

Adaptive landscape flattening allows the design of both enzyme: Substrate binding and catalytic power

Vaitea Opuu, Giuliano Nigro, Thomas Gaillard, Emmanuelle Schmitt, Yves Mechulam, Thomas Simonson

▶ **To cite this version:**

Vaitea Opuu, Giuliano Nigro, Thomas Gaillard, Emmanuelle Schmitt, Yves Mechulam, et al.. Adaptive landscape flattening allows the design of both enzyme: Substrate binding and catalytic power. PLoS Computational Biology, 2020, 16 (1), pp.e1007600. 10.1371/journal.pcbi.1007600 . hal-02493715

HAL Id: hal-02493715

https://polytechnique.hal.science/hal-02493715

Submitted on 27 Nov 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

PLOS COMPUTATIONAL BIOLOGY

RESEARCH ARTICLE

Adaptive landscape flattening allows the design of both enzyme: Substrate binding and catalytic power

Vaitea Opuu, Giuliano Nigro, Thomas Gaillard, Emmanuelle Schmitt, Yves Mechulam, Thomas Simonson*

Laboratoire de Biochimie (CNRS UMR7654), Ecole Polytechnique, Palaiseau, France

* thomas.simonson@polytechnique.fr

Check for updates

OPEN ACCESS

Citation: Opuu V, Nigro G, Gaillard T, Schmitt E, Mechulam Y, Simonson T (2020) Adaptive landscape flattening allows the design of both enzyme: Substrate binding and catalytic power. PLoS Comput Biol 16(1): e1007600. https://doi.org/10.1371/journal.pcbi.1007600

Editor: Alexey Onufriev, Virginia Tech, UNITED STATES

Received: September 20, 2019

Accepted: December 11, 2019

Published: January 9, 2020

Copyright: © 2020 Opuu et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the manuscript and its Supporting Information files.

Funding: The authors received no specific funding for this work.

Competing interests: The authors have declared that no competing interests exist.

Abstract

Designed enzymes are of fundamental and technological interest. Experimental directed evolution still has significant limitations, and computational approaches are a complementary route. A designed enzyme should satisfy multiple criteria: stability, substrate binding, transition state binding. Such multi-objective design is computationally challenging. Two recent studies used adaptive importance sampling Monte Carlo to redesign proteins for ligand binding. By first flattening the energy landscape of the apo protein, they obtained positive design for the bound state and negative design for the unbound. We have now extended the method to design an enzyme for specific transition state binding, *i.e.*, for its catalytic power. We considered methionyl-tRNA synthetase (MetRS), which attaches methionine (Met) to its cognate tRNA, establishing codon identity. Previously, MetRS and other synthetases have been redesigned by experimental directed evolution to accept noncanonical amino acids as substrates, leading to genetic code expansion. Here, we have redesigned MetRS computationally to bind several ligands: the Met analog azidonorleucine, methionyl-adenylate (MetAMP), and the activated ligands that form the transition state for MetAMP production. Enzyme mutants known to have azidonorleucine activity were recovered by the design calculations, and 17 mutants predicted to bind MetAMP were characterized experimentally and all found to be active. Mutants predicted to have low activation free energies for MetAMP production were found to be active and the predicted reaction rates agreed well with the experimental values. We suggest the present method should become the paradigm for computational enzyme design.

Author summary

Designed enzymes are of major interest. Experimental directed evolution still has significant limitations, and computational approaches are another route. Enzymes must be stable, bind substrates, and be powerful catalysts. It is challenging to design for all these properties. A method to design substrate binding was proposed recently. It used an adaptive Monte Carlo method to explore mutations of a few amino acids near the substrate. A

bias energy was gradually "learned" such that, in the absence of the ligand, the simulation visited most of the possible protein mutations with comparable probabilities. Remarkably, a simulation of the protein:ligand complex, including the bias, will then preferentially sample tight-binding sequences. We generalized the method to design binding *specificity*. We tested it for the methionyl-tRNA synthetase enzyme, which has been engineered in order to expand the genetic code. We redesigned the enzyme to obtain variants with low activation free energies for the catalytic step. The variants proposed by the simulations were shown experimentally to be active, and the predicted activation free energies were in reasonable agreement with the experimental values. We expect the new method will become the paradigm for computational enzyme design.

Introduction

One of the most important challenges in computational protein design (CPD) is to modify a protein so that it will bind a given ligand [1–4]. This is essential for problems like enzyme design, biosensor design, and building tailored protein assemblies. To design ligand binding means optimizing a free energy difference between bound and unbound states. This two-state optimization is not directly tractable by the most common CPD methods, such as simulated annealing, plain Monte Carlo (MC), or simple branch-and-bound and dead end elimination methods [4, 5]. Rather, most studies have used either heuristic methods that optimize the *bound state* energy [1–4, 6], or enumeration methods that are rigorous but expensive and explore a limited free energy range [7–10].

Recently, a new approach was proposed, using Monte Carlo simulation and importance sampling. The energy landscape in sequence space is flattened adaptively over the course of a simulation, thanks to a bias potential [11]. Flattening can be done for the bound state, the unbound state, or both [12]. Remarkably, this leads to a situation where sequence variants are sampled according to a Boltzmann distribution controlled by the *binding free energy*, exactly the quantity we want to select for. Several variations have been employed, including one that used molecular dynamics instead of MC [13]. The method allows sequences to be designed for binding affinity, but also binding *specificity*. This is especially important for enzyme design, since catalytic power is directly related to the enzyme's ability to preferentially stabilize the transition state [14]. We apply the method here to an enzyme of biological and technological importance, methionyl-tRNA synthetase (MetRS). We demonstrate that the method can be used to design an enzyme for its catalytic power.

Each aminoacyl-tRNA synthetase (aaRS) attaches a specific amino acid to a tRNA that carries the corresponding anticodon, establishing the genetic code [15]. Two reactions are catalyzed. In the first, the amino acid reacts with ATP to give aaAMP and pyrophosphate. In the second, tRNA reacts with aaAMP. For MetRS, the first reaction does not require tRNA. Several aaRSs have been engineered experimentally to bind noncanonical amino acids (ncAAs) [16–20]. Obtaining an aaRS that binds an ncAA and uses it as a substrate is a key step to allow the ncAA to become part of an expanded genetic code [17, 20, 21]. The ncAA can then be genetically encoded and incorporated into proteins by the cellular machinery. Several MetRS variants that accepted the ncAA azidonorleucine (AnL) as a substrate were obtained earlier by experimental directed evolution [22]. The AnL azide group can be used for protein labeling and imaging.

The design procedure has two stages. First, a bias potential is optimized adaptively over the course of a MC simulation of the apo protein. The adaptation method is closely analogous to

the Wang-Landau and metadynamics approaches [23, 24]. The bias is chosen so that all the allowed residue types achieve comparable probabilities at all mutating positions. This implies that the free energy landscape in sequence space has been flattened, and the bias of each sequence is approximately the opposite of its apo free energy. In the second stage, the holo state is simulated. The bias is included in the energy function, "subtracting out" the apo free energy. Thus, the method achieves positive design for the bound state and negative design for the unbound. The sequences sampled in the second stage are distributed according to their binding free energies, with tight binders exponentially enriched.

In an analogous procedure, a bias potential can be optimized for the protein bound to one ligand, say L. Then a complex with another ligand is simulated, say, L', including the bias. The sequences sampled preferentially in the second simulation are those with a strong binding free energy difference between the two ligands, *i.e.*, the most *specific* binders. Importantly, L' can be an activated, transition state ligand, while L is the non-activated substrate. In this case, the first simulation flattens the ground state landscape, while the second preferentially samples sequences that stabilize the transition state, relative to the ground state. Thus, the method can be used to select directly for low activation free energies. It is then straightforward to rank the sampled sequences based on their catalytic efficiency, the ratio between the rate constant for the catalytic step, k_{cat} and the Michaelis constant K_M.

Here, we report CPD calculations that aim to increase the binding of several ligands by MetRS. We first considered AnL. Three residues in the active site were allowed to mutate. The CPD method was tested for its ability to recover the known experimental variants [22]. The top six experimental variants were visited by the MC simulations and were highly ranked among the predicted sequences. We next considered the natural ligand methionyl adenylate (MetAMP). Another set of three residues near the ligand side chain were allowed to mutate. The wildtype sequence was highly ranked by the computational design. 17 other sequences among the top 40 predictions were tested experimentally and all found to be active. The computed binding free energy differences between variants were mostly in good agreement with the experimental values, obtained from kinetic measurements of the enzyme reaction. Next, we predicted MetRS variants that were specifically designed to bind the transition state for the enzymatic reaction Met+ATP → MetAMP+PP_i. The wildtype enzyme was highly ranked among 5832 possible variants, and for 20 variants that were characterized experimentally, the transition state binding free energies from the simulations were in good agreement with the values deduced from the experimental reaction rates. These calculations represent the first time an enzyme is specifically designed to optimize its transition state binding free energy relative to ground state binding, *i.e.*, its catalytic power. We expect the method will become the paradigm for computational design of enzymes.

Materials and methods

Theoretical approach: Designing for ligand binding

Stage 1: Adaptive apo simulation. We consider a polypeptide, with or without a bound ligand. Below, we will use a fixed backbone geometry, but the method is valid with a flexible backbone. Side chains can explore a few discrete conformations, or rotamers, and a few selected positions are allowed to mutate. In a first stage, we perform a MC exploration of the protein with no ligand, using the usual Metropolis-Hastings scheme [25–27]. We gradually increment a bias potential until all the side chain types at the mutating positions have roughly equal populations, thus flattening the free energy landscape. We number the mutating positions arbitrarily

1, . . ., p. The bias E^B at time t has the form:

$$E^B(s_1(t), s_2(t), \ldots, s_p(t); t) = \sum_i E_i^B(s_i(t); t) + \sum_{i<j} E_{ij}^B(s_i(t), s_j(t); t) \tag{1}$$

Here, $s_i(t)$ represents the side chain type at position i. The first sum is over single amino acid positions; the second is over pairs. The individual terms are updated at regular intervals of length T. At each update, whichever sequence variant ($s_1(t)$, $s_2(t)$, . . ., $s_p(t)$) is populated is penalized by adding an increment $e_i^B(s_i(t); t)$ or $e_{ij}^B(s_i(t), s_j(t); t)$ to each corresponding term in the bias. The increments have the form:

$$e_i^B(s_i(t); t) = e_0 \exp\left[-E_i^B(s_i(t); t)/E_0\right] \tag{2}$$

$$e_{ij}^B(s_i(t), s_j(t); t) = e_0 \exp\left[-E_{ij}^B(s_i(t), s_j(t); t)/E_0\right] \tag{3}$$

where e_0 and E_0 are constant energies. Thus, the increments decrease exponentially as the bias increases. This scheme is adapted from well-tempered metadynamics [24, 28, 29]. The individual bias terms depend on the system history, and can be written:

$$E_i^B(s; t) = \sum_{n; nT<t} e_i^B(s; nT)\delta_{s,s_i(nT)} \tag{4}$$

$$E_{ij}^B(s, s'; t) = \sum_{n; nT<t} e_{ij}^B(s, s'; nT)\delta_{s,s_i(nT)}\delta_{s',s_j(nT)} \tag{5}$$

where $\delta_{a,b}$ is the Kronecker delta. Over time, the bias for the most probable states grows until it pushes the system into other regions of sequence space. Two-position biases were implemented in the Proteus software [30, 31] during this work.

Stage 2: Biased holo simulation. In the second stage, the protein:ligand complex is simulated using the bias potential from stage 1. The sampled population of a sequence S is normalized to give a probability, denoted $\tilde{p}_H(S)$, where the subscript means "holo" and the tilde indicates that the bias is present. The apo state probability $\tilde{p}_A(S)$ was obtained in stage 1. Both probabilities can be converted into free energies $\tilde{G}$:

$$\begin{aligned} \tilde{p}_X(S) &= \frac{1}{Z_X} \exp\left(-\tilde{G}_X(S)/kT\right) \\ \tilde{G}_X(S) &= -kT \ln \tilde{p}_X(S) - kT \ln Z_X \end{aligned} \tag{6}$$

where $X = A$ or H and Z_X is a normalization factor that depends on X but not S. We also have a relation between the free energies with and without the bias:

$$\tilde{G}_X(S) = G_X(S) + E^B(S) \tag{7}$$

whose (straightforward) derivation is given in the Supporting Appendix (S1 File). Note that if the apo state flattening were ideal, $\tilde{p}_A(S)$ would be a constant, so that (from Eqs 6 and 7) $E^B(S) = -G_A(S)$, up to an additive constant. Thus, the ideal bias is the opposite of the apo free energy.

The binding free energy relative to a reference sequence R can be deduced from the populations. We have:

$$\begin{aligned} \Delta\Delta\tilde{G}(S) &= \left(\tilde{G}_H(S) - \tilde{G}_A(S)\right) - \left(\tilde{G}_H(R) - \tilde{G}_A(R)\right) \\ &= -kT \ln \frac{\tilde{p}_H(S)}{\tilde{p}_H(R)} + kT \ln \frac{\tilde{p}_A(S)}{\tilde{p}_A(R)} \\ &= \Delta\Delta G(S) \end{aligned} \tag{8}$$

Since the bias is the same in the bound and unbound states, it cancels out from $\Delta\Delta\tilde{G}(S)$, which is equal to the relative binding free energy *in the absence of bias*, $\Delta\Delta G(S)$. While the bias does not appear explicitly in (8), it is essential for accurate sampling. Perfect flattening, however, is not usually achieved, nor is it needed.

In the holo state, the probability of a sequence S (with bias) is:

$$\tilde{p}_H(S) \propto \exp\left(-\frac{\tilde{G}_H(S)}{kT}\right) = \exp\left(-\frac{G_H(S) + E^B(S)}{kT}\right) \approx \exp\left(-\frac{G_H(S) - G_A(S)}{kT}\right) \tag{9}$$

Thus, holo sampling follows a Boltzmann distribution governed by $G_H(s) + E^B(S)$, which is approximately the binding free energy $G_H(S) - G_A(S)$. This is exactly the quantity we want to design for. If the apo state is well-flattened, the biased holo simulation will be exponentially enriched in tight binders.

Energy function and matrix

The energy was computed using either an MMGBLK or an MMGBSA function ("molecular mechanics + Generalized Born + Lazaridis-Karplus" or "Surface Area"):

$$E = E_{\mathrm{MM}} + E_{\mathrm{GB}} + E_{\mathrm{LK|SA}} \tag{10}$$

The MM term used the Amber ff99SB force field [30, 32]. The SA term was described earlier [33–35]. The LK term and its parameterization were described earlier [35]. The GB term corresponds to a variant very similar to the one used in Amber, detailed in previous articles [33, 36, 37]. To make the calculation efficient, we compared two strategies. The first used a Native Environment Approximation (NEA), where the GB solvation radii for a given side chain were computed with the rest of the system in its native conformation [36, 38]. The second used a "Fluctuating Dielectric Boundary" (FDB) method, where the GB interaction between two residues I, J was expressed as a polynomial function of their solvation radii [39]. These were kept up to date over the course of the MC simulation, so the GB interaction could be deduced with little additional calculation [37, 39]. The solvent dielectric constant was 80; the protein one was 4.0 with the GBSA variants and 6.8 with GBLK [35]. Each solvent model is referred to by its GB variant and nonpolar term; for example, the FDBLK model combines FDB with LK.

To allow very fast MC simulations, we precomputed an energy matrix for each system [34, 40]. For each pair of residues I, J and all their allowed types and rotamers, we performed a short energy minimization (15 conjugate gradient steps) [30]. The backbone was fixed (in its crystal geometry) and the energy only included interactions between the two side chains and with the backbone. At the end of the minimization, we computed the interaction energy between the two side chains. Side chain–backbone interaction energies were computed similarly (and formed the matrix diagonal) [30].

Structural models

MetRS:AnL and MetRS:MetAMP complexes. For MetRS:AnL, we started from the crystal structure of a complex between a triple mutant of *E. coli* MetRS and AnL (PDB code 3H9B) [41]. The protein mutations were L13S, Y260L, H301L. We refer to this mutant as SLL. The protein backbone was held fixed. Side chains more than 20 Å from the ligand were held fixed. The other side chains were allowed to explore rotamers, taken from the Tuffery library, augmented to allow multiple orientations for certain hydrogen atoms [42, 43]. Side chains 13 and 301 were allowed to mutate into the following 14 types: ACDEHIKLMNQSTV; position 260 was allowed to mutate into the same types, except that Tyr replaced Asp. Thus, there were $14^3 = 2744$ possible sequences in all. Histidine protonation states at non-mutating positions were assigned by visual inspection of the 3D structure. System preparation was done using the protX module of the Proteus design software [31].

For MetRS:MetAMP, we started from a crystal complex (PDB code 1PG0) between *E. coli* MetRS and a methionyl adenylate (MetAMP) analogue [44]. The protein backbone was held fixed. Side chains more than 20 Å from the ligand were held fixed. The other side chains were allowed to explore rotamers [42, 43]. Side chains 13, 256 and 297 were allowed to mutate into all types except Gly or Pro, for a total of 5832 possible sequences in all. Histidine protonation states at non-mutating positions were assigned by visual inspection of the 3D structure.

Unfolded state. The unfolded state energy was estimated with a tri-peptide model [45]. For each mutating position, side chain type, and rotamer, we computed the interaction between the side chain and the tri-peptide it forms with the two adjacent backbone and C_β groups. Then, for each allowed type, we computed the energy of the best rotamer and averaged over mutating positions. The mean energy for each type was taken to be its contribution to the unfolded state energy. The contributions of the mutating positions were summed to give the total unfolded energy.

Ligand force field and rotamers

Force field. For the AnL azido group, we used atomic charges and van der Waals parameters obtained earlier for azidophenylalanine [46]. Parameters for the implicit solvent energy terms were assigned by analogy to existing groups. For methionyl adenylate (MetAMP), we mostly used existing Met and AMP parameters. For atoms close to the Met:AMP junction, we used atomic charges computed earlier for ThrAMP (G. Monard, personal communication) from *ab initio* quantum chemistry, in a manner consistent with the rest of the Amber force field [32]. Van der Waals parameters for atoms near the junction were assigned the same types as in Met or AMP. Parameters for bond lengths, angles and dihedrals involving junction atoms were taken from the experimental geometry of MetAMP. Stiffness parameters were assigned by analogy to existing parameters. The complete set of parameters for AnL and MetAMP is in Supplementary Material (S2 and S3 Files, respectively).

Rotamers. AnL was positioned in the protein complex so that its backbone had the position occupied in the MetRS:AnL crystal structure [41]. The ligand's side chain was allowed to explore rotamers. These were defined by the usual side chain rotamers of Met [42, 43]. We started by positioning Met in the pocket by superimposing it on AnL in the mutant MetRS:AnL crystal complex (PDB code 3H9B). We then positioned the 17 Met side chain rotamers from the Tuffery library. We extracted the AnL side chain from the experimental complex and superimposed it on each of the 17 Met rotamers, producing 17 AnL conformers. Finally, for each one, we performed a short energy minimization with the AnL backbone held fixed. The 17 minimized conformers defined the AnL rotamers. Notice that with this procedure, the azido group always had the same orientation relative to the aliphatic part of the AnL side

chain. For MetAMP, we allowed the Met rotamers from the Tuffery library, with the rest of the ligand held fixed. The ϕ and ψ dihedral angles around the MetAMP C_α were not allowed to rotate and the whole AMP moiety stayed fixed.

Modeling the MetRS transition state complex

MetRS catalyzes two reactions. In the first, Met reacts with ATP to give MetAMP and pyrophosphate. In the second, tRNA reacts with MetAMP. Here, we considered the first reaction, which occurs in the absence of tRNA. A model for the ground state ligands Met + ATP was first obtained, starting from the crystal complex between MetAMP and PP_i (PDB code 3KFL). The covalent structure was reset to that of Met + ATP and the geometry was adjusted by a short energy minimization. The complex included a magnesium ion. Next, a model for the activated ligand $[Met:ATP]^{\ddagger}$ was obtained, starting from the Met + ATP complex. First, a phosphate and carboxylate fragment were positioned in a geometry close to the expected pentacoordinate transition state arrangement [44, 47–49] and an *ab initio* energy minimization was done, including planarity constraints for the phosphorus and three oxygens. This led to a length of 2.4 Å for the P–O bonds perpendicular to the plane. Next, the molecular mechanics model was constructed. A covalent bond was introduced between the reacting Met carboxylate oxygen and the α phosphorus atom. The lengths for this bond and the symmetric one on the other side of the phosphorus were set to 2.4 Å. Planarity restraints were imposed on the phosphorus and the three α phosphate oxygens. A short energy minimization was done (with molecular mechanics). This led to an α phosphate geometry with three oxygens in plane and two perpendicular (Fig 1), as expected for in-line attack of the Met carboxylate on the

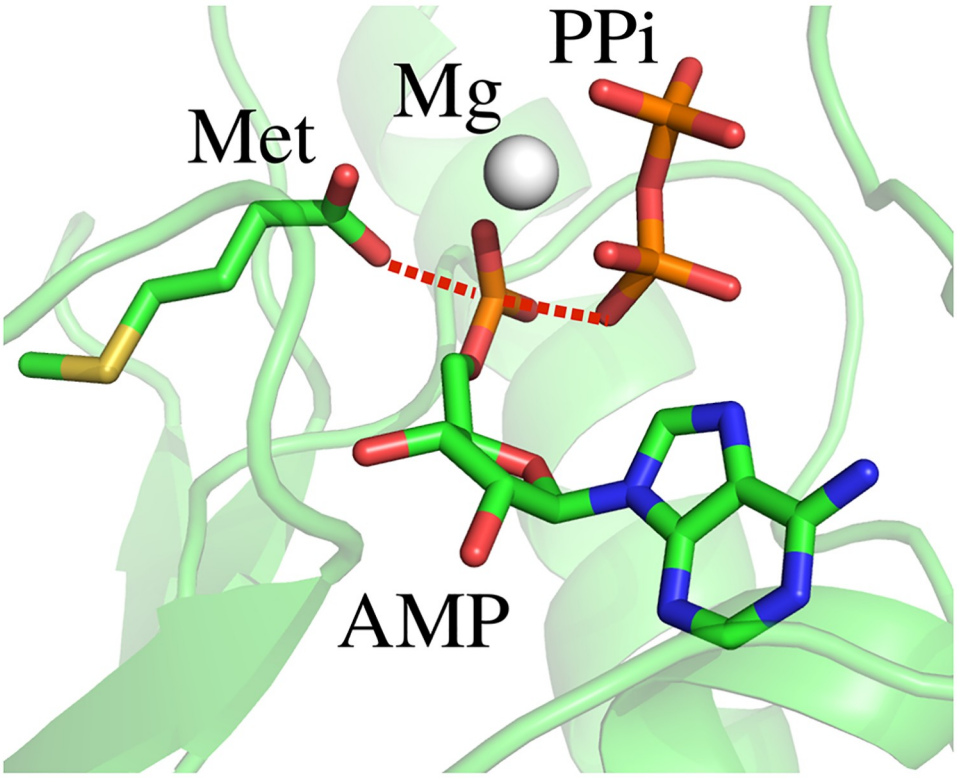


Fig 1. MetRS transition state for MetAMP formation. Closeup of the ligands.

https://doi.org/10.1371/journal.pcbi.1007600.g001

phosphate [44, 47–49]. *Ab initio* atomic charges were then computed for the entire activated ligand in this geometry, from a Merz-Kollman population analysis of the HF/6-31G* wavefunction [32], using Gaussian 9.0. The magnesium ion, which bridges the α, β and γ phosphates, was included in the calculation. The resulting charges were applied to atoms close to the α phosphate group, while other atoms kept their usual Met or ATP charges. Small manual adjustments were made to establish the correct total charge of -4. The final Mg charge was +1.5. Charges are in Supplementary Material (S3 File).

The geometry of the protein around the ligands was relaxed slightly by performing a short, restrained molecular dynamics simulation, with the ligands held fixed. The entire system was placed in a large box of explicit TIP3P water [50]. Harmonic restraints were applied to nonhydrogen atoms, with force constants that decreased gradually from 5 to 0.5 kcal/mol/Å^2 over 575 ps of dynamics, performed with the NAMD program [51]. The final protein geometry was used for the design calculations.

Monte Carlo simulations

To optimize the bias potential, we performed MC simulations of the apo state with bias updates every T = 1000 steps, with e_0 = 0.2 kcal/mol and E_0 = 50 kcal/mol [12]. During the first 10^8 MC steps, we optimized a bias potential including only single-position terms. There were p = 3 mutating positions, which all contributed to the bias. In the second stage, we ran MC or (in one case: MetAMP complex with the FDBSA solvent model) Replica Exchange MC (REMC) simulations of 5.10^8 MC steps [27], using 8 replicas with thermal energies (kcal/mol) of 0.17, 0.26, 0.39, 0.59, 0.88, 1.33, 2.0 and 3.0. Temperature swaps were attempted every 500 steps. All the replicas experienced the same bias potential. Both stages used 1- and 2-position moves.

Experimental mutagenesis and kinetic assays

Purification of wildtype and mutant MetRS. Throughout this study, we used a His-tagged M547 monomeric version of *E. coli* MetRS, fully active, both in vitro and in vivo [41]. The gene encoding M547 MetRS from pBSM547+ [52, 53] was subcloned into pET15blpa [54] to overproduce the His-tagged enzyme in *E. coli* ([55]). Site-directed mutations were generated using the QuickChange method [56], and the whole mutated genes verified by DNA sequencing. The enzyme and its variants were produced in BLR(DE3) *E. coli* cells. Transformed cells were grown overnight at 37˚C in 0.25 L of TBAI autoinducible medium containing 50 μg/ml ampicillin. They were harvested by centrifugation and resuspended in 20 ml of buffer A (10 mM Hepes-HCl pH 7.0, 3 mM 2-mercaptoethanol, 500 mM NaCl). They were disrupted by sonication (5 min, 0˚C), and debris was removed by centrifugation (15,300 G, 15 min). The supernatant was applied on a Talon affinity column (10 ml; Clontech) equilibrated in buffer A. The column was washed with buffer A plus 10 mM imidazole and eluted with 125 mM imidazole in buffer A. Fractions containing tagged MetRS were pooled and diluted ten-fold in 10 mM Hepes-HCl pH 7.0, 10 mM 2-mercaptoethanol (buffer B). These solutions were applied on an ion exchange column Q Hiload (16 mL, GE-Healthcare), equilibrated in buffer B containing 50 mM NaCl. The column was washed with buffer B and eluted with a linear gradient from 5 to 500 mM NaCl in buffer B (2 ml/min, 10 mM/min). Fractions containing tagged MetRS were pooled, dialyzed against a 10 mM Hepes-HCl buffer (pH 7.0) containing 55% glycerol, and stored at -20˚C. The homogeneity of the purified MetRS was estimated by SDS-PAGE to be higher than 95%.

Measurement of ATP-PPi exchange activity. Prior to activity measurements, MetRS was diluted in a 20 mM Tris-HCl buffer (pH 7.6) containing 0.2 mg/ml bovine serum albumin

(Aldrich) if the concentration after dilution was less than 1 μM. Initial rates of ATP-PPi exchange activity were measured at 25˚C as described [57]. In brief, the 100 μl reaction mixture contained Tris-HCl (20 mM, pH 7.6), MgCl2 (7 mM), ATP (2 mM), [^{32}P]PPi (1800-3700 Bq, 2 mM) and various concentrations (0-16 mM) of the Met amino acid. The exchange reaction was started by adding catalytic amounts of MetRS (20 μl). After quenching the reaction, ^{32}P-labeled ATP was adsorbed on charcoal, filtered, and measured by scintillation counting. kcat and K_M values were derived from iterative nonlinear fits of the theoretical equation to the experimental values using either MC-fit [58] or Origin (Origin Lab).

Results

Designing MetRS to bind azidonorleucine

As a first test, we searched for MetRS variants with strong azidonorleucine (AnL) binding. Positions 13, 260 and 301 were allowed to mutate, for comparison to the earlier experimental data [22]. 14 types were allowed at each position (see Methods), for a total of 2744 possible sequences. We compared three variants of the solvent model, which gave similar results. The first stage was to optimize a bias potential that flattened the free energy landscape in sequence space for apo MetRS. We used a bias potential including single-position terms only. After the adaptation period, we ran a further simulation of 10^8 MC steps to determine the biased populations. With the FDBLK solvent model, 2099 sequences were visited at least 1000 times, thanks to the adaptive bias. The second stage was to simulate the MetRS:AnL complex in the presence of the bias. 1957 sequences were visited at least 1000 times in both the first and second stages. For these, we used the sampled populations to deduce the AnL binding free energy (Eq 8), relative to the X-ray sequence. The overall computation time for system setup, energy matrix precalculation and both MC stages was about one day (per solvent model). Sequences sampled with and without the bias and ligand are shown in Fig 2 as logos.

Experimental directed evolution had revealed 21 active variants [22]. 13 of them were sampled by the computations and are listed in Table 1. 8 others either were not sampled or were predicted to have low stability. Each variant is referred to by the sequence of the three mutating positions; for example, the X-ray variant is SLL. The top six experimental sequences are the ones that were observed in multiple clones. The others were seen in just one clone [22]. The top six were all sampled by the computations and had good predicted stabilities and affinities (Table 1). SML was ranked the highest, 17th. The five others had lower ranks, between 45 and 104, but they were all within 1.4 kcal/mol of the top predicted variant (which was HMS). Other predicted variants may also be active, even though they were not revealed by the directed evolution experiments. For the SLL variant, the predicted rotamers for binding site residues were in good agreement with the X-ray structure (Supplementary Material; S1 File). The results in Table 1 were obtained with the FDBLK solvent treatment. The FDBSA solvent model gave similar results, while NEASA was slightly poorer (not shown), probably due to its simpler GB treatment [37].

We also searched for MetRS variants that maximized the AnL binding *specificity*, relative to Met. A bias potential was adaptively optimized for the MetRS:Met complex, then used in a simulation of the MetRS:AnL complex. The mutating positions and allowed types were the same as above. Specificity ranks are included in Table 1. Three of the top six experimental variants had high specificity ranks. The top experimental variant NLL was 36th, the next-best experimental variant SLL was 2nd, AQL was 18th, and CLL was 3rd. Thus, among the top 40 specificity ranks, there were 4 sequences that are known to be active. Evidently, selecting for specificity can help reveal active variants.

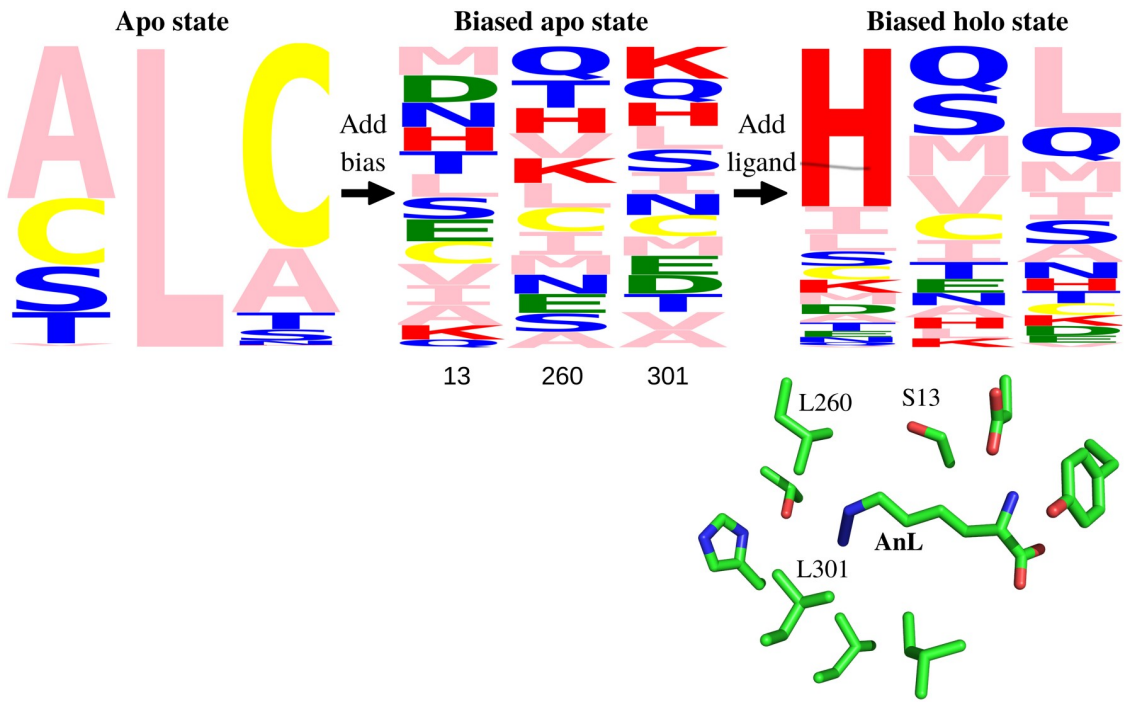


Fig 2. MetRS sequence logos. Sequences sampled without and with the AnL ligand (FDBLK solvent model) are shown in the form of logos, including the three mutating positions, 13, 260, 301. The logos represent the apo state (left), the biased apo state (middle), and the biased holo state (right). The height of each letter measures the frequency of its type. The 3D view below is a closeup of azidonorleucine (AnL) in the binding pocket, with selected side chains.

https://doi.org/10.1371/journal.pcbi.1007600.g002

Table 1. MetRS redesigned for AnL binding affinity or specificity.

a seq.	*b* pop.	*c* fold	*d* bind	*e* rank	*f* spec.	*g* rank	*a* seq.	*b* pop.	*c* fold	*d* bind	*f* rank	*e* spec.	*g* rank
NLL	62	6.7	0.3	104	5.7	36	CVL	1	6.9	-0.4	23	10.9	164
SLL	12	0.0	0.0	55	0.0	2	ACL	1	5.0	0.2	86	11.0	175
SML	4	4.6	-0.5	17	8.3	74	SCM	1	-0.9	0.4	123	18.8	589
AVL	3	6.7	-0.1	45	11.0	165	SLV	1	-2.3	1.4	688	7.4	57
AQL	2	4.2	0.0	57	3.3	18	SNL[h]	1	7.6	0.0	–	10.2	–
CLL	2	-0.6	0.1	73	1.0	3	SSL[h]	1	7.2	-0.1	–	10.3	–
							STL[h]	1	7.2	0.6	–	10.2	–

[a]Sequence at the designed positions 13, 260, 301, ranked by

[b]population among the experimental clones.

[c]Folding and

[d]binding free energies (kcal/mol) relative to the X-ray sequence SLL.

[e]Rank based on affinity or

[g]specificity.

[f]Specificity, defined by the binding free energy difference between AnL and Met (relative to SLL).

[h]Not ranked, since folding free energy is above the 7 kcal/mol threshold.

Calculations used the FDBLK solvent.

https://doi.org/10.1371/journal.pcbi.1007600.t001

Redesigning MetRS to bind MetAMP

As a second test, we searched for MetRS variants with a high affinity for the natural ligand methionyl adenylate (MetAMP). These should include the wildtype (WT) sequence and close homologs. Three positions close to the Met side chain (Fig 3), 13, 256, and 297 were allowed to mutate into all types except Gly or Pro, for a total of 5832 possible sequences. The first stage was to optimize a bias potential that flattened the free energy landscape in sequence space for apo MetRS. We performed calculations with both the FDBSA and the FDBLK variants of the solvent model, which gave similar results. We report the FDBSA results, since they were obtained first and were the basis for choosing which sequences to test experimentally. Selected FDBLK results are also reported. With the FDBSA solvent model, using Replica Exchange MC, 4178 variants were visited at least 1000 times.

The second stage was to simulate the MetRS:MetAMP complex in the presence of the bias potential. For sequences visited at least 1000 times in both stages (528 sequences), we used the sampled populations to deduce the MetAMP binding free energy (Eq 8), relative to the wild-type (WT) sequence LAI. The folding energy of each variant was also estimated (see Methods) and sequences less stable than WT by 5 kcal/mol or more were discarded. The top 20 remaining sequences, with the largest binding free energies, are shown in Table 2. The top sequence, CDV, had an Asp at position 256, positioned to form a salt bridge with the MetAMP ammonium group. Its binding free energy, relative to WT, was -1.4 kcal/mol. The next 19 variants had types similar to WT. Their computed binding free energies were close to WT, with relative values between -0.2 and 0.6 kcal/mol. The WT sequence was sixth overall. Among the top 40 variants, 17 mutants were produced experimentally. They were representative of the computational variants, while providing ease of construction (see Methods). CDV was left out, as the A256D mutation, selected for binding, might reduce the catalytic activity. All 17 tested variants had detectable activity, a 100% success rate for the design procedure. One other sequence, SAI, was tested experimentally and found to be active, but did not show up in the MC simulation. Thus the method produced one false negative along with 17 true positives.

Going further, we made a quantitative comparison between the computed and experimental binding free energies. The experimental dissociation constants were estimated from the Michaelis constants K_M. In the experimental conditions (excess ATP) and under the usual Michaelis-Menten assumptions [14, 59], K_M represents the dissociation constant for Met binding in the presence of bound ATP. Here, we computed relative binding free energies for binding MetAMP, not Met. Nevertheless, we expect that these MetAMP binding free energy changes can be compared to the experimental Met binding free energy changes; *i.e.*, we make the additional assumption that the relative effects of the mutations will be conserved going from MetAMP to Met+ATP.

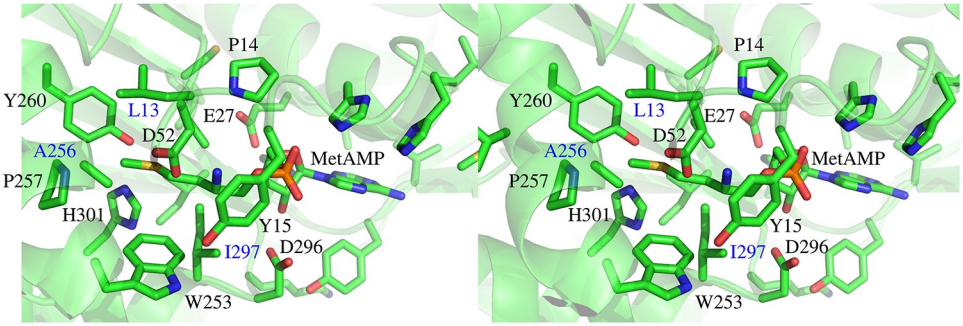


Fig 3. MetRS:MetAMP complex. Binding site closeup (stereo). Mutating side chains are 13, 256, 297.

https://doi.org/10.1371/journal.pcbi.1007600.g003

Table 2. MetRS redesigned for MetAMP binding by mutating positions 13, 256, 297.

			__binding__					___binding___	
rank	variant	[a]folding	[b]comp.	[b]exp.	rank	variant	[a]folding	[b]comp.	[b]exp.
1	CDV	4.5	-1.36		11	LAC	-0.3	0.25	1.8
2	MAV	1.3	-0.23	1.8	12	MAT	4.6	0.28	2.4
3	MAI	2.5	-0.20		13	LSV	0.4	0.29	
4	LAV	-1.3	-0.16	1.8	14	LAA	-0.6	0.31	3.8
5	MAC	2.3	-0.09	2.3	15	CAV	-8.8	0.34	2.8
6	**LAI**	0.0	0.00	0.0	16	CAI	-7.4	0.37	1.2
7	MAA	2.0	0.02		17	MSC	4.1	0.45	
8	MSV	3.1	0.11	3.4	18	MCV	1.0	0.46	
9	MSI	4.4	0.15	2.2	19	MCI	2.3	0.48	
10	LSI	1.6	0.20		20	MSA	3.8	0.56	
					21	LAT	1.8	0.59	2.2
26	CAC	-7.9	0.69	3.0	28	SAI	-3.5	0.72	1.2
51	SAC	-4.0	1.11	3.0	68	LAS	1.3	1.34	3.4
70	SSI	-1.9	1.35	2.2	81	SSC	-2.2	1.45	3.6
	MST	6.2	0.98	3.5		MSS	5.8	1.64	3.4

Calculations with the FDBSA solvent model.

[a]Folding and

[b]MetAMP binding free energies (kcal/mol) from computations and experiment, relative to the WT sequence LAI.

https://doi.org/10.1371/journal.pcbi.1007600.t002

Certain mutations at position 297 involved significant changes in the side chain volume, where the largest type, Ile or the smallest type, Ala was introduced or removed. For these, the computed binding free energies departed significantly from the experimental ones. However, if these two types were excluded, there were 25 point mutations between experimental variants, and for these, agreement was very good. The computed binding free energy differences had an rms error of just 0.52 kcal/mol and a mean unsigned error (mue) of 0.43 kcal/mol. The correlation between the experimental and computed sets was 0.52. Fig 4 shows the binding free energy changes. Note that the good agreement supports the assumption that the experimental K_M values are good proxies for the relative MetAMP binding free energies.

With the FDBLK solvent model, results were similar. The WT variant was ranked slightly lower, 20th. The top sequence was SAN, with a binding free energy of -1.3 kcal/mol relative to the WT. 7 of the 17 experimental sequences were ranked among the top 20 predictions. The computed and experimental binding free energy changes associated with point mutations are shown in Supplementary Material (Figure B in S1 File). Excluding (as above) mutations involving the types Ile or Ala at position 297, the mue and rms error were 0.76 and 0.98 kcal/mol, respectively, only slightly larger than with FDBSA.

Redesigning MetRS for catalytic power

For enzyme design, it is of great interest to select for a low activation free energy [14]. Therefore, we considered a model of the transition state complex (Fig 1). The ATP α phosphorus was bound to five oxygens: three coplanar and two perpendicular, corresponding to in-line attack of the Met carboxylate. In the first stage, we simulated a competing, ground state complex between MetRS, Met, and ATP. The same three binding pocket residues as above, 13, 256, and 297 were allowed to mutate into all types except Gly, Pro. We used the FDBLK solvent model. We optimized a bias potential during the MC simulation, flattening the free energy

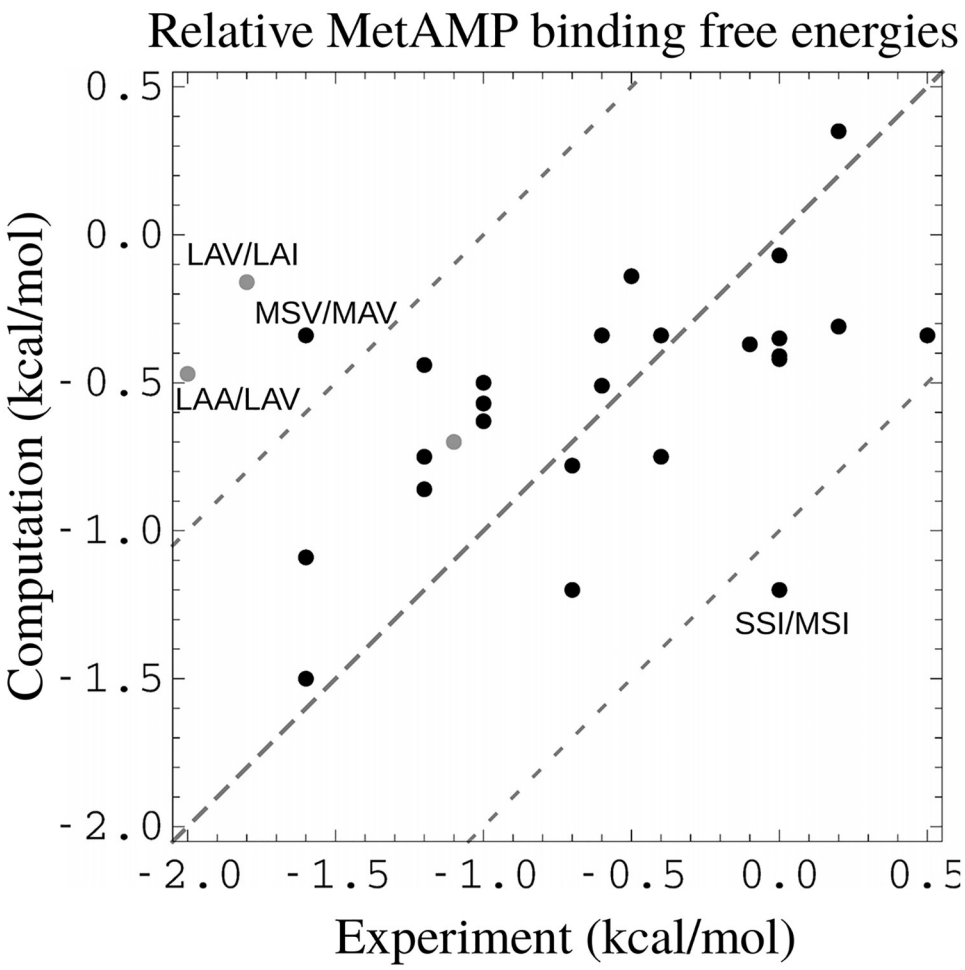


Fig 4. MetRS:MetAMP binding free energies, relative to the wildtype protein (WT). Shown are data for 28 point mutations. 3 gray points correspond to two mutations at position 297 (labeled) that change the side chain volume, plus one involving a variant (MST) that was predicted to be weakly stable (above our 5 kcal/mol threshold, see text) but was produced and measured experimentally nevertheless. Two other mutations with sizable errors are labeled.

https://doi.org/10.1371/journal.pcbi.1007600.g004

surface in sequence space. In the second stage, we simulated the transition state complex, with the bias included. All the variants that had been tested experimentally (Table 2) were sampled (WT and 19 variants, including five that Proteus had predicted (with FDBLK) to be above our 5 kcal/mol instability threshold). For each one, from the sampled populations, we deduced the free energy difference (Eq 8) between its ground state and transition state complexes, *i.e.*, its activation free energy. From transition state theory [14], this difference can be identified with the log of the catalytic reaction rate, k_{cat}. We also computed the Met dissociation free energies for the ground state complexes, which can be identified with the Met Michaelis constants, K_M. We first simulated the ground state complex with ATP but no Met, flattening its free energy surface with an adaptive bias. We then simulated the MetRS+Met+ATP complex, including the bias. From the sampled populations, we deduced the Met binding free energy of each variant, relative to WT (Eq 8). The overall protocol is schematized in Fig 5.

Fig 6 compares the k_{cat}/K_M ratios from experiment and simulations. We refer to them as catalytic efficiencies. We recall that they represent the 2nd order rate constant for the reaction of Met with the MetRS:ATP complex. Fig 6 shows the quantities $kT \log (k_{cat}/K_M) / (k_{cat}/K_M)_{WT}$,

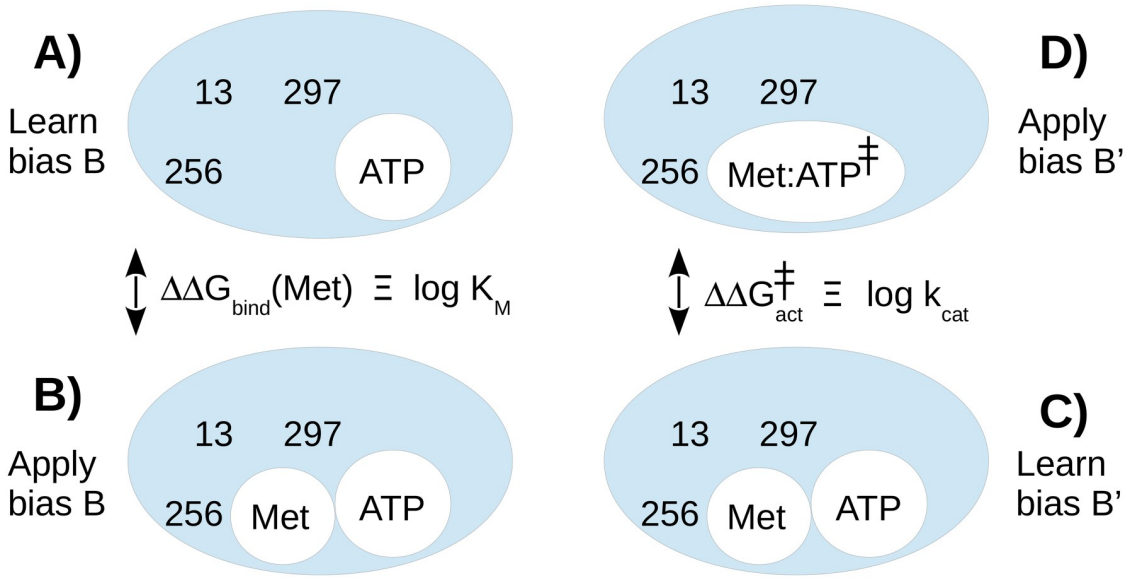


Fig 5. Computational scheme used to obtain the catalytic efficiencies k_{cat}/K_M. A) A bias B is optimized to flatten the sequence landscape of the enzyme without the Met ligand. Mutating positions are 13, 256, 297. **B)** The same bias B is used to simulate the complex including Met. Sequences are populated according to their Met binding affinities. **C)** A bias B' is optimized to flatten the sequence landscape of the complex including Met. **D)** B' is used to simulate the transition state complex. Sequences are populated according to their activation free energies. The lefthand simulations yield the predicted K_M values. The righthand simulations yield the predicted k_{cat} values.

https://doi.org/10.1371/journal.pcbi.1007600.g005

which express the catalytic efficiencies on a log scale, in thermal energy units, relative to the WT value. The figure includes WT and 19 other experimental variants. 5 of these had low predicted stabilities and are shown as gray points. WT defines the origin. For the other 14 points, agreement between calculations and experiment is quite good, with a correlation of 0.73 and mean errors of 1.36 kcal/mol (rms) and 1.18 kcal/mol (mue). Experimentally, WT has the largest efficiency. Computationally, two variants are predicted to be slightly better, by 0.9 and 0.2 kcal/mol, respectively, which is less than the mean error. Overall, by designing directly for a low activation free energy, we retrieve all the experimental variants and reproduce the catalytic efficiencies semi-quantitatively.

Discussion

Adaptive importance sampling solves the design problem for ligand binding and specificity. It applies positive design to one state (say, bound) and negative design to the other (unbound). It provides quantitative values for relative binding free energies or activation free energies. Variants sampled for one criterion, such as activation free energy (k_{cat}), can be reranked *a posteriori* based on another criterion, such as k_{cat}/K_M. *A posteriori* reranking or filtering does not leave out any important solutions; rather, the initial selection brings in too many solutions (*e.g.*, unstable variants), which are then filtered out at very little cost. In the first stage of the procedure, the sampling is very aggressive, if not exhaustive. In the second stage, it does not need to be exhaustive, since the best designs are exponentially enriched, and the unsampled variants are the ones with poor affinities or specificities. If one wants to reveal weak binders or perform reranking on another property, one can also flatten the energy landscape in the second stage. One can also use a more aggressive bias in one or both stages, including two-position biases. Replica Exchange MC can also be used to increase sampling. Using plain MC, one-position

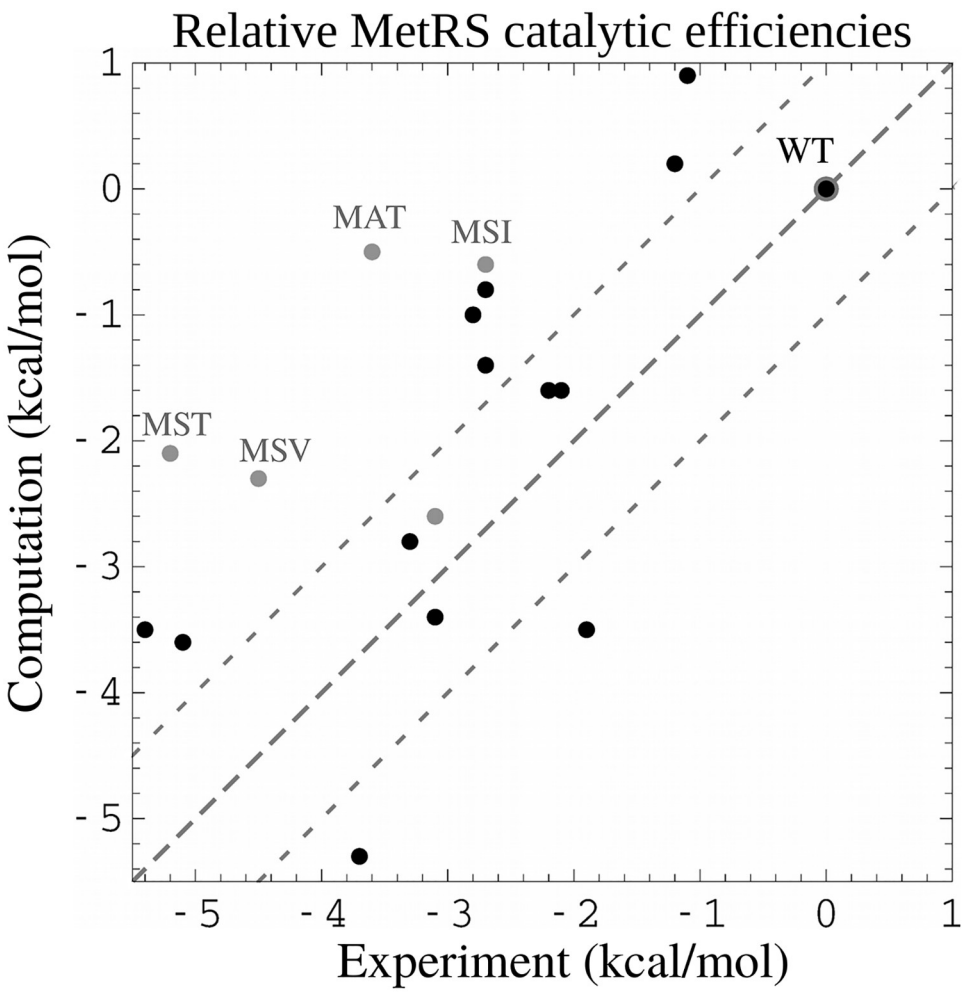


Fig 6. MetRS catalytic efficiencies kT log (k_{cat}/K_M) / $(k_{cat}/K_M)_{WT}$ relative to the wildtype (kcal/mol). Four gray points correspond to variants that were predicted to be weakly stable but were produced and measured experimentally nevertheless. Results obtained with the FDBLK solvent model.

https://doi.org/10.1371/journal.pcbi.1007600.g006

biases and no flattening of the holo state, our simulations produced 200 MetRS variants, enriched in tight binders, spanning a 7–8 kcal/mol range of binding free energies.

A difficulty when designing ligand binding is to choose one or more poses for the ligand. Here, we redesigned MetRS in cases where the ligand pose was known from an X-ray structure for one sequence: the SLL sequence in the AnL case and the wildtype sequence in the MetAMP case. For these ligands, we used the experimental ligand pose and protein backbone conformation. Three residues close to the ligand were then allowed to mutate. Not surprisingly, the calculations produced designed sequences that were homologous to the X-ray sequence. The experimental binding free energies in the MetAMP case were well-reproduced (the AnL values are not known). It is likely that other poses exist that would be compatible with other mutations, and would possibly lead to even stronger binding. The exploration of such alternate poses was left aside in this work. For the transition state complex, the position of the ligands could also be inferred with some confidence, since the enzyme achieves catalysis with little reorganization or motion of the substrates [15], and the modeled transition state geometry of the α phosphate was intermediate between that seen in two Met RS X-ray structures: the

MetRS product (adenylate) complex and the reactant (ATP) complex (PDB 4QRE). By designing the protein to stabilize this ligand pose, we may have biased the results towards native-like solutions. Here, too, the experimental relative activation free energies were well-reproduced, supporting the structural model.

Another important model component is the implicit solvent model. Here, we used a carefully-parameterized Generalized Born variant [33], a physically-plausible value of the protein dielectric constant and an "FDB" computational scheme that maintains the many-body nature of the GB model. The simpler, NEA scheme gave somewhat poorer results, similar to another recent study [35]. For nonpolar contributions to solvation, we compared a Surface Area (SA) treatment and a Lazaridis-Karplus (LK) treatment, which gave similar results. In the reported calculations, no water molecules were modelled explicitly. We also tested a model where three waters in the MetRS active site were explicitly represented: those that directly coordinate the Mg^{2+} ion in the substrate and transition state complexes for MetAMP formation. With both the FDBSA and FDBLK treatments, their explicit representation led to k_{cat} values well within the mean error of the calculations (relative to experiment). Most $kT \log (k_{cat}/K_M) / (k_{cat}/K_M)_{WT}$, values were with 0.2–0.3 kcal/mol of those reported above. Overall, the results were reasonably robust with respect to model details, with FDB giving improved performance.

Agreement with experiment was very good for three MetRS redesign test problems: redesign to bind the AnL ncAA, redesign to bind the natural intermediate MetAMP, and redesign for catalytic power for the reaction that produces MetAMP. Except for the earlier AnL data [22], the experiments were done in this work. Transition state modeling was done simply, by combining two X-ray structures and running a standard quantum chemistry protocol for atomic charges, consistent with the usual Amber force field [32]. All the procedures were carried out with the Proteus software, which is freely available to academics (https://proteus.polytechnique.fr). An entire calculation (setup, matrix calculation, MC simulations, postprocessing) lasted around one day on a 16-core desktop computer. We expect the present adaptive MC method will become the paradigm for computational enzyme design in the future.

Supporting information

S1 File. Supplementary appendix. This file includes a short theoretical derivation, some explanation of force field parameters, atomic charges for the MetRS transition state ligands, a figure showing the MetRS:AnL complex structure, and MetRS:MetAMP binding free energy results obtained with the FDBLK solvent model.
(PDF)

S2 File. Azidonorleucine force field information. This file contains the "topology" or 2D structure of AnL, including the atomic charges, followed by energy parameters for covalent bonds, angles, dihedrals, impropers, van der Waals terms and Generalized Born.
(FF)

S3 File. MetAMP force field information. This file contains the "topology" or 2D structure of MetAMP, including the atomic charges, followed by energy parameters for covalent bonds, angles, dihedrals, impropers, van der Waals terms and Generalized Born.
(FF)

Acknowledgments

We thank Christine Lazennec-Schurdevin for technical assistance, Alexandrine Daniel for preliminary MetRS:AnL calculations and Francesco Villa and David Mignon for many helpful discussions.

Author Contributions

Conceptualization: Emmanuelle Schmitt, Yves Mechulam, Thomas Simonson.

Data curation: Emmanuelle Schmitt, Yves Mechulam, Thomas Simonson.

Formal analysis: Vaitea Opuu, Thomas Simonson.

Investigation: Vaitea Opuu, Giuliano Nigro, Thomas Gaillard, Emmanuelle Schmitt, Yves Mechulam, Thomas Simonson.

Methodology: Vaitea Opuu, Emmanuelle Schmitt, Yves Mechulam, Thomas Simonson.

Project administration: Emmanuelle Schmitt, Yves Mechulam, Thomas Simonson.

Software: Vaitea Opuu, Thomas Simonson.

Supervision: Emmanuelle Schmitt, Yves Mechulam, Thomas Simonson.

Writing – original draft: Thomas Simonson.

Writing – review & editing: Emmanuelle Schmitt, Yves Mechulam, Thomas Simonson.

References

1. Malisi C, Schumann M, Toussaint NC, Kageyama J, Kohlbacher O, Höcker B. Binding Pocket Optimization by Computational Protein Design. PLoS One. 2012; 7:e52505. https://doi.org/10.1371/journal.pone.0052505 PMID: 23300688
2. Feldmeier K, Hoecker B. Computational protein design of ligand binding and catalysis. Curr Opin Chem Biol. 2013; 17:929–933. https://doi.org/10.1016/j.cbpa.2013.10.002 PMID: 24466576
3. Tinberg CE, Khare SD, Dou J, Doyle L, Nelson JW, Schena A, et al. Computational design of ligand-binding proteins with high affinity and selectivity. Nature. 2013; 501:212–218. https://doi.org/10.1038/nature12443 PMID: 24005320
4. Stoddard B, editor. Methods in Molecular Biology: Design and Creation of Ligand Binding Proteins. Springer Verlag, New York; 2016.
5. Samish I, MacDermaid CM, Perez-Aguilar JM, Saven JG. Theoretical and computational protein design. Ann Rev Phys Chem. 2011; 62:129–149. https://doi.org/10.1146/annurev-physchem-032210-103509
6. Simonson T, Ye-Lehmann S, Palmai Z, Amara N, Bigan E, Wydau S, et al. Redesigning the stereospecificity of tyrosyl-tRNA synthetase. Proteins. 2016; 84:240–253. https://doi.org/10.1002/prot.24972 PMID: 26676967
7. Shen Q, Tian H, Tang D, Yao W, Gao X. Ligand-K* sequence elimination: a novel algorithm for ensemble-based redesign of receptor-ligand binding. Trans Comp Biol Bioinf. 2014; 11:573–578. https://doi.org/10.1109/TCBB.2014.2302795
8. Viricel C, Simoncini D, Allouche D, de Givry S, Barbe S, Schiex T. Approximate counting with deterministic guarantees for affinity computation. In: LeThi HA, Dinh TP, Nguyen NT, editors. Adv. Intell. Syst. Comput. vol. 360. Springer, New York; 2015. p. 165–176.
9. Hallen MA, Donald BR. COMETS (Constrained Optimization of Multistate Energies by Tree Search): A provable and efficient protein design algorithm to optimize binding affinity and specificity with respect to sequence. J Comp Biol. 2016; 23:311–321. https://doi.org/10.1089/cmb.2015.0188
10. Karimi M, Shen Y. iCFN: an efficient exact algorithm for multistate protein design. Bioinf. 2018; 34:i811–820. https://doi.org/10.1093/bioinformatics/bty564
11. Bhattacherjee A, Wallin S. Exploring protein-peptide binding specificity through computational peptide screening. PLoS Comp Biol. 2013; 7:e1003277. https://doi.org/10.1371/journal.pcbi.1003277
12. Villa F, Panel N, Chen X, Simonson T. Adaptive landscape flattening in amino acid sequence space for the computational design of protein:peptide binding. J Chem Phys. 2018; 149:072302. https://doi.org/10.1063/1.5022249 PMID: 30134674
13. Hayes RL, Armacost KA, Vilseck JZ, Brooks CL III. Adaptive landscape flattening accelerates sampling of alchemical space in multisite lambda dynamics. J Phys Chem B. 2017; 121:3626–3635. https://doi.org/10.1021/acs.jpcb.6b09656 PMID: 28112940
14. Jencks WP. Catalysis in chemistry and enzymology. Dover, New York; 1986.

15. Ibba M, Francklyn C, Cusack S, editors. Aminoacyl-tRNA Synthetases. Landes Bioscience, Georgetown; 2005.
16. Xie J, Schultz PG. A chemical toolkit for proteins: an expanded genetic code. Nat Rev Molec Cell Biol. 2006; 7:775–782. https://doi.org/10.1038/nrm2005
17. Young TS, Schultz PG. Beyond the canonical twenty amino acids: expanding the genetic lexicon. J Biol Chem. 2010; 285:11039–11044. https://doi.org/10.1074/jbc.R109.091306 PMID: 20147747
18. Liu CC, Schultz PG. Adding new chemistries to the genetic code. Ann Rev Biochem. 2010; 79:413–444. https://doi.org/10.1146/annurev.biochem.052308.105824 PMID: 20307192
19. Neumann-Staubitz P, Neumann H. The use of unnatural amino acids to study and engineer protein function. Curr Opin Struct Biol. 2016; 38:119–128. https://doi.org/10.1016/j.sbi.2016.06.006 PMID: 27318816
20. Chin JW. Expanding and reprogramming the genetic code. Nature. 2017; 550:53–60. https://doi.org/10.1038/nature24031 PMID: 28980641
21. Wang L, Brock A, Herberich B, Schultz PG. Expanding the genetic code of *Escherichia coli*. Science. 2001; 292:498–500. https://doi.org/10.1126/science.1060077 PMID: 11313494
22. Tanrikulu IC, Schmitt E, Mechulam Y, Goddard W III, Tirrell DA. Discovery of Escherichia coli methionyl-tRNA synthetase mutants for efficient labeling of proteins with azidonorleucine in vivo. Proc Natl Acad Sci USA. 2009; 106:15285–15290. https://doi.org/10.1073/pnas.0905735106 PMID: 19706454
23. Wang FG, Landau DP. Efficient, multiple-range random walk algorithm to calculate the density of states. Phys Rev Lett. 2001; 86:2050–2053. https://doi.org/10.1103/PhysRevLett.86.2050 PMID: 11289852
24. Laio A, Gervasio F. Metadynamics: a method to simulate rare events and reconstruct the free energy in biophysics, chemistry and material science. Rep Prog Phys. 2008; 71:art. 126601. https://doi.org/10.1088/0034-4885/71/12/126601
25. Frenkel D, Smit B. Understanding molecular simulation, Chapter 3. Academic Press, New York; 1996.
26. Grimmett GR, Stirzaker DR. Probability and random processes. Oxford University Press, Oxford, United Kingdom; 2001.
27. Mignon D, Simonson T. Comparing three stochastic search algorithms for computational protein design: Monte Carlo, Replica Exchange Monte Carlo, and a multistart, steepest-descent heuristic. J Comput Chem. 2016; 37:1781–1793. https://doi.org/10.1002/jcc.24393 PMID: 27197555
28. Barducci A, Bussi G, Parrinello M. Well-tempered metadynamics: a smoothly converging and tunable free-energy method. Phys Rev Lett. 2008; 100:art. 020603. https://doi.org/10.1103/PhysRevLett.100.020603 PMID: 18232845
29. Dama JF, Parrinello M, Voth GA. Well-tempered metadynamics converges asymptotically. Phys Rev Lett. 2014; 112:art. 240602. https://doi.org/10.1103/PhysRevLett.112.240602 PMID: 24996077
30. Simonson T, Gaillard T, Mignon D, Schmidt am Busch M, Lopes A, Amara N, et al. Computational protein design: the Proteus software and selected applications. J Comput Chem. 2013; 34:2472–2484. https://doi.org/10.1002/jcc.23418 PMID: 24037756
31. Simonson T. The Proteus software for computational protein design. Ecole Polytechnique, Paris: https://proteus.polytechnique.fr; 2019.
32. Cornell W, Cieplak P, Bayly C, Gould I, Merz K, Ferguson D, et al. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. J Am Chem Soc. 1995; 117:5179–5197. https://doi.org/10.1021/ja00124a002
33. Lopes A, Aleksandrov A, Bathelt C, Archontis G, Simonson T. Computational sidechain placement and protein mutagenesis with implicit solvent models. Proteins. 2007; 67:853–867. https://doi.org/10.1002/prot.21379 PMID: 17348031
34. Gaillard T, Simonson T. Pairwise decomposition of an MMGBSA energy function for computational protein design. J Comput Chem. 2014; 35:1371–1387. https://doi.org/10.1002/jcc.23637 PMID: 24854675
35. Michael E, Polydorides S, Simonson T, Archontis G. Simple models for nonpolar solvation: parametrization and testing. J Comput Chem. 2017; 38:2509–2519. https://doi.org/10.1002/jcc.24910 PMID: 28786118
36. Polydorides S, Simonson T. Monte Carlo simulations of proteins at constant pH with generalized Born solvent, flexible sidechains, and an effective dielectric boundary. J Comput Chem. 2013; 34:2742–2756. https://doi.org/10.1002/jcc.23450 PMID: 24122878
37. Villa F, Mignon D, Polydorides S, Simonson T. Comparing pairwise-additive and many-body Generalized Born models for acid/base calculations and protein design. J Comput Chem. 2017; 38:2396–2410. https://doi.org/10.1002/jcc.24898 PMID: 28749575
38. Polydorides S, Amara N, Aubard C, Plateau P, Simonson T, Archontis G. Computational protein design with a generalized Born solvent model: application to Asparaginyl-tRNA synthetase. Proteins. 2011; 79:3448–3468. https://doi.org/10.1002/prot.23042 PMID: 21563215

39. Archontis G, Simonson T. Proton binding to proteins: a free energy component analysis using a dielectric continuum model. Biophys J. 2005; 88:3888–3904. https://doi.org/10.1529/biophysj.104.055996 PMID: 15821163

40. Dahiyat BI, Mayo SL. De novo protein design: fully automated sequence selection. Science. 1997; 278:82–87. https://doi.org/10.1126/science.278.5335.82 PMID: 9311930

41. Schmitt E, Tanrikulu IC, Yoo TH, Panvert M, Tirrell DA, Mechulam Y. Switching from an Induced-Fit to a Lock-and-Key Mechanism in an Aminoacyl-tRNA Synthetase with Modified Specificity. J Mol Biol. 2009; 394:843–851. https://doi.org/10.1016/j.jmb.2009.10.016 PMID: 19837083

42. Tuffery P, Etchebest C, Hazout S, Lavery R. A new approach to the rapid determination of protein side chain conformations. J Biomol Struct Dyn. 1991; 8:1267–1289. https://doi.org/10.1080/07391102.1991.10507882 PMID: 1892586

43. Gaillard T, Panel N, Simonson T. Protein sidechain conformation predictions with an MMGBSA energy function. Proteins. 2016; 84:803–819. https://doi.org/10.1002/prot.25030 PMID: 26948696

44. Crépin T, Schmitt E, Mechulam Y, Sampson PB, Vaughan MD, Honek JF, et al. Use of analogues of methionine and methionyl adenylate to sample conformational changes during catalysis in *Escherichia coli* methionyl-tRNA synthetase. J Mol Biol. 2003; 332:59–72. https://doi.org/10.1016/s0022-2836(03)00917-3 PMID: 12946347

45. Pokala N, Handel TM. Energy functions for protein design: adjustment with protein-protein complex affinities, models for the unfolded state, and negative design of solubility and specificity. J Mol Biol. 2005; 347:203–227. https://doi.org/10.1016/j.jmb.2004.12.019 PMID: 15733929

46. Druart K, Palmai Z, Omarjee E, Simonson T. Protein:ligand binding free energies: a stringent test for computational protein design. J Comput Chem. 2016; 37:404–415. https://doi.org/10.1002/jcc.24230 PMID: 26503829

47. Arnez JG, Augustine JG, Moras D, Francklyn CS. The first step of aminoacylation at the atomic level in histidyl-tRNA synthetase. Proc Natl Acad Sci USA. 1997; 94:7144–7149. https://doi.org/10.1073/pnas.94.14.7144 PMID: 9207058

48. Zurek J, Bowman A, Sokalski W, Mulholland A. MM and QM/MM modeling of threonyl-tRNA synthetase: Model testing and simulations. Struct Chem. 2004; 15:405–414. https://doi.org/10.1023/B:STUC.0000037896.80027.2c

49. Banik S, Nandi N. Aminoacylation Reaction in the Histidyl-tRNA Synthetase: Fidelity Mechanism of the Activation Step. J Phys Chem B. 2010; 114:12301–2311. https://doi.org/10.1021/jp910730s

50. Jorgensen WL, Chandrasekar J, Madura J, Impey R, Klein M. Comparison of simple potential functions for simulating liquid water. J Chem Phys. 1983; 79:926–935. https://doi.org/10.1063/1.445869

51. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, et al. Scalable molecular dynamics with NAMD. J Comput Chem. 2005; 26:1781–1802. https://doi.org/10.1002/jcc.20289 PMID: 16222654

52. Mellot P, Mechulam Y, Le Corre D, Blanquet S, Fayat G. Identification of an amino acid region supporting specific methionyl-tRNA synthetase:tRNA recognition. J Mol Biol. 1989; 208:429–443. https://doi.org/10.1016/0022-2836(89)90507-x PMID: 2477552

53. Schmitt E, Meinnel T, Panvert M, Mechulam Y, Blanquet S. Two acidic residues of *Escherichia coli* methionyl-tRNA synthetase act as negative discriminants towards the binding of noncognate tRNA anticodons. J Mol Biol. 1993; 233:615–628. https://doi.org/10.1006/jmbi.1993.1540 PMID: 8411169

54. Guillon L, Schmitt E, Blanquet S, Mechulam Y. Initiator tRNA binding by e/aIF5B, the eukaryotic/archaeal homologue of bacterial Initiation Factor IF2. Biochemistry. 2005; 44:15594–15601. https://doi.org/10.1021/bi051514j PMID: 16300409

55. Nigro G, Bourcier S, Lazennec-Schurdevin C, Schmitt E, Marlière P, Mechulam Y. Use of β3-methionine as an amino acid substrate of Escherichia coli methionyl-tRNA synthetase Journal of Structural Biology, in press, https://doi.org/10.1016/j.jsb.2019.107435

56. Braman J, Papworth C, A G. Site-directed mutagenesis using double-stranded plasmid DNA templates. Methods Molec Biol. 1996; 57:31–44.

57. Schmitt E, Meinnel T, Blanquet S, Mechulam Y. Methionyl-tRNA synthetase needs an intact and mobile KMSKS motif in catalysis of methionyl adenylate formation. J Mol Biol. 1994; 242:566–577. https://doi.org/10.1006/jmbi.1994.1601 PMID: 7932711

58. Dardel F. Comp App Biosci. 1994; 10:273–275.

59. Thompson D, Plateau P, Simonson T. Free energy simulations reveal long-range electrostatic interactions and substrate-assisted specificity in an aminoacyl-tRNA synthetase. ChemBioChem. 2006; 7:337–344. https://doi.org/10.1002/cbic.200500364 PMID: 16408313