Sample: Bioinformatics - Comparative Analysis

Title: Comparative analysis of Endonuclease VIII genes and encoded proteins from *Escherichia coli* and *Mus musculus.*

Objectives: The aim of this work is obtain nucleotide and amino acid sequences of Nei/Fpg family DNA glycosylases of *Escherichia coli* and *Mus musculus* and perform comparative analysis of the respective mRNAs and proteins.

Introduction:

Base excision repair is an ancient mechanism that protects both prokaryotic and eukaryotic cells from oxidative DNA damage and resulting mutagenesis (Jacobs and Schar, 2012). The enzymes named DNA glycosylases initiate this pathway by recognition and excision of damaged bases. In such a way an apurinic/apurimidinic site (AP site) is created. Further the activities of AP endonuclease, as well as DNA polymerase and ligase enzymes are needed to complete the repair process (Geacintov and Broyde, 2011). DNA glycosylases are divided into two classes, monofunctional and bifunctional, depending on the possibility to cleave DNA backbone. On the basis of structural features DNA glycosylases are divided into four groups, the uracil DNA glycosylases, the helix-hairpin-helix glycosylases, the 3-methyl-purine glycosylases, and the endonuclease VIII-like glycosylases (Kelley, 2011). The latter group forms Nei/Fpg family and consists of *Escherichia coli* Nei gene and three mammalian NEIL (Nei-like) genes (Golan *et al*., 2005). The unique feature of Nei/Fpg family members is specificity towards not only double strand DNA but single strand DNA as well as bubble DNA (Hooten *et al.,* 2012). These proteins usually contain N-terminal catalytic domain, a helix-two turn-helix (H2TH) motif and C-terminal DNA-binding region (Liu *et al.*, 2010). In this work the sequences of bacterial and mammalian homologs of Nei/Fpg family DNA glycosylases will be analysed. The conserved domains within these proteins will be determined.

Methods.

The nucleotide sequences of *E.coli* Nei gene and murine NEIL1, NEIL2, NEIL3 genes as well as the aminio acid sequences of the encoded proteins were obtained from NCBI database. The nucleotide and amino acid sequences of bacterial and murine orthologs were compared using BLAST algorithms. Conserved protein domains were determined within bacterial and murine DNA glycosylases. Structure of bacterial and murine DNA glycosylases were compared.

Results.

Nucleotide sequence of *E.coli* Nei DNA glycosylases was obtained using NCBI (National Center of Biotechnology) database (Figure 1). This sequence was used to find murine homologs of bacterial Nei gene.

ATGCCTGAAGGCCCGGAGATCCGCCGTGCAGCGGATAACCTGGAGGCGGCGATCAAAGGCAAAC
CACTAACTGATGTCTGGTTTGCCTTCCCGCAGTTAAAACCTTATCAATCACAACTTATCGGTCAACAC
GTTACCCATGTGGAAACGCGTGGTAAGGCGTTGTTAACTCATTTTTCCAACGACTTAACGCTCTACA
GCCATAATCAGCTTTACGGCGTCTGGCGCGTGGTTGATACCGGCGAAGAGCCGCAGACCACGCGA

GTATTGCGGGTAAAACTGCAAACGGCTGACAAAACCATTCTGCTTTATAGCGCCTCGGATATTGAG
ATGTTGACCCCGGAACAACTGACCACGCATCCGTTTTTACAACGCGTTGGTCCCGATGTGCTGGATC
CGAATCTGACGCCGGAGGTGGTGAAAGAACGATTATTGTCGCCGCGCTTTCGTAACCGTCAGTTTG
CTGGATTACTGCTCGATCAGGCGTTTCTGGCTGGGCTTGGCAATTATTTGCGGGTGGAGATCCTCT
GGCAGGTTGGGTTGACTGGAAATCATAAAGCGAAAGATCTCAATGCGGCGCAACTGGATGCACTC
GCACACGCGTTACTGGAGATTCCTCGATTTTCCTACGCTACGCGGGGGCAGGTGGATGAGAATAA
GCATCATGGGGCGCTGTTTCGCTTTAAGGTTTTTCATCGAGATGGCGAACCGTGCGAACGTTGTGG
CAGCATCATTGAGAAAACCACGCTGTCATCTCGCCCGTTTTACTGGTGCCCTGGCTGCCAGCACTAG

Figure 1. Nucleotide sequence of Endonuclease VIII (Nei) from *Escherichia coli* strain K12.

BLAST search was performed to find out whether nucleotide sequences of Nei genes of E.coli and M.musculus share significant level of similarity. BLAST parameters were optimized to decrease stringency. The results of the Blast search revealed no sequences or regions homologous to Nei gene in mouse genome (Figure 2). Alignment of coding sequence of Nei with coding sequences of each murine Nei-like enzymes revealed no significant homology as well.

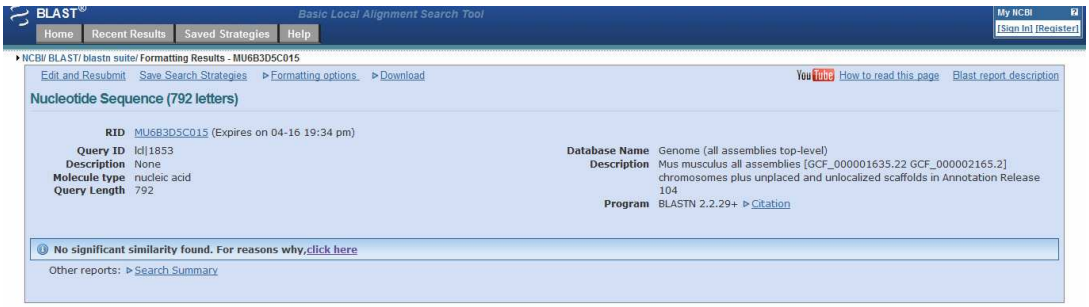

Figure 2. Results of the BLASTn search of Nei homologs within mouse genome.

To perform comparative analysis of bacterial and murine Nei DNA glycosylases amino acid sequences of respective proteins were obtained from NCBI database (Figure 3).

(a)
MPEGPEIRRAADNLEAAIKGKPLTDVWFAFPQLKPYQSQLIGQHVTHVETRGKALLTHFSNDLTLYSHN
QLYGVWRVVDTGEEPQTTRVLRVKLQTADKTILLYSASDIEMLTPEQLTTHPFLQRVGPDVLDPNLTPE
VVKERLLSPRFRNRQFAGLLLDQAFLAGLGNYLRVEILWQVGLTGNHKAKDLNAAQLDALAHALLEIPR
FSYATRGQVDENKHHGALFRFKVFHRDGEPCERCGSIIEKTTLSSRPFYWCPGCQH
(b)
MPEGPELHLASHFVNETCKGLVFGGCVEKSSVSRNPEVPFESSAYHISALARGKELRLTLSPLPGSQPPQ
KPLSLVFRFGMSGSFQLVPAEALPRHAHLRFYTAPPAPRLALCFVDIRRFGHWDPGGEWQPGRGPCVL
LEYERFRENVLRNLSDKAFDRPICEALLDQRFFNGIGNYLRAEILYRLKIPPFEKARTVLEALQQCRPSPEL
TLSQKIKAKLQNPDLLELCHLVPKEVVQLGGKGYGPERGEEDFAAFRAWLRCYGVPGMSSLRDRHGRTI
WFQGDPGPLAPKGGRSQKKKSQETQLGAEDRKEDLPLSSKSVSRMRRARKHPPKRIAQQSEGAGLQQ
NQETPTAPEKGKRRGQRASTGHRRRPKTIPDTRPREAGESSAS
(c)
MPEGPSVRKFHHLVSPFVGQKVVKTGGSSKKLHPAAFQSLWLQDAQVHGKKLFLRFDPDEEMEPLNS
SPQPIQGMWQKEAVDRELALGPSAQEPSAGPSGSGEPVPSRSAETYNLGKIPSADAQRWLEVRFGLFG
SIWVNDFSRAKKANKKGDWRDPVPRLVLHFSGGGFLVFYNCQMSWSPPPVIEPTCDILSEKFHRGQAL

EALSQAQPVCYTLLDQRYFSGLGNIIKNEALYRARIHPLSLGSCLSSSSKEALVDHVVEFSKDWLKDKFQG
KERHTQIYQKEQCPSGHQVMKETFGPPDGLQRLTWWCPQCQPQLSSKGPQNLPSS

(d)

MVEGPGCTLNGEKIRARVLPGQAVTGVRGTALQSLLGPAMSPAASLADVATSAAPMNAKDSGWKLL
RLFNGYVYSGVETLGKELFMYFGPRALRIHFGMKGSILINPREGENRAGASPALAVQLTRDLICFYDSSVE
LRNSVESQQRVRVMEELDICSPKFSFSRAESEVKKQGDRMLCDVLLDQRVLPGVGNIIKNEALFDSGLH
PAVKVCQLSDKQACHLVKMTRDFSILFYRCCKAGSAISKHCKVYKRPNCDQCHSKITVCRFGENSRMTY
FCPHCQKENPQCVQVCQLPTRNTEISWTPRGEDCFTDSVARKSEEQWSCVVCTLINRPSAKACDACLT
TRPLDSVLKNRENSIAFNNLVKYPCNNFENTHTEVKINRKTAFGNTTLVLTDLSNKSSALARKKRANHTID
GESQMFLPTDIGFSDSQHPSKEGINYITQPSNKVNISPTVCAQSKLFSSAHKKFKPAHTSATELKSYNSGL
SNSELQTNRTRGHHSKSDGSPLCKMHHRRCVLRVVRKDGENKGRQFYACSLPRGAQCGFFEWADLSF
PFCRHGKRSIMKTVLKIGPNNGKNFFVCPLEKKKQCNFFQWAENGPGMEIVPGC

Figure 3. Amino acid sequence of Nei from *Escherichia coli* strain K12 (a) and *Mus musculus* NEIL1 (b), NEIL2 (c) and NEIL3 (d) proteins.

In contrast to *E.coli* there are three Nei-like genes within mouse genome. They are located on different chromosomes, contain from 5 to 10 exons and vary in length of mRNAs and encoded proteins. These data are summarized in table 1.

Table 1. Data referred to sequences of Nei DNA glycosylase orthologs.

	Nei	NEIL1	NEIL2	NEIL3
Number of nucleotides in CDS	792 bp	1170 bp	990 bp	1820 bp
Number of amino acids in protein	263 aa	389 aa	329 aa	606 aa
Chromosomal location	-	Chromosome 9	Chromosome 14	Chromosome 8
Number of exons	-	9	5	10

Amino acid sequence of bacterial Nei protein was used for search of homologous sequences within mouse proteome. Four sequences similar to bacterial Nei protein were found using pBLAST (Figure 4). Three sequences represented murine NEIL1, NEIL2 and NEIL3 DNA glycosylases and fourth sequence was hypothetical protein (NP 001039008.1) .

The level of identity between Nei and Nei-like proteins was highest for NEIL1 and constituted 42% however this region stretched only for near 50 amino acids. Other murine Nei orthologs, NEIL2 and NEIL3, demonstrated lower identity level of 24%. Within NEIL2 stretch of near 100 amino acid residues homologous to bacterial Nei was found. Amino acid sequence of NEIL3 protein possessed the longest stretch similar to Nei of *E.coli* that comprised near 250 residues.

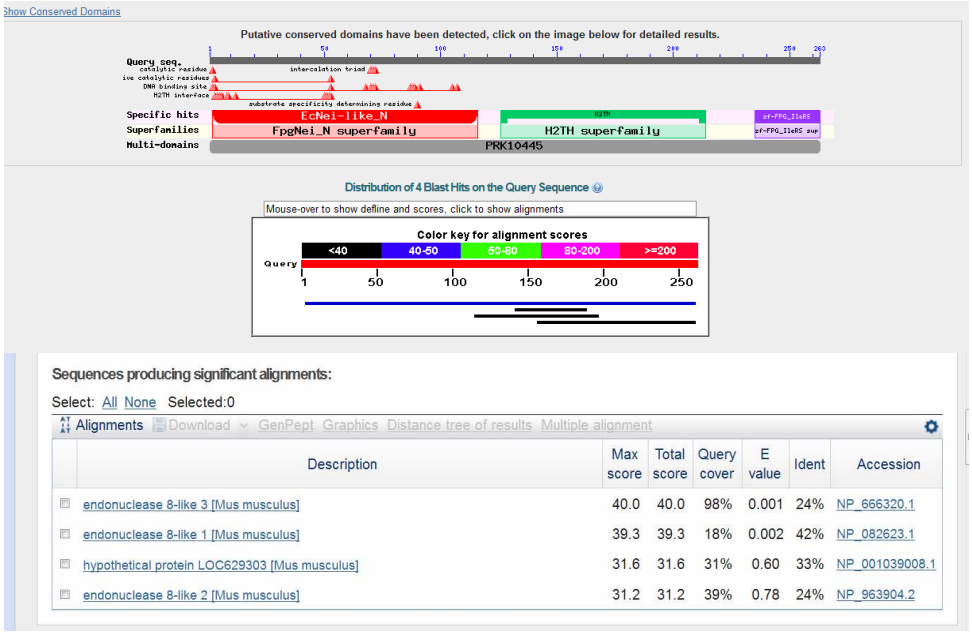

Figure 4. Results of BLASTp comparison of *E.coli* and M.musculus Nei orthologs.

Specialized BLAST tools were used to identify conserved domains within Nei and its orthologs (Figure 5). Three types of conserved domains were found in Nei and Nei-like proteins. Catalytic domain responsible for DNA damage reparation is located at the N-terminus. In the central part of protein molecule H2TH domain that mediates DNA binding was found. Members of Fpg/Nei protein family possess different number of zinc-finger motifs at their C-terminus or even zinc-finger motifs of different types. For details see Figure 5. Murine genome contains three Nei-like genes, NEIL1-3, that are paralogs. Paralogs are genes generated by duplication within a genome. Orthologs as a rule preserve the same function in the course of evolution, paralogs adopt new functions, even if these are related to the original one. Genome of *E.coli* contains related to Nei gene that encodes formamidopyrimidine-DNA glycosylase, MutM. These proteins are paralogous, they share high level of similarity and belong to Fpg/Nei protein family.

According to obtained data NEIL3 has the same domains required for DNA reparation but possesses significantly longer C-terminal tail with two additional GRF zink-finger motifs. Given the presence of enzymatic and DNA-binding domains it is possible to conclude that NEIL3 could perform similar function to other Fpg/Nei family members. However different substrate specificity for NEIL3 could not be ruled out.

Figure 5. Graphical representation of bacterial and murine Nei proteins. The FpgNei is enzymatic domain, H2TH is DNA-binding domain, zfRanB and zfGRF are two types of zinc-finger motifs.

Discussion

In this work *E.coli* gene Nei was used for search of murine genes involved in DNA damage reparation. Such approach that involves search of conserved proteins is widely used to annotate conserved genes in complex genomes. Three paralogous genes NEIL1-3 were identified in mouse genome that is in line with experimental data (Liu et al., 2010). Murine NEIL proteins have similar domain organization: enzymatic domain at the N-terminus, central H2HT domain responsible for DNA binding and additional DNA binding zinc-finger motifs. Apparently NEIL1-3 are resulted from gene duplication and subsequent subfunctionalization. For example expression of NEIL1 is upregulated during S-phase (Dou *et al*., 2008), its intrinsically disordered C-terminal tail regulates stability of the protein (Hegde *et al.,* 2012). In contrast to NEIL1, expression of NEIL2 is not regulated during cell cycle. Instead NEIL2 has mitochondrial localization and contributes to mitochondrial genome stability (Mandal *et al,* 2012). Comparison of NEIL1 and NEIL2 amino acid sequences showed that their N-terminal domains, responsible for DNA reparation differ. NEIL2 has relatively long insert in the middle of the N-terminal catalytic core domain. Recently it was shown that NEIL1 and NEIL2 have different substrate specificity (Mandal *et al.,* 2012; Katafuchi *et al.,* 2004), that could be a result of specific features of catalytic domain. The catalytic activity of NEIL3 was questioned due to presence of unconserved proline residue in its catalytic pocket. Recent data show that NEIL3 is capable to repair DNA *in vivo* and *in vitro* (Liu *et al*., 2010).

Analysis of protein amino acid sequences permits us to search for structural differences between paralogous molecules and find structural basement of their functional diversity. Bioinformatical tools that comprise comparison of protein primary structures are robust and widely used in molecular biology and biochemistry. In this work we traced common and specific features of Fpg/Nei proteins from an ancestral single prokaryotic molecule to three specialized murine proteins.

References

1. Jacobs, M. & Schär, M. (2012) DNA glycosylases: in DNA repair and beyond. *Chromosoma*, **121**, 1–20.
2. Geacintov, N. & Broyde, S. (2011) *The Chemical Biology of DNA Damage*. John Wiley & Sons, Washington.
3. Kelley, M. (2011) *DNA Repair in Cancer Therapy: Molecular Targets and Clinical Applications.* Academic Press.
4. Golan, G., Zharkov, D., Feinberg H. *et al.* (2005) Structure of the uncomplexed DNA repair enzyme endonuclease VIII indicates significant interdomain flexibility. *Nucleic Acids Research*, **33**, 5006–5016.
5. Hooten, N., Fitzpatrick, M., Kompaniez, K. *et al.* (2012) Coordination of DNA repair by NEIL1 and PARP-1: a possible link to aging. *Aging*, **4,** 674-685.
6. Liu, M., Bandaru, V., Bond, J. *et al.* (2010) The mouse ortholog of NEIL3 is a functional DNA glycosylase in vitro and in vivo. *PNAS*, **107**, 4925-4930.
7. Dou, H., Theriot, C., Das, A., *et al*. (2008) Interaction of the human DNA glycosylase NEIL1 with proliferating cell nuclear antigen. The potential for replication-associated repair of oxidized bases in mammalian genomes. *Journal of Biological Chemistry*, **6,** 3131-3140.
8. Hegde, M., Tsutakawa, S., Hegde, P. et al. (2013) The disordered C-terminal domain of human DNA glycosylase NEIL1 contributes to its stability via intramolecular interactions. *Journal of Molecular Biology*, **13**, 2359-2371.
9. Mandal, S., Hegde, M., Chatterjee, A. *et al.* (2012) Role of human DNA glycosylase Nei-like 2 (NEIL2) and single strand break repair protein polynucleotide kinase 3'-phosphatase in maintenance of mitochondrial genome. *Journal of Biological Chemistry,* **4**, 2819-2829.
10. Katafuchi, A., Matsubara, M., Terato, H. *et al.* (2004) Damage specificity of human DNA glycosylases for oxidative pyrimidine lesions. *Nucleic Acids Symp Ser*, **48**, 175-176.
11. Hazra, T., Kow, Y., Hatahet, Z. *et al*. (2002) Identification and characterization of a novel human DNA glycosylase for repair of cytosine-derived lesions. *Journal of Biological Chemistry,* **34**, 30417-30420.