Supplementary Information

Human Endonuclease III/NTH1: Focusing on the [4Fe-4S] cluster and the N-terminal domain

Elin Moe, Célia M. Silveira, Lidia Zuccarello, Filipe Rollo, Meike Stelter, Salvatore De Bonis, Catharina Kulka-Peschke, Sagie Katz, Peter Hildebrandt, Ingo Zebger, Joanna Timmins and Smilja Todorovic

Materials and Methods

Protein expression and purification

Full-length hNTH1 was expressed and purified as described previously.[1] The gene encoding the N-terminally truncated hNTH1 (residues 90-312, hNTH1Δ89) was PCR amplified from the plasmid containing the synthetic gene encoding the full-length enzyme (pET21a_hNTH1) and was cloned by restriction digestion using the NcoI and XhoI sites into pET21d for expression with a C-terminal His-tag. The conditions for the expression and purification of hNTH1Δ89 were the same as for hNTH1, however, buffers did not contain triton X-100, EDTA, β-Mercaptothanol and glycerol. The yield of the truncated protein was considerably higher than for the full-length and resulted in approximately 5 mg enzyme per liter of culture.

Raman spectroscopy

Resonance Raman (RR) spectra were acquired with a Raman spectrometer (Horiba LabRam HR-800, Edison, NJ, USA) equipped with a 1200 lines/mm grating and a liquid-nitrogen-cooled CCD detector. An Olympus 20x objective was used for laser focusing onto the sample and light collection in the backscattering geometry. Spectra were measured using 458 nm line from an argon ion laser (Coherent Innova 70, Santa Clara, CA, USA) at 77 K, from a 2 μL aliquote of the sample (0.3 – 0.5 mM) placed in a microscope stage (Linkham THMS 600, Tadworth, UK). Experiments were performed with laser power of 1.8 mW and accumulation time of 120 s. Up to 16 spectra were co-added in each measurement to improve the signal-to-noise ratio (S/N). All spectra were subjected to polynomial baseline subtraction; the positions and widths of Raman bands were determined by component analysis as described previously.[2]

SEIRA spectroscopy - immobilization

Surface enhanced infrared absorption (SEIRA) measurements were performed with a Kretschmann attenuated total reflection (ATR)-type configuration using a silicon crystal coated with a nanostructured gold film prepared via electroless deposition, as described previously.[3]

Prior to surface modification, the gold film electrode was cleaned electrochemically by cycling the potential from 0 V to +1.6 V in 0.1 M H_2SO_4 at a scan rate of 50 mV/s. The electrode was functionalized with mercaptoundecanoic acid (MUA), double-stranded DNA containing an oxidized pyrimidine, thymine glycol (Tg) (damaged DNA), with the following forward and reverse sequence: 5'-(5'thiol)AGTACGGTCATCGCG-3' and 5'-CGCGATGAC**Tg**GTACT-3´ or undamaged double-stranded DNA as described previously.[3] For functionalization with MUA (Au-MUA), the surface of the electrode was covered with 1 mL of 1 mM MUA dissolved in ethanol for around 16 h at room temperature. Functionalization with undamaged DNA- and damaged DNA-terminated SAMs required two steps. First, 100 μM undamaged DNA or damaged DNA in 50 mM Tris-HCl pH 8.0, 50 mM NaCl buffer was casted on top of the electrode and left for around 16 h at 4 °C. Then, the surface was passivated with 10 mM 6-mercaptohexanol in buffer for 30 minutes. The resulting self-assembled monolayers (SAM) were washed with 50 mM Tris-HCl pH 8.0, 50 mM NaCl prior to enzyme immobilization. Note that the concentration of DNA SAMs is one order of magnitude lower than MUA SAMs resulting in less compact films. This avoids repulsion of the DNA helices that could hamper protein attachment.[4] SAM-functionalized electrodes were incubated for 45 – 60 minutes with 0.2 – 1 μM hNTH1 or 0.48 μM hNTH1Δ89 in 50 mM Tris-HCl pH 8.0, 50 mM NaCl. Subsequently, the electrode was rinsed with buffer to remove unbound protein and 5 mL of fresh buffer was added to the SEIRA cell and used for further experiments. Immobilization was monitored by SEIRA spectroscopy using a Bruker IFS 66v S FTIR-spectrometer (Billerica, MA, USA) equipped with a liquid nitrogen cooled mercury cadmium telluride (MCT) detector. IR single channel spectra were collected at 10 °C while purging the sample compartment with nitrogen, with 2.5 mm aperture, at 20 kHz scanning velocity and 4 cm^{-1} resolution. Each spectrum is the average of 400 co-added scans. Data treatment was performed with the OPUS software version 7.5 from Bruker: the corresponding single channel of buffer was used as reference spectrum for calculating the enzymes' absorption spectra, which were subsequently baseline corrected.

SEIRA - spectroelectrochemistry

Experiments were performed using a three-electrode SEIRA spectroelectrochemical cell setup composed of Ag/AgCl (3 M, KCl) reference electrode, a platinum mesh counter electrode, and an electroless-deposited, nanostructured gold film as working electrode. The electrolyte solutions 50 mM HEPES pH 8, 50 mM K_2SO_4 (hNTH1 on MUA, hNTH1Δ89 on MUA, damaged and undamaged DNA) or 50 mM Tris-HCl pH 8, 50 mM NaCl (hNTH1 on damaged andundamaged DNA) were thoroughly purged with argon before the experiments. An EcoChemie potentiostat, model μAutolab, controlled with GPES 4.9 software (Metrohm) was used for the cyclic voltammetry measurements. Electron transfer rate constant was determined from the analysis of

scan rate dependent measurements (5-500 mV/s range) using the Laviron (m-function) method.[5] All potentials are quoted against the standard hydrogen electrode (+0.210 V *vs* Ag/AgCl).

Size-exclusion chromatography (SEC) coupled to small angle X-ray scattering (SAXS)

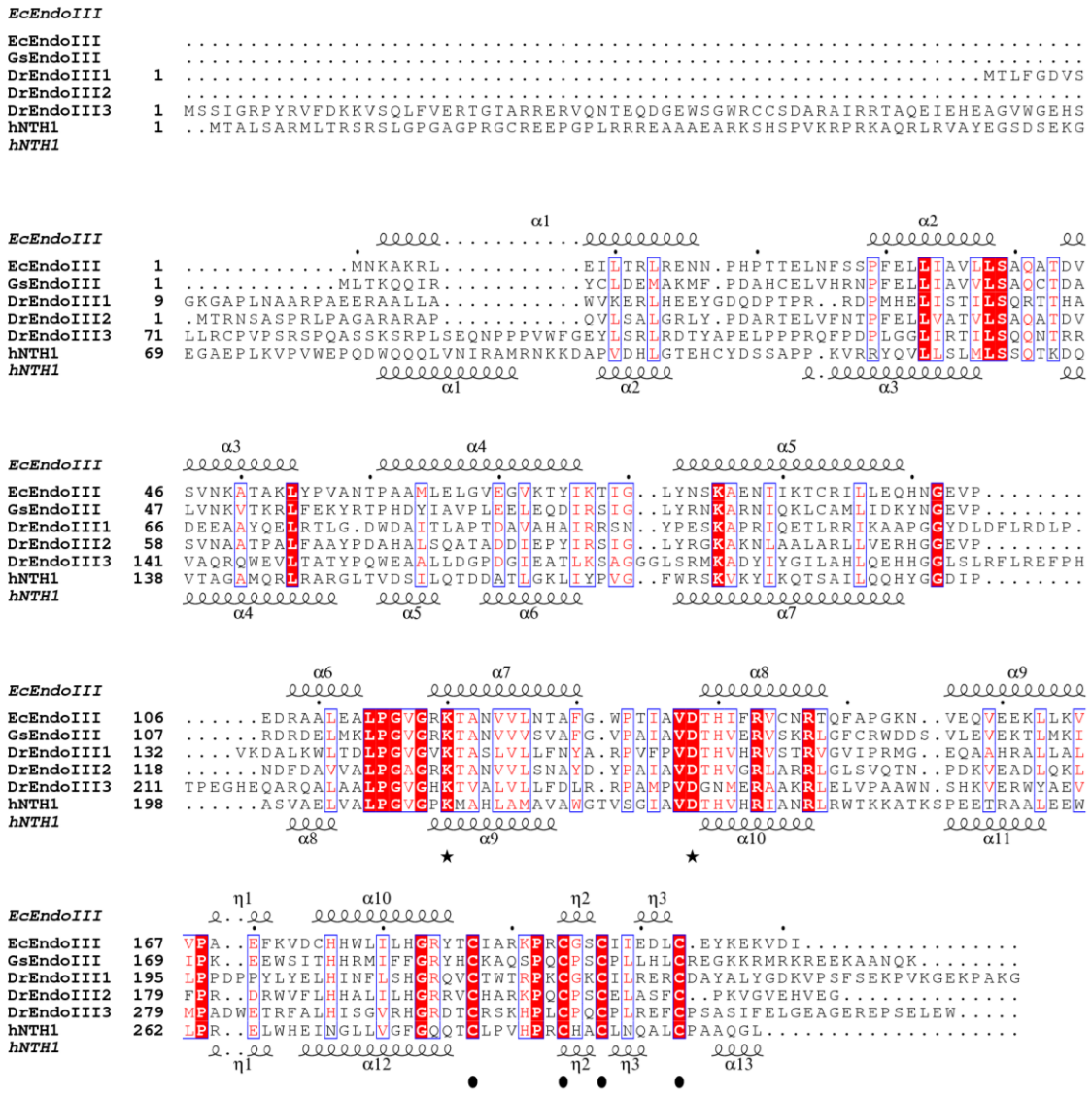
Two DNA substrates were used for SAXS measurements: (i) a 16-mer dsDNA composed of 5' CCTGTCCAXGTCTCCG-3' and 5'-ACGGAGACGTGGACAGGT-3´, and (ii) a 11-mer dsDNA composed of 5'-TGTCCAXGTCT-3' and 5'-AGACGTGGACA-3´, where X in each case is a tetrahydrofuran (THF) moiety, a stable mimic of an abasic site. 50 µL samples of hNTH1 or hNTH1Δ89 at protein concentrations of 4 to 6 mg/mL alone or in complex with a slight excess (1:1.1 molar ratio) of 16-mer (for hNTH1) or 11-mer (for hNTH1Δ89) dsDNA were injected onto a Superdex 75 5/150 GL (GE Healthcare column) connected to an on-line HPLC system (ViscotekGPCmax, Malvern Instruments) installed on the BM29 BioSAXS beamline at ESRF, Grenoble. Prior to injecting the samples, the column was equilibrated in 20 mM Tris-HCl pH 8, 100 mM NaCl, 1 mM TCEP, 5% glycerol, 0.2 mM EDTA, 0.005% Triton X-100 at a flow rate of 1 mL/min for 2 hours. Subsequently, the flow rate was set to 0.2 mL/min for the online SAXS experiments, which were performed at room temperature. In between each run, the column was equilibrated with 1.5 column volumes of buffer and the baseline was monitored. All data were collected at a wavelength of 0.9919 Å using a sample-to-detector (PILATUS1M, DECTRIS) distance of 2.81 m corresponding to an s range of 0.08–4.5 nm^{-1}. 1000 frames (1 frame/s) were collected for each run. Initial data processing was performed automatically using the EDNA pipeline[6], generating radially integrated, calibrated and normalized one-dimensional profiles for each frame. All frames were compared with the initial frame and matching frames (120) prior to the void volume of the column were merged to create the reference buffer, which was then subtracted from the subsequent frames. For each run, frames with a consistent Rg from the peak scattering intensity (typically 30 frames) were merged to yield a single averaged frame corresponding to the scattering of an individual SEC purified species. Data reduction and parameter extraction was performed using PRIMUS.[7] Evaluation of the real-space distance distribution function, *P(r)*, was performed using GNOM.[8] Subsequent fitting to theoretical SAXS curves derived from coordinate files was performed using CRYSOL.[9] Selected coordinate files were either taken from the PDB database (PDBs 7RDS and 7RDT) for the truncated form or generated by the artificial intelligence program, AlphaFold[10], in the case of the full-length protein.

Structure prediction using AlphaFold

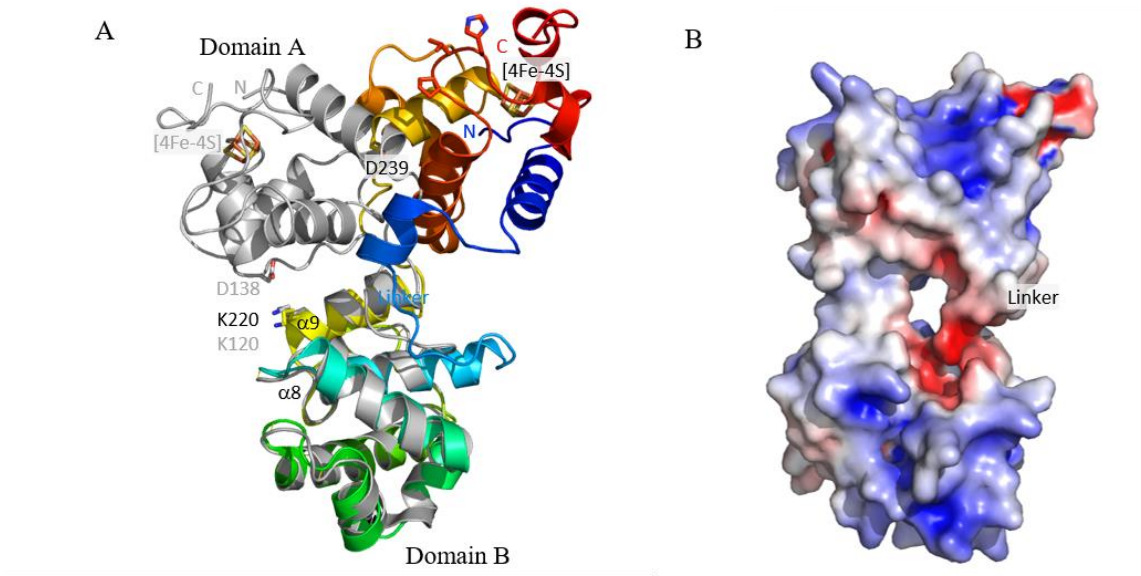
Three-dimensional structures of full-length hNTH1 were generated by AlphaFold[10] via the Colaboratory service from Google Research (https://colab.research.google.com/github/sokrypton/ColabFold/blob/main/beta/AlphaFold2_advanced.ipynb). When using all available templates in the PDB database (SI Fig. S8), AlphaFold predicted five very similar high-

confidence models of hNTH1 that all adopt a ‘closed’ conformation. The high-sequence coverage and large number of available sequences and structures of related endonuclease III proteins in the database allowed AlphaFold2 to produce highly accurate models of the catalytic domain of hNTH1 (residues 90-326) with most residues in this region exhibiting high-confidence scores with predicted lDDT (local Distance Difference Test) values >90 (SI Fig. S8B). In contrast, the NTD was consistently predicted to adopt an extended and unfolded structure but was found to adopt different conformations in each of the five models and the predicted lDDT per residue in this region was between 20 and 40. We thus relied on the fitting of the models to the SAXS data to discriminate between the proposed models. To generate a hNTH1 model in an ‘open’ state, we defined the ‘open’ crystal structure of hNTH1Δ63 as the sole template for AlphaFold prediction using the ‘custom’ template option. All other parameters were the same as for the initial prediction. In this second run, AlphaFold generated five models, three of which were of very poor quality and were thus discarded (SI Fig. S9). The two remaining models, ranked respectively 1 and 2, displayed an ‘open’ conformation and high-confidence lDDT scores for the catalytic domain, with the exception of residues 104-124 corresponding to the linker region connecting domains A and B that displayed reduced lDDT values between 55 and 75. In these two models, as in the five predicted ‘closed’ models, the NTDs were predicted to adopt extended conformations with lDDT values around 40. Plots of predicted aligned error (PAE) and illustrations of AlphaFold models colored by lDDT were prepared in ChimeraX.[11]

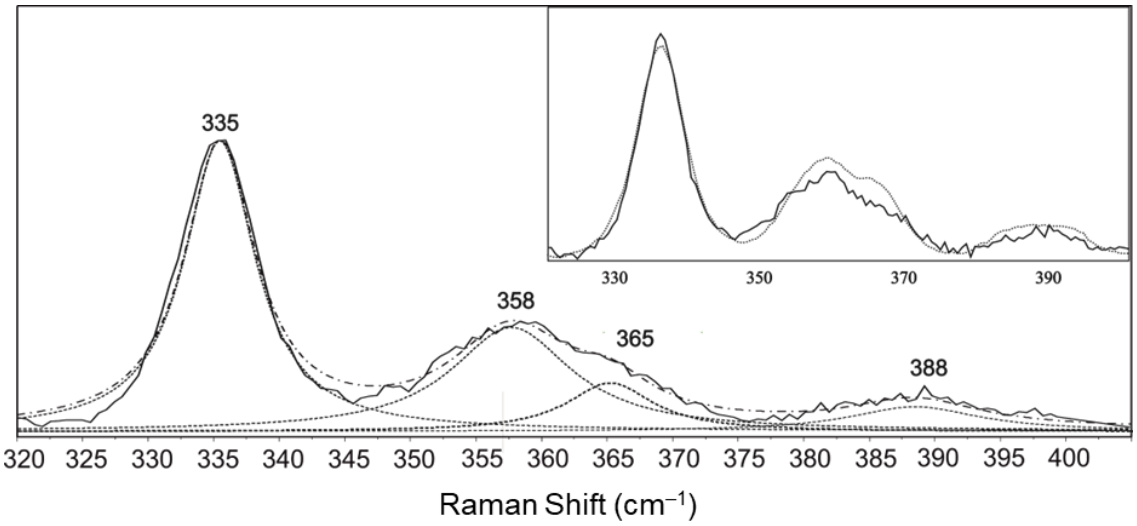
Figures and Tables



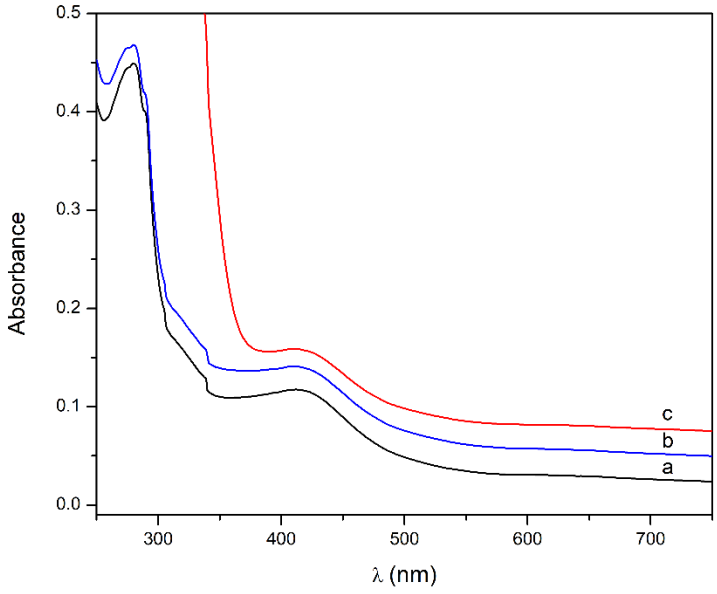
SI Fig. S1 Sequence alignment between structurally determined EcEndoIII (PDB, 2ABK), GsEndoIII (PDB, 1ORN), DrEndoIII1 (PDB, 4UOB), 2 and 3 (PDB, 4UNF), and hNTH1 (PDB, 7RDS). The black stars mark the catalytic residues, the black filled circles mark the [4Fe-4S] cluster coordinating cysteines.



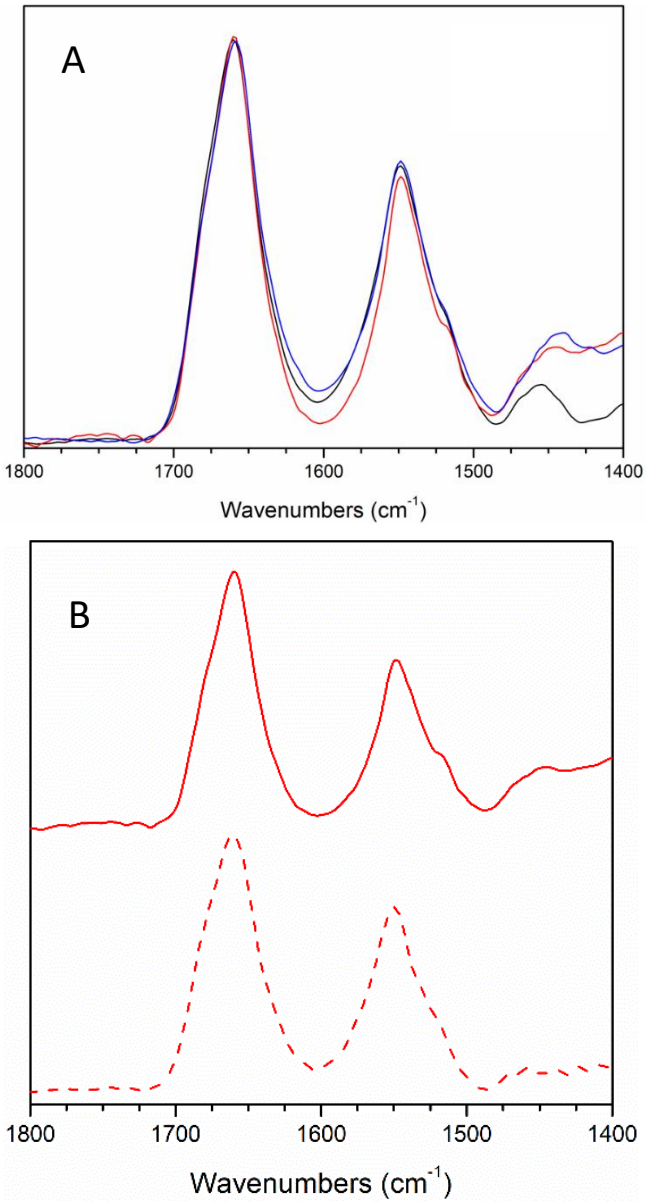
SI Fig. S2 Structural comparison of hNTH1 and EcEndoIII. (A) Overlay of hNTH1Δ63 (rainbow colored – PDB, 7RDS) and *E. coli* EndoIII grey (PDB, 2ABK) – the molecules are superimposed using domain B as a reference. hNTH1Δ63 residues are indicated in black, while those of EcEndoIII are in light grey. The [4Fe-4S] cluster is found in domain A in both cases, but due to the 'open' conformation of hNTH1Δ63, domain A and its associated cluster do not overlay with the bacterial structure. (B) Visualization of the electrostatic surface potential of hNTH1Δ63, where blue denotes positively charged potential (+5keV) and red negatively charged potential (-5keV). hNTH1Δ63 exhibits positively charged patches i) close to the [4Fe-4S] cluster in domain A and ii) in the lower part of domain B. Close to the DNA binding site (below the linker) one can observe a negatively charged area.



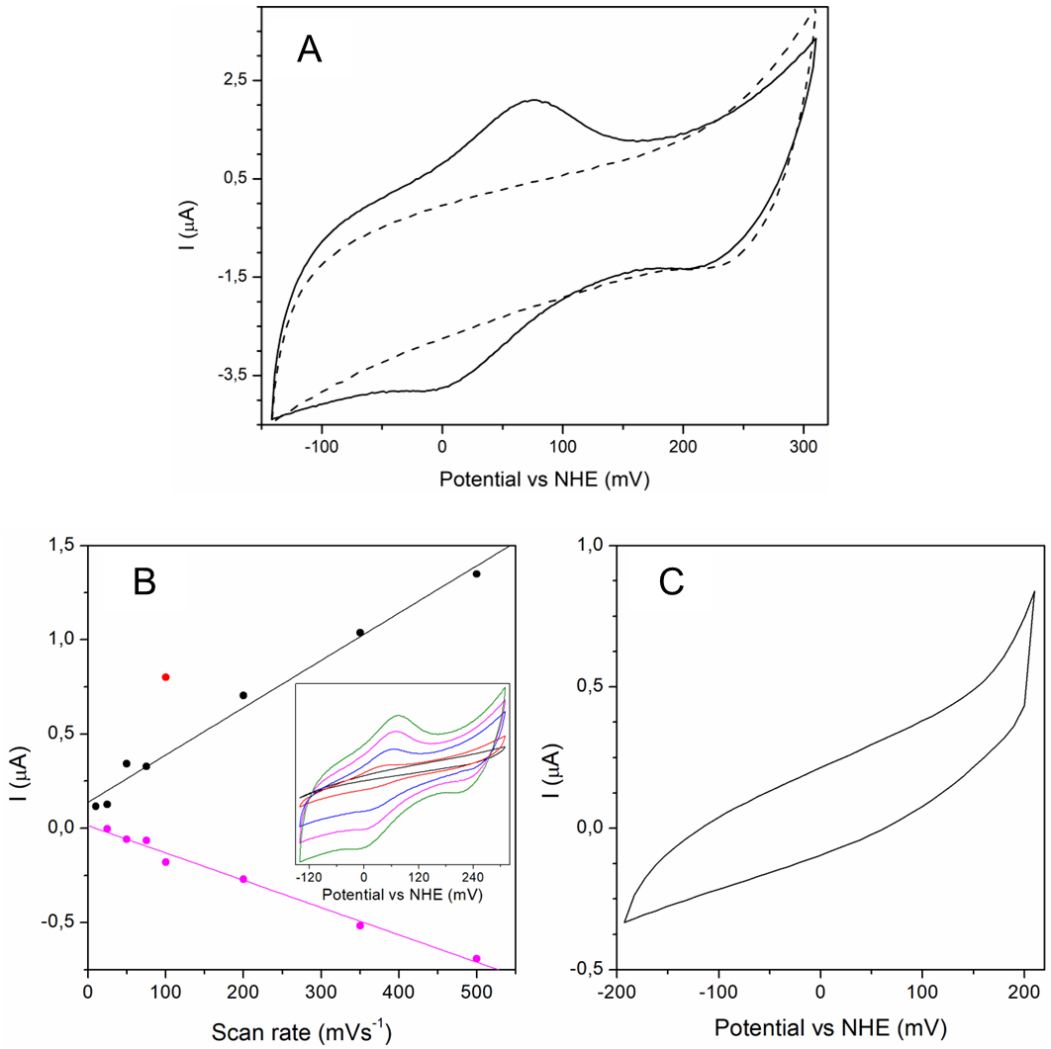
SI Fig. S3 RR spectra of hNTH1. Experimental spectrum (solid line), component spectra (dashed lines) and overall fit (dash-dotted line). Spectrum was recorded with 458 nm excitation and 1.8 mW laser power at 77 K. **Inset:** Experimental RR spectra of DrEndoIII3 (dotted line) and hNTH1 (solid line).



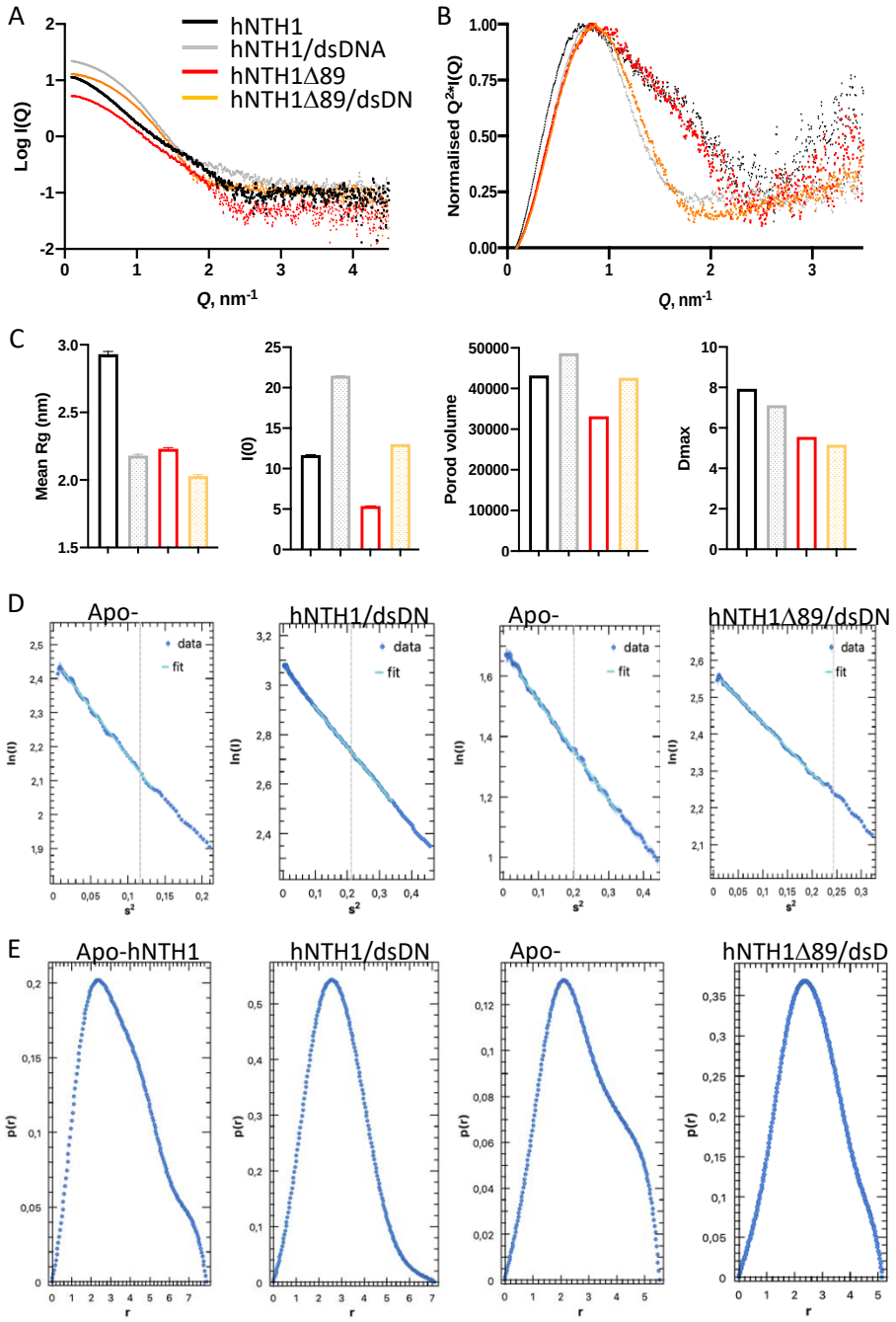
SI Fig. S4 UV-Vis spectra of hNTH1 measured in 50 mM Tris-HCl buffer pH 8 with 50 mM NaCl (a), in the presence of potassium ferricyanide (b) or sodium dithionite (c).



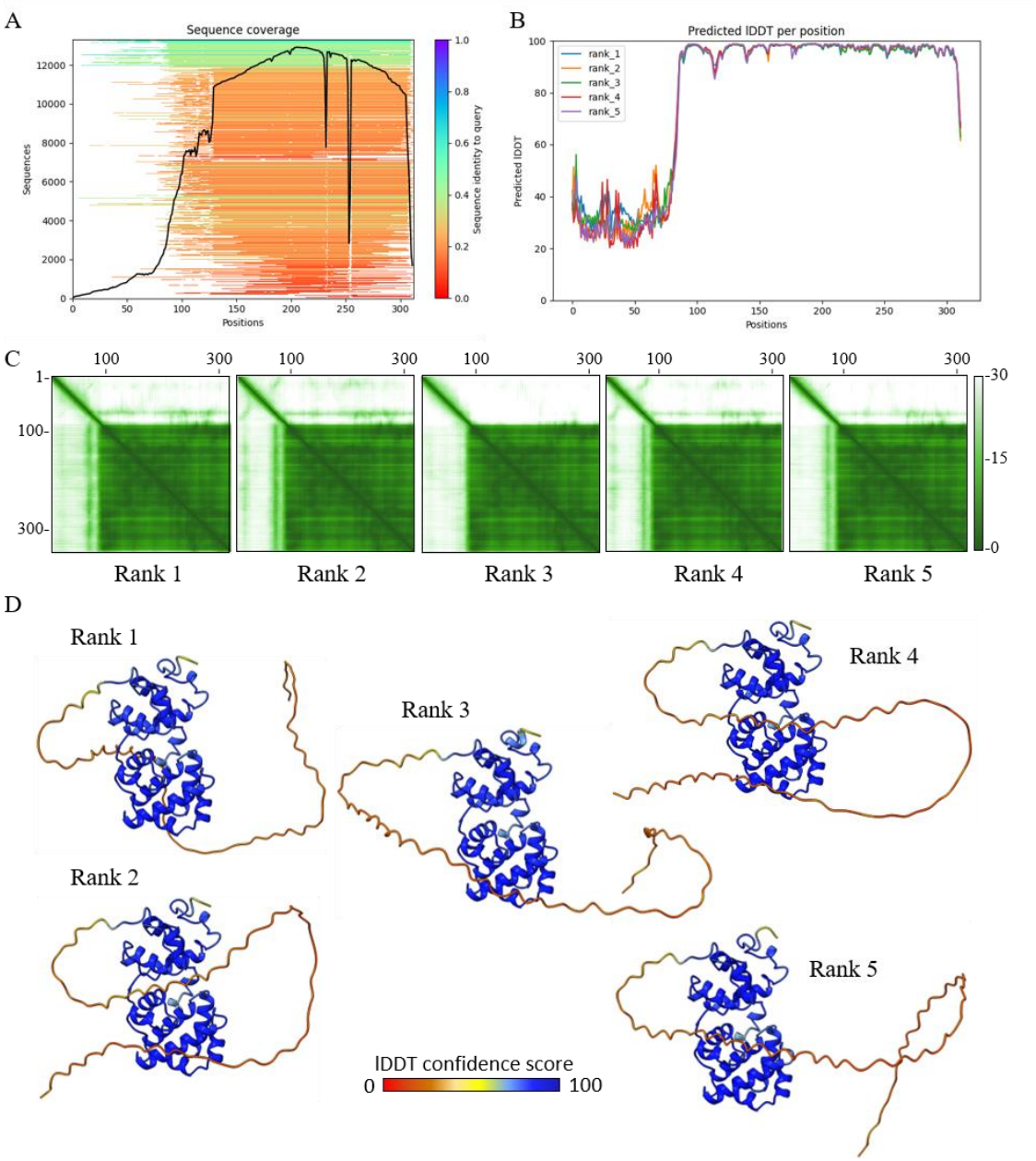
SI Fig. S5 SEIRA spectra of hNTH1 and hNTH1Δ89. (A) Overlaid spectra of hNTH1Δ89 immobilized on MUA (black), damaged DNA- (red) and undamaged DNA-terminated SAMs (blue), displayed after normalizing to amide I intensity. (B) Spectra of hNTH1Δ89 (red, solid line) and hNTH1 (red, dashed line) adsorbed on damaged DNA-terminated SAMs; spectra are normalized to amide I intensity.



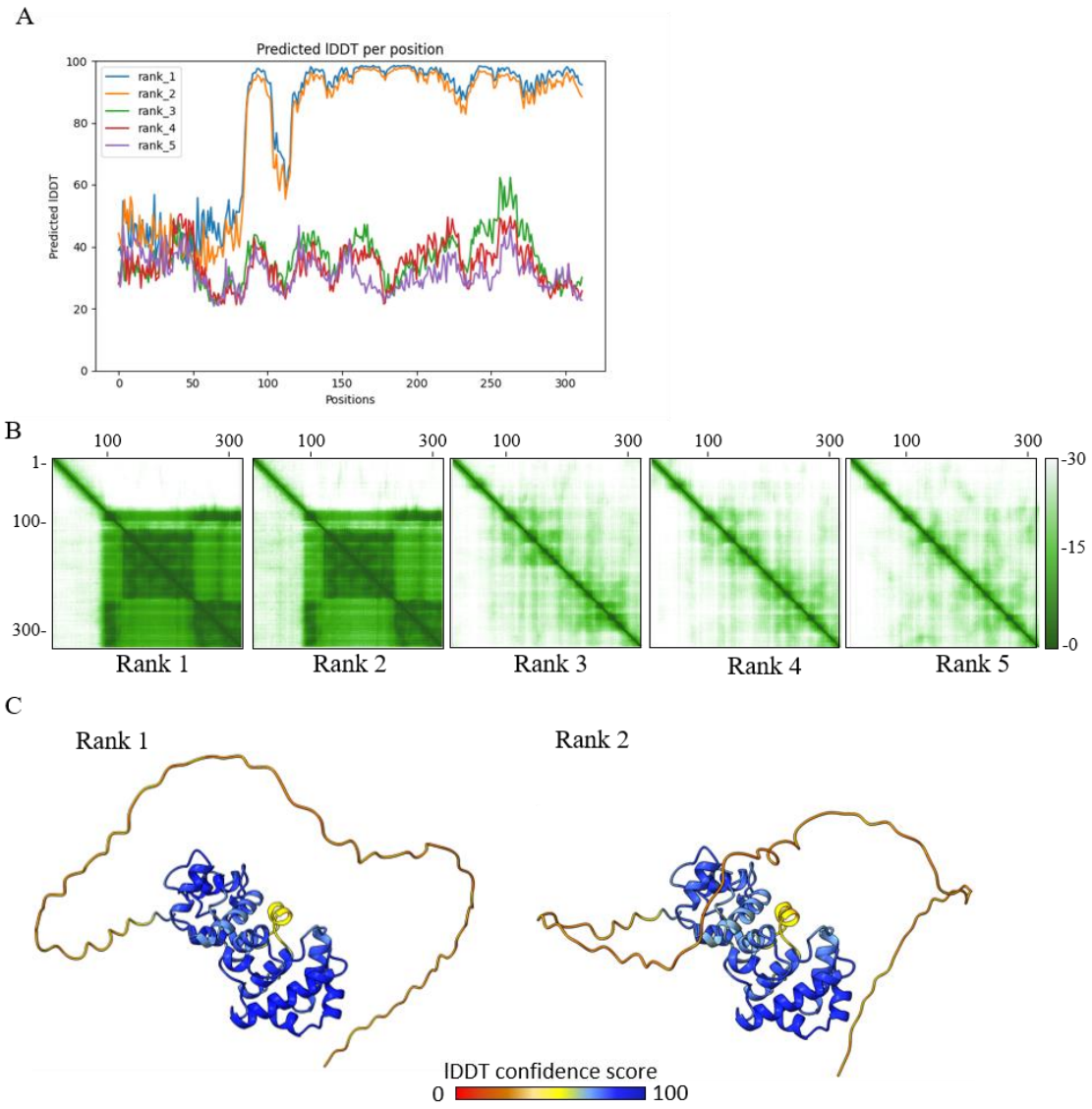
SI Fig. S6 Electrochemical characterization of hNTH1Δ89 (A and B) and hNTH1 (C) on MUA coated electrodes. (A) Cyclic voltammograms (CVs) of hNTH1Δ89 on MUA (solid line) and control without enzyme (dotted line); scan rate of 500 mV/s. (B) Scan rate linear dependence of cathodic (pink) and anodic (black) current. Red point not considered for linear fitting. **Inset:** effect of selected scan rates (black: 10 mV/s; red: 50 mV/s; blue: 200 mV/s; pink: 350 mV/s; green: 500 mV/s) on CV. (C) CV at 100 mV/s of hNTH1 on MUA.



SI Fig. S7 SAXS curves and data processing and analysis plots. (A) 1D SAXS scattering curve of hNTH1 (black and grey) and hNTH1Δ89 (red and orange) in absence and presence of DNA. (B) Kratky plots of the SAXS data. The plots are colored as in (A). (C) Plots illustrating the mean Rg, I(0), Porod volume and Dmax values derived from the SAXS data. The plots are colored as in (A). (D) Illustration of the Guinier plots used to derive the Rg values for hNTH1, hNTH1/DNA, hNTH1Δ89 and hNTH1Δ89-DNA. (E) Pair distance distribution function, P(R), of hNTH1, hNTH1/DNA, hNTH1Δ89 and hNTH1Δ89-DNA.



SI Fig. S8 AlphaFold prediction of 'closed' hNTH1 structures. (A) Sequence coverage of the query sequence (hNTH1). (B)-(C) Plots illustrating the predicted per residue confidence score, lDDT (B) and the predicted aligned error (PAE) for every pair of residues (C) for each of the generated models ranked 1 to 5. (D) Ribbon illustrations of the five predicted models of 'closed' hNTH1 colored according to their lDDT confidence score (blue, high-confidence to red, low confidence).



SI Fig. S9 AlphaFold prediction of 'open' hNTH1 structures. (A and B) Plots illustrating the predicted per residue confidence score, lDDT (A) and the predicted aligned error (PAE) for every pair of residues (B) for each of the generated models ranked 1 to 5. (C) Ribbon illustrations of the two high-confidence models of 'open' hNTH1 colored according to their lDDT confidence score (blue, high-confidence to red, low confidence).

SI Table S1 SAXS data processing parameters for determination of Rg values.

Sample	Averaged buffer frames	Merged sample frames	Rg range (data points used for Guinier fit)	Mean Rg (nm)
hNTH1	200-320	365-392	7-60	2.93 ± 0.02
hNTH1-dsDNA-THF	200-320	395-422	51-109	2.18 ± 0.01
hNTHΔ89	200-320	410-440	28-101	2.23 ± 0.01
hNTHΔ89-dsDNA-THF	200-320	385-415	8-81	2.03 ± 0.01

References

1 A. Sarre, M. Stelter, F. Rollo, S. De Bonis, A. Seck, C. Hognon, J. L. Ravanat, A. Monari, F. Dehez, E. Moe and J. Timmins, *DNA Repair (Amst).*, 2019, **78**, 45–59.

2 S. Todorovic, S. S. Leal, C. A. Salgueiro, I. Zebger, P. Hildebrant, D. H. Mugida and C. M. Gomez, *Biochemistry*, 2007, **46**, 10733–10738.

3 E. Moe, M. Sezer, P. Hildebrandt and S. Todorovic, *Chem Commun*, 2015, **51**, 3255–3257.

4 E. M. Boon, J. E. Salas J. K. and Barton, *Nat. Biotechnol.*, 2002, **20**, 282−286.

5 C. Léger and P. Bertrand, *Chem. Rev.*, 2008, 108, 2379–2438.

6 M. F. Incardona, G. P. Bourenkov, K. Levik, R. A. Pieritz, A. N. Popov and O. Svensson, in *Journal of Synchrotron Radiation*, 2009, vol. 16, pp. 872–879.

7 P. V. Konarev, V. V. Volkov, A. V. Sokolova, M. H. J. Koch and D. I. Svergun, *J. Appl. Crystallogr.*, 2003, **36**, 1277–1282.

8 A. V. Semenyuk and D. I. Svergun, *J. Appl. Crystallogr.*, 1991, **24**, 537–540.

9 D. Svergun, C. Barberato and M. H. Koch, *J. Appl. Crystallogr.*, 1995, **28**, 768–773.

10 J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, A. Bridgland, C. Meyer, S. A. A. Kohl, A. J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A. W. Senior, K. Kavukcuoglu, P. Kohli and D. Hassabis, *Nature*, 2021, **596**, 583–589.

11 E. F. Pettersen, T. D. Goddard, C. C. Huang, E. C. Meng, G. S. Couch, T. I. Croll, J. H. Morris and T. E. Ferrin, *Protein Sci.*, 2021, **30**, 70–82.