doi:10.1016/j.jmb.2011.06.037 J. Mol. Biol. (2011) 411, 837–853

Contents lists available at www.sciencedirect.com

Journal of Molecular Biology

journal homepage: http://ees.elsevier.com.jmb

Directed Evolution of Sulfotransferases and Paraoxonases by Ancestral Libraries

Uria Alcolombri, Mikael Elias* and Dan S. Tawfik*

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

Received 7 April 2011;
received in revised form
14 June 2011;
accepted 20 June 2011
Available online
24 June 2011

Edited by J. Karn

Keywords:
serum paraoxonase (PON);
cytosolic sulfotransferases;
SULT;
directed evolution;
protein phylogeny

Large libraries of randomly mutated genes are applied in directed evolution experiments in order to obtain sufficient variability. These libraries, however, contain mostly inactive variants, and the very low frequency of improved variants can only be isolated by high-throughput screening. Small but efficient libraries comprise an attractive alternative. Here, we describe the application of ancestral libraries—libraries based on mutations predicted by phylogenetic analysis and ancestral inference. We designed and constructed such libraries using serum paraoxonases and cytosolic sulfotransferases (SULTs) as model enzymes. Both of these enzyme families exhibit a range of activities in drug metabolism and detoxification of xenobiotics. The ancestral serum paraoxonase and SULT libraries were screened by low-throughput means, including HPLC, using substrates and/or reactions with which all family members exhibit low activity. The libraries showed a remarkably high frequency of highly polymorphic and functionally diverse variants. Screening of as few as 300 variants enabled the isolation of mutants with up to 50-fold higher activity than the starting point enzyme. Structural and kinetic characterizations of an evolved SULT variant show how few ancestral mutations reshaped the active site and modulated the enzyme's specificity. Ancestral libraries therefore comprise a means of focusing diversity to positions and mutations that readily trigger changes in substrate and/or reaction specificity, thereby facilitating the isolation of new enzyme variants for a variety of different substrates and reactions by medium-throughput or even low-throughput screens.

*Corresponding authors. E-mail addresses:
Mikael.Elias@weizmann.ac.il; tawfik@weizmann.ac.il.

Abbreviations used: SULT, cytosolic sulfotransferase; REAP, Reconstructed Evolutionary Adaptive Path; PON, serum paraoxonase; TBBL, 5-thiobuyl-γ-butyrolactone; TBL, γ-thiobutyrolactone; OP, organophosphate; CMP, 3-cyano-4-methyl-7-coumaryl-cyclohexyl-ester; DEPCYC, diethylphosphoryl-3-cyano-4-methylcoumarin; 3NPA, 3-nitrophenyl acetate; 4AAP, 4-acetoxy-acetophenone; IMP, *O*-isopropyl-*O*-(*p*-nitrophenyl) methyl phosphonate; PAPS, 3-phosphoadenosine 5-phosphosulfate; pNP, 4-nitrophenol; 3CyC, 3-cyanoumbelliferone; BPA, bisphenol A; L-thyroxine, 3,5,3′,5′-tetraiodothyronine; PDB, Protein Data Bank; PEG, polyethylene glycol; PAP, 3-phosphoadenosine 5-phosphate; HPLC, high pressure liquid chromatography.

Introduction

Directed evolution is traditionally performed with large libraries created by random mutagenesis. Although large libraries are essential for many tasks, small but effective libraries are of crucial importance, especially when complex or difficult-to-assay activities are needed.[1–7] In random libraries, deleterious mutations are most common, with a frequency of ≥ 0.33,[8] whereas beneficial mutations occur at frequencies around 10^{-3}.[9,10] Thus, as mutations accumulate, most of the library genes become 'nonviable,' as they do not encode folded functional proteins.[9] Methods such as family shuffling,[11] SCHEMA,[6] iterative saturation mutagenesis,[2] simultaneous

multiple-site saturation mutagenesis,[3,12–14] consensus libraries,[15] and Reconstructed Evolutionary Adaptive Path (REAP)[4] were developed with the aim of minimizing the explored sequence space while maximizing functional variability.

Here, we describe ancestral libraries that, similarly to other methods such as family shuffling, aim at exploring new combinations of substitutions that have already been explored by nature and proven functional. However, the mutational diversities explored here are borrowed from the evolutionary past and not from the present. The ancestral libraries described here are composed of active-site substitutions that were predicted to have existed at various nodes and branches of the evolutionary trajectories of a given enzyme family. These ancestral forms are long gone, but their sequences can be reconstructed from the phylogenetic tree.[16] The libraries explored all residues that are located close to, or within, the enzyme's active site and that differ between the starting point enzyme and the ancestor (or ancestors) along the trajectories that gave rise to this enzyme. Some of these substitutions are unique, but many of these are seen in certain contemporary family members. However, the libraries included only predicted ancestral substitutions, and not the entire diversity seen in existing family members. Random combinations of these ancestral mutations were generated by spiking synthetic oligonucleotides at the background of a starting point gene that encodes an existing family member.[17] Two examples are described: mammalian serum paraoxonase (PON) 1 and human cytosolic sulfotransferase (SULT) 1A1. The PON1 case provided a basis for comparison, as we have described the directed evolution of this enzyme by random and targeted mutageneses using high-throughput or even ultra-high-throughput screens,[18,19] as well as neutral drift libraries and medium-throughput screens.[1,20] In contrast, SULT1A1 comprises a new case; to our knowledge, only one directed evolution experiment has been described with this enzyme or with any other SULT (Amir Aharoni *et al.*, submitted for publication).

Here, we provide a generic protocol for the preparation of ancestral libraries, with the aim of generating 'first-generation' mutants that exhibit improved activity with the desired target substrate/reaction. Such variants could then be subjected to standard directed evolution techniques to obtain higher activity. We show that ancestral libraries comprise a relatively small array of variants that are, on average, more active and evolvable than variants obtained by random mutagenesis. Structural and kinetic characterizations of one of the evolved SULT1A1 mutants were also performed, thus indicating how few ancestral mutations can modulate an enzyme's active site.

Results

Designing ancestral PON libraries

The library design strategy is schematically described in Fig. 1.

Our first test case involved PONs, a family that includes the mammalian families PON1, PON2, and PON3 sharing 55–85% amino acid identity.[21] PONs are calcium-dependent hydrolases that catalyze the hydrolysis of a broad range of substrates such as esters, lactones, and phosphoesters. However, their primary substrates are lactones, and the remaining activities are promiscuous.[22] The different families differ in their lactonase specificity (PON1 and PON3's best substrates are lipophilic lactones, whereas PON2's best substrate is a homoserine quorum-sensing lactone) and in their promiscuous activities.[22–24] Sequences of PON family members were collected. To increase the fidelity of prediction, we removed sequences with partial lengths and sequences in which PON's characteristic sequence motifs were missing. The alignment was fine-tuned manually to improve its reliability at gap positions (the alignment is provided as Supplementary Fig. 1). The alignment was used to create a phylogenetic tree, and a bootstrapping analysis was performed to evaluate it (Supplementary Fig. 2). The tree was checked manually to assess its validity (clustering of known PON family members, compliance with the tree of life, etc.). It should be noted that most of these steps (removing sequences that might be irrelevant, refining the alignment, and assessing the tree's quality) are manual and family specific. A standard methodology cannot be described, and the only general rule is to run several iterative rounds of alignment and tree making until the tree has converged to its most convincing format.

Once the tree had been finalized, ancestral predictions were obtained for all nodes of the tree, and the most probable ancestral sequences were determined as described in Materials and Methods. To maximize the potential for changes in enzymatic specificity, we considered for library diversification all positions residing within 12 Å of the catalytic calcium that resides at the bottom of PON1's active site and all other positions of the active-site wall. This list included 94 positions, or 26% of the enzyme's total length.

The first library was based on the ancestral node N8. This ancestor combines all three mammalian PON families (PON1, PON2, and PON3) and chicken PON, whose specificity is unknown. N8 comprises the deepest ancestor in this phylogeny (i.e., most ancient) for which the reliability of prediction is still high. Later, the same process was applied on a deeper node, N6—the ancestor for all vertebrate PONs for which the prediction's fidelity is

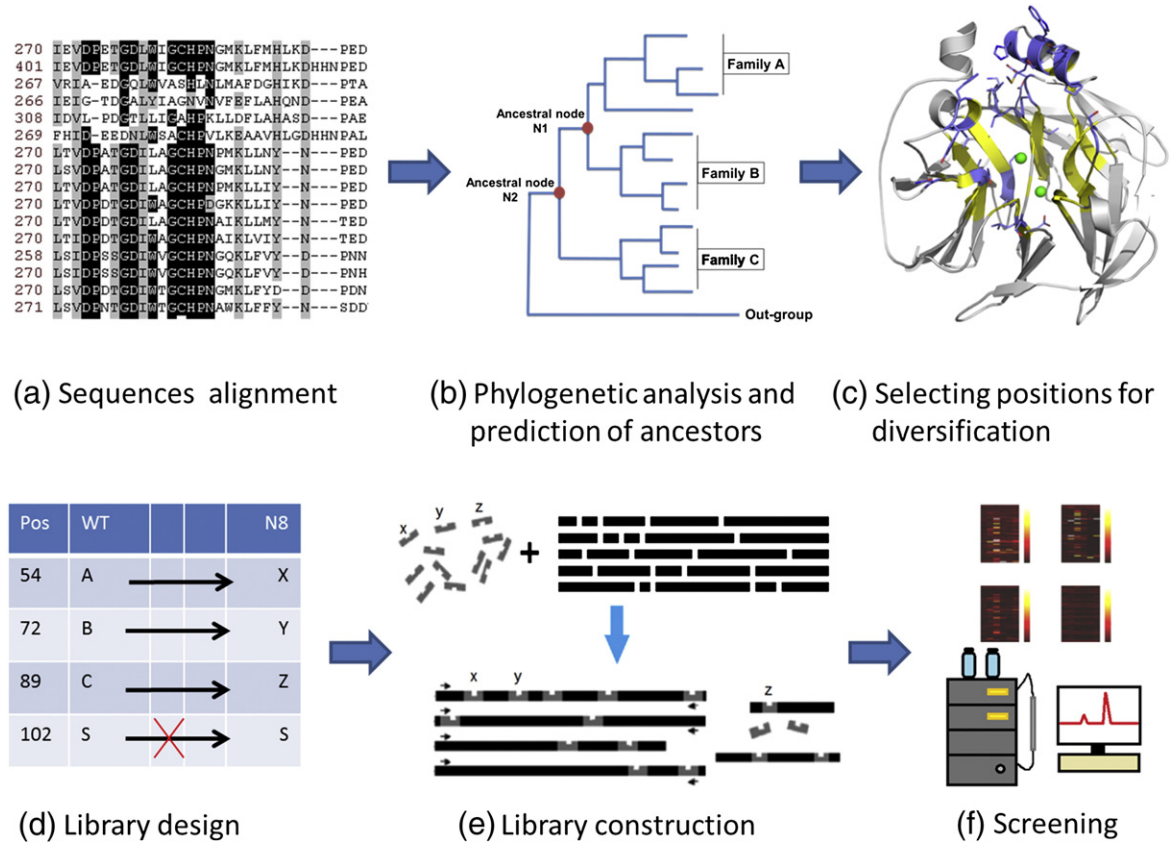

Fig. 1. A schematic representation of the ancestral library methodology. (a) Orthologues and paralogues of the target enzyme are identified, and their sequences are aligned. (b) A phylogenetic tree is generated. The ancestral nodes are assigned, and their sequences are predicted. The reference node (or nodes) for the library design (highlighted in red) typically encompasses sequences of several paralogous families that exhibit ≥50% sequence identity. (c) The diversified library positions are determined based on the enzyme's three-dimensional structure. These positions (yellow) include all positions of the active-site wall and any other position within 12 Å of the enzyme's catalytic center (e.g., the catalytic metal or a key catalytic residue). (d) The sequence of the chosen ancestral node (N8) is compared to that of the starting enzyme (WT). In most positions, the same amino acid is present (e.g., position 102). At positions where the sequence diverged, ancestral substitutions are identified (e.g., A54X, etc.). (e) The ancestral substitutions are incorporated by mutagenesis oligonucleotides that are combinatorially incorporated into the target enzyme's gene by assembly PCR. (f) The resulting gene library is cloned, transformed into *E. coli*, and screened to identify variants with the desired specificity.

much lower. Of the 94 positions considered, 21 positions differed between the starting gene and the N8 ancestor, and 27 positions differed with respect to the N6 ancestor. In the N8 library, all the ancestral substitutions could be identified in at least one contemporary family member. In contrast, in the N6 library, 9 out of 27 ancestral substitutions do not appear in any mammalian PON sequence (Table 1 and Fig. 2; Supplementary Table 1).

As our starting gene, we choose rePON1-G3C9 (dubbed here rePON1), an *Escherichia-coli*-expressed PON1 variant whose sequence is ~95% identical with rabbit PON1 and has enzymatic specificity that is essentially identical with rabbit and human PON1.[20] We chose the engineered rePON1 rather than a wild-type PON1 because we needed a bacterially expressed protein for the library screens (human PON1 and other mammalian PON1s do not express in *E. coli* in a soluble and functional form). The recombinant variant also exhibits higher stability than human PON1[25] and may thus tolerate a larger number and variety of mutations.[26]

The N8-PON library

The N8-PON library was constructed by assembly PCR using DNase-I-generated fragments of the rePON1 gene and an equimolar mixture of 21 mutagenesis oligonucleotides, each encoding one ancestral substitution (Supplementary Table 2). The library was cloned into an expression vector. Sequencing of randomly picked clones indicated 4 ± 2 ancestral substitutions per gene plus 0.1 ± 0.3 mutations that were randomly incorporated during the PCR.

To explore the viability and diversity of the N8 ancestor library, we screened a small number of

Table 1. Diversified positions in the PON ancestral libraries

Position	rePON1 (library starting point)	PON1	PON2	PON3	N8 mammalian+chick ancestor	N6 vertebrate ancestor
55	L	I/L	I	I	I	I
56	E	D/E	D	D	D	T[a]
67	S	S/A/T	S/V	S/T/N	S	T
74	I	I/L/V	L	M/L	L	L
75	M	I/K/M	K/H	P/V	K	P
78	D	Q/N/D/E	A	A	A	S[a]
136	G	D/L/H	H/E	H/N	H	H
137	S	S/A/V/Y	Q/F	M	Q	H[a]
189	P	P/F/Q	P/F	F/S/Y	F	G[a]
190	Y	Y/T	F/I	L/V/F	I	I
193	S	S/Q	Y/F	F/L/T	F	M[a]
194	W	W/L	L	L/F	L	L
196	M	V/I/M/L	M/T	I/M/L	M	V
197	H	Y/H	Y	F/I/V	Y	F
221	D	D/Y	Y/D	S/C/N	Y	Y
222	F	F/S/Y	S	S	S	F
237	I	I/V	V/I	V/I	V	V
239	E	E/D	D	D	D	D
240	L	L/I/V	I	V/A	I	L/I
241	L	L/M	L	S/T/A/F	L	L/F
244	K	K/T/R	N/E	N	N	N
271	I	L/I	L	L	L	I
272	S	S/F	S	S/T	S	E[a]
282	V	V/T	V/T	A	T	I[a]
291	I	I/L	L	L	L	L
293	Y	F/Y/S	V/Y	N/I/M	F	M
294	Y	Y	D/Y	Y	Y	H[a]
329	Q	Q/H	Q	Q	Q	I[a]
332	T	T/S	S	T/S	S	S

Listed are all positions that differ between the starting point (rePON1) and the N8 or N6 ancestors, and the amino acids that occupy these positions in the three mammalian PON families. Positions that differ between N8 and N6 are marked in boldface.

[a] Unique ancestral substitutions (i.e., substitutions to amino acids that do not appear in paralogous families).

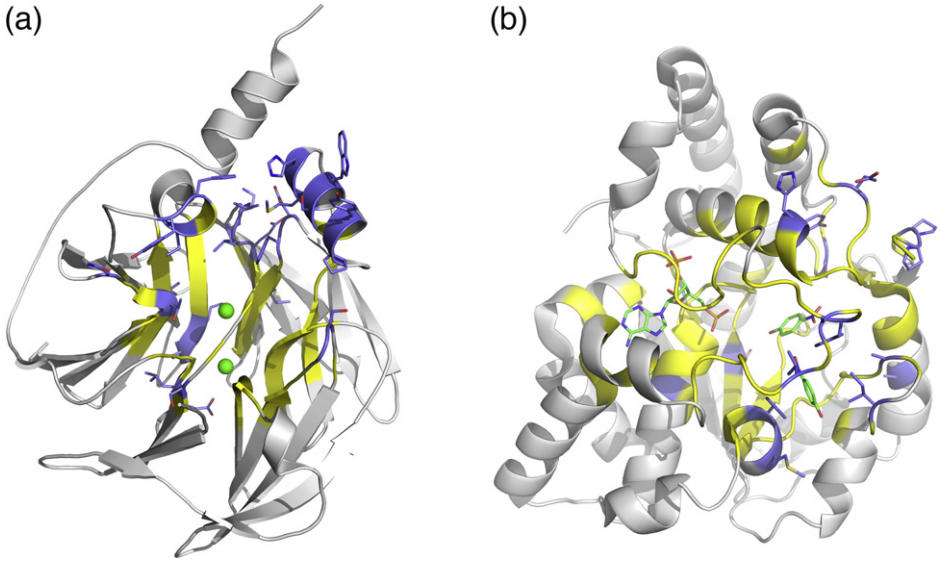

Fig. 2. Locations of the diversified library positions. (a) Backbone representation of rePON1. Marked in color are all positions considered for diversification: the active-site wall and any position within 12 Å of the catalytic calcium (top green sphere). The diversified positions in the N8 and N6 ancestral library positions are shown in blue. The remaining positions (yellow) are those in which the ancestral sequence was identical with that of rePON1. (b) Backbone representation of human SULT1A1. The diversified positions in the ancestral library positions are shown in blue, and those that do not differ between 1A1 and the ancestor are shown in yellow. The substrate (*p*-nitrophenol) and the sulfate donor's product (PAP) are shown in green.

randomly picked variants (276 clones) for six different substrates. These substrates represent three different hydrolytic activities: lactones [5-thiobuyl-γ-butyrolactone (TBBL) and γ-thiobutyrolactone (TBL)], organophosphates (OPs) [methyl phosphonic acid 3-cyano-4-methyl-7-coumaryl-cyclohexyl-ester (CMP) and diethylphosphoryl-3-cyano-4-methylcoumarin (DEPCYC)], and aryl esters [3-nitrophenyl acetate (3NPA) and 4-acetoxyacetophenone (4AAP)] (Supplementary Fig. 3). A control library that contained, on average, 3.2 ± 2.7 random mutations per gene was generated by error-prone PCR and tested for comparison (Fig. 3).

For the six substrates tested, ≥50% of the N8 library variants exhibited detectable activity (≥0.2-fold activity relative to the starting point rePON1), with each of the six substrates tested (Table 2). Moreover, 3.6–22.4% of the screened variants exhibited ≥2-fold higher activity than the wild-type-like rePON1 with one or more of the tested substrates. Furthermore, 1.3–10.7% variants exhibited ≥5-fold higher activity compared to the wild-type-like rePON1.

Differences in lysate activities stem from differences in specific activities and expression levels. Expression levels did vary between library members, but to a much lower degree than specific activities (see the text below). That most differences can be ascribed to differences in specific activities is also manifested in different activities varying to different degrees. For lactonase activity with the chromogenic lactone TBBL, a large portion of the library variants (50.8%) exhibited activities that are essentially identical with those of the wild-type-like rePON1. In addition, none of the tested N8 library variants exhibited significantly improved TBBLase activity (>2-fold higher than rePON1), and only few improved variants were found in the N6 library (1.4%). This stands in contrast to 3.6–22.4% of significantly improved variants observed with all other substrates. Furthermore, the highest improvement in TBBL seen among 552 screened variants taken from both the N8 and N6 libraries was 3-fold, relative to up to 52-fold improvements with the other substrates. This trend relates to the fact that TBBL was designed to mimic lipophilic γ-lactones and thus reflects the native activity of PON1 and, to a large degree, of all PONs.[27] The rarely observed improvements in TBBLase activity suggest that N8, and even the more the ancient ancestor N6, exhibited TBBLase activity levels similar to those of contemporary PON1.

Library variants exhibiting the highest changes in activity were sequenced and recloned to validate their activity. These variants were screened with two additional organophosphate substrates: *O*-isopropyl-*O*-(*p*-nitrophenyl) methyl phosphonate (IMP) and paraoxon. The most active clones that emerged from this screen can be divided into three major groups (Table 3): the first group (variants 2G1, 1B10, 1D2, 3D7, 1E11, and 2F6) showed higher activity than wild-type PON1 towards organophosphate

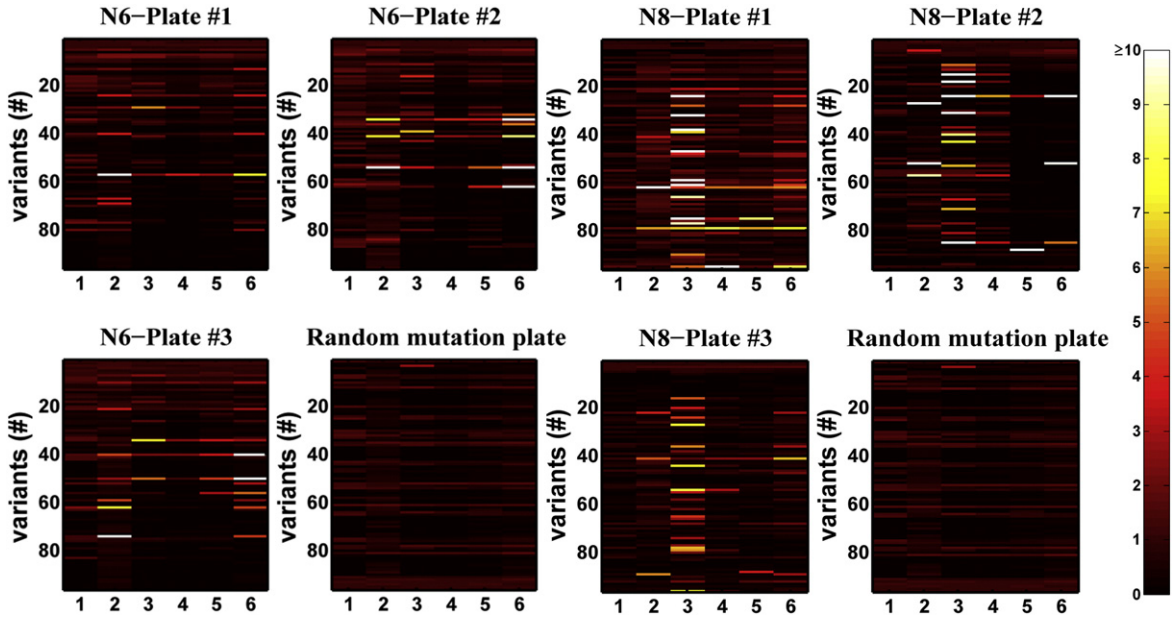

Fig. 3. Activity changes in the ancestral libraries compared to the random mutagenesis library. The libraries were screened with six different substrates: (1) TBBL, (2) TBL, (3) CMP, (4) DEPCYC, (5) 3NPA, and (6) 4AAP. The color bar represents activity compared to the starting point rePON1. The black background denotes activity levels that are lower than those of rePON1. The red-to-white scale denotes equal or higher activity relative to rePON1, whereby white denotes ≥10-fold higher activity.

Table 2. Percentage of active variants in the N8 and N6 ancestral libraries and in the random mutagenesis library

	TBBL				TBL		
Relative activity[a]	N6	N8	RM	Relative activity[a]	N6	N8	RM
Act<0.2	43.5±4.8[c]	49.2±6.2	60	Act<0.2	11.2±6.5	22.1±12.7	16
2>Act≥0.2[b]	55.1±4.1	50.8±6.2	40	2>Act≥0.2	80.4±7.6	73.6±11.7	84
Act≥2	1.4±0.7	0.0±0.0	0	Act≥2	8.4±1.3	4.3±1.1	0
	CMP				DEPCYC		
Relative activity[a]	N6	N8	RM	Relative activity[a]	N6	N8	RM
Act<0.2	56.2±3.5	31.2±5.3	65	Act<0.1[b]	51.1±6.6	22.1±17.2	58
2>Act≥0.2	38.8±2.3	46.4±0.7	34	2>Act≥0.1[b]	46.4±6.4	72.8±15.8	42
Act≥2	5.0±1.3	22.4±5.1	1	Act≥2	2.5±0.7	5.1±2.7	0
	3NPA				4AAP		
Relative activity[a]	N6	N8[b]	RM	Relative activity[a]	N6	N8[c]	RM
Act<0.2[b]	60.2±4.6	54.0±35.1	61	Act<0.2	70.3±2.5	58.4±32.0	68
2>Act<0.25[b]	36.2±5.1	42.4±32.8	39	2>Act≥0.2	14.0±3.0	33.7±25.7	32
Act≥2	3.6±2.3	3.6±2.5	0	Act≥2	7.7±2.5	7.9±6.4	0

RM = random mutagenesis library.
Noted for each substrate is the percentage of variants that show a certain activity level relative to the wild-type-like starting point rePON1.
[a] Act = activity relative to rePON1.
[b] The limit for the lowest detectable activity was defined as <0.1-fold relative to the starting point rePON1, or up to <0.25-fold, depending on the rate of background hydrolysis and the assay's variability for the tested substrate.
[c] For the N6 and N8 libraries, three plates were screened, with 92 library clones per plate. The error ranges reflect the deviations in the percentage of active clones between these three plates. Only one plate was screened from the random mutagenesis library, and error ranges could not be provided. However, for most substrates, the differences between the ancestral N6 and N8 libraries and the random mutagenesis library are well beyond the error range. The exceptions are the aryl esters 3NPA and 4AAP, which exhibit high background rates and low PON1-catalyzed rates, and therefore show relatively high error ranges.

substrates. Variants in this group contained the mutation F222S, which has also been identified from previous screens for an increased organophosphate hydrolysis activity of PON1.[1,19] These variants contained at least one other mutation (or sometimes few mutations) that is likely to mediate a higher organophosphate hydrolase activity, such as Y190F and Y293F.[1] The second group (variants 2D9, 2F10,

Table 3. Selected variants isolated from the N8-PON library

Variant	CMP	IMP	DEPCYC	PARA	TBL	TBBL	3NPA	4AAP	Substitutions
2G1	52.4±12.2	7±1.9	6.8±1.5	7.1±1.8	1.2±2.3	1.4±0.3	3.8±1.0	12.5±3.1	L55I, I74L, Y190F, F222S, F264L,[a] Y293F, T332S
1B10	46.6±8.6	22.7±3.5	6.5±1.0	11.8±2.1	1.1±0.3	1.3±0.2	1.1±0.4	2.6±0.3	D78A, F222S, T→C (309)
1D2	29.9±7.4	21.9±4.4	2.1±0.4	2.5±0.6	1±0.5	1.1±0.2	1.6±0.3	4.9±1.1	E56D, I74L, F222S, I237V
3D7	25.2±2.3	11.6±1.1	7±0.6	11±0.8	1.1±0.5	0.4±0.0	1.2±0.2	1.8±0.2	L55I, S193F, F222S, V282T, Y293F, T332S
1E11	20.7±3.4	14.3±2.0	2.2±0.4	3.2±0.3	1.4±0.8	0.7±0.1	1±0.2	1.3±0.2	L55I, S137Q, F222S, I237V, K244N, G→C(732)
2F6	16.7±4.5	11.7±3.0	1.8±0.5	2.3±0.7	0.9±0.2	0.6±0.2	0.8±0.1	1.3±0.2	L55I, S137Q, Y190I, F222S, C→T (993)
1B6	4.7±1.4	5.5±1.6	6.8±1.9	10.1±3.2	6.2±2.5	1.3±0.2	5.8±1.5	4.8±1.2	M75K,Y190I, L259S,[a] Y293F, T332S
2D4	4.4±1.0	1.9±0.2	6±1.1	7.4±1.0	6±1.0	1.2±0.1	4.1±0.6	3.2±0.6	L55I, P189F, S193F, Y293F, T332S
3E8	3.7±1.5	4.1±1.6	5.1±1.9	5.5±2.1	5.4±2.0	1.4±0.3	4.4±1.4	5.4±1.8	M75K, S137Q, I237V, T332S
1A2	3.6±1.0	6.2±2.2	9.8±3.7	6±2.2	1.2±0.4	0.9±0.3	0.7±0.2	3.9±1.2	E56D, F222S, I291L, T332S
2D9	1.4±0.2	0.4±0.1	1.1±0.1	2.4±0.3	22.8±1.9	1.7±0.1	0.8±0.3	9.9±1.2	M75K, S137Q, I237V, I291L, T332S
2F10	1±0.2	0.2±0.1	0.9±0.2	2±0.4	12±2.2	0.9±0.2	0.6±0.2	3.8±0.9	L55I, P189F, I237V, I291L, T332S, T→C (180)
3A8	0.9±0.2	0.2±0.0	0.8±0.2	1.5±0.3	5.5±1.0	0.6±0.1	0.7±0.1	2.9±0.3	P189F, I237V, I291L, T332S
1E6	0.8±0.3	0.1±0.1	0.2±0.2	0.1±0.2	2.8±1.0	1.3±0.1	0.5±0.1	3.5±0.3	L55I, I74L, V282T, I291L
rePON1-G3C9	1±0.3	1±0.1	1±0.2	1±0.1	1±0.7	1±0.0	1±0.1	1±0.3	—

Listed are the activity levels with various substrates relative to the wild-type-like rePON1. Standard deviations were obtained by measuring the activities of four individually grown colonies of the same variant.
[a] Random mutations incorporated by PCR amplifications of the library.

3A8, and 1E6) was improved mainly with the thiolactone TBL and the aryl ester 4AAP, and also exhibited a decrease in organophosphate hydrolase activity, especially with IMP. Variants belonging to this group all contained the mutation I291L that has been previously seen to affect aryl esterase and lactonase activities.[18,28] Other mutations found in various members of this group, such as I74L and S193F, are also likely to mediate new specificities.[29] The third group (variants 1B6, 2D4, 3E8, and 1A2) was improved with all different substrates tested to a similar degree, suggesting higher protein stability and solubility. Variants of this group all contained the T332S mutation, which contributed to their higher TBLase activity and is likely to mediate their higher expression levels.[18]

The N6-PON library

N6, a node that is deeper and more challenging to predict, was also examined in order to test the impact of the particular choice of node. This library includes 27 substituted positions compared to 21 positions in the N8 library (Table 1). The N6 library also included substitutions that do not appear in any of the mammalian PON family members (9 out of 27). As with the N8 library, we analyzed the activity of 276 randomly picked library clones and compared them to the random mutagenesis library (Fig. 3). The N6 library exhibited 3.6–8.4% of improved variants, depending on the substrate examined (≥2-fold higher activity compared to the starting point rePON1). The improvements observed were up to 20-fold (Table 2).

N6 library variants that exhibited the largest changes in specificity were recloned and sequenced (Table 4). A large number of N6 variants improved with different substrates. Most notable were large improvements towards aryl esters, in particular with 4AAP (e.g., 20-fold higher activity of variant 2F10 compared to rePON1), that were not observed in the N8 library. The improvements with organophosphates were, however, much lower than observed with N8 library variants (≤5-fold as opposed to ≤52-fold). These differences probably relate to mutations like F222S that appear in the N8 library but not in the N6 library, and mutations such as P189G that appear only in N6. Notably, the latter is also a mutation that appears in the ancestral node N6, but not in any of the family members that diverged from it (Table 1). Thus, exploring more than one node can further broaden the diversity of ancestral libraries.

Ancestral *versus* random libraries

In total, by screening 276 variants from each library, we identified 75 improved variants (≥2-fold higher activity with at least one of the six substrates) with the N8 library and 37 improved variants in the N6 library. Thus, on average, 1 out of 5 library variants showed improved activity with at least one substrate. In comparison, only 2 mildly improved variants were identified by screening 92 variants in

Table 4. Selected variants isolated from the N6-PON ancestral library and the random mutagenesis library

Variant	TBBL	TBL	CMP	DEPCYC	3NPA	4AAP	Substitutions
2F10	3±0.2	12±1.9	3.6±0.5	1.7±0.2	6.3±0.6	20.1±3.0	S67T, D136H, P189G,[a] T332S
3G10	1.6±0.1	1.8±0.2	3.6±0.2	1±0.02	4.3±0.4	13.5±0.4	I74L, Y190I, E239D
2E3	2.2±0.7	1.1±0.1	1.9±0.9	0.3±0.1	2.2±0.6	9.1±3.6	R32H,[b] I74L, M196V, L241F, Y293M
3D1	1.9±0.2	3.1±0.2	2.5±0.2	2.3±0.1	2.3±0.2	7.9±1.2	Y293M, T332S
2A9	1.9±0.2	4±0.9	1.9±0.2	1.5±0.1	2.2±0.2	6.3±0.8	S272E,[a] V282I,[a] Y293M, T332S
1G11	1.9±0.3	8.8±2.7	2.5±0.5	3.7±1.0	2.7±0.6	4.9±1.0	L55I, P189G,[a] T332S
2G11	1.7±0.1	5±0.5	1.4±0.4	2.8±0.2	2±0.6	4.4±0.4	S67T, Y190I, W194L, I237V, T332S
3B4	0.5±0.0	1.1±0.1	0.7±0.02	0.3±0.01	3.6±1.3	4.2±0.1	E56T,[a] I74L, S137H,[a] P189G,[a] M196V, E239D, S272E[a]
3A5	2.1±0.2	4.9±0.8	0.4±0.04	0.6±0.1	2.2±2.8	2.9±1.1	L55I, S137H,[a] L241F, V282I,[a] I291L
1F12	1.9±0.2	2.2±0.2	1.4±0.1	1.7±0.2	1.1±0.1	1.9±0.1	D136H, L241F
1C10	1±0.3	2.6±0.9	0.4±0.1	0.3±0.1	4±1.4	1.7±0.5	S137H,[a] S193M,[a] I291L, Q329I[a]
3B8	0.5±0.1	1.2±0.1	4±0.6	2.5±0.6	2.5±0.7	1.7±0.4	S67T, M75P, S137H,[a] E239D
1E1	1.5±0.1	1.9±0.1	5.3±0.2	2±0.2	1.1±0.2	1.6±0.1	L55I, S67T, M75P, D136H, Y190I, W194L, L241F, V282I[a]
1C8	0.5±0.2	1.4±0.6	1±0.4	1.1±0.4	4.6±3.6	1.5±0.2	E56T, S193M,[a] H197R,[b] V282I,[a] T332S
2B4	0.5±0.1	2.9±2.8	3.6±1.1	1.2±0.3	1.2±0.2	0.5±0.3	M75P, S137H,[a] V282I[a]
RM-A3	0.03±0.03	0.9±0.1	1.3±0.1	0.5±0.2	0.5±0.1	0.6±0.3	H115Q[b]
RM-A12	1.1±0.1	1±0.1	1.4±0.0	0.9±0.0	1±0.1	0.5±0.03	F24I[b]
rePON1-G3C9	1±0.1	1±0.1	1±0.1	1±0.1	1±0.1	1±0.3	—

RM = random mutagenesis library.

Listed are the activity levels with various substrates relative to the wild-type-like rePON1. Standard deviations were obtained by measuring the activities of four individually grown colonies of the same variant.

[a] Unique ancestral substitutions to amino acids that do not appear in any of the exiting mammalian PONs (see Table 1).

[b] Random mutations incorporated by PCR amplifications of the library.

the random mutagenesis library (Table 4). If we consider ≥5-fold at the threshold for improved activity, the average is 1 out of 10 for the N6 and N8 ancestral libraries, and nil for the random library.

The outcome with ancestral libraries is also favorable if we consider the "viability" of a library —meaning the frequency of folded and functional enzyme variants. The latter is reflected in the ability to hydrolyze the lactone substrate TBBL: 60% of "dead" variants (lactonase activity within background rates, i.e., ≤0.2 relative to rePON1) were observed in the random library compared to an average of 43.4% in the N8 and N6 libraries (Table 2).

Expression levels of library variants

The expression levels of library variants were also examined (Supplementary Fig. 4). The expression levels in all three libraries (N8, N6, and random mutagenesis library variants) were, on average, about half relative to rePON1 and more widely distributed as expected. The expression levels in library variants varied from no expression up to 1.7-fold higher expression than the starting point rePON1. As expected, most inactive clones (as determined by activity with TBBL) showed no expression or very low expression. Since the observed increases in activity in improved variants were much higher than 1.7-fold (Tables 3 and 4) and since the improvements in most cases were substrate-specific (except for variants belonging to the third group of the N8 library: 1B6, 2D4, 3E8, and 1A2; see the text above), it is clear that most improvements stem from changes in specific activities.

Ancestral SULT library

The substrates used with the PON library experiments are all colorimetric and/or fluorogenic, making it relatively easy to screen for new activities and specificities by high-throughput or even ultra-high-throughput means.[18] However, we also searched for new variants in a more challenging system where the screen is much more laborious. To this end, we explored the enzyme family of SULTs. The SULT family comprises enzymes that use 3-phosphoadenosine 5-phosphosulfate (PAPS) as a donor molecule to transfer a sulfate group to different acceptor substrates. We used two known colorimetric substrates, 4-nitrophenol (pNP) and 3-cyanoumbelliferone (3CyC), but the actual substrates of interest, bisphenol A (BPA) and 3,5,3′,5′-tetraiodothyronine (L-thyroxine), could only be screened by low-throughput assays such as high pressure liquid chromatography (HPLC) (Supplementary Fig. 5). We chose the latter two substrates because all characterized SULTs sulfonate them with low rates[32–34] and because of the potential interest in their detoxification.

Design of the ancestral SULT library

The design protocol for the SULT library followed that of the PON libraries. Candidate residues for mutagenesis included all residues within 12 Å of the reaction center of SULT1A1 [the center was defined as the hydroxyl group of the *p*-nitrophenyl substrate; Protein Data Bank (PDB) ID: 1LS6],[35] including the binding pocket cleft wall and 4 Å around the binding pocket of the donor molecule PAPS (Fig. 2b; Supplementary Table 3). From these positions (99 in total), we choose the ones that differ between wild-type human SULT1A1 (our starting point gene) and the ancestor N8. The latter comprises the ancestor of the "1s" SULT family, which includes the 1A, 1B, 1C, 1D, and 1E families (the alignment is provided as Supplementary Fig. 6, and the tree is provided as Supplementary Fig. 7). We chose N8 because the prediction for the deeper node (N7) was far less accurate (bootstrapping value of 44 out of 100) and also because the inclusion of more SULT families created too many gaps in the alignment. The ancestral

Table 5. Positions that differ between human SULT1A1 and its N8 ancestor

Residue	SULT1A1	SULT1A	SULT1B	SULT1C	SULT1D	SULT1E	SULT-N8 ancestor
10	P	P	KRQ	PQRSL	R	DTSKE	Q
11	P	P	NPDE	EKAPTSQNG	E	YF-VE	E
27	A	ANT	NGSD	NTIWM	EHQ	QYFN	N
44	S	SN	VATP	ACSV	S	VA	A
54	V	VM	IVL	TIM	VI	LVI	I
76	F	FY	TVSY	FQYHL	Y	F	Y
77	M	LIMAV	AVNESR	DHEVQKL	KR	N	N
86	A	AICVD	LVA	LRIF	I	TKNS	I
89	I	VIMACE	LQIAKV	PQLISK	FI	VLM	L
143	Y	YH	DH	SQYH	YH	IFLH	H
144	H	RHKN	LQ	RK	Q	LR	R
151	E	DENH	LAHFY	DNAH	KE	DHNP	D
168	S	CS	AS	LVSCA	AS	P	C
243	V	ILV	LVI	LVIDW	IM	LVM	I
245	Q	TPCQSER	ST–	TKSNAVG	DK	DE	S
247	F	VIFL	VLMI	IYVL	DE	VIMS	I

Table 6. Selected variants isolated from the SULT-N8 library

Variants[a]	pNP[b]	3CyC[b]	BPA[b]	l-Thyroxine[b]	Mutations
a10	2.41 ± 0.05	2.45 ± 0.2	9.4 ± 0.4	0.5 ± 0.1	S44A, M77N, V79M,[c] A86I, E151D, F247I, M260I[d]
b2	2.51 ± 0.17	6.97 ± 0.9	0.48 ± 0.1	2.76 ± 0.1	M77N, A101S,[d] Y143H, E151D, S168C, V243I, F247I
b9	3.66 ± 0.11	19.3 ± 4.4	2.09 ± 0.2	10.49 ± 1.8	P10Q, M77N, E151D, S168C, V243I, F247I
d6	4.24 ± 0.49	1.03 ± 0.3	1.11 ± 0.1	0.18 ± 0.0	P11E, Q56E,[d] L67V,[d] F76Y, A101S,[d] Y143H, M145I,[c] S168C, H213R,[d] F247I, H250Y[c]
1A1-1C5[e]	0.13 ± 0.01	2.53 ± 0.3	1.29 ± 0.08	1.58 ± 0.0	Q56E,[d] L67V,[d] A101S,[d] H213R,[d] F247I,[d] M260I[d]
1A1-1E9[e]	0.27 ± 0.02	2.07 ± 0.5	2.04 ± 0.1	1.18 ± 0.0	L67V,[d] A101S,[d] Q177K,[d] V211L,[d] F222K,[d] F247I[d]
1E1[f]	0.02 ± 0.01	0.2 ± 0.3	0.21 ± 0.1	0.33 ± 0.00	1E1
Human 1A1	1 ± 0.02	1 ± 0.01	1 ± 0.1	1 ± 0.1	1A1

[a] Listed are activities as fold change relative to human SULT1A1, as measured with purified enzyme samples. Standard deviations were obtained from three independent measurements.
[b] Assay conditions for each substrate tested (enzyme concentration, substrate concentration, and PAPS concentration, respectively): (1) $[E]_0 = 2.86$ μM, 50 μM pNP, and 0.5 mM PAPS; (2) $[E]_0 = 1.43$ μM, 1 μM 3CyC, and 100 μM PAPS; (3) $[E]_0 = 2.86$ μM, 100 μM BPA, and 500 μM PAPS; (4) $[E]_0 = 2.86$ μM, 25 μM L-thyroxine, and 100 μM PAPS.
[c] Random mutations incorporated by PCR amplifications of the library.
[d] Stabilizing mutations.
[e] Variants 1A1-1C5 and 1A1-1E9 were obtained by introducing consensus, stabilizing mutations, and selecting for higher expression levels in *E. coli* and higher thermal stability. They also contained the F247I active-site mutation (Amir Aharoni, submitted for publication) that was also included in the ancestral substitutions and may account for the changes in substrate specificity observed in variants 1A1-1C5 and 1A1-1E9. The magnitude of changes in the ancestral library variants, however, is much larger, and this mutation on its own is therefore incapable of inducing the observed changes.
[f] The activity observed with human SULT1E1 is provided for reference.

substitutions included in the library are listed in Table 5. As can be seen, all the introduced substitutions appear in at least one SULT member. To increase the mutational tolerance of the library variants, we included stabilizing consensus mutations in the library.[30,36] These mutations have been found to stabilize human 1A1 and to increase its levels of functional expression in *E. coli* (Amir Aharoni *et al.*, submitted for publication). In total, we introduced by oligonucleotide spiking and shuffling 16 ancestral mutations and 8 stabilizing mutations at an average of 5.1 ± 2.4 mutations per gene.

Screening the SULT-N8 library

We started with 184 randomly chosen library variants that were tested for quenching the fluorescence of 3CyC as an indicator for folded active enzymes. Active variants exhibiting a detectable level of sulfotransferase activity (92 in total, ~50% of the variants) were collected, regrown, and tested with colorimetric pNP. The ratio of activities with 3CyC to activities with pNP served as a preliminary indication for changes in substrate specificity. Four selected variants that exhibited the highest deviations relative to human 1A1 were subsequently collected, recloned, overexpressed, and purified. The purified enzyme variants were tested with 3CyC and pNP, and with the target substrates BPA and L-thyroxine, by monitoring the exchange of PAPS into 3-phosphoadenosine 5-phosphate (PAP) by HPLC.[32]

These four variants exhibited activity improvements compared to wild-type SULT1A1 for each of the substrates tested: pNP, ≤4.3-fold; 3CyC, ≤19.3-fold; BPA, ≤9.4-fold; and L-thyroxine, ≤10.5-fold

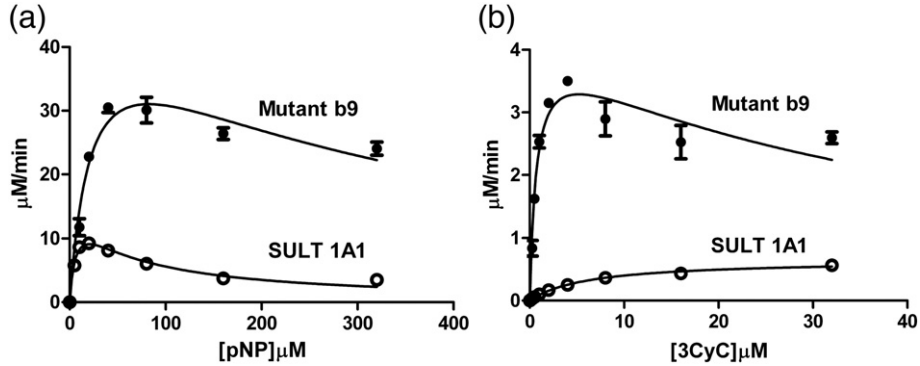

Fig. 4. Kinetic characterization of the SULT1A1-b9 mutant compared to wild-type SULT1A1. Michaelis–Menten curve for wild-type SULT1A1 (full circles) and its b9 (empty circles), with pNP (a) (left; $[E]_0 = 2.86$ μM) and 3CyC (b) (right; $[E]_0 = 1.43$ μM) fitted to the substrate inhibition model (GraphPad Prism 5). Error bars were derived from two independent measurements.

Table 7. Kinetic characterization of the SULT1A1-b9 mutant compared to wild-type SULT1A1

Kinetic parameters		pNP		3CyC	
		Wild-type 1A1	Mutant b9	Wild-type 1A1	Mutant b9
Fitted parameters	k_{cat} (min^{-1})	5.36 ± 1	17.2 ± 2.4	0.45 ± 0.03	2.94 ± 0.3
	K_m (μM)	7 ± 2	24 ± 7	6.2 ± 0.9	0.72 ± 0.2
	K_i (μM)	60. ± 14	283 ± 100	—	38 ± 12
Apparent parameters	V_{max} (μM/min)	9.25 ± 0.5	30.5 ± 1.1	0.55 ± 0.1	3.5 ± 0.02
	Activity units (nmol/min/mg)	92.5 ± 5	305 ± 11	11 ± 2	70 ± 0.4

The kinetic parameters are derived from the plots in Fig. 4.

(Table 6). These activities are also higher than the activities exhibited by 1E1 and any other SULT family, as indicated by published data.[32] The mutations that appeared in these variants included P10Q, S44A, M77N, A86I, S168C, V243I, and F247I (Table 6). Several of these mutations have been reported to be involved in substrate specificity (e.g., A86I[37] and F247I[32,38]).

As is the case with PON ancestral libraries, the active variants carried a high number of mutations, up to six ancestral mutations in variant b9 or b2 (Table 6). This high number of accommodated mutations is partly due to the parallel incorporation of compensatory consensus mutations (1.8 ± 0.67 per variant, on average).

Kinetic and structural characterizations of SULT1A1-b9

The kinetic parameters of the selected variant b9 with both pNP and 3CyC were determined and compared to those of human SULT1A1 (Fig. 4 and Table 7). SULTs, particularly human 1A1, are known to exhibit substrate inhibition. The data were

Table 8. Data collection and refinement statistics

	Human SULT1A1 variant b9+PAP+pNP	Human SULT1A1 variant b9+PAP+3CyC
Data collection		
PDB accession number	3QVU	3QVV
Number of images	125	135
Oscillation step (°)	1	1
Wavelength (Å)	1.54178	1.04628
Resolution (Å)	2.5	2.35
Space group	$P2_12_12_1$	$P2_12_12_1$
Unit cell parameters *a*, *b*, *c* (Å)	66.86, 71.54, 122.77	71.79, 71.91, 122.59
Number of observed reflections (last bin)	101,999 (11,010)	145,889 (24,812)
Number of unique reflections (last bin)	20,787 (2277)	26,761 (4512)
Completeness (last bin) (%)	98.7 (99.8)	98.4 (99.5)
R_{merge}[a] (last bin) (%)	12.6 (45.5)	7.1 (52.7)
$R_{measured}$[b] (last bin) (%)	14.0 (50.9)	7.8 (58.0)
$I/\sigma(I)$ (last bin)	10.62 (3.22)	13.36 (3.06)
Last-resolution shell	2.60–2.50	2.50–2.35
Redundancy (last bin)	4.91 (4.84)	5.45 (5.49)
Refinement statistics		
Resolution range (Å)	48.85–2.50	46.93–2.35
Number of reflections used in refinement	19,746	25,385
R_{free}/R_{work}[c]	24.7/20.8	27.6/23.0
RMSD from ideal		
Bond length (Å)	0.003	0.005
Bond angle (°)	0.587	0.947
Ramachandran plot		
Preferred regions (%)	97.1	94.2
Allowed regions (%)	2.6	5.1
Outliers (%)	0.3	0.7

[a] $R_{sym} = R_{merge} = \sum_h |\hat{I}_h - I_{h,i}| / \sum_h \sum_i^{n_h} I_{h,i}$.

[b] $R_{meas} = \sum_h \sqrt{\frac{n_h}{n_h - 1}} \sum_i^{n_h} |\hat{I}_h - I_{h,i}| / \sum_h \sum_i^{n_h} I_{h,i}$ with $\hat{I}_h = \frac{1}{n_h} \sum_i^{n_h} I_{h,i}$.

[c] $R_{work} = \sum ||F_0| - |F_c|| = \sum |F_0|$, where F_0 denotes the observed structure factor amplitude and F_c denotes the structure factor amplitude calculated from the model. R_{free} is as for R_{work} but calculated with 5% of randomly chosen reflections omitted from the refinement.

accordingly fitted to the simplest model of substrate inhibition.[39] The turnover number (k_{cat}) of the b9 mutant was higher than that of wild-type 1A1, ~3-fold higher for pNP, and 6.5-fold for 3CyC. The K_m value for pNP increased (3-fold), resulting in k_{cat}/K_m values similar to those of wild type. In contrast, the K_m value for 3CyC decreased 8.5-fold such that the k_{cat}/K_m of variant b3 with CyC increased by >50-fold relative to wild type. Substrate inhibition by pNP also decreased (the apparent K_i for wild-type 1A1 is 60 μM compared to 283 μM for the b9 mutant). With 3CyC as substrate, no substrate inhibition was observed for the wild type, whereas substrate inhibition could be seen for the b9 mutant (K_i = 37.7 μM). However, the K_i/K_m ratio was relatively high (>50 as opposed to ~10 with pNP).

To unravel the effect of the ancestral mutations, we solved the crystal structure of the b9 mutant that carried the P10Q, M77N, E151D, S168C, V243I, and F247I mutations at the background of wild-type human 1A1. We obtained structures in complex with pNP and PAP at 2.5 Å resolution (PDB ID: 3QVU), and those in complex with 3CyC and PAP at 2.35 Å resolution (PDB ID: 3QVV; Table 8).

The structures of wild-type human SULT1A1[35] and its b9 mutant are highly similar (RMSD of ~0.33 Å for 287 carbon α atoms). The binding of the acceptor pNP is also similar to that in the complexes of wild type in b9. However, significant changes could be observed in active-site volume and configuration. The acceptor's site of the b9 variant is considerably enlarged compared to wild-type SULT1A1, primarily due to the F247I mutation (Fig. 5a and b). This residue has also been identified as a key position for accommodating the binding of acceptor molecules (Amir Aharoni *et al.*,

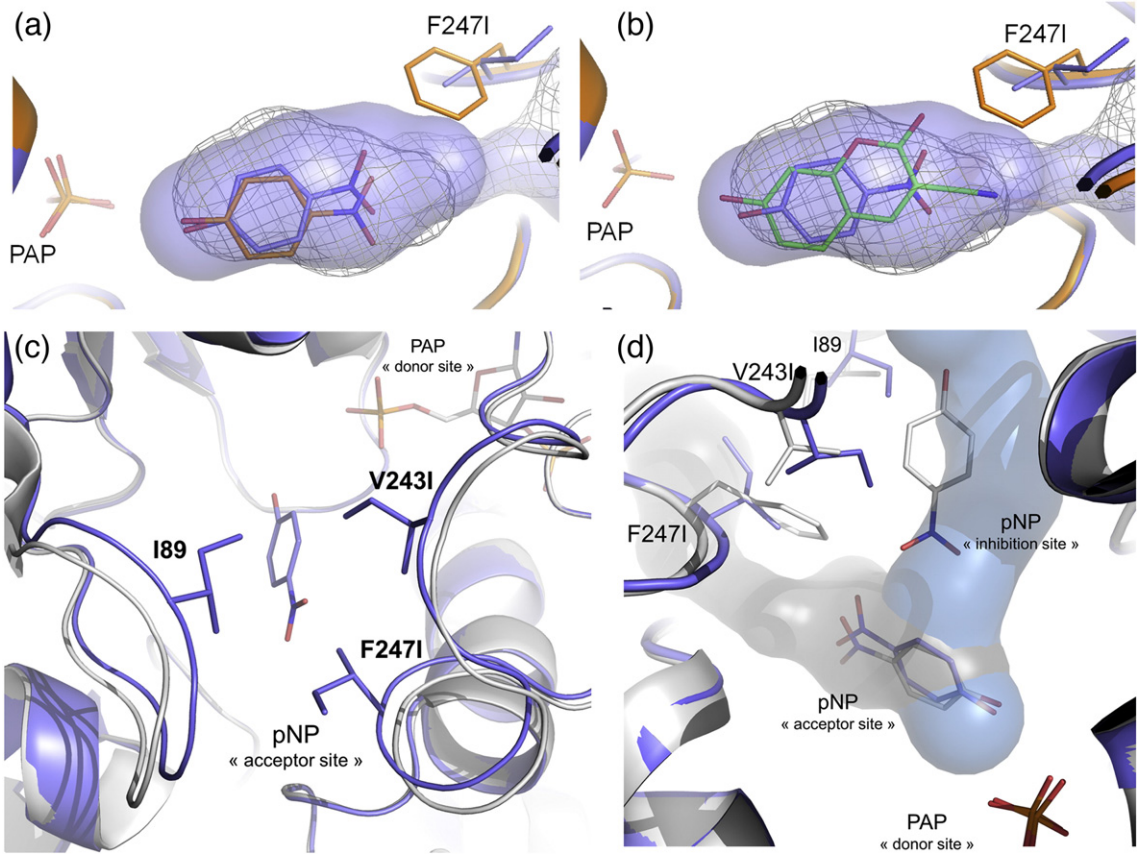

Fig. 5. Structural changes in the active site of the evolved b9 variant. (a) Superposition of the acceptor's binding pocket of the b9 variant (blue) onto that of the human 1A1 structure (PDB ID: 1LS6; gray mesh). The acceptor pockets of both wild type and b9 can readily accommodate pNP, although the wild type's pocket shows a better fit (and a 3-fold lower K_m for pNP). (b) The wild type's pocket is too small to accommodate the larger substrate 3CyC (green), whereas in b9, the F247I mutation (sticks) enlarged this pocket to accommodate 3CyC and even larger substrates. (c) The b9 mutant structure in complex with PAP and pNP (blue) superposed onto the wild-type SULT1A1 structure (gray). The incorporated ancestral mutations created a hydrophobic cluster of three isoleucines (I89, I243, and I247) that seem to have induced the closure of active-site loops by up to 1.5 Å. (d) The reshaping of the secondary acceptor's tunnel. The secondary tunnel of b9 (cyan) is displaced relative to wild type and is also narrower such that the inhibitory pNP molecule seen in human 1A1 (white sticks) cannot be accommodated.

submitted for publication). This enlarged cavity facilitates the binding of 3CyC, a bigger substrate than pNP, and accounts for the observed improvements in activity for 3CyC and with the other large substrates, L-thyroxine and BPA. Within the resolution limits, the structures of the b9 mutant in complex with 3CyC and pNP, including the reacting hydroxyl groups that coalign, appear essentially identical (Supplementary Figs. 8 and 9).

We also observed changes with respect to the secondary tunnel; this cavity, previously seen in human SULT1A1, binds a second pNP molecule in a nonreactive mode and is likely to be responsible for the observed substrate inhibition.[38] The alteration of this tunnel was mediated by two mutations (V243I and F247I). It appears that F247I, by being in van der Waals contact with I89, stabilized a new rotamer of I89. The resulting hydrophobic cluster (V243I, F247I, and I89) provokes the closing of the active loops and the narrowing of the secondary tunnel (Fig. 5c and d). In accordance, substrate inhibition in b9 is ~3-fold weaker than that in wild-type 1A1 (Fig. 4 and Table 7). Other mutations, such as M77N and S168C, are in close vicinity to the active site and might also contribute to the observed change in specificity. However, their role does not seem to be reflected in the crystal structures.

Discussion

We describe the construction of ancestral libraries as a tool for facilitating directed enzyme evolution. We exemplified this approach with mammalian PONs and SULTs. Both these enzyme families are of potential interest in drug metabolism and detoxification of xenobiotics. Although screened by low-throughput means, the ancestral PON and SULT libraries show a remarkably high frequency of polymorphic and functionally diverse variants. Screening of as few as 200–300 variants enabled the isolation of dozens of active variants with a range of different specificities and higher activities, including those with substrates with which all known family members exhibit low activity.

Ancestral libraries comprise a means of focusing diversity to positions that can readily promote changes in substrate specificity. The number of explored positions is relatively high (in the range of 20 positions or more). However, mutational diversity is highly focused on a given ancestral node or on a few nodes that could be explored in separate libraries (as here, with the N6-PON and N8-PON libraries) or combined into one library. This limited diversity is nonetheless wide enough to promote large functional variability and thereby enables the isolation of 'first-generation' variants as starting points for further directed evolution by standard methods.

Most of the incorporated mutations can be found in various family members, in particular in the case of the SULT library (Table 5). Only in certain libraries (e.g., the N6-PON library) unique substitutions not found in any known family member comprise about a third of the library (Table 1). Thus, a similar outcome could, in principle, be achieved through shuffling of various family members. However, the diversity of substitutions that appear in these positions within contemporary family members is vast (e.g., Table 5) and could not be explored within reasonable library sizes. The choice of ancestral substitution limits library sizes while maintaining sufficient functional diversity. Furthermore, family shuffling is limited to closely related sequences (≥70% amino acid identity).[7] Moreover, nearly all paralogues (i.e., members of families exhibiting different functional specificities) show sequence identity lower than 70%. Indeed, family shuffling would not be applicable for the SULT1 and PON families explored here, as their most diverse members show only 42% and 55% amino acid identities, respectively. Methods that overcome the identity barrier have been described (e.g., by shuffling fragments at predefined positions).[7] An alternative approach is to have a library based on amino acids that appear in structurally equivalent active-site positions. For example, by screening ~60,000 library variants derived from a *Pseudomonas fluorescens* esterase, variants that enantioselectively hydrolyze 3-phenylbutyric acid esters (up to 240-fold faster than wild type) were identified.[3] However, the choice of diversified library positions was largely dictated by early reports that explored various relevant positions of this enzyme. For many enzymes, however, knowledge of specificity-altering positions is limited or nonexistent. Furthermore, as indicated in Tables 1 and 5, the variability within all family members is huge; thus, family shuffling and the related methods may result in too large and insufficiently viable diversity. In the ancestral libraries, the choice of which mutations from existing family members would actually be explored is given by the ancestral sequence. In addition, a significant number of mutations and mutational compositions that are not seen in extant members are also explored.

The high viability of the ancestral libraries is particularly notable in light of the high frequency of mutations. The number of mutations in active clones from the ancestral libraries (>4 ancestral substitutions, on average, in the N8-PON and N6-PON variants, and >7 in the SULT variants) is much larger than in active variants isolated from random libraries (a single mutation in the random library screened here). This despite the average mutational rate in the unselected libraries being very similar. Indeed, the likelihood of finding a combination of four or more beneficial and/or neutral mutations in a randomly mutated library is very low.[1,40] In contrast, the ancestral libraries can easily accommodate high

mutational loads. In particular, it is clear that the accumulation of several active-site mutations, while maintaining configurational stability and enzymatic activity, is normally a laborious and lengthy process that demands several consecutive rounds of random mutagenesis and screening of a large number of library variants. In the ancestral libraries described here, variants with up to 11 mutations within and near the active site were readily identified by screening as few as 200 variants.

Other approaches that enable the incorporation of a large number of mutations while maintaining a relatively high fraction of viable variants are consensus and neutral drift libraries. Consensus libraries are composed of targeted mutations that bring the sequence of the target gene closer to the family consensus. The consensus sequence may overlap the sequence of the family ancestor (but not vice versa) or may differ. In certain positions, the ancestor and consensus sequences overlap (e.g., position 55; Table 1, first line). However, in many positions, the ancestral sequence does not match the consensus or differs altogether from the existing diversity (e.g., position 56 in the N6-PON ancestor; Table 1, second line). Furthermore, consensus mutations usually residing away from the active site exhibit mostly stabilizing effects.[41] Indeed, the main aim of consensus libraries is to obtain variants with higher stability, and not to alter the enzyme's specificity.[14] Consensus ancestor mutations were also included in our SULT library, and the recombinant PON1 variant used as starting point for the PON libraries carries several such mutations.[42] Neutral drift libraries are another related method for directed evolution. Comparison of ancestral PON libraries to previously generated neutral libraries of rePON1[1] reveals a similar frequency of improved variants and a similar magnitude of improvements (in particular for the high mutational load neutral drift library; Supplementary Fig. 10). The generation of neutral drift libraries is more laborious but requires no prior knowledge of the enzyme's structure or phylogeny. Thus, these approaches (ancestral and neutral drift) may apply to different target enzymes.

As this work progressed, Benner and Gaucher *et al.*[4] also described the application of ancestral inference to enzyme engineering. Their method called REAP was applied to engineer polymerase variants that accept nucleoside triphosphates with 3′-ONH_2 for DNA sequencing.[4] REAP was implemented to identify patterns of change in the evolutionary history of polymerases indicated by deviations from simple Markov models.[31,43,44] Because viral polymerases are likely to accept modified nucleotides, amino acid replacements that might have been responsible for the functional divergence between viral and nonviral polymerases were identified. The analysis thereby assisted the identification of candidate amino acid exchanges that could give rise to polymerase variants that accept modified nucleosides without the loss of catalytic activity. Although the general outlines for library design by REAP have been partially provided,[45] a more detailed and well-established protocol might be needed for a wider implementation. In addition, although REAP resembles our approach in the use of ancestral inference, these approaches differ in several important ways. Including positions that play a role in mediating a specific trait that appears in one branch (e.g., the nucleoside triphosphate promiscuity of viral polymerases) can help focus library diversity. However, this approach is limited to a specific functional property that existed in some form or another along the evolutionary history of the targeted protein family. We included the entire diversity found in various ancestral nodes throughout phylogeny. Our approach does not require knowledge on the specificity of the different members of the targeted protein family. Instead, it explores a large fraction of the phylogenetic sequence space by combinatorial libraries and thus obtains activities and specificities that did not exist in any of the ancestors, let alone in contemporary members of the targeted family.

In summary, the ancestral library approach described here complements other approaches to targeted mutagenesis such as family shuffling,[46] SCHEMA,[47] REAP,[4] neutral drift,[1] and simultaneous multiple-site saturation mutagenesis.[3] It has, however, its unique features and provides a new and effective way of performing protein engineering. We also describe a range of new PON and SULT variants that comprise valuable starting points for engineering, detoxifying, and decontaminating enzymes for organophosphates and other environmental hazards such as bisphenol A, or for potential therapeutic uses (e.g., prevention of organophosphate toxicity, treatment of hyperthyroidism).

Materials and Methods

Ancestral reconstruction

Sequences of PON family members were collected from the National Center for Biotechnology Information nonredundant protein sequence database (nr) using protein alignment BLAST (blastp) with its default settings. Only fully sequenced genes and those that contain all PONs' active site and other essential residues were included.[29] Redundant sequences exhibiting $\geq$90% identity to other sequences were removed using Cd-hit,[48] and the alignment was created with PRANK[49] using the F+ option and the LG substitution matrix.[50] The alignment was fine-tuned manually to improve its reliability at gap positions (Supplementary Fig. 1) and then used to create a phylogenetic tree using the Phyml program.[51] Bootstrapping analysis was performed to evaluate the tree. Ancestral predictions for the various nodes were obtained with FastML,[52] and the most probable ancestral sequence was used for library design. The SULT ancestral

predictions were performed as described above, with some modifications. Firstly, a 95% identity threshold was applied. Using a threshold of 90% resulted in the loss of all but two members of the 4A1 family, as well as the loss of family members with known structures such as human 1A2 and 1A3. Secondly, because the structures of most SULT family members are available (as opposed to only PON1), the alignment was structure-based using T-Coffee Expresso.[53] The alignment after the manual refinement of the gap positions (Supplementary Fig. 6) was used for constructing the SULT tree (Supplementary Fig. 7) and ancestral predictions, as described above.

Library preparation

We applied the ISOR (*I*ncorporating *S*ynthetic *O*ligonucleotides via Gene *R*eassembly) protocol for generating DNA libraries of random combinations of specific mutations.[17] The starting point for the PON library was a recombinant PON1 variant, rePON1-G3C9, that expresses function in *E. coli*, exhibits ~94% identity to rabbit PON1, and exhibits the same enzymatic specificity as rabbit and human PON1.[20] For each ancestral mutation, a mutagenesis oligonucleotide ~30 bp long (~15 bp from each side of the mutation) was synthesized (Supplementary Tables 2 and 4). Typically, 2 pmol of an equimolar mixture of all mutagenesis oligonucleotides and 50 ng of DNase I (Takara)-generated fragments of the starting gene (rePON1-G3C9; GenBank AY499193) were assembled as described previously,[17] followed by a nested PCR with external primers. The control library of random mutations was generated by error-prone PCR (GeneMorph® II Random Mutagenesis Kit; Stratagene, Agilent Technologies, USA) of rePON1-G3C9, and the mutation load was tuned to a similar average of ~4 mutations per gene. The various PON libraries (N6, N8, and Mu libraries) were separately ligated into a modified pET-32 expression plasmid that introduces a His-tag at PON1's C-terminus.[19] The ligated plasmids were transformed into *E. coli* BL21-DE3 cells. The transformed cells were plated on agar, and single colonies were isolated and grown on 96-deep-well plates for protein expression and activity screen.[1] The SULT library was similarly generated. As a starting point, we used two stabilized human SULT1A1 variants, dubbed hSULT1A1-1C5 and hSULT1A1-1E9, each carrying several consensus mutations (listed in Table 6). These were kindly provided by Amir Aharoni (submitted for publication; Department of Life Sciences, Ben-Gurion University of the Negev, Israel). Although their enzymatic specificities deviate from wild-type human 1A1, these deviations are much smaller than those seen in the ancestral library variants (Table 6).

Library screens

From the N8-PON library, 276 randomly picked clones were grown on 96-well plates, and crude bacterial lysates were screened with the target substrates (Supplementary Fig. 3), as described previously.[1] The lysate volume taken with each substrate varied from 5 μ to 20 μl such that initial rates could be reliably measured. The rates were corrected for background nonenzymatic rates and compared to the activity of the human-like rePON1-G3C9 starting point. From the N6-PON library, 276 randomly picked clones were similarly screened. For the SULT library screen, we used 10 μl of crude lysate in a total reaction volume of 100 μl for screening with the colorimetric pNP (50 μM; $[PAPS]_0 = 0.5$ mM, absorption measured at 405 nm) or with the fluorogenic 3CyC (1 μM; $[PAPS]_0 = 100$ μM, excitation at 408 nm and emission at 450 nm). The ratio of activity between these two substrates relative to the ratio exhibited by the stabilized human SULT1A1 variants (1A1-1C5 and 1A1-1E9) was used to identify changes in specificity. Four variants showing the highest changes in specificity were purified and assayed further.

Purification of enzyme variants and enzymatic assays

The SULT variants were purified by Ni-NTA. The purified enzymes were assayed with pNP and 3CyC as described above, and by HPLC essentially as described previously.[32] Briefly, the 100-μl reactions contained 2.9 μM purified enzyme and BPA (100 μM) and PAPS (0.5 mM), or L-thyroxine (25 μM) and PAPS (100 μM). The reactions were stopped at different time points (5 min, 10 min, and 20 min) by adding acetonitrile to 30% vol/vol (final concentration) and immediately freezing in liquid nitrogen. Samples were diluted 3-fold with buffer A (see the text below), centrifuged to remove the protein precipitate, and injected into HPLC. HPLC was performed with AKTA-Basic using an ion-exchange column (Sepax-SAX-NP5) and a gradient of buffer A (10 mM Tris, pH 7) and buffer B (buffer A plus 0.25 M NaCl)—from 0% buffer A to 50% buffer B in 10 min—and at a flow rate of 1 ml/min. The conversion of PAPS into PAP was monitored at 259 nm (retention times of 9.5 min and 7.4 min, respectively). The concentration of the produced PAP was plotted against time, and the initial rates were extrapolated. Kinetic assays with pNP and 3CyC were performed using a microtiter plate reader and 0.1-ml reactions in 10 mM phosphate buffer (pH 7.0) while monitoring optical density or florescence, respectively. The data were fitted to the equation: $v_0 = k_{cat}[E]_0[S]_0/(K_m + [S]_0(1 + [S]_0/K_i))$, whereby K_i is the apparent dissociation constant for substrate binding in the inhibitory mode.[39]

Crystallization

Freshly purified SULT1A1-b9 was concentrated to 33 mg/ml. One microliter of the donor's product PAP (100 mM in phosphate buffer, pH 7.0) and 1 μl of the acceptor pNP (200 mM in 100% dimethyl sulfoxide) were added to 30 μl of protein solution. Crystallization was performed using the hanging-drop vapor diffusion method. Equal volumes (1 μl) of protein and reservoir solutions were mixed, and the resulting drop was allowed to reach equilibrium with a 500-μl reservoir solution made of 18–22% (wt/vol) polyethylene glycol (PEG) 3350 and 50 mM Tris–HCl buffer (pH 8). Crystals appeared after 2 days at 293 K. For exchange of the acceptor molecule, crystals were rinsed for 5 min in a drop of a solution composed of the mother liquor

solution, 25% PEG 600, and 5 mM 3CyC, and then soaked for 1 h in a fresh drop of the same solution.

Data collection and structure refinement

Crystals were transferred into a cryoprotectant solution containing the reservoir solution and 25% PEG 600 for 1 min, and then flash-cooled in liquid nitrogen. X-ray diffraction data for the crystal of mutant b9 in complex with pNP and PAP (2.5 Å resolution) were collected at 100 K on a Rigaku R-AXIS IV++ imaging plate area detector mounted on a Rigaku RU-H3R generator with CuKα radiation focused by Osmic confocal mirrors. A second data set collected using a mutant b9 crystal soaked with 3CyC and PAP at 2.35 Å resolution was collected at 100 K with synchrotron radiation at ID23-1 beamline (European Synchrotron Radiation Facility, Grenoble, France) using an ADSC Quantum Q315r detector. The data were integrated and scaled using XDS and XSCALE.[54] Both crystals belonged to the same space group yet exhibited significantly different unit cell dimensions (Table 8). Molecular replacement was performed using MOLREP,[55] with the structure of human SULT1A1 (PDB ID: 2D06) used as starting model. Manual model improvement was performed using Coot.[55] Refinement of the models was performed using REFMAC5.[55] TLS refinement was applied to the 3CyC-containing structure. Figures were prepared with PyMOL (DeLano Scientific LLC), and figures displaying cavities were produced with Caver 2.1.1†.

Accession numbers

Coordinates and structure factors of the b9 mutant, in complex with PAP and pNP and in complex with PAP and 3CyC, have been deposited in the PDB under accession numbers 3QVU and 3QVV, respectively.

Supplementary materials related to this article can be found online at doi:10.1016/j.jmb.2011.06.037

Acknowledgements

We are very grateful to Hagit Bar and Tal Pupko for their assistance in performing phylogenetic analysis, to Amir Aharoni and Dotan Amar for providing the stabilized SULT1A1 variants and for assisting in SULT enzymatic assays, and to Daniel Tal for assisting in protein production and HPLC assays. Financial support by the National Institutes of Health (W81XWH-07-2-0020) and the Defense Threat Reduction Agency (HDTRA 1-07-C-0024) is gratefully acknowledged. M.E. is a fellow supported by the IEF Marie Curie Program (grant 252836). D.S.T. is the Nella and Leon Benoziyo Professor of Biochemistry.

† http://www.caver.cz

References

1. Gupta, R. D. & Tawfik, D. S. (2008). Directed enzyme evolution via small and effective neutral drift libraries. *Nat. Methods*, **5**, 939–942.
2. Reetz, M. T. & Carballeira, J. D. (2007). Iterative saturation mutagenesis (ISM) for rapid directed evolution of functional enzymes. *Nat. Protoc.* **2**, 891–903.
3. Helge, J. & Bornscheuer, U. T. (2010). Natural diversity to guide focused directed evolution. *ChemBioChem*, **11**, 1861–1866.
4. Chen, F., Gaucher, E. A., Leal, N. A., Hutter, D., Havemann, S. A., Govindarajan, S. *et al.* (2010). Reconstructed Evolutionary Adaptive Paths give polymerases accepting reversible terminators for sequencing and SNP detection. *Proc. Natl Acad. Sci. USA*, **107**, 1948–1953.
5. Lutz, S. (2010). Beyond directed evolution—semi-rational protein engineering and design. *Curr. Opin. Biotechnol.* **21**, 734–743.
6. Carbone, M. N. & Arnold, F. H. (2007). Engineering by homologous recombination: exploring sequence and function within a conserved fold. *Curr. Opin. Struct. Biol.* **17**, 454–459.
7. Sieber, V., Martinez, C. A. & Arnold, F. H. (2001). Libraries of hybrid proteins from distantly related sequences. *Nat. Biotechnol.* **19**, 456–460.
8. Wickens, M. & Cox, M. M. (2009). Critical reviews in biochemistry and molecular biology. Introduction. *Crit. Rev. Biochem. Mol. Biol.* **44**, 2.
9. Bershtein, S. & Tawfik, D. S. (2008). Ohno's model revisited: measuring the frequency of potentially adaptive mutations under various mutational drifts. *Mol. Biol. Evol.* **25**, 2311–2318.
10. Soskine, M. & Tawfik, D. S. (2010). Mutational effects and the evolution of new protein functions. *Nat. Rev. Genet.* **11**, 572–582.
11. Raillard, S., Krebber, A., Chen, Y., Ness, J. E., Bermudez, E., Trinidad, R. *et al.* (2001). Novel enzyme activities and functional plasticity revealed by recombining highly homologous enzymes. *Chem. Biol.* **8**, 891–898.
12. Amitai, G., Gupta, R. D. & Tawfik, D. S. (2007). Latent evolutionary potentials under the neutral mutational drift of an enzyme. *HFSP J.* **1**, 67–78.
13. Bloom, J. D., Romero, P. A., Lu, Z. & Arnold, F. H. (2007). Neutral genetic drift can alter promiscuous protein functions, potentially aiding functional evolution. *Biol. Direct*, **2**, 17.
14. Jochens, H., Aerts, D. & Bornscheuer, U. T. (2010). Thermostabilization of an esterase by alignment-guided focussed directed evolution. *Protein Eng. Des. Sel.* **23**, 903–909.
15. Flores, H. & Ellington, A. D. (2005). A modified consensus approach to mutagenesis inverts the cofactor specificity of *Bacillus stearothermophilus* lactate dehydrogenase. *Protein Eng. Des. Sel.* **18**, 369–377.
16. Thornton, J. W. (2004). Resurrecting ancient genes: experimental analysis of extinct molecules. *Nat. Rev. Genet.* **5**, 366–375.
17. Herman, A. & Tawfik, D. S. (2007). Incorporating Synthetic Oligonucleotides via Gene Reassembly (ISOR): a versatile tool for generating targeted libraries. *Protein Eng. Des. Sel.* **20**, 219–226.

18. Aharoni, A., Amitai, G., Bernath, K., Magdassi, S. & Tawfik, D. S. (2005). High-throughput screening of enzyme libraries: thiolactonases evolved by fluorescence-activated sorting of single cells in emulsion compartments. *Chem. Biol.* **12**, 1281–1289.
19. Gupta, R. D., Goldsmith, M., Ashani, Y., Simo, Y., Mullokandov, G., Bar, H. *et al.* (2011). Directed evolution of hydrolases for prevention of G-type nerve agent intoxication. *Nat. Chem. Biol.* **7**, 120–125.
20. Aharoni, A., Gaidukov, L., Yagur, S., Toker, L., Silman, I. & Tawfik, D. S. (2004). Directed evolution of mammalian paraoxonases PON1 and PON3 for bacterial expression and catalytic specialization. *Proc. Natl Acad. Sci. USA*, **101**, 482–487.
21. Draganov, D. I. & La Du, B. N. (2004). Pharmacogenetics of paraoxonases: a brief review. *Naunyn-Schmiedeberg's Arch. Pharmacol.* **369**, 78–88.
22. Khersonsky, O. & Tawfik, D. S. (2005). Structure–reactivity studies of serum paraoxonase PON1 suggest that its native activity is lactonase. *Biochemistry*, **44**, 6371–6382.
23. Draganov, D. I., Teiber, J. F., Speelman, A., Osawa, Y., Sunahara, R. & La Du, B. N. (2005). Human paraoxonases (PON1, PON2, and PON3) are lactonases with overlapping and distinct substrate specificities. *J. Lipid Res.* **46**, 1239–1247.
24. Teiber, J. F. & Draganov, D. I. (2011). High-performance liquid chromatography analysis of *N*-acyl homoserine lactone hydrolysis by paraoxonases. *Methods Mol. Biol.* **692**, 291–298.
25. Gaidukov, L., Bar, D., Yacobson, S., Naftali, E., Kaufman, O., Tabakman, R. *et al.* (2009). *In vivo* administration of BL-3050: highly stable engineered PON1–HDL complexes. *BMC Clin. Pharmacol.* **9**, 18.
26. Bloom, J. D., Labthavikul, S. T., Otey, C. R. & Arnold, F. H. (2006). Protein stability promotes evolvability. *Proc. Natl Acad. Sci. USA*, **103**, 5869–5874.
27. Khersonsky, O. & Tawfik, D. S. (2006). Chromogenic and fluorogenic assays for the lactonase activity of serum paraoxonases. *ChemBioChem*, **7**, 49–53.
28. Aharoni, A., Gaidukov, L., Khersonsky, O., McQ Gould, S., Roodveldt, C. & Tawfik, D. S. (2005). The 'evolvability' of promiscuous protein functions. *Nat. Genet.* **37**, 73–76.
29. Harel, M., Aharoni, A., Gaidukov, L., Brumshtein, B., Khersonsky, O., Meged, R. *et al.* (2004). Structure and evolution of the serum paraoxonase family of detoxifying and anti-atherosclerotic enzymes. *Nat. Struct. Mol. Biol.* **11**, 412–419.
30. Bershtein, S., Goldin, K. & Tawfik, D. S. (2008). Intense neutral drifts yield robust and evolvable consensus proteins. *J. Mol. Biol.* **379**, 1029–1044.
31. Gaucher, E. A., Thomson, J. M., Burgan, M. F. & Benner, S. A. (2003). Inferring the palaeoenvironment of ancient bacteria on the basis of resurrected proteins. *Nature*, **425**, 285–288.
32. Allali-Hassani, A., Pan, P. W., Dombrovski, L., Najmanovich, R., Tempel, W., Dong, A. *et al.* (2007). Structural and chemical profiling of the human cytosolic sulfotransferases. *PLoS Biol.* **5**, e97.
33. Nishiyama, T., Ogura, K., Nakano, H., Kaku, T., Takahashi, E., Ohkubo, Y. *et al.* (2002). Sulfation of environmental estrogens by cytosolic human sulfotransferases. *Drug Metab. Pharmacokinet.* **17**, 221–228.
34. Ozawa, S., Shimizu, M., Katoh, T., Miyajima, A., Ohno, Y., Matsumoto, Y. *et al.* (1999). Sulfating-activity and stability of cDNA-expressed allozymes of human phenol sulfotransferase, ST1A3∗1 ((213)Arg) and ST1A3∗2 ((213)His), both of which exist in Japanese as well as Caucasians. *J. Biochem.* **126**, 271–277.
35. Gamage, N. U., Duggleby, R. G., Barnett, A. C., Tresillian, M., Latham, C. F., Liyou, N. E. *et al.* (2003). Structure of a human carcinogen-converting enzyme, SULT1A1. Structural and kinetic implications of substrate inhibition. *J. Biol. Chem.* **278**, 7655–7662.
36. Lehmann, M. & Wyss, M. (2001). Engineering proteins for thermostability: the use of sequence alignments *versus* rational design and directed evolution. *Curr. Opin. Biotechnol.* **12**, 371–375.
37. Liu, M. C., Suiko, M. & Sakakibara, Y. (2000). Mutational analysis of the substrate binding/catalytic domains of human M form and P form phenol sulfotransferases. *J. Biol. Chem.* **275**, 13460–13464.
38. Barnett, A. C., Tsvetanov, S., Gamage, N., Martin, J. L., Duggleby, R. G. & McManus, M. E. (2004). Active site mutations and substrate inhibition in human sulfotransferase 1A1 and 1A3. *J. Biol. Chem.* **279**, 18799–18805.
39. Copeland, R. A. (2000). *Enzymes*, 2nd edit. Wiley, New York, NY.
40. Bershtein, S., Segal, M., Bekerman, R., Tokuriki, N. & Tawfik, D. S. (2006). Robustness–epistasis link shapes the fitness landscape of a randomly drifting protein. *Nature*, **444**, 929–932.
41. Lehmann, M., Kostrewa, D., Wyss, M., Brugger, R., D'Arcy, A., Pasamontes, L. & van Loon, A. P. (2000). From DNA sequence to improved functionality: using protein sequence comparisons to rapidly design a thermostable consensus phytase. *Protein Eng.* **13**, 49–57.
42. Khersonsky, O., Rosenblat, M., Toker, L., Yacobson, S., Hugenmatter, A., Silman, I. *et al.* (2009). Directed evolution of serum paraoxonase PON3 by family shuffling and ancestor/consensus mutagenesis, and its biochemical characterization. *Biochemistry*, **48**, 6644–6654.
43. Gaucher, E. A. (2007). In *Ancestral Sequence Reconstruction* (Liberles, D. A., ed.), Oxford University Press, New York, NY.
44. Benner, S. A. (2003). Interpretive proteomics—finding biological meaning in genome and proteome databases. *Adv. Enzyme Regul.* **43**, 271–359.
45. Cole, M. F. & Gaucher, E. A. (2010). Exploiting models of molecular evolution to efficiently direct protein engineering. *J. Mol. Evol.* **72**, 193–203.
46. Stemmer, W. P. (1994). DNA shuffling by random fragmentation and reassembly: *in vitro* recombination for molecular evolution. *Proc. Natl Acad. Sci. USA*, **91**, 10747–10751.
47. Meyer, M. M., Silberg, J. J., Voigt, C. A., Endelman, J. B., Mayo, S. L., Wang, Z. G. & Arnold, F. H. (2003). Library analysis of SCHEMA-guided protein recombination. *Protein Sci.* **12**, 1686–1693.
48. Li, W. & Godzik, A. (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*, **22**, 1658–1659.

49. Loytynoja, A. & Goldman, N. (2008). Phylogeny-aware gap placement prevents errors in sequence alignment and evolutionary analysis. *Science*, **320**, 1632–1635.
50. Le, S. Q. & Gascuel, O. (2008). An improved general amino acid replacement matrix. *Mol. Biol. Evol.* **25**, 1307–1320.
51. Guindon, S. & Gascuel, O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* **52**, 696–704.
52. Pupko, T., Pe'er, I., Shamir, R. & Graur, D. (2000). A fast algorithm for joint reconstruction of ancestral amino acid sequences. *Mol. Biol. Evol.* **17**, 890–896.
53. Armougom, F., Moretti, S., Poirot, O., Audic, S., Dumas, P., Schaeli, B. *et al.* (2006). Expresso: automatic incorporation of structural information in multiple sequence alignments using 3D-Coffee. *Nucleic Acids Res.* **34**, W604–W608.
54. Kabsch, W. (1993). Automatic processing of rotation diffraction data from crystals of initially unknown symmetry land cell constants. *J. Appl. Crystallogr.* **26**, 795–800.
55. Collaborative Computational Project Number 4. (1994). The CCP4 Suite: programs for protein crystallography. *Acta Crystallogr., Sect. D.: Biol. Crystallogr.* **50**, 760–763.