exchange simulations and a kinetic network model. *Proc. Natl. Acad. Sci. USA.* 102:6801–6806.
56. Wang, J., W. Huang, ..., E. Wang. 2004. Downhill kinetics of biomolecular interface binding: globally connected scenario. *Biophys. J.* 87: 2187–2194.
57. Tunbridge, I., R. B. Best, ..., M. M. Kuttel. 2010. Simulation of coarse-grained protein-protein interactions with graphics processing units. *J. Chem. Theory Comput.* 6:3588–3600.
58. Riniker, S., J. R. Allison, and W. F. van Gunsteren. 2012. On developing coarse-grained models for biomolecular simulation: a review. *Phys. Chem. Chem. Phys.* <http://dx.doi.org/10.1039/C2CP40934H> In press.
59. Rader, A. J. 2010. Coarse-grained models: getting more with less. *Curr. Opin. Pharmacol.* 10:753–759.
60. Thomas, P. D., and K. A. Dill. 1996. Statistical potentials extracted from protein structures: how accurate are they? *J. Mol. Biol.* 257: 457–469.
61. Gopal, S. M., S. Mukherjee, ..., M. Feig. 2010. PRIMO/PRIMONA: a coarse-grained model for proteins and nucleic acids that preserves near-atomistic accuracy. *Proteins.* 78:1266–1281.
62. Okazaki, K.-i., and S. Takada. 2008. Dynamic energy landscape view of coupled binding and protein conformational change: induced-fit versus population-shift mechanisms. *Proc. Natl. Acad. Sci. USA.* 105:11182–11187.
63. Levitt, M. 2009. Nature of the protein universe. *Proc. Natl. Acad. Sci. USA.* 106:11079–11084.
64. Egea, P. F., A. Mitschler, and D. Moras. 2002. Molecular recognition of agonist ligands by RXRs. *Mol. Endocrinol.* 16:987–997.
65. Jelesarov, I., and H. R. Bosshard. 1999. Isothermal titration calorimetry and differential scanning calorimetry as complementary tools to investigate the energetics of biomolecular recognition. *J. Mol. Recognit.* 12:3–18.