DNA Repair 8 (2009) 643–653

Contents lists available at ScienceDirect

DNA Repair

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

# Plant and fungal Fpg homologs are formamidopyrimidine DNA glycosylases but not 8-oxoguanine DNA glycosylases

Scott D. Kathe[a,1], Ramiro Barrantes-Reynolds[a,1], Pawel Jaruga[b], Michael R. Newton[c], Cynthia J. Burrows[c], Viswanath Bandaru[a,1,2], Miral Dizdaroglu[b], Jeffrey P. Bond[a,*], Susan S. Wallace[a,*]

[a] *Department of Microbiology and Molecular Genetics, The Markey Center for Molecular Genetics, University of Vermont, Stafford Hall, 95 Carrigan Dr., Burlington, VT 05405-0068, United States*
[b] *Chemical Science and Technology Laboratory, National Institute of Standards and Technology, Gaithersburg, MD 20899, United States*
[c] *Department of Chemistry, University of Utah, Salt Lake City, UT 84112-0850, United States*

ARTICLE INFO

*Article history:*
Received 6 November 2008
Received in revised form 19 December 2008
Accepted 20 December 2008
Available online 12 February 2009

*Keywords:*
Base excision repair
DNA glycosylases
Phylogenetics
Formamidopyrimidine DNA glycosylase
Endonuclease VIII (Nei)
*Arabidopsis thaliana*
*Candida albicans*
Guanidinohydantoin
Spiroiminodihydantoin

ABSTRACT

Formamidopyrimidine DNA glycosylase (Fpg) and endonuclease VIII (Nei) share an overall common three-dimensional structure and primary amino acid sequence in conserved structural motifs but have different substrate specificities, with bacterial Fpg proteins recognizing formamidopyrimidines, 8-oxoguanine (8-oxoG) and its oxidation products guanidinohydantoin (Gh), and spiroiminodihydantoin (Sp) and bacterial Nei proteins recognizing primarily damaged pyrimidines. In addition to bacteria, Fpg has also been found in plants, while Nei is sparsely distributed among the prokaryotes and eukaryotes. Phylogenetic analysis of Fpg and Nei DNA glycosylases demonstrated, with 95% bootstrap support, a clade containing exclusively sequences from plants and fungi. Members of this clade exhibit sequence features closer to bacterial Fpg proteins than to any protein designated as Nei based on biochemical studies. The *Candida albicans* (Cal) Fpg DNA glycosylase and a previously studied *Arabidopsis thaliana* (Ath) Fpg DNA glycosylase were expressed, purified and characterized. In oligodeoxynucleotides, the preferred glycosylase substrates for both enzymes were Gh and Sp, the oxidation products of 8-oxoG, with the best substrate being a site of base loss. GC/MS analysis of bases released from γ-irradiated DNA show FapyAde and FapyGua to be excellent substrates as well. Studies carried out with oligodeoxynucleotide substrates demonstrate that both enzymes discriminated against A opposite the base lesion, characteristic of Fpg glycosylases. Single turnover kinetics with oligodeoxynucleotides showed that the plant and fungal glycosylases were most active on Gh and Sp, less active on oxidized pyrimidines and exhibited very little or no activity on 8-oxoG. Surprisingly, the activity of AthFpg1 on an AP site opposite a G was extremely robust with a $k_{obs}$ of over 2500 $min^{-1}$.



## 1. Introduction

Cellular DNA is under constant exposure to oxidative damage from endogenous and exogenous sources and these damages must be recognized and repaired quickly and with high fidelity. Endogenous damages include single-strand breaks, deaminations, depurinations, alkylations and oxidative modifications of the DNA bases [1]. Of the more than 25,000 DNA damages a human cell receives per day about 6000 are oxidatively induced base lesions. The consequences of these damages include mutations, cancer and/or cell death. Base excision repair (BER) has evolved to recognize and repair these damages with the first step in repair being recognition of a damaged base by a DNA glycosylase (for reviews see [2–4]). The DNA glycosylase binds to the DNA, cleaves the N-glycosyl bond releasing the damaged base resulting in an apurinic or apyrimidinic (AP) site. Most oxidative DNA glycosylases also cleave the phosphodiester backbone of DNA on the 3′ side of the AP site leaving either a α,β-unsaturated aldehyde (β-elimination)

*Abbreviations:* Fpg, formamidopyrimidine DNA glycosylase; Nei, endonuclease VIII; Nth, endonuclease III; 8-oxoG, 8-oxo-7,8-dihydroguanine; 8-oxoA, 7,8-dihydro-8-oxoadenine; Gh, guanidinohydantoin; Sp, spiroiminodihydantoin; CalFpg, *Candida albicans* (Cal) Fpg DNA glycosylase; AthFpg, *Arabidopsis thaliana* (Ath) Fpg DNA glycosylase; BER, base excision repair; AP site, apurinic or apyrimidinic; ROS, reactive oxygen species; EcoNth, *Escherichia coli* endonuclease III; EcoNei, *Escherichia coli* endonuclease VIII; PCR, polymerase chain reaction; Tg, thymine glycol; 5-OHC, 5-hydroxycytosine; 5-OHU, 5-hydroxyuracil; DHT, 5,6-dihydrothymine; DHU, 5,6-dihydrouracil; UDG, uracil DNA glycosylase; FapyAde, 4,6-diamino-5-formamidopyrimidine; FapyGua, 2,6-diamino-4-hydroxy-5-formamidopyrimidine; 5OHMH, 5-hydroxy-5-methylhydantoin.

* Corresponding authors. Tel.: +1 802 656 2164; fax: +1 802 656 8749.
*E-mail addresses:* jeffrey.bond@uvm.edu (J.P. Bond), susan.wallace@uvm.edu (S.S. Wallace).
[1] Tel.: +1 802 656 2164; fax: +1 802 656 8749.
[2] Current address: Mylan Pharmaceuticals, 3711 Collins Ferry Rd, Morgantown, WV 26505, United States.

1568-7864/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.dnarep.2008.12.013

or a phosphate (β,δ-elimination). Since DNA polymerases require a 3′ hydroxyl group for synthesis the phosphodiesterase activity of an AP endonuclease is required to remove the α,β-unsaturated aldehyde or phosphate and leave a 3′ hydroxyl group. In mammalian cells, polynucleotide kinase is used to remove the phosphate generated by Fpg/Nei family glycosylases in order to leave an extendable 3′ end [5]. DNA polymerase then adds the cognate nucleotide and DNA ligase seals the gap. Eukaryotes use additional accessory proteins to maximize the efficiency of both recognition and repair (for reviews see [6,7]).

*Candida albicans* is the most prevalent human fungal pathogen. Normally it exists as a dimorphic commensal yeast found on humans, however in immune compromised individuals, such as AIDS patients and neonates, *C. albicans* can become an opportunistic pathogen. The human immune response recruits monocytes, dendritic cells and macrophages to the site of infection where genes involved with the production and release of reactive oxygen species (ROS) are induced [8]. Therefore *C. albicans* has to deal with endogenous ROS as well as the exogenous ROS produced by the host immune response. *Arabidopsis thaliana* is a vascular plant and as such it has to contend with the endogenous production of singlet oxygen from photosynthesis, internal cell signaling mediated by ROS as well as ROS from exposure to visible light [9]. Repair of oxidatively induced damage caused by treating reporter plasmid DNA with methylene blue and visible light has been demonstrated using *Arabidopsis* cell extracts [10].

The oxidative DNA glysosylases can be divided into two families based on structural and sequence homology (for reviews see [11,12]). The Nth superfamily contains *Escherichia coli* endonuclease III (EcoNth) and its yeast counterparts Ntg1 and Ntg2, that remove oxidized pyrimidines and formamidopyrimidines [4,6-diamino-5-formamidopyrimidine (FapyAde), 2,6-diamino-4-hydroxy-5-formamidopyrimidine (FapyGua)], MutY that removes A opposite 8-oxoG, Ogg1 that removes 8-oxoG and FapyGua opposite C, and AlkA that removes alkylated bases [2–4,13]. The Fpg/Nei family contains formamidopyrimidine DNA glycosylase (Fpg) also called MutM that removes 8-oxoG and FapyGua opposite C and FapyAde opposite T, and endonuclease VIII (Nei) that removes oxidized pyrimidines and FapyAde. With the exception of MutY and AlkA, which are monofunctional and only exhibit glysosylase activity, the glycosylases listed above are bifunctional and exhibit lyase activity in addition to glycosylase activity [2–4]. Members of the Nth superfamily are found throughout the three major kingdoms. Fpg proteins are primarily bacterial, while Nei and Nei-like glycosylases are found in only some of the γ-proteobacteria, actinobacteria, and metazoans [4] and we now place members of this family in acidobacteria, planctobacteria and cyanobacteria (see Fig. 2).

The BER system has redundancy with respect to overlapping substrate specificity such that the loss of a single DNA glycosylase will usually not result in a complete loss of repair for a specific oxidatively induced lesion. In addition to the glycosylases characterized in this study, *C. albicans* has the following but as yet uncharacterized putative proteins involved with the initial steps of BER: Ogg1, Ntg, Udg, AlkA and two AP endonucleases. Single mutants defective in Ogg1, Ntg1 or Apn1 are not sensitive to oxidizing agents, antifungal drugs nor are they altered in their response to macrophages [14]. *A. thaliana* has an Ogg1, MutY, MutT, Nth, six 3-methyadenine DNA glycosylases, three AP endonucleases in addition to potentially seven different forms of AthFpg that could be produced by alternative splicing [15]. We have characterized AthFpg1 in this study.

An unrooted phylogeny of the Fpg/Nei tree from a previous study from our laboratory [4] exhibited a long edge (having moderate, 52%, bootstrap support) that separated enzymes reported to have specificity for 8-oxoG from those reported to prefer oxidized pyrimidines. The phylogeny, then, supported a simple model according to which only one change in substrate preference occurred during Fpg/Nei family evolution and enzymes were thus parsimoniously named either "Fpg" or "Nei". Here we show that the evolution of substrate specificity in this family is strikingly complex. The best substrates for CalFpg and AthFpg in oligodeoxynucleotides are Sp and Gh; CalFpg also recognizes oxidized pyrimidines. Both enzymes mainly release FapyAde and FapyGua and, to some extent, several pyrimidine-derived lesions from γ-irradiated DNA containing multiple lesions (with distinct preferences) but have little or no activity on 8-oxoG.

## 2. Materials and methods

### 2.1. Sequence identification and phylogeny

Fpg/Nei homologs were identified using the PFAM domain profile (pfam06831 [16]) through the CDD database [17]. CDTree [17] was used to organize the data. All the eukaryotic sequences were kept with the exception of redundant sequences, and sequences of clades for which the edges leading to them had a length greater than 2.0 (more than 1.5 substitutions per site on average). We used this criterion because the phylogenetic analysis software was unable to place such divergent sequences accurately. These sequences included two bacteroides, and the viral and archael sequences. We further discriminated Fpg from Nei by visually inspecting sequence alignments of putative Fpg/Nei proteins, and looking for conservation of key residues. T Coffee [18,19] was used to remove nearly identical sequences (98%), MAFFT [20] for high accuracy parameters for alignment and PFAAT [21] for visualization. Seaview [22] was used to remove phylogenetically uninformative sites and the phylogenetic tree was made with Phylip's PROTDIST and NEIGHBOR [23] using SEQBOOT to create 100 bootstrap replicates.

### 2.2. Enzyme purification

*C. albicans* genomic DNA was used as a template to amplify CalFpg by the polymerase chain reaction (PCR). The gene is located on two exons but the vast majority of the gene is on one of the exons so that exon was amplified by PCR. The rest of the gene from the short exon was added with a long oligodeoxynucleotide primer of the appropriate sequence in a subsequent round of PCR. The *CalFpg* gene sequence contains three CUG codons, one at amino acid position 205, a second at position 375 and the last at position 377. The CUG codon in *C. albicans* is an exception to the universal codon code and codes for serine instead of leucine. Site directed mutagenesis using the Quikchange kit (Stratagene) was used to change these three codons so that *E. coli* would insert a serine at these positions into the CalFpg amino acid sequence. The *CalFpg* gene was cloned into pET22b with a C-terminal his tag and expressed in Rosetta DE3 plysS cells (Novagen). Transformed cells from a single colony were grown in 2 l LB under appropriate selection and expression was induced with 1 mM IPTG at an $OD_{600}$ of 0.585 and grown at 16 °C overnight with shaking at 300 rpm. The cells were harvested and flash frozen in liquid nitrogen and stored at −80 °C. The cell pellet was resuspended in 50 ml of 50 mM sodium phosphate pH 8, 100 mM NaCl, 10 mM imidazole pH 8, 10% glycerol, 5 mM β-mercaptoethanol, 1 mM PMSF and 10 mM benzamidine. The cells were lysed by sonication six times for 2 min on ice. The lysate was clarified by centrifugation and the supernatant loaded onto a 5 ml HiTrap chelating HP column charged with 100 mM nickel sulfate. The protein was eluted at a flow rate of 2 ml/min with a linear gradient of 50 mM sodium phosphate pH 8, 100 mM NaCl, 500 mM imidazole pH 8, 10% glycerol and 5 mM β-mercaptoethanol. Fractions containing CalFpg, as determined by SDS-PAGE, were dialyzed against 20 mM HEPES pH 7.6, 150 mM NaCl, 10% glycerol, 5 mM β-mercaptoethanol. The protein was further purified utilizing a