Molecular diversity in engineered protein libraries

Nora H Barakat and John J Love

Engineered protein libraries, defined here as a collection of different mutant variants of a single specific protein, are intentionally designed to be rich in molecular diversity and can span ranges from as little as 400 different variants to greater than 10^{12} members per library. The goal of engineering libraries is to generate new protein variants, identified upon screening, that possess desired novel properties. Exploitation of the natural organization of the genetic code has led to 'focused' libraries that are lower in overall complexity yet biased towards variants with preferred biophysical properties. An emerging trend, in which computational algorithms are blended with *in vivo* screens, is also leading towards greater and more rapid success in the field of protein design.

Addresses
Department of Chemistry and Biochemistry, San Diego State University, 5500 Campanile Drive, San Diego, CA 92182-1030, USA

Corresponding author: Love, John J (jlove@sciences.sdsu.edu)

Current Opinion in Chemical Biology 2007, **11**:335–341

This review comes from a themed issue on
Combinatorial chemistry and molecular diversity
Edited by Gregory A Weiss and Richard Roberts

Available online 4th June 2007

1367-5931/$ – see front matter

DOI 10.1016/j.cbpa.2007.05.014

Introduction

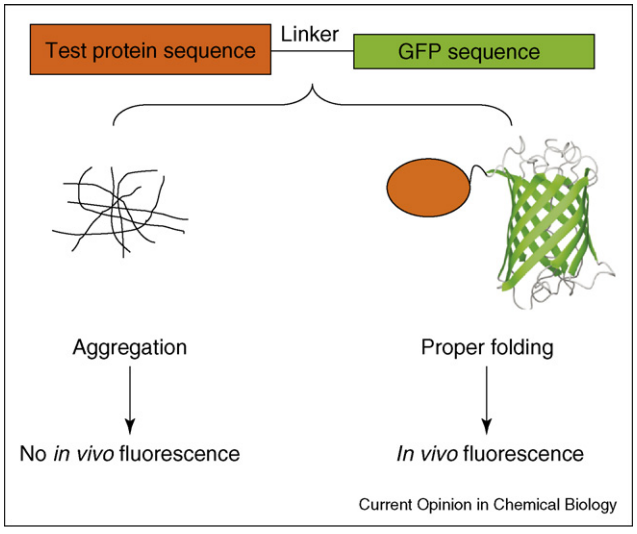
During the billions of years since life spontaneously arose, the process of evolution has resulted in significant molecular diversity at the protein level. Two key aspects of evolution are the creation (and maintenance) of molecular diversity and the selection of traits associated with specific fitness criteria (i.e. fitness conferred by particular sets of amino acids in different protein variants). Emulation of these two aspects in the laboratory, as *in vivo*, *in vitro* or genetic screens, has led to significant advances in the field of protein design. The molecular diversity inherent to engineered protein libraries, and subsequent screening, has led to designed protein variants with novel enzymatic properties, variants that have the ability to bind specifically to macromolecular targets, and variants that have increased solubility and enhanced stability. In addition, the use of protein libraries and associated screening has provided unique insights into the biophysical properties of proteins associated with certain human diseases (for example, the Aβ42 peptide associated with Alzheimer's disease). This review focuses on recent advances in the use of engineered protein libraries and the development of focused libraries designed to reduce sequence space to that which is most probable to produce variants with the desired properties. In addition, the use of computational algorithms to assist in the engineering of focused libraries is also described. As a result of the expanding breadth of this field, we focus on the use of libraries to design new protein variants utilizing three systems that have been the subject of intense engineering: green fluorescent protein (GFP) solubility screens, the Aβ42 peptide associated with Alzheimer's disease and streptococcal protein G (Gβ1) stability screens.

Enhanced solubility from engineered libraries screened as GFP fusions

The use of GFP as a fluorescence reporter has expanded significantly in many areas of research [1–6], especially in the field of protein design [7,8]. High solubility is a desired trait when expressing designed or natural proteins, and therefore a rapid means of screening monomeric variants from an engineered protein library is of high utility [9]. It was determined early on that the folding of GFP, and formation of its active chromophore, occurs relatively slowly [10,11] and that the solubility of proteins fused to the N terminus of GFP (Figure 1) can greatly impact folding and thus emitted fluorescence [12]. A close correlation between solubility and fluorescence was demonstrated for a series of test proteins from *Pyrobaculum aerophilum* fused to the N terminus of GFP [12]. In addition, directed evolution was used to engineer a variant of bullfrog H-subunit ferritin that is more soluble than the wild-type protein. The solubility of the resulting (fully functional) variant was increased and thus it was demonstrated that proteins with enhanced solubility can be screened from engineered protein libraries fused to this fluorescent reporter. However, it should be noted that a feature of the GFP fusion technology is that all mutant variants must be observed (so as to assess the degree of fluorescence each emits) and thus molecular diversity must necessarily be restricted to a range that is accessible in the laboratory.

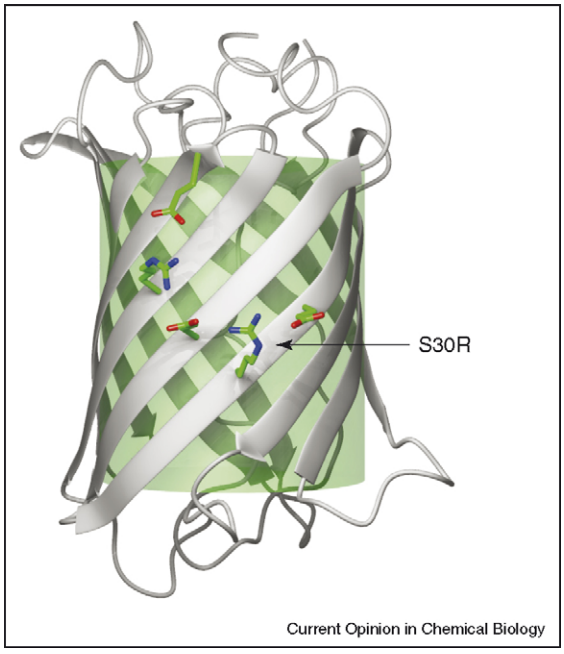
Waldo and co-workers have continued the development of the GFP solubility reporter by using intensive protein engineering (i.e. directed evolution) to generate a split GFP system in which the protein being tested is fused to a 15 amino acid GFP fragment [13,14]. When this chimeric fragment is co-expressed with a fragment consisting of the remainder of GFP, they spontaneously reassemble and form fluorescent GFP if the attached test protein remains soluble. Cabantous *et al.* used a

Figure 1



GFP solubility screen. Illustrated is the relationship between the N-terminal test protein and GFP (plus the intervening linker) for the chimeric protein generated in the screen for solubility. If the test protein is prone to aggregation (left), it will precipitate and block GFP folding and fluorescence, whereas soluble test proteins allow GFP to fold and fluoresce properly.

Figure 2



The ionic network and β-barrel topology of GFP. The five charged sidechains that form a surface-exposed ionic network are highlighted and the stabilizing mutation, S30R, is indicated with an arrow.

two-tiered approach, in which fully intact GFP was used in combination with the split GFP system to evolve higher solubility into proteins from *Mycobacterium tuberculosis* known to be insoluble and thus recalcitrant to protein expression, purification and ultimately crystallization [15].

Additional production of mutant libraries of GFP, and screening based on enhanced fluorescence, has resulted in a 'superfolder' variant that folds robustly even when fused to poorly folded test proteins [16••]. Pédelacq *et al.* started with the amino acid sequence of a well-folded GFP variant derived from a previous directed evolution cycle, which resulted in the mutations F99S, M153T, V163A [17] and two 'enhanced GFP' mutations — F64L and S65T [18]. After four rounds of DNA shuffling, a highly fluorescent 'superfolder' GFP variant was isolated that contained the following six new mutations: S30R, Y39N, N105T, Y145F, I171V and A206V. Upon solving the crystal structure of this variant (PDB accession code 2B3P), it was theorized that its faster folding kinetics and greater stability are probably a consequence of the S30R mutation [16••]. This mutation results in an intramolecular ionic network, across four adjacent β-strands, that involves a sequence of five alternating acidic and basic residues (Figure 2). This form of surface-exposed ionic network would probably be difficult to accurately predict computationally and thus reflects the value and continued importance of engineered protein libraries and associated screens.

Probing the amyloidogenicity of Aβ42 with focused protein libraries and GFP reporter fusions

To explore the biophysical properties of a peptide associated with Alzheimer's disease, Wurth *et al.* utilized GFP fusions with fluorescence-based screening of engineered peptide libraries [19,20]. Expanding on their previous work on the *de novo* design of β-sheet proteins, which exhibited high propensities to form amyloid-like fibril structures [20,21], they employed combinatorial protein libraries to probe the amyloidogenic properties of the Aβ42 peptide, which is known to be a major molecular component of the amyloid plaques associated with Alzheimer's disease [19,20,22]. When fused to GFP as an N-terminal fusion, the high propensity for Aβ42 to aggregate blocks proper GFP folding and thus bacteria harboring this chimera do not emit fluorescence. To explore the role that particular residues play in amyloidogenicity, mutant libraries of Aβ42 variants were generated using different methods for introducing mutations [19]. The resulting libraries were probed in the context of the GFP screen and 36 mutant variants of Aβ42, which are more soluble relative to the wild-type peptide, were isolated. The bulk of the mutations cluster into three hydrophobic regions and, although most agree with previous model studies, there are several conservative mutations that simple models based solely on sequence hydrophobicity would not have predicted.

Aβ42 and the shorter variant Aβ40 (identical to Aβ42 but two residues shorter) are produced in relatively equal amounts *in vivo*, yet the senile plaques in diseased brains are composed primarily of Aβ42, which also more readily forms fibrils *in vitro* [19,22]. To analyze the potential role these two terminal residues play in amyloidogenicity, a library was generated by randomizing the codons for positions 41 and 42. In addition, focused libraries were engineered in which the degenerate codon NTN (which encodes five non-polar amino acids) was incorporated at both positions for one library and NAN (which encodes six polar amino acids) incorporated at both positions for another. The resulting screen indicated that hydrophobicity, as well as β-sheet propensity, greatly influences overall solubility. Interestingly, only one colony from the hydrophilic library displayed a white phenotype and subsequent DNA sequencing revealed two arginines at positions 41 and 42. Analysis of a structural model of an Aβ42 fibril, based on solid-state NMR data [23], revealed that an arginine at the terminal positions would be in close spatial proximity to a glutamic acid at position 11, and that a putative salt bridge between these residues may function to stabilize the β-sheet structure of the Aβ42 fibril.

Kim and Hecht returned to the 'binary code' of protein structure (which specifies the pattern of polar and hydrophobic residues in protein structure [24••,25]) to engineer protein libraries for the purpose of probing another key issue associated with the amyloidogenicity of Aβ42 [26]. Recently determined structures of model amyloidogenic peptides reveal the presence of highly ordered 'steric zippers' comprising well-packed structures with specific sidechain interactions [26,27]. To ascertain if these specific sidechain interactions could be substituted for generic hydrophobic amino acids, a focused library of Aβ42 variants was engineered in which the degenerate codon NTN was used to code for five non-polar amino acids at 12 different positions. The findings demonstrate that generic hydrophobic amino acids are sufficient to drive Aβ42 to form amyloid fibril structures. The GFP reporter screen is such an effective tool that it is now being used to screen compounds from a library of different molecular diversity. Kim *et al.* used the *in vivo* Aβ42–GFP fusion system to screen a library of triazine derivatives for compounds that block the self-association of Aβ42 [28]. They demonstrated that one compound in particular effectively blocked Aβ42 self-assembly, which allowed proper GFP folding and associated fluorescence. This system will probably continue to be used as a powerful method to screen for drugs that may offset or prevent the debilitating neurodegenerative effects of Alzheimer's disease.

The use of protein design algorithms to focus protein libraries

Although significant molecular diversity in protein libraries is desirable, especially to ensure a higher likelihood of generating the optimal sequence, it is not difficult to see that many of the sequences in a large randomized library would be non-functional and potentially deleterious. The concept of using protein design algorithms [29] to virtually screen protein libraries and reduce sequence space down to a region amenable to *in vivo* or *in vitro* screening has been applied in the recent past [30–34]. For example, Hayes *et al.* used design algorithms to target specific positions in the engineering of a β-lactamase variant with improved resistance to cefotaxime [35•]. The algorithms were initially used to redesign (i.e. perform computational mutagenesis on) a subset of 19 residues in proximity to the active site of the enzyme, and resulted in one set of mutant residues, and associated rotameric positions, that represented the lowest calculated energy in the context of fixed backbone coordinates (the low-energy sequence is termed the global minimum energy conformation, or GMEC). To guide library design, they determined additional sets of amino acid mutations that had reasonable calculated energies by running Monte Carlo simulated annealing starting with the GMEC sequence. This approach effectively reduced the sequence space for the 19 positions from a possible molecular diversity of $\sim 5.2 \times 10^{24}$ down to 172 800 and, upon generation and screening in the laboratory, resulted in a β-lactamase variant that exhibited a 1280-fold increase in resistance to the targeted antibiotic [35•].

Engineered protein libraries, focused with design algorithms at specific positions, were also used to enhance the fluorescent properties of a blue variant of GFP, referred to as BFP [36,37•]. The mutation Y66H of the chromophore of GFP yields a blue fluorescent protein that has the undesirable properties of low quantum yield and relatively rapid photobleaching [38]. To improve these features, Mena *et al.* used protein design algorithms, and fluorescence activated cell sorting, to target 12 core positions within 7 Å of the imidazole ring of the fluorescent chromophore [36]. Diversity was restricted upon the use of an algorithm that determines library composition by minimizing a weighted average of conformational energies calculated for each of seven candidate libraries [37•]. Combinations of variably focused codons were incorporated into the oligonucleotides used in library engineering. The unintended incorporation of a more inclusive codon at position 224 (i.e. RBG, where R = AG and B = CTG, as opposed to the intended RTG codon) resulted in a library that encoded 3.3×10^5 unique sequences. This level of diversity is greatly reduced relative to the 4×10^{15} sequences of a library completely randomized at 12 positions. Screening of the focused libraries yielded a variant with enhanced quantum yield (0.55 versus 0.34), reduced pH sensitivity and a 40-fold increase in photobleaching half-life, thus illustrating the power of using protein design algorithms to focus library diversity to regions of sequence space most probable to include a variant with preferred properties.

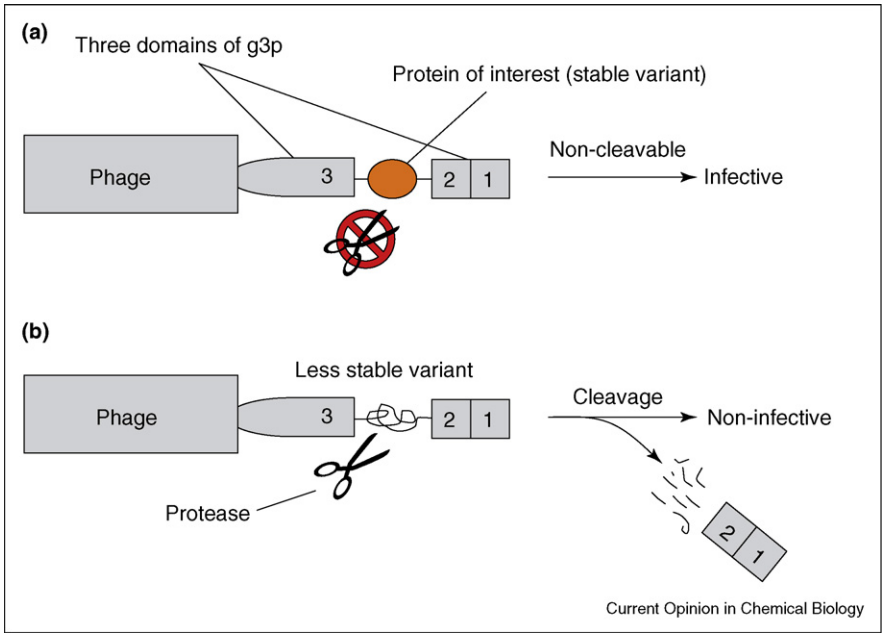
Engineered protein libraries of the Gβ1 domain

A proven workhorse in the field of protein design is the β1 domain of Gβ1, as it has been the subject of design using computational approaches [39–43], as well as *in vivo* and *in vitro* screens [44–46]. In particular, phage display has proved to be a powerful tool for selecting stable Gβ1 variants (as well as other designed proteins [47–51]) from fairly diverse engineered protein libraries. For example, in work by Schmid and co-workers, phage display was used in combination with *in vitro* proteolysis to select for Gβ1 variants with increased intrinsic stabilities [52•,53]. The genes for the Gβ1 variants were cloned into the phage gene-3-protein (G3P) and the resulting libraries were subjected to multiple rounds of *in vitro* proteolysis and amplification in bacteria (Figure 3). In one study, where this method was compared to a previous computational design of Gβ1 [42], saturation mutagenesis was used to randomize the codons at four boundary positions (molecular diversity of ~160 000) that were previously identified with computational algorithms. The genes for ~100 Gβ1 variants that exhibited high protease resistance were sequenced and revealed that amino acids that conferred greater stability were highly degenerate yet biased towards hydrophobic and aromatic residues at three of the four positions [52•]. Thermodynamic analysis of ~21 variants revealed that two were more stable than the sequence derived from the computational design. This finding is not entirely surprising, as a necessary limitation of the computational approach is that the protein backbone must be held rigid for the calculation to remain tractable.

In subsequent work, Wunderlich and Schmid used a two-step approach in which error-prone PCR was used first on the Gβ1 gene to identify candidate positions with high potential for stabilization. Interestingly, the five positions identified fall within the partially exposed 'boundary' category and none were located in the core of the protein (in fact two of the five were identical to positions identified previously with computational methods [42]). The second step entailed the saturation mutagenesis of the five positions followed by multiple rounds of selections. Manual incorporation of the most stabilizing mutations resulted in a Gβ1 variant that exhibited an increase in T_M of 35.1 °C and an increase in $\Delta\Delta G_D$ of 28.5 kJ mol^{-1} at 70 °C [53].

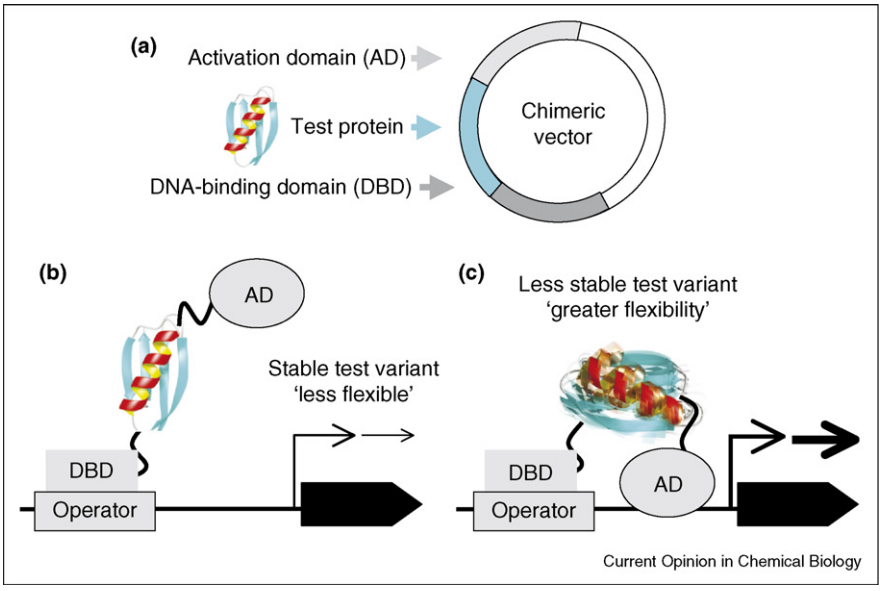
In a different system in which the Gβ1 domain was also the target of design, Barakat *et al.* developed a 'one-hybrid' screen for protein stability using transcriptional elements obtained from a bacterial 'two-hybrid' screen [54•]. A three-protein chimera was created by expressing different Gβ1 variants as fusions inserted between a DNA-binding domain at the N terminus and a transcriptional activation domain at the C terminus (Figure 4). The ability of the different chimeras to up-regulate the associated reporter gene is correlated to the intrinsic

Figure 3



Phage display combined with *in vitro* proteolysis. The protein of interest is inserted between the second N-terminal domain of the G3P protein and the third domain that anchors G3P to the phage particle. The N-terminal domains (1 and 2) are essential for phage infectivity. If the test protein is stable to proteolysis **(a)**, G3P remains intact and thus viable for bacterial infection, whereas less stable variants are more readily proteolyzed and rendered non-infective **(b)**.

Figure 4



'1-hybrid' chimeric stability screen. **(a)** The three sequential genes expressed as a single chimera from the '1-hybrid' vector. If the intervening test protein exhibits high thermal stability (and is thus less flexible), as in **(b)**, the chimera does not make as an effective transcription factor as when the intervening test protein is of lower thermal stability (and presumably greater flexibility), as in **(c)**. Variants of lower thermal stability up-regulate the reporter gene to a greater extent and therefore confer higher resistance to the reporter antibiotic (e.g. ampicillin).

stability of the intervening Gβ1 variant. Those of lower stability (and presumably higher flexibility) up-regulate the reporter to a greater extent relative to chimeras that contain more stable Gβ1 variants. Starting with a variant of low overall stability (i.e. T_M = 38 °C), the codons for three specific residue positions were randomized and the resulting library was screened for variants of higher stability. In addition, the three positions were screened virtually using the ORBIT suite of protein design algorithms [29]. The computationally derived variant was significantly stabilized (i.e. T_M increased from 38 to 64 °C) and the five variants that were obtained from the engineered library had melting temperatures greater than the starting sequence and one that was more stable than the computationally derived variant (i.e. T_M = 69 °C) [54•]. Subsequent screening of the engineered library has resulted in five additional variants more stable than the starting sequence. We intend to use the ORBIT algorithms to further analyze the ten library variants to ascertain how well calculated energies correlate to the experimental screen results. The unique use of these transcriptional elements (in combination with robust computational analysis) represents new possibilities for the creation of novel combinatorial screens that should provide yet more opportunities to rapidly and accurately explore regions of protein sequence space.

Conclusions

Significant advances in the field of protein design have come from both computational approaches and *in vivo* and *in vitro* screens. With the computational approach, molecular diversity is generated and screened virtually and normally results in a mutated sequence that reflects the best attempt to mathematically capture the physical chemistry of protein folding (usually represented by a molecular mechanics force field). A caveat to this approach is that many of the physical interactions that give rise to the final folded structure of a protein are difficult to accurately emulate (e.g. variable electrostatics) and/or do not lend themselves readily to the large-scale combinatorial analysis of the enormous number of potential sidechain interactions (e.g. pairwise deconvolution of solvation energies). Even with these caveats, computational approaches have consistently proved to be efficient at getting close to the lowest energy mutant sequence (and conformation). By contrast, engineered libraries and associated screens have also been used successfully to determine the optimal sequence for specific design goals, but have practical limitations regarding library size or identification of optimal amino acid positions for mutagenesis. The emerging trend in which robust computational algorithms are blended with powerful *in vivo* screens should lead to greater and more rapid success in the field of protein design, as each method can be used to effectively address and alleviate the inherent limitations of the other.

Update

In recently published work, Olson and Roberts [55] demonstrate a simple yet efficient method for imparting

(and assessing) relatively significant molecular diversity (i.e. 30×10^{13}) into an engineered protein library. The protein scaffold targeted for design in this work is the tenth fibronectin domain of human fibronectin (10FnIII) and the resulting protein library will ultimately be utilized for mRNA display. The main goal of this work was to assess the quality of the diverse protein library by measuring the total amount of soluble protein produced, the ratio of folded to unfolded protein, and the free energy of folding. The relevance of this work to this review is that the authors utilized the GFP solubility screen described above [12] to determine the extent of expression and folding statistics for the total amount of soluble protein produced in bacteria. A wide range of solubility for proteins of the engineered library was observed and the majority of the library (66%) expresses a good amount of soluble protein [55]. Thus, this work further demonstrates the utility of the GFP screen and the inherent value of engineered protein libraries that possess significant molecular diversity.

Acknowledgements

This material is based upon work supported by the National Science Foundation under career grant number 0448670, as well as the Donors of the American Chemical Society Petroleum Research Fund, the Blasker-Rose-Miah fund of the San Diego Foundation, and the California Metabolic Research Foundation. All opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. NB is a recipient of an Arne N Wick Pre-doctoral Research Fellowship from the California Metabolic Research Foundation.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Jackson SE, Craggs TD, Huang J-R: **Understanding the folding of GFP using biophysical techniques**. *Expert Rev Proteomics* 2006, **3**:545-559.

2. Evanko D: **Training GFP to fold**. *Nat Methods* 2006, **3**:76.

3. Miyawaki A, Nagai T, Mizuno H: **Engineering fluorescent proteins**. *Adv Biochem Eng Biotechnol* 2005, **95**:1-15.

4. Wachter RM: **The family of GFP-like proteins: structure, function, photophysics and biosensor applications. Introduction and perspective**. *Photochem Photobiol* 2006, **82**:339-344.

5. Scruggs AW, Flores CL, Wachter R, Woodbury NW: **Development and characterization of green fluorescent protein mutants with altered lifetimes**. *Biochemistry* 2005, **44**:13377-13384.

6. Zhang L, Patel HN, Lappe JW, Wachter RM: **Reaction progress of chromophore biogenesis in green fluorescent protein**. *J Am Chem Soc* 2006, **128**:4766-4772.

7. Magliery TJ, Regan L: **Reassembled GFP: detecting protein-protein interactions and protein expression patterns**. *Methods Biochem Anal* 2006, **47**:391-405.

8. Wilson CG, Magliery TJ, Regan L: **Detecting protein–protein interactions with GFP-fragment reassembly**. *Nat Methods* 2004, **1**:255-262.

9. Waldo GS: **Genetic screens and directed evolution for protein solubility**. *Curr Opin Chem Biol* 2003, **7**:33-38.

10. Heim R, Cubitt AB, Tsien RY: **Improved green fluorescence**. *Nature* 1995, **373**:663-664.

11. Cubitt AB, Heim R, Adams SR, Boyd AE, Gross LA, Tsien RY: **Understanding, improving and using green fluorescent proteins**. *Trends Biochem Sci* 1995, **20**:448-455.

12. Waldo GS, Standish BM, Berendzen J, Terwilliger TC: **Rapid protein-folding assay using green fluorescent protein**. *Nat Biotechnol* 1999, **17**:691-695.

13. Cabantous S, Terwilliger TC, Waldo GS: **Protein tagging and detection with engineered self-assembling fragments of green fluorescent protein**. *Nat Biotechnol* 2005, **23**:102-107.

14. Cabantous S, Waldo GS: ***In vivo* and *in vitro* protein solubility assays using split GFP**. *Nat Methods* 2006, **3**:845-854.

15. Cabantous S, Pedelacq JD, Mark BL, Naranjo C, Terwilliger TC, Waldo GS: **Recent advances in GFP folding reporter and split-GFP solubility reporter technologies. Application to improving the folding and solubility of recalcitrant proteins from Mycobacterium tuberculosis**. *J Struct Funct Genomics* 2005, **6**:113-119.

16. •• Pedelacq JD, Cabantous S, Tran T, Terwilliger TC, Waldo GS: **Engineering and characterization of a superfolder green fluorescent protein**. *Nat Biotechnol* 2006, **24**:79-88.
This is a comprehensive description of the use of directed evolution to generate a highly stable GFP variant. A detailed analysis of the crystal structure, and rationale for the increased stability, is provided.

17. Crameri A, Whitehorn EA, Tate E, Stemmer WP: **Improved green fluorescent protein by molecular evolution using DNA shuffling**. *Nat Biotechnol* 1996, **14**:315-319.

18. Patterson GH, Knobel SM, Sharif WD, Kain SR, Piston DW: **Use of the green fluorescent protein and its mutants in quantitative fluorescence microscopy**. *Biophys J* 1997, **73**:2782-2790.

19. Wurth C, Guimard NK, Hecht MH: **Mutations that reduce aggregation of the Alzheimer's Abeta42 peptide: an unbiased search for the sequence determinants of Aβ amyloidogenesis**. *J Mol Biol* 2002, **319**:1279-1290.

20. Wurth C, Kim W, Hecht MH: **Combinatorial approaches to probe the sequence determinants of protein aggregation and amyloidogenicity**. *Protein Pept Lett* 2006, **13**:279-286.

21. Wang W, Hecht MH: **Rationally designed mutations convert *de novo* amyloid-like fibrils into monomeric beta-sheet proteins**. *Proc Natl Acad Sci USA* 2002, **99**:2760-2765.

22. Kim W, Hecht MH: **Sequence determinants of enhanced amyloidogenicity of Alzheimer Aβ42 peptide relative to Aβ40**. *J Biol Chem* 2005, **280**:35069-35076.

23. Petkova AT, Ishii Y, Balbach JJ, Antzutkin ON, Leapman RD, Delaglio F, Tycko R: **A structural model for Alzheimer's beta-amyloid fibrils based on experimental constraints from solid state NMR**. *Proc Natl Acad Sci USA* 2002, **99**:16742-16747.

24. •• Bradley LH, Thumfort PP, Hecht MH: ***De novo* proteins from binary-patterned combinatorial libraries**. *Methods Mol Biol* 2006, **340**:53-69.
This review provides a detailed description of the application and use of binary patterning to focus molecular diversity and create combinatorial libraries of well-folded *de novo* proteins.

25. Bradley LH, Wei Y, Thumfort P, Wurth C, Hecht MH: **Protein design by binary patterning of polar and nonpolar amino acids**. *Methods Mol Biol* 2007, **352**:155-166.

26. Kim W, Hecht MH: **Generic hydrophobic residues are sufficient to promote aggregation of the Alzheimer's Aβ42 peptide**. *Proc Natl Acad Sci USA* 2006, **103**:15824-15829.

27. Nelson R, Sawaya MR, Balbirnie M, Madsen AO, Riekel C, Grothe R, Eisenberg D: **Structure of the cross-beta spine of amyloid-like fibrils**. *Nature* 2005, **435**:773-778.

28. Kim W, Kim Y, Min J, Kim DJ, Chang YT, Hecht MH: **A high-throughput screen for compounds that inhibit aggregation of the Alzheimer's peptide**. *ACS Chem Biol* 2006, **1**:461-469.

29. Dahiyat BI, Mayo SL: ***De novo* protein design: fully automated sequence selection**. *Science* 1997, **278**:82-87.

30. Voigt CA, Mayo SL, Arnold FH, Wang ZG: **Computationally focusing the directed evolution of proteins**. *J Cell Biochem Suppl* 2001, **37**:58-63.

31. Meyer MM, Silberg JJ, Voigt CA, Endelman JB, Mayo SL, Wang ZG, Arnold FH: **Library analysis of SCHEMA-guided protein recombination**. *Protein Sci* 2003, **12**:1686-1693.

32. Voigt CA, Martinez C, Wang ZG, Mayo SL, Arnold FH: **Protein building blocks preserved by recombination**. *Nat Struct Biol* 2002, **9**:553-558.

33. Patrick WM, Firth AE: **Strategies and computational tools for improving randomized protein libraries**. *Biomol Eng* 2005, **22**:105-112.

34. Denault M, Pelletier JN: **Protein library design and screening: working out the probabilities**. *Methods Mol Biol* 2007, **352**:127-154.

35. • Hayes RJ, Bentzien J, Ary ML, Hwang MY, Jacinto JM, Vielmetter J, Kundu A, Dahiyat BI: **Combining computational and experimental screening for rapid optimization of protein properties**. *Proc Natl Acad Sci USA* 2002, **99**:15926-15931.
This work represents an excellent example of the use of protein design algorithms to focus sequence space to that which is most probable to produce a mutant protein variant with the desired properties.

36. Mena MA, Treynor TP, Mayo SL, Daugherty PS: **Blue fluorescent proteins with enhanced brightness and photostability from a structurally targeted library**. *Nat Biotechnol* 2006, **24**:1569-1571.

37. • Treynor TP, Vizcarra CL, Nedelcu D, Mayo SL: **Computationally designed libraries of fluorescent proteins evaluated by preservation and diversity of function**. *Proc Natl Acad Sci USA* 2007, **104**:48-53.
In this work (in conjunction with [36]), seven different protein design algorithms are evaluated for the ability to impart a new function into a protein without destroying the protein altogether. The results of the comparison are evaluated against variants generated by error-prone PCR.

38. Shaner NC, Steinbach PA, Tsien RY: **A guide to choosing fluorescent proteins**. *Nat Methods* 2005, **2**:905-909.

39. Choi EJ, Mayo SL: **Generation and analysis of proline mutants in protein G**. *Protein Eng Des Sel* 2006, **19**:285-289.

40. Tsai H-HG, Tsai C-J, Ma B, Nussinov R: ***In silico* protein design by combinatorial assembly of protein building blocks**. *Protein Sci* 2004, **13**:2753-2765.

41. Ross SA, Sarisky CA, Su A, Mayo SL: **Designed protein G core variants fold to native-like structures: sequence selection by ORBIT tolerates variation in backbone specification**. *Protein Sci* 2001, **10**:450-454.

42. Malakauskas SM, Mayo SL: **Design, structure and stability of a hyperthermophilic protein variant**. *Nat Struct Biol* 1998, **5**:470-475.

43. Dahiyat BI, Mayo SL: **Probing the role of packing specificity in protein design**. *Proc Natl Acad Sci USA* 1997, **94**:10172-10177.

44. Distefano MD, Zhong A, Cochran AG: **Quantifying β-sheet stability by phage display**. *J Mol Biol* 2002, **322**:179-188.

45. Alexander PA, Rozak DA, Orban J, Bryan PN: **Directed evolution of highly homologous proteins with different folds by phage display: implications for the protein folding code**. *Biochemistry* 2005, **44**:14045-14054.

46. Kotz JD, Bond CJ, Cochran AG: **Phage-display as a tool for quantifying protein stability determinants**. *Eur J Biochem* 2004, **271**:1623-1629.

47. Bai Y, Feng H: **Selection of stably folded proteins by phage-display with proteolysis**. *Eur J Biochem* 2004, **271**:1609-1614.

48. Feng H, Bai Y: **Repacking of hydrophobic residues in a stable mutant of apocytochrome *b*562 selected by phage-display and proteolysis**. *Proteins* 2004, **56**:426-429.

49. Scalley-Kim M, Minard P, Baker D: **Low free energy cost of very long loop insertions in proteins**. *Protein Sci* 2003, **12**:197-206.

50. Minard P, Scalley-Kim M, Watters A, Baker D: **A ''loop entropy reduction'' phage-display selection for folded amino acid sequences**. *Protein Sci* 2001, **10**:129-134.

51. Chu R, Takei J, Knowlton JR, Andrykovitch M, Pei W, Kajava AV, Steinbach PJ, Ji X, Bai Y: **Redesign of a four-helix bundle protein by phage display coupled with proteolysis and structural characterization by NMR and X-ray crystallography**. *J Mol Biol* 2002, **323**:253-262.

52. • Wunderlich M, Martin A, Staab CA, Schmid FX: **Evolutionary protein stabilization in comparison with computational design**. *J Mol Biol* 2005, **351**:1160-1168.
In this work (as well as that described in [53]), phage display and *in vitro* proteolysis are used to generate and screen for mutant Gβ1 variants that have increased stability. The results are compared to a previous design of Gβ1 that used computational design algorithms to increase protein stability.

53. Wunderlich M, Schmid FX: ***In vitro* evolution of a hyperstable Gβ1 variant**. *J Mol Biol* 2006, **363**:545-557.

54. • Barakat NH, Carmody LJ, Love JJ: **Exploiting elements of transcriptional machinery to enhance protein stability**. *J Mol Biol* 2007, **366**:103-116.
This work describes a novel application of transcriptional elements to create an *in vivo* screen for protein stability. The results obtained from the screen are compared and contrasted to those obtained from the application of protein design algorithms.

55. Olson CA, Roberts RW: **Design, expression, and stability of a diverse protein library based on the human fibronectin type III domain**. *Protein Sci* 2007, **16**:476-484.