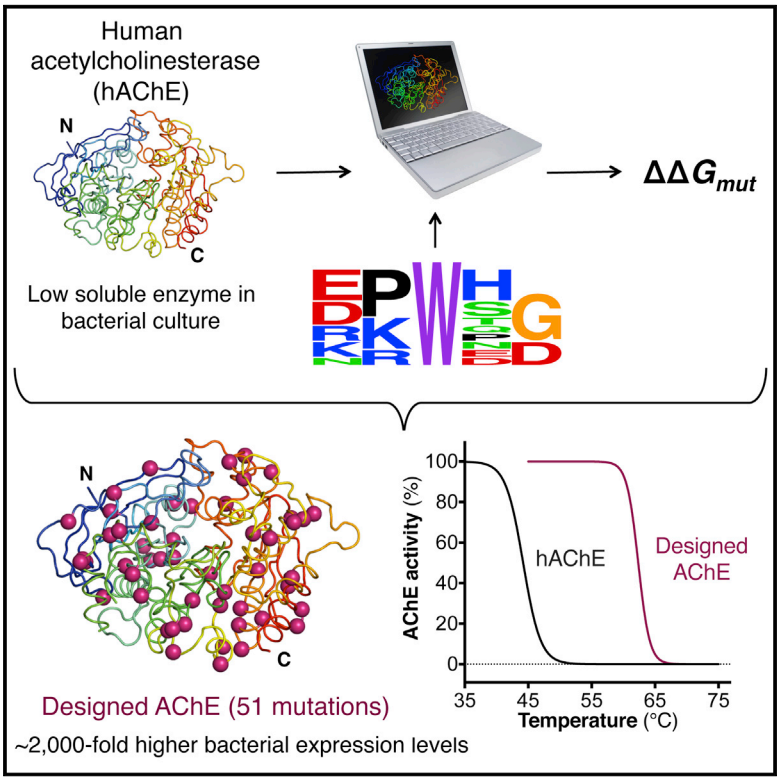
Technology

Molecular Cell

Automated Structure- and Sequence-Based Design of Proteins for High Bacterial Expression and Stability

Graphical Abstract



Authors

Adi Goldenzweig, Moshe Goldsmith, Shannon E. Hill, ..., Joel L. Sussman, Dan S. Tawfik, Sarel J. Fleishman

Correspondence

dan.tawfik@weizmann.ac.il (D.S.T.), sarel@weizmann.ac.il (S.J.F.)

In Brief

Heterologous expression of proteins and their mutants often results in misfolding and aggregation. Goldenzweig et al. (2016) developed an automated algorithm for protein stabilization requiring minimal experimental testing; for instance, the five tested variants of human acetylcholinesterase showed ≥100-fold higher soluble bacterial expression and higher melting temperatures than wild-type.

Highlights

- A new computational method is used to stabilize five recalcitrant proteins
- Designed variants show higher expression and stability with unmodified function
- A designed human acetylcholinesterase variant expresses solubly in bacteria
- The method is fully automated and implemented on a webserver

Goldenzweig et al., 2016, Molecular Cell *63*, 337–346
July 21, 2016
http://dx.doi.org/10.1016/j.molcel.2016.06.012

Molecular Cell
Technology

Automated Structure- and Sequence-Based Design of Proteins for High Bacterial Expression and Stability

Adi Goldenzweig,[1] Moshe Goldsmith,[1] Shannon E. Hill,[2] Or Gertman,[3] Paola Laurino,[1] Yacov Ashani,[1,4] Orly Dym,[4] Tamar Unger,[4] Shira Albeck,[4] Jaime Prilusky,[5] Raquel L. Lieberman,[2] Amir Aharoni,[3] Israel Silman,[6] Joel L. Sussman,[4] Dan S. Tawfik,[1,*] and Sarel J. Fleishman[1,*]

[1]Department of Biomolecular Sciences, Weizmann Institute of Science, Rehovot 7610001, Israel
[2]School of Chemistry & Biochemistry, Georgia Institute of Technology, Atlanta, GA 30332-0400, USA
[3]Department of Life Sciences, Ben-Gurion University of the Negev, P.O.B. 653, Beer-Sheva 8410501, Israel
[4]Israel Structural Proteomics Center, Weizmann Institute of Science, Rehovot 7610001, Israel
[5]Bioinformatics & Biological Computing Unit, Weizmann Institute of Science, Rehovot 7610001, Israel
[6]Department of Neurobiology, Weizmann Institute of Science, Rehovot 7610001, Israel
*Correspondence: dan.tawfik@weizmann.ac.il (D.S.T.), sarel@weizmann.ac.il (S.J.F.)
http://dx.doi.org/10.1016/j.molcel.2016.06.012

SUMMARY

Upon heterologous overexpression, many proteins misfold or aggregate, thus resulting in low functional yields. Human acetylcholinesterase (hAChE), an enzyme mediating synaptic transmission, is a typical case of a human protein that necessitates mammalian systems to obtain functional expression. We developed a computational strategy and designed an AChE variant bearing 51 mutations that improved core packing, surface polarity, and backbone rigidity. This variant expressed at ~2,000-fold higher levels in *E. coli* compared to wild-type hAChE and exhibited 20°C higher thermostability with no change in enzymatic properties or in the active-site configuration as determined by crystallography. To demonstrate broad utility, we similarly designed four other human and bacterial proteins. Testing at most three designs per protein, we obtained enhanced stability and/or higher yields of soluble and active protein in *E. coli*. Our algorithm requires only a 3D structure and several dozen sequences of naturally occurring homologs, and is available at http://pross.weizmann.ac.il.

INTRODUCTION

Most natural proteins are only marginally stable (Magliery, 2015). Thus, taken out of their natural context, through either overexpression, heterologous expression, or changes in environmental conditions, many proteins misfold and aggregate. The most general origin of overexpression challenges is low stability of the protein's native, functional state relative to alternative nonfunctional or aggregation-prone states. By designing variants with more favorable native-state energy, yields of soluble and functional protein obtained by heterologous overexpression can be dramatically increased, alongside other merits such as longer storage and usage lifetimes and enhanced engineering potential.

Engineering stable protein variants is a widely pursued goal. Methods based on phylogenetic analysis (Lehmann et al., 2000; Steipe et al., 1994) and structure-based rational or computational design (Borgo and Havranek, 2012; Jacak et al., 2012; Korkegian et al., 2005; Lawrence et al., 2007) yielded proteins with improved stability and higher functional expression (Magliery, 2015). Individual mutations, however, contribute little to stability (typically $\leq$1 kcal/mol) (Zhao and Arnold, 1999), whereas stabilizing large and poorly expressed proteins typically requires many mutations. However, since even a single severely destabilizing mutation can undermine the benefit accruing from all others, high prediction accuracy is essential. Despite improvements in accuracy, existing approaches have a relatively high probability of inadvertently introducing disruptive mutations (false-positive predictions) (Borgo and Havranek, 2012; Magliery, 2015; Potapov et al., 2009). Published efforts to stabilize large proteins therefore either incorporate only a few predicted stabilizing mutations (typically $\leq$4) at each experimental step or use library approaches to identify optimal combinations of stabilizing mutations (Khersonsky et al., 2011; Sullivan et al., 2012; Trudeau et al., 2014; Whitehead et al., 2012; Wijma et al., 2014). Such approaches are laborious and impractical for proteins without established medium-to-high throughput screens, let alone for proteins of unknown function. To address the demand for stabilizing large, recalcitrant proteins by a wide range of researchers, who lack background in computational design, we developed an automated algorithm based on atomistic Rosetta modeling and phylogenetic sequence information. We specifically aimed to develop a general method that minimizes false-positive predictions to ensure that only a few variants need to be experimentally tested to achieve high functional yields (ideally, just one variant). We applied this algorithm to four different enzymes and one protein of unknown function. In each case, up to five variants were designed as the default output, encoding from 9 to 67 mutations relative to wild-type. These

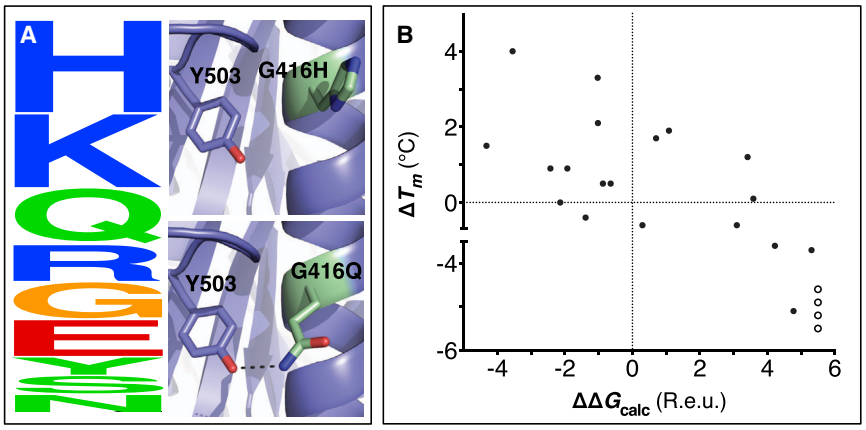


Figure 1. Eliminating Potentially Destabilizing Mutations through Homologous-Sequence Analysis and Computational Mutation Scanning

(A) (Left) Sequence logo for hAChE position Gly416. Letter height represents the respective amino acid's frequency in an alignment of homologous AChE sequences. The evolutionarily "allowed" sequence space (PSSM scores $\geq$0) at position 416 includes the nine amino acids shown. (Right) Structural models of mutations to the evolutionarily favored amino acid His, and to Gln, which is favored by Rosetta energy calculations. The His side chain is strained due to its proximity to the bulky Tyr503 aromatic ring, whereas the Gln side chain is relaxed and forms a favorable hydrogen bond with Tyr503 (dashed line).

(B) Computational mutation scanning correctly identifies all of the destabilizing mutations and $^{2}/_{3}$ of the stabilizing mutations in yeast triosephosphate isomerase (TIM). Sullivan et al. (2012) used consensus design to predict 23 mutations that would stabilize yeast TIM and measured each mutant's T_m difference relative to wild-type (ΔT_m). We used computational mutation scanning to predict the effects of each mutation on stability ($\Delta\Delta G_{calc}$). Four mutations were experimentally found to be highly deleterious, resulting in no functional expression, and therefore the T_m for these could not be measured (open circles); these four mutations have highly unfavorable $\Delta\Delta G_{calc}$ values (>4.6 Rosetta energy units; R.e.u.).
See also Table S1.

variants exhibited enhanced bacterial expression yields and stability, without sacrificing or altering activity.

DESIGN

Computational Mutation Scanning Minimizes False-Positive Mutations

To address the challenge of designing variants with a large net stabilizing effect, yet without modifying their function, our computational process began by scanning the natural sequence diversity. For every target wild-type sequence, we generated a sequence alignment, from which we computed a position-specific substitution matrix (PSSM) (Altschul et al., 2009); the PSSM represents the log-likelihood of observing any of the 20 amino acids at each position. At every amino acid position, "allowed" mutations were defined as those with a favorable PSSM score ($\geq$0). The rationale for restricting the allowed sequence space through the alignment scan is that, in general, deleterious mutations are purged by natural selection; purging is not absolute, however, and less favorable amino acids may occur at certain positions. Indeed, mutation to the most frequently observed amino acid often increases stability (the consensus effect) (Lehmann et al., 2000; Magliery, 2015; Steipe et al., 1994). However, our approach does not implement the consensus per se. Foremost, the alignment scan eliminates mutations that are rare or never seen in the natural diversity, rather than strictly selecting the most frequent amino acid. Next, we applied Rosetta computational mutation scanning (Whitehead et al., 2012), wherein each "allowed" mutation from the previous step was singly modeled against the background of the wild-type structure, and the energy difference between the wild-type and the single-point mutant was calculated ($\Delta\Delta G_{calc}$). By computing the effects of each mutation singly, rather than in combination with others, we restricted design choices to mutations that were likely to make additive contributions to stability, rather than nonadditive and context-dependent contributions, thereby minimizing the risk of false positives. We thereby defined the space of "potentially stabilizing" mutations as mutations with $\Delta\Delta G_{calc} \leq -0.45$ Rosetta energy units (R.e.u.); we chose this cutoff rather than 0 R.e.u. to further lower the risk of false positives, as indicated by the systematic evaluation described below. As a final step, we used Rosetta combinatorial sequence design to find an optimal combination of mutations within the space of potentially stabilizing mutations. Additionally, we used lower $\Delta\Delta G_{calc}$ cutoffs to select potentially stabilizing mutations prior to combinatorial sequence design, thereby generating several designs for experimental testing (see Data S1 available online).

The choice of mutations at Gly416 in human acetylcholinesterase (hAChE) illustrates the role of these two filters (alignment scan and computational mutation scanning) in pruning false positives (Figure 1A). Position 416 is located on a partially exposed helical surface, where the small and flexible amino acid Gly is likely to destabilize hAChE. Indeed, in the alignment of AChE homologs, Gly is infrequent, and His is the most prevalent amino acid. Modeling shows, however, that in the specific context of hAChE, His adopts a strained side-chain conformation; in contrast, Gln, the third most prevalent amino acid, is predicted to be most stabilizing owing to its high helical propensity and favorable hydrogen bonding with Tyr503. The combined filter therefore favors Gln over His for downstream design calculations.

To systematically evaluate the ability of our filtering method to identify stabilizing mutations, we compared its output to published experimental data on the stability effects of single-point mutations in the enzymes fungal endoglucanase Cel5A and yeast triosephosphate isomerase (TIM) (Figure 1B; Table 1) (Sullivan et al., 2012; Trudeau et al., 2014). The two filters (PSSM and Rosetta mutational scanning using $\Delta\Delta G_{calc} \leq -0.45$ R.e.u. as cutoff) eliminated all the severely destabilizing mutations (and 99.6% of all destabilizing mutations) and retained one-third and two-thirds of the experimentally verified stabilizing

Table 1. Computational Mutation Scanning versus Experimental Point Mutations in Fungal Endoglucanase 5A

Mutations	Stabilizing[a]	Destabilizing[a]	Total
True predictions[b]	12 (35%)	230 (99.6%)	242
False predictions[b]	22 (65%)	1 (0.4%)	23
Total	34	231	265

[a]Mutations were experimentally assayed by heat inactivation (Trudeau et al., 2014; raw data kindly provided by D. Trudeau).
[b]Mutations were predicted as stabilizing if they exhibited $\Delta\Delta G_{calc} < -0.45$ R.e.u. and destabilizing otherwise.

mutations in Cel5A and TIM, respectively. Thus, for both enzymes, our method identified a large fraction of the known stabilizing mutations, while excluding false positives almost entirely. The energy cutoff emphasizes the importance of minimizing false-positive predictions, that is, mutations that are experimentally destabilizing but are computationally assessed as favorable; for instance, in the case of Cel5A, using $\Delta\Delta G_{calc} \leq 0$ R.e.u. as cutoff would correctly predict four additional stabilizing mutations but would introduce eight more destabilizing mutations. False-negative predictions—that is, mutations that are stabilizing in experiment but are considered destabilizing by modeling—could not be attributed to a single modeling error (Table S1). Specifically, out of 22 false negatives in Cel5A, eight were filtered because they are unlikely according to the sequence alignment (PSSM score <0), and four showed $\Delta\Delta G_{calc}$ in the range −0.45–0 R.e.u., thereby missing the energy threshold by a small margin. Of the ten remaining false negatives, three occurred at core positions, where energy calculations typically penalize mutations, and two were mutations to proline. Improvements in the energy function and the conformation-relaxation procedure may in the future increase accuracy, and the analysis reported here provides a benchmark for such improvements.

RESULTS

Designed hAChE with Nearly 2,000-Fold Higher Bacterial Expression and Intact Activity

We applied our design strategy to human acetylcholinesterase (hAChE), a 60 kDa enzyme responsible for terminating synaptic transmission at cholinergic synapses by rapidly hydrolyzing the neurotransmitter acetylcholine (ACh) (Sussman et al., 1991). Although AChE is potentially useful for detection and detoxification of organophosphate nerve agents and pesticides, hAChE is currently produced by costly procedures involving purification from blood erythrocyte membranes, or using plant or mammalian expression systems. Previous attempts to express hAChE in bacteria resulted in extremely low levels of soluble and active protein (Fischer et al., 1993). AChE's active site is located at the bottom of a deep gorge that penetrates half way (20 Å) into the enzyme, and mutations along the gorge can reduce ACh hydrolysis rates by up to 1,000-fold (Ordentlich et al., 1995). To increase the stability and expression levels of hAChE without altering its activity, we imposed a further restriction on the allowed sequence space of the newly designed hAChE: in all Rosetta modeling simulations, the side-chain conformations of amino acids within 8 Å of the reversible inhibitor E2020, which spans the full length of the active-site gorge (Cheung et al., 2012), had to remain as in the native hAChE structure (Table S2). The latter restriction, in combination with the two above-described filters, dramatically reduced the sequence space available for design. The theoretical sequence space for hAChE, a 550-residue enzyme, is 10^{750}, a formidable number inaccessible even for advanced modeling algorithms. The size of the reduced sequence space, by contrast, was 10^{31} sequences, equivalent to complete computational design of a 24-amino acid peptide, a challenge solved already in the 1990s (Dahiyat and Mayo, 1997). We noted that the reduced sequence space led to convergence of combinatorial sequence optimization to identical, or nearly identical, sequences for any given $\Delta\Delta G_{calc}$ cutoff; this convergence, which is not usual in computational design (Fleishman et al., 2011), is a prerequisite for reproducibility and usage by nonexperts.

Given AChE's large size, we designed five alternatives using different $\Delta\Delta G_{calc}$ thresholds, with 17–67 mutations relative to hAChE (Table S3; Data S2), and subjected them to experimental testing. Synthetic genes encoding wild-type hAChE and the five designs were optimized for bacterial translation efficiency, fused to the C terminus of thioredoxin, and expressed in *E. coli* ShuffleT7express cells to facilitate disulfide-bond formation. In SDS-PAGE gels of supernatant fractions from bacterial lysates, the AChE band overlapped with other bands, precluding visual quantification. Nevertheless, because the designed variants' specific activity was found to be nearly identical to the wild-type's (Table 2), we could quantify the relative levels of soluble and active enzyme by measuring AChE activity in crude lysates. The five designs showed ≥100-fold higher lysate ACh-hydrolysis rates compared to hAChE, with the best design, dAChE4 (Figure 2A; 51 mutations), exhibiting an almost 2,000-fold higher rate (Figure 2B). Furthermore, due to its extremely low soluble expression, hAChE could not be purified from crude cell lysates, whereas the designed variants were readily purified to homogeneity, with dAChE4 yields in standard shaker flasks approaching 2 mg protein per liter of bacterial culture (Table S4).

The designed AChE mutations are scattered throughout the enzyme and show typical characteristics of stabilizing mutations, including improved core packing, higher backbone rigidity, and increased surface polarity (Figure 2A). In agreement with the design strategy and the higher levels of soluble and functional enzyme, we observed increased resistance to heat inactivation of up to 20°C relative to the *E. coli*-expressed hAChE (Figure 2C) and to hAChE expressed in mammalian cells (Table 2). The designs hydrolyzed ACh at rates that are within a 2-fold margin of hAChE rates, and displayed inactivation-rate constants by the nerve agent VX that are nearly identical to hAChE (the largest deviation was observed for dAChE3, which exhibited a 2.5-fold lower inactivation rate; Table 2).

The above observations of nearly identical activity profiles of the designed and wild-type AChEs suggested that the designed enzymes' active sites are essentially identical to that of hAChE. To verify this, we conducted crystallization trials on dAChE4, the design exhibiting the highest bacterial-expression yields. We noted that, in contrast to various natural AChEs studied by us, large crystals formed already within a few days and more

Table 2. Stability and Kinetic Parameters of Human and Designed AChE Variants

AChE Variant	Mut[a]	Normalized Activity[b]	Inactivation Temperature (°C)[c]		ACh Hydrolysis			
			Lysate	Purified	k_i (S_P-VX) × 10^7 ($M^{-1}min^{-1}$)	K_M (mM)	k_{cat} × 10^5 (min^{-1})	k_{cat}/K_M × 10^9 ($M^{-1}min^{-1}$)
hAChE (HEK293)	–	–	–	50.6 ± 0.3	7.92 ± 0.15	0.09 ± 0.01	3.8 ± 0.2	4.4 ± 0.6
hAChE (bacterial)	–	1	44 ± 0.4	ND[d]	ND[d]	ND[d]	ND[d]	ND[d]
dAChE1	17	119 ± 20	60.5 ± 0.4	ND[d]	6.48 ± 0.71	0.050 ± 0.006	4.4 ± 0.1	8.7 ± 1.1
dAChE2	30	280 ± 40	61.5 ± 0.5	67.1 ± 0.3	ND[d]	ND[d]	ND[d]	ND[d]
dAChE3	42	308 ± 44	62.3 ± 0.2	69.4 ± 0.3	2.65 ± 0.52	0.177 ± 0.010	4.4 ± 0.1	2.5 ± 0.2
dAChE4	51	1770 ± 258	62.3 ± 0.3	66.2 ± 1.2	7.60 ± 0.34	0.071 ± 0.007	2.73 ± 0.07	3.9 ± 0.4
dAChE5	67	637 ± 134	61.1 ± 0.2	68.5 ± 0.6	6.47 ± 0.82	0.104 ± 0.010	3.16 ± 0.01	3.0 ± 0.3

[a]Number of amino acid mutations relative to wild-type hAChE.
[b]Activity in crude lysates of cells expressing the AChE variants from 250 ml *E. coli* cultures.
[c]The temperature at which 50% of activity was retained. Enzyme samples were incubated for 30 min at varying temperatures and tested for AChE activity after cooling.
[d]Not determined.

reproducibly. dAChE4's structure was solved at 2.6Å resolution, thus yielding, to the best of our knowledge, the first structure of an AChE expressed in a prokaryote (Table S5). dAChE4's structure is very similar to that of wild-type hAChE, with a 0.7 Å root-mean-square deviation (rmsd) over Cα atoms. Residues at the catalytic gorge aligned particularly well, with an all-atom rmsd of only 0.125Å (Figure 2D). Thus, despite 51 mutations relative to wild-type, ~2,000-fold gain in bacterial expression levels, and 20°C higher heat tolerance, dAChE4 is virtually indistinguishable in its active site from hAChE (differences in other parts of the enzyme are shown in Figure S1 and online at http://proteopedia.org/w/Journal:Molecular_Cell:1). dAChE4 can therefore serve in future structural studies of inhibitors that target the AChE active site.

Other Designed Enzymes Show High Stability and/or Soluble Expression

We applied our algorithm to two other human enzymes—the histone deacylase SIRT6, the human DNA methyltransferase Dnmt3a, and a bacterial phosphotriesterase dubbed PTE. Being the only enzyme known to have actually evolved to degrade organophosphates (OPs), PTE is of considerable biotechnological potential for decontamination and detoxification of OPs, including nerve agents. However, PTE detoxifies most nerve agents at rates too slow for practical applications. Wild-type PTE was previously engineered for higher expression (Roodveldt and Tawfik, 2005) and hydrolysis rates toward various OPs (Bigley et al., 2015; Cherny et al., 2013). Introduction of function-altering mutations, however, destabilized the enzyme, as is often the case for laboratory-evolved enzymes (Tokuriki et al., 2008); thus, stabilization is a prerequisite for further engineering (Bloom et al., 2006). We designed three alternatives using different $\Delta\Delta G_{calc}$ thresholds, encoding 9–28 mutations relative to wild-type PTE (Table S3), and subjected them to experimental testing. Synthetic genes encoding the three designs and PTE-S5, a published variant encoding three mutations, which displays ~20-fold higher expression levels compared to wild-type (Roodveldt and Tawfik, 2005), were optimized for bacterial translation efficiency, fused to maltose-binding protein, and expressed in *E. coli* GG48 cells. All three designs displayed increased levels of soluble, functional enzyme compared to PTE-S5, which already displays higher expression levels than wild-type (Table 3). Two of the three designs showed ~10°C higher tolerance to heat inactivation relative to PTE-S5, with no significant change in activity toward the OP substrate paraoxon (Table 3; Figure S2A). Another noteworthy outcome of stabilization design was increased metal affinity. PTE is a metalloenzyme bearing two active-site metals, typically Zn^{+2} (Benning et al., 2001). Directed evolution of wild-type PTE for higher expression (PTE-S5) led to a significant decrease in metal affinity—a major practical drawback for applications in conditions in which Zn^{+2} cannot be supplemented. The designed variant dPTE2, which contained 19 mutations and exhibited the highest tolerance to heat inactivation, also exhibited a marked increase in metal affinity compared to PTE-S5, approaching the affinity of wild-type PTE (Figures 3A and S2B; Table 3).

SIRT6 is an ADP-ribosylase and NAD^+-dependent deacylase that removes acyl groups from acylated lysines, thereby regulating several essential cellular processes (Kugel and Mostoslavsky, 2014). The low bacterial-expression levels of human SIRT6 (hSIRT6), and its weak deacylation activity relative to SIRT1, limit its study. The designed variant (dSIRT6) contained 11 mutations relative to hSIRT6 (Tables S2 and S3). It was expressed in *E. coli* and purified to homogeneity with yields of 20 mg per liter culture, an ~5-fold increase relative to hSIRT6, and its deacylase activity was 60% higher than hSIRT6's (Figures S3A and S3B). We subsequently incorporated the design mutations on the background of an engineered hSIRT6 variant dubbed E1, which contains three mutations relative to hSIRT6 (our unpublished data). E1 exhibits increased deacylation activity compared to hSIRT6 but 3-fold lower expression levels in human cell lines. In contrast, the designed variant dE1 recapitulates hSIRT6's expression levels in human cell lines while maintaining high deacylation activity (Figure 3B). Given the beneficial effects of SIRT6 overexpression on longevity in mice (Kanfi et al., 2012), our designed mutant can be used to

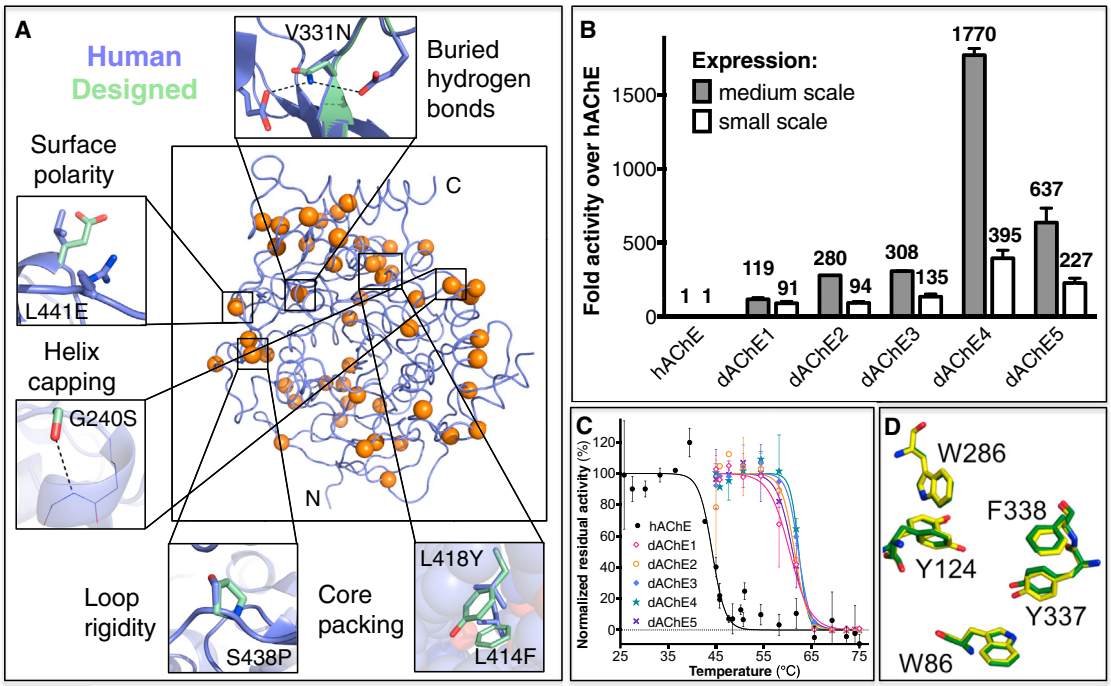


Figure 2. Design of a Stable hAChE variant and Its Functional Expression in Bacteria

(A) The structural underpinnings of stabilization in the designed variant dAChE4. Wild-type hAChE is shown in blue and 51 mutated positions, which are distributed throughout dAChE4, are indicated by orange spheres. Thumbnails highlight stabilizing effects of selected mutations.

(B) Bacterial lysate activity levels of designed AChEs normalized to hAChE activity. Crude lysates were derived from 250 ml flasks (medium scale) or 0.5 ml *E. coli* cultures grown in a 96-well plate (small scale). The higher activity levels in the designed variants reflect higher levels of soluble, functional enzyme.

(C) Designed AChE variants (colored lines) show higher resistance to heat inactivation compared to hAChE (black). Residual activities following incubation at different temperatures were measured in bacterial lysates and normalized to the activity in nontreated lysates.

(D) Sub-Ångstrom accuracy in alignment of key residues in the vicinity of the catalytic triad in the crystallographic structure of dAChE4 (PDB: 5HQ3, yellow) compared to hAChE (PDB: 4EY4, green).

See also Figure S1 and Tables S2–S5.

establish a correlation between SIRT6's deacylation rates and its physiological function.

In the case of human Dnmt3a (hDnmt3a), soluble and active fractions of the human enzyme can be obtained by *E. coli* expression, but enzyme activity is very low. We tested one design dDnmt3a containing 14 mutations relative to hDnmt3a (Tables S2 and S3). In contrast to all other designs, the designed dDnmt3a showed significantly lower expression levels than hDnmt3a in *E. coli* (Figures S3C and S3D). Nonetheless, it exhibited nearly 10-fold higher specific activity (Figure 3C), suggesting that hDnmt3a preparations contain a significant fraction of soluble but inactive or poorly active enzyme. Therefore, although Dnmt3a failed to meet our design method's success criteria (higher stability and soluble expression), it significantly increased the protein's functional yield.

A Webserver for Protein Stabilization

Encouraged by the consistently positive results presented above, we implemented the algorithm as a webserver, called the *P*rotein *R*epair *O*ne *S*top *S*hop (PROSS, http://pross.weizmann.ac.il). Following their request for assistance in solving a critical expression and stability question regarding the human myocilin OLF domain (hMyoc-OLF), two of the authors (S.E.H. and R.L.L.) were granted unsupervised access to the webserver. OLF domains are found in extracellular proteins of multicellular organisms. A number of OLF domains have been implicated in human disease, but the function(s) of most OLF domains remains elusive. The best-studied OLF domain is that of hMyoc-OLF, in which more than 100 nonsynonymous mutations are implicated in inherited forms of open-angle glaucoma. Mutations documented in patients lead to destabilized myocilin protein that forms cytotoxic aggregates (Burns et al., 2010; Yam et al., 2007). S.E.H. and R.L.L. therefore posited that mutations that confer enhanced stability might reduce hMyoc-OLF's propensity to misfold at physiological temperatures.

Three hMyoc-OLF structures (PDB: 4WXQ, 4WXS, and 4WXU) were submitted to PROSS, producing seven designs for each structure with 5–25 mutations each. From all suggested point mutations, S.E.H. and R.L.L. manually derived one variant (dMyoc-OLF) comprising 21 mutations. Although hMyoc-OLF binds Ca^{+2} (Donegan et al., 2012), the physiological role of Ca^{+2} binding is still unknown; nevertheless, the Ca^{+2}-binding positions remained unchanged, reflecting the high sequence and structural constraints acting on them. Furthermore, none of the design mutations have been implicated in human disease,

Table 3. Stability and Kinetic Parameters of PTE Variants

Variant	Mut[a]	Normalized Activity[b]	Inactivation Temperature (°C)		$T_{1/2}$ Chelator[c] (min)	K_M (mM)[d]	$k_{cat} \times 10^5$ (min^{-1}) [d]	$k_{cat}/K_M \times 10^9$ ($min^{-1}M^{-1}$) [d]
			Lysate	Purified				
PTE-S5	3	1	50.9 ± 0.7	52.4 ± 0.2	7.5 ± 0.3	0.101 ± 0.023	0.970 ± 0.076	0.96 ± 0.33
dPTE1	9	2.0	54.7 ± 3.2	ND[e]	ND[e]	ND[e]	ND[e]	ND[e]
dPTE2	19	6.1	59.2 ± 0.7	62.0 ± 0.2	51.2 ± 5.1	0.060 ± 0.014	0.70 ± 0.05	1.17 ± 0.35
dPTE3	28	2.3	47.0 ± 1.3	ND[e]	ND[e]	ND[e]	ND[e]	ND[e]

[a]Number of mutations relative to wild-type PTE.
[b]Fold increase in activity in crude *E. coli* lysates relative to PTE-S5.
[c]Chelator, 50 μM 1,10-phenanthroline.
[d]Kinetic parameters for paraoxon.
[e]ND, not determined.

supporting the notion that the disease-associated mutations are destabilizing.

dMyoc-OLF and hMyoc-OLF were expressed in *E. coli* fused to maltose-binding protein (MBP-OLF). The designed variant gave an order of magnitude higher yield than hMyoc-OLF (Table S6). In addition, whereas the purified hMyoc-OLF comprises a mixture of a soluble MBP-OLF-aggregate and the properly folded monomer (Burns et al., 2010), the designed variant was almost exclusively monomeric, and also exhibited a 16.8°C higher melting temperature (T_m) than hMyoc-OLF (Figures 3D and 3E). This is the highest T_m measured so far for Myoc-OLF, nearly the same as that of the gliomedin OLF domain that lacks a Ca^{+2}-binding site (Hill et al., 2015). Nonetheless, as indicated by the T_m increase in the presence of Ca^{+2} (Figure 3D; Table S6), Ca^{+2} binding was retained in dMyoc-OLF. The designed variant dMyoc-OLF is therefore a starting point for detailed OLF structure-function analysis.

DISCUSSION

Most existing stability-design methods focus on one element or another of protein stability. For instance, some methods introduce disulfides or prolines, while others optimize core packing (Borgo and Havranek, 2012; Korkegian et al., 2005) and increase surface polarity (Magliery, 2015) and charge (Lawrence et al., 2007). Our design algorithm, by contrast, selects for all amino acid mutations that optimize the protein's computed energy, subject to constraints inferred from homologous sequences. It thereby generates designs that improve a range of molecular parameters that are associated with stability (Table 4) while preserving function. We find, furthermore, that in the various proteins studied here, the algorithm implemented solutions that appear to selectively tackle specific defects of each protein. For instance, 17 of the 51 mutations in design dAChE4 impact core positions, a very large number compared to previous design studies that targeted protein cores (Borgo and Havranek, 2012; Korkegian et al., 2005). Indeed, natural AChE structures contain unusually large core cavities (Koellner et al., 2000). In design dPTE2, by contrast, only two core mutations were implemented, and the net negative charge increased by 13 units, shifting the isoelectic point (pI) from a near-neutral value of 6.7 for the wild-type to 5.0; indeed, near-neutral pI is often associated with poor solubility, and surface "supercharging" has been used by others to improve thermal stability (Lawrence et al., 2007). The stability considerations, however, do not come at the expense of functional constraints. In Dnmt3a, for example, the positive net charge and basic pI are fully conserved as expected for a DNA-binding protein. We conclude that the combination of evolutionary constraints, selection of mutations that individually contribute to computed stability, and combinatorial sequence optimization within the space of these mutations can address a broad range of stability defects in various proteins without affecting their original molecular functions.

The case studies discussed here represent a diverse group of proteins belonging to different fold families (TIM barrel, β-propeller, and α/β hydrolase fold) and displaying different activities. In each case, the wild-type protein and variants previously engineered using traditional methods suffer from low stability, as manifested by intolerance to high temperature and function-modifying mutations or by low solubility, heterologous expression, cofactor affinity, or specific activity. That a single fully automated method that only considers phylogenetic constraints and optimizes the native state's energy is able to address these diverse challenges without sacrificing function demonstrates that the underlying source of these problems is the native state's marginal stability. Furthermore, the stability-design method has the potential to optimize proteins used as scaffolds for de novo binder or enzyme design, where low stability and expression levels have been limiting (Fleishman et al., 2011; Khersonsky et al., 2011).

Limitations

The stability design algorithm requires a few dozen unique sequence homologs and an atomic structure of the target protein. With the burgeoning of sequence databases, the first requirement is likely to be met with few exceptions; the second, however, may be an impediment, particularly since low protein yields and stability negatively impact structure determination. To expand beyond experimentally determined structures, homology models may be used, and it remains to be seen whether they provide sufficient accuracy for atomistic stability-design calculations.

EXPERIMENTAL PROCEDURES

Computational Procedures

Phylogenetic Sequence Constraints

For every query sequence, homologous sequences were collected using the BLASTP algorithm (Altschul et al., 1990) on the nonredundant (nr) database,

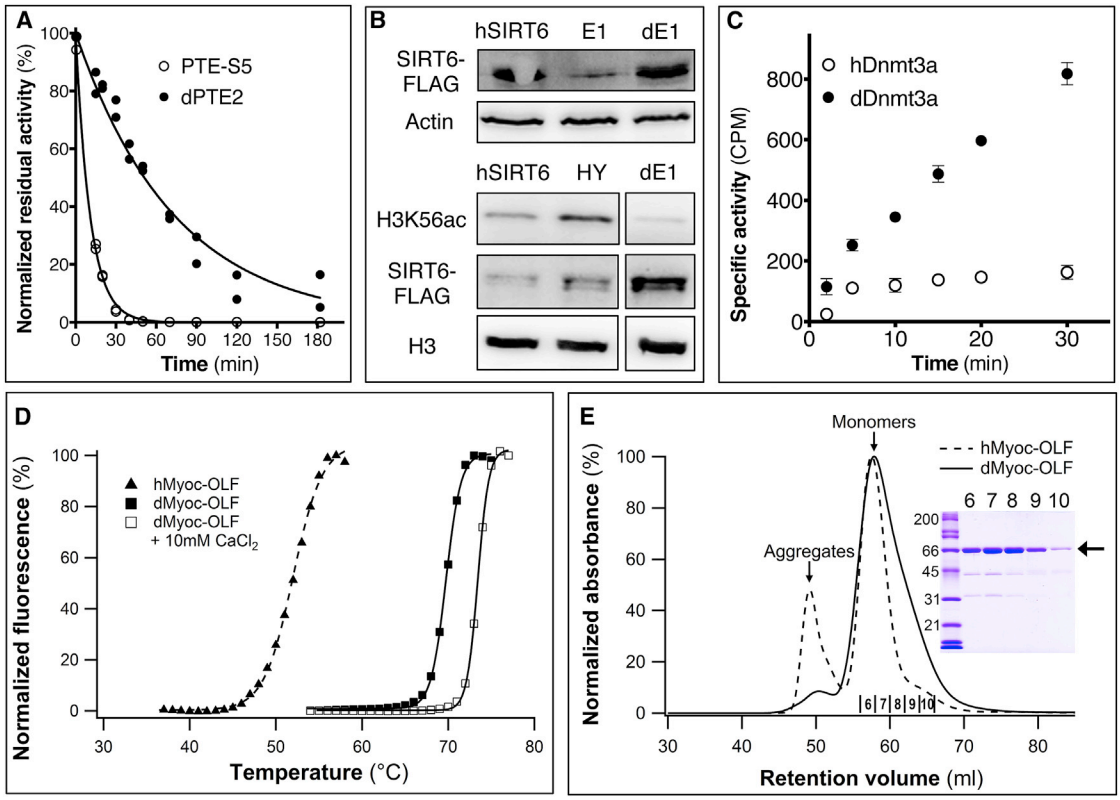


Figure 3. Higher Expression, Stability, and Activity of Designed Variants of PTE, SIRT6, Dnmt3a, and Myoc-OLF

(A) Compared to the previously engineered variant PTE-S5 (Roodveldt and Tawfik, 2005), the design dPTE2 shows higher resistance to inactivation by the metal chelator 1,10-phenanthroline (50 μM), indicating higher metal affinity and stability.

(B) (Upper panel) The previously engineered E1 variant of hSIRT6 shows a 3-fold decline in in vivo expression levels, whereas following computational design, dSIRT6-E1 (denoted as dE1) recapitulates hSIRT6's expression levels (western blot quantified using ImageJ). Actin expression levels are provided as control. (Lower panel) dE1 exhibits higher in vivo histone H3 Lys56-deacylation activity compared to hSIRT6. HY denotes a loss-of-function mutant of hSIRT6; H3 expression levels are provided as control.

(C) The designed variant dDnmt3A shows 10-fold higher DNA-methylation compared to hDnmt3a as determined by the levels of incorporation of H^3-methyl groups in the presence of equal enzyme concentrations.

(D) dMyoc-OLF is more thermostable than hMyoc-OLF, and addition of Ca^{+2} further stabilizes it.

(E) Size-exclusion chromatography of MBP-fused hMyoc-OLF and dMyoc-OLF indicated a significant aggregated fraction in hMyoc-OLF, as previously described (Burns et al., 2010), and a minor aggregated fraction in dMyoc-OLF. (Inset) SDS-PAGE analysis of dMyoc-OLF. Lane 1, molecular weight standards (kDa); lanes 2–6, size exclusion fractions as labeled in chromatogram.

See also Figures S2 and S3 and Tables S2, S3, and S6.

with the maximum number of hits set to 500 and the *e* value to 10^{-4}. All other parameters were set to default values. Hits were excluded if they covered less than 60% of the query sequence or if their sequence identity to the query sequence was lower than 30% (28% for PTE due to the low diversity of homologs at >30% sequence identity). The remaining sequences were clustered using cd-hit (Li and Godzik, 2006), with a clustering threshold of 90% and default parameters.

MUSCLE (Edgar, 2004) was used with default parameters to derive a multiple sequence alignment (MSA) from the clustered sequences. Gaps in the alignment between the query and any homolog often occur within loop regions and may reflect differences in the local backbone conformation. To reduce alignment uncertainty, we detected secondary-structure elements in the query protein structure (using DSSP [Kabsch and Sander, 1983]) and eliminated subsequences in homologs that contained gaps in loop segments that intervene between any two secondary-structure elements relative to the query. In effect, for every loop region this procedure generated a specific alignment that only comprised homologous sequences with no insertions or deletions relative to the query. Since this feature of the stability-design algorithm had not yet been implemented when we designed AChE and Dnmt3a, these two proteins were designed with default MUSCLE alignments, containing information from all homologs in loop regions.

Given the alignments of each query sequence, we computed a position-specific scoring matrix (PSSM) using the PSI-BLAST algorithm (Altschul et al., 2009). The PSSM represents the log probability of observing each of the 20 amino acids at each position in the query. Non-negative PSSM scores are considered likely to occur in evolution and define the space of allowed mutations.

Structure-Based Constraints

To prevent activity loss, residues within 8 Å of small-molecule ligands observed in the active site (hAChE and Dnmt3a), or within 5 Å of metal cofactors (PTE), DNA chains (Dnmt3a) and protein ligands (SIRT6) were held fixed throughout all Rosetta simulations. In homo-oligomer structures (PTE and AChE), residues within 5Å of the oligomer interface were held fixed. For PTE and SIRT6, some of the active-site constraints were defined using an alignment to another, ligand-bound structure, namely residues within 8Å of DPJ in 4NP7 for PTE, and residues within 5Å of chain F in 3ZG6 for SIRT6

Table 4. Structure and Sequence Features of Designed Variants Compared to Their Wild-Type Counterparts

Design Variant	aa[a]	Mut[b]	Core[c]	Salt Bridges[d]	H Bonds[d]	X → Pro[e]	Charge N/P[d,f]	pI[g] des	pI[g] WT
dAChE1	530	17 (3%)	9 (53%)	3	14	1	1/2	5.5	5.5
dAChE2		30 (6%)	15 (50%)	4	11	3	4/4	5.5	
dAChE3		42 (8%)	16 (38%)	5	18		6/6	5.5	
dAChE4		51 (10%)	17 (33%)	3	10		7/8	5.5	
dAChE5		67 (13%)	19 (28%)	10	15	5	7/9	5.6	
dPTE1	331	9 (3%)	0 (0%)	3	5	0	6/−2	5.3	6.7
dPTE2		19 (6%)	2 (11%)	8	9		11/−2	5.0	
dPTE3		28 (9%)	5 (18%)	12	7		16/−3	4.8	
dSIRT6	280	11 (4%)	1 (9%)	0	4	1	0/0	8.3	8.3
dDnmt3	272	14 (5%)	0 (0%)	4	7	1	1/1	8.4	8.6
dMyoc-OLF	259	21 (8%)	2 (10%)	2	4	0	−2/1	5.4	5.0

[a]Number of amino acids in the PDB structure (per monomer); the design algorithm ignores all protein segments that are absent from the structure.
[b]Number of amino acid mutations in the design compared to wild-type; in brackets, as percentage of total number of amino acids.
[c]Number of mutations in protein core. In brackets, as percentage of total number of mutations.
[d]Difference compared to wild-type structure. Positive values indicate higher number in design than wild-type.
[e]Number of mutations to proline.
[f]Mutations to negative (D/E) and positive (K/R) amino acids are indicated to the left and right of the slash, respectively.
[g]Isoelectric point: des refers to designed variant, and WT refers to wild-type.

(Table S2). Two positions on either side of missing density and residues at the termini (if the termini made no interactions with other parts of the protein) were not allowed to mutate. For hSIRT6, four residues that were found to increase activity in previous laboratory-evolution experiments were modeled prior to simulation, and not allowed to mutate (see Table S2). In the PROSS webserver, active-site constraints follow the same thresholds defined above, and the user may specify any other residue to be held fixed during the simulations through an interactive HTML form.

Rosetta Structure Modeling and Design

See Data S1 online for RosettaScripts, flags, and command lines.

Two energy functions were alternated throughout the simulations: the all-atom "hard-repulsive" Rosetta energy function (talaris2014) (O'Meara et al., 2015), which is dominated by van der Waals, hydrogen bonding, implicit solvation, and Coulomb electrostatics with a distance-dependent dielectric coefficient, and a "soft-repulsive" version of this energy function (soft-rep), where van der Waals overlaps and residue conformational strain are downweighted. The two energy functions were augmented with harmonic backbone coordinate restraints set to the coordinates observed in the PDB structure, and with a term that favors mutations with a higher PSSM score.

We refined each PDB structure by four iterations of side-chain packing (except in active-site residues) and side-chain and backbone minimization, saving the minimum-energy structure. For hAChE, the largest protein tested here, refinement calculations took, on average, 2.5 hr per trajectory (Data S1).

Computational mutation scanning was applied to the refined structure using the FilterScan mover in Rosetta (Whitehead et al., 2012). At every position, each allowed mutation (that is, every amino acid identity with PSSM score ≥ 0) was modeled singly against the background of the refined structure. Protein sidechains within 8 Å of the modeled mutation were repacked, and side-chain and constrained backbone minimization were used to accommodate the mutation. The energy difference between the refined structure and the optimized configuration of the single-point mutant was calculated using talaris2014 (O'Meara et al., 2015). Seven energy thresholds were used to define different mutation spaces (−0.45, −0.75, −1.0, −1.25, −1.5, −1.8, and −2.0 R.e.u.). Mutation scanning is the most time-consuming step in the calculation, taking between 30 s and 6 min per position (1 min per position on average for hAChE) using a standard CPU (Data S1). The procedure can be parallelized using a computer cluster, as implemented in the PROSS webserver.

For Dnmt3a, nine mutations were eliminated from the allowed sequence space since they showed exceptionally high and unexpected changes in their Rosetta backbone-omega energy, probably due to the poor crystallographic resolution of the input PDB structure (Table S2).

For each of the sequence spaces of potentially stabilizing mutations, combinatorial sequence optimization was implemented. Starting from the refined structure, we imposed coordinate restraints relative to the coordinates in the PDB file, and implemented four iterations of sequence design, side-chain, and backbone minimization, while alternating soft and hard repulsive potentials. For AChE, each design trajectory took ~2.5 hr on average (Data S1).

Design Model Analysis

Sequence and structural features of designed variants were compared to the sequence and structure of the wild-type protein. The differences in the number of salt bridges and hydrogen bonds were evaluated using the Rosetta FeaturesReporter suite (Leaver-Fay et al., 2013; O'Meara et al., 2015), with a salt-bridge interaction defined as two counter charges within less than 5.5Å. Buried residues were defined as residues with >22 and >75 neighboring non-hydrogen atoms within 10 Å and 12 Å, respectively. Protein isoelectric points (pI) were calculated using the ExPASy webserver (http://web.expasy.org/compute_pi).

For experimental procedures, see Supplemental Experimental Procedures online.

For DNA and protein sequences, see Data S2.

SUPPLEMENTAL INFORMATION

Supplemental Information includes three figures, six tables, two data sets, and Supplemental Experimental Procedures and can be found with this article at http://dx.doi.org/10.1016/j.molcel.2016.06.012.

AUTHOR CONTRIBUTIONS

A.G., D.S.T., and S.J.F. conceived the idea. A.G. and S.J.F. developed the algorithm. A.G. and J.P. developed the automated PROSS webserver. A.G. designed variants of AChE, PTE, Dnmt3a, and SIRT6. S.E.H. and R.L.L designed dMyoc-OLF. Experimental characterization was by M.G. and Y.A. (AChE), M.G. (PTE), S.E.H. (Myoc-OLF), O.G. (SIRT6), and P.L. (Dnmt3a). T.U. purified dAChE4 for crystallization, and O.D. solved its structure, with input from J.A.S.,

I.S., and S.J.F. The results were discussed and the manuscript written by A.G., D.S.T., and S.J.F. with input from all authors.

ACKNOWLEDGMENTS

We thank David Schreiber for analyzing structural features of design models; Ravit Netzer, Gideon Lapidoth, and Christoffer Norn for suggestions; and our collaborators for testing the PROSS webserver. Research in the Fleishman laboratory is supported by the Israel Science Foundation (ISF) through an individual grant, the Center for Research Excellence in Structural Cell Biology, and the joint ISF-UGC program, a Starter's Grant from the European Research Council, a Career Development Award from the Human Frontier Science Program and a Marie Curie Reintegration Grant, the Minerva Foundation, an Alon Fellowship, and a charitable donation from Sam Switzer and Family. Funding by a DTRA project grant (HDTRA1-11-C-0026) to D.S.T. and NIH (R01EY021205) to RLL are gratefully acknowledged. S.J.F. is a Martha S. Sagon Career Development Chair. D.S.T. is the Nella and Leon Benoziyo Professor of Biochemistry. The collaboration between the Fleishman and Tawfik laboratories is also supported by the Rothschild-Caesaria Foundation.

Received: March 15, 2016
Revised: May 18, 2016
Accepted: June 7, 2016
Published: July 14, 2016

REFERENCES

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J. Mol. Biol. *215*, 403–410.

Altschul, S.F., Gertz, E.M., Agarwala, R., Schäffer, A.A., and Yu, Y.K. (2009). PSI-BLAST pseudocounts and the minimum description length principle. Nucleic Acids Res. *37*, 815–824.

Benning, M.M., Shim, H., Raushel, F.M., and Holden, H.M. (2001). High resolution X-ray structures of different metal-substituted forms of phosphotriesterase from Pseudomonas diminuta. Biochemistry *40*, 2712–2722.

Bigley, A.N., Mabanglo, M.F., Harvey, S.P., and Raushel, F.M. (2015). Variants of phosphotriesterase for the enhanced detoxification of the chemical warfare agent VR. Biochemistry *54*, 5502–5512.

Bloom, J.D., Labthavikul, S.T., Otey, C.R., and Arnold, F.H. (2006). Protein stability promotes evolvability. Proc. Natl. Acad. Sci. USA *103*, 5869–5874.

Borgo, B., and Havranek, J.J. (2012). Automated selection of stabilizing mutations in designed and natural proteins. Proc. Natl. Acad. Sci. USA *109*, 1494–1499.

Burns, J.N., Orwig, S.D., Harris, J.L., Watkins, J.D., Vollrath, D., and Lieberman, R.L. (2010). Rescue of glaucoma-causing mutant myocilin thermal stability by chemical chaperones. ACS Chem. Biol. *5*, 477–487.

Cherny, I., Greisen, P., Jr., Ashani, Y., Khare, S.D., Oberdorfer, G., Leader, H., Baker, D., and Tawfik, D.S. (2013). Engineering V-type nerve agents detoxifying enzymes using computationally focused libraries. ACS Chem. Biol. *8*, 2394–2403.

Cheung, J., Rudolph, M.J., Burshteyn, F., Cassidy, M.S., Gary, E.N., Love, J., Franklin, M.C., and Height, J.J. (2012). Structures of human acetylcholinesterase in complex with pharmacologically important ligands. J. Med. Chem. *55*, 10282–10286.

Dahiyat, B.I., and Mayo, S.L. (1997). De novo protein design: fully automated sequence selection. Science *278*, 82–87.

Donegan, R.K., Hill, S.E., Turnage, K.C., Orwig, S.D., and Lieberman, R.L. (2012). The glaucoma-associated olfactomedin domain of myocilin is a novel calcium binding protein. J. Biol. Chem. *287*, 43370–43377.

Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. *32*, 1792–1797.

Fischer, M., Ittah, A., Liefer, I., and Gorecki, M. (1993). Expression and reconstitution of biologically active human acetylcholinesterase from Escherichia coli. Cell. Mol. Neurobiol. *13*, 25–38.

Fleishman, S.J., Whitehead, T.A., Ekiert, D.C., Dreyfus, C., Corn, J.E., Strauch, E.M., Wilson, I.A., and Baker, D. (2011). Computational design of proteins targeting the conserved stem region of influenza hemagglutinin. Science *332*, 816–821.

Hill, S.E., Donegan, R.K., Nguyen, E., Desai, T.M., and Lieberman, R.L. (2015). Molecular details of olfactomedin domains provide pathway to structure-function studies. PLoS ONE *10*, e0130888.

Jacak, R., Leaver-Fay, A., and Kuhlman, B. (2012). Computational protein design with explicit consideration of surface hydrophobic patches. Proteins *80*, 825–838.

Kabsch, W., and Sander, C. (1983). Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers *22*, 2577–2637.

Kanfi, Y., Naiman, S., Amir, G., Peshti, V., Zinman, G., Nahum, L., Bar-Joseph, Z., and Cohen, H.Y. (2012). The sirtuin SIRT6 regulates lifespan in male mice. Nature *483*, 218–221.

Khersonsky, O., Röthlisberger, D., Wollacott, A.M., Murphy, P., Dym, O., Albeck, S., Kiss, G., Houk, K.N., Baker, D., and Tawfik, D.S. (2011). Optimization of the in-silico-designed kemp eliminase KE70 by computational design and directed evolution. J. Mol. Biol. *407*, 391–412.

Koellner, G., Kryger, G., Millard, C.B., Silman, I., Sussman, J.L., and Steiner, T. (2000). Active-site gorge and buried water molecules in crystal structures of acetylcholinesterase from Torpedo californica. J. Mol. Biol. *296*, 713–735.

Korkegian, A., Black, M.E., Baker, D., and Stoddard, B.L. (2005). Computational thermostabilization of an enzyme. Science *308*, 857–860.

Kugel, S., and Mostoslavsky, R. (2014). Chromatin and beyond: the multitasking roles for SIRT6. Trends Biochem. Sci. *39*, 72–81.

Lawrence, M.S., Phillips, K.J., and Liu, D.R. (2007). Supercharging proteins can impart unusual resilience. J. Am. Chem. Soc. *129*, 10110–10112.

Leaver-Fay, A., O'Meara, M.J., Tyka, M., Jacak, R., Song, Y., Kellogg, E.H., Thompson, J., Davis, I.W., Pache, R.A., Lyskov, S., et al. (2013). Scientific benchmarks for guiding macromolecular energy function improvement. Methods Enzymol. *523*, 109–143.

Lehmann, M., Pasamontes, L., Lassen, S.F., and Wyss, M. (2000). The consensus concept for thermostability engineering of proteins. Biochim. Biophys. Acta *1543*, 408–415.

Li, W., and Godzik, A. (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics *22*, 1658–1659.

Magliery, T.J. (2015). Protein stability: computation, sequence statistics, and new experimental methods. Curr. Opin. Struct. Biol. *33*, 161–168.

O'Meara, M.J., Leaver-Fay, A., Tyka, M.D., Stein, A., Houlihan, K., DiMaio, F., Bradley, P., Kortemme, T., Baker, D., Snoeyink, J., and Kuhlman, B. (2015). Combined covalent-electrostatic model of hydrogen bonding improves structure prediction with Rosetta. J. Chem. Theory Comput. *11*, 609–622.

Ordentlich, A., Barak, D., Kronman, C., Ariel, N., Segall, Y., Velan, B., and Shafferman, A. (1995). Contribution of aromatic moieties of tyrosine 133 and of the anionic subsite tryptophan 86 to catalytic efficiency and allosteric modulation of acetylcholinesterase. J. Biol. Chem. *270*, 2082–2091.

Potapov, V., Cohen, M., and Schreiber, G. (2009). Assessing computational methods for predicting protein stability upon mutation: good on average but not in the details. Protein Eng. Des. Sel. *22*, 553–560.

Roodveldt, C., and Tawfik, D.S. (2005). Directed evolution of phosphotriesterase from Pseudomonas diminuta for heterologous expression in Escherichia coli results in stabilization of the metal-free state. Protein Eng. Des. Sel. *18*, 51–58.

Steipe, B., Schiller, B., Pluckthun, A., and Steinbacher, S. (1994). Sequence statistics reliably predict stabilizing mutations in a protein domain. J. Mol. Biol. *15*, 188–192.

Sullivan, B.J., Nguyen, T., Durani, V., Mathur, D., Rojas, S., Thomas, M., Syu, T., and Magliery, T.J. (2012). Stabilizing proteins from sequence statistics: the interplay of conservation and correlation in triosephosphate isomerase stability. J. Mol. Biol. *420*, 384–399.

Sussman, J., Harel, M., Frolow, F., Oefner, C., Goldman, A., Toker, L., and Silman, I. (1991). Atomic structure of acetylcholinesterase from Torpedo californica: a prototypic acetylcholine-binding protein. Science *253*, 872–879.

Tokuriki, N., Stricher, F., Serrano, L., and Tawfik, D.S. (2008). How protein stability and new functions trade off. PLoS Comput. Biol. *4*, e1000002.

Trudeau, D.L., Lee, T.M., and Arnold, F.H. (2014). Engineered thermostable fungal cellulases exhibit efficient synergistic cellulose hydrolysis at elevated temperatures. Biotechnol. Bioeng. *111*, 2390–2397.

Whitehead, T.A., Chevalier, A., Song, Y., Dreyfus, C., Fleishman, S.J., De Mattos, C., Myers, C.A., Kamisetty, H., Blair, P., Wilson, I.A., and Baker, D. (2012). Optimization of affinity, specificity and function of designed influenza inhibitors using deep sequencing. Nat. Biotechnol. *30*, 543–548.

Wijma, H.J., Floor, R.J., Jekel, P.A., Baker, D., Marrink, S.J., and Janssen, D.B. (2014). Computationally designed libraries for rapid enzyme stabilization. Protein Eng. Des. Sel. *27*, 49–58.

Yam, G.H.-F., Gaplovska-Kysela, K., Zuber, C., and Roth, J. (2007). Aggregated myocilin induces russell bodies and causes apoptosis: implications for the pathogenesis of myocilin-caused primary open-angle glaucoma. Am. J. Pathol. *170*, 100–109.

Zhao, H., and Arnold, F.H. (1999). Directed evolution converts subtilisin E into a functional equivalent of thermitase. Protein Eng. *12*, 47–53.

Molecular Cell, Volume *63*

Supplemental Information

Automated Structure- and Sequence-Based Design of Proteins for High Bacterial Expression and Stability

Adi Goldenzweig, Moshe Goldsmith, Shannon E. Hill, Or Gertman, Paola Laurino, Yacov Ashani, Orly Dym, Tamar Unger, Shira Albeck, Jaime Prilusky, Raquel L. Lieberman, Amir Aharoni, Israel Silman, Joel L. Sussman, Dan S. Tawfik, and Sarel J. Fleishman

Supplementary Information

Supplementary figures

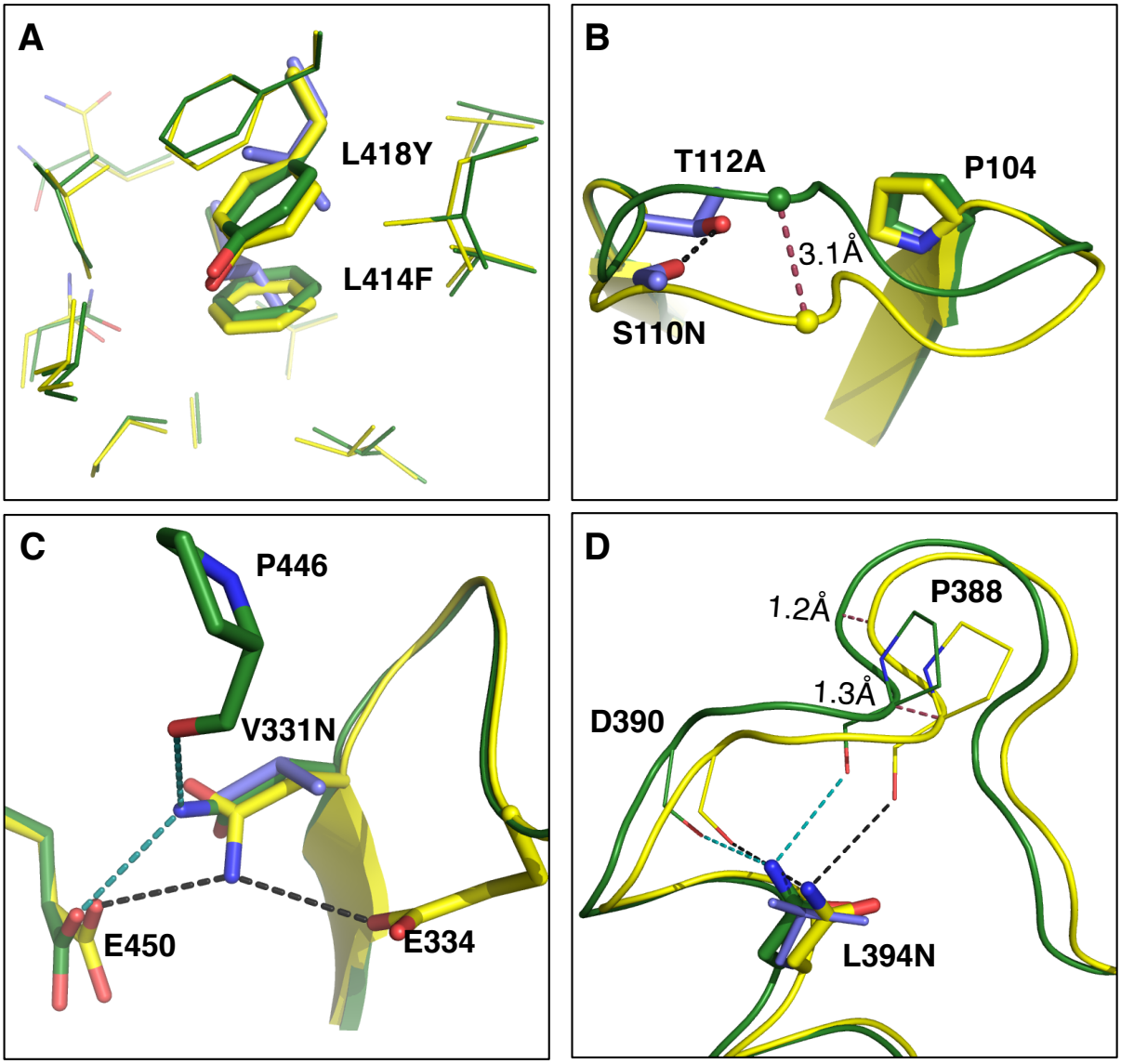


Figure S1. Comparison of the dAChE4 design model (yellow) and the solved crystal structure (PDB entry: 5HQ3, green), Related to Figure 2D. (A) Sub-Ångstrom accuracy in design of two small-to-large core mutations (wild-type leucines are shown in violet). **(B)** The maximal deviation observed between respective Cα atoms in the model and structure is 3.1Å (raspberry dashed line). This conformation change likely results from elimination of a side chain-backbone hydrogen bond between Thr112 and Ser110 due to the designed Thr112Ala mutation. **(C, D)** Comparison of designed buried hydrogen bonds. Val331Asn was predicted to form a hydrogen bond with Glu450 and another with Pro446 in the designed model; in the crystal structure, instead, Asn331 interacts with Glu334 and Glu450 (black and turquoise dashed lines for dAChE4 and the crystal structure hydrogen bonds, respectively; wild-type Val in violet). Leu394Asn forms 2 hydrogen bonds with Pro388 and Asp390, as designed.

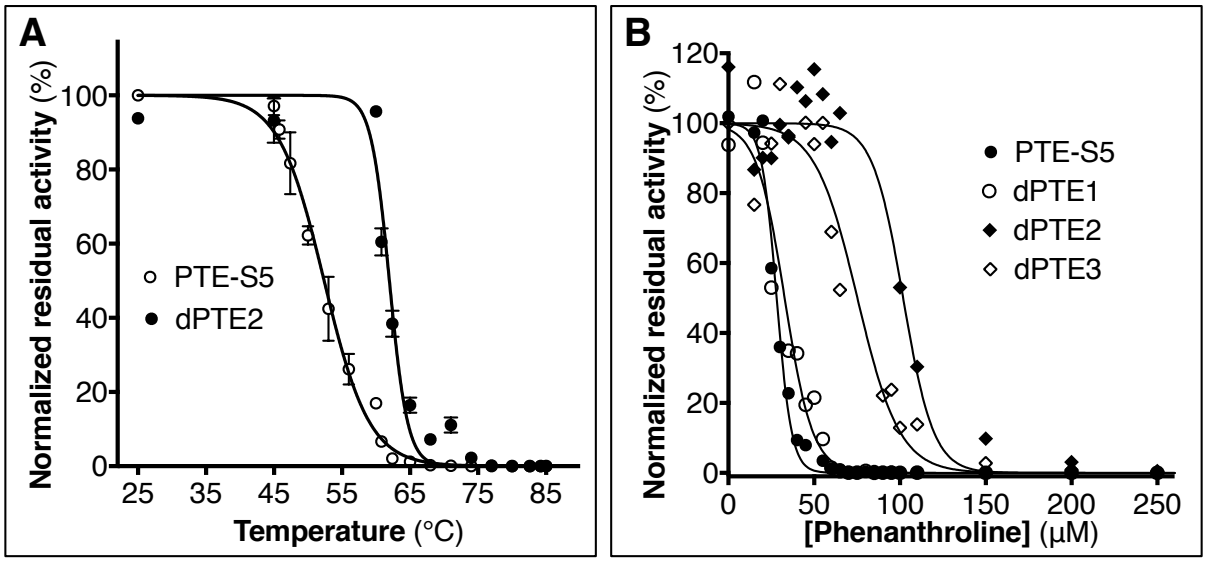


Figure S2. dPTE2 shows higher thermal stability and higher metal cofactor affinity compared to PTE-S5, Related to Figure 3A and Table 3. (A) Residual activities of the purified proteins were measured following incubation at various temperatures on the substrate ethyl-paraoxon. **(B)** Residual activities of the proteins expressed in cell lysates were measured following incubation with various concentrations of a metal chelator, 1,10-phenanthroline.

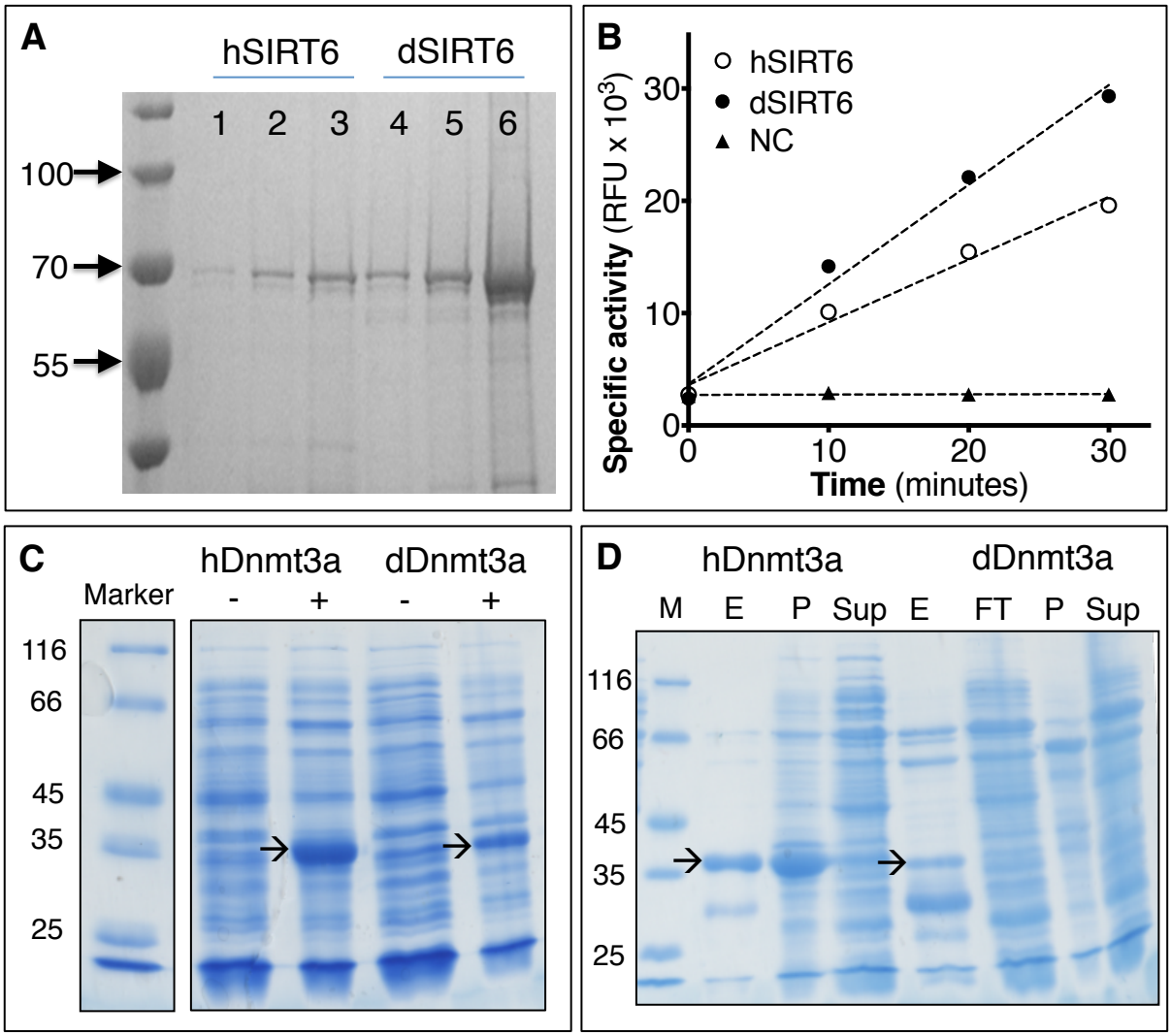


Figure S3. Additional experimental information about SIRT6 and Dnmt3a, Related to Figure 3B & C. (A) dSIRT6 shows fivefold increase in expression relative to hSIRT6. SDS-PAGE shows Ni^{2+}-NTA purified fractions of MBP-fused hSIRT6 and dSIRT6. **(B)** dSIRT6 shows 60% higher activity than hSIRT6. Deacylation activity of purified hSIRT6, dSIRT6 and a negative control (NC) were measured as a function of time. **(C, D)** dDnmt3a shows decreased expression levels compared to hDnmt3a. The left SDS-PAGE shows crude bacterial lysates of hDnmt3a and dDnmt3a before (-) and after induction (+). The right SDS-PAGE shows crude bacterial lysates and purified enzyme samples (E – eluted fraction, P – pellet, Sup – supernatant, FT – flow through). For methylation activity assays, the enzyme concentration was determined by SDS-PAGE comparing various hDnmt3a and dDnmt3a concentrations to a BSA protein standard. The major impurity in the purified dDnmt3a fraction (~30kDa) is an *E. coli* protein, probably prolyl isomerase, that co-elutes with the His-tagged Dnmt3a (Renata Jurkowska, personal communication).

Supplementary Tables

Table S1. False-negative predictions in fungal endoglucanase Cel5A and yeast triosephosphate isomerase (TIM), Related to Figure 1B and Table 1.

Cel5A (PDB entry 3QR3)			
Mutation	**$\Delta\Delta G_{calc}$ (R.e.u.)**	**ΔT_m**	**ΔPSSM[a]**
S133R	-0.14	0.4	-1
K219Q	-0.18	2.8	-4
V101I	-0.28	0.5	1
G239N	-0.33	0.7	-1
S318E	0.14	0.9	-1
S309W	0.29	0.4	5
N155E	0.95	0.5	-2
Y278F	1.05	1.0	-3
T233V	1.27	0.87	2
T80E	1.78	0.5	-4
G293A	1.78	3.7	3
S139P	2.37	2.0	2
I82M	2.71	0.3	0
N76P	3.41	2.0	-1
S318Q	Low PSSM score[b]	0.5	-2
D13E	Low PSSM score[b]	3.0	-4
T18P	Low PSSM score[b]	2.0	-6
F191V	Low PSSM score[b]	0.89	-7
V302A	Low PSSM score[b]	1.0	-4
K219A	Low PSSM score[b]	2.0	-8
S309F	Low PSSM score[b]	2.7	0
S309L	Low PSSM score[b]	1.5	-1
TIM (PDB entry 1YPI)			
I109V	0.70	1.7	1
A66C	1.08	1.9	4
Y49Q	3.42	1.2	-1
I83L	3.59	0.1	1

[a] Difference between the PSSM score of the mutated amino acid and the wild-type amino acid

[b] Mutation showed PSSM score < 0, and was therefore not subjected to Rosetta mutational scanning

Table S2. Modeling constraints, Related to Figures 2 & 3 and Tables 2-4.

Protein (PDB entry)	Unique homologues in MSA	Fixed positions	Reason for fixation/repacking only
hAChE (4EY7)	165	72-79, 82-83, 85-87, 117-122, 124-126, 130, 132-133 202-205, 229, 286-287, 291-297, 334-342, 365, 402, 439, 446-451	8Å from E2020[a]
		373-374, 376-377, 380-381, 383-386, 527, 530, 534, 538-539	5Å from chain B[a]
		256-266, 493-498	Near missing density[b]
		4-7	N terminus[b]
		298	Near catalytic gorge[c,b]
PTE (1HZY)	95	55-60, 100-104, 106, 108, 169-173, 200-202, 230, 233, 253-254, 257, 267, 270-272, 300-303, 305-306, 312-313, 317	8Å from ligand[a,d] (includes also all 5 Å from Zn^{+2} ions)
		61-65, 67-72, 131-141, 145-146, 148-149, 152-153, 159-160, 307-311	5Å from chain B[a]
SIRT6 (3K35)	278	12-13, 15-22, 49-57, 59-65 67-69, 90, 92, 110-113, 130-132, 183-186, 211-218, 220-221, 237-242, 252-259	8Å from APR[a]
		14, 58, 72, 78, 80, 84, 114-115, 134, 139, 141-142, 144, 154, 164, 166, 175, 177, 188-192, 195, 199, 219, 222-223, 225-226, 229, 244, 246-247	5Å from Zn^{+2} ion or peptide[a,e]
		167-174	Near missing density[b]
		291-295	C termini[b]
		66, 155, 210, 243	E1 engineered positions[f]
Dnmt3a (2QRV)	83	635-642, 659-662, 665, 681-684, 702-708, 710-716, 752-756, 777, 783-788, 860, 863-864, 866, 879, 882-890	8Å around SAH or DNA chains[a]
		667, 670-672, 687-689, 719-721, 723, 725-726, 728-729, 731-732, 735-736, 738,740-	5Å from chains B,

		742, 763-764, 767-770, 816-819, 848-849, 851, 853-858, 869, 872-876, 878, 881	D[a]
		823-829	Near missing density[b]
		623-627, 906-908	Termini[b]
		I651F, G669T, H673F, D743P, M775I, M775V, L801Y, P845E, F847T	Post mutation scan exclusion[g]
Myoc-OLF (4WXS/Q/U)	219	None	-

[a] Residues were allowed to minimize but not pack (sample rotamers).

[b] Residues were allowed to pack but not mutate.

[c] Near the catalytic gorge but above the automated fixation cutoffs.

[d] Ligand is DPJ from PDB entry 4NP7.

[e] Peptide is chain F from PDB entry 3ZG6.

[f] Positions engineered by authors O.G. and A.A. to generate the variant E1 (data not published)

[g] Mutations with exceptionally large difference in Rosetta omega dihedral energy were excluded.

Table S3. Computational parameters used to generate design variants, Related to Figures 2 & 3 and Tables 2-4.

Stabilization target	# designs	$\Delta\Delta G_{calc}$ cutoffs[a] (R.e.u.)
AChE	5	-2.0 (17), -1. 5 (30), -1.0 (42), -0.75 (51), -0.45 (67)
PTE	3	-2.0 (9), -1.0 (19), -0.45 (28)
Dnmt3a	1	-0.45 (14)
SIRT6	1	-0.45 (11)
Myoc-OLF	1	Combined mutations from analysis of design output using 3 PDB entries: 4WXS, 4WXQ, 4WXU (21)

[a] Cutoffs used to define the spaces of potentially stabilizing mutations in computational mutation scanning. Each defined space resulted in a single final variant that was experimentally tested. In Figures in parentheses show the number of mutations relative to starting sequence obtained by each cutoff in the final variant.

Table S4. Production and purification of AChE variants from 250 ml cultures, Related to Figure 2 and Table 2.

Protein	**Mutations relative to hAChE**	**Final Volume (µl)[a]**	**Concentration from activity (µM)[b]**	**% Purity by PAGE gel[c]**	**Yield (µg/ liter culture)**
hAChE[d]	-	850	0.000312	ND[f]	0.064
dAChE1	17	750	1.0102	34	219.2
dAChE2	30	200	0.56	31	121.2
dAChE3	42	350	1.02	72	395.2
dAChE4[e]	51	23000	0.0345	53	1974.8
dAChE5	67	200	0.92	75	118.4

a. Final volume of a pure protein solution obtained from a 250ml *E. coli* culture expressing the protein, after concentration

b. Active AChE protein concentrations obtained from activity measurements on ACh

c. AChE protein purity estimated by quantification of the 60kDa bands in SDS-PAGE gels

d. The purity and yield of hAChE expressed in *E. coli* cells (see extended experimental procedures) were very low relative to other variants

e. This variant was not concentrated following purification

f. Not determined

Table S5. Data collection and refinement statistics for dAChE4 crystallographic analysis, Related to Figure 2D.

Data collection	**dAChE4**
PDB code	5HQ3
Space group	$P4_32_12$
Cell dimensions:	
a,b,c (Å)	89.53, 89.53, 395.30
a,b,g (º)	90, 90, 90
No. of copies in a.u.	2
Resolution (Å)	49.42-.2.60
Upper resolution shell (Å)	2.74-2.60
Unique reflections	50786(7238)*
Completeness (%)	99.9 (99.8)
Multiplicity	9.7(9.9)
Average I / s(I)	6.1 (1.3)
R_{sym} (I) (%)	10.7 (57.6)
Refinement	
Resolution range (Å)	49.42-2.60
No. of reflections (I/s(I) > 0)	48279
No. of reflections in test set	2498
R-working (%) / R-free (%)	20.5 / 25.4
No. of protein atoms	8316
No. of ions/ligands atoms	36
No. of water molecules	1
Overall average B factor ($Å^2$)	41.0
Root mean square deviations:	
- bond length (Å)	0.019
- bond angle (°)	1.956
Ramachandran Plot	
Most favored (%)	88.9
Additionally allowed (%)	10.3
Disallowed (%)	0.2

* Values in parentheses refer to the data of the corresponding upper resolution shell

Table S6. Expression levels and thermal stabilities of Myoc-OLF variants, Related to Figure 3D & E.

Protein Variant	Mutations	Expression[a]	T_m (°C)
hMyoc-OLF	--	2 mg / L	53.0 ± 0.5[b]
hMyoc-OLF + 10mM $CaCl_2$	--	--	59.6 ± 0.2[b]
dMyoc-OLF	21	19 mg / L	69.8 ± 0.8
dMyoc-OLF + 10mM $CaCl_2$	21	--	73.4 ± 0.5

[a] Expression levels quantified by yield of cleaved purified myoc-OLF per liter of cell culture.

[b] Values taken from Donegan *et al*.(Donegan et al., 2012)

Data S1. RosettaScripts, flags and command lines, Related to Experimental procedures.

Data S2. Protein and DNA sequences of experimentally tested constructs, Related to Experimental procedures.

Extended experimental procedures

Experimental procedures

For RosettaScripts, flags, and command lines see **Data S1.**

DNA and protein sequences of all tested designs and their wild-type counterparts are provided in a supplemental file **Data S2**.

Cloning, expression, and purification of AChE variants

The hAChE gene (Uniprot P22303), without its 31aa N-terminal signal-peptide and its 34aa C-terminal tetramerization domain (ORF length 1,647bp), was codon-optimized for expression in *E. coli* and synthesized (Gen9, Boston). The designed hAChE variants were similarly constructed, codon-optimized and synthesized **(Data S2)**. The synthetic genes were cloned into a pET32b+ vector (Novagen, Darmstadt, Germany) in fusion with an N-terminal TRX tag, using *NcoI* and *XhoI* restriction sites. The ligated plasmids were transformed into *E. coli.* SHuffleT7Express cells (NEB, Ipswich, Massachusetts) for expression.

Single colonies expressing hAChE or its designed variants were used to inoculate starter liquid cultures (2YT plus ampicillin (100mg/l) and spectinomycin (100mg/l)) that were grown O/N at 37°C. These were used to inoculate similar cultures of 250ml, grown to OD_{600nm}= 0.5, induced with IPTG (0.45mM) and grown at 20°C for 24 h. Cells were pelleted, dried, frozen at -80°C and lysed by addition of 40ml lysis buffer (20mM Tris-HCl pH 8.0; 100mM NaCl; 10% glycerol; 0.4mg/ml lysozyme; 50U benzonase (Sigma, St. Louis); 10mM EDTA). Lysates were incubated for 15min at 37°C, sonicated (5x 30", 35% amplitude, Sonics-Vibracell VCX750), and then N-octylglucoside (0.1% W/V) and 100 ml of buffer B (10 mM Tris-HCl pH 8.0; 10% glycerol; 10 mM EDTA) were added, and the debris was pelleted. The clarified lysates were loaded onto an affinity column consisting of the affinity ligand *m*-aminophenyltrimethylammonium coupled to Sepharose 4B via a dicaproyl spacer(Sussman et al., 1988); 1ml resin/50ml culture, washed with buffer A (20mM Tris-HCl, pH 8; 30mM NaCl; 10% glycerol; 0.1% N-octylglucoside; 10mM EDTA), and eluted with buffer A supplemented with 50mM tetramethylammonium bromide plus 40mM NaCl. Fractions containing AChE were pooled, concentrated and dialyzed against dialysis buffer (20mM phosphate buffered saline, 10% glycerol, 0.1% N-octylglucoside). Protein

concentration was determined using the BCA assay (Pierce, Waltham, Massachusetts) and purity was assessed by SDS-PAGE.

AChE activity assays in bacterial cell lysates

Transformed *E. coli* colonies (as above) were picked into 96 deep-well plates containing 0.5ml 2YT plus ampicillin and spectinomycin (100 mg/l each), grown to OD_{600nm}=0.5, induced with IPTG (0.45mM), and grown at 20°C for 24 h, followed by 14 h at 16°C. Cells were pelleted, dried and frozen at -80°C. Lysis buffer containing 0.1% v/v Triton X-100 was added (300µl/well), plates were placed in a shaker-incubator (37°C, 1.5 h, 1200 RPM) and the resulting lysates were clarified by centrifugation. Samples of 10µl clear lysates were transferred to Greiner 96-well ELISA plates (Sigma, St. Louis) for activity assays. AChE activity was measured by the addition of 190µl of Ellman's reagent (0.85mM DTNB; Sigma, St. Louis), 0.55mM acetylthiocholine (Sigma, St. Louis), in PBS). Initial velocities were recorded at 412nm using a plate-reading spectrophotometer (PowerWave HT, BioTek, Winooski, Vermont).

Heat inactivation of hAChE designs

E. coli cells, expressing hAChE or its designs, were grown and lysed as described above. Samples of 20µl of clear cell lysates were transferred to a 96-well PCR plate (Axygen, Corning) using a Precision-2000 liquid handler (BioTek, Winooski), incubated in a gradient PCR (Biometra, Göttingen, Germany) for 30min at various temperatures and cooled to 4°C (10min). Samples of 10µl were transferred to 96-well ELISA plates and their AChE activity was measured as described above. The inactivation temperature (denoted as T_m) of the expressed enzymes was calculated by fitting its residual room-temp AChE activity at different temperatures to a 4-parameter Boltzmann sigmoidal curve: $A_t = A_0 + \frac{A_f - A_0}{1+e^{(Tm-T)/m}}$ where A_t corresponds to the residual activity following incubation at a given temperature T, A_0 is the activity of the unheated sample at room temperature, A_f is the activity at maximal inactivating temperature, and m is the sigmoidal slope coefficient. The heat inactivation of purified proteins was similarly measured. For heat-inactivation assays of purified proteins, hAChE (positions 32-574) was cloned in the pHLsec expression vector(Aricescu et al., 2006) and produced in large scale in HEK293T cells. The secreted protein was purified from the medium by affinity chromatography(Sussman et al., 1988) and deglycoasylated with PNGase F.

Determination of AChE kinetic parameters

Purified hAChE or designed variants were diluted in activity buffer (20mM PBS, pH 7.4, 0.1% BSA) (Sigma, St. Louis) to 0.5-3nM. Increasing amounts of the toxic isomer of the organophosphate nerve agent VX (*i.e.* S_P-VX; details of synthesis may be provided upon request) were added up to inhibition of 90% of the initial AChE activity. Residual AChE activity of aliquots was monitored using the Ellman protocol(Ellman et al., 1961). The concentration of AChE active sites was determined by plotting residual AChE activity *versus* S_P-VX concentration and extrapolating to zero AChE activity. V_{max} and K_M were determined in 50mM PBS pH 8.0 at 25^{o}C, by measuring initial velocities of acetylthiocholine hydrolysis (at 0.02 to 0.4mM acetylthiocholine (Sigma, St. Louis)) during the first 1.5min and fitting the data to the Henri-Michaelis-Menten equation(Segel, 1976). k_{cat} was calculated by dividing V_{max} by AChE active-site concentration.

dAChE4 production and purification for X-ray crystallography

Trx-dAChE4 (in *E. coli* SHuffleT7Express cells) was produced in 7.5L LB medium. Cells were grown at 30^{o}C until they reached A_{600}=0.5, and induced with 0.5 mM IPTG for 24 h at 20^{o}C. The published purification protocol for hAChE(Sussman et al., 1988) was modified as follows: The cell pellet was resuspended in lysis buffer (20mM Tris pH 8.0, 100mM NaCl, 10% glycerol, 0.2mg/ml lysozyme, 10mM EDTA, 1μg /ml DNAse) and disrupted by a cell disrupter at 4^{o}C. 0.1% (w/v) *N*-octyl glucoside was added to the cleared lysate and left to rotate for 1 h at 4^{o}C. The solution was diluted by adding 2.5x volumes of buffer (20mM Tris pH 8.0, 10% glycerol, 10mM EDTA and 0.1% (w/v) *N*-octyl glucoside) and clarified by centrifugation at 15,000 g for 30min. Trx-dAChE4 was purified on an affinity column consisting of the ligand (*m*-aminophenyl)trimethylammonium coupled to Sepharose 2B (Pharmacia) (prepared by Lilly Toker according to the manufacturer's instructions). After loading the clarified lysate, the column was washed with buffer containing 20mM Tris pH 8.0, 30mM NaCl, 10% glycerol, 10mM EDTA and 0.1% (w/v) *N*-octyl glucoside, and the enzyme was eluted with the same buffer containing 50mM tetramethylammonium bromide. Fractions containing Trx-dAChE4 were pooled and concentrated before injecting onto a Superdex 200 HR 10/30 column (GE Healthcare) equilibrated with 50mM Tris pH 8.0, 100mM NaCl. Trx-dAChE4 was concentrated to 10 mg/ml for crystallization. Spontaneous cleavage of the Trx occurred during the process.

Structure determination and refinement of dAChE4

Crystals of dAChE4 were obtained using the sitting-drop vapor-diffusion method with a Mosquito robot (TTP LabTech). The crystals were grown from 8% PEG 6000, 0.1M $MgCl_2$ and 0.1M MES pH 6.0. The crystals formed in the tetragonal space group $P4_32_12$, with 2 monomers per asymmetric unit. A complete dataset to 2.6 Å resolution was collected at 100 K using a single crystal on beamline BM14 of the European Synchrotron Radiation Facility (ESRF, Grenoble). Diffraction images were indexed and integrated using Mosflm(Leslie and Powell, 2007), and the integrated reflections were scaled using SCALA(Evans, 2006). Structure-factor amplitudes were calculated using TRUNCATE (CCP4 program suite)(French and Wilson, 1978). The structure was solved with PHASER(McCoy, 2006) using the refined structure of hAChE (PDB entry: 4EY4(Cheung et al., 2012)) as a molecular-replacement model. All atomic-refinement steps were carried out with CCP4/REFMAC5(Murshudov et al., 1997). The model was built into $2mF_{obs}$ - DF_{calc}, and mF_{obs} - DF_{calc} maps using COOT(Emsley and Cowtan, 2004). The R_{free} value is 25.4% (for the 5% of reflections not used in the refinement), and the R_{work} value is 20.5 % for all data to 2.6 Å. The dAChE4 model was evaluated with MOLPROBITY(Chen et al., 2010). Details of the refinement statistics are described in **Table S5**. The coordinates of dAChE4 were deposited in the RCSB Protein Data Bank as entry 5HQ3.

AChE inhibition by S_P-VX

The concentration of the toxic enantiomer (S_P) of VX [*O*-ethyl *S*-(2-(diisopropylaminoethyl) methylphosphonothioate] was determined by monitoring thiol released in the presence of NaF, and verified by use of the PTE variant C23(Cherny et al., 2013). The pseudo first-order rate constant (k_{obs}) for hAChE inhibition (0.2-1nM in activity buffer) by ≥10-fold stoichiometric excess of S_P-VX increased linearly with increasing S_P-VX. Since the inhibitor concentration was well below what would be expected from the dissociation constant of its reversible complex with hAChE(Berman and Leonard, 1989), the second-order rate constant for inhibition was obtained by dividing k_{obs} by S_P-VX concentration.

PTE activity assays in bacterial cell lysates

E.coli GG48 cells(Grass et al., 2001) expressing PTE-S5 or its designed variants were plated on LB plus ampicillin(100mg/l) plus kanamycin(50mg/l) agar, and grown overnight at 37°C. Colonies were randomly picked into 96 deep-well plates (Axygen, Union City, California)

containing 0.5ml 2YT plus ampicillin (100mg/l) and kanamycin (50mg/l), grown to OD_{600nm}=0.5, induced by adding IPTG (0.5mM, final concentration), and further grown at 20°C for 24 h. The cells were then pelleted, dried, and frozen at -80°C. Plates were defrosted at room temperature, and cells were lysed by addition of lysis buffer (0.1M Tris-HCl pH 8.0, 0.1M NaCl 1, 0.1% Triton X-100; 10mM Na_2CO_3, 0.2mg/ml lysozyme and 50U benzonase nuclease) followed by incubation in a shaker-incubator at 37°C, for 1.5 h at 1200 RPM. Lysates were clarified by centrifugation. Samples of clear lysates were diluted 1:200 in PTE activity buffer, and 5µl dilute samples were transferred to Greiner 96-well plates (Sigma, St. Louis) for activity assays. PTE activity on paraoxonase was measured by addition of 195µl, 0.2mM paraoxon in 0.1M Tris pH 8.0, plus 0.1M NaCl. Initial velocities were recorded at 405nm using a plate-reading spectrophotometer (PowerWave HT, BioTek, Winooski, Vermont).

PTE heat-inactivation temperature and metal-binding affinity

Heat-inactivation temperature was assayed as described above for AChE, except that residual activity was measured using paraoxon as substrate (described above). Residual activity at increasing concentrations of the metal-binding chelator 1,10-phenanthroline was assayed as described(Roodveldt and Tawfik, 2005). Briefly, clarified cell lysates were incubated for 90 min at room temperature with 10-500µM of 1,10-phenanthroline. All samples, including the untreated sample, contained up to 0.5% methanol. The samples were subsequently diluted 1:200 in activity buffer, incubated for 30 min at room temperature, and assayed with 0.25mM paraoxon. Data were fit to a 4-parameter Boltzmann sigmoid: $A_c = A_0 + \frac{A_f - A_0}{1+e^{(C_m - C)/m}}$ where A_c corresponds to the residual activity following incubation at a given 1,10-phenanthroline concentration, C, A_0 is the activity of the sample without 1,10-phenanthroline, A_f is the activity at maximal 1,10-phenanthroline concentration, C_m , is the mid-inactivation concentration, and *m* is the sigmoidal slope coefficient. Residual activity as a function of the time of incubation with 1,10-phenanthroline was assayed at 0.25µM PTE (wild-type-like PTE-S5, or the designed variant, PTE-dPTE2) and 50µM 1,10-phenantroline at room temperature. Aliquots were removed at various time points and residual activity was determined by measuring the initial rates of paraoxon hydrolysis. Residual activity at each time point was determined, and the relative activity compared to a sample incubated without the chelator was calculated. Data were fit to a single exponential decay curve: $A_t = A_0 e^{-kt}$,

where A_t corresponds to the residual activity after incubation for a given time, t, and A_0 is the activity at time zero, and *k* is the rate constant for inactivation

Expression and purification of recombinant hSIRT6 and dSIRT6

The dSIRT6 gene was synthesized by GenScript **(Data S2)** and cloned into pMAL C2x. pMAL plasmids carrying the genes hSIRT6 and dSIRT6 were transformed into Rosetta *E. coli* competent cells. Cells were grown at 37°C in TB medium containing ampicillin and chloramphenicol to A_{600}=0.6 and induced with 0.1 mM IPTG followed by 16 h incubation at 16°C. Cells were lysed using a French Press in lysis buffer (20 mM Tris-HCl pH 7.5, 100mM NaCl, and EDTA free-protease inhibitor cocktail (Calbiochem)). Cell debris was removed by centrifugation at 16,000g at 4°C for 40 min. The supernatants were loaded onto a pre-equilibrated amylose column eluted with elution buffer (20 mM Tris-HCl, 100 mM NaCl and 20 mM maltose pH 7.5). Protein purity was assessed by SDS-PAGE. The protein concentration was determined by the BCA method using bovine serum albumin as the standard.

Fluor de lys activity assay for SIRT6

Lysine-myristoylated TNFα peptide was conjugated to 4-amino-7-methylcoumarin group (AMC) at its C terminus (Peptron, Daejeon, South Korea). The fluor de lys activity assay was performed as described before(Wegener et al., 2003). Briefly, purified hSIRT6 and dSIRT6 were incubated at 37°C in reaction buffer containing 12.5 mM NAD+ and 1.25 mM TNFα myr-AMC peptide in assay buffer (50 mM Tris-HCl pH 8.0, 137 mM NaCl, 2.7 mM KCl, 1 mM MgCl2, 1 mg/ml BSA). At each time point, a 50µl aliquot was removed and mixed with 50µl of the developer solution (assay buffer+50 µM of HCl, 3 mg/ml trypsin and 10 mM nicotine amide). The quenched samples were kept at 37°C for 20 min prior to fluorescence measurements. Fluorescence readings were obtained using the TECAN infinite series 200 fluorimeter with excitation at 360nm and emission at 460nm in black 96-well solid plates (Greiner).

SIRT6 expression and activity in mammalian cell cultures

HEK293T and U2OS cell lines were obtained from the American Type Culture Collection. SIRT6 KO U2OS were generated using CRISPR-Cas9 gene disruption. Lentiviruses bearing SIRT6 CRISPR guiding sequence were produced in HEK293T cells and used to transduce

U2OS cells. Virally transduced cells were selected by the addition of 1 μg/ml puromycin (Sigma). All cells were grown in Dulbecco's modified Eagle's medium supplemented with 10% FBS, 100U/ml penicillin, 100mg/ml streptomycin and 2mM L-glutamine (all purchased from Biological Industries, Israel) and maintained at 37ºC under with 5% CO2.

Antibodies against SIRT6, H3, β-actin, and H3K56Ac were purchased from Abcam. An anti-FLAG antibody was purchased from Sigma.

Human SIRT6-FLAG variants cDNA sequences were cloned into pWZL-hygro for stable or transient transfection. Trans IT-LT1 (Mirus) was used to transfect HEK293T cells according to the manufacturer's protocol. Retroviruses bearing SIRT6-FLAG variants were produced in HEK293T cells and used to transduce a U2OS SIRT6 KO cell line. Virally transduced cells were selected for by the addition of 200μg/ml hygromycin B (Gold Biotechnology).

SIRT6 quantification by Western blot

Equal band areas on the Western membrane were used to measure the band intensities of hSIRT6, E1 and dE1 using the ImageJ software. Equal empty membrane areas around the bands were used to measure background intensities that were then subtracted from the band intensity values. The obtained values were normalized to actin levels by application of the same quantification method.

Expression and purification of Dnmt3a catalytic domains

Expression was performed essentially as described(Gowher and Jeltsch, 2001). Briefly, plasmids (peT28a) encoding hDnmt3a and dDnmt3a constructs were transformed into BL21(DE3)codon+ RIL cells. Bacteria were grown in TB medium. Protein expression was induced by addition of 0.5 mM IPTG. Following growth for another 13 h at 20°C, bacteria were harvested by centrifugation, washed with 100 mM NaCl, 1 mM EDTA, 10 mM Tris/HCl pH 8.0, and frozen at −20°C. Cells were lysed by sonication and the lysate was clarified by centrifugation for 1 h at 4°C, 20,000 rpm. The enzyme was purified on Ni-NTA agarose beads (Qiagen) at 4°C, eluted with 220 mM imidazole and dialyzed against 20mM Hepes, 200mM KCl, 0.2mM DTT, 1mM EDTA, 10% glycerol, pH 7.0. Due to the presence of contaminating bands, hDnmt3a and dDnmt3a concentrations were estimated by comparing their SDS-PAGE gel bands to various known amounts of bovine serum albumin (BSA).

Dnmt3a DNA-methylation assay

Methylation rates were measured with tritium-labeled SAM (2 μM), substrate DNA (509 bp biotinylated-DNA substrate containing 58 GpC sites; 0.4μM) and purified enzymes (0.32μM) essentially as described(Roth and Jeltsch, 2000). Aliquots taken at various time points were quenched by and transferred to streptavidin-coated ScintiPlate wells (PerkinElmer) and H^3 levels were determined using the Wallac MicroBeta TriLux counter (Perkin Elmer).

Myocilin-OLF expression, purification and thermal stability

dMyoc-OLF was synthesized and cloned into a pMal-c4x vector by GenScript **(Data S2)** utilizing MBP as a fusion protein for enhanced solubility, and a short amino acid linker as described previously for wild-type myoc-OLF(Burns et al., 2011, 2010). Fidelity of the plasmid was confirmed by DNA sequencing (MWG Operon). *E. coli* expression, purification, and cleavage from MBP, as well as determination of T_ms in 10mM Hepes pH 7.5, 200mM NaCl using differential scanning fluorimetry, were performed as described previously(Burns et al., 2011, 2010; Donegan et al., 2012).

Literature

Aricescu, A.R., Lu, W., Jones, E.Y., 2006. A time- and cost-efficient system for high-level protein production in mammalian cells. Acta Crystallogr. D. Biol. Crystallogr. 62, 1243–50. doi:10.1107/S0907444906029799

Berman, H. a., Leonard, K., 1989. Chiral reactions of acetylcholinesterase probed with enantiomeric methylphosphonothioates. Noncovalent determinants of enzyme chirality. J. Biol. Chem. 264, 3942–3950.

Burns, J.N., Orwig, S.D., Harris, J.L., Watkins, J.D., Vollrath, D., Lieberman, R.L., 2010. Rescue of glaucoma-causing mutant myocilin thermal stability by chemical chaperones. ACS Chem Biol 5, 477–487. doi:10.1021/cb900282e

Burns, J.N., Turnage, K.C., Walker, C.A., Lieberman, R.L., 2011. The stability of myocilin olfactomedin domain variants provides new insight into glaucoma as a protein misfolding disorder. Biochemistry 50, 5824–33. doi:10.1021/bi200231x

Chen, V.B., Arendall, W.B., Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J., Murray, L.W., Richardson, J.S., Richardson, D.C., 2010. MolProbity: All-atom structure validation for macromolecular crystallography. Acta Crystallogr. Sect. D Biol. Crystallogr. 66, 12–21. doi:10.1107/S0907444909042073

Cherny, I., Greisen, P., Ashani, Y., Khare, S.D., Oberdorfer, G., Leader, H., Baker, D., Tawfik, D.S., 2013. Engineering V-type nerve agents detoxifying enzymes using computationally focused libraries. ACS Chem. Biol. 8, 2394–2403. doi:10.1021/cb4004892

Cheung, J., Rudolph, M.J., Burshteyn, F., Cassidy, M.S., Gary, E.N., Love, J., Franklin, M.C., Height, J.J., 2012. Structures of human acetylcholinesterase in complex with pharmacologically important ligands. J. Med. Chem. 55, 10282–10286. doi:10.1021/jm300871x

Donegan, R.K., Hill, S.E., Turnage, K.C., Orwig, S.D., Lieberman, R.L., 2012. The glaucoma-associated olfactomedin domain of myocilin is a novel calcium binding protein. J. Biol. Chem. 287, 43370–43377. doi:10.1074/jbc.M112.408906

Ellman, G.L., Courtney, K.D., Andres, V., Featherstone, R.M., 1961. A new and rapid colorimetric determination of acetylcholinesterase activity. Biochem. Pharmacol. 7, 88–95. doi:10.1016/0006-2952(61)90145-9

Emsley, P., Cowtan, K., 2004. Coot: Model-building tools for molecular graphics. Acta Crystallogr. Sect. D Biol. Crystallogr. 60, 2126–2132. doi:10.1107/S0907444904019158

Evans, P., 2006. Scaling and assessment of data quality, in: Acta Crystallographica Section

D: Biological Crystallography. pp. 72–82. doi:10.1107/S0907444905036693

French, S., Wilson, K., 1978. On the treatment of negative intensity observations. Acta Crystallogr. Sect. A 34, 517–525. doi:10.1107/S0567739478001114

Gowher, H., Jeltsch, A., 2001. Enzymatic properties of recombinant Dnmt3a DNA methyltransferase from mouse: the enzyme modifies DNA in a non-processive manner and also methylates non-CpA sites. J. Mol. Biol. 309, 1201–1208. doi:10.1006/jmbi.2001.4710

Grass, G., Fan, B., Rosen, B.P., Franke, S., Nies, D.H., Rensing, C., 2001. ZitB (YbgR), a member of the cation diffusion facilitator family, is an additional zinc transporter in Escherichia coli. J. Bacteriol. 183, 4664–7. doi:10.1128/JB.183.15.4664-4667.2001

Leslie, A.G.W., Powell, H.R., 2007. Evolving Methods for Macromolecular Crystallography, Evolving Methods for Macromolecular Crystallography. doi:10.1007/978-1-4020-6316-9

McCoy, A.J., 2006. Solving structures of protein complexes by molecular replacement with Phaser, in: Acta Crystallographica Section D: Biological Crystallography. pp. 32–41. doi:10.1107/S0907444906045975

Murshudov, G.N., Vagin, A.A., Dodson, E.J., 1997. Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr. Sect. D Biol. Crystallogr. doi:10.1107/S0907444996012255

Roodveldt, C., Tawfik, D.S., 2005. Directed evolution of phosphotriesterase from Pseudomonas diminuta for heterologous expression in Escherichia coli results in stabilization of the metal-free state. Protein Eng. Des. Sel. 18, 51–58. doi:10.1093/protein/gzi005

Roth, M., Jeltsch, a, 2000. Biotin-avidin microplate assay for the quantitative analysis of enzymatic methylation of DNA by DNA methyltransferases. Biol. Chem. 381, 269–72. doi:10.1515/BC.2000.035

Segel, I.H., 1976. Biochemical Calculations, 2nd Edition, John Wiley and Sons Inc, NY. doi:10.1016/0307-4412(76)90087-X

Sussman, J.L., Harel, M., Frolow, F., Varon, L., Toker, L., Futerman, A.H., Silman, I., 1988. Purification and crystallization of a dimeric form of acetylcholinesterase from Torpedo californica subsequent to solubilization with phosphatidylinositol-specific phospholipase C. J Mol Biol 203, 821–823.

Wegener, D., Wirsching, F., Riester, D., Schwienhorst, A., 2003. A Fluorogenic Histone Deacetylase Assay Well Suited for High-Throughput Activity Screening. Chem. Biol.

10, 61–68. doi:10.1016/S1074-5521(02)00305-8