ANTHONY ARTALE
MED LIB NATHAN CUMMINGS CTR (S-46)
MEMORIAL SLOAN KETTERING CANCER CTR
1275 YORK AVENUE
NEW YORK, NY 10021
UNITED STATES

Client Number: DDS36699
Title: **CURRENT OPINION IN STRUCTURAL BIOLOGY.**
DB Ref. No.: IRN10485028
ISSN: ISSN0959440X
Vol./Issue: 1
Date: 1991
Pages: 630-638
Article Title: DE NOVO DESIGN OF PROTEINS
Report Number: IRN10485028
Publisher: CURRENT BIOLOGY,

De novo design of proteins

Chris Sander

European Molecular Biology Laboratory, Heidelberg, Germany

Can we design and produce novel proteins with desired structures and functions using amino acid sequences not found in nature? The design principles, abstracted from observation of natural proteins, are simplest for α-helical bundles, more complicated for β-sheet or mixed structures, and most subtle for the design of catalytic function. The simplest designs have been the most successful so far, but we have yet to encounter an experimentally determined three-dimensional structure for any *de novo* designed protein that folded as expected. The field is still in its infancy, but its ultimate potential for applications is enormous.

Current Opinion in Structural Biology 1991, **1**:630–637

The protein design cycle

Protein engineering often starts from a protein of known sequence and structure and aims to improve its properties, e.g. increase its stability or change its specificity. In *de novo* design, the rules of the game are different: the desired structure is defined, then an amino acid sequence invented that would fold up as desired. To simplify the task, the desired structure often corresponds to one of the well known structural motifs observed in nature, be it a coiled coil of α-helices, an α-helical bundle (Fig. 1), a β-sheet sandwich (Fig. 2), a regular αβαβ structure or the like (where α refers to α-helix and β to β-strand). By definition, *de novo* designed sequences are not homologous to any known natural sequence, but they are analogous to natural sequences in that, upon successful folding, they do adopt the same types of structural motif.

What are the steps in designing a protein *de novo*? There is no simple answer, as groups active in the field have developed a variety of approaches [1–3]. Let us look at one possible approach taught at a recent EMBO protein design workshop [4•]. Suppose the structural motif one has selected for the design is a globular protein consisting of a single-chain β-sheet sandwich (Fig. 2). First, one makes a rough sketch of the structure: how many strands and which approximate strand length are to be used? Are the β-sheets to cross at approximately right angles or to run essentially parallel to each other? How strongly are the sheets twisted? Then, using molecular-modelling software, one builds a backbone model, best done by picking appropriate fragments from the database of known three-dimensional structures so as not to deviate too strongly from well tested structural units.

In the crucial next step, one chooses a sequence to fit the desired structure. This step is not (yet) automated but is performed interactively using physical, statistical and intuitive criteria. For choosing individual residues, statistical tables of residue preferences are available: preferences for specific positions in secondary-structure segments, e.g. near the beginning and end of strands; preferences for turns and loops; preferences for interfaces, e.g. between two β-strands, between two β-sheets, between a β-sheet and an α-helix, between a β-sheet and solvent; and so on. Proline has its side chain covalently linked to the backbone nitrogen, so it can be used to lock in local backbone conformation or to disrupt secondary structure by interrupting hydrogen-bond formation. Glycine lacks a side chain and can be used where unusual backbone flexibility is needed. A pair of cysteines can be employed to engineer covalent disulfide links, and tryptophan can be chosen as a spectroscopic probe of folding. For choosing pairs and clusters of residues, criteria include local complementarity of shape, size and polarity, optimization of packing, hydrogen bonding and charge–charge interactions. In this part of the design process, frequent comparison of the current model with similar three-dimensional constellations in the database of known structures is extremely useful [5•]. In order to ensure good solvation properties, the overall amino acid composition of the apolar cores and solvent-exposed surfaces of globular proteins can be used as a guide.

Simply adapting a sequence to a desired static scaffold is, however, not sufficient. How does one guarantee that the sequence actually allows a sufficient number of folding pathways from an unfolded to the final fold; that it is stable with respect to conformational fluctuations; that it will not as easily or preferentially fold into another globular shape? The current theories of protein folding do not yet provide comprehensive answers to these questions. However, empirical rules based on physical approximations or statistical investigations are being developed for some of these aspects [6,7]. For example, in a helical bundle, the number of residues in a loop can be chosen so as to disfavor formation of an uninterrupted long helix instead of the desired helix-loop-helix motif [1]. Regan and DeGrado *et al.* [8] made use of repulsive ion-pair interactions to force loop formation and prevent incorrect

ISSN 0959-440X

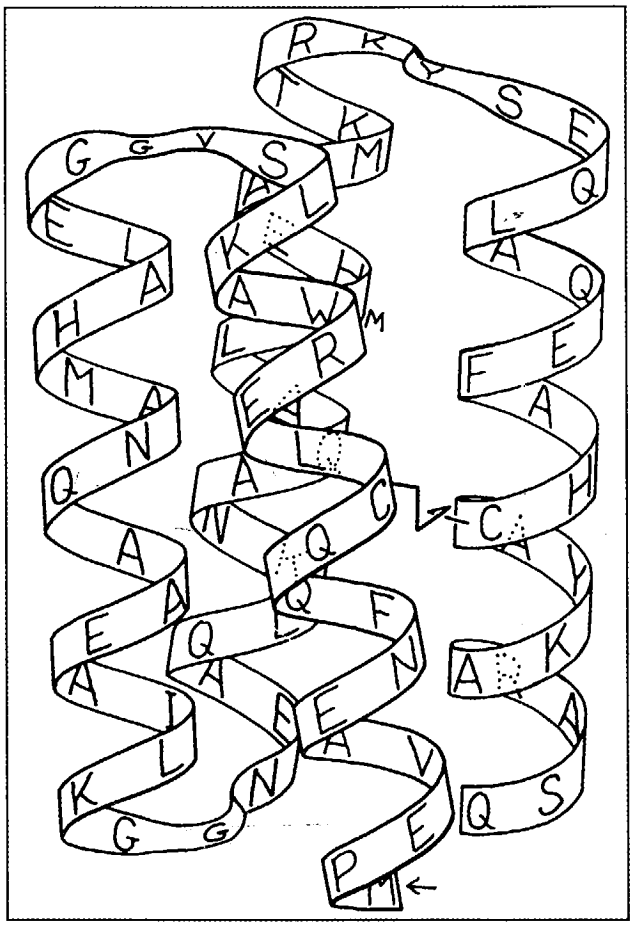

Fig. 1. Structure cartoon and amino acid sequence of Felix, a 79-residue antiparallel four-helix left-handed bundle [19] with non-symmetrical sequence. Experiments indicate that the protein has approximately the designed structure, but its stability is marginal and cooperativity of unfolding is too low for a uniquely folded structure. Published with permission [21••].

helix association, by insertion of a Pro-Arg-Arg sequence into a potential loop domain.

After an amino acid sequence has been chosen, one fully optimizes the three-dimensional model by Monte Carlo or molecular dynamics exploration, followed by energy minimization. In my experience, Monte Carlo methods are vastly preferable for this purpose [9•]. Next, one evaluates the overall quality of the designed protein according to general criteria that were perhaps not used in the construction process, e.g. estimates of solvation energy [10], transfer energy [11], packing rules (QPACK [12•]; PACANA, J Moult, personal communication) or distribution of contacts around residue types (G Vriend and C Sander, unpublished data). Such analysis will invariably lead to suggestions for point mutations in the model. Before proceeding to physical protein production in the laboratory, it is advisable to subject the chosen sequence and three-dimensional model to several or many cycles of sequence change, conformational optimization and analysis using the available computer tools.

With an optimized sequence in hand, the protein or corresponding gene can be synthesized. For *in vivo* expression of synthetic genes, the main problem in some cases has been recovering sufficient amounts of overexpressed product and purifying this into a soluble fraction. *In vitro* expression systems sidestep some of these problems (O Ptitsyn, personal communication) but suffer from low yields. In the final stage, spectroscopic methods, such as circular dichroism, tryptophan fluorescence, and one-dimensional NMR are widely used in the first tests of structure. Alternatively, in the case of a designed function that is strongly dependent on a specific three-dimensional structure, a functional assay is used. The free energy of unfolding can be estimated by following a spectral signal as a function of the concentration of denaturant. A full determination of correct three-dimensional structure requires X-ray crystallography or higher-dimensional (at least two-dimensional) NMR spectroscopy but, so far, crystal-structure determination has been completed for only one *de novo* designed protein [13••]. In general, several or many cycles of design, verification, and redesign are required to evolve a protein that is correctly folded.

De novo protein design as an experimental science became possible because of two technical advances. The first was the peptide synthesis of increasingly longer chain lengths at lower cost. Correspondingly, the first pioneers of protein design were peptide chemists. The second advance, some years later, was the perfection of gene synthesis and expression systems. Over the past year, making elegant use of these technical advances, several *de novo* design projects have been pursued actively, some with a more structural, others with a more functional orientation. These are reviewed in the following sections.

Coiled coils

A rope-like structure of two α-helices, the 'coiled coil', has a simple architecture and fairly simple sequence rules. It was first observed in tropomyosin [14]. When repeats of seven residues are laid out in a flat 'helical net' projection, there should be adjacent vertical strips of hydrophobic residues and of polar or charged residues, single-residue preferences for the helical state should be favorable [15•], and interacting residues on opposite helices should be complementary in size and shape. Indeed, a number of designed sequences of this type have been shown experimentally to assemble into coiled coils; early examples included sequences containing heptapeptides such as KLEALEG (using the amino acid one-letter code) [16].

Recently, Erickson and co-workers [17•] have used similar ideas to design coiled-coil structures as vehicles for functional groups that would interact with cell-surface receptors. For example, they synthesized 39-residue peptides that contained the repeating sequence $(RIEAIEA)_4$, a spacer (GYG) and a cell-surface ligand (GRGDSP) and showed, by a number of biophysical techniques, that the peptides formed stable coiled pairs of α-helices, as de-

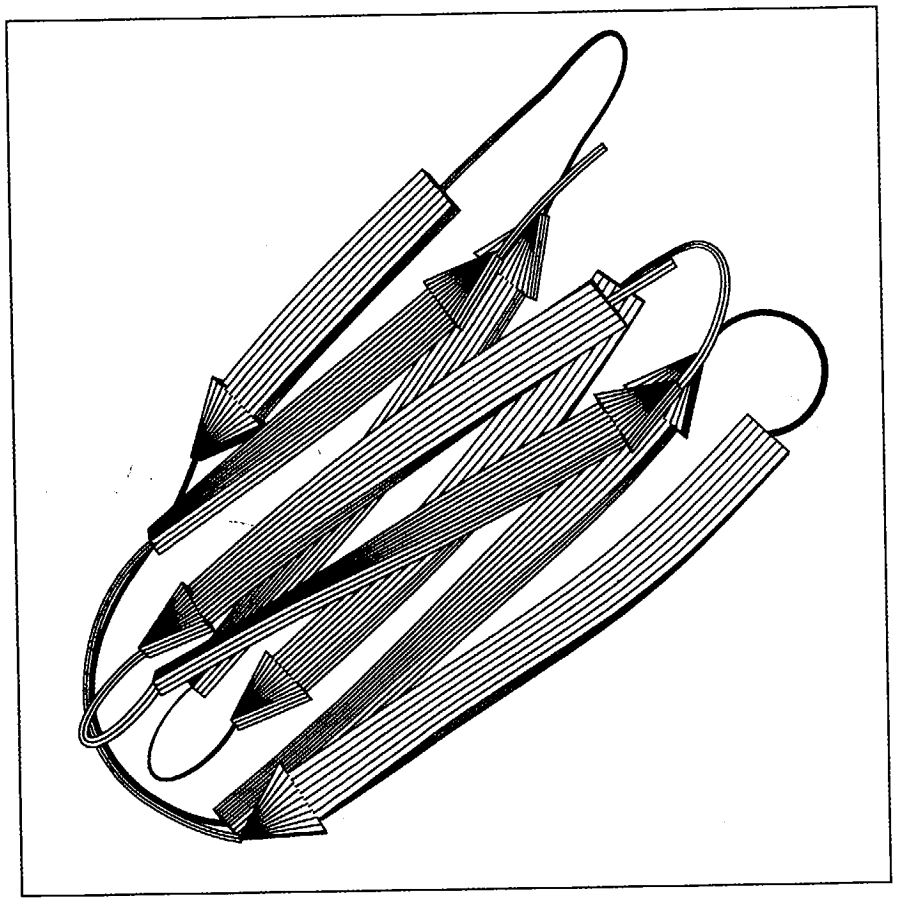

Fig. 2. Structure cartoon of Shpilka, an antiparallel sandwich of two antiparallel four-stranded β-sheets. The protein was designed [38] as a scaffold for testing various topologically different loop connections between strands. Experimental work is planned. Improvements to this sequence have been made recently by AV Finkelstein (personal communication). Drawing produced by R Schneider and U Goebel using the program RIBBONS by J Priestle.

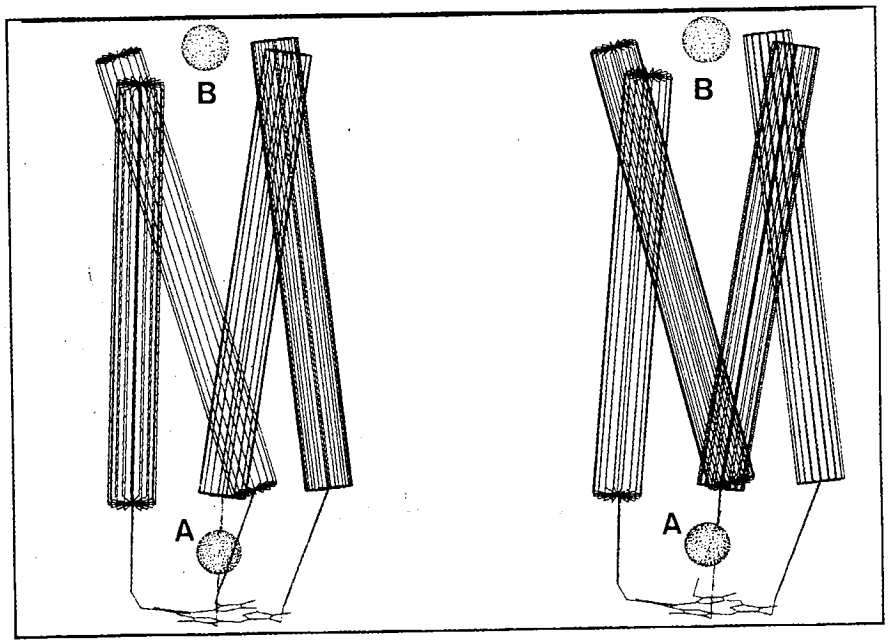

Fig. 3. Stereoscopic schematic model of a hybrid construction where protein-like α-helices in a parallel bundle are covalently linked to a template (bottom) by chemical synthesis [27]. In two recent attempts at designing enzymes *de novo* [28,29•,30••] a template-based approach was used, with the location of the active site (dotted sphere) either at the bottom (A) [28,29•] or near the top (B) [30••] of the bundle. The template drawn here is a porphyrin ring, as used in [28,29•]. Drawn by G Vriend, using WHATIF [5•], following the sketches of the original authors [28,30••].

signed. A covalently linked dimer of such coiled coils is meant to present two ligands, spaced some distance apart, for binding to two cell-surface receptors, for diagnostic or therapeutic purposes.

Hodges *et al.* [18•] recently systematically tested the influence of hydrophobic interactions (and disulfide bonds) in *de novo* designed coiled-coil dimers (2 × 35 residues). In the central peptide segment, KXEAXEG (where X represents any amino acid), stability decreased as X is changed from leucine through isoleucine and valine to phenylalanine. Such measured scales complement those derived from database statistics and may be generally useful in planning new designs. Interestingly, the result suggests that nature's use of leucine in the coiled coil (or

'zipper' [19]), used as a dimerization domain in some DNA-binding proteins, optimizes dimer stability. Coiled-coil structures are not globular, but the same priciples of helix–helix association valid for coiled-coil dimers were used in the first *de novo* design of a globular tetrameric helix bundle [20].

Helical bundles

A major step forward in *de novo* design was the successful demonstration that a designed monomeric four-helix bundle, α_4, folded up into a compact globular structure with a very cooperative unfolding transition and an unusually high stability against denaturation with guanidine hydrochloride ($\Delta G = -22$ kcal mol^{-1} for the folded–unfolded transition in water) [2,8]. The design was modular: four identical helices (sequence: GELEE LLKKL KELLK G, similar to the earlier tetrameric design [20]), were linked by three identical loops (sequence: PRR). A crystal structure of α_4 has, however, not yet been reported. It will also be interesting to see whether or not similar stability can be achieved using a non-modular design, i.e. using a sequence with no internal repeats.

A helical bundle of similar size, 'Felix', originally designed in a graduate student course held at Duke University in 1985 [21••], represents the first design of a globular four-helix bundle (Felix) with a non-repeating sequence (Fig. 1). The 79-residue design incorporates two cysteine residues positioned in the sequence in such a way that they would be near one another in the correct three-dimensional structure. It has a single tryptophan residue as a fluorescence probe of core formation and uses 19 of the 20 standard amino acids. After a number of failed attempts, the synthetic gene was expressed in *Escherichia coli* and the protein was purified from inclusion bodies using 6 M urea and allowed to fold. Physical characterization left no doubt that the protein folds up into an approximately correct structure: there is a clear monomer peak after gel filtration, the circular-dichroism spectrum is typical of an all-α-helical protein, the designed disulfide bond between the first and fourth α-helices can be induced to form; tryptophan fluorescence is blue-shifted in the folded relative to a denatured form; and, an unfolding transition is observed with increasing concentration of denaturant.

So, why only 'approximately' the correct structure? Not all attributes of this generation of Felix conform to the design: the protein is only marginally stable, has an unusually broad unfolding transition in guanidine hydrochloride, and appears to have less α-helical content than expected. There are no reports of the successful crystallization of Felix or of the determination of its solution structure by NMR. The authors conclude '... some of the interior side chains may not have only one position and may actually be fairly mobile, as has been hypothesized for the "molten globule" state' [21••]. Molten globule or not, it is clear that the current version of Felix is not yet a typical globular protein.

Binding sites

The now classical design by Gutte *et al.* [22], more than a decade ago, had as its goal a compact 34-residue ββα structure with trinucleotide-binding activity. The protein was reported to bind cytidine phosphates more strongly than trinucleotides and to have some ribonuclease activity. No three-dimensional structure was determined. An equally bold attempt involved the synthesis of a four-stranded β-sheet structure reported to bind the insecticide 2,2-bis(*p*-chlorophenyl)-1,1,1-trichloroethane (DDT) [23].

An apparently much simpler binding function to design into a protein is a metal-binding site, primarily because the liganding principles are well understood. There have been several attempts to engineer copper- and zinc-binding sites into natural proteins (not reviewed here). The challenge was recently extended to a *de novo* designed protein by Handel and DeGrado [24••] and Regan and Clarke [25••]. Both groups engineered a zinc-binding site into the 72-residues helical bundle α_4 [8], with liganding side chains placed such that the metal ion binds at a site between two of the four helices, partly protected from solvent, as in hemocyanin. Both designs exploited the fact that two side chains placed one full helical turn apart (positions i and i + 4) can form two corners of a tetrahedron appropriate for proper binding coordination. Handel and DeGrado used three histidine residues whereas Regan and Clarke used two histidine and two cysteine residues to coordinate the metal ion. Both groups were successful in demonstrating tight binding and, in both cases, the tertiary structure of the bundle was stabilized somewhat by metal binding, with respect to unfolding in guanidine hydrochloride. Regan and Clarke [23••] measured an affinity for Zn(II) of 2.5×10^{-8} M and optical-absorption spectra of the Co(II) complex fully consistent with tetrahedral coordination.

What is the possible use of engineering metal-binding sites into proteins? Metalloproteins can be powerful catalysts in which the metal ion provides part of the electrostatic environment of the active site. Thus, the design of metal-binding sites can be viewed as a step toward the *de novo* design of catalytically active proteins.

How does one identify potential liganding residues in a given protein structure? The trick is to scan various combinations of side chains that are properly positioned to form the binding site and for which the protein structure can tolerate mutation to, for example, histidine or cysteine. Scanning the very many possible combinations in a set of three-dimensional coordinates is best done by computer program and several algorithms have been invented and implemented for this purpose (METAL-SEARCH, [25••]; programs in [7,26•]; Dezymer, HW Hellinga and FM Richards, personal communication; FIND-SITE, A Godzik and C Sander, unpublished).

Design of enzymatic function?

Designing both structure and catalytic function of a protein from scratch is a tall order. Not only does the designed sequence have to fold up into a stable structure. It also has to provide the framework for precise positioning, at the crucial stage of catalysis, of active side chains to within a fraction of an Angstrom, with the active site partially shielded from solvent. Mutter [27] has described a hybrid approach to *de novo* design in which synthetic peptides are chemically linked to a template molecule in order to facilitate formation of a compact unit with protein-like tertiary structure interactions. For example, α-helices attached to a cyclic decapeptide would have an enhanced probability of forming a helical bundle. The following section describes how a template-based approach has been applied in two recent attempts to produce enzyme activity *de novo*.

Helichrome, a hydroxylase

Artificial enzymes have generally been designed by analogy to natural catalysts. On the basis of the architecture of the enzyme cytochrome P450, Sasaki and Kaiser [28] designed a four-helical bundle in which the 15-residue peptide $(AEQLLQE)_2L$ is attached through its amino-terminus to each of the four corners of the porphyrin ring of a heme molecule (Fig. 3). It was intended that the four helices would themselves arrange in parallel to form a hydrophobic core with a cavity near the heme, thus creating a possible binding or active site. The synthetic construct was indeed reported to have aniline hydroxylase catalytic activity ($k_{cat} = 0.02\,min^{-1}$, $K_m = 5.0\,mM$) [28], which is similar to that of natural hemeproteins, suggesting that substrate is indeed bound by the protein in the vicinity of the heme. Spectroscopic results indicate high α-helical content and a fairly broad unfolding transition with a midpoint at 5.2 M guanidine hydrochloride. Thus, the design appears to have worked up to a point. Whether or not the three-dimensional structure of helichrome turns out as advertized may become known if crystals obtained by J Kuriyan and L Wong (personal communication in [29•]) lead to a solved three-dimensional structure.

Chymohelizyme, an esterase

Has the recent design and synthesis of an α-helical bundle with chymotrypsin-like esterase activity been successful? Four synthetic peptides of partly similar sequence (rich in glutamic acid, cysteine and leucine residues), arranged in parallel, were covalently linked at their carboxy-terminal ends [30••]. The design called for a hydrophobic core, and properly placed serine, histidine and aspartic acid residues in the active site at the amino-terminal end of the bundle, as well as a hydrophobic pocket complementary to the aromatic side chains of chymotrypsin to provide chymotrypsin-like specificity. Circular dichroism measurements indicated about 65% helix content in water. The construct catalyses the hydrolysis of ethyl esters with an acetyltyrosine or benzoyltyrosine (but not benzoylarginine) side group — albeit at rates that were very low compared with those of chymotrypsin . Not all the evidence that would prove this design successful is available yet; there is surprisingly little primary data in the original publication [30••], and no crystals or NMR spectra have been reported. Nonetheless, the design is very promising and has already generated considerable interest in the scientific press [31,32].

Membrane ion channels

In earlier simple but elegant work [2,33], DeGrado and colleagues had shown that the synthetic peptides $(LSLLLSL)_3$ and $(LSSLLSL)_3$ formed ion channels in phospholipid bilayer membranes, presumably as parallel tetrameric α-helical bundles. These bundles had been designed on principles observed in globular helical proteins, but with reversed polarity. Non-polar residues were designed to be on the periphery of the helical bundle whereas polar residues were intended to point into the space between the helices, as would be required of a channel that is embedded in a non-polar medium and is to allow the transport of a polar entity along its central axis. Functional success of the design was demonstrated early on: channel formation varied with peptide concentration and transmembrane voltage, although only a small fraction of peptides was involved in channel formation at a given time. Detailed structural information, however, is missing, and it is not clear how the channels, which have life times in the millisecond range turn on and off. Recently, the group has obtained evidence for the α-helical structure of the peptides in phospholipid membranes from circular-dichroism spectroscopy and from the fact that the conformationally restricted amino acid α-aminoisobutyric acid can be introduced at the fourth position in each peptide repeat without greatly affecting conductance [34••]. This system may be very resistant to attempts at crystallization, but work will continue. We have barely witnessed the beginning of membrane protein design.

$(\beta\alpha)_n$ structures

One of the most common and versatile protein-structure classes is that of eightfold βα barrels of the triose phosphate isomerase (TIM) type. This class includes many enzymes of diverse function. The idea of developing a general scaffold onto which binding and active sites can be engineered, analogous to attempts to develop catalytic antibodies, has led at least two groups to pursue the design of $(\beta\alpha)_8$ barrels.

Goraj *et al.* [35••] have synthesized the gene for an eightfold repeating 31-residue βα sequence: SGLVV YLGKR PDSGT ARELL RHLVA EGDAR S. The purified protein, Octarellin, is soluble in aqueous solution, and circular dichroism indicates significant α-helical content; this, however, is not in itself evidence for tertiary structure.

Urea gradient gel electrophoresis suggests a cooperative unfolding transition; interestingly, no such transition is reported for analogous sevenfold and ninefold repeats of the same sequence. However, globularity of structure remains to be demonstrated and the three-dimensional structure has still to be verified. There is also room for improvement of the design. Fourfold symmetry, instead of the eightfold symmetry used in Octarellin, more accurately reflects the architecture of natural barrels of this type. Fourfold symmetry would be favored by using an elementary repeat unit of double the current length. Alternatively, the accumulation of 16 positive charges at one end of the barrel in the current design is not beneficial for structure formation and should be modified.

Babarellin, designed during a workshop at the European Molecular Biology Laboratory in 1986 [36], is a 178-residue $(\beta\alpha)_8$ barrel, with a fourfold repeat of a basic $\beta\alpha\beta\alpha$ unit. The corresponding gene, one of the longest synthetic genes for a *de novo* designed globular protein to date, has been synthesized and expressed (G Nyakatura, SC Emery and H Fritz, personal communication). Babarellin has a cooperative unfolding transition (FX Schmid, personal communication) and attempts at structure determination are in progress (J Richardson, personal communication; WE Eberle and C Sander, unpublished data).

Tanaka and colleagues at the Protein Engineering Research Institute in Osaka have designed several $(\beta\alpha)_8$ barrel proteins, with a basic $(\beta\alpha\beta\alpha)$ unit repeated four times to give a total length of more than 200 residues [37]. One of the variants has putative ATP-binding sites as well as cleavage sites for kallikrein and thrombin (H Nakamura, personal communication). The synthetic genes have been expressed and the proteins purified and subjected to a refolding protocol. Three of the proteins appear to have a folded monomeric structure, as judged by size-exclusion chromatography and circular-dichrosim spectroscopy. The most stable variant, TIM-1, has a cooperative unfolding curve with a midpoint at 2.9 M urea. Efforts to improve the sequences are in progress.

As nature has used the $(\beta\alpha)_8$-barrel structural motifs in a variety of enzymes, we can look forward to a series of constructs with diverse binding sites, as soon as the *de novo* designed $(\beta\alpha)_8$ barrels now being developed by several groups have attained sufficient structural stability.

Other structures

β-Sheet sandwiches

Experimental work on several other *de novo* designs is continuing or being initiated in several groups. Betabellin, a sandwich of two four-stranded β-sheets of simple up-down-up-down topology [1] is now in its 12th generation of sequence changes and the current version exhibits much improved NMR spectra (J Richardson, personal communication). Another protein, Betadoublet, of similar topology but much different sequence, has recently been designed and produced by the same group (TP Quinn *et al.*, personal communication), with promising NMR spectra in aqueous solution.

New design suggestions

The 1990 'Protein Design on Computers' workshop at the European Molecular Biology Laboratory [4•] tested available design tools and explored new design strategies. The result was the design of five proteins: Shpilka, a sandwich of two four-stranded β-sheets, a scaffold on which to explore variations in loop topology [38] (Fig. 2); Grendel, a four-helical membrane anchor, ready for fusion to water-soluble functional domains; Fingerclasp, a dimer of interdigitating $\beta\beta\alpha$ units, the simplest variant of the 'handshake' structural class; Aida, an antibody-binding surface intended to be specific for flavodoxin; and Leather, a minimal NAD-binding domain, extracted from a larger protein. Each of these are starting points for further design refinement, with the coordinates and sequences publicly available via electronic mail from netserv@emblheidelberg.de on internet (send the message 'help'). One of these, FingerClasp [39], has been synthesized by H Nakamura at the Protein Engineering Research Institute in Osaka H Nakamura (personal communication). Spectroscopic results so far indicate that the protein does not fold up as intended although some α-helical content is evident from circular dichroism. A redesigned second version of FingerClasp shows improved circular-dichroism spectra.

Why bother?

Depending on the individual author, *de novo* protein design may be said to aim at a better understanding of the principles of protein folding, at the development of simple model systems for studying structure and function, at the development of medically or industrially useful proteins, at the production of novel proteins with desired binding properties and catalytic activities, at the nanometer scale engineering of self-assembling microscopic structures, or simply at having fun learning new things. There are perhaps better approaches to achieve any of these goals. For example, new and better enzyme functions can perhaps more elegantly be evolved from natural enzymes by applying appropriate selection pressure to a diverse population of sequences in cell culture. 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 the many families of natural proteins or from mutational experiments on natural proteins. Nanometer engineering using protein material should perhaps take natural membrane proteins as its starting point. So why bother with *de novo* design?

The basic answer comes from the human desire to master the principles of natural processes not only in a passive but in an active way. Protein engineers wish to be free from the constraints and frozen accidents of natural evolution. The ability to engineer new proteins from

scratch, from first principles, would open a wide range of possible techniques and applications, transcending natural evolution. *De novo* protein design is likely to become one of the main pillars of biomolecular engineering. It only remains to be developed vigorously and applied wisely.

Acknowledgement

I thank several colleagues for generous personal communications and G Vriend, L Holm, H Nakamura, AV Finkelstein, FX Schmid, H Fritz and WF DeGrado for comments on the manuscript.

References and recommended reading

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

- • of interest
- •• of outstanding interest

1. RICHARDSON JS, RICHARDSON DC: **The *de novo* Design of Protein Structures.** *Trends Biochem Sci* 1989, 14:304–309.

2. DEGRADO WF, WASSERMANN ZR, LEAR JD: **Protein Design, a Minimalist Approach.** *Science* 1989, 243:622–628.

3. DEGRADO WF: **Design of Peptides and Proteins.** *Adv Protein Chem* 1988, 39:51–124.

4. • SANDER C, VRIEND G (EDS): **Protein Design on Computers. Five New Proteins: Shpilka, Grendel, Fingerclasp, Leather and Aida.** *Proteins* 1991, in press.
A multi-author report from an EMBO practical course in which new approaches to *de novo* design were explored and new computer methods tested.

5. • VRIEND G: **WHATIF: a Molecular Modeling and Drug Design Program.** *J Mol Graphics* 1990, 8:52–56.
This very versatile program includes combined sequence-structure database searches and a large number of tools for the analysis and manipulation of protein structures.

6. FINKELSTEIN AV, PTITSYN OB: **Why Do Globular Proteins Fit a Limited Set of Folding Patterns?** *Prog Biophys Mol Biol* 1987, 50:171–190.

7. FINKELSTEIN AV, KIRPICHNIKOV MP, PITTSYN OB, SKYABIN KG: **Design of Protein Molecules.** *Vesnik Acad Nauk SSSR (USSR)* 1988, 3:102–111.

8. REGAN L, DEGRADO WF: **Characterization of a Helical Protein Designed From First Principles.** *Science* 1988, 241:976–978.

9. • HOLM L, SANDER C: **Database Algorithm for Generating Protein Backbone and Side Chain Coordinates From a C(α) Trace. Application to Model Building and Detection of Coordinate Errors.** *J Mol Biol* 1991, 218:183–194.
A convenient and very fast program for building complete three-dimensional models from a C_α backbone trace.

10. BAUMANN G, FRÖMMEL C, SANDER C: **Polarity as a Criterion in Protein Design.** *Protein Eng* 1989, 2:329–334.

11. EISENBERG D, MCLACHLAN AD: **Solvation Energy in Protein Folding and Binding.** *Nature* 1986, 319:199–203.

12. • GREGORET LM, COHEN FE: **Novel Method for the Rapid Evaluation of Packing in Protein Structures.** *J Mol Biol* 1990, 211:959–974.
The advantage of this method is that packing can be evaluated from knowledge of the (α) trace only, using a database-derived average residue–residue contact potential.

13. •• HILL CP, ANDERSON DH, WESSON L, DEGRADO WF, EISENBERG D: **Crystal Structure of α_1: Implication for Protein Design.** *Science* 1990, 249:543–545.
The first crystal structure to be reported for a synthetic peptide designed *de novo* to form a globular protein assembly is described. The design called for an antiparallel tetramer of four α-helices. The crystal structure did exhibit largely α-helical monomers, association of α-helices and an antiparallel α-helix pair. Surprisingly, however, the most stable group of helices in the crystal was a hexamer, in disagreement with the design goal.

14. HODGES RS, SODEK LB, SMILLIE LB, JURASEK L: **Tropomyosin: Amino Acid Sequence and Coiled Coil Structure.** *Cold Spring Harbor Symp Quant Biol* 1972, 37:299–310.

15. • COHEN C, PARRY DAD: **α-Helical Coiled Coils and Bundles: How to Design an α-Helical Protein.** *Proteins* 1990, 7:1–15.
An excellent introduction to the design and properties of coiled-coil and helical-bundle structures.

16. HODGES RS, SAUND AK, CHONG PCS, ST. PIERRE SA, REID RR: **Synthetic Models for Two-Stranded α-Helical Coiled Coils.** *J Biol Chem* 1981, 256:1214–1224

17. • ENGEL M, WILLIAMS RW, ERICKSON BW: **Designed Coiled-Coil Proteins: Synthesis and Spectroscopy of Two 78-Residue α-Helical Dimers.** *Biochemistry* 1991, 30:3161–3169.
Modular design of 39-residue peptides that contain the repeat $(RIEAIEA)_4$, a spacer (GYG) and a cell-surface ligand (GRGDSP). The peptides appear to form pairs of α-helices, as designed. Eventually, a covalently linked dimer of such coiled coils is meant to present two ligands, spaced some distance apart, for binding to two cell-surface receptors, for diagnostic or therapeutic purposes.

18. • HODGES RS, ZHOU NE, KAY CM, SEMCHUK PD: **Synthetic Model Proteins: Contribution of Hydrophobic Residues and Disulfide bonds to Protein Stability.** *Pept Res* 1990, 3:123–137.
Report of an elegant method for elucidating the contributions of single residues to protein stability in the structural context of 2 × 35 residue coiled-coil dimers of α-helices. Leucine is seen to be the most useful for stabilizing homodimeric helix–helix contacts.

19. LANDSCHULZ WH, JOHNSON PF, MCKNIGHT SL: **The Leucine Zipper: a Hypothetical Structure Common to a New Class of DNA Binding Proteins.** *Science* 1988, 240:1759–1764.

20. EISENBERG D, WILCOX W, ESHITA SM, PRYCIAK PM, HO SP, DEGRADO WF: **The Design, Synthesis, and Crystallization of an α-Helical Peptide.** *Proteins* 1986, 1:16–22.

21. •• HECHT MH, RICHARDSON JS, RICHARDSON DC, OGDEN RC: ***De novo* Design, Expression and Characterization of Felix: a Four-Helix Bundle Protein of Native-Like Sequence.** *Science* 1990, 249:884–891.
This paper reports the first sucessful *de novo* design of an α-helical bundle with a non-repeating sequence. The bacterially expressed protein is primarily monomeric and α-helical, forms an intramolecular disulfide bond and buries a tryptophan side chain, consistent with the designed tertiary structure. However, the protein is only marginally stable, has an unusually broad unfolding transition and appears to have a fairly mobile core, untypical of native globular proteins.

22. GUTTE B, DÄUMINGEN M, WITTSCHIEBER E: **Design, Synthesis and Characterization of a 34-Residue Polypeptide that Interacts With Nucleic Acids.** *Nature* 1979, 281:650–655.

23. MOSER R, THOMAS RM, GUTTE B: **An Artificial Crystalline DDT-Binding Peptide.** *FEBS Lett* 1983, 157:247–251.

24. •• HANDEL T, DEGRADO WF: ***De novo* Design of a Zn^{2+}-Binding Protein.** *J Am Chem Soc* 1990, 112:6710–6711.
One of two parallel attempts to engineer a metal-binding site into the very stable *de novo* designed four-helical bundle, α_4. Protein ligands are one histidine side chain on helix 2 and two histidine side chains on helix 3. The experiment is successful in that zinc ions are bound tightly to the synthetic monomeric peptide (as well as to an analogous dimer of 2 × 2 helices). The experimental evidence comes from the sharp-

ening of histidine resonances in the one-dimensional NMR spectrum upon increasing Zn concentration.

25. •• REGAN L, CLARKE ND: **A Tetrahedal Zinc(II)-Binding Site Introduced into a Designed Protein.** *Biochemistry* 1990, 29:10878–10883.

One of two parallel attempts to engineer a metal-binding site into the very stable *de novo* designed four-helical bundle, α_4. Protein ligands are two histidine and two cysteine residues, with a His-Cys pair on each of helix 2 and 3. The experiment is successful in that zinc and cobalt ions are bound with high affinity to the bacterially produced 74-residue protein. The experimental evidence comes from gel filtration and blotting with radioactive zinc.

26. • ROBERTS VA, IVERSON BL, IVERSON SA, BENKOVIC SJ, LEMER RA, GETZOFF ED, TAINER JA: **Antibody Remodeling: a General Solution to the Design of a Metal-Coordination Site in an Antibody Binding Pocket.** *Proc Natl Acad Sci USA* 1990, 87:6654–6658.

A zinc-binding site is designed and shown to bind metal ions, with the highest affinity for copper ions. The design is based on a binding-site template from the crystal structure of the enzyme carbonic anhydrase. This is not *de novo* design, but is quoted here to illustrate the power of the 'template-based' approaches to the design of the binding sites in proteins.

27. MUTTER M: **Nature's Rules and Chemist's Tools: a Way for Creating Novel Proteins.** *Trends Biochem Sci* 1988, 13:260–265.

28. SASAKI T, KAISER ET: **Synthesis and Enzymatic Activity of a Designed Hemeprotein.** *J Am Chem Soc* 1989, 111:380–381.

29. • SASAKI T, KAISER ET: **Synthesis and Structural Stability of Helichrome as an Artificial Hemoprotein.** *Biopolymers* 1990, 29:79–88.

A follow-up to the first published design of a peptide assembly (four α-helices attached to a porphyrin ring) with a specific structure and, at the same time, enzymatic function [28]. Enzymatic activity similar to that of other heme proteins is demonstrated. Small crystals exist that may lead to a successful three-dimensional structure determination.

30. •• HAHN KW, KLIS WA, STEWART JM: **Design and Synthesis of a Peptide Having Chymotrypsin-Like Esterase Activity.** *Science* 1990, 248:1544–1547.

The second publication of a design for a peptide assembly with a specific structure and, at the same time, enzymatic function. The design is reported to have been successful in terms of specific low-level enzymatic activity. However, experiments to test correctness of three-dimensional structure have yet to be performed.

31. RINGE R: **Designing the Enzyme of the Nineties.** *Curr Biol* 1991, 1:67–68.

A review of [30••].

32. BEARDSLEY TM: **Molecular Engineers Mimic Mother Nature (News).** *Sci Am* 1990, 263:30–30.

A review of [30••]. The prospect of designing proteins to treat animal poisoning by nerve gas or pesticides is reported.

33. LEAR JD, WASSERMAN ZR, DEGRADO WF: **Synthetic Amphiphilic Peptide Models for Protein Ion Channels.** *Science* 1988, 240:1177–1181.

34. •• DEGRADO WF, LEAR DJ: **Conformationally Constrained α-Helical Peptide Models for Protein Ion Channels.** *Biopolymers* 1990, 29:205–213.

A continuation of earlier work on the first *de novo* designed ion channels [33], now with spectroscopic and other evidence that the synthetic peptides $(LSLLLSL)_3$ and $(LSSLLSL)_3$, shown earlier to form ion channels in phospholipid bilayer membranes, adopt a primarily α-helical structure.

35. •• GORAJ K, RENARD A, MARTIAL JA: **Synthesis, Purification and Initial Structure Characterization of Octarellin, a *de novo* Polypeptide Modelled on the α/βBarrel Proteins.** *Protein Eng* 1990, 3:259–266.

First published experimental test of an αβ-barrel protein designed *de novo*. The design is simple, with the basic 31-residue peptide repeated eight times. So far, there is some urea gradient electrophoresis evidence for a cooperative unfolding transition of this bacterially expressed protein, but no direct evidence yet for tertiary-structure formation.

36. RICHARDSON JS, KNELLER D, OSGUTHORPE D, SCHARF M: **Babarellin a Design for a 4-Fold Symmetric Tim Barre.** In *Protein Design Exercises, EMBL Biocomputing Technical (Documen 1)* edited by Sander C. Heidelberg: European Molecular Biology Laboratory, 1987, pp 21–39.

37. TANAKA T, ANAGUCHI H, HAYASHI M, FUKUHARA K, NAKAMURA H, IKEHARA M: ***De novo* Design and Synthesis of TIM Barrel Proteins.** In *Proceedings of the 12th American Peptide Symposium* 1991, in press.

38. FINKELSTEIN AV, LOCKHART A, MERKL R, PERRY LJ: **Shpilka Design: an Eight-Stranded Antiparallel β-Sandwich.** In *Protein Design on Computers, EMBL Biocomputing Technical Document 6* edited by Sander C, Vriend G. Heidelberg: European Molecular Biology Laboratory, 1991, pp 133–148.

39. BAZAN F, HOROWITZ A, NAKAMURA H, RIBAS L: **Design and Analysis of Novel Handshake Structure.** In *Protein Design on Computers, EMBL Biocomputing Technical Document 6* edited by Sander C, Vriend G. Heidelberg: European Molecular Biology Laboratory, 1991, pp 105–132.

C Sander, European Molecular Biology Laboratory, Meyerhofstrasse 1, D-6900 Heidelberg, Germany.