


# Diffusion of nucleotide excision repair protein XPA along DNA by coarse-grained molecular simulations

Weiwei Zhang(张伟伟) and Jian Zhang(张建)†

**Citation:**Chin. Phys. B, 2021, 30 (**10**): 108703. DOI: 10.1088/1674-1056/ac1e0e
Journal homepage: http://cpb.iphy.ac.cn; http://iopscience.iop.org/cpb

**What follows is a list of articles you may be interested in**

---

## Modeling hydrogen exchange of proteins by a multiscale method

Wentao Zhu(祝文涛), Wenfei Li(李文飞), and Wei Wang(王炜)
Chin. Phys. B, 2021, 30 (**7**): 078701. DOI: 10.1088/1674-1056/abe377

## Methods and applications of RNA contact prediction

Huiwen Wang(王慧雯) and Yunjie Zhao(赵蕴杰)†
Chin. Phys. B, 2020, 29 (**10**): 108708. DOI: 10.1088/1674-1056/abb7f3

## Improving RNA secondary structure prediction using direct coupling analysis

Xiaoling He(何小玲), Jun Wang(王军), Jian Wang(王剑), Yi Xiao(肖奕)
Chin. Phys. B, 2020, 29 (**7**): 078702. DOI: 10.1088/1674-1056/ab889d

## Molecular dynamics simulations of membrane deformation induced by amphiphilic helices of Epsin, Sar1p, and Arf1

Zhen-Lu Li(李振鲁)
Chin. Phys. B, 2018, 27 (**3**): 038703. DOI: 10.1088/1674-1056/27/3/038703

## Bending energy of a vesicle to which a small spherical particle adhere: An analytical study

Cao Si-Qin, Wei Guang-Hong, Jeff Z. Y. Chen
Chin. Phys. B, 2015, 24 (**9**): 098702. DOI: 10.1088/1674-1056/24/9/098702

---

# Diffusion of nucleotide excision repair protein XPA along DNA by coarse-grained molecular simulations*

Weiwei Zhang(张伟伟) and Jian Zhang(张建)†

*National Laboratory of Solid State Microstructures, School of Physics, Collaborative Innovation Center of Advanced Microstructures, Nanjing University, Nanjing* 210093, *China*

(Received 11 April 2021; revised manuscript received 7 July 2021; accepted manuscript online 17 August 2021)

Protein XPA plays critical roles in nucleotide excision repair pathway. Recent experimental work showed that the functional dynamics of XPA involves the one-dimensional diffusion along DNA to search the damage site. Here, we investigate the involved dynamical process using extensive coarse-grained molecular simulations at various salt concentrations. The results demonstrated strong salt concentration dependence of the diffusion mechanisms. At low salt concentrations, the one-dimensional diffusion with rotational coupling is the dominant mechanism. At high salt concentrations, the diffusion by three-dimensional mechanism becomes more probable. At wide range of salt concentrations, the residues involved in the DNA binding are similar and the one-dimensional diffusion of XPA along DNA displays sub-diffusive feature. This sub-diffusive feature is tentatively attributed to diverse strengths of XPA–DNA interactions. In addition, we showed that both binding to DNA and increasing salt concentration tend to stretch the conformation of the XPA, which increases the exposure extent of the sites for the binding of other repair proteins.

**Keywords:** nucleotide excision repair, XPA, one-dimensional diffusion along DNA, molecular simulation

**PACS:** 87.14.E−, 87.14.gk, 87.15.kj, 87.10.Tf **DOI:** 10.1088/1674-1056/ac1e0e

## 1. Introduction

Nucleotide excision repair (NER) is the central DNA repair pathway widely observed in eubacteria, eukaryotes and archaea. It is responsible for the removal of DNA lesions caused by UV irradiation, chemical mutagens, and other stress damages.[1–5] The abnormalities of NER can lead to the xeroderma pigmentosum (XP) disease, which is featured by sensitivity to sunlight and high risk of developing skin cancers.[6,7] XPA (xeroderma pigmentosum complementation group A) is a key protein in NER, and mutation of XPA gene can lead to severe disease symptoms.[6,8] It has been showed that the XPA may play multiple roles in NER.[9–12] For example, in addition to preferentially binding to various types of DNA lesions and distorted substrates,[10,13] XPA has also been found to interact with several NER related proteins during the damage recognition and verification steps, acting as a scaffold.[11,14]

The human XPA has the length of 273 amino acids and is composed of three domains, including the intrinsically disordered N-terminal region (N-region), the central DNA-binding domain (DBD), and the intrinsically disordered C-terminal region (C-region) (see Fig. 1). The flexibility of the two disordered regions were suggested to promote the binding of XPA to several partners,[15–19] while the central globular DNA-binding domain was found to bind with DNA and interact with numerous NER proteins as well.[20]

Due to its biological importance in NER, XPA has received extensive attention.[7,21] Although previous investigations mostly focused on the interactions of the XPA at the DNA lesion site during the damage recognition and verification steps, recent experimental study showed that detecting the DNA lesions may also involve the one-dimensional diffusion of the XPA at the non-damage site of DNA.[22] Therefore, more detailed investigations for the molecular dynamics of the XPA diffusion along the non-damage site of DNA are important for deeper understanding the biological function of XPA in the NER.

Although all-atom detailed molecular dynamics (MD) simulation has been widely used to directly simulate the details of the inter-molecular interactions involved in the protein-DNA recognitions, it is still extremely time-consuming to study the whole diffusion process of proteins along DNA.[23–26] Recently, the combination of the atomic-interaction-based coarse-grained model (AICG2+) for proteins[27,28] and the three-site-per-nucleotide (3SPN.2C) model for DNA[29] has shown great success in studying the dynamical and structural features of the protein-DNA interactions.[30–33] Here, based on these models, we performed extensive coarse-grained molecular simulations to study the diffusion dynamics of XPA along the DNA. From our simulations, we found that XPA diffuses along DNA with the DNA-binding residues similar to that identified from crystal structure. It was also found that the diffusion of XPA along DNA is achieved through combination of one-dimensional and three-dimensional mechanisms. At low salt concentrations, the one-

*Project supported by the National Natural Science Foundation of China (Grant Nos. 11974173 and 11774158) and the HPC center of Nanjing University.

†Corresponding author. E-mail: jzhang@nju.edu.cn

© 2021 Chinese Physical Society and IOP Publishing Ltd 

dimensional diffusion with strong rotational coupling is the dominant mechanism. At higher salt concentrations, the diffusion by three-dimensional mechanism becomes more probable. Meanwhile, the one-dimensional diffusion of XPA along DNA displays sub-diffusive behaviour at wide range of salt concentrations. In addition, the simulation showed that the XPA becomes more extended in the presence of DNA or at high salt concentration, suggesting that both DNA binding and increasing salt concentration may promote the binding of other target proteins to XPA by increasing the exposure extent of the binding sites.

## 2. Model and method

### 2.1. Coarse-grained model of protein-DNA system.

Throughout this work, to model protein (XPA), the structure-based coarse-grained model was used, where each amino acid in protein was coarse-grained to a single bead located at C$\alpha$ position. The interactions between protein beads include AICG2+ potentials and Debye–Hückel-type electrostatic interactions. The AICG2+ energy function is given by $V_{\mathrm{AICG2+}} = V_{\mathrm{local}} + V_{\mathrm{Go}} + V_{\mathrm{exv}}$, where $V_{\mathrm{local}}$, $V_{\mathrm{GO}}$, and $V_{\mathrm{exv}}$ are local potentials, non-local structure-based potentials, and excluded volume potentials respectively. The details of these terms can be found in Refs. [28,34,35], and the similar energy function has also been successfully used in the studies of protein folding, enzyme catalysis, chromatin remodeler and nucleosome dynamics in recent works.[27,36–40] Note that the structure-based interactions were omitted for the disordered regions in XPA, and effects of salt concentration are implicitly modeled in the Debye–Hückel-type interactions with the details displayed as follows:

$$V_{\mathrm{ele}} = \sum_{i<j} \frac{1}{4\pi\varepsilon(T,C)} \frac{q_i q_j \exp\left(-r_{i,j}/\lambda_{\mathrm{D}}\right)}{r_{i,j}},$$

$$\lambda_{\mathrm{D}} = \left(\frac{\varepsilon_0 \varepsilon_k k_{\mathrm{B}} T}{2N_{\mathrm{A}} e^2 I}\right)^{0.5},$$

where $\lambda_{\mathrm{D}}$ is the Debye length, $\varepsilon(T,C)$ is the dielectric constant, and $I$ is ionic strength.

To model dsDNA, the 3SPN.2C model,[29] developed by de Pablo's group, was used. In this model each nucleotide is simplified to three CG particles, representing the phosphate group (P), sugar (S), and base (B). The interactions among these particles are composed of structure-based local potentials, base pairing, intrastrand base-stacking, interstrand cross-stacking, excluded volume effects, and Debye–Hückel-type electrostatic interactions. The partial charge of the phosphate beads was set as $-0.6$e to consider the effect of local counterion condensation effect.[31,41] The 3SPN.2C model model is able to capture several properties of dsDNA and has been widely used to study the DNA-protein interactions.

For the protein-DNA interactions, only the excluded volume effects and the electrostatic interactions were considered. As the electrostatic interaction plays a dominant role in protein-DNA interactions, we paid special attention to this term in the study of protein-DNA binding. For the structured motif in proteins, the partial charges of the CG beads were parametrized by the RESPAC method,[42] which can provide more accurate description of the electrostatic potential. In describing the electrostatic interactions between the protein and DNA, the partial charge of the phosphate bead was set as $-1.0$, following the work in Ref. [31].

### 2.2. Simulation details

The model for XPA was built based on the available crystal structure (PDB entry:6j44[8]) with missing residues added by modeller9.23.[43] The sequence and structure of XPA are shown in Figs. 1(a), 1(b), and 1(c). For DNA, two random sequences with different lengths were prepared: R100, R150 (the sequences and the structures are shown in Figs. 1(d) and 1(e)). In our work, the reference structures for DNA were built with the 3DNA package.[44] For the protein-DNA system, the initial structures for DNA and protein in all the simulations were separated and placed randomly. During production runs, the center of mass (COM) of DNA was restrained to the original position and COM distance between XPA (DBD) and DNA was restrained to be smaller than 250 Å. The ionic strength was set to 50, 150, 250, 350, and 450 mM. At each ionic strength, 24 simulations were carried out independently, each lasting for $5 \times 10^8$ MD steps for XPA-R100 system and over $4 \times 10^8$ MD steps for XPA-R150 system, respectively. To characterize the structure features of XPA in the absence of DNA, we also performed several simulations with cumulative length of approximately $1 \times 10^9$ MD steps. For all the simulations, we recorded the coordinates of XPA and DNA at every $1 \times 10^4$ MD steps. The first $2 \times 10^7$ frames were omitted in the calculations of distributions during the posterior analysis.

All the CG simulations were performed by the CafeMol3.2 package.[34] The simulations were conducted by Langevin dynamics with friction coefficient $\gamma = 0.02$ at temperature $T$ of 300 K. The structures were visualized by using pymol.[45]

### 2.3. Analysis

To characterize the search modes of XPA on DNA, we defined a DNA recognition motif consisting of the residues 140, 141, 142, 151, 179, 207, 211,[46] which forms contacts with DNA in the crystal structure. Following the approach used in Ref. [41], the diffusion is considered as sliding if at least five of these residues maintain staying within 15 Å from the DNA. Correspondingly, the diffusion is considered as three-dimensional diffusion if all of the seven residues in the

recognition motif leaves the DNA surface by at least 25 Å. In comparison, the protein is considered to perform the hopping dynamics if these residues are 15 Å–25 Å away from DNA. Based on above criteria, we classified both sliding and hopping as one-dimensional diffusion mechanism following the work in Ref [41].

To describe the location of DBD searching site on DNA during one-dimensional diffusion, we used the mean position of the nucleotides that closest to the recognition motif (RM):

$$X(t) = \frac{\sum_{k\in \mathrm{RM}} \sum_{j\in \mathrm{DNA}} bp^{(j)} \delta_j^k(t) L}{N_{\mathrm{RM}}}.$$

In the above definition, $t$ is simulation time, $L$ is 3.4 Å (one bp along the strand), RM is the recognition motif, composed of the residues 140, 141, 142, 151, 179, 207, and 211, $N_{\mathrm{RM}}$ is the number of residues in recognition motif, $bp^{(j)}$ is the base pair index of particle $j$ in DNA; $\delta_j^k = 1$, if $j$ is the DNA bead closest to the protein bead $k$; otherwise $\delta_j^k = 0$.[31]

The protein bead and DNA bead form a contact if their distance is less than 7 Å. Additionally, we defined several local coordinate axes, following the procedure in Ref. [47], to characterize the rotational and translational motions of XPA around DNA.

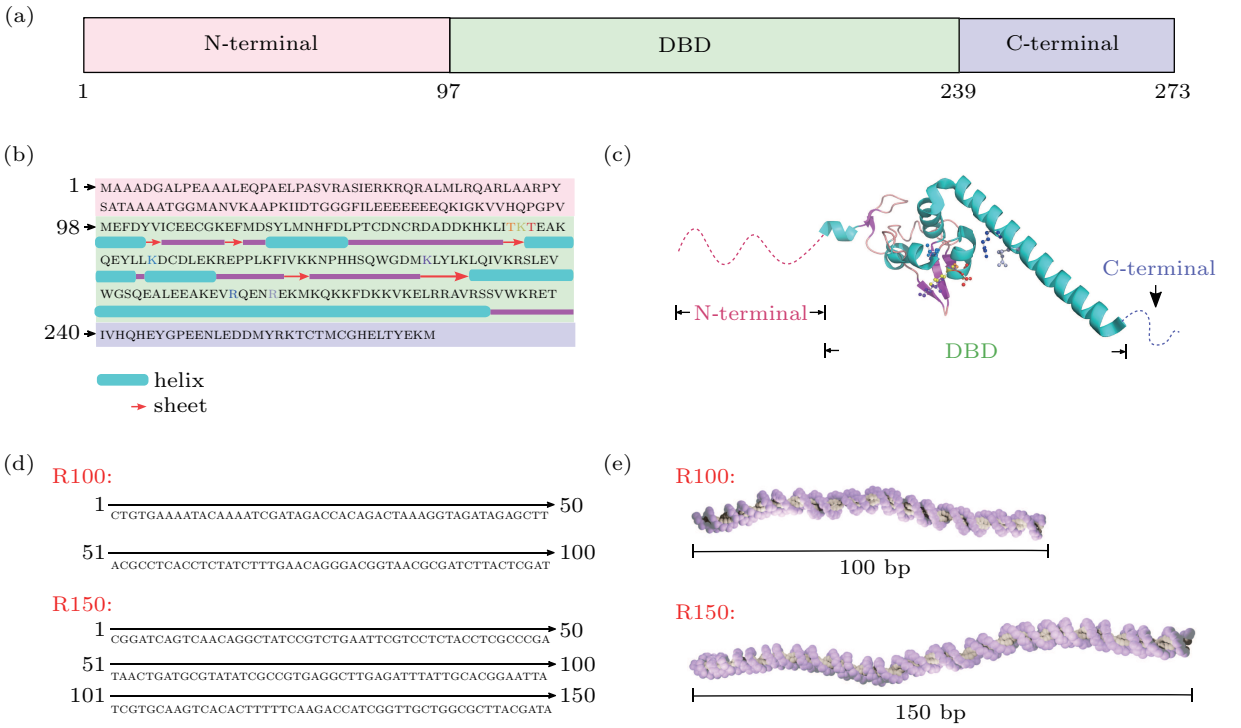



**Fig. 1.** Sequence and structure of the XPA and DNA studied in this work. (a) Domain architecture of XPA. (b) The sequence of XPA protein with each domain shadowed by the same color as in panel (a). The secondary structure features of DBD domain is also shown. (c) The available crystal sturcture of DBD is shown in cartoon representation and the color scheme for secondary structures is the same as in panel (b). The disorderd regions are represented by dotted lines with the same color as in panel (a). The key DNA-binding residues (residue 140, 141, 142, 151, 179, 207, 211) are shown in ball-and-stick representation and colored differently. The sequences (d) and structures (e) of DNA in this study.

## 3. Results

### 3.1. XPA binding and diffusion along DNA

Firstly, to quantify the structural properties of non-specific interaction interface between XPA and DNA, based on two random DNA sequences with different lengths (R100, R150), we performed MD simulations for the XPA binding to the nonspecific dsDNA site at various salt concentrations and calculated the probabilities of contact formation between the DNA and each residue of the XPA at various salt concentrations. As shown in Fig. 2, there is a striking similarity in the landscape of contact probability between R100 and R150, indicating the conserved binding interface of XPA regardless of properties of DNA. Additionally, for both R100 and R150, one can see that, besides the DBD, residues in the N-region also have high contacting probabilities with DNA, which suggests that the N-region can be involved in the XPA–DNA interactions during the searching process of damage sites. Meanwhile, the results clearly show that increasing salt concentrations tends to decrease the contact probabilities, but without altering the peak locations of the contact probability distributions, which may suggest that the XPA uses the same set of residues for DNA binding at wide range of salt concentrations. Particularly, the residues with high contact probabilities have large overlap with the DNA-binding residues observed in crystal structure of the XPA–DNA complex,[46] demonstrating that the DNA binding surface in the DBD is conserved in the specific and non-specific binding. Such behaviour of maintaining binding interface was also observed in other DNA-binding proteins,[47] and was suggested to facilitate the search of the cognate site on DNA.[48,49]

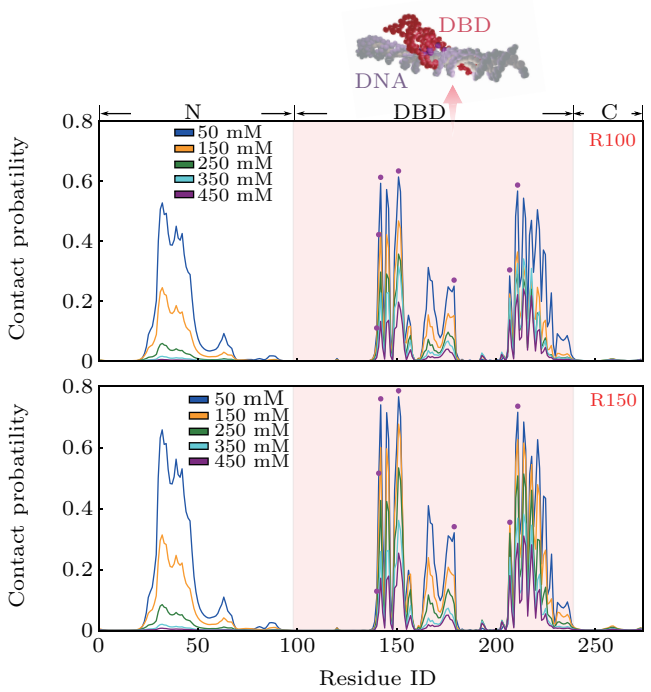



**Fig. 2.** Contact probabilities of the XPA residues with DNA at various salt concentrations. Residues in the DBD domain are shaded by pink backgound. The key DNA-binding residues in the crystal structure are highlighted by magenta circles. The result with R100 (R150) is shown in the upper (lower) panel and a representative structure with DBD bound to R100 is also displayed.

Besides the properties of binding interface, we also characterized the binding and diffusion process of XPA along DNA. Considering the relatively higher probabilities of contact formation with DNA for the residues in the DBD and the N-region, we plotted the time series of the distances between the above two protein segments and the DNA surface to illustrate the binding dynamics. The centers of mass (COM) of the DBD and that of the residues 30–50 were used to represent the positions of the DBD and the N-region relative to DNA surface. From Figs. 3(a) and 3(b), one can clearly see the attachment and detachment processes of the two XPA segments. Such attachment and detachment processes are consistent with experimental observation.[22] However, it is worthy to note that, because of the limited accessible time in our simulation, the repetition of attachment and detachment is only observed sporadically at relative high salt concentration (Fig. 3(a)). Additionally, inspection of trajectory of these two protein segments demonstrated that the attachment/detachment of the DBD and N-region are correlated, lacking of the anchoring and binding mechanism mediated by disordered region as typically observed in other DNA-binding proteins containing disordered regions.[50] After the simultaneous detachments of the N-region and the DBD, the whole XPA performs three-dimensional diffusion and then rebinds to the DNA (Fig. 3(a)). For many DNA binding proteins, rapidly finding their specific binding sites on DNA in response to external stimulus is vitally important. Thus, in addition to the random three-dimensional diffusion driven by thermal diffusion, the protein

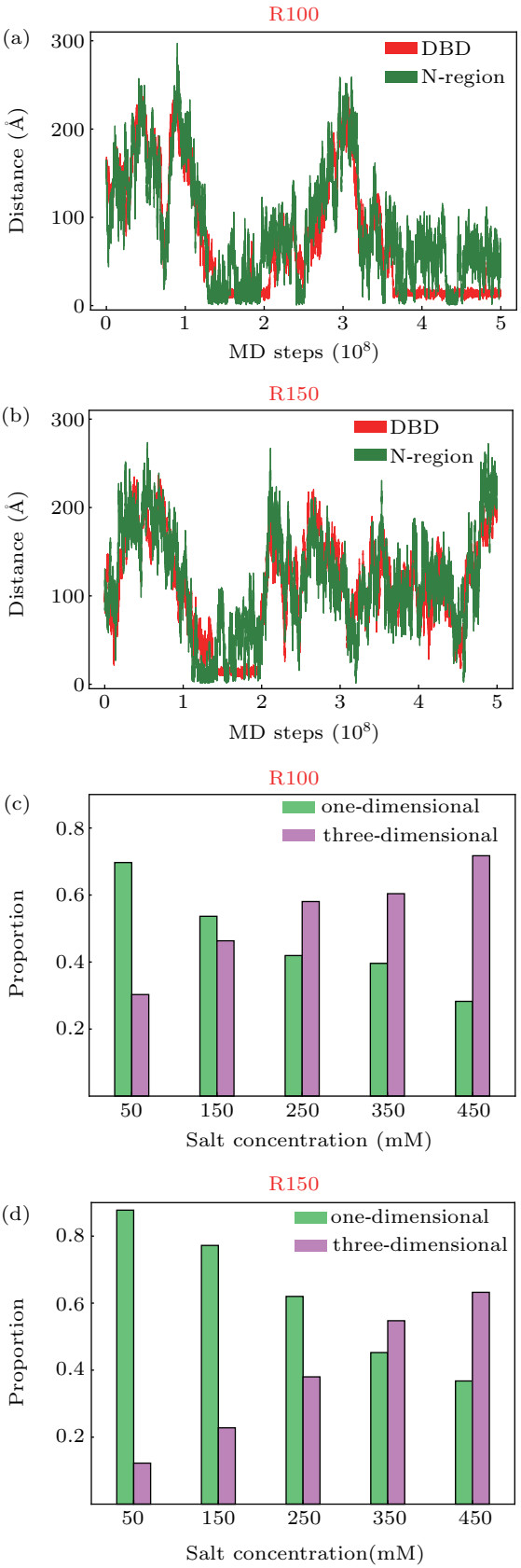



**Fig. 3.** (a) and (b) Time series of the distance between the COM of N-region and DNA surface (green) and that between the COM of DBD and DNA surface (red) at 250 mM with sequence R100 (a) and R150 (b). When calculating the distance between the COM of N-terminal and DNA surface, only the residues ranging from 30 to 50, which have high probabilities of forming contacts with DNA, were considered. (c) and (d) Probabilities of one-dimensional diffusion (green) and three-dimensional diffusion (pink) mechanism at different salt concentrations with sequence R100 (c) and R150 (d).

also uses other searching mechanisms, such as sliding,[51–53] hopping or intersegmental transfer.[54] During sliding, the protein can tightly bind to DNA and diffuse along the DNA helix using the recognition region,[55] in which electrostatic interaction between protein and DNA plays dominant role. Whereas in the hopping diffusion process, the electrostatic interactions are much weaker, which allows the protein to diffuse more freely along the DNA surface and accelerates diffusion by skipping several DNA bases.[41] Unlike three-dimensional diffusion, in which the protein can diffuse far away from DNA, the protein remain staying close to the DNA during the sliding and hopping, which tends to significantly reduce the conformational space during target site search and therefore accelerate the searching of the target site in the DNA.[51] Previous studies showed that the combination of these different search modes is needed to achieve the maximal target-search efficiency.[54,56] Therefore, we also analyzed the probabilities of the one-dimensional diffusion mechanism and three-dimensional diffusion mechanism (criteria about classifying different mechanisms can be found in Model and Method) at various salt concentrations for sequence R100 and R150, respectively (Figs. 3(c) and 3(d)). From these results, one can see that, increasing salt concentration can decrease the probability of one-dimensional searching mechanism and increase the three-dimensional searching mechanism for each DNA sequence. As shown in the Debye–Hcükel model, the effect of salt concentration is mainly achieved by changing the length of Debye radius. With the increasing of salt concentrations, the screening effect of the implicit salt ions increases, leading to the decrease of electrostatic interactions. Such results again suggest the dominate role of electrostatic interaction in the XPA–DNA binding.

Considering the important role of one-dimensional diffusion, we then analyzed the diffusion kinetics and calculated the mean square displacement (MSD) of the one-dimensional DBD translational motions along the DNA ($\mathrm{MSD}(\Delta t) = \langle (X(t+\Delta t) - X(t))^2 \rangle$, where $\langle\ \rangle$ represents the time average) as a function of time difference $\Delta t$ at various salt concentrations, and results are shown in Fig. 4. The fitting by a power law function $\mathrm{MSD}(\Delta t) = K\Delta t^{\alpha}$ were also shown in the inset by log–log scale. From our results, one can clearly see that the MSD of XPA (DBD) increases upon increasing salt concentration during one-dimensional diffusion for both R100 and R150. The fitting of the diffusion kinetics by the above power law function for R100 and R150 both give the $\alpha < 1.0$, which suggests sub-diffusive motion of DBD along DNA (inset figure in Fig. 4) at wide range of salt concentrations and may facilitate the search of target DNA sequences as suggested by other works.[57–59] Furthermore, this kind of sub-diffusion may be related to the diverse strengths of DNA–XPA interactions, which may be ascribed to the variations of the DNA conformations and the fluctuations of the DNA-protein binding. This kind of connection is consistent with those in many other sub-diffusive dynamics,[59–69] and can be described with continuous time random walk (CTRW) model.[61,69] Note that the exponents of the sub-diffusive dynamics are almost invariant at different salt concentrations for each sequence. This observation suggests that the distributions of the diverse strengths are insensitive to the variations of salt concentrations, since the conformational fluctuations are intrinsic for the the DNA-protein system. This result bring us more insights on the DNA-protein interactions.

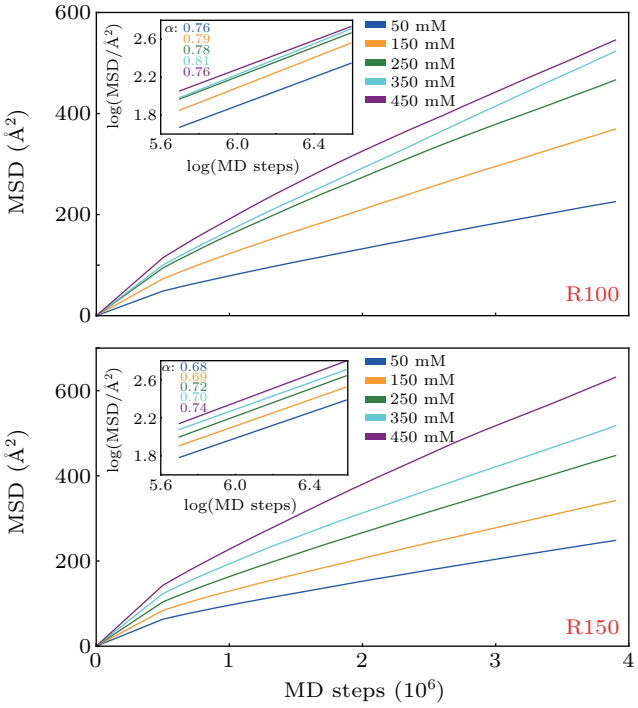



**Fig. 4.** Mean square displacement (MSD) of the one-dimensional XPA diffusion along the DNA (R100: upper panel, R150: lower panel) as a function of simulation time at different salt concentrations. The results of power function fitting $\alpha$ are also shown in the inset.

For DNA-interacting proteins, several studies have suggested that diffusion along the DNA helical path is required to achieve efficient target search.[55,70] Meanwhile, the rotation-coupled sliding was proposed to be the general mechanism for the one-dimensional search of protein on DNA.[71] Therefore, we also investigated whether the one-dimensional diffusion of XPA is along helical pitch (rotation-coupled) or not (rotation-uncoupled), and the results are shown in Fig. 5. To more appropriately describe this dynamical process, we defined several local axes following the Ref. [47]. The coordinates of the seven residues in the DNA-binding motif mentioned above were used to calculate the rotational angle and translational position along the DNA helical axis. In Fig. 5(a), we plotted the time series of rotational and translational motions of DBD along the R100 and R150 at various salt concentrations during the one-dimensional diffusion. From these trajectories, one can see that the protein can diffuse randomly along

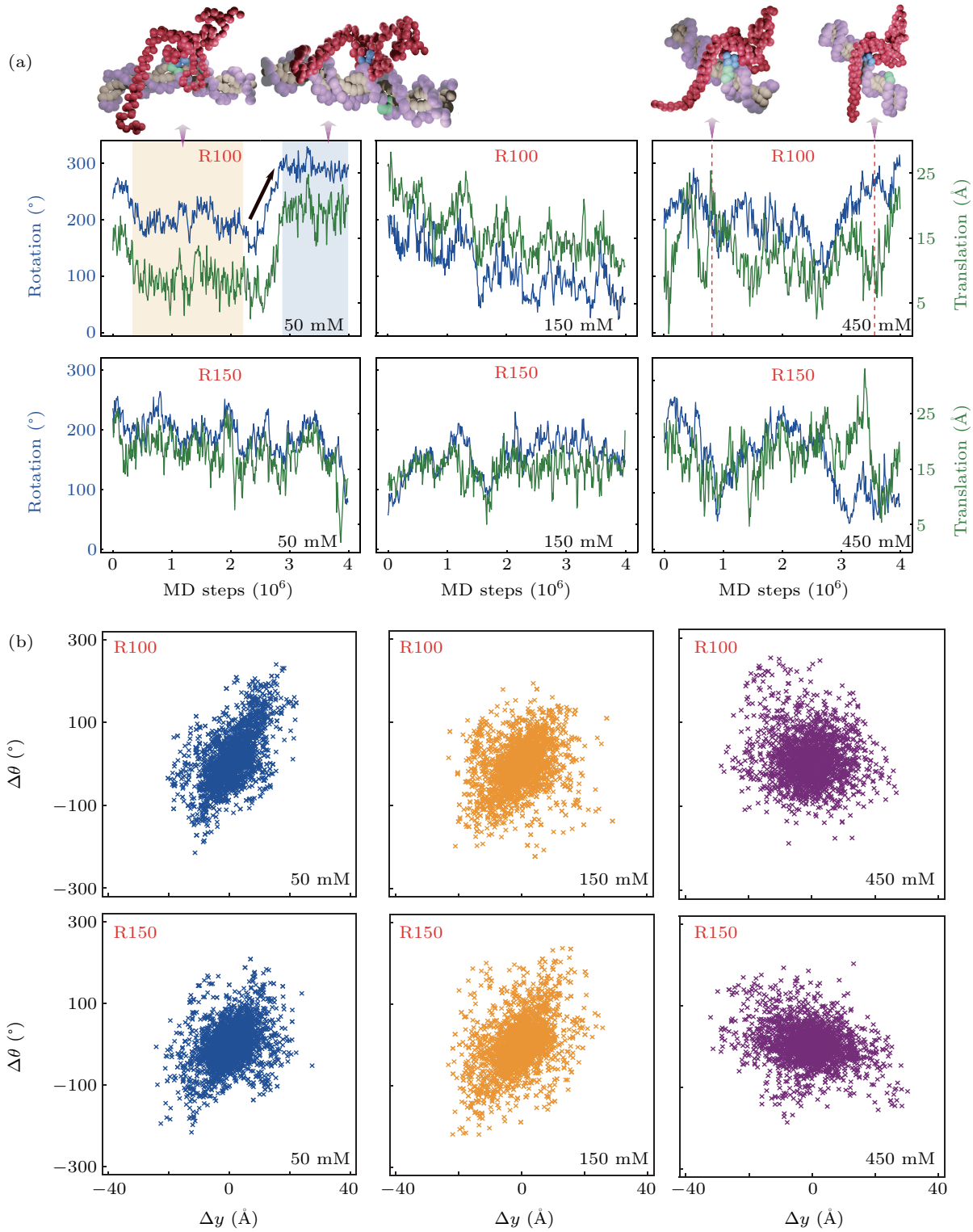



**Fig. 5.** Translational–rotational coupling in the one-dimensional diffusion of XPA along DNA. (a) Representative trajectory fragments of rotation ($\theta$) and translation ($y$) of XPA along R100 (R150) at the salt concentrations of 50 mM, 150 mM, and 450 mM are shown in upper panel (lower panel). Representive structures with DBD bound to DNA at 50 mM and 450 mM are displayed. (b) The correlation plots between $\Delta\theta$ and $\Delta y$ of XPA at 50 mM, 150 mM, and 450 mM (R100: upper panel, R150: lower panel).

DNA at these different salt concentrations. At relatively low salt concentrations (50 mM, 150 mM), where the electrostatic interactions are strong, the one-dimensional diffusion occurs dominantly by the rotation-coupled mechanism. While at high salt concentration (450 mM), the electrostatic interactions become weaker, which makes the protein move more freely along the DNA axis, leading to weak correlation between translational and rotational motions. To more clearly characterize the relationship between rotation and translation during one-dimensional diffusion, we also showed the rational-translational correlation plot in Fig. 5(b) for all trajectories at three different salt concentrations. From these results, one can

see an obvious correlation between the rotation and translation at low salt concentrations (50 mM), while it disappears at relatively high salt concentration (450 mM). Particularly, the rotational–translational correlation is still significant at the physiological relevant salt concentration (150 mM).

### 3.2. Structural properties of XPA

In addition to the diffusion features during this dynamical process, the structural properties of XPA were also studied, since its structure dynamics will surely affect the binding of XPA with other binding partners. The results for these two different DNA sequences are shown in Fig. 6. From our results, one can clearly see that both the DNA binding and varying salt concentration can change the structural properties of XPA and the difference in the effects of R100 and R150 on protein structure at each salt concentration is negligible. In detail, as shown in Fig. 6(a), at low salt concentration (50 mM), due to the strong intra-molecule electrostatic interactions in XPA, these two disordered regions pack against the DBD, leading to small radius of gyration ($R_g$) of XPA (less than 30 Å) in the absence of DNA, but it increases a lot (around 40 Å) upon binding DNA, which may be caused by the occupancy of intra-molecule interacting sites by DNA. While at high salt concentration (450 mM), the electrostatic interactions between intra and inter-molecule is decreased significantly and consequently the intra and inter molecule motions are more freely, leading to large $R_g$ of XPA with or without DNA, and the effects of DNA binding is much deprecated as well (Fig. 6(b)). What is more, in Figs. 6(c) and 6(d), we also plotted the distribution of contact number between DBD and the disordered regions. In accordance with the above analysis, there are a large number of contacts between N(C)-terminal and DBD, and it decreases much unpon binding DNA at low salt concentration (Fig. 6(c)),

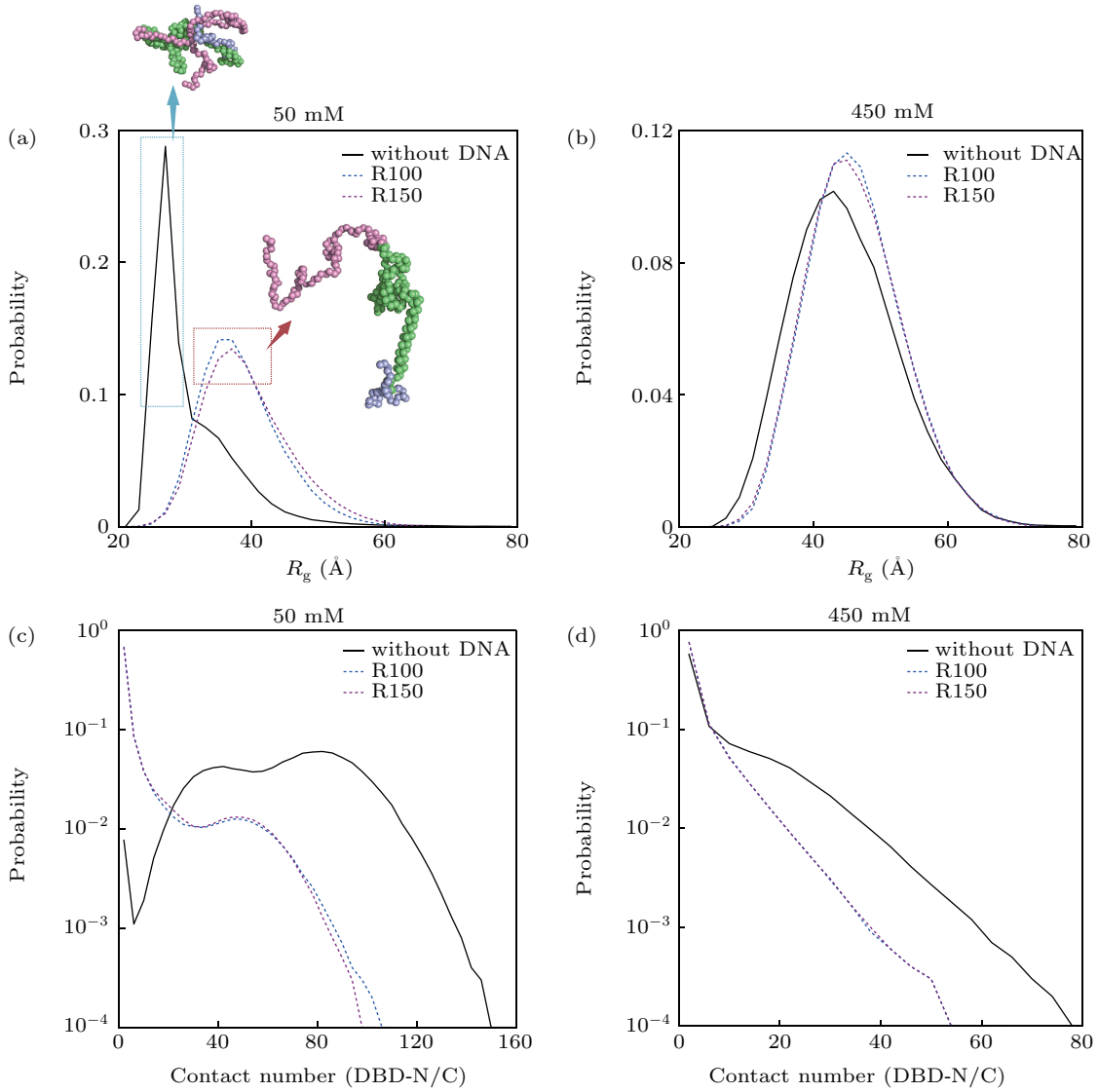



**Fig. 6.** Effect of DNA binding and salt concentration on the conformation of XPA. (a) and (b) $R_g$ distribution of XPA with (R100: blue, R150: purple) and without (black) DNA binding at the salt concentrations of 50 mM (a) and 450 mM (b). The representative structures of XPA are also shown, and the color scheme for domains is the same as in Fig. 1(a). (c) and (d) Distribution of contact number formed between the DBD domain and the two disordered regions of XPA with (R100: blue, R150: purple) and without (black) DNA binding at the salt concentrations of 50 mM (c) and 450 mM (d).

while at high salt concentration, the contact number decreases significantly with or without DNA. Notably, the contact number in the absence of DNA is still less compared with binding DNA. From the analysis of the contact number, it becomes clear that both increasing salt concentrations and DNA binding can lead to the decrease of contact number, indicating more extended conformation of XPA. Based these observations, considering the dominate role of electrostatic interactions in describing the XPA–DNA interactioins and N(C)–DBD interactions in XPA, we can conclude that, both the DNA and salt concentration can act as modulating factors of the XPA structure.

Nowdays, more and more studies have found that the conformation of protein is very sensitive to external stimulus, which play an important role in modulating the conformation of protein involed in various biological activities, like a cascade of reactions or signal transduction in cells, *etc.*[72–76] Interestingly, previously studies have also found that both salt concentration and binding DNA could affect the protein structure.[77–82] Hence, toghter with the findings that the XPA needs to cooperate with other binding partners during the NER pathway, and multiple binding sites for these partners are located at both DBD and disordered regions (N-terminal and C-terminal),[16,17,20,83–86] we can speculate that both the DNA itself and the salt concentration play a role in modulating the NER and at high salt concentration, the protein predominately adopt an extended conformation, ready for interacting with other partners, while at relative low salt concentration, the disordered regions tend to pack tightly with the DBD without DNA, however, upon DNA binding, the N(C)–DBD contacts are much reduced, which can lead to increased exposure of the binding sites in XPA for recruiting other partner proteins.

## 4. Conclusion

Here, we used a coarse-grained model to study the dynamical process of XPA diffusion along two different DNAs. Our results showed that the diffusion properties of XPA along these two DNAs are pretty similar, which indicates the robustness of dynamical behaviour of XPA along DNA. From more detailed perspective, our results showed that the XPA diffuses along the DNA with interface similar to the recognition motif found in the crystal structure, which may promote the search efficiency for the DNA target. Additionally, we also characterized the diffusion process of XPA along DNA. The XPA diffuses along DNA using combination of one-dimensional and three-dimensional search mechanisms, and increasing salt concentration can decrease the proportion of one-dimensional mechanism. During one-dimensional diffusion, the protein displays subdiffusive motions, which has also been observed in several biological system. Additionally, the correlation analysis of rotational and translational motions of DBD along DNA also showed that at relatively low salt concentration, the one-dimensional DBD translational motion is coupled to rotation, that is to say, sliding along helical pitch. Whereas at high salt concentration, the correlation disappears. All these results suggest that, like many other DNA binding proteins,[54,87,88] XPA maintains its position and orientation with respect to targeted DNA base pairs during one-dimensional diffusion at relatively low salt concentration, ensuring the efficient target search.

In addition to binding to DNA, XPA has been found to act a key scaffold protein, interacting with many proteins involved in NER.[86] Therefore, we analyzed the structural features of XPA at various salt concentrations, in the presence and absence of DNA, respectively. The results showed that both DNA and salt concentration can modulate the protein structure. At low concentration, DNA binding can significantly increase the $R_g$ of protein and decrease the contact number between DBD and disordered regions, indicating more extended conformation of XPA. While at high concentration, the weak interaction between protein and DNA and intra-molecular interaction will let the protein to be in the extensible state as well. Based on these discussions, we can speculate that even at relatively low salt concentrations, when XPA bind to DNA, the contacts between DBD and other disordered regions are reduced, which tends to increase the exposure extent of the binding sites for other repair proteins, promoting the scaffold function of the XPA.

## Acknowledgment

The authors thank the insightful discussions with Jun Wang, Wenfei Li, and Wei Wang.

## References

[1] Batty D P and Wood R D 2000 *Gene* **241** 193
[2] Shuck S C, Short E A and Turchi J J 2008 *Cell Res.* **18** 64
[3] Gillet L C and Schärer O D 2006 *Chem. Rev.* **106** 253
[4] Schärer O D 2013 *Cold Spring Harbor Perspect. Biol.* **5** a012609
[5] Hoeijmakers J H 2001 *Nature* **411** 366
[6] Fassihi H, Sethi M, Fawcett H *et al.* 2016 *Proc. Natl. Acad. Sci. USA* **113** E1236
[7] Sugitani N, Sivley R M, Perry K E, Capra J A and Chazin W J 2016 *DNA Repair* **44** 123
[8] Lian F, Yang X, Yang W, Jiang Y and Qian C 2019 *Biochem. Biophys. Res. Commun.* **514** 985
[9] Brown K L, Roginskaya M, Zou Y, Altamirano A, Basu A K and Stone M P 2010 *Nucleic Acids Res.* **38** 428
[10] Kuraoka I, Morita E H, Saijo M, Matsuda T, Morikawa K, Shirakawa M and Tanaka K 1996 *Mutat. Res.-DNA Repair* **362** 87
[11] Missura M, Buterin T, Hindges R, Hübscher U, Kaspárková J, Brabec V and Naegeli H 2001 *Embo J.* **20** 3554
[12] Fadda E 2016 *Comp. Struct. Biotechnol. J.* **14** 78
[13] Yang Z, Roginskaya M, Colis L C, Basu A K, Shell S M, Liu Y, Musich P R, Harris C M, Harris T M and Zou Y 2006 *Biochemistry* **45** 15921
[14] Li C, Golebiowski F M, Onishi Y, Samara N L, Sugasawa K and Yang W 2015 *Mol. Cell* **59** 1025
[15] Tsodikov O V, Ivanov D, Orelli B, Staresincic L, Shoshani I, Oberman R, Schärer O D, Wagner G and Ellenberger T 2007 *Embo J.* **26** 4768
[16] Tripsianes K, Folkers G E, Zheng C, Das D, Grinstead J S, Kaptein R and Boelens R 2007 *Nucleic Acids Res.* **35** 5789

[17] Li L, Elledge S J, Peterson C A, Bales E S and Legerski R J 1994 *Proc. Natl. Acad. Sci. USA* **91** 5012
[18] Sugitani N, Shell S M, Soss S E and Chazin W J 2014 *J. Am. Chem. Soc.* **136** 10830
[19] Hilton B, Shkriabai N, Musich P R, Kvaratskhelia M, Shell S and Zou Y 2014 *Biosci. Rep.* **34**
[20] Topolska-woś A M, Sugitani N, Cordoba J J *et al.* 2020 *Nucleic Acids Res.* **48** 2173
[21] Marteijn J A, Lans H, Vermeulen W and Hoeijmakers J H 2014 *Nat. Rev. Mol. Cell Biol.* **15** 465
[22] Beckwitt E C, Jang S, Detweiler I C *et al.* 2020 *Nat. Commun.* **11** 1
[23] Villa E, Balaeff A, Mahadevan L and Schulten K 2004 *Multiscale Model. Simul.* **2** 527
[24] Kirmizialtin S, Nguyen V, Johnson K A and Elber R 2012 *Structure* **20** 618
[25] Silva D, Weiss D R, Avila F P, Da L, Levitt M, Wang D and Huang X 2014 *Proc. Natl. Acad. Sci. USA* **111** 7665
[26] Hognon C, Garaude S, Timmins J, Chipot C, Dehez F and Monari A 2019 *J. Phys. Chem. Lett.* **10** 7200
[27] Li W, Wang W and Takada S 2014 *Proc. Natl. Acad. Sci. USA* **111** 10550
[28] Li W, Wolynes P G and Takada S 2011 *Proc. Natl. Acad. Sci. USA* **108** 3504
[29] Freeman G S, Hinckley D M, Lequieu J P, Whitmer J K and De pablo J J 2014 *J. Chem. Phys.* **141** 10B615_1
[30] Freeman G S, Lequieu J P, Hinckley D M, Whitmer J K and De pablo J J 2014 *Phys. Rev. Lett.* **113** 168101
[31] Tan C, Terakawa T and Takada S 2016 *J. Am. Chem. Soc.* **138** 8512
[32] Potoyan D A, Zheng W, Komives E A and Wolynes P G 2016 *Proc. Natl. Acad. Sci. USA* **113** 110
[33] Brandani G B, Niina T, Tan C and Takada S 2018 *Nucleic Acids Res.* **46** 2788
[34] Kenzaki H, Koga N, Hori N, Kanada R, Li W, Okazaki K, Yao X and Takada S 2011 *J. Chem. Theory Comput.* **7** 1979
[35] Li W, Terakawa T, Wang W and Takada S 2012 *Proc. Natl. Acad. Sci. USA* **109** 17789
[36] Li W, Wang J, Zhang J and Wang W 2015 *Curr. Opin. Struct. Biol.* **30** 25
[37] Brandani G B and Takada S 2018 *PLoS Comput. Biol.* **14** e1006512
[38] Li W, Wang J, Zhang J, Takada S and Wang W 2019 *Phys. Rev. Lett.* **122** 238102
[39] Tan C and Takada S 2020 *Proc. Natl. Acad. Sci. USA* **117** 20586
[40] Lu J, Zhang X, Wu Y, Sheng Y, Li W and Wang W 2021 *Biophys. J.* **120** 1971
[41] Dey P and Bhattacherjee A 2019 *J. Phys. Chem. B* **123** 10354
[42] Terakawa T and Takada S 2014 *J. Chem. Theory Comput.* **10** 711
[43] šali A, Potterton L, Yuan F, Van vlijmen H and Karplus M 1995 *Proteins* **23** 318
[44] Lu X and Olson W K 2008 *Nat. Protoc.* **3** 1213
[45] Delano W L 2002 *CCP4 Newsletter on protein crystallography* **40** 82
[46] Lian F, Yang X, Jiang Y, Yang F, Li C, Yang W and Qian C 2020 *Int. J. Biol. Macromol.* **148** 466
[47] Terakawa T, Kenzaki H and Takada S 2012 *J. Am. Chem. Soc.* **134** 14555
[48] Iwahara J, Zweckstetter M and Clore G M 2006 *Proc. Natl. Acad. Sci. USA* **103** 15062
[49] Iwahara J and Clore G M 2006 *Nature* **440** 1227
[50] Vuzman D and Levy Y 2012 *Mol. Biosyst.* **8** 47
[51] Von hippel P H and Berg O G 1989 *J. Biol. Chem.* **264** 675
[52] Halford S E and Marko J F 2004 *Nucleic Acids Res.* **32** 3040
[53] Von hippel P H 1994 *Science* **263** 769
[54] Givaty O and Levy Y 2009 *J. Mol. Biol.* **385** 1087
[55] Blainey P C, Van oijen A M, Banerjee A, Verdine G L and Xie X S 2006 *Proc. Natl. Acad. Sci. USA* **103** 5752
[56] Pingoud A and Wende W 2007 *Structure* **15** 391
[57] Golding I and Cox E C 2006 *Phys. Rev. Lett.* **96** 098102
[58] Guigas G and Weiss M 2008 *Biophys. J.* **94** 90
[59] Soula H, Caré B, Beslon G and Berry H 2013 *Biophys. J.* **105** 2064
[60] Havlin S and Ben-avraham D 1987 *Adv. Phys.* **36** 695
[61] Scher H and Montroll E W 1975 *Phys. Rev. B* **12** 2455
[62] Jeon J, Monne H M, Javanainen M and Metzler R 2012 *Phys. Rev. Lett.* **109** 188103
[63] Xia C, He X, Wang J and Wang W 2020 *Phys. Rev. E* **102** 062424
[64] Liu L, Cherstvy A G and Metzler R 2017 *J. Phys. Chem. B* **121** 1284
[65] Metzler R, Jeon J, Cherstvy A G and Barkai E 2014 *Phys. Chem. Chem. Phys.* **16** 24128
[66] Barr J J, Auro R, Sam-soon N *et al.* 2015 *Proc. Natl. Acad. Sci. USA* **112** 13675
[67] Saxton M J 2007 *Biophys. J.* **92** 1178
[68] Metzler R and Klafter J 2000 *Phys. Rep. -Rev. Sec. Phys. Lett.* **339** 1
[69] Golan Y and Sherman E 2017 *Nat. Commun.* **8** 1
[70] Banerjee A, Santos W L and Verdine G L 2006 *Science* **311** 1153
[71] Blainey P C, Luo G, Kou S, Mangel W F, Verdine G L, Bagchi B and Xie X S 2009 *Nat. Struct. Mol. Biol.* **16** 1224
[72] Goh C, Milburn D and Gerstein M 2004 *Curr. Opin. Struct. Biol.* **14** 104
[73] Hartl F U 1996 *Nature* **381** 571
[74] Park P S, Lodowski D T and Palczewski K 2008 *Ann. Rev. Pharmacol. Toxicol.* **48** 107
[75] Zhang M, Tanaka T and Ikura M 1995 *Nat. Struct. Biol.* **2** 758
[76] Chu X, Liu F, Maxwell B A, Wang Y, Suo Z, Wang H, Han W and Wang J 2014 *PLoS Comput. Biol.* **10** e1003804
[77] Azia A and Levy Y 2009 *J. Mol. Biol.* **393** 527
[78] Maity H, Muttathukattil A N and Reddy G 2018 *J. Phys. Chem. Lett.* **9** 5063
[79] Levy Y, Onuchic J N and Wolynes P G 2007 *J. Am. Chem. Soc.* **129** 738
[80] Liu J L, Rigolet P, Dou S, Wang P and Xi X G 2004 *J. Biol. Chem.* **279** 42794
[81] Fan J and Pavletich N P 2012 *Genes Dev.* **26** 2337
[82] Rudolph J, Mahadevan J and Luger K 2020 *Biochemistry* **59** 2003
[83] Nocentini S, Coin F, Saijo M, Tanaka K and Egly J 1997 *J. Biol. Chem.* **272** 22991
[84] Buchko G W, Daughdrill G W, De lorimier R *et al.* 1999 *Biochemistry* **38** 15116
[85] Mer G, Bochkarev A, Gupta R, Bochkareva E, Frappier L, Ingles C J, Edwards A M and Chazin W J 2000 *Cell* **103** 449
[86] Borszéková pulzová L, Ward T A and Chovanec M 2020 *Int. J. Mol. Sci.* **21** 2182
[87] Jeltsch A, Alves J, Wolfes H, Maass G and Pingoud A 1994 *Biochemistry* **33** 10215
[88] Sun J, Viadiu H, Aggarwal A K and Weinstein H 2003 *Biophys. J.* **84** 3317