Data in Brief 40 (2022) 107760

Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Data Article

Dataset for dynamics and conformational changes in human NEIL2 protein analyzed by integrative structural biology approach

Polina V. Zhdanova[a,b], Alexander A. Ishchenko[c,d], Alexander A. Chernonosov[a], Dmitry O. Zharkov[a,b], Vladimir V. Koval[a,b,*]

[a] Institute of Chemical Biology and Fundamental Medicine SB RAS, Novosibisk, Russia
[b] Novosibirsk State University, Novosibisk, Russia
[c] Groupe «Réparation de l'ADN», Equipe Labellisée par la Ligue Nationale contre le Cancer, CNRS UMR 8200, Univ. Paris-Sud, Université Paris-Saclay, Villejuif F-94805, France
[d] Gustave Roussy, Université Paris-Saclay, Villejuif F-94805, France

ARTICLE INFO

Article history:
Received 4 November 2021
Revised 10 December 2021
Accepted 15 December 2021
Available online 25 December 2021

Keywords:
DNA damage
DNA repair
Base excision repair
DNA glycosylases
NEIL2
HDX-MS
Molecular dynamics
Structural dynamics

ABSTRACT

This work presents new data on human endonuclease VIII-like 2 protein (hNEIL2), a part of DNA glycosylases of the helix–two-turn–helix structural superfamily. While X-ray structure of oNEIL2 (opossum Monodelphis) was resolved partially [1], the structure of hNEIL2 has not yet been determined. This data article describes a powerful combination of hydrogen-deuterium exchange mass spectrometry, homology modeling, and molecular dynamics simulations for protein conformational dynamics analysis.

The data supplied in this work are related to the research article entitled "Dynamics and Conformational Changes in Human NEIL2 DNA Glycosylase Analyzed by Hydrogen/Deuterium Exchange Mass Spectrometry".

Abbreviations: hNEIL2, human endonuclease VIII-like 2 protein; HDX-MS, hydrogen-deuterium exchange mass spectrometry; MD, molecular dynamics; CD, circular dichroism; LC-MS, liquid chromatography–mass spectrometry; ESI, electrospray ionization; SDS-PAGE, sodium dodecyl sulphate–polyacrylamide gel electrophoresis; PDB, Protein Data Bank; MDTRA, Molecular Dynamics Trajectory Reader & Analyzer; DTT, (2S,3S)-1,4-Bis(sulfanyl)butane-2,3-diol; HEPES, 2-[4-(2-Hydroxyethyl)piperazin-1-yl]ethane-1-sulfonic acid; TCEP, 3,3',3"-Phosphanetriyltripropanoic acid; RMSD, root-mean-square deviation; LB, Lysogeny broth; IPTG, Propan-2-yl 1-thio-β-D-galactopyranoside.

DOI of original article: 10.1016/j.jmb.2021.167334

* Corresponding author.
E-mail address: koval@niboch.nsc.ru (V.V. Koval).

https://doi.org/10.1016/j.dib.2021.107760

Specifications Table

Subject	Structural biology
Specific subject area	Structural enzymology, enzymatic catalysis, protein folding/unfolding
Type of data	Text file, Excel file figure, MD movies, raw spectra
How the data were acquired	Data was obtained using PAGE assay, hydrogen-deuterium exchange mass spectrometry, computer modeling
Data format	Raw data, Analyzed data
Description of data collection	HDX-MS data were collected using nanoACQUITY HDX manager followed by Synapt G2-Si HDMS mass spectrometer; Peptide dataset we analyzed by ProteinLynx Global Server v3.0.1 and DynamX v3.0; CD spectra were recorded on a J600 CD spectrometer; We used the AMBER 20 to perform MD simulation; USCF Chimera and PyMOL were used for data visualization
Data source location	Institute of Chemical Biology and Fundamental Medicine of SiberianBranch of the Russian Academy of Sciences, 8 Lavrentyev Ave.,Novosibirsk, 630090, Russian Federation
Data accessibility	Repository name: Mendeley DataData identification number: http://dx.doi.org/10.17632/hpr8fr7ck6.1 Direct link to the dataset: https://data.mendeley.com/datasets/hpr8fr7ck6/draft?a=fa1b452c-683c-44b1-a454-a659c8e8701e (temporary link)
Related research article	P.V. Zhdanova, A.A. Ishchenko, A.A. Chernonosov, D.O. Zharkov, V.V. Koval, Dynamics and Conformational Changes in Human NEIL2 DNA Glycosylase Analyzed by Hydrogen/Deuterium Exchange Mass Spectrometry, JMB, doi: 10.1016/j.jmb.2021.167334

Value of the Data

- The HDX-MS and MD simulation data provide information for the dynamical structure of hNEIL2 protein during substrate recognition and substrate processing. The data obtained show that free hNEIL2 is mainly in the open conformation.
- The data can be a useful guide for further studies of the structural dynamic features of proteins and their complexes since the described approaches allow us to avoid crystallization of samples and observe the processes in solution.
- The researchers studying the processes occurring during the repair of damaged DNA bases, including conformational changes in enzymes during the recognition of substrates, interaction with them, and the stage of catalytic activity, can benefit from these data.
- The data can help to study the mechanism of recognizing damage by the hNEIL2 protein, especially in terms of the relationship of NEIL2 glycosylase with various types of cancer and age-related degenerative changes.

1. Data Description

Data reported here describe the structural features of hNEIL2 protein revealed from a combination of methods, such as Hydrogen-Deuterium Exchange Mass Spectrometry, molecular dynamics simulation, and CD spectrometry.

hNEIL2 protein was expressed and purified as described in the Experimental design, materials and methods section. The purification of the recombinant protein was performed in two steps. Firstly, we purified hNEIL2 with a terminus, 6X histidine tag using a HiTrap Chelating column

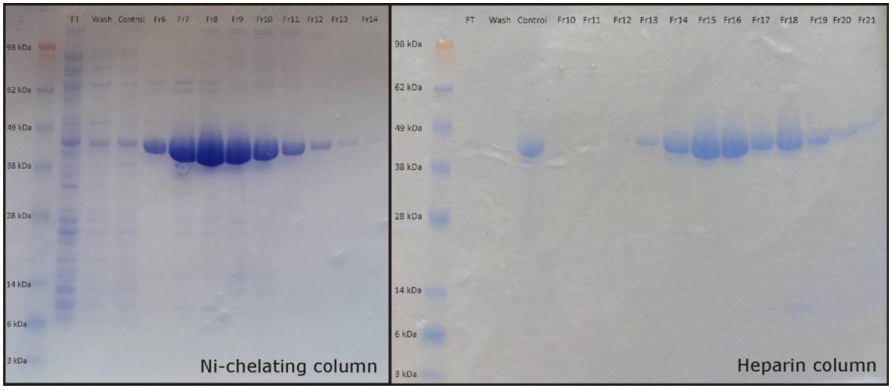

Fig. 1. The SDS–PAGE electrophoresis of the protein fractions after HiTrap™ Chelating column (A) and Heparin column (B) purification.

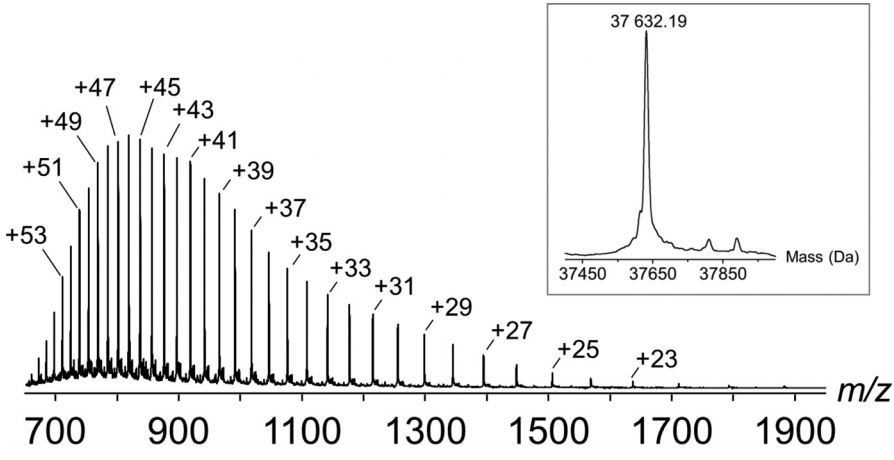

Fig. 2. Intact mass measurement of recombinant hNEIL2.

and analyzed collected fractions using SDS-PAGE (Fig. 1A). HiTrap Heparin column was used on the second purification step and the obtained fractions were analyzed by SDS-PAGE (Fig. 1B). Pure fractions contained hNEIL2 were stored at −80 °C (in buffer with 10% glycerol and 1 mM DTT).

Beforethe hydrogen-deuterium experiment, it was necessary to confirm recombinant hNEIL2 mass and its homogeneity using mass spectrometry analysis. Therefore, we determined the protein's molecular mass by LC-MS analysis as 37631.98 ± 0.11 Da (Fig. 2).

We used the characterized recombinant hNEIL2 in subsequent HDX-MS experiments. The HDX data summary is shown in Table 1. Full experiment procedures described in related paper [2].

We also have analyzed hNEIL2 unfolding after Trp fluorescence (Fig. 3) by native differential scanning fluorimetry using the Tycho NT.6 capillary and performed CD experiments in the temperature range of 20-80 °C, as described in [2].

One of the significant parts of the work was the computer modeling of the protein structure and dynamics. We used two hNEIL2 protein conformations in the MD simulations, open and close. Close protein state based on the oNEIL2 (PDB ID: 6VJI [1]) crystal structure, and open model based on the mvNEIL2/3 (PDB ID: 4MB7 [3]) structure. Residues 58–125 (the extended

Table 1
The summary of HDX-MS data.

Dataset	apo-protein
HDX reaction details	D2O buffer: 20 mM HEPES (pD 7.6), 150 mM NaCl, 1 mM TCEP prepared in 99.98% deuterium oxide; (the final deuterium content – 89%); RT; Quench buffer: 2.5% formic acid, 4 M urea; Quenched sample solution: pH ~2.5, final D2O/H2O ratio: ~0.3/~0.7.
HDX time course	10 s, 30 s, 1 min, 2 min, 5 min, 10 min, 30 min, 60 min, 120 min
Sequence coverage	96.7%
Replicates	3
Protein digestion details	10 pmols (50 µl) of labelled protein; 2 min at 100 µL/min solvent A (0.15% formic acid, pH 2.5); Immobilized pepsin column: 2.0 × 20 mm, 63 µl bed volume; 20 °C; Peptides trapped, concentrated and desalted at 0 °C, 100 µL/min solvent A; Trapping column: VanGuardTM BEH C18 pre-column, 1.7 µm, 2.1 × 5 mm.
Peptide analysis details	Column: ACQUITY UPLCTM Peptide BEH C18 analytical column, 1.7 µm, 1 × 10 mm; Gradient: 8-min linear gradient, 5%-30% solvent B (acetonitrile, 0.15% formic acid, pH 2.5) followed by a 2-min linear gradient, 30%-40% solvent B; 40 µl/min; 0 °C.
MS analysis details	Positive-ion and resolution mode; m/z range of 50–1950; Data-independent acquisition scheme (MS^E) MS^E trap collision energy ramps: 10-30 V, 15-35 V, 20-45 V, 10-45 V

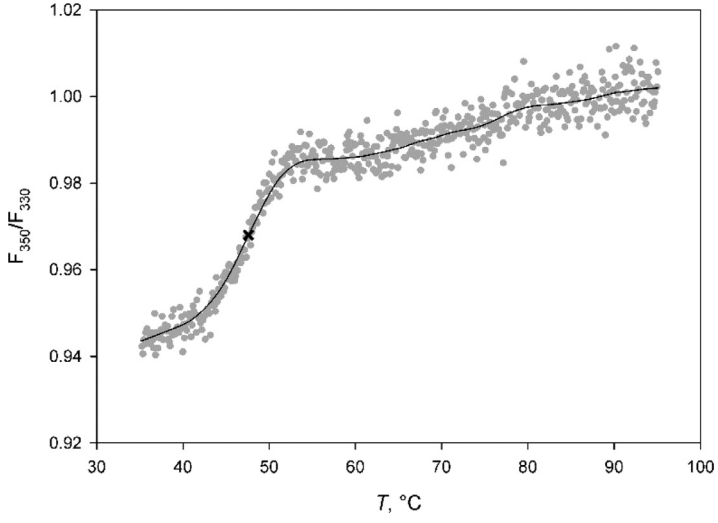

Fig. 3. Melting profile of hNEIL2 analyzed by Trp fluorescence. F_{350}/F_{330}, the ratio of emission at 350 nm and 330 nm. T_m is marked by X.

disordered loop between the β3 and β4 strands) were removed from the models, and Zn^{2+} ion was placed into the zinc finger motif prior to MD simulations. For both conformations of the protein 100 ns MD simulation was performed twice. The results of MD trajectory analysis are shown in Figs. 4 and 5. The RMSD (Fig. 4) for hNEIL2 was depicted to show the stability of the protein throughout the whole 100 ns MD trajectory. Fig. 5 shows the open-close diagram after the system reaches equilibrium (the last 20 ns of MD). The x-axis shows the difference in the distances between Cα atoms of Lys49 and Asn230 for open (pink) and close (red). Moreover, the

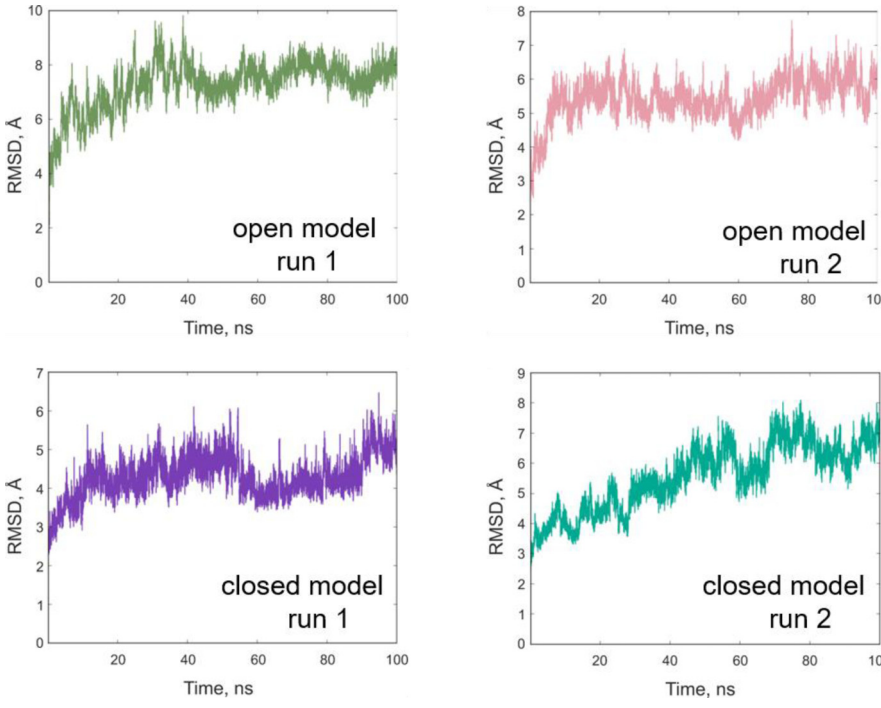

Fig. 4. RMSD fluctuations of hNEIL2 throughout the whole MD trajectory.

y-axis depicts the distance difference between Cα atoms of Glu2 and Arg309 for two models: gray for the open model and black for the close model of the protein.

In order to compare HDX and MD data to analyze the closed and open states of hNEIL2, we analyzed MD trajectories. We chose H-bond occupancy as the primary approach for detailed analysis. MD is a relevant instrument for monitoring individual H-bonds of backbone amides because of its ability to produce conformational ensembles that can be analyzed using statistical approaches. We used MDTRA to identify H-bonds [4]. Individual runs yielded remarkable inter-run correlations of intramolecular hydrogen bonds (see original article). The occupancy of each H-bond was averaged for the simulations of the two models, and those values showing the most significant distinction between the two states were chosen (Fig. 6). In this way, we obtain information about the regions that are subject to maximum rearrangement between the open and closed conformations of hNEIL2.

The initial data of experiments on circular dichroism, melting point determination, protein mass determination after isolation, and all the initial data on MD are presented in the repository (http://dx.doi.org/10.17632/hpr8fr7ck6.1). The MD movies and source of figures presented in this article also are available. For experiments on MD, the set of files in “pdb” format is available corresponding to the last 20 ns of each trajectory for both described models. Additionally, the data on MS after processing in PLGS and DynamX programs are shown, which are the set of peptides (peptide map) and its corresponding levels of exchange (see [2]).

The dataset presented in the repository consists of six folders. In the “Circular dichroism” folder one can find the eleven files with the initial data of the CD experiment. The “DynamX data” folder includes a “PepMap” subfolder and a “State.csv” file. The “State.csv” file contains the peptides and their uptake levels (the result of the HDX experiment). There are four subfolders inside “PepMap”, and they present processing the raw spectra using PLGS. “Neil2_PepMap.xlsx” is the resulting file of peptides used in the subsequent processing step. In “Figures,” there are all

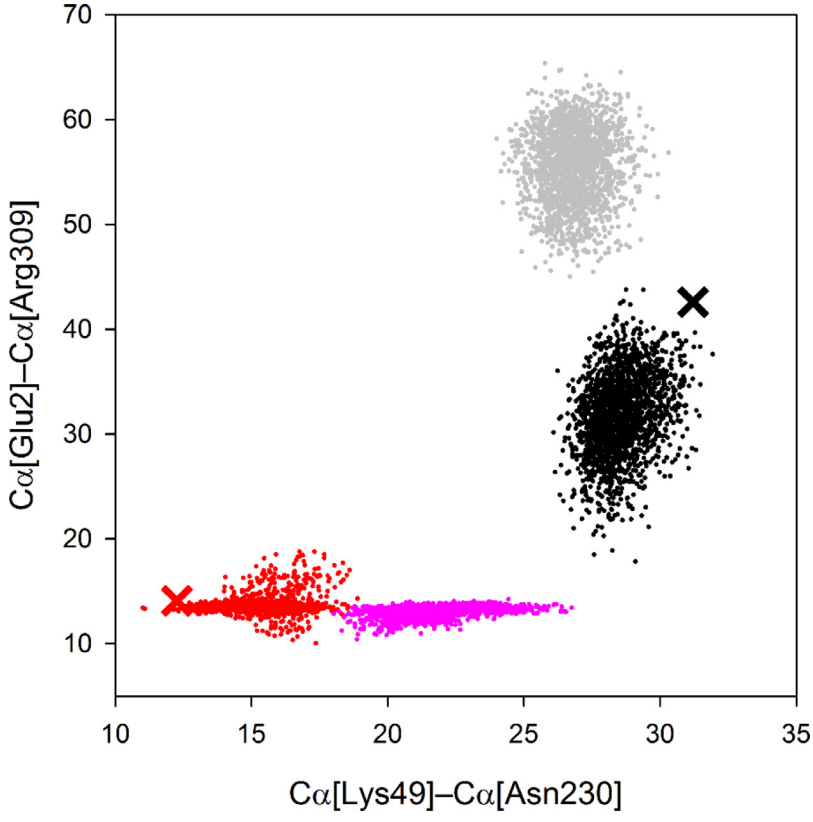

Fig. 5. Distances between Cα atoms of Lys49 and Asn230 and between Glu2 and Arg309 were observed during the last 20 ns of MD. X marks the distances in the static models that served as starting points for MD.

figures shown in this article. Content in "Molecular Dynamics" was divided into two subfolders: "MD movies" and "MD raw data." The first subfolder contains four MD movies with 100 ns trajectory dynamics. Another directory contains archives of both protein states, open and closed, with initial MD structures and scripts and protein structures in PDB format for the last 20 ns MD. The "MS C4 column" folder contains two raw MS spectra of the native protein referring to Fig. 2. The "Tycho" folder contains raw and processed melting point data.

2. Experimental Design, Materials and Methods

2.1. Buffers

Lysis buffer: 20 mM HEPES-KOH, pH 7.6, 150 mM NaCl, 1% Triton X-100, 20 mM imidazole, 5% Glycerol, 2 mM EDTA, 5 mM $MgCl_2$.

Protein purification via HiTrap^{TM} *Chelating column.*

Buffer A: 20 mM HEPES, pH 7.6, 500 mM NaCl, 30 mM imidazole, 0.025% NP-40.
Buffer B: 20 mM HEPES, pH 7.6, 500 mM NaCl, 500 mM imidazole, 0.025% NP-40.
Column was prepared by washing through with 5 mL H_2O, 5 mL EDTA 0.05 M, 5 mL H_2O, 2-4 mL $NiCl_2$ 0.1 M, 5 mL buffer A.

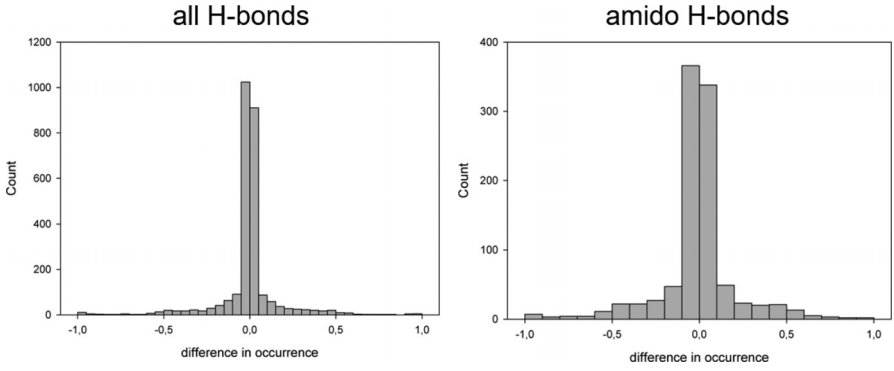

Fig. 6. Difference in the occurrence of H-bonds in the open model (positive values) and the closed model (negative values).

Protein purification via HiTrap^{TM} *Heparin column.*

Buffer A: 20 mM HEPES-KOH pH 7.6, 50 mM NaCl, 1 mM DTT.
Buffer B: 20 mM HEPES-KOH pH 7.6, 1M NaCl, 1 mM DTT.
Column was prepared by washing through with 5 mL H2O, 5 mL buffer A.

HDX MS experiments were performed in buffer containing 20 mM HEPES, 150 mM NaCl, 1 mM TCEP pH 7.6 ± 0.1.

D_2O containing buffer: 20 mM HEPES (pD 7.6 ± 0.1), 150 mM NaCl, 1 mM TCEP prepared in 99.98% deuterium oxide.

Quench buffer: 2.5% formic acid, 4 M urea.

Pepsin column wash solution: 1% formic acid, 5% acetonitrile, and 1.5 M guanidinium chloride (pH 1.7).

2.2. Expression and purification of human NEIL2 protein

We mixed 0.1 µg of E. coli expression plasmid for hNEIL2 (pET24a having C-terminal His6−tag) with 40 µL electrocompetent E. coli Rosetta2(DE3) and transformed using electroporation. The transformed cells were incubated at 37 °C overnight on the Petri dish containing LB agar with kanamycin (Kn). Then we picked a single colony into 20 mL LB, added 20 µL Kn and incubated the pre-culture at 37 °C overnight under constant shaking. After, 10 mL of the pre-culture cells and 200 µL kN were added to 190 mL LB. The cells were grown with vigorous shaking at 37 °C to OD_{600} ~0.7–0.8. 1 L of cell culture (100 mL inoculate OD600~0.7-0.8, 900 mL LB, Kn) was induced by adding IPTG to 0.5 mM, and the growth was continued at 15 °C overnight.

The cells were harvested by centrifugation, resuspended in the lysis buffer (see in *Buffers*) and lysed twice in a French press (Thermo Fisher Scientific, Waltham, MA). The lysate was cleared by centrifugation (21,700 × g) for 1 h at 4 °C. In the first step of purification, we loaded the supernatant onto a prepared HiTrap Chelating column (GE Healthcare, Chicago, IL) charged with Ni^{2+} and pre-equilibrated with Buffer A (see in *Buffers*) supplemented with 30 mM imidazole, the column was washed with the same buffer.The bound proteins were eluted with a linear 30–500 mM gradient of imidazole in Buffer A. The fractions were analyzed by SDS-PAGE (Fig. 1A). The fractions containing the protein of interest were pooled, diluted with Buffer B (see *Buffers*), and loaded onto a HiTrap Heparin column (GE Healthcare) pre-equilibrated with Buffer B containing 50 mM NaCl. Elution of hNEIL2 was performed with a linear 50–1000 mM gradient of NaCl in Buffer B. We verified protein homogeneity by SDS-PAGE (Fig. 1B). The concentration of the purified protein was determined using the Bradford assay, and the protein-containing fractions were pooled, supplemented with 10% glycerol and 1 mM DTT, and stored at −80 °C.

Table 2
The workflow and processing parameters.

PLGS	
Digestion enzyme	non-specific
False discovery rate	100
Minimum fragment ion matches peptide and protein	3 and 7
Minimum peptide matches/protein	2
Low and elevated energy thresholds	200, 40 counts
Reference mass correction window	0.25 Da at 556.2771 Da/e
DynamX (PLGS filters)	
Minimum intensity	5000
Minimum PLGS score	> 6.5
Maximum MH+ error	10 ppm
Minimum products/amino acid	0.1
File threshold	2

MS analysis. Intact mass measurement was performed on a Waters Synapt G2-Si HDMS mass spectrometer equipped with a standard ESI source. The protein concentration was adjusted to 0.20 μM in 0.15% formic acid. 10 pmols (i.e., 50 μL) were loaded onto an ACQUITY UPLC BEH C4 trap column (2.1 μm × 5 mm; Waters Corporation, Milford) and desalted for 2 min at 100 μL/min in 0.15% formic acid at room temperature. The protein was directly eluted into the mass spectrometer using a sharp linear acetonitrile gradient (supplemented with 0.15% formic acid, pH 2.5) from 5 to 90% in 2 min at 60 μL/min. The measured molecular weight (37 632.19 +/− 0.15 Da) was consistent with the expected average mass calculated from the recombinant hNEIL2 primary sequence (37 632.23 Da, $\Delta m = +0.04$ Da (+1.06 ppm)) thereby confirming the structural integrity of the protein.

2.3. HDX-MS Data analysis

Peptide assignment was performed by processing the MS^E data acquired in undeuterated samples using ProteinLynx Global Server v3.0.1 (PLGS, Waters Corporation, Milford, MA) and a user-defined database containing the hNEIL2 and pig pepsin sequences. The following workflow and processing PLGS parameters are shown in Table 2 (PLGS). Peptides were further filtered in DynamX v3.0 (Waters Corporation, Milford, MA) using the PLGS filters (Table 1, DynamX). For HDX-MS analysis, a total of 81 peptides covering 96.7% of the hNEIL2 sequence were selected.

No adjustment was made for a back-exchange level, and the results are reported as relative deuterium exchange levels expressed in either mass unit (Da) or fractional exchange (%) as described in [5].

2.4. MD simulations, disorder prediction and homology modeling

Model preparation. To forecast disorders in NEIL2 we used the ESpritz neural network [2]. This network is trained on a set of partly disordered X-ray crystal structure and helps to forecast disorders in proteins. Homology models of the hNEIL2 structure was generated using the SWISS-MODEL server [6]. As templates, we use the X-ray structures of mvNEIL2/3 (4MB7 [3]) and oNEIL2 (PDB ID 6VJI [1]). First,the unstructured loop between the β3 and β4 strands (residues 58–125) was removed, resulting in the starting structures for MD. After that, the gap was sealed, and the structure was optimized by applying Modeller [7]. The Zn^{2+} ion position in the zinc finger was checked out by CheckMyMetal [8,9].

Molecular dynamics simulation. We performed MD simulation applying Amber 20 software [8] by GPU acceleration. The TIP3P water force field was used for the Zn^{2+} ion as well as ff14SB force field for the protein molecule. The structure was minimized and then the system was gradually heated from 20 K to 300 K (100 ps). After the heating, the 1 ns equilibration step was performed before the production run. In a constant pressure ensemble, we realized the production cycle in the implicit solvent for 100 ns under 300 K, and every 1 ps saved the trajectory of the production run. To perform the trajectory analysis and generate MD movies, we used UCSF Chimera software (University of California, San Francisco). The MDTRA [4] software we used for identification of hydrogen bonds and UCSF Chimera and PyMol programs (Schrödinger, New York, NY) to visualize the structures.

CRediT Author Statement

Polina V. Zhdanova: Investigation, Validation, Formal analysis, Writing – original draft; **Alexander A. Ishchenko:** Supervision, Methodology; **Alexander A. Chernonosov:** Writing – review & editing; **Dmitry O. Zharkov:** Formal analysis, Writing – original draft; review & editing; **Vladimir V. Koval:** Supervision, Conceptualization, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This research was supported by Grant No. 20-14-00214 of the Russian Science Foundation. We thank Dr. Sébastien Brier for his valuable assistance in HDX-MS sample preparation and data acquisition and the Institut Pasteur Biological NMR Technological Platform for access to the HDX-MS instrument.

Supplementary Materials

Supplementary material associated with this article can be found in the online version at doi: 10.17632/hpr8fr7ck6.1

References

[1] B.E. Eckenroth, V.B. Cao, A.M. Averill, J.A. Dragon, S. Doublié, Unique structural features of mammalian NEIL2 DNA glycosylase prime its activity for diverse DNA substrates and environments, Structure 29 (2021) 29–42 e4, doi:10.1016/j.str.2020.08.001.

[2] P.V. Zhdanova, A.A. Ishchenko, A.A. Chernonosov, D.O. Zharkov, V.V. Koval, Dynamics and conformational changes in human NEIL2 DNA glycosylase analyzed by hydrogen/deuterium exchange mass spectrometry, J. Mol. Biol. 434 (2022) 167334, doi:10.1016/J.JMB.2021.167334.

[3] A. Prakash, B.E. Eckenroth, A.M. Averill, K. Imamura, S.S. Wallace, S. Doublié, Structural investigation of a viral ortholog of human NEIL2/3 DNA glycosylases, DNA Repair (Amst) 12 (2013) 1062–1071, doi:10.1016/j.dnarep.2013.09.004.

[4] A.V. Popov, Y.N. Vorobjev, D.O. Zharkov, MDTRA: A molecular dynamics trajectory analyzer with a graphical user interface, J. Comput. Chem. 34 (2013) 319–325, doi:10.1002/jcc.23135.

[5] T.E. Wales, J.R. Engen, Hydrogen exchange mass spectrometry for the analysis of protein dynamics, Mass Spectrom. Rev. 25 (2006) 158–170, doi:10.1002/MAS.20064.

[6] A. Waterhouse, M. Bertoni, S. Bienert, G. Studer, G. Tauriello, R. Gumienny, F.T. Heer, T.A.P. De Beer, C. Rempfer, L. Bordoli, R. Lepore, T. Schwede, SWISS-MODEL: Homology modelling of protein structures and complexes, Nucleic Acids Res 46 (2018) W296–W303, doi:10.1093/nar/gky427.

[7] A. Šali, T.L. Blundell, Comparative protein modelling by satisfaction of spatial restraints, J. Mol. Biol. 234 (1993) 779–815, doi:10.1006/JMBI.1993.1626.
[8] H. Zheng, M.D. Chordia, D.R. Cooper, M. Chruszcz, P. Müller, G.M. Sheldrick, W. Minor, Validation of metal-binding sites in macromolecular structures with the CheckMyMetal web server, Nat. Protoc. 9 (2014) 156–170, doi:10.1038/nprot.2013.172.
[9] H. Zheng, D.R. Cooper, P.J. Porebski, I.G. Shabalin, K.B. Handing, W. Minor, CheckMyMetal: a macromolecular metal-binding validation tool, Acta Crystallogr. Sect. D Struct. Biol. 73 (2017) 223–233, doi:10.1107/S2059798317001061.