2246–2257 *Nucleic Acids Research, 2020, Vol. 48, No. 5* *Published online 12 February 2020*
doi: 10.1093/nar/gkaa078

Sequence specificity, energetics and mechanism of mismatch recognition by DNA damage sensing protein Rad4/XPC

Abhinandan Panigrahi†, Hemanth Vemuri†, Madhur Aggarwal, Kartheek Pitta and Marimuthu Krishnan*

Center for Computational Natural Sciences and Bioinformatics (CCNSB), International Institute of Information Technology, Gachibowli, Hyderabad, Telangana 500032, India

Received July 08, 2019; Revised January 27, 2020; Editorial Decision January 28, 2020; Accepted February 10, 2020

ABSTRACT

The ultraviolet (UV) radiation-induced DNA lesions play a causal role in many prevalent genetic skin-related diseases and cancers. The damage sensing protein Rad4/XPC specifically recognizes and repairs these lesions with high fidelity and safeguards genome integrity. Despite considerable progress, the mechanistic details of the mode of action of Rad4/XPC in damage recognition remain obscure. The present study investigates the mechanism, energetics, dynamics, and the molecular basis for the sequence specificity of mismatch recognition by Rad4/XPC. We dissect the following three key molecular events that occur as Rad4/XPC tries to recognize and bind to DNA lesions/mismatches: (a) the association of Rad4/XPC with the damaged/mismatched DNA, (b) the insertion of a lesion-sensing β-hairpin of Rad4/XPC into the damage/mismatch site and (c) the flipping of a pair of nucleotide bases at the damage/mismatch site. Using suitable reaction coordinates, the free energy surfaces for these events are determined using molecular dynamics (MD) and umbrella sampling simulations on three mismatched (CCC/CCC, TTT/TTT and TAT/TAT mismatches) Rad4-DNA complexes. The study identifies the key determinants of the sequence-dependent specificity of Rad4 for the mismatches and explores the ramifications of specificity in the aforementioned events. The results unravel the molecular basis for the high specificity of Rad4 towards CCC/CCC mismatch and lower specificity for the TAT/TAT mismatch. A strong correlation between the depth of β-hairpin insertion into the DNA duplex and the degree of coupling between the hairpin insertion and the flipping of bases is also observed. The interplay of the conformational flexibility of mismatched bases, the depth of β-hairpin insertion, Rad4-DNA association energetics and the Rad4 specificity explored here complement recent experimental FRET studies on Rad4-DNA complexes.

INTRODUCTION

The genome integrity is safeguarded by DNA repair proteins that work in concert with other downstream factors to sense and repair DNA damage in cells (1–6). Given the large numbers of nucleotide base pairs in genomes, how these proteins search, interrogate, recognize and repair specific DNA lesions with high fidelity within a crowded cellular environment is still intriguing and unclear (7–10). Thus, deciphering the molecular mechanism of the mode of action of these proteins during DNA repair is an outstanding challenge in this area of research (11–13).

Cyclobutane pyrimidine dimer (CPD) formed by the cyclization of two adjacent thymine bases in a DNA strand is the most common UV radiation-induced DNA lesion that plays a causal role in many prevalent genetic skin-related diseases and cancers in humans (14,15). The protein xeroderma pigmentosum C (XPC) recognizes and repairs CPD and other such lesions using the nucleotide excision repair (NER) mechanism (16–18). Previous studies on the sensitivity of various DNA lesions to the human excision mechanism indicated that mismatches, small and bulky lesions including CPD are excised with different efficiencies (19).

The lack of crystal or solution structure of lesion-containing DNA-XPC complex is a major limiting factor in understanding the molecular basis of XPC-mediated DNA damage repair in humans (20). Radiation sensitive 4 (Rad4), which is the yeast orthologue of XPC, is a key protein involved in recognizing a wide variety of DNA damages in yeast cells (21,22). Owing to its functional, and structural

*To whom correspondence should be addressed. Tel: +91 40 66531447; Fax: +91 40 66531413; Email: m.krishnan@iiit.ac.in
†The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.

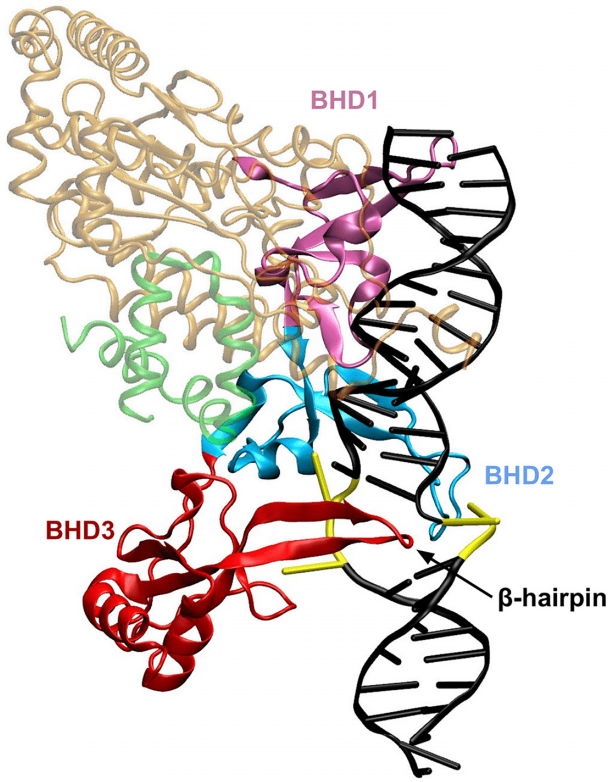

Figure 1. The crystal structure of the Rad4-DNA complex with TTT/TTT mismatch.The nucleotides at the mismatch site are shown in yellow. The image was generated using the POV-Ray renderer (www.povray.org) in VMD (33).

similarity with XPC and due to the availability of the crystal structure of DNA–Rad4 complex, Rad4 serves as a useful model to investigate XPC-mediated DNA damage repair in humans.

The crystal structure of the mismatch- or CPD-containing DNA–Rad4 complex reveals that the DNA duplex is distorted in the vicinity of the mismatch or the lesion and that the lesion or the two mismatched base pairs are entirely flipped out of the double helix such that they are exposed to water, whereas the partner bases are caught safely by Rad4 (21). The structure also shows that the three β-hairpin domains (BHDs) (BHD1, BHD2 and BHD3, as shown in Figure 1) of Rad4 are critically involved in DNA damage sensing and repair. In particular, BHD2 and BHD3 bind to a 4-bp DNA segment that contains the mismatches or CPD lesion, whereas BHD1 binds to an 11-bp segment of the matched or undamaged part of the DNA (21,22). Although the crystal structure provides only a static picture of the end-state of Rad4-DNA complexation process, it offers vital clues on the following three key molecular events that likely govern the initial stage of damage recognition and repair dynamics: (a) the association of DNA and Rad4, (b) the insertion of a β-hairpin in BHD3 into DNA duplex at the mismatch or lesion site and (c) the flipping out of the pair of nucleotide bases at the mismatch or lesion site (23–25). However, the exact mechanism and order of these events remain unclear.

The two likely mechanisms proposed for DNA lesion recognition by Rad4/XPC are the conformational capture or passive mechanism and the correlated motion or active mechanism (24–27). In the former mechanism, one or two bases opposite the lesion are spontaneously flipped out prior to binding of Rad4/XPC to DNA. These extruded bases are recognized and captured by Rad4/XPC as it scans the damaged DNA. Subsequently, the insertion of the BHD3 β-hairpin into DNA duplex forces the lesion to evict from the duplex (26). However, in the latter mechanism, both the base flipping and lesion eviction occur after or during the BHD3 β-hairpin insertion post Rad4–DNA binding (27).

A computational study by Broyde *et al.* presented data favoring conformation capture of the bases, while maintaining that the order of these events is likely to be lesion dependent (24,28). Furthermore, by resolving the crystal structure of pyrimidine–pyrimidone (6-4) photoproduct (6-4PP)-containing DNA bound to Rad4, they examined the order of events for Rad4 binding to 6-4PP and compared the excision efficiency between many structurally different DNA lesions (29,30). These computational studies indicated that the initial interaction between BHD2 of Rad4 and DNA via the minor groove and the subsequent BHD3's β-hairpin insertion into the DNA duplex can potentially discriminate between poorly- and well-recognized lesions.

A recent fluorescence lifetime study combined with MD simulations examined correlations between conformational heterogeneity, intrinsic deformability and Rad4-binding specificity of a set of 3-bp mismatched DNA (28,31). The results demonstrated that highly dynamic and conformationally heterogeneous mismatched DNA with higher deformability are recognized with high specificity by Rad4. That is, the higher the conformational heterogeneity and deformability of mismatched DNA, the higher is the specificity for Rad4. In particular, among all the 3-bp mismatched DNA studied, Rad4 exhibited high, medium, and low specificity for the CCC/CCC, TTT/TTT and TAT/TAT mismatches, respectively.

In order to understand the mechanism and molecular basis of specificity of mismatch recognition by Rad4/XPC, we investigate the structure, dynamics and energetics of three mismatched DNA-Rad4 complexes (CCC/CCC, TTT/TTT and TAT/TAT) using molecular dynamics and umbrella sampling simulations (32). Since Rad4/XPC is known to bind to these mismatched DNA, despite not being excised by the NER machinery, we chose these DNA-Rad4 complexes as model systems to explore the mechanism of mismatch recognition by Rad4. These three specific mismatch sequences were chosen on the basis of the availability of experimental FRET data and keeping in view that they span the entire specificity spectrum (31).

The primary objective of this study is to discern the ramifications of sequence-dependent specificity and associated energetics of the aforementioned three key molecular events (base flipping, Rad4-DNA association, BHD3 β-hairpin insertion). Additional molecular-level insights into the mismatch-induced deformation and Rad4's affinity for the mismatches are also provided.

MATERIALS AND METHODS

Models

The X-ray crystal structure of the Rad4-DNA complex with TTT/TTT mismatch (PDB ID: 2QSH) was used as the base template for building the CCC/CCC, TAT/TAT, and TTT/TTT mismatched DNA-Rad4 complexes (21). In this crystal structure of the complex, the β-hairpin of BHD3 domain of Rad4 is inserted into the DNA duplex and a pair of bases (denoted by uX16 and uY17 in Figure 2) on the undamaged strand of the DNA opposite to the lesion are flipped out of the duplex. The missing loop G(518)R(519)P(520)K(521) G(522)E(523)A(524)E(525) in the BHD2 domain of RAD4 was modeled using Modeller, while the two missing disordered thymine bases in the damaged strand of DNA were modelled in their extra-helical conformational states using the PSFGEN module of VMD (33,34). Following this, each system was solvated in a cubic water box using a three-point water model (TIP3P), to which 15 Na^+ ions were added to neutralize the system (35,36). The models of Rad4-free DNA with the same sequences were modelled by first generating perfectly matched B-DNA using the Nucleic Acid Builder (Amber16) tool, and then the mismatches were introduced using the Swapna module of UCSF Chimera (37,38) (refer SI for more details).

Molecular dynamics simulation

50 ns MD simulations of all Rad4-bound (referred to as REC) and Rad4-free (referred to as BDI) mismatched DNA were performed using GROMACS 2016.3 (patched with PLUMED 2.4) simulation package with the amber99bsc1 force field for the protein and DNA, and the TIP3P model for water molecules (35–36,39–43). Bond lengths were constrained using the LINear Constraint Solver algorithm (LINCS) (44). Periodic boundary conditions were employed using a periodic box of dimensions 120 × 95 × 95 Å along all the three directions. The long-range electrostatic interactions were treated with the particle mesh ewald (PME) approach with a direct space cut-off of 14 Å (45,46). Pressure was maintained at 1 bar using the Parrinello-Rahman method with a time constant of 1.0 ps, while the temperature was maintained at 300K using a velocity-rescale thermostat with a time constant of 0.1 ps (47).

Energy minimization for all the systems was done using the steepest descent algorithm for 50000 steps followed by another 50000 steps using the conjugate gradient algorithm with convergence reached when the net force on the system is <0.239 kcal mol^{-1} $Å^{-1}$ (48,49). Equilibration of the energy minimized structures was carried out in three different stages: the initial stage was carried out in an NVT ensemble and involved temperature annealing from 0 to 300 K for 140 ps with position restraints (with a harmonic spring constant of 239 kcal mol^{-1} $Å^{-2}$) on DNA–protein complex heavy atoms. This stage also involves assigning velocities from the Maxwell-Boltzmann distribution at 300 K with a random seed. The next stage was performed in an NPT ensemble with the positional restraints still in place at 300 K and 1 bar pressure for 2 ns. In the last stage of equilibration, all the positional restraints were removed and each system was subjected to 3 ns NPT equilibration followed by a 50 ns NPT production run.

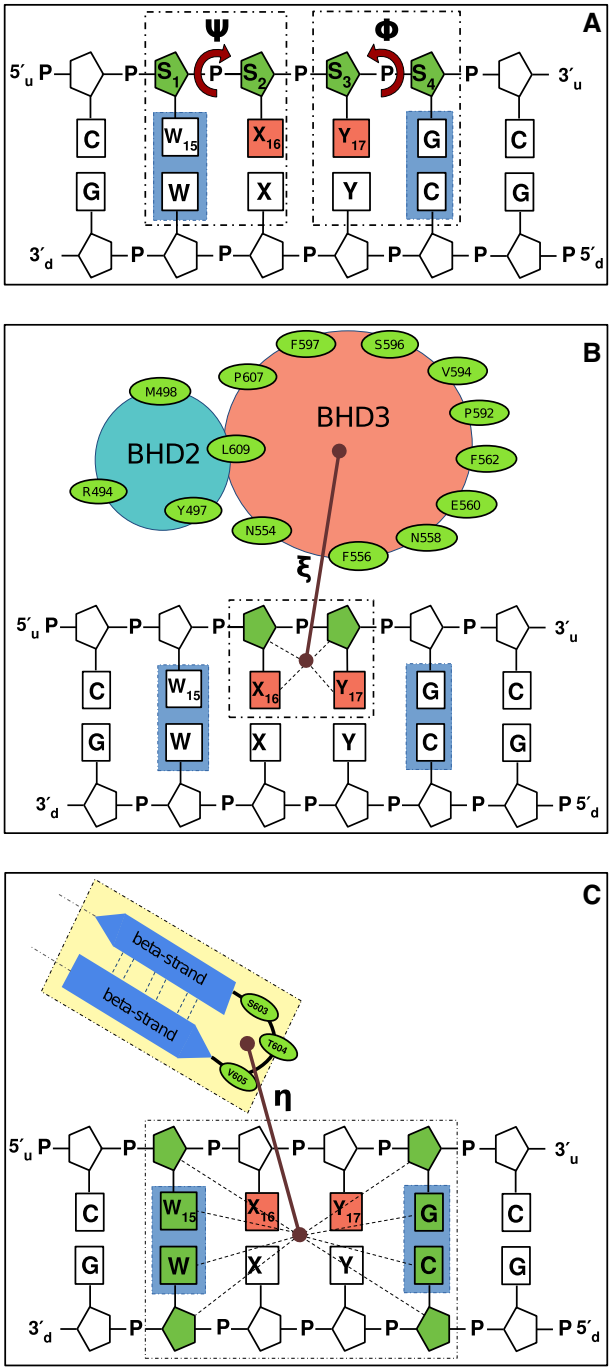

Figure 2. (**A**) Φ and Ψ are pseudo-dihedral angles that describe the rotation of the bases X and Y (red), respectively, around their corresponding sugar-sugar (green) linkages. (**B**) ξ is the distance between the DNA damage site (dotted box) and the center of mass (COM) of key residues (green ellipses) in the BHD2/3 domains of Rad4 (**C**) η is the distance between the COM of three key residues (green ellipses) of the BHD3 β-hairpin and the COM of the sugar rings and their corresponding nucleotides (green) from the binding region (dotted box).

Umbrella sampling

Using a set of reaction coordinates or collective variables (CVs) (shown in Figure 2) to describe the Rad4-DNA association, BHD3 β-hairpin insertion and the Rad4-assisted base flipping, umbrella sampling (US) simulations were performed to compute the potentials of mean force (PMFs) for these events in the chosen complexes. The conformations obtained from the unbiased simulations have been used as the starting structure for each of the following biased simulations. The Weighted Histogram Analysis Method (WHAM) (membrane.urmc.rochester.edu/?page_id=126) was used to determine PMFs from CV trajectories obtained from the US simulations (50). Each US simulation (3 ns trajectory per window) was performed in the NPT ensemble, under the same conditions as those of unbiased MD runs. The entire simulation length (excluding the first 50 ps) per window was used for the WHAM. The spring constants used for the biasing potential were 2.39 kcal mol^{-1} $Å^{-2}$ for base flipping simulation and 15.0 kcal mol^{-1} $Å^{-2}$ for insertion and association.

Additional computational details and the convergence analysis of the calculated free energy profiles (Supplementary Figures S1–S3 and S13–S19) are provided in the Supporting Information (SI).

Collective variables for base flipping. The pseudo-dihedral angles Ψ and Φ were chosen as the reaction coordinates to describe the flipping dynamics of uX16 and uY17 bases, respectively, of DNA (*d* and *u* denote the damaged and undamaged strands of DNA) (Figure 2A) (51,52). These reaction coordinates were chosen such that a clockwise rotation (Ψ: 0° → 90° → 180°) of uX16 about the pseudo-central bond between adjacent sugar moieties (refer Figure 2A) define the major-groove extrusion pathway for uX16. However, a counter-clockwise rotation of uY17 (Φ: 0° → −90° → −180°) defines its major-groove pathway in RAD4-DNA complex. The allowed range (from −180° to +180°) of Φ and Ψ was divided into 73 windows of width 5° each and US was performed in each window for 3 ns. The β-hairpin was in the inserted conformation during base flipping US simulations. To reduce the computational requirement of these biased simulations, bias was applied equally on both the pseudo-dihedral angles simultaneously. Additional runs in which bias was applied individually on the pseudo-dihedral angle demonstrated that the actual sampling of a pseudo-dihedral angle was almost independent of the bias applied on the other pseudo-dihedral angle (refer to Supplementary Figures S5 and S24 in SI for further details).

Collective variable for β-hairpin insertion. To describe the β-hairpin insertion, the distance (denoted by η) between the center of mass (COM) of the backbone heavy atoms of three residues (SER603, THR604 and VAL605) on the leading edge of the BHD3 β-hairpin and the COM of the sugar ring heavy atoms of the nucleotides (*uG*18, *dC*8, *uT*15 and *dT*11) at the lesion site was used as a suitable CV. The relevant range of η (from 1 to 9 Å) was divided into 40 windows of size 0.2 Å for US. The bases uX16 and uY17 were in their extra-helical (flipped out) conformations during the US simulations of β-hairpin insertion.

Collective variable for Rad4-DNA association. For studying the association of BHD2/BHD3 domains of Rad4 with the lesion/mismatch site, another distance based collective variable, ξ, was chosen in accordance with previous studies on the subject (24). ξ denotes the distance between the COM of the backbone heavy atoms of the two mismatched paired bases of interest and the COM of the backbone heavy atoms of the residues (R494, Y497, M498, N554, F556, N558, E560, F562, P592, V594, S596, F597, P607 and L609) of Rad4 that constitute the binding pocket for the flipped partner bases. The range of ξ (from 5 to 21 Å) was divided into 40 bins of size 0.4 Å each. Each US window for Φ, Ψ, η and ξ was sampled for 3 ns. Since our simulations were started off with Rad4 already bound to the DNA, the association of Rad4 and DNA was modelled here as a dissociation process.

RESULTS AND DISCUSSION

Base flipping

Free energy profiles. The calculated free energy profiles for flipping of uX16 and uY17 in the absence and presence of Rad4 for all systems studied are shown in Figure 3. $F_{BDI}(\Phi)$ exhibits a global minimum at $\Phi \sim 0°$ (referred to as Φ^{BDI}_{min}), indicating that uY17 prefers the intra-helical conformation in the absence of Rad4 in all the systems. The energy barrier for intra→extra-helical flipping of uY17 in BDI is highest (∼18 kcal/mol) for TAT/TAT, whereas it is ∼12 and ∼11.6 kcal/mol for TTT/TTT and CCC/CCC, respectively.

$F_{BDI}(\Psi)$ also attains its global minimum (referred to as Ψ^{BDI}_{min}) in the intra-helical region, indicating that uX16 also prefers intra-helical conformation in the absence of Rad4. However, unlike $F_{BDI}(\Phi)$, the location of the global minimum varies in the range 20° to 65° across all systems, which suggests that the most-probable conformation of uX16 is different for different sequences of BDI. The intra→extra-helical flipping barrier for uX16 is highest (∼18 kcal/mol) for TAT/TAT, followed by TTT/TTT (∼15 kcal/mol), whereas for CCC/CCC (∼5.7 kcal/mol) it is almost a factor of three less than that for other systems.

The anomalously low flipping barrier observed for uX16 in CCC/CCC suggests that uX16 is conformationally more flexible compared to uY17, and that the propensity of uX16 to flip spontaneously to extra-helical conformation in the absence of Rad4 is relatively higher in CCC/CCC than for other systems. Additionally, in Rad4-free systems, there appears to be a preferred path for the flipping of uX16, whereas the observed higher degree of symmetry of $F_{BDI}(\Phi)$ around the global energy minimum (Supplementary Figures S12 and S25) prevents us from arriving at a similar conclusion for uY17.

The binding of Rad4 to DNA significantly alters the base flipping free energy profiles. In particular, binding-induced shift in the global energy minimum was observed in both $F(\Phi)$ and $F(\Psi)$, which is indicative of alteration of the most stable conformations of uX16 and uY17; both uX16 and uY17 bases exhibited conformational transitions from the previously inhabited intra-helical region to the extra-helical region post Rad4–DNA binding. We calculated $\Delta\Phi_{min} = \Phi^{BDI}_{min} - \Phi^{REC}_{min}$ and $\Delta\Psi_{min} = \Psi^{BDI}_{min} - \Psi^{REC}_{min}$ to quantify the

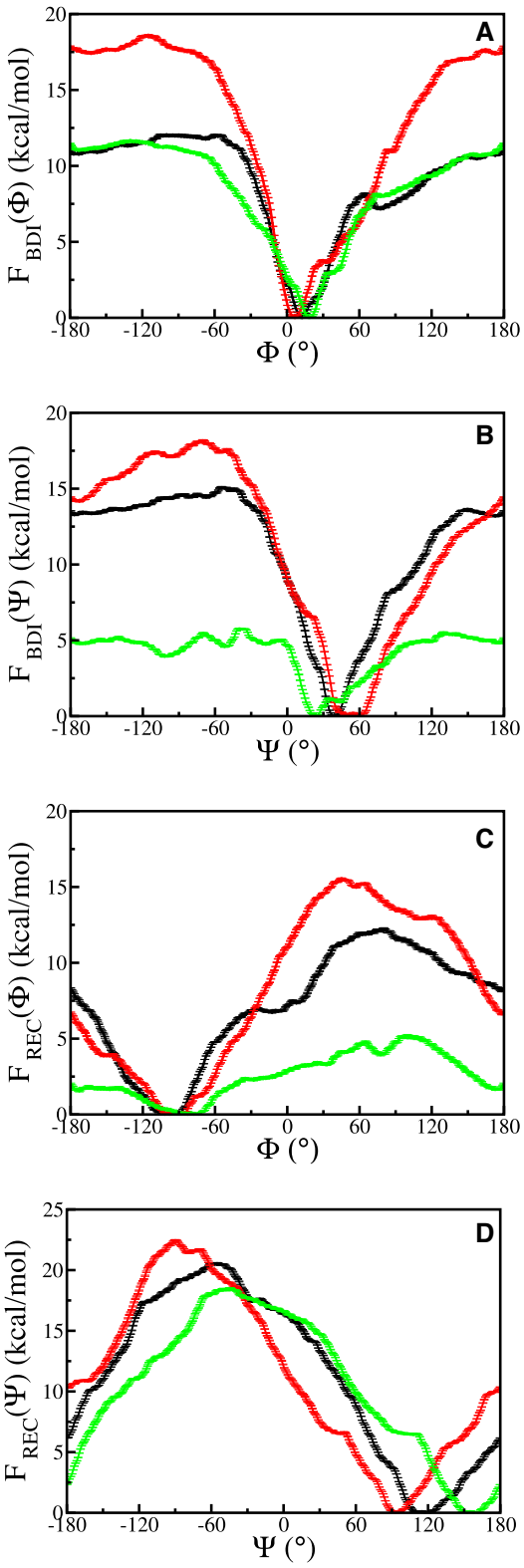

Figure 3. The base flipping free energy profiles in the absence (**A** and **B**) and presence (**C** and **D**) of Rad4 for TTT/TTT (black), TAT/TAT (red) and CCC/CCC (green).

degree of Rad4-induced shift in the global energy minimum of F(Φ) and F(Ψ), respectively. The binding-induced shift is quite pronounced for both Φ ($\Delta\Phi_{min}$ varies in the range −95° to −76°) and Ψ ($\Delta\Psi_{min}$ varies in the range 93–155°) for all systems. The structures of the Rad4–DNA complexes corresponding to the low- and high-energy regions of the base flipping free energy profiles are shown in Supplementary Figure S10.

The binding of Rad4 also alters the energy barriers separating the intra-helical and extra-helical conformational states of uX16 and uY17. Compared to the intra→extra-helical barriers in the Rad4-free state, the extra→intra-helical barriers in the Rad4-bound state decreased for uY17 and increased for uX16 across all systems. The binding-induced changes in the base flipping energetics of damaged bases are significant for CCC/CCC. In particular, the energy barrier for the base uY17 to flip from the most-favorable extra-helical state to the intra-helical state is the least for CCC/CCC. The observed low flipping barrier of uY17 in CCC/CCC can be attributed to a weaker interaction of uY17 with Rad4 in CCC/CCC compared to that in other mismatches (Supplementary Table S2). The corresponding flipping barrier for uX16 is higher than that for uY17 for all systems.

The calculated barrier heights reported in Table 1 suggest that the mismatch-induced and Rad4-induced changes in energy barriers and conformational flexibility of DNA are good predictors for the strength of the specificity with which Rad4 recognizes and binds to DNA. The conformational flexibility enables DNA to sample different *iso*-energetic conformations that differ from one another structurally. Consequently, the populations of DNA conformations that favor recognition by Rad4 increase, which, in turn, enhances the specificity for Rad4. The intermediate specificity of TTT/TTT and low specificity of TAT/TAT follow from their higher energy barriers to flipping, and thus lower conformational heterogeneity, compared to CCC/CCC. The flatter, lower barriers to flipping for CCC/CCC lead to vastly shorter base flipping times for the bases uX16 and uY17 which allow the system to sample many more structural conformations. Given activation barriers (ΔF) for base flipping, the mean flipping times (τ) for uX16 and uY17 were estimated using the transition state theory ($\tau = \frac{h}{k_b T}\exp(\frac{\Delta F}{k_b T})$) and the results are reported in Table 1.

The free energy profiles of Φ and Ψ calculated from the unbiased MD simulations are reported (Supplementary Figures S4 and S20–S23) in the SI.

Flexibility of mismatched bases. To understand the possible role of flexibility of uX16 and uY17 in Rad4 recognition and binding, we calculated the root mean square fluctuations (RMSF), root mean square deviations (RMSD) of uX16 and uY17, and the standard deviations of Φ and Ψ from unbiased MD trajectories of Rad4-free DNA and Rad4–DNA complexes (Supplementary Table S4). Since the Rad4-free DNA represents the initial stage of the mismatch recognition process prior to binding, it is reasonable to expect the flexibility of uX16 and uY17 to contribute to the initial recognition of mismatches by Rad4. On the contrary, the flexibility of uX16 and uY17 calculated for the Rad4-bound DNA is an indicator of how tightly

Table 1. The activation barriers (ΔF in kcal/mol) and the corresponding mean base flipping times (τ in ns) as estimated from transition state theory of uX16 and uY17 bases in the absence and presence of Rad4

sequence	ΔF^{uY17}_{BDI}	ΔF^{uY17}_{REC}	ΔF^{uX16}_{BDI}	ΔF^{uX16}_{REC}	τ^{uY17}_{BDI}	τ^{uY17}_{REC}	τ^{uX16}_{BDI}	τ^{uX16}_{REC}
TTT/TTT	12.02 ± 0.04	12.18 ± 0.07	15.03 ± 0.04	20.55 ± 0.05	1.05×10^{5}	1.36×10^{5}	1.69×10^{7}	1.88×10^{11}
TAT/TAT	18.57 ± 0.04	15.50 ± 0.02	18.12 ± 0.04	22.34 ± 0.06	6.68×10^{9}	3.75×10^{7}	3.09×10^{9}	3.88×10^{12}
CCC/CCC	11.64 ± 0.03	5.15 ± 0.07	5.76 ± 0.01	18.45 ± 0.06	5.47×10^{4}	9.53×10^{-1}	2.69	5.45×10^{9}

these bases are held by Rad4 in their extrahelical states. The flexibility of uX16 and uY17 also provides an entropic contribution to the overall binding affinity of the Rad4-DNA complex. Thus, we anticipate a possible correlation between the flexibility of uX16 and uY17 in Rad4-free DNA and the sequence-dependent specificity of mismatch recognition by Rad4. Since uX16 and uY17 primarily sample the low-energy regions of their respective lowest-energy basins within the time scales of the unbiased MD simulation, the aforementioned metrics merely quantify the intra-basin vibrational flexibility of these bases. The widths of the lowest-energy basins in the umbrella sampling-derived free energy profiles F(Φ) and F(Ψ) were also calculated to quantify the vibrational flexibility of uX16 and uY17 (Table 2 and Supplementary Table S1 in the SI).

The calculated values of the aforementioned quantities (Table 2) for Rad4-free DNA indicate that the flexibility of uX16 is slightly greater than that of uY17 for CCC/CCC and to a lesser extent in TAT/TAT, while uX16 and uY17 have similar flexibility in TTT/TTT. Overall, the flexibility of these bases is highest in CCC/CCC, followed by TTT/TTT and TAT/TAT. The similarity in the order of specificity of Rad4 and the order of flexibility of bases in Rad4-free DNA suggests a possible correlation between them. The observed sensitivity of uX16 to sequence variation can be attributed partly to the fact that the three nucleotides involved in the definition of Ψ (i.e. the rotation of uX16) vary across all systems, whereas for Φ only the flipping base is different. This also explains to some extent the similarity in TTT/TTT and TAT/TAT free energy profiles. However, it is not clear whether both uX16 and uY17 should be flexible or if the flexibility of one of them is sufficient for effective recognition of damage by Rad4.

The flexibility of uX16 and uY17 is altered significantly upon binding of Rad4 to DNA. In particular, uY17 becomes more flexible, while uX16 becomes more rigid relative to their Rad4-free counter parts, especially more so in CCC/CCC. This is further evident from the calculated mean Rad4–uX16 and Rad4–uY17 interaction energies for CCC/CCC, wherein Rad4–uX16 interactions are much stronger than Rad4–uY17 (Supplementary Table S2). Overall, the vibrational flexibility of uY17 is greater than that of uX16 in all the Rad4–DNA complexes investigated here. The dynamic nature of CCC/CCC (and to a lesser extent TTT/TTT and TAT/TAT) is retained even after binding with Rad4, which is in agreement with recent FRET-based experimental studies (31). The lower stacking potential and fewer hydrogen bonds of the C-C mismatch give rise to the higher flexibility of mismatched bases in CCC/CCC (53–60).

In addition to the vibrational flexibility of these bases, conformational transitions of them between different basins also contribute to the overall flexibility of DNA. The base flipping activation barriers (refer the previous section) essentially quantify the degree of conformational flexibility of these bases. As discussed in the previous section, the conformational flexibility is highest for CCC/CCC, followed by TTT/TTT and TAT/TAT, which again correlates with the order of specificity of Rad4 for these mismatches.

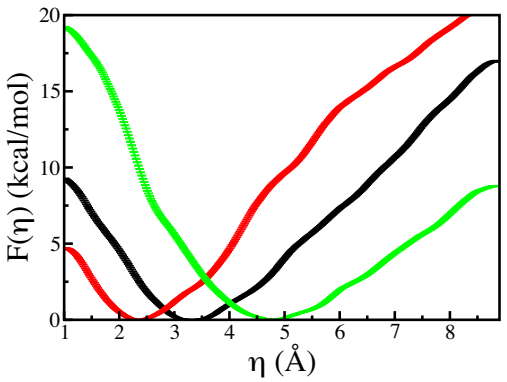

Figure 4. Free energy profiles of η for TTT/TTT (black), TAT/TAT (red) and CCC/CCC (green).

β-Hairpin Insertion

Free energy profiles. We now examine the sequence specificity of β-hairpin insertion and its correlation with mismatch-induced DNA deformation. In particular, we are concerned with whether the mismatch-induced deformation facilitates or impedes the β-hairpin insertion. To this end, we have performed umbrella sampling simulations using the distance (η) between the leading edge of the BHD3 β-hairpin and the lesion site of DNA as a suitable collective variable (Figure 2C).

As seen in Figure 4, F(η) exhibits a single asymmetric parabolic well around the energy minimum for all three sequences. However, the location of the minimum (denoted by η_{min}), which measures the most-probable distance of closest approach of the β-hairpin to the lesion site, varies significantly with the sequence (Table 3). The observed sequence-dependent variation in η_{min} suggests that the β-hairpin makes a much closer approach to the mismatch in TAT/TAT and TTT/TTT, whereas the depth of insertion of the β-hairpin is relatively less for CCC/CCC.

To quantify the ease with which the β-hairpin can be inserted into the damaged region of DNA, we used the energy difference $\Delta F_{\eta} = |F(\eta_{min} + \Delta\eta) - F(\eta_{min})|$ as a measure of the energy cost of the β-hairpin insertion into the DNA. Here, η_{min} is the location of the energy minimum

Table 2. Widths (°) of the free energy basins for Rad4-bound and Rad4-free complexes. The superscripts *b* and *ub* denote biased and unbiased runs respectively, while the subscripts REC, BDI denote Rad4-complexed and Rad4-free systems, respectively

sequence	Φ^{b}_{REC}	Φ^{ub}_{REC}	Ψ^{b}_{REC}	Ψ^{ub}_{REC}	Φ^{b}_{BDI}	Φ^{ub}_{BDI}	Ψ^{b}_{BDI}	Ψ^{ub}_{BDI}
TTT/TTT	66	76	55	88	38	37	28	39
TAT/TAT	72	60	38	71	27	31	42	41
CCC/CCC	163	159	57	48	35	25	64	92

Table 3. Location of minimum (η_{min}) on free energy profiles for η and the cost of β-hairpin insertion ΔF_{η}

sequence	η_{min} (Å)	ΔF_{η} (kcal/mol)
CCC/CCC	4.81 ± 0.31	6.28 ± 0.25
TTT/TTT	3.31 ± 0.17	8.17 ± 0.70
TAT/TAT	2.33 ± 0.21	10.49 ± 0.44

and Δη is the displacement on the high η side of the minimum. With Δη set to 3 Å, ΔF_{η} measures the energy expended by the β-hairpin to reach the minimum from a distance of 3 Å away from the minimum. The calculated β-hairpin insertion energy is the least for CCC/CCC, followed by TTT/TTT and TAT/TAT. For $\eta < \eta_{min}$, the steric clashes between DNA and Rad4 cause $F(\eta)$ to rise, whereas the disruption of favourable interactions between DNA and Rad4 lead to increase in free energy for $\eta > \eta_{min}$. For CCC/CCC, the β-hairpin is able to flip out the bases much earlier in the course of its approach, which is indicative of both the ease of insertion as well as the extent of destabilization around the lesion site. The distance at which the β-hairpin evicts the mismatch, which we take as a measure of the ease of insertion, follows the same order across all three sequences as their corresponding flipping energies for both bases (refer Figure 4).

The free energy profiles of η calculated from the unbiased MD simulations are reported (Supplementary Figures S20–S22) in the SI. The structure of the Rad-DNA complex corresponding to high-energy regions of $F(\eta)$ ($\eta > 8.5$ Å) for the TTT/TTT mismatch is shown in Supplementary Figure S11b.

Angular fluctuation of β-hairpin. To understand the origin of the observed sequence-dependent variation in the ease of β-hairpin insertion, it is important to examine the correlation between the depth of β-hairpin insertion into the DNA duplex, and the flexibility of the β-hairpin in the Rad4–DNA complex and to investigate how these attributes are influenced by the structural distortions in DNA. In particular, we seek to understand whether DNA with a higher degree of distortion facilitates deeper penetration of the β-hairpin into the duplex and to examine whether a deeply inserted β-hairpin exhibits lower flexibility than a loosely inserted β-hairpin. It is anticipated that a deeply inserted β-hairpin is likely to be relatively less flexible than a loosely inserted β-hairpin.

We define different measures to quantify the DNA distortion and the flexibility of the β-hairpin separately. To determine the flexibility of the β-hairpin, we examine the time evolution of the angles (θ and ω as shown in Figure 5) of approach of the β-hairpin to DNA. To calculate these angles, each frame of the MD trajectory was first aligned against

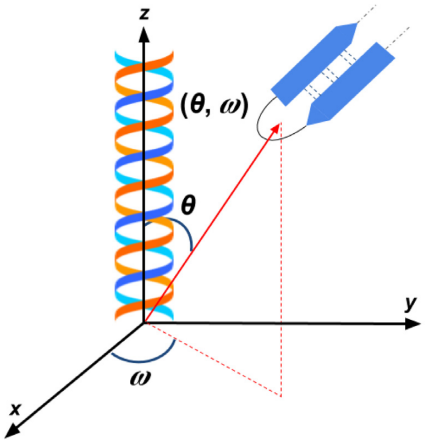

Figure 5. θ and ω to determine angle of approach of β-hairpin to the DNA.

the first frame such that the helical axis (the principal axis of least moment of inertia) of DNA always coincides with the z-axis. The angle between the long axis (the principal axis of least moment of inertia) of the β-hairpin and helical axis of DNA (i.e. the z-axis) is denoted by θ (range of θ is 0 to π), while the azimuth of the long axis of the β-hairpin is denoted by ω (the range of ω is from $-\pi$ to π).

The mean and standard deviation of θ and ω were calculated from their time series obtained from the unbiased MD trajectories. Their mean values ($\langle\theta\rangle$ and $\langle\omega\rangle$) correspond to the most-probable angles of approach of the β-hairpin, while the standard deviations ($\langle\theta^2\rangle-\langle\theta\rangle^2$ and $\langle\omega^2\rangle-\langle\omega\rangle^2$) quantify the spread in these angles. For a successful insertion, the β-hairpin must achieve the right orientation with respect to the helical axis of the DNA. For DNA sequences with lower ease of insertion, the β-hairpin is likely to approach the DNA from various orientations before it achieves the right orientation for a successful insertion. The many failed insertion attempts by the β-hairpin can lead to higher spread in the angles of approach. The standard deviations of θ and ω are expected to be low for a DNA that readily allows insertion of the β-hairpin.

The calculated mean values of θ reported in Table 4 are close to 90° suggesting that the β-hairpin approaches the damaged pocket in DNA almost perpendicularly in all the systems. However, noticeable differences in the fluctuations of θ and ω are observed across different DNA sequences. As seen from Table 4, the β-hairpin tends to explore a much larger range of θ (with a higher standard deviation) for TAT/TAT, which indicates that the β-hairpin finds it much harder to attempt a successful insertion into the damaged region of TAT/TAT; it also hints at the low specificity of Rad4 for this particular sequence. The CCC/CCC mis-

Table 4. The mean and standard deviation of angular spread (°) of β-hairpin's approach to the lesion site calculated from unbiased MD simulations

sequence	⟨ω⟩	⟨θ⟩
CCC/CCC	−185.43 ± 3.84	76.836 ± 3.56
TTT/TTT	−168.90 ± 5.13	77.756 ± 4.81
TAT/TAT	−153.46 ± 4.82	96.79 ± 6.00

Table 5. Local mean RMSD (Å) of the damaged region in the presence and absence of Rad4

Sequence	BDI	REC
CCC/CCC	0.23 ± 0.04	0.71 ± 0.03
TTT/TTT	0.25 ± 0.04	0.77 ± 0.03
TAT/TAT	0.30 ± 0.04	0.71 ± 0.03

match, which is one of the most destabilizing mismatches, potentially allows for an easier insertion by the β-hairpin. In line with Ansari's work and expanded upon by our computational studies, CCC/CCC allows for the easiest insertion of the β-hairpin, reaffirming the high specificity Rad4 for this particular sequence, while TTT/TTT demonstrates intermediate specificity (31). Along the azimuth as well, the β-hairpin is able to insert itself into the CCC/CCC mismatch by sampling fewer orientations as compared to the other two mismatches. This order of specificity is in agreement with, and further strengthens, our assertions from the study of the base flipping reaction co-ordinates, where the relative steepness and height of the energy barriers to flip out the bases (uX16, uY17) indicated that Rad4 had high specificity for CCC/CCC, intermediate for TTT/TTT and the least for TAT/TAT.

Structural distortion in DNA. The mismatch-induced local thermodynamic destabilization gives rise to structural distortions in DNA in the vicinity of the mismatch site. The higher the degree of thermodynamic destabilization, the higher is the deformation of DNA from its native B-DNA structure (61–63). The binding of Rad4 also additionally contributes to the overall distortion of DNA. It is anticipated that the ease of insertion of the β-hairpin into the damaged region of DNA is likely to be influenced by the degree of distortion of DNA. It is expected that the greater the DNA distortion, the less the steric hindrance for the β-hairpin to insert into the damaged pocket of DNA. To quantify the distortion of DNA, we calculated the local root mean square deviation (RMSD) of the heavy atoms of DNA (belonging to triple mismatched base pairs and the three base pairs on either side of the triple mismatch) that surround the damage/mismatch pocket in Rad4-DNA complex from the corresponding ideal double helical DNA. The calculated RMSD obtained from unbiased MD trajectories are reported in Table 5.

The local RMSD values are similar across all three systems, suggesting that RMSD may not be a useful measure to probe the effect of local distortion on the sequence specificity of Rad4-binding.

Correlation between β-hairpin insertion and base flipping. Thus far, we have examined base flipping and β-hairpin insertion separately and quantified the associated dynamics and energetics individually. Hence, the preceding results did not address how these molecular events are coupled with each other. It is non-trivial to elucidate the mechanism by which these events are coupled using direct MD simulation of Rad4–DNA complex for the following reasons: First, these slow rare events are very unlikely to be captured in a few hundred nanoseconds of MD simulation. Second, the crystal structure of the Rad4–DNA complex used in our simulation represents only the final state of the Rad4–DNA binding process, in which the base flipping and BHD3 β-hairpin insertion had already taken place.

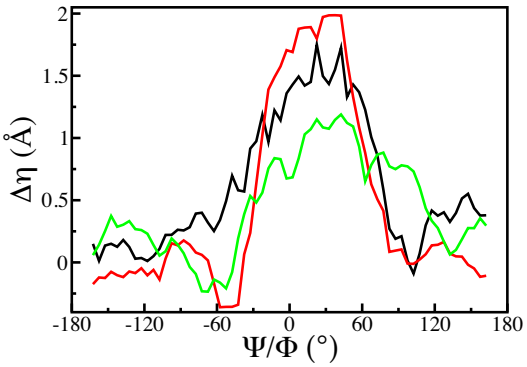

Figure 6. Variation of $\Delta\eta = \eta - \eta_{min}$ with the pseudo-dihedral angles (Φ/Ψ) associated with base flipping for TTT/TTT (black), TAT/TAT (red) and CCC/CCC (green).

A meaningful analysis of the correlation between β-hairpin insertion and base flipping involves forcefully (by applying a biasing force) changing the conformations of bases of interest such that the entire ranges of Φ and Ψ are sampled uniformly and then to examine how the β-hairpin responds to the changes in Φ and Ψ. We carried out such an analysis by changing Φ and Ψ in a correlated fashion such that $\Phi = \Psi$ at all stages. A similar analysis was performed separately on trajectories obtained from base flipping umbrella sampling simulations in which Φ and Ψ were individually biased (Supplementary Figures S8 and S9). The variation of $\Delta\eta = \eta - \eta_{min}$ with Φ (or Ψ) (Figure 6) shows that for each system Δη fluctuates around zero for $\Phi < -60°$ or $\Phi > 120°$; Δη is almost insensitive to changes in Φ (or Ψ) in this conformational region. On the contrary, η deviates significantly from η_{min} for $-50° < \Phi < 100°$; in this region, there appears to be a broad peak whose width and height vary with DNA sequence. The width of the peak is highest for CCC/CCC, followed by TTT/TTT and TAT/TAT, whereas the opposite trend is observed for the intensity of the peak (the peak intensity is highest for TAT/TAT and lowest for CCC/CCC).

Given that Δη quantifies the decrease in the depth of the β-hairpin insertion, the presence of the peak of Δη in the intra-helical region ($|\Phi| \leq 40°$) implies that (a) the β-hairpin is pushed away from the binding pocket as the bases move towards the intra-helical state and (b) the β-hairpin rebinds to the binding pocket as the bases become extra-helical again. The intensity of the peak of Δη appears to correlate with the degree of coupling between the β-hairpin insertion and base flipping. When base flipping and β-hairpin

insertion are uncoupled, η would be independent of changes in Φ and Ψ, and hence the peak intensity is expected to be zero. Higher the intensity, the farther the β-hairpin is pushed away from the damaged site as the bases become intrahelical, implying a higher degree of coupling between the two processes, and possibly hinting at the lower affinity of the β-hairpin for the damaged site. CCC/CCC has the lowest peak intensity, implying that the β-hairpin prefers to remain close to its minimum energy conformation even when the bases are intrahelical, thus hinting at the higher affinity of the β-hairpin to the damaged site.

The width of the peak is representative of the number of conformations in which the β-hairpin can interact with the damaged bases. CCC/CCC has the widest peak for $\Delta\eta$, indicating that there are many more conformations in which the damaged bases are able to interact with the β-hairpin. In general, it is seen that the higher the peak for $\Delta\eta$, the narrower it is. The relatively high η_{min} for CCC/CCC suggests that the β-hairpin faces less steric hindrance from uX16 and uY17, allowing these bases to flip back inside without displacing the hairpin too much from its equilibrium position. Similarly, because TAT/TAT makes a deeper approach into the mismatch, the bases flipping into the intrahelical region end up displacing it much more from its η_{min}.

Contrarily, the results of the variation of Φ and Ψ as a function of η (Supplementary Figure S7 in SI) obtained from the umbrella sampling simulation of the β-hairpin insertion (with bias applied only on η) did not show the coupling between β-hairpin insertion and base flipping. This is understandable in that the β-hairpin insertion was investigated using the experimental crystal structure of the Rad4-DNA complex in which uX16 and uY17 bases were already flipped out; that is, our umbrella sampling simulation examined the energetics of the β-hairpin insertion into an already evicted lesion site. Since uX16 and uY17 remained in their stable extra-helical conformational states during the entire course of these umbrella sampling simulations, the rare event of them flipping back to their intra-helical states when the β-hairpin detaches itself from the lesion site did not occur in these simulations.

Rad4–DNA association

Free energy profiles. Figure 7 and Supplementary Figures S20–S22 show the association free energy profiles $F(\xi)$ obtained from umbrella sampling simulations. The structure of the Rad–DNA complex corresponding to high-energy regions of $F(\xi)$ ($\xi > 20.5$ Å) for the TTT/TTT mismatch is shown in Supplementary Figure S11A. For all three systems studied, $F(\xi)$ exhibits an asymmetric parabolic well around the global energy minimum at ξ_{min}, followed by a plateau and a linear increase at higher ξ. The three regimes of interest in the study of Rad4-DNA association are: (i) the low-ξ regime (regime-I) in which Rad4 makes the closest approach to DNA, (ii) the intermediate-ξ regime (regime-II) in which Rad4 experiences interactions with both water and DNA. (iii) the high-ξ regime (regime-III) in which Rad4 is distantly separated from DNA and it interacts primarily with solvent molecules and only weakly with the DNA.

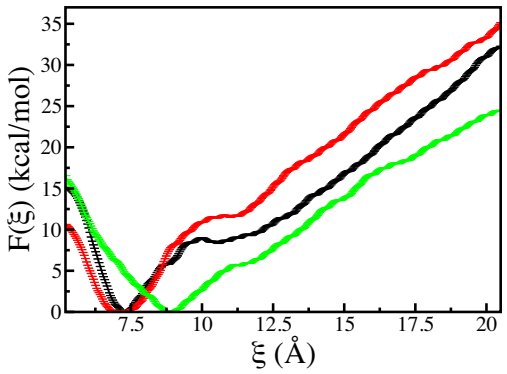

Figure 7. Free energy profiles of ξ for TTT/TTT (black), TAT/TAT (red) and CCC/CCC (green).

The interplay between Rad4–water, DNA–water and Rad4–DNA interactions and the ξ-dependent variation in their relative contributions to the overall stability of the complex determine the nature of F(ξ). Owing to close proximity of Rad4 and DNA, and their favourable interactions, the global minimum in F(ξ) is expected to lie in regime-I. The width of the energy well around the global minimum quantifies the "tightness" of Rad4–DNA association.

We examined the variation in the number of hydrogen bonds between Rad4 and DNA (Supplementary Figure S6) and the Rad4-DNA interaction energy as a function of ξ (Supplementary Figure S7). There is a general decrease in the number of hydrogen bonds and a weakening of interaction energies as the BHD2/3 domains move away from the lesion/mismatch site. The slope change observed in regime-II is a consequence of this.

In regime-III, when ξ is high, the DNA and Rad4 are not associated and are mainly surrounded by water molecules. Therefore, the Rad4–water and DNA–water interactions dominate over the Rad4-DNA interactions in this regime. Since the DNA and Rad4 are homogeneously surrounded by water, $F(\xi)$ is expected to be more or less insensitive to the variation in ξ in this regime. Although complete DNA–Rad4 dissociation would occur for $\xi > 25$ Å, the present study investigates the nature of F(ξ) only for $\xi < 20$ Å.

The location (ξ_{min}) of the global minimum is different for different sequences studied; in particular, ξ_{min} for CCC/CCC is significantly different from that for the TTT/TTT and TAT/TAT mismatches. CCC/CCC docks with the respective domains on Rad4 at a farther distance than TTT/TTT and TAT/TAT, hinting at the high specificity the protein has for it. This corresponds well with the results from the study on β-hairpin insertion, and the farther apart that Rad4 BHD domains can dock on the bases, the more space the β-hairpin has to insert itself into the lesion site and flip the bases out.

We propose that location of the energy minimum might possibly be affected by the presence of favorable interactions between Rad4 and the active site. High specificity of the CCC/CCC mismatch might signal that Rad4 experiences these favourable interactions much earlier than in the case of TAT/TAT mismatch or TTT/TTT mismatch.

We have also calculated the mean interaction energy between Rad4 and uX16/uY17 bases using the unbiased MD trajectories, and the results are reported in Supplementary Table S2. In addition, the pairwise interaction energies of uX16 and uY17 with key residues of Rad4 located in and around the lesion/mismatch site were also calculated, and the results are reported in Supplementary Table S3 in the SI.

CONCLUSION

The UV radiation-induced DNA damage is recognized and repaired by Rad4 (in yeast) and XPC (in humans) proteins. The detailed molecular mechanism of how Rad4/XPC repairs such DNA damage with high fidelity remains elusive. The present study investigates the molecular origin of sequence-dependent specificity of Rad4 for mismatched DNA using molecular dynamics and umbrella sampling simulations of three models of DNA with identical sequences except for a stretch of three contiguous mismatched base pairs.

The main objective is to understand how the conformational flexibility and flipping of the mismatched bases, the dynamics and depth of insertion of a functionally-relevant β-hairpin of Rad4 into DNA duplex, and the Rad4-DNA association energetics correlate with the specificity of Rad4 for DNA mismatches. The free energy profiles for each of these three key molecular events (Rad4-DNA association, β-hairpin insertion, and base flipping) were calculated using suitable reaction coordinates. The calculated free energy profiles for base flipping revealed that the mismatched bases prefer intra-helical conformation in the absence of Rad4; the activation barrier separating the intra- and extra-helical conformational states was at least a factor of 3 less for CCC/CCC than that for TTT/TTT and TAT/TAT. Consequently, CCC/CCC exhibited greater conformational heterogeneity and shorter mean base flipping times than that for the other systems. The observed conformational heterogeneity allows DNA to sample different isoenergetic conformers and thereby increases the populations of conformers that favor mismatch/damage recognition by Rad4. Upon binding of Rad4 to DNA, the mismatched bases underwent conformational shift to the extra-helical conformation, where they were stabilized by binding pocket residues of Rad4. The activation barriers separating the intra- and extra-helical conformational states of mismatched bases were altered by the binding of Rad4 and the binding-induced changes in the base flipping energetics of mismatched bases were significant for CCC/CCC. In addition, the mismatch-induced and Rad4-induced changes in the flexibility of mismatched bases correlate with the extent of the specificity with which Rad4 recognizes and binds to DNA.

A detailed examination of the energetics and dynamics of β-hairpin insertion into the DNA duplex revealed that the distance of closest approach of the β-hairpin to the lesion site varies significantly with the DNA sequence. The β-hairpin was able to make a much closer approach to the mismatch/lesion site in TAT/TAT and TTT/TTT, whereas its depth of insertion was relatively less for CCC/CCC. Moreover, the estimated free energy cost of β-hairpin insertion was relatively less for CCC/CCC than those for the other systems. This could be due to the higher degree of mismatch-induced distortion of DNA in CCC/CCC that facilitates favorable β-hairpin insertion.

Although the β-hairpin approaches the DNA almost perpendicularly in all the systems, a large spread in the angles of approach was observed for TAT/TAT, which suggested that the β-hairpin was relatively more flexible and it finds it hard to attain the right orientation to insert into the DNA duplex. The β-hairpin sampled relatively fewer orientations for successful insertion into the CCC/CCC mismatch as compared to the other two mismatches.

The distance between the site of mismatch and the BHD2/3 domains was relatively large for CCC/CCC than for the other systems. As a result, the β-hairpin could recognize and interact favourably with nucleotides in and around the CCC/CCC mismatch site much earlier than in the case of TAT/TAT or TTT/TTT mismatch. Consequently, it becomes easier for the β-hairpin to insert itself into the CCC/CCC mismatch and flip out the mismatched bases from the DNA duplex. The molecular origin of the order and extent of specificity of Rad4 for DNA mismatches explored in the present study complements recent experimental FRET studies on Rad4–DNA systems (31).

However, caution must be exercised in interpreting the reported free energy profiles for β-hairpin insertion as they quantify only the energetics of the β-hairpin insertion into an already evicted lesion site. Anjum Ansari *et al.* commented about overall Rad4 specificity—where the flipping out of bases uX16 and uY17 likely happens in concert with β-hairpin insertion post a successful association of the Rad4 BHD2/BHD3 domains with the DNA mismatch region (31). The base flipping free energy profiles obtained in this study (Table 1) place the order of Rad4 specificity as CCC/CCC > TTT/TTT > TAT/TAT and are in strong agreement with FRET data (31). However, the free energy profiles of η depicting the β-hairpin insertion under conditions of an already evicted lesion site, seem to suggest the opposite trend. This can be reconciled with our understanding of Rad4 specificity from previous studies by looking at the base flipping and hairpin insertion as uncoupled processes, with the base flipping event being more significant than the insertion event in determining overall specificity of Rad4 recognition to mismatched DNA sequences.

Examination of the interplay between the base flipping and β-hairpin insertion processes reveals a strong correlation between the decrease in the depth of the β-hairpin insertion into the binding pocket and the degree of coupling between the insertion of hairpin and the flipping of bases; the higher degree of coupling between the β-hairpin insertion and the base flipping, the lower the affinity of the hairpin to the binding site. This positive correlation between the insertion of the BHD2 β-hairpin and the mismatch binding specificity has been indicated in recent experimental studies on the Rad4/XPC NER complex as well (29,30). The calculated free energy profiles also indicated that the β-hairpin faces less steric hindrance from the CCC/CCC mismatch bases, allowing these bases to flip back inside without displacing the hairpin too much.

Visual inspection of the crystal structure and the unbiased MD trajectory shows the other β-hairpin of the BHD2

domain of Rad4 can also interact with the damaged bases and contribute to the variation in the flexibility between uX16 and uY17. Detailed analysis of the role of β-hairpin of BHD2 will be examined in future studies.

The free energy profiles for the association of BHD2/BHD3 domains with the mismatch site report on the interplay between Rad4–water, DNA–water and Rad4–DNA interactions in determining the overall stability of the Rad4–DNA complex. Our results showed that during association Rad4 experiences favourable interactions with the CCC/CCC mismatched bases much earlier than in the case of TAT/TAT mismatch or TTT/TTT mismatch.

In summary, the key determinants and ramifications of the specificity with which Rad4 recognizes and binds to DNA are explored using molecular dynamics and umbrella sampling simulations. The dynamic and energetic characterization of the order and extent of specificity of Rad4 for 3-bp mismatched sequences demonstrated that Rad4 is highly specific to CCC/CCC, while it recognises TTT/TTT mismatch with intermediate specificity and only poorly recognises TAT/TAT mismatch. While the present computational study provided molecular-level insights into mismatch/lesion recognition by Rad4, it is conceded that the scope of our study was largely restricted to the characterization of base flipping, β-hairpin insertion and Rad4-DNA association independently and understanding how the variation in DNA sequence affects these processes individually. However, the actual lesion recognition and Rad4–DNA binding processes are likely to be governed by a complex interplay of various coupled motions and molecular events beyond base flipping, β-hairpin insertion and Rad4–DNA association. Given the lack of an exhaustive list of all collective variables that are likely to influence these processes and the prohibitively high computational cost involved, it is practically infeasible to capture all these rare events using long unbiased MD simulations or multidimensional umbrella sampling method. Our results not only underscored the importance of the aforementioned three molecular events in DNA mismatch/damage recognition, but also hints at the need for better collective variables and multi-dimensional sampling methods to capture the coupling between these events realistically. These issues will be investigated further in our future research. In addition, a comparative study of the binding of Rad4 to DNA with *bona fide* lesions (including CPD and 6-4PP) and other mismatches to elucidate the similarities and distinctions between the lesion- and mismatch-recognition mechanisms by Rad4 will be carried out in future.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We wish to thank Dr A. Semparithi for computational assistance and support. The high-performance computing facility provided by IIIT-Hyderabad is gratefully acknowledged.

FUNDING

CSIR, Govt. of India, for a senior research fellowship (to K.P.).
Conflict of interest statement. None declared.

REFERENCES

1. Huen,M.S.Y., Sy,S.M.H. and Chen,J. (2009) BRCA1 and its toolbox for the maintenance of genome integrity. *Nat. Rev. Mol. Cell Biol.*, **11**, 138–148.
2. Shiloh,Y. (2003) ATM and related protein kinases: safeguarding genome integrity. *Nat. Rev. Cancer*, **3**, 155–168.
3. Cimprich,K.A. and Cortez,D. (2008) ATR: An essential regulator of genome integrity. *Nat. Rev. Mol. Cell Biol.*, **9**, 616–627.
4. Cleaver,J.E., Thompson,L.H., Richardson,A.S. and States,J.C. (1999) A summary of mutations in the UV-sensitive disorders: Xeroderma pigmentosum, Cockayne syndrome, and trichothiodystrophy. *Hum. Mutat.*, **14**, 9.
5. Lukas,J., Lukas,C. and Bartek,J. (2011) More than just a focus: the chromatin response to DNA damage and its role in genome integrity maintenance. *Nat. Cell Biol.*, **13**, 1161–1169.
6. Geacintov,N.E. and Broyde,S. (2010) Introduction and perspectives on the chemistry and biology of DNA damage. In: The Chemical Biology of DNA Damage, Wiley. pp. 3–20.
7. Verdine,G.L. and Bruner,S.D. (1997) How do DNA repair proteins locate damaged bases in the genome? *Chem. Biol.*, **4**, 329–334.
8. Ansari,A. and Kuznetsov,S.V. (2010) Dynamics and mechanism of DNA-bending proteins in binding site recognition. In: Biophysics of DNA-Protein Interactions. Springer, NY. pp. 107–142.
9. Sarangi,M.K., Zvoda,V., Holte,M.N., Becker,N.A., Peters,J.P., Maher,L.J. and Ansari,A. (2019) Evidence for a bind-then-bend mechanism for architectural DNA binding protein yNhp6A. *Nucleic Acids Res.*, **47**, 2871–2883.
10. Velmurugu,Y., Vivas,P., Connolly,M., Kuznetsov,S.V., Rice,P.A. and Ansari,A. (2017) Two-step interrogation then recognition of DNA binding site by Integration Host Factor: An architectural DNA-bending protein. *Nucleic Acids Res.*, **46**, 1741–1755.
11. Ciccia,A. and Elledge,S.J. (2010) The DNA damage response: making it safe to play with knives. *Mol. Cell*, **40**, 179–204.
12. Polo,S.E. and Jackson,S.P. (2011) Dynamics of DNA damage response proteins at DNA breaks: a focus on protein modifications. *Genes Dev.*, **25**, 409–433.
13. Jia,L., Kropachev,K., Ding,S., Van Houten,B., Geacintov,N.E. and Broyde,S. (2009) Exploring damage recognition models in prokaryotic nucleotide excision repair with a Benzo[a]pyrene-Derived lesion in UvrB. *Biochemistry*, **48**, 8948–8957.
14. DiGiovanna,J.J. and Kraemer,K.H. (2012) Shining a light on Xeroderma pigmentosum. *J. Invest. Dermatol.*, **132**, 785–796.
15. Kraemer,K.H. (1997) Sunlight and skin cancer: another link revealed. *Proc. Natl. Acad. Sci. U.S.A.*, **94**, 11–14.
16. Chakraverty,R.K. and Hickson,I.D. (1999) Defending genome integrity during DNA replication: a proposed role for RecQ family helicases. *Bioessays*, **21**, 286–294.
17. Lee,Y.-C., Cai,Y., Mu,H., Broyde,S., Amin,S., Chen,X., Min,J.-H. and Geacintov,N.E. (2014) The relationships between XPC binding to conformationally diverse DNA adducts and their excision by the human NER system: Is there a correlation? *DNA Repair*, **19**, 55–63.
18. Mu,H., Geacintov,N.E., Broyde,S., Yeo,J.-E. and Schärer,O.D. (2018) Molecular basis for damage recognition and verification by XPC-RAD23B and TFIIH in nucleotide excision repair. *DNA Repair*, **71**, 33–42.
19. Huang,J.C., Hsu,D.S., Kazantsev,A. and Sancar,A. (1994) Substrate spectrum of human excinuclease: repair of abasic sites, methylated bases, mismatches, and bulky adducts. *Proc. Natl. Acad. Sci. U.S.A.*, **91**, 12213–12217.
20. Zhang,E.T., He,Y., Grob,P., Fong,Y.W., Nogales,E. and Tjian,R. (2015) Architecture of the human XPC DNA repair and stem cell coactivator complex. *Proc. Natl. Acad. Sci. U.S.A.*, **112**, 14817–14822.
21. Min,J.-H. and Pavletich,N.P. (2007) Recognition of DNA damage by the Rad4 nucleotide excision repair protein. *Nature*, **449**, 570–575.
22. Chen,X., Velmurugu,Y., Zheng,G., Park,B., Shim,Y., Kim,Y., Liu,L., Van Houten,B., He,C., Ansari,A. and Min,J.-H. (2015) Kinetic gating

mechanism of DNA damage recognition by Rad4/XPC. *Nat. Commun.*, **6**, 5849.
23. Liu,Y., Reeves,D., Kropachev,K., Cai,Y., Ding,S., Kolbanovskiy,M., Kolbanovskiy,A., Bolton,J.L., Broyde,S., Van Houten,B. and Geacintov,N.E. (2011) Probing for DNA damage with β-hairpins: Similarities in incision efficiencies of bulky DNA adducts by prokaryotic and human nucleotide excision repair systems in vitro. *DNA Repair*, **10**, 684–696.
24. Mu,H., Geacintov,N.E., Zhang,Y. and Broyde,S. (2015) Recognition of damaged DNA for nucleotide excision repair: a correlated motion mechanism with a mismatched cis-syn thymine dimer lesion. *Biochemistry*, **54**, 5263–5267.
25. Mu,H., Geacintov,N.E., Min,J.-H., Zhang,Y. and Broyde,S. (2017) Nucleotide excision repair lesion-recognition protein Rad4 captures a pre-flipped partner base in a benzo [a] pyrene-derived DNA lesion: How structure impacts the binding pathway. *Chem. Res. Toxicol.*, **30**, 1344–1354.
26. Boehr,D.D., Nussinov,R. and Wright,P.E. (2009) The role of dynamic conformational ensembles in biomolecular recognition. *Nat. Chem. Biol.*, **5**, 789–796.
27. Cao,C., Jiang,Y.L., Stivers,J.T. and Song,F. (2004) Dynamic opening of DNA during the enzymatic search for a damaged base. *Nat. Struct. Mol. Biol.*, **11**, 1230–1236.
28. Velmurugu,Y., Chen,X., Sevilla,P.S., Min,J.-H. and Ansari,A. (2016) Twist-open mechanism of DNA damage recognition by the Rad4/XPC nucleotide excision repair complex. *Proc. Natl. Acad. Sci. U.S.A.*, **113**, E2296–E2305.
29. Paul,D., Mu,H., Zhao,H., Ouerfelli,O., Jeffrey,P.D., Broyde,S. and Min,J.-H. (2019) Structure and mechanism of pyrimidinepyrimidone (6-4) photoproduct recognition by the Rad4/XPC nucleotide excision repair complex. *Nucleic Acids Res.*, **47**, 6015–6028.
30. Mu,H., Zhang,Y., Geacintov,N.E. and Broyde,S. (2018) Lesion sensing during initial binding by Yeast XPC/Rad4: Toward predicting resistance to nucleotide excision repair. *Chem. Res. Toxicol.*, **31**, 1260–1268.
31. Chakraborty,S., Steinbach,P.J., Paul,D., Mu,H., Broyde,S., Min,J.-H. and Ansari,A. (2017) Enhanced spontaneous DNA twisting/bending fluctuations unveiled by fluorescence lifetime distributions promote mismatch recognition by the Rad4 nucleotide excision repair complex. *Nucleic Acids Res.*, **46**, 1240–1255.
32. Torrie,G.M. and Valleau,J.P. (1977) Nonphysical sampling distributions in Monte Carlo free-energy estimation: umbrella sampling. *J. Comput. Phys.*, **23**, 187–199.
33. Humphrey,W., Dalke,A. and Schulten,K. (1996) VMD: visual molecular dynamics. *J. Mol. Graphics*, **14**, 33–38.
34. Šali,A. and Blundell,T.L. (1993) Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.*, **234**, 779–815.
35. Jorgensen,W.L., Chandrasekhar,J., Madura,J.D., Impey,R.W. and Klein,M.L. (1983) Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.*, **79**, 926–935.
36. Price,D.J. and Brooks,C.L. (2004) A modified TIP3P water potential for simulation with Ewald summation. *J. Chem. Phys.*, **121**, 10096–10103.
37. Pettersen,E.F., Goddard,T.D., Huang,C.C., Couch,G.S., Greenblatt,D.M., Meng,E.C. and Ferrin,T.E. (2004) UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.*, **25**, 1605–1612.
38. Macke,T.J. and Case,D.A. (1997) Modeling unusual nucleic acid structures. In: Molecular Modeling of Nucleic Acid. American Chemical Society, Washington, DC. pp. 379–393.
39. Berendsen,H.J.C., van der Spoel,D. and van Drunen,R. (1995) GROMACS: a message-passing parallel molecular dynamics implementation. *Comput. Phys. Commun.*, **91**, 43–56.
40. Van Der Spoel,D., Lindahl,E., Hess,B., Groenhof,G., Mark,A.E. and Berendsen,H.J.C. (2005) GROMACS: fast, flexible, and free. *J. Comput. Chem.*, **26**, 1701–1718.
41. Ivani,I., Dans,P.D., Noy,A., Pérez,A., Faustino,I., Hospital,A., Walther,J., Andrio,P., Goñi,R., Balaceanu,A. *et al.* (2015) Parmbsc1: a refined force field for DNA simulations. *Nat. Methods*, **13**, 55–58.
42. Bonomi,M., Branduardi,D., Bussi,G., Camilloni,C., Provasi,D., Raiteri,P., Donadio,D., Marinelli,F., Pietrucci,F., Broglia,R.A. and Parrinello,M. (2009) PLUMED: A portable plugin for free-energy calculations with molecular dynamics. *Comput. Phys. Commun.*, **180**, 1961–1972.
43. Tribello,G.A., Bonomi,M., Branduardi,D., Camilloni,C. and Bussi,G. (2014) PLUMED 2: new feathers for an old bird. *Comput. Phys. Commun.*, **185**, 604–613.
44. Hess,B., Bekker,H., Berendsen,H.J.C. and Fraaije,J.G.E.M. (1997) LINCS: a linear constraint solver for molecular simulations. *J. Comput. Chem.*, **18**, 1463–1472.
45. Darden,T., York,D. and Pedersen,L. (1993) Particle mesh Ewald: an N log(N) method for Ewald sums in large systems. *J. Chem. Phys.*, **98**, 10089–10092.
46. Essmann,U., Perera,L., Berkowitz,M.L., Darden,T., Lee,H. and Pedersen,L.G. (1995) A smooth particle mesh Ewald method. *J. Chem. Phys.*, **103**, 8577–8593.
47. Parrinello,M. and Rahman,A. (1981) Polymorphic transitions in single crystals: a new molecular dynamics method. *J. Appl. Phys.*, **52**, 7182–7190.
48. Marquardt,D.W. (1963) An algorithm for least-squares estimation of nonlinear parameters. *J. Soc. Ind. Appl. Math.*, **11**, 431–441.
49. Hestenes,M. and Stiefel,E. (1952) Methods of conjugate gradients for solving linear systems. *J. Res. Nat. Bur. Stand.*, **49**, 409–436.
50. Kumar,S., Rosenberg,J.M., Bouzida,D., Swendsen,R.H. and Kollman,P.A. (1992) The weighted histogram analysis method for free-energy calculations on biomolecules I. The method. *J. Comput. Chem.*, **13**, 1011–1021.
51. Pitta,K. and Krishnan,M. (2018) Molecular Mechanism, Dynamics, and Energetics of Protein-Mediated Dinucleotide Flipping in a Mismatched DNA: A Computational Study of the RAD4-DNA Complex. *J. Chem. Inf. Model.*, **58**, 647–660.
52. Banavali,N.K. and MacKerell,A.D. (2002) Free energy and structural pathways of base flipping in a DNA GCGC containing sequence. *J. Mol. Biol.*, **319**, 141–160.
53. Tikhomirova,A., Beletskaya,I.V. and Chalikian,T.V. (2006) Stability of DNA duplexes containing GG, CC, AA, and TT mismatches. *Biochemistry*, **45**, 10563–10571.
54. Peyret,N., Seneviratne,P.A., Allawi,H.T. and SantaLucia,J. (1999) Nearest-Neighbor thermodynamics and NMR of DNA sequences with internal A·A, C·C, G·G, and T·T Mismatches. *Biochemistry*, **38**, 3468–3477.
55. Stephenson,C. and Karran,P. (1989) Selective binding to DNA base pair mismatches by proteins from human cells. *J. Biol. Chem.*, **264**, 21177–21182.
56. Ke,S.-H. and Wartell,R.M. (1993) Influence of nearest neighbor sequence on the stability of base pair mismatches in long DNA: determination by temperature-gradient gel electrophoresis. *Nucleic Acids Res.*, **21**, 5137–5143.
57. Aboul-ela,F., Koh,D., Tinoco,I. and Martin,F.H. (1985) Base-base mismatches. Thermodynamics of double helix formation for dCA3XA3G + dCT3YT3G (X, Y = A,C,G,D). *Nucleic Acids Res.*, **13**, 4811–4824.
58. SantaLucia,J. and Hicks,D. (2004) The thermodynamics of DNA structural motifs. *Annu. Rev. Biophys. Biomol. Struct.*, **33**, 415–440.
59. Jackson,B.A. and Barton,J.K. (2000) Recognition of base mismatches in DNA by 5,6-chrysenequinone diimine complexes of rhodium(III): a proposed mechanism for preferential binding in destabilized regions of the double Helix. *Biochemistry*, **39**, 6176–6182.
60. Frederico,L.A., Kunkel,T.A. and Shaw,B.R. (1993) Cytosine deamination in mismatched base pairs. *Biochemistry*, **32**, 6523–6530.
61. Jia,L., Shafirovich,V., Shapiro,R., Geacintov,N.E. and Broyde,S. (2005) Structural and thermodynamic features of spiroiminodihydantoin damaged DNA duplexes. *Biochemistry*, **44**, 13342–13353.
62. Wu,M., Yan,S., Patel,D.J., Geacintov,N.E. and Broyde,S. (2002) Relating repair susceptibility of carcinogen-damaged DNA with structural distortion and thermodynamic stability. *Nucleic Acids Res.*, **30**, 3422–3432.
63. Yan,S., Wu,M., Patel,D.J., Geacintov,N.E. and Broyde,S. (2003) Simulating structural and thermodynamic properties of carcinogen-damaged DNA. *Biophys. J.*, **84**, 2137–2148.