

Structure of a repair enzyme interrogating undamaged DNA elucidates recognition of damaged DNA

Anirban Banerjee¹, Wei Yang¹, Martin Karplus^{1,3} & Gregory L. Verdine^{1,2}

¹Departments of Chemistry and Chemical Biology, and ²Molecular and Cellular Biology, Harvard University, 12 Oxford Street, Cambridge, Massachusetts 02138, USA
³Institut de Science et d'Ingénierie Supramoléculaires (ISIS), Université Louis Pasteur, 8 allée Gaspard Monge, 67000 Strasbourg, France

How DNA repair proteins distinguish between the rare sites of damage and the vast expanse of normal DNA is poorly understood. Recognizing the mutagenic lesion 8-oxoguanine (oxoG) represents an especially formidable challenge, because this oxidized nucleobase differs by only two atoms from its normal counterpart, guanine (G). Here we report the use of a covalent trapping strategy to capture a human oxoG repair protein, 8-oxoguanine DNA glycosylase I (hOGG1), in the act of interrogating normal DNA. The X-ray structure of the trapped complex features a target G nucleobase extruded from the DNA helix but denied insertion into the lesion recognition pocket of the enzyme. Free energy difference calculations show that both attractive and repulsive interactions have an important role in the preferential binding of oxoG compared with G to the active site. The structure reveals a remarkably effective gate-keeping strategy for lesion discrimination and suggests a mechanism for oxoG insertion into the hOGG1 active site.

OxoG, produced by the attack of intracellular oxidants on G residues in DNA^{1,2}, is misread as thymine (T) by the replicative DNA polymerization machinery. The resulting oxoG•A mispair, if allowed to persist, gives rise through one further round of replication to a G•C to T•A transversion mutation. It is the role of hOGG1 to intercept the oxoG•C lesions once formed and initiate correction by the base-excision DNA repair pathway (Fig. 1a) before they are encountered by the replication machinery. Only 50,000 molecules of hOGG1 protect the entire 6×10^9 base-pair nuclear genome of a diploid human cell³, hence the enzyme must have developed an efficient mechanism for distinguishing oxoG from the four nucleobases in normal DNA.

The structural basis for recognition and removal of oxoG by hOGG1 has been intensively studied^{4–7}. The structure of recognition-competent but repair-defective mutants of hOGG1 bound to oxoG-containing DNA revealed⁴ that the target oxoG nucleoside is completely extruded from the DNA duplex and inserted deeply into a lesion recognition pocket on the enzyme. Amino acid residues lining this pocket contact oxoG directly, providing a basis for specific recognition of the lesion (Fig. 1a, inset). Because hOGG1 also makes extensive contacts to the C residue left unpaired by oxoG extrusion ('estranged C'; see below), the problem of lesion discrimination by the hOGG1 catalytic machinery is reduced to understanding how the lesion recognition pocket distinguishes oxoG from G. These two nucleobases differ in chemical composition at only two positions: C8 (O versus H) and N7 (H versus lone electron pair) (Fig. 1a). The only obvious contribution to discrimination apparent from structural studies is the hydrogen bond made by the N7 H of oxoG to the main-chain carbonyl of Gly 42 (Fig. 1a), which would be lost with G. Is this one single interaction sufficient to enable discrimination of oxoG from G, despite a roughly million-fold concentration difference between the two? To address this question, we sought to compare the available structures of hOGG1–DNA complexes presenting an oxoG lesion to the active site with a structure presenting the normal base G, and to use the set of structures as a basis for free energy difference calculations to elucidate the nature of the important interactions. Obtaining the latter structure represented a formidable technical challenge, because of necessity, DNA-binding proteins generally fail to form homogeneous complexes with DNA lacking cognate structural

features such as lesions or promoter sequences. Consequently, few structures of proteins bound nonspecifically to DNA have been determined so far^{8–11}.

In previous work^{12,13}, we reported the development of a method by which to impose structural homogeneity on otherwise transient or inhomogeneous protein–DNA interaction systems. This method, disulphide trapping¹⁴, consists of implanting a disulphide crosslink into the interface between a protein and DNA, so as to restrict the ability of the two to dissociate from each other and hence form alternative complexes. Here we report the use of disulphide trapping to observe hOGG1 in the act of presenting G to its active site. Together with the accompanying computation studies, the present structures reveal that hOGG1 uses a refined mechanism for denying G access to the active site. The structure with the normal base bound to an alternative nucleobase-binding site suggests a possible late intermediate in the base-extrusion pathway.

Overall experimental strategy

The overall experimental strategy for these studies is outlined in Fig. 1b–d. Briefly, the structure of the previously reported hOGG1–oxoG DNA lesion recognition complex⁴ was inspected in order to identify prospective crosslinking sites. Of the several candidate crosslinking positions in the protein thus identified, position 149 appeared especially attractive, because it lies completely outside the enzyme active site and yet constitutes a point of intimate contact between the protein and DNA. Specifically, the side-chain carbonyl of Asn 149 hydrogen bonds to the exocyclic amine of the estranged cytosine, the base left unpaired by extrusion of oxoG from the DNA helix (Fig. 1c; see also Supplementary Fig. S1a). We reasoned that replacement of this hydrogen-bonding interaction with a covalent crosslink (Fig. 1c) might have minimal impact on the local conformation at the site of crosslinking, while ensuring extrusion of the complementary guanine from the DNA helix. However, the feasibility of crosslinking at this site in non-lesion-containing DNA was uncertain, because the crosslinking reaction would require sustained disruption of a normal C•G pair located amidst 14 neighbouring base pairs that could equally well be disrupted.

Crosslinking chemistry and structural validation

To test the present crosslinking strategy, we started with the

catalytically inactive K249Q variant of hOGG1 and introduced a further N149C mutation to give hOGG1-QC. We constructed a series of duplex oligonucleotides having the sequence shown in Fig. 1d, with a single oxoG lesion in the upper strand and thiol-bearing tether (activated as a mixed disulphide) attached to the estranged cytosine on the lower strand (as in Fig. 1c, $n = 1$). Upon incubation of these oligonucleotides with hOGG1-QC, significant amounts of a crosslinked species were formed (Supplementary Fig. S1b); crosslinking was dependent upon the N149C mutation and was lost upon incubation with 2-mercaptoethanol (data not shown).

Inspection of the structure of the crosslinked oxoG complex refined to 2.4 Å (Supplementary Table) revealed that the active site was identical within experimental error to that of the un-crosslinked oxoG recognition complex⁴, with the extra-helical oxoG nucleoside deeply inserted into the concave lesion recognition pocket (data not shown); only minor differences were observed elsewhere, and these were localized entirely to the immediate vicinity of the crosslink (Supplementary Fig. S1a, compare the top pair of structures). Crosslinking is associated with a slight retraction of the estranged cytosine from the protein surface, but the set of hydrogen-bonding contacts to Arg 154 and Arg 204 appears to be maintained nonetheless. Electron density for the thiol-bearing tether on DNA is weaker than that for most of the surrounding atoms in the structure, indicating conformational flexibility in the crosslink itself; observations of weak electron density for similar disulphide crosslinks have been made previously^{12,13,15}. These results establish the suitability of the Cys 149/estranged C crosslink for positional trapping of hOGG1–DNA complexes.

Crosslinking hOGG1 to normal DNA

The oxoG lesion in the oligonucleotide was replaced with G, and crosslinking was performed using a version of hOGG1 containing only the N149C mutation (N149C hOGG1). Although crosslinking is considerably slower with the G-containing duplex than with oxoG (Supplementary Fig. S1b), significant amounts of crosslinked product were obtained. Crosslinking was dependent upon the N149C mutation (data not shown) and tether length (Supplementary Fig. S1c), demonstrating selectivity in the crosslinking reaction.

The global structure of hOGG1 in complex with a non-lesion-containing DNA duplex (G complex) refined to 2.50 Å resolution (Supplementary Table) bears marked overall similarity to that of the oxoG complex (Fig. 2), the most distinctive feature being the presentation of an extra-helical nucleoside to the hOGG1 active site by a drastically bent duplex ($\sim 80^\circ$ compared with $\sim 70^\circ$, respectively). The local structure at the site of crosslinking is largely unaffected by changing the extra-helical oxoG to G (Supplementary Fig. S1a). Whereas the oxoG nucleobase inserts itself deeply into the lesion recognition pocket on the enzyme (Figs 2, left, and 3a), G is rejected by the lesion recognition pocket and instead lies against the protein surface at an exo-site some 5 Å outside the pocket (Figs 2, right panel, and 3b). The G base does interact with two active-site residues, Phe 319 and His 270, but the contacts are completely different from those made with oxoG. In the oxoG complex, His 270 does not contact the oxoG nucleobase, but instead hydrogen bonds to its 5' phosphate (Fig. 3a); in the G complex, this phosphate contact is disengaged, and His 270 interacts with the π -face of the G nucleobase (Fig. 3b). The high *B*-factor for the His 270 side chain in the G complex suggests that its interaction with G is weak. These

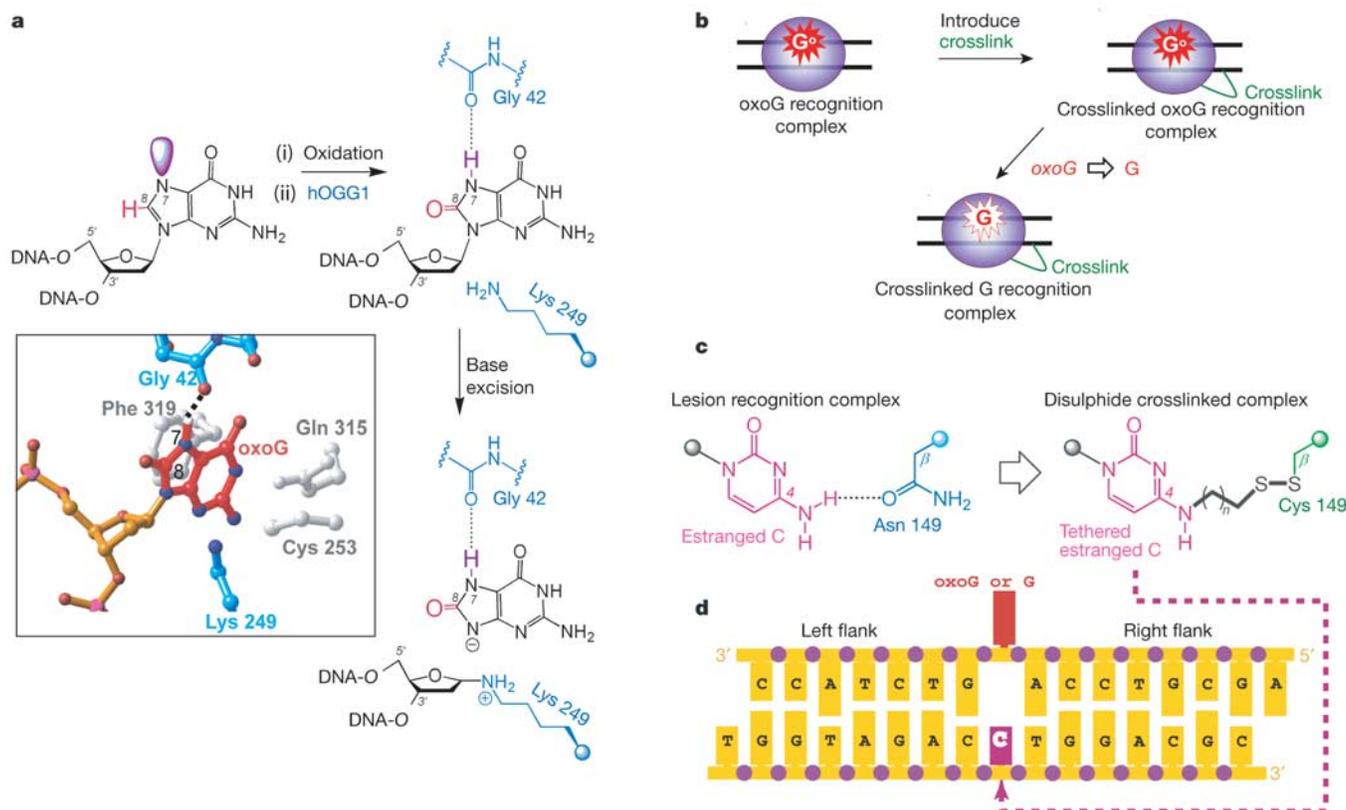


Figure 1 Generation of 8-oxoguanine (oxoG), its recognition by human 8-oxoguanine DNA glycosylase (hOGG1) and overview of the structure-based trapping strategy used here to obtain a complex of hOGG1 bound to undamaged DNA. **a**, Structural differences between G and oxoG. Inset: close-up view of the lesion recognition pocket of hOGG1 showing residues involved in recognition of oxoG and catalysis, highlighting the direct contact

between N7 H and Gly 42, and the catalytic nucleophile Lys 249. **b**, Schematic overview of the crosslinking and validation strategy. **c**, Details of the trapping chemistry. Attachment of a tether at the N4 position of cytosine is known to preserve Watson–Crick pairing in protein-unbound DNA, with the tether protruding into the major groove³¹. **d**, Sequence of the DNA duplex used in this work.

results establish that hOGG1 is able to read out the subtle structural distinctions between the oxoG lesion versus its normal counterpart G, allowing the lesion admittance to the active-site pocket while rejecting its normal counterpart.

Despite having a fully functional catalytic apparatus, the G complex undergoes no detectable base excision over the course of 10 days (data not shown). A version of the oxoG complex having Lys 249 intact (that is, constructed with N149C hOGG1) does undergo base excision, indicating that crosslinking per se does not abrogate catalysis. We thus conclude that hOGG1 is unable to cleave G from DNA at a detectable rate, even with the benefit of having G forcibly presented to the enzyme active site.

Energetic basis for discrimination of oxoG from G

The structure of the G complex presented here demonstrates the existence of an alternative nucleobase-binding site outside the lesion recognition pocket (exo-site). Because there are two sites on the enzyme that could, in principle, bind oxoG and G, the quantity of thermodynamic interest is the discrimination factor for binding oxoG versus G in the lesion recognition pocket relative to the differential binding of the two moieties in the exo-site. Free energy simulations (see Methods) were performed to determine this factor and dissect it into the important contributions. Calculations of the free energy penalties of inserting G versus oxoG in the two sites (ΔA_1 and ΔA_2 in Fig. 4a) is difficult to do directly by computational methods because of the large conformational change involved in the process. This problem was overcome by constructing a thermodynamic cycle (Fig. 4a), with which the overall discrimination free energy $\Delta\Delta A = \Delta A_1 - \Delta A_2$ can be obtained by comparing the free energy costs of bringing about an 'alchemical' transformation of oxoG to G in the active site (ΔA_3) and the exo-site (ΔA_4); such alchemical transformations can be simulated accurately by well-defined methods^{16,17}. The calculated value for $\Delta\Delta A$ was found to be $-6.8 \text{ kcal mol}^{-1}$, which corresponds to a roughly 10^5 -fold preference to insert oxoG versus G into the active site, starting with either nucleobase at the exo-site observed in the G complex. To determine whether the main contribution to the discrimination factor arises from the relative binding of oxoG versus G to the active site, the exo-site, or both, additional alchemical simulations were performed (see Supplementary Information). As

shown by a free energy simulation that uses aqueous solution as the reference, the active site favours oxoG relative to G by $8.6 \text{ kcal mol}^{-1}$. As the calculated discrimination factor is only $-6.8 \text{ kcal mol}^{-1}$, the exo-site also favours oxoG by the difference, which is equal to $1.8 \text{ kcal mol}^{-1}$.

The simulations for the relative binding of oxoG versus G to the hOGG1 active site have Lys 249 in the fully protonated state ($-\text{NH}_3^+$) and Cys 253 in the deprotonated form (S^-), with the two forming a salt bridge (Fig. 4b). Our calculations show that this configuration is more stable than the alternative neutral pair Lys 249 (NH_2)/Cys 253 (SH) or the pair having both Lys 249 and Cys 253 protonated ($-\text{NH}_3^+/\text{SH}$) (W.Y., A.B., G.L.V. and M.K., manuscript in preparation). Comparison of a quantum mechanical calculation for oxoG and G shows that the primary difference in the electrostatic potentials of the two bases is a charge inversion at positions 7 and 8 (see Fig. 4b). This creates local dipoles with opposite directions in oxoG and G. Notably, the dipole moment of the Lys 249 (NH_3^+)/Cys 253 (S^-) pair would be oriented so as to be approximately antiparallel to oxoG, but parallel to G, if G were to have a similar position as oxoG in the active site. Consequently, the Lys 249 (NH_3^+)/Cys 253 (S^-) dipole is expected to stabilize the interaction of oxoG with the active site but destabilize the interaction with G. To determine the magnitude of this effect, an alchemical simulation transforming oxoG to G was performed for the oxoG structure with the neutralized pair Lys 249 (NH_2)/Cys 253 (SH). This calculation yielded a stabilization of oxoG versus G in the oxoG active site of $5.3 \text{ kcal mol}^{-1}$, indicating that a major contribution ($3.3 \text{ kcal mol}^{-1}$) to the residue fidelity arises from the dipole-dipole interaction.

As noted above (Fig. 1a), a hydrogen bond is apparent between the carbonyl oxygen of Gly 42 and N7 H of oxoG. This attractive interaction would be replaced in the case of G by a strongly repulsive interaction with the N7 lone pair, if G and Gly 42 assumed the same positions as in the oxoG complex (Supplementary Fig. S3). A free energy perturbation analysis (see Methods) shows that the Gly 42 interaction is indeed stabilizing for oxoG versus G in the active site by about $5.3 \text{ kcal mol}^{-1}$, but this is offset in part by a destabilizing contribution of about $1.8 \text{ kcal mol}^{-1}$, due to interactions with water molecules in the active site. This result suggests that the hydrogen bond to Gly 42 makes an important contribution to stabilization of

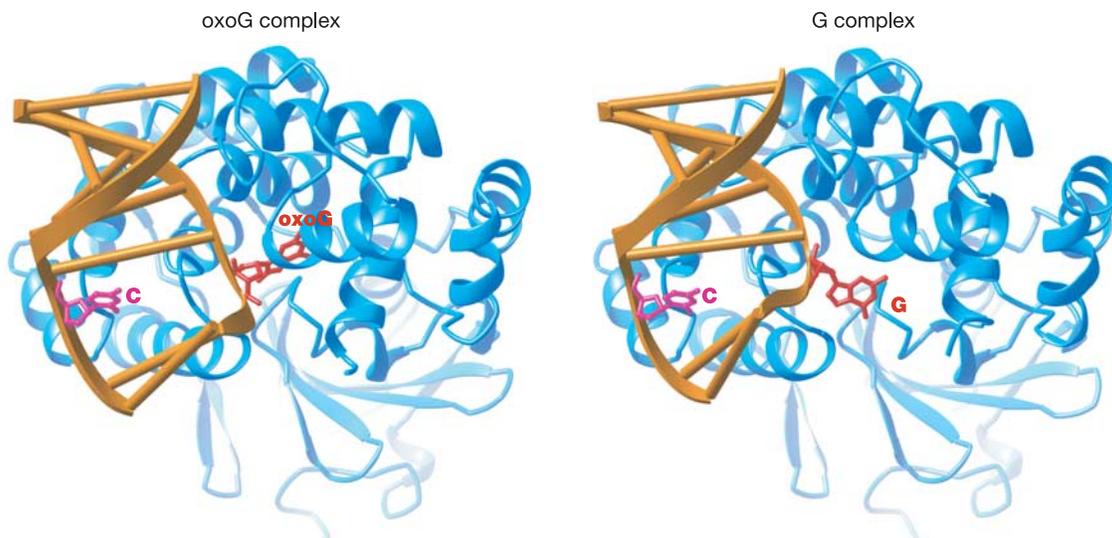


Figure 2 Comparison of the overall structures of trapped complexes obtained with oxoG-containing (left) or G-containing (right) DNA. Both protein and DNA are represented as backbone ribbon traces, with the protein in cyan and the DNA in gold. The estranged C

(magenta) and oxoG or G (red) are rendered in ball-and-stick representations. Note that oxoG is bound in the lesion recognition pocket, whereas the G is bound at the alternative extra-helical site (exo-site).

oxoG in the active site, on the order of $3.5 \text{ kcal mol}^{-1}$ (see Methods). As to the destabilizing repulsive interaction between the N7 of G and the C = O of Gly 42 (see Supplementary Fig. S3), its net contribution is less important (less than 2 kcal mol^{-1}). The reason for the small magnitude of this effect is that the hOGG1 structure is sufficiently flexible, in accord with results found more generally in simulations of proteins¹⁸, that repulsion can be relieved by a local re-orientation of Gly 42, with only a small energetic penalty (see Fig. 4c).

To explore the effect of further changes in nucleobase structure on the nature of the interactions with hOGG1, we performed alchemical simulations for two modified G analogues, 7-deazaG and 7-deaza-8-azaG (Supplementary Fig. S3), with each analogue being transmuted to G in the active and alternative sites. The results show that both 7-deazaG and 7-deaza-8-azaG, unlike G, are more stable in the active site than in the exo-site; this is in accord with the experimental structures described below. The contribution of the Lys 249 (NH_3^+)/Cys 253 (S^-) dipole is much smaller than that for oxoG, as expected; the difference in interaction energy with the ion pair and the neutral pair is less than 2 kcal. Interestingly, the major differences in $\Delta\Delta A$ for the analogues replacing oxoG arise from the poorer solvation of both analogues, relative to G, in the alternative site, where the nucleobases have significant solvent exposure.

Recognition of G analogues

To test experimentally the results of the computational studies with nucleobase analogues, we incorporated 7-deazaG into the duplex oligonucleotide, crosslinked it to hOGG1-QC, and determined the structure. As shown in Fig. 3c (see also Supplementary Fig. S4), the 7-deazaG in this structure is, like oxoG, fully inserted into the lesion recognition pocket on the enzyme. The C7 H of 7-deazaG points towards the Gly 42 carbonyl of hOGG1, and the C7 to O distance of 3.2 \AA is consistent with a weak hydrogen-bonding interaction. Similar structural findings were obtained with a second analogue, 7-deaza-8-azaG, which bears a C7 H and C8 N (see Supplementary Fig. S4). These results are consistent with the computational studies described above.

Mechanism of base extrusion

A central mystery of base-excision DNA repair concerns the mechanism by which DNA glycosylases cause the extrusion of damaged bases from the genome¹⁹. The present structure sheds light on this process. The extensive structural adjustments that take place during base extrusion cannot occur in a concerted (that is, single step) process, but rather must proceed as a pathway through a series of discrete intermediates. This being the case, then the binding mode observed in the G complex could reasonably mimic a late inter-

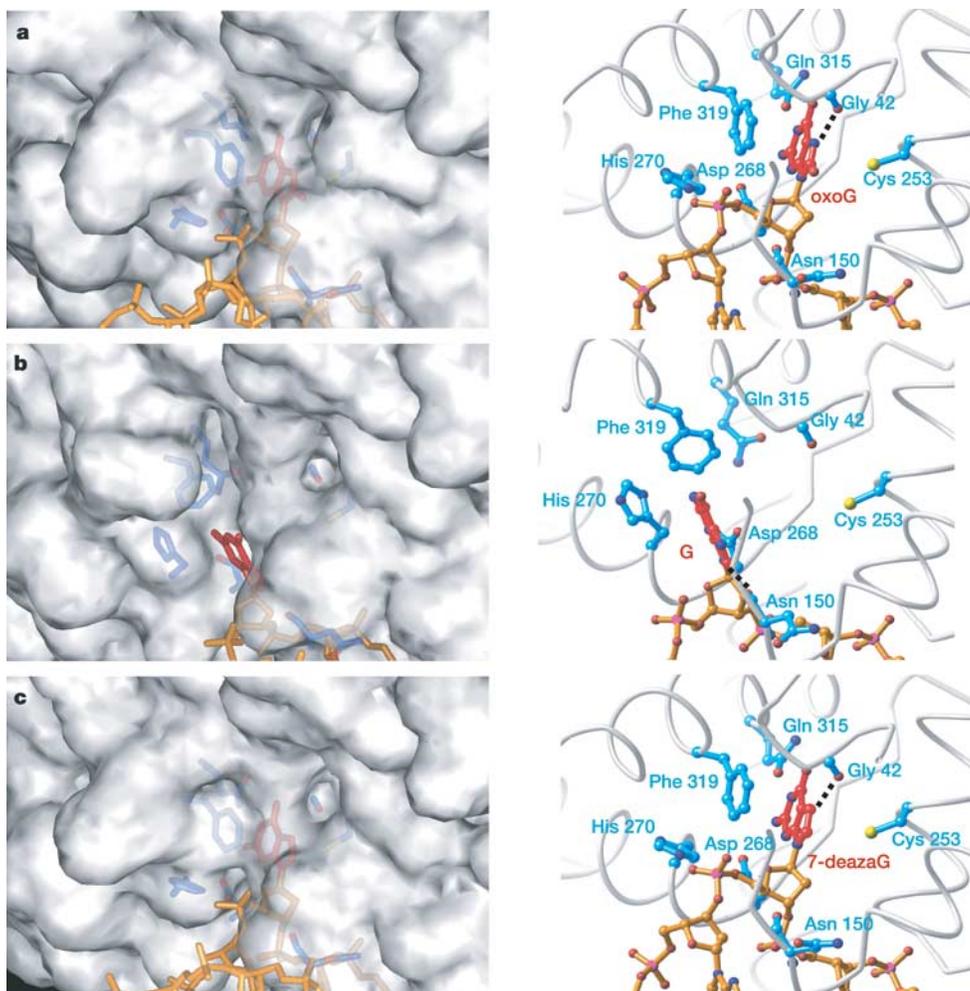


Figure 3 View of the active site region of hOGG1–DNA complexes, showing extra-helical nucleobases bound to either the lesion recognition pocket or the alternative site. **a–c**, OxoG⁺ (**a**), G (**b**) and 7-deazaG (**c**) complexes. On the left of each panel is a rendering of the solvent-accessible surface of the protein, with the DNA in framework model (DNA,

gold; oxoG/G/7-deazaG, red; key amino acid residues, blue). On the right is a detailed view of the enzyme active site, lesion recognition pocket and nearby portions of the structures (same colouring as the left panel, but with α -carbon trace of the protein backbone in grey).

mediate in the base-extrusion pathway with oxoG, just before insertion of the nucleobase into the lesion recognition pocket. Comparison of the DNA conformation in the G complex with that in the oxoG complex would thus reveal the detailed conformational changes that take place in the final stage of base extrusion, while also suggesting interactions that are established earlier in the overall extrusion pathway. Shown in Fig. 5a is a least-squares superposition of the two crosslinked complexes, using the protein component only to determine the superposition. The two cross-linked duplexes show high concordance on the 3' side of the extra-helical nucleoside (left flank), with the backbone of that strand being held in place through hydrogen-bonding contacts to the signature helix–hairpin–helix motif (Gly 245, Gln 249 and Val 250), plus a conserved electrostatic interaction with a divalent metal ion. One exception to this concordance is the 3' phosphate of the extra-helical nucleoside, which is hydrogen-bonded to Lys 249 in the G complex (predicted independently in the calculated structures); conversely, in a catalytic complex containing oxoG, Lys 249 would have to be disengaged from the 3' phosphate and swung into the active site in order to fulfil its essential role in catalysis⁶. We propose that the contacts to the left-flank—perhaps even the 3' phosphate–Lys 249 contact—are established early in the base-extrusion pathway, probably before the target base is actually extruded. On the other hand, the helix conformation at the position of the extra-helical nucleoside and on its right flank is markedly different in the two structures. Consequently, in the G complex the DNA has a more pronounced bend ($\sim 80^\circ$ compared with $\sim 70^\circ$), and the duplex on right flank is also over-rotated by $\sim 20^\circ$ about the helix axis (see Fig. 5c). Close inspection of the backbone conformation reveals that the difference results almost entirely from torsional adjustments in the extra-helical nucleoside and its 3' and 5' phosphates, and is associated with the loss of hydrogen-

bonding contacts to these phosphates by the Asn 150 backbone NH and the side-chain imidazolium NH of His 270. A divalent Ca^{2+} ion that coordinates the 3' phosphate and stabilizes the bend by inner- and outer-sphere contacts to the bases flanking the extra-helical nucleoside is also absent in the G complex (Fig. 5a). We propose that contacts to the right flank are established late in the base-extrusion pathway, only after the oxoG has achieved proper insertion into the lesion recognition pocket.

Previous comparisons of the structures of the hOGG1 lesion recognition complex⁴ and free hOGG1 (unliganded) have revealed distinct conformational states for each²⁰. Interestingly, in the G complex, the portions of hOGG1 that interact with the left flank are in the same conformation as that found in the lesion recognition complex⁴ and the oxoG complex (this work), whereas residues on the right flank and in the lesion recognition pocket, specifically His 270, Phe 319 and Gln 315, are in the same distinctive conformation as in the free enzyme. Again, this is consistent with the notion that establishment of the left-flank contacts precedes establishment of the right-flank contacts.

On the basis of the structural comparison shown in Fig. 5 and the foregoing discussion, the final stage of the base-extrusion pathway can readily be envisioned. With the left-flank interactions firmly engaged and the nucleobase positioned on the periphery of the active site, simple bond rotations about the DNA backbone at the site of the extra-helical nucleotide slot the nucleobase into the lesion recognition pocket, while relieving over-bending and over-rotation, and enabling the formation of stabilizing backbone interactions. Large rotations ($>110^\circ$) about only three bonds drive the insertion process (Fig. 5c), facilitated by smaller adjustments of flanking bonds. Only after these events have taken place, and oxoG has entered the active site, does catalysis of base excision take place. G fails to attain this final state, primarily because of the destabilizing

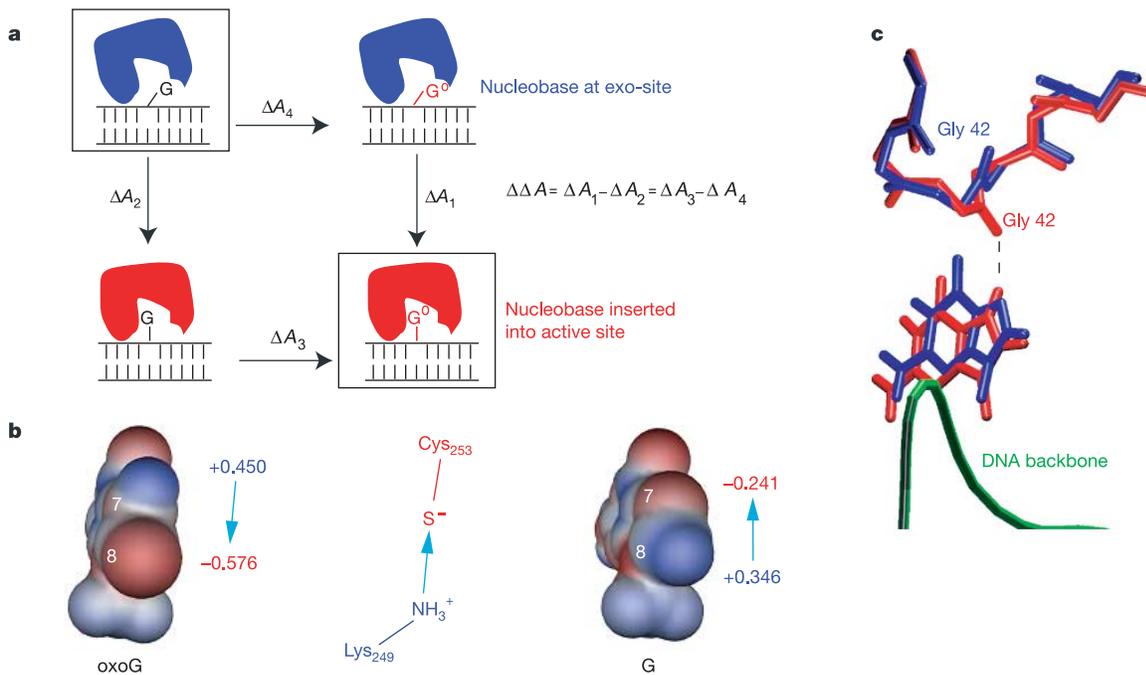


Figure 4 Computational analysis of the binding free energy difference between oxoG and G. **a**, Thermodynamic cycle for the alchemical free energy simulation of the relative binding of oxoG versus G at two binding sites. The structure containing a nucleobase at the exo-site (G versus oxoG) is schematically represented in blue, whereas that containing a nucleobase (G versus oxoG) in the active site pocket is in red. The two crystal structures used for the simulations are enclosed by a box. **b**, Electrostatic potential

difference between oxoG and G. Blue, regions of positive charge; red, regions of negative charge. Dipoles are in cyan, with Mulliken charges indicated. **c**, Orientations of the peptide group of Gly 42 in the presence of oxoG (red) versus G (blue) in the active site. The local rotations avoid a repulsive interaction between N7 of G and the carbonyl oxygen of Gly 42.

interactions it would suffer if inserted into the lesion recognition pocket; the other three bases are presumably not presented to the active site, because of the strong preference for a C at the estranged position.

The present results demonstrate that hOGG1 is capable of

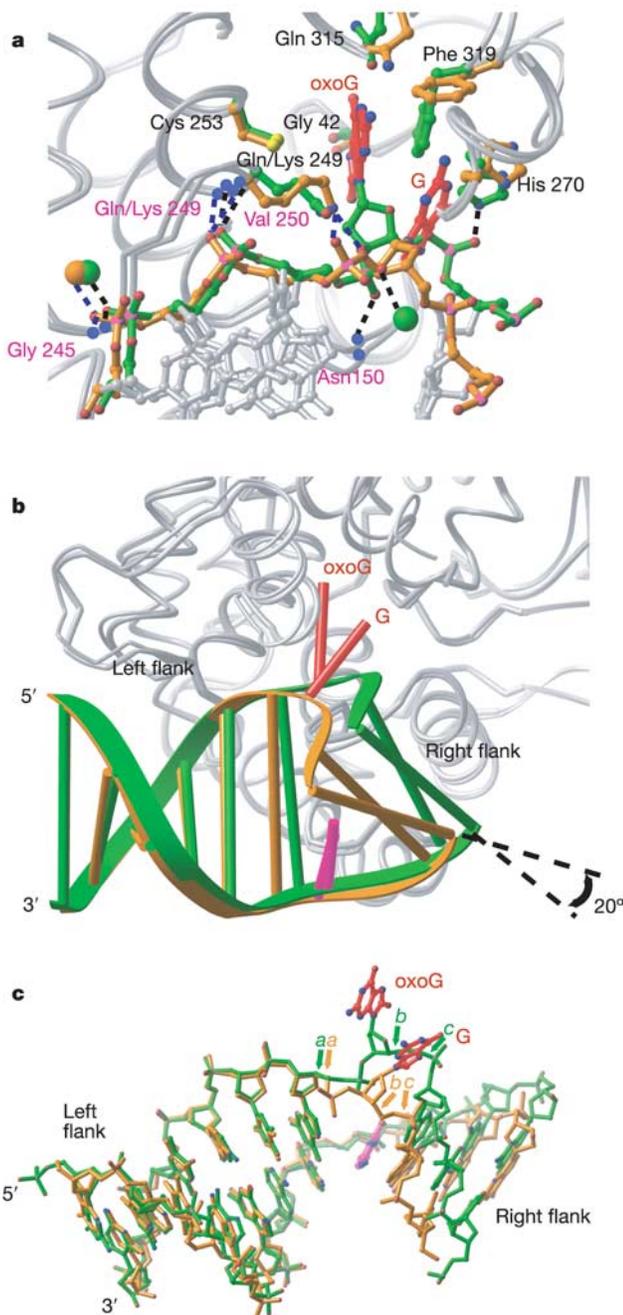


Figure 5 Superposition of the oxoG complex with the G complex in the region around the protein–DNA interface. **a**, Overlay using the protein backbone only (grey) for superposition, with the DNA backbone of the oxoG complex in green and G complex in gold. Spheres indicate Ca^{2+} ions. Residues that interact with DNA through backbone amide nitrogen atoms are denoted in magenta, whereas those that interact through side chains are in black. Dotted lines indicate hydrogen bonds. **b**, Ribbon diagram in the same orientation as **a**, but showing the whole DNA duplex. **c**, Comparison of the DNA in the two complexes, using the left flank for superposition. Arrows labelled *a*, *b* and *c* indicate bonds that have undergone significant rotations: $+110^\circ$ for *a* ($\text{C4}'\text{--}\text{C5}'$ bond of the residue 3' to oxoG/G), $+119^\circ$ for *b* ($\text{C4}'\text{--}\text{C5}'$ bond of oxoG/G) and -151° for *c* ($\text{P--O5}'$ bond of oxoG/G).

discriminating G from oxoG once the nucleobase is extra-helical (extra-helical discrimination). It remains to be determined whether hOGG1 also possesses some mechanism of distinguishing oxoG from G before extrusion from the helix (intra-helical discrimination). From a biological standpoint, the complete rejection of G from the hOGG1 active site represents a gate-keeping strategy to prevent spurious cleavage of normal bases from the genome. In this regard, hOGG1 is apparently more fastidious than its paralogue 3-methyladenine DNA glycosylase (AlkA), which does occasionally excise adenine residues from undamaged DNA^{21,22}. □

Methods

hOGG1 preparation

A polymerase chain reaction fragment of hOGG1 containing amino acids 12–327 of the human *OGG1* gene was cloned into pET30a. Megaprimer mutagenesis was performed and all new constructs were fully sequenced. Expression and purification of the wild type and N149C and N149C/K249Q hOGG1 were as described⁶.

DNA preparation, disulphide crosslinking and crystallization

Phosphoramidite derivatives of oxoG, 7-deazaG, 7-deaza-8-azaG (PPG) and O^4 -triazolyl-U were purchased from Glen Research. Oligomers 5'-AGCGTCCAXGTCTACC-3', where X denotes 8-oxoG, 7-deaza-8-azaG, 7-deazaG, or G were synthesized using standard methods. Oligomers 5'-TGGTAGACCTGGACGC-3' (where the underlined C is the thiol-tethered base; Fig. 1c) were synthesized and functionalized as described²³. The protein–DNA complex was formed by mixing the duplex DNA with hOGG1 in a twofold molar excess of the protein. The crosslinked complex was purified by MonoQ. Crystallization of the crosslinked complexes was performed using the hanging-droplet vapour diffusion method at 4 °C. The oxoG complex and the complexes containing 7-deazaG and 7-deaza-8-azaG were crystallized as described⁴. The G complex was crystallized by the hanging-droplet vapour diffusion method using a well solution containing 100 mM sodium cacodylate (pH 6.0), 150 mM CaCl_2 and 15–17% PEG 8000. Crystals were transferred to a cryoprotectant containing 100 mM sodium cacodylate (pH 6.0), 150 mM CaCl_2 , 17% PEG 8000 and 25% glycerol and then frozen in liquid nitrogen for data collection. Full details on the data collection and structure solution are contained in the Supplementary Information. Models for the oxoG complex contain 28 nucleotides and the complexes containing 7-deazaG and 7-deaza-8-azaG contain 25 nucleotides of DNA. The model for the G complex contains 20 nucleotides of DNA (density was weak for the outermost five base pairs and A overhang). The differences in electron density on the right flank for the oxoG complex versus the G complex presumably result from their differences in the helical conformation of that region, which is packed end-to-end with a neighbouring DNA duplex in the crystal. Renderings of the X-ray structures were prepared using Ribbons²⁴ and InsightII (Accelrys).

System studied by simulation

The crystal structures of hOGG1 in complex with oxoG-containing double-stranded DNA (Protein Data Bank code 1EBM)⁴ and the structure with G bound in the exo-site were used for the simulations; Lys 249 was initially built in based on the position of Gln 249 in the oxoG structure, and it moved during the simulations to interact strongly with Cys 253. The protonation states for all the residues were determined by the continuum electrostatics method²⁵; the results indicate that His 270 is doubly protonated in both structures. Lys 249 and Cys 253 were treated as Lys 249(+) / Cys 253(-) and as Lys 249/Cys 253 to study the contribution of these two residues to the active-site binding. For the alternative site binding, Lys 249(+) / Cys 253 was used.

For the simulation of the solvation free energy difference for the single bases in aqueous solution, a methyl group was added at the N9 position of oxoG, G, 7-deaza-8-azaG and 7-deazaG.

QM/MM free energy simulations

The chaperoned quantum mechanics/molecular mechanics (QM/MM) free energy simulation technique¹⁷ was used in the present study for the consistent description of oxoG, G and G analogues. The base of interest and part of the corresponding sugar ring ($\text{C1}'$, $\text{C2}'$, $\text{C4}'$, $\text{O4}'$ and hydrogen atoms attached to them) was treated quantum mechanically with the tight-binding DFT method, SCC-DFTB²⁶. The QM/MM linker scheme used in the present study is the same as that used in the calculation on UDG²⁷. The remaining system was treated classically with CHARMM27 parameters²⁸; CHARMM27 was also used for the chaperone potential (without van der Waals and Coulomb terms for the QM atoms), which serves to obtain correct geometries at the end points of the free energy simulations. Thermodynamic integration with the BLOCK module of the CHARMM program²⁹ was used to perturb only a portion of a base by sharing the common QM region in the sugar ring ($\text{C1}'$, $\text{H1}'$, $\text{C2}'$, $\text{H2}'$, $\text{H2}''$, $\text{O4}'$, $\text{C4}'$, $\text{H4}'$ and link atoms). Water was treated with the modified TIP3P water model²⁹. More details concerning the methodology, including the error estimates, are given in Supplementary Information.

Free energy perturbation analysis

To compute the contribution of each residue to the free energy difference, a free energy perturbation analysis was used. It is based on the same concept as the perturbation analysis for the energy in QM/MM studies³⁰; the contribution of each residue *j* to the overall free energy is determined by computing the change in the free energy derivative resulting from turning off all the interaction between residue *j* and other atoms.

Received 15 July 2004; accepted 14 February 2005; doi:10.1038/nature03458.

1. Michaels, M. L. & Miller, J. H. The GO system protects organisms from the mutagenic effect of the spontaneous lesion 8-hydroxyguanine (7,8-dihydro-8-oxoguanine). *J. Bacteriol.* **174**, 6321–6325 (1992).
2. Grollman, A. P. & Moriya, M. Mutagenesis by 8-oxoguanine: an enemy within. *Trends Genet.* **9**, 246–249 (1993).
3. Cappelli, E. *et al.* Rates of base excision repair are not solely dependent on levels of initiating enzymes. *Carcinogenesis* **22**, 387–393 (2001).
4. Bruner, S. D., Norman, D. P. & Verdine, G. L. Structural basis for recognition and repair of the endogenous mutagen 8-oxoguanine in DNA. *Nature* **403**, 859–866 (2000).
5. Fromme, J. C., Bruner, S. D., Yang, W., Karplus, M. & Verdine, G. L. Product-assisted catalysis in base-excision DNA repair. *Nature Struct. Biol.* **10**, 204–211 (2003).
6. Norman, D. P., Chung, S. J. & Verdine, G. L. Structural and biochemical exploration of a critical amino acid in human 8-oxoguanine glycosylase. *Biochemistry* **42**, 1564–1572 (2003).
7. Norman, D. P. G., Bruner, S. D. & Verdine, G. L. Coupling of substrate recognition and catalysis by a human base-excision DNA repair protein. *J. Am. Chem. Soc.* **123**, 359–360 (2001).
8. Luger, K., Mader, A. W., Richmond, R. K., Sargent, D. F. & Richmond, T. J. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* **389**, 251–260 (1997).
9. Viadiu, H. & Aggarwal, A. K. Structure of BamHI bound to nonspecific DNA: a model for DNA sliding. *Mol. Cell* **5**, 889–895 (2000).
10. Robinson, H. *et al.* The hyperthermophile chromosomal protein Sac7d sharply kinks DNA. *Nature* **392**, 202–205 (1998).
11. Gao, Y. G. *et al.* The crystal structure of the hyperthermophile chromosomal protein Sso7d bound to DNA. *Nature Struct. Biol.* **5**, 782–786 (1998).
12. Huang, H., Harrison, S. C. & Verdine, G. L. Trapping of a catalytic HIV reverse transcriptase•template:primer complex through a disulfide bond. *Chem. Biol.* **7**, 355–364 (2000).
13. Huang, H., Chopra, R., Verdine, G. L. & Harrison, S. C. Structure of a covalently trapped catalytic complex of HIV-1 reverse transcriptase: implications for drug resistance. *Science* **282**, 1669–1675 (1998).
14. Verdine, G. L. & Norman, D. P. Covalent trapping of protein-DNA complexes. *Annu. Rev. Biochem.* **72**, 337–366 (2003).
15. Fromme, J. C., Banerjee, A., Huang, S. J. & Verdine, G. L. Structural basis for removal of adenine mispaired with 8-oxoguanine by MutY adenine DNA glycosylase. *Nature* **427**, 652–656 (2004).
16. Simonson, T., Archontis, G. & Karplus, M. Free energy simulations come of age: protein-ligand recognition. *Acc. Chem. Res.* **35**, 430–437 (2002).
17. Yang, W., Bitetti-Putzer, R. & Karplus, M. Chaperoned alchemical free energy simulations: A general method for QM, MM, and QM/MM potentials. *J. Chem. Phys.* **120**, 9450–9453 (2004).
18. Brooks, C. L. III, Karplus, M. & Pettitt, B. M. *Advances in Chemical Physics* Vol. 71 (Wiley & Sons, New York, 1988).
19. Verdine, G. L. & Bruner, S. D. How do DNA repair proteins locate damaged bases in the genome? *Chem. Biol.* **4**, 329–334 (1997).
20. Bjoras, M., Seeberg, E., Luna, L., Pearl, L. H. & Barrett, T. E. Reciprocal “flipping” underlies substrate recognition and catalytic activation by the human 8-oxo-guanine DNA glycosylase. *J. Mol. Biol.* **317**, 171–177 (2002).
21. Berdal, K. G., Johansen, R. F. & Seeberg, E. Release of normal bases from intact DNA by a native DNA repair enzyme. *EMBO J.* **17**, 363–367 (1998).
22. O'Brien, P. J. & Ellenberger, T. The *Escherichia coli* 3-methyladenine DNA glycosylase AlkA has a remarkably versatile active site. *J. Biol. Chem.* **279**, 26876–26884 (2004).
23. MacMillan, A. M. & Verdine, G. L. Engineering tethered DNA molecules by the convertible nucleoside approach. *Tetrahedron* **47**, 2603–2616 (1991).
24. Carson, M. J. Ribbons 2.0. *J. Appl. Crystallogr.* **24**, 379–384 (1991).
25. Schaefer, M., Sommer, M. & Karplus, M. pH-dependence of protein stability: absolute electrostatic free energy differences between conformations. *J. Phys. Chem. B* **101**, 1663–1683 (1997).
26. Cui, Q., Elstner, M., Kaxiras, E., Frauenheim, T. & Karplus, M. A QM/MM implementation of the self-consistent charge density functional tight binding (SCC-DFTB) method. *J. Phys. Chem. B* **105**, 569–585 (2001).
27. Dinner, A. R., Blackburn, G. M. & Karplus, M. Uracil-DNA glycosylase acts by substrate autocatalysis. *Nature* **413**, 752–755 (2001).
28. MacKerell, A. D. Jr *et al.* All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **102**, 3586–3616 (1998).
29. Simonson, T., Archontis, G. & Karplus, M. Continuum treatment of long-range interactions in free energy calculations. Application to protein-ligand binding. *J. Phys. Chem. B* **101**, 8347–8360 (1997).
30. Bash, P. A. *et al.* Computer simulation and analysis of the reaction pathway of triosephosphate isomerase. *Biochemistry* **30**, 5826–5832 (1991).
31. Cervi, A. R., Guy, A., Leonard, G. A., Teoule, R. & Hunter, W. N. The crystal structure of N4-methylcytosine.guanosine base-pairs in the synthetic hexanucleotide d(CGCGm4CG). *Nucleic Acids Res.* **21**, 5623–5629 (1993).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We are grateful to Y. Korkhin for help in data collection and processing. We thank Enanta Pharmaceuticals for use of their X-ray instrumentation. We acknowledge the entire staff at MACCHESS, especially C. Heaton and B. Miller, and NSLS X4A for assistance in data collection and processing. We thank D. Jeruzalmi and C. Fromme for valuable discussions. This work was supported by grants from the NIH to G.L.V. and M.K.

Author contributions A.B. was responsible for performing the structural and biochemical experiments described herein, whereas W.Y. performed the computational simulations.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to G.L.V. (verdine@chemistry.harvard.edu) or M.K. (marci@tammy.harvard.edu). Atomic coordinates have been deposited in the Protein Data Bank under accession numbers 1YQK, 1YQR, 1YQL and 1YQM.