Direct observation of DNA threading in flap endonuclease complexes

Faizah A AlMalki1,6, Claudia S Flemming1, Jing Zhang2,6, Min Feng2,6, Svetlana E Sedelnikova1, Tom Ceska3, John B Rafferty1,4, Jon R Sayers2,4,5 & Peter J Artymiuk1,7

Maintenance of genome integrity requires that branched nucleic acid molecules be accurately processed to produce double-helical DNA. Flap endonucleases are essential enzymes that trim such branched molecules generated by Okazaki-fragment synthesis during replication. Here, we report crystal structures of bacteriophage T5 flap endonuclease in complexes with intact DNA substrates and products, at resolutions of 1.9–2.2 Å. They reveal single-stranded DNA threading through a hole in the enzyme, which is enclosed by an inverted V-shaped helical arch straddling the active site. Residues lining the hole induce an unusual barb-like conformation in the DNA substrate, thereby juxtaposing the scissile phosphate and essential catalytic metal ions. A series of complexes and biochemical analyses show how the substrate’s single-stranded branch approaches, threads through and finally emerges on the far side of the enzyme. Our studies suggest that substrate recognition involves an unusual ‘fly-casting, thread, bend and barb’ mechanism.

Branched nucleic acid structures arise transiently in all cells undergoing DNA replication and repair. Flaps result from extension of short RNA primers laid down during Okazaki-fragment formation on the lagging DNA strand. A 5-flap structure is created as the growing 3’ end of the lagging strand approaches the 5’ end of the RNA primer used to initiate the previous Okazaki fragment, in a process known as strand-displacement synthesis1. These flap structures must be processed efficiently to yield ligatable nicks, given that they are formed approximately every 150–1,000 nucleotides on the lagging strand2–3. Flap endonuclease (FENs) are structure-specific metalloenzymes that process flap structures4–5. FENs are required by organisms as diverse as microorganisms and mammals: FEN activity is essential for viability in bacteria6–8, and haploinsufficient (Fen1−/−) mice show rapid tumor development, whereas homozygous-knockout mice entirely lacking FEN1 (Fen1−/−) die in utero9.

FENs cleave flap substrates even when 5’-single-stranded branches of up to ~200 nt are present. Such observations have prompted suggestions that the flaps’ 5’ end threads through a hole in the doughnut-shaped FEN protein10. Subsequently, structural studies have identified a flap nuclease hole just wide enough to accommodate single-stranded DNA (ssDNA)11, thus providing some physical evidence to support the threading hypothesis. However, after studying the hydrolysis of flaps of varying lengths, Joyce and co-workers have suggested that the threading process would be energetically unfavorable12. Furthermore, heavily modified 5’ flaps, which are seemingly too large to transit such a small hole, are nonetheless cleaved by the enzyme13–15, thus casting doubt on the threading hypothesis and leading to alternative clamping or tracking models not involving a threading step12,13,16. Yet, subsequent biochemical studies have suggested that threading may occur through a hole bounded by an initially disordered region of the enzyme17. Thus, the threading model has proved controversial. Although structures of FENs with bound cleavage products have been reported18–19, to our knowledge, no structures of intact DNA substrates threading through a FEN enzyme have been published until now.

We set out to produce mutant bacteriophage T5 flap endonuclease (T5Fen) proteins that retained DNA binding but not catalytic activity, in order to obtain structural information on the enzyme–substrate complex. Here we present crystal structures and biochemical analyses that elucidate the FEN mechanism. We demonstrate that substrate-free protein exists in conformations with a small or large hole. One complex with DNA shows a single-stranded substrate poised to enter the enzyme. A second complex unequivocally shows that intact ssDNA threads through the enzyme. These structures, and an additional pseudoenzyme–product complex, provide unprecedented insight into the mechanism of these essential metalloenzymes.

RESULTS
Capturing the structure of an intact substrate–FEN complex is inherently challenging. The seven conserved acidic residues in the enzyme’s active site bind magnesium ions required for catalysis. In the absence of these cations, the negatively charged aspartate residues

1Department of Molecular Biology and Biotechnology, University of Sheffield, Sheffield, UK. 2Department of Infection, Immunity & Cardiovascular Disease, University of Sheffield, Sheffield, UK. 3UCB Pharma, Slough, UK. 4Krebs Institute, University of Sheffield, Sheffield, UK. 5Florey Institute, University of Sheffield, Sheffield, UK. 6Present addresses: Department of Biology and Biotechnology, Taif University, Taif, Saudi Arabia (F.A), MRC Weatherall Institute of Molecular Medicine, University of Oxford, Oxford, UK (J.Z.), Cancer Therapeutics and Stratified Oncology, Genome Institute of Singapore, Agency for Science, Technology, and Research (A*STAR), Biopolis, Singapore (M.F.). 7Deceased. Correspondence should be addressed to J.R.S. (j.r.sayers@shef.ac.uk) or J.B.R. (j.rafferty@shef.ac.uk).

Received 7 February; accepted 10 May; published online 6 June 2016; doi:10.1038/nsmb.3241
Figure 1 T5Fen structures and activities. (a,b) Cartoon representation of T5Fen in helical-arch (a) and looped conformations (b), with helices in magenta and β-strands in green (numbered in white and black, respectively). (c,d) Close-up views of the channel through T5Fen (orange spheres, Cat1 Mg²⁺; black spheres, Cat2 Mg²⁺). (e,f) Side chains of metal-binding residues, shown as sticks and colored according to secondary structure, for T5Fen (e) and the D153K variant (f). Electron density is contoured at 1.2σ from a final 2F – F map (gray mesh). Water molecules are shown as red spheres. (g) Cat1 sites of D153K and WT T5Fen superimposed (orange and gray backbones, respectively). The gray sphere indicates the position of magnesium M1 in e, and the magenta sphere indicates the water molecule at the M2 site in the D153K variant.

repel the DNA phosphodiester backbone, thus hindering complex formation²⁻²⁰. We reasoned that substituting an active site aspartate residue in T5Fen, such as Asp153 or Asp155, with lysine might allow its positively charged ε-amino group to mimic one of the catalytically essential metal ions. Our aim was to produce a protein that was capable of binding but not cutting DNA and thus was suitable for cocrysalization with intact substrates. We first determined crystal structures of T5Fen and its D153K variant.

Lysine’s ε-amino group mimics a metal ion in T5Fen D153K

In the absence of DNA, T5Fen and T5Fen D153K crystallized with Mg²⁺ in essentially identical forms. Each diffracted to ~1.5-Å resolution. In both cases, the crystals contained molecules in two different conformations in their asymmetric units (AUs) (Fig. 1a–d, Table 1 and Supplementary Fig. 1). All had a globular domain bearing two metal-binding regions designated Cat1 and Cat2 (Fig. 1e). In one of the equivalent conformations observed in both T5Fen and the D153K variant, residues 84–92 were present as a helix (h4, Fig. 1a,c). However, these residues adopted a looped-out conformation in the second molecule in each AU (Fig. 1b,d). Both conformations contained a hole or channel situated above the active site. This hole was more than twice as large in the looped-out conformation as in the helical-arch form (Fig. 1c,d). In wild type (WT) T5Fen, two Mg²⁺ ions (M1 and M2) separated by 3.6 Å were bound through conserved residues in the Cat1 site of the helical-arch form (Fig. 1e). A third magnesium ion (M3) was bound in Cat2. The M1 and M3 sites were occupied by Mg²⁺ ions in the molecule with the looped-out conformation (Fig. 1c,d).

The structures of the helical-arch and looped-out conformations of T5Fen D153K with Mg²⁺ were very similar to those of WT T5Fen. However, no metal ions were present in Cat1 of either conformation in the D153K variant (Fig. 1f). Instead, the ε-amino group of Lys153 occupied the analogous position to M1 in the WT (Fig. 1f,g) and interacted electrostatically with conserved aspartates in Cat1. A water molecule, rather than a cation, occupied the M2 site. The T5Fen D153K Cat2 residues bound a single Mg²⁺ ion. Otherwise, the conformations were isostructural with the WT protein. Thus, introduction of Lys153 did not perturb the overall architecture of the molecule (Supplementary Fig. 1a–c and Supplementary Video 1).

We next assessed the effects of substituting aspartate with lysine on substrate recognition and binding. WT T5Fen digested single-stranded oligonucleotide (dT₁₂), thus producing products of 3–5 nt, as expected²¹, whereas neither the T5Fen variant D153K nor a second variant, D155K, showed activity (Fig. 2a,b). By using a dual-labeled substrate, we readily detected structure-specific nuclease activity of T5Fen (Supplementary Fig. 2a) but not the variants (Fig. 2c). However, the variants bound DNA despite their negligible hydrolytic activity, and they interacted with a DNA flap more tightly than did WT protein, as indicated by their reduced dissociation constants (Supplementary Fig. 2b,c).

A prethreading T5Fen–DNA complex

We cocrysalized an 8-bp duplex DNA with four single-stranded adenosines at each 5′ end with the D153K variant and magnesium ions (Supplementary Fig. 3a). The crystals diffracted to 2.2-Å resolution (Table 1). The AU contained a protein molecule bound at each end of the DNA duplex (Supplementary Fig. 3b–d). The helical arch was intact in both molecules in the AU. The DNA did not contact the protein molecules in a symmetrical manner but revealed two distinct types of interaction (complexes C1 and C2). In complex C1, the 5′ single-stranded end was positioned just below the arch and was poised to enter it in a prethreading binding mode (Fig. 3a,b and Supplementary Fig. 3b–d). The 5′ nucleotide of one strand engaged with residues in helices 1 and 4 and was placed just in front of the hole beneath helices 4 and 5. Phe32 and His36 in helix 1 made Van der Waals (VdW) contact with the sugar and base of dA1, respectively (Fig. 3b), while its 3′ phosphate formed a salt bridge to Arg86 in helix 4. Water-mediated and direct hydrogen bonds between the dA3 phosphate group and the backbone amides of Asp155 and Gly154 helped to anchor the ssDNA in place. The ε-amino group of Lys153, introduced to mimic M1 in the Cat1 site, formed a salt bridge with the dA2 phosphate. The double-stranded DNA (dsDNA) region contacted the helix–three turn–helix motif (H3TH, Asp188–Gly225) (Fig. 3a,c). A potassium ion was bound via backbone carbonyl oxygens within this motif. It directly coordinated a phosphate group on the DNA duplex while an adjacent phosphodiester formed hydrogen bonds with backbone amides of conserved residues Lys215 and Arg216 (Fig. 3c). Further details are given in Supplementary Table 1.

A fully threaded DNA–T5Fen complex

In the second complex (C2), all four single-stranded nucleotides of the substrate threaded through the helical arch of T5Fen D153K, bringing the phosphodiester backbone into intimate contact with the active site (Fig. 4a). The last base pair of the duplex (dG5–dC12) interacted with an intact helical arch via residues Arg86 and Tyr90, and also with residues in h1, where dG5 made contact with Phe32 and
was also hydrogen-bonded to Lys35 (Fig. 4b). An unexpected barblike structure was adopted by the translocated ssDNA (Fig. 4c–e). The purine rings of nucleotides 1, 3 and 4 formed a spiral stack, but nucleotide 2 was flipped out by approximately 180°, so that its base contacted Gly70, Lys71, Leu76 and Pro80 on the distal side of the hole (Fig. 4c). The DNA backbone traced out a path resembling the letter L as it passed from the H3TH motif over the Cat1 site and through the archway (Fig. 4d, e). The distortion of the backbone was partly due to strong electrostatic interactions between the phosphodiesters of nucleotides 3–5 as they wrapped around the positively charged guanidinium group of Arg86 in helix 4. Several other residues lining the helical arch contributed to distorting the DNA backbone, including Asp87, Tyr90, Arg93 and Phe105.

A single Mg2+ ion occupied the Cat1 M1 site while the metal-mimicking ε-amino group of Lys153 was located 4 Å away at the M2 site. The scissile phosphodiester on the 3′ side of dG5 bound the Cat1 M1 Mg2+ ion (Fig. 4b, e). A Mg2+ ion bound via conserved aspartate residues that compose the Cat2 metal-binding site provided further electrostatic interactions with the phosphate of dG7 (Fig. 4e). The duplex region of the substrate made additional interactions (Supplementary Table 2) with the protein that were similar to the interactions in the C1 complex.

Structure of a pseudocleavage product–T5Fen complex

We obtained a complex (C3) with a second variant T5Fen (D155K) and a 17-residue oligonucleotide. Each protein molecule contacted different parts of three oligonucleotides in the crystal array (chains X, Y and Z; Fig. 5a and Supplementary Fig. 4a–c). The 5′ end of strand X followed the same path as the intact threaded strand in complex C2, passing over the H3TH motif and Cat1 site. The 3′ end of strand Z was stacked onto the 5′ end of strand X and continued under the helical arch. Hence, the resulting structure resembled a cleaved branched substrate lacking only the scissile phosphate (Supplementary Fig. 4b). It crystallized in the presence of calcium ions and diffracted to ~1.9-Å resolution (Table 1). Calcium ions promote binding but not catalysis in T5Fen20. One strand (X) was straddled by the helical arch, and

| Table 1 Data collection and refinement statistics |
|-----------------------------------------------|
| T5Fen (PDB 5HMM) | D153K (PDB 5HML) | D153K–5ov4 (PDB 5HNK) | D155K–3ov6 (PDB 5HP4) |
| **Data collection** | | | |
| Space group | P1 | P1 | P212121 | P41212 |
| Cell dimensions | | | | |
| a, b, c (Å) | 46.8, 58.6, 59.7 | 47.4, 58.4, 59.7 | 44.7, 109.9, 127.3 | 67.5, 67.5, 187.8 |
| α, β, γ (°) | 66.9, 79.4, 73.8 | 66.5, 79.5, 73.7 | 90, 90, 90 | 90, 90, 90 |
| Resolution (Å)* | 38.74–1.50 (1.58–1.50) | 54.60–1.48 (1.56–1.48) | 44.72–2.22 (2.28–2.22) | 67.51–1.86 (1.91–1.86) |
| Rmerge | 0.083 (0.319) | 0.057 (0.491) | 0.079 (0.654) | 0.080 (1.043) |
| Rflags | 0.093 (0.438) | 0.069 (0.600) | 0.094 (0.842) | 0.083 (1.079) |
| Rfree | 0.042 (0.296) | 0.039 (0.341) | 0.051 (0.521) | 0.022 (0.276) |
| I/σ(I) | 10.6 (2.5) | 11.4 (2.2) | 13.4 (2.2) | 30.8 (6.4) |
| Completeness (%) | 93.1 (90.4) | 93.7 (92.6) | 99.4 (94.8) | 99.8 (99.7) |
| Redundancy | 3.8 (2.2) | 3.0 (3.0) | 5.7 (3.1) | 27.9 (28.2) |
| **Refinement** | | | | |
| Resolution (Å) | 38.35–1.50 | 54.60–1.48 | 42.19–2.22 | 54.82–1.86 |
| No. reflections | 79,049 | 83,074 | 30,073 | 35,532 |
| Rwork / Rfree | 0.147 / 0.196 | 0.141 / 0.196 | 0.187 / 0.234 | 0.169 / 0.196 |
| No. atoms | 4,433 | 4,432 | 4,407 | 2,200 |
| Protein | 4,433 | 4,432 | 4,407 | 2,200 |
| DNA | 0 | 0 | 490 | 345 |
| Ligand/ion | 4 | 8 | – | – |
| Ethylene glycol | 6 | 3 | 2 | – |
| Mg2+ | 6 | 3 | 2 | – |
| Cl− | 6 | 3 | – | – |
| I− | – | 3 | – | – |
| Bis-Tris propane | – | 19 | – | – |
| Glycerol | – | – | 6 | 6 |
| K+ | – | – | 1 | 1 |
| Ca2+ | – | – | 2 | – |
| Na+ | – | – | 1 | – |
| Water | 598 | 559 | 224 | 218 |
| **B factors** | | | | |
| Protein | 16.8 | 19.0 | 34.5 | 37.5 |
| DNA | – | – | 30.8 | 48.8 |
| Ligand/ion | 26.1 | 26.5 | 65.8 | 54.4 |
| Water | 29.5 | 31.3 | 46.7 | 40.0 |
| R.m.s. deviations | | | | |
| Bond lengths (Å) | 0.0105 | 0.0102 | 0.0067 | 0.0102 |
| Bond angles (°) | 1.3289 | 1.325 | 1.060 | 1.3714 |

Data from a single crystal were used for each structure determination. *Values in parentheses are for highest-resolution shell.
its 3’-hydroxyl group (dG17) made polar contacts via Asn29 and an inner-sphere water molecule bound to Ca\(^{2+}\) in the M1 site of Cat1 (Fig. 5a–d). The M2 site was occupied by the ε-amino group of Lys155 (Supplementary Fig. 4d). It made electrostatic interactions with conserved residues in Cat1 and Cat2. It also coordinated a water molecule that was hydrogen-bonded to Ca\(^{2+}\), just as a water molecule bridged the magnesium ions in the DNA-free T5Fen structure (Fig. 5d and Supplementary Fig. 4d). The base of dG17 of chain X stacked with the recessed 5’ end of chain Y and made a base pair with a nucleotide on chain Z (Fig. 5). The threaded nucleotides made further interactions with the protein (Fig. 5b–d). These included flipping of a nucleobase such that, rather than being stacked between adjacent bases, it was rotated ~180° and was thus located adjacent to helix 3 on the distal side of the helical arch (Fig. 5b,c), similarly to the barbed conformation observed in the C2 substrate complex. Residues within the H3TH domain hosted a potassium ion that formed polar contacts with the phosphodiester backbone (Supplementary Fig. 4e).

The C3 structure also provided additional insight into interactions of the substrate’s 3’ single-stranded region with the protein (strand Y; Fig. 5). The conserved residue Arg33 in h1 contacted this region via a salt bridge (Supplementary Fig. 4f), thus explaining the results of previous mutagenesis studies in which an R33A variant has been found to bind DNA with lower affinity than that of WT\(^{22}\). Full details of the C3 complex’s molecular interactions are given in Supplementary Table 3.

**Structural changes upon substrate binding**

Comparison of the DNA-free, looped-out structure with the threaded substrate complex revealed changes upon substrate binding (Supplementary Fig. 5a). Perhaps the most striking change involved conserved amino acid Arg86. In the prethreading complex, Arg86 was located in a position very similar to that seen in the DNA-free helical-arch form (Supplementary Fig. 5b,c). Its side chain formed an angle of approximately 45° with the plane of helices 4 and 5, thereby juxtaposing the charged guanidinium group with the H3TH domain in the vicinity of Cat2. Compared with the DNA-free structure, in which residues 84–92 were looped out, the Arg86 guanidinium group was displaced by ~13 Å, and its Ca atom was displaced by ~6 Å (Supplementary Fig. 5c). Helix 1 residues involved in substrate binding underwent movements of up to 7 Å to engage with DNA (Supplementary Fig. 5d). Lys215 and Arg216 in the H3TH motif were displaced by ~3 and 6 Å, respectively, upon DNA binding (Supplementary Fig. 5e), thus explaining previous results of mutagenesis studies on these conserved residues\(^{22}\).

**Metal-ion binding in T5Fen**

A two-metal-ion mechanism for FEN catalysis has been suggested\(^{23}\), which is similar to that proposed for the Klenow proofreading 3’–5’ exonuclease\(^{24}\). However, the metal-ion separation observed in DNA-free FEN structures (5–8 Å) is generally too large for this mechanism, which requires spacings of 3–4 Å, thus casting doubt on its validity\(^{11,23,25}\). Our WT T5Fen structure revealed a pair of closely spaced metals bound in Cat1; this result was consistent with a two-metal-ion mechanism, as was the placement of these ions relative to the scissile phosphodiester (Fig. 1c and Supplementary Fig. 4d). The Cat2 metal-binding site is completely conserved in FEN domains of prokaryotic DNA polymerase I but is absent in FENs from higher organisms and some bacterial homologs\(^{6,26}\). Cat2 is not essential for flap cleavage but has been suggested to facilitate substrate binding\(^{6,20,26,27}\), as has

---

**Figure 2** T5Fen variants D153K and D155K lack enzymatic activity. (a) Reactions of radiolabeled dT12 (lane 0) with wild-type (WT) T5Fen protein or 153K or 155K variants, as indicated. M, mononucleotide ladder. The original gel image can be found in Supplementary Data Set 1. (b) Kinetics of WT T5Fen-catalyzed oligo(dT) hydrolysis (error bars, s.e.m. of 3 technical replicates). (c) Structure-specific endonuclease activity progress curves for WT (+ve), negative control (−ve) and variants, as indicated. Means (solid lines) and ranges (dashes) of 3 technical replicates are shown (a.u., arbitrary units).

---

**Figure 3** Protein-DNA interactions in prethreading T5Fen complex C1. (a) Residues within 4 Å of strand X (orange cartoon) and strand Y (green cartoon), colored cyan and magenta, respectively. Helices h4 and h5 are indicated. (b) Hydrogen bonds and salt bridges between protein and DNA (indicated by orange dashes). Residues within 4 Å are shown as sticks. (c) Strand-Y nucleotides 8–11 in the duplex region (wireframe) contacting residues (magenta sticks) in the H3TH motif via either VdW interactions or hydrogen bonds (indicated by orange dashes).
Figure 4 Structure of a fully threaded DNA complex C2. (a) Three views of T5Fen D153K in complex with Mg\(^{2+}\) (blue spheres) and DNA. (b) Contacts to the last base pair of duplex (5′-dG5–3′-dC1 in magenta and yellow carbon sticks), showing amino acids within 4 Å (gray sticks). (c) Details of protein interactions with nucleotides dA1–4. (d) Electron density for DNA (magenta mesh, contoured at 1σ). Residues within 4 Å on the helical arch are shown in cyan sticks. (e) Stereo diagram showing the threaded DNA axis distorted through ~90°. Mg\(^{2+}\) ions at M1 and M3 are shown as green spheres. Key residues involved in binding shown as sticks. Waters are shown as red spheres, and hydrogen bonds and polar contacts are shown as orange dashes.

the potassium ion bound to the H3TH motif. Our structures provide molecular details of these metal-substrate interactions (Figs. 3c,e, 4b,c and 5d and Supplementary Fig. 4d,e).

Structural conservation throughout the FEN family
T5Fen shares low primary-sequence identity with characterized homologs from bacteria and eukaryotes. Nonetheless, the structural similarities are striking and suggest common biochemistry (Fig. 6). The *Escherichia coli* Exo IX FEN homolog, crystallized as a product complex with DNA\(^{18}\), shares ~25% amino acid identity with T5Fen, yet they have 237 equivalent residues and an r.m.s. deviation of aligned Ca atoms of 2.4 Å (Fig. 6a). The active sites of these enzymes were similarly arranged with two closely spaced divalent cations (Fig. 6b,c). Perhaps more surprising was the structural similarity that T5Fen shares with human FEN1 (r.m.s. deviation of ~4.7 Å over 200 Ca atoms Fig. 6a–c). This similarity was not limited to the overall topology but also extended to the active site and H3TH motif present in T5, *E. coli* and human FEN homologs (Supplementary Fig. 4e). Thus, FENs are structurally conserved in diverse organisms. The H3TH motif hosts a potassium ion that mediates electrostatic interactions with DNA in all three enzymes\(^{18,19}\).

**DISCUSSION**

The FEN family includes structure-specific nucleases capable of cleaving flaps, nicks and gapped structures. How FEN enzymes cleave bifurcated DNA and generate sealable nicks has been the subject of debate since Lyamichev and co-workers proposed that the single-stranded 5′-flap branch might thread through a FEN protein\(^{10}\). Hydrolysis would occur once the enzyme reached the junction with dsDNA, where it would halt because of the size restriction imposed by its hole. However, it has been suggested that such a threading mechanism would be thermodynamically unfavorable\(^{12}\). Recent structural studies have added to the controversy. A study involving the FEN homolog Exo I has led to the proposal of a mechanism lacking a threading step\(^{28}\), whereas structural studies on human FEN1 in complex with cleaved products (not intact substrates)
and heavy-metal ions, have lent support to a threading hypothesis but have not provided structures showing threading\(^{12,28}\). As facilitated by our metal-mimic mutagenesis approach, a unique series of molecular snapshots characterizing the entire FEN reaction pathway has emerged.

**A fly-casting, thread, bend and barb mechanism**

The seemingly unlikely event of threading ssDNA through a small hole in a protein may be facilitated by initial binding of the flap’s dsDNA region to the FEN in a bind-then-thread mechanism\(^{12}\). After this binding, the increased local concentration of the 5’ end of the substrate in the vicinity of any hole in the enzyme would increase the likelihood of interaction. Further, it has previously been proposed that a disordered ‘arch’, with a hole larger than that previously characterized by X-ray crystallography, may facilitate passage of ssDNA and further reduce theoretical objections on the basis of thermodynamic grounds\(^{17}\). A disordered-to-ordered transition follows the threading step. The alternative conformations of residues 84–92 in the looped-out and helical-arch forms in our T5Fen provided structural evidence for this mechanism (Fig. 1b,d). The approximate dimensions of the ovoid channel through the looped-out form were 12 × 17 Å compared with 8 × 11 Å in the helical-arch conformation. Furthermore, the helical arch is entirely disordered in the crystal structure of T5Fen variant K83A, thus demonstrating flexibility of this region\(^{29}\). Similarly, the equivalent region in other published FEN structures is disordered\(^{23,25}\).

We speculate that the enzyme uses a fly-casting mechanism of molecular recognition, as proposed for transcription factor–binding site interactions in which the disordered conformation of a receptor captures ligand more efficiently than does the ordered conformation thus resulting in enhanced binding kinetics\(^{30}\).

In our threaded substrate complex, deoxyadenosines 1, 3 and 4 formed a spiral base-stack after exiting the helical arch. In contrast, dA2 was flipped 180°, so that the base contacting helix 3 was four nucleotides from the scissile bond and thus formed a structure resembling a barb. Similarly, the barb-like conformation seen in the pseudoproduct complex was located five nucleotides from the cleavage site. These findings provided a physical rationale for the observation that exonucleolytic FEN products range from 3 to 5 nt in length\(^{20,21}\). The nucleotide barb, once engaged with the hydrophobic patch, may lock down the substrate in optimal contact with the active site, thus promoting catalysis (Supplementary Video 2). DNA binding was facilitated by Arg86 in the intact substrate complex (C2). Compared with its position in free protein, Arg86 pivoted toward helix 3 and away from the Cat1 and Cat2 sites, thereby bending the DNA into an L shape through contacts with three adjacent phosphodiester groups. The importance of the equivalent arginine residue in a range of FENs in diverse organisms including phages and humans has been noted\(^{31,32}\). In the pseudoenzyme–product complex, Arg86 had swung back toward Cat1, contacting the newly revealed 5’ terminus. In a real catalytic event, this terminus would have a phosphate group carrying two negative charges. The threaded product would dissociate from the enzyme, thus allowing the new 5’ end to enter the arch for the next catalytic cycle, fueled by energy from nucleotide hydrolysis and Brownian motion\(^{33}\).

We generated a model of the molecular transitions that may occur during catalysis, based on superpositions of the structures of our DNA-free, fully threaded and pseudoproduct T5Fen complexes (Fig. 7a). Once threaded and ordered, the substrate is bound such that one of the scissile phosphates acts as an inner-sphere ligand for Mg\(^{2+}\) ions in the bound enzyme (C2). A water molecule that bridges the two Cat1 Mg\(^{2+}\) ions is poised to attack the DNA backbone at a distance of ~2.6 Å from the scissile phosphate. This scenario is similar to the two-metal-ion-mechanism proposed for Klenow proofreading exonuclease\(^{24}\). A ~1.3-Å movement would place this potential nucleophile appropriately to initiate formation of the transition state (TS). Hence, the...
nucleophile and 3′-hydroxyl leaving group would form the apical ligands of the trigonal-bipyramidal TS intermediate (Fig. 7c). As the new O-P bond forms, the 3′-O bond would concomitantly lengthen, thus leading to collapse of the TS. The attacking oxygen then would bind both M1 and M2 as part of the 5′-phosphate product (Fig. 7d).

Apart from their crucial roles in DNA replication and repair, FEN homologs such as Taq polymerase are also central to real-time quantitative PCR and clinical diagnostics and have also been proposed as potential therapeutic targets. Better understanding of FEN biology may affect these applications. We have shown a series of snapshots in the FEN-catalyzed hydrolysis reaction from free enzyme to pseudo-product complex via prethreading and fully threaded intermediates. The structures revealed that DNA can indeed pass through the channel present in a FEN family member. Our results represent a distinct improvement in understanding of the processing of branched DNA substrates, a process common to all forms of life, and highlight the likely conservation of this mechanism across phyla.

**METHODS**

Methods and any associated references are available in the online version of the paper.

**Acce ssion codes.** Coordinates and structure factors have been deposited in the Protein Data Bank under accession codes PDB 5HMM (WT T5Fen), PDB 5HML (T5Fen D153K), PDB 5HNK (T5Fen D153K–5ov4 oligonucleotide, complexes C1 and C2) and PDB 5HP4 (T5Fen D153K–3ov6 oligonucleotide, complex C3).

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

**ACKNOWLEDGMENTS**

F.A.A. was supported by a scholarship from Taif University. M.F. and C.S.F. were supported by scholarships from the University of Sheffield. Infrastructure was supported through Biotechnology and Biological Research Council (UK) grant awards 50/B19466 (J.R.S.) and RE118458 (J.R.S.).

**AUTHOR CONTRIBUTIONS**

J.R.S., T.C. and P.J.A. conceived the project. J.R.S. designed mutants and enzyme assays. M.F. and J.Z. carried out mutagenesis, biochemical experiments and kinetics analyses (supervised by J.R.S.). F.A.A., C.S.F., M.F. and J.Z. expressed proteins and, together with S.E.S., purified them. F.A.A. carried out crystallization and data collection on complexes (C1–C3). C.S.F. carried out crystallization and data collection on WT T5Fen and the D153K variant. Structure refinement was initially carried out by F.A.A. and C.S.F., under supervision of J.B.R. and P.J.A., and final refinement was performed by J.R.S. with input from J.B.R. and T.C. All authors discussed the results and commented on the manuscript. P.J.A. wrote an initial draft of the manuscript. J.R.S. wrote the manuscript with input from all authors.

**COMPETING FINANCIAL INTERESTS**

The authors declare competing financial interests: details are available in the online version of the paper. Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

© 2016 Nature America, Inc. All rights reserved.
ONLINE METHODS
Site-directed mutagenesis. Site-directed mutagenesis of the T5Fen (also known as the gene D15 product, accession number NC_000859.1, nucleotides 80,134–80,968), was conducted with oligonucleotides D153K (dATGCTCCTCTCTCTACACCACAAATGGTAAAAGTTGG) and D155K (dGTTAATAAAGTATCCATTACCATCCTGTAGATATAG), supplied by EurofinsMWG, through a previously described method36. Sequences encoding WT and mutant T5Fen were cloned into the heat-inducible pIONE X vector.

Protein production and purification. Expression of T5Fen proteins was carried out as previously described21, with the appropriate pIONEX4 derivative in E. coli strain M72(L)37. Briefly, cells were grown to mid-log phase in 1 l of rich medium containing 100 µg/ml of carbenicillin at 28 °C. Protein expression was induced by increasing the temperature to 42 °C for 3 h. Purification of proteins for biochemical assays was carried out as previously described20–22,38. Proteins for crystallography were purified by modification of these procedures and included addition of a size-exclusion-chromatography step. Briefly, initial purification was performed with a Heparin-HP chromatography column at pH 8.0 and subsequent anion-exchange chromatography on a Resource Q column at pH 6.5. Protein was applied to a Superdex-200 (GE Healthcare) size-exclusion-chromatography column (1.6 cm x 60 cm) equilibrated with 20 mM HEPES, pH 7.0, 150 mM KCl, 0.1 mM EDTA and 1 mM DTT; at a flow rate of 1 ml/min. Fractions were collected, and the protein concentration for each fraction was determined with the Bio-Rad protein assay39. Proteins were concentrated with centrifugal filtration (VivaSpin units, vivaScience) to ~15 mg/ml. Sample purity was examined by SDS–PAGE.

Real-time structure-specific endonuclease assays. Structure-specific endonuclease activity was monitored in a continuous FRET-based assay adapted from the literature40. Oligonucleotide substrate OH2P (5’-Cy3-CTCTGCGAACACACGCGCTGTTGTCTC, supplied by EurofinsMWG), labeled at a dU derivative (X) with FAM, was used as the substrate (Supplementary Fig. 2a). Oligo OHP-2 was dissolved to a final concentration of 0.1 mM in TE buffer at pH 7. Enonuclease activity was also monitored with dual-labeled poly(dT) oligonucleotide FAM_BHQ-T12 (5’-FAM-TTTTTTTTTTTTTTBHQ). Initial rates of reactions were determined at substrate concentrations ranging from 6.25 nM to 250 nM in a buffer containing 25 mM HEPES-NaOH, pH 7, 100 mM KCl, 10 mM MgCl2, 2 mM DTT and 0.1 mg/ml BSA, and the increase in fluorescence was monitored over time. The catalytic parameters of WT T5Fen were determined and compared with nonlinear regression of the Michaelis–Menten equation. Data were presented with a Hanes–Woolf plot, and the initial velocity of the reaction was determined at the linear region of the progress curve during the first few percent toward total completion of the reaction. The fluorescence emission intensity data were collected on a Hitachi F-2500 FL spectrophotometer at 496 nm/519 nm (excitation/emission wavelengths) at 1-s intervals. The mean and range of the three repeats were plotted (Fig. 2c) and compared with negative control with no added enzyme.

Binding studies. Aliquots of 199 µl of buffer containing 1 nM substrate were mixed with 1 µl of protein at different concentrations and incubated at room temperature for 20 min. The 200-µl sample was transferred into a four-window quartz cuvette and placed in a Hitachi F-2500 FL spectrophotometer with a mounted polarizer. Samples were excited with polarized light at 496 nm, and fluorescence emission readings were taken at 519 nm in both vertical and horizontal positions. The light intensities (I) were defined as follows: I1V for both the excitation and emission polarizers mounted vertically; I1H for both the excitation and emission polarizers mounted horizontally; I1V for the excitation polarizer mounted vertically and the emission polarizer mounted horizontally; and I1H for the excitation polarizer mounted horizontally and the emission polarizer mounted vertically. Anisotropy (A) is defined as A = (I1V − G1VH)/(I1V + 2 G1H), where the G factor (G) is equal to I1H/I1H (ref. 41). Experiments were repeated three times.

Crystallization and structure determination of T5Fen and T5Fen D153K proteins. T5Fen and its variants were used at a concentration of 15 mg/ml for crystallization of both protein and protein–DNA complex. Extensive screening of crystallization conditions was performed by the vapor-diffusion technique at 7 and 17 °C. Subsequently, good-quality diffraction data were collected at the Diamond Light Source, UK (X-ray beamline I03) from crystals grown in conditions optimized from those found in the screens. Crystalization of the WT T5Fen and its D153K variant incorporated 200 mM MgCl2 with the protein solution. For the WT T5Fen, the best diffraction was obtained from a crystal grown at 7 °C in 0.2 M NaBr, 0.1 M Bis-Tris propane buffer, pH 8.0, and 22% (w/v) PEG 3350. The T5Fen D153K variant also crystallized at 7 °C, with optimum conditions of 0.2 M NaI, 0.1 M Bis-Tris propane buffer, pH 7.5, and 23–25% (w/v) PEG 3350. Crystals were cryoprotected by transfer into a solution matching the growth condition plus 20% (w/v) ethylene glycol. Data collection statistics are summarized in Table 1.

Cocrystallization with DNA substrates. Two separate self-annealing oligonucleotides were used. Oligonucleotides 5ov4 (5’-dAAAAGCGTACGC-3’), which yielded an 8-bp region plus a 4-nt overhang at each 5’ end when self-annealed and 3ov6 (5’-dGATCTATATGCGCATTGG-3’), which produced a 10-bp region with two mismatches and a flipped-out nucleotide plus a 6-nt overhang at each 3’ end after annealing and mixture with the protein (Supplementary Fig. 4). The concentration of both oligonucleotides was adjusted to 1.1 mM for the duplex molecule by dissolving each one in 10 mM MES, pH 6.5, and 50 mM KCl. Anisotropy was performed by heating to 94 °C for 10 min and cooling to room temperature. Crystals with oligonucleotide 5ov4 were grown at 17 °C with the D153K variant of T5Fen. The resulting T5Fen D153K–5ov4 structure was determined from crystals grown in 0.2 M MgCl2, 0.1 M Bis-Tris buffer, pH 5.5, and 25% (w/v) PEG 3350. Crystals of the T5Fen D155K with oligonucleotide 3ov6 grew in 0.2 M CaCl2, 0.1 M sodium acetate buffer, pH 5, and 20% (w/v) PEG 6000. In all proteins, the protein–DNA complex crystals were cryoprotected by transfer into a solution matching the growth condition plus 30% (v/v) glycerol. The processing of X-ray diffraction data for all crystals with or without DNA was performed with xia2 (ref. 42) or iMosflm software43 at the Diamond Light Source (UK) beamlines I-24 (T5Fen D153K–5ov4) and I02 (T5Fen D155K–3ov6), and scaling with XDS/XSCALE44 or SCALA45 was subsequently performed. Data collection and refinement statistics are summarized in Table 1.

Structure determination and refinement. The structures of WT T5Fen and the D153K variant without DNA were determined by molecular replacement with PHASER46 with the previously determined 2.8-A-resolution structure of the enzyme in complex with Mm21 (PDB 1UTS) as a search model31. In both cases, two molecules were in the AU, and subsequent rounds of fitting and refinement were carried with COOT47 and REFMAC5 (ref. 48). Structures were validated in COOT and MOLPROBITY49. The structure of the T5Fen D153K–5ov4 complex was solved by molecular replacement with PHASER with the T5Fen structure (PDB 1EXN) as the search model31. The D153K model determined in this study was used in molecular replacement as the search model for the T5Fen D155K–3ov6 structure. Subsequent model building and refinement followed the same pattern as above for the protein alone with COOT, REFMAC and MOLPROBITY. Identification of metal cations and solvent anions in all structures was carried out on the basis of an analysis of coordination geometry, ion environment, bond lengths and refined B factors, and the CheckMyMetal web server42.

Ramachandran analysis of backbone torsion angles revealed no outliers for either the T5Fen (99% favored, remainder allowed conformations) or D153K variant models (>99% favored, remainder allowed conformations). Similarly, no outliers for either the T5Fen D153K–5ov4 (99% favored, 1% allowed conformations) or T5Fen D155K–3ov6 (99% favored, 1% allowed conformations) were observed. Data collection and refinement statistics are summarized in Table 1.

36. Sayers, J., Krekel, C. & Eckstein, F. Rapid high-efficiency site-directed mutagenesis by the phosphorothioate approach. Biotechniques 13, 592–596 (1992).
37. Sayers, J. & Eckstein, F. A single-strand specific endonuclease activity copurifies with overexpressed T5 D15 exonuclease. Nucleic Acids Res. 19, 4127–4132 (1991).
38. Garforth, S.J. & Sayers, J.R. Structure-specific DNA binding by bacteriophage T5 5’3’ exonuclease. Nucleic Acids Res. 25, 3801–3807 (1997).
39. Bradford, M.M. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein–dye binding. Anal. Biochem. 72, 248–254 (1976).
40. Hutton, R.D., Craggs, T.D., White, M.F. & Penedo, J.C. PCNA and XPF cooperate to distort DNA substrates. *Nucleic Acids Res.* **38**, 1664–1675 (2010).
41. Heyduk, T., Ma, Y., Tang, H. & Ebright, R.H. Fluorescence anisotropy: rapid, quantitative assay for protein-DNA and protein-protein interaction. *Methods Enzymol.* **274**, 492–503 (1996).
42. Winter, G., Lobley, C.M. & Prince, S.M. Decision making in xia2. *Acta Crystallogr. D Biol. Crystallogr.* **69**, 1260–1273 (2013).
43. Battye, T.G., Kontogiannis, L., Johnson, O., Powell, H.R. & Leslie, A.G. iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 271–281 (2011).
44. Kabsch, W. Integration, scaling, space-group assignment and post-refinement. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 133–144 (2010).
45. Evans, P. Scaling and assessment of data quality. *Acta Crystallogr. D Biol. Crystallogr.* **62**, 72–82 (2006).
46. McCoy, A.J. et al. Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
47. Emsley, P., Lohkamp, B., Scott, W.G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501 (2010).
48. Murshudov, G.N. et al. REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 355–367 (2011).
49. Chen, V.B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 12–21 (2010).
50. Zheng, H. et al. Validation of metal-binding sites in macromolecular structures with the CheckMyMetal web server. *Nat. Protoc.* **9**, 156–170 (2014).