

Directed enzyme evolution via small and effective neutral drift libraries

Rinkoo D Gupta & Dan S Tawfik

Small libraries for directed evolution can be obtained by neutral drifts that maintain the protein's original function, yielding highly polymorphic, stable and evolvable variants. We describe methods for preparing such libraries, using serum paraoxonase (PON1). An optimized GFP variant fused to PON1 reported levels of soluble, functional enzyme, enabling selection by flow cytometry and identification of enzyme variants exhibiting improved specific and total activities toward several substrates, including toxic organophosphates.

Traditionally, directed evolution has relied on high throughput: the larger the library, the greater the chance of finding 'hits' with the desired ligand affinity or enzymatic activity. Although large libraries are essential when high substrate selectivity is under selection, small but effective libraries are of crucial importance, especially when complex or difficult-to-assay activities are sought.

The routine mode of directed evolution is hampered by the high frequency of deleterious mutations (≥ 0.3) and the >100 -fold lower frequency of beneficial mutations. Thus, as mutations accumulate, most library genes become 'nonviable'; they do not encode a properly folded, let alone functional, protein¹. Neutral drifts comprise an alternative mode by which mutations accumulate under selection to maintain the protein's original function and structure. The result is a relatively small ensemble of highly polymorphic mutants that are all properly folded and functional. Neutral drifts are more than a mere removal of inactive library variants (purifying selection). The high mutational rates applied, and the repeated rounds of mutation and selection lead to mutational robustness and evolvability that are usually not present in the wild-type starting enzyme^{2–4}. Here we describe generally applicable methods for the preparation of neutral drift libraries with the aim of identifying 'first-generation' mutants that have improved activity with a particular substrate and then subjecting them to standard directed evolution. We demonstrate this strategy with serum paraoxonase (PON1).

Reporters such as GFP have been used in directed evolution experiments with the aim of removing from libraries enzyme variants with nonsense and highly destabilizing mutations^{5–7}. We used a GFP as a reporter not only to remove completely inactive variants but also to measure the amounts of soluble and active

enzyme. Because most mutations that mediate new functions are destabilizing and may therefore reduce enzyme levels^{8,9}, measuring specific activity (rate of product release (v_0) normalized to the amount of enzyme-GFP fusion, $[E]_0$) could enable the isolation of variants with evolutionary potential.

Fusion of an evolved GFP variant, GFPmut3, at the C terminus of PON1 previously resulted in 20-fold lower fluorescence and enzymatic activity⁴. When we fused enhanced GFP (EGFP)¹⁰ to a library of randomly mutated PON1 variants, we found poor correlations between enzymatic activity and fluorescence (Fig. 1a). We then considered using a robust folding variant of GFP, super-folder GFP (sfGFP), with nine amino-acid mutations compared to EGFP¹¹. Shuffling the *EGFP* gene in the presence of synthetic oligonucleotides individually encoding these nine mutations (Supplementary Methods online) yielded a library, the most brightly fluorescent variant in which (sfGFP-D12) contained eight out of the nine mutations in EGFP but showed the lowest correlation between fluorescence and PON1 activity (Fig. 1b and Supplementary Table 1 online). When we sorted libraries of PON1 variants fused to sfGFP-D12 by fluorescence-activated cell sorting (FACS), the majority of the sorted variants were enzymatically inactive. However, PON1 fusion with variant sfGFP-F12, containing only three sfGFP mutations (S30R, M153T and I171V) and exhibiting intermediate fluorescence levels, showed the best correlation (Fig. 1c). Unlike the highly fluorescent variants (EGFP, sfGFP-D12 and other sfGFP-related mutants), sfGFP-F12 was less soluble and did not increase expression of PON1 (Supplementary Fig. 1 and Supplementary Table 1 online; other GFP library variants are described in Supplementary Fig. 1b). Nonetheless, or because of that, it was far more sensitive to the stability and solubility of the fused PON1 variants (Fig. 1c), suggesting that a balance of stability between the GFP reporter and the reported enzyme is the key factor. Therefore, we used sfGFP-F12 (referred to hereafter as GFP*) to sort the neutral drift libraries.

Although GFP* was optimal for PON1, the range of GFP variants afforded by the library described above should enable facile optimization for other target enzymes.

As described below, neutral drift libraries can be generated by applying high or low mutational loads. High mutation rates yield highly polymorphic variants in only a few rounds of selection. But because only a small fraction of highly mutated variants remains viable, high throughput is essential. The GFP* fusion provided a facile way of sorting large libraries by FACS. Thus, we mutated the gene encoding the recombinant variant PON1 by error-prone PCR at an average of 4.8 ± 1.6 amino-acid exchanges per gene. We cloned the resulting library in fusion with the gene encoding GFP*, transformed $\sim 10^6$ variants to *Escherichia coli* and sorted the transformants by FACS. In each round, we sorted $\sim 10^6$ events, each denoting an individual *E. coli* cell expressing a different PON1

Department of Biological Chemistry, Weizmann Institute of Science, Rehovot 76100, Israel. Correspondence should be addressed to D.S.T. (tawfik@weizmann.ac.il).

RECEIVED 3 JULY; ACCEPTED 25 SEPTEMBER; PUBLISHED ONLINE 19 OCTOBER 2008; DOI:10.1038/NMETH.1262

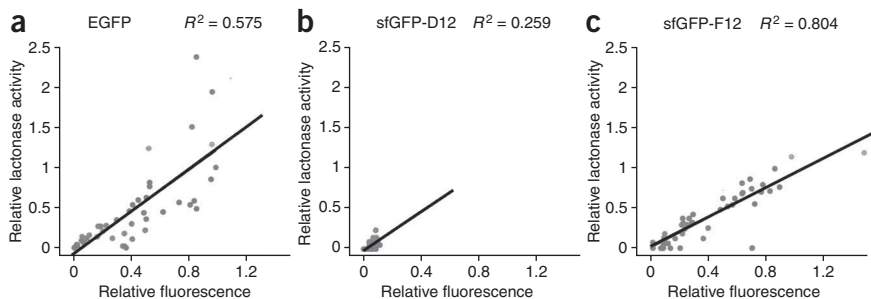


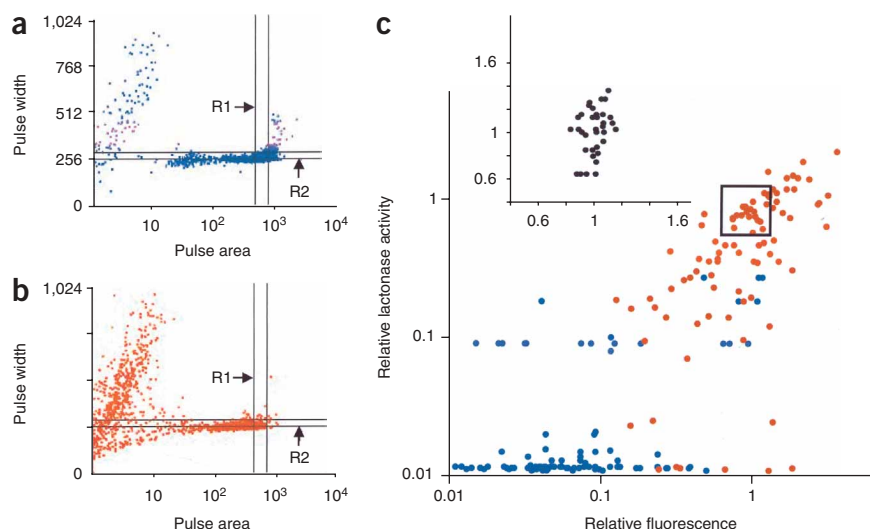
Figure 1 | Correlation of fluorescence and enzymatic activities in fusions of PON1 variants with three different GFP-variant reporters. (a–c) The same library of randomly mutated genes encoding PON1 variants was fused to the 5' end of genes encoding EGFP (a), sfGFP-D12 (b) and sfGFP-F12 (c). Forty-five randomly chosen clones were individually grown and lysed, and their fluorescence and lactonase activities (with the TBBL substrate) were measured. The resulting values were normalized to the activity of wild-type PON1 fused to the same GFP-variant reporter. The solid line represents a linear fit.

variant fused to GFP*. We sorted by the standard parameter of fluorescence intensity (pulse area), as well as by pulse width (Fig. 2). On average, the FACS sorts yielded 5,000 events each, and we recovered $\geq 70\%$ of the sorted events as individual colonies on agar plates.

We grew sorted clones individually in 96-well plates. After cell lysis, we assayed the GFP* fluorescence and PON1's primary enzymatic activity (lactonase, using 5-thiobutyl- γ -butyrolactone (TBBL) as substrate) in clarified supernatants. We calibrated the assays using the wild-type PON1-GFP* (Fig. 2c). In the drifts performed here, we considered variants exhibiting lactonase activity and GFP* fluorescence well above background ($\geq 3 \times$ s.d. of the level measured in lysates with no GFP* or PON1) positives, or 'neutral'. Comparing randomly picked samples from the library indicated $\sim 15\%$ neutral clones before sorting and $> 85\%$ neutral clones after sorting (Fig. 2c). The fraction of false positives (variants showing high fluorescence but no lactonase activity) was $\sim 10\%$ (Fig. 1c).

To continue the drift, we used ~ 300 neutral variants for the next round of mutagenesis and selection. Overall, we performed three rounds of mutagenesis, FACS and screening in 96-well plates (Table 1). This yielded a library of 494 highly polymorphic variants (7.4 ± 2.1 amino acid exchanges per gene).

Figure 2 | Library screens. (a,b) Dot plots of 5,000 representative FACS events, corresponding to single *E. coli* cells expressing wild-type PON1 (a) or PON1-GFP* library variants (b; first round of mutagenesis of the high-mutation-rate library). Pulse area and widths relate to green fluorescence (Supplementary Methods). Most wild-type events (a) fell within the R1 and R2 gates ($\sim 90\%$), in contrast to a small fraction ($\sim 3\%$) of the library events (b). The sorted cells (R1 and R2 events) were grown and subjected to a lactonase activity screen in 96-well plates. (c) Fluorescence and lactonase activity (using TBBL) in the clarified lysates of 90 randomly picked clones before (blue) and after FACS sorting (red). The observed activities were normalized to the fluorescence and lactonase activity of wild-type PON1-GFP* (inset, 35 independent repeats of wild-type PON1-GFP* analysis). Variants exhibiting relative fluorescence and enzymatic activity ≥ 0.06 were selected for the next round of neutral drift.



In cases where GFP cannot be used as a reporter, neutral drifts can be pursued by screening for enzymatic activity only. In this case, however, the mutational load must be relatively low or the number of active variants might be too small to maintain the necessary diversity. More rounds of mutagenesis and selection might also be necessary to obtain sufficient polymorphism. We accordingly drifted a low-mutation library (1 ± 0.74 amino acid exchanges per gene) and screened it in 96-well plates only. During the first two rounds of mutagenesis, the amount of positives was $\geq 80\%$ (Table 1), and we limited the screen to < 400 clones. However, as we introduced more mutations, the fraction of positives declined, and by the

fifth round we observed only 40% positives. We therefore used the fourth round products from which we isolated 500 'neutral' variants from a screen of 1,104 variants and observed considerable polymorphism (3.3 ± 1.4 amino acid exchanges per gene), although not at as many as observed under higher mutation rates.

We applied the two libraries (third generation of high-mutation-rate library and fourth generation of low-mutation-rate library; Table 1), each comprising ~ 500 'neutral' variants, as a starting point for directed evolution experiments. We screened these libraries with five different weak promiscuous substrates of PON1 that comprise candidates for directed evolution: aromatic esters (4-acetoxy-acetophenone and 3-nitrophenyl acetate), organophosphates (diethylphosphoryl-3-cyano-4-methylcoumarin (DEP-CyC) and methyl phosphonic acid 3-cyano-4-methyl-7-coumaryl cyclohexyl ester (CMP-MeCyC)), and γ -thiobutyrolactone (TBL) (Supplementary Methods). For each of these substrates, we identified ≥ 10 enzyme variants that exhibited improved total activities (v_0) and/or specific activities ($v_0/\text{fluorescence}$) (Supplementary Table 2 online). We extracted plasmids for improved clones, retransformed them and grew six individual colonies for each, lysed the cells and assayed them for both fluorescence and enzymatic activity (s.d. were $\leq 20\%$ of the average values for six

Table 1 | The creation of neutral drift libraries

Generation	Low mutation rate library		High mutation rate library (after FACS)	
	Analyzed clones	Positive clones	Analyzed clones	Positive clones
1	368	301 (81%)	368	324 (88%)
2	276	220 (80%)	276	240 (87%)
3	460	240 (52%)	644	494 ^a (77%)
4	1,104	500 ^a (45%)	–	–
5	552	221 (40%)	–	–

^aNeutral libraries used for directed evolution. –, not created.

independently grown and tested clones). In total, we identified 36 improved variants for the five substrates tested in the high-mutation-rate library and 11 improved variants in the low-mutation-rate library. These variants contained a variety of mutations, mostly distributed on PON1's surface (on average, 3.0 mutations per gene) and active site (on average, 1.3 mutations per gene; **Supplementary Table 2**). In total, we identified 59 different mutations in 18 different active-site residues (**Supplementary Fig. 2a** online). Many of the active-site mutations related to changes in substrate specificity, as indicated by the directed evolution experiments described below. Thus, repeated rounds and maintenance of sufficient diversity induced mutational tolerance and high polymorphism as observed in other neutral drift experiments^{2–4,12}.

Mutations that became enriched must also be analyzed because they are likely to be advantageous under the drift conditions and in particular under high mutation rates (**Supplementary Fig. 2b**). For example, Leu240 to Met or Ser, found in 23% of variants is likely to affect the TBBL lactonase activity; although this screening substrate (TBBL) resembles PON1's optimal lipophilic lactone substrates, it is not PON1's optimal substrate. The change of Leu16 to histidine or proline (in 17% of variants) increased the fluorescence of PON1-GFP* approximately threefold. These mutations therefore improved the stability of PON1-GFP* (and possibly of PON1 itself, as indicated by their appearance in many directly evolved variants) and thereby enabled more mutations to accumulate during the drift. However, mutations at position 16 also alter PON1's N terminus and might hamper the enzyme's physiologically relevant ability to bind lipids.

Other typically enriched mutations comprised back-to-ancestor consensus mutations (for example, F292L, A187V/T and T332S) that had been previously shown to be enriched under intense mutational drifts. These mutations increase the protein's stability, tolerance to mutations and evolvability².

To explore the use of neutral drift libraries as a starting point for directed evolution, we recombined, using DNA shuffling¹³, improved 'neutral' variants for the fluorogenic organophosphate CMP-MeCyC (21 variants; **Supplementary Table 3a** online) together with wild-type PON1 for backcrossing. We cloned the resulting library in a pET32PON1 vector (with no N-terminal GFP* and a C-terminal His tag) and screened 360 randomly selected clones for CMP-MeCyC hydrolysis. We identified several improved variants, purified and analyzed them; these variants had up to 77-fold improvement in catalytic efficiency ($k_{\text{cat}}/K_{\text{m}}$) compared to wild-type PON1 (**Supplementary Table 4** online). We identified 6 different active-site mutations that all came from

different neutral variants and recombined them by the shuffling process to generate the highly active variants.

Similarly, we shuffled 10 improved neutral variants with respect to TBL (**Supplementary Table 3b**) with wild-type PON1 and screened 360 clones from the resulting library to yield variants with up to 75-fold improvement in $k_{\text{cat}}/K_{\text{m}}$. One active-site mutation mediated an improvement in activity with this substrate (T332S; **Supplementary Table 4**).

In the case of DEPCyC, a similar procedure using 16 'neutral' variants yielded three improved variants exhibiting ≤ 14 -fold higher $k_{\text{cat}}/K_{\text{m}}$ values. The best variant contained two active-site mutations: S193T and T332S (**Supplementary Table 4**). Notably, because of a decrease in expression, the total activity of neutral variants containing the S193T mutation was lower than that of the wild-type enzyme, and this mutation therefore originated from a mutant that was selected owing to its higher specific activity (**Supplementary Table 3c**).

In summary, we demonstrated that neutral drift libraries can provide the first and most crucial step in directed evolution: the identification of different mutations that modify the enzyme's active site and improve the target function. Our most notable result was the improvement of PON1's activity with CMP-MeCyC activity of up to 44-fold, but the other four substrates tested also showed considerable variations, in both total and specific activities (4–16-fold; **Supplementary Table 2**). For each of the enzymatic functions tested, the neutral drift yielded several different active-site mutations (≥ 7) that may affect the function. The singularity and polymorphic nature of most mutations suggest that they have randomly accumulated, rather than have been selected for. The mutations were distributed throughout the active site and the highly conserved residues responsible for the primary lactonase activity remained unaffected (**Supplementary Fig. 2a**).

The utility of these small but highly polymorphic libraries was demonstrated by the directed evolution experiments. The improvements and mutational compositions we observed with CMP-MeCyC and TBL were comparable to those obtained by directed evolution of PON1 via the traditional way and using much larger libraries (10^4 – 10^7 variants)^{14,15}. The improvements with DEPCyC, however, were inferior to those previously observed¹⁵, possibly because of the dominance of S193T mutation. Overall, it appears that in comparison to the traditional manner of directed evolution, neutral drifts afford a higher diversity of mutations that can improve the enzyme's reactivity for the target substrate while screening much smaller libraries.

The key factor in making neutral drift libraries is the balance of mutation rate, screening capacity and minimal library size. We estimated the necessary minimal size library to be 200 variants, although larger repertoires are obviously beneficial (**Supplementary Note** online). To maintain this size, the mutation rate must correlate with the screening capacity: the higher the throughput, the higher the mutation rate can be. Performing the first round of mutagenesis under different mutational loads and screening samples of the resulting libraries can indicate the decay in response to increased mutational loads. As a rule, a mutation load that yields $\geq 80\%$ active variants (20% decay) after one round of mutagenesis should enable several iterative rounds to be performed while limiting the screen to $\leq 10^3$ variants. A higher throughput, for example, 10^6 variants per round screened with a GFP fusion, should allow much higher loads of mutations, and highly polymorphic libraries could be generated in fewer rounds (**Table 1**).

Note: Supplementary information is available on the Nature Methods website.

ACKNOWLEDGMENTS

Financial support by US National Institutes of Health (81XWH-07-2-0020) and Defense Threat Reduction Agency (HDTRA 1-07-C-0024) is gratefully acknowledged. We thank O. Khersonsky for substrate synthesis, D. Tal for assistance with protein purification and N. Tokuriki for help in creating the GFP constructs.

Published online at <http://www.nature.com/naturemethods/>
Reprints and permissions information is available online at
<http://npg.nature.com/reprintsandpermissions/>

1. Bershtein, S. & Tawfik, D.S. *Mol. Biol. Evol.* published online 6 August 2008 (doi:10.1093/molbev/msn174).
2. Bershtein, S. *et al. J. Mol. Biol.* **379**, 1029–1044 (2008).
3. Bloom, J.D. *et al. BMC Biol.* **5**, 29 (2007).
4. Amitai, G. *et al. HFSP J.* **1**, 67–78 (2007).
5. Waldo, G.S. *Curr. Opin. Chem. Biol.* **7**, 33–38 (2003).
6. Gerth, M.L. *et al. Protein Eng. Des. Sel.* **17**, 595–602 (2004).
7. Barakat, N.H. & Love, J.J. *Curr. Opin. Chem. Biol.* **11**, 335–341 (2007).
8. Tokuriki, N. *et al. PLOS Comput. Biol.* **4**, e1000002 (2008).
9. Wang, X. *et al. J. Mol. Biol.* **320**, 85–95 (2002).
10. Cormack, B.P. *et al. Gene* **173**, 33–38 (1996).
11. Pedelacq, J.D. *et al. Nat. Biotechnol.* **24**, 79–88 (2006).
12. Bloom, J.D. *et al. Biol. Direct* **2**, 17 (2007).
13. Raillard, S. *et al. Chem. Biol.* **8**, 891–898 (2001).
14. Aharoni, A. *et al. Chem. Biol.* **12**, 1281–1289 (2005).
15. Aharoni, A. *et al. Proc. Natl. Acad. Sci. USA* **101**, 482–487 (2004).