This is a repository copy of *Regional Conformational Flexibility Couples Substrate Specificity and Scissile Phosphate Diester Selectivity in Human Flap Endonuclease 1*.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/130528/

Version: Published Version

Article:
Bennet, I., Finger, L.D. orcid.org/0000-0002-2342-9569, Baxter, N.J. et al. (8 more authors) (2018) Regional Conformational Flexibility Couples Substrate Specificity and Scissile Phosphate Diester Selectivity in Human Flap Endonuclease 1. Nucleic Acids Research, 46 (11). pp. 5618-5633. ISSN 0305-1048

https://doi.org/10.1093/nar/gky293

Reuse
This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here: https://creativecommons.org/licenses/

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk/

Nucleic Acids Research, 2018 **1**
doi: 10.1093/nar/gky293

Regional conformational flexibility couples substrate specificity and scissile phosphate diester selectivity in human flap endonuclease 1

Ian A. Bennet[1,†], L. David Finger[1,*,†], Nicola J. Baxter[2,3], Benjamin Ambrose[1], Andrea M. Hounslow[2], Mark J. Thompson[1], Jack C. Exell[1], Nur Nazihah B. Md. Shahari[1], Timothy D. Craggs[1,*], Jonathan P. Waltho[2,3] and Jane A. Grasby[1,*]

[1]Centre for Chemical Biology, Department of Chemistry, Krebs Institute for Biomolecular Research, The University of Sheffield, Sheffield S3 7HF, UK, [2]Department of Molecular Biology and Biotechnology, Krebs Institute for Biomolecular Research, The University of Sheffield, Sheffield S10 2TN, UK and [3]Manchester Institute of Biotechnology, School of Chemistry, The University of Manchester, Manchester M1 7DN, UK

Received September 01, 2017; Revised March 26, 2018; Editorial Decision April 04, 2018; Accepted April 09, 2018

ABSTRACT

Human flap endonuclease-1 (hFEN1) catalyzes the divalent metal ion-dependent removal of single-stranded DNA protrusions known as flaps during DNA replication and repair. Substrate selectivity involves passage of the 5′-terminus/flap through the arch and recognition of a single nucleotide 3′-flap by the α2–α3 loop. Using NMR spectroscopy, we show that the solution conformation of free and DNA-bound hFEN1 are consistent with crystal structures; however, parts of the arch region and α2–α3 loop are disordered without substrate. Disorder within the arch explains how 5′-flaps can pass under it. NMR and single-molecule FRET data show a shift in the conformational ensemble in the arch and loop region upon addition of DNA. Furthermore, the addition of divalent metal ions to the active site of the hFEN1–DNA substrate complex demonstrates that active site changes are propagated via DNA-mediated allostery to regions key to substrate differentiation. The hFEN1–DNA complex also shows evidence of millisecond timescale motions in the arch region that may be required for DNA to enter the active site. Thus, hFEN1 regional conformational flexibility spanning a range of dynamic timescales is crucial to reach the catalytically relevant ensemble.

INTRODUCTION

Flap endonuclease 1 (FEN1) is a member of the 5′-nuclease superfamily that is predominantly involved in Okazaki fragment maturation, but also has roles in long-patch base excision repair and ribonucleotide excision repair (1,2). All three of these pathways create bifurcated nucleic acid structures known as 5′-flaps that must be removed precisely to create single-stranded (ss) and nicked double-stranded (ds) DNA products (Figure 1A). In line with a critical role in DNA replication, FEN is present in all organisms, from bacteriophage to mammals (3). In humans, cancer cells overexpress FEN1, and tumor aggressiveness correlates with FEN1 protein levels (4). As such, human FEN1 (hFEN1) has been postulated to be a potential cancer therapeutic target (5,6), and evidence suggests that combinatorial targeting of hFEN1 has therapeutic relevance (7). Moreover, FEN1 is the prototypical member of the 5′-nuclease superfamily whose activities span all major DNA metabolic pathways. As nucleases, specificity is paramount as unwanted hydrolysis of DNA or RNA can be deleterious; thus, how hFEN1 and paralogues achieve substrate and scissile phosphate diester specificity has been an area of considerable debate.

Earlier studies revealed that hFEN1 specificity arises from the recognition of three structural features of its substrate (Figure 1A) (2,8). Firstly, only junction dsDNAs can be bound stably by the protein as two juxtaposed dsDNA binding sites require the substrate to bend roughly 100° (Figure 1B) (9–11). Secondly, despite removing 5′-flaps, hFEN1 recognizes substrates bearing a one nucleotide (1-nt) 3′-flap (12–14). In structures of hFEN1–DNA com-

*To whom correspondence should be addressed. Tel: +44 114 222 9478; Fax: +44 114 222 9346; Email: j.a.grasby@sheffield.ac.uk
Correspondence may also be addressed to L. David Finger. Tel: +44 114 222 9364; Fax: +44 114 222 9346; Email: d.finger@sheffield.ac.uk
Correspondence may also be addressed to Timothy D. Craggs. Tel: +44 114 222 9347; Fax: +44 114 222 9346; Email: t.craggs@sheffield.ac.uk
† The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.
Present address: Jack C. Exell, Department of Chemistry, 1207 French Family Science Centre, Duke University, Durham, NC 27705, USA.

© The Author(s) 2018. Published by Oxford University Press on behalf of Nucleic Acids Research.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Downloaded from https://academic.oup.com/nar/advance-article-abstract/doi/10.1093/nar/gky293/4990015
by University of Sheffield user
on 03 May 2018

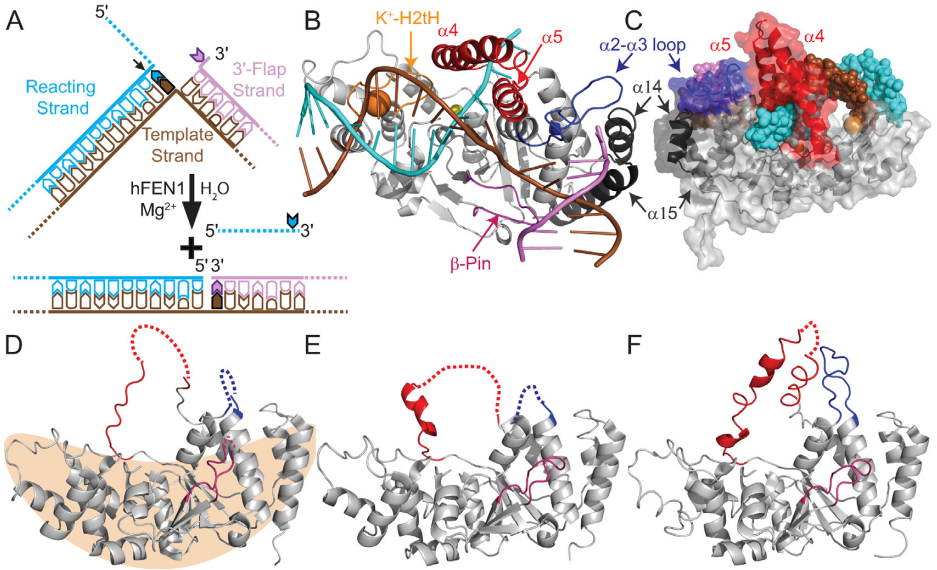

Figure 1. Regions of hFEN1 assume various conformations in crystal structures. (**A**) Representation of hFEN1-catalyzed reaction of double-flap DF#,1 creating ssDNA and nicked dsDNA products (2). # denotes a 5′-flap of any length, 1 denotes a single nucleotide 3′-flap. (**B**) The $hFEN1_{D86N}$-(DF2,1) complex (5UM9) (21) shows that the arch (α4-α5, red), α2–α3 loop (blue) and β-pin (β6-β7 loop) (pink) are ordered in the presence of DNA substrate (colored as in panel A). The α10-α11 loop (H2tH motif) that binds the reacting duplex approximately 20 Å from the scissile phosphate and the α14-loop-α15 that is unique to FEN proteins are shown in orange and black, respectively. The potassium and active site samarium cations are shown as orange and yellow spheres, respectively. (**C**) Rear view of 5UM9 (cartoon with transparent surface representation) with regions colored as in panels A and B and the DNA (DF2,1) shown as spheres. (**D**–**F**) The three crystal structures of apo-hFEN1 (1UL1) show the arch, α2–α3 loop and β-pin regions in various states of order and disorder, dashed lines indicate missing electron density (22). (D) The tan shape illustrates the saddle-region of hFEN1 that binds the dsDNA and active site metals.

plexes, the 3′-flap is cradled in a pocket created from the α14–α15 loop, the first turn of α15 and the α2–α3 loop (11). The α2–α3 loop folds into a non-regular secondary structure known as an Ω-loop (15), often associated with allosterically-regulated regions of proteins (Supplementary Figure S1) (16). Thirdly, the 5′-flap portion of the substrate must pass under the arch region (α4-α5) for biologically-relevant rates of reaction to occur (Figure 1B and C) (17,18). This latter substrate selection feature is also shared by the 5′-nuclease superfamily member human exonuclease-1 (19,20), but not by other family members that act on continuous DNAs like bubble-cutting XPG and the Holliday junction resolvase GEN1 (3). When these three substrate recognition tests are met, the DNA can untwist and enter the active site for catalysis. Of the three hFEN1 substrate feature selection processes, the threading of the 5′-flap through the arch remains the most enigmatic.

Crystal structures of hFEN1–DNA complexes (3Q8K, 3Q8L, 3Q8M, 5UM9) show the arch is composed of two α-helices above the active site (11,21). Together these create a hole in the protein that is large enough to accommodate ssDNA but not dsDNA (11,21), which could explain hFEN1 specificity for 5′-ss flaps (Figure 1C). However, this is inconsistent with the ability of hFEN1 to process 5′-flaps with some duplex secondary structure (12,18). A further puzzle relates to how threading could occur through a small hole in the absence of an energy source. In contrast, apo-hFEN1 structures show that the regions corresponding to the arch and α2–α3 loop sometimes lack discernible electron density; what has been observed is either random coil or a limited degree of secondary structure (Figure 1D–F) with high B-factors (Supplementary Figure S2) (22). Furthermore, structures of FEN proteins from a wide array of organisms show the arch region in various conformations and positions (Supplementary Figure S3) (2). It is possible that structural heterogeneity of the arch and α2–α3 loop has some functional relevance. However, direct confirmation of the structural status of the arch in free and substrate bound hFEN1 is necessary to define its mechanism.

To extend our understanding, multidimensional NMR spectroscopy was used to assess the properties and dynamics of hFEN1 alone and in complex with DNA in solution. We show that the solution conformation of hFEN1 is consistent with the crystal structure in the dsDNA binding region known as the saddle (Figure 1D), but a large discrepancy is observed in the arch region (traditionally referred to as the helical arch). Using a combination of kinetic, spin relaxation, chemical shift and single molecule FRET data, the contribution of hFEN1 conformationally dynamic regions to substrate specificity is revealed. Chemical shift perturbation mapping provides evidence that divalent cations induce conformational changes in the hFEN1–DNA complex that are not observed in hFEN1 alone. These data demonstrate that the active site communicates with regions involved in dsDNA binding and substrate feature recognition. Combined data raise the possibility that both conformational selection and induced fit mechanisms occur in the arch region in response to structural features of the substrate and reveal that the hFEN1–DNA complex exhibits millisecond timescale motions in the arch. Taken together, our analysis shows that regional conformational flexibility spanning a range of dynamic timescales is channeled into

Downloaded from https://academic.oup.com/nar/advance-article-abstract/doi/10.1093/nar/gky293/4990015
by University of Sheffield user
on 03 May 2018

the catalytically-competent conformational ensemble in response to substrate features, providing the required link between specificity and catalysis in a structure-sensing nuclease. Because the structures of other 5′-nuclease superfamily show disorder or indicate flexibility in the area that corresponds to the hFEN1 arch region (19,20,23–25), a similar conformational ensemble shift mechanism in response to other paralogue-specific triggers is likely important in their mechanisms as well.

MATERIALS AND METHODS

Protein expression

Human hFEN1-336 was expressed from *Escherichia coli* BL21(DE3)-RILP cells transformed with pET29b-hFEN1-336 plasmid, encoding for hFEN1-336 (herein referred to as hFEN1) protein containing a human rhinovirus (HRV) type 14 3C protease-cleavable $(His)_6$-tag located at the C-terminus. Natural abundance protein for kinetic analyses was expressed and purified as described previously (7). ^{15}N-labeled protein was expressed using ^{15}N autoinduction media (26). The final concentrations of ^{15}N autoinduction media components were: 50 mM Na_2HPO_4 and 50 mM KH_2PO_4 pH 7.5, 50 mM $^{15}NH_4Cl$, 5 mM Na_2SO_4, 2 mM $MgSO_4$, 0.5% glycerol, 0.05% D-glucose, 0.2% α-lactose, 12× BME vitamins (USBiological B0110), 1× Trace Metals (Teknova 1000x Trace Metals T1001), 400 μg ml^{-1} kanamycin and 34 μg ml^{-1} chloramphenicol. Cultures for ^{15}N-labeled protein expression were inoculated with a 100-fold dilution of a saturated 2× YT starter culture, and were then allowed to grow at 37°C for 4 h before the temperature was reduced to 18°C to allow for overnight expression by autoinduction.

$^{2}H,^{13}C,^{15}N$-labeled protein was prepared using a high cell-density procedure in combination with isopropyl-β-D-thiogalatopyranoside (IPTG) induction (27). The final concentrations of all media components were: 50 mM Na_2HPO_4 and 25 mM KH_2PO_4 pH 7.5, 18 mM $^{15}NH_4Cl$, 1% $^{13}C_6,^{2}H_7$-D-glucose (U-$^{13}C_6$, 99%; 1,2,3,4,5,6,6-d_7 97-98%), 0.2 mM $CaCl_2$, 5 mM $MgSO_4$, 10 mM NaCl, 0.25 × BME vitamins (USBiological B0110), 0.25 × Trace Metals (Teknova 1000x Trace Metals T1001), 400 μg.ml^{-1} kanamycin and 34 μg ml^{-1} chloramphenicol. Initially, a 100-fold dilution of a saturated culture of BL21(DE3)-RILP cells transformed with pET29b-hFEN1-336 was used to inoculate 2× YT media prepared with 50% D_2O and was grown until saturation. A 100-fold dilution of this saturated 50% D_2O 2× YT culture was then used to inoculate 2× YT media in 100% D_2O to allow the cells to acclimatize for growth. Aliquots (50 ml) of these D_2O acclimatized BL21(DE3)-RILP cells were added to 1 L of 100% D_2O 2× YT media and were grown at 37°C until the A_{600} = 5–6. The culture was pelleted by ultracentrifugation (6000 × *g* at 4°C for 30 min) and the supernatant was discarded. The pellet was then resuspended in 1 L of $^{15}N,^{2}H,^{13}C$ minimal media (as described above) and then allowed to grow at 37°C for a further 2 h to clear unlabeled metabolites. Afterward, IPTG was added to a final concentration of 0.1 mM IPTG, the temperature was reduced to 18°C, and the culture was grown for a further 72 h until A_{600} = 10–13. $^{2}H,^{15}N$-labeled protein was prepared analogously except that the 1% $^{13}C_6,^{2}H_7$-D-glucose (U-$^{13}C_6$, 99%; 1,2,3,4,5,6,6-d_7 97–98%) was replaced with 1% $^{2}H_7$-D-glucose (1,2,3,4,5,6,6-d_7 97–98%).

In all cases, cells were harvested by centrifugation (6000 × *g* at 4°C for 15 min) and the supernatant was removed. The cell pellet was resuspended in 50 ml of ice-cold phosphate buffered saline (10 mM Na_2HPO_4 and 2 mM KH_2PO_4 pH 7.4, 137 mM NaCl, 2.7 mM KCl), the cells were pelleted again at (6000 × *g* at 4°C for 15 min) and the supernatant was removed. The cells were then resuspended in 50 ml of Buffer A (20 mM Tris–HCl pH 7.4, 1 M NaCl, 5 mM imidazole) containing 5 mM β-mercaptoethanol (βME), 0.1 mg ml^{-1} lysozyme and 1 EDTA-free protease inhibitor cocktail tablet (Sigma S8830). Cells were incubated on ice for at least 1 h and were frozen at −20°C until required.

Protein purification

To purify hFEN1 protein, cells were thawed and then sonicated on ice for 60 cycles of pulsation for 5 s with 25 s cooling intervals. To each lysate, 5 ml of Buffer A containing 1% Tween-20 was added, and the lysate was clarified by centrifugation (40 000 × *g* at 4°C for 30 min). The supernatant was applied to a column (ID = 1.6 cm, length = 12 cm) containing Chelating Sepharose 6 Fast Flow agarose beads (GE Healthcare Life Sciences) that had been previously charged with Ni^{2+} ions according to the manufacturer's protocol and pre-equilibrated with 5 column volumes of Buffer A. The column was washed with 5 column volumes of Buffer A and 5 column volumes of Buffer B (25 mM Tris–HCl pH 7.4, 0.5 M NaCl, 40 mM imidazole, 0.01% Tween-20, 5 mM βME). The bound hFEN1 protein was eluted with 8 column volumes of Buffer C (250 mM imidazole pH 7.2, 0.5 M NaCl, 5 mM βME). Fractions containing hFEN1 protein were identified by SDS-PAGE, pooled and immediately diluted with an equal volume of Buffer D (20 mM Tris–HCl pH 7.4, 1 mM EDTA, 20 mM βME). The solution was then applied to a 5 mL HiTrap Q HP column (GE Healthcare Life Sciences) pre-equilibrated with the buffer D. hFEN1 protein was found in the flow through, whereas nucleic acid contamination was retained by the column and eluted using 10 column volumes of Buffer D containing a linear gradient of 0–1 M NaCl. The amount of hFEN1 protein in the flow through was estimated by measuring the absorbance at 280 nm (ε_{280} = 22 920 M^{-1} cm^{-1}) (28) using a UV-Vis NanoDrop spectrophotometer (ThermoFisher Scientific). Two units of Turbo3C (HRV3C protease) (BioVision) were added for every mg of protein, and the mixture was allowed to stand overnight at 4°C to catalyze cleavage of the $(His)_6$-tag from the hFEN1 protein. The extent of affinity tag removal was assessed by SDS-PAGE, and the solution was diluted further with an equal volume of buffer E (25 mM Tris–HCl pH 7.5, 1 mM $CaCl_2$, 20 mM βME). The hFEN1 protein solution was then applied to a 20 ml HiPrep Heparin FF 16/10 column (GE Healthcare Life Sciences) pre-equilibrated with 2 column volumes of buffer E and eluted using 15 column volumes of Buffer E containing a linear gradient of 0–1 M NaCl (hFEN1 protein eluted at 350 mM NaCl). Fractions containing hFEN1 protein were pooled and concentrated to less than 10 ml using an Am-

Downloaded from https://academic.oup.com/nar/advance-article-abstract/doi/10.1093/nar/gky293/4990015
by University of Sheffield user
on 03 May 2018

icon Ultrafiltration cell with a 10 kDa MWCO PES membrane (MerckMillipore) pressurised using nitrogen gas. The protein was exchanged into the appropriate buffer (see below) using a HiPrep 26/10 desalting column (GE Healthcare Life Sciences). The hFEN1 protein concentration was determined using the absorbance at 280 nm as described above. Unlabeled samples for kinetic measurements were exchanged into 100 mM HEPES–KOH pH 7.5, 200 mM KCl, 1 mM EDTA, 10 mM DTT and 0.04% NaN_3, and adjusted to a concentration of 200 μM. An equal volume of glycerol was added to each sample to allow optimal storage conditions (final concentration: 100 μM hFEN1, 50 mM HEPES–KOH pH 7.5, 100 mM KCl, 0.5 mM EDTA, 5 mM DTT, 0.02% NaN_3 and 50% glycerol). Samples were stored at −20°C. Isotopically labeled hFEN1 protein samples were exchanged into the appropriate NMR buffer and concentrated to 0.5 mM using a Vivaspin 20 (10 kDa MWCO) concentrator (4000 × *g* at 4°C).

All purification buffers were filtered and degassed prior to use, and an ÄKTA pure chromatography system was used for all column purification steps. Unless otherwise stated, all reagents were purchased from Sigma-Aldrich, Fisher Scientific, or VWR International and used as received.

NMR assignment

All NMR experiments were performed at 298 K (relative to d_4-methanol signal) using standard pulse sequences on either a Bruker 600 MHz Avance DRX spectrometer equipped with a 5-mm TXI cryoprobe with z-axis gradients or a Bruker 800 MHz Avance I spectrometer equipped with a 5-mm TXI probe with z-axis gradients. Both spectrometers were operated with TopSpin 2. Comparison of ^{1}H–^{15}N TROSY spectra of ^{2}H,^{15}N,^{13}C-labeled hFEN1 with ^{15}N-labeled hFEN1 showed that all backbone amide groups had back exchanged from deuterium to protium atoms. For the backbone resonance assignment of hFEN1, 0.5 ml samples contained 0.5 mM ^{2}H,^{15}N,^{13}C-labeled hFEN1 in 10 mM HEPES–KOH pH 7.5, 76 mM KCl, 0.1 mM EDTA, 4 mM NaN_3, 100 mM βME. D_2O (10% v/v) and 0.05 mM trimethylsilyl propanoic acid (TSP) were added for the deuterium lock and chemical shift reference standard, respectively. After transferring the sample into a 5 mm NMR tube, the sample was sealed with a Precision Seal® rubber septa cap (Sigma-Aldrich Z554014) and argon was passed over the sample to help minimize oxidation.

hFEN1 backbone $^{1}H^{N}$, ^{15}N, $^{13}C^{\alpha}$, $^{13}C^{\beta}$ and $^{13}C'$ were assigned using the standard suite of ^{1}H–^{15}N TROSY and 3D TROSY-based constant time experiments (HNCA, HN(CO)CA, HNCACB, HN(CO)CACB, HN(CA)CO and HNCO) (29). ^{1}H chemical shifts were referenced relative to the internal TSP signal, whereas ^{15}N and ^{13}C chemical shifts were indirectly referenced using nuclei-specific gyromagnetic ratios. Peak picking and frequency matching was performed within CCPNMR Analysis version 2.4 (30), and the backbone assignments were confirmed independently using a simulated annealing algorithm employed by the *asstools* assignment program (31). The secondary structure content and Random Coil Index-S^2 (RCI-S^2) prediction of hFEN1 was conducted by uploading the backbone $^{1}H^{N}$, ^{15}N, $^{13}C^{\alpha}$, $^{13}C^{\beta}$ and $^{13}C'$ chemical shifts to the TALOS-N webserver (32).

NMR relaxation measurements

hFEN1 samples for ^{15}N backbone fast timescale relaxation measurements were performed using ^{2}H,^{15}N-labeled hFEN1 in a 5 mm Shigemi NMR microtube assembly matched with D_2O. Sample conditions were 10 mM HEPES–KOH pH 7.5, 76 mM KCl, 4 mM NaN_3 100 mM βME, 10% D_2O, 0.05 mM TSP and 0.1 mM EDTA. Spin-lattice ^{15}N relaxation rates (R_1), rotating frame ^{15}N relaxation rates ($R_{1\rho}$) and heteronuclear steady-state ^{15}N-$\{^{1}H\}$ NOE (hNOE) values were obtained using interleaved TROSY-readout pulse sequences (33). Temperature compensation was applied in the R_1 experiment by incorporating a spin-lock pulse placed off-resonance in the inter-scan delay, equal to the longest spin-lock time and RF power of the $R_{1\rho}$ experiment. Relaxation delays of 0, 80, 240, 400, 400, 640, 800, 1200, 1760 and 2400 ms were used to calculate R_1, and delays of 1, 20, 20, 30, 40, 60, 90, 110, 150 and 200 ms were used to calculate $R_{1\rho}$. The inter-scan delay was 3.5 s and the strength of the RF spin-lock field during $R_{1\rho}$ measurement was 1400 Hz at 600 MHz and 1866.7 Hz at 800 MHz. For the hNOE measurement, two interleaved experiments were acquired with relaxation delays of 10 s. For the determination of R_1 and $R_{1\rho}$ rates, the decay of backbone amide peak intensities were fitted using an exponential function in CCPNMR (30). Relaxation parameters were obtained for 192 residues because 59 residues were omitted from further analysis due to peak overlap or poor signal to noise ratios. For hNOE experiments, the ratio of the saturated to unsaturated peak height was measured. R_2 values were calculated from R_1 and $R_{1\rho}$ rates according to Equation ((1)).

$$R_2 = R_{1\rho}/\sin^2\theta - R_1/\tan^2\theta \quad (1)$$

where $\tan\theta = \omega_1 / \Omega$, ω_1 is the spin lock RF field (1400 Hz at 600 MHz and 1866.7 Hz at 800 MHz) and Ω is the offset of the ^{15}N resonance of interest with respect to the ^{15}N carrier frequency (33).

Model-free analysis

Model-free analysis was performed using *relax* (34–38) on the Sheffield-WRGRID ICEBERG high performance computing cluster. Using R_1, R_2 and hNOE values at 600 MHz and 800 MHz and the coordinate geometry of backbone amide N-H bonds as provided by the crystal structure 3Q8K (11), model-free analysis was executed. However, residues from the arch region, α2–α3 loop and the C-terminus were excluded due to lack of coordinates in the crystal structures (1UL1) (22) or low R_2/R_1 values. Instead, these 16 residues were modeled with a spherical diffusion tensor. Both analyses were conducted for the 192 residues for which relaxation data at both field strengths was available; of these, only 179 were processed fully, as 13 were removed due to large errors or computational eliminations.

Downloaded from https://academic.oup.com/nar/advance-article-abstract/doi/10.1093/nar/gky293/4990015
by University of Sheffield user
on 03 May 2018

Titration of Ca^{2+} and Mg^{2+} cations into $hFEN1_{K93A}$

A ^{15}N-labeled hFEN1 K93A mutant ($hFEN1_{K93A}$), which hindered catalysis, was prepared in the pET29b-hFEN1-336 vector as described previously (11). The $hFEN1_{K93A}$ protein was prepared as described for the wild type above using the ^{15}N autoinduction method. ^{1}H–^{15}N TROSY spectra were separately recorded at 10 mM HEPES–KOH pH 7.5, 76 mM KCl, 0.1 mM EDTA, 4 mM NaN_3, 100 mM βME, 10% D_2O and 0.05 mM TSP with either 0 mM, 8 mM $MgCl_2$ and 8 mM $CaCl_2$ added. Absolute changes in weighted chemical shifts (ω) were determined using Equation ((2)) where the correction factor for ^{15}N was $\alpha = 0.14$.

$$\omega = \left(\Delta\delta_H{}^2 + (\alpha\Delta\delta_N)^2\right)^{1/2} \quad (2)$$

DNA preparation

DNA oligonucleotides were purchased with purification from LGC Biosearch Technologies. The sequence of the oligonucleotides are shown in Table 1. Oligonucleotide concentrations were quantified using absorbance at 260 nm and molar extinction coefficients calculated according to the 'nearest-neighbor' method (39,40).

$hFEN1_{K93A}$-DNA complex formation

DHPS1 (DNA substrate) was heated and annealed in 10 mM HEPES–KOH pH 7.5, 6 mM KCl, 4 mM NaN_3 and 0.1 mM EDTA. Initially, 0.5 mM ^{15}N-labeled $hFEN1_{K93A}$, ^{2}H,^{13}C,^{15}N-labeled $hFEN1_{K93A}$ or ^{2}H,^{15}N-labeled $hFEN1_{K93A}$, prepared as described above, in 10 mM HEPES–KOH pH 7.5, 76 mM KCl, 4 mM NaN_3 and 0.1 mM EDTA was titrated with sub-stoichiometric aliquots of lyophilized DHPS1 until the DNA was in excess of the protein (>500 μM). At 50–100 μM of DNA, the sample precipitated and the quality of the spectra decreased rapidly. To overcome protein precipitating in complex with DNA, an $hFEN1_{K93A}$-DNA complex was prepared in 10 mM HEPES–KOH pH 7.5, 6 mM KCl, 4 mM NaN_3 and 0.1 mM EDTA in the presence of excess DNA substrate (1:1.1) under dilute conditions of protein and DNA (roughly 5 μM). Low monovalent salt concentrations were used to slow dissociation of the complex. The sample was concentrated to ∼500 μM using Vivaspin 20 (10 kDa MWCO) spin columns and 100 mM βME, 10% D_2O (10% v/v) and 0.05 mM TSP were added. The same resonance assignment strategy using 3D NMR spectra was conducted with a ^{2}H,^{15}N,^{13}C-labeled $hFEN1_{K93A}$–DNA complex prepared in this manner. ^{1}H-^{15}N TROSY spectra were acquired at 0, 0.5, 1, 2, 4, 6 and 8 mM $CaCl_2$, and chemical shift differences were recorded as described in Equation (2).

Preparation of hFEN1 cysteine mutants labeled with fluorophores

Surface residues C235 and C311 were successively mutated to alanine in the pET28b-hFEN1 plasmid using Agilent site-directed mutagenesis protocols (41) and the appropriate oligonucleotides (Supplementary Table S1). The resulting double mutant hFEN1 protein (C235A/C311A) was expressed and purified as described above. This mutant did not react with maleimide dyes on a timescale of 1 h. Subsequent mutagenesis as above created $hFEN1_Q$, E120C/C235A/S293C/C311A. $hFEN1_Q$ protein was expressed and purified as described above. ESI-MS calculated: 43210.6 Da experimental: 43210.9 Da.

To stochastically label $hFEN1_Q$ with the appropriate fluorophores, 250 μl of 100 μM (25 nmol) purified $hFEN1_Q$ in 50% glycerol storage buffer was diluted 1:1 with buffer F (50 mM Tris–HCl $pH_{25°C}$ 7, 150 mM NaCl, 1 mM EDTA and 0.1 mM TCEP) to dilute the glycerol. DTT was added to the diluted $hFEN1_Q$ to a final concentration of 20 mM, and the solution allowed to incubate on ice for 15 min. The $hFEN1_Q$ solution was then subjected to size exclusion chromatography (SEC) using a Superdex 75 GL 10/300 column (GE Healthcare Life Sciences) and buffer F. Protein containing fractions were pooled and concentrated using a 10 kDa MWCO Vivaspin-20. The $hFEN1_Q$ protein concentration was quantified using the absorbance at A_{280} nm and calculated extinction coefficient (28). Stocks of 10 mM Atto 647N maleimide (Sigma-Aldrich) and 10 mM Cy3b maleimide (GE Healthcare Life Sciences) were prepared in DMSO. To label $hFEN1_Q$, the protein was added to a tube containing the appropriate volume of each stock of fluorophore maleimide conjugate to give a molar ratio of 1:2:2 of hFEN1:Atto 647N:Cy3b. Aliquots (usually 50 μl) of the reaction were removed at 2, 4, 8, 16 and 26 min and quenched by addition of excess DTT. To remove excess free fluorophore, the quenched aliquots were exchanged into buffer F using micro Bio-Spin 6 columns (BioRad) according to the manufacturer's instructions. The labeled protein was again subjected to SEC as described above. The protein containing fractions were pooled and concentrated as before. The concentrations of the protein and the covalently attached Cy3b and Atto 647N were quantified using a UV/Vis Nanodrop spectrohphotometer (ThermoFisher) at A_{280}, A_{559} and A_{646}, respectively, and the associated extinction coefficients (hFEN1 protein $\varepsilon_{280} = 22\,920\ M^{-1}\ cm^{-1}$, Cy3b $\varepsilon_{559} = 130\,000\ M^{-1}\ cm^{-1}$ and Atto 647N $\varepsilon_{646} = 150\,000\ M^{-1}\ cm^{-1}$). A_{280} values were corrected for contributions from both dyes using A_{280}/A_{556} and A_{280}/A_{646} ratios.

Kinetic measurements of hFEN1, mutants and fluorescently labeled proteins

The activities of hFEN1, $hFEN1_Q$ and $hFEN1_Q$ with attached fluorophores ($hFEN1_{QF}$) were assessed using a 5′-fluorescein (FAM) labeled bimolecular DNA oligo construct DF5,1*, which was prepared by heating and annealing F1* and T1 in a 1:1.1 ratio (Table 1), respectively, in 50 mM HEPES–KOH pH 7.5, 100 mM KCl and 0.02% NaN_3. Reactions mixtures (180 μl) containing 50 nM DF5,1* in reaction buffer (55 mM HEPES–KOH pH 7.5, 110 mM KCl, 8 mM $MgCl_2$, 0.1 mg ml^{-1} BSA. 1 mM DTT) were initiated by adding appropriate amounts of hFEN1 proteins (3–6 pM final concentration) that afforded less than ∼10% product formation in 10 min at 37°C. Aliquots (20 μl) of reaction were quenched in excess EDTA (50 μl of 500 mM EDTA) at 2, 4, 6, 8, 10 and 20 min. The amount of product formation was assessed by denaturing high performance liquid chromatography (dHPLC) as described previously (18,42). Plots

Downloaded from https://academic.oup.com/nar/advance-article-abstract/doi/10.1093/nar/gky293/4990015
by University of Sheffield user
on 03 May 2018

Table 1. Oligonucleotide sequences used herein

Name	Sequence
DHPS1	5′dTGAAAGGCAGAGCGCTAGCTCTGCCTTTCGAGCGAAGCTCC3′
*F1**	5′FAM*-dTTTTTACAAGGACTGCTCGACAC3′
F1	5′dTTTTTACAAGGACTGCTCGACAC3′
T1	5′dGTGTCGAGCAGTCCTTGTGACGACGAAGTCGTCC3′

*5′FAM is a 5′-terminal fluorescein modification produced using the single isomer 5-carboxyfluorescein-aminohexyl amidite.

of concentration of product versus time yielded the initial rate of reaction (v_0 nM min^{-1}), which could be normalized for enzyme concentration ($v_0/[E]_0$ min^{-1}).

For second order rate constant determination as a function of viscosity, reaction mixtures (180 μl) were assayed at 2.5, 5, 7.5 and 10 nM DF5,1* substrate in reaction buffer at 37°C and the indicated relative viscosity. Buffer viscosity was adjusted using either glycerol (0−36% v/v) or sucrose (0−35%). Glycerol relative viscosities and their corresponding %v/v were calculated using temperature corrected density calculations at 37°C (43). Relative viscosity was adjusted with sucrose as described previously (42). Product formation was monitored by dHPLC and normalized initial rates of reaction were generated as described above. Estimates of second order rate constants were derived from the slope of normalized initial rate (v/[E]) versus the concentration of substrate [S]. A calculated second order rate constant $(k_{cat}/K_M)_0$ value at relative viscosity of 1 was derived from the Y-intercept from a plot of k_{cat}/K_M versus relative viscosity (44).

Single molecule FRET measurements and analysis

Labeled hFEN1 ($hFEN1_{QF}$) was diluted to ∼50 pM in binding buffer (55 mM HEPES–KOH pH 7.5, 110 mM KCl, 8 mM $CaCl_2$, 0.1 mg ml^{-1} BSA, 1 mM DTT)) and incubated at room temperature for 5 min with or without 20 nM substrate DNA (i.e. DF5,1, prepared by heating and annealing F1 and T1 in a 1:1.1 ratio (Table 1)). Glass slides were passivated with 1 mg/ml BSA for 5 min prior to each measurement. Three triplicate 10-min data sets were acquired for each sample giving a total of 90 min of data for each condition, yielding ∼4000 bursts.

smFRET data were acquired using a custom built confocal microscope and alternating laser excitation (45). Two dioide lasers (515 nm and 635 nm – LuxX plus) were directly modulated (100 us, duty cycle 45%) and combined into an optical fibre. The output beam was collimated and then cropped to 2.5 mm diameter by an iris. The beam was directed into the back of the objective (Olympus UPLSAPO 60× NA = 1.35 oil immersion) using a dichroic mirror (Chroma ZT532/640 rpc 3 mm) with the fluorescence emission collected by the same objective, focussed onto a 20 um pinhole and then split (dichroic mirror: Chroma NC395323 – T640lpxr) for detection by two avalanche photodiodes (SPCM-AQRH-14 and SPCM-NIR-14, Excilitas). Photon arrival times were recorded by a national instruments card (PCIe-6353), with the acquisition controlled using custom software (LabView 7.1).

smFRET data analysis was performed using custom Jupyter notebooks, based on the open source python software, FRETBusts, described previously (46). Briefly, background rates in each channel were obtain for each 60 seconds of the acquisition by fitting a poisson distribution to inter-photon delays. This background rate was recalculated every 60 s to take account of any change in background throughout the measurement time. Bursts were identified using a dual channel burst search as previously described (47).

Apparent FRET efficiencies (E^*) and apparent stoichiometries (S^*) were calculated for each burst according to Equations (3) and (4), respectively:

$$E^* = \frac{I_{A|D}}{I_{D|D} + I_{A|D}} \quad (3)$$

$$S^* = \frac{I_{D|D} + I_{A|D}}{I_{D|D} + I_{A|D} + I_{A|A}} \quad (4)$$

where I represents the background corrected intensity in the (i) acceptor emission channel after donor excitation ($I_{A|D}$), (ii) donor emission channel after donor excitation ($I_{D|D}$) and (iii) acceptor emission channel after acceptor excitation ($I_{A|A}$). Apparent FRET efficiencies were corrected for donor leakage, acceptor direct excitation, and the detection efficiencies/quantum yields (gamma correction) according to published protocols (48) (arXiv:1710.03807), using labeled DNA standards for the gamma correction.

After correction factors were applied, bursts were selected with at least 40 photons under green excitation, and 10 photons under red excitation, with a maximum of 300 photons to filter out an acceptor heavy aggregate observed in the data set. Bursts with an observed stoichiometry between 0.5 and 0.85 were plotted to generate a histogram of relative frequency versus FRET efficiency and fitted with an unrestrained double Gaussian function.

RESULTS

NMR assignments are consistent with known secondary structure and peak intensities suggest regions with unusual dynamic properties

Human FEN1 (hFEN1) is a 380 amino acid protein that has a nuclease core domain (amino acid residues 1–336) that is sufficient for catalysis *in vitro*. As we were most interested in the dynamics associated with catalysis, the hFEN1–336 construct from X-ray crystallographic studies was used for NMR studies (11,21). Thus, the final 38 kDa hFEN1-336 construct (herein referred to as hFEN1) contains 336 native amino acids with an additional six residues present at the C-terminus from the rhinovirus 3C protease recognition site (49). Optimization of both buffer and temperature yielded stable hFEN1 samples that afforded good ^{1}H–^{15}N TROSY spectra (Figure 2A and Supplementary Figure

Downloaded from https://academic.oup.com/nar/advance-article-abstract/doi/10.1093/nar/gky293/4990015
by University of Sheffield user
on 03 May 2018

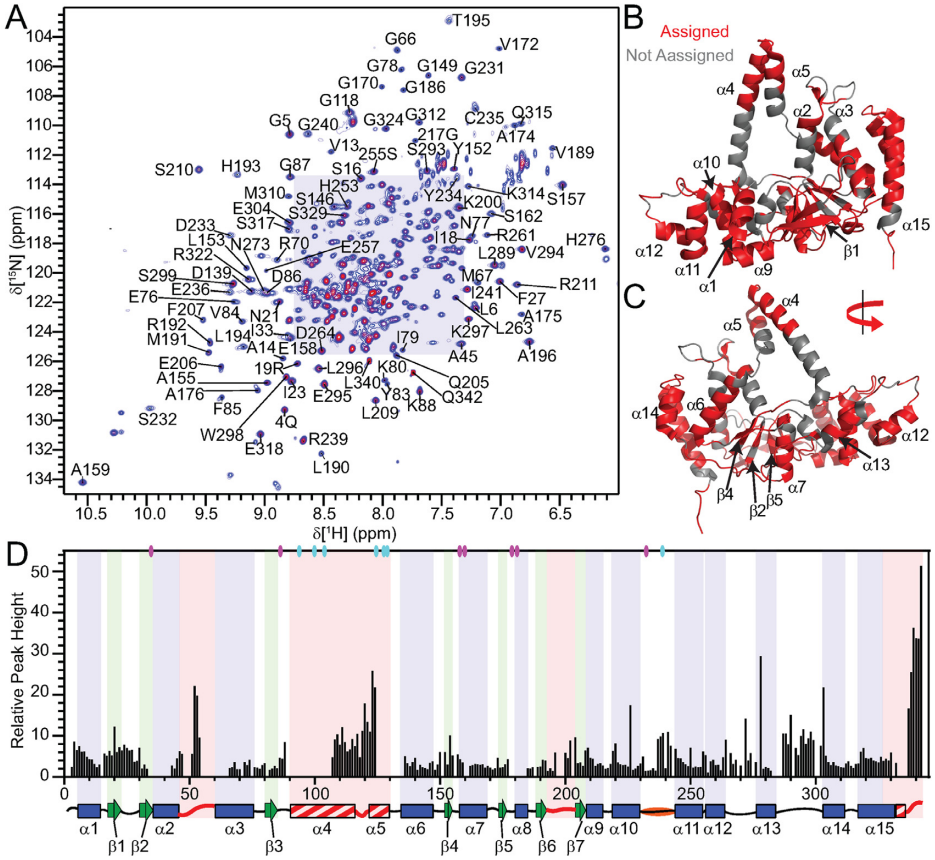

Figure 2.)) Relative peak heights suggest differences in dynamics in regions of hFEN1. (**A**) ^{1}H–^{15}N-TROSY spectrum of hFEN1 under experimental conditions. Peaks from amide side chains and the two tryptophan side-chain indole groups (H^{ε}–N^{ε}) are also observed. Expanded views of the shaded region can be found in Supplementary Figure S4. (**B**) Front and (**C**) rear views of the hFEN1 structure (3Q8K) (11) with labeled secondary structure elements and colored backbone to denote assigned (red) or unassigned (gray) residues. (**D**) Relative peak height (based on the lowest intensity peak E257) obtained from the ^{1}H–^{15}N-TROSY spectrum of hFEN1 plotted versus residue number show that loops are generally more intense and are flanked by decreasing peak intensities and sometimes missing residues. A secondary structure schematic of hFEN1 (3Q8K) is included (11). Blue rectangles, green arrows and black lines indicate α-helices, β-strands and loops, respectively. Loops known to have structural heterogeneity (1UL1 versus 3Q8K) are indicated by red lines (11,22). Red and white stripped rectangles highlight regions of structural heterogeneity. The H2tH α10–α11 loop is highlighted in orange. The location of the active site carboxylate (D34, D86, E158, E160, D179, D181 and D233) and basic residues (K93, R100, R104, K125, K128, R129 and R238) with respect to secondary structure elements are indicated by magenta and cyan ovals, respectively.

S4). Using standard ^{1}H–^{15}N TROSY and TROSY-based 3D experiments (29), 251 amino acid residues out of the 324 theoretically-detectable backbone amides (excluding prolines and initiator methionine) were assigned in the absence of divalent metal ions. Many of the unassigned residues were localized to the base of the arch region, the active site amino acids and some of the α2–α3 loop (Figure 2B and C). Furthermore, the peaks in the ^{1}H–^{15}N TROSY spectrum displayed variable intensities, suggesting that not all residues relaxed uniformly. Some of the interfaces between the assignable and unassignable regions had peaks with decreased intensity, which suggested that the missing regions were undergoing exchange on the millisecond timescale (Figure 2D).

Chemical shift analysis of protein backbone nuclei (^{15}N, $^{1}H^{N}$, $^{13}C^{\alpha}$, $^{13}C^{\beta}$ and $^{13}C'$) was conducted using TALOS-N (32). The predicted secondary structure elements agreed with the structure of the saddle region of the protein in 3Q8K and 1UL1x, thereby supporting the accuracy of our assignments (Supplementary Figure S5). Prediction of β-strand ϕ and ψ dihedral angles for residues E295−S299 were confirmed with further analysis of this region in 3Q8K and 1UL1z (11,22), which showed that the dihedral angles were consistent with the β-strand prediction despite being a loop. Furthermore, a single α-helical turn in the β1–β2 loop (residues D22−Y26) is present consistent with the TALOS-N assignment (11,22). In the arch region, the analysis for A106−A111 suggested that these residues were in an α-helical conformation. For Q112−Q115, α-helix was also predicted, but with decreasing probability. Finally, an extended loop conformation was predicted for residues A116−E124. Therefore, only the N-terminal half of the arch region was consistent with the secondary structure present in 3Q8K, whereas the C-terminal portion was not.

The observable residues of the arch region and α2–α3 loop are disordered

To characterize the dynamics of the free protein, ^{15}N spin-lattice (R_1) and spin-spin (R_2) relaxation rates and ^{15}N-$\{^{1}H\}$ heteroNOE (hNOE) values were measured using

Downloaded from https://academic.oup.com/nar/advance-article-abstract/doi/10.1093/nar/gky293/4990015
by University of Sheffield user
on 03 May 2018

$^{2}H,^{15}N$-labeled hFEN1 (33). Of the 251 residues that were assigned, only 203 and 197 were analyzed at 600 and 800 MHz (Supplementary Figure S6A-F), respectively, due to peaks either being too weak to analyze or overlapped. The data generally showed that residues in secondary structure have similar relaxation rates, indicative of the overall rotational correlation time of the molecule (τ_c). Using R_2/R_1 ratios (Supplementary Figure S6G and H), an overall τ_c was estimated to be 25 ns (50), consistent for a protein of this size (38 kDa) (51). Loops including β6–β7 (i.e. β-pin), α10–α11, α12–α13 and α13–α14 as well as the start of α15 displayed intermediate R_2/R_1 ratios. Regions that showed particularly low R_2/R_1 ratios and low hNOE values were the α2–α3 loop, the arch region and the non-native C-terminal residues. These regions were associated with higher than average *B*-factors or lacked observable electron density in crystal structures (Figure 1D–F and Supplementary Figure S2) (11,22). Furthermore, the higher peak intensities of the α2–α3 loop, the arch region and the non-native C-terminal residues (Figure 2D) are consistent with the lower R_2/R_1 ratios (Supplementary Figure S6G and H).

We used *relax* in combination with the 3Q8K protein structure to calculate model-free order parameters (52). Residues with low R_2/R_1 ratios and a lack of density in the free protein structures were treated separately as an isotropic spherical diffusion tensor (16 residues), while the other backbone amide residues were derived using an oblate rotational diffusion tensor (163 residues) (Supplementary Figure S6I and J). Although a large range of S^2 values were generated with various models (Supplementary Table S2), most residues in the protein in secondary structure elements afforded an average S^2 of ∼0.8, showing that relaxation of most residues in the saddle region of hFEN1 was influenced predominantly by the tumbling of the molecule in solution (Figure 3). Therefore, these residues were relatively rigid. The α10–α11, α12–α13 and α13–α14 loops, the C-terminal end of α10 and α14 as well as most of α15 displayed intermediate S^2 ($0.5 < S^2 < 0.8$) values indicating increased flexibility in these regions. The lowest S^2 values (<0.5) were generally observed in the C-terminal tail, arch and α2–α3 loop residues, thereby indicating that these residues were extremely flexible in solution. R_{ex} terms were reported for approximately 70% of the residues in hFEN1 that were amenable to model-free analysis (Supplementary Figure S7), indicating that millisecond timescale motions contributed to spin-spin relaxation (R_2). Most of the R_{ex} terms were associated with residues in the saddle region.

S^2 values were compared with Random Coil Index-S^2 (RCI-S^2) values predicted by TALOS-N (32) and were congruent apart from an underestimation of the degree of regional flexibility (Figure 3A). Considering the flexibility of the arch, α2–α3 loop and C-terminal tail, we decided to confirm the nature of these regions using the neighbor-corrected random coil chemical shift library for intrinsically disordered proteins (ncIDP) (53). Deviations greater than 1 ppm from ncIDP ^{15}N and ^{13}C chemical shift values were shown to be a powerful indicator of secondary structure. The observable regions of the arch, α2–α3 loop and C-terminal residues showed only small differences (≤1) (Figure 4A and B), suggesting that these residues were disordered in the free protein. The small deviations from the random coil value suggested that residues A107−Q113 were transiently sampling α-helical ϕ,ψ -space. The postulated preference for occupying α-helical-ϕ,ψ space showed a gradual decrease occurring from Q110−A114. For residues A114−A120, a progressive switch to extended ϕ,ψ -space was implied (53). These data agreed with AGADIR predictions (54) and TALOS-N analysis (32) (Figure 4C and Supplementary Figure S5A). Overall, these data suggested that the middle of α4 was poised to form α-helical structure, whereas the top of α5 was not.

The analysis of the sequence of the arch residues (P90−L130) using CIDER (55,56) (Classification of Intrinsically Disordered Ensemble Regions) categorized this IDR as a polyampholytic coil or hairpin. Whether a polyampholytic IDR prefers to be a hairpin collapsed on itself or an extended coil (Supplementary Figure S3C,D,F) was shown to depend on the linear sequence distribution of charged residues, which can be assessed by the value κ (57). A κ value of 0 indicates no self-assembly, whereas a κ value of 1 indicates full collapse. The arch region has a very low κ value of 0.12, which suggested that it prefers to adopt an extended coil rather than collapsed conformation in solution.

Divalent cations induce chemical shift changes in the vicinity of the hFEN1 active site

Because hFEN1 binds two Mg^{2+} metal ions in the active site to catalyze the hydrolysis of the scissile phosphate diester bond of the optimal substrate (1,2), we wanted to assess the effects of Mg^{2+} and Ca^{2+} metal ion coordination by hFEN1 using chemical shift perturbation mapping. Ca^{2+} was included in the analysis for studies of an hFEN1–DNA complex, because it facilitates the hFEN1-induced DNA conformational change required for scissile phosphate diester placement on the active site metal ions without substrate hydrolysis (9,58). Furthermore, the analysis was performed using $hFEN1_{K93A}$ because addition of Ca^{2+} alone was unable to completely inhibit phosphate diester hydrolysis, presumably due to small amounts of Mg^{2+} contamination (59). Removal of K93 by mutation to alanine reduces scissile phosphate diester hydrolysis by 2000-fold, because it acts as an electrophilic catalyst during scissile phosphate diester hydrolysis (42). Nonetheless, $hFEN1_{K93A}$ still facilitates substrate conformational change similarly to hFEN1 (9,58). The K93A mutation is in a region of hFEN1 that was unable to be assigned. Comparison of the $^{1}H–^{15}N$ TROSY spectra of hFEN1 and $hFEN1_{K93A}$ showed only small and localized changes in chemical shifts, thereby indicating that it had no widespread effect on the protein (Figure 5A).

Titration of either $MgCl_2$ or $CaCl_2$ (0–10 mM) into samples of hFEN1 was monitored using $^{1}H–^{15}N$ TROSY spectra and showed saturation between 8–10 mM for both divalent salts, consistent with the optimal concentrations of $MgCl_2$ used in routine kinetic assays. Furthermore, $^{1}H–^{15}N$ TROSY spectra of hFEN1 with 8 mM Ca^{2+} or 8 mM Mg^{2+} are nearly identical, but show small differences in residues near to the active site including the N-terminal residues (Supplementary Figure S8A) This indicated that Ca^{2+} binds in the active site, consistent with Ca^{2+} being competitive with respect to Mg^{2+} (60). The small deviations in chemical

Downloaded from https://academic.oup.com/nar/advance-article-abstract/doi/10.1093/nar/gky293/4990015
by University of Sheffield user
on 03 May 2018

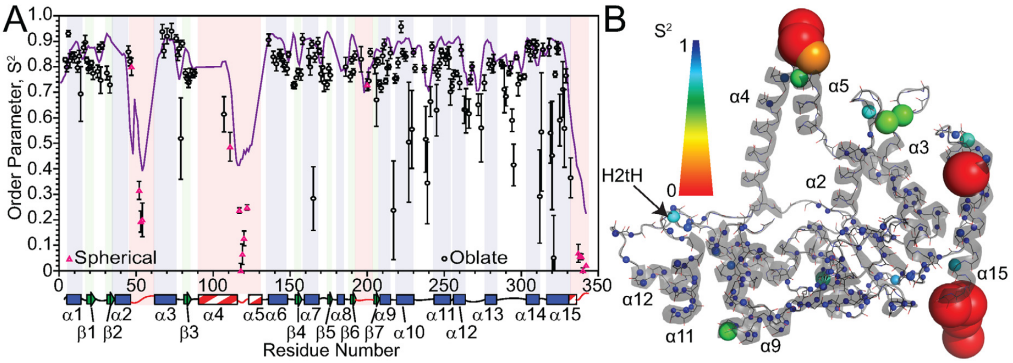

Figure 3. Model-free analysis of hFEN1 relaxation data identifies regions with increased flexibility. (**A**) Generalized order parameters (S^2) were derived from *relax* (52) using backbone ^{15}N relaxation data acquired at two field strengths and plotted versus residue number. Black circles represent data fitted to the oblate spheroid diffusion tensor, whereas pink triangles were fitted to a spherical diffusion tensor. The purple line in (**A**) represents the TALOS-N (32) random coil index S^2 prediction. Secondary structure map as described in Figure 2D. (**B**) S^2 values plotted on a cartoon depiction of the hFEN1 protein structure (3Q8K) (11). The spheres represent the nitrogen nuclei for which data were derived. The S^2 and R_{ex} spectrum bars illustrate the magnitude of S^2 and R_{ex} values with respect to sphere color and size.

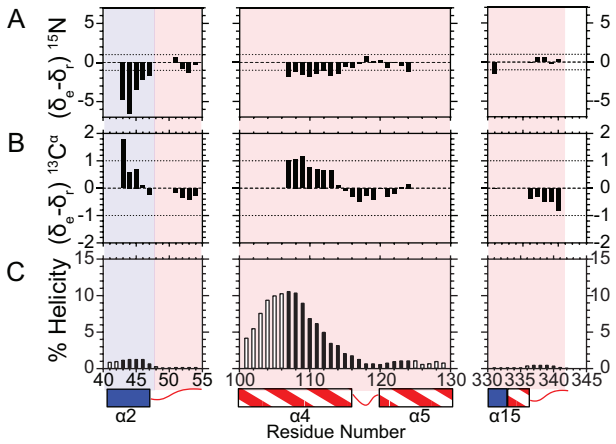

Figure 4. $^{15}N^H$ and $^{13}C^\alpha$ chemical shifts for the disordered arch region indicate that only some of the residues are sampling helical conformations. Bar graphs indicate the magnitude of the difference of experimental (δ_e) and sequence-corrected random coil chemical shifts (δ_r) for (**A**) $^{15}N^H$ and (**B**) $^{13}C^\alpha$ for the indicated regions. Differences less than one (dotted line) suggest disorder. The anti-correlated nature of $^{15}N^H$ and $^{13}C^\alpha$ is used to suggest which segments of the disordered region populate either helical or extended peptide backbone angles. Positive and negative $^{15}N^H$ and $^{13}C^\alpha$ chemical shift differences, respectively, imply extended backbone angles, whereas negative and positive $^{15}N^H$ and $^{13}C^\alpha$ chemical shift differences, respectively, suggest sampling α-helical backbone angles (53). (**C**) Agadir helical percentage predictions for the indicated residues assessed at 114 mM ionic strength, pH 7.5 and 25°C (NMR conditions) (54). Black bars indicate the residues for which NMR data is available to validate the prediction, whereas open bars indicate an absence of NMR assignments. A secondary structure map is illustrated below as described in Figure 2D.

shifts of residues closest to the active site could be due to the larger ionic radius and looser coordination geometry of Ca^{2+} compared to Mg^{2+}. Comparison of ^{1}H-^{15}N TROSY spectra of hFEN1$_{K93A}$ in the absence and presence of 8 mM Ca^{2+} showed similar chemical shift perturbations as observed in spectra of hFEN1 when Mg^{2+} or Ca^{2+} were added. The largest changes between apo-hFEN1$_{K93A}$ and Ca^{2+}-hFEN1$_{K93A}$ occurred in regions close to the metal chelating carboxylates (D34, D86, E158, E160, D179, E181, D233) and the N-terminal residues (Figure 5B–D). Other regions affected included β5, α3, α7, β4–α7 loop, the H2tH motif, the β3–α4 loop and the β6-β7 β-pin, which are all regions that are near active site residues. Some of the β-strands also showed smaller perturbations further from the active site that suggested that conformational rearrangement in the active site is propagated throughout the β-sheet. These data indicated that divalent ions predominantly affected the structure immediately surrounding the active site in agreement with their locations in X-ray crystal structures (11). Notably, the addition of Ca^{2+} did not change the higher relative intensities of the peaks associated with the arch region and α2–α3 loop, suggesting that Ca^{2+} did not significantly alter the dynamics of these regions.

Substrate binding changes the exchange regime for the arch region and α2–α3 loop

To prepare an hFEN1$_{K93A}$–DNA complex and to analyze NMR spectral changes upon binding, consecutive sub-stoichiometric amounts of DNA substrate (DHPS1; Table 1 and Supplementary Figure S9A) were added to an hFEN1$_{K93A}$ sample containing with 76 mM KCl until the DNA was in excess. However, addition of substrate induced protein precipitation in the sample suggesting non-specific DNA binding at concentrations used for NMR spectroscopy. Despite this, chemical shift changes upon addition of DNA were observed for a number of residues, consistent with locations known to interact with duplex DNA. Residues that were affected by DNA binding, but remained in the fast exchange regime (e.g. Q4, A175, L190, R192, R239 and G324) could be followed with the addition of DNA to saturation. Interestingly, the assignable α2–α3 loop and arch residues disappeared from the spectrum at sub-stoichiometric amounts of DNA, suggesting that they entered an exchange regime that prevented their detection (29). The magnitude of the exchange rate at which this happens at these field strengths should be faster than 10 s^{-1} but slower than 1000 s^{-1} as the peaks do not re-appear upon further addition of DNA to saturation. The switch from sub-nanosecond dynamics to millisecond exchange

Downloaded from https://academic.oup.com/nar/advance-article-abstract/doi/10.1093/nar/gky293/4990015
by University of Sheffield user
on 03 May 2018

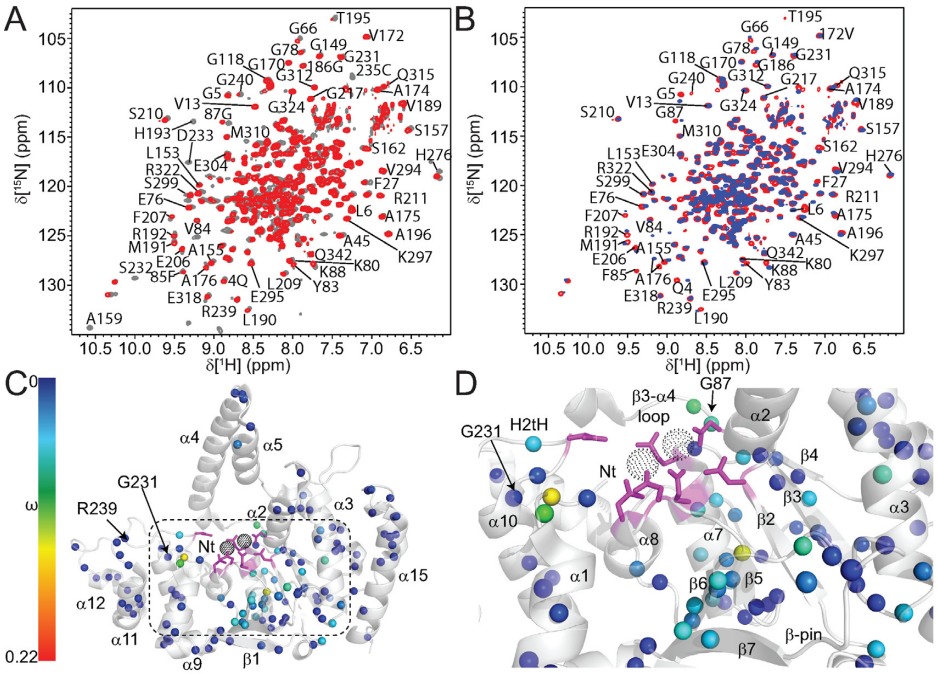

Figure 5. Addition of Ca^{2+} ions to $hFEN1_{K93A}$ perturbs chemical shifts nearest the active site. (**A**) Superposed ^{1}H–^{15}N TROSY spectra of hFEN1 (gray) and $hFEN1_{K93A}$ (red) shows only minimal changes in localized regions close to the mutation site indicating no widespread effects on global structure. (**B**) Superposed ^{1}H-^{15}N TROSY spectra of $hFEN1_{K93A}$ in the absence (red) and presence (blue) of Ca^{2+}. Well-dispersed residues are labeled accordingly. (**C**) Chemical shift changes observed on the addition of Ca^{2+} ions to $hFEN1_{K93A}$ are mapped onto a cartoon representation of the hFEN1 structure (3Q8K) (11). The black-dotted spheres indicate the locations of active site metal ions that are coordinated by the side chains of carboxylate residues shown as magenta sticks. Secondary structure elements, the H2tH motif and the N-terminus (Nt) are labeled. The locations of the G231 and R239 spheres are highlighted. The magnitudes of the nitrogen nuclei chemical shift perturbations (ω) are represented by sphere color according to the associated spectrum bar. The absence of a sphere either indicates the lack of assignment in the protein or the inability to follow chemical shift changes due to peak overlap. (**D**) Magnified view of the area indicated by the dashed box in panel C to highlight the location of residues most affected by the addition of Ca^{2+} ions to $hFEN1_{K93A}$. Labels are as described in panel C except for the omission of R239 and addition of G87.

suggested a conformational change of the α2–α3 loop and the arch upon DNA binding.

Divalent metal ions induce large chemical shift changes far from the active site

To overcome issues associated with the precipitation of $hFEN1_{K93A}$ when preparing an $hFEN1_{K93A}$–DNA complex (Supplementary Figure S9A), a different strategy was adopted where the complex was prepared under dilute conditions and at lower KCl concentrations in the presence of EDTA, followed by concentration. Preparing $hFEN1_{K93A}$-DNA complexes in this manner resulted in stable samples that afforded acceptable ^{1}H–^{15}N TROSY spectra (Supplementary Figure S8B). Comparison of ^{1}H–^{15}N TROSY spectra of $hFEN1_{K93A}$ and the $hFEN1_{K93A}$-DNA complex showed chemical shift changes for residues in or near areas known to interact with the substrate, such as the H2tH motif (e.g. G231 and R239), active site (e.g. Q4, G5, L6, G87, K88, A175 and R192), the hydrophobic wedge (e.g. A45 and G66,) and the 3′-flap binding pocket (e.g. M310, E318 and G324). This indicated that the solution complex was consistent with the hFEN1–DNA complex crystal structures (3Q8K and 5UM9) (11,21). Before assigning the $hFEN1_{K93A}$–DNA complex, Ca^{2+} was titrated into the sample (0–8 mM) and chemical shift perturbations upon each successive addition were monitored by ^{1}H-^{15}N TROSY spectra. Assignment of the $hFEN1_{K93A}$-DNA complex in 8 mM Ca^{2+} was conducted using the same methodology as for hFEN1; however, only 57% of the visible peaks (132 peaks out 230 peaks in the spectrum) could be assigned presumably due to signal attenuation in the 3D spectra. Some assignments were corroborated by following the DNA titration data (i.e. R19, G149, A175, S255, R261).

Addition of Ca^{2+} to $hFEN1_{K93A}$ and the $hFEN1_{K93A}$-DNA complex showed that residues immediately surrounding the active site were perturbed similarly (Figure 5B–D *cf.* Figure 6A and B), indicating that Ca^{2+} is coordinated in the active site in both free and bound forms. In the $hFEN1_{K93A}$–DNA complex, however, the presence of Ca^{2+} additionally resulted in the disappearance of the N-terminal residues Q4, G5 and L6 (Supplementary Figure S8C). For example, the first addition of Ca^{2+} to the complex resulted in an upfield chemical shift with peak broadening for Q4 (Figure 6C). These N-terminal residues remained observable in spectra when Ca^{2+} was added to $hFEN1_{K93A}$ (Figure 5B); thus, this unique effect was due to the presence of the DNA. Together this suggested that the addition of Ca^{2+} to the complex resulted in the N-terminal residues experiencing changes in their chemical environment due to movement of the residues themselves or movement of the DNA into the active site, thereby resulting in intermediate exchange broadening.

Downloaded from https://academic.oup.com/nar/advance-article-abstract/doi/10.1093/nar/gky293/4990015
by University of Sheffield user
on 03 May 2018

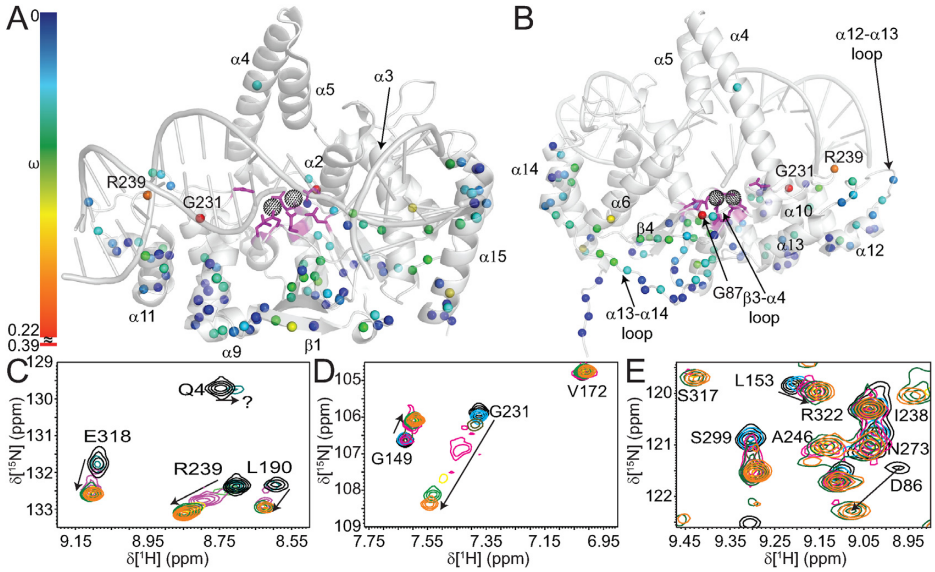

Figure 6. Addition of Ca^{2+} ions to the $hFEN1_{K93A}$–DNA complex show chemical shift perturbations (ω) throughout the protein. (**A**) Front and (**B**) rear views of a cartoon depiction of the hFEN1–DNA structure (3Q8K) (11) with the magnitudes of nitrogen nuclei chemical shift perturbations (ω) on addition of Ca^{2+} represented by sphere color according to the associated spectrum bar. Note, because the chemical shift change for G231 was larger than for most other residues, the spectrum bar is discontinuous. (C–E) Highlighted regions from Supplementary Figure S8C showing several residues with significant chemical shift perturbations that occur upon titration with Ca^{2+} at 0 mM (black), 0.5 mM (cyan), 1 mM (olive), 2 mM (magenta), 4 mM (yellow), 6 mM (green) and 8 mM (orange). The panels show a large ω for (**C**) L190, R239 and E318, (**D**) G149 and G231 and (**E**) the appearance of S317 at higher Ca^{2+} concentrations. Intermediate exchange behavior is highlighted for (C) Q4 and (E) D86.

Unlike $hFEN1_{K93A}$, the addition of Ca^{2+} to the $hFEN1_{K93A}$–DNA complex resulted in larger perturbations in regions significantly removed from the active site (Figure 5C and D *cf.* Figure 6A and B). The most pronounced weighted ^{1}H–^{15}N chemical shift change (ω) was observed for G231, which lies at the beginning of the H2tH motif. In the absence of DNA, this residue showed a slight change in chemical shift upon addition of Ca^{2+} ions (ω = 0.022 ppm, Figure 5C), whereas the change increased greatly (ω = 0.39 ppm, Figure 6D) when DNA was present. Another H2tH motif residue whose Ca^{2+} induced chemical shift change was significantly larger for the $hFEN1_{K93A}$–DNA complex was R239 (ω = 0.013 ppm, Figure 5B *cf.* ω = 0.17 ppm, Figure 6C), which lines the minor groove of the dsDNA in the reacting duplex. This showed that the H2tH DNA binding motif and residues in that region, which are approximately 20 Å from the active site metal ions, responded to the metal occupancy status of the active site in the presence of DNA. Residue G87, which is next to the divalent metal binding residue D86, also showed a larger chemical shift perturbation in response to Ca^{2+} in the $hFEN1_{K93A}$-DNA complex (ω = 0.22 ppm, Figure 6B and Supplementary Figure S8C) compared to $hFEN1_{K93A}$ (ω = 0.07 ppm, Figure 5B and Supplementary Figure S8B). In the threaded hFEN1–DNA substrate crystal structure (5UM9), residue G87, which is in the β3–α4 loop, is near the 5′-flap exit, and this larger perturbation was likely due to its proximity to the exit of the 5′-flap (Figure 1C) (21). Moreover, chemical shift changes in this region are consistent with the reported changes in the 5′-flap threading equilibrium when Ca^{2+} was added to hFEN1-DNA complexes (18). Furthermore, addition of Ca^{2+} to the $hFEN1_{K93A}$–DNA complex also gave rise to larger chemical shift perturbations than in $hFEN1_{K93A}$ in the β3–α4, α12–α13 and α13–α14 loops and helices α6, α14 and α15 (Figure 6A and B). These regions were proximal to the 5′-flap exit site and close to the 3′-flap binding pocket.

In contrast to the N-terminal residues, S317 of the $hFEN1_{K93A}$–DNA complex reappeared near the end of the Ca^{2+} titration (Figure 6E). In the hFEN1–DNA product and substrate complexes (3Q8K and 5UM9) (11,21), the S317 backbone amide group hydrogen bonds with one of the non-bridging phosphate oxygen atoms of the 3′-flap phosphate diester moiety (Supplementary Figure S1A). In ^{1}H–^{15}N TROSY spectra of $hFEN1_{K93A}$, S317 was observable in the presence and absence of Ca^{2+}; however, the peak was missing from spectra of the $hFEN1_{K93A}$-DNA complex in the absence of Ca^{2+} (Supplementary Figure S8B). The weighted chemical shift perturbation (ω) for S317 in $hFEN1_{K93A}$ and the $hFEN1_{K93A}$–DNA complex in the presence of Ca^{2+} was 0.68 ppm downfield, consistent with it forming a hydrogen bond with a non-bridging phosphate oxygen atom in the DNA. Regardless of the presence or absence of Ca^{2+}, residues associated with the arch region and α2–α3 loop were neither observable nor assignable in spectra of the $hFEN1_{K93A}$–DNA complex

Single molecule FRET measurements confirm conformational changes in the arch region

To investigate further the range of conformations adopted by FEN1 in the presence and absence of substrate DNA (Supplementary Figure S9B without the 5'-FAM label), we used single-molecule Forster Resonance Energy transfer (smFRET). Modeling of dye attachment sites for donor

Downloaded from https://academic.oup.com/nar/advance-article-abstract/doi/10.1093/nar/gky293/4990015
by University of Sheffield user
on 03 May 2018

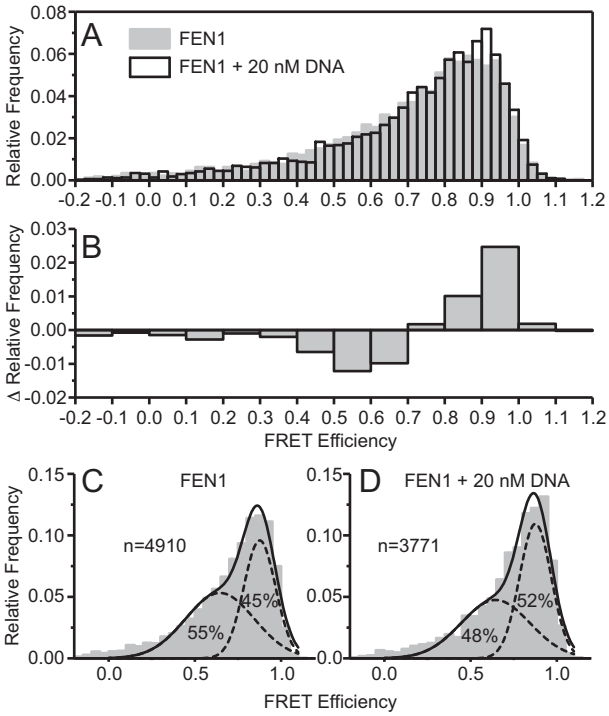

Figure 7. smFRET shows a change in conformational ensemble upon addition of DNA and reveals millisecond conformational dynamics occurring in the arch region. (A) Relative frequencies of FRET efficiencies for $hFEN1_{QF}$ alone (gray) and in complex with 20 nM DNA (white). (**B**) The difference in relative frequencies of FRET efficiencies upon addition of 20nM DNA. (**C**) $hFEN1_{QF}$ alone and (**D**) $hFEN1_{QF}$ with 20 nM DNA showing the unrestrained fit to the sum of two Gaussian functions (see Supplementary Table S4 for fitting parameters) to show that upon addition of DNA the lower FRET population decreases and the higher FRET population increases.

(Cy3B) and acceptor (Atto647N) dyes identified sites in the arch (E120) and the saddle region (S293) opposite the DNA binding site that would produce a FRET pair sensitive to the conformation and position of the arch (Supplementary Figure S10A). However, hFEN1 has two partially solvent accessible cysteines (C235 and C311) that easily label with maleimide dyes (data not shown). Thus, a quadruple mutant was generated by successive site-directed mutagenesis (E120C/C235A/S293C/C311A; $hFEN1_Q$) (41), and stochastic incorporation of the maleimide dyes was achieved at the desired postions to generate $hFEN1_{QF}$ (Supplementary Table S3) (61). Measurement of $hFEN1_Q$ and $hFEN1_{QF}$ activities showed that the mutations and attachment of the dyes had only a small effect on enzymatic activity (Supplementary Figure S10B). In the absence of DNA but with Ca^{2+}, a broad range of FRET efficiencies were observed (Figure 7A), indicative of multiple conformations and positions of the arch. In these experiments, single donor–acceptor labeled FEN1 molecules are observed as they diffuse through the confocal volume, a process which takes ∼ 1 ms. Conformations that interconvert much faster than the observation time yield a single, tight FRET efficiency signal, whereas a broad profile is indicative of multiple conformations converting on slower (> 1 ms) time scales, as was observed here (Figure 7A).

On addition of 20 nM substrate DNA, subtle changes in the FRET efficiency distribution were observed (Figure 7A); a decrease in the relative frequency of FRET efficiencies below 0.7 with a concomitant increase in FRET efficiencies above 0.7 (Figure 7B). This is indicative of a change in the conformational ensemble for FEN1 upon binding the DNA substrate. This change in the conformational ensemble could explain the unusual viscogen dependence of hFEN1 reaction (Supplementary Figure S11). The FRET efficiency distributions were well described by the sum of two gaussians, with free fits yielding the same means for the high and low FRET populations in both the presence and absence of the DNA substrate (Supplementary Table S4). On addition of DNA, a shift in the relative amplitude from the low-FRET to the high-FRET ensemble was observed (Figure 7C and D, Supplementary Table S4), showing a change in the conformational ensemble of the arch with respect to the saddle upon binding DNA. These data indicated that a change in the conformational ensemble in the arch occurred when substrate was added.

DISCUSSION

Earlier studies have suggested that hFEN1-specificity for DNA structure rather than sequence arises from the recognition of three substrate structural features: (i) the bending of the two-way dsDNA junction, (ii) 5′-flap threading and (iii) 3′-flap binding (8–12,21). The latter two substrate-feature selection steps involve regions of the protein that we show are conformationally dynamic (i.e., the arch region and the α2–α3 loop). NMR spin relaxation data for the free protein shows that residues in the top of the arch (A107, L111, A117–E120 and E122) and the middle of the α2–α3 loop (V52–Q54) are very flexible, exhibiting motions on the timescale of $\sim 10^9$ s^{-1}; hence, they are disordered. However, residues either side of these flexible regions cannot be detected in NMR experiments presumably due to intermediate exchange broadening caused by millisecond timescale motions. Consistent with this, smFRET data for the free protein suggests that the arch region is moving relative to the saddle region of hFEN1 at a rate slower than a 1000 s^{-1} (Figure 8). The highly mobile residues in the arch and α2–α3 loop that are observable by NMR may experience this millisecond motion as well, but because these regions likely experience negligible chemical shift changes, the subnanosecond timescale motions dominate the relaxation.

The observable residues in arch and α2–α3 loop undergo a change of motional timescale from $\sim 10^9$ s^{-1} in the free protein to an intermediate exchange regime (10–1000 s^{-1}) in the hFEN1–DNA complex, suggesting their chemical environment is perturbed more significantly when bound to DNA. Such perturbation can arise from changes in rates of exchange, secondary structure formation, and changes in chemical environment due to proximity to the DNA. In line with this, the smFRET data indicate a change in the conformational ensemble upon addition of DNA. Interestingly, both the smFRET and NMR data suggest that there are still millisecond time motions involving the arch and α2–α3 loop. Thus, the earlier disorder to order model upon correct substrate binding that was based on crystallographic observations (2,11,21,62,63) is an oversimplification as it suggests

Downloaded from https://academic.oup.com/nar/advance-article-abstract/doi/10.1093/nar/gky293/4990015
by University of Sheffield user
on 03 May 2018

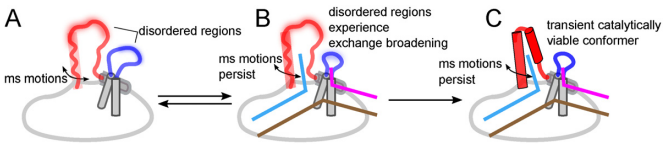

Figure 8. The flexible regions of hFEN1 respond to the appropriate structural features of the substrate, resulting in a shift to the catalytically viable ensemble. (**A**) In the absence of substrate, hFEN1 contains flexible regions in the 'arch' (red) and the α2–α3 loop (blue). The tops of these regions are disordered and display very fast motions ($\sim 10^9$ s^{-1}), whereas the flanking regions are absent presumably due to intermediate exchange broadening. The observable regions of the arch and loop are likely in fast exchange because they experience little to no change in chemical environment from the millisecond timescale motion. (**B**) Upon binding DNA, the disordered regions experience exchange broadening, likely due to a change in chemical environment and/or change in rate of motions. However, the slow millisecond timescale movements persist in both the arch and the α2–α3 loop despite the change in the conformational ensemble. (**C**) Eventually, the catalytically viable ensemble forms possibly coupling motions within the arch, the α2–α3 loop and the DNA.

there are no motions in the hFEN1–DNA complex. Rather, the complex is likely better described as an ensemble of conformations. The smFRET data shows that addition of DNA shifts the equilibrium bias of the arch to higher FRET efficiency states, implying the net position of the arch moves closer to the saddle. Interestingly, the arch has been observed in different positions relative to the saddle in hFEN1 (11,21) and hEXO1 X-ray structures (20,64). Notably, in structures where DNA has untwisted and moved into the active site (21), the arch has moved forward toward the saddle. It is possible that arch motions play an important role in the catalytic cycle, by facilitating untwisting of the DNA to transfer the scissile phosphate to the active site. Indeed, the single turnover rates of the hFEN1-catalyzed reaction are 15–30 s^{-1} (12); thus, it is possible that the slower conformational changes in the arch and α2–α3 loop regions could limit the rate of reaction.

In the hFEN1–DNA complex, residues near the active site are perturbed upon addition of Ca^{2+} as is also seen with substrate-free hFEN1. Interestingly, the N-terminal residues disappear from the spectra upon addition of Ca^{2+} to hFEN1–DNA, but not with hFEN1 alone. The recent crystal structures of the hFEN1–DNA complex (5UM9) suggests that the N-terminal amine moiety of G2 in conjunction with a metal ion and D233 activates the attacking water to generate the hydroxide ion nucleophile (21). Hence, the dynamics resulting in the disappearance of the N-terminal residues upon the addition of Ca^{2+} may be associated with this role. We also observed large chemical shift changes in response to addition of Ca^{2+} to the hFEN1–DNA complex in regions of the protein distal to the active site that were not observed at the same magnitude upon addition of Ca^{2+} to hFEN1. Several studies have shown that divalent metal ions and arch region ordering are necessary to move the scissile phosphate into the active site by untwisting (65) the reacting duplex (9,18,58,59). Thus, these large Ca^{2+} induced chemical shift changes in the H2tH motif, which directly contacts the reacting duplex 20 Å from the active site metal ions, are likely due the untwisting of the reacting duplex to move the scissile phosphate of into the active site. Remarkably, chemical shift perturbations are also propagated to the region around the 3′-flap binding pocket. The reappearance of S317 and other chemical shift changes in the 3′-flap binding region strongly suggests that the positioning of the 3′-flap duplex is altered upon the addition of Ca^{2+}. Therefore, our study provides direct evidence of DNA-mediated allostery (66) between the active site and both reacting duplex and 3′-flap recognition, thereby linking regions involved in DNA structural recognition and catalysis.

When the 5′-flap is threaded through the helical arch, the top of α4 and α5 form a mini-hydrophobic core (21). The importance of this ordered state has previously been demonstrated; it is required for untwisting of the reacting duplex DNA to place the scissile phosphate on the active site while delivering two basic residues to the active site that are essential for biologically-relevant rates of reaction (11,18,59). Chemical shift analysis (53) of the NMR-observable region corresponding to the top of α4 (A107−A116) suggests that it is transiently sampling an α-helical conformation in its disordered state, whereas the corresponding region of α5 is not. This suggests that the disorder-to-order transition for α4 arises due to conformational selection (i.e., population shift), whereas the transition in α5 results from induced fit (67–69). Having a completely induced fit model for the arch would allow for enhanced specificity but would involve large entropic penalties and slow ordering. Using a conformational ensemble shift for the larger portion of the arch (α4) reduces these penalties. A potential trigger for the ensemble population shift for the α4 region could be the catalytically-crucial contact to the +1 phosphate of the 5′-flap as well as the formation of the mini-hydrophobic core (21). On the other side of the arch, several factors are likely to induce the disorder to α-helical transition of α5. These include interaction with the α2–α3 loop when it forms an Ω-loop (15) in response to the 3′-flap and electrostatic interaction of K125, K128 and R129 with template DNA phosphate diesters (11,21). Interestingly, the distance these lysine residues are from the template phosphate diesters is shorter when the arch moves forward towards the saddle (21).

Because hFEN1 requires a disordered region that can fold in response to the appropriate ligand, the evolutionary conservation of disorder and order promoting residues to maintain the delicate balance is necessary (55,70). CIDER (56) analyses of the arch suggest that substrate-free hFEN1 has evolved to populate an extended rather than a collapsed conformation. The extended conformation is necessary to allow DNA to thread. Seemingly innocuous mutations or even interaction with small molecules may alter the conformation of the disordered state (extended coil vs. collapsed) or change the balance between disorder and order, thereby inhibiting catalysis (55). Thus, the arch region and α2–α3 loop, which are not conserved with other superfamily members, could be potential drug targets for hFEN1 (71).

Why hFEN1 has evolved to have an intrinsically disordered region and loop to achieve catalysis may seem counterintuitive initially, because enzymology has traditionally focused on structure-function paradigms (70,71). However, there is a rationale for hFEN1 evolution of intrinsically disordered regions. Firstly, because hFEN1 needs to thread

Downloaded from https://academic.oup.com/nar/advance-article-abstract/doi/10.1093/nar/gky293/4990015
by University of Sheffield user
on 03 May 2018

a 5′-flap through the helical arch, doing so when the arch is disordered would be easier due a larger capture radius and would provide the ability to accommodate 5′-flaps with some secondary structure. Secondly, the combination of induced fit and conformational selection with subsequent ensemble shifts for the arch allows fast ordering with low energy barriers in response to the appropriate substrate features necessary for specificity. However, there is also fast unfolding and release of incorrect binding partners (72). Thirdly, it allows for low affinity for the catalytically-relevant ensemble, which arises when the enthalpic gains in ordering are balanced by the entropic cost (70,71). Hence, specificity can be achieved while allowing the enzyme to disassociate once catalysis has occurred (i.e. turnover). Together, this work illustrates how conformational dynamic regions respond to appropriate substrate features, resulting in a shift in the ensemble towards the catalytically viable state (Figure 8).

Although all 5′-nuclease superfamily members recognize DNA junctions of some kind, each protein has its own specific DNA substrate (3). However, the use of flexible regions spanning a range of dynamic timescales for substrate recognition and control of active site assembly is likely to be a common theme in these structure-sensing nucleases. As hEXO1 also uses the threading mechanism to enforce substrate free 5′-ends (19,20) and requires a helical arch for active site assembly, it too is likely to use changes in protein dynamics for substrate accommodation and catalysis. In contrast, the Holiday junction resolvase GEN1 lacks the arch feature. A much shorter peptide linker that is the equivalent in GEN1 is not observable in current structures (23,24), despite containing superfamily conserved active site residues. This is the predicted dimerization interface, as two GEN1 monomers come together to sense the four-way-junction. Similarly, the very large equivalent of the arch in XPG, which also contains active site residues, presumably assembles as a result of the binding of other nucleotide excision repair protein partners and DNA (25). Thus, somewhat surprisingly, protein reversible plasticity appears to be the answer to coupling recognition with reaction for a range of nucleic acid hydrolases.

DATA AVAILABILITY

The backbone ^{1}H, ^{15}N and ^{13}C chemical shifts of the free protein in EDTA and the $hFEN1_{K93A}$–DNA complex in Ca^{2+} have been deposited in the BioMagResBank (http://www.bmrb.wisc.edu/) under the BMRB accession codes 27160 and 27404, respectively.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The author wish to acknowledge Dr Susan Tsutakawa for assistance with the graphical abstract.
Author Contributions: I.A.B. and L.D.F. prepared and purified isotopically-enriched protein samples and their DNA complexes. I.A.B., L.D.F., A.M.H. and N.J.B. conducted NMR spectroscopic measurements for assignment purposes, relaxation measurements and divalent metal titrations. I.A.B., L.D.F. and A.M.H. assigned the protein. I.A.B., J.C.E. and N.N.B.Md.S. performed enzyme kinetic measurements. I.A.B., L.D.F. and M.J.T. performed site-directed mutagenesis, protein purification and fluorophore maleimide attachment to hFEN1 for single molecule FRET measurements. I.A.B., B.A. and T.D.C. performed single molecule measurements. I.A.B., L.D.F., N.J.B., B.A., T.D.C., J.P.W. and J.A.G. contributed to the interpretation of the results and preparation of the manuscript. L.D.F., J.P.W. and J.A.G. conceived and coordinated the project. I.A.B., L.D.F., B.A., T.D.C. and J.A.G. prepared the figures and wrote the manuscript.

FUNDING

Biotechnology and Biological Sciences Research Council [BB/K009079/1 and BB/M00404X/1 to J.A.G.]; Biotechnology and Biological Sciences Research Council White Rose Doctoral Training Program in Mechanistic Biology Studentship [BB/J014443/1 to I.A.B.]. B.A. thanks the Engineering and Physical Sciences Research Council for his PhD stipend. N.N.B.Md.S. thanks the Public Service Department of Malaysia, Human Capitol Development Division for her PhD stipend. Funding for open access charge: The University Publication Fund of The University of Sheffield from RCUK.
Conflict of interest statement. None declared.

REFERENCES

1. Balakrishnan,L. and Bambara,R.A. (2013) Flap endonuclease 1. *Annu. Rev. Biochem.*, **82**, 119–138.
2. Finger,L.D., Atack,J.M., Tsutakawa,S., Classen,S., Tainer,J., Grasby,J. and Shen,B. (2012) In: MacNeill,S (ed). *The Eukaryotic Replisome: A Guide to Protein Structure and Function*. Springer, Dordrecht, pp. 301–326.
3. Grasby,J.A., Finger,L.D., Tsutakawa,S.E., Atack,J.M. and Tainer,J.A. (2012) Unpairing and gating: sequence-independent substrate recognition by FEN superfamily nucleases. *Trends Biochem. Sci.*, **37**, 74–84.
4. Zheng,L., Jia,J., Finger,L.D., Guo,Z., Zer,C. and Shen,B. (2011) Functional regulation of FEN1 nuclease and its link to cancer. *Nucleic Acids Res.*, **39**, 781–794.
5. van Pel,D.M., Barrett,I.J., Shimizu,Y., Sajesh,B.V., Guppy,B.J., Pfeifer,T., McManus,K.J. and Hieter,P. (2013) An evolutionarily conserved synthetic lethal interaction network identifies FEN1 as a broad-spectrum target for anticancer therapeutic development. *PLoS Genet.*, **9**, e1003254.
6. Ward,T.A., McHugh,P.J. and Durant,S.T. (2017) Small molecule inhibitors uncover synthetic genetic interactions of human flap endonuclease 1 (FEN1) with DNA damage response genes. *PLoS ONE*, **12**, e0179278.
7. Exell,J.C., Thompson,M.J., Finger,L.D., Shaw,S.J., Debreczeni,J., Ward,T.A., McWhirter,C., Sioberg,C.L., Martinez Molina,D., Abbott,W.M. *et al.* (2016) Cellularly active N-hydroxyurea FEN1 inhibitors block substrate entry to the active site. *Nat. Chem. Biol.*, **12**, 815–821.
8. Rashid,F., Harris,P.D., Zaher,M.S., Sobhy,M.A., Joudeh,L.I., Yan,C., Piwonski,H., Tsutakawa,S.E., Ivanov,I., Tainer,J.A. *et al.* (2017) Single-molecule FRET unveils induced-fit mechanism for substrate selectivity in flap endonuclease 1. *eLife*, **6**, e21884.
9. Algasaier,S.I., Exell,J.C., Bennet,I.A., Thompson,M.J., Gotham,V.J.B., Shaw,S.J., Craggs,T.D., Finger,L.D. and Grasby,J.A. (2016) DNA and protein requirements for substrate conformational changes necessary for human flap endonuclease-1-catalyzed reaction. *J. Biol. Chem.*, **291**, 8258–8268.

Downloaded from https://academic.oup.com/nar/advance-article-abstract/doi/10.1093/nar/gky293/4990015
by University of Sheffield user
on 03 May 2018

10. Chapados,B.R., Hosfield,D.J., Han,S., Qiu,J., Yelent,B., Shen,B. and Tainer,J.A. (2004) Structural basis for FEN-1 substrate specificity and PCNA-mediated activation in DNA replication and repair. *Cell*, **116**, 39–50.
11. Tsutakawa,S.E., Classen,S., Chapados,B.R., Arvai,A.S., Finger,L.D., Guenther,G., Tomlinson,C.G., Thompson,P., Sarker,A.H., Shen,B.I. *et al.* (2011) Human flap endonuclease structures, DNA double-base flipping, and a unified understanding of the FEN1 superfamily. *Cell*, **145**, 198–211.
12. Finger,L.D., Blanchard,M.S., Theimer,C.A., Sengerová,B., Singh,P., Chavez,V., Liu,F., Grasby,J.A. and Shen,B. (2009) The 3′-flap pocket of human flap endonuclease 1 is critical for substrate binding and catalysis. *J. Biol. Chem.*, **284**, 22184–22194.
13. Kao,H.I., Henricksen,L.A., Liu,Y. and Bambara,R.A. (2002) Cleavage specificity of Saccharomyces cerevisiae flap endonuclease 1 suggests a double-flap structure as the cellular substrate. *J. Biol. Chem.*, **277**, 14379–14389.
14. Lyamichev,V., Brow,M.A., Varvel,V.E. and Dahlberg,J.E. (1999) Comparison of the 5′ nuclease activities of taq DNA polymerase and its isolated nuclease domain. *Proc. Natl. Acad. Sci. U.S.A.*, **96**, 6143–6148.
15. Fetrow,J.S. (1995) Omega loops: nonregular secondary structures significant in protein function and stability. *FASEB J.*, **9**, 708–717.
16. Papaleo,E., Saladino,G., Lambrughi,M., Lindorff-Larsen,K., Gervasio,F.L. and Nussinov,R. (2016) The Role of Protein Loops and Linkers in Conformational Dynamics and Allostery. *Chem. Rev.*, **116**, 6391–6423.
17. AlMalki,F.A., Flemming,C.S., Zhang,J., Feng,M., Sedelnikova,S.E., Ceska,T., Rafferty,J.B., Sayers,J.R. and Artymiuk,P.J. (2016) Direct observation of DNA threading in flap endonuclease complexes. *Nat. Struct. Mol. Biol.*, **23**, 640–646.
18. Patel,N., Atack,J.M., Finger,L.D., Exell,J.C., Thompson,P., Tsutakawa,S., Tainer,J.A., Williams,D.M. and Grasby,J.A. (2012) Flap endonucleases pass 5′-flaps through a flexible arch using a disorder-thread-order mechanism to confer specificity for free 5′-ends. *Nucleic Acids Res.*, **40**, 4507–4519.
19. Shaw,S.J., Finger,L.D. and Grasby,J.A. (2017) Human Exonuclease 1 Threads 5′-Flap Substrates through Its Helical Arch. *Biochemistry*, **56**, 3704–3707.
20. Shi,Y., Hellinga,H.W. and Beese,L.S. (2017) Interplay of catalysis, fidelity, threading, and processivity in the exo- and endonucleolytic reactions of human exonuclease I. *Proc. Natl. Acad. Sci. U.S.A.*, **114**, 6010–6015.
21. Tsutakawa,S.E., Thompson,M.J., Arvai,A.S., Neil,A.J., Shaw,S.J., Algasaier,S.I., Kim,J.C., Finger,L.D., Jardine,E., Gotham,V.J.B. *et al.* (2017) Phosphate steering by Flap Endonuclease 1 promotes 5′-flap specificity and incision to prevent genome instability. *Nat. Commun.*, **8**, 15855.
22. Sakurai,S., Kitano,K., Yamaguchi,H., Hamada,K., Okada,K., Fukuda,K., Uchida,M., Ohtsuka,E., Morioka,H. and Hakoshima,T. (2005) Structural basis for recruitment of human flap endonuclease 1 to PCNA. *EMBO J.*, **24**, 683–693.
23. Lee,S.H., Princz,L.N., Klugel,M.F., Habermann,B., Pfander,B. and Biertumpfel,C. (2015) Human Holliday junction resolvase GEN1 uses a chromodomain for efficient DNA recognition and cleavage. *Elife*, **4**, e12256.
24. Liu,Y., Freeman,A.D., Declais,A.C., Wilson,T.J., Gartner,A. and Lilley,D.M. (2015) Crystal Structure of a Eukaryotic GEN1 Resolving Enzyme Bound to DNA. *Cell Rep.*, **13**, 2565–2575.
25. Miętus,M., Nowak,E., Jaciuk,M., Kustosz,P., Studnicka,J. and Nowotny,M. (2014) Crystal structure of the catalytic core of Rad2: insights into the mechanism of substrate binding. *Nucleic Acids Res.*, **42**, 10762–10775.
26. Studier,F.W. (2005) Protein production by auto-induction in high density shaking cultures. *Protein Expr. Purif.*, **41**, 207–234.
27. Sivashanmugam,A., Murray,V., Cui,C., Zhang,Y., Wang,J. and Li,Q. (2009) Practical protocols for production of very high yields of recombinant proteins using Escherichia coli. *Protein Sci.*, **18**, 936–948.
28. Gasteiger,E., Gattiker,A., Hoogland,C., Ivanyi,I., Appel,R.D. and Bairoch,A. (2003) ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res.*, **31**, 3784–3788.
29. Cavanagh,J., Fairbrother,W.J., Palmer,A.G. III, Rance,M. and Skelton,N.J. (2007) *Protein NMR Spectroscopy Principles and Practice*. 2nd edn. Elsevier Academic Press, London.
30. Vranken,W.F., Boucher,W., Stevens,T.J., Fogh,R.H., Pajon,A., Llinas,M., Ulrich,E.L., Markley,J.L., Ionides,J. and Laue,E.D. (2005) The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins*, **59**, 687–696.
31. Reed,M.A.C., Hounslow,A.M., Sze,K.H., Barsukov,I.G., Hosszu,L.L.P., Clarke,A.R., Craven,C.J. and Waltho,J.P. (2003) Effects of domain dissection on the folding and stability of the 43 kDa protein PGK probed by NMR. *J. Mol. Biol.*, **330**, 1189–1201.
32. Shen,Y. and Bax,A. (2013) Protein backbone and sidechain torsion angles predicted from NMR chemical shifts using artificial neural networks. *J. Biomol. NMR*, **56**, 227–241.
33. Lakomek,N.A., Ying,J. and Bax,A. (2012) Measurement of ^{15}N relaxation rates in perdeuterated proteins by TROSY-based methods. *J. Biomol. NMR*, **53**, 209–221.
34. d'Auvergne,E.J. and Gooley,P.R. (2003) The use of model selection in the model-free analysis of protein dynamics. *J. Biomol. NMR*, **25**, 25–39.
35. d'Auvergne,E.J. and Gooley,P.R. (2006) Model-free model elimination: a new step in the model-free dynamic analysis of NMR relaxation data. *J. Biomol. NMR*, **35**, 117–135.
36. d'Auvergne,E.J. and Gooley,P.R. (2007) Set theory formulation of the model-free problem and the diffusion seeded model-free paradigm. *Mol. Biosyst.*, **3**, 483–494.
37. d'Auvergne,E.J. and Gooley,P.R. (2008) Optimisation of NMR dynamic models II. A new methodology for the dual optimisation of the model-free parameters and the Brownian rotational diffusion tensor. *J. Biomol. NMR*, **40**, 121–133.
38. d'Auvergne,E.J. and Gooley,P.R. (2008) Optimisation of NMR dynamic models I. Minimisation algorithms and their performance within the model-free and Brownian rotational diffusion spaces. *J. Biomol. NMR*, **40**, 107–119.
39. Cantor,C.R., Warshaw,M.M. and Shapiro,H. (1970) Oligonucleotide interactions. 3. Circular dichroism studies of the conformation of deoxyoligonucleotides. *Biopolymers*, **9**, 1059–1077.
40. Cavaluzzi,M.J. and Borer,P.N. (2004) Revised UV extinction coefficients for nucleoside-5′-monophosphates and unpaired DNA and RNA. *Nucleic Acids Res.*, **32**, e13.
41. Braman,J., Papworth,C. and Greener,A. (1996) Site-directed mutagenesis using double-stranded plasmid DNA templates. *Methods Mol. Biol.*, **57**, 31–44.
42. Sengerova,B., Tomlinson,C., Atack,J.M., Williams,R., Sayers,J.R., Williams,N.H. and Grasby,J.A. (2010) Bronsted analysis and rate-limiting steps for the T5 flap endonuclease catalyzed hydrolysis of exonucleolytic substrates. *Biochemistry*, **49**, 8085–8093.
43. Cheng,N.-S. (2008) Formula for the viscosity of a Glycerol−Water mixture. *Ind. Eng. Chem. Res.*, **47**, 3285–3288.
44. Brouwer,A.C. and Kirsch,J.F. (1982) Investigation of diffusion-limited rates of chymotrypsin reactions by viscosity variation. *Biochemistry*, **21**, 1302–1307.
45. Kapanidis,A.N., Lee,N.K., Laurence,T.A., Doose,S., Margeat,E. and Weiss,S. (2004) Fluorescence-aided molecule sorting: analysis of structure and interactions by alternating-laser excitation of single molecules. *Proc. Natl. Acad. Sci. U.S.A.*, **101**, 8936–8941.
46. Ingargiola,A., Lerner,E., Chung,S., Weiss,S. and Michalet,X. (2016) FRETBursts: an open source toolkit for analysis of freely-diffusing single-molecule FRET. *PLoS One*, **11**, e0160716.
47. Nir,E., Michalet,X., Hamadani,K.M., Laurence,T.A., Neuhauser,D., Kovchegov,Y. and Weiss,S. (2006) Shot-noise limited single-molecule FRET histograms: comparison between theory and experiments. *J. Phys. Chem. B*, **110**, 22103–22124.
48. Hohlbein,J., Craggs,T.D. and Cordes,T. (2014) Alternating-laser excitation: single-molecule FRET and beyond. *Chem. Soc. Rev.*, **43**, 1156–1171.
49. Cordingley,M.G., Callahan,P.L., Sardana,V.V., Garsky,V.M. and Colonno,R.J. (1990) Substrate requirements of human rhinovirus 3C protease for peptide cleavage in vitro. *J. Biol. Chem.*, **265**, 9062–9065.
50. Kay,L.E., Torchia,D.A. and Bax,A. (1989) Backbone dynamics of proteins as studied by ^{15}N inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease. *Biochemistry*, **28**, 8972–8979.

Downloaded from https://academic.oup.com/nar/advance-article-abstract/doi/10.1093/nar/gky293/4990015
by University of Sheffield user
on 03 May 2018

51. Rossi,P., Swapna,G.V., Huang,Y.J., Aramini,J.M., Anklin,C., Conover,K., Hamilton,K., Xiao,R., Acton,T.B., Ertekin,A. *et al.* (2010) A microscale protein NMR sample screening pipeline. *J. Biomol. NMR*, **46**, 11–22.
52. Bieri,M., d'Auvergne,E.J. and Gooley,P.R. (2011) relaxGUI: a new software for fast and simple NMR relaxation data analysis and calculation of ps-ns and mus motion of proteins. *J. Biomol. NMR*, **50**, 147–155.
53. Tamiola,K., Acar,B. and Mulder,F.A. (2010) Sequence-specific random coil chemical shifts of intrinsically disordered proteins. *J. Am. Chem. Soc.*, **132**, 18000–18003.
54. Lacroix,E., Viguera,A.R. and Serrano,L. (1998) Elucidating the folding problem of alpha-helices: local motifs, long-range electrostatics, ionic-strength dependence and prediction of NMR parameters. *J. Mol. Biol.*, **284**, 173–191.
55. Das,R.K., Ruff,K.M. and Pappu,R.V. (2015) Relating sequence encoded information to form and function of intrinsically disordered proteins. *Curr. Opin. Struct. Biol.*, **32**, 102–112.
56. Holehouse,A.S., Das,R.K., Ahad,J.N., Richardson,M.O. and Pappu,R.V. (2017) CIDER: Resources to analyze Sequence-Ensemble relationships of intrinsically disordered proteins. *Biophys. J.*, **112**, 16–21.
57. Das,R.K. and Pappu,R.V. (2013) Conformations of intrinsically disordered proteins are influenced by linear sequence distributions of oppositely charged residues. *Proc. Natl. Acad. Sci. U.S.A.*, **110**, 13392–13397.
58. Finger,L.D., Patel,N., Beddows,A., Ma,L., Exell,J.C., Jardine,E., Jones,A.C. and Grasby,J.A. (2013) Observation of unpaired substrate DNA in the flap endonuclease-1 active site. *Nucleic Acids Res.*, **41**, 9839–9847.
59. Patel,N., Exell,J.C., Jardine,E., Ombler,B., Finger,L.D., Ciani,B. and Grasby,J.A. (2013) Proline scanning mutagenesis reveals a role for the flap endonuclease-1 helical cap in substrate unpairing. *J. Biol. Chem.*, **288**, 34239–34248.
60. Syson,K., Tomlinson,C., Chapados,B.R., Sayers,J.R., Tainer,J.A., Williams,N.H. and Grasby,J.A. (2008) Three metal ions participate in the reaction catalyzed by T5 flap endonuclease. *J. Biol. Chem.*, **283**, 28741–28746.
61. Kim,Y., Ho,S.O., Gassman,N.R., Korlann,Y., Landorf,E.V., Collart,F.R. and Weiss,S. (2008) Efficient site-specific labeling of proteins via cysteines. *Bioconjug. Chem.*, **19**, 786–791.
62. Tsutakawa,S.E., Lafrance-Vanasse,J. and Tainer,J.A. (2014) The cutting edges in DNA repair, licensing, and fidelity: DNA and RNA repair nucleases sculpt DNA to measure twice, cut once. *DNA Repair (Amst.)*, **19**, 95–107.
63. Tsutakawa,S.E. and Tainer,J.A. (2012) Double strand binding-single strand incision mechanism for human flap endonuclease: implications for the superfamily. *Mech. Ageing. Dev.*, **133**, 195–202.
64. Orans,J., McSweeney,Elizabeth A., Iyer,Ravi R., Hast,Michael A., Hellinga,Homme W., Modrich,P. and Beese,Lorena S. (2011) Structures of human exonuclease 1 DNA complexes suggest a unified mechanism for nuclease family. *Cell*, **145**, 212–223.
65. Kannan,S., Kohlhoff,K. and Zacharias,M. (2006) B-DNA under stress: over- and untwisting of DNA during molecular dynamics simulations. *Biophys. J.*, **91**, 2956–2965.
66. Chaires,J.B. (2008) Allostery: DNA does it, too. *ACS Chem. Biol.*, **3**, 207–209.
67. Boehr,D.D., Nussinov,R. and Wright,P.E. (2009) The role of dynamic conformational ensembles in biomolecular recognition. *Nat. Chem. Biol.*, **5**, 789–796.
68. Csermely,P., Palotai,R. and Nussinov,R. (2010) Induced fit, conformational selection and independent dynamic segments: an extended view of binding events. *Trends Biochem. Sci.*, **35**, 539–546.
69. Ma,B. and Nussinov,R. (2010) Enzyme dynamics point to stepwise conformational selection in catalysis. *Curr. Opin. Chem. Biol.*, **14**, 652–659.
70. Oldfield,C.J. and Dunker,A.K. (2014) Intrinsically disordered proteins and intrinsically disordered protein regions. *Annu. Rev. Biochem.*, **83**, 553–584.
71. Habchi,J., Tompa,P., Longhi,S. and Uversky,V.N. (2014) Introducing protein intrinsic disorder. *Chem. Rev.*, **114**, 6561–6588.
72. Nussinov,R., Ma,B. and Tsai,C.J. (2014) Multiple conformational selection and induced fit events take place in allosteric propagation. *Biophys. Chem.*, **186**, 22–30.

Downloaded from https://academic.oup.com/nar/advance-article-abstract/doi/10.1093/nar/gky293/4990015
by University of Sheffield user
on 03 May 2018

SUPPLEMENTARY DATA

Regional Intrinsic Disorder Couples Substrate Specificity and Scissile Phosphate Diester Selectivity in Human Flap Endonuclease 1

Ian A. Bennet, L. David Finger, Nicola J. Baxter, Benjamin Ambrose, Andrea M. Hounslow, Mark J. Thompson, Jack C. Exell, Nur Nazihah B. Md. Shahari, Timothy D. Craggs, Jonathan P. Waltho, Jane A. Grasby

Table S1
Oligonucleotide* sequences used for site directed mutagenesis of pET28b-hFEN1

NAME	SEQUENCE
C235A_F	5'gctaggcagtgactacgctgagagtatccggggt3'
C235A_R	5'accccggatactctcagcgtagtcactgcctagc3'
C311A_F	5'aagagctgatcaagttcatggctggtgaaaagcagttctctg3'
C311A_R	5'cagagaactgcttttcaccagccatgaacttgatcagctctt3'
E120C_F	5'aatttttccacctcctggcaggccccagcagcctgag3'
E120C_R	5'ctcaggctgctggggcctgccaggaggtggaaaaatt3'
S293C_F	5'agctccacacactctgggtccagcacc3'
S293C_R	5'ggtgctggacccagagtgtgtggagct3'

*Oligonucleotides were purchased from ThermoFisher without purification.

Table S2 (related to Figure 3)
Model selection for all 179 analyzed residues

Models	Number of Residues selected for each model
$m1 = \{S^2\}$	21
$m2 = \{S^2, \tau_e\}$	11
$m3 = \{S^2, R_{ex}\}$	62
$m4 = \{S^2, \tau_e, R_{ex}\}$	24
$m5 = \{S^2, S^2_f, \tau_s\}$	18
$m6 = \{S^2, \tau_f, S^2_f, \tau_s\}$	2
$m7 = \{S^2, S^2_f, \tau_s, R_{ex}\}$	28
$m8 = \{S^2, \tau_f, S^2_f, \tau_s, R_{ex}\}$	13
TOTAL	**179**

Table S3
Efficiency of fluorophore-maleimide labelling of $hFEN1_Q$ determined using UV/Vis spectrophotometry

Wavelength λ	$A_{\lambda max}$	$A_{\lambda max\ Corrected}$	Concentration#	[Chromophore]/[Protein] Ratio
280 nm	2.16	1.52*	66.19	1.00
559 nm	6.23	6.23‡	47.96	0.72
646 nm	7.05	7.05	47.00	0.71

*Corrected for Cy3b and ATTO 647N absorbtion at 280 nm

‡Corrected for ATTO 647N absorbtion at 559 nm

#Extinction coefficients used ($\varepsilon_{280} = 2.292\times10^4\ M^{-1}cm^{-1}$, $\varepsilon_{559} = 1.3\times10^5\ M^{-1}cm^{-1}$, $\varepsilon_{646} = 1.5\times10^5\ M^{-1}cm^{-1}$)

Table S4
Parameters from fitting the smFRET data to a two Gaussian fit model (related to Figure 7)

Gauss Curve	Mean E	Sigma	Amplitude
FEN1 Alone low FRET	0.6455	0.1989	0.5304
FEN1 Alone high FRET	0.8762	0.0924	0.4428
+20nM DNA low FRET	0.6450	0.1990	0.4732
+20nM DNA high FRET	0.8760	0.0920	0.5044

Figure S1 (Related to Figure 1)

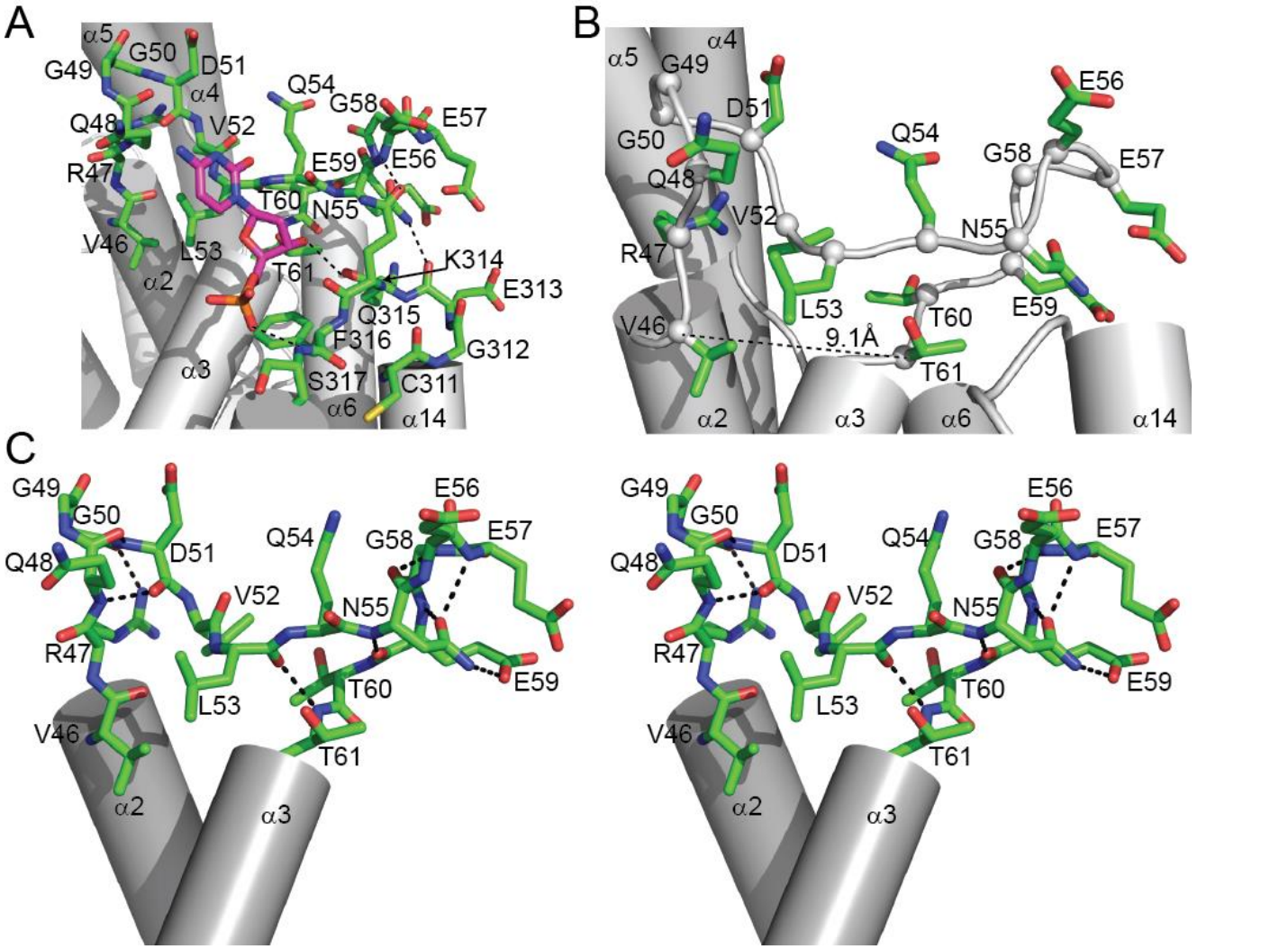

Figure S1. The 3'-flap is bound in a pocket made by the α14-α15 loop, the first turn of α15 and the α2-α3 loop, which conforms to the definition of an omega (Ω)-loop**.** (**A**) View of 3'-flap nucleotide of the DNA substrate (magenta) (3Q8K) cradled in its pocket (1). Residue identities and α-helices are indicated. Dashed lines indicate the hydrogen bonds formed between the α14–α15 and α2–α3 loops and the α14–α15 loop and the 3'-flap nucleotide. The N55 side chain amide forms a hydrogen bond with the E313 backbone carbonyl. The backbone amide of E56 forms a hydrogen bond with the side chain carbonyl of Q315. The backbone carbonyl of K314 forms a hydrogen bond with the 3'-hydroxyl moiety of the 3'-flap nucleotide. The backbone amide of S317 forms a hydrogen bond with a non-bridging phosphoryl oxygen atom of the phosphate diester moiety of the 3'-flap nucleotide. (**B**) Similar view as in panel A with the 3'-flap nucleotide and α14–α15 loop omitted for clarity. The traditional definition of an Ω-loop is a loop of 6 to 16 amino acid residues with the C^{α}–C^{α} distance of the loop termini both less than 10 Å (dashed line V46–T61 = 9.1 Å) and less than two-thirds the longest pairwise C^{α}–C^{α} distance in the loop (V46–E57 and G50–E57 = 18.4 Å). (**C**) Wall-eyed stereo-view of the α2-α3 loop to show the intra-loop network of hydrogen bonds among backbone and sidechain atoms when structured as an Ω-loop (2).

Figure S2 (Related to Figure 1)

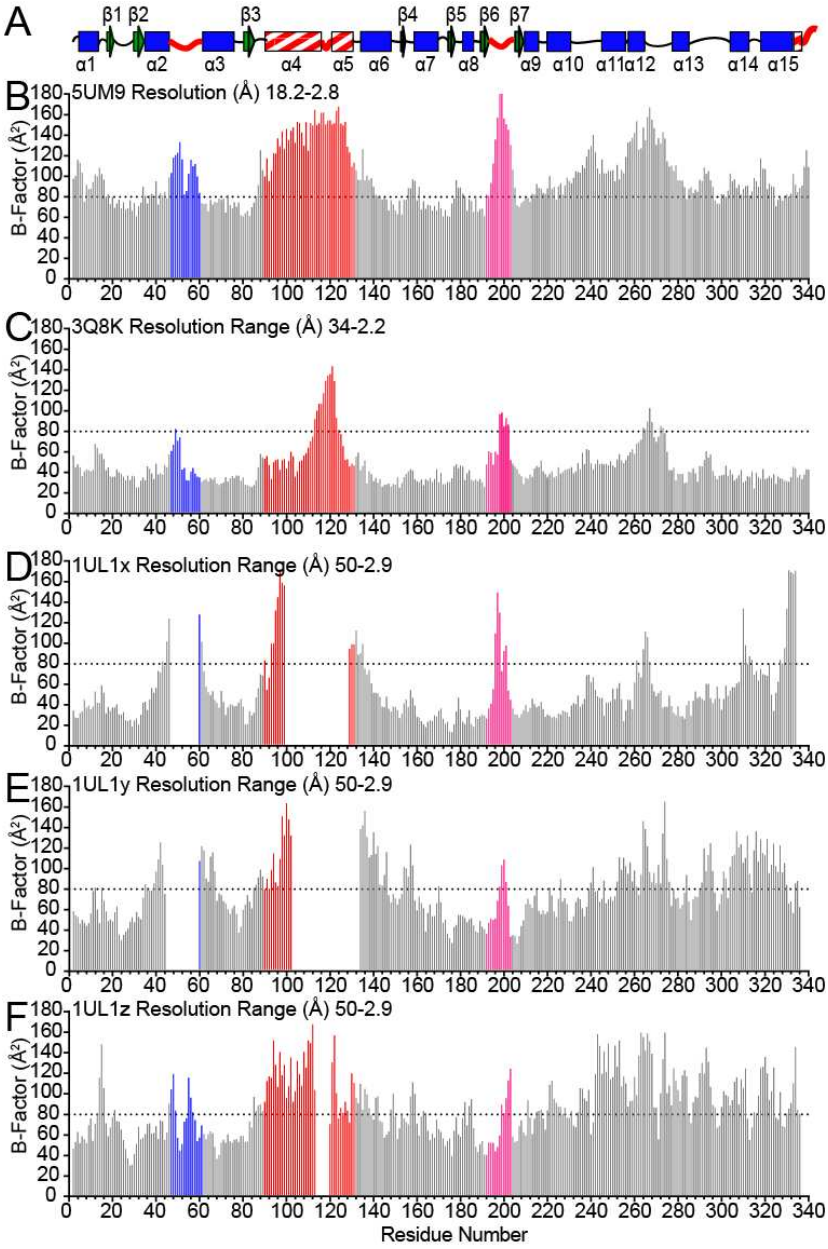

Figure S2. The α2-α3 loop, arch region, β-pin and the C-terminal residues consistently show higher than average temperature-factors (or B-factors). (**A**) Secondary structure schematic of hFEN1 from 3Q8K. Blue rectangles, green arrows and black lines indicate α-helices, β-strands, and loops, respectively. Loops known to have structural heterogeneity (1UL1 vs. 3Q8K) are indicated by red lines. Red and white stripped rectangles highlight regions where either α-helix or disorder has been observed. Plots of backbone amide nitrogen B-factors versus residue number for (**B**) 5UM9 (3), (**C**) 3Q8K (1), (**D**) 1UL1x, (**E**) 1UL1y and (**F**) 1UL1z (4). The α2-α3 loop, arch region, and β-pin are highlighted in blue, red and magenta, respectively. The dashed line indicates a B-factor of 80 Å^2, which is equivalent to a root mean square displacement of 1 Å.

Figure S3 (Related to Figure 1)

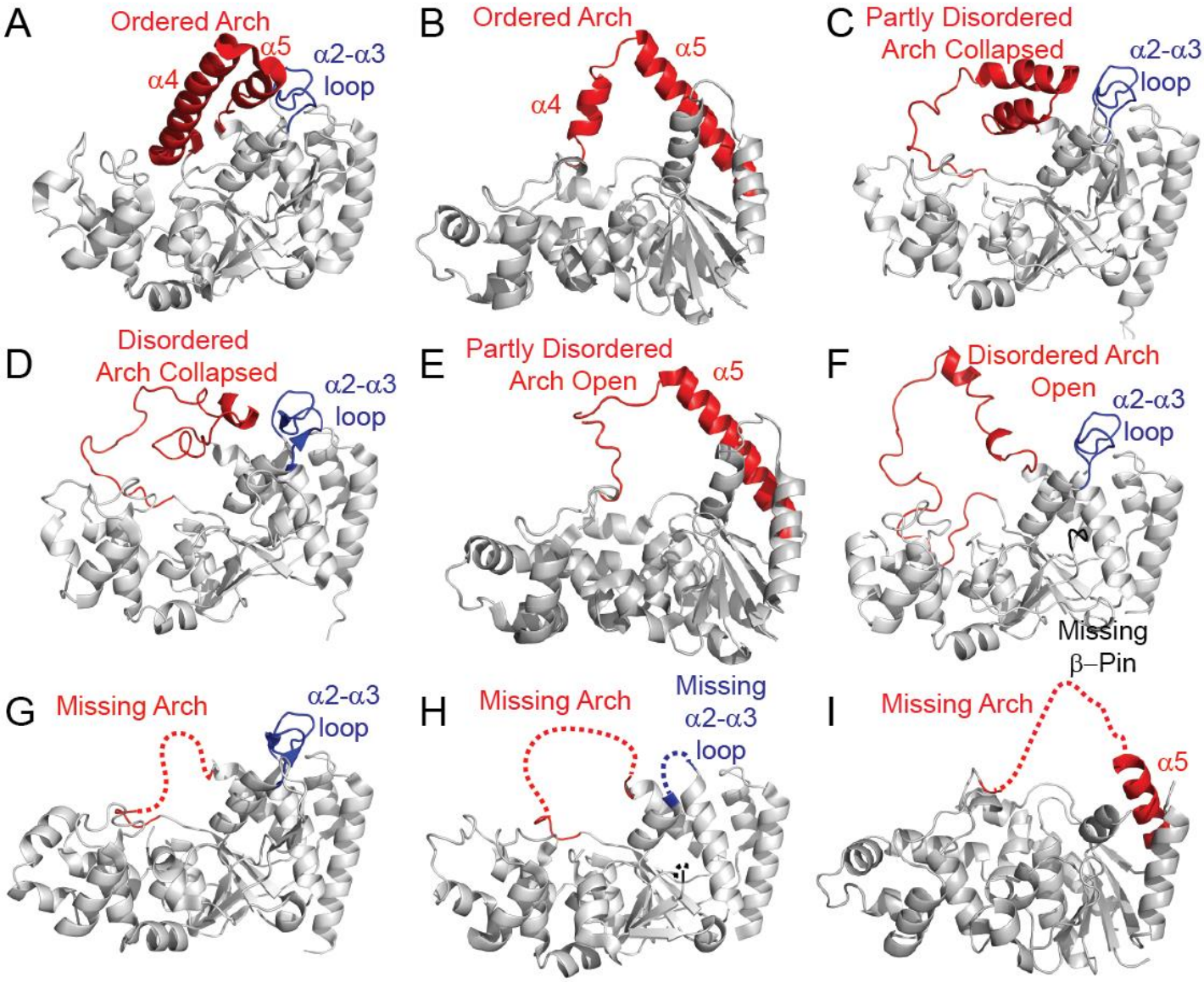

Figure S3. The arch regions and α2-α3 loops of FEN1 proteins from other organisms also show structural heterogeneity. Crystal structures of (**A**) *Desulfurococcus solfataricus* (3ORY) (7) and (**B**) several T5 FEN (1EXN, 1UT5 and 1UT8) (5,6) show an ordered arch region and α2-α3 loop. Two crystal structures of FEN1 from (**C**) *Pyrococcus furiosus* (1B43) (8) and (**D**) *Pyrococcus horikoshii* (1MC8) (9) show an ordered α2–α3 loop and an arch region with various degrees of secondary structure and random coil present, but it is collapsed over the active site. Crystal structures from (**E**) T5 FEN (5HML and 5HMM) (10) and (**F**) *Methanococcus jannaschii* (1A76 and 1A77) (11) show an ordered α2–α3 loop and an arch region with partly disordered areas, some visible secondary structure and an extended, open conformation. (**G**) The crystal structure of *Methanopyrus kandleri* (4WA8) (12) shows an ordered α2–α3 loop, but the arch region is completely missing suggesting disorder in these segments. The crystal structures of hFEN1 in complex with (**H**) an inhibitor (5FV7) (13) and (**i**) T5 FEN (1XO1) (14) lack density for both the arch region and the α2–α3 loop.

Figure S4 (Related to Figure 2)

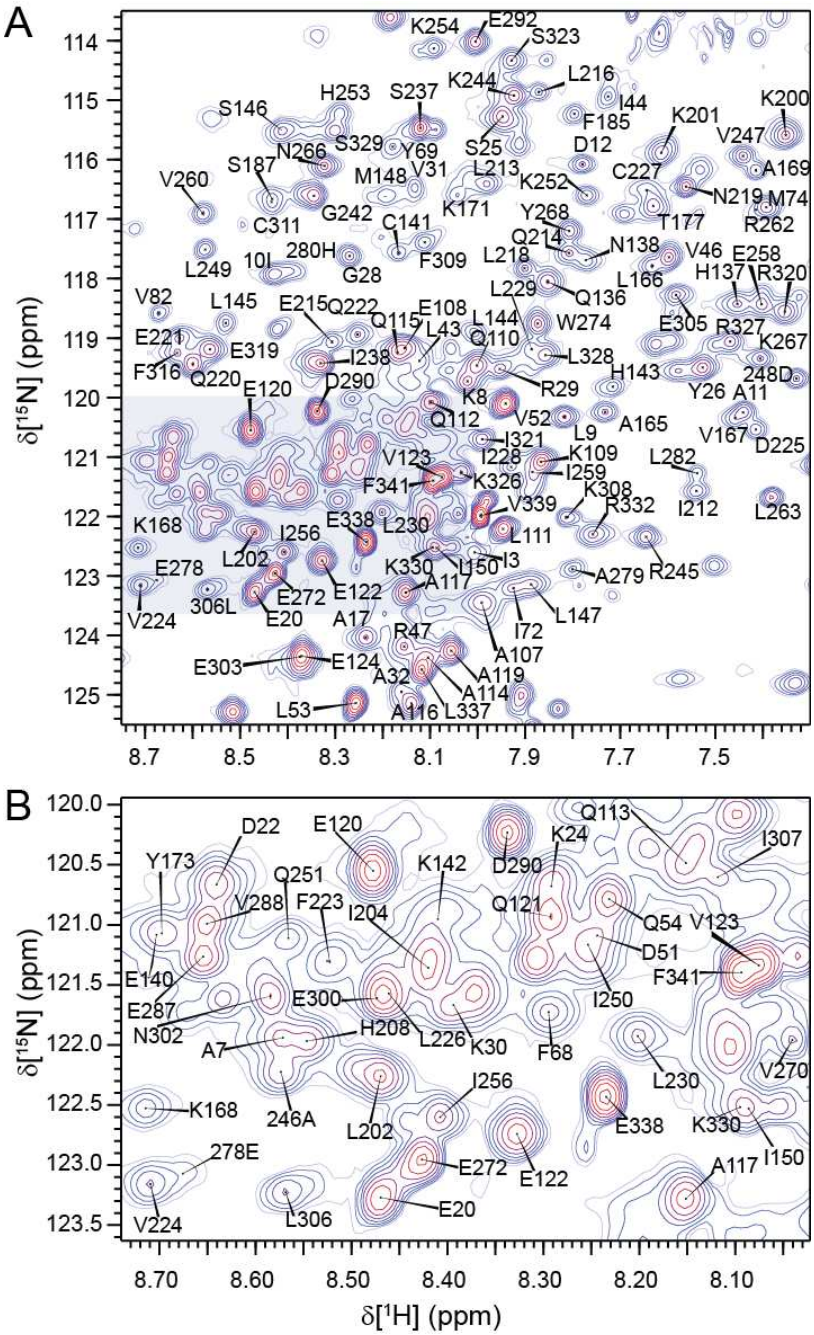

Figure S4. Expanded views of the crowded regions of the hFEN1 ^{1}H-^{15}N TROSY spectrum. (**A**) Expanded view of the shaded area in Figure 2A with assignments shown. (**B**) An expanded view of the shaded area in panel A with resonance assignments shown.

Figure S5 (Related to Figures1 and 2)

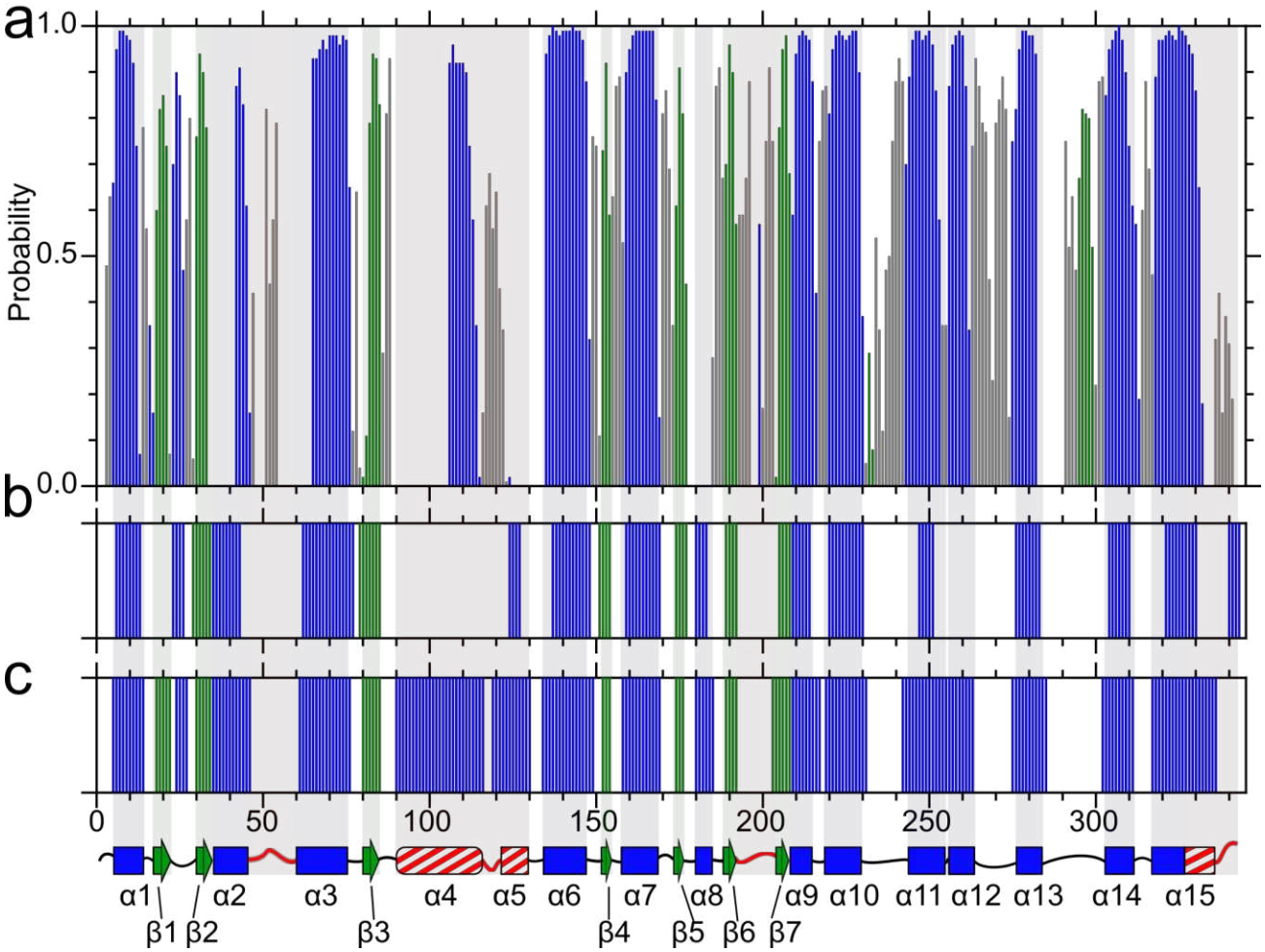

Figure S5. Secondary structure prediction for hFEN1 using backbone chemical shifts corresponds with the X-ray crystal structure of hFEN1 for the saddle region and the α2-α3 loop, but reveals differences in the arch region. (**A**) TALOS-N (15) secondary structure prediction based on assigned backbone chemical shifts ($^{1}H^{N}$, ^{15}N, $^{13}C^{\alpha}$, $^{13}C^{\beta}$, $^{13}C'$) for hFEN1. Predicted α-helices, β-strands and loops are shown as blue, green and gray bars, respectively. Absence of a bar in panel (**a**) indicates that the residue was not assigned. Secondary structure assignments for (**B**) 1UL1z (4) and (**c**) 3Q8K (1). Well-characterized α-helices and β-strands are shown, but loops are indicated by the absence of a bar in the panels (**B,C**). Secondary structure schematic derived from 3Q8K as described Figure S2A is included at the bottom of the figure.

Figure S6 (Related to Figure 3)

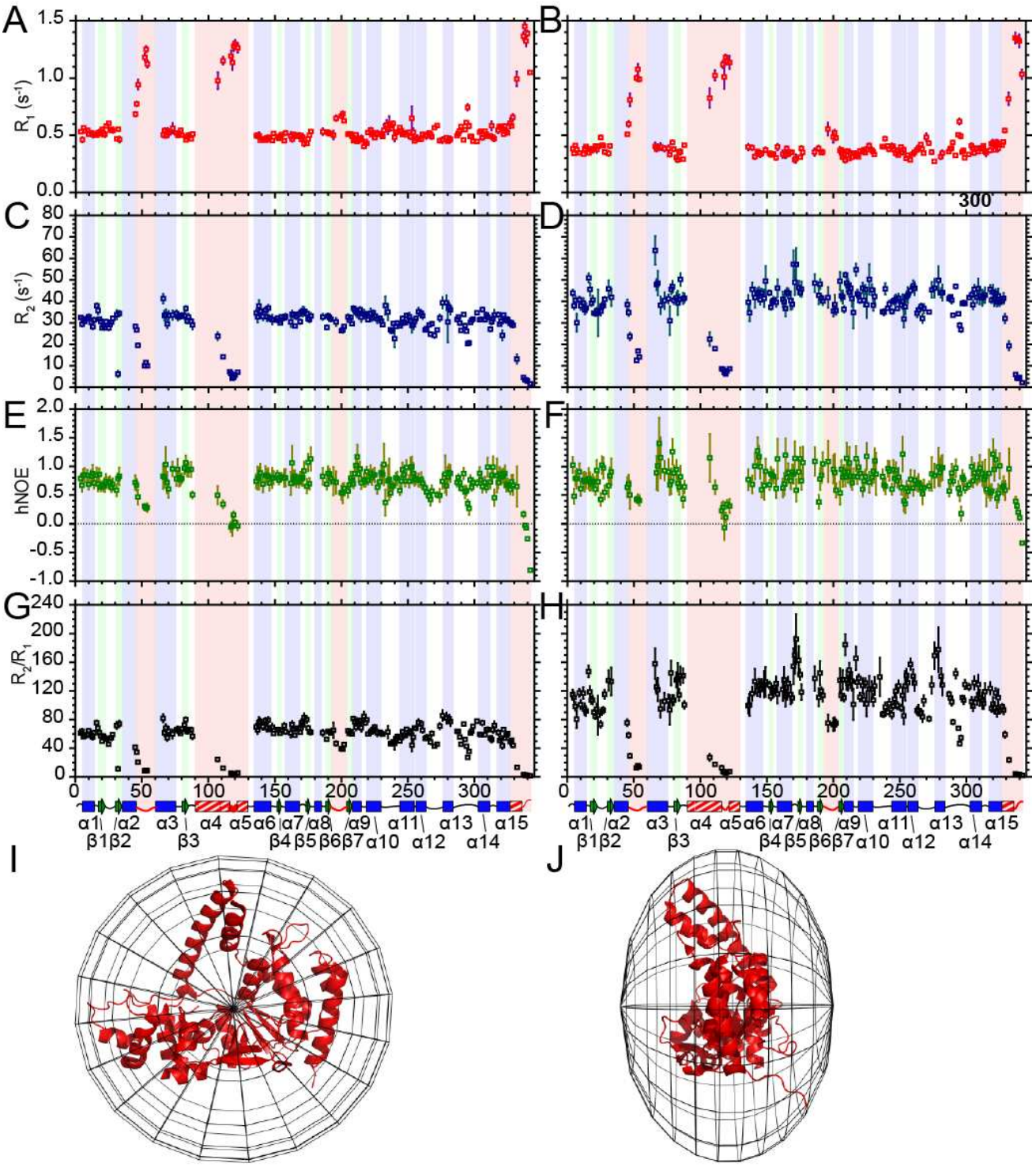

Figure S6. Experimentally-determined backbone ^{15}N relaxation parameters for hFEN1. (**A,B**) ^{15}N spin-lattice (R_1) rates, (**C,D**) spin-spin (R_2) rates and (**E,F**) ^{15}N-{^1H} NOE (hNOE) values at (**A,C,E**) 600 MHz and (**B,D,F**) 800 MHz for ^2H,^{15}N-labelled hFEN1 measured using interleaved TROSY-readout pulse sequences (16). R_2/R_1 plots generated from the data above for (**G**) 600 MHz and (**H**) 800 MHz, showing the per-residue average molecular correlation time in model-free methodology. Secondary structure maps are provided as described in Figure S2A. The oblate spheroid diffusion tensor best describes the tumbling of hFEN1 in solution. View of the diffusion tensor with the hFEN1 protein structure (3Q8K) (1) along the (**I**) z-axis and (**J**) x,y plane to illustrate agreement between the structure and the selected diffusion tensor.

Figure S7 (Related to Figure 3)

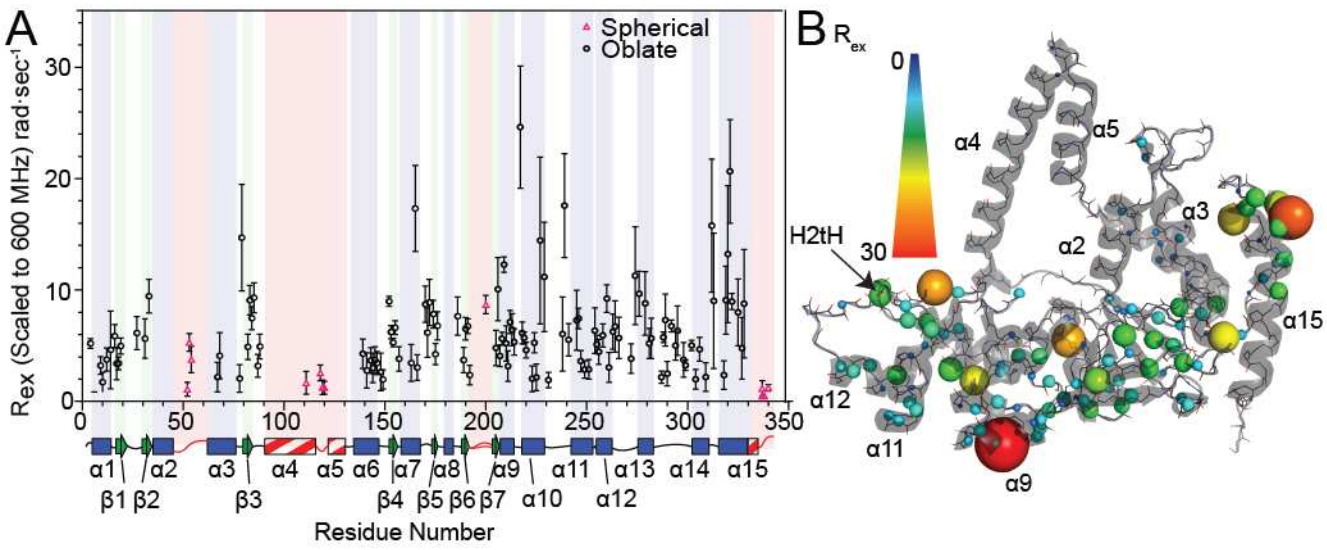

Figure S7. Model-free analysis of hFEN1 relaxation data identifies regions with millisecond timescale motions. (**A**) Chemical exchange terms (R_{ex}) were derived from *relax* (17) using backbone ^{15}N relaxation data acquired at 600 and 800 MHz (Figure S6A-F) and plotted versus residue number. Black circles represent data fitted to the oblate spheroid diffusion tensor (Figure 6I and J), whereas pink triangles were fitted to a spherical diffusion tensor. Locations of the residues with respect to protein secondary structure are illustrated below the panel as described in Figure S2A. (**B**) R_{ex} values plotted on a cartoon depiction of the hFEN1 protein structure (3Q8K) (1). The spheres represent the nitrogen nuclei for which data were derived. The R_{ex} spectrum bars illustrate the magnitude of R_{ex} values with respect to sphere color and size.

Figure S8 (Related to Figures 5 and 6)

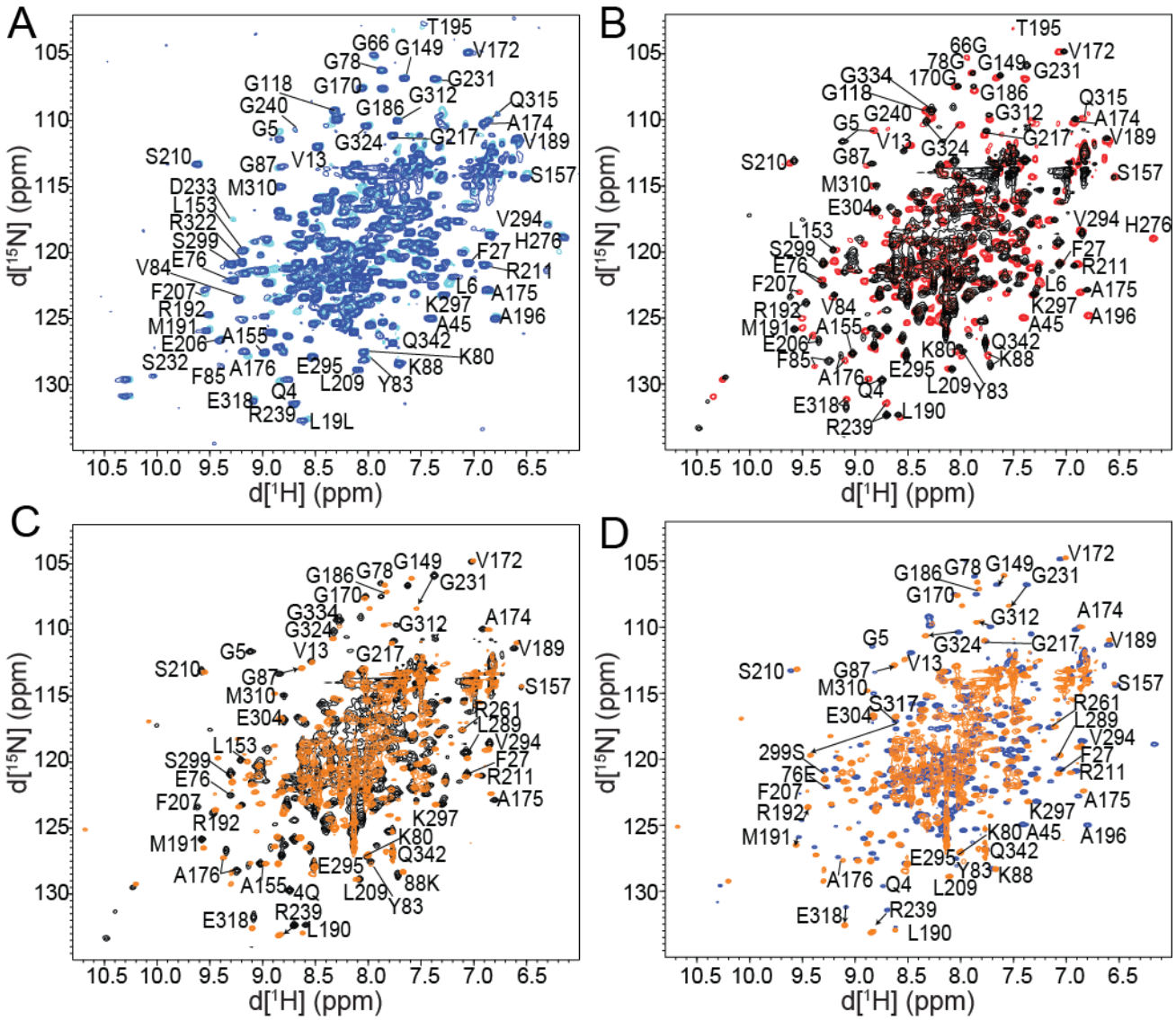

Figure S8. Superposed ^{1}H-^{15}N TROSY spectra illustrating chemical shift changes that arise from the addition of divalent cations to hFEN1, hFEN1$_{K93A}$ and the hFEN1$_{K93A}$-DNA complex. (**A**) ^{1}H-^{15}N TROSY spectra of hFEN1 in the presence of 8 mM Mg^{2+} (cyan) and 8 mM Ca^{2+} (blue) are nearly identical. Minor differences are seen for residues close to the active site (e.g., Q4, G5, L6, S157 and A176) and may arise due to the larger size and looser coordination geometry of Ca^{2+}. These small changes indicate that the Mg^{2+} and Ca^{2+} divalent metal ions are coordinated in the same active site locations. (**B**) ^{1}H-^{15}N TROSY spectra of hFEN1$_{K93A}$ (red) and the hFEN1$_{K93A}$-DNA complex (black) in the presence of 0.1 mM EDTA. (**C**) ^{1}H-^{15}N TROSY spectra of the hFEN1$_{K93A}$-DNA complex in the presence of 0.1 mM EDTA (black) or 8 mM Ca^{2+} (orange). (**D**) ^{1}H-^{15}N TROSY spectra of hFEN1$_{K93A}$ (blue) and the hFEN1$_{K93A}$-DNA complex (orange) in the presence of 8 mM Ca^{2+}. Peaks with good chemical shift dispersion are labelled.

Figure S9 (Related to Figures 6 and 7)

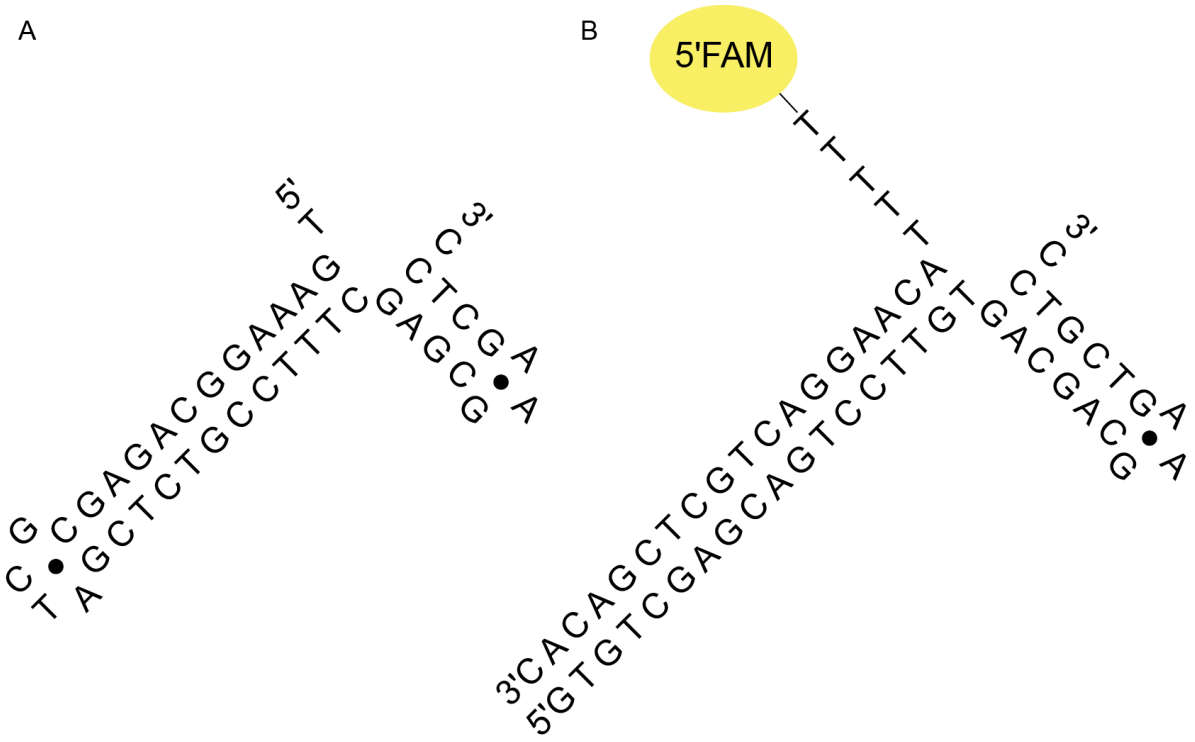

Figure S9. Sequence and secondary structure of the DNA substrate used herein. (**A**) The hFEN1$_{K93A}$-DNA complex uses a unimolecular substrate design bearing stable tri- (dGAA) (18) and tetra- (dGCTA) (19) loops to prevent hFEN1 inadvertently binding blunt-ended duplex. (**B**) The bimolecular substrate was used to assess the initial rates of reaction used to determine second order rate constants. In addition, this same substrate was used in smFRET experiments except it lacked the 5'-FAM label.

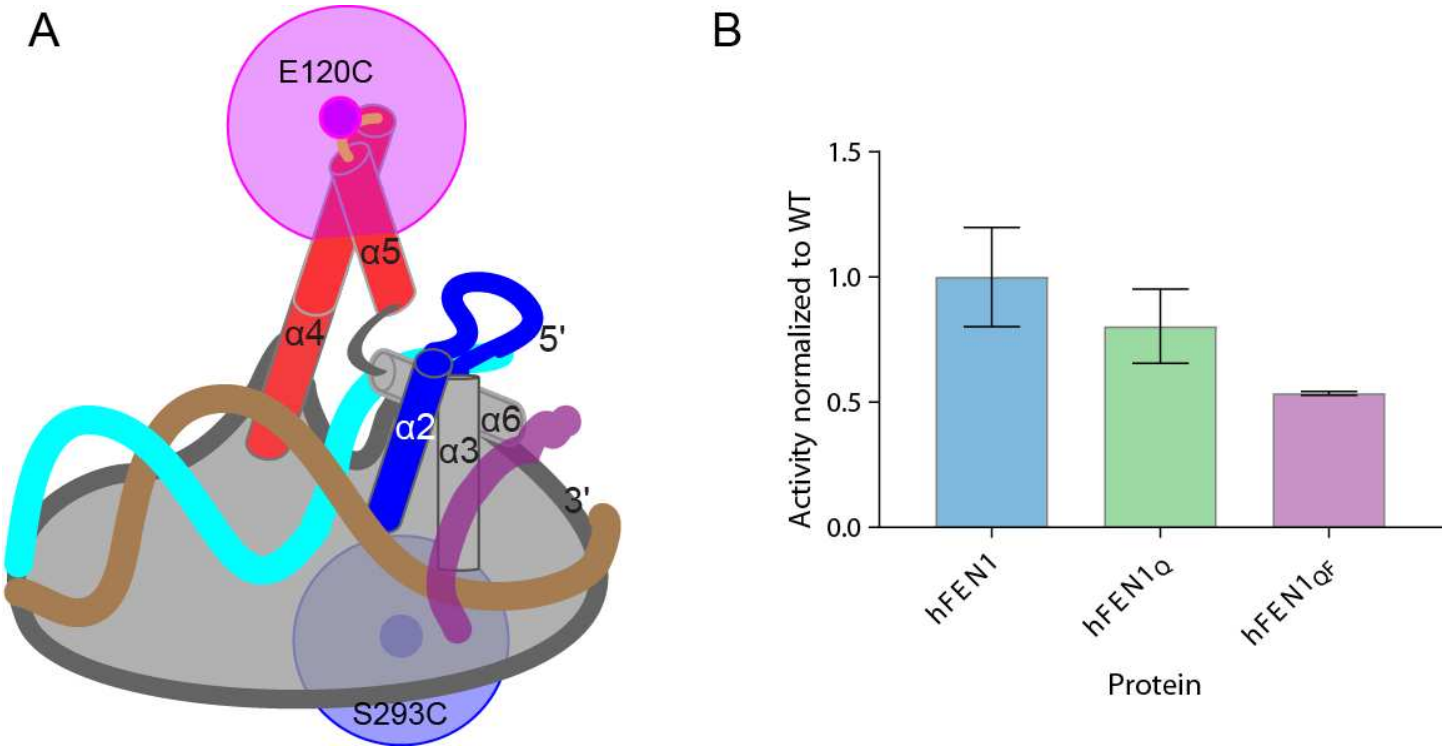

Figure S10. Relative labelling positions, activity and mass spectrometry results for hFEN1$_{Q}$ and hFEN1$_{QF}$ proteins. (**A**) Relative labelling positions of the fluorophores on hFEN1$_{QF}$ with substrate. (**B**) Activity of hFEN1 proteins assessed by normalized initial rate measurements at 50 nM substrate relative to WT protein. Values represent the average of three measurements (n=3), and error bars represent the standard error of the mean (SEM).

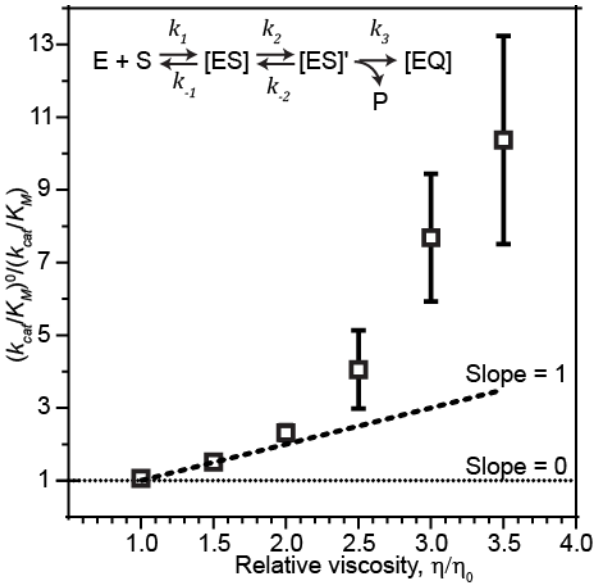

Figure S11. The effect of viscogen on relative second order rate constants is larger than expected and non-linear. Graph of mean relative second order rate constants $(k_{cat}/K_M)_0/(k_{cat}/K_M)$ for hFEN1-catalyzed hydrolysis of DF5,1 (Figure S9B) versus relative viscosity (SEM, n=3) shows a non-linear dependence. Inset: simplified scheme of the hFEN1-catalyzed reaction illustrating bimolecular complex [ES] formation, a conformational change to the catalytically-competent state [ES]' and substrate hydrolysis into ssDNA (P) and enzyme-dsDNA [EQ] products. Microrate constants (k) up to and including the first irreversible step (k_3) contribute to k_{cat}/K_M. hFEN1 second-order reaction rates (k_{cat}/K_M) on double-flap DNA substrates approached diffusion-control (10^7–10^9 $M^{-1}\cdot s^{-1}$); therefore, hFEN1 has likely achieved catalytic perfection with its optimal substrate (Figure S9B) (20). Observation of a linear dependence for the reciprocal of relative k_{cat}/K_M with respect the relative viscosity of the solution has been the standard manner by which to confirm catalytic perfection or diffusion-control for enzymes (21,22). Assuming that only the rates of enzyme and substrate association and complex disassociation are affected by viscogen, a linear relationship is expected for the reciprocal of normalized k_{cat}/K_m with respect to relative viscosity. Moreover, a diffusion-controlled reaction is expected to afford a slope of one (dashed line), whereas an enzyme that is not diffusion controlled is expected to afford a line with a slope of zero (dotted line) (21,22). Although increasing the relative viscosity of the reaction buffer with either glycerol or sucrose did increase the reciprocal of normalized k_{cat}/K_m with the optimal double flap substrate, the effects were non-linear especially at relative viscosities greater than two. Similar non-linear viscogen dependence has been reported for an intrinsically disordered protein that folds upon binding its protein partner (23). In addition to slowing diffusional encounter, microviscogens have also been shown to slow protein conformational rearrangements (24) and protein-catalyzed reactions when conformational changes in the catalytic cycle were involved (25). Indeed, the impact of viscogen on rates of hFEN1-catalyzed reaction under single-turnover conditions measured by single molecule techniques is consistent with viscogen slowing conformational change steps after enzyme-substrate complex formation (26). Therefore, the non-linear viscogen dependence on relative second order rate constants could be due to the combined effect of viscogen on the rates of bimolecular association and subsequent conformational changes. However, we cannot rule out the possibility that sucrose and glycerol are acting either as inhibitors of the reaction by some unknown mechanism or that these polyols are affecting the energy landscape of the various conformational ensembles by acting as osmolytes.

SUPPLEMENTARY REFERENCES

...1. Tsutakawa, S.E., Classen, S., Chapados, B.R., Arvai, A.S., Finger, L.D., Guenther, G., Tomlinson, C.G., Thompson, P., Sarker, A.H., Shen, B.i. *et al.* (2011) Human flap endonuclease structures, DNA double-base flipping, and a unified understanding of the FEN1 superfamily. *Cell*, **145**, 198-211.

...2. Fetrow, J.S. (1995) Omega loops: nonregular secondary structures significant in protein function and stability. *FASEB J.*, **9**, 708-717.

...3. Tsutakawa, S.E., Thompson, M.J., Arvai, A.S., Neil, A.J., Shaw, S.J., Algasaier, S.I., Kim, J.C., Finger, L.D., Jardine, E., Gotham, V.J.B. *et al.* (2017) Phosphate steering by Flap Endonuclease 1 promotes 5'-flap specificity and incision to prevent genome instability. *Nat. Commun.*, **8**, 15855.

...4. Sakurai, S., Kitano, K., Yamaguchi, H., Hamada, K., Okada, K., Fukuda, K., Uchida, M., Ohtsuka, E., Morioka, H. and Hakoshima, T. (2005) Structural basis for recruitment of human flap endonuclease 1 to PCNA. *EMBO J.*, **24**, 683-693.

...5. Ceska, T.A., Sayers, J.R., Stier, G. and Suck, D. (1996) A helical arch allowing single-stranded DNA to thread through T5 5'-exonuclease. *Nature*, **382**, 90-93.

...6. Feng, M., Patel, D., Dervan, J.J., Ceska, T., Suck, D., Haq, I. and Sayers, J.R. (2004) Roles of divalent metal ions in flap endonuclease-substrate interactions. *Nat. Struct. Mol. Biol.*, **11**, 450-456.

...7. Mase, T., Kubota, K., Miyazono, K., Kawarabayasi, Y. and Tanokura, M. (2009) Crystallization and preliminary X-ray analysis of flap endonuclease 1 (FEN1) from Desulfurococcus amylolyticus. *Acta Crystallogr Sect F Struct Biol Cryst Commun*, **65**, 923-925.

...8. Hosfield, D.J., Mol, C.D., Shen, B. and Tainer, J.A. (1998) Structure of the DNA repair and replication endonuclease and exonuclease FEN-1: coupling DNA and PCNA binding to FEN-1 activity. *Cell*, **95**, 135-146.

...9. Matsui, E., Musti, K.V., Abe, J., Yamasaki, K., Matsui, I. and Harata, K. (2002) Molecular structure and novel DNA binding sites located in loops of flap endonuclease-1 from Pyrococcus horikoshii. *J Biol Chem*, **277**, 37840-37847.

...10. AlMalki, F.A., Flemming, C.S., Zhang, J., Feng, M., Sedelnikova, S.E., Ceska, T., Rafferty, J.B., Sayers, J.R. and Artymiuk, P.J. (2016) Direct observation of DNA threading in flap endonuclease complexes. *Nat. Struct. Mol. Biol.*, **23**, 640-646.

...11. Hwang, K.Y., Baek, K., Kim, H.-Y. and Cho, Y. (1998) The crystal structure of flap endonuclease-1 from Methanococcus jannaschii. *Nat. Struct. Mol. Biol.*, **5**, 707-713.

...12. Shah, S., Dunten, P., Stiteler, A., Park, C.K. and Horton, N.C. (2015) Structure and specificity of FEN-1 from Methanopyrus kandleri. *Proteins*, **83**, 188-194.

...13. Exell, J.C., Thompson, M.J., Finger, L.D., Shaw, S.J., Debreczeni, J., Ward, T.A., McWhirter, C., Sioberg, C.L., Martinez Molina, D., Abbott, W.M. *et al.* (2016) Cellularly active N-hydroxyurea FEN1 inhibitors block substrate entry to the active site. *Nat. Chem. Biol.*, **12**, 815-821.

...14. Garforth, S.J., Ceska, T.A., Suck, D. and Sayers, J.R. (1999) Mutagenesis of conserved lysine residues in bacteriophage T5 5'-3' exonuclease suggests separate mechanisms of endoand exonucleolytic cleavage *Proc. Natl. Acad. Sci. USA*, **96**, 38-43.

...15. Shen, Y. and Bax, A. (2013) Protein backbone and sidechain torsion angles predicted from NMR chemical shifts using artificial neural networks. *J. Biomol. NMR*, **56**, 227-241.

...16. Lakomek, N.A., Ying, J. and Bax, A. (2012) Measurement of ^{15}N relaxation rates in perdeuterated proteins by TROSY-based methods. *J. Biomol. NMR*, **53**, 209-221.

...17. Bieri, M., d'Auvergne, E.J. and Gooley, P.R. (2011) relaxGUI: a new software for fast and simple NMR relaxation data analysis and calculation of ps-ns and mus motion of proteins. *J. Biomol. NMR*, **50**, 147-155.

...18. Yoshizawa, S., Kawai, G., Watanabe, K., Miura, K.-i. and Hirao, I. (1997) GNA Trinucleotide Loop Sequences Producing Extraordinarily Stable DNA Minihairpins. *Biochemistry*, **36**, 4761-4767.

...19. Nakano, M., Moody, E.M., Liang, J. and Bevilacqua, P.C. (2002) Selection for thermodynamically stable DNA tetraloops using temperature gradient gel electrophoresis reveals four motifs: d(cGNNAg), d(cGNABg),d(cCNNGg), and d(gCNNGc). *Biochemistry*, **41**, 14281-14292.

...20. Finger, L.D., Blanchard, M.S., Theimer, C.A., Sengerová, B., Singh, P., Chavez, V., Liu, F., Grasby, J.A. and Shen, B. (2009) The 3'-flap pocket of human flap endonuclease 1 is critical for substrate binding and catalysis. *J. Biol. Chem.*, **284**, 22184-22194.

...21. Brouwer, A.C. and Kirsch, J.F. (1982) Investigation of diffusion-limited rates of chymotrypsin reactions by viscosity variation. *Biochemistry*, **21**, 1302-1307.

...22. Sengerova, B., Tomlinson, C., Atack, J.M., Williams, R., Sayers, J.R., Williams, N.H. and Grasby, J.A. (2010) Bronsted analysis and rate-limiting steps for the T5 flap endonuclease catalyzed hydrolysis of exonucleolytic substrates. *Biochemistry*, **49**, 8085-8093.
...23. Rogers, J.M., Steward, A. and Clarke, J. (2013) Folding and binding of an intrinsically disordered protein: fast, but not 'diffusion-limited'. *J. Am. Chem. Soc.*, **135**, 1415-1422.
...24. Sekhar, A., Latham, M.P., Vallurupalli, P. and Kay, L.E. (2014) Viscosity-dependent kinetics of protein conformational exchange: microviscosity effects and the need for a small viscogen. *J Phys Chem B*, **118**, 4546-4551.
...25. Rauscher, A., Derenyi, I., Graf, L. and Malnasi-Csizmadia, A. (2013) Internal friction in enzyme reactions. *IUBMB Life*, **65**, 35-42.
...26. Rashid, F., Harris, P.D., Zaher, M.S., Sobhy, M.A., Joudeh, L.I., Yan, C., Piwonski, H., Tsutakawa, S.E., Ivanov, I., Tainer, J.A. *et al.* (2017) Single-molecule FRET unveils induced-fit mechanism for substrate selectivity in flap endonuclease 1. *eLife*, **6**, e21884.