

ARTICLE

Received 14 Apr 2014 | Accepted 13 Nov 2014 | Published 6 Jan 2015

DOI: 10.1038/ncomms6849

OPEN

# Kinetic gating mechanism of DNA damage recognition by Rad4/XPC

Xuejing Chen<sup>1,\*</sup>, Yogambigai Velmurugu<sup>2,\*</sup>, Guanqun Zheng<sup>3</sup>, Beomseok Park<sup>1,†</sup>, Yoonjung Shim<sup>1</sup>, Youngchang Kim<sup>4</sup>, Lili Liu<sup>5</sup>, Bennett Van Houten<sup>5</sup>, Chuan He<sup>3</sup>, Anjum Ansari<sup>2,6</sup> & Jung-Hyun Min<sup>1</sup>

The xeroderma pigmentosum C (XPC) complex initiates nucleotide excision repair by recognizing DNA lesions before recruiting downstream factors. How XPC detects structurally diverse lesions embedded within normal DNA is unknown. Here we present a crystal structure that captures the yeast XPC orthologue (Rad4) on a single register of undamaged DNA. The structure shows that a disulphide-tethered Rad4 flips out normal nucleotides and adopts a conformation similar to that seen with damaged DNA. Contrary to many DNA repair enzymes that can directly reject non-target sites as structural misfits, our results suggest that Rad4/XPC uses a kinetic gating mechanism whereby lesion selectivity arises from the kinetic competition between DNA opening and the residence time of Rad4/XPC per site. This mechanism is further supported by measurements of Rad4-induced lesion-opening times using temperature-jump perturbation spectroscopy. Kinetic gating may be a general mechanism used by site-specific DNA-binding proteins to minimize time-consuming interrogations of non-target sites.

<sup>1</sup>Department of Chemistry, University of Illinois at Chicago, 845 W. Taylor Street, Chicago, Illinois 60607, USA. <sup>2</sup>Department of Physics, University of Illinois at Chicago, 845 W. Taylor Street, Chicago, Illinois 60607, USA. <sup>3</sup>Department of Chemistry, Institute for Biophysical Dynamics, The University of Chicago, 929 E. 57th Street, Chicago, Illinois 60637, USA. <sup>4</sup>Structural Biology Center, Biosciences Division, Argonne National Laboratory, 9700 S. Cass Avenue, Argonne, Illinois 60439, USA. <sup>5</sup>Department of Pharmacology and Chemical Biology, University of Pittsburgh School of Medicine and University of Pittsburgh Cancer Institute, University of Pittsburgh, 5117 Centre Avenue, Pittsburgh, Pennsylvania 15213, USA. <sup>6</sup>Department of Bioengineering, University of Illinois at Chicago, 845 W. Taylor Street, Chicago, Illinois 60607, USA. \* These authors contributed equally to this work. † Present address: Department of Biomedical Laboratory Science, Eulji University, Seongnam 461-713, Korea. Correspondence and requests for materials should be addressed to A.A. (email: ansari@uic.edu) or to J.-H.M. (email: jhmin@uic.edu).

Control and maintenance of the genome is initiated by proteins that recognize and assemble at specific target sites on DNA. These proteins can single out a small number of target sites from a vast excess of closely related non-target sites. How proteins solve this so-called needle-in-the-haystack problem remains a central question in biology<sup>1</sup>. The mammalian global genome nucleotide excision repair (NER) factor, xeroderma pigmentosum C (XPC)–RAD23B complex (hereafter XPC), in particular, faces a difficult challenge. NER repairs diverse helix-distorting/destabilizing DNA damage caused by environmental insults<sup>2</sup>. Impaired NER leads to extreme sun sensitivity and skin cancer predisposition<sup>2</sup>. To initiate NER, XPC needs to recognize an extraordinarily wide variety of DNA lesions based on thermal fluctuations alone. Once bound to a lesion, XPC recruits the multi-subunit transcription factor IIH (TFIIH) that verifies the lesion<sup>3</sup>, which ultimately coordinates the assembly of NER factors, causing excision of a lesion-containing single-stranded DNA and repair synthesis that restores the duplex<sup>2</sup>. Although the lesion-recognition step by XPC is considered the rate-limiting step of NER<sup>4,5</sup>, much remains unknown as to how XPC achieves its specific binding efficiently and reliably.

DNA lesions that XPC recognizes include a wide variety of intra-strand crosslinks and helix-distorting adducts, which are formed by ultraviolet light (UV), air and water pollutants, and toxins<sup>2</sup>. XPC efficiently localizes to and recognizes these lesions despite its high affinity for undamaged DNA as measured by *in vitro* assays<sup>6–8</sup>. Notably, XPC resides in DNA-rich chromatin in undamaged cells<sup>9</sup>, which is unique among other mammalian NER proteins<sup>10,11</sup>. *In vitro*, XPC can also recognize artificially destabilized DNA such as 2- to 3-base-pair (bp) mismatch bubble (refs 8,12 and this study), although such mismatches fail the verification step by TFIIH and thus are not excised by NER.

The previous crystal structures of the yeast XPC orthologue, Rad4, bound to model lesions have shown that Rad4 and, by analogy, XPC recognize the damaged DNA site by inserting a  $\beta$ -hairpin into DNA duplex and flipping out damage-containing nucleotide pairs to form an ‘open’ conformation<sup>8</sup>. In this conformation, Rad4 interacts exclusively with the two nucleotides on the undamaged strand without making specific contacts with the two damaged residues of the model UV lesion. This indirect recognition mode<sup>8,13</sup> explained how Rad4/XPC is capable of binding to a variety of lesions. However, it also posed a paradox regarding the mechanism of lesion recognition since the protein can no longer ‘see’ the difference between damaged versus undamaged DNA once the ‘open’ conformation is formed. It has thus been posited that Rad4/XPC may be limited in its ability to fully ‘open’ undamaged DNA in the same way as it opens damaged DNA<sup>8</sup>. To further explore this hypothesis, we sought to solve a crystal structure of the Rad4–Rad23 complex (hereafter referred to as Rad4) bound to undamaged DNA.

Here we present the first crystal structure in which Rad4 is captured on a single register of undamaged DNA using disulphide tethering. Our structure shows that tethered Rad4 flips out normal nucleotide pairs and forms the same ‘open’ conformation as with damaged DNA. These results indicate that structural discrimination between normal and damaged DNA cannot be the basis for lesion recognition by Rad4/XPC, and instead points to a novel ‘kinetic gating’ mechanism, whereby the DNA lesion selectivity arises mainly from the kinetic competition between Rad4-induced DNA opening and the residence time of Rad4 at a given site. While prolonging residence time, for example, by tethering, induces full opening (‘recognition’) of even normal DNA, we propose that freely diffusing Rad4 has a lower probability to open undamaged DNA than damaged DNA. This mechanism is further supported by direct measurements of Rad4-induced opening times ( $\sim 7$  ms) at a helix-destabilizing

lesion, using temperature-jump perturbation spectroscopy. The opening time for stable, undamaged DNA is expected to be orders of magnitude slower than that for opening an unstable lesion, and also significantly longer than the residence time of Rad4/XPC at the site, providing keys to lesion discrimination. This ‘kinetic gating’ provides novel insights into Rad4/XPC functions within and beyond NER<sup>14–18</sup>, and it may be a general mechanism for site-specific DNA-binding proteins to facilitate target search and recognition in the genome.

## Results

### Crystal structure of Rad4–Rad23 tethered to undamaged DNA.

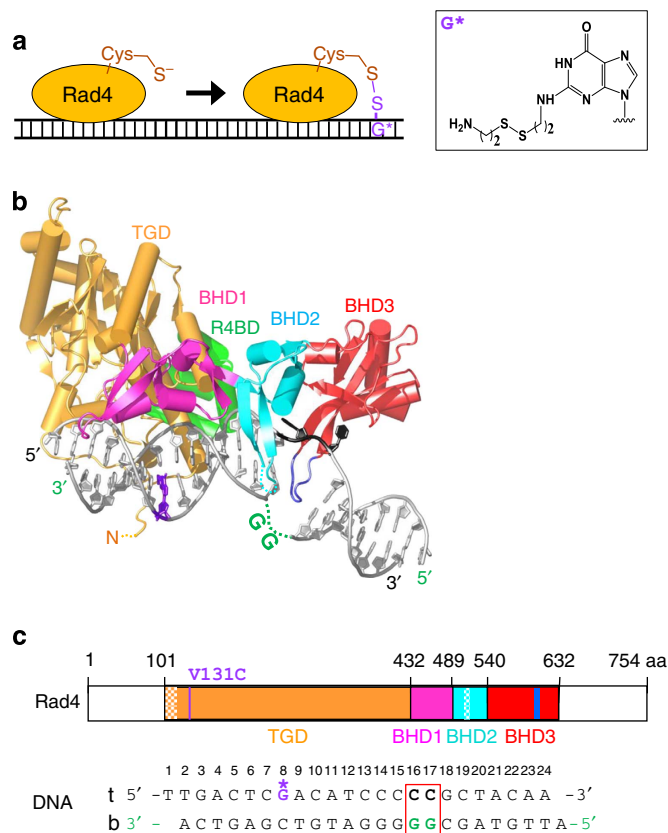
Obtaining diffracting crystals of the Rad4–undamaged DNA complex was initially challenging, presumably because Rad4 can bind in multiple different registers or binding sites on undamaged DNA. To overcome this problem, we covalently tethered Rad4 to undamaged DNA in a fixed register using a disulphide cross-linking method<sup>19</sup> and solved a 3.05 Å structure of this complex (Fig. 1; Supplementary Figs 1 and 2; Table 1). To our surprise, we found that this tethered complex formed an ‘open’ conformation that is essentially indistinguishable from the lesion-bound structure<sup>8</sup> (Fig. 2). Note that the disulphide tethering was made at a position distant from the flipped-out nucleotides (Fig. 1b,c) and, therefore, it is unlikely that the ‘open’ conformation was due to structural perturbations from crosslinking. Consistent with this notion, the electrophoretic mobility shift assays show that the modifications made in the crystallized construct do not affect the damage-specific binding of Rad4 compared with the unmodified, full-length protein (Supplementary Fig. 1). The overall root mean squared deviation<sup>20</sup> between this complex and the lesion-bound complex structures is 0.99 Å for 2,388 main-chain atoms including both protein and DNA. The distance between the crosslinked residues (C $\alpha$  of V131 in Rad4 and C2 of G\*8 in the top strand of the DNA; Fig. 1c) is 8.8 Å in this structure, only 0.3 Å shorter than the distance between equivalent atoms (9.1 Å between C $\alpha$  of C131 in Rad4 and C2 of A8 in the top strand of the DNA) in the lesion-bound structure (Protein Data Bank (PDB) code 2QSH)<sup>8</sup>, further confirming that the crosslinking did not induce structural distortions in the protein–DNA complex.

### Atomic force microscopy shows that Rad4 bends undamaged DNA.

In this Rad4–undamaged DNA structure, the DNA is bent by  $\sim 42^\circ$  as in the lesion-bound structure. Consistent with this, an atomic force microscopy (AFM) study of Rad4 bound (but not tethered) to a 514-bp undamaged DNA also showed that the DNA is bent ( $48.4 \pm 34.2^\circ$ ; Fig. 3). These results agree remarkably with a previous AFM study of human XPC bound to undamaged DNA segments ( $49^\circ \pm 36^\circ$ ; ref. 21). We also observed that a significant fraction of Rad4 binds to the ends of the DNA (Supplementary Fig. 3a). Only the fraction that was bound internally was used for analysis of the DNA bending angles (Supplementary Fig. 3b).

### Nonspecific Rad4–DNA structure suggests kinetic read-out.

To our knowledge, the presented crystal structure is the first high-resolution structure of a nonspecific DNA-bound, DNA repair protein that relies exclusively on indirect read-out and makes no direct contacts with the damaged nucleotides in the specific recognition complex. Our finding that a nonspecific complex is structurally identical to the specific complex was unexpected in light of previous studies on multiple other proteins that had shown clear differences between the two structures<sup>22–27</sup>, including those that used similar crosslinking approaches<sup>24–26</sup>. These earlier studies have been on sequence-specific DNA-binding proteins or damage-repair enzymes. For these proteins, target recognition lies, at least in part, in the ability of the proteins



**Figure 1 | Crystal structure determination of Rad4–Rad23 bound to undamaged DNA using disulphide crosslinking.** **(a)** Crosslinking scheme to tether Rad4 on DNA. A disulphide crosslink between cysteine 131 in Rad4 and a modified nucleotide (G\*) on DNA was introduced to capture the protein–DNA complex in a defined register of binding. The chemical structure of G\* is shown on the right. **(b)** Structure of Rad4–Rad23 tethered to undamaged DNA (PDB code 4U29). The transglutaminase domain (TGD, orange) and  $\beta$ -hairpin domain 1 (BHD1, magenta) of Rad4 bind to an 11-bp duplex segment of the DNA (silver), while BHD2 (cyan) and BHD3 (red) of Rad4 bind to a 4-bp segment in which two nucleotide pairs are flipped out. The tip of the long  $\beta$ -hairpin in BHD3 (residues 599–605, blue) is inserted into the DNA duplex and fills the gap created by the flipped-out nucleotides. The crosslinked, G\*8 in DNA and Cys131 in TGD are in purple. The two flipped-out cytidines (black) are bound by BHD2–BHD3, while the complementary guanines flipped out away from the protein are disordered in the crystal (green dotted line). Rad23 binds to TGD through its Rad4-binding domain (R4BD, light green). ‘N’ indicates the N terminus of Rad4. **(c; top)** Domain arrangements and boundaries of Rad4 used in this study. The colour scheme is the same as in **b**. The crystallized Rad4 construct spans residues 101–632 as before<sup>8</sup> and contains point mutations V131C (purple) and C132S. These residues are at the end of the N terminus of TGD, away from the BHD2–BHD3 bound to flipped-out nucleotides. The disordered regions in the crystal are checked. Rad23 construct is the same as in ref. 8. **(Bottom)** DNA duplex sequence used for the structure determination. The nucleotide pairs flipped out by Rad4 are boxed in red. The disulphide-modified nucleotide, G\*8, was at a position 8-bp away from the flipped-out ones (CC/GG at 16 and 17). Top strand (‘t’) corresponds to the undamaged strand and the bottom (‘b’) to the damaged strand in the lesion-bound structure<sup>8</sup> (PDB code 2QSG).

to directly accommodate target DNA sites and reject non-target sites as structural misfits. In contrast, our results reveal that such direct structural distinction cannot be the basis for the lesion recognition by Rad4. Instead, these results suggest that the key to Rad4’s recognition mechanism lies in the differential probability

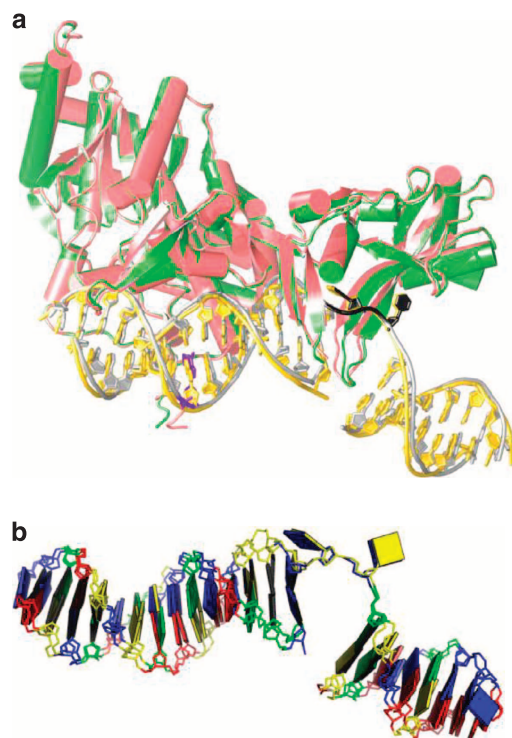
**Table 1 | Data collection and refinement statistics.**

	xc149
<b>Data collection</b>	
Space group	P 4 <sub>1</sub> 2 <sub>1</sub> 2
Cell dimensions	
<i>a</i> , <i>b</i> , <i>c</i> (Å)	79.405, 79.405, 404.366
$\alpha$ , $\beta$ , $\gamma$ (°)	90, 90, 90
Resolution (Å)	50.00–3.05 (3.10–3.05)
<i>R</i> <sub>sym</sub> or <i>R</i> <sub>merge</sub>	7.1% (79.2%)
<i>I</i> / $\sigma$ <i>I</i>	24.56 (2.3)
Completeness (%)	99.9% (100%)
Redundancy	6.7 (6.9)
<b>Refinement</b>	
Resolution (Å)	39.7–3.05
No. of reflections	25,974
<i>R</i> <sub>work</sub> / <i>R</i> <sub>free</sub> (%)	20.41/25.66
No. of atoms	5,447
Protein	4,508
DNA	939
Water	0
<i>B</i> -factors (Å <sup>2</sup> )	68.00
Protein	63.30
DNA (G47)	90.60 (91.67)
Water	0
r.m.s. deviations	
Bond lengths (Å)	0.017
Bond angles (°)	1.94
r.m.s., root mean squared.	
Values in parentheses are for highest-resolution shell.	

that Rad4 can open up a damaged versus undamaged DNA site before it diffuses away.

**Rad4-induced DNA-opening rates measured by temperature jump.** A key determinant of the ‘opening’ probability is the free energy barrier encountered by Rad4 to fully open DNA to form the recognition complex, which, in turn, determines the Rad4-induced opening rate. Here we have measured these opening rates using temperature-jump (T-jump) spectroscopy. Briefly, the Rad4 construct identical to the one used in crystallography studies<sup>8</sup> was bound to model lesions (2- or 3-bp mismatch bubbles in two different sequence contexts) in which 2-aminopurine (2AP, an adenine analogue) was incorporated as a fluorescent probe sensitive to DNA-opening events<sup>28</sup> (see Supplementary Table 1 for DNA sequences). All equilibrium and kinetics measurements reported here were carried out on untethered Rad4–DNA complexes.

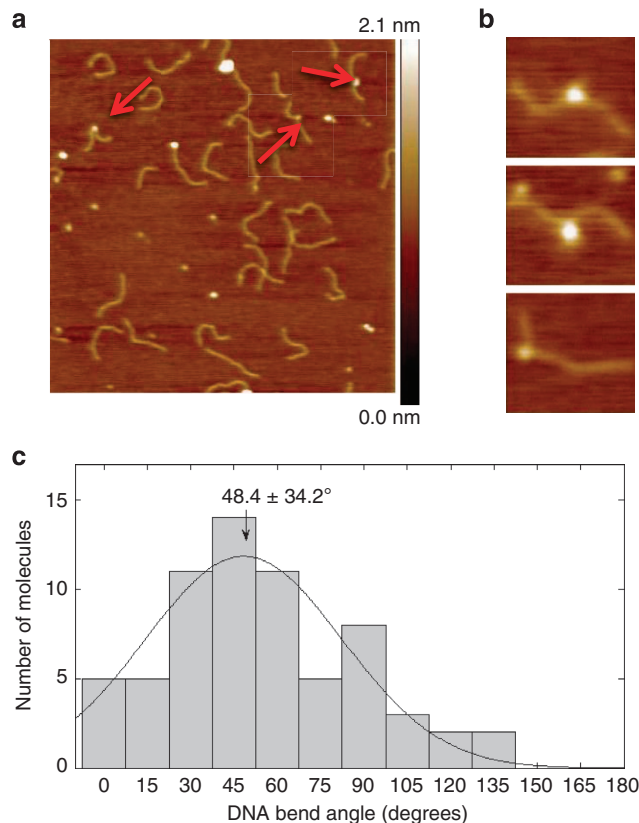
Specific binding of Rad4 to the mismatch model lesions resulted in a greater than fourfold increase in 2AP fluorescence of the DNA at equilibrium (Fig. 4a, left; Supplementary Fig. 4a). While small (less than approximately twofold) increase in 2AP fluorescence can come from unstacking of 2AP from its neighbours, for example, from partial duplex melting or unwinding, we attribute the much larger increase in the lesion-bound, specific complexes as arising from Rad4-induced DNA ‘opening’ that results in the complete flipping out of two nucleotide pairs as seen in the crystal structures<sup>8</sup>. A temperature scan from 10 to 40 °C showed that 2AP fluorescence of the specific, Rad4–mismatch DNA complexes increases as temperature is raised (Fig. 4b, left red), indicating that the extent of 2AP unstacking and/or the population of fully flipped-out conformations increases at higher temperatures. In contrast, the 2AP fluorescence intensity in mismatch DNA alone (Fig. 4b, left black) or in undamaged, matched DNA with or without Rad4 (Fig. 4b, right) decreases monotonically with



**Figure 2 | Tethered Rad4 opens up undamaged DNA nucleotide pairs as it does to damaged ones.** (a) Superposition of the Rad4-undamaged DNA (green-silver, this study) and the Rad4-damaged DNA<sup>8</sup> (red-gold, PDB code 2QSH) structures. The crosslinked G\* in DNA and Cys131 in Rad4 are indicated in purple and the flipped-out nucleotides are indicated in black in the undamaged DNA-bound structure. (b) Rad4-bound DNA conformations are nearly identical between the damaged- and the undamaged-bound conformations. Cytidines are represented in yellow boxes, thymidines in blue, guanosines in green and adenosines in red. The figure is generated by 3DNA<sup>44</sup>.

increasing temperature, which reflects a temperature-dependent decrease in 2AP quantum yield.

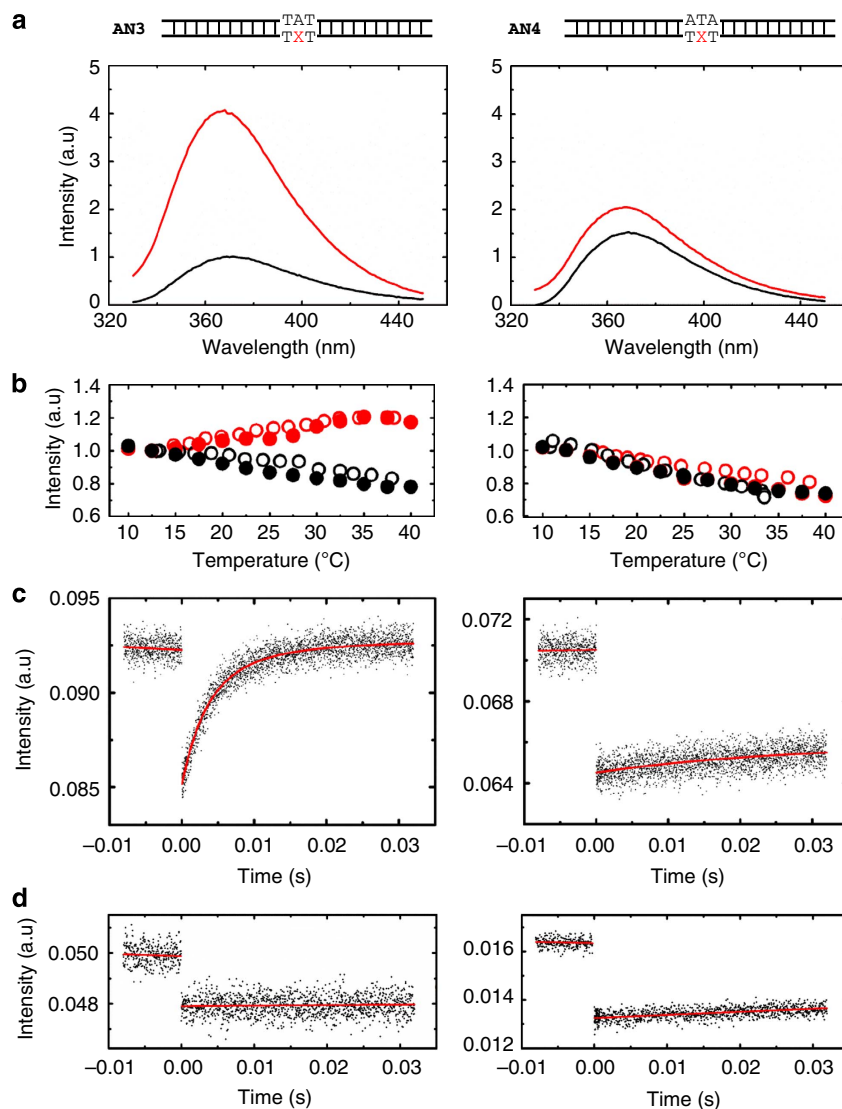
To monitor conformational relaxation kinetics of Rad4 bound to DNA, samples were rapidly perturbed by a T-jump of  $\sim 5$ – $10$  °C with a  $\sim 10$ -ns infrared laser pulse, and the temporal response of the 2AP fluorescence was monitored as the ensemble of molecules re-equilibrated to the distribution corresponding to the higher temperature. Our T-jump apparatus can monitor conformational relaxation kinetics of protein–DNA interactions in a time window from  $\sim 5$   $\mu$ s to 50 ms, as demonstrated with other DNA-binding proteins<sup>29</sup>. T-jump measurements on the ensemble of Rad4–mismatch DNA complexes revealed relaxation kinetics with time constants of  $5.5 \pm 0.5$  and  $7.7 \pm 2.0$  ms for the 3-bp and 2-bp mismatch constructs, respectively (interpolated at 25°C), indicating that the kinetics were also sequence context independent (Fig. 4c, left; Supplementary Fig. 4c). We attribute these to the Rad4-induced DNA ‘opening’ kinetics. Similar relaxation rates were also observed with full-length Rad4 complex on the 3-bp mismatch DNA (Supplementary Fig. 5). In contrast, measurements on Rad4 nonspecifically bound (but not tethered) to undamaged, matched DNA showed a much smaller increase in 2AP fluorescence on Rad4 binding and no relaxation kinetics on the millisecond timescale (Fig. 4, right). Instead, the complex showed much slower kinetics on timescales characteristic of the re-equilibration of the heated sample temperature back to the initial (bath) temperature (T-jump recovery), similar to the control samples, matched and mismatch DNA in the absence of Rad4 (Fig. 4d; Supplementary Fig. 4d). The apparent absence of



**Figure 3 | AFM on Rad4 binding to 514-bp undamaged DNA.** (a) AFM image of Rad4 with 514-bp undamaged DNA. The arrows indicate proteins bound on DNA. Scan area is  $1 \mu\text{m} \times 1 \mu\text{m}$ . (b) Zoomed images of examples of Rad4 bound on DNA. (c) Histogram of bend angle distribution for Rad4 bound to undamaged DNA. Total of 67 molecules were measured from three independent depositions. The solid-line curve represents fitting of these data to a Gaussian distribution in MATLAB, providing the centre of bend angle as  $48.4^\circ (\pm 34.2^\circ)$ . The uncertainty in the bend angle is the sample s.d. derived from the Gaussian fit.

conformational relaxation kinetics on the untethered complex with undamaged DNA was somewhat unexpected in view of the ‘open’ crystal structure of the tethered complex. The absence of kinetics can arise from heterogeneous binding registers (since Rad4-induced opening of DNA at a site some distance away from the fixed 2AP site will not be reported by 2AP fluorescence change) and/or slow relaxation kinetics not resolved in the T-jump time window.

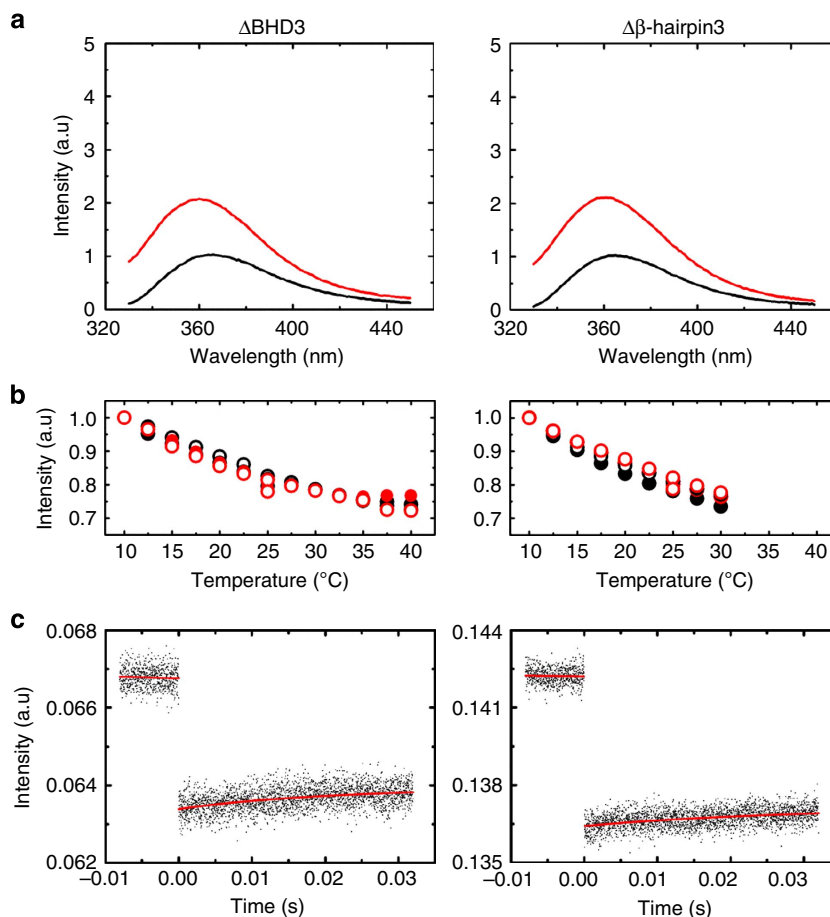
Structural studies have shown that the  $\beta$ -hairpin in the  $\beta$ -hairpin domain 3 (BHD3) inserts into the DNA duplex on forming the ‘open’ conformation (Fig. 1; ref. 8); this  $\beta$ -hairpin is also shown to be important in the ability of Rad4/XPC to recognize DNA lesions in biochemical and cellular studies (Supplementary Fig. 1; refs 18,30). To probe what role the  $\beta$ -hairpin plays in the observed increase in 2AP fluorescence in the mismatch DNA on Rad4 binding, and in the relaxation kinetics when perturbed by T-jump, we have performed corresponding 2AP fluorescence measurements with mutant Rad4 (Fig. 5) that lack either the  $\beta$ -hairpin in BHD3 ( $\Delta\beta$ -hairpin3, missing residues 599–605 indicated in blue in Fig. 1b,c) or the entire BHD3 domain ( $\Delta$ BHD3, missing residues 541–632 indicated in red in Fig. 1b,c). The 2AP fluorescence of the mismatch DNA in complex with either of the Rad4 mutants shows a temperature dependence very similar to that of free DNA (Fig. 5b). These results contrast with the behaviour of the wild-



**Figure 4 | Rad4-induced DNA-opening dynamics measured by T-jump spectroscopy using ZAP fluorescence as a probe.** (a) ZAP (X in the schematic representation of DNA substrates) was placed within 3-bp mismatch DNA (AN3) and matched DNA (AN4; Supplementary Table 1). The ZAP fluorescence emission spectra were measured for DNA alone (black) and Rad4-DNA complexes (red) with excitation at 314 nm at 25 °C (left: AN3; right: AN4). The ZAP fluorescence emission intensities increase 4.0 ( $\pm 0.7$ )-fold and 1.4 ( $\pm 0.1$ )-fold on Rad4 binding to mismatch and matched DNA, respectively. All measurements were done with untethered Rad4. Protein and DNA concentrations were 10  $\mu$ M each. (b) The maxima of the equilibrium ZAP fluorescence emission, measured at 365 nm, are plotted as a function of temperature for DNA alone (black) and Rad4-DNA complexes (red; left: AN3; right: AN4). Open and filled symbols are for two independent sets of measurements on each sample. In each panel, the intensities for free DNA and DNA in complex have been normalized to match at the lowest temperature. (c) Relaxation kinetic traces measured in response to a  $\sim 7$  °C T-jump show (left) single-exponential kinetics, with relaxation time  $5.1 \pm 0.5$  ms (at final temperature 26 °C) for Rad4-mismatch DNA, and (right) much slower kinetics, with relaxation time  $190 \pm 42$  ms (at final temperature 29 °C) for Rad4-matched DNA. The uncertainties in the relaxation times are sample s.d. from two sets of measurements. Note that the  $K_d$  values of Rad4 bound to mismatch or matched DNA are in nanomolar range (Supplementary Fig. 1), well below the 60  $\mu$ M concentrations used in the T-jump experiments. (d) Relaxation kinetics measured on the DNA-only samples: AN3 alone, in response to a 6 °C T-jump (left), and AN4 alone, in response to a 10 °C T-jump (right), exhibit much slower kinetics, with relaxation times  $341 \pm 70$  and  $240 \pm 35$  ms, respectively, consistent with the T-jump recovery kinetics (see also Supplementary Fig. 4d). The relaxation times for samples that exhibit only the slow kinetics were determined by making measurements over a longer time window, up to 80–320 ms.

type Rad4-mismatch DNA complex, which showed an increase in fluorescence emission intensity with increasing temperature (Fig. 4b, left). Furthermore, no relaxation kinetics (other than the T-jump recovery kinetics) was observed with these  $\beta$ -hairpin mutants (Fig. 5c). The results indicate that the  $\beta$ -hairpin is indeed critical for forming and stabilizing the ‘open’ conformation, consistent with other studies (Supplementary Fig. 1; refs 8,18).

The observed Rad4-induced DNA-opening kinetics of  $\sim 7$  ms on mismatch DNA are almost tenfold slower than the spontaneous base-pair ‘breathing’ measured even for a single (GT) mismatch (Fig. 6; Supplementary Fig. 5)<sup>31</sup>. These results show that the free energy barrier for forming the ‘open’ structure is higher than for the smaller ‘breathing’ motions, and thus indicate that the rate-limiting transition state for ‘opening’ entails extensive structural deformation around the lesion. Notably, NER



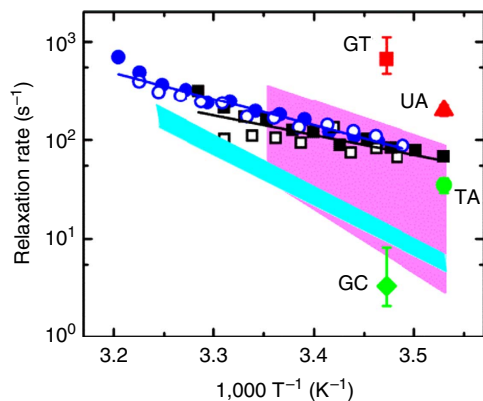
**Figure 5 | ZAP fluorescence and T-jump measurements to examine Rad4-induced DNA-opening dynamics with Rad4  $\beta$ -hairpin mutants.** Equilibrium and relaxation kinetics measurements were carried out on 3-bp mismatch DNA (AN3) bound to two  $\beta$ -hairpin mutants of Rad4,  $\Delta$ BHD3 (*left*) and  $\Delta\beta$ -hairpin3 (*right*). **(a)** The ZAP fluorescence emission spectra measured for DNA alone (black) and Rad4–DNA complexes (red) with excitation at 314 nm at 25  $^{\circ}$ C. The ZAP fluorescence emission intensities of AN3 DNA increase 2.1 ( $\pm$  0.1)-fold and 2.3 ( $\pm$  0.1)-fold on binding of  $\Delta$ BHD3 and  $\Delta\beta$ -hairpin3, respectively, compared with DNA alone. **(b)** The maxima of the equilibrium ZAP fluorescence emission measured at 365 nm are plotted as a function of temperature for DNA alone (black) and Rad4–DNA complexes (red). The open and filled symbols are for two independent sets of measurements on each sample. The intensities are normalized to match at the lowest temperature. **(c)** T-jump measurements on the  $\beta$ -hairpin mutants of Rad4–DNA complexes do not show any relaxation kinetics other than the slow recovery of the T-jump itself.

lesions often weaken base-pair hydrogen-bonding and stacking interactions and render the DNA more deformable with increased propensity for local bending/unwinding/nucleotide flipping<sup>2</sup>. Thus, the free energy barrier for Rad4-induced opening of undamaged DNA is expected to be even higher compared with damaged DNA. A recent simulations study estimated that the free energy barrier for flipping out normal DNA could be 5–8 kcal mol<sup>-1</sup> higher than for a bulky lesion<sup>32</sup>. Assuming that this entire difference appears in the free energy barrier for Rad4-induced DNA-opening kinetics, the opening times for normal DNA bases could well be >4,000-fold longer than the  $\sim$ 7 ms opening time observed for our model lesion.

Taken together, our present study strongly suggests that the kinetic barriers for forming the ‘open’ conformation are key to distinguishing damaged versus undamaged DNA, although the final ‘open’ structures appear to be nearly the same for both DNA. Figure 7a illustrates this notion. In this schematic, the free energies of the final ‘open’ structures, as well as the corresponding transition states for forming such structures, are assumed to be very similar for the damaged and undamaged DNA. However, the free energies when Rad4/XPC loosely and nonspecifically interacts with the DNA in the ‘search’ mode<sup>33</sup> are assumed to differ by an amount equal to the free energy difference between

undamaged and damaged free DNA. Although an experimental structure of Rad4/XPC in an untethered, bona fide search mode is lacking, one can anticipate that the DNA duplex in this mode remains largely in the conformation of free DNA (as modelled in Fig. 7b). Consequently, the higher free energy barrier for Rad4/XPC ‘opening’ undamaged versus damaged DNA mainly arises from the lower free energy of the Rad4/XPC–normal DNA complex in the search mode that reflects the higher stability of normal DNA than that of damaged DNA.

**Interplay between residence time and opening time.** Given a free energy barrier to open a DNA site, the overall probability of Rad4 opening the site also depends on whether this barrier can be overcome by Rad4 before it steps to the next site or falls off, that is, the residence time of Rad4 at the site. The opening probability per site then can be expressed as the ratio of the opening time ( $t_{op}$ ) and the residence time ( $t_{res}$ ), and this probability increases exponentially as the ratio decreases:  $P_{open} = 1 - \exp(-t_{res}/t_{op})$ . In our structural study, chemical tethering of Rad4 on DNA prolonged Rad4’s residence time indefinitely, thus allowing even undamaged DNA to be fully opened. However, under untethered conditions, the protein is free to diffuse away from a given site, thus decreasing

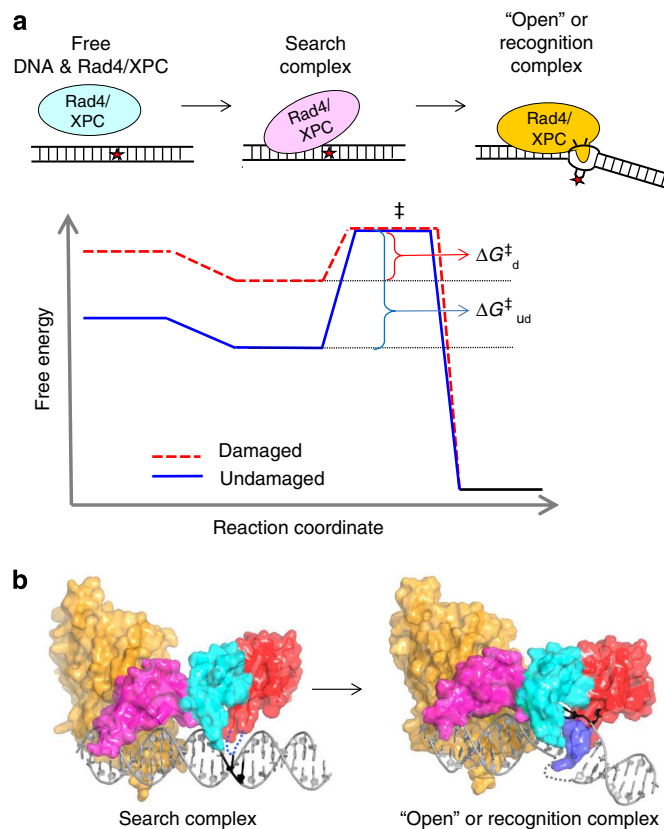


**Figure 6 | Arrhenius plots of relaxation rates measured on Rad4-mismatch DNA complexes.** Relaxation rates obtained from T-jump measurements on Rad4-mismatch DNA (AN3 in blue circles and AN21 in black squares) are plotted versus the inverse of the final temperature after the T-jump. The open and filled symbols are for two independent sets of measurements on each sample. The continuous (blue and black) lines are an Arrhenius fit to the relaxation rates for each sample; the activation enthalpies obtained from the Arrhenius fit are  $12.1 \pm 2.5$  and  $7.9 \pm 4.7$  kcal mol<sup>-1</sup>, respectively, for the two sequences, and the relaxation times interpolated at 25 °C are  $5.5 \pm 0.5$  and  $7.7 \pm 2.0$  ms, respectively. The uncertainties in the activation enthalpies and relaxation times are sample s.d. obtained from two independent sets of measurements. The pink (cyan) shaded regions represent the range of base-pair opening ('breathing') rates for A/T (G/C) base pairs, from NMR imino proton exchange measurements of Coman and Russu<sup>45</sup>. The opening rates and their ranges are also indicated for G/T mismatch (red square)<sup>31</sup>, U/A mismatch (red triangle)<sup>23</sup>, T/A base pair (green circle)<sup>23</sup>, and G/C base pair (green diamond)<sup>31</sup>.

the residence time per site. If  $t_{\text{res}} \ll t_{\text{op}}$ , this opening probability is likely to be small. While there are no direct measurements of the residence time for Rad4/XPC on DNA, several measurements have characterized diffusion of DNA repair proteins on undamaged, nonspecific DNA. These measurements yield a wide range of residence times per base pair, 0.1  $\mu$ s to 0.3 ms (Supplementary Table 2)<sup>34–36</sup>, all of which are significantly shorter than the Rad4-induced opening time expected for undamaged DNA ( $\gg 7$  ms). This estimate for opening normal DNA is based on the 7 ms opening time that we measured for mismatch DNA at 25 °C (Fig. 6) and the simulations study<sup>32</sup> that suggests significantly larger free energy barrier for flipping out normal bases in matched DNA (as illustrated in Fig. 7). Assuming that the residence time of Rad4 on undamaged DNA falls in the submillisecond range, the probability that Rad4 will open a normal site is expected to be exceedingly small. Although it remains to be determined, we speculate that the presence of an NER lesion that distorts or destabilizes DNA not only decreases the opening time, but may also increase the residence time of Rad4/XPC on the lesion compared with an undamaged site, further contributing to the selective opening and thus recognition of NER lesions. On the other hand, certain carcinogen adducts resistant to NER<sup>2,37</sup> may evade detection in part because they present very high free energy barriers for being opened by Rad4 compared with its residence time, and thus their opening probability may be similar to or worse than that of undamaged DNA.

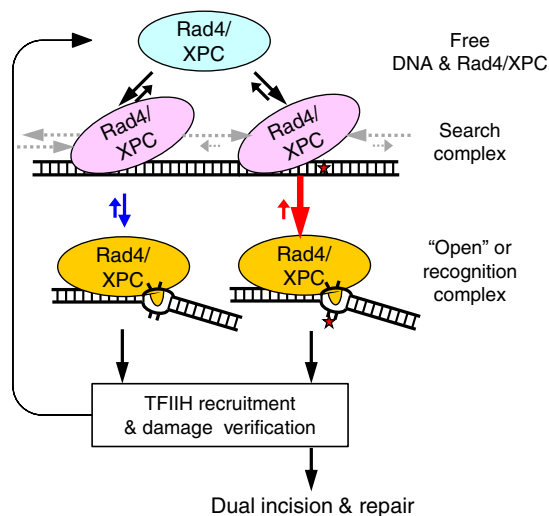
## Discussion

In summary, we propose that Rad4's lesion recognition relies on 'kinetic gating' mechanism to achieve selectively higher probability of 'opening' damaged DNA than normal DNA. In this



**Figure 7 | Free energy profile for DNA opening by Rad4/XPC at damaged and undamaged sites.** (a) A schematic representation of the free energy profile of the Rad4/XPC-induced DNA-opening process is shown: free Rad4/XPC is indicated in cyan, Rad4/XPC in the 'search' mode in pink and in the 'open' recognition complex in orange. The free energy surface for opening damaged DNA sites is in red and that for undamaged sites is in blue.  $\Delta G_d^\ddagger$  and  $\Delta G_{ud}^\ddagger$  denote the free energy barriers to open damaged and undamaged sites in the complex, respectively. (b) Structural models for Rad4 in a 'search' mode (left) and in a recognition complex (right). Here we used the apo-Rad4-Rad23 structure (PDB code 2QSF)<sup>8</sup> to model Rad4 nonspecifically bound to DNA in a search mode. The  $\beta$ -hairpin in BHD3 in this conformation can be flexible (blue dotted line), while the DNA duplex largely retains the B-DNA conformation. On the contrary, the  $\beta$ -hairpin in the recognition complex is inserted into the DNA duplex (blue surface representation) and stabilizes the 'open' structure of the DNA. The 'open' complex is formed with damaged DNA<sup>8</sup> and also with undamaged DNA when Rad4's residence time is prolonged (this study).

mechanism, lesion recognition by Rad4/XPC leading to the 'open' conformation is controlled ('gated') by the competition between two kinetic parameters, as illustrated in Fig. 8: (1) opening time to form the thermodynamically stable, 'open' or recognition complex and (2) residence time per site as determined by the one-dimensional diffusion such as sliding and hopping of protein on DNA when it is in the loosely bound, nonspecific 'search' mode<sup>33</sup>. The opening rate for damaged DNA (red arrows) is expected to be larger than the opening rate for undamaged DNA (blue arrows), because of, for instance, the weakened base stacking and hydrogen bonding within a lesion. The residence time, on the other hand, may well be shorter for undamaged than for damaged DNA. The balance between these two factors together results in a higher net probability of Rad4/XPC opening damaged rather than undamaged DNA. This difference in the opening probabilities forms the basis for lesion recognition despite little difference in the thermodynamically most stable structures.



**Figure 8 | 'Kinetic gating' mechanism for DNA damage recognition by Rad4/XPC.** This schematic illustrates that the lesion recognition by Rad4/XPC is controlled by the competition between Rad4's residence time per site when it is in the 'search' mode (pink) and its opening time to form the 'open' or recognition complex (orange). The diffusion rates that control the residence time are indicated by the grey arrows. The opening (and its reverse) rates are depicted in red and blue arrows for damaged and undamaged DNA, respectively.

How proteins find their specific targets with a large number of closely related nonspecific molecules around is not limited to DNA-binding proteins, but is prevalent in biology. It is inevitable that erroneous bindings occur, but most systems are tolerant of such thermodynamic noise owing to intermediate 'proofreading' before proceeding down a pathway. 'Kinetic proofreading' originally proposed by Hopfield<sup>38</sup> and Ninio<sup>39</sup> is one such mechanism that detects and aborts incorrect bindings using chemical energy (for example, ATP hydrolysis). Our model showcases, for the first time, that 'kinetic proofreading' can be accomplished with neither chemical energy consumption nor difference in the thermodynamically most preferred states of the protein-DNA complexes<sup>40</sup>. During NER, lesion recognition by XPC is also followed by a bona fide, ATP-dependent damage verification by TFIIH, which further augments NER accuracy<sup>3</sup>. Finally, the DNA 'opening' induced by prolonged residence time may help explain the mechanisms of XPC within and beyond NER that cannot be explained solely by lesion-binding preferences of XPC<sup>14–17</sup>. In these cases, XPC-interacting proteins (for example, UV-damaged DNA-binding protein complex or DNA glycosylases) and/or post-translational modifications on XPC can help 'stall' the protein on DNA and induce opening of otherwise non-cognizant DNA. We also posit that 'kinetic gating' is a general target-recognition mechanism even for DNA-binding proteins that ultimately rely on direct structural discrimination, as a way of reducing wasteful interrogation at each and every site. Our results may have broad implications in genome maintenance, gene regulation, and cancer biology.

## Methods

**Oligonucleotide synthesis.** Oligonucleotides containing disulphide-tethered cytosine were prepared by incorporating the 2-F-di-CE phosphoramidite (Glen Research) at the desired position during solid-phase synthesis. The conversion and deprotection of 2-F-di were performed according to Technical Bulletin provided by Glen Research ([http://www.glenresearch.com/Technical/TB\\_2-F-di.pdf](http://www.glenresearch.com/Technical/TB_2-F-di.pdf)). Briefly, 2-F-di-containing oligonucleotides were reacted with cystamine to tether the disulphide group, then deprotected with 1,8-diazabicycloundec-7-ene. All synthetic oligonucleotides were purified with denaturing polyacrylamide gel electrophoresis.

**Preparation of Rad4–Rad23 complex.** All Rad4–Rad23 complexes used in the study were overexpressed and purified from insect cells using previously described methods<sup>8</sup>. Briefly, the Hi5 insect cells co-expressing the Rad4–Rad23 complex were harvested 2 days after infection. After lysis, the proteins were purified using His-Select Nickel agarose resin (Sigma) and anion-exchange chromatography (Source Q, GE healthcare), followed by thrombin digestion and cation exchange (Source S, GE healthcare) and gel-filtration (Superdex200, GE healthcare) chromatography. Final sample was concentrated by ultrafiltration to  $\sim 13 \text{ mg ml}^{-1}$  in 5 mM bis-tris propane-HCl (BTP-HCl), 800 mM NaCl and 5 mM dithiothreitol (DTT), pH 6.8.

To site-specifically tether Rad4 on DNA using disulphide crosslink, a V131C/C132S double mutation in Rad4 was introduced in the previously crystallized construct<sup>8</sup> of the Rad4–Rad23 complex by site-directed mutagenesis. The V131C and C132S mutations were necessary and sufficient for efficient crosslinking of the purified Rad4–Rad23 complex with the DNA. The presence of seven other cysteines (C276, C354, C355, C463, C466, C509 and C572) of which the side chains were exposed on the surface of Rad4 did not affect the crosslinking, further signifying the specificity of the DNA binding and crosslinking. In fact, mutating these other cysteine residues to serines decreased the solubility of the protein and did not increase the crosslinking yield with DNA (data not shown).

The  $\beta$ -hairpin mutants of Rad4 used for T-jump experiments and electrophoretic mobility shift assays were prepared similarly. The mutant,  $\Delta$ BHD3, lacked the entire BHD3 domain (residues 541–632, red in Fig. 1b,c), but instead had three extra glutamate residues (EEE) after residue 540 to increase the solubility of the protein. The other mutant,  $\Delta\beta$ -hairpin3, lacked only the long  $\beta$ -hairpin region (residues 599–605, blue in Fig. 1b,c) in the BHD3. These mutations of Rad4 were incorporated in the context of the truncated Rad4–Rad23 complex previously described<sup>8</sup>.

## Disulphide crosslinking of Rad4–Rad23 complex with double-stranded DNA.

To crosslink the protein and DNA, DTT in the Rad4–Rad23 storage buffer was first removed from the purified Rad4–Rad23 complex through a desalting column (Zeba Spin Desalting Column, 40,000 Da molecular weight cut-off, Thermo Scientific) pre-equilibrated in 5 mM BTP-HCl and 800 mM sodium chloride (NaCl), pH 6.8. The protein complex was subsequently mixed with DNA containing disulphide-modified base (G\*, Fig. 1a,c) at 1:1 molar ratio at a final concentration of  $\sim 20 \mu\text{M}$  in crosslinking buffer (5 mM BTP-HCl, 100 mM NaCl and 10% glycerol, pH 6.8) at 4 °C overnight. The extent of the reaction was determined by SDS-PAGE under non-reducing conditions after treating the sample with 0.1 mM S-methylmethanethiosulfonate (Sigma) to quench the reaction. Typical crosslinking yield was  $\sim 50$ –70% (Supplementary Fig. 2).

## Purification of Rad4–Rad23–DNA complex.

To remove unreacted protein and DNA, the crosslink reaction mixture was subject to anion-exchange chromatography (Mono Q, GE healthcare) over a 0–2 M NaCl gradient in 5 mM BTP-HCl and 10% glycerol, pH 6.8. The buffers were degassed by purging nitrogen. Purified Rad4–Rad23–DNA eluted at 400–480 mM NaCl and was further concentrated by ultrafiltration (Amicon, Millipore) to  $\sim 30 \mu\text{M}$  or 3 mg ml<sup>-1</sup> (Supplementary Fig. 2).

## Crystallization, structure determination and refinement.

All crystals were grown by the hanging-drop vapour diffusion method at 4 °C, mixing 1  $\mu\text{l}$  of protein solution and 1  $\mu\text{l}$  of crystallization buffer. Crystals of the complex appeared after a few days at 4 °C in wells containing 50 mM BTP-HCl, 100 mM NaCl, 12–16% isopropanol and 100 mM calcium chloride (CaCl<sub>2</sub>), pH 6.8. The crystals were then harvested with a harvest buffer using 20–30% polyethylene glycol (PEG) 200 or PEG400 as cryoprotectants, and were subsequently flash frozen in liquid nitrogen (Supplementary Fig. 2). Diffraction data were collected at  $-170 \text{ }^\circ\text{C}$  and were processed with the HKL3000 suite<sup>41</sup>. The structure of the Rad4–Rad23–DNA complex was determined by molecular replacement method using the previous structure (PDB code 2QSH, chains A and X containing only Rad4 and Rad23) as the search model and refined through multiple rounds of refinement in Phenix<sup>42</sup>. The final model contains residues 126–514 and 525–632 of Rad4 and 256–308 of Rad23. Figures are generated by Pymol (the PyMOL Molecular Graphics System, Version 1.2r3pre, Schrödinger, LLC).

**AFM studies of Rad4–DNA binding.** The 514-bp undamaged DNA substrate was made by PCR from pSCW01 plasmid using the forward primer 5'-GCATTGCT GAGGGTTATTGTC-3' and reverse primer 5'-TATCCGGTAAGCGGCAGG-3'. The PCR products were purified with QIAquick PCR purification kit (QIAGEN) and eluted in filtered deionized water.

To obtain AFM images, the purified Rad4–Rad23 complex used for crystallization (2  $\mu\text{M}$ ) was incubated with the 514-bp undamaged DNA substrate (400 nM) in binding buffer of 5 mM BTP-HCl, 160 mM NaCl, 5% glycerol and 0.74 mM 3-[(3-cholamidopropyl) dimethylammonio]-1-propanesulfonate (CHAPS), pH 6.8. Reaction mixtures were incubated at room temperature for 20 min and diluted by 1:80 in AFM deposition buffer of 25 mM sodium acetate, 25 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid-potassium hydroxide (HEPES-KOH, pH 7.5) and 10 mM magnesium acetate. Ten  $\mu\text{l}$  of the dilution was



deposited onto freshly cleaved mica, rinsed with deionized water and dried under a gentle stream of nitrogen gas. Images were collected by a MultiModeV microscope (Bruker Corporation) in an E scanner in tapping mode. Images were captured at a scan size of  $1\ \mu\text{m} \times 1\ \mu\text{m}$ , scan rate of  $\sim 3\ \text{Hz}$  and resolution of  $512 \times 512$  pixels. DNA contour lengths and bend angles from the obtained AFM images were measured using ImageJ. The bend angle data were binned into histograms and fitted to a Gaussian distribution using MATLAB.

**Fluorescence measurements and T-jump spectroscopy.** Fluorescence measurements were carried out with DNA sequences containing a 3-bp mismatch (AN3) or 2-bp mismatch (AN21) in two different sequence contexts (Supplementary Table 1), with 2AP substituted for adenine at positions denoted by X. Measurements were also carried out with undamaged (matched) DNA construct (AN4; Supplementary Table 1). The 2AP-labelled DNA oligonucleotides were obtained from Oligos Etc. (Wilsonville, OR). All DNA oligonucleotides were purchased with high-performance liquid chromatography purification. All fluorescence measurements were done in phosphate-buffered saline buffer (10 mM disodium hydrogen phosphate ( $\text{Na}_2\text{HPO}_4$ ), 2 mM monopotassium phosphate ( $\text{KH}_2\text{PO}_4$ ), 137 mM NaCl and 2.7 mM potassium chloride (KCl), pH 7.4) with 1 mM DTT. The steady-state fluorescence emission spectra on 2AP-labelled DNA constructs, with and without bound Rad4, were measured on a FluoroMax4 spectrofluorimeter (Horiba Scientific, NJ). 2AP-labelled DNA duplexes were excited at 314 nm and emission spectra collected over the wavelength range 330–450 nm and temperature range 10–40 °C. Sample reversibility after the heating/cooling cycle was verified in equilibrium measurements on all samples, by comparing the fluorescence emission spectra measured at 25 °C before and after the heating/cooling cycle. For all DNA-only and protein–DNA complexes, with the exception of the complex with  $\Delta\beta$ -hairpin3 mutant, reversible behaviour was observed with heating up to at least 40 °C. For  $\Delta\beta$ -hairpin3–DNA complexes, however, irreversible behaviour was observed when the samples were heated to above 30 °C; therefore for measurements reported on this complex, the temperature range did not exceed 30 °C. The concentrations for the equilibrium measurements were 10  $\mu\text{M}$  DNA each with 1:1 ratio of protein:DNA for all complexes. The intensities of all the fluorescence emission spectra shown are scaled relative to the spectra of 3-bp mismatch DNA-only sample (AN3), which was normalized such that the maximum of the spectrum measured at 25 °C was 1 (in arbitrary units).

The kinetics measurements were performed using a home-built laser T-jump spectrometer, as described previously<sup>43</sup>. Briefly, 10 ns laser pulses at 1,550 nm, generated by Raman shifting the 1,064 nm pulses from the output of an Nd:YAG laser, are focused to  $\sim 1\ \text{mm}$  spot size on to a sample cuvette of path length 0.5 mm, which yields  $\sim 10\ \text{°C}$  T-jump at the centre of the heated volume. The magnitude of the T-jump was determined by measurements on control samples (2AP-labelled DNA only) by a comparison of the intensity observed in the kinetics experiments immediately after the T-jump with the intensity observed in the equilibrium measurements of the temperature dependence of the 2AP fluorescence for these samples. The errors in the T-jump estimates are about 10–20%. The probe source for excitation of 2AP fluorescence was a 200-W Hg–Xe lamp, with the excitation wavelengths selected by a broadband filter with transmission in the range of 300–330 nm. The probe light was focused on a  $\sim 300\text{-}\mu\text{m}$  spot in the middle of the heated volume. The fluorescence emission intensity was monitored perpendicular to the excitation direction, with a combination of a long-pass filter ( $> 352\ \text{nm}$ ) and a short-pass filter (388 nm), and measured with a Hamamatsu R928 photomultiplier tube and a 500-MHz transient digitizer. The protein and DNA concentrations for the T-jump measurements were 60  $\mu\text{M}$  each.

## References

- Berg, O. G., Winter, R. B. & von Hippel, P. H. Diffusion-driven mechanisms of protein translocation on nucleic acids. 1. Models and theory. *Biochemistry* **20**, 6929–6948 (1981).
- Gillet, L. C. & Scharer, O. D. Molecular mechanisms of mammalian global genome nucleotide excision repair. *Chem. Rev.* **106**, 253–276 (2006).
- Naegeli, H. & Sugasawa, K. The xeroderma pigmentosum pathway: decision tree analysis of DNA quality. *DNA Repair (Amst)* **10**, 673–683 (2011).
- Luijsterburg, M. S. *et al.* Stochastic and reversible assembly of a multiprotein DNA repair complex ensures accurate target site recognition and efficient repair. *J. Cell Biol.* **189**, 445–463 (2010).
- Yeo, J. E., Khoo, A., Fagbemi, A. F. & Scharer, O. D. The efficiencies of damage recognition and excision correlate with duplex destabilization induced by acetylaminofluorene adducts in human nucleotide excision repair. *Chem. Res. Toxicol.* **25**, 2462–2468 (2012).
- Hey, T. *et al.* The XPC-HR23B complex displays high affinity and specificity for damaged DNA in a true-equilibrium fluorescence assay. *Biochemistry* **41**, 6583–6587 (2002).
- Trego, K. S. & Turchi, J. J. Pre-steady-state binding of damaged DNA by XPC-hHR23B reveals a kinetic mechanism for damage discrimination. *Biochemistry* **45**, 1961–1969 (2006).
- Min, J. H. & Pavletich, N. P. Recognition of DNA damage by the Rad4 nucleotide excision repair protein. *Nature* **449**, 570–575 (2007).
- Hoogstraten, D. *et al.* Versatile DNA damage detection by the global genome nucleotide excision repair protein XPC. *J. Cell Sci.* **121**, 2850–2859 (2008).
- Luijsterburg, M. S. *et al.* Dynamic in vivo interaction of DDB2 E3 ubiquitin ligase with UV-damaged DNA is independent of damage-recognition protein XPC. *J. Cell Sci.* **120**, 2706–2716 (2007).
- Vermeulen, W. Dynamics of mammalian NER proteins. *DNA Repair (Amst)* **10**, 760–771 (2011).
- Sugasawa, K. *et al.* A multistep damage recognition mechanism for global genomic nucleotide excision repair. *Genes Dev.* **15**, 507–521 (2001).
- Buterin, T., Meyer, C., Giese, B. & Naegeli, H. DNA quality control by conformational readout on the undamaged strand of the double helix. *Chem. Biol.* **12**, 913–922 (2005).
- Sugasawa, K. *et al.* UV-induced ubiquitylation of XPC protein mediated by UV-DDB-ubiquitin ligase complex. *Cell* **121**, 387–400 (2005).
- Le May, N., Egly, J. M. & Coin, F. True lies: the double life of the nucleotide excision repair factors in transcription and DNA repair. *J. Nucleic Acids*, (doi: 10.4061/2010/616342) (2010).
- Pascucci, B., D'Errico, M., Parlanti, E., Giovannini, S. & Dogliotti, E. Role of nucleotide excision repair proteins in oxidative DNA damage repair: an updating. *Biochemistry* **76**, 4–15 (2011).
- Fong, Y. W. *et al.* DNA repair complex functions as an oct4/sox2 coactivator in embryonic stem cells. *Cell* **147**, 120–131 (2011).
- Fei, J. *et al.* Regulation of nucleotide excision repair by UV-DDB: prioritization of damage recognition to internucleosomal DNA. *PLoS Biol.* **9**, e1001183 (2011).
- He, C. & Verdine, G. L. Trapping distinct structural states of a protein/DNA interaction through disulfide crosslinking. *Chem. Biol.* **9**, 1297–1303 (2002).
- Guex, N. & Peitsch, M. C. SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis* **18**, 2714–2723 (1997).
- Janicijevic, A. *et al.* DNA bending by the human damage recognition complex XPC-HR23B. *DNA Repair (Amst)* **2**, 325–336 (2003).
- Kalodimos, C. G. *et al.* Structure and flexibility adaptation in nonspecific and specific protein–DNA complexes. *Science* **305**, 386–389 (2004).
- Parker, J. B. *et al.* Enzymatic capture of an extrahelical thymine in the search for uracil in DNA. *Nature* **449**, 433–437 (2007).
- Banerjee, A., Yang, W., Karplus, M. & Verdine, G. L. Structure of a repair enzyme interrogating undamaged DNA elucidates recognition of damaged DNA. *Nature* **434**, 612–618 (2005).
- Banerjee, A., Santos, W. L. & Verdine, G. L. Structure of a DNA glycosylase searching for lesions. *Science* **311**, 1153–1157 (2006).
- Yi, C. *et al.* Duplex interrogation by a direct DNA repair protein in search of base damage. *Nat. Struct. Mol. Biol.* **19**, 671–676 (2012).
- Viadiu, H. & Aggarwal, A. K. Structure of BamHI bound to nonspecific DNA: a model for DNA sliding. *Mol. Cell.* **5**, 889–895 (2000).
- Stivers, J. T., Pankiewicz, K. W. & Watanabe, K. A. Kinetic mechanism of damage site recognition and uracil flipping by Escherichia coli uracil DNA glycosylase. *Biochemistry* **38**, 952–963 (1999).
- Kuznetsov, S. V., Sugimura, S., Vivas, P., Crothers, D. M. & Ansari, A. Direct observation of DNA bending/unbending kinetics in complex with DNA-bending protein IHF. *Proc. Natl Acad. Sci. USA* **103**, 18515–18520 (2006).
- Camenisch, U. *et al.* Two-stage dynamic DNA quality check by xeroderma pigmentosum group C protein. *EMBO J.* **28**, 2387–2399 (2009).
- Moe, J. G. & Russo, I. M. Kinetics and energetics of base-pair opening in 5'-d(CGCGAATTCGCG)-3' and a substituted dodecamer containing G.T mismatches. *Biochemistry* **31**, 8421–8428 (1992).
- Zheng, H. *et al.* Base flipping free energy profiles for damaged and undamaged DNA. *Chem. Res. Toxicol.* **23**, 1868–1870 (2010).
- Friedman, J. I. & Stivers, J. T. Detection of damaged DNA bases by DNA glycosylase enzymes. *Biochemistry* **49**, 4957–4967 (2010).
- Schonhoft, J. D. & Stivers, J. T. Timing facilitated site transfer of an enzyme on DNA. *Nat. Chem. Biol.* **8**, 205–210 (2012).
- Blainey, P. C., van Oijen, A. M., Banerjee, A., Verdine, G. L. & Xie, X. S. A base-excision DNA-repair protein finds intrahelical lesion bases by fast sliding in contact with DNA. *Proc. Natl Acad. Sci. USA* **103**, 5752–5757 (2006).
- Gorman, J. *et al.* Dynamic basis for one-dimensional DNA scanning by the mismatch repair complex Msh2–Msh6. *Mol. Cell.* **28**, 359–370 (2007).
- Kropachev, K. *et al.* Adenine–DNA adducts derived from the highly tumorigenic Dibenzo[a,h]pyrene are resistant to nucleotide excision repair while guanine adducts are not. *Chem. Res. Toxicol.* **26**, 783–793 (2013).
- Hopfield, J. J. Kinetic proofreading: a new mechanism for reducing errors in biosynthetic processes requiring high specificity. *Proc. Natl Acad. Sci. USA* **71**, 4135–4139 (1974).
- Ninio, J. Kinetic amplification of enzyme discrimination. *Biochimie* **57**, 587–595 (1975).
- Savir, Y. & Tlusty, T. RecA-mediated homology search as a nearly optimal signal detection system. *Mol. Cell.* **40**, 388–396 (2010).

41. Otwinowski, Z. & Minor, W. in *Methods in Enzymology*. (ed Charles W., Carter Jr.) (Academic Press, 1997).
42. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221 (2010).
43. Kuznetsov, S. V., Ren, C. C., Woodson, S. A. & Ansari, A. Loop dependence of the stability and dynamics of nucleic acid hairpins. *Nucleic Acids Res.* **36**, 1098–1112 (2008).
44. Lu, X. J. & Olson, W. K. 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Res.* **31**, 5108–5121 (2003).
45. Coman, D. & Russu, I. M. A nuclear magnetic resonance investigation of the energetics of basepair opening pathways in DNA. *Biophys. J.* **89**, 3285–3292 (2005).

### Acknowledgements

We thank the staff of the Advanced Photon Source LS-CAT and SBC-CAT beamlines for help with data collection. This work was funded by the Chicago Biomedical Consortium with support from the Searle Funds at The Chicago Community Trust (to C.H. and J.-H.M.), NIH grants GM071440 (to C.H.) and 1R01ES019566 (to B.V.H.), NSF grants MCB-0721937 and MCB-1158217 (to A.A.), the Chancellor's Discovery Fund (to A.A. and J.-H.M.) and a startup fund from the University of Illinois at Chicago (to J.-H.M.). The open access publishing fee for this article is supported in part by the Research Open Access Publishing (ROAAP) Fund of the University of Illinois at Chicago.

### Author contributions

X.C. carried out protein engineering and purifications, protein–DNA crosslinking and crystallization experiments with contributions from B.P. and Y.S. G.Z. and C.H.

synthesized modified oligonucleotides for crosslinking experiments. X.C., Y.K. and J.-H.M. collected and analysed crystallographic data. J.-H.M. and Y.K. did model building and refinement. Y.V. and A.A. designed the fluorescence measurements on complexes with 2AP-labelled DNA substrates with contributions from X.C. and J.-H.M. Y.V. carried out the equilibrium and T-jump experiments and analysed the relaxation traces to obtain the DNA-opening times. L.L. and B.V.H. carried out the AFM studies. A.A. and J.-H.M. wrote the manuscript with contributions from all authors.

### Additional information

**Accession codes:** The structural factor have been deposited with the protein data bank (PDB) under accession number 4U29.

**Supplementary Information** accompanies this paper at <http://www.nature.com/naturecommunications>

**Competing financial interests:** The authors declare no competing financial interests.

**Reprints and permission** information is available online at <http://npg.nature.com/reprintsandpermissions/>

**How to cite this article:** Chen, X. *et al.* Kinetic gating mechanism of DNA damage recognition by Rad4/XPC. *Nat. Commun.* 6:5849 doi: 10.1038/ncomms6849 (2015).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>