Project DE-FG03-00ER63041

"Molecular recognition of DNA damage sites by apurinic/apyrimidinic endonucleases"

Final report, July 2005

Project period: 9/1/2000-11/30/2004

Werner A. Braun, Ph.D.
Dept. Human Biological Chemistry and Genetics, Sealy Center for Structural Biology
University of Texas Medical Branch, Galveston, TX 77555-1157
Phone: (409) 747-6810; FAX: (409) 747-4289; E-mail: webraun@utmb.edu

Summary

The DNA repair/redox factor AP endonuclease 1 (APE1) is a multifunctional protein which is known to be essential for DNA repair activity in human cells. Structural/functional analyses of the APE activity is thus been an important research field to assess cellular defense mechanisms against ionizing radiation.

We have taken a computational and experimental approach to elucidate the roles of residues in the apurinic/apyrimidinic endonuclease 1 (APE1) that mediate interactions with damaged DNA and other enzymes in the base excision repair (BER) pathway. During the project period, we

- developed a novel method to automatically detect sequence motifs defined in terms of the conserved physical-chemical properties (PCP) of residues in protein families. This methodology has been applied to the endonuclease superfamily and to DNase 1 related nucleases/phosphatases to identify common sequence motifs important for metal binding and catalytic activity[1-3].
- performed comprehensive molecular mechanics and dynamics calculations to investigate the role of the metal ion in the catalytic process. These data support a novel mechanism for cleavage of abasic sites in DNA by APE1, in which the metal ion moves from a stable position on the enzyme surface to a more solvent exposed one during DNA cleavage, aided in part by ligands from the substrate (tns in preparation).
- Generated several mutants of APE1 and performed biochemical analysis to identify amino acid side chains important for substrate/product DNA binding and to analyze the nuclear transport mechanism of APE1. In addition, we have created various APE1 mutant cDNA and recombinant proteins during this DOE project, including: (a) Cys to Ser missense mutants, (b) insertion mutants that carry Xth signature inferred for its high exonuclease activity[4, 5].
- Applied for a patent "Physical-chemical property based sequence motifs and methods regarding same" (U.S Patent application serial No. 10/817,530, inventors: W. Braun, V. Mathura, C.H. Schein) for our PCPMer method. This methodology has been implemented in the PCPMer program. The motif identification method is generally available to the scientific community via our WEB server MASIA (http://www.scsb.utmb.edu/masia/masia.html). A standalone version of the PCPMer program for motif generation and data base mining is available from our Website http://129.109.59.108/BinZhou/PCPMer/index.html.
DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.
DISCLAIMER

Portions of this document may be illegible in electronic image products. Images are produced from the best available original document.
Background and Significance

Genomic cloning has revealed that most of the enzyme families essential for maintaining cell growth have been conserved throughout evolution. However, mammalian enzymes with different functional activity, may have evolved by combining elements from several bacterial ancestral genes. Even small proteins may contain several individual domains that link them to different superfamilies. While many endonucleases share a common active site that is highly conserved across many subfamilies, identifying residues that control substrate specificity requires sophisticated analysis that combines both sequence conservation and structural data.

In our project we developed a word-based "molego" approach, to identify structural elements that control substrate specificity[2, 6]. We hypothesize that elements conserved in all the members of related protein families dictate common structures and also common "functions", i.e., individual steps in a complex reaction. Areas that affect substrate specificity will be less conserved in the superfamily than they are in subfamilies of enzymes that catalyze specific activities[7]. To provide the computational tools for this approach, we developed a novel method, PCPMer, to automatically detect sequence motifs defined in terms of the conserved physical-chemical properties (PCP) of residues in protein families [3].

We demonstrated the usefulness of our approach for the multifunctional family of DNA repair proteins, the apurinic/apyrimidinic endonucleases (APES), which have a clearly defined bacterial ancestor, *E. coli* exonuclease III (ExoIII), and several enzymes that are distantly related in activity and structure. APE1 is a multifunctional protein which was shown to be essential in an early mouse embryogenesis [8-10]. Because of its multifunctional property, it has been difficult to assess the importance of the DNA repair function, until two independent reports provided this year solid evidences that its DNA repair activity is pivotal to cells [5, 11]. Moreover, recent reports indicated that high levels of the APE1 activity in tumor tissues correlated with chemo- and radio-resistance of these tumors in adjuvant therapies [12, 13]. Structural/functional analyses of the APE activity has thus been an important research field to solve such broader issues as to seek novel approaches in cancer therapy, and to assess cellular defense mechanisms against ionizing radiation.

Results:

The PCPMer method.

We developed a novel method to automatically detect sequence motifs defined in terms of the conserved physical-chemical properties (PCP) of residues in protein families. Conservation of physical-chemical properties of amino acids in protein families is calculated from variation of five quantitative descriptors, E1 to E5, (PCP-vectors) [1]. These descriptors can reproduce the distances in the original property space with a correlation coefficient of 99%. A sequence motif for a protein family is defined as contiguous segments that are conserved in their PCP vectors. Each motif is quantitatively expressed as a physical-chemical property profile (PCP-profile) that includes the average values of the PCP-vectors and their standard deviations. The relative entropy of the actual distribution of PCP values and the 'a priori' distributions measures the statistical significance of the conservation[3].

This methodology has been implemented in the PCPMer program, which generates PCP-motifs automatically from sequence alignments. Computational tests have indicated that the method efficiently finds homologous proteins in protein databases and common motifs in distantly related proteins. A patent "Physical-chemical property based sequence motifs and methods regarding same" (U.S Patent application serial No. 10/817,530, inventors: W. Braun, V. Mathura, C.H. Schein) for our method is pending.
Identification of functional motifs in APE1

We used PCPMer to identify sequence motifs in the APE family and showed that a subset of these were common to other members of the DNase I superfamily. The method is sensitive enough to identify common motifs in distantly related proteins. By structurally annotating these motifs, the sequences were parsed into individual building blocks, molecular legos ("molegos") [2]. This word-based, molego approach can identify structural elements that control substrate specificity. We showed that molegos common to several different protein families had similar functional roles, and delineated areas of APE1 that could contribute to its specific cleavage of damaged DNA. Comparing substrate/product binding of molegos common to DNase-I showed that those distinctive for APEs are not directly involved in cleavage, but establish protein-DNA interactions 3' to the abasic site. Site directed mutagenesis showed that one area we found probably contributes to APE's ability to retain cleaved DNA after removal of the damaged bases[4]. These additional bonds enhance both specific binding to damaged DNA and the processivity of APE1.

MASIA identified 12 property-based motifs in the apurinic/apyrimidinic endonuclease (APE) family of DNA-repair enzymes within in the DNase-I superfamily. The distribution of the scores of all motifs is clearly separated for all 42 APE sequences from the distribution in the ASTRAL Database, as illustrated for motifs 1,2,11 and 12 in Fig. 1.

![Graphs showing distribution of scores for APE motifs](image)

**Fig. 1:** Distribution of the highest scoring window for APE motifs 1,2,11,12 in the 3635 proteins of the ASTRAL 40 database (continuous line) and in all proteins of the APE family.
Data mining of the ASTRAL40 database with these motifs located distantly related representatives of the DNase-I superfamily, such as Inositol 5'-polyphosphate phosphatases. Other proteins identified had no overall sequence or structural similarity. However, all were phosphatases and/or had a similar metal ion binding active site. Thus our automated method can identify discrete elements in distantly related proteins that define local structure and aspects of function. We anticipate that our method will complement existing methods to functionally annotate novel protein sequences from genomic projects[3].

Running PCPMer in parallel mode on a LINUX cluster.

Motif-mining of large sequence data bases can be time consuming. Therefore we adopted the software package in a parallel mode. We searched a non-redundant data base of 36,788 sequences in the human genome for APE related PCP-motifs. A cluster of dual-processor Pentium 4 running at 2.60GHz was used in our search. All processors are connected through a central Fast Ethernet switch, providing a flat network topology. Figure 2 depicts the CPU time and speedup obtained in term of the number of worker processor. The CPU time for database search with one processor and 18 processors are 538.1 seconds and 31.5 seconds, respectively. Thus, the speedup increased nearly linearly with the number of processors (Zhou et al., in preparation).

Figure 2: CPU times and speedup factor for the PCPMer for the parallel version of PCPMer. The parallel version uses a self-scheduling principle (SS). The SS approach in the Single Program Multiple Data (SPMD) parallel programming paradigm is based on the Master-Worker approach on distributed-memory architectures, where a single master process maintains centralized control over concurrent computations by a number of worker processes.

PCPMER complements other existing methods in data mining of sequence patterns or motifs, such as PROSITE, BLOCKS or PFAM. These are repositories of biologically significant patterns and conserved sequences that distinguish protein families, but have limitations for search of distantly related members on the superfamily level. The PROSITE motifs were primarily generated by manual scanning of the results of data base searching and sequence alignment tools. For the most part, the motifs have been scaled down to reflect only absolutely conserved residues, or those with very few substitutions, as manual methods cannot reliably estimate the statistical significance of less conserved positions. We have shown that our PCP motifs for APE1 identify more family members and known homologues than the signatures for this family in the PROSITE database. We thus suggest that PCPMer will effectively complement such databases.

Specific motifs distinctive for APE are needed for specific cleavage of damaged DNA.

To understand the interaction of APE1 with substrate/product DNA more precisely, we examined the
molego blocks in APE1 [2]. One motif was specific for the APE family and not present in the DNase superfamily. Inspection of this motif in the cocrystal structure of APE1 led to two asparagine residues, N226 and N229, at close proximity of the downstream of abasic site [2]. The H-bond network of the substrate DNA and the two Asn residues resembles that of R177, which is important for APE1's product DNA binding [4, 14]. Substituting alanine at these positions enhanced AP-endonuclease activity, as was previously observed for the R177A mutant, which binds to the same site on damaged DNA. Kinetics of these mutations showed that the enhanced reaction rate was accompanied by an increase in Km. Combining the mutations lead to progressive decreases in APE activity and ability to bind substrate DNA [4]. We examined catalytic (AP endonuclease) activity and DNA binding affinity of the missense mutant proteins, N226A and N229A, in which either of the Asn residue was changed to Ala (Table I). These mutant enzymes showed higher enzymatic activities than the wild-type (WT). Similarly, we reported that R177A mutation in APE1 enhanced its activity. Based on DNA binding assay with product DNA, we concluded that N226 and N229 are important for their product DNA binding. We conclude that the H-bond network at the DNA backbone just downstream of the abasic site is crucial for the product DNA binding, but also affect the substrate DNA binding [4].

### Table 1: Kinetics of APE1 proteins for AP-endonuclease activity

<table>
<thead>
<tr>
<th>Protein</th>
<th>Km (nM)*</th>
<th>Kcat (sec⁻¹)*</th>
<th>Kd (nM) ± STD**</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wild-type</td>
<td>4.7 ± 1.3</td>
<td>1.6 ± 0.06</td>
<td>65.7 ± 4.6</td>
</tr>
<tr>
<td>N226A</td>
<td>10.8 ± 3.4</td>
<td>3.7 ± 0.25</td>
<td>133.8 ± 40.7</td>
</tr>
<tr>
<td>N229A</td>
<td>17.3 ± 1.01</td>
<td>3.2 ± 0.46</td>
<td>159.6 ± 26.9</td>
</tr>
<tr>
<td>N226A+N229A</td>
<td>38.5 ± 16.5</td>
<td>2.1 ± 0.28</td>
<td>487.7 ± 124.6</td>
</tr>
<tr>
<td>R177A</td>
<td>-</td>
<td>-</td>
<td>410.3 ± 102.6</td>
</tr>
<tr>
<td>R177A+N226A</td>
<td>-</td>
<td>-</td>
<td>6749.4 ± 2826.2</td>
</tr>
</tbody>
</table>

*5'-32P-end labeled 43-mer DNA containing a single AP site analogue (tetrahydrofuran) was used as a substrate.

**5'-Cy5-labeled 26-mer DNA with a single tetrahydrofuran was used.

Molecular mechanics and dynamics studies on the stability of the binding sites for Mg²⁺ ions in the active site of APE1.

The functional role of the Mg²⁺ in assisting the cleavage of the damaged DNA is not clear, as the crystal structures of free APE and DNA bound APE are not conclusive. In the crystal structures of APE1 in the complex with DNA only one bound Mn²⁺ ion is observed near the active site [14], whereas crystalization of the free APE with Pb²⁺ showed two metal binding sites, positions "A" and "B" (PDB code 1E9N)[15]. In the 1DE9 structure of an APE1-DNA-Mn²⁺ complex, the DNA is cleaved and the Mn²⁺ position is equivalent to the A-site. One Sm³⁺ ion in the structure of the free protein, PDB code 1BIX, is also at the A site.

We used homology modeling and MD simulations to suggest what effects these mutations could have on metal ion binding and APE activity. Homology models of each mutant, based on the X-ray crystal structures of wt-APE1 in complex with damaged DNA determined by Mol et al. [14] (PDB-code 1DF8 and 1DE9), were used to study the overall structural consequences. Molecular dynamics simulations of wt APE and the two mutants were then used to determine the effect of the mutations on the correlated movements of residues in the catalytic site, and on the position and mobility of the metal ion(s).
In simulations with the wild type protein, the Mg$^{2+}$ ion moved towards the B-site, if the initial site is near the A-site. In the E96A mutant the Mg$^{2+}$ ion starting close to the crystal A-site, moved in one case away from the A and B-site (MD06) and in the second case to the B-site (MD10). In the double mutant simulation the Mg$^{2+}$ ion moved from the A-site to the B-site. The simulations thus consistently indicate that in both active forms of APE1, i.e. in the wild type and the double mutant form the B-site is the preferred position before the catalytic step occurs.

Further we did a systematic grid search for potential metal binding sites using 1215 different initial positions for Mg$^{2+}$ near the A and B sites. The cluster with the lowest potential energy of the Mg$^{2+}$ ions was located near the B site (Fig.4). The energy was between -571 and -550 kcal/mol. The energy of the clusters close to the A-site had energies about -503, -480, -460, and -440 kcal/mol. This result clearly showed, that the metal ion preferentially occupies the B-site before DNA cleavage occurs. Our calculations support a novel mechanism for cleavage of abasic sites in DNA by APE1, in which the metal ion moves from a stable position on the enzyme surface to a more solvent exposed one during DNA cleavage, aided in part by ligands from the substrate. A manuscript describing these results is in preparation.

Experimental tests of the single metal movement. The role of divalent cation in the endonucleolytic reaction. APE1 requires a divalent cation such as Mg$^{2+}$ or Mn$^{2+}$ as a cofactor. Use of two kinds of metals has been a successful method to probe the role of metal in nucleolytic reactions [16]. In APE1/Mg$^{2+}$ reaction, we found that Ca$^{2+}$ is inhibitory, whereas in APE1/Mn$^{2+}$ reaction Ca$^{2+}$ surprisingly increased the nuclease activity (Fig.4).
This dichotomy can be explained by assuming that there is at least an additional metal site very close to the catalytic metal pocket, which Ca^{2+} can occupy without losing the activity. Although this is consistent with the two-site model [15], this model was later rejected by the same group, and it is very unlikely that two metal ions coexist at the catalytic pocket during hydrolysis.

We rather hypothesize that the proper movement of the metal cofactor is crucial for the hydrolytic reaction. Our previous study found that a missense mutation, E96A, which loses its metal coordinating side chain, recovered the endonuclease activity by the introduction of a secondary mutation, K98R. It is possible that E96A mutation loses the control of the metal movement, and K98R intragenic suppressor forces the metal back to the normal route with its longer basic side chain. A close look at the metal pocket identified other intragenic suppressor candidate, namely, D70R, N68E, and S307R. This mutation would, like K98R, hold the metal ion at the proper site even without E96. We thus created three double-mutant proteins and examined their nuclease activities (see Figure 5). Among the new mutants, E96A/S307R double mutant gained the APE activity, similarly to the E96A K98R suppressor mutant. Both K98R and S307R mutations introduce a strong basic side chain which can block the metal expulsion from the catalytic pocket.

These data fit into our model based on the computational simulation, where the metal cofactor shifts its position during the reaction.

Analysis of nuclear localization signals (NLS) and nuclear export signals (NES) in APE.

There are reports to suggest that subcellular distribution of APE1 is regulated. In some tissues and cells APE1 is predominantly found in cytosol [17], but is accumulated after oxidative stress induction [18]. Because APE1 needs to be in nuclei to function, it is possible that cells utilize cytosol as a reservoir for APE1 which may be needed at stressful conditions. Despite the significant implications, a nuclear
localization signal of APE1 was not analyzed carefully. We examined various N-terminal deletions of APE1 fused to EGFP to identify a bipartite NLS signal [19]. One consists of the N-terminal 7 aa, and the other spans 8-13th aa. These two segments function as an NLS independently with each other, but removal of both segments drastically decreased the APE1 nuclear localization [19]. In this study, we could also assess the effect of acetylation at Lys6 and Lys7. Our results indicated that the acetylation, discovered by us as a specific modification by histone acetyltransferase p300, does not affect APE1’s nuclear localization [19]. Nuclear importin alpha 1 and 2 were found to interact with APE1 based on co-immunoprecipitation and Far-Western analysis [19]. Furthermore, an APE1-EGFP fusion protein was accumulated by leptomycin B treatment, an nuclear export inhibitor [19]. We therefore proposed that cells control the nuclear level of APE1 by nuclear import (through NLSs) and by nuclear export (through yet unidentified NESs). Identification of NES will help elucidate the mechanism by which cells regulate APE1’s level in nuclei and their sensitivity to reactive oxygen species.
List of publications supported by the grant DE-FG03-00ER63041:


Literature cited


