Request # 18592867

NOV 17, 2005

Ariel To: 64.40.17.85/140.163.217.217

Memorial Sloan-Kettering Cancer Center
Medical Library Nathan Cummings Center (METRO #146)
1275 York Avenue
New York, NY 10021

DOCLINE: Journal Copy Epayment

Title:	Trends in biotechnology.
Title Abbrev:	Trends Biotechnol
Citation:	1994 May;12(5):163-7
Article:	Design of protein structures: helix bundles and be
Author:	Sander C;
NLM Unique ID:	8310903 Verify: PubMed
PubMed UI:	7764898
ISSN:	0167-7799 (Print)
Publisher:	Elsevier Science Publishers, Barking
Copyright:	Copyright Compliance Law
Authorization:	A Artale
Need By:	NOV 21, 2005
Maximum Cost:	**$25.00**
Patron Name:	Gangi-Dino, Rita
Referral Reason:	Not owned (title)
Library Groups:	BQSIMB,EFTS,RESOURCE,METRO
Phone:	1.212.639-7441
Fax:	1.646.422-2316
Email:	ill@mskcc.org
Comments:	**Please, we prefer as PDF or Ariel. Thank you! EFTS, METRO, BQSI.**
Routing Reason:	Routed to PAUPIT in Serial Routing - cell 1
Received:	Nov 18, 2005 (08:01 AM EST)
Lender:	University of Pittsburgh/ Pittsburgh/ PA USA (PAUPIT)

Bill to: NYUMSK

Memorial Sloan-Kettering Cancer Center
Medical Library Nathan Cummings Center
1275 York Avenue
New York, NY 10021

28 Piccirilli, J. A., Krauch, T., Moroney, S. E. and Benner, S. A. (1990) *Nature* 343, 33 37
29 Bain, J. D., Chamberlin, A. R., Switzer, C. Y. and Benner, S. A. (1992) *Nature* 356, 537–539
30 Piccirilli, J. A., Moroney, S. E. and Benner, S. A. (1991) *Biochemistry* 30, 10350–10356
31 Benner, S. A. and Gerloff, D. (1991) *Adv. Enzyme Regul.* 31, 121–181
32 Benner, S. A., Cohen, M. A. and Gerloff, D. L. (1993) *J. Mol. Biol.* 229, 295–305
33 Gerloff, D. L., Jenny, T. F., Knecht, L. J., Gonnet, G. H. and Benner, S. A. (1993) *FEBS Lett.* 318, 118–124
34 Gerloff, D. L., Jenny, T. F., Knecht, L. J. and Benner, S. A. (1993) *Biochem. Biophys. Res. Commun.* 194, 560–565
35 Bazan, J. F. (1990) *Proc. Natl Acad. Sci. USA* 87, 6934–6938
36 Russell, R. B., Breed, J. and Barton, G. J. (1992) *FEBS Lett.* 304, 15–20
37 Knighton, D. R. *et al.* (1991) *Science* 253, 407–414
38 Thornton, J. M., Flores, T. P., Jones, D. T. and Swindells, M. B. (1991) *Nature* 354, 105–106
39 Ciesla, D. J., Gilbert, D. E. and Feigon, J. (1991) *J. Am. Chem. Soc.* 113, 3957–3961
40 Osterhout, J. J., Jr *et al.* (1992) *J. Am. Chem. Soc.* 114, 331–337
41 Zhou, N. E., Zhu, B. Y., Sykes, B. D. and Hodges, R. S. (1992) *J. Am. Chem. Soc.* 114, 4320–4326
42 Klaus, W. and Moser, R. (1992) *Prot. Eng.* 5, 333–341
43 Lovejoy, B., Choe, S., Cascio, D., McRorie, D. K., Degrado, W. F. and Eisenberg, D. (1993) *Science* 259, 1288–1293
44 Gutte, B., Daeumigen, M. and Wittschieber, E. (1979) *Nature* 281, 650–655

Design of protein structures: helix bundles and beyond

Chris Sander

The design of proteins or peptides with novel functions can be achieved either by modifying existing molecules or by inventing entirely new structures and sequences that are unknown in nature. Combinatorial-design strategies have led to the first *de novo* proteins, but these still lack some of the desired attributes. The most promising practical strategies for developing proteins with useful biological or chemical function combine theoretical design with experimental screening or selection systems.

Nature has evolved highly intricate and useful proteins over many millions of years, gradually optimizing protein function in response to selective pressure. When will humans be able to sidestep evolution and design novel proteins with desired structures and functions? Will the new proteins be redesigns of natural proteins, or *de novo* inventions with sequences not found in nature? The answers are not yet available, but the first steps towards them have been taken.

Re-engineering work has proven that the protein engineer has considerable latitude in modifying existing frameworks, not just in replacing surface- or active-site residues, but also in rearranging loop regions and replacing residues in the protein's interior. The rules of *de novo* design are already partially understood for simple structures, such as four-helix bundles, and the first stable proteins have been designed *de novo*. Simple functions, such as metal-binding sites, have been introduced into some designed proteins, while existing proteins have been functionally optimized by more intricate modification of the protein using, for instance, *in vitro* selection systems.

To illustrate the current capabilities and difficulties of structurally oriented protein design, this article discusses what has been achieved with a particularly simple class of protein fold – bundles of four α helices. These achievements include redesigning the topology of loop connections, redesigning the packing of the hydrophobic core, as well as creating *de novo* designs, i.e. proteins whose amino acid sequence is newly invented and not seen in nature.

A simple architecture: α-helix bundles

The architecture of the most commonly occurring type of four-helix bundle is particularly simple: each of the four helices is oriented antiparallel to its two nearest neighbors and parallel to its diagonal, more distant, neighbor. Residues on the interior helix-faces form a hydrophobic core. Examples of such structures

C. Sander is at the European Molecular Biology Laboratory, D-69012, Heidelberg Germany.

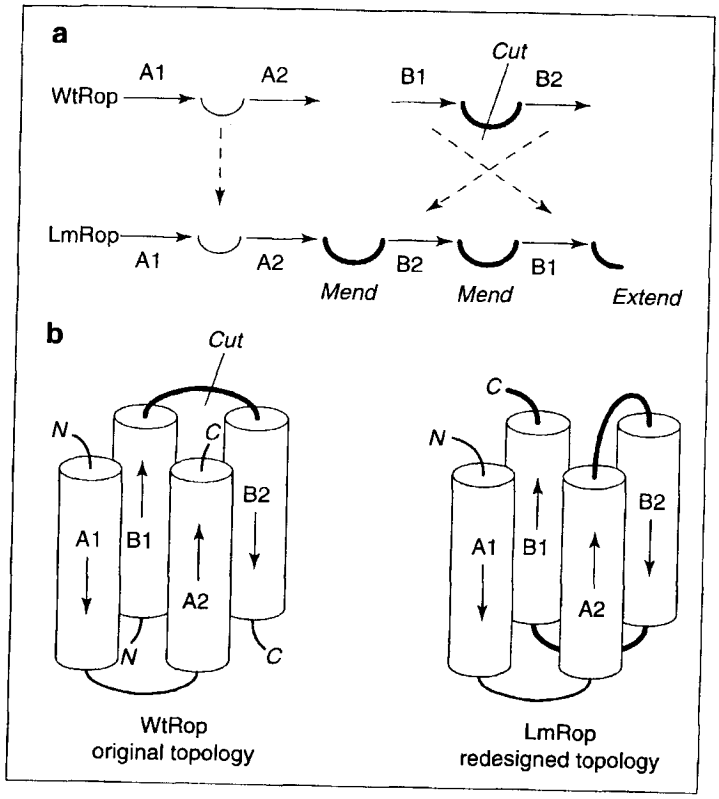

Figure 1

Design steps from wild-type Rop (Wt Rop) dimer to LmRop monomer by cut-and-link protein engineering at the gene level (**a**) and in three dimensions (**b**). Wild-type Rop has two identical subunits, A and B, and a total of four helices, A1, A2 and B1, B2 (cylinders/arrows). To achieve the desired monomeric bundle with left-handed topology and starting with helix A1, the loop between helices B1 and B2 was deleted; the order (in the chain sense) of these helices was switched from B1, B2 to B2, B1; helices A2 and B2 were connected by a new loop, and helices B2, B1 were connected by another loop; and helix B1 was completed by addition of a few residues. Loops involved in these changes are highlighted (thick lines).

include Rop protein[1] (a dimer of two helical hairpins), myohemerythrin and cytochrome c'. The amino acid sequences of these bundles appear to follow simple rules: the helical propensity is high (for example, for Glu or Leu); and there is a characteristic alternation of residue polarity with a period of 3.5 residues – polar residues populate the bundle surface, while non-polar residues pack into a hydrophobic interior. However, such amphiphilic helices can associate in a number of ways, forming dimers (coiled coils or zippers), trimers, tetramers or higher aggregates. There is, therefore, additional information in the amino acid sequence that specifies the precise relative packing of the helices.

Redesign of loop topology on a constant protein core

Given the simple architecture of the bundles, it is plausible that the formation of the unique folded native conformation is dominated by specific interaction of side chains in the hydrophobic interior, or core. According to this 'core hypothesis', the precise sequence in the loop regions and on the protein surface is of secondary importance, provided that it is consistent with loop formation and surface solvation. A test of this hypothesis was performed in a redesign experiment on the natural four-helix bundle Rop (Refs 1, 2; H. Blöcker, S. Emery and M. Sagermann, unpublished). The idea was to leave the core residues untouched, while making major alterations in the way in which the loops connect these helices. In one construct (Fig. 1), the wild-type dimer (two chains of two helices each) was turned into a monomer (one chain with four helices): the two halves of the original four-helix bundle were covalently connected and the topology of the loops connecting the four helices was rearranged. This involved deletion of one loop, as well as the addition of two new loops and a C-terminal peptide. The new, single-chain protein had 13 new residues out of a total of 120. Interhelical distance constraints, determined from 2-D NMR spectroscopy of the purified LmRop monomer and unit-cell dimensions of crystals, proved that the packing of the four helices was the same as in the wild-type protein, despite the drastic loop rearrangements.

While this result is consistent with the idea that side-chain packing in the interior is the primary determinant of the native structure, many more confirmatory experiments are needed on the road from hypothesis to generally applicable design principle. The interesting implication of protein engineering would be that loop connections in helical bundles, and perhaps in globular proteins in general, can be re-engineered, not only in sequence, but also in topology. In natural evolution, a progression leading to topologically rearranged loops requires an unlikely succession of rearrangement events. Protein engineering, however, is capable of evolutionary jumps that transcend natural evolution. The recognition that topological rewiring can be tolerated, without affecting the core structure, opens up new degrees of freedom in the design of useful proteins.

Redesign of protein cores

Structure-preserving changes in the protein core

If the packing of specific residues in the protein core is a key determinant of the native fold, then only a very small fraction of all possible core sequences should be compatible with maintenance of the fold. To test the latitude of sequence variation and the protein engineer's ability to design a uniquely folded protein core, a number of experiments have been performed. Typically, the sequence of a natural protein that has a known crystal structure is carefully altered at certain interior positions, so as to leave the specific packing intact. Alternatively, the sequence is randomized at a specific position and, using a screening or selection system, the sequences compatible with a known structure are recorded, as done for the lambda repressor[3]. The results of such experiments elegantly confirm that only selected types of non-polar amino acids are accepted as structure-preserving substitutions in the interior of proteins – a fact already evident from observation of natural variation in protein families, using multiple-sequence alignments.

Structure-altering changes in the protein core

The precise rules of such core substitutions are still somewhat elusive. A case in point is the variation of hydrophobic positions at the helix–helix interface in the GCN4 coiled coil, or leucine zipper, a 31-residue fragment known to dimerize in parallel fashion, with sequence 2-MKQLEDK VEELLSK NYHLENE VARLKKL VGE-32. Replacing residues in the 'd' (doubly underlined) position by isoleucine (I), while changing those in the 'a' (underlined) position to leucine (L), led to a radical change in helix–helix assembly[4]. The crystal structure of GCN4-LI at 2.1 Å resolution shows a regular bundle of four rather than two helices. However, contrary to the classical four-helical bundles, which have antiparallel nearest-neighbor association, the GCN4-LI bundle is fully parallel, with a continuous central channel of radius 1.0–1.3 Å. The tetramer is remarkably stable, even in boiling water. This example constitutes the successful (with hindsight) design of an apparently unique packed-bundle structure, confirms the hypothesis that packing is largely determined by the sequence in the non-polar core, and indicates that the precise nature of the interior side chains, not just their apolar character, is important. The design is fortuitous, however, in that the researchers did not anticipate that the conservative residue replacement would lead to an altered helix–helix association.

De novo design of helix bundles

An extreme goal of protein design is the ability to design amino acid sequences from scratch and to achieve a preconceived folded structure. About ten different experimental attempts at validating *de novo* designed sequences have been carried out, and are reviewed elsewhere[5–9].

Compact and stable α_4 lacks specificity of fold

A very interesting systematic series of *de novo* design experiments on four-helix bundles started with simple, partially repetitive sequences. The design was modular: four identical helices [sequence: GELEEL-LKKLKELLKG (helical portion underlined), similar to an earlier tetrameric design[10,11], were linked by three identical loops (loop sequence: PRR) (Fig. 2). The design was successful in that this monomeric four-helix bundle, α_4, folded up into a compact globular structure with a highly cooperative unfolding transition and an unusually high stability against denaturation with guanidine hydrochloride (DC = −22 kcal mol^{-1} for the folded–unfolded transition in water)[13,14]. Determination of the crystal or NMR structure of α_4, however, was not achieved. The probable reason for this is that this designed protein has properties of the folded state that are atypical of those of natural globular proteins: although the protein is compact and globular and its helices appear to be stably formed, its apolar side chains are rather mobile and not uniquely packed as in natural proteins. This mobility gives the protein added stability (entropic advantage), but less specificity of folding. In addition,

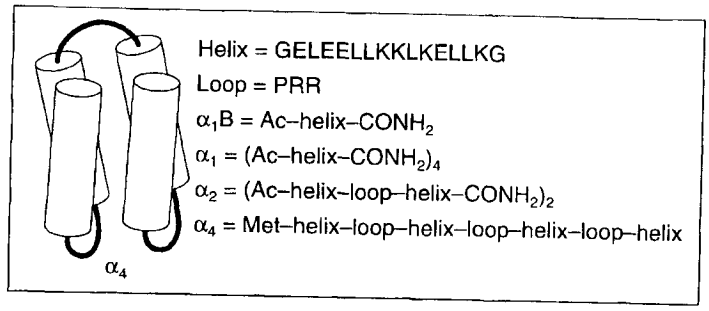

Figure 2

Cartoon structure and amino acid sequence of a family of four-helix bundles, designed by W. F. DeGrado and colleagues[12]. In this modular approach, the same 16-residue peptide (α_1B) was used to form an α-helical tetramer (α_1), a dimer of two helix–loop–helix pairs (α_2), and a monomer of four α_1B helices (α_4) linked by three ProArgArg loops. (Redrawn, with permission, from Ref. 13.)

the protein probably exists in a mixture of two alternative topographies – a right-handed and a left-handed arrangement.

Introduction of specific interactions in the α_4 designed bundle gave it properties closer to those of native natural proteins[12]. This was achieved by replacing the somewhat indiscriminate interactions of Leu residues with more geometrically restrictive Zn^{2+}-binding sites, formed by six His residues (helical sequences are GELEEL*H*KKL*H*ELLKG for helices 1 and 4, and GELEEL*H*KKLKELLKG for helices 2 and 3). This form binds two Zn^{2+} ions, one between helices 1 and 2 and one between helices 3 and 4. As a result, the NMR chemical shifts (e.g. those of the Leu methyl groups) are more dispersed, indicating that metal binding has an organizing effect on the entire protein core. Yet, even this four-helix bundle has some characteristics of a molten globule, i.e. it binds a hydrophobic dye and has amide proton–deuteron exchange rates that are rather rapid – properties that natural proteins only have in a partially non-native state.

The helix bundle α_4 and its cousins represent successful designs of compact and stable proteins[12]. However, these proteins are not folded into a specific, native-like conformation. The designers' task, therefore, is to introduce additional, position-specific, interactions in order to achieve a well-packed core that can be proven by structure determination at atomic resolution by NMR or crystallography.

Four-helical tetramer PD_1 has specific fold

Serendipity recently led to the *de novo* design of a 24-residue peptide (PD_1) that forms a tetramer with, apparently, specific native-like packing[15]. The sequence of the peptide, designed to be a peptidic detergent for solubilization of membrane proteins (and not specifically intended to form a helical bundle!) is EELLKQALQQAQQLLQQAQELAKK. Note the ends of the asymmetric sequence (E/K) that may provide the position-specific interactions required for a unique fold. The polar surface consists mainly of Gln residues, while the non-polar face was designed to be flat (Ala) in the center, and potentially interdigitating

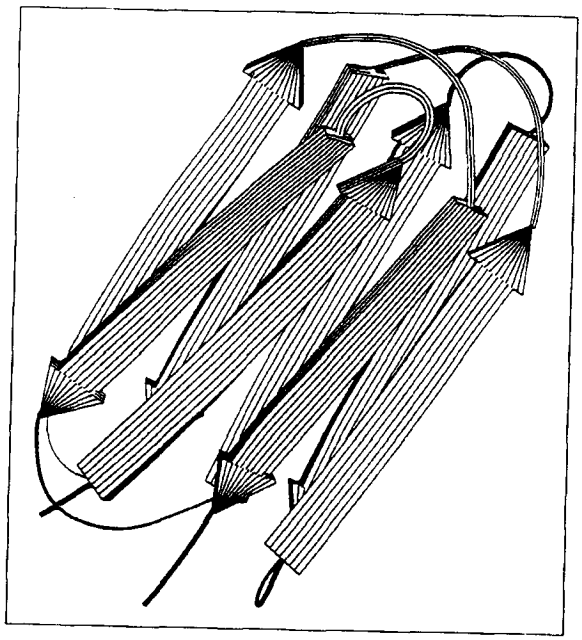

Figure 3
Structure cartoon of Shpilka, an antiparallel sandwich of two antiparallel four-stranded β-sheets. The protein was designed[22] as a scaffold for testing various topologically different loop connections between strands. Improvements to this sequence have been made by A. V. Finkelstein using physical considerations and protein expression is in progress (D. A. Dolgikh, pers. commun.). Drawn using the program 'Ribbons' by J. Priestle.

(Leu) on the sides. Evidence for a specific conformation of the tetramer comes from a 2.5 Å crystal structure (albeit with more-than-normal thermal and positional disorder) that shows a classical antiparallel helical bundle with a −20° crossing angle (except that the helices are fairly straight, rather than coiled about one another). Experience tells us that the four helices could easily be connected into a single four-helix chain that, if it also crystallizes, would be the first native-like *de novo* designed protein. Unfortunately, the assembly has not yet been characterized in terms of its stability and unfolding properties. But the fact that the accidental design of PD_1 has, for the first time, led to a well-defined antiparallel four-helix bundle is encouraging.

De novo design of all-β and α-β structures

The *de novo* design of proteins that have β-strands in their fold appears to be more difficult than that of pure helical bundles. Among the attempts at producing such proteins are: eightfold $(\beta\alpha)_8$ barrels of the triose phosphate isomerase (TIM) type[16–18]; small mixed α/β structures[19,20]; as well as sandwiches of beta sheets (Refs 21,22; Fig. 3). Some of these synthetic or recombinant proteins appear to have cooperative unfolding transitions, but none of them has stability approaching that of the designed helical bundles, nor have crystals or well-dispersed NMR spectra been obtained[9]. It is likely that none of these proteins, in their current state of design, are likely to have unique, stably packed structures. The designers have a tough job ahead.

Lessons learned

All of the experiments reviewed here leave us with the conclusion that the crucial next step is learning how to design specific intraprotein interactions that achieve predictable and unique modes of packing. Meanwhile, as a sort of spin-off effect, a number of principles of protein folding have emerged from design experiments, as well as from observations on natural protein families, including:

- hydrophobic interactions are sufficient to drive polypeptide chain folding nearly to completion, but specific interactions are required for a unique structure [23,25];
- there appears to be a trade-off between increased stability against unfolding and increased specificity of folding – both are difficult to achieve with the same sequence[4,12];
- small changes in core residues can (but usually do not) lead to alternative modes of packing and altered assembly – the shape of side chains, not just their polarity, can be an important determinant of the fold[4]; and
- loops can be configured rather freely on a constant scaffold, not just in sequence, but also in connectivity[2].

The power of large numbers

Combining design and screening

While protein designers struggle to learn packing rules, what is the most practical way of achieving a particular design goal? Probably a combination of knowledge and ignorance, of targeted design and sequence randomization. The key idea is to use the parallel-processing power of screening or selection systems to sift through a large number of candidate sequences[24]. A recent ingenious example is the successful design of four-helix bundles by screening a library of sequences constrained to fit a binary pattern of polar and non-polar residues that are typical of helical bundles (hydrophobic interior, hydrophilic surface): Hecht and colleagues[25] picked 108 clones from among a large number of sequences with degenerate codons, of which 48 had the designed amphiphilic sequence pattern, 29 were soluble and resisted proteolytic degradation, at least three were compact and monomeric, and two had cooperative unfolding transitions and stabilities similar to those of natural proteins. Whether or not these bundles have high specificity of internal packing, these experiments demonstrate that the evolutionary path from quasi-random sequences to folded proteins can be followed in engineering practice.

Combining design and in vitro *evolution*

Additional power can be harnessed when the design of partially degenerate sequences is combined with selective biological amplification. The basic idea is

simple: use our knowledge of protein structures to select a few residue positions (out of perhaps hundreds), in which mutations may have a functionally beneficial effect. Such an example is provided by the residues lining the active site of an enzyme (perhaps ten out of three hundred). These positions are then randomized in a vector carrying the original sequence, by using suitably synthesized oligonucleotides. This vector is then introduced into a rapidly growing cell line placed under selective pressure (such as from a toxic substance that must be broken down for better survival). If everything works well, cells with sequences that express a desired function will have a growth advantage, and will be represented in larger numbers in the cell population after some period of growth. Sequencing the selected genes leads to the desired result: one (or several) protein sequences with an improved phenotype.

To carry out this plan, a number of technical problems have to be solved[3]. The number of residue positions chosen for randomization must be kept reasonably small, in order to achieve good coverage of possible sequence combinations. The choice of residue positions requires considerable insight into protein architecture and functional sequence determinants. The key problem is the design of a suitable selection system. It must be chosen such that the desired phenotype implies a noticeable growth advantage[26]. This topic is addressed in detail elsewhere in this special issue[27].

Why bother?

Depending on one's point of view, structural protein design aims at a better understanding of the principles of protein folding, at the production of novel proteins with desired binding properties and catalytic activities, at the nanometer scale engineering of self-assembling microscopic structures, and the like. There are perhaps better approaches to achieving any of these goals. For example, improved enzyme functions might be evolved more directly from natural enzymes by applying appropriate selection pressure in cell culture under mutagenic conditions. The principles of protein folding can perhaps best be extracted from the tens of thousands of known sequences and hundreds of known three-dimensional structures of natural proteins. Nanometer-scale engineering using protein material should perhaps take natural membrane proteins as its starting point. So why bother with protein design from first principles?

The basic answer comes from the human desire to understand the principles of natural processes, not only in a passive, but also in an active way, to develop useful devices not only by modification, but also by construction. The better we can plan novel experiments and successfully predict their outcome, the closer we come to comprehensive understanding. In addition, the ability to engineer new proteins from scratch would open a wide range of possible applications, transcending the frozen accidents of natural evolution. Protein design, *de novo* and in combination with natural selection, is likely to become one of the main pillars of biomolecular engineering. It only remains to be developed vigorously and applied wisely.

References

1 Banner, D. W., Kokkinidis, M. and Tsernoglou, D. (1987) *J. Mol. Biol.* 196, 657
2 Sander, C. (1990) *Biochem. Soc. Symp.* 57, 25–33
3 Lim, W. A. and Sauer, R. T. (1989) *Nature* 339, 31–36
4 Harbury, P. B., Zhang, T., Kim, P. S. and Alber, T. (1993) *Science* 262, 1401–1407
5 Richardson, J. and Richardson, D. (1989) *Trends Biochem. Sci.* 14, 304–309
6 DeGrado, W. F., Raleigh, D. P. and Handel, T. (1991) *Curr. Biol.* 1, 984–993
7 Sander, C. (1991) *Curr. Opin. Struct. Biol.* 1, 630–638
8 Lovejoy, B. *et al.* (1993) *Science* 259, 1288
9 Sander, C. (1994) in *Concepts of Protein Engineering and Design* (Wrede, P. and Schneider, G., eds), Ch. 5, Walter de Gruyter
10 Eisenberg, D., Wilcox, W., Eshita, S., Pryciak, P., Ho, S. and DeGrado, W. (1986) *Proteins* 1, 16–22
11 Osterhout, J. J. *et al.* (1992) *J. Am. Chem. Soc.* 114, 331–337
12 Handel, T. M., Williams, S. A. and DeGrado, W. F. (1993) *Science* 261, 879–885
13 Regan, L. and DeGrado, W. (1988) *Science* 241, 976–978
14 DeGrado, W., Wasserman, Z. and Lear, J. (1989) *Science* 243, 622–628
15 Schafmeister, C. E., Miercke, L. J. W. and Stroud, R. M. (1993) *Science* 262, 734–738
16 Goraj, K., Renard, A. and Martial, J. (1990) *Prot. Eng.* 3, 259–266
17 Richardson, J. S., Kneller, D., Osguthorpe, D. and Scharf, M. (1987) in *Protein Design Exercises* (Sander, C., ed.), *EMBL Biocomputing Technical Document* 1, 21–39, EMBL
18 Tanaka, T. *et al.* (1990) *J. Cell. Biochem.* 14C, 233–233
19 Fedorov, A. N. *et al.* (1992) *J. Mol. Biol.* 225, 927–931
20 Sander, C. *et al.* (1992) *Proteins* 12, 105–110
21 McClain, R. D., Yan, Y., Williams, R. W., Donlan, M. E. and Erickson, B. W. (1992) in *Protein Engineering of Betabellin 12* (Smith, J. A. and Rivier, J. E., eds), pp. 364–365, ESCOM
22 Finkelstein, A. V., Lockhart, A., Merkl, R. and Perry, L. J. (1991) in *Protein Design on Computers* (Sander, C. and Vriend, G., eds), *EMBL Biocomputing Technical Document* 6, 133–148, EMBL
23 Hecht, M., Richardson, J., Richardson, D. and Ogden, R. (1990) *Science* 249, 884–891
24 Hu, J. C., O'Shea, E. K., Kim, P. S. and Sauer, R. T. (1990) *Science* 250, 1400–1403
25 Kamtekar, S., Schiffer, J. M., Xiong, H. Y., Babik, J. M. and Hecht, M. H. (1993) *Science* 262, 1680–1685
26 Wells, J. A. and Lowman, H. B. (1992) *Curr. Opin. Biotechnol.* 3, 355–362
27 Clackson, T. and Wells, J. (1994) *Trends Biotechnol.* 124, 173–184