

Research article

Open Access

Total sequence decomposition distinguishes functional modules, "molegos" in apurinic/apyrimidinic endonucleases

Catherine H Schein¹, Numan Özgün¹, Tadahide Izumi² and Werner Braun*¹

Address: ¹Sealy Center for Structural Biology, Department of Human Biological Chemistry and Genetics, University of Texas Medical Branch, Galveston TX 77555-1157, USA and ²Sealy Center for Molecular Science, Department of Human Biological Chemistry and Genetics, University of Texas Medical Branch, Galveston TX 77555-1157, USA

Email: Catherine H Schein - cathy@newton.utmb.edu; Numan Özgün - numan@bohr.utmb.edu; Tadahide Izumi - taizumi@utmb.edu; Werner Braun* - werner@newton.utmb.edu

* Corresponding author

Published: 25 November 2002

Received: 8 May 2002

BMC Bioinformatics 2002, 3:37

Accepted: 25 November 2002

This article is available from: <http://www.biomedcentral.com/1471-2105/3/37>

© 2002 Schein et al; licensee BioMed Central Ltd. This is an Open Access article: verbatim copying and redistribution of this article are permitted in all media for any purpose, provided this notice is preserved along with the article's original URL.

Abstract

Background: Total sequence decomposition, using the web-based MASIA tool, identifies areas of conservation in aligned protein sequences. By structurally annotating these motifs, the sequence can be parsed into individual building blocks, molecular legos ("molegos"), that can eventually be related to function. Here, the approach is applied to the apurinic/apyrimidinic endonuclease (APE) DNA repair proteins, essential enzymes that have been highly conserved throughout evolution. The APEs, DNase-I and inositol 5'-polyphosphate phosphatases (IPP) form a superfamily that catalyze metal ion based phosphorolysis, but recognize different substrates.

Results: MASIA decomposition of APE yielded 12 sequence motifs, 10 of which are also structurally conserved within the family and are designated as molegos. The 12 motifs include all the residues known to be essential for DNA cleavage by APE. Five of these molegos are sequentially and structurally conserved in DNase-I and the IPP family. Correcting the sequence alignment to match the residues at the ends of two of the molegos that are absolutely conserved in each of the three families greatly improved the local structural alignment of APEs, DNase-I and synaptojanin. 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. These additional bonds enhance both specific binding to damaged DNA and the processivity of APEI.

Conclusion: A modular approach can improve structurally predictive alignments of homologous proteins with low sequence identity and reveal residues peripheral to the traditional "active site" that control the specificity of enzymatic activity.

Background

Genomic cloning has revealed that most of the enzyme families essential for maintaining cell growth have been conserved throughout evolution [1]. However, mammalian enzymes with different functional activity may have evolved by combining elements from several bacterial an-

cestral genes. Even small proteins may contain several individual domains that link them to different superfamilies [2]. While many endonucleases share a common active site that is highly conserved across many subfamilies, identifying residues that control substrate specificity re-

quires sophisticated analysis that combines both sequence conservation and structural data [3–5].

In this paper we distinguish, using a word-based "molego" approach, structural elements that control substrate specificity. We postulate here 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. We have chosen to illustrate this approach using 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 are distantly related to several enzymes with varying substrate specificity.

APEs are essential for mammalian cell growth and bacterial survival in the presence of ionizing radiation and DNA mutagens [6]. They initiate repair of an abasic DNA site by cleaving the phosphodiester backbone 5' of the phosphodeoxyribose. This generates the necessary 3' hydroxyl group for DNA polymerases ($\text{pol } \beta$, δ or ϵ , in eukaryotes) to insert the correct nucleotide in later steps in the base excision repair pathway (BER-pathway) [7,8]. Recent crystal structures of huAPE1 complexed with DNA containing an abasic site [9–11], combined with sequence analysis and site-directed mutagenesis, have defined the residues that participate in metal ion based cleavage of the phosphate backbone of the DNA [12–17].

Mutations that greatly diminish the enzymatic activity of huAPE1 do not affect, and may even increase binding to damaged DNA, while non-specific DNA binding remains low [16,18]. Further, mutations that have little effect on APE activity in vitro prevent complementation of DNA repair deficient *E. coli*. As seen with other DNA repair enzymes [19], specificity determining residues, as yet unidentified for APEs, must be distinct from those involved in phosphorolysis.

To better assess which residues determine specificity, we assume that functions unique to APEs will be determined by motifs that are not conserved in a similar fashion in families with a different activity spectrum of functions. Besides cleaving the phosphate backbone, to achieve specificity APEs must coordinate a series of functions, including: interaction with target DNA in a series of small, possibly repetitive steps (scanning), locating damage sites, establishing the transition state complex, completing the cleavage, re-adjusting the charge status within the active site, and regulating release of product after interaction with the next enzyme in the BER pathway [20–25]. A finer breakdown of these functions can be achieved at the

molecular level once all the residues in the reaction mechanism are known. APEs also have RNase H, 3'-exonuclease, and 3'-phosphodiester activities that are particularly high in the bacterial members of the family [26,13].

Our web-based MASIA program [27] was used to rapidly decompose the sequences of APEs and related protein families into motifs, areas of significant conservation in members of identical function, which could then be correlated structurally using data from crystal structures. Having determined that 12 motifs were common to all APE1's, we compared the structure of the subset of these that occurred in both DNase 1 and synaptojanin, a member of the IPP family. These shared motifs had a similar 3D structure in representatives of these functionally diverse families, and we therefore called these motifs "molegos" (molecular legos). We then demonstrated that the shared molegos served a similar role in substrate binding by comparing the DNA binding profile of huAPE1 with that for the less specific enzyme DNase 1. The molegos present in both enzymes interact with target DNA in a similar fashion, while residues in molegos distinctive for APE1 control specificity by binding primarily to the bases around the apurinic site. Matching of molegos, guided by the degree of conservation of individual residues across the three families, allowed a better alignment of the individual secondary structure elements among the proteins than DALI achieved. This word based, sequence (motif) to structure (molego) to function method has clear implications for genomic analysis and template based homology modeling, as well as immediate application in recognizing specificity determinants in proteins that share active sites common to many enzymes [28].

Results

Total sequence decomposition of human Ape1 with MASIA
MASIA identified 12 motifs as conserved in all members of the APE family (Figure 1 and Table 1). As table 1, last column, illustrates, these motifs include all the residues known to be essential for DNA cleavage. Most of the highly conserved (greater than 90%) residues have been shown by previous mutagenesis studies to affect activity. The 12 motifs are also structurally conserved, as demonstrated by the low RMSD values between segments in the crystal structures of bacterial ExoIII and of huAPE1. These two proteins are only 26% identical (based on a DALI, structure based alignment) and most of the similar segments are contained in the molegos. As the third column of the table demonstrates, the backbone deviation of the segments is overall $<1 \text{ \AA}$ and for 5 of the motifs, $<0.5 \text{ \AA}$. We have chosen the name "molegos" for the structural units associated with motifs, which are presented pictorially in Figures 2 and 3. Most of the DNA and metal ion binding molegos form individual β -strands at the core of

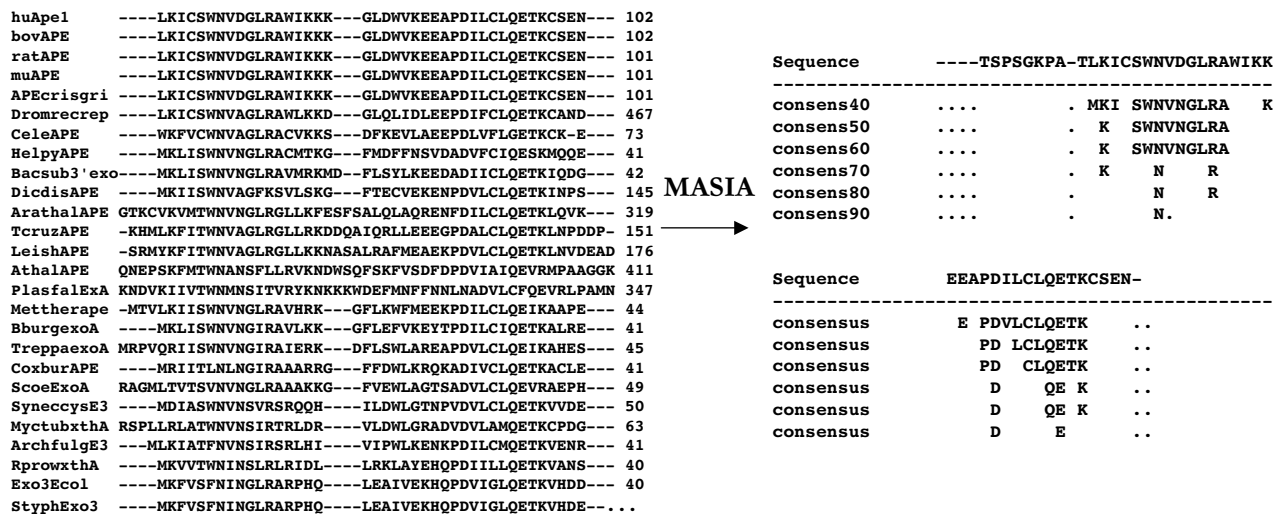


Figure 1
MASIA converts complex sequence alignments into easily readable blocks of conserved sequences. A part of the CLUSTALW alignment of the APE family used for the results in Table 1 is shown, with a section of the corresponding MASIA output that includes motifs 1 and 2.

the protein that orient the absolutely conserved residues toward the substrate, but several have a helical or hydrogen bonded coil structure.

The 12 motifs, which account for about half of the protein, are bridged by areas that vary in the different members of the APE family. These connecting regions may account for the differing activities of the bacterial and mammalian proteins. The longest molego, 7, was broken down into two areas, with the contiguous region labeled 7a. The first 7 residues of the 7a area molego are quite similar in the bacterial and mammalian APE. However, the end is differently conserved in eukaryotes. The endonuclease activity of DNase 1 is reduced many fold by integrating this loop from *E. coli* exonuclease III, but the mutant cleaves at abasic sites in DNA with low efficiency [29]. Thus additional residues in the APEs control specificity while still allowing a reasonable rate of phosphorolytic cleavage.

Finding APE molegos in the DNase 1 superfamily

In an effort to functionally annotate the molegos of APE1, we next sought to find them in other proteins that shared some structural similarity to APE. The APEs, DNase-1 and inositol 5'-polyphosphate phosphatases (IPP) have been grouped according to the SCOP database [30] as the DNase-1 like superfamily. Although DNase 1 has only 18% overall sequence identity and the IPP domain of synaptojanin, 14%, to APE1, we could show that most of the areas of identity were in molegos common to all three proteins. Motifs in other protein families were identified by genomic cross-networking with PSIBLAST (see methods for details). Our analysis identified 5 molegos that are common to the DNase 1, IPP and APE families, which roughly correspond to areas of sequence similarity identified previously [31,32]. The structural similarities of molegos 1,2, 7, 11 and 12 (i.e., the segmental RMSD's) between APE1 and representatives of the distantly related DNase 1 and IPP families are comparable to those found between members of the APE family (Figure 4 and Table 1 & 2).

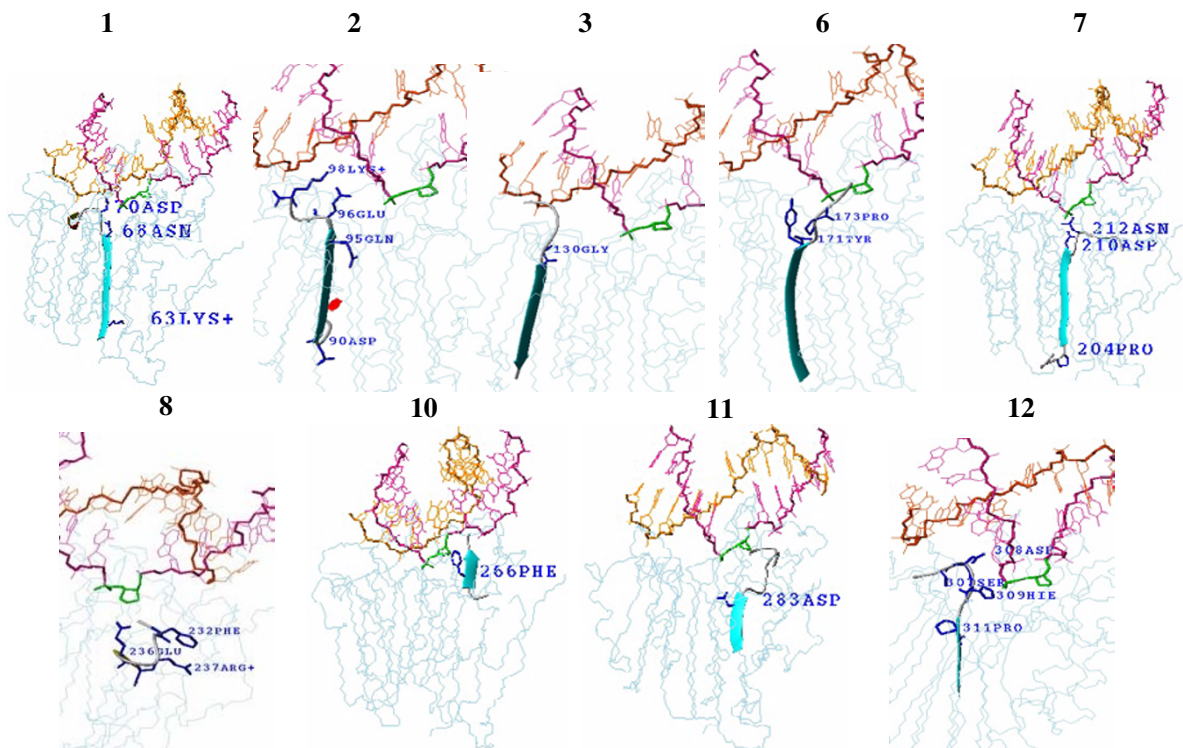


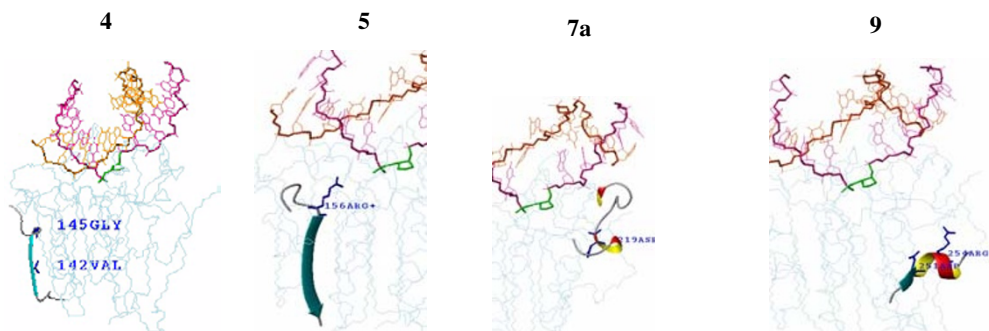
Figure 2
APE1 Molegos near the DNA. The structures for the motifs of Table 1 are taken from a minimized 1DE8 crystal structure of APE1 bound to an uncleaved 11mer DNA with an abasic site. One Mg²⁺ ion was inserted at the position seen in the IDE9 minimized structure based on IDE8 (huAPE1/Mg²⁺/11 bp (cleaved) AP-containing oligonucleotide). The molegos contain residues bind to DNA and have corresponding molegos in other proteins identified in the PSIBLAST search.

Common molegos form a similar active site in two distant relatives

The 12 conserved molegos form the β-barrel core of huAPE1. The completely conserved residues of huAPE1 concentrate, for the most part, at one end of this framework to form the metal ion binding active site (Figure 4). This core is also common to DNase 1 and synaptojanin(an IPP family member), which share the functions of metal ion based cleavage of a phosphate backbone. The shared molegos define an active site architecture conserved in all three proteins, including the orientation of the substrate toward the metal binding site.

Molegos define functional areas common to DNase 1 and APE1

A contact plot of huAPE1 with the DNA in the 1DE8 crystal structure (Figure 5) shows that motifs 1–3,5–8, and 10–12 all have residues close to the substrate, an oligonucleotide containing an abasic site (AP-DNA). The N-terminal motifs 1–3 and 5 bind primarily 5' to the apurinic site and to the 3' end of the undamaged strand. The other motifs bind more to the area 3' of the damage site. Motifs 10 and 12 span both strands of the DNA. Although motif 12 contains several highly conserved residues that, according to mutagenesis results (Table 1) contribute to APE1 activity, only His309 is very close to the abasic site in the DNA.

**Figure 3**

Other APE1 Molegos. These molegos either contain no residues that bind DNA (molegos 4 and 9) or differ significantly (5 and 7a) between the mammalian and bacterial APEs.

Molegos 4 and 9 contain no residues in contact with the DNA or metal ion. Comparing the binding of APE in a substrate complex (Figure 6, left) suggests that APE's binding to the 5' end of the DNA after cleavage (Fig. 5), especially that mediated by molego 3, is stronger, while the distance from the protein to the DNA 3' of the cleavage site increases.

The contact plots of APE1 and DNase1 with their respective substrates (Figure 6) documents that the similar molegos in the proteins serve similar functions. The N-terminal 100 residues of both proteins, including molegos 1 and 2, bind 5' of the cleavage site and to the 3' end of the opposite DNA strand. Molegos 7, 11 and 12 bind to one base 5' and the next base 3' of the cleavage site in both proteins. Overall, the pattern of protein contacts to the cleavage site, the area 5' of the cleavage site, and the 3' end of the opposite strand are common to both proteins, suggesting that the functions of forming the substrate complex and the actual phosphorolysis are similar in both proteins.

While the length of the DNA in both cases is similar, DNase 1 clearly has less binding to bases opposite and 3'

of the cleavage site. The extensive contacts that APE1 makes to these positions are mediated by molegos it does not share with DNase 1. Molegos 6, 7a and 10 all have residues within hydrogen bonding distance of the three base-pairs 3' of the AP-site. This redundancy of binding to the 3' side is unique to APE, as is its strong binding to the DNA opposite the abasic site.

The importance of such bonds for activity was shown in other work, where huAPE1's binding to the DNA backbone is only inhibited by ethylation of the phosphates two and three positions 3' to an abasic site [18]. Mutation of R177A, at the end of Molego 6, that binds to this region and to the bases opposite the AP-site had enhanced activity [11], while mutations at W280 (Molego 11) and F266 (Molego 10) [33] reduce activity and, in the latter case, substrate selectivity.

In work from this group that will be described separately, we used this analysis to generate mutants of APE1 with altered activity. An alanine substitution mutant, N226A, of a conserved residue at the end of molego 7a that forms a hydrogen bond with the second phosphate group downstream of the abasic site, had enhanced APE activity but

Table 1: Motifs characteristic of the APE family.

motif	Sequence in hu APE1 (MASIA 50% consensus) Sequence in ExoIII	RMSD, Å (# of atoms)	Known mutants
1	62L <u>K</u> I <u>C</u> S <u>W</u> N <u>V</u> D <u>C</u> L <u>R</u> A74 M <u>K</u> I <u>x</u> S <u>W</u> N <u>V</u> N <u>G</u> L <u>R</u> 1M <u>K</u> F <u>V</u> S <u>F</u> N <u>I</u> N <u>G</u> L <u>R</u> A13	0.43 (52)	C65S no Ref1, D70N no change N68A, D70A,R reduced act.
2	89P <u>D</u> I <u>L</u> C <u>L</u> Q <u>E</u> T <u>K</u> 98 P <u>D</u> I <u>L</u> C <u>L</u> Q <u>E</u> T <u>K</u> [^] 27P <u>D</u> V <u>I</u> G <u>L</u> Q <u>E</u> T <u>K</u> 36	0.30 (40)	D90A, E96A low APE, RNase H E96A, K98R wt-APE
3	125E <u>G</u> Y <u>S</u> G <u>V</u> G <u>L</u> L <u>S</u> R136 K <u>G</u> Y <u>x</u> G <u>V</u> A <u>x</u> L <u>S</u> K 60K <u>G</u> H <u>Y</u> G <u>V</u> A <u>L</u> L <u>T</u> K70	0.59 (44)	Y128A reduced APE
4	137Q <u>C</u> P <u>L</u> K <u>V</u> S <u>Y</u> G <u>I</u> G <u>D</u> E149 P <u>x</u> x <u>V</u> <u>x</u> x <u>G</u> <u>x</u> <u>G</u> 73P <u>I</u> A <u>V</u> R <u>R</u> G <u>F</u> P81	1.06 (36)	D148E no change
5	152D <u>Q</u> E <u>G</u> R <u>V</u> I <u>V</u> A <u>E</u> F <u>D</u> 163 D <u>x</u> E <u>G</u> R <u>V</u> I <u>x</u> A <u>E</u> F 86E <u>A</u> q <u>R</u> R <u>I</u> I <u>M</u> A <u>E</u> I96	2.42 (44)	R156Q low DNA binding and APE activity The RMSD for the middle 6residues (RVIVAE) is 0.46Å (24 atoms).
6	165F <u>V</u> L <u>V</u> T <u>A</u> Y <u>V</u> P <u>N</u> A175 F <u>x</u> <u>x</u> <u>x</u> <u>N</u> <u>x</u> Y <u>x</u> P <u>N</u> 103V <u>T</u> V <u>I</u> N <u>G</u> Y <u>F</u> P <u>Q</u> 112	0.62 (40)	Y171H no APE act.
7	203K <u>P</u> L <u>V</u> L <u>C</u> G <u>D</u> L <u>N</u> V K <u>P</u> V <u>x</u> x <u>C</u> G <u>D</u> <u>x</u> <u>N</u> <u>V</u> 144N <u>P</u> V <u>L</u> I <u>M</u> G <u>D</u> M <u>N</u> I154	0.86 (44)	D210A, N& N212A: no activity N212Q, D, D219A, reduced act. RMSD IS 0.35 Å for the 1 st 7 residues (28 atoms)
7a	A <u>H</u> E <u>E</u> I <u>D</u> L <u>R</u> N <u>P</u> K <u>G</u> N226 A <u>H</u> <u>x</u> E <u>I</u> D <u>L</u> <u>x</u> N <u>P</u> <u>x</u> <u>x</u> <u>N</u> S <u>P</u> T <u>D</u> L <u>D</u> I <u>G</u> I <u>G</u> E <u>E</u> N167	1.99 (52)	
8	231G <u>F</u> T <u>P</u> O <u>E</u> R237 G <u>F</u> T <u>P</u> E <u>E</u> R 178S <u>F</u> L <u>P</u> E <u>E</u> R184	0.32 (28)	R237A, slightly reduced act.
9	251D <u>S</u> F <u>R</u> H <u>L</u> Y <u>P</u> 258 D <u>T</u> F <u>R</u> <u>x</u> <u>x</u> <u>x</u> <u>P</u> [*] 197D <u>T</u> F <u>R</u> H <u>A</u> N <u>P</u> 204	0.25 (32)	H255A reduced APE activity
10	264Y <u>T</u> F <u>W</u> T <u>Y</u> M <u>M</u> N <u>A</u> R274 Y <u>T</u> W <u>w</u> <u>x</u> Y <u>R</u> <u>x</u> <u>x</u> A <u>R</u> [*] 210F <u>S</u> W <u>F</u> D <u>Y</u> R <u>S</u> K <u>G</u> F220	0.78 (44)	F266A or C: reduced APE, higher non-specific DNA cleavage. W212S (=F266) in ExoIII: no APE act. or DNA binding
11	274R <u>S</u> K <u>N</u> V <u>G</u> W <u>R</u> L <u>D</u> Y <u>F</u> L <u>L</u> S288 R <u>x</u> <u>x</u> <u>N</u> <u>x</u> G <u>W</u> R <u>I</u> D <u>Y</u> <u>x</u> L <u>x</u> S ^{**} 220F <u>D</u> D <u>N</u> R <u>G</u> L <u>R</u> I <u>D</u> L <u>L</u> L <u>A</u> S 233	0.36 (60)	W280A, reduced activity D283A: slight reduction, D283A+D308A: reduced act.
12	306G <u>S</u> D <u>H</u> C <u>P</u> I <u>T</u> L314 S <u>D</u> H <u>C</u> P <u>T</u> <u>x</u> <u>L</u> 256P <u>S</u> D <u>H</u> A <u>P</u> V <u>W</u> A264	0.92 (36)	G306A: wt APE D308A, S: reduced activity in low [Mg ²⁺] H309N or Δ 5 C-terminal residues, no activity

[^] corresponds to the PROSITE motif 1 for the APE family, [APE]D [LIVMF]₂x [LIVM]QExK * motifs correspond to PROSITE motif 2, D [ST] [FY]R [KH]-x(7,8) [FYW] [ST] [FYW]₂ **corresponds to PROSITE motif 3, NxGxR [LIVM]D [LIVMFYH]_x [LV]_xS. Completely conserved residues are bold, underlined letters are conserved to >70%. The first column shows the sequence of huApe1 with the MASIA consensus motif (50% conservation over 37 sequences) below it. The third column is the backbone RMSD between the indicated motif sequences in IBIX (huAPE1) and the IAKO (E. coli ExoIII) crystal structures (both files are for the respective protein without DNA), with the number of atoms indicated in parentheses. The last column shows the effect of mutations in the motifs on APE activity and DNA binding.

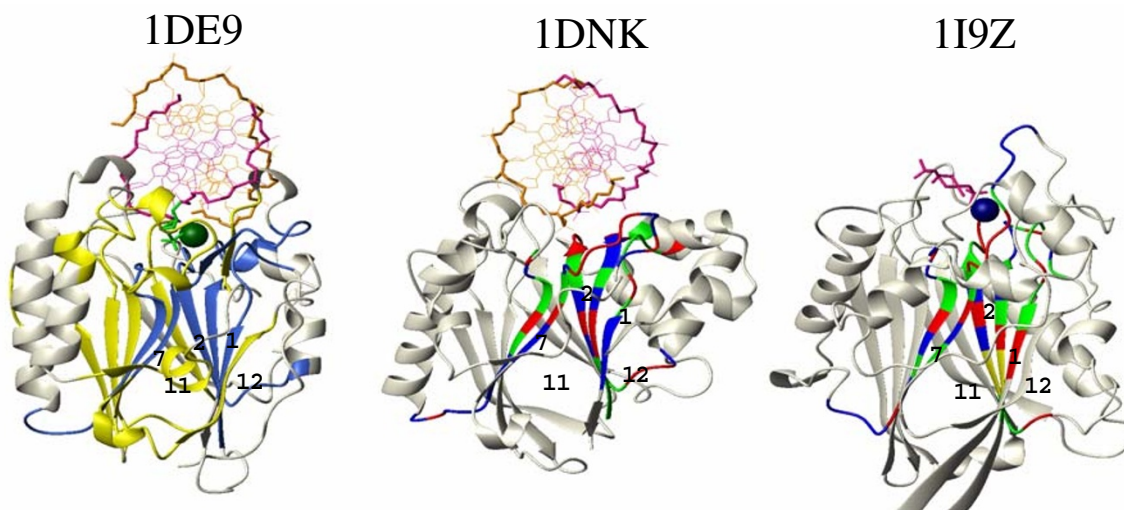


Figure 4

APE Motifs common to DNase-I and Synaptojanin have similar structure. Conserved areas of huAPE1 (from PDB file 1DE9) form the β -barrel core (left side). The motifs common to DNaseI (1DNK) and Synaptojanin (1I9Z) form a similar catalytic site, with the substrate (from the appropriate structure file) directed to the top of the β -strands and the metal ion binding site. The strands in the middle and right figures are colored according to their identity (green) or similarity (blue) to APE1. Residues that differ are red, the insertions needed for correct alignment in synaptojanin are yellow.

increased K_m and K_d values, similar to an alanine mutant of R177, which binds to the same site, reported previously [34]. A combination of the two mutants, N226A and R177A, substantially reduced the ability of APE1 to bind to DNA containing an abasic site (Izumi et al., in preparation). Thus, molegos can effectively guide the redesign of enzymes to alter specificity.

Molegos to improve structural alignment

Using molegos may also help in aligning proteins for template based modeling, by determining the end points of secondary structure elements in alignments with many gaps and insertions. According to MASIA analysis, the residues K/R and DI at the N-termini of motifs 1 and 2 are absolutely conserved in the three families, APEs, DNase1s and IPPs. However, matching these conserved residues between synaptojanin and DNase1 or APE requires a gapping that would not be consistent with CLUSTALW or a structural (DALI [35]) alignment of these proteins (Table

2). If the local alignment with synaptojanin is gapped to align these residues in the three proteins (Table 2, gapped), the RMSD for the two sections separated by the gap is much lower than that if one tries to align the whole ungapped segment. As Fig. 7 illustrates, the local environments of both conserved residue pairs DI and QE are structurally equivalent in all three proteins, indicating that a motif based alignment with a two residue gap is correct. The first two β -strand molegos in synaptojanin are 2 residues longer than in APE or DNase 1. By regarding these elements as simple lego style blocks, and recognizing the connectivity, molego based alignment correctly defined the changing length of the secondary structure elements.

Discussion

Previously, a "lego[®] block" approach described for organic synthesis [36–38] was used to describe the reshuffling of large sections of plant genomes [39] and as a rational

Table 2: Aligning sequences according to MASIA motif analysis improves the correlation between structures.

Motif	PDB file	Sequence	Residue # used for alignment	RMSD (Å) to APE1	RMSD (Å) to DNase1
1	1DE9	LKICSWNVVDGLRA	62-74	0	
	1DNK	LKIAAFNIRTFGE	1-13	1.6	0
	1I9Z	(VK)IFVASYNLNGCSA	560-572	2.0	2.1
Gapped:					
	1DE9	LKI -- CSWNVVDGLR A	62-64, 65-73	0	
	1DNK	LKI -- AAFNIRTFG E	1-3, 4-12	0.2, 0.9	0
	1I9Z	VKI FV ASYNLNGCS A	558-560, 563-571	0.3, 1.1	0.2 1.2
2	1DE9	(EA)PDILCLQETK	89-98	0	
	1DNK	(RR)YDIVLIQEVN	32-41	1.3	
	1I9Z	(AD)IYVVGFOEIV	590-599	1.9	2.0
Gapped:					
	1DE9	PDI --L CLQETK	89-91, 93-98	0	
	1DNK	YDI --V LIQEVN	32-34, 36-41	0.2, 0.5	0
	1I9Z	ADI YVV GFQEIV	588-590, 594-599	0.1, 0.9	0.2, 1.0
For comparison:					
12	1DE9	GSDHCPITL	306-314		
	1DNK	ISDHYPVEV	249-257	0.9	
	1I9Z	YSDHRPIYA	836-844	0.3	0.9

The segmental RMSD between homologous motifs from huAPE1 (from PDB file 1DE9), DNase I (1DNK) and synaptojanin (I9Z) are shown first ungapped, from a simple alignment, and then gapped to match conserved residues (Motifs 1 and 2) and allow insertions in the secondary structure elements. Motif 12 is shown as an example of the RMSD where the molegos are spelled in a similar fashion in all three proteins.

method to build novel protein structures in the lab [36,40]. Here we demonstrate that the concept is also useful to define the structural and functional role of conserved amino acids. Combining the MASIA decomposition approach to sequence analysis (Figure 1 and Tables 1 and 2) with data from crystal structures and site directed mutagenesis (Figure 2,3,4,5,6,7) showed that molegos can indicate areas of the protein that control individual functions that contribute to enzymatic activity and improve alignments for template based modeling of homologues with low identity (Table 2 and Figure 7). Most of the MASIA-motifs were near the DNA, metal ions, or both in the co-crystal structures of huAPE1.

Further, functional roles could be assigned to motifs based on their occurrence in related protein families. The similarity of the 3D structures of these motifs in three distantly related proteins and even their modes of binding substrate (Table 2, Figs. 2,3,4,5,6,7) imply that these molegos will be found in even more distantly related proteins. Several molegos can contribute to the same interacting surface and can thus define domains that are

not linearly located in the protein sequence. The combined structural/sequence definition allows much more flexibility in defining a functional element than is possible with purely sequence based approaches such as PROSITE [41].

Is the specificity of APE determined by binding 3' to an abasic site?

Crystal structure data, coupled with molego analysis, outlined the areas of APE1 that distinguish its mode of DNA binding from the less specific DNase 1. Contact maps (Figure 6) illustrate how the conserved motifs direct DNA binding in the distributive (i.e., rapidly releasing substrate/product), relatively non-specific DNase1 as opposed to the processive, highly specific huAPE1. Both enzymes cleave only one DNA strand in a duplex and bacterial Xth cleaves ssDNA containing an abasic site [42]. The additional contacts huAPE1, compared to DNase 1 (Figure 6), makes 3' to the damage site and to the opposite strand lower its turnover rate and its potential to cleave normal DNA. The residues contacting the region 3' to the abasic site come from three different uniquely conserved

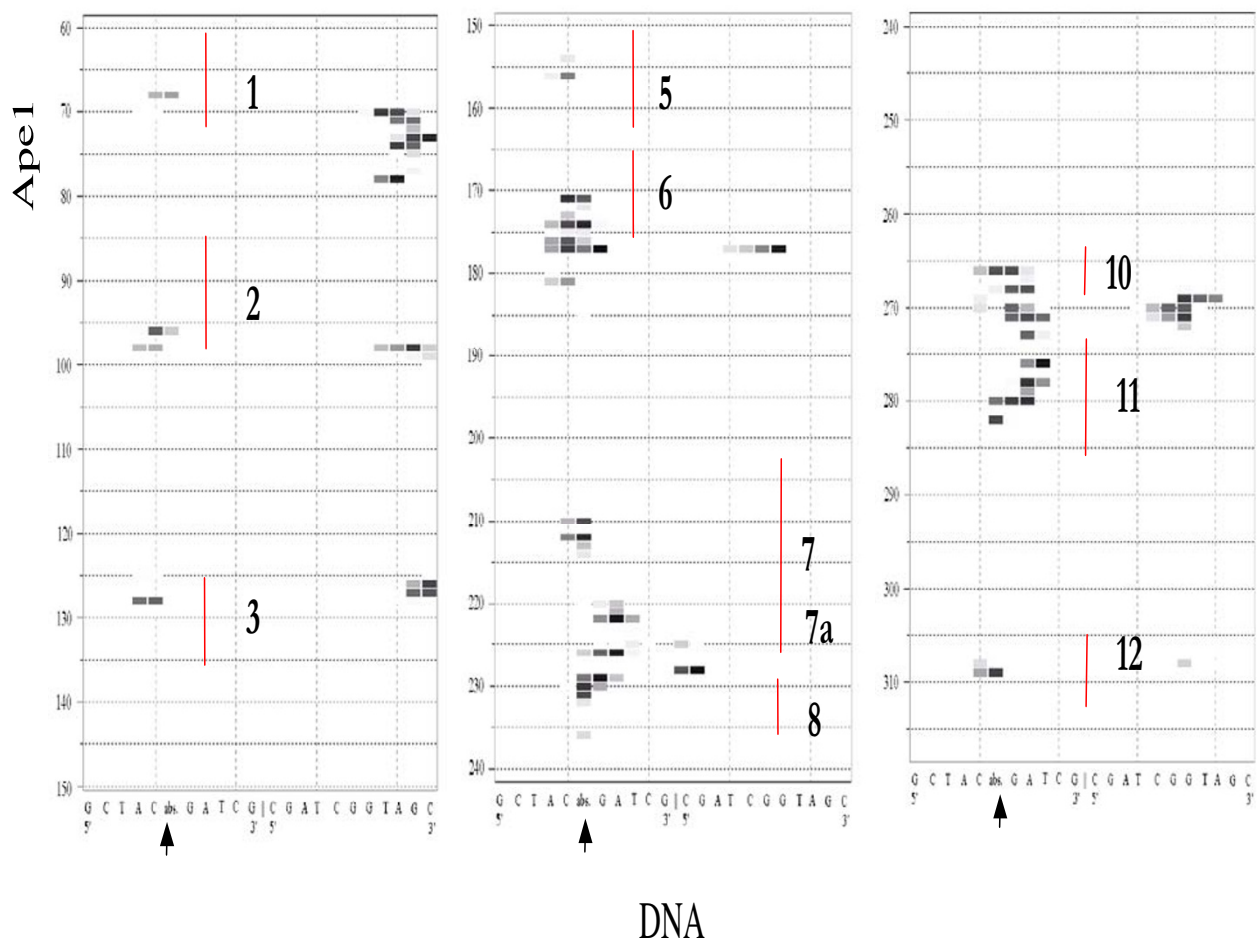


Figure 5
Protein DNA contact plots for huAPE1. Protein/DNA contact plots for huAPE1 binding to substrate DNA taken from the minimized IDE8 structure described in Figure 2. Blocks, black at 1.5 Å, lighten with increasing distance of residues from the DNA up to 7.5 Å. Most of the MASIA motifs (Table I) for the APE family are near the DNA interface. The motifs are indicated by numbers; the abasic site at position 6 is indicated by arrows.

areas of APEs (molegos 6,7a, 10) as well as 11, a molego that is similar to that in DNase-1. These observations, coupled with DNA ethylation data [18], indicates that 3' binding is a key element in specific recognition by APEs.

This is confirmed by site directed mutagenesis studies. Of the four protein areas that bind to the DNA 3' of the abasic site, mutating F266 (molego 10) or W280 (middle of molego 11) decreases APE activity [33]. The F266 mutation is particularly interesting, as the mutants at this position had reduced substrate specificity and enhanced 3'-exonuclease activity. However, an R177A mutant had enhanced APE activity [11], as do mutations at N226 (Izumi et al., in preparation). Combining these mutations however great-

ly decreases substrate binding (Izumi et al., in preparation). The 3' approach to the DNA [34] and the wide area covered by the protein on both sides of the abasic site [14] are both consistent with the need to hold the product until the correct polymerase moves in 5' to 3' to complete the repair [25]. This implies that the mammalian enzyme has evolved to be processive, to facilitate more efficient functioning of the overall BER pathway, and may not be optimized for simple catalysis. Processivity is an important facet of the activity of enzymes that function in complex pathways [43]. Reduced processivity may explain, for example, the repair deficits in Xeroderma pigmentosum (XPA) cells [44]. Our molego approach provides a basis

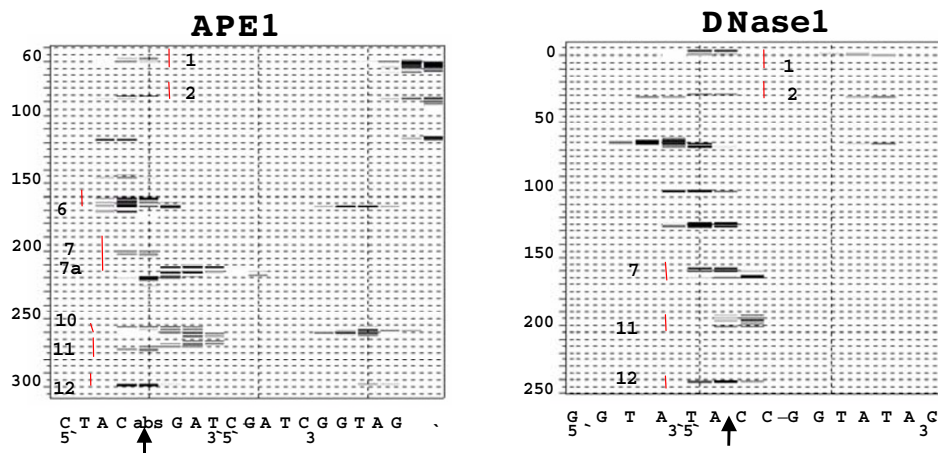


Figure 6
Comparison of DNA contacts in APE1 and DNase I. Comparison of the DNA-contact plots of huAPE1 with product (cleaved DNA, from the IDE9 structure) and DNase I (uncleaved substrate, from IDNK) illustrates the similar mode of binding by the areas conserved between the two proteins. The scissile bonds in both DNAs, 5' of the abasic site in APE1, C315 in IDNK [53], are indicated by arrows.

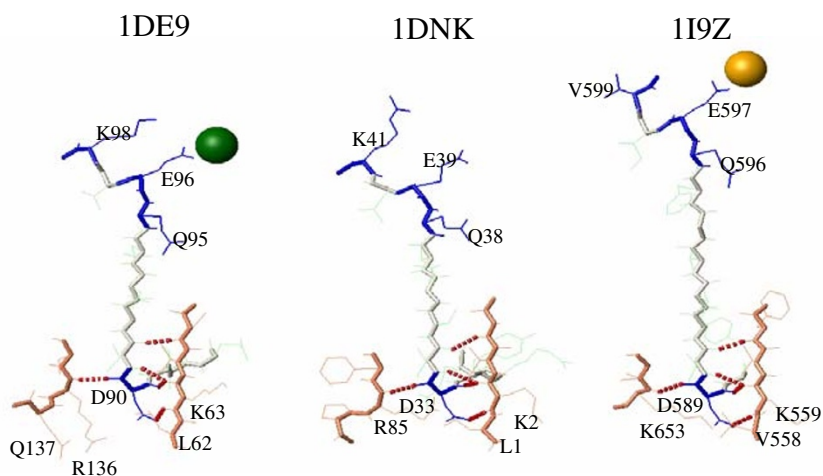
for exploring the role of segments of the protein in its functions, rather than relying only on data from missense mutations.

Using molegos to detect structural and functional homologues

We have demonstrated here the derivation and uses of molegos for analyzing the specificity of enzymes, based on those derived from the APE family. The methodology can be used to complement searches with programs such as PSIBLAST and PROSITE [41] to detect distantly related functional or structural homologues in sequences revealed by genome sequencing. PSIBLAST searches often reveal areas of local similarity in proteins that have no significant overall sequence identity. Molego analysis could be useful to analyze the significance of such findings. The combined sequence and structure definition makes molegos more flexible for defining shared protein elements than methods such as PROSITE that require a strict one-dimensional definition. An improved motif definition

method, based on physical property similarity [42], which has been incorporated into our MASIA tool] also promises to enhance the usefulness of the method. This may eventually lead to a method to find functional relationships between proteins with even lower overall sequence similarity.

Another potential area for applying the molego approach is in homology modeling. Molegos may prove useful in to check alignments for template based modeling of homologues with low identity (Table 2 and Figure 7), if the "anchoring" residues are conserved in sequence or property across the members of both subfamilies. Our molego approach is closest in principle to that of the ROSETTA program [45] whereby the latter seeks only to connect structure, not function, to a sequence element. We are currently testing the usefulness of the molego approach in modeling in the CASP5 competition.

**Figure 7**

Molegos can improve structural alignments. Molegos 1 and 2 have a similar structure in all three proteins, but the β -strand is 2 amino acids longer in synaptojanin. The amino acids around the N-terminal ends of the molegos that form hydrogen bonds to the conserved aspartate (D90 in huAPE1) are shown.

Conclusions

The MASIA program can parse sequences into discrete blocks of significant conservation. The motifs identified in the APE family could be structurally annotated using crystal data to derive molegos, words in the protein sequence that correlate with structural elements. These molegos could in turn be functionally annotated by comparing the DNA binding profile of APE1 with that of the less specific nuclease DNase 1. This analysis indicated that residues binding 3' to the site of phosphorolytic cleavage control the substrate specificity of APE1. These results indicate that molegos can provide a useful basis for identifying specificity determining regions in enzymes with similar active sites but different activity spectra [46,28,3]. Site directed mutagenesis based on these results can define the function of the unique elements of the APEs, and aid in the design of enzymes with altered specificity.

Materials and Methods

Sequence alignment

A BLAST [47] <http://www.ncbi.nlm.nih.gov/BLAST/> search of the "non-redundant" protein database using the whole sequence of human APE1 yielded over 100 related sequences. Some sequence entries represented the same protein, called by different names or isolated in different screens, including many entries for huAPE1, *Drosophila* Rrp 1 protein (~40% identical to the mammalian APE1 in the C-terminal third of the protein), Xth from *E. coli*, and counterparts of this and exodeoxyribonuclease (exo A) sequences from many bacteria, which are about 25% identical to mammalian APE1. The mammalian sequences are highly conserved, with only 6 non-conservative residue variations between the human and murine sequences, 5 of which occur in the apparently unstructured N-terminus. Several proteins with more distant relationship to APE1, such as mammalian and yeast APEIIs, and the CRC protein from *Pseudomonas*, which has no APE activity [48] were in the BLAST list, but were not used for this analysis. To derive functional motifs, the BLAST list was culled to

37 unique sequences with identity ranging from 25% to 98%. These were aligned using the default parameters of CLUSTALW [49,50] <http://www2.ebi.ac.uk/clustalw/>. Sequences were the APE1 protein from human, bovine, monkey, rat murine, *Arabidopsis thaliana* (Mouse-ear cress) *Dictyostelium discoideum* (Slime mold), *Schizosaccharomyces pombe* (Fission yeast), *Caenorhabditis elegans*, *Saccharomyces cerevisiae* (Baker's yeast), *Thermoplasma acidophilum*, *Neisseria meningitidis*, *Methanobacterium thermoautotrophicum*, *Leishmania major*, *Trypanosoma cruzi*, *Coxiella burnetii*; the Rrp1 protein of *Drosophila*; exonuclease III from *E. coli*, *Bacillus subtilis*, *Mycobacterium tuberculosis*, *Haemophilus influenzae*, *Salmonella typhimurium*, *Helicobacter pylori*, *Rickettsia prowazekii*, *Archaeoglobus fulgidus*, *Actinobacillus actinomycetemcomitans*, *Streptomyces coelicolor*, *Synechocystis* sp. PCC 6803, *Haemophilus influenzae*; Exonuclease A from *Streptococcus pneumoniae*, *Treponema pallidum*, *Borrelia burgdorferi*, *Plasmodium falciparum*. Different set conditions and sequence lists were tested in CLUSTALW for their effect on alignment and subsequent motif definition with MASIA. Areas peripheral to the endonuclease domain, such as the 50 amino acid mammalian and the 428 residue *Drosophila* Rrp-1 N-terminal regions were eliminated to improve the consensus.

Identification of motifs using MASIA

Motifs were identified in the aligned sequences using the MASIA consensus macro <http://www.scsb.utmb.edu/masia/masia.html>. Motifs start when at least 3 of 4 consecutive positions are more than 40% conserved according to the dominant criterion [51], and extend until at least 2 positions in a row are less than 40% conserved. To allow for mistakes in the alignment of all the sequences, essential residues are those >90% conserved by MASIA criteria over all sequences in the alignment.

Genomic cross-networking with PSIBLAST

A PSIBLAST search, using huAPE1 as the founder sequence, with an e-value of 0.1 per iteration, did not converge after 6 iterations, but few new sequences were added in the last 2 cycles. Searches with an e-value of 0.01/iteration had similar results, but members of several families were not included until later cycles. Members of the DNase 1, LINE-1 repeats, inositol 5'-polyphosphate phosphatase, Nocturnin, CCR4, cytolethal distending toxin, neutral sphingomyelin phosphodiesterase, and amino acid methyltransferase families were found with expectation values of 10^{-4} or less to be significantly similar to APE1. To determine the presence of motifs in these relatives, a CLUSTALW alignment of at least 5 representatives of a protein family was prepared and analyzed with MASIA for significant areas of conservation. In some cases, alignments taken from literature references (e.g., for IPPs [32]) were used to confirm MASIA results. The motifs common to these families were compared with the APE

motifs of Table 1. Criteria for inclusion (presence of motif) included conservation of residues >90% conserved (side chains shown in blue in the tables) and patterns of polarity (as determined with a macro included in the MASIA packet as a user specified feature).

Molego building and comparison

The drawings of "molego blocks" and structures (Figures 2,3,4 and 7), contact plots of protein/DNA (Figures 5,6), and calculation of the RMSD between similar segments (Table 2) were done with MOLMOL <http://www.mol.biol.ethz.ch/wuthrich/software/molmol/>[52]. The RMSD values in Table 1 were calculated using SwissPDB viewer and the "fit selected residues" option.

Authors' contributions

CHS developed the molego concept, performed the sequence analysis and prepared the manuscript. NO analyzed the structural properties of the molegos, and prepared figures 2,3,4,5,6,7. TI prepared mutants of human APE1 protein to test the conclusions of this paper, and he provided background and list of mutations affecting APE function. WB designed and developed the MASIA program and the modular analysis approach to enzyme functions.

Additional material

Additional File 1

Sequence alignment and MASIA printout for the APE family The original alignment used for the APE family and the MASIA consensus macro output. Note that the *A. thaliana* sequence is consistently an outlier in the file, due to the CLUSTAL-W misalignment of the nuclease area of the protein when the N-terminal sequence is not removed. The real position of the motifs in the *A. thaliana* sequence are underlined.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2105-3-37-S1.doc>]

Acknowledgements

We thank our coworkers Venkatarajan Mathura, Sankar Mitra, and Ovidiu Ivanciuc for helpful discussions. Funding for this project was provided by grants from the Department of Energy (DE-FG03-00ER63041), the Texas Higher Education Coordinating Board (ARP 004952-0084-1999) and the John Sealy Memorial Endowment Fund (2535-01).

References

- Lander ES, Linton LM BB, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, Funke R, Gage D, Harris K, Heaford A, Howland J, Kann L, Lehoczy J, LeVine R, McEwan P, McKernan K, Meldrim J, Mesirov JP, Miranda C, Morris W, Naylor J, Raymond C, Rosetti M, Santos R, Sheridan A, Sougnez C, Stange-Thomann N, Stojanovic N, Subramanian A, Wyman D, Rogers J, Sulston J, Ainscough R, Beck S, Bentley D, Burton J, Clee C, Carter N, Coulson A, Deadman R, Deloukas P, Dunham A, Dunham I, Durbin R, French L, Grafham D, Gregory S, Hubbard T, Humphray S, Hunt A, Jones M, Lloyd C, McMurray A, Matthews L, Mercer S, Milne S, Mullikin JC, Mungall A, Plumb R, Ross M, Showkeen R, Sims S, Waterston RH, Wilson

- RK, Hillier LW, McPherson JD, Marra MA, Mardis ER, Fulton LA, Chinwalla AT, Pepin KH, Gish WR, Chissoe SL, Wendl MC, Delehaunty KD, Miner TL, Delehaunty A, Kramer JB, Cook LL, Fulton RS, Johnson DL, Minx PJ, Clifton SW, Hawkins T, Branscomb E, Predki P, Richardson P, Wenning S, Slezak T, Doggett N, Cheng JF, Olsen A, Lucas S, Elkin C, Uberbacher E, Frazier M, Gibbs RA, Muzny DM, Scherer SE, Bouck JB, Sodergren EJ, Worley KC, Rives CM, Gorrell JH, Metzker ML, Naylor SL, Kucherlapati RS, Nelson DL, Weinstock GM, Sakaki Y, Fujiyama A, Hattori M, Yada T, Toyoda A, Itoh T, Kawagoe C, Watanabe H, Totoki Y, Taylor T, Weissbach J, Heilig R, Saurin W, Artiguenave F, Brottier P, Bruls T, Pelletier E, Robert C, Wincker P, Smith DR, Doucette-Stamm L, Rubenfield M, Weinstock K, Lee HM, Dubois J, Rosenthal A, Platzer M, Nyakatura G, Taudien S, Rump A, Yang H, Yu J, Wang J, Huang G, Gu J, Hood L, Rowen L, Madan A, Qin S, Davis RW, Federspiel NA, Abola AP, Proctor MJ, Myers RM, Schmutz J, Dickson M, Grimwood J, Cox DR, Olson MV, Kaul R, Raymond C, Shimizu N, Kawasaki K, Minoshima S, Evans GA, Athanasiou M, Schultz R, Roe BA, Chen F, Pan H, Ramser J, Lehrach H, Reinhardt R, McCombie WR, de la Bastide M, Dedhia N, Blocker H, Hornischer K, Nordsiek G, Agarwala R, Aravind L, Bailey JA, Bateman A, Batzoglu S, Birney E, Bork P, Brown DG, Burge CB, Cerutti L, Chen HC, Church D, Clamp M, Copley RR, Doerks T, Eddy SR, Eichler EE, Furey TS, Galagan J, Gilbert JG, Harmon C, Hayashizaki Y, Haussler D, Hermjakob H, Hokamp K, Jang W, Johnson LS, Jones TA, Kasif S, Kasprzyk A, Kennedy S, Kent WJ, Kitts P, Koonin EV, Korfi I, Kulp D, Lancet D, Lowe TM, McLysaght A, Mikkelsen T, Moran JV, Mulder N, Pollara VJ, Ponting CP, Schuler G, Schultz J, Slater G, Smit AF, Stupka E, Szustakowski J, Thierry-Mieg D, Thierry-Mieg J, Wagner L, Wallis J, Wheeler R, Williams A, Wolf YI, Wolfe KH, Yang SP, Yeh RF, Collins F, Guyer MS, Peterson J, Felsenfeld A, Wetterstrand KA, Patrinos A, Morgan MJ, Szustakowski J, de Jong P, Catanese JJ, Osoegawa K, Shizuya H, Choi S, Chen YJ and International Human Genome Sequencing Consortium **Initial sequencing and analysis of the human genome.** *Nature* 2001, **409**:860-921
2. Ramos A and Pastore A **Dissecting nucleases into their structural and functional domains: mapping the RNA-binding surface of RNase III by NMR.** In: *Methods in Molecular Biology* (Edited by: Schein CH) Humana Press, Totowa NJ 2001, 237-248
 3. Jeltsch A and Pingoud AM **Methods for determining activity and specificity of DNA binding and DNA cleavage by Class II restriction endonucleases.** In: *Methods in Molecular Biology* (Edited by: Schein CH) Humana Press, Totowa NJ 2001, 287-308
 4. Lichtarge O, Yamamoto KR and Cohen FE **Identification of functional surfaces of the zinc binding domains of intracellular receptors.** *J Mol Biol* 1997, **274**:325-337
 5. Sowa ME, He W, Wensel TG and Lichtarge O **A regulator of G protein signaling interaction surface linked to effector specificity.** *Proc Natl Acad Sci* 2000, **97**:1483-1488
 6. Demple B and Harrison L **Repair of oxidative damage to DNA: enzymology and biology.** *Ann Rev Biochem* 1994, **63**:915-948
 7. Mitra S, Hazra TK, Roy R, Ikeda S, Biswas T, Lock J, Bolodogh I and Izumi T **Complexities of DNA base excision repair in mammalian cells.** *Mol Cells* 1997, **7**:305-312
 8. Mitra S, Izumi T, Bolodogh I, Ramana CV, Hsieh CC, Saito H, Lock J and Papaconstantinou J **Repair of oxidative DNA damage and aging: central role of AP-endonuclease.** In: *Proceedings of the NATO Asi Monograph Plenum Press Publishers, New York, NY* 1999, 295-311
 9. Barzilay G, Mol CD, Robson CN, Walker LJ, Cunningham RP, Tainer JA and Hickson ID **Identification of critical active-site residues in the multifunctional human DNA repair enzyme HAPI.** *Nat Struct Biol* 1995, **2**:561-568
 10. Gorman MA, Morera S, Rothwell DG, de La Fortelle E, Mol CD, Tainer JA, Hickson ID and Freemont PS **The crystal structure of the human DNA repair endonuclease HAPI suggests the recognition of extra-helical deoxyribose at DNA abasic sites.** *EMBO J* 1997, **16**:6548-6558
 11. Mol CD, Izumi T, Mitra S and Tainer JA **DNA-bound structures and mutants reveal abasic DNA binding by APE1 DNA repair and coordination.** *Nature* 2000, **403**:451-456
 12. Erzberger JP and Wilson DM **The role of Mg²⁺ and specific amino acid residues in the catalytic reaction of the major human abasic endonuclease: new insights from EDTA-resistant incision of acyclic abasic site analogs and site-directed mutagenesis.** *J Mol Biol* 1999, **290**:447-457
 13. Izumi T, Malecki J, Chaudhry MA, Weinfeld M, Hill JH, Lee JC and Mitra S **Intragenic suppression of an active site mutation in the human apurinic/aprimidinic endonuclease.** *J Mol Biol* 1999, **287**:47-57
 14. Nguyen LH, Barsky D, Erzberger JP and Wilson DM **Mapping the protein-DNA interface and the metal-binding site of the major human apurinic/aprimidinic endonuclease.** *J Mol Biol* 2000, **298**:447-459
 15. Rothwell DG and Hickson ID **Asparagine 212 is essential for abasic site recognition by the human DNA repair endonuclease HAPI.** *Nucleic Acids Res* 1996, **24**:4217-4221
 16. Masuda Y, Bennett RAO and Demple B **Dynamics of the interaction of human apurinic endonuclease (Ape1) with its substrate and product.** *J Biol Chem* 1998, **273**:30352-30359
 17. Rothwell DG, Hang B, Gorman MA, Freemont PS, Singer B and Hickson ID **Substitution of Asp-210 in HAPI(APE/Ref-1) eliminates endonuclease activity but stabilizes substrate binding.** *Nucleic Acids Res* 2000, **28**:2207-2213
 18. Wilson DM, Takeshita M and Demple B **Abasic site binding by the human apurinic endonuclease, Ape, and determination of the DNA contact sites.** *Nucleic Acids Res* 1997, **25**:933-939
 19. Zhu H, Schein CH and Braun W **Homology modeling and molecular dynamics simulations of PBCV-1 glycosylase complexed with UV-damaged DNA.** *J Mol Model* 1999, **5**:302-316
 20. Hill JW, Hazra TK, Izumi T and Mitra S **Stimulation of human 8-oxoguanine-DNA glycosylase by AP-endonuclease: potential coordination of the initial steps in base excision repair.** *Nucleic Acids Res* 2001, **29**:430-438
 21. Parikh SS, Mol CD, Slupphaug G, Bharati S, Krokan HE and Tainer JA **Base excision repair initiation revealed by crystal structures and binding kinetics of human uracil-DNA glycosylase with DNA.** *EMBO J* 1998, **17**:5214-5226
 22. Waters TR, Gallinari P, Jiricny J and Swann PF **Human thymine DNA glycosylase binds to apurinic sites in DNA but is displaced by human apurinic endonuclease I.** *J Biol Chem* 1999, **274**:67-74
 23. Yang H, Clendenin WM, Wong D, Demple B, Slupka MM, Chiang JH and Miller JH **Enhanced activity of adenine-DNA glycosylase (Myh) by apurinic/aprimidinic endonuclease (Ape1) in mammalian base excision repair of an A/GO mismatch.** *Nucleic Acids Res* 2001, **29**:743-752
 24. Klungland A and Lindahl T **Second pathway for completion of human DNA base excision-repair: reconstitution with purified proteins and requirement for DNase IV (FEN1).** *EMBO J* 1997, **16**:3341-3348
 25. Wilson SH and Kunkel TA **Passing the baton in base excision repair.** *Nature Struct Biol* 2000, **7**:176-178
 26. Barzilay G, Walker LJ, Robson CN and Hickson ID **Site-directed mutagenesis of the human DNA repair enzyme HAPI: identification of residues important for AP endonuclease and RNase H activity.** *Nucleic Acids Res* 1995, **23**:1544-1550
 27. Zhu H, Schein CH and Braun W **MASIA: recognition of common patterns and properties in multiple aligned protein sequences.** *Bioinformatics* 2000, **16**:950-951
 28. Gerlt JA and Babbitt PC **Divergent evolution of enzymatic function: mechanistically diverse superfamilies and functionally distinct suprafamilies.** *Ann Rev Biochem* 2001, **70**:209-246
 29. Cal S, Tan KL, McGregor A and Connolly BA **Conversion of bovine pancreatic DNase I to a repair endonuclease with a high selectivity for abasic sites.** *EMBO J* 1998, **17**:7128-7138
 30. Murzin AG, Brenner SE, Hubbard T and Chothia C **SCOP: a structural classification of proteins database for the investigation of sequences and structures.** *J Mol Biol* 1995, **247**:536-540
 31. Furano AV **The biological properties and evolutionary dynamics of mammalian LINE-1 retrotransposons.** *Prog Nucleic Acid Res Mol Biol* 2000, **64**:255-294
 32. Whisstock JC, Romero S, Gurung R, Nandurkar H, Ooms LM, Bottomley SP and Mitchell CA **The inositol polyphosphate 5-phosphatases and the apurinic/aprimidinic base excision repair endonucleases share a common mechanism for catalysis.** *J Biol Chem* 2000, **275**:37055-37061
 33. Hadi MZ, Ginalski K, Nguyen LH and Wilson DM **Determinants in Nuclease Specificity of Ape1 and Ape2, Human Homologues of Escherichia coli Exonuclease III.** *J Mol Biol* 2002, **316**:853-866
 34. Mol CD, Hosfield DJ and Tainer JA **Abasic site recognition by two apurinic/aprimidinic endonuclease families in DNA base excision.**

- cision repair: the 3' ends justify the means. *Mut Res* 2000, **460**:211-229
35. Holm L and Sander C **Touring protein fold space with Dali/FSSP**. *Nucleic Acids Res* 1998, **26**:316-319
 36. Karle IL, Das C and Balaram P **De novo protein design: crystallographic characterization of a synthetic peptide containing independent helical and hairpin domains**. *Proc Natl Acad Sci* 2000, **97**:3034-3037
 37. Valetti F, Sadeghi SJ, Meharena YT, Leliveld SR and Gilardi G **Engineering multi-domain redox proteins containing flavodoxin as bio-transformer: preparatory studies by rational design**. *Biosens Bioelectron* 1998, **13**:675-685
 38. Wittingham MJ and Sogah DY **A modular approach to polymer architecture control via catenation of biomolecular Lego(R) sets. Polymers containing templated beta-sheets**. *Abstr Pap Am Chem Soc* 1997, **213**:88-PMSE
 39. Messing J and Llaca V **Importance of anchor genomes for any plant genome project**. *Proc Natl Acad Sci* 1998, **95**:2017-2020
 40. Sadeghi SJ, Meharena YT, Fantuzzi A, Valetti F and Gilardi G **Engineering artificial redox chains by molecular 'Lego'**. *Faraday Discuss* 2000, **116**:135-153
 41. Hofmann K, Bucher P, Falquet L and Bairoch A **The PROSITE database, its status in 1999**. *Nucleic Acids Res* 1999, **27**:215-219
 42. Shida T, Noda M and Sekiguchi J **Cleavage of single- and double-stranded DNAs containing an abasic site by Escherichia coli exonuclease III (AP endonuclease VI)**. *Nucleic Acids Res* 1996, **24**:4572-4576
 43. Lloyd RS **Processivity of DNA repair enzymes**. In: *Nuclease methods and protocols* (Edited by: Schein CH) Humana Press, Totowa NJ 2001, 3-14
 44. Feng S, Parrish DD and Lambert MW **A processive versus a distributive mechanism of action correlates with differences in ability of normal and xeroderma pigmentosum group a endonucleases to incise damaged nucleosomal DNA**. *Carcinogen* 1997, **18**:279-286
 45. Simons KT, Bonneau R, Ruczinski I and Baker D **Ab Initio protein structure prediction of CASP III targets using ROSETTA**. *Proteins* 1999, **37**:171-176
 46. Armstrong RN **Mechanistic diversity in a metalloenzyme superfamily**. *Biochemistry* 2000, **39**:13625-13632
 47. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W and Lipman DL **Gapped BLAST and PSI-BLAST: a new generation of protein database search programs**. *Nucleic Acids Res* 1997, **25**:3389-3402
 48. Macgregor CH, Arora SK, Hager PW, Dahl MB and Phibbs PV **The nucleotide sequence of the Pseudomonas aeruginosa pyrE-crc-rph region and the purification of the crc gene product**. *J Bacteriol* 1996, **178**:5627-5635
 49. Thompson JD, Higgins DG and Gibson TJ **CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice**. *Nucleic Acids Res* 1994, **22**:4673-4680
 50. Thompson JD, Plewniak F and Poch O **A comprehensive comparison of multiple sequence alignment programs**. *Nucleic Acids Res* 1999, **27**:2682-2690
 51. Hanggi G and Braun W **Pattern recognition and self-correcting distance geometry calculations applied to myohemerythrin**. *FEBS Lett* 1994, **34**:147-153
 52. Koradi R, Billeter M and Wuthrich K **MOLMOL: a program for display and analysis of macromolecular structures**. *J Mol Graph* 1996, **14**:51-55
 53. Weston SA, Lahm A and Suck D **X-ray structure of the DNase I-d(GGTATACC)2 complex at 2.3 Å resolution**. *J Mol Biol* 1992, **226**:1237-1256

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp

