THE JOURNAL OF PHYSICAL CHEMISTRY Letters

Letter

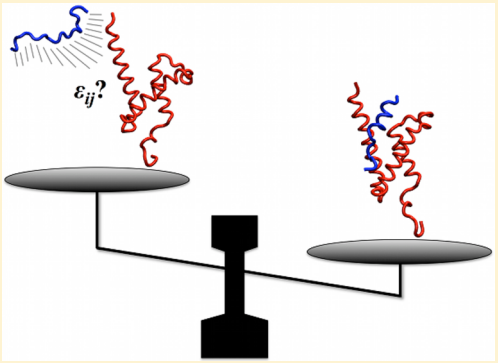
pubs.acs.org/JPCL

Hamiltonian Mapping Revisited: Calibrating Minimalist Models to Capture Molecular Recognition by Intrinsically Disordered Proteins

Sean M. Law,[†] Logan S. Ahlstrom,[†] Afra Panahi,[†] and Charles L. Brooks, III*[,†,‡]

[†]Department of Chemistry and [‡]Biophysics Program, The University of Michigan, Ann Arbor, Michigan 48109, United States

ABSTRACT: Molecular recognition by intrinsically disordered proteins (IDPs) plays a central role in many critical cellular processes. Toward achieving detailed mechanistic understanding of IDP−target interactions, here we employ the "Hamiltonian mapping" methodology, which is rooted in the weighted histogram analysis method (WHAM), for the fast and efficient calibration of structure-based models in studies of IDPs. By performing reference simulations on a given Hamiltonian, we illustrate for two model IDPs how this method can extrapolate thermodynamic behavior under a range of modified Hamiltonians, in this case representing changes in the binding affinity (K_d) of the system. Given sufficient conformational sampling in a single trajectory, Hamiltonian mapping accurately reproduces K_d values from direct simulation. This method may be generally applied to systems beyond IDPs in force field optimization and in describing changes in thermodynamic behavior as a function of external conditions for connection with experiment.



SECTION: Biophysical Chemistry and Biomolecules

Molecular recognition involving intrinsically disordered proteins (IDPs) is a key component in signaling and regulation networks. Moreover, IDPs are increasingly implicated in human diseases, such as cancer and neurodegeneration.[1] Delineating the mechanisms by which IDPs interact with their biological partners thus has the potential to expand our knowledge of protein interactions in the cell and to treat human diseases.

Biomolecular simulation methods provide a means to characterize the atomic-level details of IDP recognition. However, the long biological time scale of binding events is beyond the reach of conventional and most enhanced sampling simulation approaches. Recently, this issue has been addressed through the application of structure-based models to study IDP interactions.[2−10] Since IDPs commonly fold into a stable structure upon interaction with a target protein, in these studies, an experimental structure of an IDP−target protein complex serves as a reference for defining a native residue−residue contact map. The success of structure-based models in characterizing IDP recognition requires a delicate balance between the strength of the interactions governing both the native intermolecular binding and intramolecular folding contacts.[4] A reasonable balance of forces is typically achieved by tuning the strength of the contacts such that the binding affinity (K_d) and secondary structure content from simulation match those values from experiment.[4,8] While tuning the model is essential for capturing the proper mechanism of molecular recognition, it often necessitates running many additional simulations until the model is properly calibrated, which can be both computationally expensive and time-consuming.

In 1998, Brooks and co-workers[11] developed a framework for exploring the space of Hamiltonians in the "vicinity" of a given reference system in the context of investigating the folding free-energy landscape of a coarse-grained model. This framework builds upon the weighted histogram analysis method (WHAM) of Swendsen[12] and has been applied in a number of related contexts to examine the influence of denaturants[13−15] and pH,[16] as well as other factors that affect the folding landscape of proteins and peptides.[17] In the present work, we utilize for the first time, to our knowledge, this formalism to explore a timely question of tuning a coarse-grained force field to recapitulate binding affinities between two protein partners. The approach is general and can also be applied to higher-order complexes.[8] Given sufficient conformational sampling, this approach extrapolates changes in the thermodynamic behavior of a system under modified Hamiltonians from just a single reference trajectory. We illustrate the utility of "Hamiltonian mapping" by comparing theoretically computed K_d values to direct simulations and to experiment for two model IDPs (Figure 1). The approach is general and can be used to study molecular recognition in other IDP systems and to help advance other biophysical modeling efforts.

Following recent work,[4,8] we performed reference simulations of the KIX−*c*-Myb and MLL−KIX binary complexes[18] (Figure 1) using an intermolecular scaling factor, λ, equal to 1.1 and at three different time lengths (1.5, 15, and 450 μs). The intermolecular interactions from these reference simulations were then reweighted using the Hamiltonian mapping approach (see the Computational Methods section) by increasing λ. For

Received: August 27, 2014
Accepted: September 19, 2014
Published: September 19, 2014

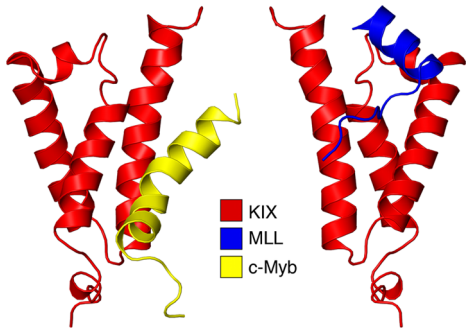


Figure 1. KIX−*c*-Myb (left) and MLL−KIX (right) binary complexes (PDB ID: 2AGH).

each reweighted ensemble (i.e., representing a given λ value), we computed the binding affinity (K_d) and compared this value with the corresponding K_d value from 450 μs long direct simulations with the same scaling factor. Overall, Figure 2 shows that as λ increases, binding becomes tighter (lower log K_d values). K_d values calculated from Hamiltonian mapping using a λ = 1.1 reference ensemble matched the K_d values calculated from the direct simulations (Figure 2A−F). We also examined the scenario in which we increased λ to 1.5 in our reference simulation and decreased λ for the mapping (Figures 2G−I). With the exception of the 1.5 μs long simulation (Figure 2G), which failed Hamiltonian mapping due to the simulation being too short to sample unbound states, we also obtained accurate results in this case. In general, given sufficient sampling, the Hamiltonian mapping method was able to accurately reweight K_d values spanning several orders of magnitude and provide a reliable estimate of λ to quickly and efficiently match simulated K_d values to experiment.

In the earliest work on the subject by Brooks and co-workers,[11] it was demonstrated that the thermodynamics of a large range of modified Hamiltonians could be extrapolated by simply reweighting one simulation using a revised expression of the WHAM equations. As noted previously and reiterated again here, it is important that the sampling in the reference simulation adequately covers the configuration space of interest. In other words, the sampling in the reference simulation must overlap significantly with the sampling of the modified Hamiltonian in order for the extrapolation method to be effective. This is generally not a problem with much longer coarse-grained simulations using modern computers. In all of the cases presented in the current work, the 15 μs long reference simulations, which took about 1 week to perform on a single CPU, provided a reliable estimate of K_d following reweighting. Moreover, it is important to point out that Hamiltonian mapping is not strictly limited to reweighting a single reference simulation as presented here. In fact, the sampling from multiple reference simulations with different λ scaling factors could be trivially combined and reweighted simultaneously using the modified WHAM equations. For example, we assume the scenario in which λ = 0.5 is the scaling factor that yields a K_d in agreement with the direct simulation and that two independent reference simulations produce ensembles that are either predominantly bound (λ = 1) or unbound (λ = 0). Then, instead of reweighting either reference simulation alone, which would produce a poor estimate of λ due to insufficient overlap in sampling, one could combine information from both reference simulations for the mapping to obtain an accurate estimate of the K_d.

To our knowledge, this is the first systematic study that demonstrates the true efficacy of the Hamiltonian mapping method through direct comparisons with reference simulations where the phase space was exhaustively sampled. We illustrate

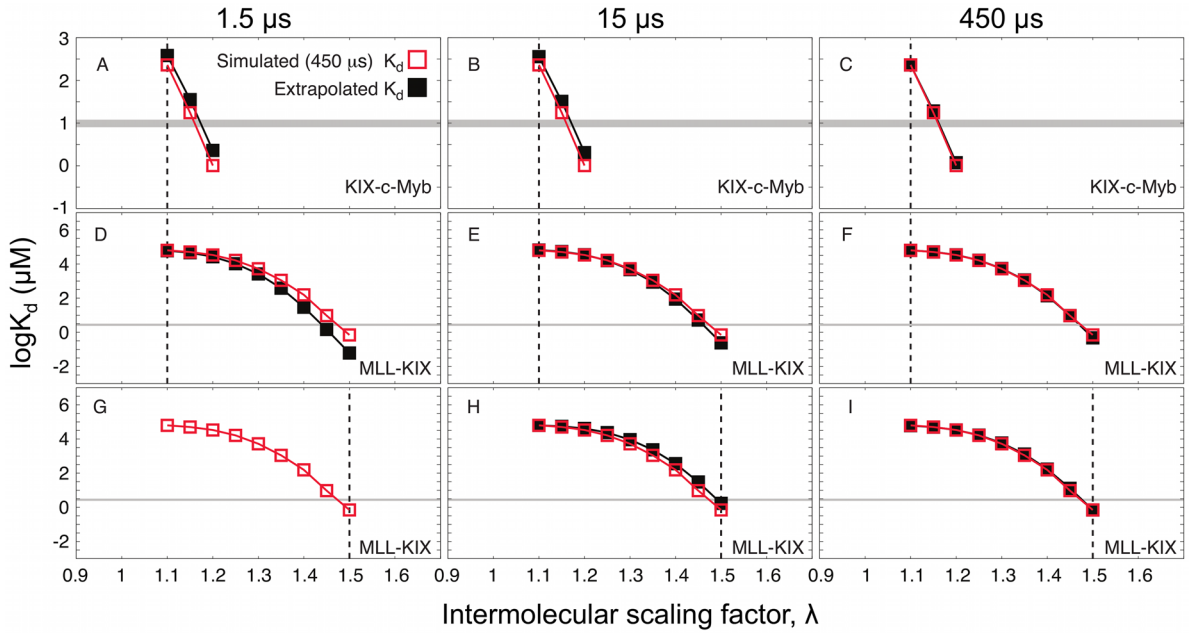


Figure 2. Binding affinities for (A−C) KIX−*c*-Myb and (D−I) MLL−KIX extrapolated from Hamiltonian mapping (black boxes) compared with reference simulations at each λ value. The columns are organized by the time length of the reference simulations, which are indicated above the first row. In each panel, the dotted vertical line denotes the reference simulation to which Hamiltonian mapping was applied, and the gray horizontal bars correspond to the experimental K_d value (10 ± 2 μM for KIX−*c*-Myb and 2.8 ± 0.4 μM for MLL−KIX).[19] Note that the simulated K_d values in (D−F) are identical to those in (G−I), but the reference simulations for Hamiltonian mapping are different.

dx.doi.org/10.1021/jz501811k | J. Phys. Chem. Lett. 2014, 5, 3441−3444

the utility of the method by examining the thermodynamics of binding in coarse-grained simulations of two model IDP systems. Given sufficient conformational sampling, K_d values extrapolated from Hamiltonian mapping are in excellent agreement with those achieved from direct simulation. Furthermore, we find that a moderate simulation time with the coarse-grained model and low computational cost are needed to obtain a reliable reference ensemble for Hamiltonian mapping, such that this approach could be practically applied to the rapid and efficient calibration of other IDP systems. We anticipate that the quantitative assessment of binding thermodynamics afforded by the method will play a valuable role in characterizing mechanisms of IDP recognition.

Beyond studies of IDPs, Hamiltonian mapping has the potential to facilitate other biophysical modeling efforts. For instance, the mapping procedure could significantly expedite the parameterization of force fields[4,8] (both coarse-grained and all-atom) or any study in which the Hamiltonian is being modified (e.g., Hamiltonian replica exchange simulations[20−23]). The approach also presents a practical tool for modeling changes in thermodynamics under varying external conditions (e.g., temperature, pH, denaturant concentration, etc.).[14−17] The unification of such efforts within the framework of Hamiltonian mapping in studies of protein interactions and folding would provide a robust link between simulation and experiment.

COMPUTATIONAL METHODS

Hamiltonian Mapping Formalism. Hamiltonian mapping[11] demonstrates how the standard WHAM formalism[12] can be modified and used to extrapolate information about a biased Hamiltonian from the unmodified reference Hamiltonian, $H_0(x)$, where x represents the molecular coordinates. WHAM is typically used along with umbrella sampling,[24] in which a set of L additional restraining potentials, $\{V\} = V_1(x), ..., V_L(x)$, are added to $H_0(x)$ in order to enhance conformational sampling in rarely visited regions of phase space. In this case, the modified Hamiltonian takes the form

$$H_{\{\lambda\}}(x) = H_0(x) + \sum_{i=1}^{L} \lambda_i V_i(x) = \sum_{i=0}^{L} \lambda_i V_i(x) \tag{1}$$

where $\{\lambda\}$ is the set of coupling parameters used for scaling individual biasing potentials, $\lambda_0 = 1$, and $H_0(x) = V_0(x)$. Then, for R independent simulations, each performed at a temperature $T = 1/k_B\beta$ (where k_B is the Boltzmann constant) and with differing $\{\lambda\}$, the bias is removed by solving the following pair of WHAM equations self-consistently[12]

$$P_{\{\lambda\},\beta}(\{V\}, \xi) = \frac{\sum_{k=1}^{R} N_k(\{V\}, \xi) \exp(-\beta \sum_{j=0}^{L} \lambda_j V_j)}{\sum_{m=1}^{R} n_m \exp(f_m - \beta_m \sum_{j=0}^{L} \lambda_{j,m} V_j)} \tag{2}$$

and

$$\exp(-f_j) = \sum_{\{V\},\xi} P_{\{\lambda\}_j,\beta_j}(\{V\}, \xi) \tag{3}$$

T may be constant or vary across the R simulations. (In the current work, T = 300 K for all simulations.) ξ is a particular value along a progress variable of interest, n_m is the total number of configurations in the mth simulation, and $N_k(\{V\},\xi)$ is the histogram count of configurations with $\{V\}$ and ξ. Finally, f_m is related to the Helmholtz free energy for the mth simulation.[11,12]

In contrast to standard WHAM, in the Hamiltonian mapping framework, we proceed in the opposite direction; we perform one or more simulations of the system under the original Hamiltonian $H_0(x)$ and then add a bias to the statistical averages computed from $H_0(x)$ to examine thermodynamic behavior under a modified Hamiltonian. For this scenario, eq 2 is recast as

$$P_{\{\lambda\},\beta}(\{V\}, \xi) = \frac{\sum_{k=1}^{R} N_k(H_0(x), \xi) \exp(-\beta \sum_{j=0}^{L} \lambda_j V_j)}{\sum_{m=1}^{R} n_m \exp(f_m - \beta_m H_0(x))} \tag{4}$$

Essentially, the histogram counts N_k now come from the sampling of $H_0(x)$, and the sum of the scaled restraining potentials in the denominator reduces to $H_0(x)$. Equations 4 and 3 are then iterated as above for standard WHAM.

Coarse-Grained Potential. To model binding and folding events of the KIX−*c*-Myb and MLL−KIX binary complexes, we employ the sequence-flavored Gō-like model of Karanicolas and Brooks.[25] Briefly, in this model, each residue is represented by a single bead centered at the Cα position and with the mass of the corresponding amino acid. An additive potential describes the bonded and non-bonded interactions between the beads. Virtual bonds and angles are defined by harmonic potentials with reference values determined by the Cα coordinates in the experimental structure. A potential for virtual dihedral angles is based upon backbone dihedral angle probability distributions from the Protein Data Bank (PDB) for the 400 possible amino acid pairs and thus is independent of the specific topology of the system. Residue pairs separated by three or more bonds interact through the following potential

$$V_{ij} = \varepsilon_{ij}\left[13\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - 18\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{10} + 4\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] \tag{5}$$

where r_{ij} is the distance between residues i and j, σ_{ij} corresponds to the distance between the two residues at which V_{ij} is minimum, and ε_{ij} is the interaction strength at σ_{ij}. Residues located in close proximity in the experimental structure (i.e., forming a "native contact") interact favorably with ε_{ij} values based on the statistical contact energies of Miyazawa and Jernigan,[26] while residue pairs not in close proximity experience a slight repulsive interaction that takes the form of a typical 12−6 Lennard-Jones potential. Further description of the Gō-like model and details of the simulation setup can be found in refs 25 and 8, respectively.

Tuning Intermolecular Binding Forces. We further consider the non-bonded interaction potential (eq 5) for native contacts between the IDP and the receptor in tuning the model to match the experimental K_d. For this set of intermolecular contacts, we scale the interaction strength, ε_{ij}, by a factor λ so as to control the proportion of bound and unbound states throughout simulation. From the fraction of unbound states, p_u, the K_d is computed as

$$K_d = [\text{protein}]\frac{p_u^2}{1 - p_u} \tag{6}$$

where [protein] is the concentration of protein in moles per liter (10^{-3} mol/L in the current study).[4,8] A configuration is considered unbound if zero intermolecular native contacts are

dx.doi.org/10.1021/jz501811k | J. Phys. Chem. Lett. 2014, 5, 3441−3444

formed and if the distance between the centers of mass of the IDP and of the target protein is greater than the cutoff for calculating the non-bonded interactions (25 Å). We also note that the strength of intramolecular contacts within the IDP are tuned in a similar manner to reproduce the experimental estimate of residual helical structure in the unbound state.[8]

AUTHOR INFORMATION

Corresponding Author

*E-mail: brookscl@umich.edu.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors would like to acknowledge support from the NIH (GM107233). L.S.A. acknowledges support from an NIH Ruth L. Kirschstein NRSA Postdoctoral Fellowship (GM108298). S.M.L. would like to acknowledge valuable scientific discussions with Drs. Michael Feig and Bin Zhang.

REFERENCES

(1) Uversky, V. N.; Oldfield, C. J.; Dunker, A. K. Intrinsically Disordered Proteins in Human Diseases: Introducing the D2 Concept. *Annu. Rev. Biophys.* **2008**, *37*, 215−246.

(2) Ahlstrom, L. S.; Dickson, A.; Brooks, C. L., III. Binding and Folding of the Small Bacterial Chaperone HdeA. *J. Phys. Chem. B* **2013**, *117*, 13219−13225.

(3) De Sancho, D.; Best, R. B. Modulation of an IDP Binding Mechanism and Rates by Helix Propensity and Non-Native Interactions: Association of Hif1α with CBP. *Mol. Biosyst.* **2012**, *8*, 256−267.

(4) Ganguly, D.; Chen, J. Topology-Based Modeling of Intrinsically Disordered Proteins: Balancing Intrinsic Folding and Intermolecular Interactions. *Proteins* **2011**, *79*, 1251−1266.

(5) Ganguly, D.; Otieno, S.; Waddell, B.; Iconaru, L.; Kriwacki, R. W.; Chen, J. Electrostatically Accelerated Coupled Binding and Folding of Intrinsically Disordered Proteins. *J. Mol. Biol.* **2012**, *422*, 674−684.

(6) Ganguly, D.; Zhang, W.; Chen, J. Synergistic Folding of Two Intrinsically Disordered Proteins: Searching for Conformational Selection. *Mol. Biosyst.* **2012**, *8*, 198−209.

(7) Ganguly, D.; Zhang, W.; Chen, J. Electrostatically Accelerated Encounter and Folding for Facile Recognition of Intrinsically Disordered Proteins. *PLoS Comput. Biol.* **2013**, *9*, e1003363.

(8) Law, S. M.; Gagnon, J. K.; Mapp, A. K.; Brooks, C. L., III. Prepaying the Entropic Cost for Allosteric Regulation in KIX. *Proc. Natl. Acad. Sci. U.S.A.* **2014**, *111*, 12067−12072.

(9) Lu, Q.; Lu, H. P.; Wang, J. Exploring the Mechanism of Flexible Biomolecular Recognition with Single Molecule Dynamics. *Phys. Rev. Lett.* **2007**, *98*, 128105.

(10) Turjanski, A. G.; Gutkind, J. S.; Best, R. B.; Hummer, G. Binding-Induced Folding of a Natively Unstructured Transcription Factor. *PLoS Comput. Biol.* **2008**, *4*, e1000060.

(11) Shea, J. E.; Nochomovitz, Y. D.; Guo, Z. Y.; Brooks, C. L., III. Exploring the Space of Protein Folding Hamiltonians: The Balance of Forces in a Minimalist β-Barrel Model. *J. Chem. Phys.* **1998**, *109*, 2895−2903.

(12) Kumar, S.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A.; Rosenberg, J. M. The Weighted Histogram Analysis Method for Free-Energy Calculations on Biomolecules. I. The Method. *J. Comput. Chem.* **1992**, *13*, 1011−1021.

(13) Liu, Z. X.; Reddy, G.; O'Brien, E. P.; Thirumalai, D. Collapse Kinetics and Chevron Plots from Simulations of Denaturant-Dependent Folding of Globular Proteins. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 7787−7792.

(14) O'Brien, E. P.; Brooks, B. R.; Thirumalai, D. Molecular Origin of Constant *m*-Values, Denatured State Collapse, and Residue-Dependent Transition Midpoints in Globular Proteins. *Biochemistry* **2009**, *48*, 3743−3754.

(15) O'Brien, E. P.; Ziv, G.; Haran, G.; Brooks, B. R.; Thirumalai, D. Effects of Denaturants and Osmolytes on Proteins Are Accurately Predicted by the Molecular Transfer Model. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 13403−13408.

(16) O'Brien, E. P.; Brooks, B. R.; Thirumalai, D. Effects of pH on Proteins: Predictions for Ensemble and Single-Molecule Pulling Experiments. *J. Am. Chem. Soc.* **2012**, *134*, 979−987.

(17) Thirumalai, D.; Liu, Z. X.; O'Brien, E. P.; Reddy, G. Protein Folding: From Theory to Practice. *Curr. Opin. Struct. Biol.* **2013**, *23*, 22−29.

(18) De Guzman, R. N.; Goto, N. K.; Dyson, H. J.; Wright, P. E. Structural Basis for Cooperative Transcription Factor Binding to the CBP Coactivator. *J. Mol. Biol.* **2006**, *355*, 1005−1013.

(19) Goto, N. K.; Zor, T.; Martinez-Yamout, M.; Dyson, H. J.; Wright, P. E. Cooperativity in Transcription Factor Binding to the Coactivator Creb-Binding Protein (CBP). The Mixed Lineage Leukemia Protein (Mll) Activation Domain Binds to an Allosteric Site on the KIX Domain. *J. Biol. Chem.* **2002**, *277*, 43168−43174.

(20) Bergonzo, C.; Henriksen, N. M.; Roe, D. R.; Swails, J. M.; Roitberg, A. E.; Cheatham, T. E. Multidimensional Replica Exchange Molecular Dynamics Yields a Converged Ensemble of an RNA Tetranucleotide. *J. Chem. Theory Comput.* **2014**, *10*, 492−499.

(21) Law, S. M.; Feig, M. Base-Flipping Mechanism in Postmismatch Recognition by MutS. *Biophys. J.* **2011**, *101*, 2223−2231.

(22) Roe, D. R.; Bergonzo, C.; Cheatham, T. E. Evaluation of Enhanced Sampling Provided by Accelerated Molecular Dynamics with Hamiltonian Replica Exchange Methods. *J. Phys. Chem. B* **2014**, *118*, 3543−3552.

(23) Su, L.; Cukier, R. I. Hamiltonian Replica Exchange Method Study of *Escherichia coli* and *Yersinia pestis* HPPK. *J. Phys. Chem. B* **2009**, *113*, 16197−16208.

(24) Torrie, G. M.; Valleau, J. P. Monte-Carlo Free-Energy Estimates Using Non-Boltzmann Sampling — Application to Subcritical Lennard-Jones Fluid. *Chem. Phys. Lett.* **1974**, *28*, 578−581.

(25) Karanicolas, J.; Brooks, C. L., III. The Origins of Asymmetry in the Folding Transition States of Protein L and Protein G. *Protein Sci.* **2002**, *11*, 2351−2361.

(26) Miyazawa, S.; Jernigan, R. L. Residue-Residue Potentials with a Favorable Contact Pair Term and an Unfavorable High Packing Density Term, for Simulation and Threading. *J. Mol. Biol.* **1996**, *256*, 623−644.