CHAPTER 10

Sliding Dynamics Along DNA: A Molecular Perspective

AMIR MARCOVITZ AND YAAKOV LEVY*

Department of Structural Biology, Weizmann Institute of Science Rehovot, 76100, Israel
*Email: Koby.Levy@weizmann.ac.il

10.1 Introduction

DNA-binding proteins execute many different genetic tasks, such as transcription, repression, replication, and the repair of damaged DNA. Given the size of the genome, the large number of molecular species *in vivo*, and the crowded environment in which these processes take place, the remarkable efficiency and specificity of protein–DNA recognition present a major theoretical puzzle. In particular, two fundamental issues of great interest in protein–DNA recognition are: how proteins achieve a high degree of selectivity among a huge background of competing nonspecific DNA sequences and the aspect of the rapid recognition rate of the DNA target sequence by the protein.

Over forty years ago, it was suggested that the efficiency of bimolecular interactions could be increased significantly by reducing the dimensionality of the diffusion-based reaction.[1] This hypothesis has been experimentally supported by measuring the association rate of the *lac*-repressor protein to its DNA binding site, which is much higher than the protein–DNA association rate predicted by three-dimensional (3D) diffusion in solution.[2] This result implies that the search for the target site is facilitated by an alternative search mechanism that combines, in addition to 3D random diffusion in solution,

RSC Biomolecular Sciences No. 24
Innovations in Biomolecular Modeling and Simulations: Volume 2
Edited by Tamar Schlick

Published by the Royal Society of Chemistry, www.rsc.org

one-dimensional (1D) sliding of the protein along the DNA, during which the protein binds the DNA nonspecifically.[3–7] Since the pioneering work of Berg *et al.*,[8] it has become well accepted that a protein's search for its target sequence comprises both 1D search (sliding) and 3D search, where the protein dissociates from the DNA into the surrounding solution and reassociates randomly with the DNA at a distant position from the initial dissociation point. In addition to these search mechanisms, the protein may also translocate along the DNA *via* short-range microscopic dissociations to neighboring DNA sites (hopping) or through intersegmental transfers in which the protein directly transfers to a different DNA location by a looped DNA strand, presumably *via* a doubly bound intermediate.[9]

The task of probing the structural and molecular details of nonspecific protein–DNA interactions, which govern the search for the target DNA sequence, is a challenge because of the elusive and transient nature of nonspecific protein–DNA complexes. Nevertheless, several crystal structures of nonspecific or semi-specific protein–DNA complexes have been resolved. In non-specific protein–DNA complexes (*e.g.* the EcoRV,[10] BamHI,[11] *lac*-repressor,[12,13] and BstYI[14] systems), the DNA maintains its canonical B form,[12] and the proteins interact mostly with negatively charged phosphate groups.[15] In general, DNA binding proteins have substantial regions of positive electrostatic potential at their DNA binding interface that form a complementary electrostatic patch to the negatively charged DNA double-helix.[16–18] The importance of electrostatic interactions in dominating nonspecific protein–DNA interactions[19] is supported by a stronger dependence of nonspecific protein–DNA interactions on salt concentration than that seen in specific interactions[20] and by the observation that the protein–DNA interface is more hydrated in the nonspecific complex than in the specific complex.[21,22]

Using a variety of approaches, several studies in recent years concluded that there is a high degree of similarity between the specific and nonspecific binding of DNA binding proteins with DNA. In recent NMR measurements, the structural and kinetic aspects of the nonspecific interaction of the HoxD9 homeodomain with DNA were successfully characterized.[23,24] These studies established that proteins may make use of similar binding interfaces for both nonspecific and specific DNA binding, utilizing the positive electrostatic patch on their surface to maintain a similar orientation with the DNA in the two binding modes. In addition, single-molecule experiments that investigated the linear diffusion of DNA binding proteins along DNA have reported diffusion coefficients for several proteins that are consistent with rotation coupled translocation dynamics, in which the protein propagates one-dimensionally along the DNA while rotating along the helical contour of the DNA.[25] In such a rotation-coupled sliding fashion, a secondary structural element from the protein is able to probe the base-pair content of the DNA grooves and transit to specific binding that involves additional types of interactions (such as hydrogen-bonds with the bases and Van der Waals interactions) once the target DNA sequence is reached.[26]

Recently, a few single-molecule experiments have been designed to investigate the Brownian nature of the translocation of proteins along DNA. These experiments indicated the broad distributions of the 1D diffusion coefficient (D_1) and the diffusion length.[27,28] The D_1 coefficient ranges from $\sim 10^2\,nm^2\,s^{-1}$ to $\sim 10^5\,nm^2\,s^{-1}$, is much smaller than the 3D diffusion coefficient, which is about $10^8\,nm^2\,s^{-1}$. The relationship between the 1D and 3D DNA search mechanisms has been studied for the heterodimeric restriction enzyme BbvCI, which recognizes and cleaves an asymmetrical target site.[5] By probing the cleavage of a DNA substrate containing two sites placed in different orientations to each other and with varying separation distances between the two, the authors have indicated a typical sliding distance of about 50 bp under physiological conditions. This typical sliding distance has been also demonstrated for other proteins, such the restriction enzyme EcoRV.[29]

Here, we review a simplified computational approach presented earlier by Givaty and Levy[30] to explore the mechanism and the molecular details of the search process adopted by the protein to find a DNA target. We elaborate more on the distributions of sliding durations and provide a direct and quantitative indication for rotation-coupled sliding. On account of the elusive nature of nonspecific protein–DNA interactions, we use a reduced model in which the interactions between DNA-binding proteins and DNA are governed solely by electrostatic forces. In particular, we address here the structural details characterizing sliding and hopping, measure 1D diffusion coefficients (D_1) and distributions of sliding durations, and directly observe a rotation-coupled movement of a protein along DNA.

10.2 A Computational Model for Nonspecific Protein–DNA Interactions

10.2.1 A Coarse-grained Model for Simulations of Protein Search along DNA

To explore relatively long timescales of protein translocation on DNA by molecular dynamics simulations, we used a reduced model that allows sliding, hopping and 3D diffusion to be captured from both molecular and mechanistic perspectives. A canonical B-DNA molecule is modeled by three beads per nucleotide, representing the phosphate, sugar, and base groups. Each bead is positioned at the geometric center of the group it represents. The protein is represented by one bead for each residue at the C_α position of that residue, and is placed with the DNA inside a box with dimensions of $200 \times 200 \times 500$ Å. We used a 100 bp DNA molecule that was maintained in a static position at the center of the box and aligned with the Z-axis throughout the simulation. While the DNA remained frozen during the simulations, the protein was flexible and diffused freely within the boundaries of the box. The protein was simulated with a native topology-based model[31–33] that excludes non-native interactions and uses the Lennard-Jones potential to represent native contact interactions.

The dynamics of the protein–DNA system is simulated with the Langevin equation:[31–33]

$$m_i \dot{\nu}_i = \mathbf{F}_i - \gamma m_i \nu_i + \mathbf{R}_i(t) \tag{10.1}$$

where m_i, ν_i and $\dot{\nu}_i$ are the mass, velocity, and acceleration of the i'th bead, respectively. $\mathbf{F}_i$ is the force applied on the i'th bead, and $\mathbf{R}_i$ is a stochastic variable drawn from a Gaussian distribution with zero mean and variance:

$$\left\langle \mathbf{R}_i(t)\mathbf{R}_j(t+\tau) \right\rangle = 2m_i\gamma k_B T\delta(\tau) \tag{10.2}$$

In this study, γ was set to 0.01. We note that the random walk motion of the proteins during sliding along the DNA is not achieved when random noise is excluded from the simulations (*i.e.*, with pure Newtonian dynamics).

On account of the elusive nature of non-specific interactions between protein and DNA, which are central to the DNA search process by proteins, we used a simplified representation for non-specific DNA binding in which the interactions between proteins and DNA are governed solely by electrostatic forces. Beads representing the charged amino acids (Arg, Lys, Asp, and Glu) and the DNA phosphate groups were charged in the model. We employed the Debye-Hückel potential to describe the electrostatic interactions between the protein and DNA:[34]

$$U_{Debye\text{-}Huckel} = {K_{Coulomb}}^{B(\kappa)} \sum_{i,j} \frac{q_i q_j \exp(-\kappa r_{ij})}{\varepsilon r_{ij}} \tag{10.3}$$

The Debye-Hückel theory predicts the range of the electrostatic influences of an ion to be the Debye screening length κ^{-1}. Linearization of the Poisson-Boltzmann equation yields the relation:

$$\kappa^2 = \frac{8\pi N_A e^2 \rho_A C_s}{1000\ \varepsilon k_B T} \tag{10.4}$$

where N_A is Avogadro's number, ρ_A is the solvent density, e is the proton charge, ε is the solvent dielectric constant, and C_s is the ionic concentration in molar units ($C_s = 0.5\sum c_i {q_i}^2$, where c_i is the molar concentration of ion with charge q_i). In equation 10.3, q_i is a point charge of the i'th bead, r_{ij} is the distance between two charged beads (bearing either the same or opposite charges), and $K_{\mathrm{Coulomb}} = 4\pi\varepsilon_0 = 332$ kcal mol^{-1}. $B(\kappa)$ is a salt dependent coefficient that is related to the Debye screening length and the ion radius (a_r) by $\exp(\kappa a_r)/(1+\kappa a_r)$, giving rise to $B(\kappa) \approx 1$ for dilute solutions. The Debye-Hückel model has been used in the past to address the energetics and dynamics of various biomolecular systems, such as RNA folding;[35] the conformational stability of long DNA;[36] protein–DNA binding;[37] and DNA–DNA association.[38] While

the Debye-Hückel model is a powerful means of introducing the salt effect of screening electrostatic interactions into the Coulomb potential, one should be aware of its approximations. The model is valid for relatively dilute solutions, as it approximates that the potential energy of an ion is determined by pairwise interactions with other neighboring ions. The detailed effects of higher salt concentrations and of ion condensation on DNA have to be studied using the nonlinearized Poisson-Boltzmann equation, as well with atomistic simulations that can elucidate the dynamics of the ionic layer during sliding on the DNA.

Since the model is simplified, the distances between the charged beads of the protein and the charged DNA beads are larger (as not all atoms are represented and the charges are placed at the phosphate and C_α beads). Consequently, the salt concentrations reported in this review are several times smaller than the experimental ionic strengths. The sliding dynamics was explored at salt concentrations in the range of 0.01–0.3 M in simulations of 10^7 time steps. We point out that due to coarse-graining of the model, one cannot easily convert the time steps to realistic time. Nevertheless, our model reproduces the ratio $D_3/D_1 \sim 100$ as characterized experimentally as well as other kinetic and mechanistic features such as the dependence of the linear diffusion on salt concentration. 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 non-specific complex than in the specific complex.[21,22]

10.2.2 Structural Classification of Protein Sliding, Hopping, and 3D Diffusion

During the simulation, the protein was categorized as performing a sliding, hopping, or 3D search of DNA. One can imagine a search with continuous transitions from sliding to hopping, or from hopping to 3D search. To differentiate *in silico* between protein sliding, hopping, and free 3D diffusion, it is necessary to provide a clear definition for each of them. A simulation frame was considered to show 3D diffusion if the protein was farther than 32 Å from the main DNA axis, since the electrostatic energy drops at this distance to about 2% of the energy in sliding conformations at low salt concentrations. A snapshot was classified as showing a sliding search mode if three criteria were met. The first criterion was that 70% of the recognition region must be in contact with a single groove (one can determine if a given protein atom resides in the major or minor groove of the DNA based on the distance between the two phosphates from each DNA chain that are closest to the atom of interest). The second criterion was that the distance of the center of mass of the recognition region from the DNA must be up to 10 Å longer than that measured in the crystal structure. The third criterion required that the orientation angle be less than 90° (where the orientation angle to DNA was defined as the angle between the geometric center of the recognition region of the protein, the geometric center of the protein, and the point on the DNA's main axis perpendicular to the protein's geometric center). If the protein was at a distance of

less than 32 Å from the DNA and did not meet the criteria for the sliding mode, the frame was classified as representing protein hopping along the DNA.

10.2.3 DNA Bead Radius Parameterization: The Effect of the Major Groove Width on Sliding

In addition to the electrostatic forces between all charged residues and phosphate beads, each bead of the protein had a repulsion potential from all of the DNA beads modeled by $(\sigma_{ij}/r_{ij})^{12}$,[30] where σ_{ij} equals the sum of the hard-sphere repulsion radii of the protein and DNA beads ($\sigma_{ij} = C_{\mathrm{rep,prot}} + C_{\mathrm{rep,DNA}}$). We set $C_{\mathrm{rep,prot}}$ and $C_{\mathrm{rep,DNA}}$ to be 2 Å and 3.7 Å respectively (*i.e.*, $\sigma_{ij} = 5.7$ Å, according to the shortest distance between the C_α atoms and the DNA beads found in a survey of crystallographic protein–DNA complexes in which the DNA is recognized *via* an α-helix).[30]

The repulsion radius for the protein beads is similar to those used in previous coarse-grained simulations.[31–33] The choice of the repulsion radius for the DNA beads may influence the sliding behavior of the protein. Figure 10.1a shows the effect of the DNA repulsion radius $C_{\mathrm{rep,DNA}}$ and the salt concentrations on the total fraction of sliding across the entire search process undertaken by the Sap1 protein. For DNA beads with a small repulsion radius of 1 or 2 Å, the protein may protrude into the DNA grooves relatively easily through its α-helical recognition region, resulting in a high fraction of sliding events (>0.75) even at relatively high salt concentrations of 90 mM. The high propensity for sliding dynamics is accompanied by very few hopping events

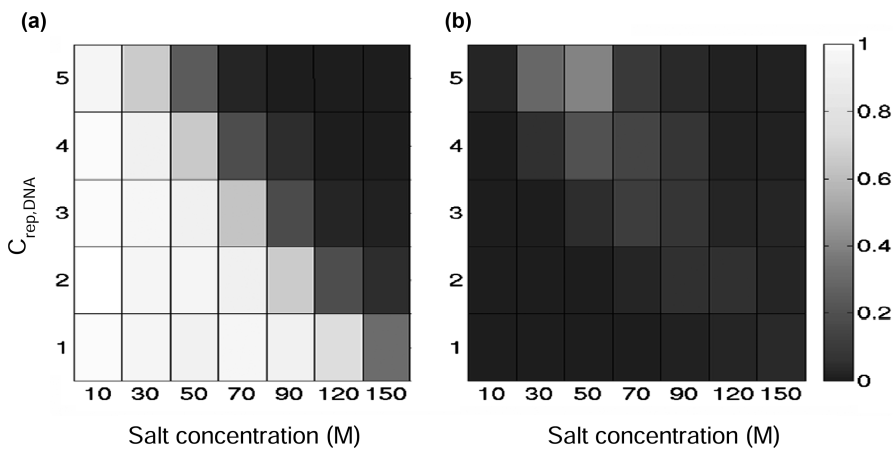


Figure 10.1 Color-coded maps for the fractions of sliding and hopping occurring under varying salt concentrations for the Sap1 protein. The maps were obtained by averaging 10 simulations for each system. Lighter colors indicate a larger fraction of the protein in the particular search mode. (a) Fractions of sliding for systems with different DNA repulsion radii, $C_{\mathrm{rep,DNA}}$, and (b) Fractions of hopping.

(Figure 10.1b). When the DNA repulsion radius $C_{rep,DNA}$ is in the range of 3–4 Å, the transition of the protein from the sliding mode to a 3D diffusive mode with increasing ionic strength is smoother (Figure 10.1a) and the protein is shown to dissociate from the DNA at an ionic strength of 70–90 mM, where the fraction of sliding events drops to less than 0.2. Moreover, for these values of $C_{rep,DNA}$, simulation snapshots that correspond to hopping events could be distinguished at intermediate salt concentrations of 50–70 mM (the fraction of hopping performed was ~0.2, Figure 10.1b). A larger repulsion radius of 5 Å, which could mimic the existence of a water layer, shows an overall similar sliding behavior to that observed for $C_{rep,DNA}$ values of 3.5–4 Å, with the transition from pure sliding (fraction of sliding >0.75) to hopping + 3D diffusion occurring at an ionic strength of 30–50 mM (Figure 10.1a). A more significant fraction of hopping events (>0.4) is observed for a $C_{rep,DNA}$ of 5 Å compared to hopping in systems with lower DNA repulsion radii values (Figure 10.1b).

Figure 10.2a–c shows the 15 residue recognition helix of Sap1 (red spheres with $C_{rep,prot} = 2$ Å) during its interaction with the major groove of a B-DNA molecule in a coarse-grained phosphate-sugar-base model, with $C_{rep,DNA}$ of 1, 3 and 5 Å shown by black spheres with the correspondingly modeled radii (Figure 10.2). For each DNA molecule, the real molecular surface is shown in a transparent gray color. One can observe that DNA with $C_{rep,DNA}$ of 3 Å

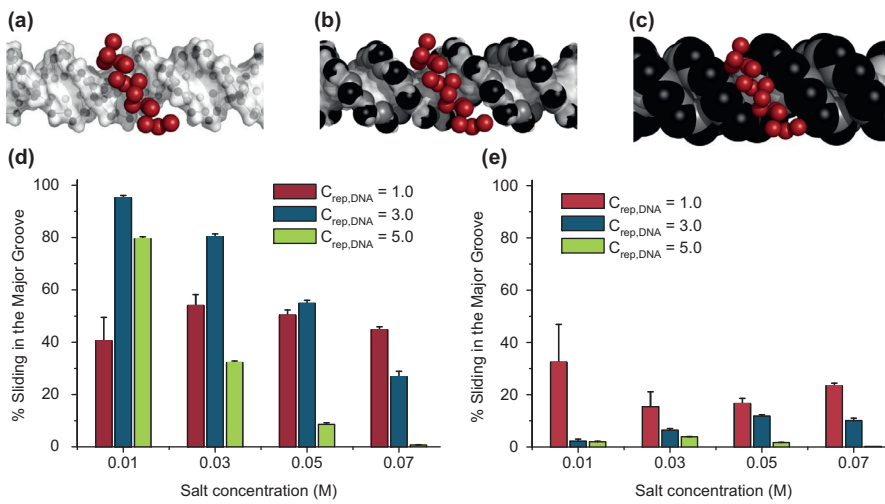


Figure 10.2 Effects of DNA atom repulsion radii ($C_{rep,DNA}$) on the statistics of sliding and major/minor groove sampling by the Sap1 protein. (a–c) Illustrations from a coarse-grained protein–DNA model showing the red spheres of the recognition helix of the protein interacting with the major groove of a B-DNA molecule. Repulsion radii ($C_{rep,DNA}$) are 1 Å (a), 3 Å (b) and 5 Å (c), shown in black. The real molecular surface of the DNA is shown in transparent gray. (d, e) Histograms showing the percentage of sliding via the major (d) or minor (e) grooves (average of 10 simulations at each salt concentration and DNA model).

(Figure 10.2b) resembles more closely the real molecular DNA surface than $C_{rep,DNA}$ of 1 Å, which is much less bulky than real DNA (Figure 10.2a) or $C_{rep,DNA}$ of 5 Å, which significantly exceeds the real DNA molecular surface (Figure 10.2c).

For the Sap1 protein, we expect that, during sliding, the recognition helix will probe the DNA major groove.[30] For each of the DNA models shown in Figures 10.2a–c (*i.e.*, with $C_{rep,DNA}$ of 1, 3 and 5 Å), we checked the extent to which the protein may distinguish between major and minor DNA grooves during simulations. For simulations carried out in salt concentrations of 10–70 mM, we calculated the fraction of the snapshots in which the protein samples the major and the minor groove of the DNA. In each snapshot, we determined in which groove(s) each of the 15 residues belonging to the recognition helix of Sap1 resides (for further details on major/minor groove determination see section 10.2.2 above), and the distance of the bead at the middle of the helix from the main DNA axis. We classified a snapshot as belonging to a major/minor groove sampling if more than 85% percent of the recognition helix beads were in the major/minor groove, the distance of the bead from the middle of the helix to the main DNA axis was <16 Å, and the overall snapshot was classified as a sliding snapshot according to the definition provided above (section 10.2.2).

Figures 10.2d and 10.2e show the fractions of major- and minor-groove sampling (respectively) in the three DNA systems with $C_{rep,DNA}$ of 1, 3, and 5 Å. For $C_{rep,DNA}$ of 1 Å, major groove sampling is relatively low (less than half of the total search) and is relatively unaltered by increasing ionic strength (Figure 10.2d). The fraction of minor-groove sampling is also low (ranging from $\sim$15% to $\sim$35% of the total search, Figure 10.2e), indicating that widening of both grooves due to the low repulsion radius of the DNA beads enables the recognition helix to protrude into both grooves indiscriminately. The constant sampling of DNA grooves with $C_{rep,DNA}$ of 1 Å in close proximity to the main DNA axis, which is favored by the attractive negative charges of the phosphate beads, results in an over classification of snapshots as sliding (Figure 10.1a). For protein–DNA system with $C_{rep,DNA}$ of 3 Å, sampling of the major groove predominates, as is expected for Sap1[30], and the fraction of major-groove sampling decreases with increasing ionic strength (Figure 10.2d). The overall sampling of the minor groove by the recognition helix is transient and is more probable at higher salt concentrations (Figure 10.2e). The overall major-groove sampling trends observed for protein–DNA systems with $C_{rep,DNA}$ of 5 Å are similar to those observed for $C_{rep,DNA}$ of 3 Å, although the absolute values are lower (Figure 10.2d). Sampling of the minor groove in this bulky DNA model is very unlikely (Figure 10.2e).

We conclude from this analysis that the choice of 3.7 Å for $C_{rep,DNA}$ is a reasonable approximation as it resembles the distance observed in crystal structures.[30] Specifically, it captures the essential search properties (such as sliding fraction and major/minor-groove discrimination) for DNA–binding proteins that interact with the DNA using an α-helical region, and it resembles the expected behavior when a water layer is present around the DNA (*i.e.*, of DNA with $C_{rep,DNA}$ of 5 Å) although this aspect is not directly addressed in this review.

10.3 Results and Discussion

10.3.1 Structural Characterization of Protein during Sliding and Hopping on DNA

10.3.1.1 Protein Sliding: Electrostatic Interactions Dictate Sliding Conformation

Several kinetic experiments in solution have provided evidence for a search mechanism that involves both 1D and 3D components.[39] These experiments show that the rate of specific DNA target binding significantly increases when the nonspecific DNA segments surrounding the target are longer[5,23,24,29] and support the role of sliding in facilitating the target search process.[3,40] Recent developments in single-molecule techniques that allow a direct observation of tagged proteins moving along DNA have made it possible to address additional details of the process, such as obtaining lengths of protein translocation on DNA and diffusion coefficients of protein movement along DNA.[25,41] However, most experiments are limited to providing structural criteria that distinguish between protein sliding, hopping, and intersegmental transfer. Computational tools have the potential to structurally characterize the sliding conformation of many DNA-binding proteins, a task which is formidable to undertake by traditional X-ray crystallography approaches.

A recent computational study has explored the molecular details of sliding, its driving forces, and the interplay between DNA search by hopping and sliding using molecular-dynamics simulations in which protein–DNA interactions are represented solely by electrostatic forces.[30] In the study, the properties of protein sliding along canonical B-DNA were studied for several α-helical DNA-binding proteins (engrailed homeodomain HoxD9, Sap1, and Skn1), as well as for an RNA-binding protein (Barnase) and were compared to a non-DNA-binding protein (SH3 domain). It was shown that, during the simulations, the DNA-binding proteins as well as the RNA-binding protein (Barnase) remain in proximity to the main axis of the DNA while randomly diffusing along it. In contrast, the SH3 domain randomly diffuses in the bulk and is not attracted at all to the DNA. This indicates that the electrostatic interactions are sufficient to dominate nonspecific protein–DNA interactions. The ability of Barnase to bind DNA is explained by the charge distribution on its surface. The narrow distribution of the distances of each protein from the DNA (Figure 10.3a) illustrates the electrostatic attraction of nucleic acid binding proteins to DNA.

In a kinetic study of the translocation of the BbvCI restriction enzyme that recognizes and cleaves an asymmetrical target site, Gowers *et al*. have shown that a DNA substrate that contains two target sites in the same orientation is cleaved more efficiently when the separation between the two sites is <50 bp.[5,39] This experiment shows that, in addition to a 3D search in solution, the enzyme one-dimensionally slides along the DNA while maintaining the same orientation relative to the DNA axis. NMR experiments also support the notion that the proteins use a similar interface when sliding on DNA and when

they interact specifically with the target site.[23,24,42] In molecular dynamics simulations, HoxD9, Sap1 and Skn1 were also shown to maintain a relatively fixed orientation relative to the DNA during sliding.[30] The observation that the orientation angle remains relatively fixed throughout a sliding event indicates that the DNA-binding proteins interact with DNA through a recognition region that faces the DNA during the sliding simulation. Accordingly, DNA binding proteins exhibit a narrow distribution of orientation angles with the DNA (Figure 10.3b), which implies the existence of a structural preference for non-specific DNA binding. In comparison with the three DNA-binding proteins, Barnase (RNA-binding protein) shows a wider distribution, and SH3 shows no orientational specificity at all with respect to the DNA, as reflected by an angle distribution that is almost random.

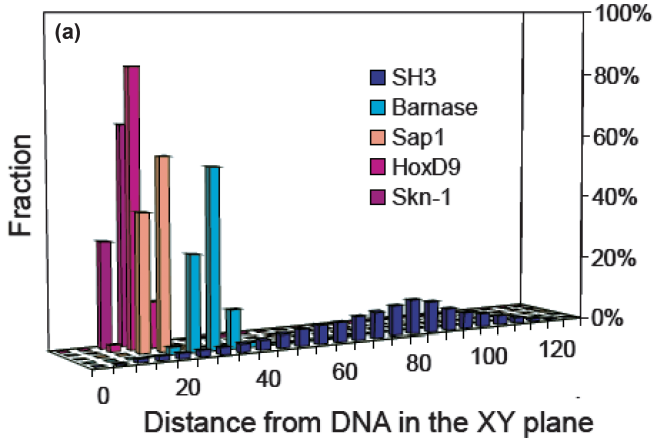
Many DNA binding proteins utilize an alpha-helical motif that probes the major groove of the double stranded DNA during the formation of specific protein–DNA interactions.[43] During sliding, the protein may use non-specific interactions with the DNA that share a certain degree of similarity with the interactions that construct the complex with the specific DNA target. Figure 10.3c shows the percentage of the recognition region that is situated in the major groove for the DNA-binding proteins and for Barnase during the simulations.[30] More than 80% of the recognition region of the three DNA-binding proteins resides in the major groove during the simulations, whereas Barnase shows no clear tendency toward any groove.

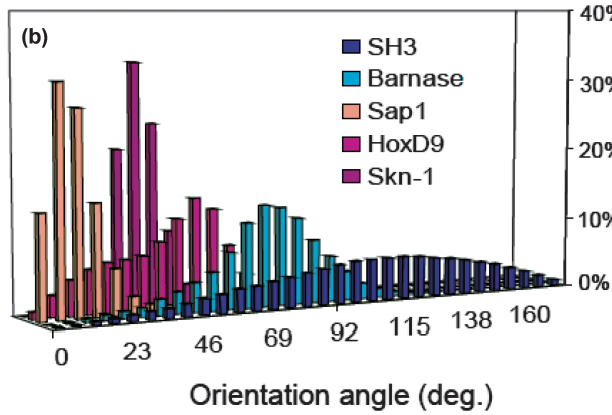
Figure 10.4a illustrates the similarity that the sliding configuration of the Sap1 protein shares with the configuration of the specific complex in the crystal structure (green and gray lines, respectively). The plot indicates the distances of each of the protein C_α atoms from their nearest DNA backbone phosphate atom. The two distance profiles are very similar (with $R^2 = 0.75$), in particular in the recognition region, which is indicated by a green bar (residues 55–69). Figure 10.4b shows a typical sliding snapshot (left) where the recognition helix (in green) is shown to face the DNA and protrude into the major groove, in agreement with the measure presented in Figure 10.3c.

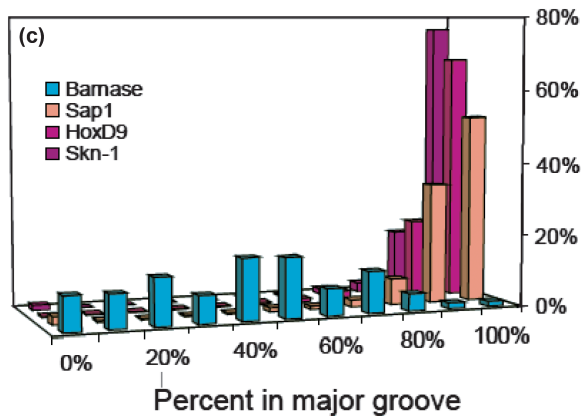
10.3.1.2 Structural Characterization of Protein–DNA Interactions during Hopping

During 1D protein translocation along the main axis of the DNA using the sliding mode, a protein moves bidirectionally and displays properties consistent with a one-dimensional random walk driven by thermal diffusion. It was earlier suggested that an optimal pathway for a target search would involve repetitive dissociations of the protein from the DNA that are followed by reassociations at other locations on the nucleic acid.[4] This scenario enables the protein to scan a greater proportion of the DNA and avoid redundant sequence scanning by sliding.

An important ingredient in the search process is hopping, in which the protein remains in the vicinity of the DNA and performs short-range dissociations to neighboring DNA sequences. There are typically less


(a)
SH3
Barnase
Sap1
HoxD9
Skn-1
Fraction
100%
80%
60%
40%
20%
0%
0
20
40
60
80
100
120
Distance from DNA in the XY plane


(b)
SH3
Barnase
Sap1
HoxD9
Skn-1
40%
30%
20%
10%
0%
0
23
46
69
92
115
138
160
Orientation angle (deg.)


(c)
Barnase
Sap1
HoxD9
Skn-1
80%
60%
40%
20%
0%
0%
20%
40%
60%
80%
100%
Percent in major groove

charge-charge interactions between the protein and the DNA during hopping compared to sliding and therefore 1D diffusion increases with increasing ionic strength.[8,30,44] Figure 10.4a shows that although it is close to the DNA, the configuration of the Sap1 protein (blue line) relative to the DNA is less similar to the specific protein–DNA configuration during hopping than sliding (with $R^2 = 0.39$). In addition, the recognition helix is mostly excluded from the major DNA groove during hopping, giving rise to a one-dimensional diffusion of the protein that is not bound to the DNA helical path. Figure 10.4b shows a random hopping snapshot where the protein is slightly more distant from the DNA than during a sliding snapshot and the recognition helix (in green) is outside the DNA major groove.

The structural properties of HoxD9, Sap1, and Skn1 were characterized from molecular dynamics simulations to determine the extent to which hopping as well as dissociation modes are governed by the specific protein–DNA interface.[30] Figures 10.4c and 10.4d describe the configuration of the Sap1 protein relative to the DNA (distance from the DNA axis and orientation angle) during hopping and dissociation in comparison with sliding. During hopping, the protein adopts a conformation in which the binding region mostly faces the DNA, but the orientation is much less restricted compared to that displayed during sliding. In hopping, the protein is attracted to the DNA but does not follow the helical backbone rail as occurs in sliding. The detected hopping events reported in the simulations of HoxD9, Sap1, and Skn1[30] mostly occur on short timescales and span only a few DNA base pairs. Therefore, such transient hopping events can be viewed as bridging two consecutive sliding events on neighboring DNA fragments. There is, however, experimental evidence for proteins, such as the UL42[45] processivity factor, that predominately use hopping rather than sliding to translocate one-dimensionally along DNA.

10.3.2 DNA Search Efficiency: Combining 1D and 3D Search Modes Enhances DNA Scanning

10.3.2.1 Effects of Salt Concentration on Search Efficiency

Electrostatic interactions dominate non-specific protein–DNA interactions[15,19] as supported by a salt concentration dependence that is stronger than that

Figure 10.3 Protein structural characteristics during sliding on DNA. (a) Histograms showing the location of the protein's center of mass relative to the DNA axis (Z-axis). While nucleic acid binding proteins remain in close proximity to the nucleic acid, the SH3 domain is mostly detached from the DNA as indicated by its wide distribution of distances. (b) Histograms of the orientations angles between the protein and the DNA. The distributions displayed by DNA-binding proteins indicate that the proteins scan the DNA using a distinct interface. Barnase and SH3 domains, in contrast, present much wider distributions. (c) Percentage of the recognition region situated in the major groove. For DNA binding proteins, more than 80% of the protein's recognition region is situated in the major groove, while Barnase shows no groove preference.

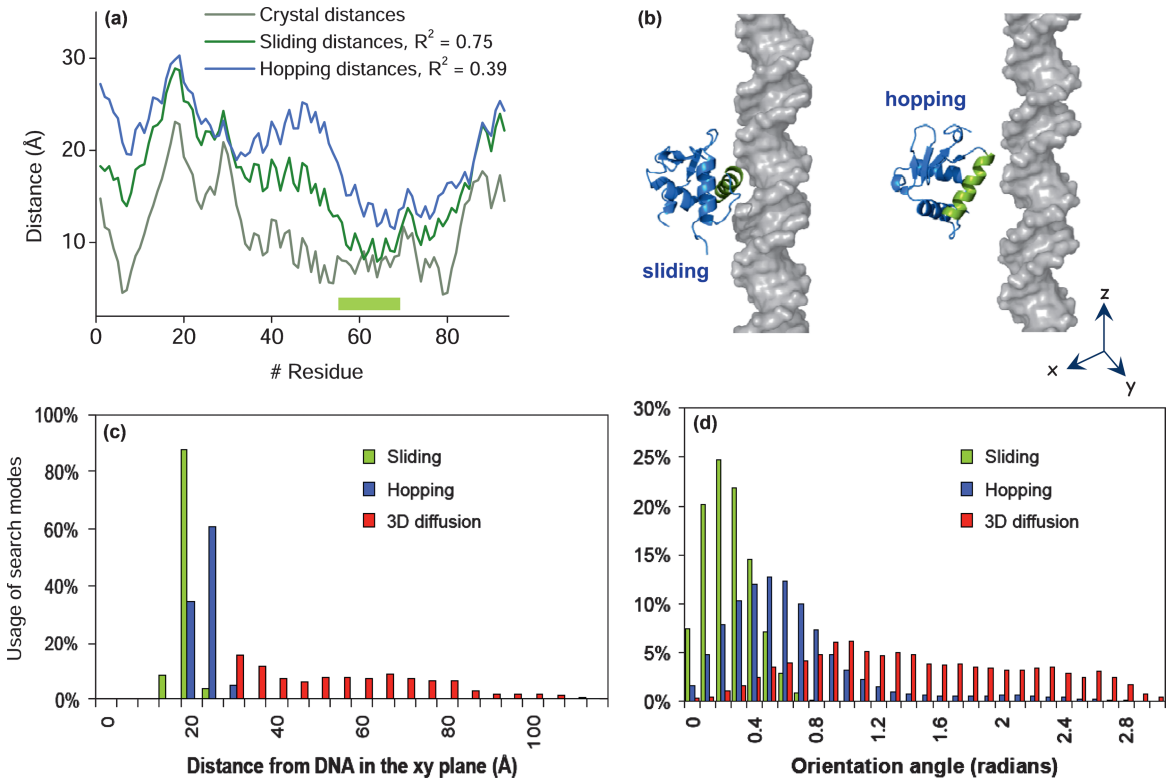


Figure 10.4 Structural properties of the Sap1 protein during sliding, hopping and 3D diffusion. (a) Average distances of each protein residue from the closest DNA phosphate bead in the crystal structure (gray line), during sliding (green line), and during hopping (blue line). The lower green bar indicates the sequence position of the protein's recognition helix. (b) Representative snapshots of the protein during sliding (left) and hopping (right). Distributions of the distances between the protein's center of mass and the main DNA axis (c) and the orientation angle of the protein to the DNA (d) during the three search modes.

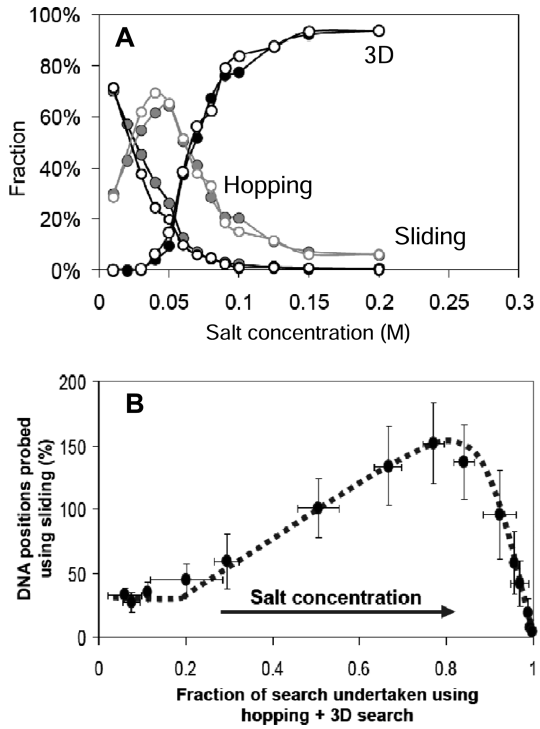


Figure 10.5 Effects of protein flexibility and salt concentration on the interplay between sliding, hopping, and 3D diffusion for the Sap1 protein. (a) The balance between the three search mechanisms for the Sap1 protein in different salt concentrations at $0.9T_F$ (where T_F is the folding temperature of flexible Sap1). Empty circles represent a flexible protein with a 10–12 van der Waals potential representing all protein native contacts. Full circles represent a rigid protein model with harmonic constraints applied to all native protein contacts to restrict their fluctuations. (b) Percentage of positions probed by the protein using sliding during the simulation as a function of the fraction of non-sliding conformations adopted (*i.e.*, as a fraction of the sum of hopping and 3D diffusion).

which exists in specific interactions.[20] Therefore, the balance between DNA sliding, hopping, and 3D search might be affected by either salt concentration or temperature as both can decrease the strength of charge–charge interactions. Simulations of the Sap1 protein at a wide range of salt concentrations and at two temperatures (both below the equilibrium folding temperature of the protein) show that, as the salt concentration increases, the protein becomes more dissociated and the usage of sliding decreases[30] (Figure 10.5a). For the Sap1 protein, the authors have also addressed the interplay between protein flexibility and the properties of the DNA search mechanism by quantifying the fraction of sliding, hopping, and 3D diffusion in a protein model that is completely rigid, and found that the partitioning between the three search modes under each ionic strength condition is similar for both the flexible protein model and the rigid model[30] (Figure 10.5a).

To explore the relationship between the ionic strength of the environment and the search efficiency of the DNA, Givaty and Levy defined a measure called Probed Position, which indicates the number of new DNA sites that are sampled by sliding (*i.e.*, when a direct readout of local DNA sequence by the protein helical recognition site is enabled).[30] In this measure, any new DNA site (base-pair equivalent) that is visited by sliding is added to the Probed Position measure unless it was already been scanned earlier in the same sliding cycle. When the protein dissociates from the DNA, marked positions are erased, while the number of Probed Positions is left unchanged. The marked positions are erased upon dissociation because the probability of the protein reassociating at the exact same position on the DNA *in vivo* is very low. Since the model does not include a specific site, the Probed Position value serves as a measure for the fraction of DNA sites that are scanned using sliding.

Figure 10.5a shows search efficiency as a function of the percentage of nonsliding events performed during the search. When the protein utilizes sliding in most of the search (at a low salt concentration), search efficiency is low because the search is mostly performed locally. At a high salt concentration, the efficiency of the DNA search is also low as most of the electrostatic interactions are screened and the protein mostly diffuses three-dimensionally around the DNA rather than sliding on it. At a moderate salt concentration that gives rise to ~20% sliding (as a proportion of the total search), the DNA-binding protein searches the DNA by an optimal combination that yields an efficient search. This result is in agreement with theoretical models[4,7] suggesting that an optimal search strategy requires a combination of the different search mechanisms and indicates the importance of hopping and dissociation events for an efficient search. Halford *et al.*, for example, have indicated that an optimal target association rate is achieved with alternating cycles of 1D and 3D diffusion where the lengths sampled by each sliding event are ~10% of the target size.[4] Measurements of the translocation of the BbvCI restriction enzyme along the DNA[5] indeed occur for short distances of 30 base-pairs and suggest that sliding is crucial for local DNA scanning, while hopping and dissociation play important roles in reaching distant DNA regions. Below, we further discuss the spatial and temporal durations of individual sliding events obtained from molecular dynamics simulations under varying ionic strength conditions.

10.3.2.2 Distributions of Sliding Durations and Protein Translocation along DNA during Sliding

Under high salt concentrations, dissociation events of the protein from the DNA become dominant. The efficiency of the DNA search by the protein (as measured by the Probed Position in Figure 10.5b) is therefore low, as the overall number of base pairs probed by the protein is small. Under low salt conditions, however, the observed Probed Position from the simulations is also low as the partition between 1D search and 3D search is far from optimal. We examine here the distributions of the sliding durations and their dependence on

the salt concentration, and observe the redundancy with which DNA sites are visited by the protein under varying salt conditions.

For non-specific binding, the motion of the protein along DNA can be modeled as a random walk driven by thermal diffusion.[4] The mean position of the protein is at the initially bound site and the distribution of visited sites around the mean broadens as the sliding duration increases at a rate of $N^{0.5}$, where N is the number of steps. Figure 10.6a on the left hand side illustrates such a scenario for a prolonged sliding event, where the protein repeatedly re-probes sites it has already visited. The right hand side of Figure 10.6a illustrates an analogous scenario in which the duration of the sliding event is shorter, and thus the protein performs no more than one repeat over visited DNA sites. The histograms obtained from the simulations of Sap1 shown in Figure 1.6b demonstrate that at a lower salt concentration of 0.03 M, the search process is dominated by relatively prolonged sliding events in comparison with the length of sliding events observed at higher salt concentrations of 0.07, 0.09, and 0.15 M. At high salt concentrations of 0.09 M and 0.15 M, the sliding events are relatively short, and the overall proportion of 1D search events is lower (see the decrease in the area under the histograms with increasing salt concentration).

To quantify the extent of search redundancy during sliding, we examined each sliding event from the simulations individually and calculated two values: MSD_z and d_z. MSD_z is the overall distance travelled by the protein along the DNA axis (*i.e.*, the sum of all the projections on the Z-axis during a single sliding event), while d_z, is the distance between the maximal and minimal positions on the Z-axis visited by the protein during the sliding event (see Figure 10.6a). Thus, a higher ratio of MSD_z/d_z reflects a sliding event in which there is more redundancy in terms of repeated visits to the same DNA site. Figure 10.6c shows a scatter plot of the probing redundancy (measured as MSD_z/d_z) against the duration of the sliding event for salt concentrations of 0.03, 0.07, 0.09 and 0.15 M. At a salt concentration of 0.03 M, there is a considerable increase in the probing redundancy during long sliding events as the random walk nature of the motion tends to increase MSD_z, but not d_z, linearly over time. The probing redundancies at salt concentrations of 0.07 and 0.09 M are significantly lower than those observed for a salt concentration of 0.03 M as the durations of the sliding events are shorter. Under these conditions, the protein may scan a short DNA segment and dissociate until it randomly approaches another DNA site by non-specific binding, and therefore, 'wasteful' repeat visits to DNA sites are avoided. At a high salt concentration of 0.15 M, the sliding events become transient, such that there are hardly any repeat visits to DNA sites. However, the efficiency of the search is low under such high salt conditions for the reasons discussed above.

Figure 10.6d summarizes the results for various salt conditions by showing averages of MSD_z and d_z per unit time step of the simulation (namely, $<MSD_z/\tau_s>$ and $<d_z/\tau_s>$) for all the sliding events. The former has no dependence on the salt concentration and is consistent with the fact that the one-dimensional diffusion coefficient during sliding is not affected by salt concentration (see further discussion in the sections below). The value of

$<d_z/\tau_s>$ increases with increasing salt concentration until saturation is achieved at about 0.09 M.

These results provide a mechanistic explanation for the low efficiency of DNA search (*i.e.*, low Probed Position, see Figure 10.5B) under low salt

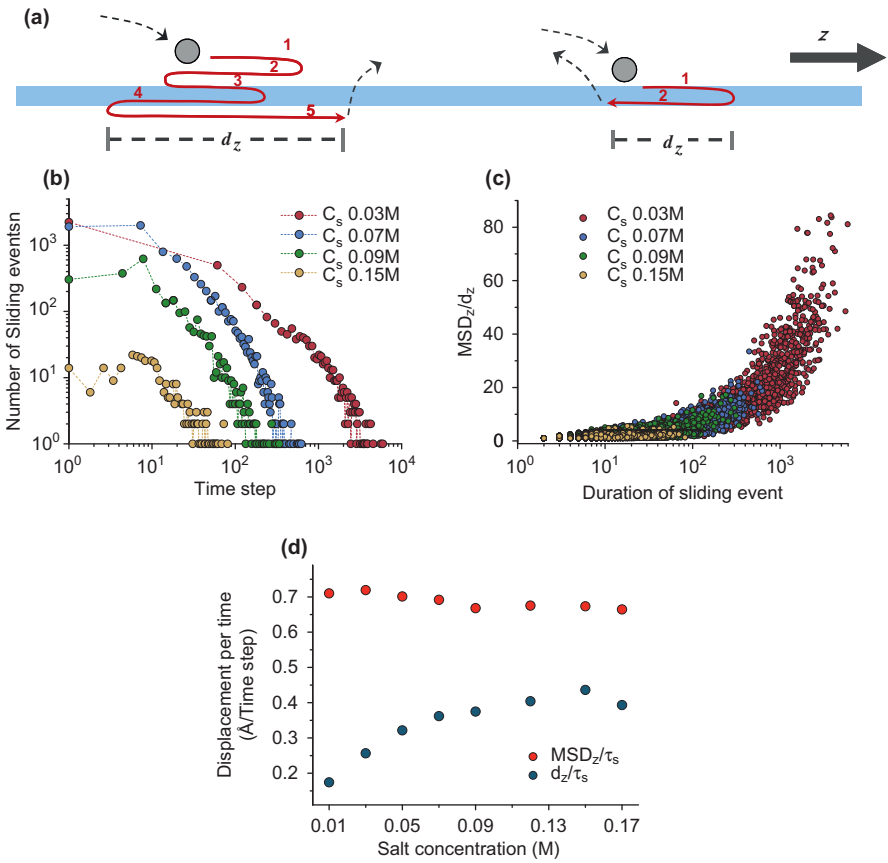


Figure 10.6 The effects of salt concentration on the distributions of sliding durations for the Sap1 protein. (a) A prolonged sliding event (left) in which DNA sites are sampled several times by the protein, and a shorter siding event (right) in which DNA sites are visited fewer times. (b) Distributions of the sliding durations for varying salt concentrations: 0.03 M (red), 0.07 M (blue), 0.09 M (green), and 0.15 M (orange) (averages from 10 simulations). (c) Scatter plot of MSD_z/d_z (where MSD_z is the total distance traveled by the protein along the DNA axis and d_z is the distance between the maximal and minimal Z-axis positions visited by the protein during the sliding event, see (a), left) against the sliding durations for each of the individual sliding events simulated at salt concentrations of 0.03, 0.07, 0.09, and 0.15 M. (d) Averages of MSD_z per time step ($<MSD_z/\tau_s>$, blue) and d_z per time step ($<d_z/\tau_s>$, red) calculated over all individual sliding events simulated at salt concentrations ranging from 0.01 to 0.17 M.

conditions. First, the fact that the durations of the individual sliding events are relatively long gives rise to a search mechanism that is more local because the lower frequency of dissociation events prevents the protein from accessing remote regions of the DNA. Second, because of the one-dimensional random walk motion of the protein along DNA, individual sliding events under low salt conditions are 'wasteful' because a significant amount of the search time is spent on visiting sites that were already scanned. An optimal search process is obtained at an ionic strength of about 0.07 M in which the protein spends $\sim$20% of the search time in sliding, and probing redundancy is relatively moderate with $MSD_z/d_z < 20$.

10.3.3 Protein Dynamics during Target Search: Influence of Helical Sliding and Hopping on 1D Diffusion Coefficient

10.3.3.1 1D Diffusion Coefficient (D_1) Increases with Ionic Strength during Hopping but not during Sliding

The dynamics of a protein as it moves along DNA as well its microscopic structural details, which govern sliding and hopping, can be studied computationally from the trajectories collected under various ionic strength conditions and temperatures. The left panels on Figures 10.7a and 10.7b shows the time-dependent position of the protein along the Z-axis (*i.e.*, the axis of the static DNA) for the recognition helix of the Sap1 protein at salt concentrations of 0.01 M and 0.07 M. According to the expression of mean square displacement derived by Qian *et al.*,[46] one can evaluate the 1D diffusion coefficient using equation 10.5:

$$MSD_z(n,N) = \sum_{i=1}^{N-n} \frac{(Z_{i+n} + Z_i)^2}{N-n} = 2D_1 n\Delta t \tag{10.5}$$

where N is the number of time steps measured, n is the measurement window ranging from 1 to N, Δt is the time interval between two consecutive steps, and D_1 is the 1D diffusion coefficient. To compute 1D diffusion along the DNA we use the distances that are projected along the Z-axis of a protein atom (C_α) that is selected from the center of the recognition helix. Similarly, we use the projections along the X and Y axis, to compute MSD profiles (denoted as MSD_x and MSD_y) and diffusion constants along these axes. For Brownian diffusion, the MSD(n,N) at n below a cutoff n_c is linear with a slope of $2D_1\Delta t$.

Many single molecule experiments have used the same approach to evaluate the 1D diffusion coefficient of different proteins such as RNA polymerases, DNA repair proteins, and transcription factors as they allow the movements of an individual fluorescently labeled protein along DNA to be observed.[39] In most experiments, the DNA is stretched in the microscopy field and images of the protein that one-dimensionally traverses the DNA are collected.

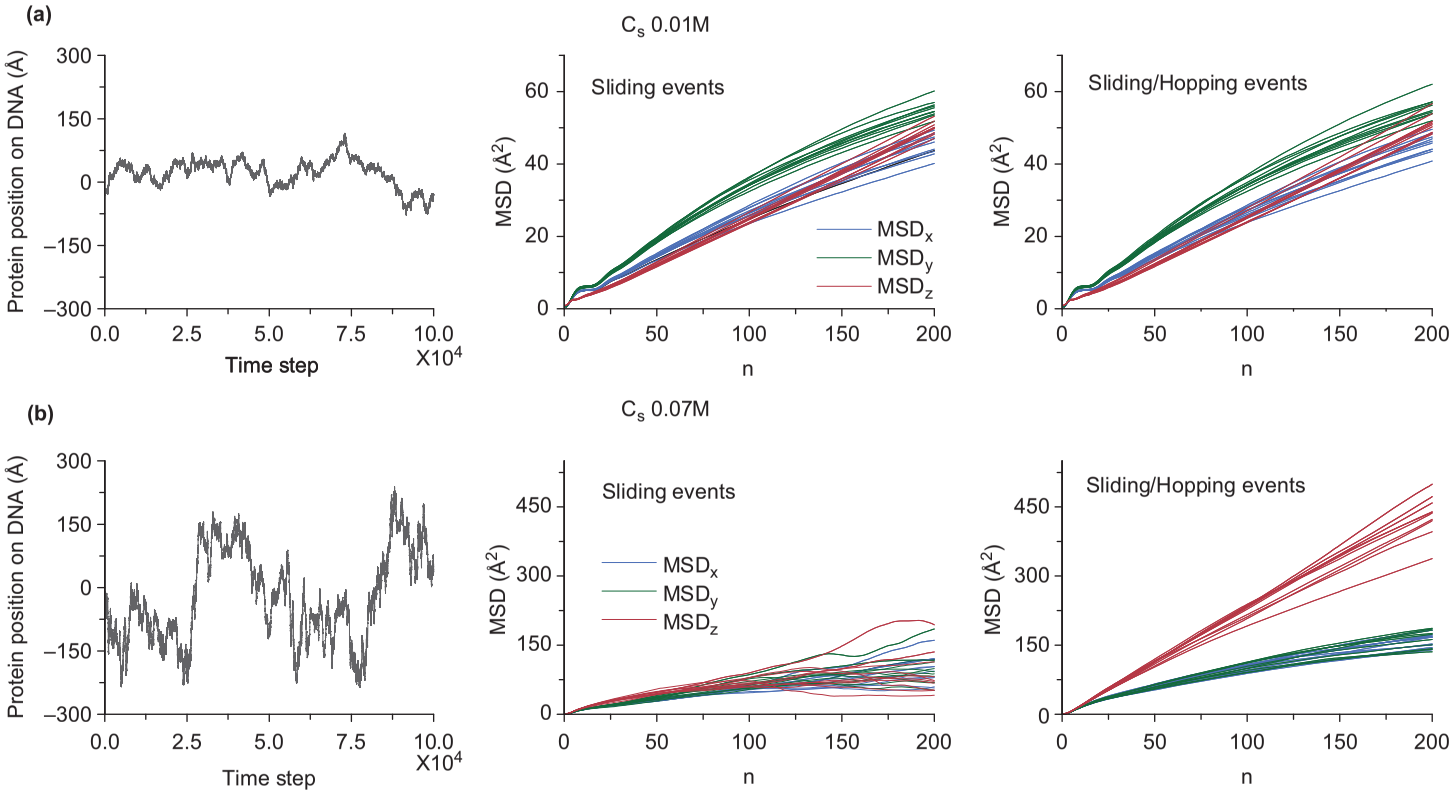


Figure 10.7 Calculation of the 1D diffusion coefficient (D_1) for protein movement along DNA. (a) Representative trajectory of the movement of the Sap1 protein along DNA at a salt concentration of 0.01 M is shown on the left (the position of a residue from the protein recognition helix along the Z-axis is recorded). MSD profiles for 10 trajectories at 0.01 M are shown, calculated for sliding only (middle panel) and for sliding + hopping (right panel). (b) Same as in (a) but for a salt concentration of 0.07 M.

This technique also enables a determination of whether a given protein predominately uses sliding or hopping to move along DNA by measuring the diffusion coefficient at increasing salt concentrations. If hopping dominates protein translocation along the DNA, then at higher salt conditions the protein will spend more time in solution and therefore the measured diffusion coefficient will increase. During sliding, however, the protein maintains electrostatic contact with the DNA continuously and therefore the diffusion coefficient is independent of the salt concentration. By employing this approach, Kim *et al.* have demonstrated that the diffusion coefficient of RNA polymerase from the T7 bacteriophage is independent of salt concentration, suggesting that sliding is the underlying mechanism for the movement of the protein along DNA.[41] There is also an experimental example for an increase in the diffusion coefficient with increasing salt concentration for the UL42 processivity factor, indicating that hopping dominates the movement of protein along the DNA.[45]

The middle panel of Figure 10.7a, shows MSD(n,N) profiles for 10 trajectories calculated only for the sliding periods along the X, Y and Z axes (denoted as MSD_x, MSD_y, MSD_z in blue, green and red, respectively) for simulations of Sap1 protein movement along a 100 bp dsDNA at a low ionic strength of 0.01 M. The convergence of the MSD profiles for all three directions may suggest that the recognition helix of the protein not only progress along the Z-axis but also fluctuates equally in the X and Y directions. This observation results from a protein motion that is not restricted only to fluctuations along the Z-axis but also to motions of the protein in directions that are perpendicular to the DNA. The middle panel of Figure 10.7b shows the MSD profiles for sliding at a salt concentration of 0.07 M. The average slope is only slightly greater than those at 0.01 M. In the right panels of Figure 10.7a and b, the MSD profiles for the same trajectories are shown for all the periods in which the protein is bound to the DNA by either sliding or hopping (since hopping events are rather short, it is impractical to estimate the D_1 value of a hopping motion alone). At a low salt concentration of 0.01 M (Figure 10.7a, right), the profiles resemble those obtained for sliding only, as the population of hopping events under these conditions is negligible. However, as the ionic strength is increased to 0.07 M (Figure 10.7b, right), the MSD_z profiles exhibit larger slopes than the MSD_x and MSD_y profiles, indicating that the protein traverses along the DNA main axis more rapidly in comparison to its diffusion in the directions perpendicular to it.

Figure 10.8a shows a summary of the comparison between a pure sliding mode and a bound mode in which the protein engages the DNA in either sliding or hopping. In agreement with the experimental observation discussed above, the 1D diffusion coefficient D_1 along the DNA axis ($D_{1,z}$) increases with increasing salt concentration when the protein uses sliding + hopping to move along the DNA but remains unaffected for a pure sliding motion of the protein. A similar observation is obtained from simulations of the HoxD9 and Skn1 proteins.[30] To examine the effects of temperature on the 1D diffusion coefficient, the authors measured D_1 for both modes at different temperatures and at different salt concentrations. As expected, an increase of the D_1 coefficients with salt concentration as well as with temperature is observed but with a more

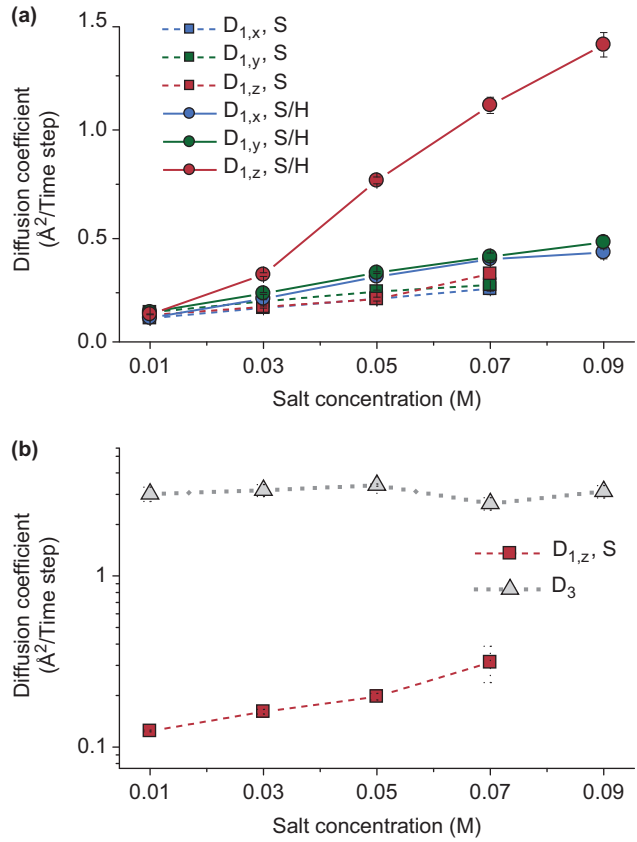


Figure 10.8 Diffusion coefficients (D_1) for sliding and bound (sliding + hopping) conformations of Sap1 protein. (a) The values of D_1 increase with increasing salt concentration for MSD profiles calculated over the bound conformations only for 1D diffusion along the DNA ($D_{1,z}$), but are relatively fixed for MSD profiles calculated over the sliding conformations only. (b) A comparison of 1D diffusion coefficient during sliding ($D_{1,z}$) with 3D diffusion coefficient of the protein in bulk (D_3).

pronounced effect for the bound conformation. Figure 10.8b shows a comparison of the calculated 1D diffusion coefficient D_1 along the DNA axis ($D_{1,z}$) from sliding, with the 3D diffusion coefficients calculated at various salt concentrations for the protein in bulk (D_3). The plot shows that D_3 is more than an order of magnitude greater than $D_{1,z}$.

10.3.3.2 During Sliding Proteins may undergo Rotation-Coupled Diffusion along DNA

Although it was demonstrated that proteins hop and slide along DNA, the path traced by the protein during sliding along the DNA surface has not been

directly observed yet. Biophysical and biochemical models have taken into consideration both linear translocations along the DNA and helical motion along the DNA grooves as plausible assumptions. A model proposed by Schurr[47] for rotation-coupled diffusion along DNA suggests that the diffusion coefficient has a stronger dependence on protein size ($D_{1,Rotation} \sim R^{-3}$) during rotation-coupled diffusion than it exhibits during a pure translational motion ($D_{1,Translation} \sim R^{-1}$). This discrepancy may give rise to a difference of up to three orders of magnitude between diffusion coefficients from pure linear translation compared to rotation-coupled diffusion. Bagchi *et al.*, have incorporated a parameter that describes the minimal distance of the protein's center of mass from the DNA axis (R_{OC}) into Schurr's model[48] and performed a series of elegant single-molecule measurements indicating that DNA-binding proteins indeed undergo rotation-coupled diffusion on a rugged free-energy landscape with barriers of $\sim 1.1\, k_B T$ separating adjacent sites on DNA.[25] The rotation along the helical path of the DNA enables the protein to continuously probe the base-pair content in the DNA major groove. Additional experimental evidence for rotation-coupled protein diffusion was obtained for proliferating cell nuclear antigen (PCNA), whose diffusion properties were measured under varying solvent viscosities and by attaching quantum dots that change the protein's size. The change in the diffusion coefficient for different protein sizes suggested that the clamp moves along DNA while rotationally tracking the helical path of the double-stranded DNA.[49]

While providing indirect evidence for coupling between rotational diffusion and linear translocation along DNA, the current experimental strategies cannot directly observe the rotational motion because of spatial resolution limitations.[39] Using the computational model described above, we were able to explicitly follow the curved linear motion of a protein along DNA. Figure 10.9a (on the right) shows the paths taken by a residue located at the center of the recognition helix of the Sap1 protein as it moves along the DNA surface during several sliding events at a low salt concentration of 0.01 M. The image demonstrates that the major groove of the DNA is substantially populated during sliding and that the protein motion is coupled with the helical structure of the double stranded DNA. Occasionally, the protein may undergo a rapid transition along the DNA that is decoupled from the helical pitch of the major groove but still in close proximity with the surface of the DNA. The coupling between rotation and translation is also exhibited in the plots of Figure 10.9a, which shows the angle of the protein against its position along the Z-axis in several sliding events of a single trajectory. The plots show a clear linear relationship ($<$Correlation coefficient$> = -0.87$) between rotation and translation with an average slope of -0.16 radians/Å. This value is very close to $2\pi/34$ Å that is the helical pitch in a canonical B-DNA molecule (*i.e.*, 1 turn per 10 base-pair rise). The emergence of several parallel lines results from transient decouplings between the rotation angle and position, in which the protein may for example, traverse between two neighboring grooves. The protein thus propagates along the Z-axis but not rotate around the DNA. Figure 10.9b shows the results for a salt concentration of 0.07 M. Sliding events under these conditions are much more transient and fast transitions of the

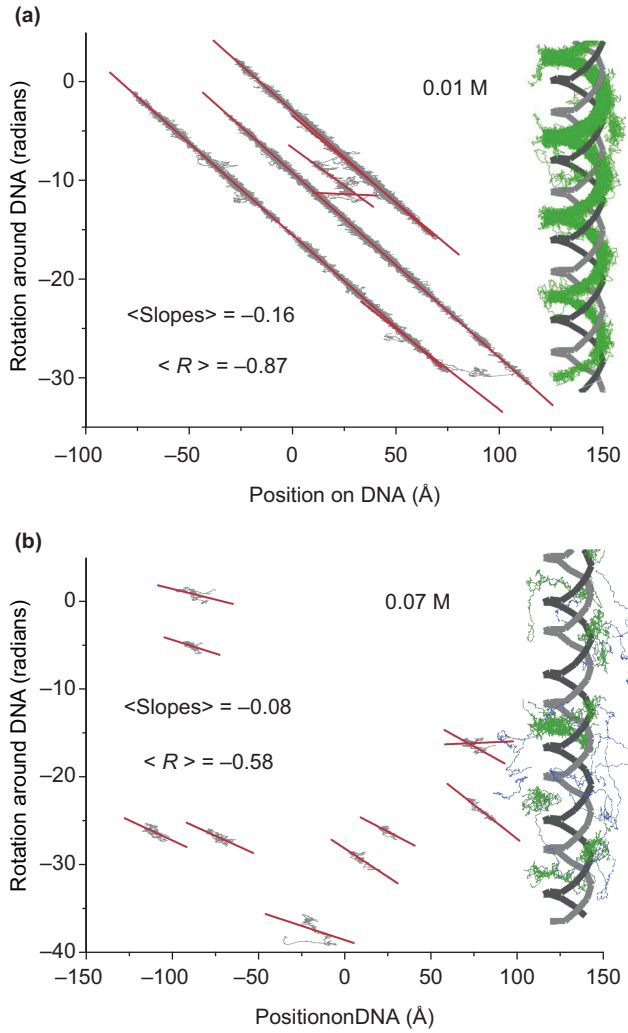


Figure 10.9 Observation of rotation-coupled sliding in Sap1 protein. (a) Low salt concentration of 0.01 M. The protein is shown to rotate around the DNA while sliding and maintain its interactions with the major groove (right, the trace of an atom from the recognition helix is shown in green during sliding). The rotation versus translation linear scatter plots on the left (the average correlation between the rotation and translation motions, $<R> = -0.87$) indicate average slopes ($<Slope> = -0.16$) that are consistent with the geometrical properties of a B-DNA molecule (*i.e.*, $2\pi/34$ Å) indicating a strong coupling between protein translation and rotation and the helical pitch of dsDNA. (b) Higher salt concentration of 0.07 M. The sliding events are shorter and more hopping events occur (right, green and blue lines indicate the recognition helix trace during sliding and hopping, respectively). The rotation-translation scatter plots on the left ($<R> = -0.58$), with average slopes of -0.08, indicate that the coupling of the rotation to the translation of the protein along the DNA is weaker.

recognition helix along the minor groove towards an adjacent major groove are much more frequent. Several hopping events (blue lines) are shown in which the protein also remains close to the DNA but exhibits no preference for the DNA groove or backbone but rather 'glides' randomly near its surface. Since the sliding events are significantly shorter and are less committed to the major groove track as the salt concentration increases, the correlation between the linear translocation and the rotation angle around the DNA axis is much weaker ($<$Correlation coefficient$> = -0.58$) with an average slope of -0.08 radians/Å. These results may suggest that adoption of a helical bound protein motion in the DNA major groove at low salt conditions, while enabling the protein to scan each DNA site rigorously, comes at the expense of search efficiency. By contrast, at higher salt concentrations the protein samples major groove positions locally, and rapidly translocates to other sites. This detailed mechanistic viewpoint is in agreement with the observation reported before (Figure 10.5b) that the Position Probed measure is low at low salt conditions. We point out that this coarse-grained simulation approach offers the potential to determine the degree of rotation coupled diffusion in many other DNA binding proteins as well as determining the preference of the protein for major-groove search, minor-groove search, or a search that is coupled to the DNA backbone.

10.4 Concluding Remarks

Many genomic processes, such as transcription, DNA repair, and recombination rely on DNA binding proteins that locate and bind their DNA targets remarkably quickly and efficiently. Understanding the nature of interactions between regulatory proteins and nonspecific DNA sequences is a fundamental step in deciphering all physicochemical mechanisms that underlie the protein–DNA recognition process.

In this review, we described a computational model to characterize the structural and dynamic features of DNA recognition by proteins during DNA search. In our simplified model, the protein was flexible but the DNA remained rigid, and the protein–DNA interactions were modeled by electrostatic forces only. This modeling is in accord with structural and thermodynamic studies indicating that electrostatics, rather than intimate interactions between the protein and the DNA bases, govern the nonspecific binding mode.

Using molecular dynamics simulations, our results indicate that, during sliding, the proteins perform a bidirectional 1D random walk movement driven by thermal diffusion. In 1D sliding, the movement of the protein is coupled with a rotational motion along the helical double-stranded DNA pitch, in which the protein utilizes an interface and adopts an orientation to the DNA that is very similar to that found in the crystal structure of the specific complex. This result is in agreement with recent NMR measurements on the HoxD9 protein, which indicate the signature of the specific protein–DNA recognition in nonspecific sliding.[23,24,42] Our measurements for the average slope of the plot describing

protein rotation versus protein translation indicate a value of $\sim 2\pi/10$ bp, which is consistent with the helical pitch of a B-DNA molecule.

Our observations for the distributions of sliding durations under varying salt conditions combined with a measure for the efficiency of the DNA search indicate that at low salt concentrations, sliding events are longer and less efficient because the protein redundantly rescans DNA sites. At higher salt concentrations, the protein is more detached from the DNA and may remain in the vicinity of the DNA (undertaking a linear search *via* hopping). The hopping search mode is accompanied by a higher diffusion coefficient and results in an enhancement of DNA scanning efficiency. Further increase of the salt concentration will result in a significant increase of the 3D search mode in solution at the expanse of 1D DNA scanning and will therefore significantly reduce the efficiency of the search. Accordingly, we suggest that there is an optimal salt concentration at which the DNA binding protein combines sliding, hopping, and 3D diffusion to search the DNA target. This optimal interplay between the various search mechanisms of DNA may be different for proteins having different DNA binding affinities.

Although we succeeded in capturing many of the key structural and dynamic properties of protein sliding with a model that relates only to the electrostatic interactions between proteins and DNA, it is clear that higher-resolution models are also essential. Specifically, such models are required to understand additional components of protein-DNA interactions, such as, the role of water molecules and ions in sliding,[50–52] water release from the protein–DNA interface upon specific binding,[22] and the role of DNA conformational changes and flexibility.[53,54] Nevertheless, the methodology presented here may serve as a predictive tool to study relatively long biological timescales in protein–DNA recognition as well as complementing experimental data. This computational approach has been applied recently to study how various aspects of the molecular architecture of DNA-binding proteins may affect the mechanism of DNA search. It was found that the oligomerization state of the protein can significantly affect sliding speed and the overall search kinetics.[55,56] Splitting a DNA-binding protein into two or more structural domains may assist the search[57–60] by promoting intersegment transfer. Furthermore, not all DNA-binding proteins perform sliding using the interface found in the specific protein-DNA recognition, and consequently a switching transition is involved when the target site is localized.[61]

References

1. G. Adam and M. Delbruck, *Reduction of Dimensionality in Biological Diffusion Processes*, San Francisco, 1968.
2. A. D. Riggs, S. Bourgeois and M. Cohn, *J Mol Biol*, 1970, **53**, 401–417.
3. P. H. von Hippel and O. G. Berg, *J Biol Chem*, 1989, **264**, 675–678.
4. S. E. Halford and J. F. Marko, *Nucleic Acids Res*, 2004, **32**, 3040–3052.
5. D. M. Gowers, G. G. Wilson and S. E. Halford, *Proc Natl Acad Sci U S A*, 2005, **102**, 15883–15888.

6. F. W. Dahlquist, *Nat Chem Biol*, 2006, **2**, 353–354.
7. M. Slutsky and L. A. Mirny, *Biophys J*, 2004, **87**, 4021–4035.
8. O. G. Berg, R. B. Winter and P. H. von Hippel, *Biochemistry*, 1981, **20**, 6929–6948.
9. D. Vuzman, A. Azia and Y. Levy, *J Mol Biol*, **396**, 674–684.
10. F. K. Winkler, D. W. Banner, C. Oefner, D. Tsernoglou, R. S. Brown, S. P. Heathman, R. K. Bryan, P. D. Martin, K. Petratos and K. S. Wilson, *Embo J*, 1993, **12**, 1781–1795.
11. H. Viadiu and A. K. Aggarwal, *Mol Cell*, 2000, **5**, 889–895.
12. C. G. Kalodimos, N. Biris, A. M. Bonvin, M. M. Levandoski, M. Guennuegues, R. Boelens and R. Kaptein, *Science*, 2004, **305**, 386–389.
13. P. H. von Hippel, *Science*, 2004, **305**, 350–352.
14. S. A. Townson, J. C. Samuelson, Y. Bao, S. Y. Xu and A. K. Aggarwal, *Structure*, 2007, **15**, 449–459.
15. M. C. Mossing and M. T. Record, Jr., *J Mol Biol*, 1985, **186**, 295–305.
16. B. Honig and A. Nicholls, *Science*, 1995, **268**, 1144–1149.
17. E. W. Stawiski, L. M. Gregoret and Y. Mandel-Gutfreund, *J Mol Biol*, 2003, **326**, 1065–1079.
18. A. Szilagyi and J. Skolnick, *J Mol Biol*, 2006, **358**, 922–933.
19. V. K. Misra, J. L. Hecht, A. S. Yang and B. Honig, *Biophys J*, 1998, **75**, 2262–2273.
20. M. T. Record, Jr., J. H. Ha and M. A. Fisher, *Methods Enzymol*, 1991, **208**, 291–343.
21. L. Jen-Jacobson, *Biopolymers*, 1997, **44**, 153–180.
22. T. Lundback and T. Hard, *Proc Natl Acad Sci U S A*, 1996, **93**, 4754–4759.
23. J. Iwahara and G. M. Clore, *Nature*, 2006, **440**, 1227–1230.
24. J. Iwahara, M. Zweckstetter and G. M. Clore, *Proc Natl Acad Sci U S A*, 2006, **103**, 15062–15067.
25. P. C. Blainey, G. Luo, S. C. Kou, W. F. Mangel, G. L. Verdine, B. Bagchi and X. S. Xie, *Nat Struct Mol Biol*, 2009, **16**, 1224–1229.
26. P. H. von Hippel, *Annu Rev Biophys Biomol Struct*, 2007, **36**, 79–105.
27. A. Graneli, C. C. Yeykal, R. B. Robertson and E. C. Greene, *Proc Natl Acad Sci U S A*, 2006, **103**, 1221–1226.
28. Y. M. Wang, R. H. Austin and E. C. Cox, *Phys Rev Lett*, 2006, **97**, 048302.
29. I. Bonnet, A. Biebricher, P. L. Porte, C. Loverdo, O. Benichou, R. Voituriez, C. Escude, W. Wende, A. Pingoud and P. Desbiolles, *Nucleic Acids Res*, 2008, **36**, 4118–4127.
30. O. Givaty and Y. Levy, *J Mol Biol*, 2009, **385**, 1087–1097.
31. C. Clementi, H. Nymeyer and J. N. Onuchic, *J Mol Biol*, 2000, **298**, 937–953.
32. Y. Levy, S. S. Cho, J. N. Onuchic and P. G. Wolynes, *J Mol Biol*, 2005, **346**, 1121–1145.
33. Y. Levy, J. N. Onuchic and P. G. Wolynes, *J Am Chem Soc*, 2007, **129**, 738–739.
34. T. Schlick, *Molecular modeling and simulation: An interdisciplinary guide,* Springer, New York, 2000.

35. C. Hyeon and D. Thirumalai, *Proc Natl Acad Sci U S A*, 2005, **102**, 6789–6794.
36. T. Schlick, B. Li and W. K. Olson, *Biophys J*, 1994, **67**, 2146–2166.
37. D. A. Beard and T. Schlick, *Structure*, 2001, **9**, 105–114.
38. A. Savelyev and G. A. Papoian, *J Am Chem Soc*, 2007, **129**, 6060–6061.
39. A. Tafvizi, L. A. Mirny and A. M. van Oijen, *Chemphyschem*, 2011, **12**, 1481–1489.
40. J. Gorman, A. J. Plys, M. L. Visnapuu, E. Alani and E. C. Greene, *Nat Struct Mol Biol*, **17**, 932–938.
41. R. G. Larson and J. H. Kim, *Nucleic acids research*, 2007, **35**, 3848–3858.
42. J. Iwahara and G. M. Clore, *J Am Chem Soc*, 2006, **128**, 404–405.
43. C. W. Garvie and C. Wolberger, *Molecular Cell*, 2001, **8**, 937–946.
44. P. C. Blainey, A. M. van Oijent, A. Banerjee, G. L. Verdine and X. S. Xie, *Proceedings of the National Academy of Sciences of the United States of America*, 2006, **103**, 5752–5757.
45. D. M. Coen, G. Komazin-Meredith, R. Mirchev, D. E. Golan and A. M. van Oijen,*Proceedings of the National Academy of Sciences of the United States of America*, 2008, **105**, 10721–10726.
46. H. Qian, M. P. Sheetz and E. L. Elson, *Biophysical Journal*, 1991, **60**, 910–921.
47. J. M. Schurr, *Biophysical Chemistry*, 1979, **9**, 413–414.
48. B. Bagchi, P. C. Blainey and X. S. Xie, *Journal of Physical Chemistry B*, 2008, **112**, 6282–6284.
49. A. M. van Oijen, A. B. Kochaniak, S. Habuchi, J. J. Loparo, D. J. Chang, K. A. Cimprich and J. C. Walter, *Journal of Biological Chemistry*, 2009, **284**, 17700–17710.
50. M. Fuxreiter, M. Mezei, I. Simon and R. Osman, *Biophys J*, 2005, **89**, 903–911.
51. Y. Levy and J. N. Onuchic, *Annu Rev Biophys Biomol Struct*, 2006, **35**, 389–415.
52. V. Dahirel, F. Paillusson, M. Jardat, M. Barbi and J. M. Victor, *Phys Rev Lett*, 2009, 102.
53. T. Hu, A. Y. Grosberg and B. I. Shklovskii, *Biophysical Journal*, 2006, **90**, 2731–2744.
54. R. Rohs, H. Sklenar and Z. Shakked, *Structure*, 2005, **13**, 1499–1509.
55. A. Marcovitz and Y. Levy, *Biophys J*, 2009, **96**, 4212–4220.
56. N. Khazanov and Y. Levy, *Journal of Molecular Biology*, 2011, **408**, 335–355.
57. D. Vuzman, A. Azia and Y. Levy, *J Mol Biol*, 2010, **396**, 674–684.
58. D. Vuzman and Y. Levy, *Molecular BioSystems.*, 2012, **8**, 47–57.
59. Y. Takayama and G. M. Clore, *Proceedings of the National Academy of Sciences of the United States of America*, 2011, **108**, E169–E176.
60. M. Doucleff and G. M. Clore, *Proceedings of the National Academy of Sciences of the United States of America*, 2008, **105**, 13871–13876.
61. A. Marcovitz and Y. Levy, *Proceedings of the National Academy of Sciences of the United States of America*, 2011, **108**, 17957–17962.