1240–1255 *Nucleic Acids Research, 2018, Vol. 46, No. 3* *Published online 18 December 2017*
doi: 10.1093/nar/gkx1216

Enhanced spontaneous DNA twisting/bending fluctuations unveiled by fluorescence lifetime distributions promote mismatch recognition by the Rad4 nucleotide excision repair complex

Sagnik Chakraborty[1], Peter J. Steinbach[2], Debamita Paul[3], Hong Mu[4], Suse Broyde[4], Jung-Hyun Min[3,*] and Anjum Ansari[1,5,*]

[1]Department of Physics, University of Illinois at Chicago, Chicago, IL 60607, USA, [2]Center for Molecular Modeling, Center for Information Technology, National Institutes of Health, Bethesda, MD 20892, USA, [3]Department of Chemistry, University of Illinois at Chicago, Chicago, IL 60607, USA, [4]Department of Biology, New York University, New York, NY 10003, USA and [5]Department of Bioengineering, University of Illinois at Chicago, Chicago, IL 60607, USA

Received May 17, 2017; Revised November 07, 2017; Editorial Decision November 21, 2017; Accepted December 12, 2017

ABSTRACT

Rad4/XPC recognizes diverse DNA lesions including ultraviolet-photolesions and carcinogen-DNA adducts, initiating nucleotide excision repair. Studies have suggested that Rad4/XPC senses lesion-induced helix-destabilization to flip out nucleotides from damaged DNA sites. However, characterizing how DNA deformability and/or distortions impact recognition has been challenging. Here, using fluorescence lifetime measurements empowered by a maximum entropy algorithm, we mapped the conformational heterogeneities of artificially destabilized mismatched DNA substrates of varying Rad4-binding specificities. The conformational distributions, as probed by FRET between a cytosine-analog pair exquisitely sensitive to DNA twisting/bending, reveal a direct connection between intrinsic DNA deformability and Rad4 recognition. High-specificity CCC/CCC mismatch, free in solution, sampled a strikingly broad range of conformations from B-DNA-like to highly distorted conformations that resembled those observed with Rad4 bound; the extent of these distortions increased with bound Rad4 and with temperature. Conversely, the non-specific TAT/TAT mismatch had a homogeneous, B-DNA-like conformation. Molecular dynamics simulations also revealed a wide distribution of conformations for CCC/CCC, complementing experimental findings. We propose that intrinsic deformability promotes Rad4 damage recognition, perhaps by stalling a diffusing protein and/or facilitating 'conformational capture' of pre-distorted damaged sites. Surprisingly, even mismatched DNA specifically bound to Rad4 remains highly dynamic, a feature that may reflect the versatility of Rad4/XPC to recognize many structurally dissimilar lesions.

INTRODUCTION

Control and maintenance of the genome is carried out by proteins that assemble at specific DNA target sites. These biological processes require efficient recognition of target sites, which typically involves conformational reorganizations in both the protein and the DNA in a specific manner, as shown by X-ray and nuclear magnetic resonance (NMR) structural studies. Target DNA, in particular, often becomes deformed from the canonical B-DNA conformation when forming a specific complex with a partner protein. Currently, much remains unclear as to how these changes are initiated and funneled to bring about recognition of the target sites, and how rapidly the conformations adjust upon binding.

NMR measurements have revealed microsecond-to-millisecond timescale conformational fluctuations for DNA and proteins non-specifically interacting with each other, indicating that preliminary interrogation may encompass these timescales (1–3). Rates of conformational changes in these molecules as the complex transitions from non-specific to specific binding have been more difficult to capture although there are notable exceptions, for example from stopped-flow studies (4–6), micromanipulation exper-

*To whom correspondence may be addressed. Tel: +1 312 996 8735; Fax: +1 312 996 9016; Email: ansari@uic.edu
Correspondence may also be addressed to Jung-Hyun Min. Tel: +1 312 355 0838; Fax: +1 312 996 0431; Email: jhmin@uic.edu

iments (7), and laser temperature-jump (T-jump) perturbation approach (8–13). With T-jump, we could also dissect multiple steps attributable to DNA bending/twisting during preliminary interrogation and DNA kinking and/or nucleotide-flipping dynamics during site-specific recognition (11,13). Despite the successes in structural and dynamical studies on protein–DNA complexes, our understanding of how the conformational adaptabilities of DNA and proteins influence the recognition remains limited, particularly for large protein–DNA complex systems that are not amenable to NMR. These limitations persist because the techniques applicable to such large systems (e.g. X-ray crystallography structures and 'ensemble' dynamics measurements such as stopped-flow or T-jump) cannot capture the complex, heterogeneous conformations that coexist and interconvert in solution under equilibrium. Measurements that can reveal the equilibrium conformational distributions of DNA, for instance, can provide valuable information on the deformability of DNA as well as intrinsic DNA distortions that are thermally accessible. The deformability/distortions of DNA are likely used for target recognition by site-specific proteins, especially those that rely on 'indirect readout' (14). Here, we combined picoseconds-resolved fluorescence lifetime measurements with a maximum entropy analysis to visualize the population distributions of DNA conformations as probed by a unique set of fluorescent probes (tC° and tC_{nitro}). This approach could effectively address key questions, including what conformations are thermally accessible to the DNA molecules and whether these conformations in DNA alone resemble the conformations adopted in the specific complex.

The lifetime approach (15) has several advantages over other techniques such as single-molecule FRET (smFRET) which can also yield information on conformational distributions (16–19): it can detect minor populations as this technique samples Avogadro's number of molecules instead of a few thousands; it has higher sensitivity that enables the use of intrinsic probes such as tryptophan (Trp) residues (in proteins) (20) or fluorescent nucleotide analogs (in DNA) (21–23) with relatively low quantum yield instead of the high quantum-yield probes needed for smFRET; it provides snapshots of a conformational distribution that are captured over a much shorter time-window, typically less than tens of nanoseconds and characteristic of the excited state lifetime of the fluorophore, compared with the milliseconds time-window required in smFRET measurements.

The fluorescent probes used in this study, tC° and tC_{nitro}, are cytosine analogs which can be incorporated in the DNA as a donor and an acceptor, respectively, for fluorescence resonance energy transfer (FRET) (24,25). These probes exhibit normal Watson–Crick pairing with guanine with minimal perturbation of DNA structure and stability (13,24). Furthermore, the rigid exocyclic ring and its base stacking interactions hold these nucleotide analogs in relatively fixed orientations within the DNA helical structure, making the FRET between them sensitive to DNA helicity, and hence to even subtle distortions in DNA conformations that alter their separation and/or relative orientation (26–28). These probes have proven suitable to monitor transitions between B-DNA and Z-DNA conformations (29). In our previous work, we applied them to measure protein-induced DNA unwinding dynamics using T-jump (13).

Here, we take advantage of the fluorescence-lifetime approach to understand how DNA damage is recognized by the Rad4–Rad23 nucleotide excision repair (NER) complex. Rad4–Rad23, the yeast ortholog of the human xeroderma pigmentosum C (XPC)–RAD23B complex (reviewed in (30,31)) is a key damage recognition factor in the global genome NER pathway: upon DNA damage, it localizes to the lesions scattered around the genomic DNA in cells and is strictly required for the recruitment of the transcription factor IIH complex (TFIIH) containing XPD and XPB helicases. TFIIH in turn verifies the presence of a bulky lesion and recruits other subsequent factors (32–34). Successful lesion recognition and proofreading leads to the excision of the lesion-containing single-stranded DNA (24–32 nt) followed by repair synthesis and nick sealing (30,35). Inherited mutations in *XPC* in humans can cause the xeroderma pigmentosum syndrome, which entails high sun sensitivity and predisposition to skin cancers due to the inability to repair sunlight-induced DNA lesions (36,37). *In vitro*, Rad4 and XPC bind specifically to various lesions repaired by NER; these include the 6–4 photoproduct (6–4PP) induced by UV light and adducts derived from polycyclic aromatic hydrocarbons, aromatic and heterocyclic aromatic amines, which are widely present in environmental pollutants, notably automobile exhaust and in cigarette smoke and grilled meat (38–41). While some of these lesions are helix distorting and destabilizing, others are not (42). Rad4/XPC can also bind to artificially destabilized DNA such as 2 to 3 base pair (bp) mismatch sites with an affinity similar to that of a *bona fide* substrate, 6–4PP (43,44). Notably, the recognition and repair efficiencies of NER lesions can vary widely (30,45–57), and the lesions that are recognized less efficiently and thus repaired more slowly, tend to be more toxic and mutagenic compared with the repair-proficient lesions (58–65). Thus, understanding the basis of such variance in repair and mutagenicity has been a key issue in the field (30,42). The crystal structures of Rad4 bound to UV-induced as well as mismatched DNA model lesions have shown that Rad4 unwinds and bends DNA at these lesion sites, flips out two damage-containing nucleotide pairs from both strands, and inserts a β-hairpin in the opened DNA to stabilize this recognition complex (Figure 1) (44). Notably, in this 'open' structure, Rad4 does not make any direct contact with the damaged nucleotides (cyclobutane pyrimdine dimer (CPD) UV lesion) but interacts exclusively with the normal nucleotides flipped out from the complementary strand. Such an indirect mode of recognition provides insights into how Rad4 may serve as a common sensor for diverse lesions to be removed by NER (44,66). We previously characterized the DNA dynamics during the recognition process using a series of T-jump measurements on Rad4 bound to mismatched DNA, used as model substrates for Rad4. These studies demonstrated that Rad4 non-specifically unwinds the DNA on ∼100–500 μs time scale before specifically recognizing (and flipping out) the mismatched nucleotides, which takes several milliseconds (12,13). However, little is known about the underlying conformational landscapes of the abnormal versus normal

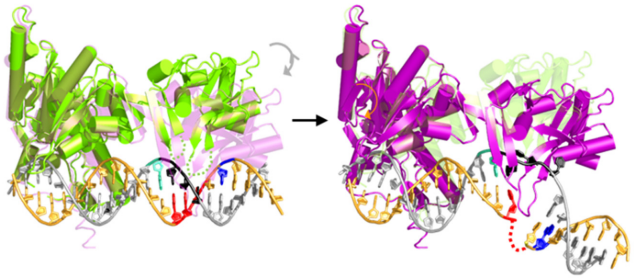

Figure 1. Rad4 and DNA undergo conformational changes upon specific binding. (*Left*) Rad4 (green) bound to B-DNA in a non-specific binding mode. To model this non-specific binding mode, we used the apo-Rad4–Rad23 structure (PDB code: 2QSF) superposed on a model B-DNA structure. The 3 bp mismatched nucleotides whose sequences were varied in this study are indicated in red and black. The positions of the donor (tC°, blue) and acceptor (tC_{nitro}, cyan) are also shown. (*Right*) Rad4 (purple) bound to a UV-damaged/mismatched DNA in a specific manner, forming an 'open' recognition complex (PDB code: 2QSG and 2QSH). In this structure, the DNA is unwound and bent, with two damage/mismatch-containing nucleotide pairs flipped out from the duplex DNA. Rad4 inserts a β-hairpin into the DNA duplex to stabilize this unwound 'open' structure of the DNA. Two flipped-out nucleotides (black) from the undamaged strand (gray) make direct contacts with the protein; however, two (red) on the damaged strand (gold) are flipped out from the protein and are disordered in the crystals, thus indicated only schematically. The third mismatched base pair remains intrahelical. The partially transparent structures indicate the specific complex structure (purple) on the left, and the non-specific complex model (green) on the right. Gray arrow indicates the motion of the protein during the recognition process. The figure was made using PyMOL version 1.7.6.6 (Schrodinger, LLC).

DNA under equilibrium and how these are altered upon Rad4-binding.

This fluorescence lifetime study addresses these outstanding questions by examining the range of conformations accessible to a set of 3 bp mismatched DNAs and tracking their changes in the presence of bound Rad4. We particularly took advantage of our recent finding that certain mismatched sequences are recognized by Rad4 more specifically than other mismatches (13), which may reflect aspects of repair-proficient and -resistant NER lesions in cells. Our approach revealed that mismatched DNA is intrinsically more dynamic than matched DNA and exhibits a broader range of deformations that likely include unwinding and bending. Most importantly, even in the absence of Rad4, the range of these deformations is larger for mismatched DNA recognized with high specificity by Rad4 compared with DNA with low or little specificity. These experimental observations were further supported and complemented by molecular dynamics (MD) simulations. The differences in these *thermally* accessible conformations may help the repair proteins distinguish DNA lesions from normal DNA by, for instance, preferentially stalling the repair proteins at damaged sites during a one-dimensional (1-D) diffusional search, as well as by guiding the subsequent recognition through 'conformational selection/capture'. Another surprising revelation from these studies was that even the most specific complex remained highly dynamic and heterogeneous, a feature that may reflect the promiscuous nature of Rad4 (and XPC, by analogy) and its ability to recognize a wide range of structurally dissimilar lesions.

MATERIALS AND METHODS

Preparation and characterization of Rad4–DNA complexes

DNA oligonucleotides, 24 nt long, containing the fluorescent probes were purchased from TriLink Biotechnologies with reverse-phase HPLC purification. All oligonucleotides appeared as a single band on denaturing polyacrylamide gels, indicating high length-purity (>90%) of the oligonucleotides. Annealing to make duplex DNA was done in phosphate-buffered saline (PBS) (10 mM Na_2HPO_4, 2 mM KH_2PO_4, 137 mM NaCl, 2.7 mM KCl, pH 7.4) as described in Supplementary Methods 1.1. All Rad4–Rad23 complexes were prepared as previously described (12,13) (Supplementary Methods 1.2). Rad4–Rad23–DNA complexes were prepared by combining purified Rad4–Rad23 and annealed duplex DNA substrates in a 1:1 molar ratio at 5 μM each in PBS containing 1 mM dithiothreitol (DTT). When analyzed by native gel electrophoresis, 83 ± 6% of protein and DNA in the samples were bound for highly specific mismatched DNA and 77 ± 4% for non-specific matched DNA (Supplementary Methods 1.3 and Figure S1). The apparent binding affinities ($K_{d,app}$) were also determined by competition electrophoretic mobility shift assays (EMSA) carried out essentially as before in EMSA buffer (5 mM BTP-HCl, 75 mM NaCl, 5 mM DTT, 5% glycerol, 0.74 mM CHAPS, 500 μg ml^{-1} BSA, pH 6.8) (Supplementary Methods 1.4) (13). The thermal stabilities of the DNA duplexes were measured as described in Supplementary Methods 1.5.

Picoseconds-resolved fluorescence spectroscopy

Fluorescence decay curves were measured with a PicoMaster fluorescence lifetime instrument (HORIBA-PTI, London, Ontario, Canada) equipped with time-correlated single photon counting (TSCPC) electronics (67). Further details of the apparatus and data acquisition are in Supplementary Methods 1.6. For all FRET measurements, decay traces were measured for donor-only (tC°-labeled, DNA_D) duplexes without acceptor (tC_{nitro}) as well as donor–acceptor (tC°-tC_{nitro}-labeled, DNA_DA) duplexes.

Analysis of the fluorescence decay traces

Decay curves were analyzed either as a sum of discrete exponentials (DE; Supplementary Methods 1.7) convolved with the instrument response function (IRF; obtained as described in Supplementary Methods 1.6), or using a maximum entropy method (MEM; Supplementary Methods 1.8), in which the effective distribution of log-lifetimes $f(\log \tau)$ was inferred from the decay traces using the program MemExp (68,69) (available online). The MEM distributions were further analyzed as a sum of Gaussian distributions (Supplementary Methods 1.9). The FRET values between tC° and tC_{nitro}, obtained as described in Supplementary Methods 1.7–1.9, were compared with theoretical predictions based on DNA structures, as described in Supplementary Methods 1.10.

Molecular modeling and molecular dynamics (MD) simulations

We performed 2 μs of MD simulations starting from the initial models of the mismatched duplexes and their matched controls. We prepared the initial models with Discovery Studio (*Dassault Systèmes* BIOVIA) by choosing 12-bp sequences with centrally positioned mismatched or matched base pairs (detailed in Supplementary Methods 1.11). We used the AMBER14 (70) package with ff14SB force field (71), explicit water and counterions. The structural properties were calculated using the cpptraj module of AMBER14 (70) and plotted using MATLAB 7.10.0 (The MathWorks, Inc.). The principal component analyses for obtaining structural clusters were performed with the Bio3D package (72) in R (73). All molecular structures obtained from the simulations were rendered using PyMOL 1.3.× (Schrodinger, LLC.). Full details of force field, molecular modeling, MD simulation and analysis protocols are given in Supplementary Methods 1.11–1.13.

RESULTS

Fluorescence lifetime measurements combined with maximum entropy analyses provide a robust snapshot of conformational distributions

In the present study, we use the FRET efficiency between tC° (donor) and tC_{nitro} (acceptor) placed within a duplex DNA to probe the DNA conformations. Such FRET efficiency (E) can be obtained directly from the lifetimes of the excited donor fluorophore, as $E = 1 - \frac{\tau_{DA}}{\tau_D}$, where τ_{DA} and τ_D are the donor lifetimes in the presence and absence of the acceptor, respectively. Fluorescence lifetime measurements offer a more robust way to obtain FRET than steady-state measurements that rely on fluorescence intensities, as the lifetimes depend neither on the excitation intensity nor on the fluorophore concentrations. More importantly, the picosecond time-resolution of our instrumentation enables us to capture the distribution of conformations that co-exist in solution, if that distribution is mirrored in different FRET values between the probes. Here, the fluorescence lifetime of tC° in the absence of tC_{nitro}, i.e., τ_D measured in DNA constructs harboring only the donor (DNA_D), is relatively insensitive to its macromolecular environment (74). However, when tC° is in the proximity of tC_{nitro} to which it can transfer energy, as in DNA constructs containing both the donor and the acceptor (DNA_DA), its lifetime τ_{DA} is shortened due to the FRET. As the FRET efficiency depends strongly on the relative distance and orientation between the donor and acceptor, the FRET efficiency and the corresponding τ_{DA} provide a sensitive measure of the DNA conformations harboring the probes. A single conformation with a fixed donor–acceptor distance/orientation is expected to exhibit a single FRET efficiency and hence a single-exponential fluorescence decay trace with a unique τ_{DA}. If the sample consists of multiple conformations, then this heterogeneity can be reflected in a distribution of FRET efficiencies yielding a multi-exponential decay trace with multiple values of τ_{DA}.

Extracting multiple lifetime components (and thus conformations) from a complex, multi-exponential decay trace can be a challenge (75), which we have effectively overcome by utilizing the MEM. The MEM inverts fluorescence decay traces into lifetime distributions without the need to specify the number of exponential terms (76). The program MemExp (68,69,77) has been extended here to account for the zero-time shift between the instrument response and the kinetics (see Supplementary Methods 1.8), to address the ambiguity in parameters that can arise in a DE analysis of the experiments reported.

A

Construct	Sequence	$K_{d,app}$ (nM)
CH10_NX	5′-TTGACTCGACATCC**CCC**GCTACAA -3′ 3′- ACTGAGCTGTAGG**CCC**CGATGTTA -5′	54 ± 2
CH7_NX	5′-TTGACTCGACATCC**CCC**GCTACAA -3′ 3′- ACTGAGCTGTAGG**GGG**CGATGTTA -5′	344 ± 19
CCC/CCC_I	5′-TTGACTCGACATC**P**CCCGCTACAA -3′ 3′- ACTGAGCTGTAGG**CCC**C**D**ATGTTA -5′	70 ± 3
GGG/CCC_I	5′-TTGACTCGACATC**P**GGGGCTACAA -3′ 3′- ACTGAGCTGTAGG**CCC**C**D**ATGTTA -5′	355 ± 28
TTT/TTT_I	5′-TTGACTCGACATC**P**TTTGCTACAA -3′ 3′- ACTGAGCTGTAGG**TTT**C**D**ATGTTA -5′	116 ± 3
AAA/TTT_I	5′-TTGACTCGACATC**P**AAAGCTACAA -3′ 3′- ACTGAGCTGTAGG**TTT**C**D**ATGTTA -5′	377 ± 9
TAT/TAT_I	5′-TTGACTCGACATC**P**TATGCTACAA -3′ 3′- ACTGAGCTGTAGG**TAT**C**D**ATGTTA -5′	504 ± 36
ATA/TAT_I	5′-TTGACTCGACATC**P**ATAGCTACAA -3′ 3′- ACTGAGCTGTAGG**TAT**C**D**ATGTTA -5′	534 ± 44

B

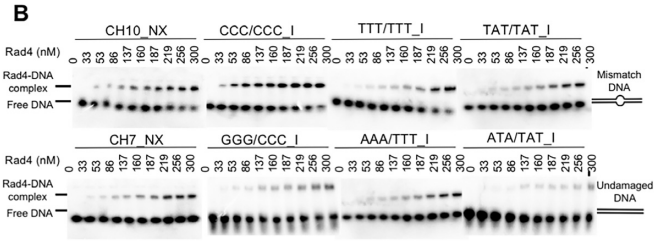

Figure 2. Apparent Rad4-binding affinities of design I DNA constructs measured by competition gel-shift assays. (**A**) The construct names and DNA sequences of the 24-bp DNA substrates are shown along with the apparent Rad4-binding affinities ($K_{d,app}$) of each DNA as determined by the gel-shift assays. Mismatched or the corresponding matched base pairs are underlined; tC° (D) or tC_{nitro} (P) are also indicated. (**B**) Typical gel images are shown from the assays for the DNA constructs indicated.

Conformational distortions in mismatched DNA substrates for Rad4 were probed using tC°-tC_{nitro} FRET pair

The choices of the DNA substrates and probe positions in this study were motivated by previous studies of Rad4-DNA binding and are shown in Figure 2. As mentioned, duplex-destabilizing 2 or 3 bp mismatches are useful model DNA substrates for Rad4/XPC *in vitro* (12,43,44). The structure of Rad4-bound to mismatched DNA was found to be identical to the 'open' Rad4–DNA complex structure in the presence of a UV lesion (44). Notably, the relative strength in Rad4-binding affinity (i.e., specificity) of mismatched DNA depends on the mismatch sequence (13): competition gel-shift assays showed that Rad4 exhibited high specificity for a CCC/CCC mismatch, intermediate specificity for a TTT/TTT mismatch, and little specificity for a TAT/TAT mismatch, with apparent binding affinities $K_{d,app}$ of 70 ± 3 nM, 116 ± 3 nM, and 504 ± 36 nM, respectively (Supplementary Methods 1.4; Figure 2 and Supplementary Figure S2) (13). The binding affinity measurements indicated that DNA duplexes with the mismatched sites CCC/CCC, TTT/TTT and TAT/TAT bound more tightly than their matched counterparts by a factor of ∼5.1, ∼3.3

and ∼1.1, respectively. Thus, in this study, we have used these three mismatched duplexes along with their matched counterparts and examined their conformational distributions, as probed by FRET between tC° and tC_{nitro} probes.

We previously showed that tC°-tC_{nitro} probes incorporated in DNA do not much affect the thermal stabilities or the Rad4-binding affinities of DNA duplexes (13). Incorporation of these probes on either side of a TTT/TTT mismatch, at positions indicated as in Figure 2 (denoted as design I constructs), could detect DNA distortions induced by specific Rad4 binding, as measured by FRET changes in the mismatched but not the matched constructs (13). These FRET measurements were consistent with the specific binding of TTT/TTT to wild-type Rad4 demonstrated in gel-shift experiments and crystal structures (13,44) (Figure 2). T-jump studies on these tC°-tC_{nitro}-labeled DNA constructs (hereafter referred to as 'TTT/TTT_I') further revealed ∼10 ms kinetics when specifically bound to wild-type Rad4 as well as ∼100–500 μs kinetics when non-specifically bound to Rad4 mutants (13). The ∼10 ms relaxation times were similar to those obtained using a 2-aminopurine probe (2AP, a fluorescent analog of adenine) incorporated as a part of 2 or 3 bp mismatch sites, thereby probing full nucleotide flipping kinetics directly at the mismatch (12). The ∼10 ms kinetics were thus attributed to the opening of the damaged site that accompanies full nucleotide-flipping, as seen in the 'open' crystal structures of the specific complex; the ∼100–500 μs kinetics were attributed to unwinding of DNA during non-specific interrogation (13). These results pointed to the ability of the tC°-tC_{nitro} probes in design I to detect various conformational states of DNA along the Rad4–DNA binding trajectory. Here, we used the same placement of tC° and tC_{nitro} (design I) in each of the aforementioned three mismatched/matched DNA pairs to examine the DNA conformational distributions under equilibrium using fluorescence-lifetime-based FRET measurements.

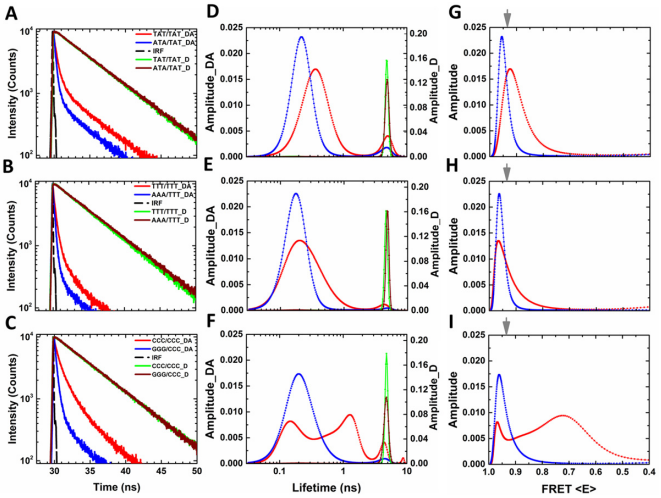

Figure 3. Fluorescence lifetime measurements on mismatched and matched DNA constructs. (**A**–**C**) Fluorescence intensity decay curves for DNA_DA (labeled with donor and acceptor) with excitation of donor (tC°) for mismatched (red) and matched (blue) DNA constructs: TAT/TAT and ATA/TAT (A); TTT/TTT and AAA/TTT (B); CCC/CCC and GGG/CCC (C). In each panel, corresponding fluorescence intensity decay curves for DNA_D (labeled only with donor) are shown for mismatched (green) and matched (brown) constructs. The instrument response function (IRF) is shown in black. (**D**–**F**) The distribution of lifetimes that best describe the intensity decay profiles, as obtained from the MEM, are shown. The amplitudes from the MEM analysis, normalized to add up to one, are indicated on the left y-axis for DNA_DA and on the right y-axis for DNA_D. (**G-I**) FRET distributions obtained from the lifetime distributions of panels D–F are shown for mismatched (red) and the corresponding matched (blue) DNA. The gray arrows indicate the value of FRET expected for design I constructs as computed from models of B-DNA conformations for each of the matched sequences, as described in Supplementary Methods 1.10.

Mismatched DNA construct with the highest specific binding to Rad4 exhibited the largest deviations from canonical B-DNA conformations in the absence of Rad4

The fluorescence lifetime measurements of donor-only (DNA_D) and donor–acceptor-labeled (DNA_DA) design I constructs at 20°C, are shown in Figure 3. When analyzed by DE fits, each fluorescence decay trace of the DNA_D constructs could be described in terms of a single-exponential decay with a time constant in the range of 4.64.9 ns, consistent with previous results by the Wilhelmsson group (74) (Supplementary Methods 1.7 and Figure 3A–C). In contrast, the DNA_DA constructs exhibited significant deviations from a single-exponential decay, requiring at least two or sometimes three time constants (Supplementary Figures S3 and 4; Table S1).

Consistent with the DE analyses, the MEM analyses of the decay traces showed single sharp distributions of lifetimes for donor-only DNA_D and broader distributions for donor–acceptor-labeled DNA_DA (Figure 3D–F). The broader distributions of lifetimes of the DNA_DA samples translated into correspondingly broad FRET distributions (Figure 3G–I and Supplementary Figure S5). The reproducibility of the distributions obtained from the MEM analyses from four independent lifetime measurements on each sample are illustrated in Supplementary Figure S6. The data presented in Figure 3 are for one representative from this set. To determine the minimal number of components that describe the lifetime distributions from the MEM analyses, we fitted the measured distributions to a sum of Gaussians (Supplementary Figure S4). The peak position of each Gaussian component was used to calculate the average FRET representing that component and the area under the Gaussian curve was taken as a measure of the fractional population of that component (Figure 4A–C).

For all the donor–acceptor-labeled DNA_DA constructs, we observed broad distributions extending from ∼50 ps to 10 ns (Figure 3D–F and Supplementary Figure S4). Notably, every DNA construct exhibited a minor but distinct long-lifetime component centered at ∼5 ns; this component overlapped with the predominantly single lifetime component measured for the donor-only DNA_D samples, resulting in an apparent 'zero-FRET' component of varying amplitudes in different samples. This component could arise from donor-only samples that have a missing or inactive acceptor strand and/or from real conformations sampled by DNA. A similar zero-FRET component is routinely observed in smFRET measurements and typically discarded as an artefact (19). While we did control experiments to rule

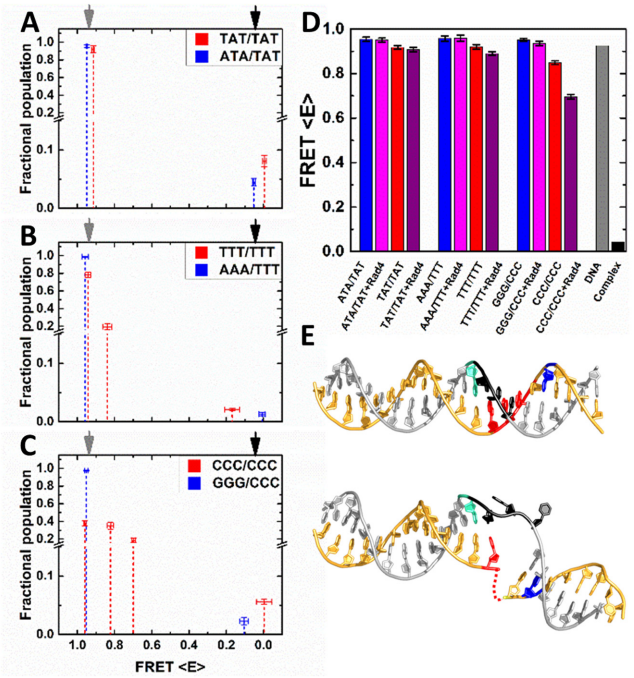

Figure 4. FRET values for distinct DNA conformations obtained from fluorescence lifetime measurements on DNA only samples. The fractional populations versus the average FRET computed for each of the distinct conformations, as obtained from Gaussian fits to the fluorescence lifetime distributions of DNA_DA are shown for matched (blue) and mismatched (red) DNA constructs: TAT/TAT and ATA/TAT (**A**); TTT/TTT and AAA/TTT (**B**); CCC/CCC and GGG/CCC (**C**). The arrows indicate the computed values of FRET for B-DNA conformation (gray) and for the DNA conformation in the Rad4-bound specific complex (black), computed as described in Supplementary Methods 1.10. (**D**) The average FRET ⟨E⟩ for all design I constructs obtained from fluorescence lifetime decays are shown for DNA alone (matched: blue; mismatched: red) and Rad4–DNA (matched: magenta; mismatched: purple). The value of FRET expected for design I constructs are shown for B-DNA conformation (gray) and for Rad4–DNA complex (black), computed as described in Supplementary Methods 1.10. (**E**) The corresponding DNA structures are also shown below: (top) model B-DNA structure; (bottom) structure of DNA in the Rad4-bound specific complex (from PDB ID: 2QSH). The positions of 3 bp mismatched nucleotides are indicated in red and black. The positions of the donor (tC°, blue) and acceptor (tC_{nitro}, cyan) are also shown.

out excess donor-only strands in our measurements (Supplementary Figure S7), we could not conclusively ascertain whether this component was an artefact or reflected small populations of distorted DNA conformations. Further discussion of this component is in Supplementary Discussion and Figures S7–9.

Here we focus primarily on the dominant distributions of lifetimes in the 50 ps – 3 ns window. The lifetime distributions of all the matched constructs in this window exhibited a single predominant conformation with average FRET efficiencies of 0.95 ± 0.01 and fractional populations of 96–99%; we denote these as the 'high-FRET' conformations (Figures 3D–I and 4A–C; Supplementary Figure S4; uncertainties in the fractional populations are ∼0–3%; the uncertainties associated with these numbers, also summarized in the Supplementary Tables S1–3, are standard error of the mean (s.e.m.) from four independent sets of measurements). These FRET values from fluorescence lifetime measurements are consistent with previously reported FRET on design I constructs measured by steady-state fluorescence (13). Assuming a canonical B-DNA conformation, the FRET value of these matched DNA constructs is computed to be 0.93, based on the distance and relative orientation of the probes (78); thus, the experimental FRET of ∼0.95 for the dominant component of matched DNA is in good agreement with the computed FRET.

On the other hand, the distributions of the mismatched constructs were broader and shifted to longer lifetimes. Importantly, the deviations from the matched DNA profiles were more pronounced for the mismatched sequences with higher specific binding to Rad4. In the non-specific mismatched TAT/TAT construct, the dominant component was still well described by a single Gaussian (Figure 3D and G; Supplementary Figure S4D) though the average FRET for this 'high-FRET' component shifted to 0.92 ± 0.01 (91%) (Figure 4A). The medium specificity TTT/TTT construct exhibited slight asymmetry and required two Gaussians consisting of a high-FRET component at 0.94 ± 0.01 (78%) and an additional 'medium-FRET' component at 0.84 ± 0.01 (20%) (Figures 3E, H and 4B; Supplementary Figure S4E). Notably, the highest specificity CCC/CCC construct exhibited the broadest distribution and required three Gaussians in the 50 ps – 3 ns window; the high-FRET conformation appeared in roughly the same position as the predominant conformations of the matched DNA constructs, at 0.96 ± 0.00, but with decreased amplitude (38%), while two additional medium-FRET conformations appeared at 0.82 ± 0.01 (35%) and 0.70 ± 0.01 (22%) (Figures 3F, I and 4C; Supplementary Figure S4F and Table S2).

The broader distributions of lifetimes (and thus FRET) in mismatched DNA demonstrate that these DNA constructs sample a larger range of multiple, distinct conformations in comparison with matched DNA, which indicates higher deformability in the mismatched DNA (Figures 3 and 4A–C). The range of conformations sampled by mismatched DNA includes conformations that resemble B-DNA like conformations as well as those that deviate quite significantly from B-DNA. Deviations from B-DNA-like conformations in mismatched DNA are already evident in the overall average FRET (averaged over the 50 ps – 3 ns window of the lifetime measurements as described in Supplementary Methods 1.8), with smaller average FRET observed for the mismatched constructs compared with their matched counterpart (Figure 4D). These deviations were the largest for the high specificity CCC/CCC construct. Notably, this trend correlated with the change in thermal stability observed for each of the mismatched/matched pair, with melting temperatures for the TAT/TAT, TTT/TTT and CCC/CCC constructs lower than for their matched counterparts by 8.8 ± 1.0°C, 10.6 ± 1.2°C and 18.0 ± 2.1°C, respectively (Supplementary Methods 1.5 and Figure S10).

Rad4-binding significantly altered the conformational distribution of mismatched DNA, but not matched DNA

Next, we examined each of the design I sequences (Figure 2A) in the presence of equimolar Rad4. First, Rad4 binding to the matched DNA revealed little change in the lifetime distributions (Supplementary Figure S11) or in the average

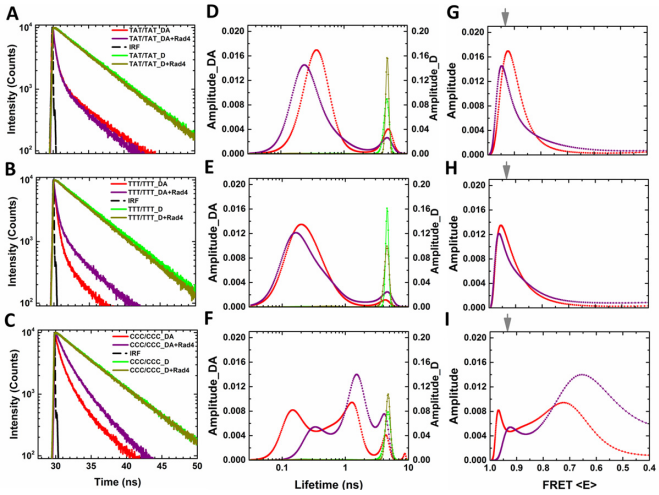

Figure 5. Fluorescence lifetime measurements on mismatched DNA constructs in the presence of Rad4. The data in each panel are as described for Figure 3, for mismatched DNA_DA alone (red) and Rad4–DNA_DA (purple). The corresponding donor-only data are also shown: DNA_D (green) and Rad4–DNA_D (olive green).

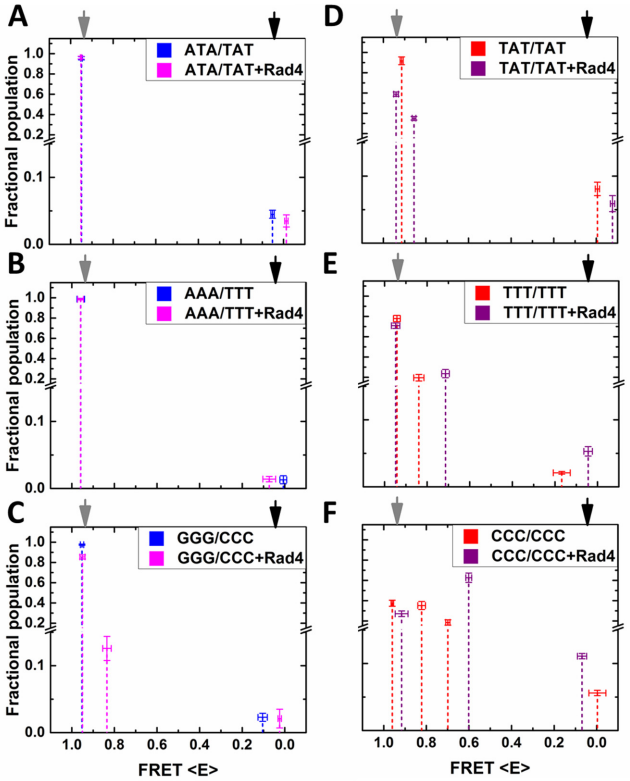

Figure 6. FRET values for distinct DNA conformations obtained from fluorescence lifetime measurements on DNA samples, with and without Rad4. The fractional populations versus the average FRET, as described in Figure 4, are shown for (**A**–**C**) matched DNA and (**D**–**F**) mismatched DNA. (A–C): DNA only (blue) and Rad4–DNA complexes (magenta), for ATA/TAT (A), AAA/TTT (B) and GGG/CCC (C); (D–F): DNA only (red) and Rad4–DNA complexes (purple), for TAT/TAT (D), TTT/TTT (E) and CCC/CCC (F).

FRET values (Figure 4D) compared with free DNA. However, the lifetime distributions (and thus the DNA conformations) of the mismatched DNA were altered upon Rad4 binding as detailed below. Some of these changes could not be detected using steady-state or averaged FRET measurements alone, which reflect only 'ensemble' properties. We note here that under the conditions of our lifetime experiments, at 5 μM each of protein and DNA, we expect ∼70–90% of DNA bound to Rad4, based on our measured $K_{d,app}$ values for each of the DNA constructs (Supplementary Figure S2B). The fraction bound was directly tested for the highly specific CCC/CCC construct and its matched counterpart (Supplementary Figure S1), and was found to be 83 ± 6% for CCC/CCC and 77 ± 4% for GGG/CCC, consistent with the estimated fractions.

The mismatched TAT/TAT construct in the presence of Rad4 revealed an increase in the dynamical fluctuations as indicated by a broader distribution of lifetimes that now required two Gaussians to describe the major peak (Figure 5D and Supplementary Figure S4D), with FRET of 0.94 ± 0.01 (59 ± 1%) and 0.86 ± 0.01 (35 ± 1%), respectively (Figure 6D). These results indicate larger, more varied distortions in mismatched TAT/TAT than in matched ATA/TAT upon Rad4 binding, despite no detectable change in average FRET for either of these constructs in the presence of Rad4 (Figure 4D), and little specific binding of Rad4 for mismatched TAT/TAT compared with its matched counterpart in gel-shift assays (Figure 2). Both TAT/TAT and ATA/TAT were >70% bound under the conditions of these measurements. Therefore, while some contribution of the B-DNA-like component in these complexes could be from DNA molecules that do not have a protein bound, this fraction of unbound DNA was estimated to be <30% for these constructs.

In the case of mismatched TTT/TTT, for which the fraction bound is estimated to be ∼86%, we saw a similar shift in the lifetime distributions as with TAT/TAT (Figure 5E; also compare Supplementary Figures S4E and K), and observed a decrease in the amplitude of the high-FRET component, from 78 ± 3% in the absence of Rad4 to 71 ± 3% with Rad4, and an increase in the medium-FRET component, from 20 ± 4% to 24 ± 3% (Figure 6E and Supplementary Table S2). Furthermore, while the high-FRET component retained the characteristic of the B-DNA conformation (FRET of 0.95 ± 0.01) even in the presence of Rad4, the medium-FRET component exhibited a shift in its average FRET from 0.84 ± 0.01 to 0.71 ± 0.01 upon Rad4 binding (Figure 6E). These results indicate that Rad4 binding might further distort the already distorted DNA conformations that showed propensity for deviations from B-DNA in the absence of Rad4.

These conclusions were further reinforced with the CCC/CCC construct that exhibited the highest specificity for Rad4. For the CCC/CCC construct, the average FRET decreased significantly, from 0.80 ± 0.01 to 0.64 ± 0.02 upon binding to Rad4 (Figure 4D and Supplementary Table S3). In addition, Rad4 had a dramatic effect on the shape of the distribution of lifetimes of the mismatched DNA (Figure 5F; also compare Supplementary Figures S4F and L). In the absence of Rad4, the CCC/CCC construct already revealed significant amplitude in the distorted

medium-FRET conformations in addition to the B-DNA-like conformation. In the presence of Rad4, the high-FRET component shifted from 0.96 ± 0.01 to 0.092 ± 0.02 and diminished in amplitude (from 38 ± 2% to 27 ± 1%), while the two medium-FRET conformations coalesced into one broad distribution with a FRET of 0.60 ± 0.01, and gained in overall amplitude to 63 ± 2% (Figure 6F and Supplementary Figure S4). Thus, in the high-specificity CCC/CCC construct, we observed primarily a shift in the distribution to longer lifetimes upon Rad4 binding, with perhaps some narrowing of the underlying Gaussian components of the distribution (see Supplementary Figure S4). Thus, for DNA that is most specifically bound, Rad4 binding increased the overall deformation, as anticipated from the crystal structures. Surprisingly, however, specific binding of Rad4 to CCC/CCC did not push the equilibrium into a narrow range of DNA conformations but rather the specific complex remained highly dynamic. It is important to note here that while there could be as much as 10–20% unbound CCC/CCC DNA in these measurements that could contribute to the apparent heterogeneity, the absence of any significant amplitude in the B-DNA-like features of the lifetime distributions of Rad4-bound CCC/CCC complex indicated that the unbound fraction in these complexes was negligible.

An alternative explanation for this high heterogeneity in the conformations sampled by Rad4-bound CCC/CCC DNA could be from a heterogeneous distribution of Rad4 along the length of the DNA, with some fraction of proteins bound non-specifically and the remaining fraction bound specifically. To further examine whether our distributions reflected a highly dynamic specific complex or Rad4 simply distributed between specific and non-specific sites along the DNA, we carried out measurements with a 3-fold longer DNA construct (72 bp) containing one CCC/CCC site, to increase the number of non-specific sites (Supplementary Figure S12). We anticipated that the former scenario would leave the distribution unchanged while the latter would increase the population of B-DNA-like conformations at the expense of the more distorted conformations. Our measurements showed that the distribution of conformations remained quite similar (Figure 7), thus supporting our conclusion of a highly dynamic-specific complex.

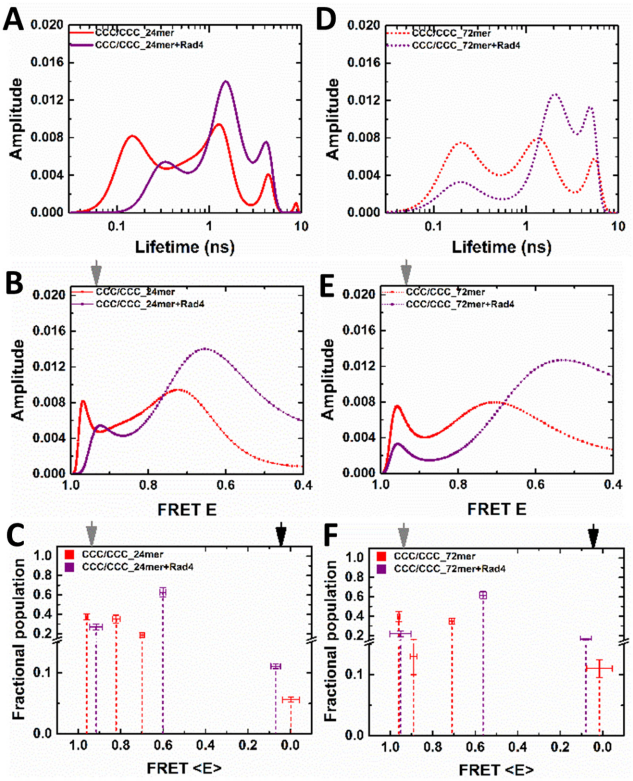

Figure 7. Comparison of fluorescence lifetime measurements on 24-mer and 72-mer DNA with CCC/CCC mismatch, with and without Rad4. Fluorescence lifetime distributions (**A** and **D**), FRET distributions (**B** and **E**) and discrete FRET components from Gaussian fits (**C** and **F**), obtained as previously described in Figures 5 and 6, are shown for 24-mer CCC/CCC construct (A–C) and 72-mer CCC/CCC construct (D–F), for DNA_DA alone (red) and Rad4–DNA_DA (purple).

Distorted conformations in mismatched DNA increased with increasing temperature

The data presented thus far were all obtained at 20°C. To examine how the range of conformations thermally accessible to DNA and Rad4–DNA complexes varied with temperature, we measured the fluorescence lifetime decays for each of the samples at three additional temperatures: 10, 30 and 40°C. As before, four independent measurements were carried out at each temperature for all the samples. The fluorescence lifetime distributions obtained from the MEM analysis on one representative decay trace from each dataset are shown in Figure 8 for all three mismatched DNA constructs in the absence and presence of Rad4, and in Supplementary Figure S13 for the corresponding matched DNA. The lifetime distributions measured for matched DNA, both free and Rad4-bound, were essentially unchanged over the 10−40°C temperature range, and exhibited predominantly a single B-DNA-like conformation at all temperatures, in the 50 ps – 3 ns range. Similar behavior, with very little temperature-dependence in the shape or the peak position of the dominant component, was observed for the non-specific TAT/TAT mismatched construct in the absence of Rad4 (Figure 8A).

In contrast, both the medium-specificity TTT/TTT construct and the high-specificity CCC/CCC construct exhibited significant changes in the lifetime distributions as the temperature was raised (Figure 8B and C). The TTT/TTT construct showed an increase in the asymmetry of the dominant peak, with a shoulder appearing and increasing in amplitude at ∼0.6–0.7 ns, and a minor peak appearing at ∼1.5 ns (Figure 8B). Notably, these effects of increasing temperature resembled those caused by Rad4-binding (Figure 8E). The temperature-dependent effects were even more striking for the CCC/CCC construct than for the TTT/TTT construct, with significant increase in the medium FRET (∼1.5 ns lifetime component) and a shift in the B-DNA-like short lifetime component (Figure 8C). Temperature increase on Rad4-bound DNA further shifted and broadened the lifetime/FRET distributions (Figure 8E and F). These trends were also reflected in the average FRET values which decreased with increasing temperature and with Rad4 bind-

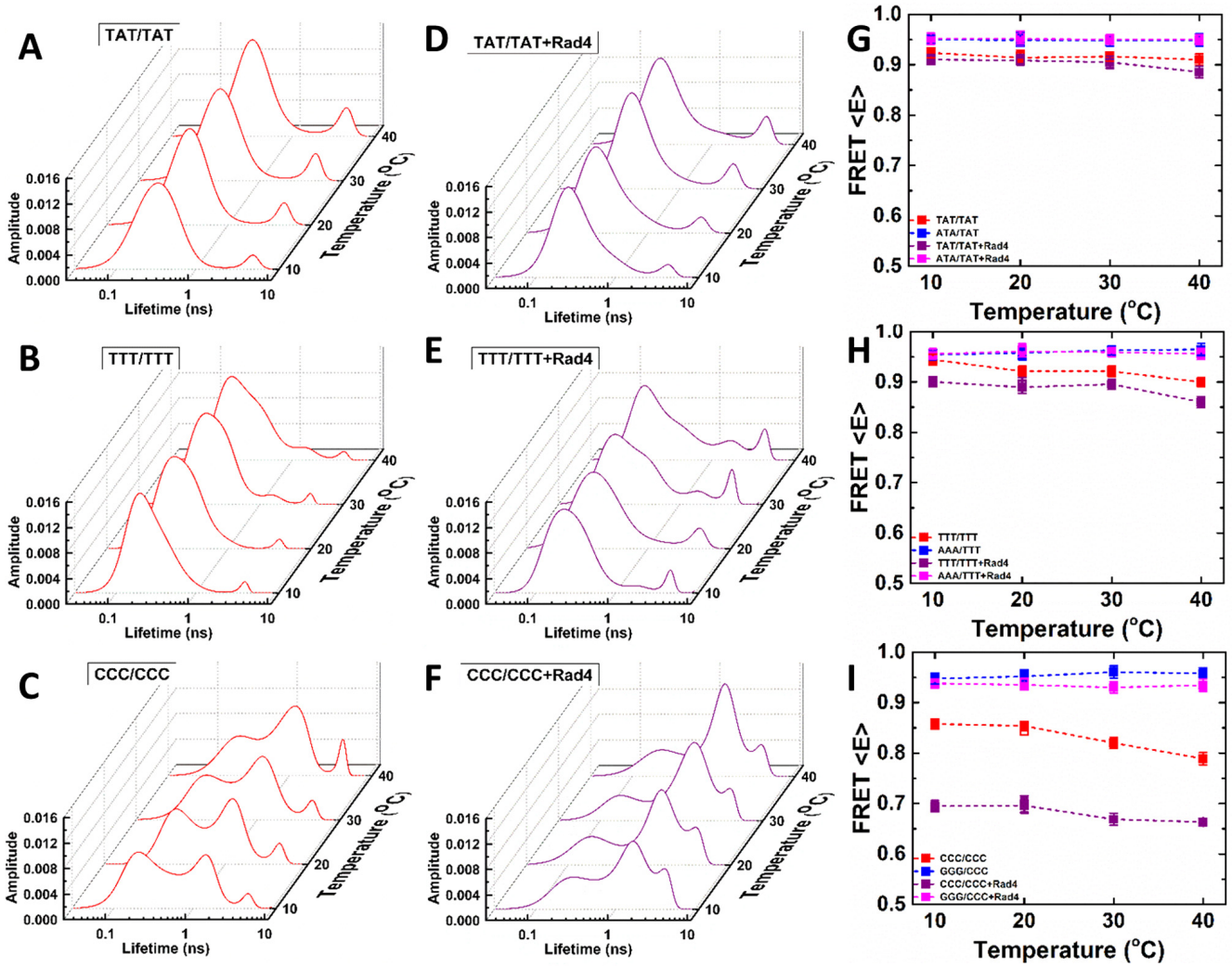

Figure 8. Temperature dependence of fluorescence lifetime distributions and average FRET for mismatched DNA samples, with and without Rad4. (**A**–**F**) The fluorescence lifetime distributions are shown at four different temperatures for mismatched DNA_DA samples in the absence (red; A–C) and presence (purple; D–F) of Rad4. (**G**–**I**) Average FRET values are plotted as a function of temperature for mismatched DNA (red), matched DNA (blue), Rad4–mismatched DNA complex (purple), and Rad4–matched DNA complex (magenta). Data are shown for (G) TAT/TAT and ATA/TAT; (H) TTT/TTT and AAA/TTT; and (I) CCC/CCC and GGG/CCC. The reversibility of the samples after the heating/cooling cycle was checked by first heating up from 10 to 40°C and then cooling back down to 20°C; the measurements after cooling back are shown as open squares.

ing (Figure 8G–I), with the largest changes observed for the high specificity CCC/CCC construct.

We expect that the intrinsic deformability of DNA must be reflected in the range and fractional population of thermally accessible conformations and that the deformability should increase with increasing temperatures. In these DNA constructs, it was the high-FRET populations that broadened and shifted to lower FRET populations with increasing temperature. Thus, these data further support the notion that the high FRET components correspond to less-deformed B-DNA-like conformations, while lower FRET components indicate deviations from B-DNA. Taken together, the temperature-dependent changes reaffirmed that high intrinsic deformability correlated with high specificity in Rad4 binding and illustrated the exquisite capability of the fluorescence lifetime approach in detailing the complex conformational landscapes and their shifts in solution.

MD simulations revealed conformational heterogeneity in mismatched sequences

In order to explore the conformations of the mismatched sequences, we performed 2 μs of MD simulations for the 12-mer DNA duplexes containing each of the three mismatches and for their matched controls. Structural clusters along each trajectory were obtained based on the conformations of the mismatched sequences, as detailed in Supplementary Methods 1.13. The distance between the cytosines (C_A and C_D in Supplementary Figures S14–16, which were replaced with tC_{nitro} (acceptor) and tC° (donor), respectively, in the FRET experiments) was measured between the N4 atoms of the two cytosines. The probability densities of the distance between C_A and C_D for the mismatched sequences and their matched controls, as well as examples of structures, are shown in Figure 9. For the TTT/TTT duplex, the distance between C_A and C_D showed one peak at 15.7 Å,

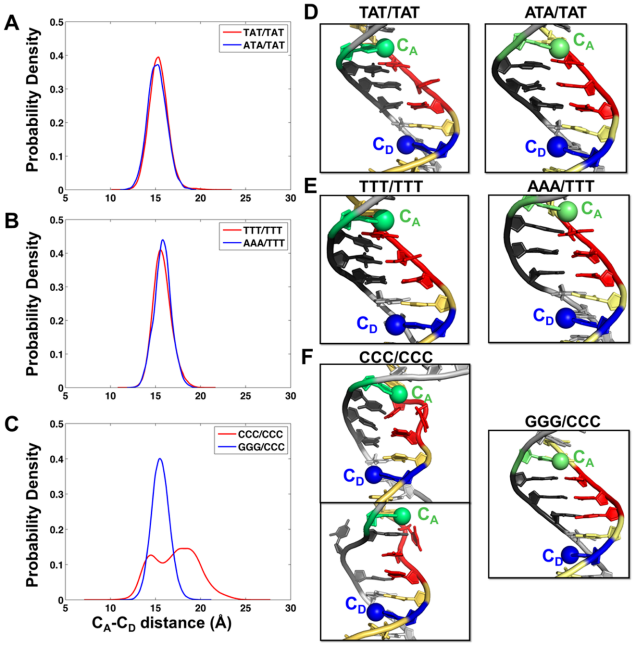

Figure 9. MD simulations of matched and mismatched 12-mer duplexes. (**A**–**C**) The distribution of the C_A–C_D distances for the mismatched duplexes (red) and their matched controls (blue). The C_A and C_D bases were replaced by tC° (donor) and tC_{nitro} (acceptor) for the FRET experiments. (**D**–**F**) Examples of structures for the mismatched duplexes (left) and their matched controls (right). The DNA duplexes are shown in cartoon. The heavy atoms of each base are shown in sticks. The N4 atoms of C_A and C_D bases are shown in spheres. The bases are colored according to the color codes in Figure 1. Full details for the structure clusters of the mismatched duplexes are given in Supplementary Figures S14–16.

very near that of the matched control. The structures also oscillated around one conformation, where the T:T pairs were stacked-in and adopt hydrogen bonds shown in Figure 9 and Supplementary Figure S14. For the TAT/TAT duplex, the distance between C_A and C_D also showed one peak at 15.3 Å (Figure 9), close to the matched control sequence, with a very rare excursion to 18.2 Å (Supplementary Figure S15). However the structures were dynamic between two major clusters, where the A:A base pair adopted two different hydrogen bonding patterns although both patterns exhibited the same C_A and C_D distance (Figure 9 and Supplementary Figure S15). The rare excursion corresponded to a conformation where the A:A bases partially stack with each other (Supplementary Figure S15). Consistent with the experimental data, the CCC/CCC duplex exhibited the highest heterogeneity. The C_A and C_D distance showed two peaks at 14.4 Å and 18.5 Å, while the matched control showed one peak at 15.3 Å. However, the mismatched structures were very dynamic and transited among four major clusters and three minor clusters (Supplementary Figure S16), where the C:C pairs were not stably stacked-in and the duplex often contained one or more flipped out C bases (Figure 9 and Supplementary Figure S16). Mostly, a single C-C hydrogen bond in one of the three C:C mismatches was found. The seven conformational families fell into three distance populations (Supplementary Figure S16).

DISCUSSION

This paper aims to examine how intrinsic DNA conformational fluctuations impact DNA damage recognition, which remains a poorly understood first step in DNA repair. The damage recognition proteins in both mismatch repair and NER systems are not designed to recognize a well-defined structural motif that distinguishes the damaged from the undamaged site; rather, they are tasked with recognizing a wide range of damaged sites against the background of normal DNA. Structures of damage recognition proteins bound to lesion sites show significantly distorted DNA in the complexes, suggesting that recognition likely relies on detecting intrinsic distortions in the helical geometry of damage containing DNA and/or differences in the intrinsic flexibility/deformability of the damaged versus undamaged sites (79,80). Indeed, a large body of NMR studies on DNA containing various damages have shown enhanced propensity of base dynamics in the vicinity of the lesion sites as well as evidence for conformational heterogeneity in damaged DNA (81–87). Similarly, measurements using fluorescent nucleotide analogs placed adjacent to various kinds of mismatches have also revealed increased exposure of these nucleotides to solvent compared to matched DNA (23,88).

In the context of NER, damage-induced helix destabilization often correlates with the lesion recognition and repair propensity (56,89). The presence of unpaired bases opposite a lesion has also been shown to increase the lesion's recognition by XPC and the repair rate by NER (43,66). Although global helix destabilization as indicated by the DNA duplex melting temperatures does not necessarily correlate with its recognition by XPC/Rad4 and repair efficiency by NER, such correlations are frequently observed (42,90,91). MD simulations on lesions derived from polycyclic aromatic chemicals have further probed how lesion-induced structural perturbations may affect thermal stability and lesion recognition, and have suggested that helix-destabilizing distortions can be counteracted by stabilizing van der Waals interactions locally which may influence lesion recognition (59,92). The importance of local dynamics and sequence context in the vicinity of the lesion have been shown in NMR and excision repair experiments as well as in MD simulations for bacterial NER (57,93) and for eukaryotic NER (46–48,94).

Here, we uncovered the distribution of DNA conformations by analyzing the fluorescence lifetime decay traces of tC°-tC_{nitro}-labeled DNA with the maximum entropy method. This powerful approach enabled us not only to detect but also to quantify the relative populations of alternative conformations that co-exist in solution under thermal equilibrium. Our approach also readily lent itself to examining how these conformational distributions were altered in the presence of a large bound protein, a distinct advantage over NMR. Our experimental results together with MD simulations revealed strikingly varied conformational landscapes of DNA containing 3 bp mismatch sites used as model substrates for Rad4, which suggested a direct link between intrinsic DNA deformability and DNA mismatch recognition by this protein.

tC°-tC$_{nitro}$ FRET probes placed close to the mismatched site (as in design I) discerned DNA deformability and conformational variations that correlated with the Rad4-binding specificity of the DNA

As stated before, this study mainly focused on design I DNA constructs (Figure 2 and Supplementary Table S4). Previous equilibrium FRET and T-jump studies using TTT/TTT_I and AAA/TTT_I have suggested that the probes placed as in design I can sense the specific binding of TTT/TTT_I to wild-type Rad4. Here, we expanded the DNA substrates to include two more sets of mismatched/matched DNA, and found that the sequence identity of the mismatched site had a profound effect on the conformational distributions accessible to DNA: the high- and medium-specificity CCC/CCC and TTT/TTT mismatch constructs exhibited more significant distortions from B-DNA-like conformations even in the absence of the protein when compared with the non-specific TAT/TAT construct. Importantly, the degree of these intrinsic DNA distortions correlated with their Rad4 binding specificities. Specific binding to Rad4 further broadened and increased the distorted DNA populations. Notably, our approach could detect even subtle changes in non-specific TAT/TAT_I mismatched DNA when bound to Rad4, which evaded detection using other 'ensemble' approaches. We speculate that the predominantly B-DNA-like conformation undergoing some distortion (as indicated by a minor population in medium FRET) when Rad4 binds to TAT/TAT or TTT/TTT may indicate conformational switching involved in interrogation of the mismatched site, while the populations with larger distortions observed when Rad4 binds to CCC/CCC may reflect conformational states closer to final recognition modes.

The broad range of distortions thermally accessible to the specific, mismatched DNA reflects high intrinsic flexibility/deformability of these DNA substrates; these distortions are likely coupled with increased spontaneous breathing of mismatched base pairs, as observed by NMR at the single base-pair levels (81,95). Taken together, our results demonstrated how the fluorescence lifetime approach can be used to detect subtle differences in the intrinsic DNA dynamics that correlate with specific binding to Rad4.

Probe positions influence the sensitivities to detecting specific versus non-specific conformational changes

As discussed above, the current study used constructs harboring FRET probes at positions designated as design I. With this design, we could not detect any significant changes in the average FRET for the non-specific TAT/TAT or its matched counterpart ATA/TAT, upon binding by Rad4. Previously, we had also carried out FRET studies on TAT/TAT and ATA/TAT, but with probes placed differently from the design I constructs (13). In these 'design II' constructs, hereafter referred to as TAT/TAT_II and ATA/TAT_II, respectively, both the donor and acceptor probes were placed further away from the 3 bp mismatch site in comparison with design I constructs (Supplementary Table S4). Intriguingly, Rad4 binding to both mismatched TAT/TAT_II and matched ATA/TAT_II had revealed a detectable change in FRET that was comparable (13); we attributed these changes in FRET induced by Rad4 in these non-specific substrates to non-specific DNA unwinding by Rad4.

Our observations that tC°-tC$_{nitro}$ probes did not detect any average FRET changes upon Rad4 binding to design I constructs TAT/TAT_I or ATA/TAT_I (Figure 4D), but in fact detected similar average FRET changes upon Rad4 binding to design II constructs TAT/TAT_II and ATA/TAT_II, indicate that the placement of the probes is critical in determining the types of conformational changes detected by these probes. Placing the probes closer together, as in the design I construct, enhances their ability to detect Rad4-induced distortions at the mismatched sites; however, it may also diminish their ability to detect small scale and longer range unwinding induced by Rad4 in its non-specific binding mode, which is picked up when the probes are placed further apart, as in the design II constructs.

Can Rad4 select for pre-distorted conformations to recognize DNA damage?

Numerous studies have shown that both protein and DNA deform their structures when bound specifically to each other, forming specific complexes. Such 'induced-fit' structures have long invoked the question whether specific recognition is initiated by the protein (as it binds and subsequently deforms the DNA) or there exists a pool of pre-distorted DNA conformations that the protein can 'select' to recognize ('conformational selection') (96). Conformational flexibility near or at a lesion site has been hypothesized as a critical feature that proteins sense to pick out potential damaged sites from a vast excess of normal sites (79,80). Although NMR studies have previously yielded data on the dynamics and flexibility of lesion-containing DNA and have provided ample evidence of conformational heterogeneity in these DNAs in the absence of bound protein (82–84), how this heterogeneity is altered when protein binds has been difficult to ascertain. In particular, an open question in the field of site-specific recognition is whether cognate DNA in the absence of protein can adopt pre-distorted conformations that resemble the conformations adopted by DNA in the specific complex.

Our lifetime studies directly address this issue. Our results, particularly those with the highly specific CCC/CCC revealed significant populations of pre-distorted conformations that overlapped with and thus resembled the conformations sampled by DNA in the specific complex. The fact that these conformations were accessible to DNA even in the absence of a bound protein provides experimental support for 'conformational selection' as one of the potential mechanisms for lesion recognition by Rad4 (Figure 10). Computational studies have previously revealed that such conformational capture could be a mechanism for recognizing 10*R*-(+)-*cis*-*anti*-benzo[*a*]pyrene-N^2-dG (*cis*-B[*a*]P-dG), a lesion derived from the environmental carcinogen benzo[*a*]pyrene (97), but not for a thymine dimer lesion derived from UV light (98). Similar enzymatic capture mechanisms have been proposed for discrimination between thymine and uracil by the base excision repair enzyme uracil DNA glycosylase (95) and for alkylated DNA base recognition by an AlkD glycosylase (99).

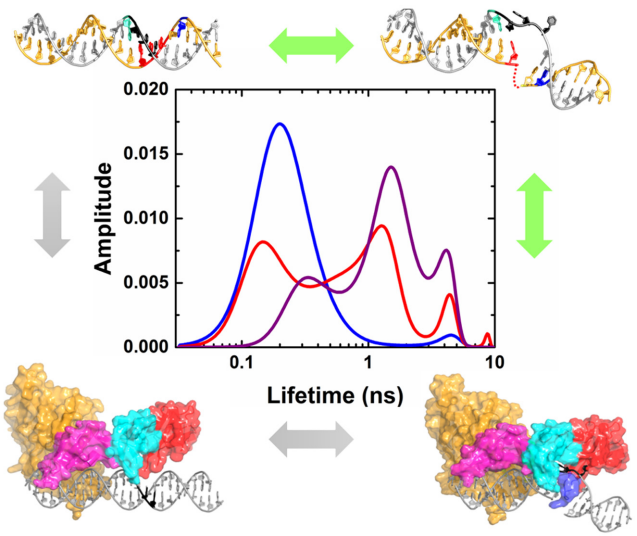

Figure 10. Conformational capture mechanism of DNA damage recognition by Rad4/XPC. Fluorescence lifetime studies on mismatched DNA provide evidence for thermal fluctuations in DNA, in the absence of Rad4, between B-DNA conformations similar to those in matched DNA and distorted DNA conformations that resemble protein-bound conformations. DNA binding to Rad4 further shifts the equilibrium toward the distorted conformations. These results suggest that conformational capture of pre-distorted DNA conformations (indicated by green arrows) is a plausible path for DNA damage recognition by Rad4/XPC in some cases. Fluorescence lifetime distributions are shown for the high specificity CCC/CCC construct (red: DNA alone; purple: Rad4–DNA complex) and for the matched GGG/CCC construct (blue).

MD simulations support the conformational capture mechanism for recognition of certain mismatches by Rad4

Our MD simulations helped provide molecular details underlying the experimental findings. The number of components in the fluorescence lifetime measurements for the mismatched duplexes is consistent with our observed number of C_A-C_D distance domains (Figure 9 and Supplementary Figures S14–16). We observed one C_A-C_D distance domain for the TTT/TTT and one for the TAT/TAT sequences. On a conformational level, the TTT/TTT mismatched duplex revealed a single family of structures. For the TAT/TAT case, however, two different hydrogen bonding schemes were observed for the A:A base pair within the same C_A-C_D distance domain. For the CCC/CCC case, there were three C_A-C_D distance domains that corresponded conformationally to very diverse and unstable structures in which mismatched C bases fluctuate between stacked-in and extruded states (Supplementary Figure S16). C/C mismatches have been shown to be one of the most helix-destabilizing single mismatches (100,101) and MD simulations on CC/CC double mismatches have also shown multiple conformations for the mismatches (102). Our simulations provide molecular interpretations of the observed binding data that reveals that the CCC/CCC duplex binds Rad4 with greater specificity than the other mismatched duplexes. The significant populations of extruded bases in the CCC/CCC duplex are states that are capture-ready for Rad4 binding. This role of extruded bases is similar to that of the extruded partner base C opposite the *cis*-B[*a*]P-dG lesion, which is conformationally captured by Rad4 in binding pathway studies (97). The severe destabilization of the CCC/CCC duplex observed in the melting studies is also consistent with the large population of extruded conformations. The binding studies also showed that TTT/TTT binds more specifically than the TAT/TAT duplex, which is also consistent with slightly more destabilizing TTT/TTT than TAT/TAT, from melting studies. We hypothesize that extrusion of mismatched T bases in the presence of Rad4, which is required for binding, would be more facile with neighboring pyrimidines than purines, which may explain the binding differences.

Specifically bound Rad4–DNA complexes are highly dynamic

It is noteworthy that, in solution, we observe a very broad range of conformations for DNA even after binding to Rad4 with high specificity (Figures 3, 6 and 7). The observed conformations span DNA structures that are not too perturbed from B-DNA all the way to highly distorted conformations with an apparent FRET efficiency near zero, close to the calculated FRET value (0.04) based on the crystal structures (Figure 1 and Supplementary Method 1.10). Intriguingly only a very small fraction of these broad conformations corresponded to this zero-FRET conformation expected of the crystal structure. These results suggest that the DNA conformations remain highly dynamic and heterogeneous even after binding to Rad4 and that structural constraints imposed by crystallization may be capturing only the highly distorted conformations that do not represent the entirety of the accessible conformations in solution. Furthermore, increasing the non-specific DNA sites by extending either sides of the CCC/CCC mismatch (thereby increasing the total length of DNA to 72-mer from 24-mer) did not diminish the broadly distributed low-FRET populations observed in the 24-mer DNA, indicating that contributions from non-specific binding are modest. Altogether, it is clear from our data that specific binding of Rad4 to CCC/CCC does not push the equilibrium into a narrow range of DNA conformations but rather that the specific complex remains highly dynamic. This dynamical behavior is consistent with the structural observation that Rad4 does not make very many specific contacts with the lesions themselves in the specific complex, although there may be specific interactions during interrogation *en route* to recognition. Though unexpected at first, the dynamic flexibility of the specific Rad4–DNA complexes may be key to understanding the biological functions of Rad4—how Rad4 can recognize a wide variety of structurally and chemically dissimilar NER lesions and recruit downstream factors. These notions and the dynamical nature of the specific complex merit further investigation.

Highly dynamic DNA conformations may stall a damage-searching protein, initiating lesion recognition

The efficiency with which DNA-binding proteins search and locate their target sites on long stretches of DNA is found to be optimized with some combination of 3-D diffusional search in bulk solution with 1-D diffusion along the DNA (103,104). These DNA-binding proteins, includ-

ing many examples of DNA repair proteins, have been observed *in vitro* to undergo relatively rapid 1-D diffusion as measured on long stretches of DNA, with the measured diffusion constants translating into residence times per DNA site that range from 0.1−300 μs (105–112). In contrast, site-specific recognition as measured on a few systems on which it has been directly observed occurs on timescales of a few milliseconds (6,8,10,12,13,113,114), which suggests the need for an additional mechanism by which these rapidly diffusing proteins can intermittently slow down. The protein-centric view of such a mechanism has invoked transient conformational switching from a rapidly diffusing 'search' mode to an intermittently stalled 'interrogation' mode (3,115,116). Indeed, a recent single-molecule study that examined the diffusional dynamics of Rad4 on DNA provided evidence for at least three different binding modes of the protein to DNA, a rapidly diffusing mode, a constrained diffusion mode, and a non-motile mode (117). The relative populations in these different binding modes changed when specific damage sites were introduced in the DNA, with increased probability in the non-motile mode at the expense of the randomly diffusing mode.

Previously, we had shown that Rad4, when tethered to undamaged DNA in a specific orientation, could flip out even undamaged nucleotide pairs in a manner very similar to how it flips out damaged nucleotide pairs (12). To reconcile these structural studies with the ability of Rad4 to discriminate between damaged and undamaged sites, we proposed a 'kinetic gating' mechanism for lesion discrimination, whereby the probability that a freely diffusing Rad4 would flip out damaged nucleotide pairs was predicted to be much higher than the probability of it flipping out undamaged nucleotide pairs (12). This differential probability, we argued, came from the kinetic competition between the rates of flipping out damaged/undamaged nucleotide pairs versus the rates at which the protein could diffuse away from that site. Although the rates at which nucleotides in undamaged DNA are flipped out by Rad4 have not yet been measured, free energy calculations suggest that the barrier for such motions is significantly (5–8 kcal/mol) higher than for damaged DNA (118), indicating that nucleotide flipping at undamaged sites could be sufficiently slow as to be improbable on the ∼0.6−180 μs residence times estimated for this protein, based on the single-molecule diffusion measurements (117).

Missing from these previous studies are the mechanism by which damaged sites stall and trap repair proteins as well as a measurement of how much the protein slows down as it stumbles upon a damaged site. Our results presented here strongly suggest that high conformational fluctuations accompanying DNA damage, as seen in the case of the high-specificity CCC/CCC construct, could severely impair the ability of Rad4 to diffuse, thus trapping it long enough to enable it to efficiently form the specific complex. Based on our previous measurements of ∼10 ms for full nucleotide-flipping by Rad4 at model lesion sites, we argue that intrinsic conformational fluctuations at these sites could increase the effective residence times of the protein at that site to a few milliseconds. In the absence of these highly unusual conformational fluctuations, as in the case of the TAT/TAT mismatch, the damaged site appears no different from undamaged DNA, and the protein likely wanders away before it has a chance to latch on and form the recognition complex. We envision that certain DNA lesions that are not recognized and repaired efficiently by NER in cells may share similarity with the TAT/TAT mismatch, with conformational distortions and fluctuations significantly suppressed in comparison with those repaired efficiently. Further investigation of the conformational distributions of DNA with real NER damage, with the kinds of approaches used here, will shed further light on the role that intrinsic DNA conformational dynamics plays in damage recognition mechanisms. Repair-resistant lesions tend to persist through cell cycles and thus are more mutagenic than repair-proficient ones (60–64). If intrinsic DNA deformability and/or distortions in damaged DNA indeed correlate with efficient recognition by Rad4/XPC and repair proficiency, then this sensitive approach could serve as a useful tool for predicting the Rad4/XPC binding propensities of various *bona fide* DNA lesions for NER and help identify the mechanisms underlying persistent DNA lesions that could also serve as biomarkers for cancers.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Saroj Baral for his help with melting studies of DNA constructs in this study. We also thank Marcus Wilhelmsson and Søren Preus for their assistance in calculating the expected FRET values from their FRETMATRIX program.

FUNDING

National Science Foundation (NSF) [MCB-1158217 to A.A., MCB-1412692 to J.-H.M]; LAS Award for Faculty in the Sciences from the University of Illinois at Chicago (to A.A., J.-H.M); National Institutes of Health [R01-ES025987, R01-CA75449 to S.B.]; Intramural Research Program of the National Institutes of Health (NIH, CIT); NSF Grant [MCB060037 to S.B.]; High performance computing resources of New York University (NYU-ITS). Funding for open access charge: NSF [MCB-1158217, MCB-1412692].
Conflict of interest statement. None declared.

REFERENCES

1. Kalodimos,C.G., Biris,N., Bonvin,A.M., Levandoski,M.M., Guennuegues,M., Boelens,R. and Kaptein,R. (2004) Structure and flexibility adaptation in nonspecific and specific protein-DNA complexes. *Science*, **305**, 386–389.
2. Iwahara,J., Zweckstetter,M. and Clore,G.M. (2006) NMR structural and kinetic characterization of a homeodomain diffusing and hopping on nonspecific DNA. *Proc. Natl. Acad. Sci. U.S.A.*, **103**, 15062–15067.
3. Friedman,J.I., Majumdar,A. and Stivers,J.T. (2009) Nontarget DNA binding shapes the dynamic landscape for enzymatic recognition of DNA damage. *Nucleic Acids Res.*, **37**, 3493–3500.
4. Stivers,J.T., Pankiewicz,K.W. and Watanabe,K.A. (1999) Kinetic mechanism of damage site recognition and uracil flipping by Escherichia coli uracil DNA glycosylase. *Biochemistry*, **38**, 952–963.

5. Sugimura,S. and Crothers,D.M. (2006) Stepwise binding and bending of DNA by *Escherichia coli* integration host factor. *Proc. Natl. Acad. Sci. U.S.A.*, **103**, 18510–18514.
6. Sharma,A., Doucette,C., Biro,F.N. and Hingorani,M.M. (2013) Slow conformational changes in MutS and DNA direct ordered transitions between mismatch search, recognition and signaling of DNA repair. *J. Mol. Biol.*, **425**, 4192–4205.
7. van den Broek,B., Noom,M.C. and Wuite,G.J. (2005) DNA-tension dependence of restriction enzyme activity reveals mechanochemical properties of the reaction pathway. *Nucleic Acids Res.*, **33**, 2676–2684.
8. Kuznetsov,S.V., Kozlov,A.G., Lohman,T.M. and Ansari,A. (2006) Microsecond dynamics of protein-DNA interactions: direct observation of the wrapping/unwrapping kinetics of single-stranded DNA around the *E. coli* SSB tetramer. *J. Mol. Biol.*, **359**, 55–65.
9. Vivas,P., Kuznetsov,S.V. and Ansari,A. (2008) New insights into the transition pathway from nonspecific to specific complex of DNA with Escherichia coli integration host factor. *J. Phys. Chem. B*, **112**, 5997–6007.
10. Vivas,P., Velmurugu,Y., Kuznetsov,S.V., Rice,P.A. and Ansari,A. (2012) Mapping the Transition State for DNA Bending by IHF. *J. Mol. Biol.*, **418**, 300–315.
11. Ansari,A. and Kuznetsov,S.V. (2010) Dynamics and mechanism of DNA-bending proteins in binding site recognition. In: Williams,MC and Maher,LJ (eds). *Biophysics of DNA-Protein Interactions*. Springer, NY, pp. 107–142.
12. Chen,X., Velmurugu,Y., Zheng,G., Park,B., Shim,Y., Kim,Y., Liu,L., Van Houten,B., He,C., Ansari,A. *et al.* (2015) Kinetic gating mechanism of DNA damage recognition by Rad4/XPC. *Nat. Commun.*, **6**, 5849.
13. Velmurugu,Y., Chen,X., Slogoff Sevilla,P., 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.
14. Lawson,C.L. and Berman,H.M. (2008) Indirect readout of DNA sequence by proteins. In: Rice,PA and Correll,CC (eds). *Protein-Nucleic Acid Interactions*. Royal Society of Chemistry, Cambridge, pp. 66–90.
15. Klostermeier,D. and Millar,D.P. (2001) Time-resolved fluorescence resonance energy transfer: a versatile tool for the analysis of nucleic acids. *Biopolymers*, **61**, 159–179.
16. Roy,R., Kozlov,A.G., Lohman,T.M. and Ha,T. (2009) SSB protein diffusion on single-stranded DNA stimulates RecA filament formation. *Nature*, **461**, 1092–1097.
17. Sass,L.E., Lanyi,C., Weninger,K. and Erie,D.A. (2010) Single-molecule FRET TACKLE reveals highly dynamic mismatched DNA-MutS complexes. *Biochemistry*, **49**, 3174–3190.
18. Pirchi,M., Ziv,G., Riven,I., Cohen,S.S., Zohar,N., Barak,Y. and Haran,G. (2011) Single-molecule fluorescence spectroscopy maps the folding landscape of a large protein. *Nat. Commun.*, **2**, 493.
19. Schuler,B., Lipman,E.A. and Eaton,W.A. (2002) Probing the free-energy surface for protein folding with single-molecule fluorescence spectroscopy. *Nature*, **419**, 743–747.
20. Xu,J., Toptygin,D., Graver,K.J., Albertini,R.A., Savtchenko,R.S., Meadow,N.D., Roseman,S., Callis,P.R., Brand,L. and Knutson,J.R. (2006) Ultrafast fluorescence dynamics of tryptophan in the proteins monellin and IIAGlc. *J. Am. Chem. Soc.*, **128**, 1214–1221.
21. Lin,S., Horning,D.P., Szostak,J.W. and Chaput,J.C. (2009) Conformational analysis of DNA repair intermediates by time-resolved fluorescence spectroscopy. *J. Phys. Chem. A*, **113**, 9585–9587.
22. Neely,R.K., Tamulaitis,G., Chen,K., Kubala,M., Siksnys,V. and Jones,A.C. (2009) Time-resolved fluorescence studies of nucleotide flipping by restriction enzymes. *Nucleic Acids Res.*, **37**, 6859–6870.
23. Moreno,A., Knee,J. and Mukerji,I. (2012) Applying 6-methylisoxanthopterin-enhanced fluorescence to examine protein-DNA interactions in the picomolar range. *Biochemistry*, **51**, 6847–6859.
24. Borjesson,K., Preus,S., El-Sagheer,A.H., Brown,T., Albinsson,B. and Wilhelmsson,L.M. (2009) Nucleic acid base analog FRET-pair facilitating detailed structural measurements in nucleic acid containing systems. *J. Am. Chem. Soc.*, **131**, 4288–4293.
25. Preus,S., Borjesson,K., Kilsa,K., Albinsson,B. and Wilhelmsson,L.M. (2010) Characterization of nucleobase analogue FRET acceptor tCnitro. *J. Phys. Chem. B*, **114**, 1050–1056.
26. Shi,Y., Dierckx,A., Wanrooij,P.H., Wanrooij,S., Larsson,N.G., Wilhelmsson,L.M., Falkenberg,M. and Gustafsson,C.M. (2012) Mammalian transcription factor A is a core component of the mitochondrial transcription machinery. *Proc. Natl. Acad. Sci. U.S.A.*, **109**, 16510–16515.
27. Xia,S., Wood,M., Bradley,M.J., De La Cruz,E.M. and Konigsberg,W.H. (2013) Alteration in the cavity size adjacent to the active site of RB69 DNA polymerase changes its conformational dynamics. *Nucleic Acids Res.*, **41**, 9077–9089.
28. Posse,V., Hoberg,E., Dierckx,A., Shahzad,S., Koolmeister,C., Larsson,N.G., Wilhelmsson,L.M., Hallberg,B.M. and Gustafsson,C.M. (2014) The amino terminal extension of mammalian mitochondrial RNA polymerase ensures promoter specific transcription initiation. *Nucleic Acids Res.*, **42**, 3638–3647.
29. Dumat,B., Larsen,A.F. and Wilhelmsson,L.M. (2016) Studying Z-DNA and B- to Z-DNA transitions using a cytosine analogue FRET-pair. *Nucleic Acids Res.*, **44**, e101.
30. Gillet,L.C. and Scharer,O.D. (2006) Molecular mechanisms of mammalian global genome nucleotide excision repair. *Chem. Rev.*, **106**, 253–276.
31. Puumalainen,M.R., Ruthemann,P., Min,J.H. and Naegeli,H. (2016) Xeroderma pigmentosum group C sensor: unprecedented recognition strategy and tight spatiotemporal regulation. *Cell Mol. Life Sci.*, **73**, 547–566.
32. Uchida,A., Sugasawa,K., Masutani,C., Dohmae,N., Araki,M., Yokoi,M., Ohkuma,Y. and Hanaoka,F. (2002) The carboxy-terminal domain of the XPC protein plays a crucial role in nucleotide excision repair through interactions with transcription factor IIH. *DNA Repair (Amst)*, **1**, 449–461.
33. Yokoi,M., Masutani,C., Maekawa,T., Sugasawa,K., Ohkuma,Y. and Hanaoka,F. (2000) The xeroderma pigmentosum group C protein complex XPC-HR23B plays an important role in the recruitment of transcription factor IIH to damaged DNA. *J. Biol. Chem.*, **275**, 9870–9875.
34. Sugasawa,K., Ng,J.M., Masutani,C., Iwai,S., van der Spek,P.J., Eker,A.P., Hanaoka,F., Bootsma,D. and Hoeijmakers,J.H. (1998) Xeroderma pigmentosum group C protein complex is the initiator of global genome nucleotide excision repair. *Mol. Cell*, **2**, 223–232.
35. Riedl,T., Hanaoka,F. and Egly,J.M. (2003) The comings and goings of nucleotide excision repair factors on damaged DNA. *EMBO J.*, **22**, 5293–5303.
36. Friedberg,E.C. (2001) How nucleotide excision repair protects against cancer. *Nat. Rev.*, **1**, 22–33.
37. Friedberg,E.C. (2006) *DNA Repair and Mutagenesis*. 2nd edn. ASM Press, Washington, D.C.
38. Heflich,R.H. and Neft,R.E. (1994) Genetic toxicity of 2-acetylaminofluorene, 2-aminofluorene and some of their metabolites and model metabolites. *Mutat. Res.*, **318**, 73–114.
39. Bostrom,C.E., Gerde,P., Hanberg,A., Jernstrom,B., Johansson,C., Kyrklund,T., Rannug,A., Tornqvist,M., Victorin,K. and Westerholm,R. (2002) Cancer risk assessment, indicators, and guidelines for polycyclic aromatic hydrocarbons in the ambient air. *Environ. Health Perspect.*, **110**(Suppl. 3), 451–488.
40. Hecht,S.S. (2017) Oral cell DNA adducts as potential biomarkers for lung cancer susceptibility in cigarette smokers. *Chem. Res. Toxicol.*, **30**, 367–375.
41. Turesky,R.J. (2010) Aromatic amines and heterocyclic aromatic amines: from tobacco smoke to food mutagens. In: Broyde,S and Geacintov,NE (eds). *The Chemical Biology of DNA Damage*. Wiley-VCH, Weinheim, pp. 157–184.
42. Geacintov,N.E. and Broyde,S. (2017) Repair-resistant DNA lesions. *Chem. Res. Toxicol.*, **30**, 1517–1548.
43. Sugasawa,K., Okamoto,T., Shimizu,Y., Masutani,C., Iwai,S. and Hanaoka,F. (2001) A multistep damage recognition mechanism for global genomic nucleotide excision repair. *Genes Dev.*, **15**, 507–521.
44. Min,J.H. and Pavletich,N.P. (2007) Recognition of DNA damage by the Rad4 nucleotide excision repair protein. *Nature*, **449**, 570–575.
45. Wood,R.D. (1999) DNA damage recognition during nucleotide excision repair in mammalian cells. *Biochimie*, **81**, 39–44.
46. Cai,Y., Patel,D.J., Geacintov,N.E. and Broyde,S. (2009) Differential nucleotide excision repair susceptibility of bulky DNA adducts in

different sequence contexts: hierarchies of recognition signals. *J. Mol. Biol.*, **385**, 30–44.
47. Kropachev,K., Kolbanovskii,M., Cai,Y., Rodriguez,F., Kolbanovskii,A., Liu,Y., Zhang,L., Amin,S., Patel,D., Broyde,S. *et al.* (2009) The sequence dependence of human nucleotide excision repair efficiencies of benzo[a]pyrene-derived DNA lesions: insights into the structural factors that favor dual incisions. *J. Mol. Biol.*, **386**, 1193–1203.
48. Cai,Y., Kropachev,K., Xu,R., Tang,Y., Kolbanovskii,M., Kolbanovskii,A., Amin,S., Patel,D.J., Broyde,S. and Geacintov,N.E. (2010) Distant neighbor base sequence context effects in human nucleotide excision repair of a benzo[a]pyrene-derived DNA lesion. *J. Mol. Biol.*, **399**, 397–409.
49. Mu,H., Kropachev,K., Wang,L., Zhang,L., Kolbanovskiy,A., Kolbanovskiy,M., Geacintov,N.E. and Broyde,S. (2012) Nucleotide excision repair of 2-acetylaminofluorene- and 2-aminofluorene-(C8)-guanine adducts: molecular dynamics simulations elucidate how lesion structure and base sequence context impact repair efficiencies. *Nucleic Acids Res.*, **40**, 9675–9690.
50. Batty,D., Rapic'-Otrin,V., Levine,A.S. and Wood,R.D. (2000) Stable binding of human XPC complex to irradiated DNA confers strong discrimination for damaged sites. *J. Mol. Biol.*, **300**, 275–290.
51. Hey,T., Lipps,G., Sugasawa,K., Iwai,S., Hanaoka,F. and Krauss,G. (2002) The XPC-HR23B complex displays high affinity and specificity for damaged DNA in a true-equilibrium fluorescence assay. *Biochemistry*, **41**, 6583–6587.
52. Kusumoto,R., Masutani,C., Sugasawa,K., Iwai,S., Araki,M., Uchida,A., Mizukoshi,T. and Hanaoka,F. (2001) Diversity of the damage recognition step in the global genomic nucleotide excision repair in vitro. *Mutat. Res.*, **485**, 219–227.
53. Reardon,J.T., Mu,D. and Sancar,A. (1996) Overproduction, purification, and characterization of the XPC subunit of the human DNA repair excision nuclease. *J. Biol. Chem.*, **271**, 19451–19456.
54. Bunick,C.G., Miller,M.R., Fuller,B.E., Fanning,E. and Chazin,W.J. (2006) Biochemical and structural domain analysis of xeroderma pigmentosum complementation group C protein. *Biochemistry*, **45**, 14965–14979.
55. Trego,K.S. and Turchi,J.J. (2006) Pre-steady-state binding of damaged DNA by XPC-hHR23B reveals a kinetic mechanism for damage discrimination. *Biochemistry*, **45**, 1961–1969.
56. Ding,S., Kropachev,K., Cai,Y., Kolbanovskiy,M., Durandina,S.A., Liu,Z., Shafirovich,V., Broyde,S. and Geacintov,N.E. (2011) Structural, energetic and dynamic properties of guanine(C8)-thymine(N3) cross-links in DNA provide insights on susceptibility to nucleotide excision repair. *Nucleic Acids Res.*, **40**, 2506–2517.
57. Cai,Y., Patel,D.J., Geacintov,N.E. and Broyde,S. (2007) Dynamics of a benzo[a]pyrene-derived guanine DNA lesion in TGT and CGC sequence contexts: enhanced mobility in TGT explains conformational heterogeneity, flexible bending, and greater susceptibility to nucleotide excision repair. *J. Mol. Biol.*, **374**, 292–305.
58. Sidorenko,V.S., Yeo,J.E., Bonala,R.R., Johnson,F., Scharer,O.D. and Grollman,A.P. (2012) Lack of recognition by global-genome nucleotide excision repair accounts for the high mutagenicity and persistence of aristolactam-DNA adducts. *Nucleic Acids Res.*, **40**, 2494–2505.
59. Kropachev,K., Kolbanovskiy,M., Liu,Z., Cai,Y., Zhang,L., Schwaid,A.G., Kolbanovskiy,A., Ding,S., Amin,S., Broyde,S. *et al.* (2013) Adenine-DNA adducts derived from the highly tumorigenic Dibenzo[a,l]pyrene are resistant to nucleotide excision repair while guanine adducts are not. *Chem. Res. Toxicol.*, **26**, 783–793.
60. Phillips,D.H., Hewer,A., Seidel,A., Steinbrecher,T., Schrode,R., Oesch,F. and Glatt,H. (1991) Relationship between mutagenicity and DNA adduct formation in mammalian cells for fjord- and bay-region diol-epoxides of polycyclic aromatic hydrocarbons. *Chem. Biol. Interact.*, **80**, 177–186.
61. Amin,S., Krzeminski,J., Rivenson,A., Kurtzke,C., Hecht,S.S. and el-Bayoumy,K. (1995) Mammary carcinogenicity in female CD rats of fjord region diol epoxides of benzo[c]phenanthrene, benzo[g]chrysene and dibenzo[a,l]pyrene. *Carcinogenesis*, **16**, 1971–1974.
62. Leavitt,S.A., George,M.H., Moore,T. and Ross,J.A. (2008) Mutations induced by benzo[a]pyrene and dibenzo[a,l]pyrene in lacI transgenic B6C3F1 mouse lung result from stable DNA adducts. *Mutagenesis*, **23**, 445–450.
63. Cai,Y., Kropachev,K., Kolbanovskii,M., Kolbanovskii,A., Broyde,S., Patel,D. and Geacintov,N.E. (2010) Recognition and removal of bulky DNA lesions by the nucleotide excision repair system. In: Geacintov,NE and Broyde,S (eds). *The Chemical Biology of DNA Damage*. WILEY-VCH VerlagGmbH & Co., Weinheim, pp. 261–298.
64. Dreij,K., Seidel,A. and Jernstrom,B. (2005) Differential removal of DNA adducts derived from anti-diol epoxides of dibenzo[a,l]pyrene and benzo[a]pyrene in human cells. *Chem. Res. Toxicol.*, **18**, 655–664.
65. Rosenquist,T.A. and Grollman,A.P. (2016) Mutational signature of aristolochic acid: Clue to the recognition of a global disease. *DNA Repair (Amst)*, **44**, 205–211.
66. Buterin,T., Meyer,C., Giese,B. and Naegeli,H. (2005) DNA quality control by conformational readout on the undamaged strand of the double helix. *Chem. Biol.*, **12**, 913–922.
67. O'Connor,D.V. (1984) *Time-correlated Single Photon Counting*. Academic Press, London.
68. Steinbach,P.J., Ionescu,R. and Matthews,C.R. (2002) Analysis of kinetics using a hybrid maximum-entropy/nonlinear-least-squares method: application to protein folding. *Biophys. J.*, **82**, 2244–2255.
69. Steinbach,P.J. (2012) Filtering artifacts from lifetime distributions when maximizing entropy using a bootstrapped model. *Anal. Biochem.*, **427**, 102–105.
70. Case,D.A., Babin,V., Berryman,J.T., Betz,R.M., Cai,Q., Cerutti,D.S., Cheatham,T.E. III, Darden,T.A., Duke,R.E., Gohlke,H. *et al.* (2014) *AMBER14*. University of California, San Francisco, CA.
71. Maier,J.A., Martinez,C., Kasavajhala,K., Wickstrom,L., Hauser,K.E. and Simmerling,C. (2015) ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. *J. Chem. Theory Comput.*, **11**, 3696–3713.
72. Skjaerven,L., Yao,X.Q., Scarabelli,G. and Grant,B.J. (2014) Integrating protein structural dynamics and evolutionary analysis with Bio3D. *BMC Bioinformatics*, **15**, 399–409.
73. R Core Team (2017) R. A language and environment for statistical computing. *R Foundation for Statistical Computing*, Vienna, https://www.R-project.org/.
74. Sandin,P., Borjesson,K., Li,H., Martensson,J., Brown,T., Wilhelmsson,L.M. and Albinsson,B. (2008) Characterization and use of an unprecedentedly bright and structurally non-perturbing fluorescent DNA base analogue. *Nucleic Acids Res.*, **36**, 157–167.
75. Provencher,S.W. (1992) Low-bias macroscopic analysis of polydispersity. In: Harding,SE, Sattelle,DB and Bloomfield,VA (eds). *Laser Light Scattering in Biochemistry*. The Royal Society of Chemistry, Cambridge, pp. 92–111.
76. Livesey,A.K. and Brochon,J.C. (1987) Analyzing the distribution of decay constants in pulse-fluorimetry using the maximum entropy method. *Biophys. J.*, **52**, 693–706.
77. Steinbach,P.J. (2002) Inferring lifetime distributions from kinetics by maximizing entropy using a bootstrapped model. *J. Chem. Inf. Comput. Sci.*, **42**, 1476–1478.
78. Preus,S., Kilsa,K., Miannay,F.A., Albinsson,B. and Wilhelmsson,L.M. (2013) FRETmatrix: a general methodology for the simulation and analysis of FRET in nucleic acids. *Nucleic Acids Res.*, **41**, e18.
79. Isaacs,R.J. and Spielmann,H.P. (2004) A model for initial DNA lesion recognition by NER and MMR based on local conformational flexibility. *DNA Repair (Amst)*, **3**, 455–464.
80. Yang,W. (2006) Poor base stacking at DNA lesions may initiate recognition by many repair proteins. *DNA Repair (Amst)*, **5**, 654–666.
81. Moe,J.G. and Russu,I.M. (1992) Kinetics and energetics of base-pair opening in 5′-d(CGCGAATTCGCG)-3′ and a substituted dodecamer containing G.T mismatches. *Biochemistry*, **31**, 8421–8428.
82. Cho,B.P., Beland,F.A. and Marques,M.M. (1994) NMR structural studies of a 15-mer DNA duplex from a ras protooncogene modified with the carcinogen 2-aminofluorene: conformational heterogeneity. *Biochemistry*, **33**, 1373–1384.
83. Schwartz,J.L., Rice,J.S., Luxon,B.A., Sayer,J.M., Xie,G., Yeh,H.J., Liu,X., Jerina,D.M. and Gorenstein,D.G. (1997) Solution structure

of the minor conformer of a DNA duplex containing a dG mismatch opposite a benzo[a]pyrene diol epoxide/dA adduct: glycosidic rotation from syn to anti at the modified deoxyadenosine. *Biochemistry*, **36**, 11069–11076.
84. Pradhan,P., Tirumala,S., Liu,X., Sayer,J.M., Jerina,D.M. and Yeh,H.J. (2001) Solution structure of a trans-opened (10S)-dA adduct of (+)-(7S,8R,9S,10R)-7,8-dihydroxy-9,10-epoxy-7,8,9,10-tetrahydrobenzo[a]pyrene in a fully complementary DNA duplex: evidence for a major syn conformation. *Biochemistry*, **40**, 5870–5881.
85. Isaacs,R.J., Rayens,W.S. and Spielmann,H.P. (2002) Structural differences in the NOE-derived structure of G-T mismatched DNA relative to normal DNA are correlated with differences in (13)C relaxation-based internal dynamics. *J. Mol. Biol.*, **319**, 191–207.
86. Bhattacharya,P.K., Cha,J. and Barton,J.K. (2002) 1H NMR determination of base-pair lifetimes in oligonucleotides containing single base mismatches. *Nucleic Acids Res.*, **30**, 4740–4750.
87. Cao,C., Jiang,Y.L., Krosky,D.J. and Stivers,J.T. (2006) The catalytic power of uracil DNA glycosylase in the opening of thymine base pairs. *J. Am. Chem. Soc.*, **128**, 13034–13035.
88. Nag,N., Rao,B.J. and Krishnamoorthy,G. (2007) Altered dynamics of DNA bases adjacent to a mismatch: a cue for mismatch recognition by MutS. *J. Mol. Biol.*, **374**, 39–53.
89. Geacintov,N.E., Broyde,S., Buterin,T., Naegeli,H., Wu,M., Yan,S. and Patel,D.J. (2002) Thermodynamic and structural factors in the removal of bulky DNA adducts by the nucleotide excision repair machinery. *Biopolymers*, **65**, 202–210.
90. Jing,Y., Kao,J.F. and Taylor,J.S. (1998) Thermodynamic and base-pairing studies of matched and mismatched DNA dodecamer duplexes containing cis-syn, (6-4) and Dewar photoproducts of TT. *Nucleic Acids Res.*, **26**, 3845–3853.
91. Ruan,Q., Kolbanovskiy,A., Zhuang,P., Chen,J., Krzeminski,J., Amin,S. and Geacintov,N.E. (2002) Synthesis and characterization of site-specific and stereoisomeric fjord dibenzo[a,l]pyrene diol epoxide-N(6)-adenine adducts: unusual thermal stabilization of modified DNA duplexes. *Chem. Res. Toxicol.*, **15**, 249–261.
92. Cai,Y., Geacintov,N.E. and Broyde,S. (2012) Nucleotide excision repair efficiencies of bulky carcinogen-DNA adducts are governed by a balance between stabilizing and destabilizing interactions. *Biochemistry*, **51**, 1486–1499.
93. Ruan,Q., Liu,T., Kolbanovskiy,A., Liu,Y., Ren,J., Skorvaga,M., Zou,Y., Lader,J., Malkani,B., Amin,S. *et al.* (2007) Sequence context- and temperature-dependent nucleotide excision repair of a benzo[a]pyrene diol epoxide-guanine DNA adduct catalyzed by thermophilic UvrABC proteins. *Biochemistry*, **46**, 7006–7015.
94. Rodriguez,F.A., Cai,Y., Lin,C., Tang,Y., Kolbanovskiy,A., Amin,S., Patel,D.J., Broyde,S. and Geacintov,N.E. (2007) Exocyclic amino groups of flanking guanines govern sequence-dependent adduct conformations and local structural distortions for minor groove-aligned benzo[a]pyrenyl-guanine lesions in a GG mutation hotspot context. *Nucleic Acids Res.*, **35**, 1555–1568.
95. Parker,J.B., Bianchet,M.A., Krosky,D.J., Friedman,J.I., Amzel,L.M. and Stivers,J.T. (2007) Enzymatic capture of an extrahelical thymine in the search for uracil in DNA. *Nature*, **449**, 433–437.
96. 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.
97. 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.
98. 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.
99. Rubinson,E.H., Gowda,A.S., Spratt,T.E., Gold,B. and Eichman,B.F. (2010) An unprecedented nucleic acid capture mechanism for excision of DNA damage. *Nature*, **468**, 406–411.
100. Peyret,N., Seneviratne, Allawi,P.A., SantaLucia,H.T. and Jr,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.
101. 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.
102. Zhang,Y., Roland,C. and Sagui,C. (2017) Structure and Dynamics of DNA and RNA Double Helices Obtained from the GGGGCC and CCCCGG Hexanucleotide Repeats That Are the Hallmark of C9FTD/ALS Diseases. *ACS Chem. Neurosci.*, **8**, 578–591.
103. Halford,S.E. and Marko,J.F. (2004) How do site-specific DNA-binding proteins find their targets? *Nucleic Acids Res.*, **32**, 3040–3052.
104. Porecha,R.H. and Stivers,J.T. (2008) Uracil DNA glycosylase uses DNA hopping and short-range sliding to trap extrahelical uracils. *Proc. Natl. Acad. Sci. U.S.A.*, **105**, 10791–10796.
105. Wang,Y.M., Austin,R.H. and Cox,E.C. (2006) Single molecule measurements of repressor protein 1D diffusion on DNA. *Phys. Rev. Lett.*, **97**, 048302.
106. Blainey,P.C., van Oijen,A.M., Banerjee,A., Verdine,G.L. and Xie,X.S. (2006) A base-excision DNA-repair protein finds intrahelical lesion bases by fast sliding in contact with DNA. *Proc. Natl. Acad. Sci. U.S.A.*, **103**, 5752–5757.
107. Gorman,J., Chowdhury,A., Surtees,J.A., Shimada,J., Reichman,D.R., Alani,E. and Greene,E.C. (2007) Dynamic basis for one-dimensional DNA scanning by the mismatch repair complex Msh2-Msh6. *Mol. Cell*, **28**, 359–370.
108. Tafvizi,A., Huang,F., Leith,J.S., Fersht,A.R., Mirny,L.A. and van Oijen,A.M. (2008) Tumor suppressor p53 slides on DNA with low friction and high stability. *Biophys. J.*, **95**, L01–L03.
109. Bonnet,I., Biebricher,A., Porte,P.L., Loverdo,C., Benichou,O., Voituriez,R., Escude,C., Wende,W., Pingoud,A. and Desbiolles,P. (2008) Sliding and jumping of single EcoRV restriction enzymes on non-cognate DNA. *Nucleic Acids Res.*, **36**, 4118–4127.
110. Gorman,J., Plys,A.J., Visnapuu,M.L., Alani,E. and Greene,E.C. (2010) Visualizing one-dimensional diffusion of eukaryotic DNA repair factors along a chromatin lattice. *Nat. Struct. Mol. Biol.*, **17**, 932–938.
111. Schonhoft,J.D. and Stivers,J.T. (2012) Timing facilitated site transfer of an enzyme on DNA. *Nat. Chem. Biol.*, **8**, 205–210.
112. Nelson,S.R., Dunn,A.R., Kathe,S.D., Warshaw,D.M. and Wallace,S.S. (2014) Two glycosylase families diffusively scan DNA using a wedge residue to probe for and identify oxidatively damaged bases. *Proc. Natl. Acad. Sci. U.S.A.*, **111**, E2091–E2099.
113. van den Broek,B., Noom,M.C. and Wuite,G.J. (2005) DNA-tension dependence of restriction enzyme activity reveals mechanochemical properties of the reaction pathway. *Nucleic Acids Res.*, **33**, 2676–2684.
114. Friedman,J.I. and Stivers,J.T. (2010) Detection of damaged DNA bases by DNA glycosylase enzymes. *Biochemistry*, **49**, 4957–4967.
115. Slutsky,M. and Mirny,L.A. (2004) Kinetics of protein-DNA interaction: facilitated target location in sequence-dependent potential. *Biophys. J.*, **87**, 4021–4035.
116. Zhou,H.X. (2011) Rapid search for specific sites on DNA through conformational switch of nonspecifically bound proteins. *Proc. Natl. Acad. Sci. U.S.A.*, **108**, 8651–8656.
117. Kong,M., Liu,L., Chen,X., Driscoll,K.I., Mao,P., Bohm,S., Kad,N.M., Watkins,S.C., Bernstein,K.A., Wyrick,J.J. *et al.* (2016) Single-molecule imaging reveals that Rad4 employs a dynamic DNA damage recognition process. *Mol. Cell*, **64**, 376–387.
118. Zheng,H., Cai,Y., Ding,S., Tang,Y., Kropachev,K., Zhou,Y., Wang,L., Wang,S., Geacintov,N.E., Zhang,Y. *et al.* (2010) Base flipping free energy profiles for damaged and undamaged DNA. *Chem. Res. Toxicol.*, **23**, 1868–1870.