



ELSEVIER

Engineering by homologous recombination: exploring sequence and function within a conserved fold

Martina N Carbone¹ and Frances H Arnold^{1,2}

In nature similar protein folds accommodate distant sequences and support diverse functions. This observation coupled with the recognition that proteins can tolerate many homologous substitutions inspires protein engineers to use recombination to search for new functions within sequences encoding structurally related molecules. These searches have led to proteins with novel activities, diversified specificities and greater stabilities. Computational methods that exploit structural and evolutionary information are being used to design highly mutated yet still natively folded chimeric proteins and protein libraries.

Addresses

¹ Division of Chemistry and Chemical Engineering, California Institute of Technology, Mail code 210-41, Pasadena, CA 91125, USA

² Biochemistry and Molecular Biophysics, California Institute of Technology, Mail code 210-41, Pasadena, CA 91125, USA

Corresponding author: Arnold, Frances H (frances@cheme.caltech.edu)

Current Opinion in Structural Biology 2007, 17:454–459

This review comes from a themed issue on
Engineering and design
Edited by Steve Mayo and Andreas Pluckthun

0959-440X/\$ – see front matter
© 2007 Elsevier Ltd. All rights reserved.

DOI 10.1016/j.sbi.2007.08.005

Introduction

The most intricate human-designed machines pale in comparison with the complexity and stunning functionality of the proteins created by evolution. The bewildering complexity of how protein primary sequences encode these remarkable functions, such as catalyzing in a few seconds chemical reactions that would otherwise take millions of years, poses a significant challenge to would-be designers of new proteins. This review discusses recent efforts to engineer new proteins by recombining related proteins, efforts that borrow from nature's toolbox of mutational moves and attempt to both circumvent and remedy our profound ignorance of the details that dictate function. Non-homologous recombination is beyond the scope of this review; we refer the interested reader to reference [1].

Homologous recombination distinguishes itself from other protein engineering strategies (such as point mutagenesis) in that it explores distant regions of sequence space: proteins that differ in many tens or even hundreds of

amino acids from known proteins yet still fold and function can be constructed. The inspiration to use homologous recombination in protein engineering comes from the observation that proteins with identical folds can diverge greatly not only in sequence but also in function. Thus, laboratory-constructed protein 'chimeras' with fragments swapped between two parent sequences have been used widely to pinpoint which sequence differences determine functional differences. More recently, however, the recognition that recombination can generate a high level of sequence diversity while conserving fold has been exploited in more exploratory studies in which recombination is used to make large libraries of new proteins. These studies address the following questions: (1) When is recombination not disruptive to structure (and function)? (2) What novel functions are accessible by the (largely conservative) mutations made by recombination? (3) Can the novel sequences generated by recombination be used to probe relationships between sequence and structure or function?

The first part of this review discusses recombination as an evolutionary search strategy. This is followed by a brief survey of recent studies that explore the functional diversity accessible to chimeric proteins and includes examples of recombination leading to sequence–structure–function inferences. Finally, we discuss structure-guided and evolutionarily guided recombination and efforts to expand the search for new function to more distant regions of sequence space.

Recombination as an evolutionary search strategy

The 'ruggedness' of the fitness landscape (the fitness as a function of sequence) determines whether recombination can benefit an evolutionary fitness optimization. In the limit of perfectly rugged landscapes, the fitness values of adjacent sequences are not correlated, and recombination cannot aid an evolutionary search. Recombination requires that local optima be clustered on the fitness landscape such that the region of sequence space enclosed by the parental sequences has, on average, higher fitness than a randomly selected portion [2].

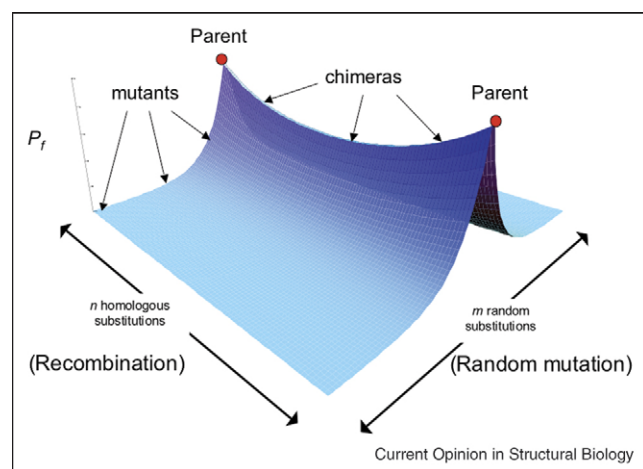
In a recent study, Drummond *et al.* [3*] compared random mutation to recombination, investigating how the probability of retaining fold (or parental function) depends on the number of mutations introduced. Random mutations cause a steep, exponential decay in this probability. As is well appreciated by protein engineers and protein scientists, most mutations are deleterious. As chimeras migrate

from one functional native sequence to the next, however, the likelihood of preserving structure or function follows a parabolic curve whose initial slope is much less steep (Figure 1). With data from chimeric mutated and randomly mutated β -lactamases, Drummond *et al.* showed that recombination is much more conservative than random mutation, leading to a folding probability that is many orders of magnitude greater at the highest mutation levels. These results provide direct evidence that natural protein fitness landscapes are well suited for recombination. By exploiting the conservative nature of mutations introduced into a structure that has already proven to tolerate them, recombination creates chimeric enzymes that are distant from one another in sequence with minimal loss in their probability of folding.

Recombination of homologous proteins creates novel and improved phenotypes

Using information inherent in native sequences, recombination preserves fold while diversifying sequence. Recent studies have demonstrated how this sequence diversification enables the discovery of new functional proteins, such as enzymes with new specificities or catalytic activities. Thulasiram *et al.* [4^{*}] created 11 single-crossover chimeras of farnesyl diphosphate synthase and chrysanthemyl diphosphate synthase from *Artemisia tridentata* ssp. *spiciformis* sharing 75% sequence identity. The enzymes synthesize isoprenoid compounds from chain elongation, cyclopropanation and branching reactions. Their chimeras acquired the ability to produce isoprenoid compounds from a cyclobutanation activity absent in the parents.

Figure 1



Chimeras occupy a functionally enriched ridge in sequence space. Surface height represents the probability of retaining fold as a function of random and homologous substitutions. Substituting amino acids that already exist in a homologous protein is much more conservative of structure and function than random substitutions. Figure reproduced from reference [3^{*}]. Copyright 2005 National Academy of Sciences, U.S.A.

Construction of complex synthetic cellular networks requires macromolecular building blocks such as DNA-binding proteins and protein interaction domains. Giesecke *et al.* [5] recombined cys_2his_2 zinc finger domains from eight homologous transcription factors to create libraries of chimeric zinc finger proteins exhibiting novel finger–finger interaction specificities. The authors showed that their chimeric proteins could mediate activation of a reporter gene in bacterial cells and of an endogenous gene in human cells more efficiently than the wild-type zinc fingers.

Understanding the mechanics of proteins provides the opportunity to engineer protein-based materials and use proteins as building blocks for the construction of nano-mechanical devices [6]. Sharma *et al.* generated four chimeras of the 27th and 32nd immunoglobulin domains from human cardiac titan, which share 42% sequence identity, and measured their mechanical properties by single molecule force microscopy. Each chimera was properly folded. Their resistance to stretching forces applied at the N-termini and C-termini differed from those of the parents and exhibited a wide distribution of values [6].

Landwehr *et al.* [7] studied the diversification of function within a library of cytochrome P450 enzymes assembled by the structure-guided recombination (see below) of *Bacillus megaterium* P450BM3 and two homologs. They measured the ability of the parents and 14 chimeric P450s to hydroxylate a set of 11 substrates, including 4 human drugs. The best enzyme on each compound was always a chimera, and some chimeras accepted substrates not accepted by any of the parents. Soluble, bacterial P450 chimeras that can produce drug metabolites may be useful for drug metabolic profiling and lead diversification.

Enzymes that retain a large fraction of their native residues while exhibiting a desired exogenous activity are useful for various biomedical applications. Glutathione transferases (GSTs) play a key role in cellular detoxification and may be useful as clinical therapeutics. Georgiou and co-workers [8^{*}] isolated a chimera from human and rat GSTs (54% identical) containing only 35 rat residues (out of 240) and exhibiting a 3.5-fold and 300-fold higher k_{cat} with 7-amino-4-chloromethyl coumarin relative to rat and human GST, respectively. Singh *et al.* [9] recombined the cysteine proteases falcipain-2 and berghepain-2 isolated from human and rodent malaria parasites, respectively (53% identical). Cysteine protease inhibitors block parasite development *in vitro* and cure mice with malaria. The authors identified a chimera that exhibited inhibitor specificities more similar to the human enzyme but differed from the rodent enzyme by only eight substitutions. Rodent malaria parasites engineered to exhibit biochemical properties more similar to their human parasite counterparts can facilitate drug discovery [9].

'Synthetic protein families' made in the laboratory can be used to explore determinants of structure and function free from the constraints of natural selection. Otey *et al.* [10[•]] used logistic regression to identify the sequence elements that contribute to folding and catalytic activity of chimeric P450s. Their analysis identified previously uncharacterized specificity-determining residues, whose significance was then verified experimentally. Taly *et al.* [11] created chimeras of mouse CYP1A1 and rabbit CYP1A2 and human CYP1A1 and CYP1A2 enzymes and tested their activities on three different substrates. Some of the chimeras exhibited increased specificities towards certain substrates, and the authors were able to associate specific sequence elements with the specificities. Mannervik and co-workers [12[•]] created a library of chimeric human GSTs from parents sharing 84% sequence identity to explore the divergence of specificity on eight alternative substrates. The authors clustered their variants by specificity and were able to identify individual amino acids responsible for the observed substrate profiles.

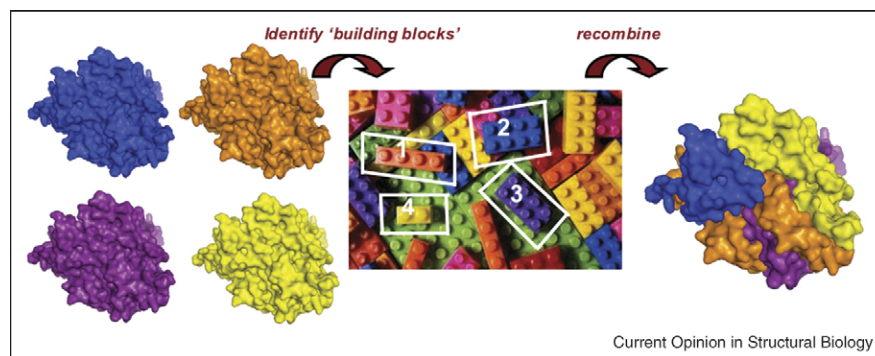
Expanding the search to more distant chimeras: exploiting information from protein structure

Crossovers introduced by recombination can separate native clusters of interacting amino acids. This phenomenon, known as crossover disruption, should be given careful consideration when designing chimeric proteins. Crossover disruption becomes a significant concern as the parents diverge. In fact, randomly recombining sequences with less than ~70% identity generates mostly unfolded proteins [13,14]. More distant sequences, however, are thought to offer the greatest opportunity for functional innovation because their offspring contain more mutations, particularly in functionally important parts of the molecule [8[•]]. Structure-guided recombination attempts to overcome this hurdle by directing crossovers to the least disruptive locations.

To minimize the number of broken native interactions, crossovers should partition the structure into structurally and functionally independent 'building blocks'. Thus, the challenge is to identify the boundaries that define independent structural units that, in analogy to Lego parts, preserve their functionality in different backgrounds (Figure 2). Several groups have proposed that 'contiguous peptide chains forming compact structures' or secondary structural elements could act as independent, recombinable structural units [15–17]. More recently, however, it has become clear that such elements cannot be defined without consideration of the parent sequences. Voigt *et al.* developed a computational algorithm that uses the parental structure and sequence identities to score the number of broken native pair-wise interactions in a chimera [18]. An optimization algorithm then directs crossovers to locations that minimize the average disruption in the library [19]. SCHEMA analyses on several proteins showed that in certain cases crossovers are less disruptive in the middle rather than at the ends of an alpha helix [10[•]], showing that the best partitions are not intrinsic to the tertiary fold but depend on the spatial organization of conserved residues. According to this framework, the interfaces between independent structural units are defined by interfaces composed primarily of conserved residues.

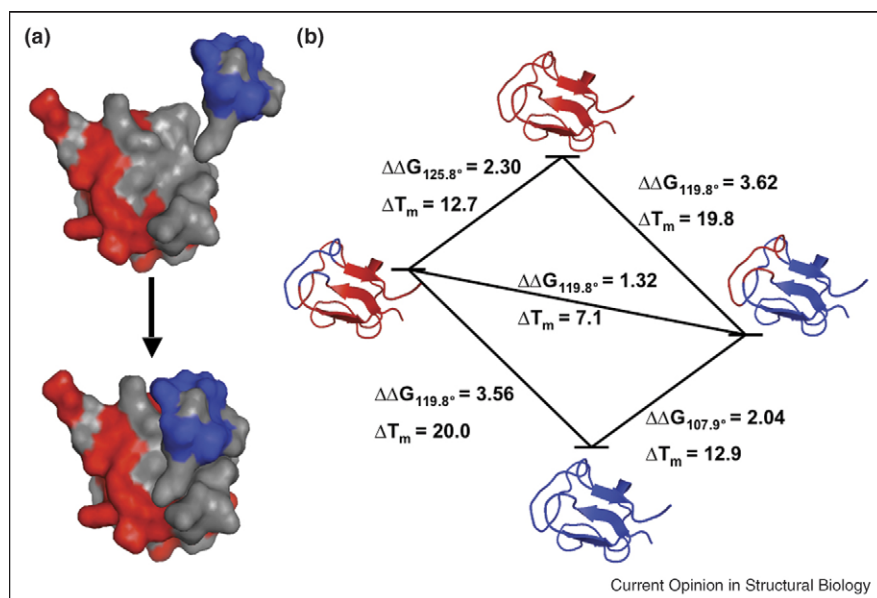
In the limit of perfectly conserved interfaces, such structural building blocks have been verified to act independently. Hernandez and LeMaster developed the algorithm HybNat, which partitions residues into mutually exclusive clusters of interacting amino acids [20]. Unlike SCHEMA, HybNat interactions are atom-based rather than residue-based, and conservative mutations do not contribute to disruption. The interfaces between the clusters are perfectly conserved, eliminating disruption upon recombination [21[•]]. LeMaster and Hernandez identified a perfectly conserved interface in the structures of the highly thermostable *Pf* rubredoxin and its mesophilic counterpart *Cp*

Figure 2



Identifying structural constituents of natively folded chimeric proteins. Lego representation of homologous recombination. Successful recombination requires identification of recombinable structural units or 'building blocks' whose integrity is preserved in different backgrounds.

Figure 3



Conserved interfaces partition a structure into energetically independent building blocks. **(a)** A conserved interface in the core of *Cp* and *Pf* rubredoxins partitions the proteins into two fragments that can be swapped with minimal disruption. Gray, conserved residues; red and blue, non-conserved residues from different parents. **(b)** Thermodynamic cycle for the parental and chimeric rubredoxins indicates that conserved interfaces identify building blocks that make additive contributions to free energy of folding and thermal transition temperature. $\Delta\Delta G$ values in kcal/mol. ΔT_m values in Celsius. (b) Reproduced with permission from reference [21*].

rubredoxin (59% identical) that partitions the proteins into two fragments (Figure 3a). Recombining these fragments showed that their contributions to free energy of folding and thermal transition temperature were additive (Figure 3b) [20,21*].

Using structure-guided SCHEMA recombination, Otey *et al.* [10*] partitioned the heme domains of cytochrome P450BM3 and homologs sharing 61–64% sequence identity into eight blocks and recombined those to make thousands of chimeric P450s. About 47% of the library encoded a properly folded P450, and of those more than 75% were functional. Functional chimeras differed from any known parent by up to 101 amino acid mutations (out of 466). Meyer *et al.* [22] recombined three β -lactamases sharing as low as 32% identity to create a SCHEMA library with 20% functional members that differed from their closest parent by up to 86 mutations (out of 265). This simultaneously high level of sequence diversity and frequency of function is not accessible by point mutagenesis or random recombination. Meyer *et al.* [22] also showed that, among chimeras with similar numbers of mutations relative to the closest parent, those with lower SCHEMA disruption were more likely to fold and function. This validates the physical significance of the SCHEMA score and supports the assumption that new interactions are, on average, disruptive to structure and function. As described in the previous section, various P450 chimeras with improved stabilities and activities

relative to the best parents were identified from this library, indicating that newly formed interactions enable functional evolution [7,10*].

Expanding the search to more distant chimeras: exploiting information from protein evolution

The information contained in a natural multiple sequence alignment (MSA) can also be used to minimize crossover disruption for recombination of distant sequences [23,24,25*,26]. Structure-guided and evolutionary-based recombination algorithms both seek to identify and preserve native interactions. The latter use the statistical covariance, or the pair-wise conservation of amino acids, within a natural MSA to identify important interactions. Evolutionary information in natural MSAs has been used to identify potential stabilizing mutations [27] in ‘consensus stabilization’ approaches, which assume that highly conserved residue positions represent energetically important sites and that the most frequent amino acids at those sites are the most stabilizing [28,29]. Similarly, co-evolving or highly conserved pairs of residues in a MSA may represent an evolutionary trace of energetically important interactions. Double-mutant thermodynamic cycles have, in fact, verified energetic coupling between co-evolving residues [30]. The FamClash algorithm developed by Maranas and co-workers uses a natural MSA to identify pairs of positions that exhibit conserved amino acids properties (hydrophobicity, charge and volume; ‘HCV’) and then

scores chimeras according to the total number of ‘clashes’, where a clash designates a pair of residues with physical properties that deviate from those found at the corresponding positions in the MSA [23]. Recently, Pantazes *et al.* [24] developed a hybrid of SCHEMA and FamClash that penalizes broken interactions, defined by a structural distance cutoff, in proportion to how dissimilar the HCV properties of their components are from those in the family sequence alignment. Ye *et al.* developed another algorithm that uses both evolutionary and structural information to score chimeras and design chimeric libraries [26]. Unlike other methods, their model takes into account higher than second order interactions. All of these algorithms can be used to direct crossovers to locations that optimize their scoring metric.

Ranganathan and co-workers [25•] developed an evolution-based algorithm for generating libraries whose residue statistics resemble those of the natural MSA. Co-evolving residues are identified by a perturbation method known as statistical coupling analysis (SCA) that stores the information in the SCA matrix [30]. Design proceeds by shuffling the columns of the natural MSA while minimizing the differences between the natural and artificial SCA matrix, thus creating protein libraries with conserved independent and pair-wise residue statistics. The resulting proteins contain homologous substitutions and are chimeras, albeit with many crossovers. Using this approach, Socolich *et al.* [25•] designed a library of 43 chimeric WW domains from 120 wild-type parental sequences. Twelve of the 43 members of this library were natively folded and differed on average by 14 amino acids to their closest natural WW domain (out of 35). Members of the library also preserved native-like activity [31]. The authors showed that preserving the independent residue statistics in the recombined MSA is insufficient to design folded sequences, validating the significance and necessity of conserving the SCA pair-wise statistics. Merely shuffling large numbers of homologous substitutions leads to significant crossover disruption, which is alleviated by conserving key pair-wise interactions.

Closely related proteins are robust to homologous mutations because the substitutions are recruited in an environment close to one that has already proven to tolerate them. Distant proteins are less robust to homologous mutations because they are recruited in an environment of novel interactions, which on average are deleterious to structural integrity. Algorithms that maximize the conservation of interactions enable the recombination of more distant parents. This conceptual framework is also applicable to larger biological systems. Recently, Martin *et al.* [32] recombined the genomes of various strains of the Maize streak virus and showed that genes involved in complex networks of interactions cannot tolerate many nucleotide substitutions when swapped with corresponding genes from different species. On the

contrary, modular genes are more robust to recombination. These results suggest that a known genetic network of interactions can be used to guide the design of a recombinant genome much like residue interactions guide the design of chimeric proteins.

Conclusion

The protein products of homologous recombination can acquire novel and useful biophysical features. Examples covered in this review suggest that chimeras will find uses in the biosynthesis of human metabolites, construction of advanced biomaterials and the regulation of cellular function, among others. It is likely that structure-guided and evolutionarily-guided recombination will allow access to ever more distant protein sequences, with important consequences for understanding how sequence encodes function.

Acknowledgements

This work is supported by the National Institutes of Health (R01 GM068664-0) and a National Science Foundation Predoctoral Fellowship (to MC).

References and recommended reading

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

- of special interest
- of outstanding interest

1. Ostermeier M, Benkovic SJ: **Evolution of protein function by domain swapping.** *Adv Protein Chem* 2000, **55**:29-77.
2. Kauffman S: **Biological implications of rugged fitness landscapes.** In: *The Origins of Order.* Oxford University Press; 1993: 65-120.
3. Drummond DA, Silberg JJ, Meyer MM, Wilke CO, Arnold FH: **On the conservative nature of intragenic recombination.** *Proc Natl Acad Sci USA* 2005, **102**:5380-5385.
4. Thulasiram HV, Erickson HK, Poulter CD: **Chimeras of two isoprenoid synthases catalyze all four coupling reactions in isoprenoid biosynthesis.** *Science* 2007, **316**:73-76.
5. Giesecke AV, Fang R, Joung JK: **Synthetic protein-protein interaction domains created by shuffling Cys₂His₂ zinc-fingers.** *Mol Syst Biol* 2006, **2**:2011.
6. Sharma D, Cao Y, Li H: **Engineering proteins with novel mechanical properties by recombination of protein fragments.** *Angew Chem Int Ed Engl* 2006, **45**:5633-5638.
7. Landwehr M, Carbone M, Otey CR, Li Y, Arnold FH: **Diversification of catalytic function in a synthetic family of chimeric cytochrome p450s.** *Chem Biol* 2007, **14**:269-278.
8. Griswold KE, Kawarasaki Y, Ghoneim N, Benkovic SJ, Iverson BL, Georgiou G: **Evolution of highly active enzymes by homology-independent recombination.** *Proc Natl Acad Sci USA* 2005, **102**:10082-10087.

A chimera of rat and human theta-class GSTs, containing mostly human residues and possessing a 3.5-fold and 300-fold higher k_{cat} with 7-amino-

4-chloromethyl coumarin (relative to rat and human GSTs, respectively), was isolated. GSTs play a crucial role in cellular detoxification and enzymes engineered to retain a large fraction of human residues while exhibiting a desired exogenous activity may be useful as clinical therapeutics.

9. Singh A, Walker KJ, Sijwali PS, Lau AL, Rosenthal PJ: **A chimeric cysteine protease of *Plasmodium berghei* engineered to resemble the *Plasmodium falciparum* protease falcipain-2.** *Protein Eng Des Sel* 2007, **20**:171-177.
10. Otey CR, Landwehr M, Endelman JB, Hiraga K, Bloom JD,
 - Arnold FH: **Structure-guided recombination creates an artificial family of cytochromes P450.** *PLoS Biol* 2006, **4**:e112.
 Using SCHEMA recombination the authors create thousands of folded and functional chimeras of three bacterial cytochrome P450s. The functional chimeras differ from their closest parent by as many as 101 amino acids. This study illustrates how structure-guided recombination can create diverse libraries of folded proteins that can be used to probe sequence–structure–function relationships.
11. Taly V, Urban P, Truan G, Pompon D: **A combinatorial approach to substrate discrimination in the P450 CYP1A subfamily.** *Biochim Biophys Acta* 2007, **1770**:446-457.
12. Kurtovic S, Runarsdottir A, Emren LO, Larsson AK, Mannervik B:
 - **Multivariate-activity mining for molecular quasi-species in a glutathione transferase mutant library.** *Protein Eng Des Sel* 2007, **20**:243-256.
 The authors recombine two human glutathione transferases and study the diversification of function with eight alternative substrates. The paper surveys the different multivariate analyses, including *k*-means and hierarchical clustering and principle component analysis, which can be used to group chimeras by specificity. Furthermore, the authors are able to associate specific amino acids with substrate specificity illustrating how clustering by specificity can lead to sequence–function inferences.
13. Ostermeier M, Shim JH, Benkovic SJ: **A combinatorial approach to hybrid enzymes independent of DNA homology.** *Nat Biotechnol* 1999, **17**:1205-1209.
14. Lutz S, Ostermeier M, Moore GL, Maranas CD, Benkovic SJ: **Creating multiple-crossover DNA libraries independent of sequence identity.** *Proc Natl Acad Sci USA* 2001, **98**:11248-11253.
15. Tsuji T, Onimaru M, Yanagawa H: **Towards the creation of novel proteins by block shuffling.** *Comb Chem High Throughput Screen* 2006, **9**:259-269.
16. O'Maille PE, Tsai MD, Greenhagen BT, Chappell J, Noel JP: **Gene library synthesis by structure-based combinatorial protein engineering.** *Methods Enzymol* 2004, **388**:75-91.
17. Go M: **Protein structures and split genes.** *Adv Biophys* 1985, **19**:91-131.
18. Voigt CA, Martinez C, Wang ZG, Mayo SL, Arnold FH: **Protein building blocks preserved by recombination.** *Nat Struct Biol* 2002, **9**:553-558.
19. Endelman JB, Silberg JJ, Wang ZG, Arnold FH: **Site-directed protein recombination as a shortest-path problem.** *Protein Eng Des Sel* 2004, **17**:589-594.
20. Hernandez G, LeMaster DM: **Hybrid native partitioning of interactions among nonconserved residues in chimeric proteins.** *Proteins* 2005, **60**:723-731.
21. LeMaster DM, Hernandez G: **Additivity in both thermodynamic stability and thermal transition temperature for rubredoxin chimeras via hybrid native partitioning.** *Structure* 2005, **13**:1153-1163.

This study shows that fragments separated by perfectly conserved interfaces make additive contributions to the free energy of folding and thermal transition temperature when recombined. Algorithms that can identify such independent sequence elements will prove valuable in probing sequence–structure–function relationships.
22. Meyer MM, Hochrein L, Arnold FH: **Structure-guided SCHEMA recombination of distantly related beta-lactamases.** *Protein Eng Des Sel* 2006, **19**:563-570.
23. Saraf MC, Horswill AR, Benkovic SJ, Maranas CD: **FamClash: a method for ranking the activity of engineered enzymes.** *Proc Natl Acad Sci USA* 2004, **101**:4142-4147.
24. Pantazes R, Saraf MC, Maranas CD: **Optimal protein library design using recombination or point mutagenesis based sequence based scoring functions.** *Protein Eng Des Sel* 2007, online early access.
25. Socolich M, Lockless SW, Russ WP, Lee H, Gardner KH,
 - Ranganathan R: **Evolutionary information for specifying a protein fold.** *Nature* 2005, **437**:512-518.
 Using an evolution-based recombination algorithm, the authors create 12 natively folded chimeras sharing on average 63% sequence identity to their closest relative. This study illustrates how the information contained in a natural MSA can be used to construct diverse libraries with natively folded members.
26. Ye X, Friedman A, Bailey-Kellogg C: **Hypergraph model of multi-residue interactions in proteins: sequentially-constrained partitioning algorithms for optimization of site-directed recombination.** *Lect Notes Comput Sci* 2006, 15-29.
27. Lehmann M, Kostrewa D, Wyss M, Brugger R, D'Arcy A, Pasamontes L, van Loon AP: **From DNA sequence to improved functionality: using protein sequence comparisons to rapidly design a thermostable consensus phytase.** *Protein Eng* 2000, **13**:49-57.
28. Steipe B, Schiller B, Pluckthun A, Steinbacher S: **Sequence statistics reliably predict stabilizing mutations in a protein domain.** *J Mol Biol* 1994, **240**:188-192.
29. Finkelstein AV, Badretdinov A, Gutin AM: **Why do protein architectures have Boltzmann-like statistics?** *Proteins* 1995, **23**:142-150.
30. Lockless SW, Ranganathan R: **Evolutionarily conserved pathways of energetic connectivity in protein families.** *Science* 1999, **286**:295-299.
31. Russ WP, Lowery DM, Mishra P, Yaffe MB, Ranganathan R: **Natural-like function in artificial WW domains.** *Nature* 2005, **437**:579-583.
32. Martin DP, van der Walt E, Posada D, Rybicki EP: **The evolutionary value of recombination is constrained by genome modularity.** *PLoS Genet* 2005, **1**:e51.