Biophysical Chemistry 147 (2010) 13–19

Contents lists available at ScienceDirect

Biophysical Chemistry

journal homepage: http://www.elsevier.com/locate/biophyschem

Using multi-objective computational design to extend protein promiscuity

Maria Suarez [a], Pablo Tortosa [a], Maria M. Garcia-Mira [b], David Rodríguez-Larrea [b], Raquel Godoy-Ruiz [b,1], Beatriz Ibarra-Molero [b], Jose M. Sanchez-Ruiz [b,*], Alfonso Jaramillo [a,c,*]

[a] Synth-Bio Group, Universite d'Evry Val d'Essonne-Genopole-CNRS UPS3201. Batiment Geneavenir 6. Genopole Campus 1. 5, rue Henri Desbruères. 91030 Evry Cedex, France
[b] Facultad de Ciencias, Departamento de Quimica Fisica, Universidad de Granada, 18071-Granada, Spain
[c] Ecole Polytechnique, 91128 Palaiseau, France

ARTICLE INFO

Article history:
Received 22 October 2009
Received in revised form 3 December 2009
Accepted 3 December 2009
Available online 19 January 2010

Keywords:
Protein design
Protein promiscuity
Computational chemistry
Multipurpose catalysts
Pareto Set

ABSTRACT

Many enzymes possess, besides their native function, additional promiscuous activities. Proteins with several activities (multipurpose catalysts) may have a wide range of biotechnological and biomedical applications. Natural promiscuity, however, appears to be of limited scope in this context, because the latent (promiscuous) function is often related to the evolved one (sharing the active site and even the chemical mechanism) and its enhancement upon suitable mutations usually brings about a decrease in the native activity. Here we explore the use of computational protein design to overcome these limitations. The high-plasticity positions close to the original ("native") active-site are the most promising candidates for mutations that create a second active-site associated to a new function. To avoid compromising protein folding and native activity, we propose a minimal-perturbation approach based on the combinatorial optimization of, both the *de novo* catalytic activity and the folding free-energy: essentially, we construct the Pareto Set of optimal stability/promiscuous-function solutions. We validate our approach by introducing a promiscuous esterase activity in *E. coli* thioredoxin on the basis of mutations at positions close to the native-active-site disulfide-bridge. Native oxidoreductase activity is not compromised and it is, in fact, found to be 1.5-fold enhanced, as determined by an insulin-reduction assay. This work provides general guidelines as to how computational design can be used to expand the scope and applications of protein promiscuity. From a more general viewpoint, it illustrates the potential of multi-objective optimization as the computational analogue of multi-feature natural selection.

1. Introduction

Enzymes are natural protein catalysts that can enhance reaction rates up to 23 orders of magnitude [1] with an impressive affinity and/or specificity. Their applications are enormous, and enzyme design holds the promise to provide important impact in society areas like medicine (i.e. treatment of neurodegenerative diseases [2], design of therapeutically antibodies to bind tumor-associated antigenic determinants while maintaining a small immunogenicity, development of peptide-based vaccines), biotechnology (biosensors [3,4], biocatalysts with activity in non-natural environments) and bioremediation (design of enzymes that would reduce waste by-products and toxicity [5]).

In the last few years, studies on enzyme evolvability have given convincing evidence on the mechanism of evolution of protein functions from cross-reactive proteins [6–10], starting from the observation that most proteins possess, besides their native function, additional promiscuous functions with specificities in the range of $k_{cat}/K_m \sim 10^{-2}$–10^{6} [10]. In the proposed mechanism, a weak promiscuous function arises due to neutral evolution, protein robustness (the ability of proteins to tolerate mutations without compromising fitness) and plasticity (the ability of gaining new functions by a reduced number of mutations). Under the right selection pressure, natural selection can improve the new function once it has arisen until, at some point, the protein may become specialized for the new function.

In the classic view, selection constraints on the native function were believed to be determinant in the process (if a given function appeared by natural selection, there must be a large penalty associated to the loss of that function) reducing the allowed evolution scenario to gene duplication at an early stage of the specialization. However, modern studies have shown that in most cases the coupling of the nascent and the original functions is smaller than expected (see [10] and references therein). This allows for several generations of "generalist" proteins [11] that are able to perform both functions, and suggests that gene duplication acts *after* the new function has appeared [8] and not before.

Proteins with several activities (multipurpose enzymes) may have a wide range of biotechnological applications related (but not limited)

* Corresponding authors. Sanchez-Ruiz is to be contacted at Tel.: +34 958 243189; fax: +34 958 272879. Jaramillo, Tel.: +33 1 6947 4444; fax: +33 1 6947 4437.
E-mail addresses: sanchezr@ugr.es (J.M. Sanchez-Ruiz), Alfonso.Jaramillo@polytechnique.fr (A. Jaramillo).
[1] Present address: Department of Chemistry and Biochemistry and Center for Biomolecular Structure and Organization, University of Maryland, College Park, Maryland 20742, USA.

0301-4622/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.bpc.2009.12.003

to industrial organic synthesis and metabolic engineering [12–17]. However, natural protein promiscuity (as described in the two preceding paragraphs) may perhaps be of limited use in the development of these multipurpose catalysts. First of all, natural promiscuous activities are often related to the evolved ones, sharing the same active sites and even the basic chemical mechanisms and, in general, bearing a significant resemblance to the original function [10]. Secondly, development of the promiscuous activity upon suitable mutations usually brings about a decrease in the evolved activity (see Table 2 in [10]). We explore in this work the use of computational design to overcome these limitations.

We have selected thioredoxin from *E. coli* (PDB code 2trx, 1.5 Å resolution [18]) as scaffold for our studies. It is a small (108 residues) general disulphide oxidoreductase found in all the kingdoms of living organisms; it is a common model for protein design studies because of its high stability and good expression properties [19]. We aim at introducing an esterase activity [the nucleophilic hydrolysis of the p-nitrophenyl acetate (PNPA) into p-nitrophenol (PNP) and acetate] in *E. coli* thioredoxin, following an approach similar to that used in ref [20]. Unlike these previous studies [20], however, we intend to preserve the natural thioredoxin activity (see below for details).

Natural promiscuous activities appear to be shaped by residues at the "wall and perimeter" of the native active site [10]. These residues show high-plasticity, likely because they do not belong to the protein scaffold or the native catalytic machinery [10] and provide a suitable target for the introduction of new, non-natural promiscuous activities [21,22]. However, designing a new active-site implies introduces "unsatisfied" destabilizing interactions (which will hopefully be satisfied upon ligand and/or transition-state binding). Designing a new active-site in close proximity to the native one poses the additional problem that the introduced destabilizing interactions may disrupt the native active site and affect the original activity. Our computational design approach, therefore, is based on a multi-objective optimization. Both a measure of protein stability and *de-novo* catalytic activity are simultaneously optimized by using two competing score functions for folding free energy and binding free energy of the protein-ligand (i.e. transition-state-model) complex , obtaining the Pareto Set [23] of optimal stability/promiscuous-function solutions. The goal of this procedure is the development of the new activity with the lowest possible stability cost, which has two advantages. First of all, the need to introduce additional stabilizing mutations (to compensate the destabilizing effect of the new active-site mutations) is minimized. This is an important point, since the original (native) active-site would be one obvious target for stabilizing mutations (active-site residues are optimized for function, not for stability) and, in this case, we aim at preserving the original activity. Indeed, in a previous work [20], the catalytic D26 residue was mutated to isoleucine, a change which enhances stability but which will impair the natural thioredoxin oxidoreductase activity. Secondly and most important, since our designed active-site is spatially close to the original one, the low-stability–cost strategy guarantees the minimal perturbation required for maintaining the original activity.

Our computational design approach is based upon the DESIGNER software [24,25], which optimizes protein sequence for a given target structure. This procedure uses atomic models and rotamer libraries to represent side-chain conformations. The free energies of the different models are calculated with the CHARMM force-field [26] and a free-energy solvation term proportional to the surface area [27]. Calculations on a reference state (taken to represent the unfolded state) allow a quantity akin to the unfolding free energy to be computed for each model.

The DESIGNER program was originally developed to address the inverse folding problem. However, as we show in this work, it can also be used to design new active-sites. This requires that the design is approached as an optimization in sequence-space of a protein structure which includes a model of the transition state of the chemical reaction (the tetrahedral PNPA intermediate in the case of interest here) with reference to the structure without the transition-state model bound, thus yielding a quantity akin to the free energy barrier of the reaction. Furthermore, since DESIGNER also leads to a quantity akin to the unfolding free-energy, the multi-objective stability-function optimization is indeed feasible.

2. Materials and methods

2.1. Molecular modeling of the new active site

2.1.1. General approach

We aim at introducing a promiscuous esterase activity in *E. coli* thioredoxin: the nucleophilic hydrolysis of the p-nitrophenyl acetate (PNPA) into p-nitrophenol (PNP) and acetate. The general approach we use is similar to that described in ref. [20]; i.e., we choose a histidine residue as a nuclophile for the reaction and we model the tetrahedral transition state for the reaction as PNPA-histidine structure constructed as a generalized rotamer of the histidine. An initial exploration of the available positions to place the anchor nucleophile, and their different energies lead us to fix our search to position 39. This position has additional advantages: good accessibility properties and proximity to the native thioredoxin active site [28], a high plasticity region. Also, position 39 is close to wild-type charged residues (36 K, 57 K) which could in principle be electrostatically complementary to p-nitrophenyl acetate. However, this does not represent a bias in our computational analysis, as all mutations are allowed at both positions and lysine is not favored in the rotamer construction process. All subsequent computations described in this work have been carried out on the A39H background.

The new binding site is in principle defined by the positions neighboring to 39, excluding those with wild-type proline or glycine: 1, 4, 26, 28, 30, 36, 38, 42, 43, 46, 55, 57, 59 and 99, which are allowed to mutate to any amino acid except proline, cysteine and methionine (see Fig. 1A). All other positions were forced to conserve their wild-type amino acid but were free to change rotamer to avoid border effects. No previous stabilizing mutation was introduced. This gave a total of 15 designed positions and 93 fixed amino acid positions, yielding a sequence space of 10^{18}.

2.1.2. Side-chain rotamer library enhancement for tetrahedral transition-state modeling

We do not introduce in the force field biasing extra terms in order to obtain interactions known to favor the reaction. On the other hand, we have enhanced the rotamer library to improve the modeling of protein-ligand (transition state model) interactions. We have used the backbone-dependent rotamer library by Dunbrack and Karplus [29] including all rotamers with probability $P>90\%$. The rotamer of the combined histidine ($N\delta$ as nucleophile) and tetrahedral intermediate was modeled with the same initial dihedral angles as in [20]: χ_1 and χ_2 from the histidine in the library, χ_3 (± 30, ± 90, ± 150), χ_4 (± -60, ± 180), and χ_5 (± 90). To provide a better docking of the ligand to thioredoxin, we have added a refined version of the standard library following the protocol described in [30]. This procedure refines the library specifically in the ligand-protein interface by searching only for conformations that provide a better interaction and improve ligand recognition. For each pair of rotamers involved in binding we first minimize the structure (100 iterations of steepest descent) of the interacting rotamers. A low-dielectric environment ($\varepsilon=1$) in CHARMM was found to significatively speed-up calculations by biasing the search towards H-bond formation. During minimization, backbone charges are set to zero to restrict the optimization to rotamer–rotamer interactions. In a second step, we minimize the resulting structure (100 steps of ABNR, $\varepsilon=8$) to obtain the refined conformation of both residues, which are stored as new rotamers for further use. To simplify the size of the calculations, we required a maximum interacting energy of -5.0 kcal/

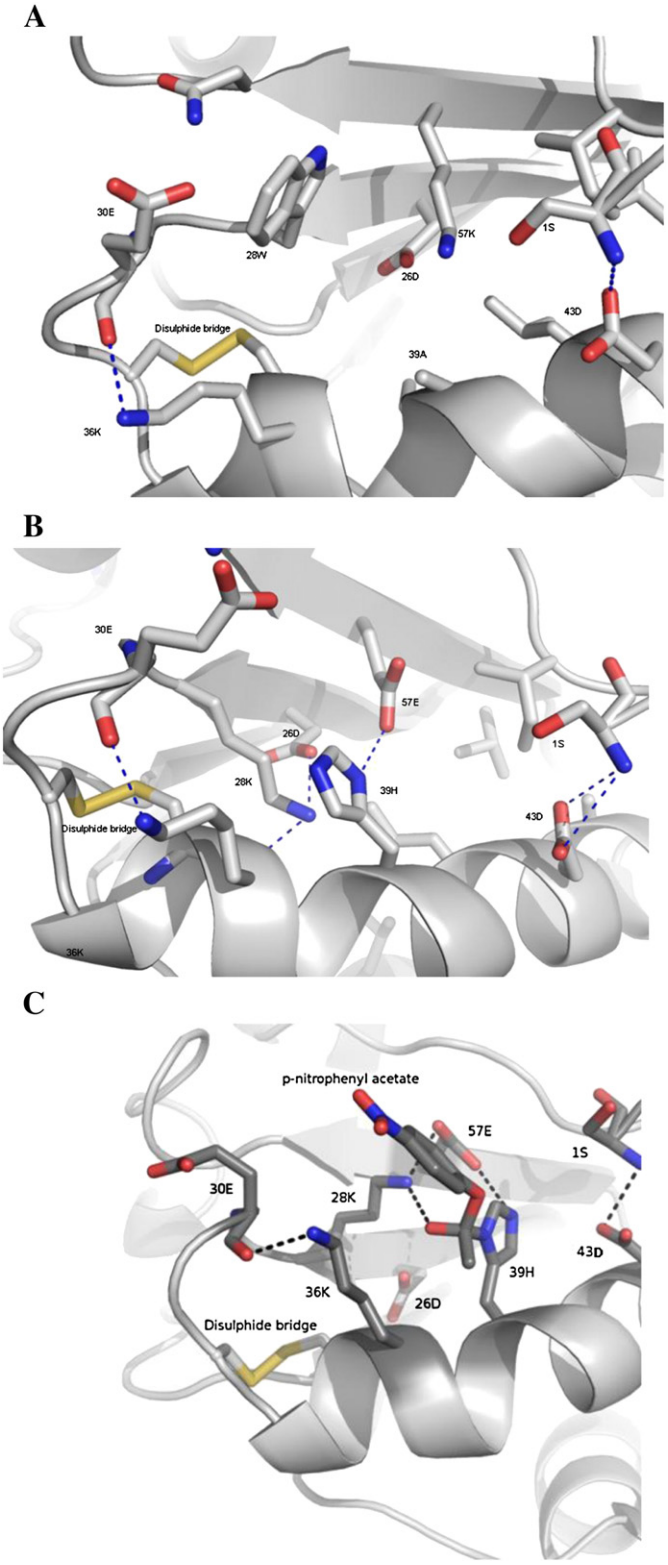

Fig. 1. Structural analysis of the design of a promiscuous activity in thioredoxin. A) Structure of the wild-type thioredoxin active site (C32-C35 disulfide bridge and D26). Wild type residues considered in the design procedure are shown. The entrance to the native active site is from the other side of the central β sheet. B) Structure of the designed thioredoxin active site for esterase activity in the absence of the substrate, PNPA. The designed mutations are W28K, A39H and K57E (V3 variant). Relevant interactions (1S-43D, 28K-26D, 30E-36K, 39H-57E) are shown. C) Detail of the modeling of the PNPA transition state in the V3 mutant. Relevant interactions are shown: binding interactions with the ligand (36 K, 28 K) and stabilizing interactions (1S-43D, 28K-57E, 30E-36K, 39H-57E). Note that the native function active-site interacts with the promiscuous function active site.

mol and performed a clustering in dihedral angle with a granularity factor of 10°. This resulted in a total number of 2886 rotamers and a size of the combinatorial space of 10^{95} possible structures.

2.1.3. Computational analyses

We performed the initial exploration of the available positions to place the anchor nucleophile which included the enhancement of the library described above and resulted in a CPU time consumption of 8000 h in processors Power4 and Power4+ (1.3 GHz and 1.7 GHz) using the facilities provided by IDRIS supercomputer center. Once we fixed the anchor nucleophile position (H39) and performed the clustering in dihedral angle of the resulting library, we proceeded to recalculate the interaction energies of the remaining 2886 rotamers, (2000 CPU hours provided by the BSC supercomputer center: Power PC 970MP processors at 2.3 GHz). The optimization procedure took 1000 CPU hours at BSC.

2.2. Site-directed mutagenesis, protein expression, purification and sample preparation

Oligonucleotides used for mutagenesis were purchased from Invitrogen Corp. Mutations were introduced in the sequence of wild-type thioredoxin from *E. coli*, previously cloned into pET-30a, using the QuikChange (TM) Site-Directed Mutagenesis method from Stratagene©. The V3 triple mutant (W28K/A39H/K57E) was over expressed in BL21 (DE3) pLys during 4 h. Cells were harvested and resuspended in 30 mM Tris, 1 mM EDTA, pH 8.3 (TE buffer), lysed by sonication and centrifuged at 10,000 × *g* for 10 min. Nucleic acids in the soluble fraction were precipitated overnight by addition of a streptomycin sulphate 10% solution (Sigma-Aldrich) prepared in TE buffer and removed by centrifugation at 100,000 × *g* for 30 min. The protein was purified by size exclusion chromatography and anion exchange chromatography as previously described [31]. The purity of the protein was checked by SDS-PAGE electrophoresis. Samples were prepared by exhaustive dialysis against 10 mM sodium phosphate, pH 6.95 buffer. Protein concentration was determined by UV absorbance ($\varepsilon_{280} = 8310\ M^{-1}\ cm^{-1}$).

2.3. Activity measurements

Two kinds of activity measurements were carried out in the present study. The nucleophilic hydrolysis of p-nitrophenyl acetate (PNPA) into p-nitrophenol (PNP) was performed as previously described in [20] with minor modifications. All experiments were carried out at 22 °C. Stock solutions of 75 mM PNPA were freshly prepared in acetonitrile by weight. The reaction was initiated by the addition of PNPA to a solution of thioredoxin previously tempered. PNP formation was followed by absorbance at 400.5 nm ($\varepsilon_{400.5} = 12400\ M^{-1}\ cm^{-1}$). In all cases, kinetics were measured against PNPA solutions in buffer, used as a blank to take into account the spontaneous hydrolysis of the reagent. Initial rates were calculated from the slope of the time-dependence of the absorbance and data were analyzed by non-linear least-squares fitting of the Michaelis–Menten equation (see Fig. 3) using SigmaPlot (SPSS Inc.). Additionally, the natural oxidoreductase activity was determined using a turbidimetric assay consisting of the thioredoxin-catalyzed rate of reduction of insulin disulfides by dithiothreitol [32]. Reaction mixtures contained thioredoxin (up to 6–8 μM) in potassium phosphate 0.1 M, EDTA 2 mM pH 6.5 and insulin 0.5 mg/mL. Reaction was initiated by addition of DTT, 1 mM final concentration, previously dissolved in water. The increase in the A_{650} due to insulin aggregation upon thioredoxin mediated reduction was monitored for 1 h at 25 °C (see Fig. 4). Activities were calculated from the slopes of the maximum dA_{650}/dt versus thioredoxin concentration. In all cases, absorbance measurements were corrected for the value given by a sample lacking thioredoxin and DTT.

2.4. Stability measurements

The thermal unfolding transitions were followed by differential scanning calorimetry of V3 against dialysis buffer at a scan rate of 90 $K \cdot h^{-1}$. A VP-DSC microcalorimeter (Microcal Inc.) was used. Rescans of V3 were done to check the reversibility of the transition. Experimental data were analyzed on the basis of the two-state model as previously described [33].

3. Results and discussion

3.1. Multi-objective optimization and construction of the Pareto Set

We approach enzyme design as a two-objective problem. We consider as one objective the effect on stability of the designed region, which is coupled to the primary function of thioredoxin due to spatial proximity. It is estimated from an approximation to the folding free energy as defined previously [24] and does not involve the presence of the PNPA tetrahedral intermediate. The other objective is a quantity akin to the activation free energy of the reaction which is estimated from calculations on a protein structure which includes a model of the transition state of the chemical reaction (the tetrahedral PNPA intermediate in the case of interest here) with reference to the structure without the transition-state model bound. Note that the values of K_m and k_{cat} are not specifically addressed by this procedure, since they are given by complex expressions including the equilibrium and rate constants for the several steps of the catalytic mechanism: see Fig. 3 in ref [20].

To rank the sequences by simultaneously using two competing scores (akin to the folding free energy and the binding free energy of the transition-state complex) we use a method previously presented [34], based on standard techniques in computer science for multi-objective optimization. In a single objective protein design problem, the goal is to find the solution associated with the lowest cost function (the global minimum). However, in a multi-objective optimization problem there is no single optimal solution, but a set of optimal solutions, each of them representing a different trade-off between the objectives. To compare two different solutions, a relation of dominance is defined as follows: *A* dominates *B* if, *for all objectives, A* has a better performance than *B*; as a result *A* has to be kept while *B* has to be discarded (even though B may lead to an active foldable protein). The set of all non-dominated solutions is called Pareto globally optimal set (or, simply, the Pareto Set) and finding this set is the goal of the multi-objective computational optimization problem [23]. Note that, in the Pareto Set, only those solutions showing an optimal trade-off between the different objectives are kept. As a result, the complexity of the problem is dramatically reduced without loss of relevant information (see illustration in Fig. 2A).

In order to construct the Pareto Set, we have followed the procedure we have recently described in detail [34]. Briefly, we have used a protocol of Monte Carlo-Simulated Annealing (MCSA) cycles [35,36], combined with an exponential cooling schedule (initial and final temperature $T_0 = 1$ R kcal/mol and $T_f = 0.01$ R kcal/mol) to minimize a weighted sum of the scores corresponding to the two objectives:

$$\Delta G^{\text{scalar}}(\lambda) = \lambda \cdot \Delta G^{\text{stability}} + (1-\lambda) \cdot \Delta G^{\text{newfunction}} \quad (1)$$

where $\Delta G^{\text{stability}}$ and $\Delta G^{\text{newfunction}}$ are the computational measures akin to the free-energies of folding and binding of the transition-state model, respectively. The weight, λ, was allowed to take many different values in the range [0.05, 0.95] to ensure the exploration of different design regions. The rationale behind this procedure is that each of the different λ values leads to a given point belonging to the Pareto Set of solutions (although neighboring λ values may lead to the same point). However, the use of λ values different from 0.5 produces solutions in a sub-optimal conformation for that sequence, so an additional minimization by an unbiased MCSA (λ = 0.5) was performed for each of the obtained sequences to ensure they reach their minimum energy conformation. The Pareto Set of non-dominated solutions thus obtained is the result of the two-objective optimization problem and is shown in Fig. 2B. Note that the trade-off between the two competing objectives is clearly apparent in the Pareto Set, the most stable sequences showing poor binding and the best binding sequences being highly unstable.

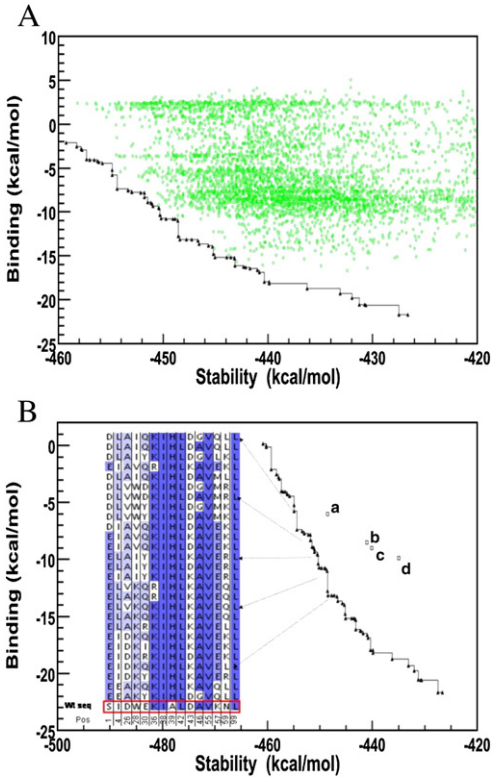

Fig. 2. Multi-objective computational optimization and construction of the Pareto Set. A) Illustration of the Pareto Set for a stability/binding multi-objective optimization. Each dot represents a given sequence. The Pareto Set of non-dominated solutions is shown in black. B) Pareto Set of sequences corresponding to the design of esterase activity on a thioredoxin scaffold. The stability corresponds to a measure akin to the folding free energy of the enzyme, while binding corresponds to a measure akin to the free energy of interaction of the enzyme with the tetrahedral intermediate of the reaction. Points a, b, c and d correspond to the V3 mutant considering each of the four possible conformers. The alignment of the sequences with binding energy in the (−10, −24) kcal/mol is shown (note that only those residues considered in the design are shown).

3.2. Design of a thioredoxin variant with a promiscuous activity

The Pareto Set provides an ordered and non-redundant list of solutions, each of them with a different trade-off between the chosen objectives. We have examined this list in the intermediate stability/binding range (see Fig. 2B) to identify relevant mutations for a three-mutation candidate for experimental validation. A consensus analysis immediately suggests mutations S1E, W28K and K57E (see alignment shown as Inset in Fig. 2B). However, the Pareto Set shown in Fig. 2 has been constructed on the basis of the average of the four possible conformers associated to the A39H mutation and the subsequent side-chain modeling (i.e., two different chiralities of the acetate and two possible ways of histidine attack). A more detailed examination reveals that the performance of the mutation S1E depends strongly of the conformer associated to the A39H mutation, while mutations W28K and K57E are more robust in this regard. These two mutations, together with the nucleophile insertion (A39H) were modeled with DESIGNER and seen to be reasonably close to the Pareto Set (see Fig. 2B). We therefore chose this triple mutant variant (subsequently referred to as V3) for experimental validation. The structure predicted

by DESIGNER for our mutant V3 is shown in Fig. 1C, with our three designed mutations (W28K, A39H and K57E) together with surrounding wild-type residues. It illustrates the balance between binding interactions with the ligand (36 K, 28 K) and stabilizing interactions with neighboring residues (1S-43D, 28K-57E, 30E-36K, 39H-57E).

3.3. Native and promiscuous activities of the designed thioredoxin variant

We have determined the promiscuous esterase activity by following the hydrolysis of PNPA into PNP catalyzed by our designed thioredoxin variant, V3, at 22 °C in sodium phosphate buffer 10 mM, pH 6.9. We carried out the experiments using several concentrations of V3 as well as PNPA. The formation rate of PNP was found to be linear over a wide time interval. The reaction initial rates measured at different enzyme concentration values have been represented in Fig. 3A. No deviations from the expected linear dependences with the enzyme concentration were found. The dependence of the reaction rate on PNPA concentration is consistent with a simple Michaelis–Menten mechanism (see Fig. 3B) with $K_m = 1.198 \cdot 10^{-3}$ M and $k_{cat} = 3.69 \cdot 10^{-4}\ s^{-1}$ (for the uncatalyzed reaction $k_{uncat} = 2.5 \cdot 10^{-6}\ s^{-1}$, ref. [20]). It must be noted that the values of K_m and k_{cat} are given by complex expressions including the equilibrium and rate constants for the several steps of the catalytic mechanism: see Fig. 3 in ref [20]. As expected, wild type thioredoxin does not show any significant esterase activity.

In addition, the native oxidoreductase activity of thioredoxin and V3 was experimentally determined (Fig. 4) by following the enzymatic-catalyzed reduction of insulin by DTT as described by Holmgren [32]. Reduced insulin aggregates with a concomitant increase in solution turbidity that is reflected in the increase in absorbance at 650 nm observed after a certain lag-time (Fig. 4A). Activities were calculated as the maximum of the dA_{650}/dt slopes. These values change linearly with protein concentration (in the 0.2–3 μM concentration range). V3 is not only found to have native reductase activity but, remarkably, it shows a 1.5-fold enhancement in native activity with respect to wild type thioredoxin (Fig. 4B). This enhancement is also evident in the lag-time values, which are significantly reduced in the case of V3 (Fig. 4C). Recent single-molecule force-clamp spectroscopy studies [37] have revealed two alternative forms of the catalytic mechanism of *E. coli* thioredoxin catalysis, the first requiring a re-orientation of the substrate disulfide bond and the second elongating the substrate disulfide bond. Given this complexity in the catalytic process, it does not seem possible at this stage to propose a reliable molecular interpretation of mutation effects on bulk-phase activity (such as the 1.5-fold enhancement observed with V3). The important point in the context of this work is, however, that the introduction of the promiscuous esterase activity in thioredoxin has not compromised the native oxidoreductase activity. This result does support that our multi-objective stability/binding design is a minimal-perturbation approach, less likely to lead to mutations affecting the original function, even when the high-plasticity positions close to the original active-site are used to create the second active-site associated to the new function (as we have done in this work).

3.4. The balance between stability and promiscuous function in the designed thioredoxin variant

The thermal stability of V3 was determined by differential scanning calorimetry (DSC) at pH 7.0 in 10 mM phosphate buffer (see Fig. 3C). The calorimetric transition was found to be reversible and well-described by the two-state reversible model. Denaturation temperature (71.8 °C) is significantly lower than that for wild type thioredoxin under the same solvent conditions: 86.6 °C [33]. Unfolding enthalpy is also found to be lower: 71 kcal/mol in V3 versus 114 kcal/mol in wild type thioredoxin. However, these unfolding enthalpy values belong to the corresponding denaturation temperatures and, therefore, a significant part of the difference is to be attributed to the temperature dependence of the unfolding enthalpy.

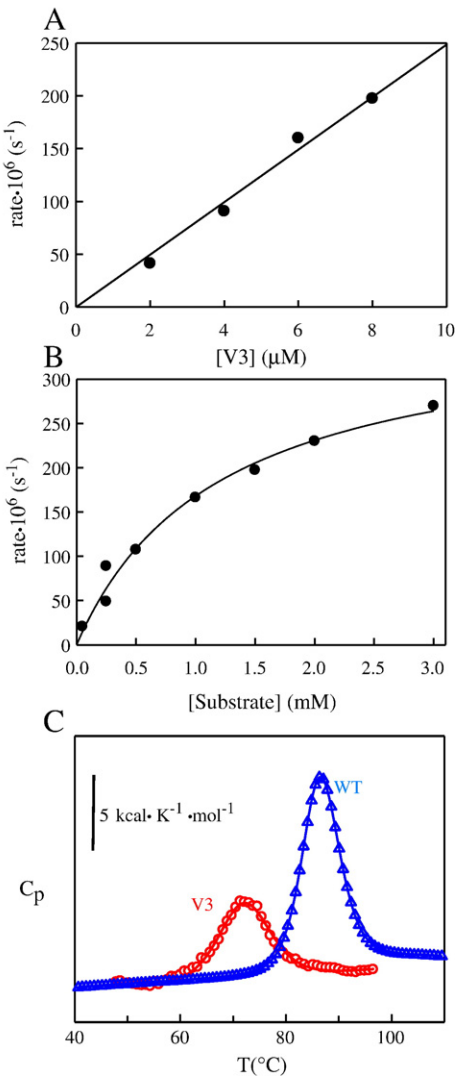

Fig. 3. Promiscuous activity and stability of the designed V3 variant of thioredoxin. A) Hydrolysis rates of PNPA at different concentrations of the thioredoxin variant V3. All measurements were carried out at an initial substrate concentration of 1.5 mM. B) Dependence of the initial rates of hydrolysis of PNPA as a function of the substrate concentration. The filled circles are the experimental values obtained using a protein concentration of 8 μM and the continuous line is the best fit of the Michaelis–Menten equation. C) Differential scanning calorimetry thermograms for wild-type thioredoxin and the V3 variant at pH 7.0, 10 mM phosphate buffer. The symbols stand for the heat capacity data and the continuous lines for the best fits of the two-state equilibrium model.

Creating a new active-site in a protein involves introducing potential interactions that will only be satisfied upon substrate (or transition-state) binding. These unsatisfied interactions will have a destabilizing effect in the non-ligated protein. However, for a design based on the Pareto Set of optimal stability/binding solutions, the stability decrease observed should correspond to the achieved binding affinity (reflecting an optimal trade-off between the competing objectives). We show below that this correspondence does qualitatively hold for the designed esterase activity in the V3 variant of thioredoxin.

Consider the binding of a ligand to a protein. The binding free energy is related to dissociation constant (K_d) through:

$$\Delta G^0_{BINDING} = -RT \cdot \ln\left({}^{1}/_{K_d}\right) \quad (2)$$

where the superscript ^{0} indicates a standard state free-energy change. Standard state is defined by the pressure and concentration units associated to K_d [38]. As is customary, we use K_d in molar units for

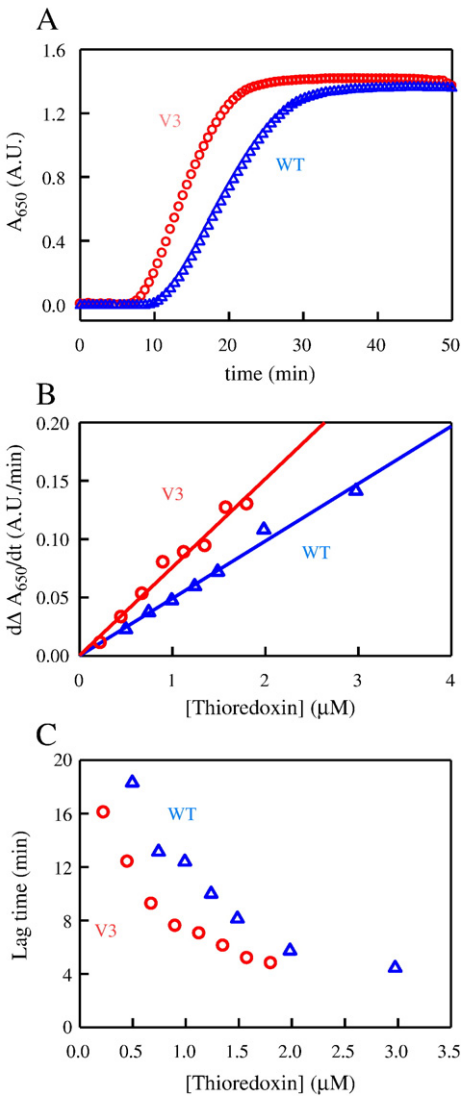

Fig. 4. Native oxidoreductase activity of the wild-type form and the designed V3 variant of thioredoxin as determined from insulin-reduction assays. A) Representative plots of absorbance at 650 nm versus time monitoring the aggregation resulting from the reduction of insulin by DTT catalyzed by thioredoxin. The concentration of the wild-type and V3 forms was 1.5 μM. B) Plot of oxidoreductase activity versus enzyme concentration for the wild-type and V3 forms of thioredoxin. Activities are calculated as the maximum values of the slopes of plots of absorbance at 650 nm versus time, such as those shown in panel A. C) Plot of lag-time in the insulin-reduction assays (see panel A) versus enzyme concentration for the wild-type and V3 forms of thioredoxin.

free-energy calculations and, therefore, our standard concentration state is 1 mol.

The binding free-energy can be split into two contributions, corresponding to the newly formed interactions in the protein-ligand complex (ΔG_{INT}) and to the loss of translational entropy associated to reduction in the number of particles in solution upon binding (ΔG^0_{TR}):

$$\Delta G^0_{BINDING} = \Delta G_{INT} + \Delta G^0_{TR} \quad (3)$$

where only the translational contribution depends on the standard-state concentration unit. Freire, Amzel and coworkers [38,39] have provided evidence that translational free-energies can be calculated to an acceptable approximation on the basis of Kauzmann's cratic entropies [40]:

$$\Delta G^0_{TR} = -T \cdot \Delta S_{CRAT} \quad (4)$$

where, for a P + L → PL process and 1 mol standard state, ΔS_{CRAT} is -8 cal·K^{-1}·mol^{-1}.

Eqs. (2)–(4) allow us to estimate the free-energy associated to protein-ligand interactions (ΔG_{INT}) from the experimental value of the dissociation constant (K_d). Thus, using as K_d the Michaelis constant value determined from the fitting shown in Fig. 3B ($K_m = 1.2 \cdot 10^{-3}$ M) we obtain -6.4 kcal/mol as an estimate of the free energy associated to the PNPA-V3 interactions. We assume now that those interactions are not satisfied in the non-ligated V3 variant and, therefore, that they contribute to its destabilization with respect to wild type thioredoxin. Schellman and coworkers [41] showed many years ago that, to a first approximation, the effect of a free energy perturbation (g) on equilibrium denaturation temperature can be obtained from:

$$\Delta T_m = \frac{g \cdot T_m}{\Delta H_m} \quad (5)$$

where T_m is the "non-perturbed" denaturation temperature value and ΔH_m is the non-perturbed denaturation enthalpy at the temperature T_m. For thioredoxin under the solvent conditions of the stability determinations reported here, $T_m = 86.6$ °C = 359.8 K and $\Delta H_m = 114$ kcal/mol. Then, using -6.4 kcal/mol as the value of the free-energy perturbation, we obtain from Eq. (5) that $\Delta T_m = -20°$. This value is in excellent qualitative agreement with the fact that the denaturation temperature for V3 is 15° lower than that for wild type thioredoxin (see Fig. 3B).

The above calculation is meant mainly for illustration. Thus for instance, it neglects unfolded-state effects and it has been carried using the PNPA binding free energy derived from the Michaelis constant value, which may in fact be given by a complex expression including equilibrium and rate constants for the several steps of the catalytic mechanism (see Fig. 3 in ref. [20]). Furthermore, it would have been desirable to perform the calculation on the basis of the activation free-energy value (i.e. the free energy value for transition-state "binding"). This, however, would have required the assumption of a value for the front-factor in transition-state-theory rate equation, which would have added considerable uncertainty to the calculation. Caveats notwithstanding, the calculation suggests in a clear manner an important general point, which is elaborated below:

We may expect a clear difference between re-designing a native active site for a new function [42] and introducing a second active-site for the new function. In the first case, destabilizing interactions consistent with the new function are substituted for those corresponding to the original function. In the second case, the destabilizing interactions associated to the new-function are introduced without eliminating those involved in the original "native" active site (which is to remain functional) and, therefore, creating a second active site will very likely cause a significant decrease in protein stability which may compromise folding. Note that, according to the Schellman equation (Eq. (5)) a free-energy perturbation of the order of several kcal/mol (required for order-of-magnitude enhancement in binding affinity or catalytic rate) translates into a decrease in denaturizing temperature of several tens of degrees. Certainly, design of protein promiscuity should benefit from previous stability enhancement but, as illustrated by this work, it also requires the minimization of the associated stability penalty by using an efficient computational multi-objective methodology that selects the optimal trade-off solutions.

3.5. Concluding remarks

During evolution, many different protein features may simultaneously be under selective pressure; these include properties related to stability, catalysis, interactions, solubility, prevention of aggregation, etcetera. Another example of the interplay between multiple features in evolution is the development of a new function from a pre-existing weak promiscuous function; in this case, to avoid gene duplication at early stages of the development of the new activity, the

protein has to maintain both activities during a certain evolutionary stage.

In contrast, computational protein design usually employs a single-objective scoring function to rank the different sequences in the optimization procedure. However, the need to satisfy multiple objectives in protein engineering has always been realized, but rarely exploited, perhaps because of the lack of general methodological guidelines to the multi-feature optimization problem. This work makes proposals regarding such guidelines and provides the relevant experimental validation. In particular, this work shows with a specific experimental scenario how the computational analysis of several competing objectives can be posed to lead to a set of optimal solutions (not a single global minimum), with each possible criterion to choose a specific solution representing a necessary trade-off between the competing objectives.

From a different viewpoint, multi-objective analyses, such as that described here, can offer valuable insight on the diversity of the solution space, the choices that natural evolution confronts when selecting enzymes under different pressures and the role of enzyme promiscuity in shaping evolution.

Acknowledgments

P.T. acknowledges an EMBO long-term fellowship. A.J. acknowledges support from HPC-EUROPA2 (project 228398) and the use of the BSC and IDRIS supercomputer facilities to perform the calculations reported here. This research was supported by Feder Funds and Grant BIO2006-07332 (Spanish Ministry of Education and Science) to J.M.S.-R. Grant CVI-1668 (Junta de Andalucía) to B.I.-M and Grants BioModularH2 (FP6-NEST-043340), ATIGE (Genopole), TARPOL (FP7-KBBE-212894) to A.J.

References

[1] D.A. Kraut, K.S. Carrol, D. Herschlag, Challenges in enzyme mechanism and energetics, Annu. Rev. Biochem. 72 (2003) 517–571.
[2] L. Lecanu, V. Papadopoulos, Cutting-edge patents in Alzheimer's disease drug discovery: anticipation of potential future treatments, Recent Patents on CNS Drug Discovery 2 (2007)113–123.
[3] J.S. Marvin, E.E. Corcoran, N.A. Hattangadi, J.V. Zhang, S.A. Gere, H.W. Hellinga, The rational design of allosteric interactions in a monomeric protein and its applications to the construction of biosensors, Proc. Natl. Acad. Sci. U. S. A. 94 (1997) 4366–4371.
[4] L.L. Looger, M.A. Dwyer, J.J. Smith, H.W. Hellinga, Computational design of receptor and sensor proteins with novel functions, Nature 423 (2003) 185–190.
[5] E.L. Anga, H. Zhaoa, J.P. Obrad, Recent advances in the bioremediation of persistent organic pollutants via biomolecular engineering, Enzyme Microb. Technol. 37 (2005) 487–496.
[6] R.A. Jensen, Enzyme recruitment in evolution of new function, Annu. Rev. Microbiol. 30 (1976) 409–425.
[7] P.J. O'Brien, D. Herschlag, Catalytic promiscuity and the evolution of new enzymatic activities chemistry and biology, Chem. Biol. 6 (1999) R91–R105.
[8] A. Aharoni, L. Gaidukov, O. Khersonsky, M.S. Gould, C. Roodveldt, D.S. Tawfik, The 'evolvability' of promiscuous protein functions, Nat. Genet. 37 (2005) 73–76.
[9] S.D. Copley, Enzymes with extra talents: moonlighting functions and catalytic promiscuity, Curr. Opin. Chem. Biol. 7 (2003) 265–272.
[10] O. Khersonsky, C. Roodveldt, D.S. Tawfik, Enzyme promiscuity: evolutionary and mechanistic aspects, Curr. Opin. Chem. Biol. 10 (2006) 498–508.
[11] S.C. Rothman, J.F. Kirsch, How does an enzyme evolved in vitro compare to naturally occurring homologs possessing the targeted function? Tyrosine aminotransferase from aspartate aminotransferase, J. Mol. Biol. 327 (2003) 593–608.
[12] C. Jürgens, A. Strom, D. Wegener, S. Hettwer, M. Wilmanns, R. Sterner, Directed evolution of a ($\beta\alpha$)8-barrel enzyme to catalyze related reactions in two different metabolic pathways, Proc. Natl. Acad. Sci. U. S. A. 97 (2000) 9925–9930.
[13] R.J. Kazlauskas, Enhancing catalytic promiscuity for biocatalysis, Curr. Opin. Chem. Biol. 9 (2005) 195–201.
[14] K.A. Canada, S. Iwashita, S. Shim, T.K. Wood, Directed evolution of toluene ortho-monooxygenase for enhanced 1-naphthol synthesis and chlorinated ethene degradation, J. Bacteriol. 184 (2002) 344–349.
[15] W. Zheng, K.A. Scheibner, K.A. Ho, P.A. Cole, Mechanistic studies on the alkyltransferase activity of serotonin N-acetyl transferase, Chem. Biol. 8 (2001) 379–389.
[16] R. Spreitzer, M.E. Salvucci, RUBISCO: structure, regulatory interactions, and possibilities for a better enzyme, Annu. Rev. Plant Biol. 53 (2002) 449–475.
[17] U.T. Bornscheuer, R.J. Kazlauskas, Catalytic promiscuity in biocatalysis: using old enzymes to form new bonds and follow new pathways, Angew. Chem. Int. Ed. Engl. 43 (2004) 6032–6040.
[18] S.K. Katti, D.M. Le Master, H. Eklund, Crystal structure of thioredoxin from *Escherichia coli* at 1.68 Å resolution, J. Mol. Biol. 212 (1990) 167–184.
[19] J.E. Ladbury, R. Wynn, H.W. Hellinga, J.M. Sturtevant, Stability of oxidized *Escherichia coli* thioredoxin and its dependence on protonation of the aspartic acid residue in the 26 position, Biochemistry 32 (1993) 7526–7530.
[20] D.N. Bolon, S.L. Mayo, Enzyme-like proteins by computational design, Proc. Natl. Acad. Sci. 98 (2001) 14274–14279.
[21] Y. Yoshikumi, T.E. Ferrin, J.D. Keasling, Designed divergent evolution of enzyme function, Nature 440 (2006) 1078–1082.
[22] R. Bone, J.L. Silen, D.A. Agard, Structural plasticity broadens the specificity of an engineered protease, Nature 339 (1989) 191–195.
[23] P.Y. Papalambros, D.J. Wilde, Principles of optimal design, Cambridge University Press, New York, 2000.
[24] A. Jaramillo, L. Wernisch, S. Hery, S.J. Wodak, Folding free energy function selects native-like protein sequences in the core but not on the surface, Proc. Natl. Acad. Sci. U. S. A. 99 (2002) 13554–13559.
[25] L. Wernisch, S. Hery, S.J. Wodak, Automatic protein design with all atom force-fields by exact and heuristic optimization, J. Mol. Biol. 301 (2000) 713–736.
[26] A.D. MacKerell, D. Bashford Jr., R.L. Bellott, R.L. Dunbrack, J.D. Evanseck Jr., M.J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F.T.K. Lau, C. Mattos, S. Michnick, T. Ngo, D.T. Nguyen, B. Prodhom, W.E.I.I.I. Reiher, B. Roux, M. Schlenkrich, J.C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin, M. Karplus, All-atom empirical potential for molecular modeling and dynamics studies of proteins, J. Phys. Chem. B 102 (1998) 3586–3616.
[27] T. Ooi, M. Oobatake, G. Nemethy, H.A. Scheraga, Accessible surface areas as a measure of the thermodynamic parameters of hydration of peptides, Proc. Natl. Acad. Sci. U. S. A. 84 (1987) 3086–3090.
[28] S. Vohnik, C. Hanson, R. Tuma, J.A. Fuchs, C. Woodward, G.J. Thomas, Conformation, stability, and active-site cysteine titrations of *Escherichia coli* D26A thioredoxin probed by Raman spectroscopy, Protein Sci. 7 (1998) 193–200.
[29] R.L. Dunbrack Jr., M. Karplus, Backbone-dependent rotamer library for proteins: application to side-chain prediction, J. Mol. Biol. 230 (1993) 543–574.
[30] P. Tortosa, A. Jaramillo, Active sites by computational protein design, AIP Conf. Proc. 851 (2006) 96–101.
[31] R. Perez-Jimenez, R. Godoy-Ruiz, B. Ibarra-Molero, J.M. Sanchez-Ruiz, The efficiency of different salts to screen charge interactions in proteins: a Hofmeister effect? Biophys. J. 86 (2004) 2414–2429.
[32] A. Holmgren, Thioredoxin catalyzes the reduction of insulin disulfides by dithiothreitol and dihydrolipoamide, J. Biol. Chem. 254 (1979) 9627–9632.
[33] R.E. Georgescu, M.M. Garcia-Mira, M.L. Tasayco, J.M. Sanchez-Ruiz, Heat capacity analysis of oxidized *Escherichia coli* thioredoxin fragments (1–73, 74–108) and their noncovalent complex. Evidence for the burial of apolar surface in protein unfolded states, Eur. J. Biochem. 5 (2001) 1477–1485.
[34] M. Suarez, P. Tortosa, J. Carrera, A. Jaramillo, Pareto optimization in computational protein design with multiples objectives, J. Comput. Chem. 29 (2008) 2704–2711.
[35] V. Gardiner, J.G. Hoffman, N. Metropolis, Digital computer studies of cell multiplication by Monte Carlo methods, J. Natl. Cancer Inst. 17 (1956) 175–188.
[36] S. Kirkpatrick, C.D. Gelatt, M.P. Vecchi, Optimization by simulated annealing, Science 220 (1983) 671–680.
[37] A.P. Wiita, R. Perez-Jimenez, K.A. Walter, F. Gräter, B.J. Berne, A. Holmgren, J.M. Sanchez-Ruiz, J.M. Fernandez, Probing the chemistry of thioredoxin catalysis with force, Nature 450 (2007) 124–127.
[38] K.P. Murphy, D. Xie, K.S. Thompson, M.L. Amzel, E. Freire, Entropy in biological binding processes: estimation of translational entropy loss, Proteins 18 (1994) 63–67.
[39] X. Siebert, L.M. Amzel, Loss of translational entropy in molecular associations, Proteins 54 (2004) 104–115.
[40] W. Kauzmann, Some factors in the interpretation of protein denaturation, Adv. Protein Chem. 14 (1959) 1–63.
[41] W.J. Becktel, J.A. Schellman, Protein stability curves, Biopolymers 26 (1987) 1859–1877.
[42] M. Allert, S.S. Rizk, L.L. Looger, H.W. Hellinga, Computational design of receptors for an organophosphate surrogate of the nerve agent soman, Proc. Natl. Acad. Sci. U. S. A. 101 (2004) 7907–7912.