Arc-Repressor Dimerization on DNA: Folding Rate Enhancement by Colocalization

Amir Marcovitz and Yaakov Levy*

Department of Structural Biology, Weizmann Institute of Science, Rehovot, Israel

ABSTRACT   Multimeric proteins are ubiquitous in many cellular processes that require high levels of regulation. Eukaryotic gene expression is often regulated by a mechanism of combinatorial control that involves the binding of dimeric transcription factors to DNA together with the coordinated activity of additional proteins. In this study, we investigated the dimerization of the Arc-repressor on DNA with the aim of achieving microscopic insight into the possible advantages of interacting with DNA as a complex rather than as a monomeric single-domain protein. We used a computational coarse-grained model in which the protein dynamics was governed by native interactions and protein-DNA interactions were dictated by electrostatic forces. Inspired by previous experimental work that showed an enhanced refolding rate for the Arc-repressor in the presence of DNA and other polyanions, we focused on the mechanism and kinetics of the assembly of Arc monomers in the presence of single- (ssDNA) and double-stranded DNA (dsDNA) molecules in a low-salt concentration environment. The electrostatic interactions that attract the protein to the dsDNA were shown to be fundamental in colocalizing the unfolded Arc chains and in accelerating refolding. Arc monomers bind the dsDNA efficiently and nonspecifically, and search for each other via one-dimensional diffusion. The fastest folding of Arc is observed for DNA of 30 bp. Longer DNA is significantly less efficient in accelerating the Arc refolding rate, since the two subunits search distinct regions of the one-dimensional DNA and are therefore much less colocalized. The probability that the two unfolded chains will meet on 200 bp DNA is similar to that in the bulk. The colocalization of Arc subunits on ssDNA results in much faster folding compared to that obtained on dsDNA of the same length. Differences in the rate of Arc refolding, cooperativity, and the structure of its transition state ensemble introduced by ssDNA and dsDNA molecules demonstrate the important role of colocalization in biological self-assembly processes.

INTRODUCTION

Protein self-association to form dimers and higher-order oligomeric assemblies is ubiquitous in many cellular processes. In addition to providing structural and functional advantages, such as higher stability and specificity, protein assemblies expand the opportunities for regulation. For example, dimerization and oligomerization provide combinatorial control (1–4), form interfaces between adjacent subunits (which are potential sites for allostery), and may control activation and inhibition of enzymatic activity (5,6).

Regulation of gene expression requires synchronous binding of several different proteins to DNA to achieve appropriate repression or activation of the downstream gene. This type of transcriptional regulation is believed to be a key factor in the more complex regulatory networks in higher eukaryotes (6,7). Many transcription factors often form dimeric complexes with DNA in which the two monomers interact with each other while each recognizes the DNA. These dimeric proteins may exist as monomers or dimers in the absence of DNA. The rate and mechanism of such a 2:1 complex assembly (protein-protein-DNA) influence the identity and order of addition of other transcriptional proteins and therefore determine the composition and function of the supramolecular complex that forms at the promoter (8).

The assembly of the 2:1 complex requires a ternary collision (two protein monomers and DNA) that is relatively unlikely and therefore probably occurs through a two-step process. Two pathways may describe the mechanism of these complex assemblies: the protein may bind to the DNA as a dimer after the assembly of the two monomers (dimeric pathway), or each monomeric subunit may bind the DNA independently and associate with the other subunit while bound to the DNA (monomeric pathway) (8). It was elegantly shown in a series of experimental studies by the Schepartz group that complexes formed by the basic region leucine zipper (bZIP) and the basic region Helix-Loop-Helix zipper (bHLHZip) follow the monomeric pathway (8–13). The mediation of folding and assembly of multimeric DNA binding proteins by DNA was reported by other groups as well (14–19).

A proposed physical explanation for the dominance of the monomeric pathway is based on electrostatic guidance in the early stage of the reaction as both the first and second steps in the pathway are promoted by strong electrostatic interactions between protein and DNA. Another fundamental aspect in DNA binding is the flexibility of the protein, which was recently shown to be synergistically coupled to the electrostatic steering effect in the binding process of the Ets-DNA complex (20). Protein flexibility can facilitate binding through fly-casting mechanisms (21). A relatively unstructured protein has a greater capture radius for a specific binding site than does a folded protein, which is much more restricted in its conformational freedom.
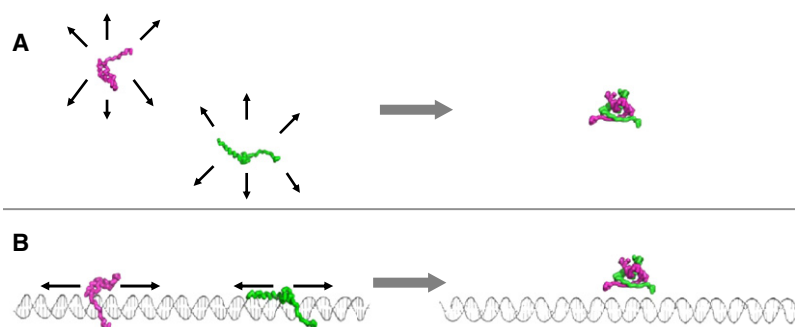
FIGURE 1 Nucleic acid effect on the Arc-repressor refolding. (*A*) In the absence of DNA, the molecular recognition process of the unfolded chains is performed in 3D space. The relatively slow refolding kinetics corresponds to high translational and rotational degrees of freedom and the need for an extensive search for a chain-chain association. (*B*) In the presence of DNA, the molecular recognition search space is shifted from 3D to 1D. The localization of the two chains significantly reduces the translational and rotational entropy of the Arc subunits.

Electrostatic interactions and other fundamental forces, such as compartmentalization and hydrophobic binding, target proteins to specific locations in the cell where they are colocalized with other proteins (22). The highly charged and complementary interface of the protein-DNA complex can steer the protein to the DNA. Furthermore, the tidal electrostatic force can also accelerate folding of monomeric proteins if, for example, the unfolded protein is attracted to the DNA and consequently becomes more structured. When nonspecific binding to DNA alters the nature of the unfolded state, the protein thermodynamics and kinetics may be affected. A thorough investigation of the effect of DNA on protein folding was performed by Rentzeperis and co-workers (23) on the Arc-repressor (Arc), which is a two-state homodimer. They reported that the refolding rate of the Arc dimer can be accelerated 30-fold or more by negatively charged polymers, including double- (dsDNA) and single-stranded DNA (ssDNA), but not by neutral or positively charged polymers. The kinetics of enhanced refolding of Arc are consistent with a model in which unfolded Arc monomers bind rapidly and nonspecifically to the polyanion and complete the formation of the folded dimer in the bound state. Rentzeperis et al. also showed that the refolding rate of Arc strongly depends on the concentration of ssDNA and dsDNA as well as on the ionic strength. In particular, they observed that Arc refolding on ssDNA at low ionic strength increases until an optimal DNA concentration is reached, but decreases for higher DNA concentration.

In this study, we used coarse-grained models to explore the molecular details of bimolecular refolding of Arc in the presence of dsDNA and ssDNA molecules in comparison with its folding characteristics in the bulk (Fig. 1). We applied electrostatic forces between the protein and the DNA to simplify the elusive nature of the nonspecific protein-DNA interactions. We aimed to provide microscopic details of the kinetics and thermodynamics of Arc refolding under the effects of varying lengths of ssDNA or dsDNA molecules, and to complement the macroscopic experimental results of the effect of DNA concentration on Arc refolding. Furthermore, we addressed the detailed mechanism of Arc refolding by exploring the existence of a monomeric pathway in this system, investigating the cooperativity of folding, and analyzing the transition state ensemble.

MATERIALS AND METHODS

Arc-repressor

P22 Arc-repressor is a small homodimeric β-sheet DNA binding protein (24,25) (Protein Data Bank (PDB) identification code 1ARR). Each monomeric chain contains 11 positively charged amino acid residues (Lys and Arg) and seven negatively charged residues (Asp and Glu). The Arc-repressor is a so-called two-state dimer, since no intermediate is detected during its folding under equilibrium conditions (26) and it shows a sharp transition from two unfolded subunits to a folded dimer (27–29).

Model

To allow a long simulation timescale such that several transitions of folding-unfolding or DNA binding-unbinding events could be sampled, we studied the Arc-DNA systems using a coarse-grained model. The protein was described as "beads on a string" where each amino acid was represented by a single bead centered at the Cα position. The DNA was represented by three beads for each nucleotide: the phosphate group (*P*), the ribose sugar group (*S*), and the base (*B*). Each bead was located at the geometric center of the group it represented. Positively charged residues (Lys and Arg) were assigned a point charge of $(+1)$ and negatively charged residues (Asp and Glu), as well as the phosphate beads, were assigned a negative charge of (-1).

We simulated the system with a native topology-based model (the Gō-model) in which native protein interactions are attractive and all other interactions are repulsive. Native topology-based models correspond to an unfrustrated model with a perfectly funneled energy landscape where the native state is dominant and unique (30–32). Several studies have shown that, owing to the minimal frustration principle, binding mechanisms are robust and, just as for protein folding, are governed primarily by the protein's native topology (28,29,33–41). In particular, the folding mechanism of Arc-repressor was shown to be well captured using the native topology-based model (28,29,40).

The native protein interactions were modeled by a Lennard-Jones potential without any discrimination based on the chemical properties of the interactions. Nonnative protein-protein and protein-DNA interactions were represented by the hard-sphere repulsion term $(\sigma_{i,j}/r_{i,j})^{12}$, where $\sigma_{i,j} = 4$ Å for a Cα-to-Cα collision, and $\sigma_{i,j} = 5.7$ Å for a Cα-to-DNA bead collision. The finer details of the Hamiltonian that describes the protein, including the bonds and angular degrees of freedom, can be found elsewhere (20,42,43). The electrostatic potential between charged beads (44,45) q_i, q_j was modeled by the Debye-Hückel interaction, which accounts for the ionic strength of a solute immersed in aqueous solution (46), Simulations (with 10^9 time steps) were performed with a salt concentration of 0.01 M and a dielectric constant of 70. The Debye-Hückel interaction has been used to study the energetics and dynamics of various biomolecular systems (45,47–50), and was recently used to obtain dynamics and structural characterization of proteins sliding along DNA in various salt concentrations (51). Although the model successfully introduces the salt effect of the screening electrostatic interactions to the coulomb pairwise interactions, it is valid mainly for dilute
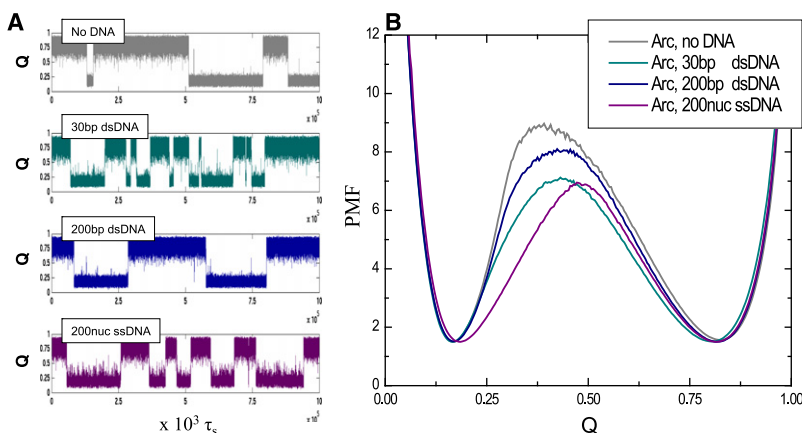
FIGURE 2 Enhancement of the Arc refolding rate by dsDNA and ssDNA molecules. (A) Typical T_f trajectories of the folding of the Arc-repressor simulated without DNA and with molecules of 30 bp dsDNA, 200 bp dsDNA, and 200 nuc ssDNA. The two-state Arc dimer undergoes folding-unfolding transitions in which two unfolded chains concomitantly fold and dimerize. The refolding rate of each system is also indirectly related to the number of transitions in the simulations of the same time period. (B) Potential of mean force (PMF) profiles as a function of Q (the fraction of native contacts). The free-energy barrier $\Delta G^{\ddagger}$ is the highest for refolding in the absence of DNA (gray) and the lowest for refolding in the 30 bp dsDNA (dark cyan) and 200 nuc ssDNA (purple) systems. The free-energy barrier of the 200 bp dsDNA (dark blue) lies in the middle.

solutions. Detailed effects of higher salt concentrations and condensation of cations on DNA have to be studied with the nonlinearized Poisson-Boltzmann equation.

The dynamics of the system was simulated by the Langevin equation, which introduces a dissipative drag force (with $\gamma = 0.01$) and an additional random force that represents stochastic collisions between the solvent molecules and the solute. The weighted histogram analysis method was executed on the multiple trajectories to obtain the thermodynamic values of the systems (52).

Arc refolding simulations with dsDNA

Nonspecific dsDNA B-DNA molecules of varying lengths (6, 15, 30, 50, 70, 150, 200 bp) were used to explore the effect of DNA length on Arc refolding. The protein and the DNA were placed in a box with the DNA molecule placed at the center of the box and aligned with the Z-axis. The dimensions of the box were $250 \times 250 \times L$ Å^3, where L is proportional to the length of the DNA molecule. Although the DNA was rigid and static during the simulations, the protein was flexible and freely diffused within the box and could undergo folding-unfolding and DNA binding-unbinding events. Only a nonspecific repulsive force was considered between the box boundaries and the protein beads, and the dimensions of the system were set such that any perturbations to the dynamics were negligible, if they occurred at all. As a control system, we calculated the effect of various length of DNA on the refolding rates of SH3, a non-DNA-binding protein; Antp, a DNA-binding protein; and the dimer of the p53 tetramerization domain, which is a non-DNA-binding protein.

Arc refolding simulations with ssDNA

To study the effects of DNA flexibility, the refolding of Arc was studied in the presence of ssDNA molecules of varying lengths (6, 15, 30, 50, 70, 150, and 200 bp) similar to those used to study the effect of dsDNA molecules. A cubic box was used to confine the system, and its dimensions were set to match the volume of the box used in the equivalent Arc + dsDNA simulations. A recent NMR study on the structural organization of ssDNA in aqueous solution showed a substantial population of right-handed helical structures in hexameric ssDNA (53), which is reminiscent of their helical conformation in the double-stranded complex. Yet, ssDNA and ssRNA molecules often form hairpin loops and other structures, and in several experimental and computational works investigators have endeavored to decipher the folding mechanism and structural characteristics of such structures (47,54–56). In contrast to the rigid and static representation of dsDNA in our simulation as described above, in this experiment the ssDNA molecule was flexible and represented as a self-avoiding chain in which the beads of the DNA backbone were covalently linked $[(P-S)_n]$ and the base was covalently linked to the ribose center. Bond angles and dihedral angles formed by the four beads $B_i S_i S_{i+1} B_{i+1}$ were considered as well. The values of the native bond lengths and angles were obtained from the PDB structure of the helical

structure that ssDNA possesses in the duplex form. In our simulations, we added a Lennard-Jones potential between bases that were at least three nucleotides apart from each other ($B_i–B_j$, $|i-j|>3$) and excluded all phosphate-phosphate electrostatic interactions. All other DNA-DNA and DNA-protein interactions were repulsive. The parameters of the Hamiltonian were calibrated such that the ssDNA molecule was relatively flexible but compact due to transient basepairing interactions. When we plotted the logarithm of the radius of gyration for the different lengths of ssDNA studied using our model versus the number of nucleotides, N, we obtained a slope of ~0.6 (a slope of 0.6 is typical of random coil polymers and of denatured proteins with residual structures (57,58)). For comparison, when the ssDNA model included only electrostatic repulsion between the phosphates and no transient basepairing was allowed, the slope was close to unity. We note that this type of modeling does not aim to accurately describe the features of ssDNA, but rather seeks to include its basic features (i.e., a dynamic polymer that is stabilized by transient interactions and has a smaller persistence length than dsDNA). The dynamic ssDNA molecule was kept at the center of the cubic box by transforming the coordinates of all the beads in each time step with respect to the displacement of the center of mass of the ssDNA from the origin.

Refolding kinetics measurements of Arc-repressor

The reversible two-state folding of Arc-repressor yields equilibrium trajectories with multiple transitions at the folding temperature, T_f. The refolding rate of Arc (k_f) was estimated by calculating the mean passage time for folding transition from the unfolded state at the respective T_f.

RESULTS AND DISCUSSION

dsDNA and ssDNA molecules accelerate Arc refolding

To explore the effect of DNA molecules on the refolding of Arc-repressor, we carried out long folding/binding simulations of the protein in the presence of dsDNA and ssDNA molecules of varying lengths and at a relatively low ionic strength of 0.01 M (see Materials and Methods section). In agreement with Rentzeperis et al. (23), we observed a significant enhancement of the Arc refolding rate in the presence of DNA. Fig. 2 A shows representative trajectories of the folding of Arc-repressor in the bulk (i.e., in the absence of DNA) and in the presence of 30 bp dsDNA, 200 bp dsDNA, and 200 nucleotides ssDNA molecules at their respective folding temperatures (T_f, defined as the temperature at which

the protein's stability (ΔG) equals zero). The number of folding-unfolding transitions is lowest in the absence of DNA and highest in the presence of the 30 bp dsDNA molecule. A mild enhancement of the refolding rate is observed in the presence of the 200 bp dsDNA molecule; however, a much more significant rate enhancement is observed in the presence of its equivalent 200 nuc ssDNA molecule.

The free-energy profiles of Arc in the bulk and in the presence of DNA, with Q (the fraction of native contacts) as a reaction coordinate, are shown in Fig. 2 B. Although the protein-DNA interactions are nonspecific (i.e., are governed by electrostatic forces only), the presence of dsDNA can remarkably affect the free-energy landscape of Arc folding. A substantial decrease in the free-energy barrier for folding, $\Delta G^{\ddagger}$, is seen in the presence of a 30 bp dsDNA molecule compared to refolding in the bulk. Increasing the dsDNA length from 30 bp to 200 bp diminishes the rate enhancement effect, yet the free-energy barrier of refolding in the 200 bp system remains lower than that observed in the bulk. In contrast, the free-energy profile of Arc refolding in the presence of 200 nuc ssDNA resembles that of 30 bp dsDNA, with equal heights of the free-energy barrier. This is consistent with the enhanced folding-unfolding kinetics for these systems observed in Fig. 2 A. A careful estimation of the transition state ensemble of the free-energy profiles reveals a shift toward higher values of Q in the presence of DNA.

The enhanced refolding rate of the dimer in the presence of a nucleic acid may be due to the unfolded Arc chains nonspecifically binding the DNA molecules. A dsDNA binding event reduces the three-dimensional (3D) search space for molecular recognition to a single dimension (Fig. 1), and thus may increase the probability that the Arc chains will encounter each other. Long dsDNA molecules, however, cause a decrease in the efficiency of the one-dimensional (1D) search, since the two subunits are localized in distant regions of the long dsDNA, which reduces the probability of their associating, leading to slower refolding kinetics. ssDNA molecules that are longer than 50 nucleotides are more effective in accelerating Arc refolding than are dsDNA molecules with the same number of nucleotides per DNA chain. It appears that ssDNA systems with lengths of >50 nucleotides colocalize the unfolded chains much more efficiently than do long dsDNA systems in which the 1D search space is less efficient.

The long-range electrostatic field introduced by the charged phosphate groups of the nucleic acid steers the unfolded Arc chains to the DNA surface. At a relatively low ionic strength, the unfolded Arc chains bind rapidly and nonspecifically to the DNA, where they are colocalized to a 1D molecular recognition search space. We found that an increased ionic strength of 0.1 M reduces the electrostatic interactions range and results in an equal sampling of the entire 3D conformational space and to refolding kinetics equal to that in the bulk (see Fig. S1 in the Supporting Material). The colocalization may thus lead to the observed enhanced kinetics of Arc-repressor (Fig. 3) in a DNA length-dependent manner. The refolding
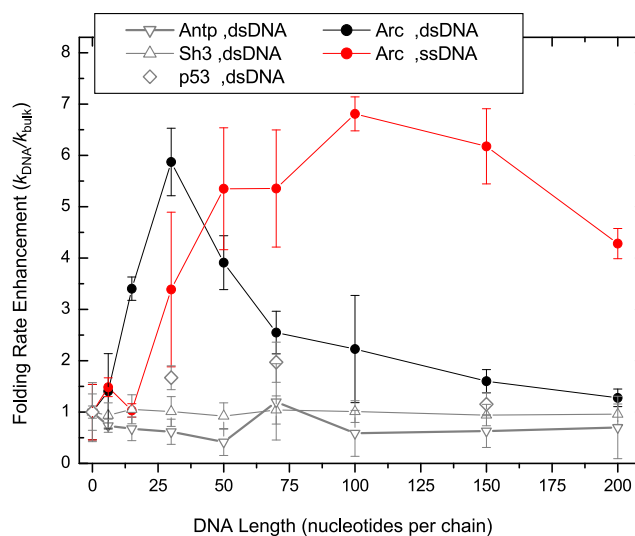

FIGURE 3 Refolding rate enhancement $(k_{f,DNA}/k_{f,bulk})$ of the Arc-repressor with dsDNA molecules (*black*) and ssDNA molecules (*red*) and a comparison with rate enhancements of the SH3 domain with dsDNA (*triangles*), a monomeric non-DNA-binding protein; Antp (*inverted triangles*), a monomeric DNA binding protein; and a dimer of the p53 tetramerization domain (*diamonds*). The rate increases in the ssDNA systems and becomes substantially higher than the rate in dsDNA when the DNA molecules are longer than 30–50 nucleotides. No significant rate enhancement is observed in the other protein-DNA systems.

rates of the control proteins (Sh3 domain, Antennapedia Homeodomain-DNA complex (Antp), and the dimer from the p53 tetramerization domain) are not significantly affected by the presence of dsDNA. This indicates the important role of electrostatic guidance in the colocalization of unfolded subunits of a dimeric DNA-binding protein such as Arc-repressor, and the significant refolding rate enhancement of a DNA-binding dimeric protein by a nucleic acid.

Colocalization of Arc subunits by ssDNA and dsDNA molecules

Long ssDNA molecules (≥ 30 nucleotides) are more effective than dsDNA molecules of the same length in accelerating Arc-repressor refolding (Fig. 3). Most likely, the colocalization of the unfolded Arc chains is more significant in ssDNA systems than on the surface of a dsDNA molecule.

To probe the chain colocalization effect in the different protein-DNA systems, we measured the separation distances between the centers of mass of the two Arc chains $[R_{cm}(A) - R_{cm}(B)]$ in the unfolded state. Arc chains near short dsDNA segments of 6–15 bp diffuse within an average separation distance (Fig. 4 A) that exceeds the length of the respective dsDNA molecule (Fig. 4 B, *inset*). This indicates that the protein-DNA electrostatic attraction is insufficient and that the chains diffuse in the bulk. The separation distance between the two chains decreases to a minimum for a dsDNA segment of ~50 bp. In the presence of 50 bp dsDNA molecules, the separation distance between the two subunits of Arc does not exceed the length of the DNA molecule, hence
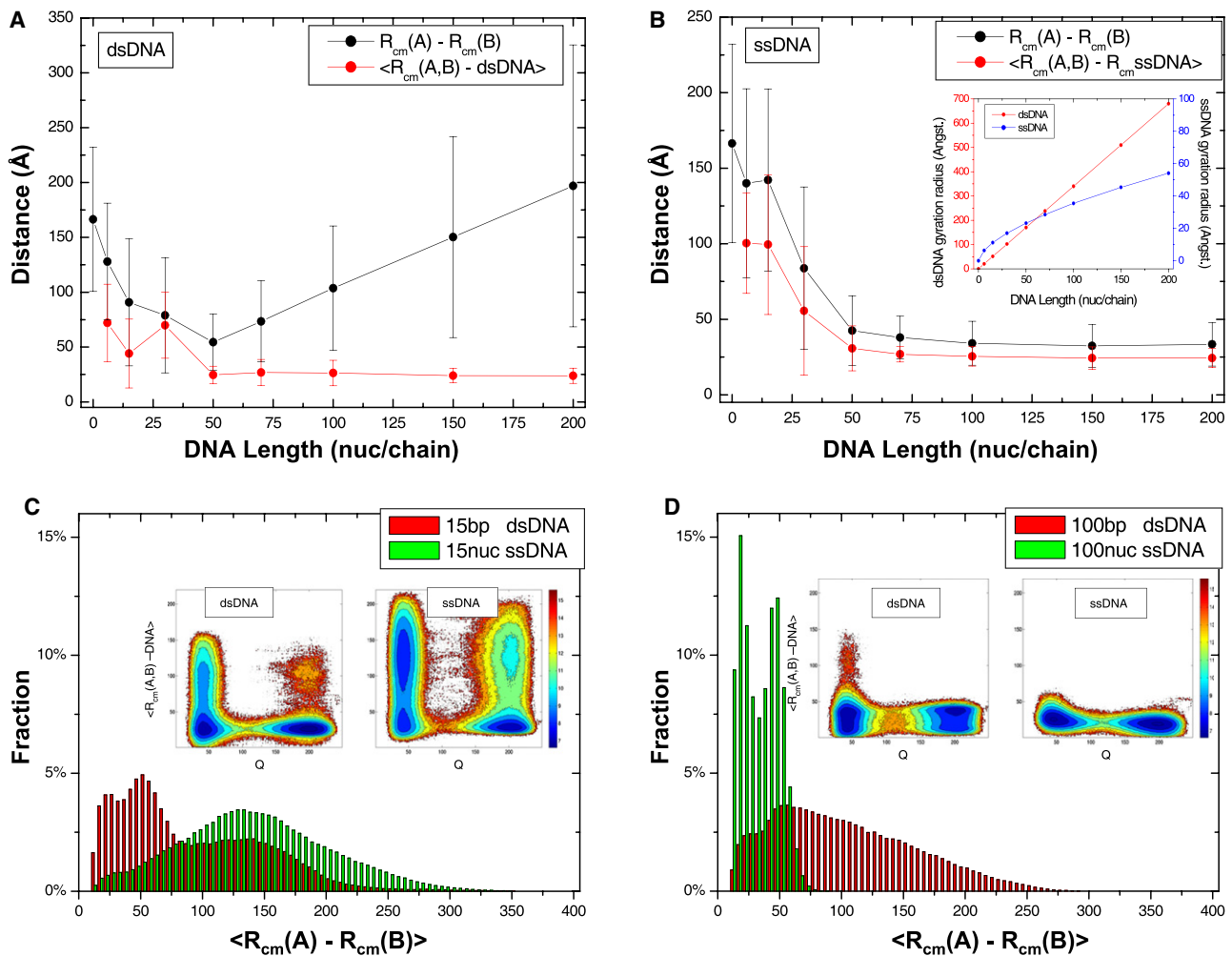
FIGURE 4 Separation distance (*black line*) between Arc-repressor chains in the unfolded state and their average distance to the DNA (*red line*) when studied with dsDNA (*A*) and ssDNA (*B*). The inset shows the gyration radius of the dsDNA and ssDNA molecules (*red and blue*, respectively) as a function of the number of nucleotides. It is evident that the chain separation distance does not exceed the length of the DNA or the gyration radius of the respective DNA systems, due to diffusion in the bound state. (*C*) Histograms of the chain separation distances of the unfolded state in dsDNA (*red bars*) and ssDNA (*green bars*) systems with a length of 15 nucleotides. A narrower distribution around short distances is observed near dsDNA molecule at this length. The free-energy surfaces of Arc refolding shown in the inset demonstrate that in the presence of 15 bp dsDNA chains, the folding is coupled to the protein-DNA association. (*D*) The same as in *C*, in dsDNA and ssDNA systems with 100 nucleotides. A narrow distribution is observed in the ssDNA system and a broad distribution is seen in the dsDNA system. The free-energy surfaces in the inset show that long ssDNA (*right*) introduces tighter localization than does dsDNA (*left*) with the same number of nucleotides per chain.

the chains are bound to the DNA and the diffusion of the chains is performed along the DNA molecule surface in a quasi 1D space. This 1D localization is the origin of the faster folding rate. For longer dsDNA, the separation distance between the Arc chains increases, thereby reducing the probability that the two chains will encounter each other. It is noteworthy that the protein in the 50 bp dsDNA system does not present the fastest refolding rate constant; rather, that constant is associated with the 30 bp dsDNA system (Fig. 3). It is possible that the most efficient molecular recognition search is achieved when the two chains perform a combination of both 1D diffusion along the DNA and local disassociation events (hopping). This combination is observed in the 30 bp dsDNA system, but not in the 50 bp system from which the Arc subunits do not dissociate during the simulations.

The chain separation distance near ssDNA molecules (Fig. 4 *B*) decreases as the length of the DNA molecule increases. The separation distance near long ssDNA molecules is much shorter than that near long dsDNA. Thus, the colocalization effect in the presence of ssDNA molecules is much more substantial. The dynamics of the ssDNA molecule allows it to adopt more compact conformations that are far from being linear, such that the diffusion of the protein chains in the vicinity of the ssDNA is performed in a confined geometry that resembles a two-dimensional (2D)-3D space. For long ssDNA systems, the mean distance of the center of mass of the chains from the ssDNA does not exceed the gyration radius of the DNA (Fig. 4 *B*, *inset*).

The chain colocalization effect of ssDNA and dsDNA is illustrated in Fig. 4, *C* and *D*, for two DNA lengths: 15

and 100 nucleotides per chain. Each panel shows the histogram of the separation distance between the two Arc chains in the unfolded state. In the presence of a short, 15 nucleotide DNA, the distance between the chains may be long for both the ssDNA and dsDNA systems due to poor electrostatic attraction, with a relatively better chain colocalization near the dsDNA molecule reflected by a narrower distribution around short distances. This is also supported by the free-energy surfaces shown in the inset of Fig. 4 C. Free-energy surfaces of Arc in the presence of 15 bp dsDNA (*right*) or 15 nucleotide ssDNA (*left*) are projected onto Q and the minimum distance between the center of mass of the monomers and the DNA axis (for dsDNA) or DNA center of mass (for ssDNA). In the presence of a 15 bp dsDNA molecule, Arc dimerization takes place on the DNA, whereas for ssDNA of the same length, the Arc monomers assemble both on the DNA and at a distance from it.

In the presence of a 100 nucleotide DNA molecule, an inverse of the colocalization effect is observed in which the chain separation distance distribution becomes significantly narrower for the ssDNA system and broader in the dsDNA system. This reflects the significant effect of the flexibility of ssDNA in Arc refolding acceleration. The tighter binding of Arc chains to 100 nucleotides ssDNA compared to 100 bp dsDNA is also reflected in the free-energy surfaces in the inset.

The unfolded Arc chains perform a mostly 1D diffusion along the surface of long dsDNA molecules, and a quasi 3D diffusion in the vicinity of long ssDNA molecules. To appreciate the enhanced folding kinetics induced by ssDNA compared to dsDNA, we considered two noninteracting random walkers on 1D, 2D, and 3D lattices. In the 1D lattice simulation, each random walker moved randomly to either one of two possible directions (left or right), whereas in the 2D and 3D lattice simulations, each move was performed in any one of the four or six possible directions, respectively. The random walkers were constrained to the boundaries of the lattices (i.e., reflecting boundaries) and double population of the sites was restricted. We found that the probability of one walker encountering the other (the encountering probability) decreases as the number of lattice sites increases, with the decay being faster in 1D lattices than in 2D and 3D lattices (Fig. 5). This phenomenon explains the slow Arc dimerization for long dsDNA (Fig. 3). Moreover, the encountering probability in a 3D lattice is higher than that in a 1D lattice for all lattice sizes. The higher encountering probability in 3D lattices may explain the preferential chain colocalization mechanism in ssDNA systems compared to dsDNA systems, and the faster refolding kinetics in the ssDNA systems.

Nucleic acid decreases the cooperativity of Arc-repressor unfolding

The free-energy profiles of Arc refolding, as well as their melting curves, imply that Arc folding proceeds more cooper-
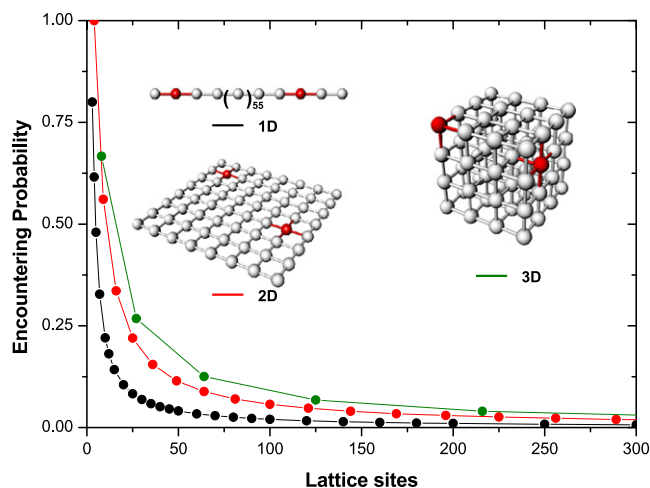

FIGURE 5 Two noninteracting random walkers on 1D, 2D, and 3D lattices without double population of sites. The fastest decrease in the encountering probability is observed in 1D lattices (*black*), and the slowest decrease is observed in 3D lattices (*green*). For all lattice sizes, the encountering probability in the 3D lattices is higher than that in the 1D lattices. Representative lattices of 64 sites are shown with the two random walkers labeled in red.

atively in the absence of DNA. Inspired by previous studies by Hyeon and Thirumalai (47) and Kouza et al. (59), we measured the cooperativity (Ω_c) of the unfolding transition using the dimensionless cooperativity index with respect to T by $(\Omega_c^T) = T_{max}^2/\Delta T |dQ/dT|_{max}$, where ΔT is the peak width at the half-maximum of the derivative $|(dQ/dT)|$, and T_{max} is the temperature at the maximum of dQ/dT. We carried out the calculation on a system with no DNA and on two representative dsDNA and ssDNA systems comprising 100 nucleotides. We found that the cooperativity of unfolding in the bulk was the highest, with $\Omega_{c, No DNA} = 1041.55$, and that the cooperativity of unfolding in the dsDNA and ssDNA systems was 928.19 and 794.42, respectively (Fig. 6). This result indicates a less cooperative transition in the ssDNA system, which is consistent with the lower free-energy barriers in the presence of ssDNA compared to Arc folding in the bulk and in the presence of dsDNA (Fig. 2).

Transition-state analysis for Arc dimerization

The regions of the transition-state ensemble in the free-energy profiles are shifted toward higher values of Q in the presence of dsDNA than in the bulk (Fig. 2 B). In the case of ssDNA, the effect is more pronounced. To explore the microscopic effect of DNA on the transition state ensemble, we performed a Φ-value analysis of Arc folding with no DNA and in the presence of 100 nucleotide ssDNA at the respective T_f of each system. For each of the total number of native contacts of Arc (n = 248), the contact Φ-value was estimated by calculating (29) $\phi_{ij} = ((P_{ij}^{TS} - P_{ij}^U)/(P_{ij}^F - P_{ij}^U))$, where P_{ij} is the probability of contact being made between residues i and j, and superscripts TS, U, and F stand for the transition state, unfolded state, and folded state, respectively. Fig. 7 A shows the Φ-value
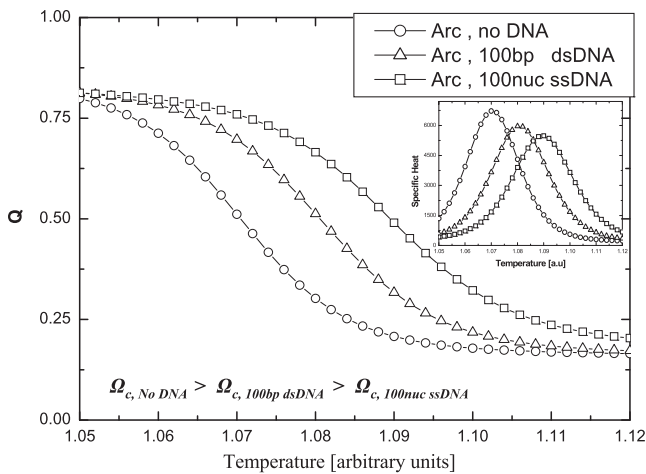
FIGURE 6 Cooperativity (Ω_c) of Arc-repressor folding. Sigmoidal curves of the number of native contacts for folding (Q) against temperature (T), and specific heat plots (*inset*), which are equivalent to the derivative dQ/dT, are shown. Data for both plots were obtained by the weighted-histogram analysis method. The presence of a nucleic acid, especially ssDNA, makes the unfolding transition less cooperative with $(\Omega_{c,100\mathrm{bp\ ds}})/(\Omega_{c,\mathrm{No\ DNA}}) = 0.89$, and $(\Omega_{c,100\mathrm{nuc\ ssDNA}})/(\Omega_{c,\mathrm{No\ DNA}}) = 0.76$.

contact map of the Arc dimer for refolding in the absence of DNA (above the main diagonal) and refolding in the presence of a 100 nucleotide ssDNA (below main diagonal). The blueprint of the contact map indicates structured regions of the dimer in both systems, with overall higher Φ-values in the ssDNA system.

To investigate the structural details of the ssDNA effect on the transition-state ensemble, we subtracted the background

Φ-values (i.e., of the system with no DNA) from the Φ-values of the ssDNA system and considered only contacts with $\Phi_{ij,100\mathrm{nuc\ ssDNA}} - \Phi_{ij,\mathrm{No\ DNA}} > 0.15$. The contact map in Fig. 7 *B* shows the results using a ribbon representation of the Arc-repressor. The most dominant contacts in this map are interfacial contacts (*red* regions) in the β-sheet DNA binding motif (marked with an *arrow*) as well as local interfacial contacts between the two helices in the C-termini. Some intrinsic monomer contacts in the helices following the recognition motif are present as well. These findings suggest that in the presence of a nucleic acid, the transition state of the Arc-repressor is more folded, and that DNA may stabilize structural motifs that are required for Arc dimerization.

CONCLUSIONS

In this study, we used the native topology-based model supplemented with nonspecific electrostatic interactions to investigate the folding and dimerization mechanisms of the Arc-repressor in the presence of DNA molecules. In agreement with experimental results from Rentzeperis and co-workers (23), we showed that the refolding rate of the Arc-repressor can be significantly enhanced by both ssDNA and dsDNA. Our study provides microscopic insights as well as predictive tools to estimate the effects of DNA length and flexibility, and of the salt concentration on folding acceleration induced by nucleic acids.

Our model for Arc refolding in the presence of dsDNA, in which the protein was flexible but the DNA remained rigid, showed that the refolding rate depends on DNA length and decreases with increasing DNA length beyond ~30 bp. The
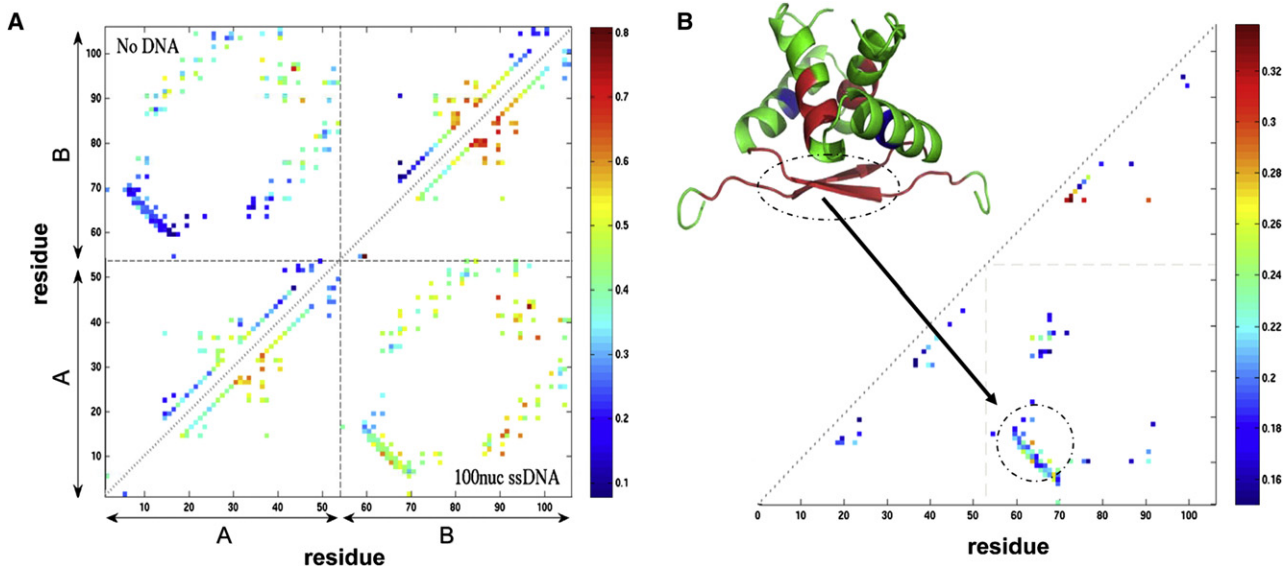

FIGURE 7 Arc-repressor transition-state ensemble analysis. (*A*) Contact Φ-value map at the folding temperature, T_f, in the presence of a 100 nucleotide ssDNA molecule (*lower triangle*) and in the absence of DNA (*upper triangle*). Straight dashed black lines separate the monomers A and B. (*B*) The difference between the contact Φ-values of the transition-state ensemble of Arc studied in the presence of ssDNA and in the bulk, $\Phi_{ij,100\mathrm{nuc\ ssDNA}} - \Phi_{ij,\mathrm{No\ DNA}}$. Significantly structured interfacial regions are colored red in the ribbon figure representation of Arc. The spots along the map diagonal imply local intrachain contact formation in the helices, examples of which are colored blue.

Arc-repressor exhibited higher refolding rate constants in the presence of long ssDNA molecules compared to long dsDNA molecules. The attractive electrostatic forces between the Arc monomers and the DNA (especially for long DNA molecules) result in colocalization of the proteins and restriction of their translational motions. The colocalization of the Arc monomers on 1D molecules enhances their dimerization. For long DNA, the acceleration of Arc folding, in comparison to its folding in the bulk, is less significant since the two monomers are localized in distinct regions and their encountering probability becomes very low. This finding is consistent with the slower refolding kinetics of Arc in a high concentration of ssDNA reported by Sauer and colleagues (23), which presumably results from localization of the monomers on separate DNA molecules. For ssDNA, the acceleration is more pronounced than for dsDNA of the same length. Because of its flexibility, one may view the ssDNA as a molecule with complex geometry whose dimensionality is between 2D and 3D. Colocalization on ssDNA may therefore result in a more efficient Arc folding than on 1D dsDNA.

A recent view of the evolution of protein interactions and allostery suggests that the natural processes of protein colocalization in the cell, which effectively increase the local concentration of the molecules, may change improbable evolutionary events into probable ones (22). The prominent attraction of the protein to the surface of the nucleic acid and the nonspecific binding that follows are driven by electrostatic forces between the negatively charged DNA phosphate groups and the positive residues of the protein. An increase in the salt concentration perturbs the protein-DNA attraction and significantly reduces the DNA effect on the dimerization process. This natural colocalization process is essential for transcriptional control and, apparently, is consistent with the assembly mechanism of the Arc-repressor dimer.

Arc-repressor dimerization on DNA follows the monomeric pathway. In the presence of DNA molecules composed of a few tens of nucleotides, the folding takes place between two unfolded Arc subunits that nonspecifically bind to the DNA and can linearly diffuse along it. In this assembly pathway, which is typical of many other dimeric transcription factors, a monomer diffuses more rapidly than a dimer along the DNA. We believe that this in turn may suggest an evolutionary mechanism that on one hand allows DNA-binding protein to perform a fast target recognition, but on the other hand requires a ternary encounter of monomer-monomer-DNA target that satisfies the regulatory demands of gene expression or repression.

The unfolded monomers of the Arc-repressor are more structured when they are bound to the DNA, which yields less cooperative folding. Investigation of the transition-state ensemble has shown that during the molecular recognition search in the presence of DNA, Arc monomers are partially structured, and some regions, such as the β-sheet DNA binding motif, are significantly more structured than they are in the transition state in the bulk. It is therefore plausible, as was previously suggested

(23), that the binding of the protein to DNA may stabilize a structural nucleus that is required for Arc dimerization.

We believe that the characteristic kinetics and thermodynamics of Arc-repressor dimerization on DNA presented here and in previous works are also common to other dimeric transcription factors. Furthermore, the dependence of the dimer assembly mechanism and rate on the environment provided by the nucleic acid demonstrates the high regulatory demands that must be fulfilled in processes of gene expression. The specific binding of the Arc-repressor and other transcription factors to DNA must be explored in the future to determine whether the widespread evolutionary use of multimeric DNA-binding proteins provides a strategy for a more efficient and well-regulated search.

SUPPORTING MATERIAL

A figure is available at http://www.biophysj.org/biophysj/supplemental/S0006-3495(09)00614-6.

REFERENCES

1. Lamb, P., and S. L. McKnight. 1991. Diversity and specificity in transcriptional regulation: the benefits of heterotypic dimerization. *Trends Biochem. Sci.* 16:417–422.

2. Ptashne, M., and A. Gann. 1997. Transcriptional activation by recruitment. *Nature.* 386:569–577.

3. Halford, S. E., and J. F. Marko. 2004. How do site-specific DNA-binding proteins find their targets? *Nucleic Acids Res.* 32:3040–3052.

4. Buchler, N. E., U. Gerland, and T. Hwa. 2003. On schemes of combinatorial transcription logic. *Proc. Natl. Acad. Sci. USA.* 100:5136–5141.

5. Marianayagam, N. J., M. Sunde, and J. M. Matthews. 2004. The power of two: protein dimerization in biology. *Trends Biochem. Sci.* 29:618–625.

6. Remenyi, A., H. R. Scholer, and M. Wilmanns. 2004. Combinatorial control of gene expression. *Nat. Struct. Mol. Biol.* 11:812–815.

7. Wolberger, C. 1999. Multiprotein-DNA complexes in transcriptional regulation. *Annu. Rev. Biophys. Biomol. Struct.* 28:29–56.

8. Kohler, J. J., and A. Schepartz. 2001. Effects of nucleic acids and polyanions on dimer formation and DNA binding by bZIP and bHLHZip transcription factors. *Bioorg. Med. Chem.* 9:2435–2443.

9. Metallo, S. J., and A. Schepartz. 1997. Certain bZIP peptides bind DNA sequentially as monomers and dimerize on the DNA. *Nat. Struct. Biol.* 4:115–117.

10. Kohler, J. J., S. J. Metallo, T. L. Schneider, and A. Schepartz. 1999. DNA specificity enhanced by sequential binding of protein monomers. *Proc. Natl. Acad. Sci. USA.* 96:11735–11739.

11. Kohler, J. J., and A. Schepartz. 2001. Kinetic studies of Fos center dot Jun center dot DNA complex formation: DNA binding prior to dimerization. *Biochemistry.* 40:130–142.

12. Schneider, T. L., and A. Schepartz. 2001. Hepatitis B virus protein pX enhances the monomer assembly pathway of bZIP center dot DNA complexes. *Biochemistry.* 40:2835–2843.

13. Chin, J. W., J. J. Kohler, T. L. Schneider, and A. Schepartz. 1999. Gene regulation: protein escorts to the transcription ball. *Curr. Biol.* 9:R929–R932.

14. Bjornson, K. P., K. J. M. Moore, and T. M. Lohman. 1996. Kinetic mechanism of DNA binding and DNA induced dimerization of the *Escherichia coli* Rep Helicase. *Biochemistry.* 35:2268–2282.

15. Wendt, H., R. M. Thomas, and T. Ellenberger. 1998. DNA-mediated folding and assembly of Myo-E47 heterodimers. *J. Biol. Chem.* 273:5735–5743.

16. Berger, C., L. Piubelli, U. Haditsch, and H. R. Bosshard. 1998. Diffusion-controlled DNA recognition by an unfolded, monomeric bZIP transcription factor. *FEBS Lett.* 425:14–18.

17. Wu, X., C. Spiro, W. G. Owen, and C. T. McMurray. 1998. cAMP response element-binding protein monomers cooperatively assemble to form dimers on DNA. *J. Biol. Chem.* 273:20820–20827.

18. Cranz, S., C. Berger, A. Baici, I. Jelesarov, and H. R. Bosshard. 2004. Monomeric and dimeric bZIP transcription factor GCN4 bind at the same rate to their target DNA site. *Biochemistry.* 43:718–727.

19. Guarnaccia, C., B. Raman, S. Zahariev, A. Simoncsits, and S. Pongor. 2004. DNA-mediated assembly of weakly interacting DNA-binding protein subunits: in vitro recruitment of phage 434 repressor and yeast GCN4 DNA-binding domains. *Nucleic Acids Res.* 32:4992–5002.

20. Levy, Y., J. N. Onuchic, and P. G. Wolynes. 2007. Fly-casting in protein-DNA binding: frustration between protein folding and electrostatics facilitates target recognition. *J. Am. Chem. Soc.* 129:738–739.

21. Shoemaker, B. A., J. J. Portman, and P. G. Wolynes. 2000. Speeding molecular recognition by using the folding funnel: The fly-casting mechanism. *Proc. Natl. Acad. Sci. USA.* 97:8868–8873.

22. Kuriyan, J., and D. Eisenberg. 2007. The origin of protein interactions and allostery in colocalization. *Nature.* 450:983–990.

23. Rentzeperis, D., T. Jonsson, and R. T. Sauer. 1999. Acceleration of the refolding of Arc repressor by nucleic acids and other polyanions. *Nat. Struct. Biol.* 6:569–573.

24. Raumann, B. E., M. A. Rould, C. O. Pabo, and R. T. Sauer. 1994. DNA recognition by beta-sheets in the Arc repressor-operator crystal structure. *Nature.* 367:754–757.

25. Bonvin, A. M. J. J., H. Vis, J. N. Breg, M. J. M. Burgering, R. Boelens, et al. 1994. Nuclear-magnetic-resonance solution structure of the Arc repressor using relaxation matrix calculations. *J. Mol. Biol.* 236:328–341.

26. Bowie, J. U., and R. T. Sauer. 1989. Equilibrium dissociation of the Arc repressor dimer. *Biochemistry.* 28:7139–7143.

27. Papoian, G. A., and P. G. Wolynes. 2003. The physics and bioinformatics of binding and folding—an energy landscape perspective. *Biopolymers.* 68:333–349.

28. Levy, Y., P. G. Wolynes, and J. N. Onuchic. 2004. Protein topology determines binding mechanism. *Proc. Natl. Acad. Sci. USA.* 101:511–516.

29. Levy, Y., S. S. Cho, J. N. Onuchic, and P. G. Wolynes. 2005. A survey of flexible protein binding mechanisms and their transition states using native topology based energy landscapes. *J. Mol. Biol.* 346:1121–1145.

30. Clementi, C., H. Nymeyer, and J. N. Onuchic. 2000. Topological and energetical factors: what determines the structural details of the transition state ensemble and "En-route" intermediate for protein folding? An investigation of small globular proteins. *J. Mol. Biol.* 298:937–953.

31. Onuchic, J. N., Z. Luthey-Schulten, and P. G. Wolynes. 1997. Theory of protein folding: the energy landscape perspective. *Annu. Rev. Phys. Chem.* 48:539–594.

32. Onuchic, J. N., and P. G. Wolynes. 2004. Theory of protein folding. *Curr. Opin. Struct. Biol.* 14:70–75.

33. Shea, J. E., J. N. Onuchic, and C. L. Brooks. 1999. Exploring the origins of topological frustration: design of a minimally frustrated model of fragment B of protein A. *Proc. Natl. Acad. Sci. USA.* 96:12512–12517.

34. Levy, Y., and J. Onuchic. 2006. Mechanisms of protein assembly: lessons from minimalist models. *Acc. Chem. Res.* 39:284–290.

35. Shea, J.-E., and C. L. Brooks. 2001. From folding theories to folding proteins: a review and assessment of simulation studies of protein folding and unfolding. *Annu. Rev. Phys. Chem.* 52:499–535.

36. Prieto, L., D. de Sancho, and A. Rey. 2005. Thermodynamics of Gō-type models for protein folding. *J. Chem. Phys.* 123:154903.

37. Wang, J., Q. Lu, and H. P. Lu. 2006. Single-molecule dynamics reveals cooperative binding-folding in protein recognition. *PLoS Comput. Biol.* 2:e78.

38. Wang, J., K. Zhang, H. Lu, and E. Wang. 2006. Dominant kinetic paths on biomolecular binding-folding energy landscape. *Phys. Rev. Lett.* 96:168101.

39. Lu, Q., H. P. Lu, and J. Wang. 2007. Exploring the mechanism of flexible biomolecular recognition with single molecule dynamics. *Phys. Rev. Lett.* 98:128105.

40. Turjanski, A. G., J. S. Gutkind, R. B. Best, and G. Hummer. 2008. Binding-induced folding of a natively unstructured transcription factor. *PLoS Comput. Biol.* 4:e1000060.

41. Okazaki, K., and S. Takada. 2008. Dynamic energy landscape view of coupled binding and protein conformational change: induced-fit versus population-shift mechanisms. *Proc. Natl. Acad. Sci. USA.* 105:11182–11187.

42. Clementi, C., P. A. Jennings, and J. N. Onuchic. 2000. How native-state topology affects the folding of dihydrofolate reductase and interleukin-1 b. *Proc. Natl. Acad. Sci. USA.* 97:5871–5876.

43. Koga, N., and S. Takada. 2001. Roles of native topology and chain-length scaling in protein folding: a simulation study with a Go-like model. *J. Mol. Biol.* 313:171–180.

44. Savelyev, A., and G. A. Papoian. 2006. Electrostatic, steric, and hydration interactions favor Na+ condensation around DNA compared with K+. *J. Am. Chem. Soc.* 128:14506–14518.

45. Savelyev, A., and G. A. Papoian. 2007. Inter-DNA electrostatics from explicit solvent molecular dynamics simulations. *J. Am. Chem. Soc.* 129:6060–6061.

46. Schlick, T. 2000. Molecular Modeling and Simulation: An Interdisciplinary Guide. Springer, New York.

47. Hyeon, C., and D. Thirumalai. 2005. Mechanical unfolding of RNA hairpins. *Proc. Natl. Acad. Sci. USA.* 102:6789–6794.

48. Schlick, T., B. Li, and W. K. Olson. 1994. The influence of salt on the structure and energetics of supercoiled DNA. *Biophys. J.* 67:2146–2166.

49. Beard, D. A., and T. Schlick. 2001. Computational modeling predicts the structure and dynamics of chromatin fiber. *Structure.* 9:105–114.

50. Beard, D. A., and T. Schlick. 2001. Modeling salt-mediated electrostatics of macromolecules: the discrete surface charge optimization algorithm and its application to the nucleosome. *Biopolymers.* 58:106–115.

51. Givaty, O., and Y. Levy. 2008. Protein sliding along DNA: dynamics and structural characterization. *J. Mol. Biol.* 385:1087–1097.

52. Kumar, S., J. M. Rosenberg, D. Bouzida, R. H. Swendsen, and P. A. Kollman. 1992. The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J Comput. Chem.* 13:1011–1021.

53. Isaksson, J., S. Acharya, J. Barman, P. Cheruku, and J. Chattopadhyaya. 2004. Single-stranded adenine-rich DNA and RNA retain structural characteristics of their respective double-stranded conformations and show directional differences in stacking pattern. *Biochemistry.* 43:15996–16010.

54. Liphardt, J., B. Onoa, S. B. Smith, I. Tinoco, and C. Bustamante. 2001. Reversible unfolding of single RNA molecules by mechanical force. *Science.* 292:733–737.

55. Goddard, N. L., G. Bonnet, O. Krichevsky, and A. Libchaber. 2000. Sequence dependent rigidity of single stranded DNA. *Phys. Rev. Lett.* 85:2400–2403.

56. Kannan, S., and M. Zacharias. 2007. Folding of a DNA hairpin loop structure in explicit solvent using replica-exchange molecular dynamics simulations. *Biophys. J.* 93:3218–3228.

57. Fitzkee, N. C., and G. D. Rose. 2004. Reassessing random-coil statistics in unfolded proteins. *Proc. Natl. Acad. Sci. USA.* 101:12497–12502.

58. Kohn, J. E., I. S. Millett, J. Jacob, B. Zagrovic, T. M. Dillon, et al. 2004. Random-coil behavior and the dimensions of chemically unfolded proteins. *Proc. Natl. Acad. Sci. USA.* 101:12491–12496.

59. Kouza, M., M. S. Li, E. P. O'Brien, Jr., C. K. Hu, and D. Thirumalai. 2006. Effect of finite size on cooperativity and rates of protein folding. *J. Phys. Chem. A.* 110:671–676.