

## Directed evolution of recombinant serum paraoxonase (PON) variants

Amir Aharoni and Dan S. Tawfik\*

Department of Biological Chemistry, Weizmann Institute of Science, Rehovot 76100, Israel

\* Correspondence: tawfik@weizmann.ac.il

Owing to their detoxifying functions, and roles in drug metabolism as well as the prevention of atherosclerosis, mammalian or serum paraoxonases (PONs) are an intriguing subject of research and a prime therapeutic and engineering target. Initially identified in mammals, PON and PON-related genes have now been found in fowls, zebra fish, and even in invertebrates such as *C. elegans*. The more closely-related PON genes are divided into three classes or sub-families: PON1, PON2 and PON3, that share 60-70% sequence identity (Draganov and La Du, 2003). PONs are calcium-dependent hydrolases that catalyze the hydrolysis of a broad range of esters and lactones. PON1, which is by far the most investigated member of this family, also catalyzes, albeit at much lower rates, the hydrolysis and thereby inactivation of various organophosphates (OPs), including the nerve agents sarin and soman. PON1 is also involved in drug metabolism and is used for drug inactivation. In recent years, it has become apparent that PONs also play an important role in the prevention of atherosclerosis. The levels of PON1 in the blood and its catalytic proficiency appear to have a major impact both on the individual's susceptibility to pollutants and insecticides, and to atherosclerosis. Furthermore, mice lacking the PON1 gene are highly susceptible to atherosclerosis and to OP poisoning. PON1 and PON3 reside in the high-density lipoprotein cholesterol-carrying particles known as HDL ("good cholesterol"). HDL has two key roles: mediation of cholesterol efflux, e.g., from macrophage foam cells in atherosclerotic lesions, and limitation of lipid oxidation in LDL. PONs have been implicated in both activities (Resenberg et al., 2003).

Despite intense research, the structure and mechanism of action of PONs remain enigmatic. The name, paraoxonase, is purely historical, since PON is a hydrolase fam-

ily with one of the broadest specificities known. PON1 is a proficient esterase towards several synthetic substrates, but the paraoxonase activity of PON1 is rather weak. PON2 and PON3 have almost no paraoxonase activity. Moreover, all these activities towards man-made chemicals are promiscuous activities of PONs rather than their primary function. A variety of physiological roles have been proposed for PON1, including phospholipase A2 action (in hydrolysis of platelet activating factor (PAF) as well as of oxidized lipids), and hydrolysis and inactivation of homocysteine thiolactone -- a known risk factor for atherosclerotic vascular disease. Yet the fact that PONs exhibit very low catalytic efficiencies towards these substrates is casting doubts on the physiological relevance of these activities.

Thus far, PONs have been primarily isolated from sera. However, access to recombinant PONs affords many benefits: It would greatly facilitate structural and functional studies and would enable their engineering for improved catalytic efficiency and specificity. For example, the activity of PONs in the hydrolysis of paraoxon as well as other and more toxic organophosphates is too low for in vivo detoxification, and so is their homocysteine thiolactonase activity. However, protein engineering is entirely dependent on the ability to amply express and screen many variants. Initial attempts to express PONs in *E. coli* failed, presumably due to misfolding and aggregation. We have recently obtained, however, the first PON1 and PON3 variants that express in a soluble and active form in *E. coli*. Directed evolution led to PON variants with kinetic parameters similar to those reported for PONs purified from sera that show dramatically increased activities. In particular, we have evolved PON1 variants with OP-hydrolyzing activities 40-fold higher than wild-type (Aharoni et al., 2004).

Conventional approaches to express soluble and active proteins in heterologous expression systems (primarily in *E. coli*) include low-temperature expression, promoters with different strengths, a variety of solubility enhancing tags and modified growth media. These approaches are sometimes successful but suffer from a major drawback: as they do not modify the intrinsic stability of the protein. Thus, even when a soluble fraction of the protein is obtained, the protein may still be inactive or aggregate during purification and storage. This is especially pronounced for structural studies that demand concentrated protein samples. The alternative approach is to engineer the target protein and improve its intrinsic stability and

solubility (Waldo, 2003). Engineering by rational design of site-directed mutations is an option, yet the rules that dictate protein folding and stability are poorly defined, and detailed structural information may not be available for the target protein. Indeed, as the structure of PONs is completely unknown, a different approach had to be taken.

In the past several years, directed evolution has emerged as an alternative approach to rational design, thus enabling the reshaping of functional and structural properties of proteins in the absence of any structural knowledge (Tao and Cornish, 2002). Rather than designing a limited number of site-directed mutants, directed evolution implements an iterative Darwinian optimization process, whereby, the fittest variants are selected from a population of random mutations. Diversity in the gene of interest is typically created by random mutagenesis, or by recombination of closely-related genes using a process called DNA shuffling. Improved variants are identified by screening or selection for the properties of interest. Genes encoding improved variants are then used as parent genes for the next round of evolution, and this process of mutation and selection is repeated until an optimal protein is obtained. Directed evolution has been applied by many laboratories to engineer enzymes for improved robustness (e.g., at extreme temperatures, pH and in organic co-solvents) and changes in their reaction and substrate specificity (Tao and Cornish, 2002). This approach is highly advantageous, requiring no prior knowledge of the protein's structure or mechanism, and was found to be particularly effective for the generation of recombinant PON variants.

We first created a gene library by shuffling four closely-related PON1 genes cloned from human, rabbit, mouse and rat. DNA shuffling mimics homologous recombination in nature. It has proven to be a powerful technique for evolving new proteins in the laboratory, and it was shown that libraries created by DNA shuffling accelerate the evolutionary process relative to libraries created by completely random mutagenesis (Cramer et al., 1998). This is primarily due to the fact that the recombinants are comprised of a pre-selected diversity of functional proteins. Thus, beneficial mutations from one parental gene can randomly recombine with beneficial mutations from other parental genes, to create new, and potentially superior, genes. Shuffled genes are created from randomly distributed, short fragments of the parental genes that are assembled by primerless PCR. Although generally limited to the recombination of genes with >75% homology, techniques

for low-homology recombination were also developed.

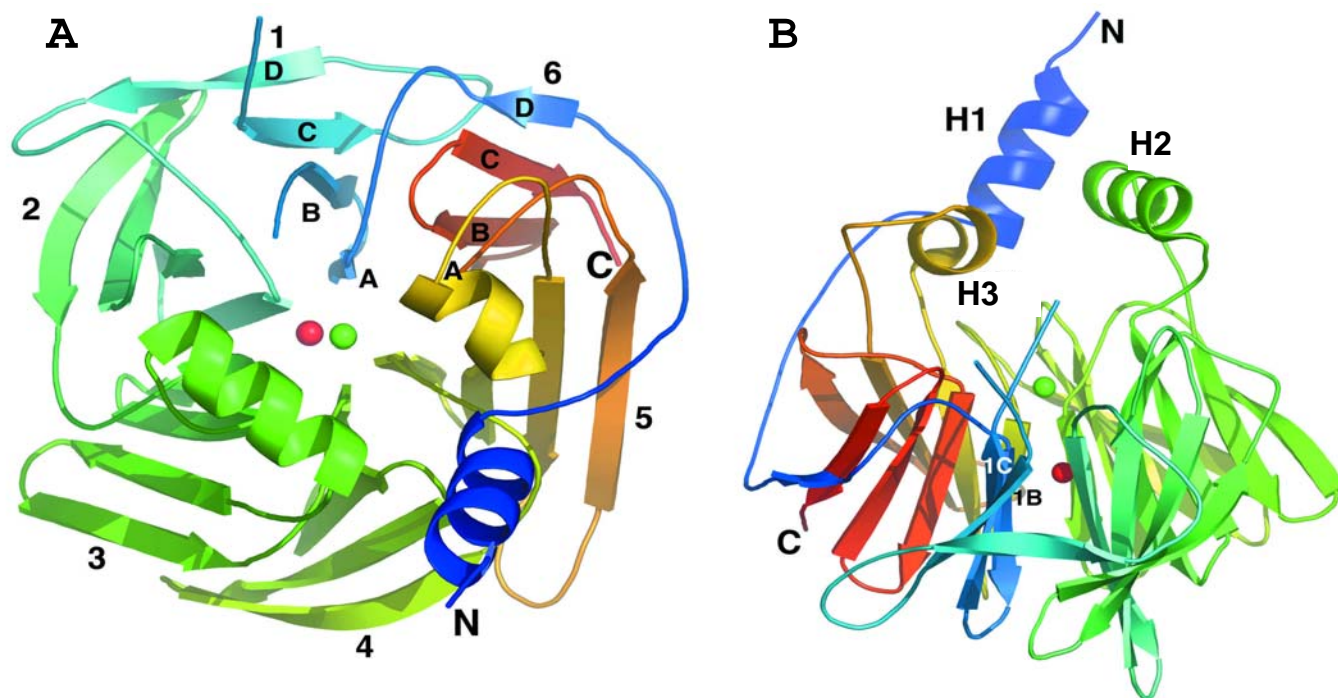
Screening of the shuffled gene library for protein variants that express in *E. coli* in an active form and at large quantities could be done by several ways. Several approaches have been developed that rely on expression of the target protein fused to the N-terminus of a reporter protein with a selectable function. An insoluble target protein then leads to the aggregation of the reporter protein and loss of function (Waldo, 2003). For example, cells expressing the green fluorescent protein (GFP) fused to soluble proteins are brighter than those expressing GFP fused to insoluble proteins. In the case of chloramphenicol acetyl transferase fusion proteins, selection for target protein solubility is performed by antibiotic resistance. These techniques are generally applicable and do not depend on the function of the target protein. They suffer, however, from a major drawback: evolution for increased solubility may be accompanied by significant changes in activity that may even render the soluble protein inactive. We have therefore chosen to screen the library directly for an increase in enzymatic activity. By monitoring several activities of PON1 in parallel, we ensured the selection of soluble PON1 variants with enzymatic properties identical to the wild-type protein.

The library was initially screened on agar plates for esterase activity using 2-naphthylacetate and an azo dye (Fast Red) that reacts with the released 2-naphthol to generate an insoluble red product.  $10^3$ - $10^4$  individual agar colonies were screened in each round. Typically, 200-500 clones exhibiting the highest rate of color formation were picked from a replica plate, grown in liquid medium (in 96-well plates), and the crude cell lysate assayed spectrophotometrically for the hydrolysis of various substrates including paraoxon and 2-naphthylacetate. The twenty best clones were sub-cloned and used as templates for further rounds of DNA shuffling and screening. Following three rounds of shuffling and screening, improved clones of PON1 were isolated. The crude lysates of these clones exhibited a parallel increase of ~20-fold higher rates with all substrates tested. The sequences of the selected variants revealed a clear convergence towards the rabbit PON1 gene. However, eight conserved changes were identified in all the selected variants that originated from the other parental genes. These changes were located in only two regions of the protein and were quite subtle in nature (primarily replacements of one hydrophobic residue into another). The selected variants were expressed in *E. coli* to levels of >20 mg per liter culture.

Studies of the purified proteins indicated almost no differences in the kinetic parameters between variants of the first, second or third generations, and wild-type, serum-purified PON1. This indicated that the directed evolution process led to increased protein solubility only, with no alterations in enzymatic properties.

Once bacterial expression of PON1 has been obtained (and of PON3; see Aharoni et al., 2004), we turned our attention to the structure of PON1 (see below) and its enzymatic properties. PON1, being a hydrolase of very broad specificity, ranging from phosphotriesters to esters (of carboxylic acids) and lactones, we aimed at catalytic specialization for a specific substrate. Reshaping an enzyme's substrate specificity is not a trivial task. Twenty years of protein engineering made it clear that although local, active site interactions dictate substrate specificity, they are not the sole determinant. This is exemplified by the difficulties encountered in converting the substrate

specificity of trypsin to chymotrypsin. These two enzymes have almost identical active sites but they cleave peptides on the carboxylate side of either positively charged residues (trypsin) or hydrophobic residues (chymotrypsin). This was attributed to a single amino acid difference at the specificity pocket of these enzymes. However, sixteen residue mutations, many of which are not in contact with the substrate, were required to engineer the substrate specificity of trypsin to chymotrypsin. This engineering exercise demonstrates the complexity of protein design (Brannigan and Wilkinson, 2002). Thus, as with bacterial expression, we chose to apply directed evolution for which no prior knowledge of the structure and reaction mechanism is required. Gene libraries were prepared by random mutagenesis of a gene encoding a highly soluble recombinant PON1 variant. We used an error-prone PCR method based on substituted wobble base analogs (either dPTP (a 5'-triphosphate nucleoside analogue) or 8-oxo-dGTP) that lead to mismatch and muta-



**Figure 1.** The overall structure of PON1. (A) A view of the 6-bladed  $\beta$ -propeller from its top. Shown are the N- and C-termini, the six blades (labeled 1-6) each of which is comprised of 4  $\beta$ -strands (labeled A-D), and the two calcium atoms in the central tunnel of the propeller (green and red spheres). (B) A side view of the propeller, including the three helices at the top of the propeller (H1-H3). The top calcium ion (green sphere) is a key part of PON1's active site, and the three helices comprise an active-site lid. Helices H1 (the N-termini of PON1) and H2 are also thought to participate in the binding of PON1 to HDL.

tion. The number of mutations per gene was adjusted by varying the nucleoside analog's concentration and number of PCR cycles. The average number of mutations per gene is of crucial importance. High load of mutations leads to libraries encoding a pool of enzymes that are mostly inactive due to stop codons and global effects on the protein's folding. Hence, the probability of selecting positive clones is low. However, some modifications of function require a simultaneous change of more than one residue. Thus, when a high throughput screening method is available, the use of gene libraries with a relatively high rate of mutation could be advantageous. In contrast, low- or medium- throughput screens of  $10^3$ - $10^5$  clones are more compatible with libraries carrying few mutations per gene.

Five different libraries were generated: two dPTP libraries with 98% transitions and 2% transversions, an average of 5 and 14 mutations per gene, and 20 or 6% residual activity in the pools of genes, respectively; and three 8-oxo-dGTP libraries with an average of 85% transversions and 15% transitions, 3 to 10 mutations per gene, and 7 to 13% residual activity. In addition to the plate screen for 2-naphthyl acetate hydrolysis described above, the libraries were directly screened on agar plates for the hydrolysis of a fluorogenic OP substrate, 7-*O*-diethylphosphoryl-3-cyano-7-hydroxycoumarin (DEPCyC). Two variants with significantly increased activity towards DEPCyC were isolated. Both exhibited ~40-fold higher catalytic efficiency ( $k_{cat}/K_M$ ) towards DEPCyC compared to wild-type PON1, and a 50-fold reduction in catalytic efficiency towards phenyl acetate. Thus, a shift of >2,000 fold in specificity was observed: whilst wild type PON1 is essentially an esterase with very weak phosphotriesterase activity ( $k_{cat}/K_M$  for phenyl acetate is >200-fold higher than for paraoxon or DEPCyC), the directly evolved variants were converted into a phosphotriesterase with weak esterase activity ( $k_{cat}/K_M$  for DEPCyC is 10-fold higher than for phenyl acetate).

Using the same libraries, and similar screening methods, we have evolved several other PON1 variants, which specialize in the hydrolysis of thiolactones (and  $\gamma$ -butyrolactone in particular), *O*-acetoxycoumarin (a fluorogenic ester substrate), and esters of long-chain (C<sub>8</sub> to C<sub>12</sub>) carboxylic acids (Harel et al., 2004).

The bacterial expression of PON1 has also enabled us, in close collaboration with the groups of Joel Sussman and Israel Silman, to crystallize and solve the structure of a recombinant PON1 variant at 2.2Å resolution (Harel et

al., 2004). This turns out to be the first crystal structure of a PON family member, and as such, it reveals fundamental aspects of this important enzyme family. PON1 is a 6-bladed  $\beta$ -propeller with a unique active-site lid, which is also involved in HDL binding (Figure 1). Several proteins are associated with HDL, including lecithin:cholesterol acyl transferase (LCAT), but their mode of binding to HDL is still under investigation. PON1 is the first HDL-associated protein whose 3D-structure has been solved, and thereby affords a model for HDL-binding.

The 3D-structure and mechanistic studies also enabled a detailed description of PON1's active site. Its catalytic mechanism, is reminiscent of secreted phospholipase A<sub>2</sub>, in having a catalytic calcium ion which stabilizes the negatively-charged, oxyanionic intermediates of ester, lactone and phosphotriester hydrolysis, and a histidine dyad which presumably deprotonates a water molecule to generate the attacking hydroxide ion. The structure also reveals a deep and hydrophobic active site that may account for PON1's very broad substrate range. The active-site lid is in a close proximity to the part presumed to mediate the binding of PON1 to HDL. This finding suggests that PON1 may be subjected to interfacial activation whereby binding to HDL significantly alters its enzymatic properties. Finally, the 3D-structure and the directed-evolution studies provided insights into the routes by which PON family members, which have now been identified in mammals and other vertebrates, as well as in invertebrates, diverged towards different substrate and reaction selectivities.

In conclusion, the outcome of the directed evolution process is a range of recombinant PON variants with a broad range of enzymatic properties. Expression in *E. coli* provides an ample source of PON1 and PON3 free from other serum proteins, and greatly facilitated the mechanistic and structural characterization of these important enzymes. Finally, the demonstration that PONs are amenable to the powerful tool of directed evolution opens new prospects for improving their detoxification activity toward toxic OPs, and their potency of preventing atherosclerosis. In the longer term, the application of engineered PON variants as pharmaceuticals may be considered (Vellard, 2003).

## References

Aharoni A, Gaidukov L, Yagur S, Toker L, Silman I, Tawfik DS.

Directed evolution of mammalian paraoxonases PON1 and PON3 for bacterial expression and catalytic specialization.

*PNAS* 101:482-487, 2004.

**Summary:** The results described in this paper are reviewed above.

Brannigan JA, Wilkinson AJ.

Protein engineering 20 years on.

*Nat Rev Mol Cell Biol* 3:964-970, 2002.

**Summary:** This very readable review, written as a perspective view, describes the advances in the field of protein engineering since 20 years ago when the first site-directed mutagenesis of a protein was performed. It shows how protein engineering has shaped our understanding of the function and structure of proteins, and yielded new and improved proteins for the industry and for treatment of human diseases.

Cramer A, Raillard SA, Bermudez E, Stemmer WP.

DNA shuffling of a family of genes from diverse species accelerates directed evolution.

*Nature* 391(6664):288-291, 1998.

**Summary:** The article demonstrates the power of family DNA shuffling, showing that libraries created by family DNA shuffling are superior to libraries created by random mutagenesis. The authors have compared the efficiency of obtaining moxalactamase activity from four cephalosporinase genes. A round of random mutagenesis and screening yielded only an 8-fold improvement in the separately evolved genes, whereas shuffling of these four genes, leads to a 270- to 540-fold improvement.

Draganov DI, La Du BN.

Pharmacogenetics of paraoxonases: a brief review.

*Naunyn Schmiedebergs Arch Pharmacol* (369):78-88, 2003.

**Summary:** This comprehensive and up-to-date review of the PON family highlights the natural polymorphism in PONs and its relation to cardiovascular disease, and summarizes research on the possible roles of PONs in the prevention of atherosclerosis and other diseases.

Harel M, Aharoni A, Gaidukov L, Brumshtein B, Khersonsky O, Meged R, Dvir H, Ravelli RBG, McCarthy A, Toker L, Silman I, Sussman JL, Tawfik DS. Structure and evolution of the serum paraoxonase family of detoxifying and anti-atherosclerotic enzymes.

*Nature Struct Mol Biol* 11:412-419, 2004.

**Summary:** The article describes the first structure of a PON family member (a directly evolved variant of

PON1). The 3D-structure and directed-evolution studies permit a detailed description of PON1's active site and catalytic mechanism, and of the routes by which PON family members diverged towards different substrate and reaction selectivities. The structure also provides a model for PON1's anchoring onto HDL.

Rozenberg O, Shih DM, Aviram M.

Human serum Paraoxonase 1 decreases macrophage cholesterol biosynthesis.

*Arterioscler Thromb Vasc Biol* 23:461, 2003.

**Summary:** This article shows that human PON1 significantly decreases the cholesterol biosynthesis in macrophages harvested from PON1-deficient mice. It was also shown that, upon incubation of the macrophages with human PON1, a significant increase in the amount of lysophosphatidylcholine is observed, evidence of PON1 phospholipase A2 activity.

Tao H, Cornish VW.

Milestones in directed enzyme evolution.

*Curr Opin Chem Biol* 6(6):858-864, 2002.

**Summary:** This review summarizes the recent advances in directed evolution of enzymes, listing examples for the evolution of enzymes towards higher activity and stability, and of changing reaction and substrate specificities.

Vellard M.

The enzyme as drug: application of enzymes as pharmaceuticals.

*Curr Opin Biotechnol* 14:444-450, 2003.

**Summary:** This review discusses the use of enzymes as drugs. It lists recent examples in the treatment of several diseases including: Gaucher, Cystic Fibrosis, Fabry's diseases and leukemia. It highlights the advantages of using enzymes as drugs -- their specificity and catalytic turnover, in view of the drawbacks associated with the usage of therapeutic proteins.

Waldo GS.

Genetic screens and directed evolution for protein solubility.

*Curr Opin Chem Biol* 7:33-38, 2003.

**Summary:** This review summarizes new methods for the heterologous expression of proteins with an emphasis on directed evolution for soluble expression in *E. coli*. The review outlines different genetic screens based on fusion of a reporter protein to the target protein and immunological detection.