Exploring the Free Energy Landscape of Nucleosomes

Bin Zhang,[†] Weihua Zheng,[†] Garegin A. Papoian,[‡] and Peter G. Wolynes*,[†,§]

[†]Department of Chemistry and Center for Theoretical Biological Physics and [§]Department of Physics and Astronomy, Rice University, Houston, Texas 77005, United States

[‡]Department of Chemistry and Biochemistry and Institute for Physical Science and Technology, University of Maryland, College Park, Maryland 20742, United States

S *Supporting Information*

ABSTRACT: The nucleosome is the fundamental unit for packaging the genome. A detailed molecular picture for its conformational dynamics is crucial for understanding transcription and gene regulation. We investigate the disassembly of single nucleosomes using a predictive coarse-grained protein DNA model with transferable force fields. This model quantitatively describes the thermodynamic stability of both the histone core complex and the nucleosome and predicts rates of transient nucleosome opening that match experimental measurements. Quantitative characterization of the free-energy landscapes reveals the mechanism of nucleosome unfolding in which DNA unwinding and histone protein disassembly are coupled. The interfaces between H2A-H2B dimers and the $(H3-H4)_2$ tetramer are first lost when the nucleosome opens releasing a large fraction but not all of its bound DNA. For the short strands studied in single molecule experiments, the DNA unwinds asymmetrically from the histone proteins, with only one of its two ends preferentially exposed. The detailed molecular mechanism revealed in this work provides a structural basis for interpreting experimental studies of nucleosome unfolding.
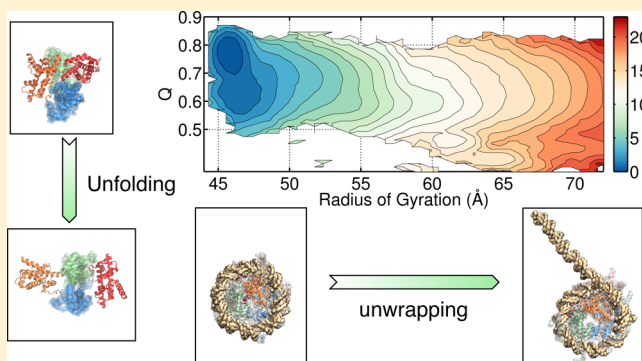
INTRODUCTION

The genome, the blueprint of life, contains nearly all the information needed to build and maintain an entire organism. In higher organisms, at the chromosomal level, the three-dimensional structural organization of the genome is crucial for its function.[1−3] Large scale chromosome folding can bring into proximity regulatory elements, i.e., enhancers and promoters, that are separated far away in sequence in order to control gene expression.[4,5] At a finer nanometer scale, the packaging of the genome plays an important role in gene regulation as well.[6] For eukaryotic cells, the fundamental unit of DNA organization is the so-called nucleosome, in whose crystal structure the DNA wraps approximately 1.7 times around a core made of histone proteins.[7] We investigate the stability and conformational dynamics of single nucleosomes by computing the free-energy landscapes for nucleosome disassembly using a coarse-grained model that includes a transferable protein force field suitable for structure prediction[8,9] and a DNA force field that successfully predicts its elastic properties.[10,11]

The nucleosome structure itself presents a steric barrier for gene transcription.[6] Approximately 147 base pairs of duplex DNA wrap around each histone octamer, which is formed from two copies of the histone heterodimers $(H2A-H2B)_{\alpha,\beta}$ and $(H3-H4)_{\alpha,\beta}$. The linker length between neighboring nucleosomes ranges from 20 to 90 base pairs long,[12] and approximately 75% of the DNA is sterically occluded by being bound in nucleosomes.[13,14] In order for other proteins,

including transcription factors and RNA polymerases, to access their binding sites, the tightly bound DNA must at least partially unwind from the histone core. Numerous mechanisms inside the cell regulate the stability of the nucleosome and thereby fine-tune the amount of accessible DNA. These mechanisms range from passive histone modifications to active remodeling that uses ATP.[15,16] A detailed characterization of the molecular mechanism for nucleosome assembly will not only improve our understanding of genome packaging but will also shed light on the various pathways that regulate gene expression kinetically.

Many experimental studies already have provided insight into how the nucleosome assembles. Single molecule stretching experiments using optical traps suggest that the DNA may unwrap from the histone core following a three step process that includes first (i) the release of the outer turn, next (ii) the release of the inner turn, and finally (iii) irreversible dissociation of the histone core.[17,18] When the three stage picture was proposed, the histone core was assumed to be rather rigid retaining a stable octamer conformation. Recently, however single molecule Förster resonance energy transfer (FRET) experiments have revealed that the protein core is rather flexible and disassembles with a loss of the $(H3-H4)_2$ tetramer/(H2A-H2B) dimer interface as the DNA unwraps.[7,19]

Furthermore, in contrast to the symmetric unwinding of the two DNA ends initially proposed, recent experiments on short DNA suggest that the DNA unwraps asymmetrically with one end being predominantly exposed. These experiments used an assay that merges single molecule FRET together with optical tweezers.[20] We show here that computational modeling of nucleosome disassembly further elucidates the molecular mechanism and provides a quantitative theoretical foundation that is needed to unify these experiments.

Using computer simulations of a coarse-grained protein−DNA model, we provide a comprehensive characterization of free-energy landscapes for the histone complex and nucleosome disassembly. We find that the histone core complex without the DNA is unstable at physiological conditions. Instead an intermediate state in which the $(H3\text{-}H4)_2$ tetramer is sandwiched between the two H2A-H2B dimers with non-specific interactions is favored thermodynamically. Though the octamer structure is stable when the DNA is bound, a similar intermediate state which has lost the $(H3\text{-}H4)_2$ tetramer/H2A-H2B dimer interface is observed as the DNA unwinds, supporting the idea that DNA unwrapping and histone core complex unfolding are coupled processes. Finally, the free-energy landscape of nucleosome disassembly supports asymmetric conformations of DNA unwrapping that predominantly expose only one of the two ends. Comparing the free-energy landscapes of intact and tailless nucleosomes, we show that this asymmetric unwrapping mainly arises from electrostatic interactions between histone tails and the DNA, and we find that the tails of histone H3 have the most profound effect. The combined chemical accuracy and computational efficiency of the coarse-grained model thus enables a rigorous energy landscape analysis for a single nucleosome and paves the way for further investigation of higher order structures formed by oligonucleosomes.

■ METHODS

Coarse-Grained Protein−DNA Model. Computational modeling promises to provide a detailed molecular characterization for the assembly of a single nucleosome as well as the higher order structures formed by oligonucleosomes. In fact, atomistic simulations have already provided structural insight into the transient DNA unwrapping near its entry/exit sites[21−23] and revealed the effect of post-translational modifications and different histone variants on nucleosomal dynamics.[22,24] The minimal systems are large in size and involve a complex ensemble of molecular players having intricate physicochemical interactions. These features make the modeling of nucleosomes a challenge[25] and limit the time scale currently accessible from fully atomistic simulations to microseconds.[22,23,26,27] This limitation constrains the application of all atom models currently providing a comprehensive landscape characterization. Instead, we adopt a coarse-grained modeling approach, which has already proven fruitful in investigating a wide range of biological systems.[28−30]

To investigate protein−DNA interactions in the nucleosome, we combine the associative memory, water-mediated, structure and energy model (AWSEM) for protein[9] with an improved version of the three site per nucleotide model (3SPN.2C) for DNA.[10,11] Each amino acid in AWSEM is modeled with three atoms, C_α, C_β, and O, and the transferable interactions among amino acids are parametrized following the energy landscape theory prescription to maximize the ratio of folding temperature over glass transition temperature for a set of training proteins.[31−34] AWSEM has been shown to predict monomer structures reasonably well from sequence alone and to predict protein−protein interfaces in dimers with remarkable accuracy when monomer structures are known.[8,35] The coarse-grained DNA model developed by de Pablo and co-workers quantitatively reproduces the persistence length of double-stranded DNA at varying ionic concentrations and for different DNA sequences.[10,11] Thus, it encodes DNA's elastic properties in a predictive fashion. When we combine the protein and DNA models, we preserve the original fine-tuned force fields for protein−protein and DNA−DNA interactions by themselves. In the present model, we introduce additional protein−DNA interactions at a nonspecific level using a screened Debye−Hückel potential for the electrostatics along with a Lennard-Jones potential for excluded volume (see Supporting Information (SI) for details). A typical dielectric constant of 78.0 for water and an ionic concentration of 100 mM at the physiological condition are used for the electrostatic interactions in this study. Using this nonspecific simplification of the protein−DNA interactions is not unreasonable for the nucleosome assembly problem because the X-ray structure indicates the lack of base-specific interactions between histone proteins and the DNA as well as the predominance of water molecules at the interface of the two.[7,36] We note this kind of simple treatment of the direct protein−DNA interaction has already been applied successfully to study protein−DNA interactions in a wide range of biological systems,[37−39] including some studies of nucleosomes.[40,41]

We introduce two modifications to the original AWSEM force field presented in ref 8 in order to improve modeling the chemical complexity of histone proteins. First, to better characterize long-range electrostatic interaction among histone proteins, we explicitly include Debye−Hükcel potential among charged amino acid residues following ref 42. Unfortunately, such a simple treatment of electrostatics, though useful in capturing long-range interactions, may give rise to some double counting at short-range. This double counting arises since AWSEM, in its original form, already includes short-range direct contact potentials between charged residues that implicitly involve electrostatics. Another modification we employ remedies to some extent the double counting issue. Additional weak nonadditive Gō-potentials derived from the octamer conformation in the nucleosome crystal structure were introduced for the histone protein core. These partially counteract the repulsion among positively charged residues in short range. Nonadditive Gō-potentials are further helpful in quantitatively reproducing energetic barriers and sufficient cooperativity while tuning an already funneled energy landscape more completely toward native conformations.[43] It is important to note that the strength of the nonadditive Gō-potential employed here is small (<30% in the native state of the physically motivated AWSEM contact potentials, i.e., $\lambda_c V_{\text{contact}}$ in eq S1 of the SI), so the emergence of a basin of attraction in calculated free-energy landscapes should be attributed mostly to the original physical potentials. Most of the features seen in the simulation also appear in simulations completely lacking the nonadditive Gō term, albeit with somewhat less clarity (see Figure S1). Details of the definition and the parametrization of the nonadditive Gō-potential are provided in the SI.

Reaction Coordinates and Free-Energy Calculations. We determine the free-energy profiles using reaction coordinates Q and R_{DNA} to monitor the disassembly of the histone core and the global DNA portion of the nucleosome, respectively. The fraction of native contacts Q is defined as

$$Q = \frac{2}{(N-2)(N-3)} \sum_{i<j-2} \exp\left[-\frac{(r_{ij} - r_{ij}^N)^2}{2\sigma^2}\right] \tag{1}$$

with $\sigma = 3$ Å and N being the total number of nontail residues from all eight histone proteins. The summation in eq 1 only includes C_α atoms. The native separation r_{ij}^N is the distance between the two C_α atoms from amino acids i and j calculated using the coordinates from the crystal structure.[36] Similar measures can be defined to study protein−protein interfaces when only intermolecular contacts are included in the summation. Q ranges from 0 to 1, with a higher value corresponding to greater similarity to the native structure. For funneled surfaces, Q has been shown to provide an excellent characterization of the progression of folding for single proteins[44,45] and binding for protein−protein complexes.[46] To study how the DNA becomes unwrapped when the nucleosome unfolds, we use the radius of gyration of the DNA R_{DNA} defined as

$$R_{\text{DNA}} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (\mathbf{r}_i - \mathbf{r}_{\text{com}})^2} \tag{2}$$

where $\mathbf{r}_{\text{com}}$ is the center of mass, and the summation is conducted over all the coarse-grained sugar beads of the DNA.

We used umbrella sampling together with replica exchange techniques to enhance conformation sampling for free-energy calculations.[47,48] Harmonic potentials $1/2K_q(Q - Q_o)^2$ and $1/2K_r(R_{\text{DNA}} - R^o_{\text{DNA}})^2$ were introduced to restrain constant temperature molecular dynamics simulations toward reference values, with $K_q = 1000$ kcal/mol and $K_r = 0.8$ kcal/mol/Å². The reference values for Q_o are equally spaced from 0.2 to 0.8 with a step size 0.1. For R^o_{DNA}, 12 references were chosen from 45 to 72.5 Å with an increment of 2.5 Å. Twelve replicas were used for each umbrella window with temperature ranging from 260 to 370 K with a step of 10 K. Data from different windows were stitched together with the weighted histogram analysis method (WHAM) to construct free-energy landscapes.[49]

Simulation Details. All simulations were performed using the software Large-Scale Atomic/Molecular Massively Parallel Simulator (LAMMPS). Initial configurations of the simulation and the DNA sequence are obtained from the crystal structure with PDB ID: 1KX5 (see Figure S2). Molecular dynamics trajectories were performed at constant temperature and volume without periodic boundary conditions for 5 million steps with a time step of 20 fs, and exchanges among different replicas were attempted at every 100 steps. Due to the coarse graining, we note the simulation time scale cannot be converted precisely into real time units.[50] In any event, we checked the convergence of the simulations by performing rigorous error analysis of the calculated free-energy profiles (see SI Section: Convergence of the Simulation for Details). Since the simulations were performed without periodic boundary conditions, we introduced a constraint on the radius of gyration of the histone core complex when the protein assembly is studied without the presence of the DNA in order to prevent molecules from diffusing too far away from each other unproductively. Definition of this constraint and additional simulation details are provided in the SI.

■ RESULTS AND DISCUSSION

Free-Energy Landscape of the Assembly of the Histone Protein Core. Several coarse-grained models have already been used for the investigation of nucleosome dynamics.[40,41,51,52] In most of these studies, the histone core complex was restrained to having the octamer conformation found in the crystal structure because the models that were employed lack a transferable force field for protein molecules. These structural restraints prohibit large-scale conformational changes away from the crystal structure, whether they are artifactual and unphysical or physical and mechanistically required. In our view, allowing protein flexibility is essential because the histone octamer structure is unstable under physiological conditions in the absence of the DNA.[53−55] Partial disassembly of the histone core complex as the nucleosome unfolds has also been observed in single molecule FRET experiments.[19,56] Conformational flexibility of the histone core complex must therefore play a crucial role in nucleosome dynamics. We first investigate whether the histone core would assemble in the absence of the DNA using free-energy landscape analysis.

Figure 1 presents the free-energy profile as a function of the fraction of native contacts Q, with representative structures of the protein complex at various Q values shown at the top. There are three free-energy basins in this landscape at approximately $Q = 0.35$, 0.45, and 0.75, respectively. At $Q = 0.35$, the system has disassembled into a tetramer $(\text{H3-H4})_2$ and two H2A-H2B dimers, and no specific contacts between H2A-H2B and the tetramer are present. Throughout our
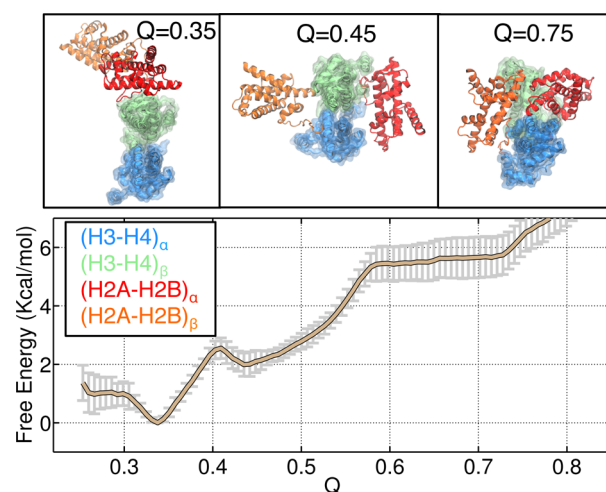

Figure 1. Free energy profile as a function of the fraction of native contacts Q for the folding of the histone protein core. Error bars shown in gray represent the standard deviation of the mean. Example configurations of the histone complex at various values of Q are shown in the top panel, with the two H3-H4 dimers drawn in blue and green and the two H2A-H2B dimers in red and orange. Histone tails are not displayed for clarity.

simulation, we have not observed complete dissociation of any of the four dimers into monomer structures. The stability of this low Q basin is sensitive to protein concentrations due to the entropic contributions from the free diffusion of proteins in the solution.[54] At $Q = 0.45$, the two dimers H2A-H2B begin to assemble around $(\text{H3-H4})_2$ from two sides, and the tetramer is seen to be sandwiched in between H2A-H2B dimers, as illustrated in the top panel. Finally, specific contacts form between the dimers and the tetramer at $Q = 0.75$, and the histone core complex adopts the conformation captured in the nucleosome crystal structure that contains DNA. Average contact maps of protein structures at various Q values are provided in Figure S3.

Mechanistic insight about the assembly process can be obtained by following the formation of interfacial contacts as a function of Q. As shown in Figure 2A, the formation of contacts between the two H3-H4 dimers (blue) precedes the formation of contacts between H3-H4 and H2A-H2B (red and yellow). The partially disassembled state at small Q values thus consists of a $(\text{H3-H4})_2$ tetramer and two H2A-H2B dimers. As the two distinct interfaces between H3-H4 and H2A-H2B are chemically identical, it is reassuring that their average number of contacts are found to be the same within numerical accuracy.

The attachment of the two H2A-H2B dimers to the $(\text{H3-H4})_2$ tetramer occurs largely sequentially, as shown in Figure 2B. The free-energy profile as a function of the two interfacial contact numbers exhibits two parallel reaction channels from the disassembled state to the octamer conformation, as indicated by the arrows. Along either one of the reaction pathways, the formation of each of the two interfaces is decoupled from the formation of the other, suggesting two parallel serial mechanisms.

The molecular picture for the assembly of histone proteins revealed from our simulation is consistent with prior experimental observations. For example, in agreement with Figure 2A, the histone proteins are known to stabilize into a $(\text{H3-H4})_2$ tetramer and two H2A-H2B dimers when disassembled.[53−55,57] Furthermore, in support of the sequential
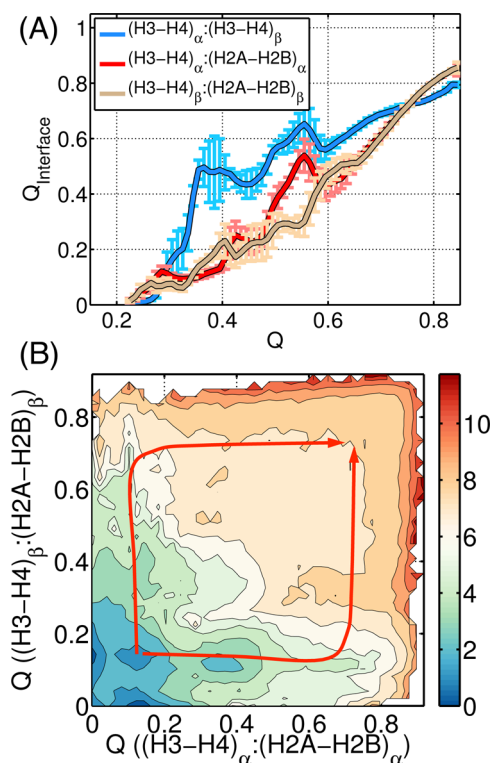
Figure 2. Formation of different protein–protein interfaces as the histone protein core assembles. (A) Average fraction of native contacts for various protein–protein interfaces as a function of the global Q. (B) Two-dimensional free-energy profile for the two chemically identical interfaces formed between H3-H4 and H2A-H2B hetero-dimers. The arrows indicate the two parallel reaction channels for folding. Standard deviation is provided in Figure S4.

Figure 3. Free-energy profile as a function of the DNA radius of gyration R_{DNA} for the unfolding of the nucleosome (yellow). Error bars shown in gray represent the standard deviation of the mean. The blue line measures the average number of DNA base pairs bound to histone proteins. Examples of nucleosome configurations at various values of R_{DNA} are shown in the top panel, with the DNA colored in yellow and the same coloring scheme as in Figure 1 for proteins.

pathway shown in Figure 2B, an intermediate hexameric structure consisting of the H3-H4 tetramer and one copy of the H2A-H2B dimer has been observed in ref 54.

Free-Energy Landscape of Full Nucleosome Disassembly. The free-energy landscape for the histone core complex quantifies the stability of the octamer state and provides a detailed molecular pathway for the assembly process. The remarkable agreement between the molecular mechanism predicted from simulation and that proposed from experimental observations on the histone complex by itself encourages the application of AWSEM for studying full nucleosome disassembly. We therefore now turn to investigate the coupling between DNA unwrapping and the conformational changes of the histone core complex as the nucleosome unfolds.

Figure 3 presents the free-energy profile (yellow) as a function of the radius of gyration of the DNA (R_{DNA}). Unfolded nucleosome conformations with exposed DNA can be seen from example snapshots shown in the top panel together with Figures S5 and S6. The free-energy landscape exhibits a single basin around the crystal structure at R_{DNA} = 45 Å, where the DNA is tightly bound to histone proteins. The model thus reproduces the expected stability of the nucleosome as a packaging unit for the genome. The free-energy cost of unwrapping the outer layer of the DNA in intact nucleosomes has been estimated to be around 7 to 10 kcal/mol.[17,18,58] As detailed in the SI (Section: Thermodynamics and Kinetics of DNA Unwrapping), by carefully defining the state in which the outer layer DNA has been unwound, our simulation predicts the free-energy cost for unwinding the DNA to be 8 kcal/mol.
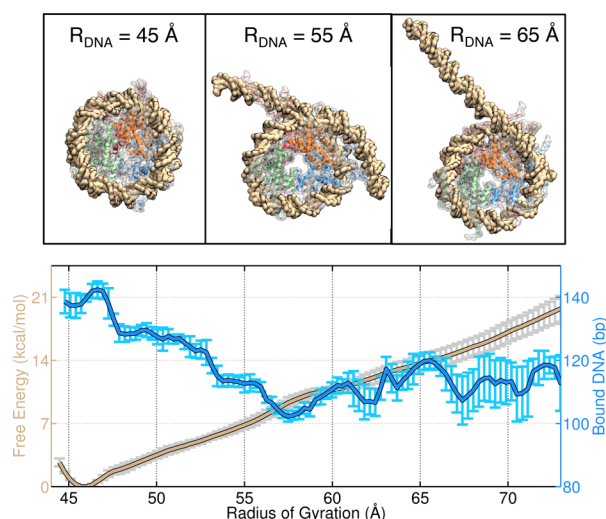
Furthermore, using a diffusion constant of $D = 5500$ bp^2/s estimated experimentally,[59] we find the rate for unwrapping the outer layer DNA for the intact nucleosome to be approximately 3.6×10^{-4} s^{-1}, which is in good agreement with reported rate 0.00038 s^{-1} from single molecule pulling experiments.[18]

The average number of DNA base pairs bound to histone proteins (see the SI for a rigorous definition) as a function of R_{DNA} is also shown in Figure 3 as a blue curve. As R_{DNA} increases, we find that the number of bound DNA base pairs changes in a stepwise manner. For example, most of the DNA base pairs remain bound over the range of extension 45 < R_{DNA} < 47.5 Å, which is followed by a sudden drop of ~10 bp, followed again by another plateau region 48 < R_{DNA} < 52 Å. The stepwise unwinding is even clearer at lower temperature, as shown in Figure S7. Step-wise DNA unwinding is expected from the periodic contacts that form between histone proteins and the DNA at a 10−11 bp frequency;[36,57,60,61] see SI Section: Periodicity of Histone DNA Contacts for a detailed discussion.

Figure 4A presents a two-dimensional free-energy landscape as a function of the DNA radius of gyration R_{DNA} and the fraction of native contacts Q. At small R_{DNA} when the DNA is fully bound, the histone core complex is highly stable around the octamer conformation that is captured in the crystal structure with $Q \sim 0.8$. As the DNA unwraps at large R_{DNA}, the free-energy of the low and high Q conformations become comparable, and the histone proteins begin to fall away from the core and begin to deviate from the octamer X-ray structure. Figure 4B further characterizes in detail the unfolding of various protein–protein interfaces. The interface between the two copies of H3-H4 forming the tetramer remains stable throughout the entire range of R_{DNA} studied. On the other hand, the two interfaces between H3-H4 and H2A-H2B gradually disappear as the nucleosome unfolds. We note the loss of the interface between H3-H4 and H2A-H2B is consistent with the stability of different interfaces determined from the free-energy landscape for histone core assembly shown in Figure 2. The coupling between histone disassembly
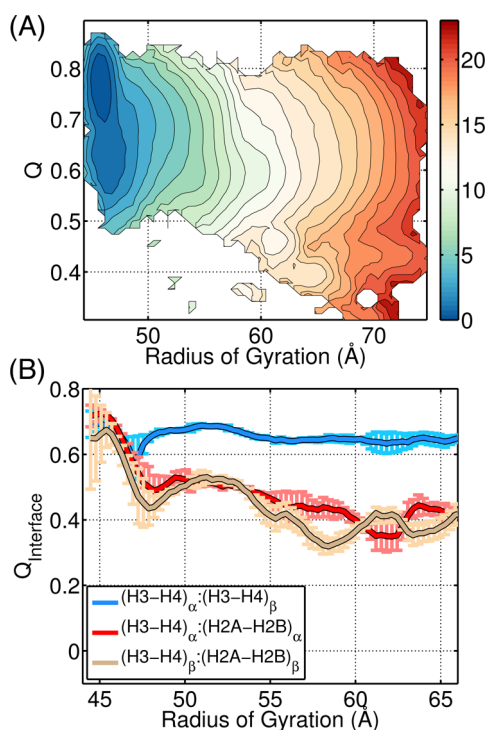
Figure 4. Coupling between DNA unwrapping and histone protein core disassembling. (A) Two-dimensional free-energy profile as a function of the DNA radius of gyration (R_{DNA}) and the fraction of native contacts for the histone protein core (Q). Energies in kcal/mol. (B) Average fraction of native contacts for various protein–protein interfaces as a function of R_{DNA}.

Figure 5. Asymmetric unwrapping of the two DNA ends as the nucleosome unfolds. (A) Example asymmetric nucleosome conformations with the first DNA segment colored in purple and the second in yellow. The coloring scheme for proteins is identical to Figure 1. (B) Two-dimensional free-energy profile for the radius of gyration of the two chemically identical DNA segments separated at the nucleosome dyad. (C) Average protein–DNA electrostatic interactions as a function of the DNA radius of gyration R_{DNA} and the asymmetry measure ξ. Energy scales in both part (B) and (C) are kcal/mol.

and DNA unwinding revealed in Figure 4 suggests that histone proteins will remain bound to the DNA by disrupting protein–protein interfaces at large R_{DNA}, thus explaining the plateau of average bound DNA base pairs shown in Figure 3.

As illustrated in the top panel of Figure 3B, the DNA molecule unwraps asymmetrically, and only one of two ends is preferentially exposed at large R_{DNA}. Figure 5 further consolidates this observation with a free-energy landscape as a function of the R_{DNA} for each of the two equally divided DNA segments separated at the nucleosome dyad. This landscape clearly illustrates that the energetic cost of opening the two ends simultaneously, i.e., moving along the diagonal of the landscape, is much higher than opening only one end following the pathways highlighted with arrows. Two examples of sampled half open nucleosome structures are shown in Figure 5A, with the two segments of the DNA colored in yellow and purple, respectively. It is important to point out that the short DNA sequence employed in our simulation is palindromic, and thus the two segments of the DNA are identical chemically. There is thus no preference for either one DNA end to unwrap first or the other, as reflected in the symmetry of the two reaction channels on the free-energy landscape.

The asymmetric DNA conformation having only one of its two ends unwrapped allows the other end to interact more favorably with histone proteins. This energetic preference can be seen from Figure 5C, which plots the average protein–DNA electrostatic interaction energy of various nucleosome conformations. The y-axis of this figure is a measure for DNA asymmetry defined as $\xi = (R_{DNA}^{first})/(R_{DNA}^{first} + R_{DNA}^{second})$. This definition is motivated by the observation that the unwrapped end has larger radius of gyration. From Figure 5C, we see that
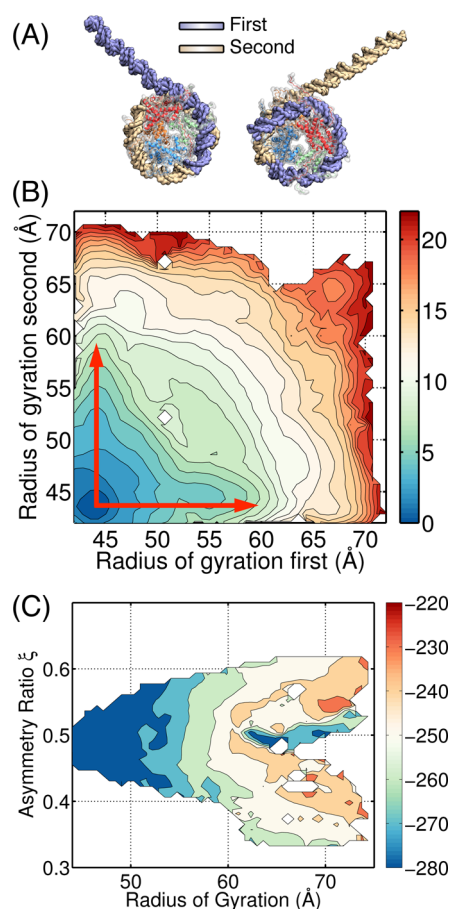
as we pull the DNA apart, for R_{DNA} between 58 and 62 Å, the protein–DNA electrostatic interaction energy is indeed lower either for $\xi < 0.5$ or for $\xi > 0.5$ compared with the symmetric configuration with $\xi \sim 0.5$.

Sampling becomes difficult at large distances, but we note that at the $R_{DNA} = 62$ Å, a new set of configurations, in which the DNA folds back onto itself along with a completely dissociated histone protein core occasionally appears in the simulation (see Figure S8). These configurations also have a low electrostatic energy. This multiplicity of observed structures at large R_{DNA} suggests that in vivo unwrapping is likely to be mechanically coupled to large scale DNA motions that are promoted by motor proteins.[62]

The detailed molecular model for nucleosome disassembly put forward by our simulation is well supported with experimental single molecule studies. For example, the predicted loss of the H3-H4 tetramer/H2A-H2B dimer interface along with DNA unwinding was indeed observed in single molecule FRET experiments.[19,56] Similarly, the sequential asymmetric unwrapping of the two DNA ends has been detected in recent single molecule pulling experiments.[20]

Effect of Histone Tails on Nucleosome Stability. The terminal regions of histone proteins, often called "histone tails" due to their lack of pronounced secondary structure, are known to play a crucial role in nucleosome stability and gene regulation.[16] Histone tails are highly positively charged and bind tightly with the DNA, thus keeping the genomic content from being exposed. The importance of histone tails for gene regulation has been demonstrated by showing that a variety of covalent modifications of the charged residues in the tails in vivo, including acetylations and methylations, provide a delicate scheme of regulation along the developmental course of higher organism.[63] We now investigate the effect of these histone tails on nucleosome stability in our model and study how the unfolding mechanism is modified by computing the free-energy landscape of a tailless nucleosome.

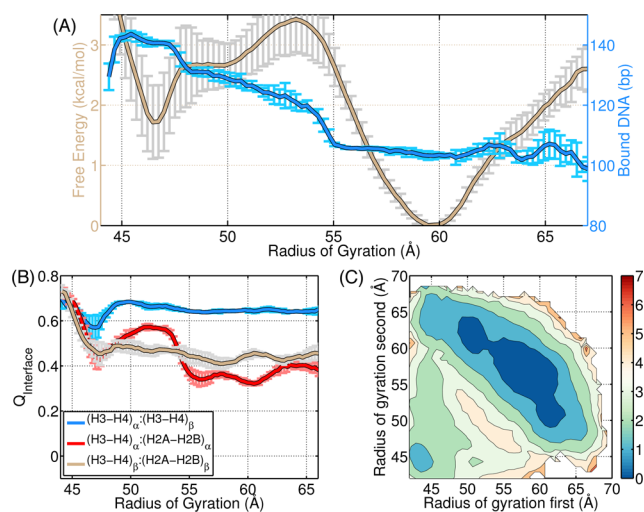Figure 6A presents the free-energy profile for DNA unwrapping from the tailless nucleosome. The difference

Figure 6. Energy landscape analysis for the tailless nucleosome. (A) Free-energy profile as a function of the radius of gyration R_{DNA} for the unfolding of the tailless nucleosome (yellow). The blue line measures the average number of DNA base pairs bound to histone proteins. (B) Average fraction of native contacts for various protein−protein interfaces as a function of R_{DNA}. (C) Two-dimensional free-energy profile for the radius of gyration of the two chemically identical DNA segments separated at the nucleosome dyad. Energies in kcal/mol.

between this profile and that for the intact nucleosome shown in Figure 3 is dramatic. Besides the free-energy minima near the crystal structure at R_{DNA} = 45 Å, an additional basin emerges at large R_{DNA} toward conformations with partially unwrapped DNA. The average number of bound base pairs as a function of R_{DNA} shown in blue indicates that around 40 base pairs are exposed at the basin with R_{DNA} = 60 Å. Just as for the intact nucleosome, for the tailless nucleosome, disassembly of the histone proteins is observed as the DNA unwinds. As shown in Figure 6B, the two H2A-H2B dimer/H3-H4 tetramer interfaces are lost at large R_{DNA} while the tetramer itself remains stable.

For the tailless nucleosome, the preference for asymmetric DNA unwinding is seen to be less prominent when we compare the configurations of this system with those of the intact nucleosome as it unwinds. As shown in Figure 6C, the free-energy costs of the symmetric and the asymmetric DNA unwinding pathways are comparable. Similarly, the electrostatic energies for symmetric and asymmetric DNA conformations do

not differ significantly (see Figure S9). Therefore, electrostatic interactions among the histone tails and the DNA seem to play a crucial role in stabilizing the asymmetric pathway.

In analyzing the results of the intact histone, we try to dissect the contribution from each histone tail. The individual contributions from each histone tail have been computed using free-energy perturbation theory. We find that the H3 tails have the most profound effect, as shown in Figure S10. Quantitative results are discussed in the SI Section: Effect of Individual Histone Tails on Nucleosome Stability.

The significant drop in the energetic cost of unwinding the DNA when comparing the free-energy landscapes shown in Figure 6 and Figure 3 predicts a faster DNA unwrapping rate for tailless nucleosomes. Recent single molecule experiments indeed confirm such a prediction and report a significant enhancement when comparing the unwrapping rate of the outer layer DNA between tailless and intact nucleosomes.[64] Furthermore, the significant drop in free-energy cost also explains the reduction of the binding affinity between the outer layer DNA and the histone octamer for the tailless nucleosome, as detected by Wang and co-workers using force−extension curves from single molecule pulling experiments.[65]

■ CONCLUSIONS

We have introduced a chemically accurate and computationally efficient coarse-grained protein−DNA model to investigate quantitatively the molecular mechanism of nucleosome disassembly. Proteins in this model are described using the AWSEM force field, the parameters of which were determined from statistical optimization algorithms that sculpt a funnelled energy landscape for natural proteins in the known structural database. AWSEM has been applied successfully in previous studies to predict monomer structures and dimeric protein− protein interfaces. Here we have extended its application to large protein complexes and demonstrated that AWSEM accurately describes the thermodynamic pathways for the assembly of the histone octamer. Intermediate states observed along these simulated pathways, in which interfacial contacts between H2A-H2B dimers and the (H3-H4)₂ tetramer are only partially formed, have indeed been indicated in prior experimental studies.[54]

When combined with the 3SPN.2C DNA model, AWSEM not only quantitatively reproduces the thermodynamics and kinetics of nucleosome unfolding but also provides a detailed molecular picture for the nucleosome assembly pathway. The coupling between DNA unwinding and histone core disassembly observed in the simulations has been suggested previously by single molecule FRET experiments.[19] The asymmetric DNA unwrapping predicted from our simulation is consistent with the recent single molecule pulling experiment on short DNA segment as well.[20] The qualitative and quantitative agreement between simulation and experiment for single nucleosomes demonstrates the usefulness of this model in studying nucleosomal dynamics and makes us look forward to its application for simulation studies of large nucleosomal assemblies like those in the intact chromosome.

■ ASSOCIATED CONTENT

⑤ Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/jacs.6b02893.

Detailed simulation protocols and supplementary trajectory analysis (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*pwolynes@rice.edu

Notes

The authors declare no competing financial interest.

■ REFERENCES

(1) Bickmore, W. A.; van Steensel, B. *Cell* **2013**, *152*, 1270−1284.
(2) Gorkin, D.; Leung, D.; Ren, B. *Cell Stem Cell* **2014**, *14*, 762−775.
(3) Zhang, B.; Wolynes, P. G. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 10185−10190.
(4) Sanyal, A.; Lajoie, B. R.; Jain, G.; Dekker, J. *Nature* **2012**, *489*, 109−113.
(5) Rao, S. P.; Huntley, M.; Durand, N.; Stamenova, E.; Bochkov, I.; Robinson, J.; Sanborn, A.; Machol, I.; Omer, A.; Lander, E.; Aiden, E. *Cell* **2014**, *159*, 1665−1680.
(6) Workman, J. L.; Kingston, R. E. *Annu. Rev. Biochem.* **1998**, *67*, 545−579.
(7) Andrews, A. J.; Luger, K. *Annu. Rev. Biophys.* **2011**, *40*, 99−117.
(8) Davtyan, A.; Schafer, N. P.; Zheng, W.; Clementi, C.; Wolynes, P. G.; Papoian, G. A. *J. Phys. Chem. B* **2012**, *116*, 8494−8503.
(9) Schafer, N. P.; Kim, B. L.; Zheng, W.; Wolynes, P. G. *Isr. J. Chem.* **2014**, *54*, 1311−1337.
(10) Hinckley, D. M.; Freeman, G. S.; Whitmer, J. K.; de Pablo, J. J. *J. Chem. Phys.* **2013**, *139*, 144903.
(11) Freeman, G. S.; Hinckley, D. M.; Lequieu, J. P.; Whitmer, J. K.; de Pablo, J. J. *J. Chem. Phys.* **2014**, *141*, 165103.
(12) Jiang, C.; Pugh, B. F. *Nat. Rev. Genet.* **2009**, *10*, 161−172.
(13) Szerlong, H. J.; Hansen, J. C. *Biochem. Cell Biol.* **2011**, *89*, 24−34.
(14) van Holde, K. *Chromatin (Springer Series in Molecular Biology)*; Springer-Verlag: New York, 1988.
(15) Becker, P. B.; Horz, W. *Annu. Rev. Biochem.* **2002**, *71*, 247−273.
(16) Bowman, G. D.; Poirier, M. G. *Chem. Rev.* **2015**, *115*, 2274−2295.
(17) Brower-Toland, B. D.; Smith, C. L.; Yeh, R. C.; Lis, J. T.; Peterson, C. L.; Wang, M. D. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 1960−1965.
(18) Mihardja, S.; Spakowitz, A. J.; Zhang, Y.; Bustamante, C. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 15871−15876.
(19) Bohm, V.; Hieb, A. R.; Andrews, A. J.; Gansen, A.; Rocker, A.; Toth, K.; Luger, K.; Langowski, J. *Nucleic Acids Res.* **2011**, *39*, 3093−3102.
(20) Ngo, T. T.; Zhang, Q.; Zhou, R.; Yodh, J. G.; Ha, T. *Cell* **2015**, *160*, 1135−1144.
(21) Ettig, R.; Kepper, N.; Stehr, R.; Wedemann, G.; Rippe, K. *Biophys. J.* **2011**, *101*, 1999−2008.
(22) Winogradoff, D.; Zhao, H.; Dalal, Y.; Papoian, G. A. *Sci. Rep.* **2015**, *5*, 17038.

(23) Shaytan, A. K.; Armeev, G. A.; Goncearenco, A.; Zhurkin, V. B.; Landsman, D.; Panchenko, A. R. *J. Mol. Biol.* **2016**, *428*, 221−237.
(24) Winogradoff, D.; Echeverria, I.; Potoyan, D. A.; Papoian, G. A. *J. Am. Chem. Soc.* **2015**, *137*, 6245−6253.
(25) Materese, C. K.; Savelyev, A.; Papoian, G. A. *J. Am. Chem. Soc.* **2009**, *131*, 15005−15013.
(26) Zhang, B.; Miller, T. F. *J. Am. Chem. Soc.* **2012**, *134*, 13700−13707.
(27) Wang, C. Y.; Miller, T. F. *J. Biol. Chem.* **2014**, *289*, 30868−30879.
(28) Zhang, B.; Miller, T. F. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 5399−5404.
(29) Zhang, B.; Miller, T. *Cell Rep.* **2012**, *2*, 927−937.
(30) Van Lehn, R. C.; Zhang, B.; Miller, I.; Thomas, F. *eLife* **2015**, *4*, e08697.
(31) Hardin, C.; Eastwood, M. P.; Luthey-Schulten, Z.; Wolynes, P. G. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97*, 14235−14240.
(32) Eastwood, M. P.; Hardin, C.; Luthey-Schulten, Z.; Wolynes, P. G. *J. Chem. Phys.* **2002**, *117*, 4602−4615.
(33) Hardin, C.; Eastwood, M. P.; Prentiss, M. C.; Luthey-Schulten, Z.; Wolynes, P. G. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 1679−1684.
(34) Papoian, G. A.; Ulander, J.; Eastwood, M. P.; Luthey-Schulten, Z.; Wolynes, P. G. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 3352−3357.
(35) Zheng, W.; Schafer, N. P.; Davtyan, A.; Papoian, G. A.; Wolynes, P. G. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 19244−19249.
(36) Davey, C. A.; Sargent, D. F.; Luger, K.; Maeder, A. W.; Richmond, T. J. *J. Mol. Biol.* **2002**, *319*, 1097−1113.
(37) Terakawa, T.; Kenzaki, H.; Takada, S. *J. Am. Chem. Soc.* **2012**, *134*, 14555−14562.
(38) Vuzman, D.; Levy, Y. *Isr. J. Chem.* **2014**, *54*, 1374−1381.
(39) Potoyan, D. A.; Zheng, W.; Komives, E. A.; Wolynes, P. G. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113*, 110−115.
(40) Kenzaki, H.; Takada, S. *PLoS Comput. Biol.* **2015**, *11*, e1004443.
(41) Freeman, G. S.; Lequieu, J. P.; Hinckley, D. M.; Whitmer, J. K.; de Pablo, J. J. *Phys. Rev. Lett.* **2014**, *113*, 168101.
(42) Tsai, M.-Y.; Zheng, W.; Balamurugan, D.; Schafer, N. P.; Kim, B. L.; Cheung, M. S.; Wolynes, P. G. *Protein Sci.* **2016**, *25*, 255−269.
(43) Eastwood, M. P.; Wolynes, P. G. *J. Chem. Phys.* **2001**, *114*, 4702−4716.
(44) Cho, S. S.; Levy, Y.; Wolynes, P. G. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 586−591.
(45) Schafer, N. P.; Hoffman, R. M. B.; Burger, A.; Craig, P. O.; Komives, E. A.; Wolynes, P. G. *PLoS One* **2012**, *7*, e50635.
(46) Levy, Y.; Wolynes, P. G.; Onuchic, J. N. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 511−516.
(47) Sugita, Y.; Okamoto, Y. *Chem. Phys. Lett.* **1999**, *314*, 141−151.
(48) Souaille, M.; Roux, B. *Comput. Phys. Commun.* **2001**, *135*, 40−57.
(49) Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. *J. Comput. Chem.* **1992**, *13*, 1011−1021.
(50) Takada, S.; Kanada, R.; Tan, C.; Terakawa, T.; Li, W.; Kenzaki, H. *Acc. Chem. Res.* **2015**, *48*, 3026−3035.
(51) Dobrovolskaia, I. V.; Arya, G. *Biophys. J.* **2012**, *103*, 989−998.
(52) Fan, Y.; Korolev, N.; Lyubartsev, A. P.; Nordenskiold, L. *PLoS One* **2013**, *8*, e54228.
(53) Eickbush, T. H.; Moudrianakis, E. N. *Biochemistry* **1978**, *17*, 4955−4964.
(54) Ruiz-Carrillo, A.; Jorcano, J. L. *Biochemistry* **1979**, *18*, 760−768.
(55) Ruiz-Carrillo, A.; Jorcano, J. L.; Eder, G.; Lurz, R. *Proc. Natl. Acad. Sci. U. S. A.* **1979**, *76*, 3284−3288.
(56) Gansen, A.; Valeri, A.; Hauger, F.; Felekyan, S.; Kalinin, S.; Toth, K.; Langowski, J.; Seidel, C. A. M. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 15308−15313.
(57) Luger, K.; Mader, A. W.; Richmond, R. K.; Sargent, D. F.; Richmond, T. J. *Nature* **1997**, *389*, 251−260.
(58) Kruithof, M.; van Noort, J. *Biophys. J.* **2009**, *96*, 3708−3715.
(59) Mochrie, S. G.; Mack, A. H.; Schlingman, D. J.; Collins, R.; Kamenetska, M.; Regan, L. *Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys.* **2013**, *87*, 012710.

(60) Hall, M. A.; Shundrovsky, A.; Bai, L.; Fulbright, R. M.; Lis, J. T.; Wang, M. D. *Nat. Struct. Mol. Biol.* **2009**, *16*, 124−129.

(61) Chereji, R. V.; Morozov, A. V. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 5236−5241.

(62) Sheinin, M. Y.; Li, M.; Soltani, M.; Luger, K.; Wang, M. D. *Nat. Commun.* **2013**, *4*, 2579.

(63) Voigt, P.; Tee, W.-W.; Reinberg, D. *Genes Dev.* **2013**, *27*, 1318−1338.

(64) Bintu, L.; Ishibashi, T.; Dangkulwanich, M.; Wu, Y.-Y.; Lubkowska, L.; Kashlev, M.; Bustamante, C. *Cell* **2012**, *151*, 738−749.

(65) Brower-Toland, B.; Wacker, D. A.; Fulbright, R. M.; Lis, J. T.; Kraus, W. L.; Wang, M. D. *J. Mol. Biol.* **2005**, *346*, 135−146.