Bowhead NEIL1: molecular cloning, characterization, and enzymatic properties

Signe Holm, Rikke Møller Larsen, Camilla Myrup Holst, Mads Peter Heide-Jørgensen, John Fleng Steffensen, Tinna Stevnsner, Knud Larsen

Department of Molecular Biology and Genetics, Aarhus University, Universitetsbyen 81, DK-8000, Aarhus C, Denmark
Greenland Institute of Natural Resources, Strandgade 91, 2, DK-1401, Copenhagen K, Denmark
Marine Biological Section, University of Copenhagen, Strandpromenaden 5, DK-3000, Helsingør, Denmark

ABSTRACT

Nei Like DNA Glycosylase 1 (NEIL1) is a DNA glycosylase, which specifically processes oxidative DNA damage by initiating base excision repair. NEIL1 recognizes and removes bases, primarily oxidized pyrimidines, which have been damaged by endogenous oxidation or exogenous mutagenic agents. NEIL1 functions through a combined glycosylase/AP (apurinic/apyrimidinic)-lyase activity, whereby it cleaves the N-glycosyl bond between the DNA backbone and the damaged base via its glycosylase activity and hydrolysis of the DNA backbone through beta-delta elimination due to its AP-lyase activity. In our study we investigated our hypothesis proposing that the cancer resistance of the bowhead whale can be associated with a better DNA repair with NEIL1 being upregulated or more active. Here, we report the molecular cloning and characterization of three transcript variants of bowhead whale NEIL1 of which two were homologous to human transcripts. In addition, a novel NEIL1 transcript variant was found. A differential expression of NEIL mRNA was detected in bowhead eye, liver, kidney, and muscle. The A-to-I editing of NEIL1 mRNA was shown to be conserved in the bowhead and two adenosines in the 242Lys codon were subjected to editing. A mass spectroscopy analysis of liver and eye tissue failed to demonstrate the existence of a NEIL1 isoform originating from RNA editing. Recombinant bowhead and human NEIL1 were expressed in E. coli and assayed for enzymatic activity. Both bowhead and human recombinant NEIL1 catalyzed, with similar efficiency, the removal of a 5-hydroxyuracil lesion in a DNA bubble structure. Hence, these results do not support our hypothesis but do not refute the hypothesis either.

© 2022 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

1. Introduction

Reactive oxygen species are generated as byproducts of respiration and by exogenous sources such as UV-radiation, ionizing radiation, carcinogenic compounds, and environmental toxins [1]. These compounds and free radicals are genotoxic and can cause damage to biological molecules such as DNA by creating strand breaks and inducing a surplus of base lesions [2]. Many of these lesions are mutagenic and toxic, and can lead to increased genetic instability, inflammation, uncontrolled proliferation, cell death, and apoptosis [3,4]. The majority of the damaged bases are repaired via the base excision repair (BER) pathway. Recognition and removal of the oxidized bases in the DNA is initiated by DNA glycosylases with broad substrate preferences [5]. In mammals, the DNA glycosylases involved in repair of oxidized bases include NEIL1, NEIL2, NEIL3, OGG1, and NTH1 (reviewed by Dizdaroglu et al.) [6]. MUTYH is a glycosylase indirectly involved in the removal of oxidized bases that is essential for the repair of the OG:A base repair [7]. The three NEIL homologs initiating BER have functional activities as DNA glycosylases [6], however, their modes of strand incision differ [8,9]. NEIL1, NEIL2, and NEIL3 possess broad and overlapping substrate specificities and the preferred substrates for all are spiroiminodihydantoin and guanidinohydantoin, which are both highly mutagenic DNA lesions [10]. With these DNA glycosylase activities, the NEIL enzymes have the ability to regulate mutation frequency in cells. Specific variations in mammalian NEIL1 and NEIL2 have been shown to lead to an elevated mutation frequency, indicating that these enzymes can suppress mutations in cells [11–13]. NEIL1

https://doi.org/10.1016/j.biochi.2022.10.014
0300-9084/© 2022 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).
has AP (apurinic/apyrimidinic) lyase activity and introduces nicks in the DNA strand. NEIL1 and NEIL2 cleave the DNA backbone through beta-delta elimination to generate a single-strand break at the site of the removed base, with both 3' and 5'-phosphates. Additionally, NEIL1 possesses DNA glycosylase/lyase activity towards mismatched uracil and thymine, in particular in U:C and T:C mismatches [14].

The NEIL1 gene has been identified and shown to be cloned by several groups simultaneously [14–18]. The human NEIL1 gene is approximately 10 kb long and is located on chromosome 15 (15:75,346,955–75,357,115). The gene is organized in 10 exons (nine coding) separated by nine introns.

At least 14 different protein-coding splicing variants of human NEIL1 have been reported (Ensembl). The human NEIL1 transcript is highly expressed in liver, pancreas, and thymus and moderately expressed in brain, spleen, and ovary [15]. Human NEIL1 and NEIL3 have been demonstrated to be cell-cycle regulated and the expression of both reaches the maximum at S/G2. In contrast, human NEIL2 is very likely constitutively expressed [19,20]. NEIL1 deficiency has been associated with cancer and the NEIL1 gene promoter is hypermethylated in head and neck squamous cell carcinoma (HNSCC), non-small cell lung cancer (NSCLC) tumors, and colorectal cancer [21–23].

NEIL1 pre-mRNA is subject to A-to-I editing mediated by adenosine deaminase acting on double-stranded RNA 1 (ADAR1). Editing occurs at amino acid codon 242 and results in recoding of a lysine residue to arginine [24,25]. Editing of NEIL1 is strongly enhanced by interferon α, a well-known regulator of ADAR1 expression [24,26]. A-to-I editing of NEIL1 pre-mRNA affects both enzyme kinetics and substrate specificity of the NEIL1 enzyme. Yeo et al. [24] demonstrated that the K242 NEIL1 isoform removes thymine glycol approximately 30 times faster than the edited R242 isoform of NEIL1. In contrast, the edited form reacts nearly three times faster with guanidinohydratoin and spiranomohydratoin in duplex DNA than the unedited form. In a recent study, Minko et al. [27] showed clear differences in substrate specificity of unedited and edited NEIL1. The DNA lesions imidazole ring-opened 4,6-diamino-5-formamidopyrimidine (FapyAdo) and 2,6-diamino-4-hydroxy-5-formamidopyrimidine (FapyGua) were preferentially released by both NEIL1 enzymes, with K242R being 1.3- and 1.2-fold more efficient than K242 in excision of FapyAdo and FapyGua, respectively. Significant differences (7.5–12-fold) were measured in the excision of thymine glycol from genomic DNA by the unedited versus edited NEIL1. In contrast, the edited NEIL1 was more efficient (3- to 5-fold) in release of 5-hydroxycytosine.

NEIL2 was first cloned and characterized by Hazra et al. [15,16]. The NEIL2 gene is localized to the human chromosome 8 (8p23.1), is approximately 18 kb, and consists of five exons. NEIL2 is highly expressed in skeletal muscle and testis and, unlike NEIL1, its expression is independent of the cell cycle stage in fibroblasts [15]. The NEIL2 enzyme appears to primarily associate with DNA repair during transcription and acts preferentially on cytosine-derived lesions, with the highest affinity for 5-hydroxycytosine (5-OHU) and lower affinity for 5-hydroxycytosine and 5,6-dihydroxycytosine. NEIL2 contains an N-terminal catalytic domain, a hinge region, and a C-terminal DNA-binding domain with helix-two-turn-helix and zinc finger motif, which is an unusual CHCC-type motif and distinct from that of NEIL1 [28]. The zinc finger motif is essential for maintaining the structural integrity and activity of NEIL2 [29]. NEIL2 shows a unique preference for excision of lesions from a DNA bubble [30]. Mutations in the zinc finger abolish NEIL2 enzyme activity [29]. Abnormal expression of NEIL2 has been associated with somatic mutation loads in diverse cancers [31–33]. The level of NEIL2 transcript expression in tumor tissue compared with normal tissue has been shown to be significantly reduced in four of 13 cancer types previously examined [33]. Similarly, NEIL2 protein expression levels have been found to be significantly lower in 50% of cancer tissues [31]. A functional analysis, via knockdown of NEIL2 expression, led to a dramatic increase in the accumulation of non-homologous ends in cell cultures. In that study, the down-regulation of NEIL2 resulted in a 6–7-fold increase in spontaneous mutation frequency [12]. The specialized NEIL2 functions in multiple cellular processes were recently reviewed by Sarkar et al. [34].

NEIL3 is the third member of the NEIL gene family. In previous studies, NEIL3 genes have been cloned from mouse and human and the functions of their recombinant proteins have been investigated [14,17,18,35–37]. NEIL3 has been found to differ from NEIL1 and NEIL2 in temporal and spatial expression. While NEIL1 and NEIL2 possess a beta-delta-elimination activity, thereby generating a one-nucleotide gap, NEIL3 incises damaged bases through beta-elimination [30]. NEIL3 has poor lyase activity [9] and is also involved in damage removal from non-canonical DNA structures [38], and unhooking interstrand DNA crosslinks [39].

NEIL3 has shown increased excision of methylformamidopyrimidines (methyl-FapyG) from duplex DNA [18]. NEIL3 transcript expression is restricted to thymus, spleen, and bone marrow in cerebral stem cells in both humans and mice [36,40,41]. Additionally, the expression of NEIL3 mRNA is age dependent [40]. In 13 cancer types examined by Shinmura et al. [33], the NEIL3 transcript expression level was shown to be significantly increased in tumor tissue compared with normal tissue.

The bowhead whale is cancer resistant, which makes it very interesting to study given the insights that can be gained regarding understanding how evolution has suppressed the development of cancer. Theoretically, large-bodied animals should have a greater risk of developing cancer as they potentially undergo more cell divisions and thereby introduce somatic mutations, however, no such association has been found. This means that incidence of cancer does not correlate with the number of cells in an organism. This lack of correlation between body size and cancer risk at the species level has been named Peto’s paradox [42]. A recent study with zoo mammals confirmed Peto’s paradox and demonstrated key species-specific anti-cancer mechanisms [43].

Another molecular hypothesis proposed to address Peto’s paradox was presented by Abegglen et al. [44], who showed that elephants have a lower-than-expected rate of cancer very likely because they have multiple (least 20) copies of the tumor suppressor TP53. This clearly demonstrates that there are different molecular strategies that contribute to cancer resistance in large and long-lived organisms [44].

We hypothesize that the bowhead whale possesses one or more
specialized mechanisms to avoid development of cancer. One such mechanism could be DNA repair. We propose that that NEIL1 activity could be higher and remaining stable in expression for a longer time in age development in the bowhead than in humans. As NEIL1, NEIL2, and NEIL3 are important enzymes involved in the repair of mutagenic bases and thereby capable of suppressing mutations, while simultaneously playing a role in epigenetics, we cloned and characterized these three enzymes from bowhead whale for this study. We also investigated enzymatic activity with bowhead and human NEIL1.

2. Materials and methods

2.1. Ethics

Professor Mads-Peter Heide-Jørgensen of the University of Copenhagen and Greenland Institute of Natural Resources provided bowhead samples. The CITES permission number 12GL1003387 was used for this project.

2.2. Biological material

Bowhead whale organ and tissue samples were from four individuals: #322, #323, #501, and #502. Characteristics of the bowhead individuals such as estimated age, gender, size, and age of capture were described by Heide-Jørgensen et al. [45]. Professor John Fleng Steffensen provided shark samples.

2.3. Isolation of nucleic acids

Tissues from bowhead and shark were dissected and homogenized in liquid nitrogen. Total RNA was isolated using the RNeasy method (Qiagen). For the cloning of human NEIL1 coding sequence (cDNA), RNA was isolated from a human blood sample using the PureLink RNA mini kit. DNA was isolated from biological samples according to standard protocols [46]. Synthesis of the cDNA used for cloning was performed as described in Henriksen et al. [47]. cDNAs used for expression analysis were synthesized from RNA isolated from various sources (as indicated below) using random primers (Roche) and according to the manufacturer’s protocol. The integrity of the RNA samples was verified through ethidium bromide staining of the ribosomal RNA on 1% agarose gels. The quality of the DNA isolated from bowhead tissues was examined by agarose gel electrophoresis.

2.4. Molecular cloning of bowhead NEIL1 cDNA sequences

The cDNA fragments encoding NEIL1 were amplified through PCR using a bowhead retina cDNA and NEIL1-specific primers derived from The Bowhead Whale Genome Resource (http://www.bowhead-whale.org/). Two independent PCR reactions with two NEIL1 primer sets were performed. The PCR reaction mix contained 2.5 μl cDNA, 1.5 mM MgCl₂, 0.2 mM dNTP, and 0.5 μM of each primer NEIL1-SF combined with NEIL1-SR, NEIL1-LF, and NEIL1-LR (Table 1) and 1 U Phusion DNA polymerase (Finnzymes) for a total volume of 25 μl. The PCR conditions were as follows: 95 °C for 2 min, 10 touchdown cycles of 95 °C for 20 s, 60–55 °C for 30 s, 72 °C for 45 s, followed by 25 cycles of 95 °C for 20 s, 65 °C for 30 s, 72 °C for 45 s, and finally an elongation at 72 °C for 5 min. Two PCR products of approximately 1300 bp and 1200 bp were visualized and isolated from an ethidium bromide stained 1% agarose gel. The recovered cDNA amplicons were cloned directly into the PCR TOPO 2.1 vector (Invitrogen) and sequenced in both directions.

Human NEIL1 cDNA was RT-PCR cloned using cDNA synthesized from whole blood using the unique SepMate™ peripheral blood mononuclear cell isolation tubes. RNA was purified from the peripheral blood mononuclear cells using the PureLink RNA Mini Kit, as described above. The PCR reaction mix contained 0.5 μM of the primers HSNEIL1-EX-F and HSNEIL1-EX-R (Table 1), 0.2 mM dNTP, and SuperFi Green DNA polymerase. The PCR conditions were as follows: 98 °C for 30 s, 40 cycles of 98 °C for 10 s, 68 °C for 15 s, 72 °C for 30 s, and finally 72 °C for 7 min.

The shark NEIL1 cDNA sequence was PCR amplified using cDNA synthesized from RNA isolated from Greenland shark muscle and spiny dogfish spleen. The PCR reaction mix was as follows: 0.5 μM of the primers SHARKNEIL1-F and SHARKNEIL1-AS (Table 1), 0.2 mM dNTP, and SuperFi Green DNA polymerase. The PCR conditions were as follows: 98 °C for 30 s, 40 cycles of 98 °C for 10 s, 68 °C for 15 s, 72 °C for 30 s, and 72 °C for 7 min.

2.5. Molecular cloning of bowhead NEIL2 cDNA and NEIL3 genomic DNA sequences

The NEIL2 cds was PCR amplified using the primer set BMNEIL2-S and BMNEIL2-AS (Table 1) with bowhead retina cDNA. The PCR reaction mix was as follows: 0.5 μM of the primers, 0.2 mM dNTP, and SuperFi Green DNA polymerase. The PCR conditions were as follows: 98 °C for 30 s, 40 cycles of 98 °C for 10 s, 68 °C for 15 s, 72 °C for 30 s, and 72 °C for 7 min. One PCR product of approximately 1000 bp was visualized and isolated from an ethidium bromide stained 1% agarose gel. The recovered cDNA amplicon was cloned directly into the pcDNA3 TOPO 2.1 vector (Invitrogen) and sequenced in both directions.

Several attempts to RT-PCR clone NEIL3 from different bowhead tissues were unsuccessful. Therefore, NEIL3-specific primers derived from a sequence (bmy_03959) retrieved from The Bowhead Whale Genome Resource (http://www.bowhead-whale.org/) were used to PCR amplify genomic sequences. Sequenced ampli- cons were assembled into a partial cds for NEIL3. The NEIL3 exon primers are listed in Table 1.

RT-PCR cloning of partial shark NEIL1 cDNAs was performed with RNA isolated from Greenland shark (Somniosus microcephalus) red muscle and RNA from spiny dogfish (Squalus acanthias) spleen. PCR was performed with Invitrogen Platinum SuperFi Green DNA polymerase and 0.5 μM primers: SHARKNEIL1-F and SHARKNEIL1-AS (Table 1). The following PCR program was used to amplify the shark NEIL1 amplicon of approximately 900 bp: 98 °C for 30 s followed by 35 cycles of 98 °C for 10 s and 72 °C for 30 s and then 72 °C for 7 min.

2.6. Expression analysis

Expression was determined through quantitative RT-PCR analysis and performed as described by Larsen and Heide-Jørgensen [48]. A 66 bp amplicon obtained via PCR with the primers NEIL1-QF and NEIL1-QR (Table 1), both located within exon 2 of the NEIL1 gene, was amplified and probed using NEIL1-probe #17 (Table 1) from the Human Probe Library at Roche. PCR controls with RNA as the template did not yield any product, indicating no contamination with chromosomal DNA.

2.7. PCR amplification of edited sequences for NEIL1 mRNAs from bowheads

A reverse transcriptase-polymerase chain reaction (RT-PCR) was performed to quantify the A-to-I editing in bowhead NEIL1 pre-mRNA. The sequences of the oligonucleotides used in the PCR reaction, BMNEIL1-EF1 and BMNEIL1-ER1, are listed in Table 1. PCR was performed with Invitrogen Platinum SuperFi Green DNA
polymerase and 0.5 μM primers. The following PCR program was used to amplify porcine, bowhead, and shark amplicons of 318 bp, 238 bp, and 322 bp, respectively: 98 °C for 15 s and 3′-end of the coding sequence through PCR with the primers HSNEIL1-EXF and HSNEIL1-EXR (Table 1). The amplification product was digested with XhoI and ligated into the expression vector pET22b. The resulting constructs were bowhead and human NEIL1 containing a His-tag at the carboxy-terminal end.

2.8. Estimation of A-to-I editing degree

The A-to-I editing degree was estimated by measuring the top heights of the adenosine and guanine signals in the sequencing electropherograms. The editing degree was calculated as $\frac{[G \text{ (mm)}]}{[A \text{ (mm)} + G \text{ (mm)}]} \times 100$.

2.9. Expression and purification of recombinant bowhead and human NEIL1 enzymes

The cloned bowhead NEIL1 coding sequence was codon optimized for expression in E. coli and cloned into the expression vector pET-30a(+) using the restriction enzymes NdeI and Hind III. In order to express recombinant human NEIL1, the coding sequence was RT-PCR cloned using RNA isolated from blood. Sequencing of the NEIL1 clone revealed a complete homology identity with the sequence deposited at GenBank (NM_024608). Linkers containing recognition sequences for NdeI and XhoI were added to the 5′- and 3′-end of the coding sequence through PCR with the primers HSNEIL1-EXF and HSNEIL1-EXR (Table 1). The amplified NEIL1 product was digested with NdeI and XhoI and ligated into the expression vector pET22a. The resulting constructs were bowhead and human NEIL1 containing a His-tag at the carboxy-terminal end. E. coli BL21 Star (DE3) competent cells were transformed with the recombinant plasmid. A single colony was inoculated into LB medium containing kanamycin and the cultures were incubated at 16 °C and 37 °C at 200 rpm. A volume of 5 ml overnight culture was used to monitor the expression. The Western blot was generated using anti-His antibody (GenScript, Cat. No.A00186).

2.10. Purification of recombinant NEIL1 proteins

Bowhead and human NEIL1 were cloned into the pET-22b vector, resulting in a recombinant C-terminal His6-tagged human NEIL1 protein. This protein was purified using anti-His antibody (GenScript, Cat. No.A00186).
NEIL1 expression plasmid. *E. coli* transfected with the plasmid was grown at 37 °C in LB medium supplemented with 500 mM d-sorbitol, 1 mM betaine, and 100 µg/mL ampicillin. Expression of the fusion protein was induced in log phase cells at OD = 0.8 via addition of 0.1 mM IPTG. Cells were harvested 4 h after induction through centrifugation at 6500 × g for 20 min. The lysate was applied to a 10 ml Ni2⁺-column (Ni-NTA Superflow, Qiagen) pre-equilibrated with sonication buffer. The column was washed in three volumes of sonication buffer followed by two volumes of sonication buffer containing 50 mM imidazole. Protein was eluted with 300 mM imidazole containing 40 mM Tris-HCl pH 8.0, 4 mM MgCl₂, 30 mM NaCl, 5 mM β-mercaptoethanol, and 10 mM imidazole). The cells were then lysed using sonication and insoluble particles removed through centrifugation at 20,000 × g for 20 min. The column was washed in three volumes of sonication buffer followed by two volumes of sonication buffer containing 50 mM imidazole. Protein was eluted with 300 mM imidazole in three column volumes of sonication buffer and collected in fractions. The fractions were analyzed on an SDS-PAGE gel and dialyzed against dialysis buffer (10 mM Tris-collected in fractions. The fractions were analyzed on an SDS-PAGE containing 50 mM imidazole. Protein was eluted with 300 mM sonication buffer. The column was washed in three volumes of sonication buffer followed by two volumes of sonication buffer containing 50 mM imidazole. Protein was eluted with 300 mM imidazole in three column volumes of sonication buffer and collected in fractions. The fractions were analyzed on an SDS-PAGE protein gel and dialyzed against dialysis buffer (10 mM Tris-HCl pH 7.0, 50 mM NaCl, and 10 mM β-mercaptoethanol). The fractions containing the highest concentrations of NEIL2 were pooled and applied to a HiTrapSP XL 1 ml column (Amersham) and eluted with a 50 mM–2 M gradient of NaCl in 10 mM Tris–HCl (pH 7.0) and 10 mM β-mercaptoethanol. The fractions with the purest and most concentrated NEIL1 protein were pooled. NEIL1 protein concentration was determined through SDS-PAGE using BSA as standard, followed by total protein staining with Imperial Protein Stain (Thermo Scientific). A protein size marker (Spectra multicolor, Fermentas) was run in parallel on the same gel.

2.11. Radiolabeling and annealing of oligonucleotide

The oligonucleotide 5-OHU B11 (Table 1) used in the inclusion assay was purchased from Sigma and the 5'-end was labeled through incubation with [γ-32P]-ATP (Perkin-Elmer) and T4 polynucleotide kinase (Thermo Scientific) for 90 min at 37 °C in 1 x forward buffer A (Thermo Scientific), as outlined by Ferrando et al. [49]. The reaction was stopped via incubation at 90 °C for 1 min. Free [γ-32P]-ATP was removed using G50 microspin columns (GE Healthcare) according to the manufacturer’s protocol. The labeled oligonucleotide was annealed to partially complementary oligonucleotide (B11 complementary strand in Table 1) in 175 mM KCl and 100 mM EDTA with 5 min incubation at 90 °C, followed by gradual cooling overnight. Annealing with the B11 complementary strand resulted in a double-stranded substrate with the 5-OHU lesion in the center of an internal 11 nt single-stranded bubble structure.

2.12. Incision assay with recombinant NEIL1 proteins

To perform the incision assay, 2–50 nM recombinant human or bowhead NEIL1 protein was incubated with 2 nM substrate (double-stranded oligonucleotide with 11 nt bubble structure containing a 5-OHU lesion or C as control) for 15 min at 37 °C in a buffer containing 40 mM Tris–HCl pH 8.0, 4 mM MgCl₂, 30 mM NaCl, 5 mM DTT, 0.5 mM β-mercaptoethanol, 100 µg/mL BSA, 100 µM EDTA, and 5% glycerol. Reactions were stopped through the addition of 2x formamide loading buffer (80% formamide with 100 mM NaOH to cleave any residual AP sites) and boiled for 5 min at 95 °C. Substrate and product were separated using a 20% denaturing polyacrylamide gel and visualized via phosphoimaging (BioRad). Percentage incision was calculated as the amount of product relative to the sum of the substrate and product.

2.13. Bioinformatic analyses

The ORF Finder (http://www.ncbi.nlm.nih.gov/orf/fig.cgi) was used to identify open reading frames. Sequence analysis was performed using software online at NCBI (http://ncbi.nlm.nih.gov) and Expasy (http://expasy.org). The putative amino acid sequences were deduced using the Expasy translate tool (http://expasy.org/translate). Homologs of CBLNs were retrieved from NCBI using blastx. ClustalW (http://www.genome.jp/tools/clustalw/) was used for sequence alignment. The theoretical isoelectric point (pI) and molecular weight (Mw) for CBLNs were estimated using Compute pi/Mw software (http://web.expasy.org/compare_pi/).

The sorting intolerant from tolerant (SIFT) algorithm was used to predict the potential impact of amino acid substitutions on the protein function of the gene products identified in this study (https://sift.bii.a-star.edu.sg/www/SIFT_seq_submit2.html).

3. Results and discussion

3.1. Molecular cloning of bowhead NEIL1, human NEIL1, and shark NEIL1

In order to perform an in-depth analysis of the amino acid sequence of bowhead NEIL1 and to compare with counterparts from other species we RT-PCR cloned the NEIL1 cDNA. For this study, three variants of bowhead NEIL1 were RT-PCR cloned from retina cDNA. A long variant (NEIL1-l) was cloned using the primers BMMNEIL1-LF and BMMNEIL1-LR (Table 1). The DNA sequence of this long NEIL1 variant consists of 1272 nucleotides encoding a protein of 411 amino acids. Compared with the shorter variant, NEIL1-l has an additional 21 amino acids in the N-terminal end. A shorter bowhead NEIL1 variant (NEIL1-s) was also identified, consisting of 1217 nucleotides coding for a protein of 390 amino acids. This shorter bowhead variant is most similar to the human NEIL1 transcript variant 2 (GenBank NM_024608). Compared with NEIL1-l, NEIL1-s differs in the 5’ UTR, lacks a portion of the 5’ coding region, and initiates translation at a downstream start codon. The encoded isoform NEIL1-s has a distinct N-terminus and is shorter than the NEIL1-l isoform. Further, during the cloning procedure we identified a third variant with an insertion in exon 5 (NEIL1-i5). The reading frame was kept intact, resulting in a protein very similar to NEIL1-s but with a five amino acid (PSREN) insertion. A Blast p analysis did not reveal any similar NEIL1 isoforms in other organisms. Hence, this third variant identified must be regarded as a novel bowhead-specific NEIL1 variant.

An amino acid sequence alignment of the bowhead NEIL1 isoforms and the human counterpart is presented in Fig. 1A. A high amino acid homology was observed for the NEIL1 protein from the two species, with an identity of 84%. A comparison of the human and the bowhead NEIL1-s isoform revealed 62 amino acid differences distributed through the entire amino acids sequence. The highest homology was seen in the N-terminal end and the middle part of NEIL1, whereas there was a much higher variability in the carboxy-terminal end of the protein. The human and bowhead NEIL1 carboxy-terminus showed 33 amino acid differences. All amino acid residues important for NEIL1 catalytic activity—Pro2, Glu3, Glu6, Lys54, D167, and E181, and the three amino acid residues important for NEIL1 catalytic activity

NEIL1 contains three serine residues—Ser61, Ser207, and Ser306—which have been shown to be phosphorylated [49]. Of
### A

<table>
<thead>
<tr>
<th>Protein</th>
<th>Sequence</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>BmSNL1</td>
<td>MHTF4HLGCLSSPTQGFPFHRK60</td>
<td>60</td>
</tr>
<tr>
<td>BmSNL1-s</td>
<td>MHTF4HLGCLSSPTQGFPFHRK</td>
<td>39</td>
</tr>
<tr>
<td>BmSNL1-15</td>
<td>MHTF4HLGCLSSPTQGFPFHRK</td>
<td>39</td>
</tr>
<tr>
<td>HeSNL1</td>
<td>MHTF4HLGCLSSPTQGFPFHRK</td>
<td>39</td>
</tr>
<tr>
<td>BmSNL1</td>
<td>FBSAYHIALAASQELITLMILQVQQPLPQLTVPFKTSPQCLVLQSLALPHKLRH120</td>
<td>120</td>
</tr>
<tr>
<td>BmSNL1-s</td>
<td>FBSAYHIALAASQELITLMILQVQQPLPQLTVPFKTSPQCLVLQSLALPHKLRH</td>
<td>99</td>
</tr>
<tr>
<td>BmSNL1-15</td>
<td>FBSAYHIALAASQELITLMILQVQQPLPQLTVPFKTSPQCLVLQSLALPHKLRH</td>
<td>99</td>
</tr>
<tr>
<td>HeSNL1</td>
<td>FBSAYHIALAASQELITLMILQVQQPLPQLTVPFKTSPQCLVLQSLALPHKLRH</td>
<td>99</td>
</tr>
<tr>
<td>BmSNL1</td>
<td>KRTUPSPPLALPCVDKFGHGKDQVPRQPCVLLSYEQPREN---VLQNLAD</td>
<td>175</td>
</tr>
<tr>
<td>BmSNL1-s</td>
<td>KRTUPSPPLALPCVDKFGHGKDQVPRQPCVLLSYEQPREN---VLQNLAD</td>
<td>154</td>
</tr>
<tr>
<td>BmSNL1-15</td>
<td>KRTUPSPPLALPCVDKFGHGKDQVPRQPCVLLSYEQPREN---VLQNLAD</td>
<td>159</td>
</tr>
<tr>
<td>HeSNL1</td>
<td>KRTUPSPPLALPCVDKFGHGKDQVPRQPCVLLSYEQPREN---VLQNLAD</td>
<td>154</td>
</tr>
</tbody>
</table>

#### Helix-two-turn helix

<table>
<thead>
<tr>
<th>Protein</th>
<th>Sequence</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>BmSNL1</td>
<td>KAFDRICZELGQFQRTGQYLYRLYKDFPEFAAFLQVLQSLALQQRSELILQ</td>
<td>235</td>
</tr>
<tr>
<td>BmSNL1-s</td>
<td>KAFDRICZELGQFQRTGQYLYRLYKDFPEFAAFLQVLQSLALQQRSELILQ</td>
<td>214</td>
</tr>
<tr>
<td>BmSNL1-15</td>
<td>KAFDRICZELGQFQRTGQYLYRLYKDFPEFAAFLQVLQSLALQQRSELILQ</td>
<td>219</td>
</tr>
<tr>
<td>HeSNL1</td>
<td>KAFDRICZELGQFQRTGQYLYRLYKDFPEFAAFLQVLQSLALQQRSELILQ</td>
<td>214</td>
</tr>
</tbody>
</table>

#### Recog-loop

<table>
<thead>
<tr>
<th>Protein</th>
<th>Sequence</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>BmSNL1</td>
<td>KIRKQNFQDLPDLCHSVPQKVQG8RTOPSTKEDAAPOAMNDEAOPQPSLQR</td>
<td>295</td>
</tr>
<tr>
<td>BmSNL1-s</td>
<td>KIRKQNFQDLPDLCHSVPQKVQG8RTOPSTKEDAAPOAMNDEAOPQPSLQR</td>
<td>274</td>
</tr>
<tr>
<td>BmSNL1-15</td>
<td>KIRKQNFQDLPDLCHSVPQKVQG8RTOPSTKEDAAPOAMNDEAOPQPSLQR</td>
<td>279</td>
</tr>
<tr>
<td>HeSNL1</td>
<td>KIRKQNFQDLPDLCHSVPQKVQG8RTOPSTKEDAAPOAMNDEAOPQPSLQR</td>
<td>274</td>
</tr>
</tbody>
</table>

#### Zn-less finger

<table>
<thead>
<tr>
<th>Protein</th>
<th>Sequence</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>BmSNL1</td>
<td>QRTTFQDQDLPDLCHSVPQKVQG8RTOPSTKEDAAPOAMNDEAOPQPSLQR</td>
<td>355</td>
</tr>
<tr>
<td>BmSNL1-s</td>
<td>QRTTFQDQDLPDLCHSVPQKVQG8RTOPSTKEDAAPOAMNDEAOPQPSLQR</td>
<td>334</td>
</tr>
<tr>
<td>BmSNL1-15</td>
<td>QRTTFQDQDLPDLCHSVPQKVQG8RTOPSTKEDAAPOAMNDEAOPQPSLQR</td>
<td>339</td>
</tr>
<tr>
<td>HeSNL1</td>
<td>QRTTFQDQDLPDLCHSVPQKVQG8RTOPSTKEDAAPOAMNDEAOPQPSLQR</td>
<td>334</td>
</tr>
</tbody>
</table>

#### Prot-Prot interact.

<table>
<thead>
<tr>
<th>Protein</th>
<th>Sequence</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>BmSNL1</td>
<td>TAQQPFBTSLGQFPQPVVEKGRHSSRQATSGRSGFQIKKDTPSESEPSTAS</td>
<td>411</td>
</tr>
<tr>
<td>BmSNL1-s</td>
<td>TAQQPFBTSLGQFPQPVVEKGRHSSRQATSGRSGFQIKKDTPSESEPSTAS</td>
<td>390</td>
</tr>
<tr>
<td>BmSNL1-15</td>
<td>TAQQPFBTSLGQFPQPVVEKGRHSSRQATSGRSGFQIKKDTPSESEPSTAS</td>
<td>395</td>
</tr>
<tr>
<td>HeSNL1</td>
<td>TAYKQFBTSLGQFPQPVVEKGRHSSRQATSGRSGFQIKKDTPSESEPSTAS</td>
<td>390</td>
</tr>
</tbody>
</table>

---

### B

- **Bowhead**
- **Minke whale**
- **Beluga**
- **Narwhale**
- **Yangtze finless porpoise**
- **Killer**
- **Pacific white-sided dolphin**
- **Physeter**
- **Common bottlenose dolphin**
- **Cow**
- **Pig**
- **Human**
- **Mus**
these serine residues, only Ser61 and Ser207 were conserved in the bowhead NEIL1 sequences. In addition, Tyr263 and Ser269 were conserved. Both these residues were phosphorylated in human NEIL1. Phosphorylation of Tyr263 is possibly detrimental for NEIL1 activity [50]. Four of five possible phosphorylation sites with serine and tyrosine residues were conserved in the bowhead NEIL sequence: Ser61, Ser207, Tyr263, and Ser269. Only one residue, Ser306, was not conserved, but phosphorylation of this site had little effects on the enzyme activity of human NEIL1 [50]. However, phosphorylation of Ser306 had a slightly diminishing effect on enzyme turnover, therefore, the lack of modification probably had no effect. Within the amino acid sequences from 1 to 142, only a small number (20) of mismatches were found.

The effect of non-matching amino acid residues was evaluated using the predictive tools PolyPhen-2 [51,52] and SIFT [53] using the bowhead NEIL1 protein as the functional sequence on account of the hypothesis that bowhead DNA repair activity is more efficient. The SIFT scores are shown in Table S1. Among the small number of mismatches (amino acids 1–142), few or none seemed to interact in a manner that would cause significant structural changes. However, in the human NEIL1 protein, Arg95 and Arg122 may be in the range to interact weakly with the DNA backbone (Fig. 2A), whereas the two bowhead equivalents, Pro95 and His122, may be too short for this hypothetical interaction. For NEIL1, this could be an advantage, meaning improved DNA substrate binding, or a disadvantage, making the substrate release more difficult. The serine residue found at position 82 in the human NEIL1 sequence was substituted with a threonine in the bowhead NEIL1 sequence. Prakash et al. [54] reported a reduced enzymatic activity for the substrate 5,6-thymine glycol of the S82C human NEIL1 variant. The SIFT core value of 0.92 for T82S could indicate that this substitution might impact the enzymatic activity.

When K242 was unedited, the recognition loop (Fig. 1A) consisting of 14 amino acids was conserved in 12 of these positions. The A-to-t edited lysine codon (AAA) was found here. The Lys242 was only found in the bowhead NEIL-l variant, whereas NEIL1-s and NEIL-l5 had an arginine residue at position 242. In both bowhead and human NEIL1, A-to-t RNA editing resulted in substitution of a lysine to an arginine residue at position 242. The two other mismatches were calculated to be tolerable for NEIL1 function, however, considering the significance of the loop, the changes could potentially affect substrate recognition. In essence, the two mismatches can be interpreted as a hydroxyl-group (in Thr249) moving up three positions (Ser246) and being replaced by a non-polar amino acid. Unfortunately, the crystallization of the NEIL1 recognition loop was difficult, especially when bound to DNA. The recognition loop for the unbound human NEIL1 with approximate conformations of Arg242, Ser246, and Gly249 is shown in Fig. 2B. A threonine residue in position 249 may not have much impact in the unbound conformation, however, when bound, Lys/Arg-242 flipped about 180°, meaning that the amino/guanidine group might come closer to a 249 hydroxyl group than a 246 amino acid, which could affect lesion recognition and/or base excision. Additionally, a zinccenter finger motif consisting of 28 amino acids was highly conserved (boxed in Fig. 1A). This motif contained the essential Arg277. The amino acid sequence with amino acids 263–290, and containing the zinccenter finger motif, was highly conserved. R272Q and Q275H were confirmed as tolerable and these two residues were virtually perpendicular to the highly conserved Arg277, suggesting that their side chains had little effect on the enzymatically important residue. In addition, while there was less similarity between threonine and methionine, the 265 residue was even further from the DNA and not in any apparent position to affect the rest of the protein.

The majority of the observed mismatches between bowhead

---

**Fig. 1.** A) Alignment of bowhead NEIL1 sequences and human NEIL1 generated through Clustal W (https://www.genome.jp/tools-bin/clustalw). NEIL1 sequences used for the alignment were: Bowhead NEIL1 (GenBank Access Nos. MZ055365, MZ055366, MZ055370) and human NEIL1 (GenBank Access. No. NM_024608). Functional motifs are indicated by boxes and names. Helix-two-turn-helix (HT2H); recognition loop, zinccenter finger motif, and protein-protein interaction domain. Amino acid residues important for catalytic activity are shown with bold underlined letters. Potential phosphorylation sites are shown in shaded boxes. Three “gap-filling” residues (Met81, Arg118, and Phe120) are boxed. Two lysine clusters are underlined. The essential Arg277 is indicated by an e. Identical amino acids at each position are indicated by asterisks. Amino acids subjected to A-to-I RNA editing are indicated by arrows. B) Phylogenetic tree (rooted UPGMA) of porcine NEIL1 and 12 other mammalian NEIL1 proteins. The tree was constructed using the clustal method based on amino acid similarities of the full sequences (https://www.genome.jp/tools-bin/clustalw). The GenBank accession numbers for the NEIL1 sequences used were: minke whale (XM_007197675), beluga whale (XM_02388076), narwhal (XM_029239116), Yangtze finless porpoise (XM_024761807), bottlenose dolphin (XM_033852011), killer whale (XM_033440112), pacific white-sided dolphin (XM_027123659), Physeter (XM_028495069), cow (NM_001014917), pig (XM_021099028), human (NM_024608), and mouse (NM_028347).
and human NEIL1 were found in the carboxy-terminal end. While none of the single substitutions were reported as problematic by SIFT or PolyPhen-2, the fact that a high number of them were gathered in a small space was likely to have an effect on the protein. Unfortunately, no crystallization of the carboxy-terminal was available to evaluate and compare interactions between residues visually. The C-terminal is not essential for enzymatic activity in vitro, which might explain the low degree of conservation, but it has been shown to hold an important site for interaction with other BER and replication proteins among those PARP1. The many variations may indicate distinct protein–protein interactions between the two NEIL1 enzymes. Perhaps bowhead NEIL1 shows improved coordination with other DNA repair proteins, which would result in increased efficiency. The two lysine clusters were well-conserved overall, with only the second (amino acids 356 to 361 implicated in protein fold-back) showing variation between sequences. Arginine and lysine should be almost interchangeable, but the glutamic acid–lysine discrepancy in position 356 is more interesting, as these two amino acids are oppositely charged. As the lysine cluster is thought to interact with an acidic cluster in the core domain, a shift from positive to negative charge from the glutamic acid may mean a weakening of the intramolecular interactions. This could result in a more tightly bound enzyme, but might also facilitate zinceeded release on DNA binding and thereby possibly faster initiation of enzyme activity.

The sequence responsible for protein-protein interaction with PARP-1 (BRCT region) is found in the carboxy-terminal. This motif also showed some conservation. Finally, two lysine clusters (Fig. 1A) were partially conserved in the bowhead NEIL-s sequence.

All mismatches between the bowhead NEIL1-s isoform and the human NEIL1 were assessed using PolyPhen2 and SIFT. None of the mismatches (Table S1) were predicted to have deleterious effects, but the replacement of an arginine with cysteine on position 370 did have a very low predicted tolerance score (SIFT score 0.16). As the amino acid at position 370 is located in the carboxy-terminal end, it is not possible to predict changes on interaction with other residues, but in general it can be problematic to incorporate or eliminate cysteines as they might contribute to disulfide bridge formation. None of these seemed to interact in a way that would cause significant changes. However, in the human NEIL1 protein, Arg95 and Arg122 may be in range to interact weakly with the DNA backbone, whereas the bowhead equivalents (Pro95 and His122) may be too short for these hypothetical interactions. For human NEIL1, this interaction with the DNA backbone could be an advantage, meaning improved DNA binding and thereby possibly faster initiation of enzyme activity.

In our study, we also RT-PCR cloned two partial NEIL1 cDNA sequences from Greenland shark muscle RNA and spiny dogfish spleen RNA. The cDNAs were of 938 bp for the Greenland shark NEIL1 (Fig. S5) and 925 bp for the spiny dogfish counterpart (Fig. S6). The two predicted shark NEIL1 protein sequences were aligned (Fig. S7). The amino acid sequences covered 306 residues and contained the amino-terminal and middle part of the NEIL1 protein, whereas part of the carboxy-terminal end was lacking. The alignment of the shark NEIL1 sequences revealed a high homology, with an amino acid identity of 74%. This homology was particularly high within the characteristic motifs PE-helix, H2TH, and the zincless finger, as well as in the recognition loop (Fig. S7). The PE-helix was conserved between the two shark NEIL1 sequences and the human and bowhead counterparts. Further, the helix-two-turn-helix, the recognition loop, and the zincless finger motifs were highly conserved between the shark NEIL1 sequences. However, when comparing the overall homology between the shark and human and bowhead NEIL1 sequences, the value was low, with an amino acid identity of 57%.

### 3.2. Phylogenetic tree

To investigate the genetic relationship of NEIL1 from different species, we generated a phylogenetic tree based on the species’ NEIL1 sequences, which included human, land-living mammals, and baleen and toothed whales (Fig. 1B). As expected, the bowhead whale NEIL1 was closely related to the sequence for NEIL1 in the minke whale, which is a baleen whale. The bowhead whale NEIL1 was shown to be least genetically related to the mouse and human NEIL1 proteins. The other whales and dolphins in this phylogenetic tree were all tooth whales and, as expected, were genetically further from the bowhead NEIL1 than other baleen whales.

### 3.3. Spatial expression of NEIL1 transcript

The spatial expression of NEIL1 mRNA in bowhead whale was determined through RT-qPCR analysis. A differential NEIL1 mRNA expression was seen in the four analyzed organs and tissues with a high expression in eye, moderate expression in liver and kidney, and low expression in muscle (Fig. 3). The monitored NEIL1 mRNA expression was only indicative as few individuals were sampled and some displayed varying RNA quality. NEIL1 mRNA expression was only determined in eye, liver, kidney and muscle. NEIL1 deficiency has been shown to be associated with head and neck squamous cell carcinoma, non-small cell lung cancer, and colorectal cancer. Therefore, it would have been interesting and relevant to examine NEIL1 transcript expression in bowhead neck, lung and colorectal tissues. However, these organs and tissues were not sampled from bowhead, and unfortunately, it is not trivial to obtain samples from arctic animals. The bowhead’s conservation status is listed as “least concern” overall, but near Greenland populations are endangered. This limits the possibilities to obtain tissue and organs samples from this species. Hazra et al. [15] performed a NEIL1 expression analysis of human organs, reporting the highest transcript expression in the liver, pancreas, and thymus; a moderate expression in the brain, spleen, prostate, and ovary; and a relatively low level in the testis and leukocytes. A comparison between expression of different human glycosylases in the same tissues implied that these enzymes can function as each other’s backup and/or that different tissues have different mechanistic requirements in DNA repair. These different mechanistic requirements in DNA have been observed in the brain, which has low OGG1 expression but relatively high expression of NEIL1 [15]. Additionally, expression has been found to be cell-cycle regulated, with NEIL1 transcript reaching its maximum approximately 24 h after release from serum starvation and NEIL1 protein reaching its maximum 27 h after release, where the maximum levels are 4- and

---

**Fig. 3.** Spatial expression of bowhead NEIL1 determined by RT-qPCR.


6-fold higher than G-phase, respectively. Both time points corresponded to the S-phase of the cell cycle, in which DNA replication takes place, suggesting a link between NEIL1 activity and cell replication [15]. In addition, high levels of oxidative stress have been shown to have a positive effect on NEIL1 expression, as a rational response to the expected increase in oxidized DNA bases [24].

3.4. Analysis of A-to-I editing of NEIL1 mRNA

As A-to-I RNA editing of the human NEIL1 transcript has been shown to create a novel isoform of the NEIL1 protein with enzymatic properties different from the wildtype protein, we wanted to investigate the A-to-I editing of the bowhead NEIL1 mRNA. We have previously reported on A-to-I RNA editing of NEIL1 in bowhead retina and optical nerve [48]. We extended the analysis with more tissues and demonstrated editing in the eye, liver, and retina. However, no editing of NEIL1 was observed in the muscle and kidney (Fig. 4A). The editing ratio estimated from the results of Sanger sequencing of the PCR products showed 100% editing in the second adenosine in the AAa codon for K242 and approximately 30% editing in the adenosine in the third position in the eye NEIL1 transcript. Lower editing degrees were found in the retina (46%) and liver (33%), as shown in Fig. 4B.

To search for the NEIL1 protein isoform arising from the A-to-I editing of NEIL1 mRNA, we performed mass spectroscopy (MS) analyses, which included a global analysis and a targeted analysis for NEIL1 in bowhead liver and eye tissue. However, we did not identify any NEIL1 isoforms among the identified peptides in the global MS analysis. This might be explained by a low NEIL1 protein expression in liver and eye. Similarly, the targeted MS analysis failed to identify any NEIL1 peptide sequences.

Yeo et al. [24] discovered that a codon shift from AAA (Lys) to AIA (Arg), as a result of A-to-I RNA editing of the NEIL1 pre-mRNA, leads to a functional variant with markedly different substrate specificity. The Arg-242 variant excises thymine glycol (Tg) 30 times slower than the Lys-242 variant, while Gh and Sp are excised approximately three and two times faster, respectively [24]. These findings were supported by Minko et al. [27], who showed that edited NEIL1 excises 5-hydroxycytosine (5-OHC) five times as efficiently but observed no significant difference in excision for the foraminidopyrimidines FaPyA and FaPyG when all lesions (and Tg) were present in the same high-molecular weight genomic DNA. The difference in substrate specificity for Tg between the two NEIL1 isoforms seems to be associated with base excision rather than substrate recognition, as binding affinities are similar [27]. The codon for K242 may also be edited at the third adenosine position (AAA to AAI), but this does not result in an amino acid substitution.

3.5. Expression constructs and production of recombinant NEIL1—bowhead and human

With the aim of comparing the enzymatic properties of the bowhead and the human NEIL1 we cloned and expressed both recombinantly. Bowhead and human NEIL1 cDNAs were RT-PCR cloned as described. For expression of recombinant NEIL1, the coding sequences were codon optimized for expression in E. coli, and a sequence encoding a 6 x His-tag was added to the carboxy-terminal end. The synthetic NEIL1 sequence was cloned into the expression vector pET30a using the restriction enzymes Nde I and Hind III (performed by GenScript). The codon optimized NEIL1 sequences from bowhead and human were cloned into expression vectors and transformed into E. coli BL21 (DE3) cells. Expression of recombinant protein was induced by IPTG. Cells were lysed through sonication and recombinant NEIL1 was purified from crude extracts. The expression of bowhead NEIL1 was shown using SDS-PAGE and Western blotting (Fig. S1). The apparent molecular mass of the recombinant NEIL1 was approximately 50 kDa, which agrees with the theoretical calculated value. Recombinant NEIL1 protein was purified from cell lysates using Ni-chelate chromatography and concentrated using a Hi-TRAP column. Representative elution profiles from the Ni-chelate affinity purification were shown for both bowhead and human NEIL1 (Figs. S2A–E). The purity of the bowhead recombinant NEIL1 was low, as depicted in Figs. S2A and B. The purity of the human NEIL1 preparations were >95%, as judged by Coomassie staining (Figs. S2C and D). The concentrated fractions from the HiTRAP columns were pooled and subsamples were analyzed using SDS-PAGE (Fig. S2E). The concentration of recombinant protein was determined and these samples were used for NEIL1 activity assays.

3.6. Enzymatic activity of recombinant NEIL1 proteins

To examine whether the recombinant bowhead and human NEIL1 proteins displayed enzymatic activity, we conducted a NEIL1 incision assay. Recombinant NEIL1 was incubated with a double-stranded DNA substrate containing an internal 11 nt bubble structure with a 5-OHU lesion in the center of the bubble. Enzymatic activity of recombinant NEIL1 resulted in removal of the 5-OHU lesion and cleavage of the DNA backbone. Due to a radiolabel on the 5'-end of the same strand as the 5-OHU lesion, repair products generated by NEIL1 activity could be separated from the substrate in a denaturing gel and visualized through phosphoimaging. Both the bowhead and human NEIL1 proteins displayed incision activity towards the 5-OHU, containing bubble substrate with similar efficiency (Fig. 5). As expected, an increased number of NEIL1 proteins resulted in an increased amount of product (Fig. 5A), which was visualized through the quantification analysis (Fig. 5B). Hence, the functionality of the recombinant proteins was confirmed. However, it is still very difficult to compare NEIL1 enzymatic activities between the two species. Such a comparison would at least demand a significantly more detailed enzyme kinetic study, including different substrates and determination of kinetic parameters. We plan to include such analyses in future studies.
3.7. Cloning and characterization of bowhead NEIL2 cDNA and NEIL3 genomic sequence

NEIL2: Together with NEIL1, the mammalian NEIL2 and NEIL3 proteins belong to the formamidopyrimidine-DNA glycosylase/endonuclease VII superfamily (Fpg/Nei). Members of this family are characterized based on their typical structures, which include N-terminal and C-terminal domains connected by a flexible linker. The N-terminal domain is a two-sheet beta-sandwich consisting of eight antiparallel strands. The PE-helix is an alpha-helix containing the catalytic Pro2 and Glu3 residues in its N-terminus. The initial methionine is removed upon polypeptide maturation, therefore, the numbering starts at the adjacent proline residue. The carboxy-terminal element contains a DNA-binding helix-two-turn-helix motif, characteristic of the Fpg/Nei superfamily, and a Cys4-type zinc finger motif consisting of two antiparallel beta-strands. The NEIL2 protein bears the closest resemblance to NEIL1, at only 58 residues shorter. Both NEIL1 and NEIL2 prefer oxidized pyrimidines as substrates and bind AP sites by forming a Schiff base between the N-terminal proline and DNA [56]. However, the specificity for substrates varies between the two enzymes. NEIL2 prefers substrate lesions within dsDNA bubbles and D-loops and has low affinity for 5-OHC, Tg, and 8-oxoG, instead favoring 5-OHU and DHU [15,57]. NEIL2 also shares its ability to function on ssDNA and dsDNA bubble structures with NEIL1. In addition, given NEIL2’s binding to RNA polymerase II, among other transcription-associated proteins, it is thought to function in transcription [58].

In our study, we RT-PCR cloned the bowhead NEIL2 cDNA using bowhead retina RNA. An amino acid alignment between NEIL2 proteins from bowhead, minke whale, human, and mouse revealed highly conserved sequence stretches, especially in the N-terminal and carboxy-terminal parts of NEIL2, whereas a variable sequence was found from amino acids 61 to 125 (Fig. S3). This particular sequence of NEIL2 is a disordered region unique to this member of the NEIL gene family [59]. The disordered region is located within the glycosylase fold covering beta-sheets β-3, β4, and β-5. Compared with the human counterpart, the lack of six amino acid residues in the whale NEIL2 sequences is notable. An insert of 8–10...
residues (residues 155–165 in the human NEIL2) unique to NEIL2 is located between β-strands β 5 and β 6. In both the human and the bowhead sequences, the insert contains three lysines and two arginines. The amino acid sequence for this insert sequence could influence activity on the unique substrate range for NEIL2 [39].

Through comparison of the four NEIL2 sequences, seven bowhead-specific amino acid variations were observed: Cys111, Glu122, Phe156, Ala182, Ser302, Phe303, and Arg304 in the bowhead NEIL2 sequence (Fig. S3). An amino acid identity of 71% was observed between the bowhead and human NEIL2 proteins. The low amino acid identity between bowhead NEIL2 and the other NEIL2 proteins included in the alignment is not surprising as low homology for members of the Fpg/Nei superfamily has been reported previously [8]. However, a very high homology was seen within the PE-helix and both Pro1 and Glu2 were completely conserved in the four species (Fig. S3). The alpha amino group of the Pro1 residue acts as a nucleophile in the DNA glycosylation reaction and forms a Schiff base with the C1’ atom [60,61]. Further, the Glu2 residue necessary for the DNA glycosylation activity was conserved in all four NEIL2 protein sequences [62,63].

In addition, conservative motifs were located in the short loops connecting beta-strands 2 and 3 of the N-terminal domain. The lysine59 (Lys59) and lysine60 residues, which are involved in the coordination of the 5’-phosphate group of the damaged deoxynucleoside, were conserved in the NEIL2 sequences from all four species [61].

In the carboxy-terminal end, the conservative motifs were a DNA-binding helix-two-turn-helix motif and a zinc finger, whereby the latter made contact with the major groove of DNA. The helix-two-turn-helix motif was highly conserved with only one amino acid difference at position 233 (238 in the human NEIL2 sequence), with a phenylalanine present in the whale NEIL2 and a tryptophan in the human sequence. Zinc finger motifs in NEIL proteins may vary. The NEIL2 protein contains an unusual CHCC zinc finger motif, which has been shown, through inductively coupled plasma mass spectrometry, to contain a zinc atom [29]. Interestingly, the homology within the zinc finger motif—amino acids 285 to 313 in the bowhead NEIL2 sequence—was moderate. In the bowhead NEIL2 sequences, the zinc finger protein CHCC-type motif in the C terminus was constituted of Cys-286, His-290, Cys-310, and Cys-313, which are candidate residues for coordinating Zn$^{2+}$.

Several post-translational modification sites have been predicted in the human NEIL2 protein, among which are two phosphorylation sites: Ser58 and Thr485 (Fig. S4). The Ser58 residue has been shown to be conserved in the human mouse, rat, pig, and cow protein NEIL2 sequences. However, in the bowhead NEIL2 protein, position 68 is occupied by an asparagine residue. The amino acid position 70 seems to be much more variable and in the species mentioned above. Ser, Leu, Arg, and Glu, and a Thr residue in the human NEIL2 protein. Hence, no phosphorylation is possible on the amino acid positions 68 and 70 in the bowhead NEIL2 sequence.

Bakhiet et al. [28] demonstrated that Lys49 and Lys153 are the major acetylation sites in NEIL2. Acetylation of Lys49 was shown to be conserved among Nei orthologs and its mutation to Arg to inactivate both base excision and AP lyase activities, while acetylation of Lys153 had no effect. Reversible acetylation of Lys49 was found to regulate the repair activity of NEIL2 in vivo [28]. Lys49 and Lys153 were conserved in the bowhead NEIL2 protein.

Recent reports have suggested that some mutations in the human NEIL2 are associated with cancer development. Kahkarova et al. [64] summarized the NEIL2 polymorphism associations with different cancer types and reported the biochemical characterization of two human NEIL2 variants, R103W and P304T, both as polymorphisms in the general population. The P304T variant had 5-fold lower catalytic activity compared with the wild-type NEIL2 enzyme. Eckenroth et al. [59] presented the first crystal structure of NEIL2 and reported three cancer variants of human NEIL2: S140 N, G230W, and G303R. In addition, Benitez-Buelga et al. [65,66] identified a NEIL2 SNP as cancer risk modifier for BRCA2 mutation carriers. In this current study, all three amino acids (S140, G230, and G303) were conserved in the bowhead NEIL2 protein.

**NEIL3**: Despite several attempts to clone the full-length cDNA encoding bowhead NEIL3, we did not succeed. However, we managed to PCR clone a partial bowhead NEIL3 sequence. The partial sequence lacked 51 amino acids in the amino-terminal end of the encoded protein. The deduced amino acid sequence for the bowhead NEIL3 protein was in alignment with the NEIL3 counterparts from narwhale, cow, and human (Fig. S4). Across the four compared species, a high degree of sequence conservation was observed for NEIL3. Seven bowhead-specific amino acid differences were found in the alignment of NEIL3 protein sequences: Ala238 (Val287 in human), Thr261 (Met310 in human), Glu282 (Lys331 in human), Leu299 (Ser348 in human), Ala339 (Thr388 in human), Phe435 (Ser485 in human), and Lys487 (Arg537). Only one of these amino acid substitutions, Q282, was found within important
conserved motif sequences, in this case zinc-finger motif 2.

NEIL3 is 215 amino acid residues longer than NEIL1 and has a valine in place of the catalytic Pro-2 found in NEIL1 and NEIL2. NEIL3 prefers much the same oxidized pyrimidines as NEIL1 and can function on both dsDNA bubble strand and ssDNA, and therefore also during replication, but where NEIL1 seems to prefer lesions in dsDNA around the replication fork, NEIL3 seems to prefer ssDNA. In addition, NEIL3 co-precipitates with the replisome. Together, these characteristics could indicate that NEIL3 and NEIL1 act together to remove oxidized bases from the replication fork in replication-coupled repair [67]. Biochemical evidence from a study by Albelazi et al. [67] suggested that NEIL3 is the primary DNA glycosylase associated with the excision of oxidized bases within ssDNA and dsDNA. NEIL1 was shown to complement by serving as the primary DNA glycosylase associated with dsDNA prior to replication fork unwinding. Additionally, NEIL3 was found to exhibit poor lyase activity and use mostly β,ε-elimination, contrary to NEIL1 and NEIL2, which utilize β,δ-elimination [58,67]. Genetic variants in the human NEIL3 gene have been associated with different diseases: NEIL3 polymorphisms have been associated with cancer types such as glioblastoma and prostate cancer [68,69], SNPs have been associated with impulsivity [70], and genetic variants in NEIL3 have been associated with increased risk of myocardial infarction [71].

4. Conclusion

NEIL1 is an essential member of the group of DNA repair enzymes involved in the base excision repair pathway (BER). NEIL1 can detect and remove DNA damage generated by oxidation of adenine, guanine, and thymine. Previous studies have detected at least 996 NEIL1 polymorphisms, which range from none to severe effects on enzyme activity. The reduction in DNA repair competence due to impaired catalytic activity can increase the risk of developing cancer 20–30-fold [58]. Both NEIL1 deficiency and haploinsufficiency may result in predisposition to cancer and also in metabolic syndrome [54,72–74]. Large whales, such as the bowhead whale, with over 1000 times more cells than humans, do not exhibit an increased cancer risk [75]. This is possibly due to the existence of natural mechanisms that can suppress cancer more effectively in these animals, whereby one such mechanism could be DNA repair. Analyses of the genome sequence of the bowhead whale have identified genes under positive selection and bowhead-specific mutations in genes linked to cancer and aging [76]. We have hypothesized that the bowhead DNA repair enzyme NEIL1 has a greater capacity and perhaps longer lasting expression and enzymatic activity during life-span than its human counterpart. Here, we have presented the cloning and characterization of NEIL1, NEIL2, and NEIL3 from bowhead, which are all important DNA repair enzymes with activity that likely prevents the development of cancer. Additionally, we have demonstrated enzymatic activity of recombinant bowhead NEIL1 expressed in E. coli. Our study represents the initial steps to clarify our hypothesis, where we propose that NEIL1 is connected to cancer resistance and increased longevity in the bowhead whale, that is, that NEIL1 activity would be expected to be higher in the bowhead than in humans. Based on the alignment analysis of the bowhead and the human NEIL1, there is limited evidence to confidently predict a difference in activity. Most notable is the shift of a hydroxyl group in the recognition loop from position 246 to 249. This could bring it in closer proximity to the essential lys/Arg-242 amino/guanidino, thereby affecting binding interactions. However, both the 242 residue and the 249 hydroxyl group were protonated at physiological pH, meaning that their ability to function as hydrogen bond donors may cause interference or repulsion of each other or perhaps the establishment of additional hydrogen bonds to the substrate, resulting in more stable binding. This could be clarified through site-directed point mutation of Ser246/Thr249 codons. In further studies, we will focus on a more comparative enzyme kinetics analysis of the bowhead and the human recombinant NEIL1 proteins, both with respect to substrate specificity and determination of kinetic parameters, for example, affinity towards different substrates.

Sequences

The sequences of the bowhead NEIL1 and NEIL2 cDNAs were submitted to GenBank under the following accession numbers: NEIL1 cDNA transcript variant 1 (GenBank ID: MZ055365), NEIL1 cDNA transcript variant 2 (GenBank ID: MZ055366), NEIL1 cDNA transcript variant 3 (GenBank ID: MZ055370), and NEIL2 cDNA (GenBank ID: MZ055369). A partial bowhead NEIL3 cDNA sequence was submitted to GenBank and obtained accession number OL770101. Partial NEIL1 cDNA sequences for Greenland shark and spiny dogfish were submitted to GenBank under the accession numbers GenBank ID: MZ055367 and GenBank ID: MZ055368, respectively.

Author contributions and Agreement

Conception of the work: KL, TS, and CMH. Experimental work and collection of data: KL, SH, RML, CMH, MPHJ, and JFS. Analysis of data: KL, SHL, RML, and CMH. TS, Writing of manuscript: KL, RML, CMH and TS. All authors have approved the final version of the manuscript.

Acknowledgements

We thank laboratory technicians Hanne Jørgensen, Ulla Birk Henriksen, and Bente Flügel for their excellent technical assistance. We also thank the Brødrene Hartmanns Fond for providing financial support for Knud Larsen.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.biochi.2022.10.014.

References


[28] A. Prakash, V.B. Cao, S. Doubilé, Phosphorylation sites identified in the human DNA glycosylase are potential targets for the JNK1 kinase, PLoS One 11 (2016), e0157860.


