Advanced Review

Coarse-grained models for studying protein diffusion along DNA

Arnab Bhattacherjee,[1]* Dana Krepel[2] and Yaakov Levy[2]*

Understanding the molecular mechanism and the fast kinetics of DNA target site recognition by a protein is essential to decipher genetic activity in the cell. The speed of searching DNA may depend on the structural complexity of the proteins and the DNA molecules as well as the cellular environment. Coarse-grained (CG) molecular dynamics simulations are powerful means to investigate the molecular details of the search performed by protein to locate the target sites. Recent studies showed how different proteins scan DNA and how the search efficiency can be enhanced and regulated by the protein properties. In this review, we discuss computational approaches to study the physical chemistry of DNA search processes using CG molecular dynamics simulations and their advantage in covering long time-scale biomolecular processes.

How to cite this article:
WIREs Comput Mol Sci 2016, 6:515–531. doi: 10.1002/wcms.1262

INTRODUCTION

How DNA binding proteins (DBPs) search and specifically bind to their target sites on DNA is the crux to unravel many complex cellular regulatory processes, such as repression, transcription, replication, and the repair of damaged DNA.[1–3] The search mechanism operates in an extremely crowded cellular environment containing bound protein obstacles that must be by-passed. Furthermore, a genomic landscape contains vast numbers of nonspecific genomic sequences that share partial or complete sequence similarity with the target site and that may therefore misguide the DBPs into binding at an incorrect location. Experiments, however, suggest that DBPs typically circumvent all the obstacles and locate their target sites, which are often known as cognate sites, with remarkable efficiency and accuracy. However, the biophysics behind the facilitated target search mechanism of DBPs remains the subject of keen interest.

Recent advancements in experimental techniques and theoretical modeling suggest that DBPs adopt a facilitated diffusion mechanism during their target search. According to this mechanism, DBPs lower the dimension of the search space by performing fast 1D sliding and 2D hopping dynamics along the DNA contour in addition to slow 3D diffusion.[4–8] This mechanism has been examined indirectly by investigating DNA cleavage using DNA repair enzymes in bulk assays and monitoring their inter-site processivity.[9–13] However, with the advent of nuclear magnetic resonance (NMR) spectroscopy and single-molecule techniques, it has become feasible to visualize the 1D diffusion of a protein along DNA.[14–18] For example, in a series of NMR experiments, Clore and coworkers[19–21] have demonstrated that, during sliding, proteins perform a spiral motion along the major DNA groove while interacting nonspecifically with it via electrostatic attractions between commonly occurring positively charged patches on the DBPs and the negatively charged backbone of the DNA. Other modes of protein translocation on the genomic DNA have also been

*Correspondence to: Koby.Levy@weizmann.ac.il, arnab@jnu.ac.in

[1]Center for Computational Biology and Bioinformatics, School of Computational and Integrative Sciences, Jawaharlal Nehru University, New Delhi, India

[2]Department of Structural Biology, Weizmann Institute of Science, Rehovot, Israel

Conflict of interest: The authors have declared no conflicts of interest for this article.

suggested, including rapid intermolecular jumps (hopping) and intersegmental transfer.[22,23]

Bulk experimental techniques that measure the observables as the average of an ensemble of many rapidly exchanging states can be supplemented by single-molecule experiment that can elucidate the individual interactions between protein and DNA. Notwithstanding this wealth of information, obtaining complete microscopic and mechanistic understanding is limited. Achieving this goal necessitates the development of an *in silico* approach for studying protein–DNA interactions. Using this approach, transcription factor binding and sliding were studied recently using all-atom computer simulations.[24,25] An *in silico* study of the interactions between DNA and the sex determining region Y (SRY) protein found that the DNA molecule switches its conformation upon associating with the SRY protein.[26] Although atomistic simulations can provide a complete structural description of the system of interest and capture hidden steps in a complex cellular process, they typically require gigantic computational efforts. The associated computational cost increases exponentially with the size of the system and, consequently, long time-scale investigations, such as the diffusion of DBPs on genomic DNA, are rather difficult to achieve. A way to enhance the efficiency of the atomistic simulations is to perform umbrella-sampling simulations along a predefined reaction coordinate of the protein sliding (e.g., the major groove track).[24,25] Coarse-grained (CG) models are a means to solving this problem.[27,28] In CG modeling, less-essential atomic details are ignored to develop a model that effectively captures the biophysics of biomolecules while imposing reasonable computing demands. CG approaches are tailored to address certain types of questions to the exclusion of others and, therefore, their power is restricted to the question at hand. This approach has seen a tremendous amount of success in probing the search dynamics and interaction kinetics of DBPs. For example, rotation-coupled sliding dynamics has been experienced through a CG computational study.[29] Other *in silico* studies using similar CG models have provided many other important insights regarding protein–DNA interactions. For example, *in silico* studies have shed light on how the DNA search mechanism is affected in dimeric and monomeric proteins, in the presence or absence of the disordered tails associated with DBPs, by changing the electrostatic properties of the protein interface and frustration in protein–DNA binding, and by the presence of crowder molecules on genomic DNA.[29–44]

In this perspective, we discuss the development of CG models and their application to studying nonspecific protein–DNA interactions. The first part of the review presents a comparative study of different CG models to demonstrate how using slightly different models may alter the kinetics and dynamics of the protein target search mechanism and, therefore, utmost care should be taken while developing CG models. The second part of the review demonstrates how a simple CG model can be used to understand the role of DNA conformations in the search dynamics of DBPs, and hence provides otherwise hard-to-obtain insights into a complex bio-system. This article reviews some previously published data on CG modeling of protein diffusion along DNA as well as additional analysis in which the parameter space of the CG model is further explored.

METHODS FOR DEVELOPING COARSE-GRAIN (CG) MAPS OF PROTEIN AND DNA INTERACTIONS

One of the major challenges in transforming an atomically detailed system into a CG representation is to capture the key features of the system of interest while eliminating the atomic details that are of less importance to the specific question to be addressed. The most commonly employed technique for achieving this outcome is 'system mapping', in which a group of atoms are represented by a single pseudo atom (a bead) located at the center of mass of the corresponding groups, such that the pseudo super atom describes the chemical identity of the group it represents. Imaginary bonds are then constructed between the beads to describe the overall conformation. The success of such mapping is evaluated in terms of its efficiency, accuracy, and transferability. The fewer the number of pseudo atoms in a CG model, the greater the efficiency of the simulation because of the smaller number of degrees of freedom (proportional to N) and especially, because having a smaller number of pseudo atoms reduces the number of pairwise interactions (proportional to N^2). (This argument is still valid even when comparing to atomistic simulations where mostly pairwise interactions between neighboring atoms are considered because the number of pairwise interactions in the CG models will be lower.) In contrast, higher number of beads in the model typically corresponds to more accurate representation of the system.

Here, we adopted a CG model (similar to that used in the pioneering studies of Levitt and Warshel[45,46] to represent protein conformations) in which a single bead represents an amino acid. The model shares the 'teleological' assumption that the strength

of non-bonded interactions between protein residues is dictated by the folded three-dimensional structure (native interactions) of a protein[47–53] whereas non-native interactions (which are purely repulsive in nature) make a negligible contribution to describing the excluded volume effects. The local geometry of the reference structure is maintained by restricting the harmonic bond and the angular and dihedral potentials. Such native-centric models provide an idealized 'funnel'-like energy landscape for folding in which the energetic frustrations because of non-native interactions are absent. Despite this simplicity, such models enjoy many successes. They are able to describe the conformational transitions between multiple structures,[54–56] address the effects of denaturants and osmolytes,[57–59] and explain experimentally determined data regarding the free energies of mutation.[60] They are also used successfully to model conformational transitions in biomolecular complexes,[61] investigate the effects of confinement and crowding upon folding,[62,63] and even to study the remarkable mechanism by which intrinsically disordered proteins couple their folding to ligand binding.[64,65]

COARSE-GRAINED SIMULATIONS OF NONSPECIFIC PROTEIN–DNA INTERACTIONS

To explore the mechanism of protein translocation along DNA, we performed Langevin dynamics simulations with sufficiently small time steps and with the friction coefficient γ set to 0.01 (proportional to velocity of the residue). The choice of time step is such that the error caused by discretization of the equations of motion is small and the correct ensemble is sampled.[66] Furthermore, the model has to include random noise in order to incorporate molecular fluctuation due to the solvent and to the random walk motion of the proteins during sliding.

The nonspecific interactions between protein and DNA molecules were modeled by electrostatic energy (U_{el}) and by excluded volume (U_{ev}). The major component of the nonspecific interactions between protein and DNA molecules is the electrostatic forces between the charged amino acids [Arg, Lys (negatively charged); Asp, and Glu (positively charged)] and the negatively charged DNA phosphate groups. We employed the Debye–Hückel potential to describe the electrostatic interactions between the protein and DNA. The Debye–Hückel model has been used successfully to address the energetics and dynamics of various biomolecular systems, such as RNA folding,[67–69] protein–DNA binding,[34,70] and chromatin assembly.[71] While the Debye–Hückel model is a powerful means of introducing the salt effect (which screens electrostatic interactions) into the Coulomb potential, one should be aware of its approximations. The model is valid for relatively dilute solutions of monovalent ions and does not take ion condensation into account. In particular, it fails for an ion concentration >0.5 M.[72] We also note that the strength of electrostatic interactions is strongly correlated with the resolution of the CG model and the location of charges. This aspect is discussed later in this review. We used a dielectric constant of 70–80, which is the typical range in water, as the protein–DNA interface is much more hydrated in the nonspecific complex than in the specific complex. During the simulations, we kept the protein molecule dynamic, because 100 bp linear DNA molecules are fairly rigid (large persistence length), while we kept the DNA conformations frozen, which enabled us to neglect DNA dynamics. In this manner, we were able to investigate explicitly the role of DNA geometry (e.g., canonical B-DNA, curved or bent DNA) on search dynamics by proteins. Recent studies showed that the internal dynamics of a DNA molecules has a small effect on the biophysical characterization of sliding.[73,74] In addition to the DNA fluctuations, its local geometry (e.g., bending or kinking) may also affect protein dynamics.

The other energetic contribution that affects protein–DNA interactions is the excluded volume energy, which is given by, $E_{ev} = \sum_{i<j-3}^{\text{non-native}} \varepsilon_{ev}\left(\sigma_{ij}/r_{ij}\right)^{12}$, where, r_{ij} is the distance between the protein and DNA residues, ε_{ev} denotes the strength of the interaction and is typically set to 0.1, and σ_{ij} is a model-dependent parameter (discussed later) that measures the closest possible distance between two interacting sites.

ROLE OF PROTEIN COARSE GRAINING ON THE TARGET SEARCH MECHANISM(S) ADOPTED BY DBPs

To investigate how the parameters of the CG model affect outcomes, we designed four distinct but closely related CG models (see Figure 1) of a 93-residue transcription factor protein Sap-1 (PDB id: 1BC8) interacting with DNA.

1. Model 1 featured a single bead per amino acid residue, located at the respective C_α positions. The radius of each protein bead was 2 Å. The

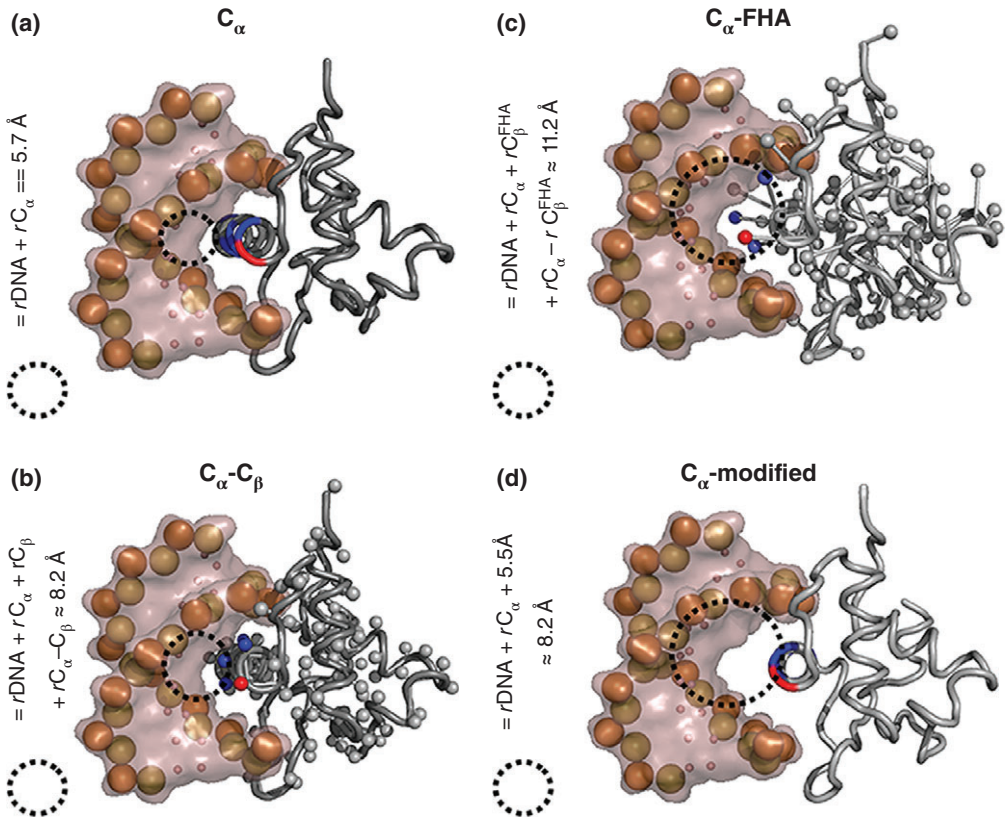

FIGURE 1 | Coarse-grained representations of the protein–DNA complex of the 93-residue transcription factor protein Sap-1 (PDB id: 1bc8) with DNA modeled using the (a) C_α, (b) C_α–FHA, (c) C_α–C_β, and (d) C_α-modified models. The dotted circles represent the specified distance between the protein and the DNA. The recognition region of Sap-1 is labeled with blue and red colors, where the blue region is positively charged and red region is negatively charged. Each DNA nucleotide is presented through negatively charged phosphate beads (orange color), yellow colored sugar beads, and small pink colored bases.

radius of a DNA phosphate bead was 3.7 Å, which fixed the closest possible distance between them (σ_{ij} in term of excluded volume interaction), at 5.7 Å (C_α model, Figure 1(a)).

2. Model 2 utilized a second bead located at the C_β position in addition to the bead located at C_α. The second bead represented the amino acid side chain. The bead was smaller (1.55 Å) than the C_α carbon atom (1.7 Å) and resembled the size of a nitrogen/oxygen atom (C_α–C_β model). The value of σ for this model was ~8.2 Å (see Figure 1(b)) because of the presence of the additional C_β atom and the associated C_α—C_β bond distance.
3. In Model 3, the bead corresponding to the amino acid side chain was placed on the furthest heavy atom (FHA, excluding the hydrogen atoms) rather than at the C_β position. Therefore, σ for this C_α–FHA model (Figure 1(c)) was ~11.2 Å because of the greater linear distance between C_α and the position of the FHA. Although it is more realistic in terms of charge placement, this model lacked rotational degrees of freedom for the amino acid side chains.
4. Model 4 was a modified-C_α model (Figure 1(d)), in which the protein chain was artificially constrained to maintain a location ~8.2 Å away from the DNA molecule. This was similar to the σ of the C_α–C_β model but differed from the later in terms of the resolution and orientation of the charges. Like the previous C_α model, here also a single bead represented an amino acid and therefore, charges at the recognition region were clustered and in close proximity to each other, unlike the C_α–C_β and C_α–FHA models.

These four models can address questions regarding the fine geometry of the interatomic distances at the protein–DNA interface. Side chain rotations could be beneficial in order to minimize the steric clashes among amino acid side chains by

orienting them in different directions. The developed models may shed light on the effect of the orientation of charges at the recognition region of DBPs and whether the distance between the protein and the DNA controls the target search mechanism.

Coarse Graining Affects Intercommunications Between Protein and DNA Molecules

To investigate how the details of coarse graining impact inter-communication between protein and DNA molecules, we performed molecular dynamics simulations of a 100 bp nonspecific DNA molecule with Sap-1, modeled using the four models described above. Ten independent runs were carried out at varying salt concentrations ranging from 0.02 to 0.3 M to achieve statistically significant results for each model. The extent of inter-communications between protein and DNA was estimated from the distributions of the nonspecific energies (E_{ev}, the excluded volume energy; and E_{el}, the electrostatic energy) under the 0.02 M salt condition and is shown in Figure 2.

Our result suggest that, when the amino acid side chains are included in the model of Sap-1 (as in the C_α–C_β and C_α–FHA models), the excluded volume interaction increases (Figure 2(a)). This can be explained by the fact that, in such higher-resolution CG models, the presence of side chain atoms means that the protein lies in close proximity to the DNA beads, which led to an approximately two times increase in the average excluded volume energy compared with the C_α model. In the case of the modified-C_α model, where the protein beads were forced to stay at least ~8.2 Å from any DNA beads, the average excluded volume energy was, as expected, lower (~5-fold) compared with the C_α model. A different trend, however, was obtained for the distribution of electrostatic energy between the interacting Sap-1 and DNA molecules (Figure 2(b)). Clearly, by placing the charges on side chain atoms in the C_α–C_β model, the positively charged recognition helix of Sap-1 had the ability to approach the negatively charged phosphate beads of the DNA molecule more closely than in the C_α model of Sap-1. This resulted in approximately two times greater stabilization of the average electrostatic interactions in the C_α–C_β and C_α–FHA models compared with that obtained using the C_α and modified-C_α models, in which identical distributions of electrostatic energies were found.

Modeling side chain atoms guarantees greater stability in electrostatic interactions, however, the magnitude of the stabilization may depend on the geometry of the beads that represent the side chains. In the C_α–FHA model, side chain beads were connected to their respective C_α beads by longer bonds, yet, showed ~3% less stabilization in electrostatic interactions compared with the C_α–C_β model. This is because in the C_α–FHA model, the side chain beads were oriented in different directions in order to minimize steric clashes with neighboring amino acid side chains, and this resulted in a dispersed distribution of charges at the recognition region. As a result, the effective distances between the charged sites of the protein and DNA molecules were longer and thus, the electrostatic interactions were weaker. By contrast, the side chain beads of the C_α–C_β model were connected to their respective C_α atoms by shorter bonds, which prevented the charges on the side chains from scattering much. This resulted in a comparatively stronger overlap between the recognition helix of Sap-1 and the negatively charged phosphate beads of DNA, leading to a rise in electrostatic

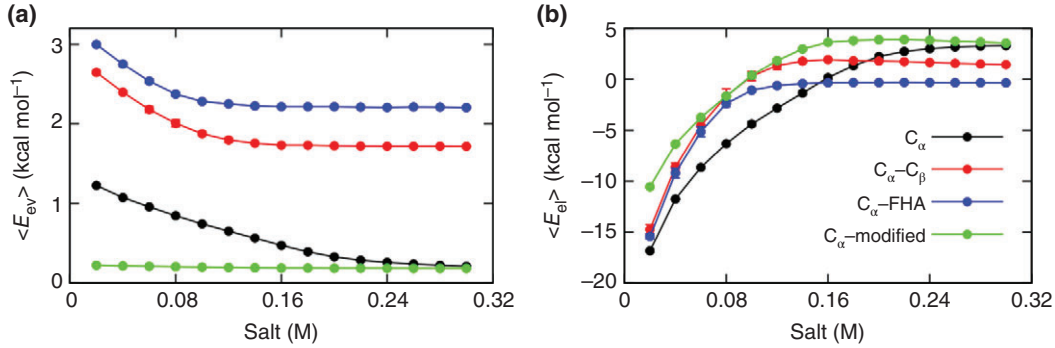

FIGURE 2 | The energetics of the interface in protein–DNA complexes using different coarse-grained models. (a) The excluded volume energy (E_{ev}) and (b) the electrostatic energy (E_{el}) calculated by using the Debye–Hückel potential in molecular dynamics simulations of a 100 bp nonspecific DNA molecule with Sap-1, modeled using the C_α, C_α–C_β, C_α–FHA, C_α-modified protein–DNA models. The mean energy is shown for different salt concentrations.

stabilization when the C_α–C_β model was used, in comparison with the C_α–FHA model.

Correlation Between Coarse-Graining and Target Search Dynamics of Proteins Under Various Salt Conditions

Having seen that the degree of coarse graining impacts the strength of the nonspecific interactions (both electrostatic and excluded volume interactions) between interacting protein and DNA molecules, we examined how it affects the related target search mechanism of Sap-1 under varying salt conditions. The role of the salt condition in regulating the target search dynamics of DBPs was underscored in our previous study.[29] To discriminate between different search mechanisms, we monitored three parameters, namely: the distance (r) between the center of the recognition region in Sap-1 and the center of the closest DNA base pairs; the orientation angle (θ), defined as the angle between the geometric center of the recognition region of the interacting protein, the geometric center of the protein, and the point of association on the DNA surface; and the fraction of the recognition region that resided within the major groove of the DNA. Details can be found in our previous publication.[29] We note that some of the parameters of sliding may depend on the resolution of the CG protein model. Therefore, to precisely characterize the structural variations in search modes, we drew contour plots of r versus θ, as shown in Figure 3. Based on these plots, we defined a snapshot as 3D diffusion if r was greater than 32 Å, where the average electrostatic energy was weak enough to affect the search dynamics. For the C_α model, a sliding mechanism was defined if at least 70% of the recognition helix of Sap-1 was inside the DNA major groove, r was <14 Å and θ was <25 Å (see Figure 3). However, for the C_α–C_β, C_α–FHA, and modified-C_α model systems, considering r < 14 Å seemed to be inadequate and imposing such a condition could lower the sliding propensity by skewing the actual result. Corresponding contour plots helped to choose the right distance criteria, namely, 16, 18, and 18 Å for the C_α–C_β, C_α–FHA, and C_α-modified model systems, respectively. If the protein was found at a distance of less than 32 Å from the DNA and did not satisfy any of the sliding criteria, the snapshot was considered to be representing hopping dynamics. Following this prescription, we estimated the propensities of the protein to adopt each of the three search modes and presented them as functions of salt concentration in Figure 4.

With an increase in salt concentration, the sliding propensity decreases and 3D diffusion increases

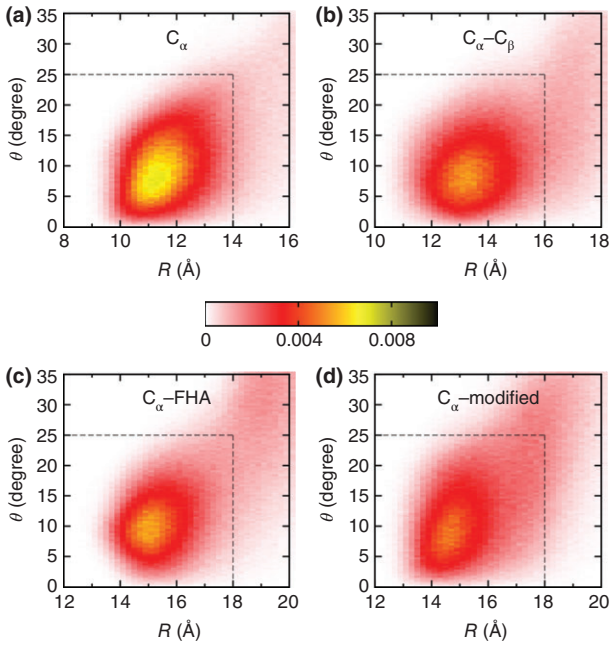

FIGURE 3 | Conditions required to satisfy sliding criteria by Sap-1 in the (a) C_α, (b) C_α–C_β, (c) C_α–FHA, and (d) C_α-modified models where R denotes the distance between the center of mass of the recognition helix of Sap-1 and the center of the closest DNA base pairs and θ represents the orientation angle between them.

for all the four models. The rationale behind this is that higher salt concentrations weaken the electrostatic attraction between the protein and the DNA, leading to a rise in the number of events in which the protein dissociates from the DNA surface. In contrast, the hopping search mode, during which the protein molecule remains bound to the DNA but not necessarily in the major groove, is most populated at moderate salt concentrations. A similar trend was observed previously.[29] Variation in the usage of the three searching modes at different salt conditions corresponded to the different strengths of the electrostatic interactions depending on the coarse-graining. For example, at low salt concentration such as 0.02 M, the sliding propensity in the C_α model was ~82% followed by ~72% in the C_α–C_β model and ~60% in the C_α–FHA model. The hopping and 3D diffusion propensities followed the reverse order and were direct consequences of the associated excluded volume interactions (Figure 2) in these models. The trend, however, was reversed at a high salt condition. While under the low salt condition, the presence of charges on side chains in the C_α–C_β model helped to create a stronger electrostatic association between the protein and DNA molecules (and therefore, a high sliding propensity), under higher salt conditions, where the electrostatic interaction was weaker, steric

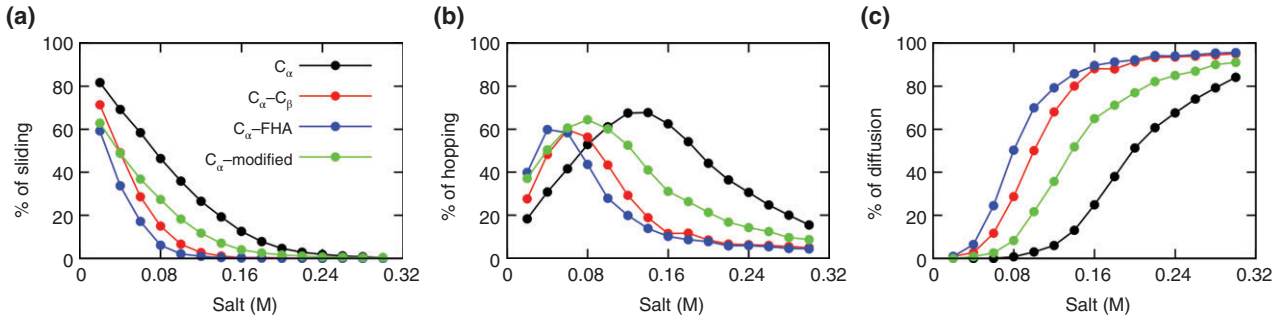

FIGURE 4 | Effect of salt concentration on the interplay between (a) sliding, (b) hopping, and (c) 3D diffusion in modeling undertaken using the C_α, C_α–C_β, C_α–FHA, and C_α-modified protein–DNA models.

clashes between the side chains and DNA beads dominated. This resulted in a fall in the sliding propensity determined using the C_α–C_β model compared with the modified-C_α model from which side chains were absent, and the associated excluded volume interaction was lowest in the C_α–C_β model (Figure 2).

Mechanism Used by Proteins to Search Target DNA Sites and Its Efficacy

The search mechanism is expected to depend on various parameters, including the molecular characteristics of the searcher protein,[32,33,41,75,76] the sequence and structure of the DNA,[37] and cellular conditions.[39,77–79] We illustrate the details of the target search mechanism by investigating the parameters of the sliding events within it. Two parameters, the cumulative distance (MSDz) traveled during a single sliding event by the protein along the contour of DNA and the number of sliding events, were measured for each system under the 0.02 M salt condition. In Figure 5(a) and (b), we present the probability distribution of MSDz and the cumulative number of sliding events as a function of the number of base pairs for all the four CG models.

The higher sliding propensity in CG models that lack protein sidechains (C_α and C_α-modified) is correlated with more continuous and long sliding events and larger MSDz values. In the presence of protein side chains that caused steric clashes with DNA beads (C_α–C_β and C_α–FHA models), sliding events were frequently disrupted. For example, the probability of performing continuous sliding events for more than 5 Å was ~65% in the absence of side chains (C_α model) compared with only ~55% in the presence of side chains (C_α–C_β model). Furthermore, the number of cumulative sliding events was higher (>40,000, Figure 5(b)) if the side chains were not modeled compared with the CG models that incorporated side chains (~14,000 sliding events).

These results raise the question of how these mechanistic variations in target search dynamics that arise from subtle differences between the CG models regulate the search efficiency of the modeled protein.

We enquired this issue by estimating DNA positions probed, which is defined as the number of new DNA sites scanned by the protein via sliding motion (Figure 5(c)). Each DNA site was counted only once unless the protein performs 3D diffusion. After a spell of 3D diffusion, the memory of previously scanned DNA sites was wiped out. This scheme was adopted on the basis that the probability to associate at last visited site after dissociation is negligible, thus, upon reassociation, the protein would be searching an unprobed piece of DNA. Clearly, when the side chains are not modeled, the protein performs smooth sliding dynamics along the DNA contour, which enables it to probe the maximum number of DNA sites. For example, in the absence of side-chains, the maximum number of visited DNA sites is ~5.5 times higher than if the side chains are modeled as in C_α–C_β and C_α–FHA models. The differences in efficiency to scan the DNA sites are also reflected in the corresponding measurements of 1D diffusion coefficients (Figure 5(d)) for different CG models. The CG protein models with side chains diffuse along the DNA contour more slowly (~3–4 times) compared with the C_α and modified-C_α model. This indicates that the flanking side chains of proteins are engaged in more 'cross-talks' with the DNA atoms because of their close proximity and spatial orientation. While this reduces the overall sliding speed and may retard the overall translocation of the protein along the DNA contour compared with model proteins without side chains, it may on the other hand, provide enough time to the protein to orient/switch its conformation in order to form specific contacts when it reaches the target site.

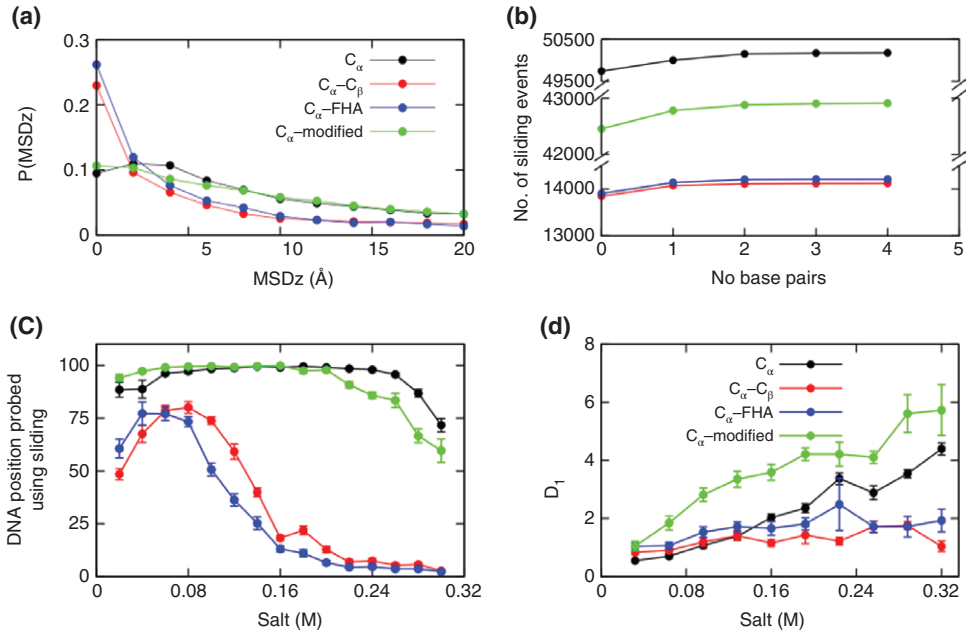

FIGURE 5 | The extent of search dynamics by sliding under the C_α, C_α–C_β, C_α–FHA, and C_α-modified models are measured by calculating (a) MSDz, which is the cumulative distance traveled during a single sliding event by the protein along the *Z*-axis of the DNA and (b) the cumulative number of sliding events in terms of the total number of DNA base pairs probed. The effect of conformational space on the efficiency of DNA search is measured by (c) the number of DNA base pairs probed using sliding dynamics and (d) the 1D diffusion coefficient (D_1), which is obtained from the linear behavior of the mean square displacement of Sap-1.

For many protein–DNA complexes, this is the rate-determining step.[36] Interestingly, all four models showed that, during sliding, the searching protein spun along the DNA axis (Figure 6(a)–(d)), which is along the line of observations from previous experiments.[15] However, examples of rotation-uncoupled sliding (which can be viewed as the hopping search mode) were also reported for multi-domain proteins[31] and proteins with disordered tails.[73]

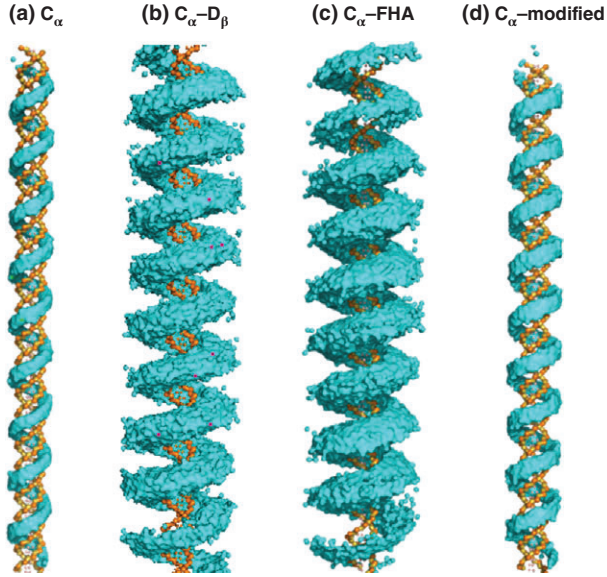

FIGURE 6 | Traces (cyan color) of the rotation-coupled sliding search path of Sap-1 modeled through different coarse graining prescriptions: (a) C_α, (b) C_α–C_β, (c) C_α–FHA, and (d) modified-C_α models.

The Effect of DNA Flexibility on Facilitated Diffusion Mechanism

A similar 'system mapping' approach is also found to be useful in developing a reduced description of the double-stranded (ds) DNA molecule in which each nucleotide is represented by three beads (corresponding to the sugar, phosphate groups, and nitrogenous bases) located at the geometric centers of the associated group. This model produces a canonical B-DNA structure with distinct major and minor grooves and, coupled with a suitable energy function, it captures the intrinsic biophysics of dsDNA in situations such as DNA melting and renaturation, and in describing nanotweezers.[80–83] Incorporating such CG models for DNA flexibility in computational studies of DNA search by proteins is essential to examine the role of the DNA dynamics on the biophysical characterization of the facilitated diffusion. Recent CG models show only a small effect of DNA flexibility on the search mechanism.[73,74] Figure 7 shows a comparison between protein dynamics along a rigid and flexible DNA molecule (the DNA flexibility was modeled using the 3SPN.1 model).[82] Searching DNA via the hopping search mode is slightly increased at the expense of reduced sliding propensity. Consequently, the 1D diffusion coefficient (D_1) is larger when searching a flexible DNA. In accordance with the earlier publications, the search mechanism is not markedly different when the DNA is flexible. The validity of representing the DNA as a rigid molecule stems from the persistence length of the DNA which may exceed 50 nm. Yet, the DNA flexibility may influence the biophysics of the search and regulate specific recognition.[84,85]

ROLE OF THE DNA MOLECULE ON THE TARGET SITE SEARCH MECHANISM ADOPTED BY DBPS—INSIGHTS FROM CG MODELS

Although toilsome, efforts to develop a suitable CG model are worthwhile because CG models are well suited to studying those aspects of biological processes that remain hidden from other approaches. For example, using experimental tools to understand the role of DNA conformation on the target search dynamics of DBPs is a daunting task because of many confounding factors, such as cellular environment, DNA dynamics, and salt conditions. The DNA structure is obviously important for specific protein–DNA recognition and although isolated instances were reported in which a protein bound with severely deformed DNA structures,[78,86,87] usually even small nuances in DNA conformations can change its affinity to proteins.[84,88,89] However, the effect of DNA structure (e.g., width and depth of the grooves) is expected to be smaller for nonspecific than for specific binding.[90] To investigate the effect of the DNA parameters on the search, researchers can separate the contribution of the DNA into geometry and internal dynamics components. We recently achieved this goal by investigating protein dynamics and kinetics on circular DNA having different curvatures.[31,44] The starting atomistic DNA structures were

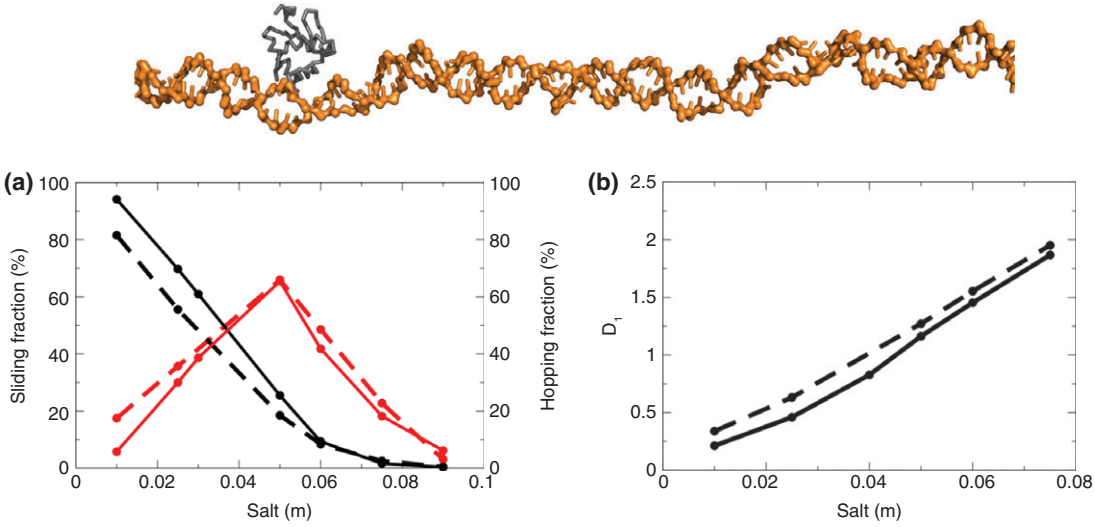

FIGURE 7 | Sliding on a flexible DNA. The effect of DNA flexibility on the sliding and hopping propensities (a) as well as on the D_1 diffusion coefficient (b).

generated by w3DNA web server[91] and NAB software,[92] respectively. Circular DNA was tested because of the abundant presence of curved, bent, and looped DNA structures in a circular geometry.

Differences in the Conformational Geometry of Circular and Linear DNA Molecules

We began by extensively characterizing circular and linear B-DNA structures and estimated the associated electrostatic potential around the molecules using the Curves[93] and DelPhi web servers,[94] respectively. Shifting from linear B-DNA to circular DNA, the curvature, defined as an inverse function of the radius of the DNA axis, changes from 0 to ~0.03 per Å (as estimated by Curves).[93] Also, in circular DNA, the major groove width (W) differs between the inside ($W_{in} \sim 9.4$ Å) and outside ($W_{out} \sim 11.5$ Å) of the molecule (see Figure 8(a)). In sharp contrast, linear B-DNA exhibits a constant major groove width $W_{B\text{-}DNA} = 10.8$ Å with infinite radius. A high curvature value signifies that the negatively charged phosphate atoms are in close proximity to each other inside the circular DNA compared with outside the circle. As a result, the associated electrostatic potential, which stems from the orientation of the DNA phosphate beads, also varies from the interior to the exterior of circular DNA. By estimating the differences in both the major groove width ($\Delta W = W_{out} - W_{in}$) and the electrostatic potential ($\Delta EP = EP_{out} - EP_{in}$) for 15 circular DNA conformations comprised of 50–200 bp (and thus, exhibiting different curvatures), we showed (Figure 8(b)) that a linear relationship exists between DNA curvature and the two parameters mentioned above. The higher the DNA curvature is, the larger the differences between the major groove widths and, consequently, the greater the change in the electrostatic potential between the outside and inside of a DNA surface. Similarly, the DNA helical twist was also found to correlate with DNA groove geometry and, thus, with electrostatic potential. Typically, for a relaxed double helical B-DNA molecule, two strands twist around the helical axis once every 10.4–10.5 bp of sequence. However, for supercoiled DNA structures, the degree of helicity varies. Here, by varying the change in linking number, Δlk, between −2 and 5, we generated eight circular DNA conformations with varying degrees of helical twist and found that, as the degree of twisting in the DNA double helix rises, DNA curvature increases and W_{in} decreases considerably (Figure 8(c)). In particular, major grooves become narrower in order to accommodate the extra turns caused by an increase in the Δlk values. As a result, negatively charged phosphate atoms on the DNA backbone tend

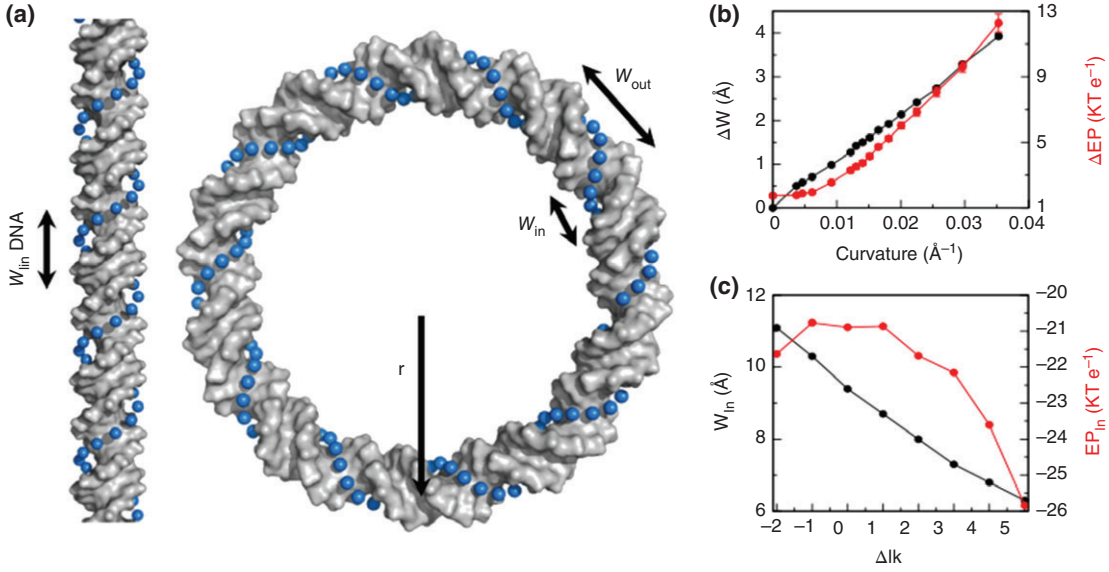

FIGURE 8 | (a) Structural characterization of linear (100 bp) and circular (100 bp) DNA molecules. Unlike the linear B-DNA, whose major groove widths are constant, circular DNA shows wider major groove widths on the exterior of the circle (W_{out}) and narrower major groove widths inside the circle (W_{in}). (b) Correlations between the changes in major groove width (ΔW; black) and the changes in potential energy (ΔEP; red) as functions of DNA curvature. $\Delta W = W_{out} - W_{in}$ and $\Delta EP = EP_{out} - EP_{in}$. Curvature is defined as the inverse of the radius (r) of the DNA, which is zero for a linear DNA molecule. (c) The variations of W_{in} (black) and corresponding electrostatic potential (EP_{in}; red) as functions of the change in DNA linking number (Δlk).

to cluster inside the circular DNA, leading to a strong electrostatic potential.

Protein Dynamics on Linear and Circular DNA as a Function of Salt Concentration

Before proceeding further, it was necessary to identify a suitable salt condition at which the protein–DNA bound state could be characterized. We therefore performed molecular dynamics simulations of Sap-1 paired with linear B-DNA and circular DNA, separately, under a wide range of salt concentrations (0.02–0.20 M) and estimated the propensity of Sap-1 to adopt different search modes (sliding, hopping, and 3D diffusion) as a function of salt concentration. Sliding was defined as $r < 14$ Å, $\theta < 25$ Å, and at least 70% of the recognition region is within the DNA major groove, as described previously.

Our results (Figure 9) suggest that the major difference between the search modes on circular and linear DNA is that while sliding was the preferred mode of translocation on linear DNA under low salt conditions, a circular DNA geometry promoted the hopping of Sap-1 over a wide range of salt conditions (0.02–0.12 M). Despite this variation, we identified 0.02 M as a suitable choice of salt condition at which Sap-1 remained associated with the DNA (3D diffusion at this concentration is only ~1%) irrespective of DNA geometry.

Role of DNA Curvature and Helical Twist in Regulating Target Search Dynamics of Proteins

Having seen that both DNA curvature and helical twist regulate the DNA groove geometry and thereby, modulate the associated electrostatic potential around the molecule, we enquired if these have any impact on the dynamics of the interacting protein. In other words, do DBPs scan DNA differently if the geometry of DNA conformation varies? In Figure 10(a) and (b), we present how the sliding, hopping, and 3D diffusion propensities of the searching protein change with changes in ΔW and Δlk, respectively. The common feature that both the plots share is that 3D diffusion is negligible because of the choice of salt condition (0.02 M). In addition, the sliding propensity decreases and hopping intensity increases with a rise in ΔW and Δlk. This can be explained on the basis that, as curvature increases, ΔEP rises ($\Delta EP \geq 5.0\ k_BT$ for $\Delta W \geq 1.8$ Å), which makes it difficult for the protein to slide from the inside to the outside of the DNA circle. In this context, one should note that the suggested average free energy barrier for sliding dynamics by DBPs is only $-1.1 \pm 0.2\ k_BT$.[15] On the other hand, inside the highly curved DNA structures, the close proximity of the negative charges on the DNA backbone promotes hopping, as can be seen from the steep rise in hopping frequency inside the DNA at high ΔW values (Figure 11(a)–(d)). This result is in agreement with the study of van den Broek et al., who, using optical tweezers and a fast buffer exchange system, observed an enhanced hopping frequency for *Eco*RV on coiled DNA (high curvature) compared with a linear DNA structure.[78] The relationship between search dynamics and Δlk was also found to be driven by the geometry of the DNA groove. For example, sliding is the preferred mode of translocation in under-twisted DNA structures (Δlk = −5 and −2, see Figure 11(e) and (f)), whose major groves are sufficiently wide to easily accommodate the recognition region of Sap-1. The hopping mode of dynamics, however, increases with increased helical twisting, as proxied by Δlk, because of the highly negative electrostatic potential inside the major groove of the over-twisted DNA. Up to a certain point (beyond which the DBP cannot enter the DNA major groove, see Figure 11(h)) the highly negative interior electrostatic potential acts as a promoter of hopping events as it is strong enough to enable mild attraction between the DBP

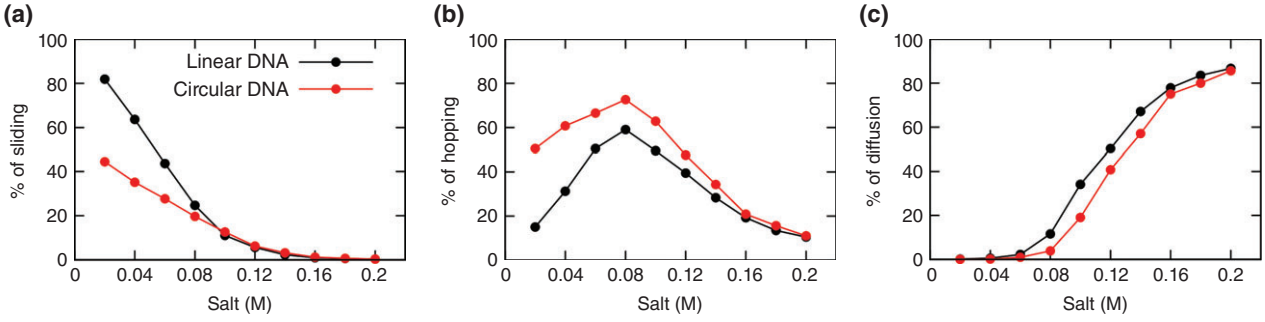

FIGURE 9 | Effects of salt concentration on (a) sliding, (b) hopping, and (c) 3D diffusion for linear and circular DNA.

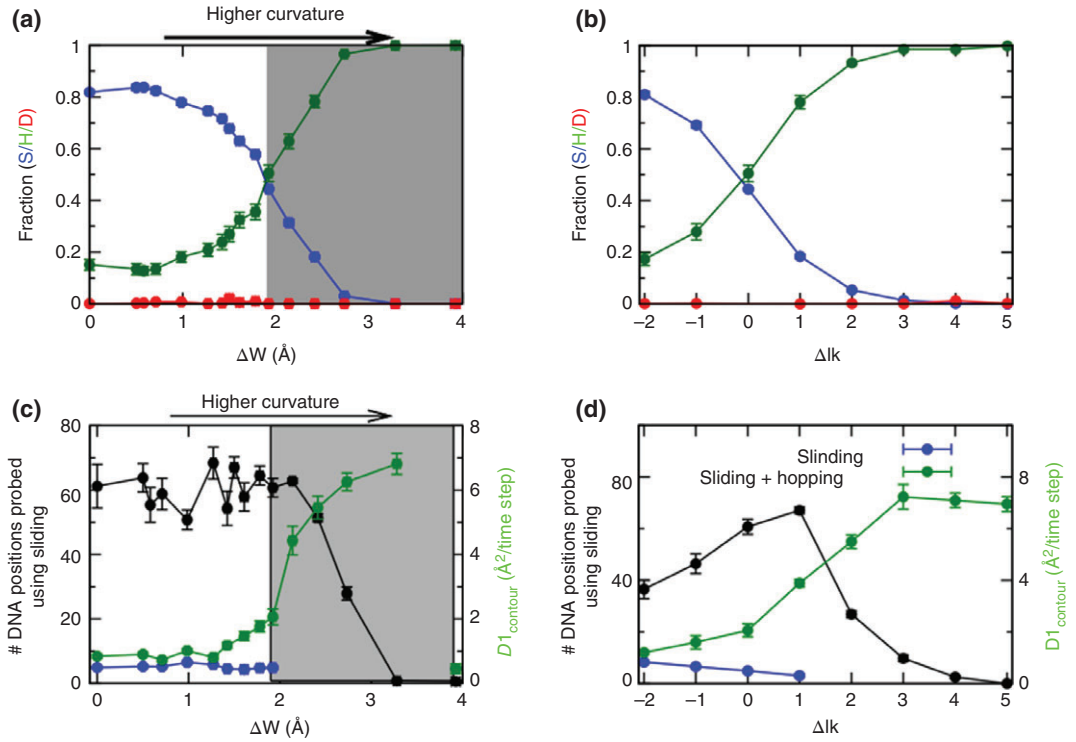

FIGURE 10 | Effect of DNA curvature and supercoiling on the search dynamics of DBPs are measured by calculating proportions of sliding (S), hopping (H), and 3D diffusion (D) adopted by the searching protein as a function of (a) the change in major groove width (ΔW) for circular DNA (circumference, 50–500 bp) and (b) varying numbers of helical twists (Δlk) on 100 bp circular DNA at a 0.02 M salt concentration. The corresponding search efficiency was measured by DNA position probed and the 1D diffusion coefficient (D_1) of the interacting protein. (c) The number of DNA base pairs probed by Sap-1 using sliding dynamics varies with the change in DNA curvature (ΔW). The one-dimensional diffusion coefficient D_1 was calculated for the portions of the simulation during which Sap-1 scanned the DNA contour via pure sliding (blue circles) and for the portions during which Sap-1 was bound to the DNA (green circles) by either sliding or hopping dynamics. As sliding frequency decreases sharply for $\Delta W \geq 1.8$ Å, D_1 values were not calculated beyond this point. Gray shaded region denotes DNA minicircles of circumference ≤100 bp. (d) Variation of same DNA position probed and D_1 as function of change in linking number, Δlk.

and DNA, but not so strong as to prevent dissociation of the DBP from the DNA surface.

The variation in search mode with ΔW and Δlk has far reaching consequences in regulating the target search efficiency of the interacting protein. In order to quantify the effects, we estimated the number of positions probed on DNA by Sap-1 and the related D_1. Corresponding results are shown in Figure 10 (c) and (d), which suggest that for DNA with $0 \leq \Delta W < 1.8$ Å (corresponds to circular DNA with $100 < N_{bp} \leq 500$; i.e., to minimally curved DNA structures), the number of visited sites shows a roughly constant value of 50–70. This means that Sap-1 scanned all these DNA molecules with roughly the same efficiency. However, in highly curved DNA minicircles, where $\Delta W \geq 1.8$ Å ($N_{bp} \leq 100$), the number of probed positions decreases sharply because of a decrease in sliding, even though hopping-assisted diffusion increases along the DNA contour. Along the line, as Δlk increases, the number of probed DNA position increases initially and then, drops for Δlk > 1. Both of these results suggest that the search mechanism does not reflect naive sliding, but rather a tradeoff between sliding and hopping dynamics is essential to attain maximum target search efficiency for DBPs. The balance in turn is strongly correlated with DNA geometry. For example, optimal search efficiency was achieved for $\Delta W \sim 2.1$ Å (where the curvature of the corresponding DNA is 0.02 per Å). Under such conditions, Sap-1 diffuses relatively fast by hopping ($D_1 = 4.43$ Å^2 time step^{-1}) yet still scanned ~62 bp (out of 90) by sliding. Similarly, Sap-1 could scan the maximum number of base pairs (~67 bp out of 100) in circular DNA when the DNA is slightly over-twisted (Δlk = 1, Figure 11(d)). Optimal efficiency was achieved through 78% hopping (fast diffusion, $D_1 \sim 3.88$ Å^2 time step^{-1}) and only 18% sliding dynamics.

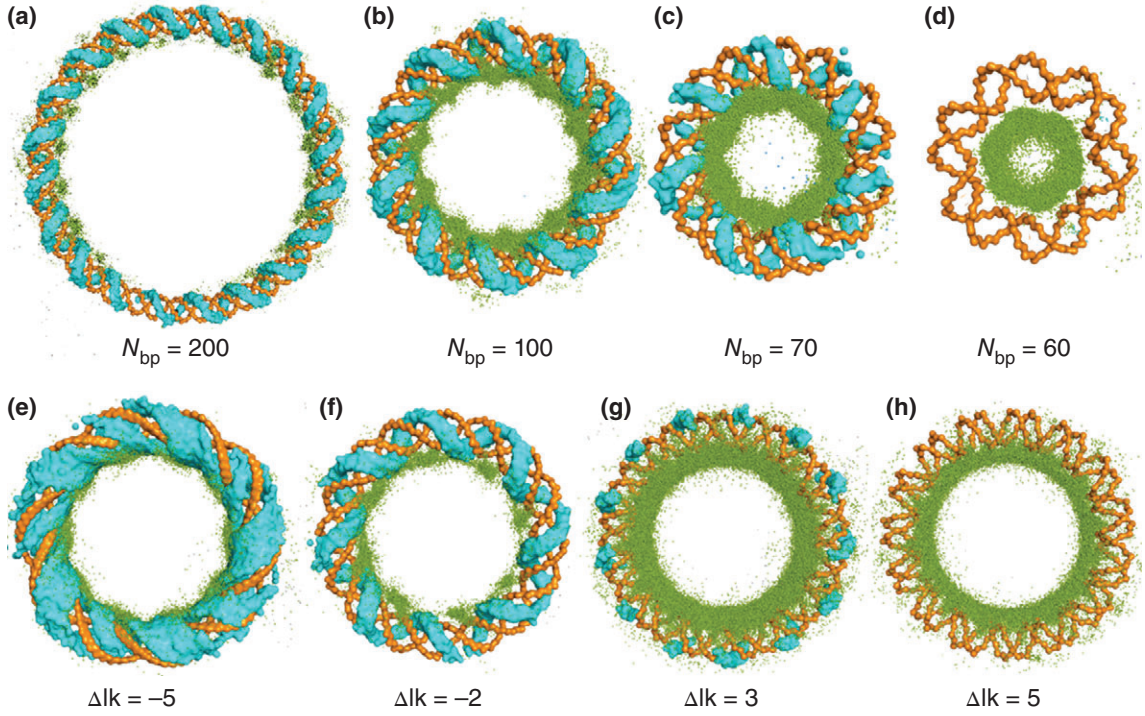

FIGURE 11 | The traces of the search path of Sap-1 around circular DNA of (a) 200 bp, (b) 100 bp, (c) 70 bp, and (d) 60 bp at a salt concentration of 0.02 M. (e–h) The same search paths on 100 bp circular DNA twisted to various extents around the double helix. A negative value of ΔIk denotes an under-twisted DNA structure and a positive value of ΔIk denotes an over-twisted DNA structure. DNA is colored orange, while cyan and green represent the sliding and hopping modes, respectively.

CONCLUSIONS

CG models have provided simplified yet predictive and powerful tools for investigating the molecular determinants that underpin the target search mechanism of DBPs. Successful modeling requires that the researcher select the appropriate level of coarse graining to meet the objective of study. It is thus instructive to pay attention to the subtle details while developing CG models. The CG model should be not only simple enough to capture the question at hand but also should represent the essential physical and chemical properties of the studied systems. We showed that, despite its simplicity, the Debye–Hückel model is well able to represent protein–DNA interactions and to study the linear diffusion of proteins along DNA. Supplementing the CG model by adding solvent effects and ion-specific protein–DNA interactions is expected to further capture the biophysical properties of DNA search.[95] We discuss that, although different CG models that capture the protein molecule at different resolutions may affect the sliding dynamics, they all qualitatively portray the interaction dynamics with DNA molecules. For example, in nonspecific protein–DNA interactions, apparently simple modeling of protein side chains poses a complexity. With the addition of side chains, the effective location of point charges on the corresponding amino acids is shifted by a few angstroms, which may alter the associated strength of the electrostatic potential and, thereby, the search dynamics of DBPs. Indeed, the sliding propensity and the diffusion coefficient can be changed by these coarse-graining nuances. In addition to the geometrical aspects of CG models for nonspecific protein–DNA interactions, there is scope for improving the associated model and energy functions for calculating effective charge–charge interactions beyond the Debye–Hückel potential. An approach to refine the electrostatic interactions is by mapping the energetics from atomistic simulations and accordingly to tune the parameters of the CG model.[96,97]

Researchers must keep in mind that while CG models lack some details which are often included in atomistic models, their simplicity lies in their power. A CG model can become more realistic not only by increasing its molecular resolution but also by calibrating its parameters to fit experimental observables or those obtained from atomistic modeling turning them to be more quantitative. While the degree of coarse-graining may affect the sliding characteristics

to some extent, they all provide insightful analyses on the search mechanisms. The details of the CG model are expected to be more critical for specific protein-recognition where the molecular resolution of the interface is central to specificity and affinity.

ACKNOWLEDGMENTS

This work is supported by DST Inspire Faculty Grant (DST/INSPIRE/04/2013/000100), Government of India, by the Kimmelman Center for Macromolecular Assemblies and the Minerva Foundation, with funding from the Federal German Ministry for Education and Research. Y.L. is the Morton and Gladys Pickman professional chair in Structural Biology.

REFERENCES

1. Garvie CW, Wolberger C. Recognition of specific DNA sequences. *Mol Cell* 2001, 8:937–946.
2. Lavery R, Zakrzewska K. Towards a molecular view of transcriptional control. *Curr Opin Struct Biol* 2012, 22:160–167.
3. von Hippel PH. From "simple" DNA-protein interactions to the macromolecular machines of gene expression. *Annu Rev Biophys Biomol Struct* 2007, 36:79–105.
4. Berg OG, Winter RB, Hippel PH. Diffusion-driven mechanisms of protein translocation on nucleic acids. 1. Models and theory. *Biochemistry* 1981, 20:6929–6948.
5. Halford SE. An end to 40 years of mistakes in DNA-protein association kinetics? *Biochem Soc Trans* 2009, 37:343–348.
6. Hammar P, Leroy P, Mahmutovic A, Marklund EG, Berg OG, Elf J. The lac repressor displays facilitated diffusion in living cells. *Science* 2012, 336:1595–1598.
7. Riggs AD, Bourgeois S, Cohn M. The lac repressor-operator interaction. 3. Kinetic studies. *J Mol Biol* 1970, 53:401–417.
8. von Hippel PH, Berg OG. Facilitated target location in biological systems. *J Biol Chem* 1989, 264:675–678.
9. Gowers DM, Wilson GG, Halford SE. Measurement of the contributions of 1D and 3D pathways to the translocation of a protein along DNA. *Proc Natl Acad Sci U S A* 2005, 102:15883–15888.
10. Wang YM, Austin RH, Cox EC. Single molecule measurements of repressor protein 1D diffusion on DNA. *Phys Rev Lett* 2006, 97:048302.
11. Stivers JT, Schonhoft JD. Timing facilitated site transfer of an enzyme on DNA. *Nat Chem Biol* 2012, 8:205–210.
12. Rowland MM, Schonhoft JD, McKibbin PL, David SS, Stivers JT. Microscopic mechanism of DNA damage searching by hOGG1. *Nucleic Acids Res* 2014, 42:1–9.
13. Porecha RH, Stivers JT. Uracil DNA glycosylase uses DNA hopping and short-range sliding to trap extrahelical uracils. *Proc Natl Acad Sci U S A* 2008, 105:10791–10796.
14. Gorman J, Greene EC. Visualizing one-dimensional diffusion of proteins along DNA. *Nat Struct Mol Biol* 2008, 15:768–774.
15. Blainey P, Luo G, Kou S, Mangel W, Verdine G, Bagchi B, Xie XS. Nonspecifically bound proteins spin while diffusing along DNA. *Nat Struct Mol Biol* 2009, 16:1224–1229.
16. Bustamante C. Facilitated target location on DNA by individual *Escherichia coli* RNA polymerase molecules observed with the scanning force microscope operating in liquid. *J Biol Chem* 1999, 274:16665–16668.
17. Elf J, Li GW, Xie XS. Probing transcription factor dynamics at the single-molecule level in a living cell. *Science* 2007, 316:1191–1194.
18. Graneli A, Yeykal CC, Robertson RB, Greene EC. Long-distance lateral diffusion of human Rad51 on double-stranded DNA. *Proc Natl Acad Sci U S A* 2006, 103:1221–1226.
19. Iwahara J, Clore GM. Detecting transient intermediates in macromolecular binding by paramagnetic NMR. *Nature* 2006, 440:1227–1230.
20. Iwahara J, Clore GM. Direct observation of enhanced translocation of a homeodomain between DNA cognate sites by NMR exchange spectroscopy. *J Am Chem Soc* 2006, 128:404–405.
21. Iwahara J, Zweckstetter M, Clore GM. NMR structural and kinetic characterization of a homeodomain diffusing and hopping on nonspecific DNA. *Proc Natl Acad Sci U S A* 2006, 103:15062–15067.
22. Doucleff M, Clore GM. Global jumping and domain-specific intersegment transfer between DNA cognate sites of the multidomain transcription factor Oct-1. *Proc Natl Acad Sci U S A* 2008, 105:13871–13876.
23. Takayama Y, Clore GM. Intra- and intermolecular translocation of the bi-domain transcription factor Oct1 characterized by liquid crystal and paramagnetic NMR. *Proc Natl Acad Sci U S A* 2011, 108: E169–E176.

24. Marklund EG, Mahmutovic A, Berg OG, Hammar P, van der Spoel D, Fange D, Elf J. Transcription-factor binding and sliding on DNA studied using micro- and macroscopic models. *Proc Natl Acad Sci U S A* 2013, 110:19796–19801.
25. Echeverria I, Papoian GA. DNA exit ramps are revealed in the binding landscapes obtained from simulations in helical coordinates. *PLoS Comput Biol* 2015, 11:e1003980.
26. Bouvier B, Zakrzewska K, Lavery R. Protein-DNA recognition triggered by a DNA conformational switch. *Angew Chem Int Ed Engl* 2011, 50:6516–6518.
27. Takada S, Kanada R, Tan C, Terakawa T, Li WF, Kenzaki H. Modeling structural dynamics of biomolecular complexes by coarse-grained molecular simulations. *Acc Chem Res* 2015, 48:3026–3035.
28. Dans P, Walther J, Gómez H, Orozco M. Multiscale simulation of DNA. *Curr Opin Struct Biol* 2016, 37:29–45.
29. Givaty O, Levy Y. Protein sliding along DNA: dynamics and structural characterization. *J Mol Biol* 2009, 385:1087–1097.
30. Chen C, Pettitt BM. The binding process of a nonspecific enzyme with DNA. *Biophys J* 2011, 101:1139–1147.
31. Bhattacherjee A, Levy Y. Search by proteins for their DNA target site: 1. The effect of DNA conformation on protein sliding. *Nucleic Acids Res* 2014, 42:12404–12414.
32. Khazanov N, Levy Y. Sliding of p53 along DNA can be modulated by its oligomeric state and by cross-talks between its constituent domains. *J Mol Biol* 2011, 408:335–355.
33. Khazanov N, Marcovitz A, Levy Y. Asymmetric DNA-search dynamics by symmetric dimeric proteins. *Biochemistry* 2013, 52:5335–5344.
34. Levy Y, Onuchic JN, Wolynes PG. Fly-casting in protein-DNA binding: frustration between protein folding and electrostatics facilitates target recognition. *J Am Chem Soc* 2007, 129:738–739.
35. Levy Y, Vuzman D, Takayama Y, Iwahara J, Zandarashvili L, Esadze A, Sahu D. Asymmetrical roles of zinc fingers in dynamic DNA-scanning process by the inducible transcription factor Egr-1. *Proc Natl Acad Sci U S A* 2012, 109: E1724-E1732.
36. Marcovitz A, Levy Y. Frustration in protein–DNA binding influences conformational switching and target search kinetics. *Proc Natl Acad Sci U S A* 2011, 108:17957–17962.
37. Marcovitz A, Levy Y. Weak frustration regulates sliding and binding kinetics on rugged protein-DNA landscapes. *J Phys Chem B* 2013, 117:13005–13014.
38. Marcovitz A, Levy Y. Sliding dynamics along DNA: a molecular perspective. In: *Innovations in Biomolecular Modeling and Simulation*. RSC, Biomolecular Sciences, vol. 10. 2010, 237–262.
39. Marcovitz A, Levy Y. Obstacles may facilitate and direct DNA search by proteins. *Biophys J* 2013, 104:2042–2050.
40. Tóth-Petróczy A, Fuxreiter M, Levy Y. Disordered tails of homeodomains facilitate DNA recognition by providing a trade-off between folding and specific binding. *J Am Chem Soc* 2009, 131:15084–15085.
41. Vuzman D, Azia A, Levy Y. Searching DNA via a "Monkey Bar" mechanism: the significance of disordered tails. *J Mol Biol* 2010, 396:674–684.
42. Vuzman D, Levy Y. DNA search efficiency is modulated by charge composition and distribution in the intrinsically disordered tail. *Proc Natl Acad Sci U S A* 2010, 107:21004–21009.
43. Vuzman D, Hoffman Y, Levy Y. Modulating protein-DNA interactions by post-translational modifications at disordered regions. *Pac Symp Biocomput* 2012, 17:188–199.
44. Bhattacherjee A, Levy Y. Search by proteins for their DNA target site: 2. The effect of DNA conformation on the dynamics of multidomain proteins. *Nucleic Acids Res* 2014, 42:12415–12424.
45. Levitt M. A simplified representation of protein conformations for rapid simulation of protein folding. *J Mol Biol* 1976, 104:59–107.
46. Levitt M, Warshel A. Computer simulation of protein folding. *Nature* 1975, 253:694–698.
47. Kaya H, Chan HS. Solvation effects and driving forces for protein thermodynamic and kinetic cooperativity: how adequate is native-centric topological modeling? *J Mol Biol* 2003, 326:911–931.
48. Clementi C, Nymeyer H, Onuchic JN. Topological and energetic factors: what determines the structural details of the transition state ensemble and "en-route" intermediates for protein folding? An investigation for small globular proteins. *J Mol Biol* 2000, 298:937–953.
49. Onuchic JN, Wolynes PG. Theory of protein folding. *Curr Opin Struct Biol* 2004, 14:70–75.
50. Whitford PC, Noel JK, Gosavi S, Schug A, Sanbonmatsu KY, Onuchic JN. An all-atom structure-based potential for proteins: bridging minimal models with all-atom empirical forcefields. *Proteins* 2009, 75:430–441.
51. Cho SS, Weinkam P, Wolynes PG. Origins of barriers and barrierless folding in BBL. *Proc Natl Acad Sci U S A* 2008, 105:118–123.
52. Rao VVHG, Gosavi S. In the multi-domain protein adenylate kinase, domain insertion facilitates cooperative folding while accommodating function at domain interfaces. *PLoS Comput Biol* 2014, 10:e1003938.
53. Gosavi S, Whitford PC, Jennings PA, Onuchic JN. Extracting function from a beta-trefoil folding motif. *Proc Natl Acad Sci U S A* 2008, 105:10384–10389.

54. Zuckerman DM. Simulation of an ensemble of conformational transitions in a united-residue model of calmodulin. *J Phys Chem B* 2004, 108:5127–5137.
55. Best RB, Chen YG, Hummer G. Slow protein conformational dynamics from multiple experimental structures: the helix/sheet transition of arc repressor. *Structure* 2005, 13:1755–1763.
56. Okazaki K, Koga N, Takada S, Onuchic JN, Wolynes PG. Multiple-basin energy landscapes for large-amplitude conformational motions of proteins: structure-based molecular dynamics simulations. *Proc Natl Acad Sci U S A* 2006, 103:11844–11849.
57. O'Brien EP, Ziv G, Haran G, Brooks BR, Thirumalai D. Effects of denaturants and osmolytes on proteins are accurately predicted by the molecular transfer model. *Proc Natl Acad Sci U S A* 2008, 105:13403–13408.
58. Liu Z, Reddy G, Thirumalai D. Theory of the molecular transfer model for proteins with applications to the folding of the src-SH3 domain. *J Phys Chem B* 2012, 116:6707–6716.
59. Linhananta A, Hadizadeh S, Plotkin SS. An effective solvent theory connecting the underlying mechanisms of osmolytes and denaturants for protein stability. *Biophys J* 2011, 100:459–468.
60. Matysiak S, Clementi C. Optimal combination of theory and experiment for the characterization of the protein folding landscape of S6: how far can a minimalist model go? *J Mol Biol* 2004, 343:235–248.
61. Hyeon C, Lorimer GH, Thirumalai D. Dynamics of allosteric transitions in GroEL. *Proc Natl Acad Sci U S A* 2006, 103:18939–18944.
62. Cheung MS, Klimov D, Thirumalai D. Molecular crowding enhances native state stability and refolding rates of globular proteins. *Proc Natl Acad Sci U S A* 2005, 102:4753–4758.
63. Klimov DK, Newfield D, Thirumalai D. Simulations of beta-hairpin folding confined to spherical pores using distributed computing. *Proc Natl Acad Sci U S A* 2002, 99:8019–8024.
64. Levy Y, Wolynes PG, Onuchic JN. Protein topology determines binding mechanism. *Proc Natl Acad Sci U S A* 2004, 101:511–516.
65. Bhattacherjee A, Wallin S. Coupled folding-binding in a hydrophobic/polar protein model: impact of synergistic folding and disordered flanks. *Biophys J* 2012, 102:569–578.
66. Holzgräfe C, Bhattacherjee A, Irbäck A. Hybrid Monte Carlo with non-uniform step size. *J Chem Phys* 2014, 140:044105.
67. Pincus DL, Cho SS, Hyeon C, Thirumalai D. Minimal models for protein and RNA: from folding to function. *Prog Mol Biol Transl Sci* 2008, 84:203–250.
68. Biyun S, Cho S, Thirumalai D. Folding of human telomerase RNA pseudoknot using ion-jump and temperature-quench simulations. *J Am Chem Soc* 2011, 133:20634.
69. Hyeon C, Thirumalai D. Mechanical unfolding of RNA hairpins. *Proc Natl Acad Sci U S A* 2005, 102:6789–6794.
70. Marcovitz A, Levy Y. Arc-repressor dimerization on DNA: folding rate enhancement by colocalization. *Biophys J* 2009, 96:4212–4220.
71. Schlick T, Perisić O. Mesoscale simulations of two nucleosome-repeat length oligonucleosomes. *Phys Chem Chem Phys* 2009, 11:10729–10737.
72. Li R, Ge HW, Cho SS. Sequence-dependent base-stacking stabilities guide tRNA folding energy landscapes. *J Phys Chem B* 2013, 117:12943–12952.
73. Terakawa T, Kenzaki H, Takada S. p53 searches on DNA by rotation-uncoupled sliding at C-terminal tails and restricted hopping of core domains. *J Am Chem Soc* 2012, 134:14555–14562.
74. Mondal A, Bhattacherjee A. Searching target sites on DNA by proteins: role of DNA dynamics under confinement. *Nucleic Acids Res* 2015, 43:9176–9186.
75. Zandarashvili L, Esadze A, Vuzman D, Kemme CA, Levy Y, Iwahara J. Balancing between affinity and speed in target DNA search by zinc-finger proteins via modulation of dynamic conformational ensemble. *Proc Natl Acad Sci U S A* 2015, 112:E5142–E5149.
76. Vuzman D, Levy Y. Intrinsically disordered regions as affinity tuners in protein-DNA interactions. *Mol Biosyst* 2012, 8:45–57.
77. Lomholt MA, van den Broek B, Kalisch SMJ, Wuite GJL, Metzler R. Facilitated diffusion with DNA coiling. *Proc Natl Acad Sci U S A* 2009, 106:8204–8208.
78. van den Broek B, Lomholt MA, Kalisch SMJ, Metzler R, Wuite GJL. How DNA coiling enhances target localization by proteins. *Proc Natl Acad Sci U S A* 2008, 105:15738–15742.
79. Lange M, Kochugaeva M, Kolomeisky AB. Dynamics of the protein search for targets on DNA in the presence of traps. *J Phys Chem B* 2015, 119:12410–12416.
80. Knotts TA, Rathore N, Schwartz DC, De Pablo JJ. A coarse grain model for DNA. *J Chem Phys* 2007, 126:084901.
81. Dans PD, Zeida A, Machado MR, Pantano S. A coarse grained model for atomic-detailed DNA simulations with explicit electrostatics. *J Chem Theory Comput* 2010, 6:1711–1725.
82. Sambriski EJ, Schwartz DC, De Pablo JJ. A mesoscale model of DNA and its renaturation. *Biophys J* 2009, 96:1675–1690.
83. Freeman GS, Hinckley DM, de Pablo JJ. A coarse-grain three-site-per-nucleotide model for DNA with explicit ions. *J Chem Phys* 2011, 135:165104.

84. Rohs R, West SM, Sosinsky A, Liu P, Mann RS, Honig B. The role of DNA shape in protein-DNA recognition. *Nature* 2009, 461:1248–1253.

85. Fogg JM, Randall GL, Pettitt BM, Sumners DL, Harris SA, Zechiedrich L. Bullied no more: when and how DNA shoves proteins around. *Q Rev Biophys* 2012, 45:257–299.

86. Chen Z, Yang H, Pavletich NP. Mechanism of homologous recombination from the RecA-ssDNA/dsDNA structures. *Nature* 2008, 453:489–484.

87. Halford SE, Marko JF. How do site-specific DNA-binding proteins find their targets? *Nucleic Acids Res* 2004, 32:3040–3052.

88. Barozzi I, Simonatto M, Bonifacio S, Yang L, Rohs R, Ghisletti S, Natoli G. Coregulation of transcription factor binding and nucleosome occupancy through DNA features of mammalian enhancers. *Mol Cell* 2014, 54:844–857.

89. Zhou T, Yang L, Lu Y, Dror I, Dantas Machado AC, Ghane T, Di Felice R, Rohs R. DNAshape: a method for the high-throughput prediction of DNA structural features on a genomic scale. *Nucleic Acids Res* 2013, 41:W56–W62.

90. Chen C, Pettitt BM. DNA shape versus sequence variations in the protein binding process. *Biophys J* 2016, 110:534–544.

91. Zheng G, Lu X-J, Olson WK. Web 3DNA—a web server for the analysis, reconstruction, and visualization of three-dimensional nucleic-acid structures. *Nucleic Acids Res* 2009, 37:W240–W246.

92. Macke TJ, Case DA. Modeling unusual nucleic acid structures. In: Leont NB, ed. *Molecular Modeling of Nucleic Acids*. Washington, DC: American Chemical Society; 1998.

93. Lavery R, Moakher M, Maddocks JH, Petkeviciute D, Zakrzewska K. Conformational analysis of nucleic acids revisited: Curves+. *Nucleic Acids Res* 2009, 37:5917–5929.

94. Li L, Li C, Sarkar S, Zhang J, Witham S, Zhang Z, Wang L, Smith N, Petukh M, Alexov E. DelPhi: a comprehensive suite for DelPhi software and associated resources. *BMC Biophys* 2012, 5:9.

95. Marcovitz A, Naftaly A, Levy Y. Water organization between oppositely charged surfaces: implications for protein sliding along DNA. *J Chem Phys* 2015, 142:085102.

96. Terakawa T, Takada S. RESPAC: method to determine partial charges in coarse-grained protein model and its application to DNA-binding proteins. *J Chem Theory Comput* 2014, 10:711–721.

97. Savelyev A, Papoian GA. Chemically accurate coarse graining of double-stranded DNA. *Proc Natl Acad Sci U S A* 2010, 107:20340–20345.

