13. Fuentes-Prior, P. *et al.* Structure of the thrombin complex with triabin, a lipocalin-like exosite-binding inhibitor derived from a triatomine bug. *Proc. Natl Acad. Sci. USA* **94,** 11845–11850 (1997).
14. Zushi, M. *et al.* Aspartic acid 349 in the fourth epidermal growth factor-like structure of human thrombomodulin plays a role in its Ca^{2+}-mediated binding to protein C. *J. Biol. Chem.* **266,** 19886–19889 (1991).
15. Campbell, I. D. & Bork, P. Epidermal growth factor-like modules. *Curr. Opin. Struct. Biol.* **3,** 385–392 (1993).
16. Downing, A. K. *et al.* Solution structure of a pair of calcium-binding epidermal growth factor-like domains: implications for the Marfan syndrome and other genetic disorders. *Cell* **85,** 597–605 (1996).
17. Knobe, K. E. *et al.* Probing the activation of protein C by the thrombin-thrombomodulin complex using structural analysis, site-directed mutagenesis, and computer modeling. *Proteins* **35,** 218–234 (1999).
18. Glaser, C. B. *et al.* Oxidation of a specific methionine in thrombomodulin by activated neutrophil products blocks cofactor activity. A potential rapid mechanism for modulation of coagulation. *J. Clin. Invest.* **90,** 2565–2573 (1992).
19. Clarke, J. H. *et al.* The short loop between epidermal growth factor-like domains 4 and 5 is critical for human thrombomodulin function. *J. Biol. Chem.* **268,** 6309–6315 (1993).
20. Nagashima, M., Lundh, E., Leonard, J. C., Morser, J. & Parkinson, J. F. Alanine-scanning mutagenesis of the epidermal growth factor-like domains of human thrombomodulin identifies critical residues for its cofactor activity. *J. Biol. Chem.* **268,** 2888–2892 (1993).
21. Light, D. R. *et al.* The interaction of thrombomodulin with Ca^{2+}. *Eur. J. Biochem.* **262,** 522–533 (1999).
22. Adler, M. *et al.* The structure of a 19-residue fragment from the C-loop of the fourth epidermal growth factor-like domain of thrombomodulin. *J. Biol. Chem.* **270,** 23366–23372 (1995).
23. Sampoli Benitez, B. A., Hunter, M. J., Meininger, D. P. & Komives, E. A. Structure of the fifth EGF-like domain of thrombomodulin: An EGF-like domain with a novel disulfide-bonding pattern. *J. Mol. Biol.* **273,** 913–926 (1997).
24. Mather, T. *et al.* The 2. 8 Å crystal structure of Gla-domainless activated protein C. *EMBO J.* **15,** 6822–6831 (1996).
25. Hogg, P. J., Ohlin, A. K. & Stenflo, J. Identification of structural domains in protein C involved in its interaction with thrombin-thrombomodulin on the surface of endothelial cells. *J. Biol. Chem.* **267,** 703–706 (1992).
26. Martin, P. D., Malkowski, M. G., Box, J., Esmon, C. T. & Edwards, B. F. New insights into the regulation of the blood clotting cascade derived from the X-ray crystal structure of bovine meizothrombin des F1 in complex with PPACK. *Structure* **5,** 1681–1693 (1997).
27. Gerlitz, B. & Grinnell, B. W. Mutation of protease domain residues Lys37-39 in human protein C inhibits activation by the thrombomodulin-thrombin complex without affecting activation by free thrombin. *J. Biol. Chem.* **271,** 22285–22288 (1996).
28. Vincenot, A., Gaussem, P., Pittet, J. L., Debost, S. & Aiach, M. Amino acids 225-235** of the protein C serine-protease domain are important for the interaction with the thrombin-thrombomodulin complex. *FEBS Lett.* **367,** 153–157 (1995).
29. Parry, M. A. *et al.* The ternary microplasmin-staphylokinase-microplasmin complex is a proteinase-cofactor-substrate complex in action. *Nature Struct. Biol.* **5,** 917–923 (1998).
30. Leslie, A. in *Crystal. Computing V* (eds Moras, D., Podjarny, A. D. & Thierry, J. C.) 27–38 (Oxford Univ. Press, Oxford, 1991).
31. Collaborative Computational Project No. 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50,** 760–763 (1994).
32. Navaza, J. An automated package for molecular replacement. *Acta Crystallogr. A* **50,** 157–163 (1994).
33. Brünger, G. J. *XPLOR (version 3. 1) A System for X-ray Crystallography and NMR* (Yale Univ. Press, New Haven, Connecticut, 1993).
34. Meininger, D. P., Hunter, M. J. & Komives, E. A. Synthesis, activity, and preliminary structure of the fourth EGF-like domain of thrombomodulin. *Protein Sci.* **4,** 1683–1695 (1995).
35. Weisel, J. W., Nagaswami, C., Young, T. A. & Light, D. R. The shape of thrombomodulin and interactions with thrombin as determined by electron microscopy. *J. Biol. Chem.* **271,** 31485–31490 (1996).
36. Nicholls, A., Bharadwaj, R. & Honig, B. GRASP- graphical representation and analysis of surface properties. *Biophys. J.* **64,** A166 (1993).
37. Grinnell, B. W., Gerlitz, B. & Berg, D. T. Identification of a region in protein C involved in thrombomodulin-stimulated activation by thrombin: potential repulsion at anion-binding site I in thrombin. *Biochem. J.* **303,** 929–933 (1994).

Acknowledgements

We thank J. McRobbie, Berlex Biosciences, for help and advice on the production of the medium containing TME456, G. Bourenkov for help during data collection at DESY, R. Mentele for sequencing the crystallized material, and T. Mather and M. Bauer for initial crystallization attempts. W.B. acknowledges the financial support of the EU research programs 'Biomed', 'Training and Mobility' and 'Biotechnology', HSFP, the Fonds der Chemischen Industrie and the Sonderforschungsbereich 469.

Correspondence and requests for material should be addressed to P.F-P. and W.B. (e-mail: fuentes@biochem.mpg.de, bode@biochem.mpg.de). Atomic coordinates have been deposited in the Protein Data Bank under accession code 1dx5.

correction

The thymine glycosylase MBD4 can bind to the product of deamination at methylated CpG sites

Brian Hendrich, Ulrike Hardeland, Huck-Hui Ng, Josef Jiricny & Adrian Bird

Nature **401,** 301–304 (1999)

It has been drawn to our attention that we inadvertently miscited work leading to the purification and cloning of the UV endonuclease from *Micrococcus luteus*. Reference 19 should have been to Piersen, C. E., Prince, M. A., Augustine, M. L., Dodson, M. L. & Lloyd, R. S. *J. Biol. Chem.* **270,** 23475–23484 (1995). □

erratum

DNA-bound structures and mutants reveal abasic DNA binding by APE1 and DNA repair coordination

Clifford D. Mol, Tadahide Izumi, Sankar Mitra & John A. Tainer

Nature **403,** 451–456 (2000)

The title of this Letter contained an error. The correct title is as printed above.

The thymine glycosylase MBD4 can bind to the product of deamination at methylated CpG sites

Brian Hendrich*†, Ulrike Hardeland†‡, Huck-Hui Ng*, Josef Jiricny‡ & Adrian Bird*

** Institute of Cell and Molecular Biology, University of Edinburgh, The King's Buildings, Mayfield Road, Edinburgh EH9 3JR, UK*
‡ Institute of Medical Radiobiology, August Forel-Strasse 7, CH-8008 Zürich, Switzerland
† These authors contributed equally to this work.

In addition to its well-documented effects on gene silencing, cytosine methylation is a prominent cause of mutations. In humans, the mutation rate from 5-methylcytosine (m^5C) to thymine (T) is 10–50-fold higher[1–4] than other transitions and the methylated sequence CpG is consequently under-represented[5]. Over one-third of germline point mutations associated with human genetic disease[6] and many somatic mutations leading to cancer[7,8] involve loss of CpG. The primary cause of mutability appears to be hydrolytic deamination. Cytosine deamination produces mismatched uracil (U), which can be removed by uracil glycosylase[9,10], whereas m^5C deamination generates a G·T mispair that cannot be processed by this enzyme. Correction of m^5CpG·TpG mismatches may instead be initiated by the thymine DNA glycosylase, TDG[11,12]. Here we show that MBD4, an unrelated mammalian protein that contains a methyl-CpG binding domain[13,14], can also efficiently remove thymine or uracil from a mismatches CpG site *in vitro*. Furthermore, the methyl-CpG binding domain of MBD4 binds preferentially to m^5CpG·TpG mismatches—the primary product of deamination at methyl-CpG. The combined specificities of binding and catalysis indicate that this enzyme may function to minimize mutation at methyl-CpG.

MDB4 (ref. 14) was identified in a database search for proteins with a methyl-CpG-binding domain (MBD) resembling that of the transcriptional repressor MeCP2 (refs 13, 15). Hypothetical translation of the full-length human and mouse MBD4 complementary DNAs revealed an amino-terminal methyl-CpG-binding domain and a carboxy-terminal region that was closely related to the glycosylase/endonuclease domains of bacterial repair proteins (Fig. 1a). Related proteins include the 8-oxoG·A mispair-specific adenine glycosylase MutY of *Escherichia coli*[16], the G·T mismatch-specific thymine glycosylase Mig of *Methanobacterium thermoautotrophicum*[17], the thymine glycol glycosylase EndoIII of *E. coli*[18] and the photodimer-specific UV-endonuclease of *Micrococcus luteus*[19] (Fig. 1b). Mouse and human MBD4 proteins are 86% and 95% identical in the methyl-CpG-binding domain and the C-terminal glycosylase-like domain, respectively, but are less conserved elsewhere (Fig. 1a).

On the basis of these relationships, we asked whether human MBD4 has DNA glycosylase or endonuclease activity. The presence of a methyl-CpG-binding domain led us to postulate that MBD4 might be involved in the processing of DNA damage associated with methylated CpG sites. Upon incubation of recombinant human MBD4 with a fluorescently labelled oligonucleotide duplex containing a single G·T mismatch in a CpG context, an abasic site was generated by the removal of the mispaired pyrimidine by glycosidic bond cleavage (Fig. 2a). In addition to the G·T mispair, G·U substrates were also cleaved (Fig. 2b), but the processing of other base/base mismatches, including C·A, C·C and G·G, was not observed (data not shown). That MBD4 is a mismatch specific T/U DNA glycosylase was confirmed by high-performance liquid chromatography (HPLC) analysis of the product mixture. Free T or U could be detected in reaction mixtures which contained MBD4 and G·T or G·U mismatches, respectively (data not shown). The sugar-phosphate backbone of DNA was not cleaved unless treated with hot alkali or the human AP-endonuclease HAP1 (ref. 20; and Fig. 2a). HAP1 cleaves oligonucleotide substrates at the 5′ end of abasic sites. In the assay shown in Fig. 2a, HAP1 generates two fragments: a 5′ fluorescein-labelled 23-mer terminated at its 3′ end with a hydroxy group; and a 37-mer that has at its 5′ end the baseless sugar phosphate from which the mispaired T has been removed by the action of MBD4. Neither glycosylase nor endonuclease activity could be detected on perfectly matched substrates containing symmetrically methylated, hemimethylated or non-methylated CpG sites (Fig. 2b). From these experiments, we conclude that MBD4 is a mismatch-specific T/U glycosylase that lacks endonuclease activity.

A CpG sequence context for the mismatch was preferred, but was not absolutely required, as G·T and G·U mismatches in other sequences were also processed, albeit at a reduced rate (Fig. 2c;

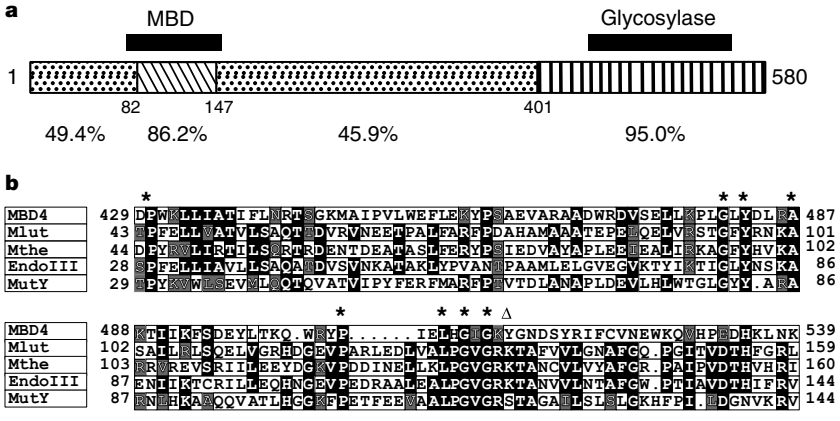

Figure 1 MBD4 contains a conserved glycosylase-like domain and a methyl-CpG-binding domain. **a**, Diagram of MBD4 showing the N-terminal MBD and the C-terminal glycosylase domain. Percentage conservation between human and mouse proteins is shown below each shaded domain. **b**, Amino-acid sequence alignment of human MBD4 (amino acids 429–539) with the *E. coli* thymine glycol glycosylase EndoIII, the 8-oxoG·A-specific adenine glycosylase MutY, the Mig G·T glycosylase of *M. thermoautotrophicum* (Mthe) and the *M. luteus* UV endonuclease (Mlut). Asterisks mark residues found in all the proteins; triangle indicates the residue that, when basic (K), is diagnostic of an endonuclease (glycosylase/lyase).

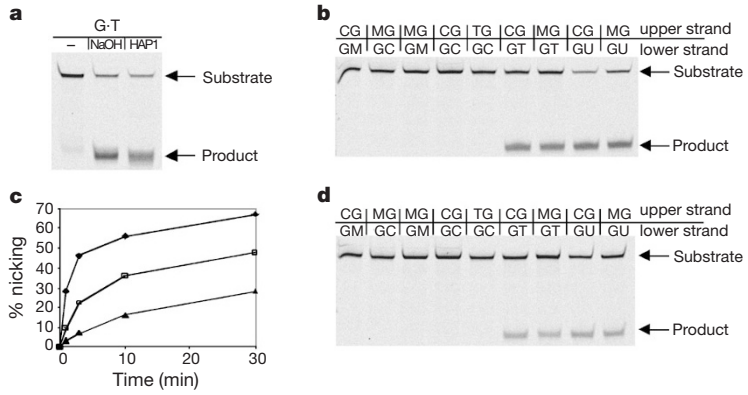

Figure 2 MBD4 is a mismatch-specific T/U DNA glycosylase. **a**, MBD4 does not cleave the sugar-phosphate backbone of a mismatched DNA substrate (lane –). Cleavage occurs only upon the addition of hot alkali (lane NaOH) or human HAP1 endonuclease (lane HAP1). **b**, G·T and G·U mispairs are substrates for T glycosylase processing by full-length MBD4. Oligonucleotide substrate JJ, containing the appropriate mismatches or the corresponding perfectly matched, methylated or unmethylated controls, was incubated with MBD4 (see Methods). Apyrimidinic sites were then cleaved with hot alkali. Substrate and product bands are indicated. **c**, Kinetics of MBD4 glycosylase action in CG and non-CG sequence contexts. G·U (diamonds) or G·T (squares) in a CG context is compared with G·T in a GG context (triangles). **d**, A C-terminal fragment of MBD4 that includes the glycosylase domain (amino acids 379–580; see Fig. 4a) can process G·U or G·T mismatches. **a**, **b** and **d** are fluorescence scans of denaturing 15% polyacrylamide gels. M denotes m^5C.

and data not shown). The methylation status of cytosine in CpG·TpG or CpG·UpG mismatches did not affect the glycosylase reaction in this assay (Fig. 2b). Substrate specificity *in vitro* is therefore similar to that of the other known mammalian G·T mismatch-specific DNA glycosylase, TDG[11,12]. As anticipated from the amino-acid sequence alignment (Fig. 1b), the glycosylase activity resides in the C-terminal region of MBD4, as a polypeptide corresponding to amino acids 379–580 could process G·T and G·U mispairs (Fig. 2d), whereas polypeptides in which the C terminus was deleted were inactive (data from Figs 2 and 3 are summarized in Fig. 4a; and data not shown). The substrate specificity and flanking sequence preference of full-length MBD4 and its C-terminal fragments were similar (compare Fig. 2b, d), indicating that the methyl-CpG-binding domain (see below) did not influence site preference in this assay.

Next, we investigated the DNA-binding specificity of MBD4. Mouse MBD4 binds to densely methylated DNA molecules *in vitro* and localizes to heavily methylated foci in mouse cell nuclei when overexpressed[14]. Bandshift assays with probes that contained non-base-paired regions, however, revealed additional specificity. Human MBD4 gave a strong complex with a probe that had three m^5CpG·TpG mismatches but bound relatively weakly when m^5CpG·m^5CpG pairs were substituted at these sites (Fig. 3a, left panel). The probe with non-methylated, mismatched CpG·TpG sites also gave a weak complex. We conclude that both G·T mismatch and m^5C moieties are important for recognition by MBD4. Dependence of mismatch binding on the methyl-CpG-binding domain was shown by bandshift assays with truncated versions of MBD4. The N-terminal 165 residues, which include the DNA-binding domain, retained specificity for the methylated mismatch (Fig. 3a, right panel), whereas fragments that excluded all or part of the domain could not bind any of the probes (Fig. 4a; polypeptides 1–580 and 379–580). The results show that MBD4 preferentially binds to the mismatch that results from hydrolytic deamination of m^5C in mammalian genomic DNA.

The complexes in Fig. 3a were formed under conditions in which processing of the mismatched site is undetectable (data not shown). After incubation for 20 min at 37 °C, MBD4–DNA complexes reproduced the specificity of the glycosylase reaction, which prefers mismatches in a CpG context, but is indifferent to methylation (Fig. 2b). The full-length protein and a C-terminal fragment lacking the methyl-CpG binding domain (379–580; Fig. 3c) formed a complex with each of the mismatched probes (Fig. 3b), but a truncation lacking the C-terminal glycosylase-like domain did not form the complex (data not shown). As other DNA glycosylases bind tightly to the abasic site resulting from their catalytic action[21], we suspected that the observed complexes also represented product binding by MBD4. Accordingly, we found that the glycosylase domain (379–580) and the intact form of MBD4 bound efficiently to abasic sites under the same conditions (Fig. 3c; and data not shown). Thus the N terminus of MBD4 binds to the MG·TG mismatch that is the substrate for base removal, whereas the C terminus, following processing, binds to the abasic product of the reaction.

MBD4 shows no apparent amino-acid sequence similarity to TDG, a protein that also excises mispaired T and U from G·T and G·U mismatches[11,12]. The two proteins probably evolved from different ancestors, as TDG is related to a bacterial G·U-processing enzyme, Mug[22], whereas MBD4 seems to have arisen through the fusion of a glycosylase domain, such as that of bacterial MutY, to an N terminus that may target the enzyme to methylated mismatches. TDG does not contain a methyl-CpG-binding domain. Despite these sequence differences, MBD4 and TDG show very similar catalytic specificities *in vitro*, both enzymes preferring G·T and G·U mismatches within a CpG context[23]. In addition, both proteins bind to the apyrimidinic site that arises following mismatch processing (ref. 21; and Fig. 3b, c).

MED1 (identical to MBD4) is reportedly a binding partner of MLH1, a protein implicated in mismatch repair[24]. It was proposed that MBD4/MED1 is an endonuclease that binds to hemi-methylated CpG sites and directs mismatch repair via a single strand nick towards the newly replicated strand, in a manner analogous to bacterial MutH[25,26]. Against this hypothesis, we find that MBD4 lacks endonuclease activity. The weak binding of MBD4 to hemi-methylated CpG sites compared with symmetrically methylated CpG[24] or m^5CpG·TpG mismatches also argues against the MutH homologue model.

We propose instead that the function of MBD4 is to counteract the mutability of m^5C by initiating conversion of m^5CpG·TpG mismatches back to m^5CpG·CpG. Among the sequences that we have tested, only m^5CpG·TpG shows both preferential binding by the N terminus of MBD4 and processing by its catalytic domain

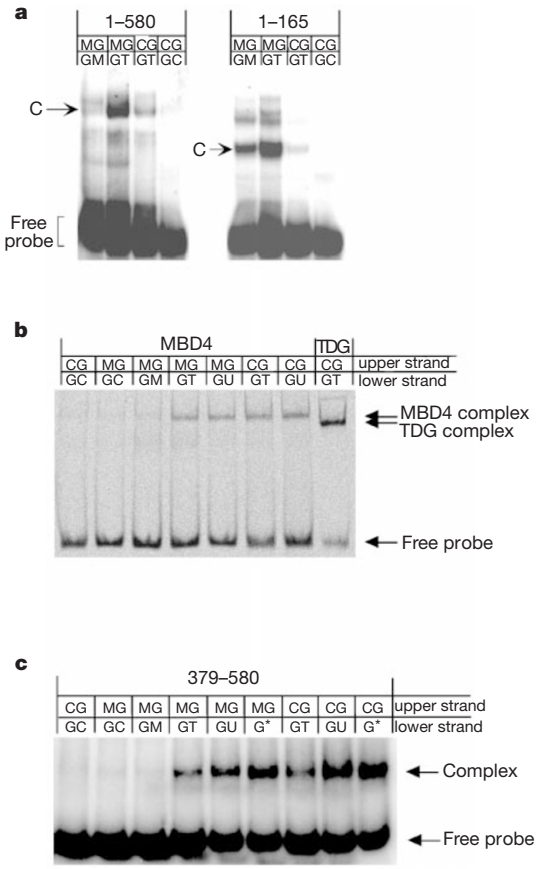

Figure 3 DNA-binding specificities of MBD4 and its truncated derivatives to DNA containing methylated and non-methylated duplexes with or without G·U or G·T mismatches. **a**, Binding of full-length MBD4 (left) and the N-terminal 165 amino acids of MBD4 (right) to probe 'BH' (see Methods), which contains three potential CpG mismatch sites. Within each probe, all three sites were CpG·CpG duplex (CG·CG), m^5CpG·m^5CpG duplex (MG·MG), m^5CpG·TpG mismatch (MG·TG) or CpG·TpG mismatch (CG·TG). Major complexes (C) are indicated. The slower migrating complexes in the MG·MG and MG·TG lanes of the 1–165 panel may be due to binding of more than one protein molecule to the probe. **b**, Full-length MBD4 incubated under conditions that favour the glycosylase reaction with fluorescently labelled probe JJ, in which a single CpG site was methylated or mismatched as shown (see Methods). Complexes were formed with G·U or T·G mismatches regardless of methylation status. For comparison, the complex between CG·TG mismatch and the T glycosylase TDG is shown (lane TDG). **c**, Under the same conditions, a C-terminal fragment (amino acids 379–580; see Fig. 4a) that can carry out the glycosylase reaction complexed with all mismatches in probe JJ, which was mismatched or methylated as shown. Asterisk denotes an abasic site.

(Fig. 4b). Although processing of mismatches *in vitro* was not accelerated by methylation, it is possible that *in vivo* the N-terminal domain targets MBD4 to these sites as they arise by hydrolytic deamination of m^5C (Fig. 4c). Repair of the resulting abasic site may involve interaction with MLH1 (ref. 24). Given the known importance of CpG to TpG transitions in genetic disease and cancer, it will be of interest to test the effect of MBD4 loss on mutation frequency and cancer predisposition. □

Methods

Protein expression.

MBD4 deletion constructs were made by PCR amplification from an MBD4 cDNA and cloning the resulting products into the pET6H vector. Recombinant proteins were expressed and purified as described[14]. N-terminal constructs were made by amplifying with the 5′ primer 5′-CCTGCTCCATGGGCACGACTGGGCTG-3′ and one of the

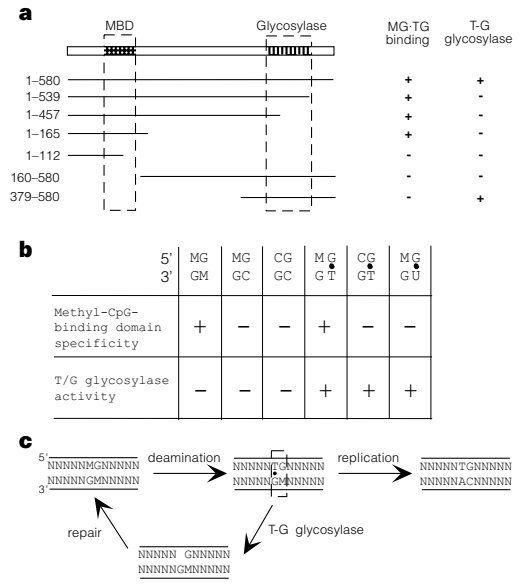

Figure 4 The properties of MBD4 suggest a role in initiating repair of m^5CpG·TpG mismatches. **a**, Summary of DNA-binding and T glycosylase activity of truncated MBD4 polypeptides against m^5CpG·TpG (data is from Figs 2 and 3; and data not shown). The failure of the 160–580 protein to process mismatches is unexplained (aberrant protein folding?). **b**, Of the DNA sequences tested, only the methylated mismatch m^5CpG·TpG is both bound by the N terminus of MBD4 and processed by its glycosylase domain. **c**, A scheme for the role of T·G glycosylases in initiation of repair of deaminated methyl-CpG pairs. Solid circles in **b** and **c** indicate interstrand base mismatches.

following 3′ primers: (580) 5′-GCGGGATCCTGAGCTTGAAAGCTGCAG-3′; (539) 5′-GCGGGTACCAATCCCATGAAGCTC-3′; (165) 5′-GCGGATCCGATGTC-AGGGCTGCCATG-3′; (112) 5′-GCGCAGCTGATAAAGTACACACTC-3′. The 1–457 construct was made by cloning the N-terminal *Nco*I fragment of the 1–580 construct into pET6H. C-terminal constructs were made by amplifying with the 580 primer (above) and one of the following 5′ primers: (160) 5′-CGCCATGGCAGCCCTGACATC-3′; (379) 5′-GCGCCATGGGCTCTGAAATGGACAACAAC-3′.

Enzymatic activity assay.

The enzymatic activity of the wild-type and mutant recombinant proteins was monitored using a 'nicking assay'. A 60-mer double-stranded oligonucleotide substrate containing different mismatches was prepared by annealing oligonucleotides to give the 'JJ' duplex: 5′-TAGACATTGCCCTCGAGGTACCATGGATCCGATGTZRACCTCAAACCTAGACGAATTCCG-3′ 3′-ATCTGTAACGGGAGCTCCATGGTACCTAGGCTACAGYTGGAGTTTGGATCTGCTTAAGGC-F-5′ where F is fluorescein and R = G or A, Y = C, m^5C, T or U, and Z = C or m^5C (see underlined dinucleotide). A 0.5 μM solution of the labelled lower strand and 1 μM solution of the unlabelled upper strand were heated for 5 min at 95 °C in 10 mM Tris-HCl pH 8.0, 50 mM NaCl, and the solution was allowed to cool slowly to room temperature. In standard 'nicking reactions', 50 nM substrate and 50 nM protein were incubated for 1 h at 37 °C in a total volume of 20 μl in 1× nicking buffer (50 mM Tris-HCl pH 8.0, 1 mM DTT, 0.1 mg ml^{-1} BSA, 1 mM EDTA). The reaction was stopped by adding NaOH to a final concentration of 90 mM and heating for 10 min at 99 °C. The NaOH treatment cleaves the AP sites resulting from the removal of the mispaired bases by the glycosylase. After adding 0.5 μl 10 mg ml^{-1} transfer RNA, 1/10 vol. 3 M NaOAc pH 5.2 and 3 vol. cold ethanol (−20 °C) to the reaction mixture, the DNA was precipitated at −20 °C for 1 h, pelleted by centrifugation and washed with 80% ethanol (−20 °C). The dried DNA was resuspended in 10 μl formamide buffer (90% formamide, 1× TBE) and heated for 5 min at 99 °C before loading on a 15% denaturing polyacrylamide gel. The fluorescein-labelled DNA fragments were detected using the blue fluorescence mode of the PhosphorImager (Storm 860, Applied Biosystems). The HAP1 endonuclease was expressed and purified from a clone donated by I. Hickson.

In some experiments, glycosylase activity was also assayed by incubating 1–10 ng recombinant protein with a radiolabelled 27-mer oligonucleotide,

5′-TCAGATTCGCGCMGGCTGCGATAAGCT

3′-AGTCTAAGCGCGGTCGACGCTATTCGA

(where M signifies m^5C), in 25 mM sodium phosphate pH 7.2, 10 mM EDTA, 50 mM NaCl and 1 mg ml^{-1} BSA for 30 min at 37 °C. Piperidine was added to 1 M and the reaction was incubated at 90 °C for 30 min, after which the reactions were completely dried under vacuum. Samples were resuspended in 90% formamide gel loading buffer[27], denatured at 95 °C for 5 min, and separated on a 15% denaturing polyacrylamide gel.

Bandshift assays.

Standard bandshift reactions (Fig. 3a) utilized ^{32}P-labelled 'BH' probes that contained three CpG sites (underlined below) whose methylation/mismatch status was varied. The BH probes used in Fig. 3c were

5′TCAGATTCGCGCZGGCTGCGATAAGCTGZGCGGATCCZGGGAATTCAGCT3′

3′AGTCTAAGCGCGGYCGACGCTATTCGACGYGCCTAGGGYCCTTAAGTCGA5′

where Y = C, m^{5}C, T or U, and Z = C or m^{5}C (see underlined dinucleotides). The three CpG sites were identically modified/mismatched within each probe. Binding reactions were carried out at room temperature for 30 min in 20 mM HEPES pH 7.9, 25 mM NaCl, 10 mM β-mercaptoethanol, 1 mM EDTA, 4% glycerol, 1% digitonin and 50 ng sonicated *E. coli* DNA. The glycosylase reaction does not occur under these conditions (data not shown). Complexes were electrophoresed through 6% polyacrylamide gels in 0.5 × TBE at 4 °C.

Complexes formed under conditions favourable to the glycosylase reaction (Fig. 3b, c) utilized the fluorescent or ^{32}P-labelled JJ oligonucleotide (see above) as a probe. In standard gel-retardation reactions, 200 nM protein was incubated with 66 nM labelled oligonucleotide substrate and 333 nM unlabelled homoduplex oligonucleotide in 50 mM Tris–HCl pH 8.0, 1 mM DTT, 5% glycerol, 1 mM EDTA at 37 °C for 20 min. The samples were electrophoresed immediately through a 6% native 0.5 × TBE polyacrylamide gel for 45 min at 100 V. A probe with an abasic site was generated by treatment of oligonucleotide JJ containing a MG·UG mismatch with the enzyme uracil DNA glycosylase.

Received 25 June; accepted 21 July 1999.

1. Duncan, B. K. & Miller, J. H. Mutagenic deamination of cytosine residues in DNA. *Nature* **287,** 560–561 (1980).
2. Britten, R. J., Baron, W. F., Stout, D. B. & Davidson, E. H. Sources and evolution of human Alu repeated sequences. *Proc. Natl Acad. Sci. USA* **85,** 4770–4774 (1988).
3. Sved, J. & Bird, A. The expected equilibrium of the CpG dinucleotide in vertebrate genomes under a mutation model. *Proc. Natl Acad. Sci. USA* **87,** 4692–4696 (1990).
4. Bulmer, M. Neighboring base effects on substitution rates in pseudogenes. *Mol. Biol. Evol.* **3,** 322–329 (1986).
5. Bird, A. P. DNA methylation and the frequency of CpG in animal DNA. *Nucleic Acids Res.* **8,** 1499–1594 (1980).
6. Cooper, D. N. & Youssoufian, H. The CpG dinucleotide and human genetic disease. *Hum. Genet.* **78,** 151–15 (1988).
7. Hollstein, M. *et al.* Database of p53 gene somatic mutations in human tumors and cell lines. *Nucleic Acids Res.* **22,** 3551–3555 (1994).
8. Jones, P. A., Rideout, W. M., Shen, J.-C., Spruck, C. H. & Tsai, Y. C. Methylation, mutation and cancer. *BioEssays* **14,** 33–36 (1992).
9. Lindahl, T. An N-glycosidase from *Escherichia coli* that releases free uracil from DNA containing deaminated cytosine residues. *Proc. Natl Acad. Sci. USA* **71,** 3649–3653 (1974).
10. Lindahl, T., Karran, P. & Wood, R. D. DNA excision repair pathways. *Curr. Opin. Genet. Dev.* **7,** 158–169 (1997).
11. Wiebauer, K. & Jiricny, J. *In vitro* correction of G·T mispairs to G·C pairs in nuclear extracts from human cells. *Nature* **339,** 234–236 (1989).
12. Neddermann, P. *et al.* Cloning and expression of human G/T mismatch-specific thymine-DNA glycosylase. *J. Biol. Chem.* **271,** 12767–12774 (1996).
13. Nan, X., Meehan, R. R. & Bird, A. Dissection of the methyl-CpG binding domain from the chromosomal protein MeCP2. *Nucleic Acids Res.* **21,** 4886–4892 (1993).
14. Hendrich, B. & Bird, A. Identification and characterization of a family of mammalian methyl-CpG binding proteins. *Mol. Cell. Biol.* **18,** 6538–6547 (1998).
15. Nan, X., Campoy, J. & Bird, A. MeCP2 is a transcriptional repressor with abundant binding sites in genomic chromatin. *Cell* **88,** 471–481 (1997).
16. Michaels, M. L., Pham, L., Nghiem, M., Cruz, C. & Miller, J. H. MutY, an adenine glycosylase active on G-A mispairs, has homology to endonuclease III. *Nucleic Acids Res.* **18,** 3841–3845 (1990).
17. Horst, J. P. & Fritz, H. J. Counteracting the mutagenic effect of hydrolytic deamination of DNA 5-methylcytosine residues at high temperature: DNA mismatch N-glycosylase Mig.Mth of the thermophilic archaeon *Methanobacterium thermoactotrophicum* THF. *EMBO J.* **15,** 5459–5469 (1996).
18. Asahara, H., Wistort, P. M., Bank, J. F., Bakerian, R. H. & Cunningham, R. P. Purification and characterization of *Escherichia coli* endonuclease III from the cloned nth gene. *Biochemistry* **28,** 444–4449 (1989).
19. Shiota, S. & Nakayama, H. UV endonuclease of *Micrococcus luteus*, a cyclobutane pyrimidine dimer-DNA glycosylase/abasic lyase: cloning and characterization of the gene. *Proc. Natl Acad. Sci. USA* **94,** 593–598 (1997).
20. Robson, C. N. & Hickson, I. D. Isolation of cDNA clones encoding a human apurinic/apyrimidinic endonuclease that corrects DNA repair and mutagenesis defects in *E. coli* xth (exonuclease III) mutants. *Nucleic Acids Res.* **19,** 5519–5523 (1991).
21. Scharer, O. D., Nash, H. M., Jiricny, J., Laval, J. & Verdine, G. L. Specific binding of a designed pyrrolidine abasic site analog to multiple DNA glycosylases. *J. Biol. Chem.* **273,** 8592–8597 (1998).
22. Gallinari, P. & Jiricny, J. A new class of uracil-DNA glycosylases related to human thymine-DNA glycosylase. *Nature* **383,** 735–738 (1996).
23. Sibghat-Ullah *et al.* Base analog and neighboring base effects on substrate specificity of recombinant human G:T mismatch-specific thymine DNA glycosylase. *Biochemistry* **35,** 12926–12932 (1996).
24. Bellacosa, A. *et al.* MED1, a novel human methyl-CpG-binding endonuclease, interacts with DNA mismatch repair protein MLH1. *Proc. Natl Acad. Sci. USA* **96,** 3969–3974 (1999).
25. Modrich, P. & Lahue, R. Mismatch repair in replication fidelity, genetic recombination and cancer biology. *Annu. Rev. Biochem.* **65,** 101–133 (1996).
26. Jiricny, J. Replication errors: cha(lle)nging the genome. *EMBO J.* **17,** 6427–6436 (1998).
27. Maniatis, T., Fritsch, E. F. & Sambrook, J. *Molecular Cloning: A Laboratory Manual* 196–198 (Cold Spring Harbor, New York, 1982).

Acknowledgements

We thank P. Hunziger for HPLC analysis; P. Schär, R. Wood and T. Lindahl for fruitful discussions; I. Hickson for the HAP1 endonuclease; and S. Tweedie for comments on the manuscript. This work was supported by grants from the Wellcome Trust (B.H. and A.B.), the Schweizerische Krebsliga (U.H. and J.J.) and by a Darwin Trust Scholarship to H-H.N.

Correspondence and requests for materials should be addressed to A.B. (e-mail: A.Bird@ed.ac.uk).

erratum

Full sintering of powdered-metal bodies in a microwave field

Rustum Roy, Dinesh Agrawal, Jiping Cheng & Shalva Gedevanishvili

Nature **399,** 668–670 (1999)

In the penultimate paragraph of this Letter, there are two instances where "electric field" has been used when "magnetic field" is meant: one is in the fourth sentence and the other in the last. □

molar excess of γ^{32}P-labelled top strand in 10 mM MOPS, pH 6.5, 1 mM EDTA. The solution was heated to 95 °C for 2 min and cooled to room temperature over 90 min. Duplexes were separated from single-strand RNA by native PAGE, visualized by radiolytic scanning (Packard Instant Imager) and excised from the gel.

Unwinding reactions

Unwinding reactions were performed at room temperature in 30 µl of 40 mM Tris-HCl buffer, pH 8.0, and 4 mM $MgCl_2$ or $CoCl_2$. The reaction also contained 25 mM NaCl, which was introduced with the protein storage buffer. In a typical reaction, 1–2 nM RNA substrate was incubated with 10–15 nM NPH-II in reaction buffer without ATP at room temperature for 5–7 min. Longer pre-incubation time did not change the reaction kinetics. Saturation of substrate with protein before the reaction was verified by gel-shift analysis[15]. The unwinding reaction was initiated by adding the ATP to a final concentration of 3.5 mM, unless otherwise stated. Aliquots at the respective time points were quenched with two volumes of stop buffer (25 mM EDTA, 0.4% SDS, 0.05% BPB, 0.05% XCB, 10% glycerol) containing 200 nM of unlabelled top-strand RNA to prevent re-annealing of unwound duplexes during electrophoresis. Reannealing of unwound duplexes during the reaction did not affect unwinding kinetics under any reaction conditions, as verified by independent measurements of association rate constants of the substrate strands. Trap RNA (200 nM final concentration, added with the ATP) was a single-strand of 39 nucleotides of unrelated sequence. Higher concentrations of trap RNA did not change observed reaction amplitudes.

Unwinding reactions were also monitored in real time using fluorescence energy transfer experiments on substrates labelled at the 3′ and 5′ ends of the duplex terminus farthest from the single-stranded 3′ overhang. These measurements were superimposable with results from gel-shift measurements and confirmed the first-order nature of the time courses as well as the rate constants.

Kinetic analysis

Data were fit numerically using the FITSIM software package[21]. Time courses were corrected for the reaction amplitudes, and initial kinetic parameters were determined using the KINSIM software[22]. Unwinding step size and rate constants were determined from time courses in Co^{2+} using a kinetic model in which the unwinding reaction consists of N consecutive first-order reactions, where N is the number of kinetic steps. Rate constants (except for the first step, see below) were linked, that is, forced to yield the same value in the fitting. The linked rate constants and the first rate constant were allowed to float simultaneously. On the basis of the observation that an initial step is slower than subsequent steps in Mg^{2+} (see Fig. 3b and text), the first kinetic step was defined as having a different rate constant (k_I) than subsequent kinetic steps (k_U). Using this model, increasing numbers of steps (N) were examined iteratively for each duplex until k_I and k_U approached constant values for the different duplexes. Convergence occurred when $k_I = 2.3 \pm 0.2\,min^{-1}$ and $k_U = 13.4 \pm 0.8\,min^{-1}$.

Note that k_U is an apparent rate constant that depends on the ATP concentration according to $k_U = k_{U[S]}[ATP/(ATP + K'_d)]$, where $k_{U[S]}$ is the respective rate constant at ATP saturation. The step size (m) was calculated according to $m = L/(N - y)$, where y is the number of kinetic steps that do not contribute to the actual strand displacement and L is the duplex length. A step size of $m = 6$ and a value of $y = 1$ were consistently calculated for all duplexes including multi-piece substrates.

The lower limit for the translocation rate constant in Mg^{2+} was estimated by considering that a lag phase was not observed with the longest (83 bp) duplex (Fig. 1b). Assuming an unwinding step size identical to that in Co^{2+} (6 bp), and that k_U must be equal or greater than a value that does not cause an apparent lag phase in 14 unwinding steps, the translocation rate constant was estimated as $k_U > \sim 350\,min^{-1}$.

Equation (1) was derived from the expression describing the dependence of reaction amplitude (A) on the processivity (P)[13]:

$$A = (1 - x)\,P^{L/m} \qquad (2)$$

where x is the fraction of protein that dissociates from the substrate before unwinding occurs. P is then substituted with a term that relates processivity and ATP concentration under the conditions specified by the scheme in Fig. 5.

$$P = \frac{k_{U[S]}\left(\frac{[ATP]}{K'_d + [ATP]}\right)}{k_{U[S]}\left(\frac{[ATP]}{K'_d + [ATP]}\right) + k_F\left(1 - \frac{[ATP]}{K'_d + [ATP]}\right)} \qquad (3)$$

where K'_d is the apparent dissociation constant of ATP from the enzyme substrate complex.

Received 29 July; accepted 28 October 1999.

1. de la Cruz, J., Kressler, D. & Linder, P. Unwinding RNA in *Saccharomyces cerevisiae*: DEAD-box proteins and related families. *Trends Biochem. Sci.* **24,** 192–198 (1999).
2. Kadare, G. & Haenni, A. L. Virus encoded RNA helicases. *J. Virol.* **71,** 2583–2590 (1997).
3. Wagner, J. D. O., Jankowsky, E., Company, M., Pyle, A. M. & Abelson, J. N. The DEAH-box protein Prp22 is an ATPase that mediates ATP dependent mRNA release from the spliceosome and unwinds RNA duplexes. *EMBO J.* **17,** 2926–2937 (1998).
4. Staley, J. P. & Guthrie, C. Mechanical devices of the spliceosome: motors clocks springs and things. *Cell* **90,** 1041–1050 (1998).
5. Shuman, S. Vaccinia Virus RNA helicase: an essential enzyme related to the DE-H family of RNA dependent NTPases. *Proc. Natl Acad. Sci. USA* **89,** 10935–10939 (1992).
6. Shuman, S. Vaccinia Virus RNA helicase. *J. Biol. Chem.* **268,** 11798–11802 (1993).
7. Gross, C. H. & Shuman, S. Vaccinia virus RNA helicase: nucleic acid specificity in duplex unwinding. *J. Virol.* **70,** 2615–2619 (1996).
8. Ali, J. A. & Lohman, T. M. Kinetic measurement of the step size of DNA unwinding by *Escherischia coli* UvrD helicase. *Science* **275,** 377–380 (1997).
9. Shimamoto, N. One-dimensional diffusion of proteins along DNA. *J. Biol. Chem.* **274,** 15239–15296 (1999).
10. Kelemen, B. R. & Raines, R. T. Extending the limits to enzymatic catalysis: diffusion of ribonuclease A in one dimension. *Biochemistry* **27,** 5302–5307 (1999).
11. Jeltsch, A. & Pingoud, A. Kinetic characterization of linear diffusion of the restriction endonuclease EcoRV on DNA. *Biochemistry* **24,** 2160–2169 (1998).
12. Santoro, S. W. & Joyce, G. F. A general purpose RNA-cleaving DNA enzyme. *Proc. Natl Acad. Sci. USA* **94,** 4262–4266 (1997).
13. Lohman, T. M. & Bjoernson, K. P. Mechanisms of helicase catalyzed DNA unwinding. *Annu. Rev. Biochem.* **65,** 169–214 (1996).
14. Schnapp, B. Molecular Motors. Two heads are better than one. *Nature* **373,** 655–666 (1995).
15. Gross, C. H. & Shuman, S. Mutational analysis of vaccinia virus nucleoside triphosphate hydrolyase II, a DExH box RNA helicase. *J. Virol.* **69,** 4727–4736 (1995).
16. Jankowsky, E. & Schwenzer, B. Efficient improvement of hammerhead ribozyme mediated cleavage of long substrates by oligonucleotide facilitators. *Biochemistry* **37,** 15313–15321 (1996).
17. Xiang, Q., Qin, P., Michels, W. J., Freeland, K. & Pyle, A. M. Sequence specificity of a group II intron ribozyme: multiple mechanisms for promoting unusually high discrimination against mismatched targets. *Biochemistry* **37,** 3839–3849 (1998).
18. Wincott, F. *et al.* Synthesis, deprotection, analysis and purification of RNA and ribozymes. *Nucleic Acids Res.* **23,** 2677–2684 (1995).
19. Pyle, A. M., Chu, V. T., Jankowsky, E. & Boudvillain, M. Using DNAzymes to cut, process and map RNA molecules for structural studies or modification. *Methods Enzymol.* **317,** 140–146 (2000).
20. Moore, M. J. & Sharp, P. A. Site specific modification of pre-messenger RNA—the 2′-hydroxyl groups at the splice sites. *Science* **256,** 992–997 (1992).
21. Zimmerle, C. & Frieden, C. Analysis of progress curves by simulations generated by numerical integration. *Biochem. J.* **258,** 381–387 (1989).
22. Barshop, B. A. & Frieden, C. Analysis of numerical methods for computer simulation of kinetic processes: development of KINSIM—a flexible portable system. *Anal. Biochem.* **130,** 134–145 (1983).

Acknowledgements

We thank the members of the Pyle laboratory, particularly M. Boudvillain, J. B. Green and P. Z. Qin for many insightful discussions. E.J. was supported by a Curt-Engelhorn postdoctoral fellowship from the German Center for Cancer Research. A.M.P. is an assistant investigator of the Howard Hughes Medical Institute.

Correspondence and requests for materials should be addressed to A.M.P.

DNA-bound structures and mutants reveal abasic DNA binding by APE1 DNA repair and coordination

Clifford D. Mol*, Tadahide Izumi†, Sankar Mitra† & John A. Tainer*

** Skaggs Institute for Chemical Biology, and the Department of Molecular Biology, MB-4, The Scripps Research Institute, 10550 North Torrey Pines Rd., La Jolla, California 92037-1027, USA*

† Sealy Center for Molecular Science, Department of Human Biological Chemistry and Genetics, University of Texas Medical Branch, Galveston, Texas 77555-1079, USA

Non-coding apurinic/apyrimidinic (AP) sites in DNA are continually created in cells both spontaneously and by damage-specific DNA glycosylases[1]. The biologically critical human base excision repair enzyme APE1 cleaves the DNA sugar-phosphate backbone at a position 5′ of AP sites to prime DNA repair synthesis[2–4]. Here we report three co-crystal structures of human APE1 bound to abasic DNA which show that APE1 uses a rigid, pre-formed, positively charged surface to kink the DNA helix and engulf the AP-DNA strand. APE1 inserts loops into both the DNA major and minor grooves and binds a flipped-out AP site in a pocket that excludes DNA bases and racemized β-anomer AP sites. Both the APE1 active-site geometry and a complex with cleaved AP-DNA and Mn^{2+} support a testable structure-based catalytic mechanism. Alanine substitutions of the residues that penetrate the DNA helix unexpectedly show that human APE1 is

structurally optimized to retain the cleaved DNA product. These structural and mutational results show how APE1 probably displaces bound glycosylases and retains the nicked DNA product, suggesting that APE1 acts *in vivo* to coordinate the orderly transfer of unstable DNA damage intermediates between the excision and synthesis steps of DNA repair.

Despite the essential and central role of APE1 in human base excision repair (BER)[2,5], there have been no clear hypotheses concerning how APE1 coordinates human DNA repair between the first damage-specific stage carried out by distinct DNA glycosylases, and the subsequent damage-general stage of DNA repair synthesis. To resolve the molecular basis for substrate recognition, catalysis and progression of BER steps coordinated by APE1, we determined two co-crystal structures of human APE1 bound to 11-base-pair (bp) and 15-bp DNA substrates with synthetic abasic sites, at 2.95Å and 2.65Å resolution, respectively. We also determined the 3.0Å resolution ternary product complex of APE1, nicked abasic DNA and a divalent metal ion (Fig. 1a, b; see Methods).

APE1 electrostatically orients a rigid, pre-formed DNA-binding face and penetrates the DNA helix from both the major and minor grooves, stabilizing an extrahelical conformation for target abasic nucleotides and excluding normal DNA nucleotide and racemized abasic site binding (Fig. 2a, b, and see animation in Supplementary information). Although nucleotide flipping to expose damaged bases is characteristic of DNA glycosylases, it is not obviously required by APE1 and other DNA-backbone-cleaving endonucleases, and is not seen for the very short patch DNA repair endonuclease (VSR)[6]. However, abasic nucleotide flipping also occurs in a structurally unrelated AP endonuclease, endonuclease IV (ref. 7), resembling that seen in these APE1–DNA complexes, and suggesting that nucleotide flipping may be a property of AP endonucleases.

Unlike the open-to-closed conformational transition that occurs in DNA glycosylases and polymerases, the APE1 structure is pre-formed for abasic DNA recognition. The r.m.s. deviation is only ~0.7Å between Cα atoms of the free APE1 enzyme[8] and those of DNA-bound structures. The overall DNA-binding position resembles the model suggested by the unbound enzyme[8]; however, the APE1-bound DNA is severely distorted, with the DNA bent ~35° and the helical axis kinked or displaced ~5Å, where the two inserted loops cap the extrahelical AP site (Fig. 2a, b). In agreement with biochemical results[9], the enzyme–DNA interface encompasses both DNA strands but is dominated by contacts with the AP-DNA strand.

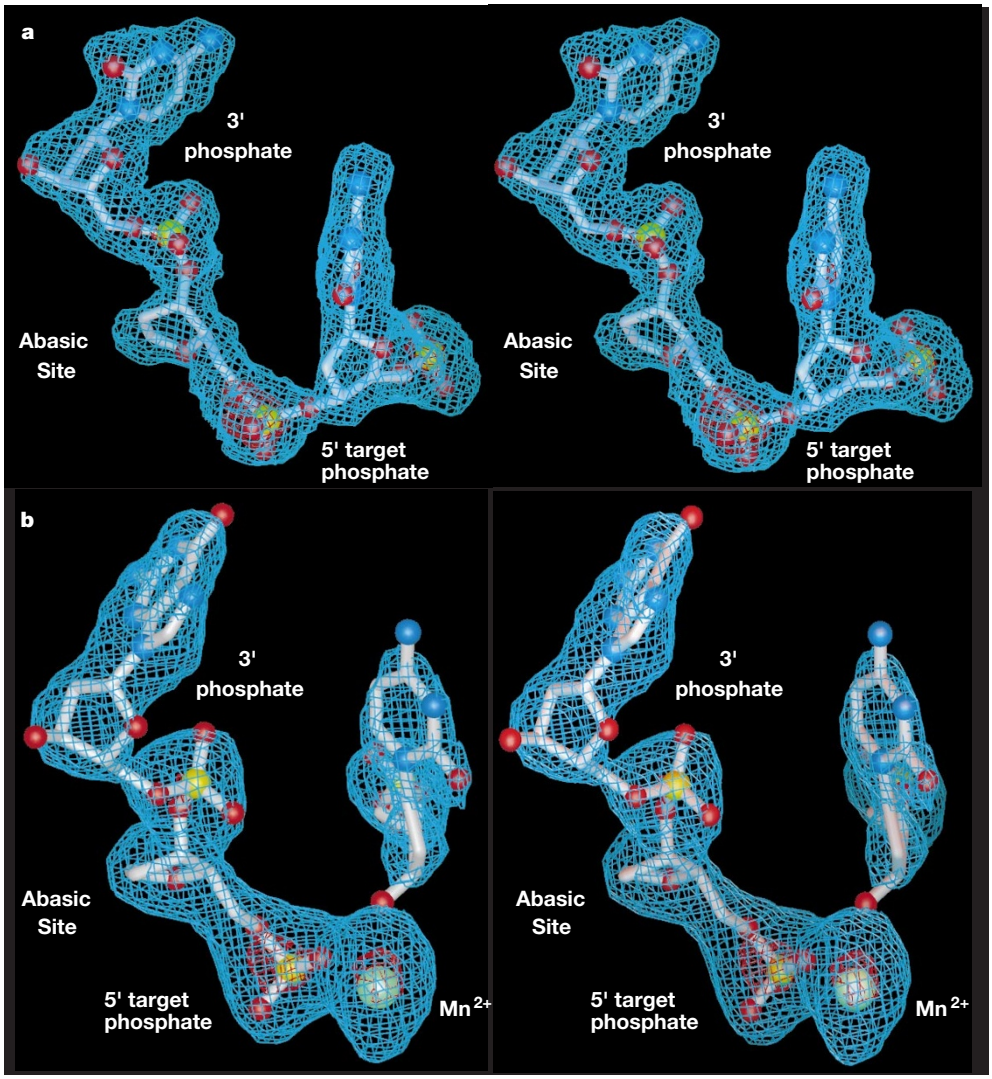

Figure 1 Stereo views of simulated-annealed omit electron density for APE1–DNA complexes showing the flipped-out abasic site and phosphodiester bond cleavage. Electron-density maps were calculated excluding the abasic and flanking nucleotides, and the divalent metal ion in **b**. APE1-bound DNA is shown with phosphorus (yellow), oxygen (red), nitrogen (blue) and carbon polytubes (white) highlighted. **a**, The APE1 active site from the 2.65Å resolution APE1–DNA complex. Omit electron-density contours are shown at 3σ (blue) and 9σ (red). **b**, The APE1 active site from the ternary complex with bound Mn^{2+} (green sphere) and nicked DNA. Omit electron-density contours are shown at 4σ (blue) and 10σ (red).

The positively charged, grooved APE1 DNA-binding face (Fig. 2b) buries nearly 1,600Å^2 of DNA solvent-accessible surface area, with ~1,000Å^2 contributed by the AP-DNA strand, and ~600Å^2 by the opposing strand. The interface is centred on the flipped-out abasic nucleotide and its flanking phosphates, which account for ~500Å^2 of the total buried DNA surface area.

APE1 stabilization of the kinked abasic DNA is mediated by residues emanating from four loops and from one α-helix (Fig. 2a). As AP sites fit within a B-DNA helix in solution[10], APE1 DNA contacts facilitate and stabilize the extrahelical AP site. 5′ of the AP site, three APE1 helical residues (Arg 73, Ala 74 and Lys 78) contact three consecutive DNA phosphates of the opposing strand, and Tyr 128 and Gly 127 span and widen the minor groove ~2Å. These 5′ contacts may anchor the DNA for the extreme kinking caused by three APE1 loops 3′ of the extrahelical abasic site (Fig. 2a). Immediately 3′ of the AP site, APE1 binds two phosphates and braces the AP-DNA backbone for double-loop insertion. Across from the flipped-out AP site, Met 270 inserts through the minor groove to pack against the orphan base partner of the abasic site and to occupy the space where it would be in unkinked B-DNA. The APE1-bound DNA is so severely distorted, however, that the orphan base still stacks with the 5′ base (Fig. 2a). Above the abasic site, Arg 177 inserts through the DNA major groove and delivers a hydrogen bond to the AP site 3′ phosphate (Fig. 3a). This DNA major groove interaction is unusual for a DNA repair enzyme, and, as the sequence and conformation of the Arg 177 loop is unique to APE1, it probably reflects specific APE1 functions. The double-loop penetration of the DNA helix, from the insertion of Arg 177 in the major groove and Met 270 in the minor groove, caps the DNA-bound enzyme active site, stabilizes the extrahelical AP site conformation and effectively locks APE1 onto the AP-DNA.

Our APE1 substrate and product complexes suggest a reaction mechanism distinct from previous predictions[8,11,12], in which the target phosphate is polarized about the scissile bond and attacked by

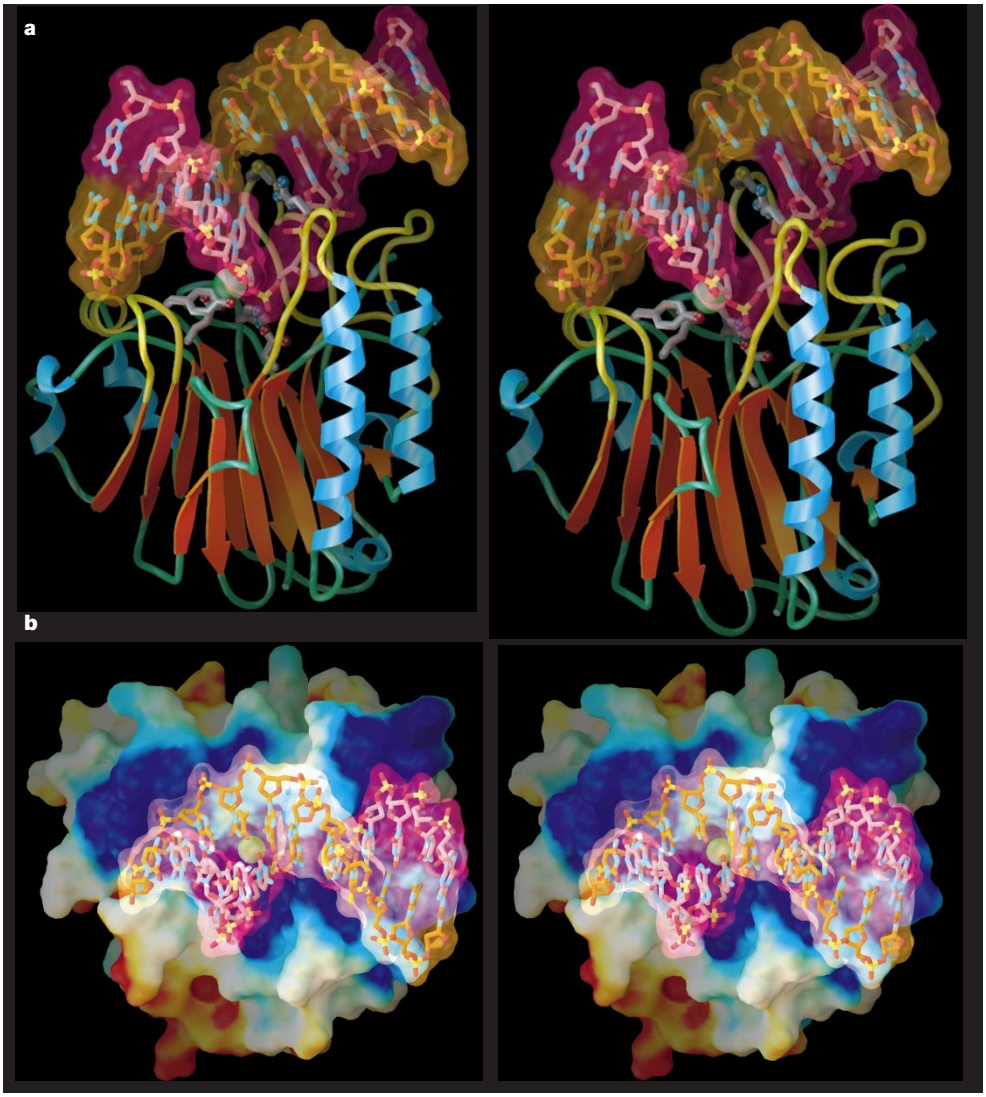

Figure 2 Stereo views of the APE1 complex with AP-DNA showing double-loop penetration of the DNA helix, plus charge and surface complementarity. **a**, DNA (top) binding to APE1 (bottom) from the co-crystal structure of APE1 bound to the 11-bp abasic DNA. The complex is viewed from the side that is roughly perpendicular to the kinked DNA helix axis oriented with the AP strand running 5′ (left) to 3′ (right), and shows the two central, six-stranded APE1 β-sheets (orange arrows), α- and 3_{10}-helices (blue coils), and coils (green tubes). Five APE1 DNA-binding regions (yellow; residues 73–79, 126–129, 174–178, 221–228 and 268–272) contact both the AP-DNA strand (pink carbon polytubes and transparent surface) and the opposing strand (orange polytubes and surface). The position of the divalent metal ion (green sphere) which is 5′ of the AP site as derived from the ternary product complex, and the side chains for key residues are shown. **b**, DNA binding to APE1 viewed looking down on the enzyme–DNA interface, rotated ~90° from the view in **a**. Arg 177 penetrates the DNA major groove (bottom) and Met 270 inserts through the DNA minor groove (top), capping the DNA-bound APE1 active site. The APE1-bound DNA, which is bent ~35° with the helix axis kinked ~5Å, complements the positively charged DNA-binding surface of APE1, coloured red (−1.5 kT e^{-1}) to blue (+1.5 kT e^{-1}) according to its electrostatic potential (calculated including the Mn^{2+} ion), which engulfs the AP-DNA strand (pink) around the flipped-out AP site.

an aspartate-activated hydroxyl nucleophile, with the metal ion stabilizing the transition state and the leaving group of the phosphodiester cleavage reaction. On the 3′ side of the APE1–DNA complex, Trp 280 contacts a DNA phosphate, and a chain of three asparagine residues, Asn 222, Asn 226 and Asn 229, lines the back of the active site and contacts two 3′ AP-DNA phosphates. On the 5′ side of the active site, the side chains of Asn 174 and Asn 212 form hydrogen bonds with O5′ and O2P of the target phosphate, and the metal ion, bound by Glu 96, ligates the O3′ atom (Fig. 3a). In our structure-based reaction scheme, which agrees with known mutagenesis results[12,13], the conserved Asp 283/His 309 pair makes a direct hydrogen bond from His 309 Nϵ2 to the remaining O1P atom (Fig. 3b, and animation in Supplementary information). These APE1–phosphate contacts orient and polarize the scissile P-O3′ bond with the Asp 210 side chain favourably aligned to activate an attacking nucleophile. Positioned by hydrogen bonds with the backbone amide of Asn 212 and Asn 68 Nδ2, the pK_a of buried Asp 210 is probably elevated, and along with the Asp 283/His 309 phosphate interaction, probably results in the observed maximal APE1 activity at pH 7.5, with a pH 6.7–9 activity profile[14]. Protonation of Asp 210 can generate the proposed attacking hydroxyl nucleophile (Fig. 3b). The divalent metal ion may facilitate the O3′ leaving group, either by direct ligation or through a water in the first hydration shell of a Mg^{2+} ion. Specific binding of APE1 to extrahelical AP sites derives from a hydrophobic pocket bordered by Phe 266, Trp 280 and Leu 282, which pack with the hydrophobic side of the abasic deoxyribose (Fig. 3a). This tight packing excludes the binding of normal DNA nucleotides, and is specific for the α-anomer of natural AP sites generated as the direct product of DNA glycosylases and not the β-anomer obtained by racemization of AP sites in solution. Thus, this APE1 specificity indicates that *in vivo* the AP site may be passed directly to APE1 from the damage-specific DNA glycosylases, thereby avoiding the release of repair intermediates more toxic than the original damage.

To investigate the role of double-loop penetration in AP-site detection and APE1 function, we determined the effects on enzyme activity of site-directed alanine substitutions for Met 270, Met 271 and Arg 177. Unexpectedly, none of these residues is required to promote abasic nucleotide flipping (Fig. 4), indicating that APE1 may be a structure-specific nuclease that detects and productively binds DNA that can adopt the kinked conformation and can present a flipped-out AP site to the APE1 recognition pocket. Both the Met270Ala and Met271Ala mutant enzymes have essentially wild-type levels of activity, whereas the specific activity is, surprisingly, increased for the Arg177Ala mutant enzyme (Fig. 4). Under the appropriate assay conditions of excess substrate, these data support the hypothesis that the Arg 177 intercalation and phosphate interaction slow APE1 dissociation from the cleaved product[15]. Moreover, kinetic analysis of the Arg177Ala mutant confirms that it is

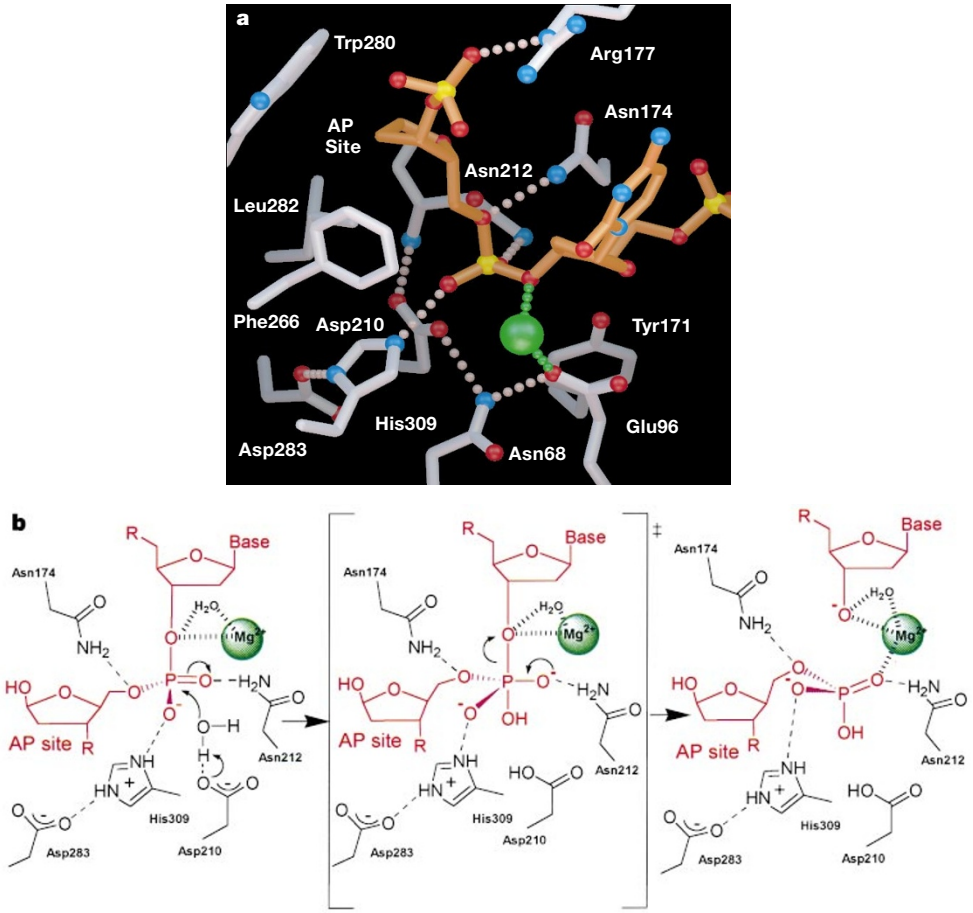

Figure 3 APE1 interactions with the flipped-out AP site provide damage specificity and suggest a specific reaction mechanism for phosphodiester bond cleavage. **a**, APE1 active-site interactions with the flipped-out abasic deoxyribose and target 5′ phosphate. The hydrophobic face of the extrahelical AP site packs within a complementary APE1 pocket formed by the side chains of Phe 166, Trp 280 and Leu 282. The Arg 177 side chain inserts through the kinked DNA major groove to form a hydrogen bond to the non-target AP site 3′ phosphate. The target AP site 5′ phosphate is oriented by hydrogen bonds with Asn 174, Asn 212 and His 309. Asp 210 is oriented by hydrogen bonds from the Asn 212 backbone amide and Asn 68 Nδ2. The metal ion position (derived from the ternary product complex) stabilizes the transition state and facilitates the O3′ leaving group, perhaps through a first hydration shell water molecule of a Mg^{2+}ion. **b**, Structure-based reaction mechanism for phosphodiester bond cleavage. Substrate AP-DNA is oriented by the bound divalent metal ion and APE1 active-site residues for attack of an hydroxyl nucleophile activated by buried Asp 210 (left panel). Collapse of the pentacovalent transition state (middle panel) leads to cleavage of the scissile P–O3′ bond, with the transition state and O3′ leaving group stabilized by the metal ion, leading to bond cleavage and inversion of the phosphate configuration (right panel).

over 25% more efficient than wild-type APE1, with an increased Michaelis constant (K_m) offset by a larger increase in the rate constant (k_{cat}) (Fig. 4). As evolutionary selection occurs at the level of biological function rather than single enzyme rates, this apparent dichotomy supports a biological role for tight product binding by APE1. Thus, the biological need to coordinate APE1 with downstream DNA BER enzymes[16,17] has evidently guided the evolution of the APE1 structure, rather than maximizing catalytic efficiency and allowing rapid release of toxic, incised DNA products.

Together these APE1–DNA complex structures and mutagenesis results suggest a coherent and testable model for the orderly transfer of DNA damage in BER. First, we propose that the extensive double-strand DNA (dsDNA)-binding interaction surface of APE1 and pronounced kinking of the DNA helix as compared with DNA glycosylases allow APE1 to stage human DNA repair by displacing bound glycosylases from AP sites, consistent with the observed APE1 enhancements of human uracil- and thymine-DNA glycosylase activities[18,19]. Second, by being optimized to retain its nicked DNA product and to also interact with the repair polymerase Polβ in the presence of DNA, the kinked APE1–DNA complex may recruit Polβ to sites of DNA damage. The high relative amounts of APE1 per cell (350,000 to 7,000,00)[20] compared with Polβ (~50,000 molecules)[21] suggest that functionally relevant interactions with Polβ may be mediated by recognition of the distorted APE1–DNA complex. Through the known interactions[22,23] of Polβ, XRCC1 and DNA ligase III, the entire DNA repair synthesis machinery may then be assembled. Most DNA BER synthesis replaces only the single damaged nucleotide, and the religated and repaired DNA should disfavour the ~90° bend for Polβ binding, and promote enzyme dissociation after repair synthesis. Thus, these structural and mutational analyses of human APE1 reveal adaptations that are counterintuitive in terms of single enzyme catalysis and provide insights into the evolutionary optimization of APE1 biological function for maintaining DNA integrity through the efficient transfer of DNA damage in BER. □

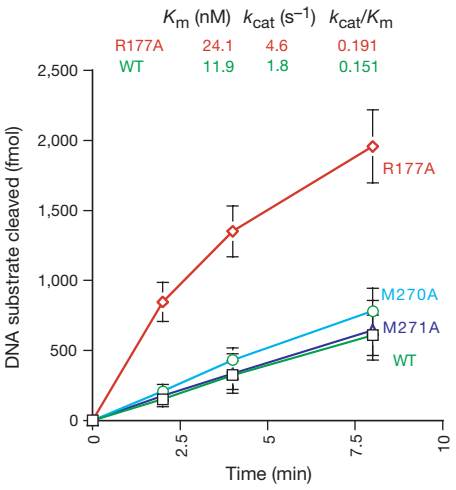

Figure 4 Activities and kinetics of wild-type and mutant human APE1 enzymes. AP endonuclease activities of wild-type APE1 (green), and Met 270 (cyan), Met 271 (blue) and Arg 177 (red) site-directed alanine mutant enzymes. Enzyme activity was assayed using a 43-bp oligonucleotide (160 nM) containing a single AP site analogue[4] with a [^{32}P]ATP 5′ end label and 100 pg of purified protein. Wild-type APE1 and mutant enzymes were incubated at 37° with ~160 nM substrate in a volume of 20 μl containing 60 mM Tris pH 8.0, 1 mM $MgCl_2$, 0.2 mM EDTA, 1 mM DTT and 100 μg ml^{-1} bovine serum albumin. The reaction was stopped at the indicated time points, separated on 20% polyacrylamide gels containing 7 M urea, and quantified by phosphorimage analyses. Results are the average of three independent experiments with ranges as indicated by the error bars. Enzyme kinetic analyses of Arg177Ala were performed as described for wild-type APE1 (ref. 13).

Methods

Biochemistry and crystallization

Expression and purification of full-length APE1 and of a fully active, N-terminal truncated enzyme containing residues 40–318 (NΔ40APE1) were done as described[13], with expression of the NΔ40 APE1 gene in *E. coli* TZ1 (xth nfo DE3) to create an N-terminal 6xHis tag. The enzyme was treated with thrombin (1:1,000 for 6 h at 37°C) to remove the 6xHis tag; and purified recombinant human NΔ40APE1 enzyme was concentrated to 20 mg ml^{-1}, and mixed with an equal volume of ~2mM 11-bp dsDNA (5′-GCTACFGATCG-3′) containing a synthetic tetrahydrofuran (F) AP site analogue (Midland Certified Reagent Co.) This AP site analogue is more stable than, and binds to and is cleaved by APE1 at rates comparable to, a deoxyribose with a C1′ hydroxyl group[3,12,24,25].

Co-crystals were grown by mixing equal volumes of enzyme–DNA solution with reservoir solutions of 20% polyethylene glycol monomethylether (MPEG) 5000 (Sigma), 50 mM 2-(*N*-morpholino)ethanesulphonic acid buffer pH 6.5 and 200 mM Li_2SO_4. Crystals grew to full size (0.5 × 0.2 × 0.06 mm) in ~2 weeks, belong to the body-centred, tetragonal space group $I4_1$, a=b= 123.4, c= 107.1Å, and contain two APE1–DNA complexes in the asymmetric unit. X-ray diffraction data were collected from single crystals flash-cooled at −170° at beamline 9-1 of the Stanford Synchrotron Radiation Laboratory (SSRL) on a 34.5 cm MARresearch imaging plate and processed with DENZO and SCALEPACK[26] to yield 42,400 observations of 15,572 unique reflections to 2.95Å resolution (93% complete) with an average $I/\sigma I$ of 12.7 and an overall R_{sym} between symmetry related reflections of 0.050 (94% complete, $I/\sigma I$= 2.2, R_{sym} =0.302; for the 2.95 to 3.06Å resolution shell). The APE1–DNA–Mn^{2+} complex crystals were obtained by co-crystallization with 25 mM $MnCl_2$ and the same 11-bp dsDNA, and belong to space group $P2_12_12_1$, a= 90.1, b=98.3, c= 101.1Å, with 41,650 observations of 16,996 unique reflections to 3.0Å (91% complete), an overall $I/\sigma I$ of 8.6, and R_{sym} =0.086 (92% complete, $I/\sigma I$= 2.3, R_{sym} =0.424; for the 3.0–3.11Å resolution shell). Crystals of APE1 bound to a 15-bp dsDNA oligonucleotide (5′-GCGTCCFCGACGACG-3′) are grown using full-length, wild-type APE1 from well solutions containing 100 mM Cacodylate (pH 6.5), 200 mM Li_2SO_4 and 27% MPEG 2000 (Sigma). These crystals also have two APE1–DNA complexes in the asymmetric unit, and belong to space group $P2_1$, a = 71.2, b = 72.4, c= 93.7Å, β = 94.3, with 61,212 observations of 24,575 unique reflections to 2.65Å (93% complete), $I/\sigma I$= 7.8, and R_{sym} = 0.087 (92% complete, $I/\sigma I$= 2.2, R_{sym} =0.367; for the 2.65–2.75Å resolution shell).

Structure solution and refinement

The 2.95Å resolution APE1–DNA structure was solved by molecular replacement with AMoRe[27] using free APE1 (ref. 8) as a search model and all data from 4.0–15.0Å resolution. The correct solutions gave a correlation coefficient of 0.36 and an R value of 0.47, with the corresponding values for the next highest peak 0.25 and 0.51. This model was refined with X-PLOR 3.8 (ref. 28) and electron density for the DNA was clearly visible in σA-weighted[29] $2F_o–F_c$, $F_o–F_c$ and simulated-annealed omit electron-density maps. Cycles of model building and stereochemically restrained simulated annealing and positional refinement allowed all the DNA to be placed, with manual inspection and rebuilding of the structure performed with XFIT[30]. The APE1–DNA complex structure contains 4,364 non-hydrogen protein atoms and 874 DNA atoms and has been refined to an R value of 0.192 (R_{free} =0.308) for 12,453 reflections from 20.0 to 2.95Å resolution with $F > 3.0\sigma F$ (74% complete), with r.m.s. deviations from ideality of 0.010Å for bond lengths, and 1.48° for bond angles. This structure was used to solve the ternary product complex with AMoRe (correlation coefficient =0.60, R=0.37), which was refined as with the APE1–DNA complex and data from 20.0 to 3.0Å resolution (13,766 reflections, 75% complete), to an R=0.181 (R_{free} =0.274), and r.m.s. deviations of 0.006Å for bond lengths and 1.48° for bond angles. The 2.65Å resolution APE1–DNA complex structure was also solved with AMoRe and refined with X-PLOR 3.8 using data from 20.0 to 2.65Å with $F>2.0\sigma F$ (25,211 reflections, 90% complete) to yield an R value of 0.203 (R_{free} =0.291) with 0.007Å deviation in bond lengths and 1.40° for bond angles. The ~40 N-terminal residues and the last base pair are disordered in this structure, and all three structures contain a *cis* peptide bond betwen Val 247 and Pro 248.

Received 23 September; accepted 15 November 1999.

1. Lindahl, T. Instability and decay of the primary structure of DNA. *Nature* **362,** 709–715 (1993).
2. Wilson, D. M. & Thompson, L. H. Life without DNA repair. *Proc. Natl Acad. Sci. USA* **94,** 12754–12757 (1997).
3. Wilson, D. M., Takeshita, M., Grollman, A. P. & Demple, B. Incision activity of human apurinic endonuclease (Ape) at abasic site analogs in DNA. *J. Biol. Chem.* **270,** 16002–16007 (1995).
4. Izumi, T. & Mitra, S. Deletion analysis of human AP-endonuclease: minimum sequence required for the endonuclease activity. *Carcinogenesis* **19,** 525–527 (1998).
5. Lindahl, T. & Wood, R. D. Quality control in DNA repair. *Science* **286,** 1897–1905 (1999).
6. Tsutakawa, S. E., Jingami, H. & Morikawa, K. Recognition of a TG mismatch: the crystal structure of very short patch repair endonuclease in complex with a DNA duplex. *Cell* **99,** 615–623 (1999).
7. Hosfield, D. J., Guan, Y., Haas, B. J., Cunningham, R. P. & Tainer, J. A. Structure of the DNA repair enzyme endonuclease IV and its DNA complex: double-nucleotide flipping at abasic sites and three-metal-ion catalysis. *Cell* **98,** 397–408 (1999).
8. Gorman, M. A. *et al.* The crystal structure of the human DNA repair endonuclease HAP1 suggests the recognition of extra-helical deoxyribose at DNA abasic sites. *EMBO J.* **16,** 6548–6558 (1997).
9. Wilson, D. M., Takeshita, M. & Demple, B. Abasic site binding by the human apurinic endonuclease, Ape, and determination of the DNA contact sites. *Nucleic Acids Res.* **25,** 933–939 (1997).
10. Withka, J. M., Wilde, J. A. & Bolton, P. H. Characterization of conformational features of DNA heteroduplexes containing aldehydic abasic sites. *Biochemistry* **30,** 9931–9940 (1991).
11. Mol, C. D., Kuo, C. –F., Thayer, M. M., Cunningham, R. P. & Tainer, J. A. Structure and function of the multifunctional DNA-repair enzyme exonuclease III. *Nature* **374,** 381–386 (1995).
12. Erzberger, J. P. & Wilson, D. M. The role of Mg^{2+} and specific amino acid residues in the catalytic

reaction of the major human abasic endonuclease: new insights from EDTA-resistant incision of acyclic abasic site analogs and site-directed mutagenesis. *J. Mol. Biol.* **290,** 447–457 (1999).
13. Izumi, T. *et al.* Intragenic suppression of an active site mutation in the human apurinic/apyrimidinic endonuclease. *J. Mol. Biol.* **287,** 47–57 (1999).
14. Kane, C. M. & Linn, S. Purification and characterization of an apurinic/apyrimidinic endonuclease from HeLa cells. *J. Biol. Chem.* **256,** 3405–3414 (1981).
15. Masuda, Y., Bennett, R. A. & Demple, B. Dynamics of the interaction of human apurinic endonuclease (Ape1) with its substrate and product. *J. Biol. Chem.* **273,** 30352–30359 (1998).
16. Bennett, R. A. O., Wilson, D. M., Wong, D. & Demple, B. Interaction of human apurinic endonuclease and DNA polymerase β in the base excision repair pathway. *Proc. Natl Acad. Sci. USA* **94,** 7166–7169 (1997).
17. Prasad, R. *et al.* Specific interaction of DNA polymerase β and DNA ligase I in a multiprotein base excision repair complex from bovine testis. *J. Biol. Chem.* **271,** 16000–16007 (1996).
18. Parikh, S. S. *et al.* Base–excision repair initiation revealed by crystal structures and DNA–binding kinetics of human uracil–DNA glycosylase bound to DNA. *EMBO J.* **17,** 5214–5226 (1998).
19. Waters, T. R., Gallinari, P., Jiricny, J. & Swann, P. F. Human thymine DNA glycosylase binds to apurinic sites in DNA but is displaced by human apurinic endonuclease 1. *J. Biol. Chem.* **274,** 67–74 (1999).
20. Chen, D. S., Herman, V. & Demple, B. Two distinct human DNA diesterases that hydrolyze 3′–blocking deoxyribose fragments from oxidized DNA. *Nucleic Acids Res.* **19,** 5907–5914 (1991).
21. Horton, J. K., Srivastava, D. K., Zmudzka, B. Z. & Wilson, S. H. Strategic down–regulation of DNA polymerase beta by antisense RNA sensitizes mammalian cells to specific DNA damaging agents. *Nucleic Acids Res.* **23,** 3810–3815 (1995).
22. Kubota, Y. *et al.* Reconstitution of DNA base excision–repair with purified human proteins: interaction between DNA polymerase β and the XRCC1 protein. *EMBO J.* **15,** 6662–6670 (1996).
23. Caldecott, K. W., McKeown, C. K., Tucker, J. D., Ljungquist, S. & Thompson, L. H. An interaction between the mammalian DNA repair protein XRCC1 and DNA ligase III. *Mol. Cell. Biol.* **14,** 68–76 (1994).
24. Takeshita, M., Chang, C. N., Johnson, F., Will, S. & Grollman, A. P. Oligodeoxynucleotides containing synthetic abasic sites. Model substrates for DNA polymerases and apurinic/apyrimidinic endonucleases. *J. Biol. Chem.* **262,** 10171–10179 (1987).
25. Erzberger, J. P., Barsky, D., Scharer, O. D., Colvin, M. E. & Wilson, D. M. Elements in abasic site recognition by the major human and *Escherichia coli* apurinic/apyrimidinic endonucleases. *Nucleic Acids Res.* **26,** 2771–2778 (1998).
26. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276,** 307–325 (1997).
27. Navaza, J. AMoRe: an automated package for molecular replacement. *Acta Crystallogr. A* **50,** 157–163 (1994).
28. Brünger, A. T., Kuriyan, J. & Karplus, M. Crystallographic R factor refinement by molecular dynamics. *Science* **235,** 458–460 (1987).
29. Read, R. J. Improved Fourier coefficients for maps using phases from partial structures with errors. *Acta Crystallogr. A* **42,** 140–149 (1986).
30. McRee, D. E. XtalView/Xfit—a versatile program for manipulating atomic coordinates and electron density. *J. Struct. Biol.* **125,** 156–165 (1999).

Supplementary information is available on *Nature*'s Wolrd-Wide Web site (http://www.nature.com).

Acknowledgements

We thank S. S. Parikh, C. D. Putnam, D. S. Daniels and D. J. Hosfield for helpful discussions, and the staff and facilities at SSRL. This work was supported by the NIH, by a Laboratory Directed Research and Development grant from the Lawrence Berkeley Laboratory, by a cancer research supplement administered through LBNL from the National Cancer Institute (to J.A.T., S.M. and P. Cooper), the Skaggs Institute for Chemical Biology and a Special Fellowship from the Leukemia Society of America (to C.D.M.).

Correspondence and requests for materials should be addressed to J.A.T. (e-mail: jat@scripps.edu). Coordinates have been deposited in the RCSB Protein Data Bank under accession codes 1DE8, 1DE9 and 1DEW.

Design of single-layer β-sheets without a hydrophobic core

Shohei Koide*, Xiaolin Huang*, Karl Link*, Akiko Koide*, Zimei Bu† & Donald M. Engelman†

** Department of Biochemistry and Biophysics, University of Rochester Medical Center, Rochester, New York 14642, USA*
† Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, Connecticut 06511, USA

The hydrophobic effect is the main thermodynamic driving force in the folding of water-soluble proteins[1,2]. Exclusion of nonpolar moieties from aqueous solvent results in the formation of a hydrophobic core in a protein, which has been generally considered essential for specifying and stabilizing the folded structures of proteins[1–6]. Outer surface protein A (OspA) from *Borrelia burgdorferi* contains a three-stranded β-sheet segment which connects two globular domains[7]. Although this single-layer β-sheet segment is exposed to solvent on both faces and thus does not contain a hydrophobic core, the segment has a high conformational stability[8]. Here we report the engineering of OspA variants that contain larger single-layer β-sheets (comprising five and seven β-strands) by duplicating a β-hairpin unit within the β-sheet. Nuclear magnetic resonance and small-angle X-ray scattering analyses reveal that these extended single-layer β-sheets are formed as designed, and amide hydrogen–deuterium exchange and chemical denaturation show that they are stable. Thus, interactions within the β-hairpin unit and those between adjacent units, which do not involve the formation of a hydrophobic core, are sufficient to specify and stabilize the single-layer β-sheet structure. Our results provide an expanded view of protein folding, misfolding and design.

The crystal structure of OspA revealed that this protein has an unusual architecture[7]: it is dumbbell-shaped and contains a three-stranded 'single-layer' β-sheet in the centre (Fig. 1). The single-layer β-sheet segment is rich in polar amino acids, and its amino-acid sequence does not follow the canonical, alternating hydrophobic/hydrophilic pattern usually found in amphipathic β-sheets (Fig. 2). Our previous studies showed that the solution structure of OspA is similar to the crystal structure[9,10], and that the protein, including the unusual single-layer β-sheet region, is highly stable[8,11].

The OspA single-layer β-sheet, although it does not contain a hydrophobic core, buries nonpolar surfaces to a degree comparable to that found in small globular proteins that do contain a hydrophobic core[8]. We attributed this effective burial of nonpolar surfaces to strategically placed amino acids with long side chains in the β-sheet. Despite the lack of a hydrophobic core in this region, the single-layer β-sheet architecture can gain sufficient stabilization from the hydrophobic effect, generally considered the dominant factor in protein stability[1,2]. However, β-strands 8 and 10 of the three-stranded single-layer β-sheet make extensive interactions with the globular domains (Fig. 1), and thus these two strands

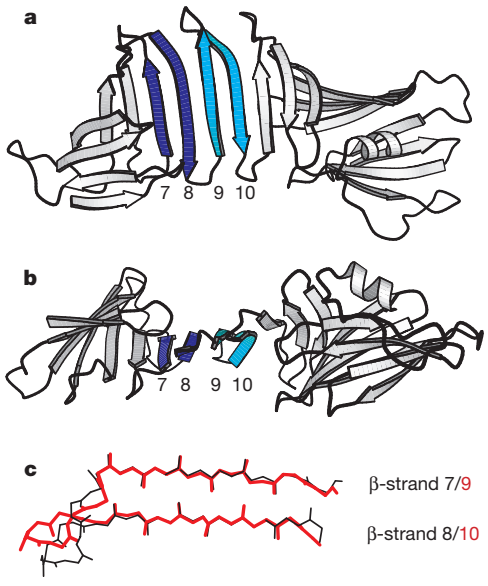

Figure 1 Structure of wild-type OspA. **a**, **b**, Drawings of the OspA structure[7]. The molecule in **b** is rotated along a horizontal axis relative to the molecule in **a**. The homologous β-hairpins (β-strands 7 and 8, and 9 and 10) in the single-layer β-sheet region are labelled and shown in blue and cyan, respectively. **c**, Superposition of the β-hairpin units. β-strands 7 and 8 (residues 95–118; black lines) and 9 and 10 (119–141; red lines) have been superimposed using C_α atoms in the β-strand regions. The root-mean-squared deviation for the backbone atoms in the β-strands is 0.51 Å.