PROTEINS: Structure, Function, and Genetics 12:105–110 (1992)

Protein Design on Computers. Five New Proteins: Shpilka, Grendel, Fingerclasp, Leather, and Aida

Chris Sander, Gerrit Vriend, Fernando Bazan, Amnon Horovitz, Haruki Nakamura, Luis Ribas, Alexei V. Finkelstein, Andrew Lockhart, Rainer Merkl, L. Jeanne Perry, Stephen C. Emery, Christine Gaboriaud, Cara Marks, John Moult, Christophe Verlinde, Marc Eberhard, Arne Elofsson, Tim J.P. Hubbard, Lynne Regan, Jay Banks, Roberto Jappelli, Arthur M. Lesk, and Anna Tramontano
European Molecular Biology Laboratory, D-6900 Heidelberg, Federal Republic of Germany

ABSTRACT **What is the current state of the art in protein design? This question was approached in a recent two-week protein design workshop sponsored by EMBO and held at the EMBL in Heidelberg. The goals were to test available design tools and to explore new design strategies. Five novel proteins were designed: Shpilka, a sandwich of two four-stranded β-sheets, a scaffold on which to explore variations in loop topology; Grendel, a four-helical membrane anchor, ready for fusion to water-soluble functional domains; Fingerclasp, a dimer of interdigitating β–β–α units, the simplest variant of the "handshake" structural class; Aida, an antibody binding surface intended to be specific for flavodoxin; Leather—a minimal NAD binding domain, extracted from a larger protein. Each design is available as a set of three-dimensional coordinates, the corresponding amino acid sequence and a set of analytical results. The designs are placed in the public domain for scrutiny, improvement, and possible experimental verification.**

Key words: protein structure, protein sequences, protein design de novo, protein engineering, computer algorithms

INTRODUCTION

The forward protein folding problem, that of calculating structure from sequence, remains basically unsolved. The inverse problem, that of designing sequences to achieve desired structural properties, has been the subject of considerable effort over the last few years (for reviews, see refs. 1 and 1a). We have attempted to promote the development of new techniques and to encourage the design of new types of proteins by organizing intensive workshops, first in 1986[2] and, more recently, in 1990.[3]

The problem posed to the participants was this: specify a model protein structure (or structural property) and then invent a protein sequence that will lead to the desired structure. In the same spirit, in 1986 two idealized 4*βαβα barrel structure scaffolds (Babarellin, Tiny Tim), two β/α folds (Betalphacin, Idealized Flavodoxin), a bundle of four α-helices (Bundle), and a Cu-binding variant of a natural protein (CuRop) were constructed, and the corresponding protein sequences invented.[2] Since then, Babarellin and CuRop have been synthesized and purified (G. Nyakatura, H.-J. Fritz, and S. C. Emery, personal communications) and structural tests are in progress (F.X. Schmid, W. Eberle, J. Richardson, and M. Sagermann, personal communications). The five new proteins designed in the recent 1990 workshop are described below by each of the working groups.

TECHNIQUES USED

Typically, the design procedure followed these steps: (1) analyze known protein structures and identify structural units that might be used as building blocks, e.g., αβ units; (2) sketch out the secondary structure elements, their relative orientations, and the topology of loop connections, e.g., a four-stranded antiparallel β-sheet packed against two α-helices crossing the strands; (3) construct the protein scaffold by building explicit backbone coordinates, first for the structural core, consisting primarily of elements of secondary structure, then for loops; (4) choose an appropriate amino acid sequence in interior and surface regions, e.g., Glu on the surface of a helix near the N-terminus, Val or Ile at a β–β or β–α interface and so on; (5) optimize the model in interactive mode using visual inspection or in automatic mode using molecular mechanics software, i.e., vary backbone and side chain degrees of freedom, with simple energetics as a guide; primary goals are to regularize covalent geometry, remove clashes, avoid holes, optimize hydrogen bonding as well as charge–charge and protein–solvent interactions; (6) check the quality of the model by analyzing, e.g., solvation, electrostatic, or interior packing

Received April 16, 1991; accepted June 17, 1991.
Address reprint requests to Chris Sander, European Molecular Biology Laboratory, Meyerhofstrasse 1, D-6900 Heidelberg, Federal Republic of Germany.

bility or critical during folding. Alternatively, they may just be the remains of previous evolution, without a critical role. If parts could be removed, functional molecules of a more manageable and more economical size would be obtained.

To this end we looked for a protein system where a ligand binding site was defined by residues from relatively few secondary structural units. We chose lactate dehydrogenase[21] as the basis for a minimal NAD binding protein. Natural NAD binding sites have been probed with compounds mimicking parts of the NAD molecule, interacting with only part of the extended binding site. The same compounds could be useful in testing a designed protein that failed to bind NAD to investigate if at least parts of the binding site were present.

To obtain the minimal design, the C-terminal (catalytic) domain and one β–α unit of the N-terminal (NAD-binding) domain were removed (Fig. 2D). However, a C-terminal α-helix, essential for the binding site, was retained by reconnecting it to the old N-terminus using a new loop. A disulfide bond was introduced to stabilize the other end of this helix (the new N-terminus). Residues newly exposed to solvent were mutated to "solubilize" the protein and a single Trp residue was incorporated into the hydrophobic core of the protein to act as a spectroscopic probe for folding. The final designed sequence has 131 residues compared to 329 for lactate dehydrogenase.

AIDA: AMERICAN–ITALIAN DESIGNER ANTIBODY

Can we predict sequences of antibodies that will bind specified epitopes on proteins?

The antigen-binding sites of immunoglobulins are created by six loops between strands of β-sheet in the variable domains of light and heavy chains. At least five of the six loops show a limited range of mainchain conformations called canonical structures;[22,23] these main chain conformations are determined by a few particular residues in the sequence. Other residues in the loops are relatively free to vary to create a variety of surface topographies and charge distributions in the binding site without changing the main chain conformations of the loops.

Because we know which residues we can change and which we must leave alone, in order to retain the main chain conformation of the antigen binding site, we can attempt to design modifications of a known antibody–antigen interface to substitute one antigen for another. In this exercise we started with the known structure of the Fab HyHEL-5/Hen egg white lysozyme complex,[24] and tried to substitute flavodoxin for lysozyme.

We began by placing the flavodoxin in the same region of space, relative to the antibody, that lysozyme occupied. Then we stripped off the side chains of the residues in the antibody binding site, and tried to rebuild the interface. The goal was a well-fitting interface which would form without change in main chain conformation of either protein. The techniques used included careful study of the interface, at each stage, using computer graphics software, followed by many cycles of mutations and refinement by energy minimization.

The project allowed us to evaluate the difficulty of designing an interface complementary to an antigen. The goal is of practical interest because, if successful, it would make is possible to design an antibody against a pathogen. In this respect the design techniques may take their place in the general context of artificial antibody construction, although it will certainly be a long time before theoretical methods can compete with the ability of the immune system to create mature antibodies of finely tuned affinity and specificity.

ACKNOWLEDGMENTS

We wish to thank the European Molecular Biology Organization as well as Silicon Graphics, Evans and Sutherland, Hoechst AG, and Hoffmann-LaRoche for financial and material support; Reinhard Schneider, Ulrike Goebel, and John Priestle for ribbon plots; and the EMBL Computer Group for system support. A.L. thanks the Kay Kendall Foundation for support. Authors are from various laboratories in Europe, the U.S.A. and Japan (addresses not given). Working groups were as follows. Fingerclasp: F. Bazan, A. Horovitz, H. Nakamura and L. Ribas; Shpilka: A.V. Finkelstein, A. Lockhart, R. Merkl, L.J. Perry; Grendel: S. Emery, C. Gaboriaud, C. Marks, J. Moult, C. Verlinde; Leather: M. Eberhard, A. Elofsson, T.J.P. Hubbard, L. Regan; Aida: J. Banks, R. Jappelli, A.M. Lesk, A. Tramontano. Teachers were A.V. Finkelstein, T. Hubbard, A. Lesk, J. Moult, H. Nakamura, C. Sander, A. Tramontano, and G. Vriend.

REFERENCES

1. Richardson, J., Richardson, D.C. The de novo design of protein structures. TIBS 163:304–309, 1989.
1a. Sander, C. De novo design of proteins. Curr. Opin. Struc. Biol. 1:630–637, 1991.
2. Protein Design Exercises 86, EMBL BIOcomputing Technical Document 1, C. Sander, ed., 1987.
3. Protein Design 90, EMBL BIOcomputing Technical Document 6, C. Sander and G. Vriend, eds., 1991.
4. Gregoret, L.M., Cohen, F.E. Novel method for the rapid evaluation of packing in protein structures. J. Mol. Biol. 211:959–974, 1990.
5. Novotny, J., Bruccoleri, R., Karplus, M. An analysis of incorrectly folded protein models. J. Mol. Biol. 177:787–818, 1984.
6. Baumann, G., Froemmel, C., Sander, C. Polarity as a criterion in protein design. Protein Eng. 2:329–334, 1989.
7. Bjorkman, P.J., Saper, M.A., Samraoui, B., Bennett, W.S., Strominger, J.L., Wiley, D.C. Structure of the human class I histocompatibility antigen HLA-H2. Nature (London) 329:506–512, 1987.
8. St. Charles, R., Walz, D.A., Edwards, B.F.P. The three-dimensional structure of bovine platelet factor 4 at 3.0 Å resolution. J. Biol. Chem. 264:2092–2099, 1989.

```
A
        ....;....1....;....2....;....3....;..
fing    NDVKVDVRSNGGTKTEWKADSADLEKFAKQLKEKHGA
        _EEEEEEESTTTEEEEE_SSSHHHHHHHHHHHHHT__

B
        ....;....1....;....2....;....3....;....4....;....5....;....6....;....7....;....8....;....9....;.
shpi    GIKMTITLELGTTKHSIPIETGSPGGIRMTITLELGTTRHSVPVEGGTKHSVPVDTGGGGGVKWTVTMDLGTTRHSTPVDTSGSPNVRMTVTVDLG
        _EEEEEEEEETTEEEEEEEEEESS___EEEEEEEEETTEEEEEEEEEESSEEEEEEEE___S_EEEEEEEEETTEEEEEEEEES__S_EEEEEEEE__

C
        ....;....1....;....2....;....3....;....4....;....5....;....6....;....7....;....8....;....9....;....0
gren    PEWIWLALGTALMGLLTLYFLVKG!KKFYAITTLVPAIAFTMFLSMLLKGGTDG!GNPWYAITTLVPAIAFTMLLFMLLMTAAAGG!GNPKVAVIILIAL
        _THHHHHHHHHHHHHHHHHHHHH_!_HHHHHHHHHHHHHHHHHHHHHHHHHH___!__HHHHHHHHHHHHHHHHHHHHHHHHHHHHH_!__HHHHHHHHHHH

gren    LTALMGLLTLYSLV
        HHHHHHHHHHHHH_

D
        ....;....1....;....2....;....3....;....4....;....5....;....6....;....7....;....8....;....9....;....0
leat    MGCGTSDAQGAKVADDAETKMNGGVNKITVVGYGAVGAASAASILAKDLADQVAYVDGGSDYSDSAGSKLVVITAGARQSCGESRENLVQRNVNIAKYIA
1LDM     KGY--W-I-LS---L---I-     ------V----M-C-I---M-----E--L--  K---V---------------QE----L---------F-F-I
        ______HHHHHHHHHHHHHHHHTS__SEEEEE__SHHHHHHHHHHHHTT__SEEEEE_TTS_GGGGTT_SEEEE__S____TT__GGGGHHHHHHHHHHH

leat    PNIQKQSPEVTILIVSNPSDALTEVWKKNSG
1LDM    ---V-H--DCI--V----V-V--Y-AW-L--
        HHHHHH_TT_EEEE_SSSHHHHHHHHHHHH_

E
heavy (H) and light (L) variable domain, residues 1-60 and 87-119
        <....;....1....;....2....;....3....;....4....;....5....;....6> <...9....;....0....;....1....;....>
aida-L  <DIVLTQSPAIMSASPGEKVTMTCSASSSVSQMIWFQQKSGTSPKRWIQNTSKLASGVPVR> <CQQMSSNPTFGGGTKLEIKRADAAPTVSIFPPS>
2HFL-L  <-----------------------------NY-Y-Y-----------YD----------> <---WGR---------------------------->
        <___EEEE_SEEEE_TT__EEEEEEEEEEES__EEEEE_SSS__EEEEETTEE_TT__TT> <EE__SSS__BB__EEEEE____B__EEEEE___>

        <....;....1....;....2....;....3....:....4....:....5....:....6> <...9....;....0....:....1....;....>
aida-H  <QVQLQQSGAELMKPGASVKISCKASGYTFSNSHIDWVKQRPGHGLEWIGVINPGSGSTRY> <TSEDSGVYYCLHGHYQFDGWGQGTTLTVSSAKT>
2HFL-H  <-----------------------------DYW-E-------------E-L------N-> <-------------N-D----------------->
        <__EEEE___EEE_TT__EEEEEEEESS_GGGS_EEEEEE_TTS_EEEEEEEEGGGTEEEE> <_STT_EEEEEE__SSS___B___EEEEE_SS__>
```

Fig. 1. Amino acid sequences of the designed proteins. fing, FingerClasp (**A**); shpi, Shpilka (**B**); gren, Grendel (**C**); leat, Leather (**D**); and aida, Aida (**E**). Below the sequences, secondary structure are shown as extracted from the model 3-D coordinates by the program DSSP[25]: H,G, helix; E,B, β (extended) structure; T, H-bonded turn; S, bend; —, extended loop; !, chain break (the loops between the four helices in Grendel were not modeled). Finger, Grendel, and Shpilka are de novo designed sequences. Leather is constructed from rearranged regions of lactate dehydrogenase (Protein Data Bank code 1LDM) plus two new sequence pieces: 1LDM residues 242-261/NGGVN/1LDM 22-72/GG/1LDM 81-152, followed by point mutations, where / means concatenation. In line 1LDM, residues kept in Leather are marked by—while mutated ones are given in one letter code. Aida is the result of designed point mutations in the antilysozyme antibody (Protein Data Bank code 2HFL), so only sequence regions which contain mutations relative to the original antibody are given (1–60 and 87–119 of both light and heavy variable domains); residues 61–86 in both domains of Aida are exactly as in 2 HFL. In line 2HFL, residues kept in Aida are marked by —, while mutated residues are given in one letter code. The protein coordinates can be obtained as electronic mail from netserv@EMBL-Heidelberg.DE on Internet (send the messages "help" and "dir proteindata").

5. edge strands that are restricted in backbone H-bond formation on one specified side; and
6. close packing of side chains in the interior.

The designed sequence of Shpilka (Fig. 1B) contains eight regions enriched in β-forming residues, separated by β-breaking connections. Nonpolar and polar residues alternate to form well-defined inner and outer surfaces for β-strands but not for helices. Inner (nonpolar) surfaces of the strands contain Val, Ile, and Met residues, the best β-formers. The inner surfaces of edge strands contain β-forming Thrs, which are partly polar, making contact with the hydrophobic core as well as with solvent. The outer surfaces of all strands include charged groups that can form salt bridges. Pros, which have no backbone NH group, are incorporated into the edge strands so that only one side of an edge strand can form a continuous hydrogen bond net.

Turns and loops are made of β-breaking residues; to help ensure unambiguous folding, they have minimal length, just enough to connect the strands in the designed structure. A Trp was included on the interior face of the sheet to serve as an experimental marker for correct folding. At the inner surfaces of internal strands bulky side chains alternate with small Alas in order to form complementary surfaces for close packing. We did our best to improve the close packing by "point mutations" using computer graphics software as well as CPK models. According to the theory described in refs. 12 and 13, the designed sequence would form a β-sheet sandwich with the desired fold.

GRENDEL: A FOUR HELIX MEMBRANE ANCHOR

The recent determination of two membrane protein structures by X-ray crystallography (the photoreaction center[14]) and electron microscopy (bacteriorhodopsin BRH),[15] together with progress in understanding the basis of their stability and

A

```
         ....;....1....;....2....;....3....;..
fing     NDVKVDVRSNGGTKTEWKADSADLEKFAKQLKEKHGA
         _EEEEEEESTTTEEEEE_SSSHHHHHHHHHHHHHT__
```

B

```
         ....;....1....;....2....;....3....;....4....;....5....;....6....;....7....;....8....;....9....;.
shpi     GIKMTITLELGTTKHSIPIETGSPGGIRMTITLELGTTRHSVPVEGGTKHSVPVDTGGGGGVKWTVTMDLGTTRHSTPVDTSGSPNVRMTVTVDLG
         _EEEEEEEEETTEEEEEEEEESS___EEEEEEEEETTEEEEEEEEESSEEEEEEEE___S_EEEEEEEEETTEEEEEEEEES__S_EEEEEEE__
```

C

```
         ....;....1....;....2....;....3....;....4....;....5....;....6....;....7....;....8....;....9....;....0
gren     PEWIWLALGTALMGLLTLYFLVKG!KKFYAITTLVPAIAFTMFLSMLLKGGTDG!GNPWYAITTLVPAIAFTMLLFMLLMTAAAGG!GNPKVAVIILIAL
         _THHHHHHHHHHHHHHHHHHHHH_!_HHHHHHHHHHHHHHHHHHHHHHHHH___!__HHHHHHHHHHHHHHHHHHHHHHHHHHHHH_!__HHHHHHHHHH

gren     LTALMGLLTLYSLV
         HHHHHHHHHHHHH_
```

D

```
         ....;....1....;....2....;....3....;....4....;....5....;....6....;....7....;....8....;....9....;....0
leat     MGCGTSDAQGAKVADDAETKMNGGVNKITVVGYGAVGAASAASILAKDLADQVAYVDGGSDYSDSAGSKLVVITAGARQSCGESRENLVQRNVNIAKYIA
1LDM      KGY--W-I-LS---L---I-     ------V----M-C-I---M-----E--L--  K---V---------------QE----L---------F-F-I
         ______HHHHHHHHHHHHHHHHTS__SEEEEE__SHHHHHHHHHHHHTT__SEEEEE_TTS_GGGGTT_SEEEE__S____TT__GGGGHHHHHHHHHHHH

leat     PNIQKQSPEVTILIVSNPSDALTEVWKKNSG
1LDM     ---V-H--DCI--V----V-V--Y-AW-L--
         HHHHHH_TT_EEEE_SSSHHHHHHHHHHHH_
```

E

```
heavy (H) and light (L) variable domain, residues 1-60 and 87-119
       <....;....1....;....2....;....3....;....4....;....5....;....6> <...9....;....0....;....1....;....>
aida-L <DIVLTQSPAIMSASPGEKVTMTCSASSSVSQMIWFQQKSGTSPKRWIQNTSKLASGVPVR> <CQQMSSNPTFGGGTKLEIKRADAAPTVSIFPPS>
2HFL-L <------------------------------NY-Y-Y------------YD----------> <---WGR--------------------------->
       <__EEEE_SEEEE_TT__EEEEEEEEEES__EEEEE_SSS__EEEEETTTEE_TT__TT> <EE__SSS__BB__EEEEE____B__EEEEE___>

       <....;....1....;....2....;....3....:....4....:....5....:....6> <...9....;....0....:....1....;....>
aida-H <QVQLQQSGAELMKPGASVKISCKASGYTFSNSHIDWVKQRPGHGLEWIGVINPGSGSTRY> <TSEDSGVYYCLHGHYQFDGWGQGTTLTVSSAKT>
2HFL-H <------------------------------DYW-E--------------E-L------N-> <-------------N-D----------------->
       <__EEEE___EEE_TT__EEEEEEEEESS_GGGS_EEEEEE_TTS_EEEEEEEEGGGTEEEE> <_STT_EEEEEE__SSS___B___EEEEE_SS__>
```

Fig. 1. Amino acid sequences of the designed proteins. fing, FingerClasp (**A**); shpi, Shpilka (**B**); gren, Grendel (**C**); leat, Leather (**D**); and aida, Aida (**E**). Below the sequences, secondary structure are shown as extracted from the model 3-D coordinates by the program DSSP[25]: H,G, helix; E,B, β (extended) structure; T, H-bonded turn; S, bend; —, extended loop; !, chain break (the loops between the four helices in Grendel were not modeled). Finger, Grendel, and Shpilka are de novo designed sequences. Leather is constructed from rearranged regions of lactate dehydrogenase (Protein Data Bank code 1LDM) plus two new sequence pieces: 1LDM residues 242-261/NGGVN/1LDM 22-72/GG/1LDM 81-152, followed by point mutations, where / means concatenation. In line 1LDM, residues kept in Leather are marked by—while mutated ones are given in one letter code. Aida is the result of designed point mutations in the antilysozyme antibody (Protein Data Bank code 2HFL), so only sequence regions which contain mutations relative to the original antibody are given (1–60 and 87–119 of both light and heavy variable domains); residues 61–86 in both domains of Aida are exactly as in 2 HFL. In line 2HFL, residues kept in Aida are marked by —, while mutated residues are given in one letter code. The protein coordinates can be obtained as electronic mail from netserv@EMBL-Heidelberg.DE on Internet (send the messages "help" and "dir proteindata").

5. edge strands that are restricted in backbone H-bond formation on one specified side; and
6. close packing of side chains in the interior.

The designed sequence of Shpilka (Fig. 1B) contains eight regions enriched in β-forming residues, separated by β-breaking connections. Nonpolar and polar residues alternate to form well-defined inner and outer surfaces for β-strands but not for helices. Inner (nonpolar) surfaces of the strands contain Val, Ile, and Met residues, the best β-formers. The inner surfaces of edge strands contain β-forming Thrs, which are partly polar, making contact with the hydrophobic core as well as with solvent. The outer surfaces of all strands include charged groups that can form salt bridges. Pros, which have no backbone NH group, are incorporated into the edge strands so that only one side of an edge strand can form a continuous hydrogen bond net.

Turns and loops are made of β-breaking residues; to help ensure unambiguous folding, they have minimal length, just enough to connect the strands in the designed structure. A Trp was included on the interior face of the sheet to serve as an experimental marker for correct folding. At the inner surfaces of internal strands bulky side chains alternate with small Alas in order to form complementary surfaces for close packing. We did our best to improve the close packing by "point mutations" using computer graphics software as well as CPK models. According to the theory described in refs. 12 and 13, the designed sequence would form a β-sheet sandwich with the desired fold.

GRENDEL: A FOUR HELIX MEMBRANE ANCHOR

The recent determination of two membrane protein structures by X-ray crystallography (the photoreaction center[14]) and electron microscopy (bacteriorhodopsin BRH),[15] together with progress in understanding the basis of their stability and bility or critical during folding. Alternatively, they may just be the remains of previous evolution, without a critical role. If parts could be removed, functional molecules of a more manageable and more economical size would be obtained.

To this end we looked for a protein system where a ligand binding site was defined by residues from relatively few secondary structural units. We chose lactate dehydrogenase[21] as the basis for a minimal NAD binding protein. Natural NAD binding sites have been probed with compounds mimicking parts of the NAD molecule, interacting with only part of the extended binding site. The same compounds could be useful in testing a designed protein that failed to bind NAD to investigate if at least parts of the binding site were present.

To obtain the minimal design, the C-terminal (catalytic) domain and one β–α unit of the N-terminal (NAD-binding) domain were removed (Fig. 2D). However, a C-terminal α-helix, essential for the binding site, was retained by reconnecting it to the old N-terminus using a new loop. A disulfide bond was introduced to stabilize the other end of this helix (the new N-terminus). Residues newly exposed to solvent were mutated to "solubilize" the protein and a single Trp residue was incorporated into the hydrophobic core of the protein to act as a spectroscopic probe for folding. The final designed sequence has 131 residues compared to 329 for lactate dehydrogenase.

AIDA: AMERICAN–ITALIAN DESIGNER ANTIBODY

Can we predict sequences of antibodies that will bind specified epitopes on proteins?

The antigen-binding sites of immunoglobulins are created by six loops between strands of β-sheet in the variable domains of light and heavy chains. At least five of the six loops show a limited range of mainchain conformations called canonical structures;[22,23] these main chain conformations are determined by a few particular residues in the sequence. Other residues in the loops are relatively free to vary to create a variety of surface topographies and charge distributions in the binding site without changing the main chain conformations of the loops.

Because we know which residues we can change and which we must leave alone, in order to retain the main chain conformation of the antigen binding site, we can attempt to design modifications of a known antibody–antigen interface to substitute one antigen for another. In this exercise we started with the known structure of the Fab HyHEL-5/Hen egg white lysozyme complex,[24] and tried to substitute flavodoxin for lysozyme.

We began by placing the flavodoxin in the same region of space, relative to the antibody, that lysozyme occupied. Then we stripped off the side chains of the residues in the antibody binding site, and tried to rebuild the interface. The goal was a well-fitting interface which would form without change in main chain conformation of either protein. The techniques used included careful study of the interface, at each stage, using computer graphics software, followed by many cycles of mutations and refinement by energy minimization.

The project allowed us to evaluate the difficulty of designing an interface complementary to an antigen. The goal is of practical interest because, if successful, it would make is possible to design an antibody against a pathogen. In this respect the design techniques may take their place in the general context of artificial antibody construction, although it will certainly be a long time before theoretical methods can compete with the ability of the immune system to create mature antibodies of finely tuned affinity and specificity.

ACKNOWLEDGMENTS

We wish to thank the European Molecular Biology Organization as well as Silicon Graphics, Evans and Sutherland, Hoechst AG, and Hoffmann-LaRoche for financial and material support; Reinhard Schneider, Ulrike Goebel, and John Priestle for ribbon plots; and the EMBL Computer Group for system support. A.L. thanks the Kay Kendall Foundation for support. Authors are from various laboratories in Europe, the U.S.A. and Japan (addresses not given). Working groups were as follows. Fingerclasp: F. Bazan, A. Horovitz, H. Nakamura and L. Ribas; Shpilka: A.V. Finkelstein, A. Lockhart, R. Merkl, L.J. Perry; Grendel: S. Emery, C. Gaboriaud, C. Marks, J. Moult, C. Verlinde; Leather: M. Eberhard, A. Elofsson, T.J.P. Hubbard, L. Regan; Aida: J. Banks, R. Jappelli, A.M. Lesk, A. Tramontano. Teachers were A.V. Finkelstein, T. Hubbard, A. Lesk, J. Moult, H. Nakamura, C. Sander, A. Tramontano, and G. Vriend.

REFERENCES

1. Richardson, J., Richardson, D.C. The de novo design of protein structures. TIBS 163:304–309, 1989.
1a. Sander, C. De novo design of proteins. Curr. Opin. Struc. Biol. 1:630–637, 1991.
2. Protein Design Exercises 86, EMBL BIOcomputing Technical Document 1, C. Sander, ed., 1987.
3. Protein Design 90, EMBL BIOcomputing Technical Document 6, C. Sander and G. Vriend, eds., 1991.
4. Gregoret, L.M., Cohen, F.E. Novel method for the rapid evaluation of packing in protein structures. J. Mol. Biol. 211:959–974, 1990.
5. Novotny, J., Bruccoleri, R., Karplus, M. An analysis of incorrectly folded protein models. J. Mol. Biol. 177:787–818, 1984.
6. Baumann, G., Froemmel, C., Sander, C. Polarity as a criterion in protein design. Protein Eng. 2:329–334, 1989.
7. Bjorkman, P.J., Saper, M.A., Samraoui, B., Bennett, W.S., Strominger, J.L., Wiley, D.C. Structure of the human class I histocompatibility antigen HLA-H2. Nature (London) 329:506–512, 1987.
8. St. Charles, R., Walz, D.A., Edwards, B.F.P. The three-dimensional structure of bovine platelet factor 4 at 3.0 Å resolution. J. Biol. Chem. 264:2092–2099, 1989.

PROTEINS: Structure, Function, and Genetics 12:105–110 (1992)

Protein Design on Computers. Five New Proteins: Shpilka, Grendel, Fingerclasp, Leather, and Aida

Chris Sander, Gerrit Vriend, Fernando Bazan, Amnon Horovitz, Haruki Nakamura, Luis Ribas, Alexei V. Finkelstein, Andrew Lockhart, Rainer Merkl, L. Jeanne Perry, Stephen C. Emery, Christine Gaboriaud, Cara Marks, John Moult, Christophe Verlinde, Marc Eberhard, Arne Elofsson, Tim J.P. Hubbard, Lynne Regan, Jay Banks, Roberto Jappelli, Arthur M. Lesk, and Anna Tramontano
European Molecular Biology Laboratory, D-6900 Heidelberg, Federal Republic of Germany

ABSTRACT **What is the current state of the art in protein design? This question was approached in a recent two-week protein design workshop sponsored by EMBO and held at the EMBL in Heidelberg. The goals were to test available design tools and to explore new design strategies. Five novel proteins were designed: Shpilka, a sandwich of two four-stranded β-sheets, a scaffold on which to explore variations in loop topology; Grendel, a four-helical membrane anchor, ready for fusion to water-soluble functional domains; Fingerclasp, a dimer of interdigitating β–β–α units, the simplest variant of the "handshake" structural class; Aida, an antibody binding surface intended to be specific for flavodoxin; Leather—a minimal NAD binding domain, extracted from a larger protein. Each design is available as a set of three-dimensional coordinates, the corresponding amino acid sequence and a set of analytical results. The designs are placed in the public domain for scrutiny, improvement, and possible experimental verification.**

Key words: protein structure, protein sequences, protein design de novo, protein engineering, computer algorithms

INTRODUCTION

The forward protein folding problem, that of calculating structure from sequence, remains basically unsolved. The inverse problem, that of designing sequences to achieve desired structural properties, has been the subject of considerable effort over the last few years (for reviews, see refs. 1 and 1a). We have attempted to promote the development of new techniques and to encourage the design of new types of proteins by organizing intensive workshops, first in 1986[2] and, more recently, in 1990.[3]

The problem posed to the participants was this: specify a model protein structure (or structural property) and then invent a protein sequence that will lead to the desired structure. In the same spirit, in 1986 two idealized 4*βαβα barrel structure scaffolds (Babarellin, Tiny Tim), two β/α folds (Betalphacin, Idealized Flavodoxin), a bundle of four α-helices (Bundle), and a Cu-binding variant of a natural protein (CuRop) were constructed, and the corresponding protein sequences invented.[2] Since then, Babarellin and CuRop have been synthesized and purified (G. Nyakatura, H.-J. Fritz, and S. C. Emery, personal communications) and structural tests are in progress (F.X. Schmid, W. Eberle, J. Richardson, and M. Sagermann, personal communications). The five new proteins designed in the recent 1990 workshop are described below by each of the working groups.

TECHNIQUES USED

Typically, the design procedure followed these steps: (1) analyze known protein structures and identify structural units that might be used as building blocks, e.g., αβ units; (2) sketch out the secondary structure elements, their relative orientations, and the topology of loop connections, e.g., a four-stranded antiparallel β-sheet packed against two α-helices crossing the strands; (3) construct the protein scaffold by building explicit backbone coordinates, first for the structural core, consisting primarily of elements of secondary structure, then for loops; (4) choose an appropriate amino acid sequence in interior and surface regions, e.g., Glu on the surface of a helix near the N-terminus, Val or Ile at a β–β or β–α interface and so on; (5) optimize the model in interactive mode using visual inspection or in automatic mode using molecular mechanics software, i.e., vary backbone and side chain degrees of freedom, with simple energetics as a guide; primary goals are to regularize covalent geometry, remove clashes, avoid holes, optimize hydrogen bonding as well as charge–charge and protein–solvent interactions; (6) check the quality of the model by analyzing, e.g., solvation, electrostatic, or interior packing

Received April 16, 1991; accepted June 17, 1991.
Address reprint requests to Chris Sander, European Molecular Biology Laboratory, Meyerhofstrasse 1, D-6900 Heidelberg, Federal Republic of Germany.