HFSP Journal

ISSN: 1955-2068 (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/tfls19

Latent evolutionary potentials under the neutral mutational drift of an enzyme

Gil Amitai , Rinkoo Devi Gupta & Dan S. Tawfik

To cite this article: Gil Amitai , Rinkoo Devi Gupta & Dan S. Tawfik (2007) Latent evolutionary potentials under the neutral mutational drift of an enzyme, HFSP Journal, 1:1, 67-78, DOI: 10.2976/1.2739115/10.2976/1

To link to this article: https://doi.org/10.2976/1.2739115/10.2976/1

Latent evolutionary potentials under the neutral mutational drift of an enzyme

Gil Amitai,[1] Rinkoo Devi Gupta,[1] and Dan S. Tawfik[1]

[1]Department of Biological Chemistry, the Weizmann Institute of Science, Rehovot 76100, Israel

(Received 21 December 2006; accepted 20 April 2007; published online 21 May 2007)

Biological systems exhibit mutational robustness, or neutrality, whereby the impact of mutations is minimized. Does neutrality hamper their ability to adapt in the face of changing environments? We monitored changes in genotype and phenotype that occur within a neutral mutational network of an enzyme, experimentally and computationally (see accompanying article). Using the enzyme PON1 as a model, we performed random mutagenesis and purifying selection to purge deleterious mutations. We characterized ~300 variants that are apparently neutral, or close to neutral, with respect to PON1's levels of expression and native lactonase activity. Their activities with promiscuous substrates and ligands indicated significant changes in adaptive potentials. Almost half of the variants exhibited changes in promiscuous activities, specificities, or inhibition, and several of these were found to be one or two mutations, closer to potentially new phenotypes: aryl esterase, thiolactonase, phosphotriesterase, or drug resistance. This empirical measure of phenotypic changes under neutrality supports the notion that sequence changes that are neutral, i.e., non-adaptive, in a current context can facilitate adaptation under changing circumstances, by both expanding the activity range of existing enzymes and thus providing an immediate advantage, and by reducing the number of mutations required for divergence of new functions.
[DOI: 10.2976/1.2739115]

CORRESPONDENCE
Dan S. Tawfik: tawfik@weizmann.ac.il

Most mutations are potentially deleterious, and biological systems and molecules exhibit appreciable tolerance to their effects (mutational robustness, or neutrality). If the deleterious effects of mutations are generally suppressed, then, because most adaptive mutations (mutations conferring a new protein function and phenotype) undermine protein stability and folding (Beadle and Shoichet, 2002; Meiering *et al.*, 1992), they might only emerge once a gene has been relieved from the burden of carrying its existing function. Indeed, the prevailing view surmises that gene duplication is a prerequisite for the divergence of a new function, and that following duplication, mutations conferring a new function will typically occur at a frequency of about 10^{-9} per site (Graur and Li, 2000). If the evolution of a new function depends, for example, on the simultaneous substitution of two amino acids, the frequency would become 9 orders of magnitude lower, or 10^{-18} (Gressel and Levy, 2006). The rarity of adaptive events of this type has been recently highlighted (Behe and Snoke, 2004; Lynch, 2005). However, if mutations with adaptive potential accumulate prior to duplication as neutral, namely, with no or little effect on fitness under a current state, adaptive events may become more frequent. Thus, although neutrality can be seen as an opposing force to evolutionary innovation, adaptive evolutionary change may be expedited by mutations that accumulate under neutral, nonadaptive conditions (Kirschner and Gerhart, 1998; Nei, 2005; Wagner, 2005a).

Neutrality relates to numerous genotypes accommodating the very same phenotype (sequence redundancy) and neutral networks are induced by redundant mapping in which sets of genotype that produce the same phenotype are

connected by single point mutations (Bornberg-Bauer, 1997; Govindarajan and Goldstein, 1997; Huynen *et al.*, 1996; Lipman and Wilbur, 1991; Schuster *et al.*, 1994). Having the same phenotype might imply that all sequences in a neutral network have the same evolutionary potential. However, neutrality might be defined by certain aspects of a system, primarily to those traits under selection at a given time (Kimura, 1986). A change in other aspects might still be possible, thus increasing the potential for future adaptation (Wagner, 2005a; Wagner, 2005b). In other words, "by moving neutrally something does vary: the potential for change" (Schuster and Fontana, 1999).

The last decade has seen many attempts to validate the above hypothesis. Computational studies showed how neutral mutations help explore new structures, and thereby new phenotypes (Aita *et al.*, 2003; Fontana and Schuster, 1998; Huynen *et al.*, 1996; Schuster and Fontana, 1999). Numerous studies suggest that proteins may promiscuously interact with ligands and substrates other than those they evolved for (Aharoni *et al.*, 2005b; Bornscheuer and Kazlauskas, 2004; Copley, 2003; O'Brien and Herschlag, 1999). In doing so, proteins and RNA may adopt alternative conformations that differ from the conformation that mediates their primary function (James and Tawfik, 2003; Ma *et al.*, 2002; Schultes and Bartel, 2000). The existence of this latent pleiotropy—a range of promiscuous functions and conformational isomers that were neither selected for, nor against—may facilitate evolution by providing ample starting points for new functions while retaining the primary activity. Indeed, laboratory evolution experiments indicate that, under selection, such promiscuous functions dramatically increase (typically by 10–1000-fold, and sometimes by up to 10^6-fold), often without severely compromising the protein's native activity (Aharoni *et al.*, 2005a; Khersonsky and Tawfik, 2006a).

Despite the above-described findings there is no direct evidence, nor a biophysical model, that indicate whether and how latent promiscuous functions and conformations develop under neutral drift while preserving the protein's primary function. This paper aims at an experimental examination of latent activity changes within a large set of apparently neutral enzyme variants. The neutral set was generated by three rounds of random mutagenesis and selection to maintain the enzyme's native activity and expression levels. We then determined changes in activity and selectivity for various promiscuous substrates and sequence changes that occurred within this sample of neutral variants. The accompanying article provides a biophysical model indicating how alternative protein conformations, and hence functions, emerge under neutrality for the dominant or native function (Wroe *et al.*, 2007).

Our model protein is an enzyme called serum paraoxonase, or PON1. For several decades, PON1's primary activity remained unknown, but a large number of substrates, including aromatic lactones, esters, and phospotriesters (e.g., the pesticide paraoxon that endowed its name), were discovered that enable its detection in sera. Recently, it has been discovered that PON1 is a lactonase, with lipophylic lactones being its prime substrates, and all other activities being merely promiscuous (Draganov *et al.*, 2005; Khersonsky and Tawfik, 2005). Thus, PON1 represents an enzyme whose native and promiscuous functions are known and easily monitored. Previous works also demonstrated that these promiscuous activities could amply evolve (30–200-fold, in response to 1–4 mutations), typically at the expense of large decreases in other promiscuous functions (100–1000-fold), yet with much smaller effects (0–22-fold decreases) on the native lactonase function (Aharoni *et al.*, 2005b; Khersonsky *et al.*, 2006). These experiments therefore defined several putative evolutionary intermediates, or transition points, possibly leading from a lactonase (wild-type PON1) to a range of new enzymes that stem from PON1's promiscuous activities.

Here we examined PON1's potential to drift towards these evolutionary transition points under selection to retain its primary function. Our approach mimicked a random mutational drift up to seven amino acid substitutions from the wild type's sequence, and combined detailed and reproducible measurements of enzymatic activities and expression levels with sequence analysis. This approach enabled us to sample changes that occur within the neutral network of PON1. Although our data set is significant, it obviously represents a miniscule sample of the actual neutral network of PON1. The accompanying study complements this drawback, and describes the exhaustive computational mapping of protein sequences onto structures (Wroe *et al.*, 2007).

RESULTS

The experimental model

Our experimental model was based on *in vitro* measurements, and included the following assumptions: (1) PON1's native function and neutrality or apparent neutrality are defined by the levels of its primary enzymatic activity (lipolactonase) and expression. (2) The promiscuous activities (aryl esterase, phosphotriesterase, and aromatic lactonase) are latent (have no current physiological or evolutionary role), and represent potential evolutionary starting points. (3) The active-site inhibitor simulates a drug whose action is to inhibit PON1, and mutations that enable PON1 to function in the presence of high inhibitor concentrations simulate drug resistance. Because proteins may have more than one native or physiological function, and what seems to be neutral for one function might be associated with changes in other traits (Wagner, 2005a; Wagner, 2005b), we reinforced our model by including the wild type's expression level as an additional parameter of neutrality. The PON1 gene was therefore fused at its carboxy-terminus to green fluorescent protein (GFP) to give a construct named here "wild-type PON1," and GFP emission levels were monitored as a measure of the concentration of soluble, properly folded PON1 (Waldo *et al.*,

1999). By monitoring expression levels, our selection procedure purged any mutation that undermines the stability, folding, and solubility of the PON1 protein. Due to the complexity of living systems, one cannot assume that the above assumptions comply with physiologically relevant scenarios, certainly not in their entirety. Obviously, our *in vitro* measurements are only indicative of neutrality, but only fitness measurements in living organisms could ensure the actual neutrality of these mutations. Nevertheless, together they represent a consistent and reasonable experimental model for examining the neutrality model at the molecular level.

PON1's open reading frame was subjected to random mutagenesis *in vitro* (incorporating, on average, 1.8 mutations per gene or 1.7×10^{-3} mutations/bp), and a selection, using *in vitro* compartmentalization (IVC) in double emulsions and FACS (Aharoni *et al.*, 2005a), aimed at isolating apparently neutral PON1-GFP variants exhibiting wild-type-like lipo-lactonase activity and expression levels. Approximately 100 apparently neutral variants isolated from the first round of mutation and selection, were subjected to a second round of mutagenesis and selection, and this process was repeated twice to form a third round. Altogether, we characterized 311 variants that repetitively exhibited lactonase activity and GFP levels well above the background level (>3 standard deviation (SD) above background mean level; see Materials and Methods). We defined *neutral* variants as variants possessing lactonase activity and expression level within two standard deviations (≤2SD) of the wild-type mean (Fig. 1). As discussed above, these variants are neutral only to the context of our experimental system, and thus re referred to as apparently neutral. To address a scenario where the selective trait is associated primarily with enzymatic activity, we classified as *nearly neutral* another set of variants that possessed wild-type-like enzymatic activity, but deviated in expression levels. Finally, PON1 variants retaining activity outside the wild-type range yet significantly above mean background level (>3SD), might be still neutral in terms of organismal fitness. These were classified as *possibly neutral*, and simulate a drift to the perimeter of our putative neutral network.

Phenotypic characterization

The phenotypic characterization included a native-like lipophylic lactone substrate (5-thiobutyl butyrolactone; TBBL), GFP emission, and five different promiscuous substrates representing three different hydrolytic activities: paraoxon and DEPCyC (phosphotriesterase); 2NA and 7AcC (aryl esterase); and DHC (aromatic lactonase) (Aharoni *et al.*, 2005a). We also tested the lactonase activity in the presence of PON1's active-site inhibitor 2ODQ ($K_i\approx6\ \mu$M) (Khersonsky and Tawfik, 2005). [Substrates and reactions are detailed in Fig. S1 (see EPAPS material)].

We analyzed changes in evolutionary potentials according to four criteria as detailed below:

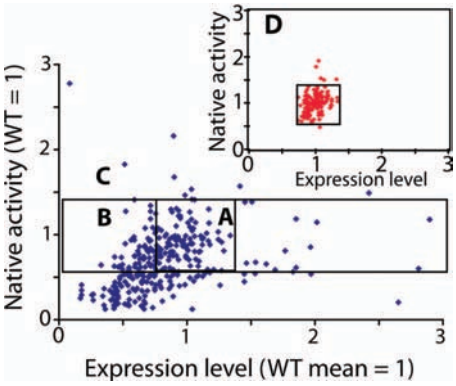

Figure 1. Defining the neutrality categories according to the native lactoanse activity and expression level. The lactonase activity (with TBBL as substrate), and expression level (GFP emission), were determined for each variant by averaging 2–8 independent measurements, and normalized to wild-type values. Variants were divided into three groups according to these categories: (A). *Neutral*—both lactonase activity and expression levels within 2SD of the wild type's values. (B). *Nearly neutral*—only lactonase activity within 2SD of the wild type's value. [Note that, the *neutral* variants [group (A)] comprise a subset of the *nearly neutral* variants group (B)]. (C) *Possibly neutral*—lactonase activity and expression levels that deviate from wild-type PON1, yet are well above background (≥3SD). (D) The lactonase activity and expression levels for multiple independent repeats of wild-type PON1 ($N=147$). Noted (black box) are the activity and expression levels that define the neutral set [category (A)].

Total activity. The first criterion by which we segregated variants is changes in the overall activity towards each promiscuous substrate (rate of product formation relative to wild-type PON1). Many variants (~75%, 232/311) exhibited a decrease in one or more promiscuous activity, but only 2.2% (7/311) of variants, belonging to the *nearly neutral* and *possibly neutral* sets, displayed more than a two-fold increase in at least one activity (Table I, Fig. 2).

Specific activity relates to the level of activity per protein molecule. Total activities (rate of product formation) were divided by expression levels (GFP emission), and normalized to the wild type's specific activity with the same substrate. About 9% (29/311) of variants showed improvements (2–36-fold) in the specific activity towards at least one promiscuous substrate. Overall, 6% of the *nearly neutral*, and 23% of the *possibly neutral* variants, displayed ≥ twofold increases in specific activity (Table I).

Selectivity concerns changes in the ratios of activities between different pairs of substrates. All activities were normalized relative to wild type, and the ratios for all possible pairs of substrates were calculated. Changes in selectivity were most frequent: ≥ fivefold selectivity changes were observed in 11.9% of all variants, including 7.2% (8/111) of the *neutral*, and 7.7% (14/183) of the *nearly neutral* variants. The magnitude of *selectivity* changes reached 50-fold, and was much higher relative to changes in *total*, and *specific activities* (Fig. 3). To integrate the entire range of *selectivity*

Table I. Summary of changes observed within the studied variants. Variants were divided to various categories ("neutral," "nearly neutral," and "possibly neutral") in accordance with the levels of native lactonase activity and expression, as described in Fig. 1. Wild type repeats refers to a set of 147 independent measurements of wild-type PON1 aimed at determining the error range, and the extent of expected false positives in the mutants dataset. Errors and false positives arise due to differences in growth rate, lysis, expression levels, and errors in activity measurements.

	Neutral		Nearly neutral		Possibly neutral		All variants		WT repeats	
	#	%	#	%	#	%	#	%	#	%
Total activity[a]	0	0.0	2	0.0	5	3.9	7	2.3	2	1.4
Specific activity[b]	0	0.0	6	3.3	23	13.2	29	9.3	3	2.0
Selectivity[c] (⩾five-fold changes)	8	7.2	14	7.7	23	13.2	37	11.9	1	0.7
Selectivity[c] (⩾ten-fold changes)	5	4.5	8	4.4	7	5.5	15	4.8	0	0.0
Inhibitor resistance[d]	3	2.7	4	2.2	2	1.6	6	1.9	0	0.0
Number of sequenced variants (genotypes)	13	11.7	29	15.8	28	21.9	57	18.3		
Mean of amino acid mutations per gene	1.7±2.2		1.9±1.8		2.9±1.7		2.3±1.8		0.0	
Number of variants analyzed	111		183		128		311		147	

[a]Noted is the number (#), and percentage (out of all variants in the same category), of variants that exhibit total activity levels (rate of product release) that is ⩾two-fold higher than wild type with at least one of the tested promiscuous substrates.
[b]Noted is the number and percentage of variants exhibiting, relative to wild type, ⩾two-fold increase in specific activities (rate of product release divided by GFP emission), with at least one of the tested promiscuous substrates.
[c]Noted is the number and percentage of variants exhibiting, ⩾five-fold, or ⩾ten-fold, changes in selectivity—i.e., the ratio of activities between any pair of substrates, relative to the ratio observed for wild-type PON1 with the same pair.
[d]Noted is the number and percentage of variants exhibiting ⩽25% inhibition of the wild type's level.

changes observed in each variant, we applied a measure termed "overall deviation from wild-type phenotype," or d value. For each variant, selectivity factors derived from all possible combinations of substrates were calculated, normalized to the wild-type ratios, and summed up as follows:

$$d = \sum^{n} \left| \log\left(\frac{a_i^{\text{variant}}/a_i^{WT}}{aj^{\text{variant}}/aj^{WT}} \right) \right| \Big/ n,$$

where a is the activity level of either substrate i or j ($i \neq j$); and n is the number of nonredundant combinations [i.e., for substrates i and j, the ratio i/j reciprocates the ratio j/i, and thus only the first ratio (i/j) is included] for all possible activity ratios (for details see EPAPS meterial Scheme S1). To reduce errors derived from multiple sampling, the variants data was compared with a multiple-sampling set of the wild type (147 independent repeats). Indeed, the deviation values (d) of the variants were significantly different from those of the wild-type set ($p < 0.001$) (Supplementary Fig. S2).

Inhibition. The fourth criterion related to the ratio of lactonase activity with and without the inhibitor. Wild-type PON1 displayed an inhibition of 89% at 100 μM 2ODQ. At least 5% of variants (6/311) held the potential for avoiding the inhibitor (<25% of inhibition) and thereby exhibited lactonase activities that are significantly higher than wild type in the presence of the inhibitor (Fig. 4). Notably, out of these inhibitor-resistant variants, half were within the *neutral* set of variants. Thus, 3.7% (3/81) of all *neutral* variants became tolerant to the inhibitor, or "drug-resistant" by our model.

Genotype-phenotype characterization

Representative fractions of variants with wild-type-like phenotypes ($d < 0.2$; 25 variants), and with significantly altered phenotypes ($d \geqslant 0.2$; 29 variants), were sequenced, and the phenotype verified by six independent repetitions (Supplementary Table S1). Variants of the former group ($d < 0.2$) were found to have mostly no, and sometimes one, active site mutation, whereas variants with altered phenotypes possessed up to three active site mutations. The latter group was also characterized by a higher total number of mutations (average of 3 per gene, vs 1.3 for wild-type-like variants) (Supplementary Table S1; Figure S3). The observed changes in activity and selectivity therefore stem from mutations in and around the active site [Fig. 5 (upper panel)] However, none of the observed mutations is in residues that take a direct part in catalysis and comprise the active-site core (including the His-dyad, and residues ligating the catalytic calcium ion (Harel *et al.*, 2004; Khersonsky and Tawfik, 2006b).

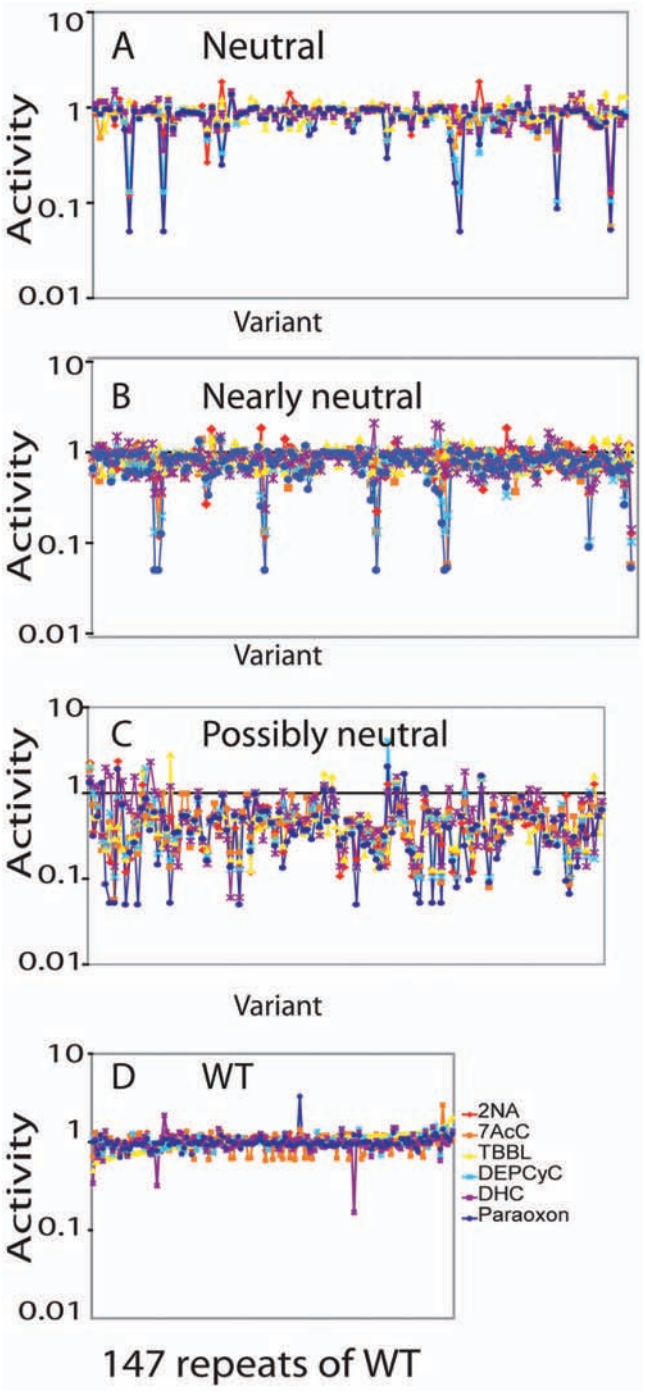

Figure 2. Changes in total activities within apparently neutral variants. Variants were categorized as detailed in Fig. 1. Presented here are the total activities (rate of product formation relative to wild type) for each variant with all substrates. (D) The total activities measured for a control set of 147 repeats of wild-type PON1.

Rather, the mutated residues are in areas that can be broadly described as the active site wall and perimeter.

A variety of phenotypic changes that affect PON1's evolutionary potential were observed. For example, variant 13F10 illustrates how frequent is an improvement amongst a population of variants that had never been under selection for higher aryl esterase activity. It has six amino acid mutations, including two in the active site (F222L and T332A). Nevertheless, it retains ~50% of the wild type's expression level, and ~30% of its total lactonase activity (its specific activity is therefore 60% of wild type's) (Supplementary Table S1). Its specific activity towards the aryl ester 2NA, and paraoxon, is ten and fourfold higher, respectively). Thus, under an explicit selection for a new function, variants such as 13F10 can provide an initial selective advantage, and then adapt via fewer mutations. The latter is supported by the observation that several of the mutations now seen in neutral variants were previously observed when PON1 gene libraries were selected towards an increase in promiscuous activities, such as the phosphotriesterase, aryl esterase, or thiolactonase (Aharoni *et al.* 2005a; Aharoni *et al.*, 2005b; Harel *et al.*, 2004). For example, variant 12B9 shows higher selectivity for DEPCyC and carries the active site mutation S193P (Supplemenatry Table S1). This apparently neutral mutation was also identified in several variants that were previously selected for improved rates with DEPCyC (Aharoni *et al.*, 2005b; Harel *et al.*, 2004). Similarly, variant 11C11 displays higher aryl esterase activity with 2NA and carries the active site mutation Y293D. The same mutation appeared in a variant evolved towards a similar aryl ester (Aharoni *et al.*, 2005b). Thus, changes in activity and selectivity observed in the neutral variants under negative, purifying selection to maintain the native activity, are consistent with improvements towards the same substrate under positive, adaptive selection for increased activity with that substrate.

The variants exhibiting lower susceptibility to the inhibitor 2ODQ illustrate the high potential for the evolution of drug resistance under neutrality. In the presence of nearly saturating concentrations of the inhibitor, these variants exhibit the same level of lactonase activity as the uninhibited wild type (Fig. 4). The high frequency of inhibitor-resistant variants is probably due to a whole range of mutations that occur in an area relatively remote from PON1's catalytic core (4–14 Å away from the catalytic calcium, and 9–18 Å from the catalytic His) and confer resistance (Supplementary Fig. S4). Interestingly, the inhibitor-tolerant variants also lost their promiscuous phosphotriesterase activity (with both paraoxon and DEPCyC), and were scored amongst those most deviating from wild-type phenotype ($d > 0.39$). One variant, 31A1, has completely lost its aryl esterase as well as phosphotriesterase activities (Supplementary Table S1).

Neutral variants and natural diversity in PONs

If the phenotypic changes in the variants characterized here are primarily neutral, sequence changes observed in them should be found in natural homologues of PON1. The below describes evidence in support of this hypothesis.

The mammalian serum paraoxonase/arylesterase gene (PON1) is a member of a family consisting of three mamma-

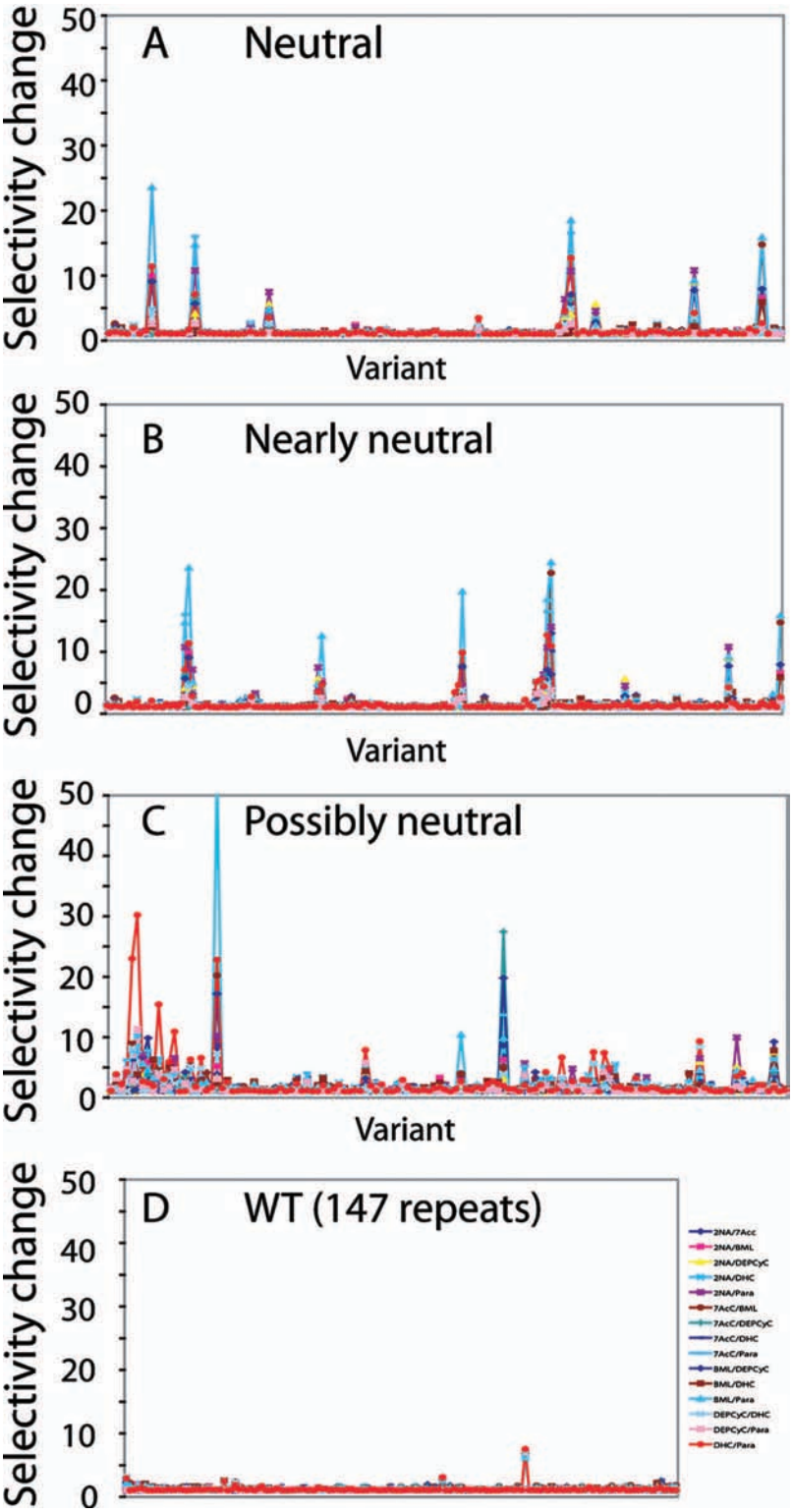

Figure 3. Changes in selectivity within apparently neutral variants. Selectivity values were determined by calculating for each variant the ratios of activities between different pairs of substrates and then normalizing to wild type. (D) The selectivity measured for a control set of 147 repeats of the wild type.

lian paralogs (PON1, 2, and 3) that were all identified as lactonases (Draganov *et al.*, 2005). To compare the residue variance in each position along the three paralogs we identified the conserved nongapped blocks of the PON family. These blocks allowed us to look at specific positions and their variance in nature. Within these blocks there is a range of evolutionary conservation ranging from compleletly to mildly conserved positions. Three evolutionary conserved blocks could be detected in this gene family (27 known vertebrate sequences) that cover about a third of the gene [Fig. 5 (lower

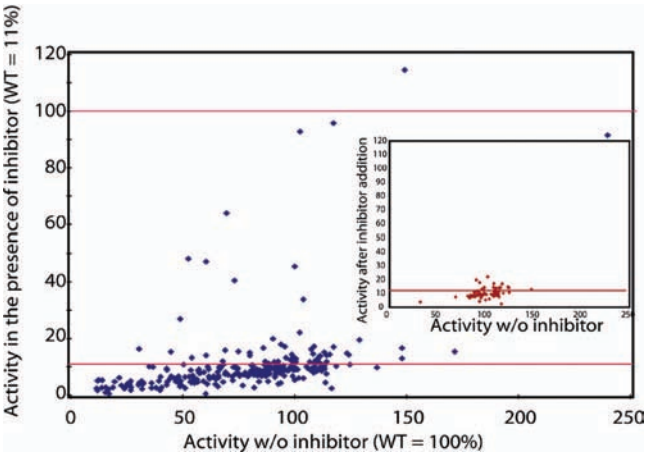

Figure 4. Changes in susceptibility to inhibition within apparently neutral variants. Presented is the relative level of lipolactonase activity in the presence of 100 μM of the inhibitor 2ODQ, for the entire set of 311 apparently neutral variants, and 147 repeats of the wild type (inset). The wild-type lactonase activity in the absence of inhibitor (100%), and in its presence (11%) are marked with red lines.

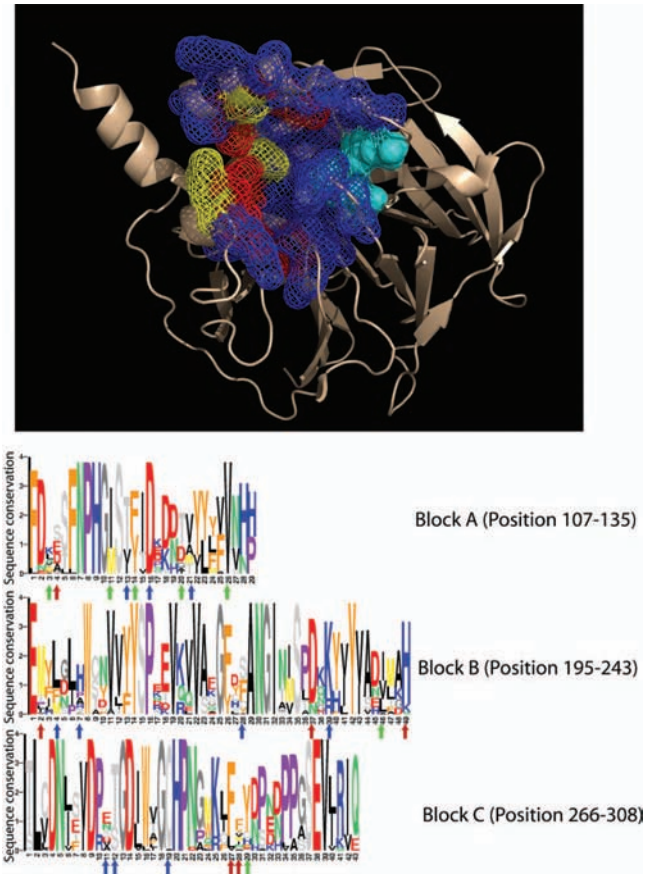

Figure 5. The apparently neutral sequence changes with respect to PON1's tertiary and primary structures. (A) A ribbon representation of the 3D structure of PON1. The active site pocket comprises 32 residues and is presented in mesh. The highly conserved active site core (catalytic histidine-dyad H115 and H134) is shown in cyan spheres. Active site residues found to be mutated in *neutral* and *nearly neutral* variants are colored in red and *possibly neutral* variants in yellow. (B) Mutations found in "neutral" PON1 variants correlates with the natural variability of the PON family. The multiple sequence alignment of the PON gene family indicated three conserved blocks (marked *A*, *B*, and *C*). The conservation of each position and residue is illustrated as its height in bits (LOGO presentation). Residues found to be mutated in *neutral* PON1 variants are indicated by a red arrow, *nearly neutral* in green, and *possibly neutral* in blue.

panel)]. Within these blocks, there are 32 residues that are absolutely conserved. Only two out of 23 mutations identified in these conserved blocks are within an absolutely conserved residue (Supplemenatry Table S1, V132A in variant 12C2, and D122G in variant 33A11), although most of the substitutions observed within these sites (16/23) are for amino acids that differ from those found in natural PON genes. Further support for the neutral nature of mutations is given by the identification of several variants carrying the mutations K192R and K192N (Supplementary Table S1). These variations are analogous to the nonpathological human 192R/Q polymorphism and also show a similar phenotype [K192N significantly reduces the phosphotriesterase activity as in the 192Q polymorph of human PON1 (Davies *et al.*, 1996)].

We also observed multiple changes within a set of 15 residues that are known to affect the substrate selectivity of PONs (Harel *et al.*, 2004). Because all three PON paralogs are lactonases (Draganov *et al.*, 2005), the variability in these residues primarily affects the various man-made promiscuous substrates. Indeed, these vary dramatically between the natural paralogs [e.g., PON1 exhibits considerable phosphotriesterase activity ($k_{cat}/K_M \leqslant 10^4$ M^{-1} s^{-1}), PON2 has none, and PON3's is barely detectable (Draganov and La Du, 2004)]. We found mutations in 11 out of the 15 residues that comprise this set of selectivity determining residues, primarily in variants exhibiting changes in activity and selectivity ($d > 0.2$). Thus, by virtue of maintaining the primary lactonase function, and altering the promiscuous functions, changes within these residues may mirror neutral changes that had occurred during the evolutionary drift of natural PONs.

DISCUSSION

Darwinian theory surmises that, in general, evolutionary processes occur gradually while maintaining organismal fitness throughout. Thus, prior to its fixation, a "new" trait is typically represented at low frequency, or at low level, amongst an existing species. At the molecular level, this rationale is hard to reconcile with the paradigm of "one sequence–structure–one function." However, the "new view" of proteins surmises that, in addition to their primary functions and structures, biomolecules may possess a range of alternative structures, and functions, that have never been selected for (James and Tawfik, 2003). These conformational

isomers and activities, which are referred to as promiscuous, comprise latent evolutionary potentials (Aharoni *et al.*, 2005b; Copley, 2003; James and Tawfik, 2003; O'Brien and Herschlag, 1999). Indeed, promiscuous activities were shown to comprise ample starting points in laboratory evolution experiments performed both *in vitro* (Khersonsky *et al.*, 2006) and *in vivo* (Hall, 1999; Miller and Raines, 2005). There also exists evidence (which is inevitably indirect) indicating that promiscuity may have driven the divergence of new receptors (Bridgham *et al.*, 2006) and enzymes (Afriat *et al.*, 2006) in nature.

The notion of neutrality, or neutral networks, is directly related to the "new view" [see also (Wroe *et al.*, 2007)]. Neutrality ascribes a role to neutral drifts or nonadaptive evolution in prompting adaptive evolution. It is surmised that genetic diversity acquired under neutral drift, as fixated mutations or polymorphism, can facilitate the transition towards a new function (Wagner, 2005a; Wagner, 2005b). The notion of neutral networks, initially developed in the context of RNA folding maps (Schuster *et al.*, 1994), has provided key theoretical grounds for studying how evolutionary potentials are imbedded in neutral drifts. Reaching an optimal conformation or new function need not depend on exploring all possibilities. A landscape guided exploration may funnel evolutionary adaptation in a manner similar to protein folding. The genotype-phenotype landscape properties described by Fontana and Schuster allow movements from one phenotype to another by diffusion through the neutral network. If these conclusions, which regard RNA neutral network, are also relevant to proteins, then promiscuity of protein function and conformation may highlight the overlapping regions or transition points between the neutral networks of different phenotypes (Fig. 6).

The available studies of neutral networks have focused on computational simulations of structures, and primarily of RNA structures (Aita *et al.*, 2003; Fontana and Schuster, 1998; Huynen *et al.*, 1996; Schuster and Fontana, 1999). But structure is just the means by which a biomolecule attains function and highly similar structures often display different functions. However, a neutral network layout that illustrates a linkage between genotype and function, can, at present, be derived only by experiments. Indeed, this study provides the first empirical measurements of the frequencies and magnitudes of changes in latent activities within a seemingly neutral network.

Our experimental system comprises a model of no physiological relevance, nor is it possible to establish that our apparently neutral variants would have no fitness effects while replacing wild-type PON1 in a living organism. Even natural polymorphs or allozymes of the same enzyme that appear to be neutral can be the subject of selection under different conditions or genetic backgrounds (Hartl and Dykhuizen, 1981; Kimura, 1986). Nevertheless, we have attempted to illustrate a feasible and consistent scenario. Our model enzyme (PON1) has one native or primary function that is under selection, and the apparently neutral variants studied here possess wild-type-like levels of the primary activity, as well as wild-type-like expression levels. Evolutionary potentials are manifested in changes in the levels of various promiscuous activities. Several different activities that deviate from PON1's native lactonase function (EC 3.1.1.25) were monitored: aryl esterase (EC 3.1.1.2), phosphotriesterase (EC 3.1.8), aromatic lactonase [that proceeds via a mechanism that differs from the aliphatic lactones that comprise PON1's primary substrates (Khersonsky and Tawfik, 2005), and the binding of an active site inhibitor. We assume that PON1 has never been under selection for catalyzing the hydrolysis of these synthetic substrates, nor for binding the man-made inhibitor used here.

The inhibitor illustrates the case of drug resistance—a very common scenario of adaptive evolution. From an evolutionary view, drug binding is inherently promiscuous—despite the fact that drugs mimic natural substrates or transition states, no enzyme has actually evolved to bind, let alone be inhibited by a synthetic drug (Aharoni *et al.*, 2005b). Whilst this view has been criticized (Fernandez *et al.*, 2005), the linkage between loss of certain promiscuous activities and loss of inhibition (Table S1) suggests a common theme between the recognition of promiscuous substrates and that of synthetic inhibitors. Quite frequently, loss of either, or both, is brought about by mutations that do not hamper PON1's native activity. At any rate, our results demonstrate the facility by which drug resistance may emerge from a repertoire of largely neutral mutations (Fig. 4).

Random mutagenesis of PON1's gene, and selection, gave a range of variants that are, on average, three mutations away from wild type. The phenotype of 311 apparently neutral variants was characterized in detail, and the genotype of 54 of those was acquired. All parameters, be they enzymatic activities, inhibition, or expression levels, were described by continuous measures [Eq. (1); Figs. 1–4]. Analysis of this set of apparently neutral variants indicated that, all variants that exhibited changes in specificity had at least one mutation within the active site. Still, the residues mediating the catalytic functionality were not mutated, but rather the wall and perimeter of the active site. The subtlety of these modifications is also reflected in the fact that almost none of these changes had occurred in a completely conserved residue (Fig. 5), and some of these changes are seen amongst natural polymorphs of PON1. Moreover, a notable number of apparently neutral variants shared one or more active site mutations with variants that were previously evolved towards a new function (e.g., S193P, T332A, Y293D; Supplementary Table S1). Indeed, when this set of 311 apparently neutral variants was screened with other promiscuous substrates, variants with large rate improvements could be identified. For example, a screen with a promiscuous phosphonate substrate (methylphosphonic acid 3-cyano-4-methyl-2-oxo-2

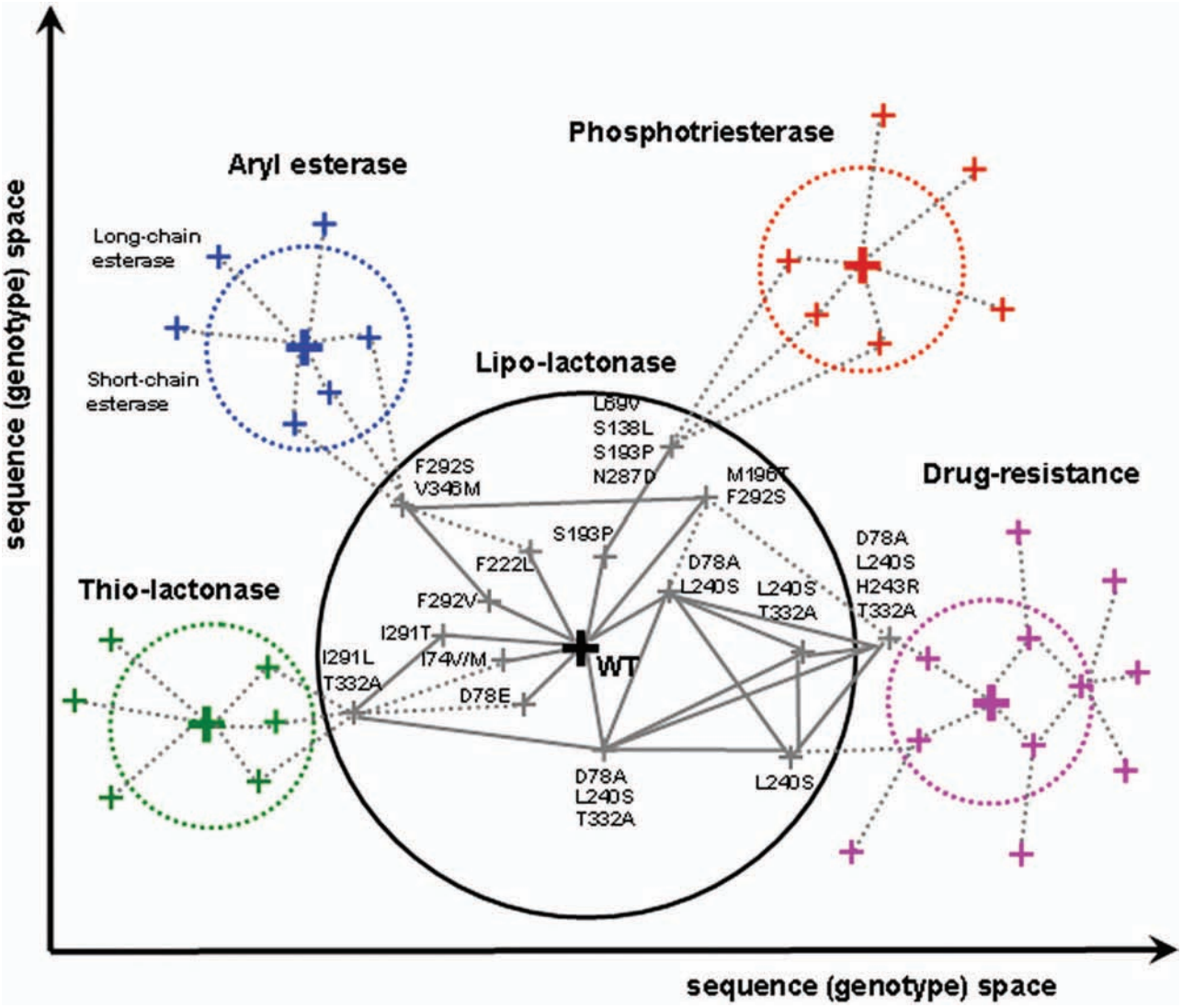

Figure 6. Schematic representation of PON1 putative neutral network. The scheme is based on the neutral variants described here (supplementary Table S1), and on the mutations and transition phenotypes identified in a previous directed evolution experiment and in natural paralogs of PON1 (Aharoni *et al*., 2005a; Aharoni *et al*., 2005b; Harel *et al*., 2004). The length of the edges corresponds to their neutrality state. The *neutral* variants are scattered in a distance (edge length) of 1 from the WT. *Nearly neutral* variants are scattered around the WT in a distance of 2, and *possibly neutral* in a distance of 3. The direction is the tendency towards a specific latent activity. The other neutral regions (e.g., aryl esterase, thio-lactonase, etc.) were arbitrarily placed in some point in the sequence space scheme, although some phenotypes might be closer to others (e.g., thio-lactonase and lipo-lactonase). The large circle denotes the boundaries of the neutral network of PON1 native phenotype (lipo-lactonase). The dashed borders and edges are hypothetical. Filled edges are based on common mutations. Each node corresponds to a specific sequence with one or more mutations that do not appear at the point of origin (WT sequence).

H-coumarin-7-yl ester cyclohexyl ester) identified two apparently neutral variants (33B4 and 35B3; Supplementary Table S1) that when expressed and purified (without the GFP fusion) exhibited 13 and 72-fold, improved k_{cat}/K_M values relative to wild-type PON1. Thus, neutral drifts of the kind described here may also provide a powerful starting point for directed enzyme evolution *in vitro*.

Our data thus provide empirical evidence in support of the hypothesis that neutrality enables the formation of latent changes or "latent adaptation." We found that latent evolutionary potentials are indeed very frequent within a neutral set of related genotypes, and that these potentials are most often seen as changes in specificity for one or more promiscuous substrates. The initial manifestation of such "latent adaptation" might be in providing an immediate selective advantage by expanding the range of activities of an existing enzyme. Alternatively, the potential for change might be revealed with subsequent changes, thus shortening the adaptive path by one, or possibly more, mutations. For example, in most variants tested, the potential for change is manifested by changes in specificity. Subsequent changes in regulation, or even duplication of the evolving gene, may lead to in-

creased expression, and thus increases the total level of activity to provide a distinct selective advantage.

Combined with the computational analysis in the accompanying article (Wroe *et al.*, 2007) our results show how gradual and close in genotype space the transitions between one phenotype to another can be, thus forming a continuous phenotype space in conjunction with Maynard-Smith's concept (Maynard-Smith, 1970). The neutral active site mutations identified here bring neutral variants much closer to a transition point from which an existing phenotype could switch to a new one (Fig. 6). This scenario differs from that of a fitness landscape punctuated with high peaks, each corresponding to a different phenotype separated by valleys of low or no fitness. In a continuous space, random diffusion through the neutral network brings different variants to different localities. Some of these variants appear closer to adjacent neutral networks corresponding to other phenotypes, and promiscuity underlines the trajectories between the different networks, each corresponding to a different phenotype (Fig. 6). Our knowledge of these landscapes is still rather limited. For example, what would be the trajectories leading from one new phenotype to another (e.g., "phsphotriesterase" into "thiolactonase")? By extrapolating from PON1's behavior one may assume that newly emerging phenotypes would eventually possess their own neutral networks and range of promiscuous activities, and these may allow smooth transitions between them. However, the rate, and the extent by which a neutral network of a newly emerging protein is created, remain unknown.

MATERIALS AND METHODS

Constructing the PON1 gene library. The *NcoI* site within the GFPmut3 gene (Clontech) was removed by a synonymous mutation, and the gene was inserted in pET32b (+) plasmid from which the thioredoxin fusion protein and peptide tags were truncated using *NotI/XhoI* sites. The starting point for our neutral drift was a recombinant PON1 variant, which exhibits 95% sequence identity, and enzymatic parameters that are essentially identical to wild-type rabbit PON1 (Aharoni *et al.*, 2004). The "wild-type PON1" template gene was constructed by inserting the recombinant PON1 variant G3C9 (Aharoni *et al.*, 2004) into the GFPmut3-pET32 plasmid (Novagen) using the *NcoI/NotI* sites (for a detailed map see Supplementary see Fig. S5). The resulting PON-GFP fusion plasmid was used as template for library construction by error-prone PCR using the dPTP and 8-oxo-dGTP wobble base analogs (Zaccolo *et al.*, 1996). Primers used for PON1 amplifications are listed elsewhere (Aharoni *et al.*, 2004). After cleaving the amplified gene at the *NcoI/NotI* sites, the mutated PON1 open reading frame was recloned into the pET32-GFP plasmid, thus keeping the sequence of both the GFP and its upstream linker constant. Sequencing of the libraries prior to selection indicated an average of 1.88 mutations per gene, 0% of synonymous mutations, 80% transitions, and 20% transversions.

Expression of PON-GFP library variants. Ligated plasmids were electrotransformed into *E. cloni* competent cells (Lucigen) to give $\sim 3 \times 10^6$ individual transformants. Plasmid DNA extracted from these colonies was then transformed into BL21(DE3) cells for the expression of PON-GFP of the T7 promoter. Transformed BL21(DE3) cells were grown (while shaking at 250 RPM) in 5 ml 2xYT media containing 100 μg/ml ampicillin and 1 mM $CaCl_2$, for 12 h at 30 °C, followed by another 24 h at 20 °C.

Sorting of gene libraries. Libraries were sorted by compartmentalization of single *E. coli* cells, each expressing an individual library variant in double emulsion droplets, and sorting these droplets by fluorescent activated cell sorter (FACS), essentially as decribed (Aharoni *et al.*, 2005a; Miller *et al.*, 2006). Briefly, following PON-GFP expression, cells were centrifuged at 3000 g for 10 min at 4 °C, resuspended in 2xYT, and kept for 15 min at room temperature for recovery. Cells were then rinsed once in 0.1 M Tris-HCl, 1 mM $CaCl_2$, 0.1 M NaCl, *p*H 8.3, resuspended in the same buffer, and filtered through a 5 μm filter (Sartorius). Filtered cells were compartmentalized in the first emulsion (water-in-oil), and 270 mM solutions of TBBL and the thiol-detecting dye CPM (molecular probes) were added (Khersonsky and Tawfik, 2006a) to the oil phase (1.9 μl, to a final concentration of 0.3 mM and 50 μM, respectively). The production of the second emulsion (water-in-oil-in-water) and sorting were performed as described (Miller *et al.*, 2006). Events were sorted according to GFP emission (530 nm) and TBBLase activity (450 nm) as exemplified in Supplementary Fig. S6. The sorted cells were plated on LB-agar plates containing 100 μg/ml ampicillin and 20 mM glucose, and grown overnight at 30 °C.

Growth of and lysis of PON-GFP variants. Following FACS sorting, approximately 200 individual colonies were grown in duplicates in 96-deep-wells plates in a 0.5 ml 2xYT medium. Several repeats of wild type were also grown as standards. Growth conditions were as in the expression of PON-GFP library variants. Following OD measurements at 600 nm for monitoring cell density, the plates were centrifuged at 3000 g for 15 min at 4 °C, and pellets were kept at −20 °C for 24 h. The pellets were lysed with 200 μl of lysis buffer: 0.1 M Tris-HCl *p*H 8.0, 1 mM $CaCl_2$, 10 μg/ml Lysozyme (Sigma), 0.2% Triton x-100, and 5 units/ml Benzonase (Novagen), by shaking at 1300 RPM for 30 min at 37 °C. The plates were centrifuged at 3000 g for 15 min at 4 °C; the supernatant was transferred to a new set of plates and stored at 4 °C.

Activity measurements. GFP emission ($488_{ex}/516_{em}$) and apparent rates (k_{obs}) for six different substrates were measured in a plate reader (Synergy-HT BioTek). Substrate concentration and wavelengths were as follows: 0.5 mM/405 nm, paraoxon; 0.12 mM/365 nm, O-acetoxy-

7-hydroxycoumarin (7AcC); 0.1 mM/408 nm, 7-O-diethylphosphoryl-3-cyano-7-hydroxycoumarin (DEPCyC); 0.5 mM/320 nm, 2-naphtylacetate (2NA); 0.5 mM/270 nm, dihydrocoumarin (DHC); 0.05 mM (plus 0.6 mM DTNB)/412 nm, 5-thiobutyl butyrolactone (TBBL). All kinetic measurements were performed at the linear range of product release. Background product formation rates (rates in the presence of lysates of cells expressing GFP only), and GFP emission (emission in cells expressing PON1 only), were subtracted from the k_{obs} and GFP emission values obtained with the measured variants, and the resulting "net" values were normalized to the values of wild-type PON1-GFP. The background levels varied for each measurement, comprising 8.6% of the wild type's signal for TBBL, 5.5% for 2NA, 14.5% for 7AcC, 16.5% for DepCyC, 17% for DHC, 5.6% for paraoxon, and 6.6% for GFP. For each activity, the coefficient of variance (CV) for the duplicates was calculated to assess the repeatability (the difference between duplicate measurements divided by the mean). Variants exhibiting a CV value of >0.5 with any of the measured enzymatic activities, or GFP emission, were discarded from the final analysis. The expected CV of each activity was calculated from 77 repeats of the wild type, which were independently inoculated, grown, and lysed (Supplementary Fig. S7). 3SD above mean background noise level was set as the threshold for noise. Therefore, all substrate activities and GFP emission levels below this threshold were considered background noise.

Inhibition assays. A 100 mM DMSO solution of 2ODQ (2-oxo-dihydro-quinoline) was added to lysates to a final concentration of 100 μM), and incubated at room temperature for 15 min. The substrate (0.05 mM TBBL plus 0.6 mM DTNB) was added, and k_{obs} measured as above. Rates of a sample with no 2ODQ (under the very same conditions, including DMSO concentration) were measured in parallel, and the percentage of inhibition determined from the ratio of the "net" rates observed in the two samples.

ACKNOWLEDGMENTS

We are grateful to Dr. Hillary Voet for valuable advice regarding statistical controls and to Dr. Tony Dean for his friendly comments. We thank the Israel Ministry of Science and Technology, the Horowitz Fund, and the Center for Systems Biology at the Weizmann Institute for financial support.

REFERENCES

Afriat, L, Roodveldt, C, Manco, G, and Tawfik, DS (2006). "The latent promiscuity of newly identified microbial lactonases is linked to a recently diverged phosphotriesterase." *Biochemistry* **45**, 13677–13686.

Aharoni, A, Amitai, G, Bernath, K, Magdassi, S, and Tawfik, DS (2005a). "High-throughput screening of enzyme libraries: thiolactonases evolved by fluorescence-activated sorting of single cells in emulsion compartments." *Chem. Biol.* **12**, 1281–1289.

Aharoni, A, Gaidukov, L, Khersonsky, O, Mc, QGS, Roodveldt, C, and Tawfik, DS (2005b). "The 'evolvability' of promiscuous protein functions." *Nat. Genet.* **37**, 73–76.

Aharoni, A, Gaidukov, L, Yagur, S, Toker, L, Silman, I, and Tawfik, DS (2004). "Directed evolution of mammalian paraoxonases PON1 and PON3 for bacterial expression and catalytic specialization." *Proc. Natl. Acad. Sci. U.S.A.* **101**, 482–487.

Aita, T, Ota, M, and Husimi, Y (2003). "An *in silico* exploration of the neutral network in protein sequence space." *J. Theor. Biol.* **221**, 599–613.

Beadle, BM, and Shoichet, BK (2002). "Structural bases of stability-function tradeoffs in enzymes." *J. Mol. Biol.* **321**, 285–296.

Behe, MJ, and Snoke, DW (2004). "Simulating evolution by gene duplication of protein features that require multiple amino acid residues." *Protein Sci.* **13**, 2651–2664.

Bornberg-Bauer, E (1997). "How are model protein structures distributed in sequence space?." *Biophys. J.* **73**, 2393–2403.

Bornscheuer, UT, and Kazlauskas, RJ (2004). "Catalytic promiscuity in biocatalysis: using old enzymes to form new bonds and follow new pathways." *Angew. Chem., Int. Ed. Engl.* **43**, 6032–6040.

Bridgham, JT, Carroll, SM, and Thornton, JW (2006). "Evolution of hormone-receptor complexity by molecular exploitation." *Science* **312**, 97–101.

Copley, SD (2003). "Enzymes with extra talents: moonlighting functions and catalytic promiscuity." *Curr. Opin. Chem. Biol.* **7**, 265–272.

Davies, HG, Richter, RJ, Keifer, M, Broomfield, CA, Sowalla, J, and Furlong, CE (1996). "The effect of the human serum paraoxonase polymorphism is reversed with diazoxon, soman and sarin." *Nat. Genet.* **14**, 334–336.

Draganov, DI, and La, DuB N (2004). "Pharmacogenetics of paraoxonases: a brief review." *Naunyn-Schmiedeberg's Arch. Pharmacol.* **369**, 78–88.

Draganov, DI, Teiber, JF, Speelman, A, Osawa, Y, Sunahara, R, and La Du, BN (2005). "Human paraoxonases (PON1, PON2, and PON3) are lactonases with overlapping and distinct substrate specificities." *J. Lipid Res.* **46**, 1239–1247.

EPAPS document No. E-HJFOA5-1-008701 for supplemental material. This doucument can be reached through a direct link in the online articles HTML reference section or via the EPAPS home page (http://www.aip.org/pubservs/epaps.html).

Fernandez, A, Tawfik, DS, Berkhout, B, Sanders, R, Kloczkowski, A, Sen, T, and Jernigan, B (2005). "Protein promiscuity: drug resistance and native functions—HIV-1 case." *J. Biomol. Struct. Dyn.* **22**, 615–624.

Fontana, W, and Schuster, P (1998). "Shaping space: the possible and the attainable in RNA genotype-phenotype mapping." *J. Theor. Biol.* **194**, 491–515.

Govindarajan, S, and Goldstein, RA (1997). "Evolution of model proteins on a foldability landscape." *Proteins* **29**, 461–466.

Graur, D, and Li, W-H (2000). Fundamentals of molecular evolution, 2nd ed., Sinauer Associates, Inc., Massachusetts.

Gressel, J, and Levy, AA (2006). "Agriculture: the selector of improbable mutations." *Proc. Natl. Acad. Sci. U.S.A.* **103**, 12215–12216.

Hall, BG (1999). "Experimental evolution of Ebg enzyme provides clues about the evolution of catalysis and to evolutionary potential." *FEMS Microbiol. Lett.* **174**, 1–8.

Harel, M, *et al.* (2004). "Structure and evolution of the serum paraoxonase family of detoxifying and anti-atherosclerotic enzymes." *Nat. Struct. Mol. Biol.* **11**, 412–419.

Hartl, DL, and Dykhuizen, DE (1981). "Potential for selection among nearly neutral allozymes of 6-phosphogluconate dehydrogenase in Escherichia coli." *Proc. Natl. Acad. Sci. U.S.A.* **78**, 6344–6348.

Huynen, MA, Stadler, PF, and Fontana, W (1996). "Smoothness within ruggedness: the role of neutrality in adaptation." *Proc. Natl. Acad. Sci. U.S.A.* **93**, 397–401.

James, LC, and Tawfik, DS (2003). "Conformational diversity and protein evolution—a 60-year-old hypothesis revisited." *Trends Biochem. Sci.* **28**, 361–368.

Khersonsky, O, Roodveldt, C, and Tawfik, DS (2006). "Enzyme promiscuity: evolutionary and mechanistic aspects." *Curr. Opin. Chem. Biol.* **10**, 498–508.

Khersonsky, O, and Tawfik, DS (2005). "Structure-reactivity studies of serum paraoxonase PON1 suggest that its native activity is lactonase." *Biochemistry* **44**, 6371–6382.

Khersonsky, O, and Tawfik, DS (2006a). "Chromogenic and fluorogenic assays for the lactonase activity of serum paraoxonases."

ChemBioChem **7**, 49–53.
Khersonsky, O, and Tawfik, DS (2006b). "The histidine 115-histidine 134 dyad mediates the lactonase activity of mammalian serum paraoxonases." *J. Biol. Chem.* **281**, 7649–7656.
Kimura, M (1986). "DNA and the neutral theory." *Philos. Trans. R. Soc. London, Ser. B* **312**, 343–354.
Kirschner, M, and Gerhart, J (1998). "Evolvability." *Proc. Natl. Acad. Sci. U.S.A.* **95**, 8420–8427.
Lipman, DJ, and Wilbur, WJ (1991). "Modelling neutral and selective evolution of protein folding." *Proc. R. Soc. B.* **245**, 7–11.
Lynch, M (2005). "Simple evolutionary pathways to complex proteins." *Protein Sci.* **14**, 2217–2225 ("Discussion." 2226–2217).
Ma, B, Shatsky, M, Wolfson, HJ, and Nussinov, R (2002). "Multiple diverse ligands binding at a single protein site: a matter of pre-existing populations." *Protein Sci.* **11**, 184–197.
Maynard-Smith, J (1970). "Natural selection and the concept of a protein space." *Nature (London)* **225**, 563–564.
Meiering, EM, Serrano, L, and Fersht, AR (1992). "Effect of active site residues in barnase on activity and stability." *J. Mol. Biol.* **225**, 585–589.
Miller, BG, and Raines, RT (2005). "Reconstitution of a defunct glycolytic pathway via recruitment of ambiguous sugar kinases." *Biochemistry* **44**, 10776–10783.
Miller, OJ, Bernath, K, Agresti, JJ, Amitai, G, Kelly, BT, Mastrobattista, E, Taly, V, Magdassi, S, Tawfik, DS, and Griffiths, AD (2006). "Directed evolution by in vitro compartmentalization." *Nat. Methods* **3**, 561–570.
Nei, M (2005). "Selectionism and neutralism in molecular evolution." *Mol. Biol. Evol.* **22**, 2318–2342.
O″Brien, PJ, and Herschlag, D (1999). "Catalytic promiscuity and the evolution of new enzymatic activities." *Chem. Biol.* **6**, R91–R105.
Schultes, EA, and Bartel, DP (2000). "One sequence, two ribozymes: implications for the emergence of new ribozyme folds." *Science* **289**, 448–452.
Schuster, P, and Fontana, W (1999). "Chance and necessity in evolution: lessons from RNA." *Physica D* **133**, 427–452.
Schuster, P, Fontana, W, Stadler, PF, and Hofacker, IL (1994). "From sequences to shapes and back: a case study in RNA secondary structures." *Proc. R. Soc. London, Ser. B* **255**, 279–284.
Wagner, A (2005a). Robustness and Evolvability in Living Systems: Princeton Studies in Complexity, Princeton University Press, New Jersey.
Wagner, A (2005b). "Robustness, evolvability, and neutrality." *FEBS Lett.* **579**, 1772–1778.
Waldo, GS, Standish, BM, Berendzen, J, and Terwilliger, TC (1999). "Rapid protein-folding assay using green fluorescent protein." *Nat. Biotechnol.* **17**, 691–695.
Wroe, R, Chan, HS, and Bornberg-Bauer, E (2007). "A structural model of latent evolutionary potentials underlying neutral networks in proteins." *HFSP J.* **1**(1), 79–87.
Zaccolo, M, Williams, DM, Brown, DM, and Gherardi, E (1996). "An approach to random mutagenesis of DNA using mixtures of triphosphate derivatives of nucleoside analogues." *J. Mol. Biol.* **255**, 589–603.