THE JOURNAL OF PHYSICAL CHEMISTRY B

Article

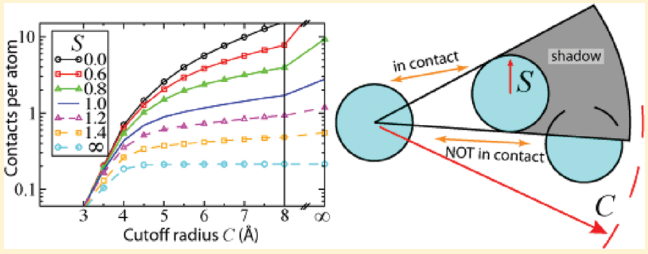
pubs.acs.org/JPCB

The Shadow Map: A General Contact Definition for Capturing the Dynamics of Biomolecular Folding and Function

Jeffrey K. Noel, Paul C. Whitford, and José N. Onuchic*

Center for Theoretical Biological Physics and Department of Physics, Rice University, Houston, Texas 77005, United States

ABSTRACT: Structure-based models (SBMs) are simplified models of the biomolecular dynamics that arise from funneled energy landscapes. We recently introduced an all-atom SBM that explicitly represents the atomic geometry of a biomolecule. While this initial study showed the robustness of the all-atom SBM Hamiltonian to changes in many of the energetic parameters, an important aspect, which has not been explored previously, is the definition of native interactions. In this study, we propose a general definition for generating atomically grained contact maps called "Shadow". The Shadow algorithm initially considers all atoms within a cutoff distance and then, controlled by a screening parameter, discards the occluded contacts. We show that this choice of contact map is not only well behaved for protein folding, since it produces consistently cooperative folding behavior in SBMs but also desirable for exploring the dynamics of macromolecular assemblies since, it distributes energy similarly between RNAs and proteins despite their disparate internal packing. All-atom structure-based models employing Shadow contact maps provide a general framework for exploring the geometrical features of biomolecules, especially the connections between folding and function.



1. INTRODUCTION

Structural biology techniques, such as nuclear magnetic resonance (NMR), X-ray crystallography, and cryogenic electron microscopy (cryo-EM), have provided extraordinary insights into the details of the functional configurations of biomolecular systems. Recent experimental advances have enabled the structural characterization of large and diverse molecular assemblies that are composed of heterogeneous parts: DNA, RNA, proteins, and small molecules. Some notable examples include the ribosome,[1] proteasome,[2] and spliceosome.[3] Molecular simulations allow one to connect these static pictures with dynamical experimental data such as single molecule Förster resonance energy transfer (FRET).[4] To bridge static and dynamic structural data, it is essential that we establish robust theoretical models that are able to accurately describe the dynamics of complex biomolecules.

Long-range communication between spatially distant residues in these assemblies are, to first approximation, controlled by the geometry of the molecular complexes. The ability of structural biologists to capture detailed structural information means that these folded structures are in low free-energy configurations. Energy landscape theory[5−7] and the *principle of minimal frustration*[8] explain that robust folding and assembly implies that these low free energy native structures are composed of consistent or "minimally frustrated" native interactions. This organization leads to a funnel-shaped energy landscape, where the overall energetic drive to the native structure is much larger than competing traps stemming from non-native interactions. While originally developed in the context of protein folding, energy landscape theory has been extended to account for biomolecular oligomerization[9] and functional transitions.[10] The class of theoretical models that probe the dynamics that emerge from a molecular geometry is called a structure-based model (SBM).[11,12] SBMs represent the dynamics of minimally frustrated systems through the approximation that all native interactions are stabilizing and include non-native interactions only to maintain proper excluded volume.

We recently introduced an all-atom SBM, which explicitly represents the atomic geometry of a biomolecule.[12] This SBM is a baseline model that can be used to fully discern the role of biomolecular geometry. Additionally, by introducing additional energetic complexity, one may also uncover the extent to which detailed energetics contribute to biomolecular structure, folding, and function. While our initial study showed the robustness of the all-atom SBM Hamiltonian to changes in many of the energetic parameters,[12] an important aspect, which has not been explored previously, is the definition of native interactions. Each native interaction, or "native contact", is formed by an atom−atom pair (or residue−residue pair in a C_α representation) interaction that is proximate in the native state. The set of native contacts is called a *contact map* and is a ubiquitous tool in the analysis of internal biomolecular interaction networks.[13−15]

Special Issue: Macromolecular Systems Understood through Multi-scale and Enhanced Sampling Techniques

Received: January 26, 2012
Revised: April 23, 2012
Published: April 26, 2012

   dx.doi.org/10.1021/jp300852d | *J. Phys. Chem. B* 2012, 116, 8692−8702

The definitions of contact maps in the literature are nearly as diverse as their applications. The simplest algorithms define contacts between atom (or residue) pairs that are within a cutoff radius of each other.[16] More complicated algorithms additionally consider, for example, solvent accessibility[17,18] or atomic chemistry.[19] For protein folding studies, contacts have often been defined through the atomic geometry, by choosing residue pairs that have heavy atoms within a cutoff distance (4.0−6.5 Å)[20−23] or atom pairs that shield each other's solvent accessibility.[11,24] In a SBM, the native contact map is an integral part of the Hamiltonian, since it defines the distribution of stabilizing energy in the biomolecule. Therefore, as SBMs are being explored on multiple levels of detail and are being applied to increasingly diverse and heterogeneous systems, a consistent method for choosing contact maps is desirable.

In this study, we propose a general definition for generating atomically grained contact maps called "Shadow" (Figure 1). It

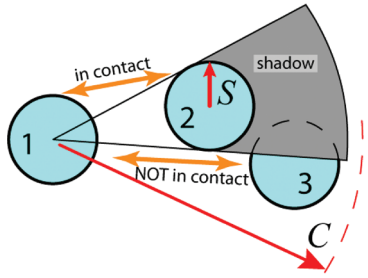


Figure 1. The Shadow contact map screening geometry. Only atoms within the cutoff distance C are considered. Atoms 1 and 2 are in contact because they are within C and have no intervening atom. To check if atoms 1 and 3 are in contact, one checks if atom 2 shadows atom 1 from atom 3. The three atoms are viewed in the plane, and all atoms are given the same shadowing radius S. Since a light shining from the center of atom 1 causes a shadow to be cast on atom 3, atoms 1 and 3 are not in contact.

is motivated by the need to satisfy two mutually incompatible features of a simple heavy atom cutoff contact map: to include relevant contacts at distances of at least 6 Å without introducing nonphysical next-nearest neighbor contacts. Long cutoffs enable the map to capture atomic contacts across structural waters or heavy metals that are not explicitly represented. At long cutoff distances, though, contacts will be introduced between atom pairs that we do not wish to model, specifically, those that have an intervening atom. The Shadow algorithm initially considers all atoms within a cutoff distance C and then, controlled by a screening parameter S, discards the occluded contacts. We compare two classes of contact maps: (1) maps based on a simple cutoff distance C and $S = 0$ and (2) maps with $S > 0$. They are compared dynamically by measuring the folding thermodynamics of well studied two-state proteins, the thermodynamics of an RNA hairpin, and the native basin fluctuations of the ribosome. We find that the Shadow contact map gives a consistent definition of atomically grained native interactions from small proteins up to macromolecular assemblies. Two-state proteins and RNA hairpins show reliably cooperative folding transitions. Also, Shadow contact maps distribute energy similarly between RNAs and proteins despite their disparate internal packing. All-atom structure-based models employing Shadow contact maps are a general framework for exploring the geometrical features of biomolecules, especially the connections between folding and function.

2. METHODS

2.1. The All-Atom Structure-Based Model. The all-atom (AA) SBM[12,25,26] for proteins discussed in this study has recently been used to study proteins, nucleic acids, and ligands for both dynamics[4,27,28] and molecular modeling.[29−31] All heavy (non-hydrogen) atoms are included, and each atom is represented as a single bead of unit mass. Bond lengths, bond angles, improper dihedrals, and planar dihedrals are maintained by harmonic potentials. Dihedral energy functions are defined such that each angle is a minimum in the native configuration. Nonbonded atom pairs that are in contact in the native state between residues i and j, where $i > j + 3$, are given an attractive potential, while all other nonlocal interactions are repulsive, which ensures the atoms have a defined excluded volume. The functional form of the potential is

$$V = \sum_{\text{bonds}} \varepsilon_b (r - r_0)^2 + \sum_{\text{angles}} \varepsilon_\theta (\theta - \theta_0)^2 + \sum_{\text{impropers/planar}} \varepsilon_\chi (\chi - \chi_0)^2 + \sum_{\text{backbone}} \varepsilon_{BB} F_D(\phi - \phi_0) + \sum_{\text{side chains}} \varepsilon_{SC} F_D(\phi - \phi_0) + \sum_{\text{contact map}} \varepsilon_C^{ij} C(r_{ij}, r_0^{ij}) + \sum_{\text{non-contacts}} \varepsilon_{NC} \left(\frac{r_{NC}}{r_{ij}}\right)^{12} \quad (1)$$

where F_D is a traditional dihedral potential $F_D(\phi - \phi_0) = [1 - \cos(\phi - \phi_0)] + {}^1/_2[1 - \cos(3(\phi - \phi_0))]$. C_{ij} is a contact potential, an effective short-range interaction between atoms i and j that are in contact in the native state (see section 2.4). The definition of the native contacts is considered in detail in section 2.3. Consistent with previous studies,[12,25] three criteria define the values of ε_{BB}, ε_{SC}, and ε_C for a given molecular complex: (1) ε_{BB} and ε_{SC} are scaled so that $\varepsilon_{BB}/\varepsilon_{SC} = R_{BB/SC}$. (2) The energetic weight of each dihedral and contact is also scaled, such that the ratio of total contact energy to total dihedral energy, $\sum\varepsilon_C/(\sum\varepsilon_{BB} + \sum\varepsilon_{SC}) = R_{C/D}$, is satisfied. (3) The total stabilizing energy is set, such that $\sum\varepsilon_C + \sum\varepsilon_{BB} + \sum\varepsilon_{SC} = \varepsilon N_{atoms}$, where ε is the reduced energy unit. Here, $R_{BB/SC} = 2$ for protein and $R_{BB/SC} = 1$ for RNA and $R_{C/D} = 2$. This means that, as the contact map is varied, even though the number of contacts may vary, the net energy contribution from the contacts is constant at ${}^2/_3\varepsilon N_{atoms}$. The energy per contact though will vary. This allows for careful comparison between the different native contact maps. r_0^{ij} is the native distance separation between atoms i and j. $\varepsilon_b = 100\varepsilon$, $\varepsilon_\theta = 20\varepsilon$, $\varepsilon_\chi = 10\varepsilon$, and $\varepsilon_{NC} = \varepsilon$. r_0, θ_0, χ_0, ϕ_0, and r_0^{ij} are given the values found in the native state, and $r_{NC} = 1.7$ Å. We note that previous implementations of SBMs of RNA[25] reduced the strength of contacts between stacked bases by a factor of 3 when using cutoff maps (also see section 3.4).

2.2. Simulation Details. AA structure-based simulations were performed using the Gromacs software package.[32] Protein simulations were typically performed on four cores, and the ribosome simulations were performed on 128, or 256, cores each. The Gromacs source code was modified to include the Gaussian interaction (available at http://smog.ucsd.edu); no further modifications were necessary. The Gromacs topology files were generated with the smog@ctbp webserver.[33] Reduced

dx.doi.org/10.1021/jp300852d | J. Phys. Chem. B 2012, 116, 8692−8702

units were used. The time step was 0.0005. Temperature was controlled through stochastic dynamics with a coupling constant of 2. For all systems simulated in this paper, several constant temperature trajectories were obtained. In the case of folding, temperatures varied from the protein being always folded to always unfolded, and trajectories contained many folding transitions (>20). The weighted histogram analysis method[34,35] was used to combine data from multiple temperatures into single free energy profiles. Each ribosome simulation was performed for 2×10^7 time steps, with the second half used for data analysis. Fluctuations in proteins are calculated from 2×10^7 time steps of data. Convergence of native-state fluctuations was reached by 10^7 time steps, since doubling the data gives no discernible difference in the averages.

2.3. Contact Maps. Atoms that are spatially near in the native state are considered *contacts*, and together, the set of all contacts composes a *native contact map*. A contact map encodes which atom pairs *ij* are given attractive interactions in the SBM potential. In the context of a SBM, the native contact map sets the distribution of renormalized stabilizing enthalpy in the native state.

Here, we propose an algorithm for determining atomic contacts, called Shadow. It uses a heavy-atom cutoff distance together with geometric occlusion. There are two parameters in the algorithm, the cutoff distance C and the screening radius S (Figure 1). The algorithm can be metaphorically described: if a light source were located at the center of atom i, and all other atoms were opaque, then all atoms within the cutoff C that have no shadow cast upon them would be considered contacts. To keep the bonded atoms from overlapping, S is maintained at ≤0.5 Å when screening a bonded neighbor. To put shadowing in the context of other approaches, we compare it to the commonly used simple heavy-atom cutoff distance ($S = 0$). We denote a contact map with cutoff distance C and screening radius S as M_C^S. C and S are given in units of Å. Related geometric occlusion methods have been employed by Wu et al.[36] and by Veloso et al.[19]

2.3.1. The Shadow Map. The parameter set recommended for general use is termed "the Shadow map", and refers to M_6^1. Both the Shadow map and contact maps with variable S and C can be generated using the smog@ctbp webserver (http://smog.ucsd.edu).[33]

2.4. Contact Potential $C(r_{ij}, r_0^{ij})$. All of the pair interactions defined in the native contact map interact through an attractive potential, denoted in the SBM potential by $C(r_{ij}, r_0^{ij})$. The contact potential has a minimum at the distance between the pair in the native state r_0^{ij}. Traditionally, a contact is defined through a Lennard-Jones (LJ) type potential

$$C_{\mathrm{LJ}}(r_{ij}, r_0^{ij}) = \left(\frac{r_0^{ij}}{r_{ij}}\right)^{12} - 2\left(\frac{r_0^{ij}}{r_{ij}}\right)^{6} \tag{2}$$

The LJ potentials are well tested and perform well for many systems, but they introduce an excluded volume that scales with the contact distance (Figure 2). Since the effective volume of two atoms in contact grows with r_0^{ij}, this can lead to complications for certain applications. The variable repulsion introduces heterogeneity into coarse-grained beads, which allows the model to capture effective excluded volume effects. However, it is less clear that this feature is appropriate for all-atom SBMs, since the excluded volume should already be explicitly captured by the all-atom geometry.

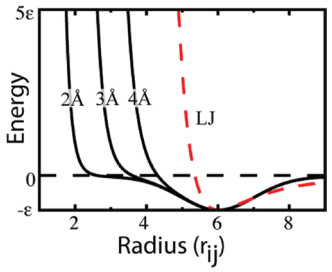


Figure 2. The versatile Gaussian contact potential. The repulsive part is constrained to shift with the position of the minimum in the Lennard-Jones potential, which introduces extraneous excluded volume for each native contact. In contrast, the excluded volume can be independently set relative to the contact minimum with the Gaussian contact potential. This decouples the energetics of the contact map from the protein geometry.

Here, we employ a contact potential that allows independent control of the excluded volume. By decoupling the protein geometry from the energetics, the contact map definition is independent of the excluded volume. Without this feature, the consequences of varying the contact map will be obscured by the entropic effects of the varying excluded volume. As used elsewhere,[27,37] we included contact interactions through the use of an attractive Gaussian well coupled to a fixed LJ repulsion

$$C_{\mathrm{G}}(r_{ij}, r_0^{ij}) = \left(1 + \left(\frac{r_{\mathrm{ex}}}{r_{ij}}\right)^{12}\right)(1 + G(r_{ij}, r_0^{ij})) - 1 \tag{3}$$

where

$$G(r_{ij}, r_0^{ij}) = -\exp[-(r_{ij} - r_0^{ij})^2/(2\sigma_{ij}^2)] \tag{4}$$

This functional form ensures that the depth of the minimum is −1 (scaled by ε_C in eq 1), and r_{ex} sets the excluded volume. r_{ex} has the same function as r_{NC} in eq 1. If $r_{\mathrm{ex}} = r_{\mathrm{NC}}$, all atomic interactions have an equal excluded volume. For consistency with the LJ potentials, the width of the Gaussian well σ_{ij} models the variable width of the LJ potential. $C_{\mathrm{LJ}}(1.2r_0^{ij}, r_0^{ij}) \sim -1/2$, so σ_{ij} is defined such that $G(1.2r_0^{ij}, r_0^{ij}) \sim -1/2$, giving $\sigma_{ij}^2 = (r_0^{ij})^2/(50 \ln 2)$. If r_{ex} is significantly smaller than r_0^{ij}, eq 3 reduces to a more transparent form

$$C_{\mathrm{G}}(r_{ij}, r_0^{ij}) \rightarrow \left(\frac{r_{\mathrm{ex}}}{r_{ij}}\right)^{12} + G(r_{ij}, r_0^{ij}) \quad \text{for} \quad r_{\mathrm{ex}}, \sigma_{ij} \ll r_0^{ij} \tag{5}$$

2.5. Thermodynamics. Folding experiments on small globular proteins have long shown evidence of thermodynamic and kinetic cooperativity, which indicates a phenomenon similar to a first-order phase transition between native and denatured states.[38,39] To quantify the thermodynamics and cooperativity of the SBM, the heat capacity was calculated. Two different dimensionless measures of cooperativity are considered: the width of the peak in the heat capacity κ_1 and the van't Hoff criterion κ_2. Both are applicable for describing the cooperativity of two-state transitions.[24,40,41]

$$\kappa_1 = \frac{\sigma_{1/2}}{T_{\max}} \tag{6}$$

where $\sigma_{1/2}$ is the full width at half-maximum of the heat capacity $C_V \equiv \partial\langle H\rangle/\partial T$ and $T_{\max}$ is the temperature corresponding to the peak in C_V (Figure 3). κ_1 is interpreted

dx.doi.org/10.1021/jp300852d | J. Phys. Chem. B 2012, 116, 8692−8702

as a measure of the temperature range over which the transition occurs, where smaller κ_1 indicates a higher degree of cooperativity.

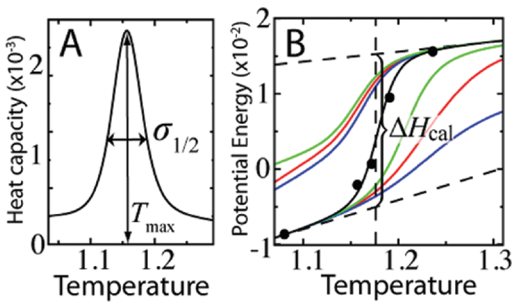


Figure 3. Measures of cooperativity. (A) $\kappa_1 = \sigma_{1/2}/T_{max}$. $\sigma_{1/2}$ is the width of the heat capacity at half the maximum. (B) κ_2 is a measure of the enthalpy change associated with the transition relative to the total enthalpy change ΔH_{cal}. The behavior of the enthalpy in the folded and unfolded states is modeled linearly (dotted lines). The vertical dotted line marks T_F. Weighted histogram analysis gives the continuous lines. Black dots show $\langle H \rangle$ during constant temperature molecular dynamics. The blue, red, and green lines show the folded and unfolded state enthalpy for different rmsd cutoffs, d_c, of 5, 6, and 7 Å, respectively.

The van't Hoff criterion κ_2 is a measure of cooperativity that is based on the enthalpy distribution during the transition. A cooperative transition has a well-defined energy separation between unfolded U and folded F ensembles. With $K_{eq} = [F]/[U]$ as the equilibrium constant of the folding reaction, the van't Hoff criterion is defined at the midpoint of the transition, given by $K_{eq} = {}^1/_2$.

$$\kappa_2 = \frac{-k_B T^2}{\Delta H_{cal}} \frac{\partial \ln K_{eq}(T)}{\partial T}\bigg|_{K_{eq}=1/2} = \frac{\langle H \rangle_U - \langle H \rangle_F}{\Delta H_{cal}}\bigg|_{K_{eq}=1/2} \quad (7)$$

where ΔH_{cal} is the calorimetric enthalpy change of the transition and $\langle H \rangle_X$ is the average enthalpy of ensemble X. ΔH_{cal} is the integral of C_V over the transition region and is determined by extrapolating the unfolded state enthalpy and the folded state enthalpy to $T_{1/2}$, the temperature where $K_{eq} = {}^1/_2$ (Figure 3). These extrapolations, known as *baselines*, approximate the temperature dependence of the enthalpy in the absence of the protein transition.[41] The baselines, H_U and H_F, isolate the heat change of the transition, $\Delta H_{cal} = H_U(T_{1/2}) - H_F(T_{1/2})$. Determination of $T_{1/2}$ requires a definition of the unfolded and folded ensembles. In this investigation, a cutoff in root-mean-square deviation from the native state (rmsd) d_c is used to partition configurations.[42,43] The "proper" d_c may be determined simply by maximizing κ_2, i.e., $\partial\kappa_2/\partial d_c = 0$. Note that simplifying the calculation by fixing $\langle H \rangle_F = H_F$ will overestimate κ_2, since $\langle H \rangle_F > H_F$.

3. RESULTS AND DISCUSSION

Since SBMs are applied to diverse biomolecular systems, the present study encompasses a broad range of biomolecular systems, in particular, globular proteins, RNA, and the ribosome. First, we discuss the effects of geometric occlusion on the number and distribution of native contacts in globular proteins and in RNA secondary structure. Then, we analyze the sensitivity of both folding thermodynamics and native state fluctuations to the choice of native contacts in model protein and RNA systems. Lastly, we examine the sensitivity of

Table 1. Statistics on Various Contact Maps of Model Globular Proteins

protein	M_4^0	$M_4^{0.7}$	M_4^1	M_5^1	M_6^1	M_5^0	M_6^0
				total contacts			
UBQ[a]	387	322	262	625	874	1504	3510
CI2[b]	280	229	188	466	597	1117	2566
SH3[c]	325	266	185	409	532	1397	3155
BDPA[d]	179	153	115	222	329	575	1345
				contacts per atom			
CI2	0.56	0.45	0.37	0.92	1.2	2.2	5.1
SH3	0.71	0.58	0.41	0.89	1.2	3.1	6.9
BDPA	0.49	0.41	0.31	0.60	0.90	1.6	3.6
NHGP[f]	0.70	0.60	0.43	0.89	1.2	2.6	5.7
				dispersion in contacts between residues[g]			
NHGP	2.50	2.49	2.05	1.83	1.63	2.06	1.67
				C_V criterion, κ_1			
CI2	3-state	0.038	0.031	0.024	0.032	0.063	0.12
SH3	0.027	0.025	0.021	0.025	0.028	0.033	0.047
BDD	0.14	0.087	0.050	0.046	0.046	0.072	0.10
				van't Hoff criterion, κ_2			
CI2	3-state	0.66	0.73	0.93	0.84	0.70	0.72
SH3	0.82	0.91	0.93	0.96	0.94	0.90	0.89
BDD	0.77	0.71	0.76	0.79	0.84	0.70	0.73

[a]PDB code 1UBQ. [b]PDB code 1YPA. [c]PDB code 1FMK, residues 84−140. [d]PDB code 1BDD. [e]Contacts per residue with map M_6^1. [f]Average over a set of 33 nonhomologous globular proteins.[47] [g]Atom−atom contacts per residue−residue contact; dispersion is normalized by average.

dx.doi.org/10.1021/jp300852d | J. Phys. Chem. B 2012, 116, 8692−8702

fluctuations to the choice of contact map in a large molecular assembly: the ribosome.

3.1. Protein Contact Maps. Protein native structures, as determined by structural biology techniques, are compact and densely packed structures stabilized by both short- and long-range interactions.[44,45] The all-atom SBM encodes the stability imparted by these interactions with short-range attractive potentials between pairs of atoms. These interactions drive the protein toward the low free energy native configuration. The short-range atomic interactions in proteins are on the Å length scale. The closest pairs are the hydrogen bonding interactions between the carboxyl O and amino N found throughout α-helices and β-sheets. The N−O are commonly separated by 2.6−3.0 Å. In the hydrophobic core, carbon pairs are separated by 3.5−4.5 Å. This longer distance is a consequence of the larger van der Waals radius of carbon compared to oxygen and nitrogen. Salt-bridges exist in protein cores with separations up to 5.5 Å.[46] Indirect pair interactions mediated through water molecules, either surface or buried, can vary between 5 and 7 Å[45] and are a source of enthalpic stabilization.[47]

An algorithm to generate protein contact maps that includes all of the above-mentioned short-range interactions must accommodate pair separations up to at least 6 Å, or more. While a simple cutoff distance criterion will capture all of the essential interactions, there will also be many occluded pair interactions that we are not seeking to model (Figure 1). The occluded interactions represent three-body interactions, and their effects should be considered higher-order corrections. These higher-order, occluded contacts can be identified, and discarded, by using the shadowing geometric criteria described in section 2.3. The parameter choices of C = 6 Å and S = 1 Å, or M_6^1, define the contact map henceforth called "the Shadow map".

3.1.1. Removal of Contacts through Geometric Occlusion. The abundance of occluded contacts is checked by constructing various native contact maps M_C^S, where S and C are measured in Å and are described in Figure 1. S is the screening strength, which sets the radius of each shadowing atom, and C is the cutoff radius that sets the maximum separation allowed between contacts. Results are summarized for four proteins in Table 1. To quantify average values and statistical variability in the contact map calculations, we use a standard library of 33 nonhomologous globular proteins (NHGP) often used in structure prediction.[47] Figure 4A shows the number of contacts per atom $N(M_C^S)$ as a function of the cutoff radius for different shadowing sizes, averaged over NHGP. M_6^0 has nearly 6 contacts per atom, but for M_6^1, it drops to 1.2 contacts per atom. Thus, geometric occlusion removes 80% of the contacts if shadowing atoms are given a radius of 1 Å. Interestingly, shadowing has a significant effect, even at cutoff distances as small as 4 Å, $N(M_4^0)$ = 0.70 while $N(M_4^1)$ = 0.43.

While shadowing removes contacts that are occluded by intervening atoms, longer distance contacts that are separated by buried (implicit) solvent are maintained. Figure 5A and B shows the contact networks in regions of disrupted secondary structure, where buried waters satisfy the leftover backbone hydrogen bonds. The black dotted lines highlight contacts that are separated by more than 4.5 Å but that are included in the Shadow map. In both cases shown, there are waters sufficiently localized to be detected in X-ray crystallography, which are depicted by yellow spheres. The water molecules sit in voids in the protein interior, and provide stabilization to a configuration that would otherwise be enthalpically costly. Although the solvent is not explicitly modeled in SBMs, choosing contacts through shadowing automatically fills these open pockets with

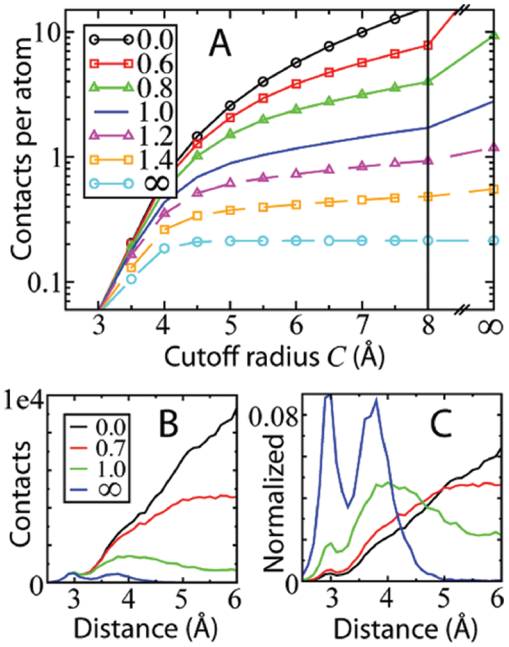


Figure 4. The removal of contacts through shadowing. (A) Average number of contacts per atom $N(M_C^S)$ as a function of cutoff radius C. Each curve is for a different shadowing size S. Each $N(M_C^S)$ represents an average over all the atoms of the proteins in NHGP. C, S = ∞ means C, S greater than the diameter of the protein. (B) Contact distance histogram for different S. Curves represent a sum over all the proteins in NHGP. (C) Contact distance histogram normalized by total contacts, $N_C^S \times N_{atoms}$. This corresponds to the distribution of contact energy as a function of contact distance. S = 0 is peaked at the cutoff limit, whereas S = 1 peaks at an intermediate distance of 4 Å.

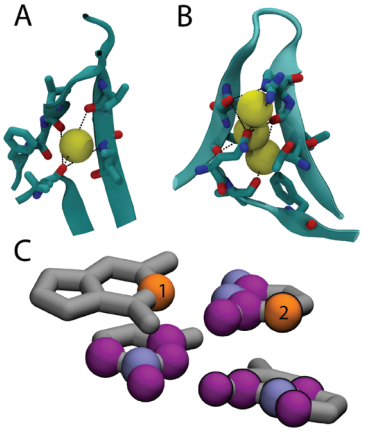


Figure 5. Shadow automatically includes contacts where ligands, metal clusters, and buried waters are not explicitly represented. (A) Buried water in flavodoxin (PDB code: 2FCR) coordinated by VAL87:O, VAL121:O, LEU143:O, and ILE89:N. Black dotted lines show three contacts longer than 4.5 Å involving LEU143:O that are included in M_6^1. (B) Buried waters in interleukin 1-β (PDB code: 1L1B) where three β-sheets come together. Several contacts included in M_6^1 and longer than 4.5 Å are shown. (C) Shadowed contacts in an RNA helix. Native contacts within 4 Å (M_4^0) of the two numbered atoms (orange) are shown in blue and purple. The contacting atoms that are shadowed (i.e., not in M_4^1) are shown in purple. Atom 1 has both stacking interactions and base pairing interactions, while atom 2 has only stacking interactions.

dx.doi.org/10.1021/jp300852d | J. Phys. Chem. B 2012, 116, 8692−8702

compensating contacts because there are no occluding atoms in the void left by the solvent.

There are global differences between the distributions of stabilizing contact energy between cutoff and shadowing maps, i.e., $S = 0$ and $S = 1$. The most obvious difference is the significant reduction in the total number of contacts when $S = 1$. This reduction in contacts is strongest for the longest distance contacts, since they are more likely to be occluded. This alters the contact radial distribution function (Figure 4B,C). The distribution becomes more heavily weighted toward short-range contacts. Peaks at 3 and 4 Å become visible in M_6^1 and are more pronounced for M_∞^∞ (only nearest neighbors). For all contact maps, the 3 Å peak is due to the hydrogen bonding interactions along the secondary structure and the 4 Å peak results from hydrophobic interactions. A more subtle difference is that shadowing tends to smooth the distribution of stabilizing energy between residues. There is a reduction in residue−residue contact energy variance for $S > 0$ (Table 1). Residue−residue contact energy is defined as the sum of atom−atom contacts shared between two residues. These differing contact energy distributions will be seen to alter the thermodynamics of protein folding (discussed in section 3.3).

A quantity that shows no systematic variation with contact map is the relative contact order (CO). Averaged over the proteins in NHGP, $\langle \Delta CO \rangle = \langle CO_i(M_6^1) - CO_i(M_6^0) \rangle \approx 0$, and there is little variation from protein to protein, since $\langle |\Delta CO| / CO \rangle = 0.04$. The constant CO shows that the ratio of long-range to short-range (in sequence) contacts is constant.

3.1.2. Parameter Reduction: $C \to \infty$ and $S \to \infty$. By increasing $C \to \infty$, a cutoff-invariant definition of contacts is obtained. This corresponds to including as contacts any unshadowed atoms regardless of distance. As mentioned above, any protein interior contacts so generated are likely enthalpically important, since an absence of mediating atoms is entropically unlikely. $N(M_\infty^1)$ increases by 1.7 over M_6^1 to 2.9, but for a slightly larger shadow size, $N(M_\infty^{1,2})$ only increases by 0.3 contacts per atom over $N(M_6^{1,2})$. The amount of free space rapidly decreases for $S > 1$. These additional contacts generated with long cutoffs are dominated by atoms near the protein surface interacting through multiple waters. In order to separate out the desired interior contacts, we would need to introduce a burial parameter, and this is left to future studies.

A parameter-less contact map results from $C \to \infty$ and $S \to \infty$. Since an atom k can only shadow a contact between atom pair ij if r_{ik}, $r_{jk} < r_{ij}$, M_C^∞ only includes the nearest neighbor pairs, $N(M_\infty^\infty) \sim 0.2$. While M_∞^∞ can be used to find nearest neighbor pairs, nearly all interactions longer than 4.5 Å are excluded (Figure 4C) and it does not result in cooperative folding (data not shown).

3.2. Decoupling the Protein Geometry from the Contact Energy Distribution. If the contact potentials introduce additional excluded volume between native atomic pairs (Figure 2), different contact maps will have different amounts of excluded volume. To probe the effects of introducing additional excluded volume in the native contacts, the thermodynamics of CI2 was calculated (Figure 6A). Heat capacity (C_V) was compared for M_6^0 with varied native repulsive distances and a constant repulsive distance (r_{NC}) between non-native beads of 1.7 Å. For example, the black curve labeled "4 Å" includes a Lennard-Jones-type repulsion (σ_{NC}) at 4 Å between all native pairs of 4 Å or larger, and at the native position for those closer than 4 Å. C_V becomes sharper (more cooperative) with increasing native repulsion. Also, since the folded basin is being destabilized relative to the unfolded basin, the folding temperature T_F (i.e., the temperature at the peak in C_V) decreases. This excluded volume effect makes the Hamiltonian with Lennard-Jones contact potentials ("LJ") markedly more cooperative and less stable than the equivalent Hamiltonian with Gaussian contact potentials ("1.7 Å").

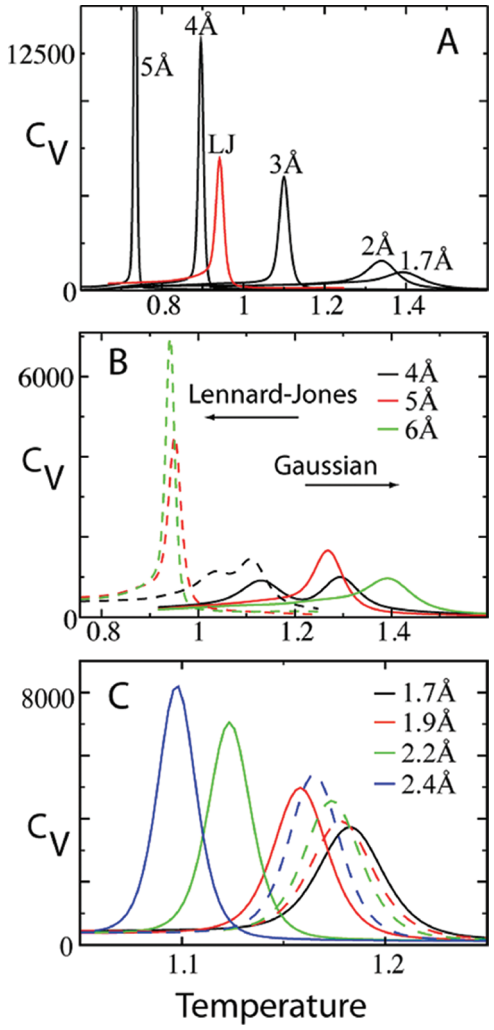


Figure 6. Excluded volume imparts cooperativity. (A) Heat capacity of the folding transition for M_4^0 and atoms of diameter 1.7 Å. As the excluded volume of only the native contacts is increased, both the folding temperature T_F and cooperativity are dramatically affected. The red curve "LJ" uses Lennard-Jones contact potentials. Though it would seem LJ should be less stable than the 4 and 5 Å curves, the tighter width of the Gaussian potential causes an additional destabilization beyond that of the additional native excluded volume in LJ. (B) After removing the extraneous excluded volume, increasing cutoff distance has the opposite effect on T_F. Increased cutoff results in increased cooperativity with Lennard-Jones, whereas the Gaussians show a maximum at intermediate cutoff. (C) The black curve denotes the Shadow map with standard parameters and Gaussian contact potentials. Dotted curves correspond to excluded volume altered only between non-native pairs. Solid curves correspond to excluded volume altered between all pairs. $\kappa_1 = 0.032$ for the standard Shadow map, and $\kappa_1 = 0.018$ when the atomic repulsive size is increased to 2.4 Å. All data is from CI2.

The tendency of native excluded volume to alter cooperativity and stability has opposite thermodynamic behavior between Lennard-Jones and Gaussian potentials with M_4^0, M_5^0, and M_6^0 (Figure 6B). The Lennard-Jones potentials

dx.doi.org/10.1021/jp300852d | J. Phys. Chem. B 2012, 116, 8692−8702

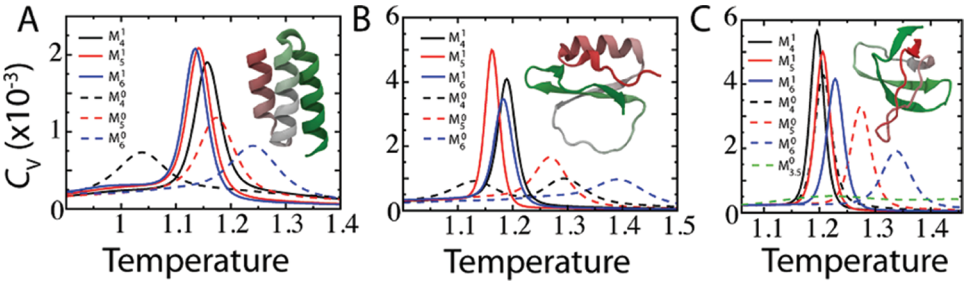


Figure 7. Heat capacity is consistent and folding is cooperative as the cutoff parameter C is varied with the Shadow algorithm. Three proteins are shown: (A) BDPA, (B) CI2, and (C) SH3. Their native structures are shown as insets. The solid lines are shadowing maps with $S = 1$, and the dotted lines are cutoff maps ($S = 0$).

decrease protein stability, since increasing the contact map cutoff C introduces more native contacts, and thus more native excluded volume. The increased excluded volume decreases the entropy of the native basin relatively more than the unfolded basin, and therefore decreases the stability of the native state. In contrast, the Gaussian potentials isolate the effects of changing the contact energy distribution by maintaining a constant native excluded volume of 1.7 Å between all atoms. The Gaussian potentials show an opposite behavior, where protein stability is increased as C is increased. Now the dominant effect is the increased entropy of the native state as more contacts are introduced. This stabilizing effect will be further discussed in the next section.

Independent of the contact map and contact potential, the repulsive size of the atoms also affects the folding cooperativity and stability. The Gaussian potential allows us to also isolate the effects of changing the atomic repulsion between either only the non-native atomic pairs or all atomic pairs (Figure 6C). The Shadow map (M_6^1) is used, non-native excluded volume is controlled by r_{NC} (eq 1), and native excluded volume is controlled by r_{ex} (eq 3). Increasing the size of all the atoms has a similar effect as only increasing the repulsion between native pairs (Figure 6A), where κ_1 increases and stability decreases. Since the native state is denser and has more atomic collisions than the unfolded configurations, the entropy of the native basin is relatively smaller when the atoms are larger. Somewhat surprising is that increasing the repulsive size of only the non-native interactions follows the same trend as well. While one might surmise that a larger excluded volume of non-native interactions lowers the entropy of the unfolded basin more than the folded basin, the destabilizing effect shows that in fact non-native interactions are more frequently encountered in the *folded* basin of the all-atom model. This is opposite to the effect seen in a closely related coarse-grained C_α model.[37] While the C_α atoms in the backbone are similarly constrained to their native positions in both the coarse-grained and all-atom models,[12] the all-atom model introduces close-packed side chains that encounter many non-native atomic collisions. In addition to the close atomic distances, there are less native restraints on each atom, since M_6^1 gives 1.2 contacts per atom versus 2.6 contacts per residue. We note that the ability to encounter non-native collisions is enhanced by the smooth energy landscape. Previous work showed that an all-atom SBM makes comparatively more non-native contacts in the folded basin than an explicit solvent transferable potential like OPLS.[12]

3.3. Shadowing Tends to Increase Folding Cooperativity. In this section, we explore the effects on folding cooperativity of changing the largest energetic component of the SBM, the native contact map. The native contact map defines the distribution of tertiary stabilizing energy. The effects of changing this distribution are isolated by using a Gaussian contact potential that maintains a constant excluded volume across contact maps (Figure 2). The model proteins are three small, fast-folding globular proteins: B-domain of protein A (BDPA), chymotrypsin inhibitor 2 (CI2), and the sh3 domain of c-src kinase (SH3). These three proteins, which we studied previously,[12] are well studied both experimentally[39,48,49] and theoretically[11,20,50] and represent simple to complicated folds, respectively.[15] Differential scanning microcalorimetry has shown that small globular proteins like BDPA, CI2, and SH3 fold cooperatively in a two-state manner with singly peaked heat capacity at the folding transition and $\kappa_1 < 0.05$ and $\kappa_2 > 0.95$.[39,40,51]

We find that using a contact map generated with geometric occlusions consistently increases folding cooperativity relative to a map generated with a cutoff distance. Figure 7 shows the heat capacity calculated for two sets of contact maps and three proteins. The first set of maps used a direct cutoff (M_4^0, M_5^0, and M_6^0), while the second set have $S = 1$ (M_4^1, M_5^1, and M_6^1). In every case, the map with $S = 1$ has a smaller κ_1 than the corresponding cutoff map (Table 1). In addition to consistently higher folding cooperativity, the thermal stabilities for $S = 1$ vary little in the same protein (<5%) and between proteins (<10%). The Shadow map (M_6^1) dependably gives folding temperatures near 1.2 for globular proteins. Proteins (PDB codes) not in Figure 7 that have been folded with the default all-atom SBM are 3MLG, 1RIS, 2A3D, and 2EFV, and have folding temperatures of 1.12, 1.21, 1.18, and 1.15, respectively.

The thermodynamics of the cutoff contact maps shows some interesting features. First, as C increases, the protein becomes more thermally stable seen by the movement of T_F. This is because of two effects: (1) as C increases, the contacts are on average wider and (2) the stabilizing energy is more diffuse. Both of these effects increase the entropy of the native state and hence increase stability. The cutoff map contact distance distribution is skewed toward C, and therefore, the average native distance between contacts $\langle r_0^{ij} \rangle$ increases with C (Figure 4). A larger native distance produces a wider contact potential, since $(r_0^{ij}) \propto \sigma_{ij}$ (eq 4). The energy distribution becomes more diffuse because at higher C there are more total contacts. The total energy available for the contacts is held fixed, so each contact has a smaller share of stabilizing energy. Interestingly, κ_2 does not follow the trend of κ_1 as $C \rightarrow 6$ Å, instead staying constant or even increasing. This implies that the increase in κ_1 is not from the introduction of intermediate states but rather from the slow conversion of well-defined unfolded and folded ensembles. Second, there is a minimum cutoff distance, below which the protein no longer makes a cooperative transition.

dx.doi.org/10.1021/jp300852d | J. Phys. Chem. B 2012, 116, 8692−8702

Remarkably, at C = 4 Å, CI2 becomes a three-state folder and the heat capacity shows a thermodynamic intermediate[36] (Figure 7B). At C = 3.5 Å, SH3 resembles a downhill folder (Figure 7C). Last, since cooperativity vanishes at both low and high C, there is a peak in cooperativity at an intermediate range of 4 Å < C < 5 Å. The thermal stability T_F of the most cooperative cutoff maps is near the stability of the Shadow map. This property, that the contact maps with similar stabilities have similar cooperativities, was seen to hold among the many variations of M_C^S tested for this paper. It implies that there is an optimal temperature to have a cooperative transition. Perhaps, the Shadow map consistently achieves this stability and thus is cooperative.

3.4. Dynamics of RNA and Macromolecular Assemblies. There are many new and exciting areas ripe for exploring through the lens of energy landscape theory, the foundation upon which structure-based models are built. These theoretical tools are already being applied to the study of RNA folding[25,52,53] and the dynamics of molecular machines composed of either protein, such as kinesin,[54] or RNA−protein complexes, like the ribosome.[4,55] In this section, we look beyond protein folding, and show that Shadow contact maps provide a consistent treatment for heterogeneous systems, and thus a solid framework for addressing the geometrical features of molecular machines.

3.4.1. RNA Contact Maps. RNA has three main types of contacts, Watson−Crick (WC) base-pairing, base-stacking (BS) interactions, and tertiary backbone contacts (Figure 5C). WC pairs are the hydrogen bonding interactions between complementary RNA bases (i.e., A·U and G·C). BS interactions refer to $\pi-\pi$ stacking: attractive, noncovalent interactions between the aromatic rings of stacked bases that are adjacent in sequence. Maintaining proper energetic balance between these interactions will be important to the performance of RNA models.

Short-range cutoff contact maps have been shown to overweight the BS interactions relative to WC pairs and tertiary contacts. To maintain a proper balance between secondary and tertiary structure in the study of the folding of the mRNA SAM-I riboswitch with a SBM,[25] BS interactions were scaled by a factor of 1/3 when using a 4 Å contact map M_4^0. Here, we denote the cutoff contact map including scaled BS interactions as M_4^0*. The overstabilization of BS interactions in M_4^0 arises from the geometry of closely packed rings. As seen in Figure 5C, atoms 1 and 2 are each within 4 Å of five atoms in the adjacent stacked base. This is the case for every atom in the ring, and for every stacked ring in the riboswitch. Interestingly, if geometric occlusion is considered, due to the close packing, the overcounting is avoided. Introducing shadowing with S = 1 Å, atoms 1 and 2 each have only a single stacking interaction.

Shadowing naturally gives rise to the approximate 1/3 scaling in stacking interactions. Table 2 compares M_4^0* to M_6^1 for an RNA helix. Base-stacking interactions relative to WC pairs are decreased by a factor of 0.59/1.74 = 0.34. Relative to all contacts, the BS contacts are decreased by a factor of 0.18/0.48 = 0.37, which is in surprising agreement with the previous conjecture of 1/3.[25] Thus, the energy distributions between M_4^0* and M_6^1 are similar in RNA but vary by a factor of 2.5 in the number of total contacts. The heat capacity of the isolated 16-residue P2 helix of the SAM-I riboswitch[25] was calculated for the two contact maps (Figure 8). M_6^1 is more cooperative, while M_4^0* is more stable. These trends are in line with those observed in section 3.3 for proteins. M_6^0 in protein is the analogue of M_4^0* in RNA. It introduced an excess of contacts that increased stability, while the shadow map M_6^1 was less stable but more cooperative. Thus, while the Shadow map gives a reasonable distribution of energy within RNA, to be applied to RNA−protein assemblies, the shared energy distribution with proteins must also be balanced.

Table 2. Comparison of Cutoff versus Shadowing Contact Maps in RNA Systems (The RNA Helix Is Helix P2 from the SAM-I Riboswitch[25] Minus the 4 Turn Residues)

	M_4^0*	M_6^1
RNA Helix		
Watson−Crick contacts (WC)	137	61
base-stacking contacts (BS)	232	35
total contacts (All)	480	190
WC/BS	0.59	1.74
BS/All	0.48	0.18
Ribosome		
$E_{rna-rna}$ contacts	77 529	57 355
$E_{pro-rna}$ contacts	8045	14 771
$E_{pro-pro}$ contacts	15 053	28 510
E_{rna}^{D} (per RNA atom)[a]	0.37	0.37
E_{pro}^{D} (per protein atom)	0.26	0.26
E_{rna}^{C+D} (per RNA atom)[b]	1.18	1.01
E_{pro}^{C+D} (per protein atom)	0.64	0.97
$\sigma_{E^C}/\overline{E^C}$ in RNA atoms[c]	0.27	0.26
$\sigma_{E^C}/\overline{E^C}$ in protein atoms	0.63	0.49

[a]Dihedral energy in RNA (protein) divided by the number of RNA (protein) atoms. [b]Total contact and dihedral energy in RNA (protein) divided by the number of RNA (protein) atoms. [c]E^C represents the contact energy per atom by residue.

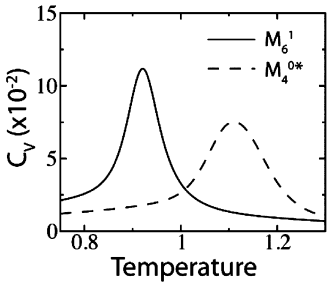


Figure 8. Folding an RNA hairpin with the all-atom SBM (16-residue helix P2 of the SAM-I riboswitch, PDB code: 2GIS). The heat capacity C_V is compared between two native contact maps. M_6^1 refers to the normal Shadow contact map, and M_4^0* refers to a cutoff map where the energy of all base-stacking interactions are scaled by 1/3. κ_1 is 0.10 for M_6^1 and 0.15 for M_4^0*.

3.4.2. Shadowing in Heterogeneous Assemblies. Tables 1 and 2 indicate that the atomic packing is very different between RNA and protein. In the RNA hairpin, M_4^0* has 150% more contacts than M_6^1, whereas, in proteins, M_6^1 has ∼70% more contacts than M_4^0. The regularity of base-stacking dominates the short-range contacts in RNA. Proteins, in contrast, have no regular residue packing, since the amino acid side chains have a diversity of shapes. This difference in packing causes short-range cutoff maps to skew the distribution of stabilizing energy in favor of RNA.

The contact energy per atom by residue E^C in the ribosome is shown in Figure 9. Even with the BS contacts scaled by 1/3 in M_4^0*, the contact energy in RNA is double that in protein, $\overline{E_{rna}^C}/\overline{E_{pro}^C}$ = 2.1, where $\overline{E^C}$ is E_X^C averaged over all residues of

dx.doi.org/10.1021/jp300852d | J. Phys. Chem. B 2012, 116, 8692−8702

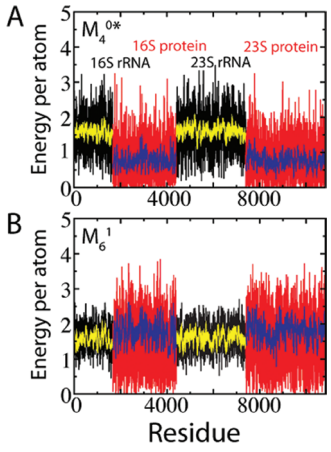


Figure 9. Distribution of native contact energy between proteins and RNA in the ribosome. Energy per atom of residue i is calculated as $\varepsilon_C N_C^i / N_A^i$, where N_C^i is the number of atomic contact pairs involving atoms of residue i and N_A^i is the number of atoms in residue i. The previously used[4,25] cutoff contact map M_4^{0*} (A) is compared to the Shadow contact map M_6^1 (B). Black regions correspond to the 16S and 23S RNA, and red regions correspond to the proteins associated with the 16S and 23S. Yellow and blue are running averages over 100 residues as a guide for the eye.

type X. The Shadow map gives a much closer division, $\overline{E_{rna}^C} / \overline{E_{pro}^C} = 0.86$. Since RNA has a higher density of dihedrals than protein, if the dihedral and contact energy are summed and compared, the Shadow map gives an equal distribution of energy, $\overline{E_{rna}^{C+D}} / \overline{E_{pro}^{C+D}} = 1.0$. This feature is desirable when simulating heterogeneous molecular assemblies. Fluctuations in the ribosome for two different contact maps, M_4^{0*} and M_6^1, are compared to the fluctuations predicted from the experimental B-factors (Figure 10A). For the 23S ribosomal RNA, the correlation between experiment and the SBM with the Shadow map is 0.78. On a smaller scale, to highlight the variability between contact maps, fluctuations are shown for three proteins at $\sim 0.75 T_F$ (Figure 10B). While the correlation is high between the two maps, deviations can be seen. Future work will have to explore how robust these fluctuations are, since deviations in fluctuations between related proteins have been predicted to have functional consequences.[28]

4. CONCLUSIONS

We have proposed a general algorithm for generating atomically grained contact maps called "Shadow" (Figure 1). This algorithm enables sufficient contact cutoff distances to capture atomic contacts across structural waters or heavy metals that are not explicitly represented, without introducing contacts between atom pairs that one does not wish to model, specifically, those that have an intervening atom. The Shadow algorithm initially considers all atoms within a cutoff distance C = 6 Å and then, controlled by a screening parameter S = 1 Å, discards the occluded contacts. We showed that this choice of contact map is not only well behaved for protein folding, since it produces consistently cooperative folding behavior, but also desirable in exploring the dynamics of macromolecular assemblies, since it distributes energy similarly between RNAs and proteins despite their disparate internal packing.

The study of the connection between the contact distribution and folding cooperativity highlighted that many components of the SBM Hamiltonian affect cooperativity, especially the

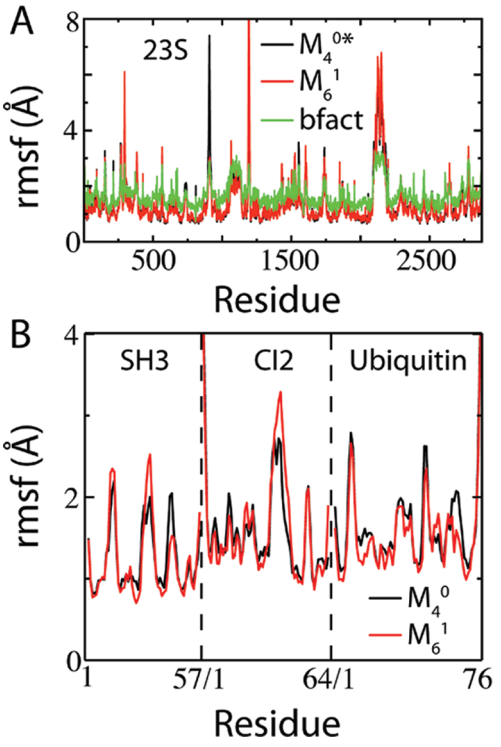


Figure 10. Structure-based models capture near-native-state fluctuations of both small proteins and large macromolecular assemblies. (A) Comparison of simulated root mean squared fluctuations (rmsf) of each residue in the 23S ribosomal RNA (PDB codes: 3F1E, 3F1F) between the scaled cutoff map M_4^{0*} and the Shadow map M_6^1. Overlaid are the rmsf by residue, calculated using the experimental B-factors (using the isotropic approximation, rmsf = $(\langle r_i^2 \rangle)^{1/2} \sim (3/8\pi^2) B_i$, where r_i is the displacement of atom i and B_i is the experimental B-factor of atom i[62]). A residue rmsf is computed as the arithmetic average of its constituent atoms' rmsf. Overall rmsf agreed with the B-factors for M_4^{0*} at T = 0.46 and M_6^1 at T = 0.71. The discrepancy in stability is likely due to M_6^1 including double the Mg^{2+}–RNA contacts, which are modeled as harmonic restraints instead of Gaussian contact potentials. Pearson correlation r between M_6^1 and B-factors is 0.78. (B) Simulated rmsf of the C_α atoms for three globular proteins at T = 0.9. The overall agreement is very close ($r > 0.9$) between two different contact maps, M_4^0 and M_6^1.

geometric components. We showed how the Lennard-Jones contact interaction mixes the geometric and energetic parts of the Hamiltonian by changing the excluded volume of native interactions. By decoupling the geometric and energetic parts with the Gaussian contact potential, it became clear that the increased cooperativity obtained through additional Lennard-Jones native contacts was caused by the extra excluded volume. Further, the decoupling showed that the innate cooperativity of the Shadow map was purely an effect of the contact energy distribution. In the case of CI2, the energetic effect of changing contact maps from M_6^0 to M_6^1 decreased κ_1 from 0.12 to 0.032, while the geometric effect of increasing the diameter of the atoms from 1.7 Å to a more realistic 2.4 Å brought κ_1 even further down to 0.018 (experimental range was $\kappa_1 < 0.05$). Other studies have shown that, for example, excluded volume,[56,57] backbone stiffness,[12,58] contact potential width (e.g., σ_{ij} in eq 4),[37,59,60] and many-body effects[24,57,61] affect the cooperativity of protein folding models.

Structure-based models will continue to be an important tool in the characterization of molecular machines and macro-

dx.doi.org/10.1021/jp300852d | J. Phys. Chem. B 2012, 116, 8692–8702

molecular assemblies. They are baseline models that can be used to fully discern the role of biomolecular geometry. Going forward, all-atom structure-based models employing Shadow contact maps provide a general framework for exploring the geometrical features of biomolecules, especially the connections between folding and function.

AUTHOR INFORMATION

Corresponding Author

*E-mail: jonuchic@rice.edu.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

J.K.N. wishes to thank Shachi Gosavi for helpful discussion and enthusiasm. This work was supported by the Center for Theoretical Biological Physics sponsored by the NSF (Grant PHY-0822283) and by NSF-MCB-1214457. J.N.O. is a CPRIT Scholar in Cancer Research sponsored by the Cancer Prevention and Research Institute of Texas. This research was also supported in part by the NSF through TeraGrid resources provided by TACC under grant number TGMCB110021. J.K.N. was supported in part by an NIH Molecular Biophysics Training Grant while at UCSD (Grant T32 GM08326).

REFERENCES

(1) Yusupov, M. M.; Yusupova, G. Z.; Baucom, A.; Lieberman, K.; Earnest, T. N.; Cate, J. H.; Noller, H. F. *Science* **2001**, *292* (5518), 883−896.

(2) Bochtler, M.; Ditzel, L.; Groll, M.; Hartmann, C.; Huber, R. *Annu. Rev. Biophys. Biomol. Struct.* **1999**, *28*, 295−317.

(3) Wahl, M. C.; Will, C. L.; Lührmann, R. *Cell* **2009**, *136* (4), 701−718.

(4) Whitford, P. C.; Geggier, P.; Altman, R. B.; Blanchard, S. C.; Onuchic, J. N.; Sanbonmatsu, K. Y. *RNA* **2010**, *16* (6), 1196−1204.

(5) Bryngelson, J.; Wolynes, P. *J. Phys. Chem.* **1989**, *93*, 6902−6915.

(6) Leopold, P. E.; Montal, M.; Onuchic, J. N. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89* (18), 8721−8725.

(7) Onuchic, J. N.; Wolynes, P. G. *Curr. Opin. Struct. Biol.* **2004**, *14* (1), 70−75.

(8) Bryngelson, J.; Wolynes, P. *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *84*, 7524.

(9) Levy, Y.; Wolynes, P. G.; Onuchic, J. N. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101* (2), 511−516.

(10) Miyashita, O.; Onuchic, J. N.; Wolynes, P. G. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100* (22), 12570−12575.

(11) Clementi, C.; Nymeyer, H.; Onuchic, J. N. *J. Mol. Biol.* **2000**, *298* (5), 937−953.

(12) Whitford, P. C.; Noel, J. K.; Gosavi, S.; Schug, A.; Sanbonmatsu, K. Y.; Onuchic, J. N. *Proteins* **2009**, *75* (2), 430−441.

(13) Miyazawa, S.; Jernigan, R. *Macromolecules* **1985**, *18* (3), 534−552.

(14) Tirion, M. *Phys. Rev. Lett.* **1996**, *77* (9), 1905−1908.

(15) Plaxco, K. W.; Simons, K. T.; Baker, D. *J. Mol. Biol.* **1998**, *277* (4), 985−994.

(16) Silveira, C. H. D.; Pires, D. E. V.; Minardi, R. C.; Ribeiro, C.; Veloso, C. J. M.; Lopes, J. C. D.; Meira, W.; Neshich, G.; Ramos, C. H. I.; Habesch, R.; et al. *Proteins* **2009**, *74* (3), 727−743.

(17) Sobolev, V.; Sorokine, A.; Prilusky, J.; Abola, E. E.; Edelman, M. *Bioinformatics* **1999**, *15* (4), 327−332.

(18) Sułkowska, J. I.; Cieplak, M. *Biophys. J.* **2008**, *95* (7), 3174−3191.

(19) Veloso, C. J. M.; Silveira, C. H.; Melo, R. C.; Ribeiro, C.; Lopes, J. C. D.; Santoro, M. M.; Meira, W. *Genet. Mol. Res.* **2007**, *6* (4), 799−820.

(20) Shea, J.-E.; Onuchic, J. N.; Brooks, C. L., III. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96* (22), 12512−12517.

(21) Koga, N.; Takada, S. *J. Mol. Biol.* **2001**, *313* (1), 171−180.

(22) Shen, T.; Zong, C.; Portman, J. J.; Wolynes, P. G. *J. Phys. Chem. B* **2008**, *112* (19), 6074−6082.

(23) Zhang, Z.; Chan, H. S. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107* (7), 2920−2925.

(24) Kaya, H.; Chan, H. S. *J. Mol. Biol.* **2003**, *326* (3), 911−931.

(25) Whitford, P. C.; Schug, A.; Saunders, J.; Hennelly, S. P.; Onuchic, J. N.; Sanbonmatsu, K. Y. *Biophys. J.* **2009**, *96* (2), L7−9.

(26) Noel, J. K.; Onuchic, J. N.; In *Computational Modeling of Biological Systems*; Dokholyan, N., Ed.; Springer: New York, 2012; Chapter 2.

(27) Noel, J. K.; Sulkowska, J. I.; Onuchic, J. N. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107* (35), 15403−15408.

(28) Nechushtai, R.; Lammert, H.; Michaeli, D.; Eisenberg-Domovich, Y.; Zuris, J. A.; Luca, M. A.; Capraro, D. T.; Fish, A.; Shimshon, O.; Roy, M.; et al. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108* (6), 2240−2245.

(29) Jamros, M. A.; Oliveira, L. C.; Whitford, P. C.; Onuchic, J. N.; Adams, J. A.; Blumenthal, D. K.; Jennings, P. A. *J. Biol. Chem.* **2010**, *285* (46), 36121−36128.

(30) Ratje, A. H.; Loerke, J.; Mikolajka, A.; Brünner, M.; Hildebrand, P. W.; Starosta, A. L.; Dönhöfer, A.; Connell, S. R.; Fucini, P.; Mielke, T.; et al. *Nature* **2010**, *468* (7324), 713−716.

(31) Schug, A.; Weigt, M.; Onuchic, J. N.; Hwa, T.; Szurmant, H. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106* (52), 22124−22129.

(32) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. *J. Chem. Theory Comput.* **2008**, *4* (3), 435−447.

(33) Noel, J. K.; Whitford, P. C.; Sanbonmatsu, K. Y.; Onuchic, J. N. *Nucleic Acids Res.* **2010**, *38*, W657−61.

(34) Ferrenberg, A.; Swendsen, R. *Phys. Rev. Lett.* **1988**, *61* (23), 2635−2638.

(35) Ferrenberg, A.; Swendsen, R. *Phys. Rev. Lett.* **1989**, *63* (12), 1195−1198.

(36) Wu, L.; Zhang, J.; Qin, M.; Liu, F.; Wang, W. *J. Chem. Phys.* **2008**, *128* (23), 235103.

(37) Lammert, H.; Schug, A.; Onuchic, J. N. *Proteins* **2009**, *77* (4), 881−891.

(38) Privalov, P. L.; Khechinashvili, N. N. *J. Mol. Biol.* **1974**, *86* (3), 665−684.

(39) Jackson, S. E.; Fersht, A. R. *Biochemistry* **1991**, *30* (43), 10428−10435.

(40) Privalov, P. L.; Potekhin, S. A. *Methods Enzymol.* **1986**, *131*, 4−51.

(41) Kaya, H.; Chan, H. S. *Proteins* **2000**, *40* (4), 637−661.

(42) Clementi, C.; García, A. E.; Onuchic, J. N. *J. Mol. Biol.* **2003**, *326* (3), 933−954.

(43) Cho, S.; Levy, Y.; Wolynes, P. G. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103* (3), 586−591.

(44) Faure, G.; Bornot, A.; de Brevern, A. G. *Biochimie* **2008**, *90* (4), 626−639.

(45) Williams, M. A.; Goodfellow, J. M.; Thornton, J. M. *Protein Sci.* **1994**, *3* (8), 1224−1235.

(46) Rashin, A. A.; Honig, B. *J. Mol. Biol.* **1984**, *173* (4), 515−521.

(47) Papoian, G. A.; Ulander, J.; Eastwood, M. P.; Luthey-Schulten, Z.; Wolynes, P. G. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101* (10), 3352−3357.

(48) Sato, S.; Religa, T. L.; Daggett, V.; Fersht, A. R. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101* (18), 6952−6956.

(49) Viguera, A. R.; Martínez, J. C.; Filimonov, V. V.; Mateo, P. L.; Serrano, L. *Biochemistry* **1994**, *33* (8), 2142−2150.

(50) Hoang, T.; Cieplak, M. *J. Chem. Phys.* **2000**, *113* (18), 8319−8328.

(51) Schafer, H.; van Gunsteren, W. F.; Mark, A. E. *J. Comput. Chem.* **1999**, *20* (15), 1604−1617.

(52) Sorin, E. J.; Nakatani, B. J.; Rhee, Y. M.; Jayachandran, G.; Vishal, V.; Pande, V. S. *J. Mol. Biol.* **2004**, *337* (4), 789−797.

(53) Hyeon, C.; Dima, R. I.; Thirumalai, D. *Structure* **2006**, *14* (11), 1633−1645.
(54) Hyeon, C.; Onuchic, J. N. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104* (44), 17382−17387.
(55) Whitford, P. C.; Ahmed, A.; Yu, Y.; Hennelly, S. P.; Tama, F.; Spahn, C. M. T.; Onuchic, J. N.; Sanbonmatsu, K. Y. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108* (47), 18943−18948.
(56) Qi, X.; Portman, J. J. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104* (26), 10841−10846.
(57) Suzuki, Y.; Noel, J. K.; Onuchic, J. N. *J. Chem. Phys.* **2011**, *134* (24), 245101.
(58) Prieto, L.; Rey, A. *J. Chem. Phys.* **2007**, *126* (16), 165103.
(59) Prieto, L.; de Sancho, D.; Rey, A. *J. Chem. Phys.* **2005**, *123*, 154903.
(60) Suzuki, Y.; Noel, J. K.; Onuchic, J. N. *J. Chem. Phys.* **2008**, *128* (2), 025101.
(61) Eastwood, M.; Wolynes, P. *J. Chem. Phys.* **2001**, *114* (10), 4702.
(62) Garcia, A.; Krumhansl, J.; Frauenfelder, H. *Proteins* **1997**, *29* (2), 153−160.