Chapter 10

Strategies for Increasing Protein Stability

Peter G. Chandler, Sebastian S. Broendum, Blake T. Riley, Matthew A. Spence, Colin J. Jackson, Sheena McGowan, and Ashley M. Buckle

Abstract

The stability of wild-type proteins is often a hurdle to their practical use in research, industry, and medicine. The route to engineering stability of a protein of interest lies largely with the available data. Where high-resolution structural data is available, rational design, based on fundamental principles of protein chemistry, can improve protein stability. Recent advances in computational biology and the use of nonnatural amino acids have also provided novel rational methods for improving protein stability. Likewise, the explosion of sequence and structural data available in public databases, in combination with improvements in freely available computational tools, has produced accessible phylogenetic approaches. Trawling modern sequence databases can identify the thermostable homologs of a target protein, and evolutionary data can be quickly generated using available phylogenetic tools. Grafting features from those thermostable homologs or ancestors provides stability improvement through a semi-rational approach. Further, molecular techniques such as directed evolution have shown great promise in delivering designer proteins. These strategies are well documented and newly accessible to the molecular biologist, allowing for rapid enhancements of protein stability.

Key words Protein stability, Rational design, Consensus design, Ancestral reconstruction, Semi-rational design, Directed evolution

1 Introduction

The ability to tailor the properties of a protein for extended working life and improved activity in unfavorable conditions has numerous applications for technology, biotechnology, and biomedicine. Additionally, successfully manipulating a protein's physicochemical properties or activity often requires a protein scaffold that is thermodynamically stable. Therefore, a major hurdle for protein engineering is the introduction or improvement of protein stability to the target protein. However, there is no one-size-fits-all strategy for enhancing protein stability, and many tools are now available.

Juliet A. Gerrard and Laura J. Domigan (eds.), *Protein Nanotechnology: Protocols, Instrumentation, and Applications*, Methods in Molecular Biology, vol. 2073, https://doi.org/10.1007/978-1-4939-9869-2_10,

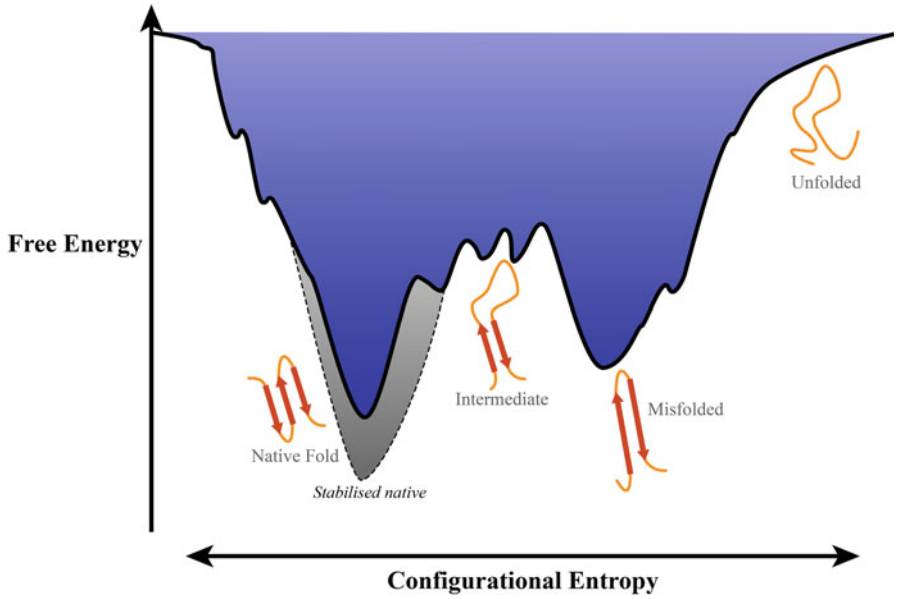

Fig. 1 Stabilization of the protein native state. Folding states of a protein exist along a landscape characterized by the free energy of each fold. Improving stability involves stabilizing the native state so it will remain functional in harsh environments

Proteins exist in a dynamic equilibrium populated by non-native species, partially folded intermediates, and fully folded native species (Fig. 1). Each of these species exists on an energy landscape, and a protein will transition between these states as it unfolds and refolds. Preservation of functional protein in harsh conditions, such as high temperatures or denaturing chemicals, can be achieved by ensuring the functional protein fold is the most favorable in these conditions (Fig. 1). Since in the vast majority of cases only the native state is functional, protein stabilization targets two broad goals: increasing the population of natively folded protein by stabilizing a native state or decreasing the population of non-native folded protein by destabilizing intermediates and misfolded states [1]. Although the primary aim of stabilization is focused on the native state, the strategies detailed may also reduce occurrence of misfolded states.

In this review, we aim to highlight some of the most accessible methods that can be used to enhance protein stability and provide an overview of the supporting theory behind each approach.

2 Methods

2.1 Rational, Structure-Based Approach

Evolution cares only about protein function; there is no selective pressure to maintain protein stability beyond what is sufficient for function [2]. As a result, proteins are generally metastable and often

contain regions that are not optimized for stability. Rational engineering aims to identify these unstable regions and install new structural features that will contribute to stability in a predictable manner. These approaches rely heavily on good-quality, high-resolution structural data for the target protein to successfully add new structural features which improve protein stability.

2.1.1 Find Sites with Suboptimal Contribution to Stability

There are multiple ways of identifying target regions of a protein for rational improvement of protein stability. Commonly used methods include analyzing the flexibility of the structure, assessing the available evolutionary data through phylogenetic analysis, and using in silico predictors for residues that contribute to folding stability.

Protein flexibility is often detrimental to stability, and flexible regions may, therefore, be good targets for rational intervention. One way of identifying regions of increased flexibility is to analyze crystallographic temperature factors (B-factors) in a protein structure, as flexibility is often correlated with high B-factors [3, 4]. The B-FIT computational tool can identify and report regions with high B-factors from any available coordinate file (https://www.kofo.mpg.de/en/research/biocatalysis) [5]. A more computationally expensive method to identify flexible regions within a protein is molecular dynamics (MD) simulations. Analysis of the root mean square deviation (RMSD) and the root mean square fluctuation (RMSF) of MD trajectories can identify overall protein dynamics (RMSD) or even identify individual residues that show higher than average flexibility (RMSF).

Conservation of sequence and structure generally implies that the region makes a significant contribution to fold or function. Therefore, regions that show little conservation or high divergence may be good targets for rational design. The ConSurf [6] online tool takes the protein structure of interest, performs a search for structural homologs, and allows a visual inspection of a multiple sequence alignment (MSA) as well as mapping the conservation back onto the target structure. Other computational tools can simply analyze the stability of the protein structure and suggest improvements without reference to homologs or sequence conservation. The FoldX [7] and RosettaDesign [8] computational tools calculate the Gibbs free energy of folding for the protein of interest and use an energy function to identify amino acids that are important for stability. The impact on the stability of the protein from designed mutations can be predicted from calculation of the changes in energy after mutation [7, 9].

2.1.2 Design a Novel Structural Feature for Stabilization

Several features can be integrated into a protein to increase its stability. Analysis of thermostable proteins shows that these proteins often have more salt bridges, hydrogen bonds, and improved hydrophobic packing [10, 11], features that can be introduced into

the protein of interest. The addition of salt bridges and/or disulfide bond "staples" between appropriately spaced residues, may reduce flexibility [12]. Shortening or increasing the proline content of loops can also rigidify highly flexible regions [3]. However, the stability benefit of adding disulfide bonds may be complicated by the requirement for more complex systems required for correct expression.

While protein "staples" can significantly increase the stability of a protein, the effective use of disulfide staples depends on the number and location of native cysteine residues [12]. If the protein of interest has functional cysteine residues or even a high content of cysteine residues, it is often not practical to introduce a disulfide pair. However, recent advances in molecular and chemical biology make use of other chemical functionalities and allow crosslinking "staples" to be introduced via targeted incorporation of nonnatural amino acids (NNAs) into the protein [13]. Compared to native cysteine staples, NNA staples provide better specificity, by removing the potential for off-target crosslinks, reducing the likelihood of misfolding, and adding a degree of temporal control [14]. To allow precise control of where the staple is located, the use of two NNAs, introduced at selected locations, allows the protein to be stapled by various click chemistries (as available on the chosen NNAs). The location, spatial separation, and stereochemistry of the NNAs to be introduced should be selected to minimize the impact on the overall structure.

Computational assistance for the positioning and design of protein staples (either disulfides or NNA chemistries) can be obtained using RosettaMatch software, which will also provide a theoretical change in Gibbs free energy (ΔG) for each generated solution [12]. It is also worthwhile, once a new staple has been designed, to integrate the feature into the protein structure using MODELLER [15]. The theoretical new construct can then be visually inspected with PyMOL [16] or with other computational tools to predict the effect of the new structural element on protein stability.

2.1.3 *Validating Changes in Stability*

Applied on a large scale, computational methods can be used as a powerful triage tool. In silico methods can provide predictions on a vast array of candidate stabilizing mutations to funnel them into a manageable number of candidates for translation into laboratory experiments. Mutations that increase the size of a sidechain may result in severe clashes with surrounding residues that need to be gradually relaxed. Such modifications can be tested computationally using protocols whereby the nonnatural amino acid is gradually inserted into MD simulations [17]. Ultimately, validation of any of the approaches outlined above requires experimental data. Once in the laboratory, the stability of mutant proteins can be characterized

by measuring several properties. The first, and most important, indicator of stability is that the protein can be expressed and purified. Once protein has been produced, traditional methods for assessment of protein stability include the measurement of the thermal melting point (T_m). T_m can be easily and rapidly measured by circular dichroism (CD) spectroscopy [18] or differential scanning fluorimetry (DSF) [19]. More rigorous and accurate assessment of stability comes from the measurement of the free energy of folding ($\Delta G_{D\rightarrow N}$), using equilibrium unfolding experiments with chemical denaturants [20].

2.2 Sequence-Based Approaches

Billions of years of diversification and natural selection has produced a plethora of amino acid sequences. These sequences each encode the structure of a naturally occurring protein fold [21], enabling a functionally viable protein. In mapping the relationship between protein sequence, structure, and function, modern bioinformatics databases contain a wealth of information to be exploited. Phylogenetic techniques capture the evolutionary relationships contained within these databases and allow these relationships to be used in the design and engineering of proteins for selected properties, including stability.

2.2.1 Generating Stable Proteins from Phylogenetic Data

Two well-established techniques—consensus design and ancestral sequence reconstruction (ASR)—use sequence information to design stable homologs. From a sequence alignment and phylogenetic tree of extant proteins, these techniques either reconstruct the most likely sequence of a common ancestral protein or perform a simple statistical analysis producing a consensus sequence [22] (Fig. 2).

2.2.2 Produce a Multiple Sequence Alignment

The essential step in both consensus design and ancestral reconstruction methods is the construction of a multiple sequence alignment (MSA) for the protein family of interest. Since the quality of alignment required varies between these techniques, it is recommended to try different methods of alignment as well as removing duplicate species or sequences that do not align well. In an ideal MSA, each sequence is aligned optimally to all other sequences. MSAs are, therefore, a concise description of the evolutionary relatedness and sequence similarity within a protein family.

ASR and consensus methods use the information encoded in an MSA differently. ASR uses an MSA to determine phylogenetic relationships between sequences and "resurrect" the most likely ancestor of the given set of sequences. Consensus design, on the other hand, uses an MSA to identify the most frequently observed amino acid at each position to build a consensus sequence (Fig. 2). Depending on the protein family of interest, creating reliable MSAs can be nontrivial. In this section, we have provided only an overview of important considerations to be made in the construction of

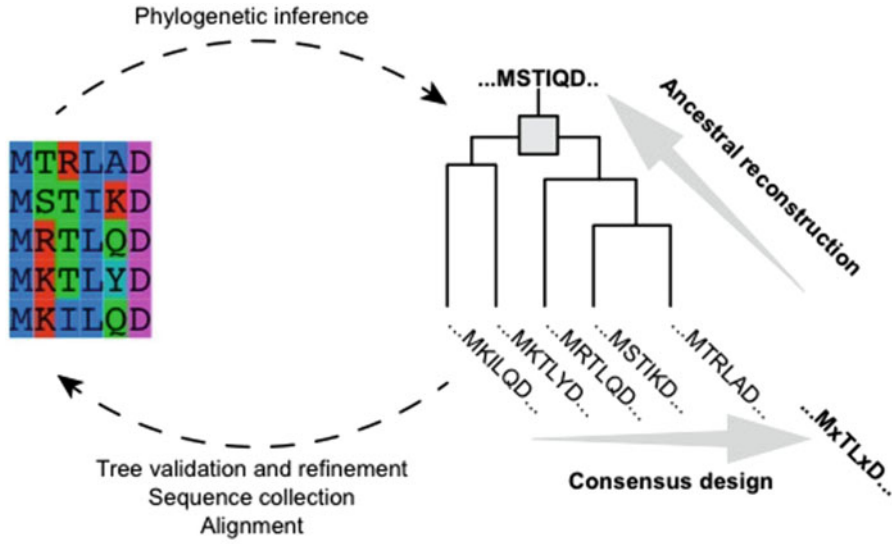

Fig. 2 Both ASR and consensus design leverage phylogenetic information from contemporary protein families. ASR infers a putative ancestral sequence from a phylogenetic tree of extant proteins, whereas consensus design yields the most common amino acid from each position in an alignment of extant proteins while making no evolutionary inferences. In either case, the often iterative process of producing a robust sequence alignment (and phylogenetic tree for ASR) is imperative for reliable results [23]

an MSA for ASR/consensus. Previous editions of this series have included instructions for creating a reliable MSA [24].

The production of an MSA can be broken down into three steps: collection of sequences, filtering, and alignment. The genomics era has led to an explosion of non-curated protein sequences in databases [25]. From a single query sequence, or seed, homologous sequences can be identified and collated from the NCBI databases using the BLAST server [26]. The traditional BLASTp will identify close homologs, while the Position-Specific Iterated (PSI)-BLAST server can be helpful in identifying distantly related proteins that may be otherwise overlooked by the standard BLASTp. Alternatively, for many protein families, highly annotated databases like PFAM [27], PROSITE [28], and SUPERFAMILY [29] contain small, manually curated seed alignments. In concert with the HMMER webserver [30], these seed alignments can be used to generate a hidden Markov model (HMM) sequence profile for the family. This curated HMM profile can then be used to collect more closely related sequences from larger databases, such as UniProtKB's Swiss-Prot and TrEMBL [25].

For ASR and consensus design, it is important to consider the effect of sampling from large databases (e.g., NCBI) on sequence diversity within the MSA. From a practical standpoint, a more diverse set of sequences will present a greater challenge for alignment. However, from a theoretical standpoint, a more diverse set of

sequences samples a larger sequence space and hence also samples a deeper evolutionary history from which consensus and ancestral techniques can derive information. Another consideration, when collecting sequences, is that larger sets of sequences may also contain significant bias toward sequences from a particular clade or organism [31]. To minimize biases caused by sampling errors, sequence redundancy should be removed using CD-HIT [31], typically to a similarity threshold of 90% sequence identity prior to the construction of the initial MSA. Sequence similarity networks that cluster sequences are particularly helpful in refining sequence datasets collected straight from PFAM. These networks can be generated from a PFAM number using the EFI-EST server [32] and visualized as a force-directed graph in Cytoscape [33].

Once the sequences have been collected and initially filtered, they can be aligned using a variety of alignment algorithms which differ in their computational complexity and accuracy [34]. The choice of alignment software is largely dependent on the phylogenetic diversity of the dataset. A collection of closely related sequences that typically share >35% sequence identity may be adequately aligned using a fast and computationally efficient alignment algorithm such as MUSCLE [35], whereas alignment of evolutionarily divergent sequences that share <35% sequence identity could necessitate the use of more computationally intensive alignment algorithms, for example T-COFFEE [36]. For datasets that feature highly divergent sequences that fail to align accurately even in the default mode of T-COFFEE, the EXPRESSO mode of T-COFFEE incorporates putative structural information from the PDB to improve the accuracy of the alignment. The final MSA should be benchmarked against biological and structural features to verify the accuracy of the alignment. Residues known a priori to be conserved across homologs (such as active site residues) should be aligned to the produced MSA and verified to be in the correct position. A structural alignment of representative crystal structures (if available) can also be helpful in validation. Servers such as NCBI's VAST+ [37] are useful in identifying homologous crystal structures for structural benchmarking.

2.2.3 Produce a Consensus Sequence

Procedurally, consensus design is simple—once all sequences in a given dataset have been aligned, the most frequently occurring amino acid in each column (a consensus residue) is taken to form a consensus sequence [21, 38]. As a result of evolutionary pressures on protein stability, the most common residue at any position is more likely to be stabilizing (50%) [39] than a random mutation (8–29%) [40]. Consensus design exploits this emergent phenomenon to increase protein stability [41, 42]. Once constructed, a consensus sequence can be used in one of two ways: as a de novo sequence [42–50] or, in a more conservative approach, as a method

for identifying "back-to-consensus" stabilizing mutations for a target protein [39, 51–53]. Both of these techniques have proven successful in increasing protein stability [23].

2.2.4 Produce an Ancestral Sequence

ASR is a computational method for inferring the most likely sequence of a precursor protein from a sequence alignment of its extant descendants. A feature of ASR is the generation of consistently thermostable ancestral proteins that frequently unfold at temperatures exceeding their contemporary counterparts [23, 54–57]. It is unclear whether the elevated thermostability of reconstructed proteins is a reflection of harsh prehistoric climates that favored thermophilic organisms and thermostable proteins [22, 58] or an artificial exaggeration stemming from systematic biases in reconstruction algorithms [22, 55, 59]. Nevertheless, ASR is a valuable tool for increasing the stability of a protein scaffold.

The quality of reconstructed ancestral sequences is contingent on a robust phylogenetic tree. The maximum likelihood (ML) statistical framework for tree inference produces the single most probable phylogenetic tree from a sequence alignment and is the most appropriate method for ASR [60]. The ML phylogenetic tree can be reconstructed by IQ-TREE, which also fits a sequence evolution model to the alignment data [61, 62]. Once computed, the ML phylogenetic tree can be viewed in FigTree [63]. As with the initial alignment, tree inference is often a lengthy process that may require multiple iterations of evaluation, sequence collection and alignment, and recalculation under various sequence evolution models before a satisfactory tree is produced. Interested readers should refer to Yang and Rannala [64] for a more extensive review on the principles of molecular phylogenetics.

An inherent challenge in phylogenetic inference is tree validation. Nonparametric bootstrapping is a popular method for assessment of the statistical support for each branch reconstructed in the tree; however, this method can introduce bias if model assumptions are severely violated [65]. The complexity of bootstrapping simulations may also create a computational bottleneck for tree inferences, particularly for larger datasets. An alternative to nonparametric bootstrapping is ultrafast bootstrap approximation, which reduces the computational demand of nonparametric bootstrapping while maintaining robustness to slight model violations [61, 66]. For reliable results, nonparametric bootstrapping and ultrafast bootstrap approximations should be performed to no less than 100 and 1000 replicates, respectively.

In general, an ancestral sequence can be confidently reconstructed from a phylogenetic node that has a nonparametric bootstrap of 80 or higher or an ultrafast bootstrap approximation 95 or higher. Bootstrap values below these indicate poor phylogenetic

support and possible misplacement of sequences within the clade. Approximated likelihood ratio test (SH-aLRT) statistics can also be calculated faster than both bootstrapping methods to quickly assess single branch support throughout the tree. Ancestral sequences are reconstructed using a combination of Bayesian and ML statistics in a method most accurately described as empirical Bayesian [67]. The reconstruction can be performed with CodeML from the PAML software package [68]. The output from the reconstruction is a single ML ancestral amino acid sequence and a posterior probability distribution of all possible amino acids at each position for each ancestral node in the phylogenetic tree. It is highly likely that enhanced thermostability will be a property of the ML sequence. A more detailed review on the theory and limitations of ASR can be found in Joy et al. [67].

The statistical confidence of a reconstruction can be measured as the posterior probability averaged over each residue in the reconstructed sequence. Ancestral sequences are generally considered poorly supported if their mean posterior probability is 70–80%, moderate if 80–90%, and highly supported if >90% [69]. Confidence in both the phylogenetic node (nonparametric or ultrafast bootstrap) and the reconstruction (mean posterior probability) should be considered independently when assessing the overall quality of reconstructed ancestral sequences, as the two measures may not necessarily correlate.

2.3 Future Methods: Ab Initio Generation of Stable Folds

Existing protein sequences, and those which can be produced by laboratory experiments (Subheading 2.6), are generated by evolution from other extant sequences. The iterative process of evolution renders certain sequences inaccessible, some of which may have increased protein stability. Ab initio computational methods allow researchers to create these inaccessible sequences by relying on rational principles of protein stability [70]. This approach generates sequences without evolutionary bias and folds them to match a given protein backbone and then ranks the sequences according to an energy scoring function. A number of tools have been built for this application including RosettaDesign [8], ORBIT [71], Liang and Grishin's algorithm [72], or EGAD [73]. This method should find that some wild-type regions are already the most optimal for stability, as a rule of thumb promising ab initio sequences display over 35% similarity to the wild-type protein [74].

However, researchers should be mindful of what they expect from this technique. Previous investigators found that only half of their ab initio proteins display increased stability [75] and the active site may need to remain unchanged to ensure function [76]. The relative ease of providing a protein fold to an algorithm does not consider required work afterward for validation and rescue of function. As a result, the authors submit that this method may require

further development before it can be considered a robust technique for protein stabilization.

2.4 Semi-rational Approaches

Rational approaches described previously are useful for identifying regions of a protein that can be targeted for stability enhancement. Ways in which these regions can be improved can be found through analysis of extant sequences and phylogenetic inference of thermostable sequences or through a directed evolution experiment.

2.5 Grafting from Thermostable Homologs

Grafting refers to the production of proteins carrying subsequences that have been deliberately transferred from another protein. To improve protein stability, grafting allows regions of suboptimal stability to be removed and replaced with an equivalent region from a thermostable homolog. Thermophilic organisms have experienced a selective pressure to create proteins with higher thermostability [77]. As the grafted region is from an extant stable protein from the same protein family, there is a higher chance that the graft will produce a protein with improved stability.

2.5.1 Source Information of a Thermostable Homolog to the Protein

Thermostable homologs can be found in nature [78, 79] or can be designed [80–84]. Information on homologs from thermophilic organisms is readily and publicly available [85], with sequences available from curated databases such as ProTherm (http://www.abren.net/protherm/protherm.php). Ancestral or consensus sequences can be generated as described in Subheading 2.2 of this review.

2.5.2 Design Grafts of Thermostable Features

Regions to graft can be identified, as described in Subheading 2.1.1, by comparison of Gibbs free energy of folding, visual inspection of structural features, or sequence comparison. Thermophilic proteins often have design features such as shortened loops or an increased number of salt bridges [78, 85] which can be promising candidates for grafting and can be identified in a structure or sequence alignment. Sequence comparison between the extant protein and thermostable relative(s) will also identify promising candidates for an ancestral or back-to-consensus graft [83]. These grafts can span single residues [80] and entire sections [86], and some approaches work in reverse by grafting an active site onto the thermophile [79]. Once the residues for grafting have been decided upon, FoldX or Rosetta should be used to determine a theoretical change in the Gibbs free energy of folding compared to the complete graft ($\Delta\Delta G_{D\rightarrow N}$).

2.5.3 Characterize Changes in Dynamics

Changes to protein dynamics, such as improved rigidity or packing of a protein core, can be simulated to ensure a graft is compatible with the target protein. Given relatively small changes to the target sequence, packages such as MODELLER [15] can produce highly plausible models of the new structure. Molecular dynamics

simulations over a short timescale (up to ~10 ns, run with at least three replicates) have been used to describe changes to dynamics [3, 84, 87], although longer simulations will provide more comprehensive data. These changes to dynamics can be quantified through the average structural change across the entire protein domain (RMSD) or through changes in rigidity at a residue level (RMSF) to identify flexible regions.

2.5.4 Automated Process with Online Tools

As this semi-rational grafting approach contains multiple generalizable or automated steps, multiple computational tools have been developed to automate the semi-rational design process. Modern online tools such as iRDP [88], FireProt [89], or PROSS [90] have been experimentally verified and are available to use via public webservers.

2.6 Directed Evolution and Beyond

Directed evolution (DE) is a laboratory-based technique that mimics the process of natural selection to create new proteins with improvements in a specific, user-defined trait. Traditionally, DE is based on an iterative two-step protocol. In the first step, the target gene is mutated or recombined to generate a diverse gene library. Then, the library is tested to identify variants with improvements in the desired trait. These variants then form the starting point for the next round of DE. DE has proven to be highly effective and broadly applicable for optimizing the activity, specificity, and stability of proteins [91–95], the latter of which is the focus of this chapter. However, careful experimental design is important to increase the chances of success and minimize time spent in the laboratory [96].

Two of the major considerations when initiating a DE experiment are library design—in particular, library size—and the method to be used in the subsequent testing of the library. These two factors are not independent; the larger the library, the more clones will have to be analyzed to identify variants with increased stability. Hence, limitations in the number of variants that can be tested should be taken into consideration when deciding the library size.

2.6.1 Design Libraries

The primary consideration for library design is the available information. If no prior information is available, a library of randomly mutated variants is often the method of choice. Random libraries are routinely generated using error-prone PCR [97], which is simple to perform and allows sampling of a very large number of different variants [98]. However, the frequency of mutations that increase protein stability in such random libraries is low. Further, the subsequent testing of these libraries generally requires high-throughput methods that are not available in many cases. To address these problems, many researchers have employed new strategies to design smaller, focused libraries [96].

Often referred to as semi-rational or smart library design, these approaches combine information from protein sequence and structure with computational predictive algorithms to identify residues that can be targeted for mutagenesis. Examples of established computational tools include PROSS [90], FireProt [89], HotSpot Wizard [99], and FRESCO [100], which can all be used to predict promising sites for optimization of protein stability. In addition, many of the previously discussed tools such as B-factor analysis or targeting non-conserved residues [84] can also be applied. Targeting individual amino acids reduces library size, and incorporation of sequence and structural information increases the potential of each variant to increase protein stability.

Once specific residues of interest have been identified, site-saturation mutagenesis can be performed. Site-saturation mutagenesis involves randomizing a single codon at the desired position to generate codons for all possible amino acids. Several different protocols can be used for site-saturation mutagenesis, the easiest being PCR-based mutagenesis using oligonucleotides that encode degenerate codons at positions corresponding to targeted residues. Mutating multiple residues simultaneously provides access to combinations of mutations that are epistatic, which can be difficult to access with most computational methods [101]. However, the number of unique sequences in the library increases exponentially with the number of randomized sites. Concurrently, mutating multiple residues greatly increases library size, which may not be desirable.

2.6.2 Determine Selection or Screening Strategy

The principal dilemma facing researchers when deciding on a method to analyze their DE library is whether to use selection or screening. Selection methodologies, mirroring natural selection, link survival of a variant to fitness of a trait. This typically enables high throughput (10^6 to $>10^{12}$ variants) but only identifies improved variants. Screening methodologies are capable of lower throughput (10^2 to 10^4 variants) but evaluate every variant in the library [96, 102]. As previously mentioned, the choice here depends largely on the library design. High-throughput methods become more advantageous when analyzing larger libraries, such as from random mutagenesis or simultaneous site-saturation mutagenesis of several positions. Examples of high-throughput strategies that have been successfully used to select variants with increased stability include Proside [103], β-lactamase loop insertion [104], and GFP-fusion systems [105–107]. One disadvantage of high-throughput selection methods is that they need to be optimized for the individual protein, which can be a project in itself. Furthermore, because selection only identifies improved variants, it is more prone to artifacts that improve the measured trait but not the desired quality [96], such as enhanced expression levels

or increased activity. Therefore, screening rather than selecting is often preferred when analyzing smaller libraries.

A straightforward approach to screen smaller libraries for variants with increased thermostability is to monitor protein activity in the cell lysate, either at elevated temperatures or after heating [105, 108]. This approach has been extensively used to identify enzymes with increased stability [92–94, 109–111]. Unfortunately, the approach is limited to proteins that possess an activity which can be easily assayed in the context of the cell lysate, and there is no clear correlation between thermal inactivation and protein stability [105]. There are many other methods to screen for protein stability which have been discussed in detail in several excellent reviews [98, 105, 108, 112]. As no single screen is universally suitable, the screening strategy will have to be designed on a case-by-case basis considering the specific protein, the library size, and the available resources.

2.6.3 *Combine Libraries to Combine Improvements*

Well-designed DE experiments allow for parallel exploration of mutations at more than one position to find combinatorial configurations (Fig. 3). Such parallel exploration is challenging with most computational methods [101]. Once the best-performing variants in each saturation mutagenesis library are identified, they can be combined to assess potential additive and cooperative effects on stability. Searching within this promising sequence space for

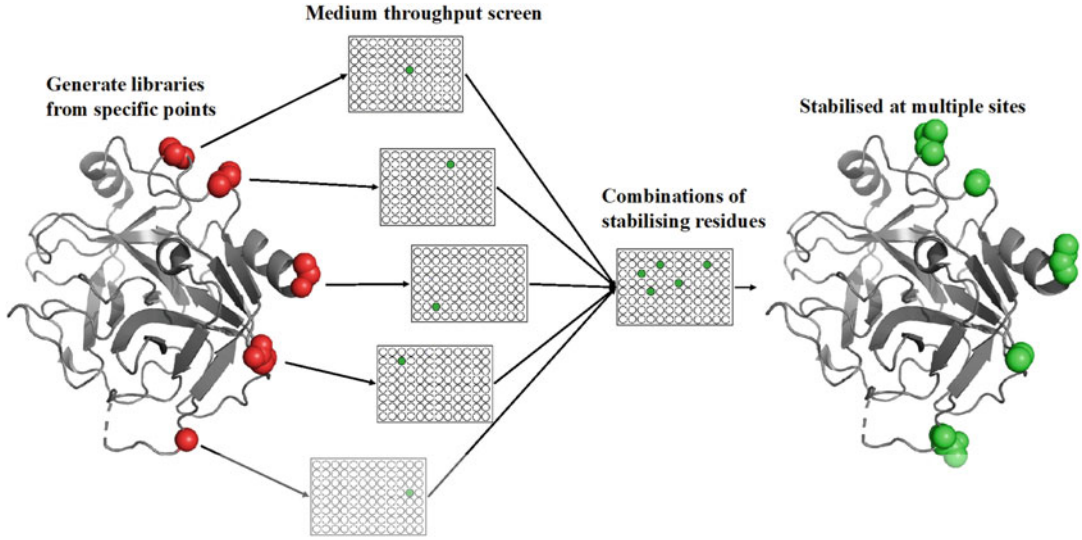

Fig. 3 Example of directed evolution with a smart library and recombination. Small smart libraries generate diversity at identified sites. The example shown here has a library design that allows for medium-throughput screening in a 96-well plate format with common laboratory equipment. Subsequent to the primary DE libraries, beneficial mutations are recombined in sets. The most beneficial set of mutations is combined into a final stabilized design

combined mutations, with a potential for additive effects on stability, can compensate for the small library size in this experiment.

2.7 Conclusion

Over the past decade, our understanding of what determines protein stability has drastically increased, resulting in large improvements in the ability to optimize protein stability in silico [2]. Concurrently, there has been substantial development of new and improved experimental techniques to engineer, evolve, and screen for protein stability. Combined, these approaches form a set of powerful and easily applicable tools for increasing protein stability.

References

1. Costanzo JA, O'Brien CJ, Tiller K et al (2014) Conformational stability as a design target to control protein aggregation. Protein Eng Des Sel 27:157–167. https://doi.org/10.1093/protein/gzu008
2. Goldenzweig A, Fleishman S (2018) Principles of protein stability and their application in computational design. Annu Rev Biochem 87:105–129. https://doi.org/10.1146/annurev-biochem
3. Yu H, Huang H (2014) Engineering proteins for thermostability through rigidifying flexible sites. Biotechnol Adv 32:308–315. https://doi.org/10.1016/j.biotechadv.2013.10.012
4. Kim SJ, Lee JA, Joo JC et al (2010) The development of a thermostable CiP (Coprinus cinereus peroxidase) through in silico design. Biotechnol Prog 26:1038–1046. https://doi.org/10.1002/btpr.408
5. Reetz MT, Carballeira JD (2007) Iterative saturation mutagenesis (ISM) for rapid directed evolution of functional enzymes. Nat Protoc 2:891–903. https://doi.org/10.1038/nprot.2007.72
6. Glaser F, Pupko T, Paz I et al (2003) ConSurf: identification of functional regions in proteins by surface mapping of phylogenetic information. Bioinformatics 19:163–164. https://doi.org/10.1093/bioinformatics/19.1.163
7. Guerois R, Nielsen JE, Serrano L (2002) Predicting changes in the stability of proteins and protein complexes: a study of more than 1000 mutations. J Mol Biol 320:369–387. https://doi.org/10.1016/S0022-2836(02)00442-4
8. Liu Y, Kuhlman B (2006) RosettaDesign server for protein design. Nucleic Acids Res 34:235–238. https://doi.org/10.1093/nar/gkl163
9. Buß O, Rudat J, Ochsenreither K (2018) FoldX as protein engineering tool: better than random based approaches? Comput Struct Biotechnol J. https://doi.org/10.1016/j.csbj.2018.01.002
10. Dill KA (1990) Dominant forces in protein folding. Biochemistry 29:7133–7155. https://doi.org/10.1021/bi00483a001
11. Vogt G, Argos P (1997) Protein thermal stability: hydrogen bonds or internal packing? Fold Des 2:S40–S46. https://doi.org/10.1016/S1359-0278(97)00062-X
12. Moore EJ, Zorine D, Hansen WA et al (2017) Enzyme stabilization via computationally guided protein stapling. Proc Natl Acad Sci U S A 114:12472–12477. https://doi.org/10.1073/pnas.1708907114
13. Hodgson DRW, Sanderson JM (2004) The synthesis of peptides and proteins containing non-natural amino acids. Chem Soc Rev 33:422–430. https://doi.org/10.1039/b312953p
14. Link AJ, Mock ML, Tirrell DA (2003) Non-canonical amino acids in protein engineering. Curr Opin Biotechnol 14:603–609. https://doi.org/10.1016/j.copbio.2003.10.011
15. Eswar N, Webb B, Marti-Renom MA, et al (2006) Comparative protein structure modeling using Modeller Curr Protoc Bioinformatics Chapter 5:Unit-5.6
16. DeLano WL (2002) The PyMOL molecular graphics system. Version 2.0 Schrödinger, LLC
17. Aronica PGA, Verma C, Popovic B et al (2016) The Parasol Protocol for computational mutagenesis. Protein Eng Des Sel 29:253–261. https://doi.org/10.1093/protein/gzw009
18. Greenfield NJ (2006) Using circular dichroism collected as a function of temperature to

determine the thermodynamics of protein unfolding and binding interactions. Nat Protoc 1:2527–2535. https://doi.org/10.1038/nprot.2006.204

19. Johnson CM (2013) Differential scanning calorimetry as a tool for protein folding and stability. Arch Biochem Biophys 531:100–109. https://doi.org/10.1016/j.abb.2012.09.008
20. Matouschek A, Fersht AR (1991) Protein engineering in analysis of protein folding pathways and stability. Methods Enzymol 202:82–112
21. Porebski BT, Buckle AM (2016) Consensus protein design. Protein Eng Des Sel 29:245–251. https://doi.org/10.1093/protein/gzw015
22. Wheeler LC, Lim SA, Marqusee S, Harms MJ (2016) The thermostability and specificity of ancient proteins. Curr Opin Struct Biol 38:37–43. https://doi.org/10.1016/j.sbi.2016.05.015
23. Risso VA, Gavira JA, Gaucher EA, Sanchez-Ruiz JM (2014) Phenotypic comparisons of consensus variants versus laboratory resurrections of Precambrian proteins. Proteins 82:887–896. https://doi.org/10.1002/prot.24575
24. Do CB, Katoh K (2008) Protein multiple sequence alignment. In: Thompson JD, Ueffing M, Schaeffer-Reiss C (eds) Methods in molecular biology. Humana Press, Totowa, NJ, pp 379–413
25. Bateman A, Martin MJ, O'Donovan C et al (2017) UniProt: the universal protein knowledgebase. Nucleic Acids Res 45:D158–D169. https://doi.org/10.1093/nar/gkw1099
26. Altschul SF, Gish W, Miller W et al (1990) Basic local alignment search tool. J Mol Biol 215:403–410. https://doi.org/10.1016/S0022-2836(05)80360-2
27. Finn RD, Coggill P, Eberhardt RY et al (2016) The Pfam protein families database: towards a more sustainable future. Nucleic Acids Res 44:D279–D285. https://doi.org/10.1093/nar/gkv1344
28. Sigrist CJA, De Castro E, Cerutti L et al (2013) New and continuing developments at PROSITE. Nucleic Acids Res 41:344–347. https://doi.org/10.1093/nar/gks1067
29. Wilson D, Pethica R, Zhou Y et al (2009) SUPERFAMILY - sophisticated comparative genomics, data mining, visualization and phylogeny. Nucleic Acids Res 37:380–386. https://doi.org/10.1093/nar/gkn762
30. Finn RD, Clements J, Arndt W et al (2015) HMMER web server: 2015 update. Nucleic Acids Res 43:W30–W38. https://doi.org/10.1093/nar/gkv397
31. Li W, Godzik A (2006) Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics 22:1658–1659. https://doi.org/10.1093/bioinformatics/btl158
32. Gerlt JA, Bouvier JT, Davidson DB et al (2015) Enzyme function initiative-enzyme similarity tool (EFI-EST): a web tool for generating protein sequence similarity networks. Biochim Biophy Acta 1854:1019–1037. https://doi.org/10.1016/j.bbapap.2015.04.015
33. Shannon P, Markiel A, Owen O et al (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 13:2498–2504. https://doi.org/10.1101/gr.1239303.metabolite
34. Simossis V, Kleinjung J, Heringa J (2003) An overview of multiple sequence alignment. Curr Protoc Bioinformatics. https://doi.org/10.1002/0471250953.bi0307s03
35. Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32:1792–1797. https://doi.org/10.1093/nar/gkh340
36. Di Tommaso P, Moretti S, Xenarios I et al (2011) T-Coffee: a web server for the multiple sequence alignment of protein and RNA sequences using structural information and homology extension. Nucleic Acids Res 39:13–17. https://doi.org/10.1093/nar/gkr245
37. Madej T, Lanczycki CJ, Zhang D et al (2014) MMDB and VAST+: tracking structural similarities between macromolecular complexes. Nucleic Acids Res 42:297–303. https://doi.org/10.1093/nar/gkt1208
38. Steipe B, Schiller B, Plückthun A, Steinbacher S (1994) Sequence statistics reliably predict stabilizing mutations in a protein domain. J Mol Biol 240:188–192
39. Polizzi KM, Chaparro-Riggers JF, Vazquez-Figueroa E, Bommarius AS (2006) Structure-guided consensus approach to create a more thermostable penicillin G acylase. Biotechnol J 1:531–536. https://doi.org/10.1002/biot.200600029
40. Tokuriki N, Stricher F, Schymkowitz J et al (2007) The stability effects of protein mutations appear to be universally distributed. J Mol Biol 369:1318–1332. https://doi.org/10.1016/j.jmb.2007.03.069

41. Lehmann M, Wyss M (2001) Engineering proteins for thermostability: the use of sequence alignments versus rational design and directed evolution. Curr Opin Biotechnol 12:371–375. https://doi.org/10.1016/S0958-1669(00)00229-9
42. Lehmann M, Pasamontes L, Lassen SF, Wyss M (2000) The consensus concept for thermostability engineering of proteins. Biochim Biophys Acta Protein Struct Mol Enzymol 1543:408–415. https://doi.org/10.1016/S0167-4838(00)00238-7
43. Porebski BT, Nickson AA, Hoke DE et al (2015) Structural and dynamic properties that govern the stability of an engineered fibronectin type III domain. Protein Eng Des Sel 28:67–78. https://doi.org/10.1093/protein/gzv002
44. Porebski BT, Keleher S, Hollins JJ et al (2016) Smoothing a rugged protein folding landscape by sequence-based redesign. Sci Rep 6:1–14. https://doi.org/10.1038/srep33958
45. Jacobs SA, Diem MD, Luo J et al (2012) Design of novel FN3 domains with high stability by a consensus sequence approach. Protein Eng Des Sel 25:107–117. https://doi.org/10.1093/protein/gzr064
46. Dai M, Fisher HE, Temirov J et al (2007) The creation of a novel fluorescent protein by guided consensus engineering. Protein Eng Des Sel 20:69–79. https://doi.org/10.1093/protein/gzl056
47. Pantoliano MW, Whitlow M, Wood JF et al (1989) Large increases in general stability for subtilisin BPN' through incremental changes in the free energy of unfolding. Biochemistry 28:7205–7213. https://doi.org/10.1021/bi00444a012
48. Blatt LM, Davis JM, Klein SB, Taylor MW (1996) The biologic activity and molecular characterization of a novel synthetic interferon-alpha species, consensus interferon. J Interferon Cytokine Res 16:489–499
49. Vazquez-Figueroa E, Yeh V, Broering JM et al (2008) Thermostable variants constructed via the structure-guided consensus method also show increased stability in salts solutions and homogeneous aqueous-organic media. Protein Eng Des Sel 21:673–680. https://doi.org/10.1093/protein/gzn048
50. Sullivan BJ, Durani V, Magliery TJ (2011) Triosephosphate isomerase by consensus design: dramatic differences in physical properties and activity of related variants. J Mol Biol 413:195–208. https://doi.org/10.1016/j.jmb.2011.08.001
51. Nikolova PV, Henckel J, Lane DP, Fersht AR (1998) Semirational design of active tumor suppressor p53 DNA binding domain with enhanced stability. Proc Natl Acad Sci U S A 95:14675–14680. https://doi.org/10.1073/pnas.95.25.14675
52. Wang Q, Buckle AM, Foster NW et al (1999) Design of highly stable functional GroEL minichaperones. Protein Sci 8:2186–2193. https://doi.org/10.1110/ps.8.10.2186
53. Ferreiro DU, Cervantes CF, Truhlar SME et al (2007) Stabilizing IκBα by "consensus" design. J Mol Biol 365:1201–1216. https://doi.org/10.1016/j.jmb.2006.11.044
54. Gaucher EA, Thomson JM, Burgan MF, Benner SA (2003) Inferring the palaeoenvironment of ancient bacteria on the basis of resurrected proteins. Nature 425:285–288. https://doi.org/10.1038/nature01977
55. Gaucher EA, Govindarajan S, Ganesh OK (2008) Palaeotemperature trend for Precambrian life inferred from resurrected proteins. Nature 451:704–707. https://doi.org/10.1038/nature06510
56. Clifton BE, Jackson CJ (2016) Ancestral protein reconstruction yields insights into adaptive evolution of binding specificity in solute-binding proteins. Cell Chem Biol 23:236–245. https://doi.org/10.1016/j.chembiol.2015.12.010
57. Perez-Jimenez R, Inglés-Prieto A, Zhao Z et al (2011) Single-molecule paleoenzymology probes the chemistry of resurrected enzymes. Nat Struct Mol Biol 18:592–596. https://doi.org/10.1038/nsmb.2020
58. Akanuma S (2017) Characterization of reconstructed ancestral proteins suggests a change in temperature of the ancient biosphere. Life 7:33. https://doi.org/10.3390/life7030033
59. Williams PD, Pollock DD, Blackburne BP, Goldstein RA (2006) Assessing the accuracy of ancestral protein reconstruction methods. PLoS Comput Biol 2:0598–0605. https://doi.org/10.1371/journal.pcbi.0020069
60. Zhang J, Nei M (1997) Accuracies of ancestral amino acid sequences inferred by the parsimony, likelihood, and distance methods. J Mol Evol 44:139–146. https://doi.org/10.1007/PL00000067
61. Nguyen LT, Schmidt HA, Von Haeseler A, Minh BQ (2015) IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol 32:268–274. https://doi.org/10.1093/molbev/msu300
62. Kalyaanamoorthy S, Minh BQ, Wong TKF et al (2017) ModelFinder: fast model

selection for accurate phylogenetic estimates. Nat Methods 14:587–589. https://doi.org/10.1038/nmeth.4285

63. Rambaut A (2006) FigTree. Institute of Evolutionary Biology, University of Edinburgh, Edinburgh. http://tree.bio.ed.ac.uk/software/figtree
64. Yang Z, Rannala B (2012) Molecular phylogenetics: principles and practice. Nat Rev Genet 13:303–314. https://doi.org/10.1038/nrg3186
65. Anisimova M, Gil M, Dufayard JF et al (2011) Survey of branch support methods demonstrates accuracy, power, and robustness of fast likelihood-based approximation schemes. Syst Biol 60:685–699. https://doi.org/10.1093/sysbio/syr041
66. Minh BQ, Nguyen MAT, Von Haeseler A (2013) Ultrafast approximation for phylogenetic bootstrap. Mol Biol Evol 30:1188–1195. https://doi.org/10.1093/molbev/mst024
67. Joy JB, Liang RH, McCloskey RM et al (2016) Ancestral reconstruction. PLoS Comput Biol 12:1–20. https://doi.org/10.1371/journal.pcbi.1004763
68. Yang Z (2007) PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol 24:1586–1591. https://doi.org/10.1093/molbev/msm088
69. Eick GN, Bridgham JT, Anderson DP et al (2017) Robustness of reconstructed ancestral protein functions to statistical uncertainty. Mol Biol Evol 34:247–261. https://doi.org/10.1093/molbev/msw223
70. Li Z, Yang Y, Zhan J et al (2013) Energy functions in de novo protein design: current challenges and future prospects. Annu Rev Biophys 42:315–335. https://doi.org/10.1146/annurev-biophys-083012-130315
71. Dahiyat BI, Sarisky CA, Mayo SL (1997) De novo protein design: towards fully automated sequence selection. J Mol Biol 273:789–796. https://doi.org/10.1006/jmbi.1997.1341
72. Liang S, Grishin NV (2004) Effective scoring function for protein sequence design. Proteins 54:271–281
73. Pokala N, Handel TM (2005) 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 347:203–227. https://doi.org/10.1016/j.jmb.2004.12.019
74. Dai L, Yang Y, Kim HR, Zhou Y (2010) Improving computational protein design by using structure-derived sequence profile. Proteins 78:2338–2348. https://doi.org/10.1002/prot.22746
75. Dantas G, Kuhlman B, Callender D et al (2003) A large scale test of computational protein design: folding and stability of nine completely redesigned globular proteins. J Mol Biol 332:449–460. https://doi.org/10.1016/S0022-2836(03)00888-X
76. Korkegian A, Black ME, Baker D, Stoddard BL (2005) Computational thermostabilization of an enzyme. Science 308:857–860. https://doi.org/10.1126/science.1107387
77. Vieille C, Zeikus GJ (2001) Hyperthermophilic enzymes: sources, uses, and molecular mechanisms for thermostability. Microbiol Mol Biol Rev 65:1–43. https://doi.org/10.1128/MMBR.65.1.1-43.2001
78. Xiong H, Fenel F, Leisola M, Turunen O (2004) Engineering the thermostability of Trichoderma reesei endo-1,4-β-xylanase II by combination of disulphide bridges. Extremophiles 8:393–400. https://doi.org/10.1007/s00792-004-0400-9
79. Kapoor D, Kumar V, Chandrayan SK et al (2008) Replacement of the active surface of a thermophile protein by that of a homologous mesophile protein through structure-guided "protein surface grafting". Biochim Biophys Acta 1784:1771–1776. https://doi.org/10.1016/j.bbapap.2008.05.007
80. Pedone E, Saviano M, Rossi M, Bartolucci S (2001) A single point mutation (Glu85Arg) increases the stability of the thioredoxin from Escherichia coli. Protein Eng 14:255–260. https://doi.org/10.1093/protein/14.4.255
81. Vázquez-Figueroa E, Chaparro-Riggers J, Bommarius AS (2007) Development of a thermostable glucose dehydrogenase by a structure-guided consensus concept. ChemBioChem 8:2295–2301. https://doi.org/10.1002/cbic.200700500
82. Loening AM, Fenn TD, Wu AM, Gambhir SS (2006) Consensus guided mutagenesis of Renilla luciferase yields enhanced stability and light output. Protein Eng Des Sel 19:391–400. https://doi.org/10.1093/protein/gzl023
83. Amin N, Liu AD, Ramer S et al (2004) Construction of stabilized proteins by combinatorial consensus mutagenesis. Protein Eng Des Sel 17:787–793. https://doi.org/10.1093/protein/gzh091
84. Anbar M, Gul O, Lamed R et al (2012) Improved thermostability of Clostridium thermocellum endoglucanase Cel8A by using consensus-guided mutagenesis. Appl Environ

Microbiol 78:3458–3464. https://doi.org/10.1128/AEM.07985-11

85. Kumar S, Nussinov R (2001) How do thermophilic proteins deal with heat? Cell Mol Life Sci 58:1216–1233. https://doi.org/10.1007/PL00000935
86. Feldwisch J, Tolmachev V, Lendel C et al (2010) Design of an optimized scaffold for affibody molecules. J Mol Biol 398:232–247. https://doi.org/10.1016/j.jmb.2010.03.002
87. Badieyan S, Bevan DR, Zhang C (2012) Study and design of stability in GH5 cellulases. Biotechnol Bioeng 109:31–44. https://doi.org/10.1002/bit.23280
88. Panigrahi P, Sule M, Ghanate A et al (2015) Engineering proteins for thermostability with iRDP web server. PLoS One 10:1–20. https://doi.org/10.1371/journal.pone.0139486
89. Bednar D, Beerens K, Sebestova E et al (2015) FireProt: energy- and evolution-based computational design of thermostable multiple-point mutants. PLoS Comput Biol 11:1–20. https://doi.org/10.1371/journal.pcbi.1004556
90. Goldenzweig A, Goldsmith M, Hill SE et al (2016) Automated structure- and sequence-based design of proteins for high bacterial expression and stability. Mol Cell 63:337–346. https://doi.org/10.1016/j.molcel.2016.06.012
91. Mate DM, Gonzalez-perez D, Mateljak I et al (2017) The pocket manual of directed evolution: tips and tricks. Elsevier Inc., Amsterdam
92. Martinez R, Jakob F, Tu R et al (2013) Increasing activity and thermal resistance of Bacillus gibsonii alkaline protease (BgAP) by directed evolution. Biotechnol Bioeng 110:711–720. https://doi.org/10.1002/bit.24766
93. Giver L, Gershenson A, Freskgard P-O, Arnold FH (1998) Directed evolution of a thermostable esterase. Proc Natl Acad Sci U S A 95:12809–12813. https://doi.org/10.1073/pnas.95.22.12809
94. Salazar O, Cirino PC, Arnold FH (2003) Thermostabilization of a cytochrome P450 peroxygenase. ChemBioChem 4:891–893. https://doi.org/10.1002/cbic.200300660
95. Tokuriki N, Jackson CJ, Afriat-Jurnou L et al (2012) Diminishing returns and tradeoffs constrain the laboratory optimization of an enzyme. Nat Commun 3:1257–1259. https://doi.org/10.1038/ncomms2246
96. Goldsmith M, Tawfik DS (2012) Directed enzyme evolution: beyond the low-hanging fruit. Curr Opin Struct Biol 22:406–412. https://doi.org/10.1016/j.sbi.2012.03.010
97. Copp JN, Hanson-Manful P, Ackerley DF, Patrick WM (2014) Error-prone PCR and effective generation of gene variant libraries for directed evolution. Methods Mol Biol 1179:3–22
98. Packer MS, Liu DR (2015) Methods for the directed evolution of proteins. Nat Rev Genet 16:379–394. https://doi.org/10.1038/nrg3927
99. Bendl J, Stourac J, Sebestova E et al (2016) HotSpot Wizard 2.0: automated design of site-specific mutations and smart libraries in protein engineering. Nucleic Acids Res 44:W479–W487. https://doi.org/10.1093/nar/gkw416
100. Wijma HJ, Floor RJ, Jekel PA et al (2014) Computationally designed libraries for rapid enzyme stabilization. Protein Eng Des Sel 27:49–58. https://doi.org/10.1093/protein/gzt061
101. Reetz MT, Carballeira JD, Vogel A (2006) Iterative saturation mutagenesis on the basis of B factors as a strategy for increasing protein thermostability. Angew Chem Int Ed 45:7745–7751. https://doi.org/10.1002/anie.200602795
102. Leemhuis H, Kelly RM, Dijkhuizen L (2009) Directed evolution of enzymes: library screening strategies. IUBMB Life 61:222–228. https://doi.org/10.1002/iub.165
103. Martin A, Schmid FX, Sieber V (2003) Proside: a phage-based method for selecting thermostable proteins. Methods Mol Biol 230:57–70. https://doi.org/10.1385/1-59259-396-8:57
104. Foit L, Morgan GJ, Kern MJ et al (2009) Optimizing protein stability in vivo. Mol Cell 36:861–871. https://doi.org/10.1016/j.molcel.2009.11.022
105. Magliery TJ, Lavinder JJ, Sullivan BJ (2011) Protein stability by number: high-throughput and statistical approaches to one of protein science's most difficult problems. Curr Opin Chem Biol 15:443–451. https://doi.org/10.1016/j.cbpa.2011.03.015
106. Lindman S, Hernandez-Garcia A, Szczepankiewicz O et al (2010) In vivo protein stabilization based on fragment complementation and a split GFP system. Proc Natl Acad Sci U S A 107:19826–19831. https://doi.org/10.1073/pnas.1005689107
107. Seitz T, Thoma R, Schoch GA et al (2010) Enhancing the stability and solubility of the glucocorticoid receptor ligand-binding

domain by high-throughput library screening. J Mol Biol 403:562–577. https://doi.org/10.1016/j.jmb.2010.08.048

108. Socha RD, Tokuriki N (2013) Modulating protein stability - directed evolution strategies for improved protein function. FEBS J 280 (22):5582–5595
109. Shivange AV, Serwe A, Dennig A et al (2012) Directed evolution of a highly active Yersinia mollaretii phytase. Appl Microbiol Biotechnol 95:405–418. https://doi.org/10.1007/s00253-011-3756-7
110. Buettner K, Hertel TC, Pietzsch M (2012) Increased thermostability of microbial transglutaminase by combination of several hot spots evolved by random and saturation mutagenesis. Amino Acids 42:987–996. https://doi.org/10.1007/s00726-011-1015-y
111. García-Ruiz E, Maté D, Ballesteros A et al (2010) Evolving thermostability in mutant libraries of ligninolytic oxidoreductases expressed in yeast. Microb Cell Factories 9:1–13. https://doi.org/10.1186/1475-2859-9-17
112. Xiao H, Bao Z, Zhao H (2015) High throughput screening and selection methods for directed enzyme evolution. Ind Eng Chem Res 54:4011–4020. https://doi.org/10.1021/ie503060a