

Protein redesign

Lynne Regan

It has been demonstrated that, for a number of proteins, it is possible to dramatically alter the connectivities between elements of secondary structure. Remarkably large loop insertions are tolerated and many redesigns have generated proteins that successfully fold to stable, active structures. Some redesigns have been entirely the choice of the investigators, whereas others have incorporated a randomization and selection step to identify optimal sequences. These studies have provided basic guidelines for the rational manipulation of protein structure and stability, they have allowed the dissection of folding pathways and they have generated proteins with the potential for practical therapeutic applications.

Addresses

Yale University, Department of Molecular Biophysics and Biochemistry,
266 Whitney Avenue, New Haven, CT 06520, USA;
e-mail: lynne@csbgly.csb.yale.edu

Current Opinion in Structural Biology 1999, **9**:494–499

<http://biomednet.com/eleceref/0959440X00900494>

© Elsevier Science Ltd ISSN 0959-440X

Abbreviations

CD	circular dichroism
CI-2	chymotrypsin inhibitor-2
DHFR	dihydrofolate reductase
DsbA	thiol-disulfide oxidoreductase
GCSF	granulocyte colony stimulating factor
IL-3	interleukin-3
SH	Src homology

Introduction

Recently, there have been a number of significant accomplishments in the area of 'protein redesign'. By this phrase, I refer to designs that start with a protein of known structure and dramatically modify key features of its construction. I distinguish such wholesale redesigns from protein engineering studies involving point mutations and also from designs that are completely *de novo*. The particular examples that are the focus of this review are redesigns that introduce dramatic changes in the connectivity between elements of secondary structure: varying loop length and creating circular permutations. As a result of space limitations, I focus on examples for which the aim is to investigate the structural, energetic and folding consequences of the redesigns, rather than on those whose goal is to alter the specificity of protein–ligand interactions or enzymatic activity. The results of these studies provide important insights into the balance of forces that stabilizes natural proteins, they aid in the dissection of key stages in protein folding and they have the potential to fine-tune a protein's properties for therapeutic applications.

Varying loop length

It is of interest to investigate the role that loops between elements of secondary structure play in specifying the

structure of a particular protein: can large insertions be tolerated? For a remarkable number of proteins, the answer is yes. One would predict that, as the loop length increased, protein stability would decrease, as constraining the loop ends in the folded protein became increasingly entropically costly. This effect is seen and agrees well with the observation that thermophilic proteins tend to have shorter surface loops than their mesophilic counterparts. A final point of interest is the effect of loop insertions on the kinetics of protein folding. Can the behavior of these redesigned proteins shed light upon the nature of the transition state and the rate-determining step or the nature of folding intermediates? Below, I discuss some specific protein redesign examples to illustrate these points.

Rop

Nagi *et al.* [1,2**] inserted (Gly)_n linkers (where n ranged from 1 to 10) between two α helices in the four-helix bundle protein Rop (Figure 1). They found that all the loop length insertions gave rise to folded, native proteins that bind Rop's RNA substrate with wild-type-like affinity. As the length of the loop increased, the stability of the protein decreased and this effect could be modeled quite well by considering the entropic cost of loop closure [1]. An interesting follow-up to this work was to investigate the effect of increasing loop length on the kinetics of protein folding [2**]. As the loop length increased, the rate of formation of a folding intermediate decreased, suggesting that bringing distant parts of the molecule together was important in its formation. In addition, a study of the effect of loop length on the stability of the intermediate showed effects that mirrored the effects on the native protein: the stability of the intermediate decreased with increasing loop length. The effect of loop length on the stability of the intermediate supports a model for the intermediate in which the helix-loop-helix hairpins have associated, but the hydrophobic core has not yet achieved its final well-packed state.

Chymotrypsin inhibitor-2

Ladurner and Fersht [3] investigated the effects of inserting glycine, alanine and glutamine repeats, varying in length from 7 to 13 amino acids, at a suitable position within a large natural loop that is fairly centrally located in the protein chymotrypsin inhibitor-2 (CI-2) (Figure 2). All the redesigned proteins folded into a structure that, by CD, was indistinguishable from the wild-type protein, but as the length of the inserted loop increased, the stability of the protein decreased. All the mutants were found to preserve the two-step folding and unfolding kinetics observed for wild-type CI-2; however, they all folded somewhat more slowly and unfolded somewhat more rapidly. The greatest effects were seen for loop insertions involving amino acids with longer sidechains. In this example,

because increasing loop length had a relatively small effect on the folding rate, the authors concluded that the diffusion of the N and C termini towards each other is not the rate-determining step in the folding of CI-2 and suggested that the effect of the loop is to increase the configurational entropy of the transition state.

Arc

In the above examples, the identities of the amino acids in the inserted loops were glycine — the most flexible amino acid, alanine — the smallest nonglycine amino acid and glutamine — chosen with a view to mimicking glutamine insertion diseases. All of these are completely justifiable and effective choices, but how can an optimal loop length and composition be determined? Robinson and Sauer [4] took a systematic approach to this question. In their initial studies, the wild-type Arc dimer was converted into a single-chain variant using a 15-residue linker to connect the two subunits. This redesign was successful and the single-chain version of the protein was both stable and functional. But could the redesign be further optimized? To address this question, Robinson and Sauer [5**] created a 'library' of linkers, in which a 16-residue stretch within a 19-residue linker was randomized with combinations of alanine/glycine or serine/glycine. They found that, in the serine/glycine library, the most stable protein had seven serines and nine glycines, whereas in the alanine/glycine library, the most stable protein had 11 alanines and 5 glycines. Distinct differences in the behavior of proteins with different linker compositions were observed. Alanine-rich linkers stabilized the protein mainly by accelerating folding, whereas serine-rich linkers stabilized predominantly by slowing unfolding. There was also an optimum ratio of Ala : Gly or Ser : Gly that was required to generate proteins of maximum stability. These results emphasize that, although glycine is commonly used in linkers because of its high flexibility, this flexibility comes with an energetic cost: too much is detrimental and optimal linkers may need to incorporate more sequence variability.

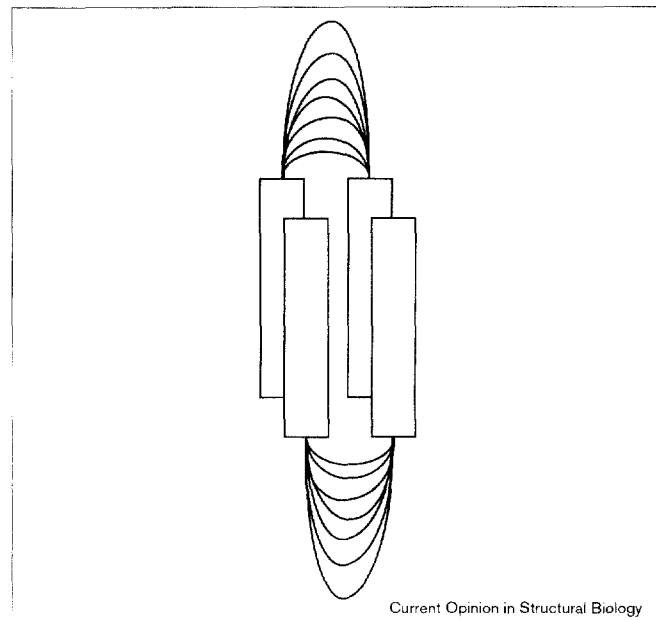
Circularly permuted proteins

Taking loop variation to further extremes, over the past few years, there have been a number of studies in which novel loops have been inserted into proteins in order to link the N and C termini, while breaks were made at various places to produce circularly permuted proteins [6–14]. Can such proteins fold? How does their stability compare with that of the wild-type? What is the influence of the permutation on the kinetics of folding and on the nature of the transition state? I will concentrate my discussion on a few of the most recent examples.

Src homology 3 domain

Viguera and Serrano [15] made variable-length loop insertions within three circular permutants of an α -spectrin SH3 domain, an orthogonal β -sandwich protein. They found that, as increasing numbers of glycine residues were inserted, protein stability decreased until a plateau was reached. They

Figure 1



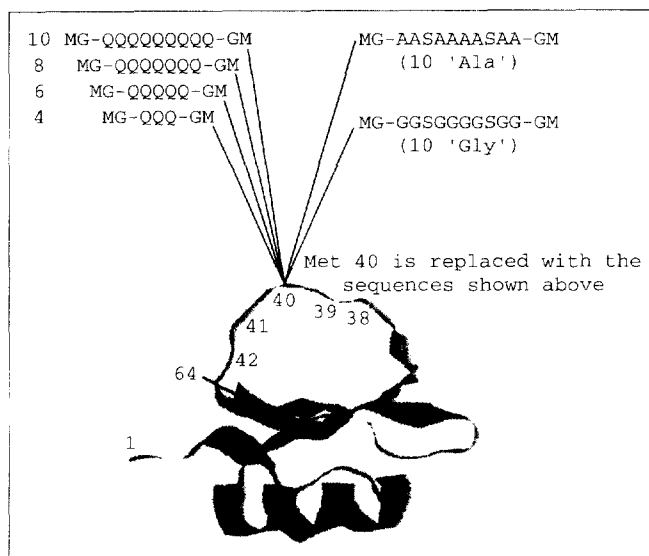
Schematic illustration of the loop-length variants of Rop [1].

found that the effective loop length appeared to behave as a loop that was somewhat longer than that expected from the number of glycine residues inserted and suggested that loop elongation as a result of the flexibility of the neighboring residues may be a component in this system.

It was of particular interest to study the effect of the linker insertions on the kinetics of folding. If the transition state involves making contacts between residues that are far away in the linear sequence, then the introduction of long unstructured loops should slow the rate of folding. They found that the rate of folding did indeed decrease as the number of glycine residues in the loop increased. They concluded that the diffusion of different parts of the protein with respect to each other occurred on going from the denatured state to the transition state and that the transition state was perhaps more homogeneous than had been proposed previously.

In a follow-up paper, the thermodynamics of this system was investigated in greater detail [16**]. These investigations suggested that the introduction of the first glycine actually slightly enhanced the protein stability, suggesting the presence of some intrinsic strain in the wild-type. Additional glycine insertions destabilized the protein, as originally reported. Investigation of the temperature dependence of the folding and unfolding transitions of the wild-type versus circularly permuted variants suggested that their transition states were different. In particular, differences in their heat capacities suggest that the transition state of the permutants is more solvent accessible than that of the wild-type.

Figure 2



Schematic illustration of the position and identity of the loop-length variants of CI-2. Reproduced with permission from [17**].

Chymotrypsin inhibitor-2

A different result was obtained from studies with a circular permutant of CI-2 [17**]. In this case, the permutation was accomplished by the introduction of two cysteine residues towards the N and C termini of the protein. Oxidation to form a disulfide generates a circular form of the protein that is more stable than the wild-type and that folds sevenfold more rapidly. A circularly permuted version of the protein can then be conveniently generated by CNBr cleavage at a unique methionine in the middle of the protein sequence. Mutagenic analysis, involving the independent substitution of 11 residues at key positions to alanine or glycine, was performed and the results for the wild-type, circular protein and circular permutant were compared. By studying the effect of these mutations on the proteins' folding, it was possible to perform Φ -value analysis in order to probe the nature of the transition state for each protein. In contrast to the SH3 study, the authors found that the folding nucleus was conserved in all three proteins, with only subtle differences among them. One difference was that the constraint of the disulfide seemed to impose a more uniformly native-like transition state on the circular protein, in comparison with the other two proteins, but, basically, the folding core is conserved. The authors suggest that the difference between the behavior of circular permutants of CI2 (folding nucleus retained) and SH3 (folding nucleus altered) may reflect the fact that, in the SH3 studies, the protein was actually cleaved at a point within the proposed folding nucleus.

β 2-crystallin

Wielgmann *et al.* [18] produced one of the few, if not only, examples of a circularly permuted protein for which the circular permutant displays enhanced stability relative to

the parent protein. The two vertebrate eye lens polypeptides β 2-crystallin and γ B-crystallin fold into similar β -sheet structures, but the β -crystallins are oligomeric, whereas the γ -crystallins are monomeric. The aim of this circular permutation was to create a monomeric version of β 2, re-engineering it to adopt a γ -crystallin-like structure. The circular permutation did not achieve the initial aim of the redesign; the protein was not converted to a monomer, but still formed dimers. Interestingly, however, the circularly permuted protein was somewhat more resistant to chemical denaturation than the wild-type protein.

Thiol-disulfide oxidoreductase

An interesting example of the consequences of circular permutation is seen in the conversion of the catalytic disulfide bond of thiol-disulfide oxidoreductase (DsbA) into a structural bond [19]. The protein has a thioredoxin-like fold, with a single disulfide formed between the cysteine residues within the sequence Cys30-Pro31-His32-Cys33. In the natural protein, the presence of this catalytic disulfide destabilizes the protein by about 16 kJ. A circularly permuted variant of this protein was designed, in which the N and C termini are connected by a Gly₃-Thr-Gly linker, and the protein was cleaved between residues 31 and 32. The protein adopts a wild-type-like structure, but is catalytically inactive as dithiol oxidase. This observation can be explained as follows. The wild-type protein is strongly oxidizing because of the low pKa of Cys30 (pKa about 3.4) and the formation of the disulfide bond severely destabilizes the protein. In the permuted variant, the second and penultimate residues are joined, such that the entire mainchain of the protein forms a large loop. The pKa of both cysteine residues is now greater than 8 and the protein is more stable when the disulfide is present than when it is reduced. The stabilizing effect of disulfide formation is, however, significantly less than predicted from a consideration of the entropic restriction of such a cross-link on the protein's denatured state, suggesting that the permutation has introduced steric strain or other unfavorable interactions into the protein.

Streptavidin-biotin

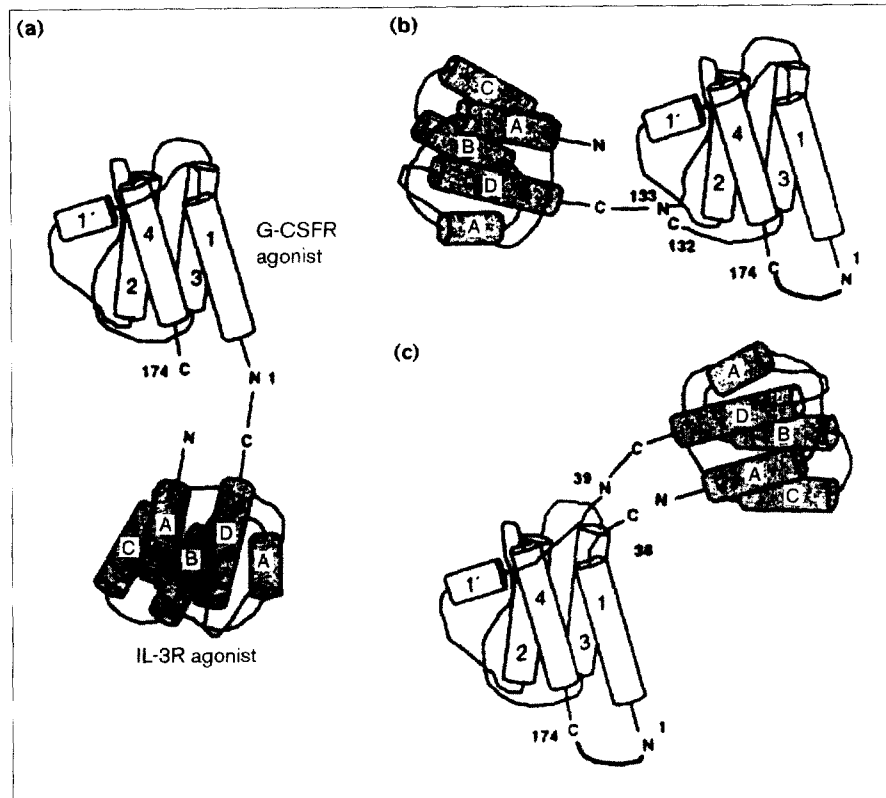
There is a large flexible loop in streptavidin that undergoes a significant conformational change when biotin is bound. Chu *et al.* [20] designed a circular permutant of streptavidin, in which a four amino acid linker joins the N and C termini and four residues are removed from the loop. Although the circularly permuted protein had a very similar overall structure to the wild-type, its affinity for biotin was greatly decreased. This result emphasizes the importance of the presentation of interacting residues within the constraints of a loop. The energetics of the interaction of a particular sequence can be very different if presented as a linear string, rather than in a loop.

Dihydrofolate reductase

Iwakura and Nakamura [21] recognized the importance of the connecting linker joining the N and C termini in

Figure 3

Schematic illustration of the relative orientations of GCSF and IL-3 (dark gray) when different circular permutants of GCSF are fused. (a) Fusion of wild-type GCSF. (b,c) Fusion of two different permutants. Reproduced with permission from [23**].



circularly permuted proteins. They therefore investigated the effect of using a variable length (Gly) n linker to connect the N and C termini of circularly permuted dihydrofolate reductase (DHFR), keeping the new N terminus constant at position 16. The characterization of the proteins showed that (Gly) 5 was optimal, though even this displayed only about 20% of the activity of the wild-type protein.

Granulocyte colony stimulating factor

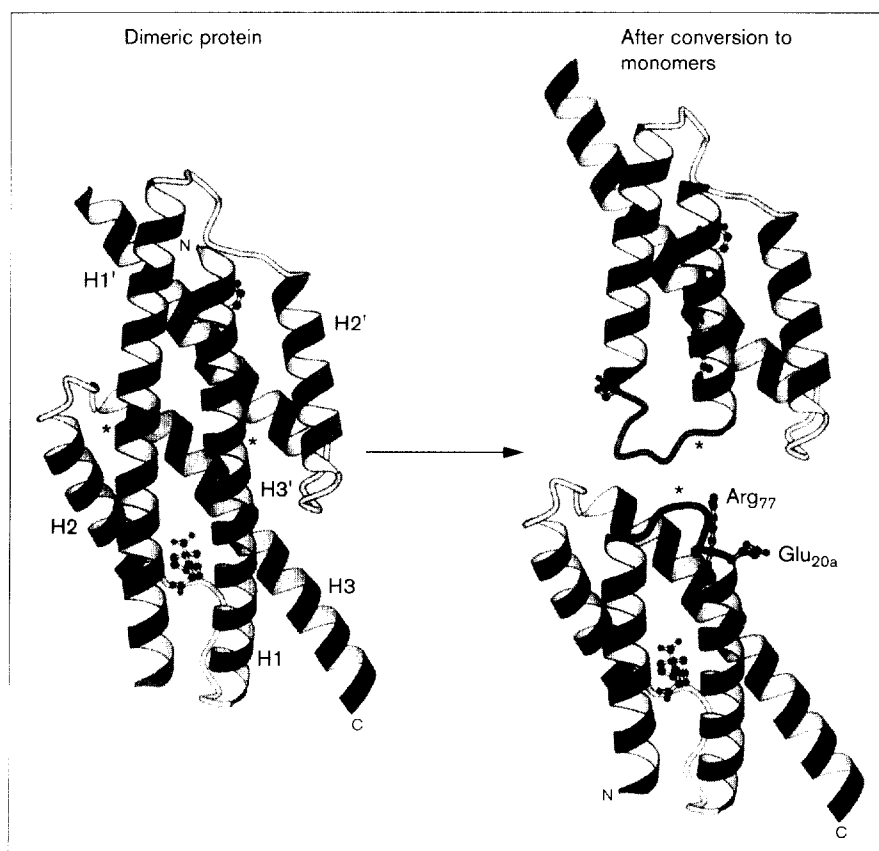
Granulocyte colony stimulating factor (GCSF) is a hematopoietic cytokine that promotes cell growth and differentiation. Therapeutically, it is important as a drug that is administered to patients undergoing chemotherapy to promote neutrophil production. GCSF is a four-helix bundle protein with two long 'overhand' connections. Feng *et al.* [22**] created a fairly extensive set of circularly permuted variants of GCSF, connecting the N and C termini and introducing breakpoints at a variety of positions. The goals of this study were twofold. First, to investigate the relationship between stability and activity, and to determine the importance of the linker and the location of the breakpoints [22**]. Second, to provide a set of proteins that could be fused to interleukin (IL)-3 with a variety of different relative spatial relationships between the two proteins [23**] (Figure 3). A fusion of wild-type GCSF with wild-type IL-3 has been already engineered to give a 'myelopoietin' that displays enhanced activity, in comparison

with the single agonists, in its ability to promote the growth and maturation of hematopoietic cells. Although all circularly permuted GCSF molecules were destabilized to some extent, they exhibited a range of potencies, from about 10% to greater than 100% of that of wild-type GCSF. In addition, some of the fusions had altered ratios of GCSF : IL-3 activity, suggesting that this redesign approach has the potential to provide a route to an expanded range of therapeutically active myelopoietins.

Chorismate mutase

Chorismate mutase catalyzes the conversion of chorismate to prephenate in the biosynthetic pathway of tyrosine and phenylalanine. The enzyme from *E. coli* is a homodimeric four-helix bundle protein. A long N-terminal helix spans both domains, contributing residues to both active sites of the dimer. MacBeath *et al.* [24**] attempted an adventurous redesign, inserting a loop into the middle of this helix to induce a 180° turn, such that the dimer could be converted into a monomer (Figure 4). The authors took advantage of the power of genetic selection to identify sequences that were capable of inducing this dramatic transformation. Chorismate mutase deficient strains of bacteria are available that can be complemented by libraries of protein variants expressed from exogenously introduced plasmids. On appropriate media, only bacteria expressing an active enzyme survive. Using this strategy, a library containing six randomized positions at the desired

Figure 4



Schematic illustration of the conversion of dimeric chorismate mutase into a monomer. On the left is the wild-type dimer, with the position of an inhibitor bound in the active site shown in ball-and-stick form. The asterisks indicate the positions of the loop insertions. On the right are the two monomeric halves of the molecule after the introduction of a loop connection into helix H1. The asterisks indicate the positions of the new loops. Adapted from [24**].

turn location could be screened to identify the turn sequences that gave rise to a functional protein. Although active, some of the proteins identified by this strategy formed mixtures of monomers and dimers, but one clear success behaved as an authentic monomer. Moreover, the redesigned monomer displayed catalytic activity that is comparable to that of the wild-type enzyme.

This final experiment again illustrates the strength of combining design with selection. The main caveat is that the number of systems for which a selection step can be incorporated is limited.

Conclusions

I have discussed a variety of examples that illustrate how extremely tolerant natural proteins are to large-scale redesign. The loop length and circular permutation variants illustrate the power of the driving force for the correct association of elements of secondary structure, regardless of how they are connected. Not all redesigns work equally well and few of the results described could have been predicted in detail *a priori*. In the work of Otzen and Fersht [17**], for example, a circular permutant was successfully created by cleavage between residues 40 and 41, but cleavage at any other site was severely destabilizing. Similarly, in the work of MacBeath and colleagues [24**], the successful selection of a monomeric variant of chorismate

mutase only succeeded using the enzyme from the thermophile *Methanococcus jannaschii*; initial attempts with a mesophilic version were unsuccessful. Finally, the papers by McWherter and colleagues [22**,23**] illustrate that not only are we acquiring basic redesign rules, but practical implementation of protein redesign is also underway and likely will significantly increase in the future.

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. Nagi AD, Regan L: **An inverse correlation between loop length and protein stability in a four helix bundle protein.** *Fold Des* 1997, **2**:67-75.
2. Nagi AD, Anderson KS, Regan L: **Using loop length variants to dissect the folding pathway of a four-helix-bundle protein.** *J Mol Biol* 1999, **286**:257-265.

The study of the behavior of loop-length variants aids in the dissection of a folding pathway.

3. Ladurner AG, Fersht AR: **Glutamine, alanine and glycine repeats inserted into the loop of a protein have minimal effects on stability and folding rates.** *J Mol Biol* 1997, **273**:330-337.
4. Robinson CR, Sauer RT: **Equilibrium stability and sub-millisecond folding of a designed single chain Arc repressor.** *Biochemistry* 1996, **35**:13878-13884.
5. Robinson CR, Sauer RT: **Optimizing the stability of single-chain proteins by linker length and composition mutagenesis.** *Proc Natl Acad Sci USA* 1998, **95**:5929-5934.

Genetic selection reveals energetically optimal loop sequences and provides guidelines for future redesigns.

6. Goldenberg DP, Creighton TE: **Folding pathway of a circular form of bovine pancreatic trypsin inhibitor.** *J Mol Biol* 1984, **179**:527-545
7. Luger K, Hommel U, Hofsteenge J, Kirschner K: **Correct folding of circularly permuted variants of a beta alpha barrel enzyme *in vivo*.** *Science* 1989, **243**:206-210.
8. Buchwalder A, Szadkowski H, Kirschner K: **A fully active variant of dihydrofolate reductase with a circularly permuted sequence.** *Biochemistry* 1992, **31**:1621-1630.
9. Horlick RA, George HJ, Cooke GM, Tritch RJ, Newton RC, Dwivedi A, Lischwe M, Salemme FR, Weber PC, Horuk R: **Permutins of interleukin 1 beta—a simplified approach for the construction of permuted proteins having new termini.** *Protein Eng* 1992, **5**:427-431.
10. Zhange T, Bertelsen E, Benvegno D, Alber T: **Circular permutation of T4 lysozyme.** *Biochemistry* 1993, **32**:12311-12318.
11. Ay J, Hahn M, Decanniere K, Piotukh K, Borriss R, Heinemann U: **Crystal structures and properties of de novo circularly permuted 1,3-1,4-beta-glucanases.** *Proteins* 1998, **30**:155-167.
12. Puri RK, Hoon DS, Leland P, Snoy P, Rand RW, Pastan I, Kreitman RJ: **Preclinical development of a recombinant toxin containing circularly permuted interleukin 4 and truncated Pseudomonas exotoxin for therapy of malignant astrocytoma.** *Cancer Res* 1996, **56**:5631-5637.
13. Vinais ML, Corbier C, Mulliert G, Branlant C, Branlant G: **Circular permutation within the coenzyme binding domain of the tetrameric glyceraldehyde-3-phosphate dehydrogenase from *Bacillus stearothermophilus*.** *Protein Sci* 1995, **4**:994-1000.
14. Uversky VN, Kutysenko VP, Protasova N, Rogov VV, Vassilenko KS, Gudkov AT: **Circularly permuted dihydrofolate reductase possesses all the properties of the molten globule state, but can resume functional activity by interaction with its ligands.** *Protein Sci* 1996, **5**:1844-1851.
15. Viguera AR, Serrano L: **Loop length, intramolecular diffusion and protein folding.** *Nat Struct Biol* 1997, **4**:939-946.
16. Martinez JC, Viguera AR, Berisio R, Wilmanns M, Mateo PL, Filimonov VV, Serrano L: **Thermodynamic analysis of beta-spectrin SH3 and two of its circular permutants: discerning the reasons for rapid folding in proteins.** *Biochemistry* 1999, **38**:549-559.
A detailed thermodynamic analysis of circular permutants of an SH3 domain.
17. Otzen DE, Fersht AR: **Folding of a circularly permuted chymotrypsin inhibitor 2: retention of the folding nucleus.** *Biochemistry* 1998, **37**:8139-8146.
A completely circular variant, a circular permutant and the wild-type protein were shown to fold by the same pathway and to have closely similar transition states.
18. Wieligmann K, Norledge B, Jaenicke R, Mayr E-M: **Eye lens beta-crystallin: circular permutation does not influence the oligomeric state but enhances the conformational stability.** *J Mol Biol* 1998, **280**:721-729.
19. Hennecke J, Glockshuber R: **Conversion of a catalytic into a structural disulfide bond by circular permutation.** *Biochemistry* 1998, **37**:17590-17597.
20. Chu V, Freitag S, Le Trong I, Stenkamp RE, Stayton PS: **Thermodynamic and structural consequences of flexible loop deletion by circular permutation of the streptavidin-biotin system.** *Protein Sci* 1998, **7**:848-859.
21. Iwakura M, Nakamura T: **Effects of the length of a glycine linker connecting the N- and C-termini of a circularly permuted dihydrofolate reductase.** *Protein Eng* 1998, **8**:707-713.
22. Feng Y, Minnerly JC, Zurfluh LL, Joy WD, Hood WF, Abegg AL, Grabbe ES, Shieh J-J, Thurman TL, McKearn JP, McWherter CA: **Circular permutation of granulocyte colony-stimulating factor.** *Biochemistry* 1999, **38**:4553-4563.
See annotation to [23**].
23. McWherter CA, Feng Y, Zurfluh L, Klein BK, Baganoff MP, Polazzi JO, Hood WF, Paik K, Abegg AL, Grabbe ES *et al.*: **Circular permutation of the granulocyte colony-stimulating factor receptor agonist domain of myelopoietin.** *Biochemistry* 1999, **38**:4564-4571.
This paper (together with [22**]) describes circular permutants of GCSF, a protein that is currently administered to patients undergoing chemotherapy. Several are successful and, when fused to IL-3, create hybrid molecules with an expanded range of biological activities.
24. MacBeath G, Kast P, Hilvert D: **Redesigning enzyme topology by directed evolution.** *Science* 1998, **279**:1958-1961.
An illustration of the power of genetic selection. Enzyme topology was successfully redesigned by identifying the 1 in 2000 loop sequences that could function as desired.