DNA Repair 22 (2014) 77–88

Contents lists available at ScienceDirect

DNA Repair

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

Phylogenetic analysis and evolutionary origins of DNA polymerase X-family members

Rachelle J. Bienstock, William A. Beard, Samuel H. Wilson*

Laboratory of Structural Biology, National Institute of Environmental Health Sciences, National Institutes of Health, 111 T.W. Alexander Drive, Research Triangle Park, NC 27709, United States

ARTICLE INFO

Article history:
Received 5 May 2014
Received in revised form 25 June 2014
Accepted 9 July 2014

Keywords:
DNA polymerase
Evolution
Genomics
Phylogenetic
Structure-function
X-family

ABSTRACT

Mammalian DNA polymerase (pol) β is the founding member of a large group of DNA polymerases now termed the X-family. DNA polymerase β has been kinetically, structurally, and biologically well characterized and can serve as a phylogenetic reference. Accordingly, we have performed a phylogenetic analysis to understand the relationship between pol β and other members of the X-family of DNA polymerases. The bacterial X-family DNA polymerases, *Saccharomyces cerevisiae* pol IV, and four mammalian X-family polymerases appear to be directly related. These enzymes originated from an ancient common ancestor characterized in two *Bacillus* species. Understanding distinct functions for each of the X-family polymerases, evolving from a common bacterial ancestor is of significant interest in light of the specialized roles of these enzymes in DNA metabolism.

Published by Elsevier B.V.

1. Introduction

Phylogenetic analysis is a computational method for quantifying evolutionary changes and relationships between protein sequences from different species over time. This type of analysis contributes to our understanding the functional development of a specific enzyme and the relationship between enzymes within cellular pathways, as well as the origin of enzymatic pathways within a cell. Additionally, evolutionary pathway analysis can provide insight into the origin of a particular disease pathway, as well as the relationship between genes in a model organism, as compared to humans.

DNA polymerases catalyze DNA synthesis during repair, replication, and recombination of DNA, as well as specialized DNA synthesis functions during viral replication and antibody gene maturation. Human cells have at least 16 distinct DNA polymerases, now characterized as members of different groups or "families" (designated A, B, X and Y) based on primary protein sequences and putative structural motifs [1,2]. DNA polymerases are usually multi-domain proteins that include an accessory domain in addition to the polymerase domain [3]. The accessory domain (e.g., proofreading exonuclease) can complement the biological function of the polymerase. The nucleotidyl transferase super-family structural fold consists of a catalytic subdomain with a β-sheet and 2 α-helices [4]. The catalytic subdomain includes three carboxylate side chains that coordinate two divalent metal cations (usually Mg^{2+}) within the polymerase catalytic site.

The human DNA polymerase X-family, or family-X, is comprised of DNA polymerase (pol) β, pol λ, pol μ and terminal deoxynucleotidyl transferase (Tdt). Only vertebrates possess members of all four of these family-X DNA polymerases, with plants, fungi and simpler organisms having only one or two family members; in some cases X-family members are not present at all (e.g., *Caenorhabditis elegans* and *Drosophila melanogaster*) [5,6]. Tdt has only been identified in vertebrates. Family-X DNA polymerases also include the African swine fever virus pol X, *Saccharomyces cerevisiae* pol IV and a large and emerging series of recently identified pol X-family members in bacterial systems [7,8].

Human pol β has been kinetically, structurally, and biologically characterized [9], and in the current work serves as a reference for comparison with the bacterial X-family polymerases. Bioinformatic and phylogenetic analyses were used to establish an evolutionary, structural and functional relationship between human pol β and bacterial DNA polymerase X-family members. These *in silico* studies may facilitate use of bacterial systems as models in understanding

Abbreviations: AP, apurinic/apyrimidinic; ASFV, African swine fever virus; BER, base excision repair; dNTP, deoxynucleoside triphosphate; dRP, deoxyribose phosphate; kDNA, kinetoplast DNA; NHEJ, non-homologous end-joining; PHP, polymerase and histidinol phosphatase; Tdt, terminal deoxynucleotidyl transferase.

* Corresponding author. Tel.: +1 919 541 4701; fax: +1 919 541 2843.
E-mail address: wilson5@niehs.nih.gov (S.H. Wilson).

http://dx.doi.org/10.1016/j.dnarep.2014.07.003
1568-7864/Published by Elsevier B.V.

DNA transactions in more complex organisms with pol X-family members and/or provide insights into the role of the bacterial enzymes in their native environment.

Aravind and Koonin [10] characterized a broad group of nucleotidyl transfer enzymes that included DNA polymerase X-family members as well as members from related families: archaeal and bacterial CCA-adding enzymes/polyA polymerases, protein nucleotidyltransferases, antibiotic nucleotidyltransferases, and proteobacterial adenylyl cyclases. All of these enzymes transfer a nucleotide to an acceptor hydroxyl group, and their common active site suggested an evolutionary relationship.

Analyses of phylogenetic relationships of X-family members have been reported more recently. Uchiyama et al. [6] suggested that all X-family members evolved from a single pol λ-like gene involved in non-homologous end-joining (NHEJ) and that the other X-family member polymerases arose due to gene duplication of this pol λ-like gene. Similarly, Kodera et al. [11] concluded that since the most basal phylum (e.g., *Echinodermata* in metazoans) contained three X-family DNA polymerase genes (i.e., β, λ, and Tdt/μ-like), it is likely that the eukaryotic pol X-family diverged from a single pol λ-like coelenterate phylum gene. In contrast, the computational analyses presented here suggest that all X-family members evolved from a polymerase nucleotidyl transfer catalytic core protein present in ancient bacterial organisms, and gene duplication and alterations occurred over time providing increasing complexity and organelle differentiation within species. This work was aimed at achieving an understanding of the functional and chronological development of different X-family members, especially in relation to pol β.

DNA polymerase X-family members are present and conserved throughout many of the oldest and most varied forms of life. We applied several established methods for sequence alignment followed by phylogenetic analysis to assess the hypothesis that the various X-family polymerases evolved from a DNA polymerase X present in ancient bacterial species. The analysis was more extensive than that in previously published studies. For example, one such study [12] included only 27 X-family polymerase sequences. With recent advances in genomic DNA sequencing, the present study represents a sampling of more than 100 diverse species' sequences. Additionally, as crystal structures for many of the polymerase X-family members have been solved, including bacterial representatives such as from *Deinococcus radiodurans* [13], *Thermus thermophilus HB8* [14], and the African swine fever virus (ASFV) pol X [15,16], many structure-function relationships important in phylogenetic considerations of DNA polymerase X-family members could be evaluated.

2. Materials and methods

Several established algorithms were utilized for sequence alignment and analysis of the assembled DNA polymerase X-family phylogenetic tree. For creating the phylogenetic tree, we used the following: Phylome DB v. 3.0 [17], PhyML v 3.0 [18,19], ETE [20], iTOL [21,22], Phylemon 2.0 [23], Archaeopteryx [24], and TreeGraph2 [25]. MUSCLE v. 3.7 was used as the sequence alignment algorithm [26]. JalView was used for creating visual images of sequence alignments produced by MUSCLE [27].

Phylogenetic analysis was initiated from an alignment of 111 identified X-family DNA polymerase sequences. The resulting phylogenetic tree was developed by beginning with the defined "Phylome" deposited in the PhylomeDB v3.0 [17] for human pol β. PhylomeDB (http://orthology.phylomedb.org), is a database of complete collections of gene phylogenies (phylomes) including a number of model species. Phylome uses orthology prediction on a genomic scale to determine the evolutionary relationship between genes from multiple independent phylogenetic trees. The Phylome DB contains 717 species and provides redundant orthology and paralogy predictions. To the original 63 sequences present in this defined phylogenetic pol β tree in the Phylome DB v. 3.0, were added an additional 48 sequences; 23 identified bacterial DNA X-family polymerases, 5 trypanosomatid sequences, and 20 eukaryote (animal) sequences belonging to the X-family. These sequences were then realigned with the original group defining the "Phylome" using the MUSCLE v. 3.7 sequence alignment algorithm [26] and a new phylogeny tree was generated using PhyML v. 3.0 [18,19]. All the identified pol β sequences analyzed in Asagoshi et al. [5], including bacterial DNA polymerase X sequences, and all the bacterial sequences analyzed by Banos et al. [28] were included in the alignment. Many of these sequences are annotated by GenBank as "hypothetical proteins," "phosphotransferase domain containing proteins," and "DNA polymerase X or X-family DNA polymerase" or "predicted to be or similar to DNA polymerase β."

Eggnog 3.0 [29] was also searched to identify 376 protein members of the COG1796 DNA polymerase IV X-family group or "cog" when searched with the human pol β sequence in 261 species. However, most of these orthologs are based on sequence similarity and polymerase enzymatic function has not been confirmed within these species. KOG2534 includes 198 enzymes in the DNA polymerase X-family category from 86 different species. The DNA polymerase X-family members used in the sequence alignments and subsequent phylogenetic analysis are tabulated in Table 1.

3. Results

3.1. Phylogenetic relationships

The comprehensive phylogenetic tree developed from the DNA polymerase X-family sequences included in this study is illustrated in Fig. 1. The detailed sequence alignment used to develop this tree is prohibitively large (Supplementary FASTA file). Nevertheless, the quality and consensus of a portion of the alignment is illustrated in Fig. 2. This figure illustrates the sequence conservation exhibited over the human pol β sequence (residues 1–335). Key residues of pol β involved in substrate binding, catalysis, conformational changes, and its deoxyribose phosphate (dRP) lyase activity are tabulated in Table 2.

From the standpoint of a broad overview, the bacterial X-family polymerases are the evolutionary ancestors of all family-X polymerases (including the eukaryotic cellular enzymes designated pol β, pol λ, pol μ, TdT, and pol IV) (Figs. 1 and 3). The ancient bacterial X-family polymerases evolved into an ancestral pol IV from *Kluyveromyces lactis* and later *S. cerevisiae* pol IV. After this point, two distinct branches arose: (1) the Tdt and pol μ branch (two gram-negative rod shaped bacteria also have polymerases clustered with this group, as well as several fungi), and the (2) the pol β and λ branch. The pol β branch includes a sub-branch of slime mold and trypanosomatid pol β; and the pol λ sub-branch includes the plant *Arabidopsis thaliana* pol X and some fungal polymerases.

As illustrated in Figs. 1 and 3, the evolutionary origin of the DNA polymerase X-family members begins with the most ancient pol X from the *Bacillus* bacterial species, *B. subtilis and B. amyloliquefaciens* and then *B. pumilus.* These are gram-positive bacteria with protective endospores permitting the organisms to tolerate extreme environmental conditions. The oldest ancestral *Bacillus* polymerases X are followed by the *Listeria* and *Staphylococcus* pol X that lead to the *Desulfotomaculum reducens* polymerase and then to the *Thermoplasma* and *Thiobacillus*, followed by the thermophilic species, *T. thermophilius* and *T. aquaticus* pol X; these are followed by the *D. radiodurans*, *Mycobacterium tuberculosis*, and *Nematostella*