Protein Design: A Hierarchic Approach
Author(s): James W. Bryson, Stephen F. Betz, Helen S. Lu, Daniel J. Suich, Hongxing X. Zhou, Karyn T. O'Neil, William F. DeGrado
Source: *Science*, New Series, Vol. 270, No. 5238 (Nov. 10, 1995), pp. 935-941
Published by: American Association for the Advancement of Science
Stable URL: http://www.jstor.org/stable/2888100
Accessed: 23/06/2009 14:17

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at http://www.jstor.org/page/info/about/policies/terms.jsp. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at http://www.jstor.org/action/showPublisher?publisherCode=aaas.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is a not-for-profit organization founded in 1995 to build trusted digital archives for scholarship. We work with the scholarly community to preserve their work and the materials they rely upon, and to build a common research platform that promotes the discovery and use of these resources. For more information about JSTOR, please contact support@jstor.org.

American Association for the Advancement of Science is collaborating with JSTOR to digitize, preserve and extend access to *Science.*

http://www.jstor.org

Protein Design: A Hierarchic Approach

James W. Bryson, Stephen F. Betz, Helen S. Lu, Daniel J. Suich, Hongxing X. Zhou,
Karyn T. O'Neil, William F. DeGrado*

The de novo design of peptides and proteins has recently emerged as an approach for investigating protein structure and function. Designed, helical peptides provide model systems for dissecting and quantifying the multiple interactions that stabilize secondary structure formation. De novo design is also useful for exploring the features that specify the stoichiometry and stability of α-helical coiled coils and for defining the requirements for folding into structures that resemble native, functional proteins. The design process often occurs in a series of discrete steps. Such steps reflect the hierarchy of forces required for stabilizing tertiary structures, beginning with hydrophobic forces and adding more specific interactions as required to achieve a unique, functional protein.

During protein folding, a random coil with a large number of rapidly interconverting conformers assembles into a fully folded and unique three-dimensional structure by molecular recognition. The side chains facing the interior of proteins are well-packed and almost invariably occupy single conformations; indeed, this precise orientation of functional groups is essential for function (for example, catalysis or signal transduction). Folded proteins come in sizes and shapes as numerous as their diverse functions. This complexity stands in contrast to the simplicity of the forces that determine their folds; clearly, the same van der Waals, electrostatic, hydrophobic, and hydrogen-bonded interactions that determine intermolecular recognition of small molecules also dictate the intramolecular folding process (*1*). Thus, considerable effort is now being directed to the de novo prediction of protein conformations with the use of computational methods (*2*) as well as the design of proteins from scratch (*3–5*). The lessons learned from the study of designed proteins depend on the rigor of their experimental characterization (*5*). Here, we review a hierarchic approach to protein design, which begins with highly simplistic minimalist (*3, 6, 7*) models and builds in complexity only as needed to achieve a uniquely folded, functional protein.

Helix Stabilization

The α helix represents one of the earliest recognized and fundamental building blocks used in the construction of natural proteins. Simple model systems have contributed much to our understanding of the forces stabilizing this secondary structure. Although in general peptides spanning the sequences of helices in proteins are nonhelical as monomers in water, it is possible to design peptides that form short α helices. Strategies used to create these helices include (i) the incorporation of many helix-stabilizing Ala residues (*8, 9*); (ii) adding salt bridges between residues separated by one α-helical turn (*8, 10–13*); (iii) incorporating covalent macrocycles (*14*); and (iv) adding nonpeptide templates to initiate helix formation (*15*). The sequences of these minimal models have been systematically varied to elucidate helix-stabilizing interactions.

The thermodynamic basis for the distinct α-helical propensities of amino acids has been evaluated by systematically varying solvent-exposed positions near the center of model helices (*10, 12, 16–20*). The results are generally expressed as the difference in the Gibbs free energy (ΔG°) for forming a helix with a given residue at the variable position, relative to the corresponding process with a reference amino acid. There is good agreement in both the rank order and magnitude of the helical propensities determined with different model systems, with the correlation coefficients ranging from 0.81 to 0.94 (Table 1). The difference between the most (Ala) and the least (Gly, excepting Pro) helix-stabilizing amino acid is approximately 1.0 kcal/mol for every scale, with one exception (Table 1) (*20*). The energetic range of the helical propensities (1 kcal/mol) is small by comparison to other interactions (Fig. 1), so it is tempting to assume that factors such as helix propensity are insignificant. However, helical propensities accumulate along the entire length of the helix, amounting to a sizable energetic driving force. These findings have stimulated numerous theoretical studies that

Table 1. Helix propensities of the individual amino acids. In general, there is good agreement between the experimental scales (±0.3 kcal/mol). The outliers presumably reflect contextual effects in either the native or unfolded states. Lys., lysate.

	Values for ΔΔ *G*° (kcal/mol) determined with different methods						
Amino acid	Statistical		Experimental in peptides		Experimental in proteins		Composite
	Helix*	PhiPsi*	Coiled coils†	Monomeric helices‡	T4 lys. (site 44)§	Barnase (site 32)‖	AGADIR¶
Ala	0	0	0	0	0	0	0
Arg	0.1	0.1	0.1	0.2	0.2	0.1	0.06
Asp	0.4	0.3	0.6	0.9	0.5	0.7	0.59
Asn	0.5	0.5	0.7	0.9	0.6	0.7	0.60
Cys	0.5	0.5	0.5	0.8	0.5	0.8	0.60
Gln	0.2	0.2	0.4	0.6	0.2	0.5	0.32
Glu	(−0.3)#	0.1	0.5	0.7	0.4	0.6	0.34
Gly	0.7	0.7	0.8	1.9	1.0	0.9	1.10
His	0.4	0.4	0.7	0.8	0.4	0.8	0.62
Ile	0.1	0.3	0.5	0.7	0.1	0.8	0.35
Leu	0.1	0.1	0.1	0.3	0.0	0.3	0.19
Lys	0.2	0.2	0.1	0.4	0.2	0.2	0.15
Met	0.0	0.1	0.3	0.5	0.1	0.3	0.21
Phe	0.3	0.4	0.4	0.9	0.4	0.7	0.47
Ser	0.5	0.4	0.4	0.8	0.4	0.4	0.52
Thr	0.5	0.4	0.7	1.3	0.4	0.8	0.57
Tyr	0.3	0.4	0.6	0.7	0.2	0.8	0.47
Trp	0.3	0.3	0.3	0.8	0.4	0.8	0.47
Val	0.3	0.3	0.6	1.0	0.3	0.9	0.51
Correlation to AGADIR	0.93	0.93	0.86	0.94	0.91	0.81	

*For description, see (*105*). †For description, see (*16*). ‡For description, see (*19*). §For description, see (*18*). ‖For description, see (*17*). ¶For description, see (*20*). #Glu was excluded from the correlation analysis for this scale.

The authors are at DuPont Merck Pharmaceutical Company, P.O. Box 80328, Wilmington, DE 19880, USA.

*To whom correspondence should be addressed. E-mail: degradwf@lldmpc.dnet.dupont.com

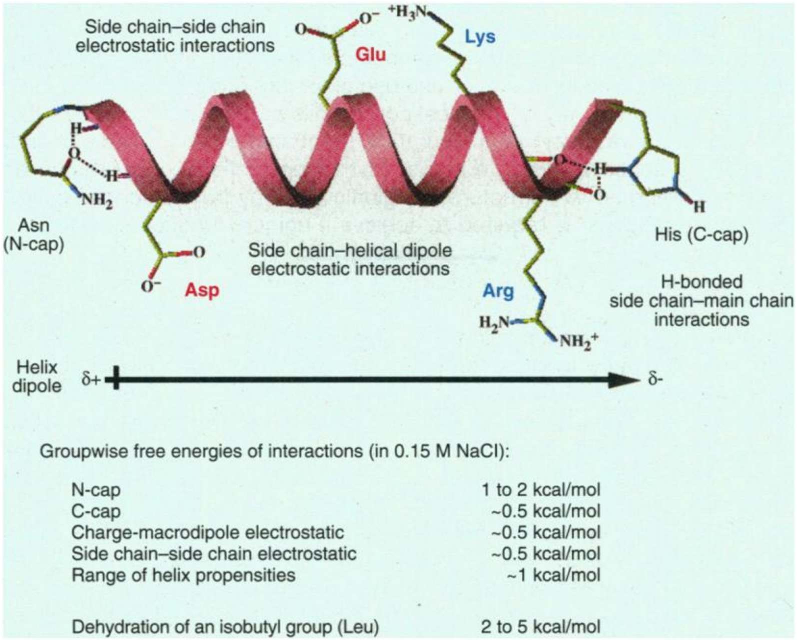
have shown that conformational entropy is a major determinant of the helical propensities of amino acids (*16, 21*). Helix formation also affects the solvent accessibility of amino acid side chains, which further affects helical propensities (*22*). The extent of this effect will vary depending on the structural and sequential context of the site (*23*) and can be evaluated with atomic solvation parameters (*24*).

The α helix has a large macrodipole (Fig. 1), and appropriately charged residues near the NH_2- and COOH-termini can electrostatically stabilize the structure by approximately 0.5 kcal/mol (*13, 20, 25*). Specific hydrogen-bonded interactions that "cap" the ends of helices (*26*) have been investigated with minimal models (*27*) and natural proteins (*28*) (Fig. 1). These interactions serve as punctuation marks in the syntax of protein folding. Ser, Asn, Gly, Asp, or Thr residues can hydrogen-bond to exposed amide protons at the NH_2-terminus of a helix, stabilizing it by 1 to 2 kcal/mol (*27*). In addition, hydrogen bonding and electrostatic interactions between amino acid side chains separated by a single α-helical turn stabilize a helix by up to 0.5 kcal/mol (*11, 13*).

Coiled Coils

The α-helical coiled coil (Fig. 2) represents a structure of intermediate complexity, bridging the gap between simple monomeric helices and native proteins. Coiled coils figure largely in the structures of fibrous proteins and certain DNA binding proteins and also serve as a convenient model system for the study of protein folding. Coiled coils have a seven-residue geometric repeat, which led Hodges and co-workers to design repeating heptapeptides as models for two-stranded coiled coils. In the prototype, $(Leu_aGlu_bAla_cLeu_dGlu_eGly_fLys_g)_n$, apolar Leu residues at positions "a" and "d" of the heptad hydrophobically stabilize the structure (*29*). Glu and Lys residues at "e" and "g" also stabilize the structure only if the helices pack parallel to one another. Antiparallel homodimers (*30*) or parallel heterodimers (*31, 32*) have also been engineered by manipulating the charges of residues at "e" and "g."

This heptad repeat formed the basis for the design of a 29-residue peptide (*16*) that was used to determine the helical propensities of various amino acids substituted at a solvent-exposed position of the helix (Table 1). Subsequent determination of the crystal structure of this peptide provided a trimeric structure (*33*) (Fig. 2B). As in the design, the trimer contains a parallel pair of α helices, but a third helix docks against this dimer in an antiparallel manner. This finding stimulated a reexamination of the solution behavior of the peptide, which had previously been assumed to be dimeric. The peptide changes its association state progressively from monomer to dimer to trimer as the concentration is increased from a very dilute solution to levels approaching those used in the crystallization (*34*). A careful determination of the helical propensities with the use of a monomer-dimer-trimer equilibrium resulted in a scale not significantly different from that published earlier (*34*).


Side chain–side chain electrostatic interactions
O
O⁻
⁺H₃N
Glu
Lys
H
O
H
NH₂
Asn (N-cap)
O
O⁻
Asp
Side chain–helical dipole electrostatic interactions
O
H
O
H
Arg
H₂N
NH₂⁺
His (C-cap)
H-bonded side chain–main chain interactions
Helix dipole
δ+
δ-

Groupwise free energies of interactions (in 0.15 M NaCl):

N-cap	1 to 2 kcal/mol
C-cap	~0.5 kcal/mol
Charge-macrodipole electrostatic	~0.5 kcal/mol
Side chain–side chain electrostatic	~0.5 kcal/mol
Range of helix propensities	~1 kcal/mol
Dehydration of an isobutyl group (Leu)	2 to 5 kcal/mol

Fig. 1. Idealized α helix, highlighting stabilizing intrahelical interactions. The α helix has a large macrodipole arising from the alignment of the amide bonds such that there is an effective +1/2 or −1/2 charge near the NH_2- and COOH-termini, respectively (*106*). Dehydration of an isobutyl group is included as an example of the hydrophobic effect for comparative purposes.

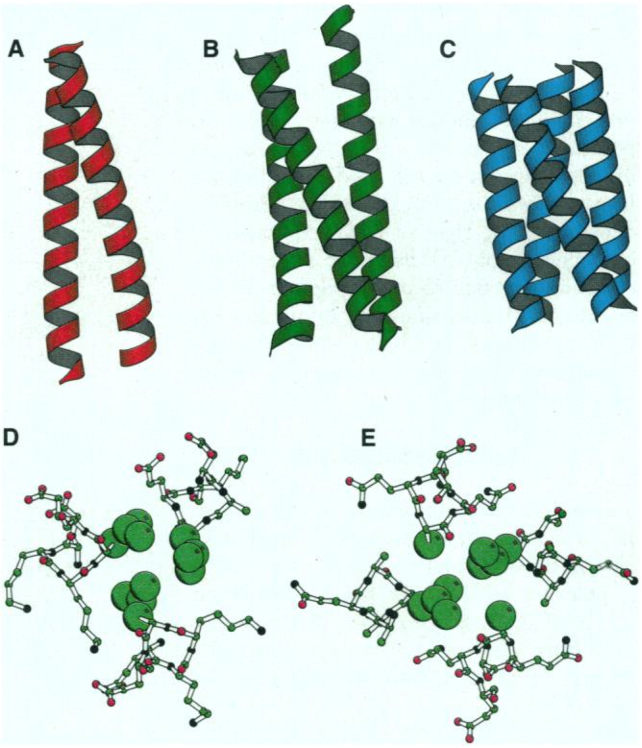


Fig. 2. MOLSCRIPT (*107*) diagrams of crystallographically determined α-helical bundles of associating peptides. Side chain packing interactions determine the geometry of two-, three-, and four-stranded helical bundles. The parallel two-stranded coiled coil GCN4 (**A**) contains Val at "a" and Leu at "d," as well as a critical Asn at a central "a" (*36*). The de novo–designated antiparallel coil-Ser trimer (**B**) contains Leu at both "a" and "d" (*33*). The antiparallel four-stranded α-helical bundle PD1 (**C**), is not strictly a coiled coil. However, near the center of the structure small (Ala) and large (Leu) residues at "a" and "d," respectively (*75*), alternate. (**D** and **E**) A representative 5 Å axial slice of coil-Ser (D) and a similar slice from PD1 (E). In (D) and (E), carbon atoms are green, nitrogen atoms are black, and oxygen atoms are pink. The side chain atoms of hydrophobic core residues are displayed as space-filling models with an atomic radius of 1.2 Å.

The free energies of the two association steps in the monomer-dimer-trimer equilibrium are about the same, which is indicative of a relatively noncooperative assembly process. However, specific changes to the sequence can stabilize either the three-helical or the two-helical state. For instance, addition of loops between the helices leads to single-chain, three-helix bundles (*35*). A noncovalent dimer has also been engineered by consideration of the two-stranded coiled coil (*36*) from GCN4, which contains a hydrophilic Asn residue on the hydrophobic surface of its helices. Asn at an equivalent position of the designed peptide similarly specifies dimers. Conversely, changing this Asn residue to Val in GCN4 led to trimers (*37, 38*). In crystal structures of GCN4, this Asn side chain forms a hydrogen bond between the carboxamide proton of one monomer and the carboxamide carbonyl of its interhelical neighbor. Importantly, the remaining polar functionality of each Asn residue remains accessible for hydration by water. However, in a trimer the Asn residue would be fully buried without the formation of hydrogen bonds to all its polar functional groups. Thus, this break in hydrophobic periodicity provides specificity to the structure, albeit at the price of thermodynamic stability as this substitution destabilizes the free energy of dimerization by about 3 kcal per mole of monomer (*34, 39*). Other substitutions that affect the hydrophobic periodicity appear to have similar effects (*39, 40*).

Kim, Alber, and co-workers have studied variants of the GCN4 coiled coil in which each "a" or "d" was replaced by a single amino acid. Peptides with Leu at "d" and Val or Leu at "a" exist in a monomer-dimer-trimer equilibrium (*37*). By contrast, Ile at "d" and Leu at "a" specifies tetramers, whereas Ile at both "a" and "d" specifies trimers (*37, 38*). The specific steric properties of the amino acids dictate the aggregation state (*37, 38*); each peptide adopts an aggregation state that provides the best side chain packing while simultaneously maintaining low-energy side chain conformations. The residues at positions other than "a" and "d" also help to determine the aggregation state. For example, peptides with Leu at "a" and "d" form dimers (*31*), trimers (*33*), tetramers (*35, 41*), pentamers (*42*), or hexamers (*43*), depending on the hydrophobicity and steric properties of the residues at "e" and "g." Thus, hydrophobicity can drive the formation of secondary structure and self-assembly, but the specific topology depends on side chain packing, hydrogen bonding, and elements included to selectively destabilize alternative folds.

Various researchers have also designed monomolecular coiled coils or helical pairs. Two-stranded structures can be entropically stabilized by forming a disulfide between opposing Cys residues at "a" or "d" (*40*) or by linking two helices via peptide linkers (*44*). The low-resolution solution structure of a de novo–designed 38-residue peptide with an antiparallel pair of helices (*45*) has been determined by nuclear magnetic resonance (NMR) spectroscopy. Similarly, the solution structure of an antiparallel helical dimer ALIN has also been determined by NMR techniques (*46*).

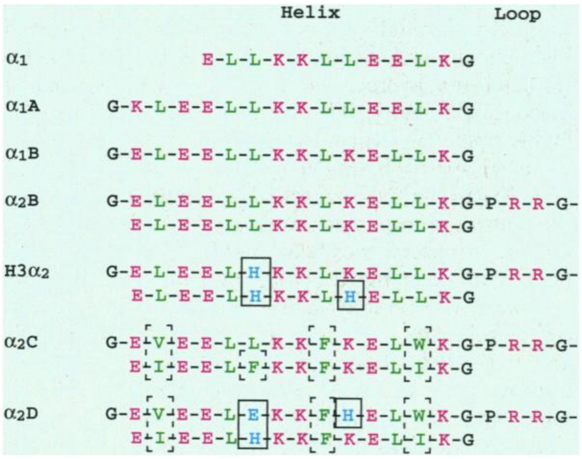
Design of Four-Helix Bundles

The simplicity, approximate symmetry, and functional diversity of the four-helix bundle—found in cytochromes, lymphokines, enzymes, and phage coat proteins (*47*)—has made it an attractive target for such protein design. A class of four-helix bundles referred to as TASP proteins (*48*) (template-assembled synthetic peptides) has been designed in which four helical peptides are attached to a macrocyclic template, such as cyclic peptides (*48*) or porphyrin derivatives (*49, 50*). A number of TASP proteins have been prepared (*51*), including models for protein folding (*52*), ion channels (*50, 53*), and redox-active porphyrin derivatives (*49, 54*). New methods (*54, 55*) for ligating peptides to templates allow the construction of asymmetric arrays of either parallel or antiparallel helices, greatly expanding the complexity of structures that can be engineered.

Whereas TASP proteins use a rigid template to force the helices into juxtaposition, natural proteins use weak, noncovalent forces to specify their fold. In an attempt to mimic this process, we, in collaboration with Eisenberg and co-workers, designed a series of sequences (Fig. 3) that self-assemble into four-helix structures, including tetrameric helices (designated α_1A and α_1B), helix-loop-helix dimers (α_2), and single-chain, four-helix proteins (α_4) (*41, 56, 57*). These minimal abstractions of natural four-helix bundles contain Leu as the sole hydrophobic residue comprising the apolar core. More specific long-range interactions such as interhelical hydrogen bonding or buried salt bridges were not included.

One feature that helps to shape the geometry of the four-helix bundle is the length of the helices: helices of four or more turns pack optimally in elongated bundles, whereas shorter helices can pack into a number of other geometries (*58*). Therefore, we were interested in determining the structure of a 12-residue amphiphilic α-helical peptide (α_1; Fig. 3) that was isolated as a truncated side product during the course of synthesis of the longer helical peptides (*41*). In aqueous solution, this peptide equilibrates between random coil monomers and α-helical aggregates of tetramers, hexamers, and higher order aggregates (*35, 41, 59*). In crystals grown at low pH, it adopts a novel hexameric state stabilized by the packing of the hydrophobic Leu residues (Fig. 4) (*60*), whereas crystallization at pH 7 provides a different structure.

Fig. 3. Sequence progression of the α_1 and α_2 family of four-helix bundle proteins (*96*). The sequences are listed to illustrate the hierarchic, iterative design from simple to more complex molecules. Green residues are hydrophobic, pink residues are polar, and boxed blue residues represent designed metal-binding sites. Dashed boxes indicate apolar residues that have been engineered to replace the all-leucine core with a more native-like, well-packed interior. References for the sequences are as follows: α_1 and α_1A, (*56*); α_1B and α_2B, (*41*); H3α_2, (*69*); α_2C, (*71*); and α_2D, (*72*).


Helix
Loop
α1 E-L-L-K-K-L-L-E-E-L-K-G
α1A G-K-L-E-E-L-L-K-K-L-L-E-E-L-K-G
α1B G-E-L-E-E-L-L-K-K-L-K-E-L-L-K-G
α2B G-E-L-E-E-L-L-K-K-L-K-E-L-L-K-G-P-R-R-G-
E-L-E-E-L-L-K-K-L-K-E-L-L-K-G
H3α2 G-E-L-E-E-L-H-K-K-L-K-E-L-L-K-G-P-R-R-G-
E-L-E-E-L-H-K-K-L-H-E-L-L-K-G
α2C G-E-V-E-E-L-L-K-K-F-K-E-L-W-K-G-P-R-R-G-
E-I-E-E-L-F-K-K-F-K-E-L-I-K-G
α2D G-E-V-E-E-L-E-K-K-F-H-E-L-W-K-G-P-R-R-G-
E-I-E-E-L-H-K-K-F-K-E-L-I-K-G

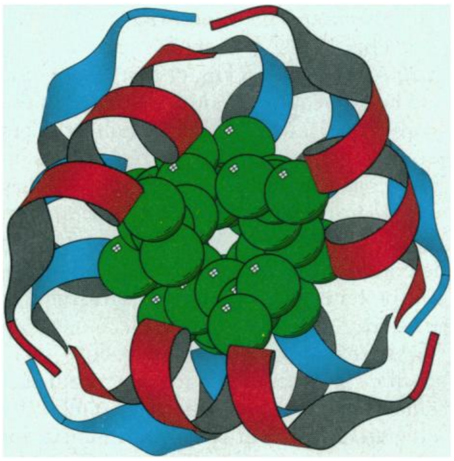
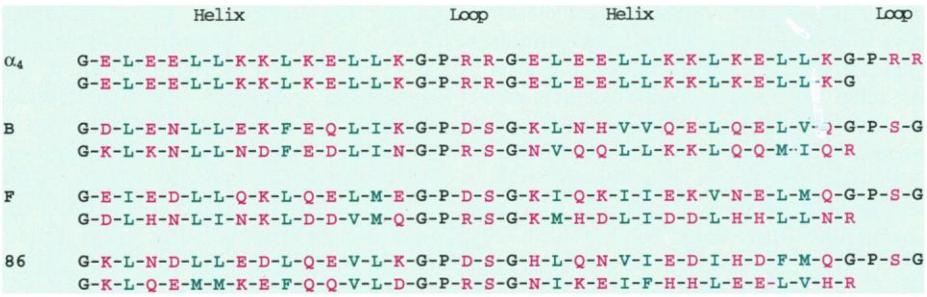


Fig. 4. MOLSCRIPT (*107*) diagram of the crystal structure of the 12-residue peptide α_1 (*60*) (see Fig. 3 for sequence). The hexameric bundle structure of α_1 illustrates the interior packing of leucine side chains in the hydrophobic core. The side chain carbon atoms are displayed as space-filling models with an atomic radius of 1.6 Å.



```
            Helix                   Loop          Helix                  Loop

α4    G-E-L-E-E-L-L-K-K-L-K-E-L-L-K-G-P-R-R-G-E-L-E-E-L-L-K-K-L-K-E-L-L-K-G-P-R-R
      G-E-L-E-E-L-L-K-K-L-K-E-L-L-K-G-P-R-R-G-E-L-E-E-L-L-K-K-L-K-E-L-L-K-G

B     G-D-L-E-N-L-L-E-K-F-E-Q-L-I-K-G-P-D-S-G-K-L-N-H-V-V-Q-E-L-Q-E-L-V-Q-G-P-S-G
      G-K-L-K-N-L-L-N-D-F-E-D-L-I-N-G-P-R-S-G-N-V-Q-Q-L-L-K-K-L-Q-Q-M-I-Q-R

F     G-E-I-E-D-L-L-Q-K-L-Q-E-L-M-E-G-P-D-S-G-K-I-Q-K-I-I-E-K-V-N-E-L-M-Q-G-P-S-G
      G-D-L-H-N-L-I-N-K-L-D-D-V-M-Q-G-P-R-S-G-K-M-H-D-L-I-D-D-L-H-H-L-L-N-R

86    G-K-L-N-D-L-L-E-D-L-Q-E-V-L-K-G-P-D-S-G-H-L-Q-N-V-I-E-D-I-H-D-F-M-Q-G-P-S-G
      G-K-L-Q-E-M-M-K-E-F-Q-Q-V-L-D-G-P-R-S-G-N-I-K-E-I-F-H-H-L-E-E-L-V-H-R
```

Fig. 5. Comparison of semirandom proteins (*67*) to α_4 (*96*). The color coding of the residues is the same as in Fig. 3. The free energy of denaturation of α_4 is 15 kcal/mol (*70*), whereas that of proteins B, F, and 86 is 1 to 5 kcal/mol. This stability gap can be partially accounted for by differences in helix propensity (≈4 to 6 kcal/mol; see Table 1) and possibly also by intrahelical electrostatic interactions (≈4 kcal/mol; see Fig. 1).

In contrast to α_1, the full-length α_1B peptide cooperatively assembles into four-helix bundles that show some of the characteristics of native proteins: the α_1B tetramer is compact and globular, and its backbone is structurally well defined and helical as assessed by NMR (*61*). However, between the conception of a designed protein and its realization lies the molten globule (*62*)—an energy well of surprising depth and breadth that must be overcome en route to the final goal (*5, 63*). Molten globules are nonnative states of proteins with dynamically averaging conformations, often containing poorly packed hydrophobic cores. Both the α_1B tetramer and the full-length α_4 protein show some characteristics of this state. Several other designed four-helical bundle proteins (*64*), including one protein, Felix (*65*), with a native-like sequence, also show some characteristics of the molten globule (*5, 63*).

In each of these designs, the association of hydrophobic side chains provides a powerful driving force for the formation and association of helices. This conclusion is consistent with lattice simulations, which show that sequences of the proper hydrophobic and hydrophilic periodicity can rapidly assume compactly folded structures (*66*). However, both lattice models as well as early design attempts lack the diversity of stabilizing interactions and specificity found in natural proteins, which we believe are essential for stabilizing native-like folds and function.

Toward this goal, Hecht and co-workers (*67*) introduced a combinatorial approach as a powerful tool for de novo protein design. They used mixed oligonucleotides to encode a library of partially random proteins with polar-nonpolar patterns similar to those used in α4. Hydrophobic side chains were drawn randomly from a pool consisting of Met, Val, Leu, Phe, or Ile; hydrophilic residues were Asn, Asp, Gln, Glu, His, or Lys; hydrophilic and helix-breaking residues were used in the loops (Fig. 5). Most of the partially random proteins were expressed at high levels in *Escherichia coli*, which suggests that they achieved structures protecting them from rapid proteolysis. It will be interesting to determine what percentage of the randomized proteins adopt native-like structures in solution.

Alternatively, one can introduce more specific interactions into a minimalist framework such as α_4, thereby establishing a hierarchic order of principles essential for achieving a unique fold. Regan and Clarke (*68*) introduced a Cys_2His_2, tetrahedral Zn^{2+}-binding site into α_4, and Handel and co-workers (*69, 70*) introduced one or two Zn^{2+}-binding sites into α_2 (Fig. 3) and α_4. These proteins were considerably less stable than α_4 in the absence of metal ions, but Zn^{2+} restored their thermodynamic stabilities while also decreasing the conformational mobility of the hydrophobic core. Thus, it has been possible to replace much of the nonspecific hydrophobic driving force with geometrically restrictive metal ion–ligand interactions. However, for all the improvements, H6α_4 still lacked a cooperative thermal transition indicative of a fully native conformation.

We also investigated the effect of varying the steric bulk of the interior side chains to introduce more geometric complementarity into the helix-helix interfaces. Seven simultaneous changes were made in the hydrophobic core of α_2B, resulting in a dimeric protein (α_2C; Fig. 3) that underwent a temperature-dependent transition from a native-like state to a molten globule-like state with a transition midpoint near room temperature (*71*). We next introduced a metal-binding site into this protein and found a surprise characteristic of progress in protein design: The resulting dimeric protein, α_2D, proved to be native-like, even in the absence of metal ions (*72*). Subsequent synthesis of α_2D analogs showed that the improvements were a consequence of changing two hydrophobic residues to hydrophilic residues (originally included to bind metal ions), which might destabilize unwanted folds in a manner similar to the above-mentioned Asn in position "a" of coiled coils. Thus, conformational specificity demands the correct balance of hydrophobic and hydrophilic interactions (*73*): too few hydrophobic residues lead to inadequate stability, but too many lead to highly stable but dynamic structures.

In an approach similar to de novo design, Regan and co-workers have idealized the entire hydrophobic interior (24 residues) of a natural four-helix bundle, ROP (*74*), while keeping the polar interactions invariant. This protein tends to have a heptad repeat with small hydrophobes at "a" and large hydrophobes at "d," and these residues stack in layers consisting of two small and two large side chains per stack. This tendency was idealized by making all the "a" and "d" residues Ala and Leu, respectively. The resulting protein is highly stable and behaves in all respects examined like a native protein. A similar pattern of Ala and Leu residues has recently been found in the crystal structure of a helical peptide (*75*) designed to solubilize membrane proteins (Fig. 2, C and E). These findings provide a paradigm for the design of uniquely folded, four-helix bundle proteins. The alternation of small and large residues at "a" and "d" should destabilize rotationally symmetrical parallel coiled coils; the Ala residues would group in layers, leading to cavernous holes. On the other hand, in antiparallel structures the small and large residues combine with considerable geometric complementarity: large Leu side chains pack into holes created by smaller Ala side chains.

β Sheet and Mixed α-β Structures

A number of experimental difficulties have beset the development of models for β sheet formation. Unlike α helices, where there is a regular succession of hydrogen bonds between amides four residues apart in sequence, β sheets are formed by residues at variable and often distant positions in the sequence. Also, the exposed amides at the edge of β sheets can hydrogen-bond to other sheets, leading to insoluble aggregates. Nevertheless, considerable progress has been made through the design of dibenzofurans and other templates that stabilize β sheets (*51, 76*). In addition several groups have modified natural β sheet proteins to obtain model systems for determining β sheet propensity scales (*77*). These scales appear to be more contextually dependent than the analogous α-helical scales, so their utility for protein design has yet to be established.

Betabellin was one of the first de novo designed proteins (*78*). It is intended to fold into a sandwich of two identical four-stranded, antiparallel β sheets. Iterative improvements on the original design increased the solubility of the protein, whereas type I′ β turns were introduced with the use of D-amino acids (*78*). Recently, Quinn and co-workers (*79*) designed betadoublet, which is similar to betabellin but contains only naturally encoded amino acids. The protein possesses a cooperative thermal denaturation and a well-dispersed amide NMR spectrum. However, its aliphatic NMR spectrum is less well dispersed, which suggests that its core is poorly ordered. The most recent version of betabellin also shows a cooperative thermal denaturation and a reasonably well dispersed NMR spectrum (*80*).

Pessi and co-workers made extensive modifications to the sequence of the variable heavy domain of the antibody McPC603 to create the "minibody" (*81*), a six-stranded β sandwich. The utility of this molecule as a scaffold for molecular recognition was demonstrated by its ability to selectively bind Cu^{2+} and Zn^{2+} over other metals. These studies on minibodies and betabellins suggest that the design of β sandwich proteins with unique, native-like conformations is a realizable goal that may be achieved through iterative cycles of design and structure determination. In the most ambitious design attempts to date, several groups have engineered α-β proteins (*82, 83*), two of which represent the familiar $(\beta\text{-}\alpha)_8$ barrels typified by triose phosphate isomerase (*83*). Biophysical characterization of the proteins indicates that they fold into compact, monomeric structures with the appropriate secondary structure, but they also display characteristics of the molten globule (*83*).

Functional Proteins

Recent progress in designing structural proteins has set the stage for the engineering of functional proteins. Unidirectional electron transfer is one particularly attractive function because of its biological importance and mechanistic simplicity (*84*). Minimal requirements for design of an electron transfer protein include the binding of multiple redox-active centers in a predetermined array; tuning the midpoint potential of these centers; and controlling the rigidity and solvent accessibility of the environment to minimize the reorganization energy.

Dutton and co-workers have begun designing minimalist peptides, which, like the small, three-dimensional maquettes used by architects during the design of full-scale buildings, are intended to serve as stepping stones en route to the construction of molecular machines. In initial work, a heme-binding site with bis-His ligation was introduced into the interior of a dimeric four-helix bundle based on α_2 (*85*). The resulting protein binds protoporphyrin IX with a dissociation constant in the nanomolar range, and its redox and spectroscopic properties are close to those anticipated from the design. Studies with control peptides indicated that specific heme binding requires the presence of two strongly ligating His residues for coordinating iron and a hydrophobic cavity to accommodate the macrocycle. The initial designs show some of the characteristics of the molten globule state and will require iterative improvement to provide the rigid interiors characteristic of cytochromes.

Next, a 62-residue helix-loop-helix peptide that binds two hemes with the use of bis-histidine ligation was prepared (*86*). The peptide assembles into a dimeric four-helix bundle protein with four bound hemes. The spectroscopic properties of the model protein are consistent with the design, and redox titrations indicate that the hemes are electrochemically coupled. In related work, Suslick and Huffman have designed single α-helical peptides that contain a single His residue on the face of an amphiphilic α helix (*87*). These and related peptides (*88*) provide attractive systems for exploring the role of the protein matrix in electron transfer.

Designed ion channel peptides and proteins have also contributed to our understanding of the mechanism of transmembrane ion conductors (*89*). In particular, careful characterization of minimally designed channels composed of amphiphilic α helices has shown how the α-helical macrodipole—in the absence of other charged groups—can give rise to marked asymmetry of ion conduction (rectification) (*90*).

The design of enzymes provides an even greater challenge for protein design. Although several designed peptides and proteins have been reported to have substantial enzymatic activity (*91*), the structural basis for this activity has been established for only one peptide. Benner and co-workers designed a short, self-associating, Leu-Lys–rich peptide that accelerated the rate of decarboxylation of oxaloacetate by a factor of 10 to 100 (*92*). This peptide catalyzes decarboxylation by means of a Schiff's base intermediate between substrate and an amine with an electrostatically depressed pK_a (where K_a is the acid constant). The peptide is helical, and this secondary structure is important for activity. Recently, a combinatorial approach has been taken to improve this peptide, resulting in a fivefold increase in the rate accelerations (*93*). In other research, a four-helical bundle peptide was designed to mimic the activity of α-chymotrypsin. Although the initially reported high level of esterase activity was later shown to be in error, this protein does show some catalysis toward activated 4-nitrophenyl esters (*94*). In these two examples, the catalysts show many of the features of molten globule proteins, which suggests that they may be assuming a large ensemble of conformations, only a limited number of which are active. With an appropriate selection, it might be possible to evolve more active catalysts from such molten globule precursors through the accrual of random mutations that increasingly stabilize the catalytically active conformation.

DNA Binding Proteins

Minimalist design principles have been applied to the study of four classes of DNA binding proteins, including the basic region, leucine zipper (bLZ) motif. The bLZ consists of a coiled coil segment responsible for dimerization and a helical basic region responsible for DNA binding. We designed an idealized bLZ peptide based on a prediction of the structure of a DNA-GCN4 complex (*95*). The coiled coil consisted of the repeating polyheptapeptide $(LEEKLKA)_4$ (*96*). The basic region contained the features required for specific binding to DNA: four residues thought to contact the bases were retained; Arg was placed at each position believed to electrostatically interact with the phosphodiester backbone; and the helix-stabilizing residues Glu and Ala were placed at solvent-exposed sites. The peptide indeed bound with high affinity to DNA, and subsequent crystallographic studies of GCN4 bound to DNA supported the main features of the predicted model (*97*). Fisher and co-workers applied a similar approach to map out the residues believed to be important for DNA binding in the related basic helix-loop-helix (bHLH) motif, which consists of a basic DNA binding region followed by a helix-loop-helix dimerization motif (*98*). Again, crystallographic studies of the bHLH domain of USF bound to DNA confirmed the essential features inferred from construction of minimal models (*99*).

Minimalist design has also been applied to probe the role of specific residues in the folding and function of DNA binding motifs of known structure. Berg and co-workers (*6*) designed a model zinc finger, a motif whose conformation is determined by interactions between conserved His and Cys residues and a tetrahedrally ligated Zn^{2+} ion. A Cys_2His_2 peptide was designed and synthesized, and its metal binding and structural properties were characterized (*6*). This 26-residue peptide was based on a polyalanine sequence and contained only seven of the most conserved residues for TFIIIA-type zinc fingers. The designed peptide exhibited properties similar to those of native zinc fingers, which demonstrated that all of the information required for correct folding and metal coordination is contained within the seven conserved residues. However, the peptide lacked specificity as it formed both 2:1 and 1:1 complexes with metal ions.

Similarly, initial attempts to design minimal peptides based on the homeodomain motif resulted in structures with fluctuating tertiary structures (*7*). Subsequent designs served to identify several exposed residues that are important for specific folding and DNA binding.

Other researchers have used the zinc finger motif as a framework for creating proteins with novel DNA binding preferences. Because each finger module recognizes three base pairs, it is theoretically possible to identify zinc fingers that bind each of the 64 possible codons. It would then be possible to design multimeric zinc fingers to bind to any DNA sequence (*100*). With this goal in mind, a prototype three finger sequence from Zif268 was fused to pIII, a minor coat protein from bacteriophage M13 (*101*). After the residues known to be important for DNA binding in one finger module were randomized, phages displaying mutant fingers were selected for their ability to bind to novel DNA sequences. In a particularly elegant contribution, Klug and co-workers (*102*) created a three–zinc finger protein that bound a nine–base pair target sequence spanning the fusion point of p190$^{BCR\text{-}ABL}$. In vivo expression of the peptide blocked transcription of DNA containing the target sequence.

Outlook

De novo protein design has provided a powerful methodology for investigating protein folding. The gross features that direct the collapse of the protein chain into compact structures have been demonstrated, and the subtleties that specify unique structures are being elucidated. A variety of uniquely folded helical bundles have been designed; similar progress on other folds will certainly follow. As designers turn their attention to functional proteins with less symmetry, the blending of combinatorial approaches with rational design should provide impressive results. For instance, one might convert a minimalist framework to a redox active protein by first introducing side chains that ligate a desired transition metal ion. The ligating side chains could then be stabilized in the appropriate conformation through interactions with surrounding residues introduced with an experimental (*103*) or computational (*104*) combinatorial approach. In a similar manner, it is easy to envision the design of sites capable of binding transition-state analogs or proteins of therapeutic importance.

REFERENCES AND NOTES

1. T. E. Creighton, *Biochem. J.* **270**, 1 (1990).
2. J. Skolnick, A. Kolinski, C. L. Brooks III, A. Godzik, A. Rey, *Curr. Biol.* **3**, 414 (1993); J. U. Bowie and D. Eisenberg, *Curr. Opin. Struct. Biol.* **3**, 437 (1993); R. Srinivasan and G. D. Rose, *Proteins Struct. Funct. Genet.* **22**, 81 (1995).
3. W. F. DeGrado, Z. R. Wasserman, J. D. Lear, *Science* **243**, 622 (1989).
4. J. S. Richardson and D. C. Richardson, *Trends Biochem. Sci.* **14**, 304 (1989).
5. W. F. DeGrado, D. P. Raleigh, T. Handel, *Curr. Opin. Struct. Biol.* **1**, 984 (1991); S. F. Betz, D. P. Raleigh, W. F. DeGrado, *ibid.* **3**, 601 (1993); S. F. Betz, J. W. Bryson, W. F. DeGrado, *ibid.*, in press.
6. S. F. Michael, V. J. Kilfoil, M. H. Schmidt, B. T. Amann, J. M. Berg, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 4796 (1992).
7. Z. Shang *et al.*, *ibid.* **91**, 8373 (1994).
8. S. Marqusee, V. H. Robbins, R. L. Baldwin, *ibid.* **86**, 5286 (1989).
9. J. M. Scholtz, E. J. York, J. M. Stewart, R. L. Baldwin, *J. Am. Chem. Soc.* **113**, 5102 (1991).
10. P. C. Lyu, M. I. Liff, L. A. Marky, N. R. Kallenbach, *Science* **250**, 669 (1990).
11. J. M. Scholtz, H. Qian, V. H. Robbins, R. L. Baldwin, *Biochemistry* **32**, 9668 (1993).
12. S. H. Park, W. Shalongo, E. Stellwagen, *ibid.*, p. 7048.
13. B. M. P. Huyghues-Despointes, J. M. Scholtz, R. L. Baldwin, *Protein Sci.* **2**, 80 (1993).
14. A. M. Felix *et al.*, *Int. J. Pept. Protein Res.* **32**, 441 (1988); D. Y. Jackson, D. S. King, J. Chmielewski, S. Singh, P. G. Schultz, *J. Am. Chem. Soc.* **113**, 9397 (1991); C. Bracken, J. Gulyás, J. W. Taylor, J. Baum, *ibid.* **116**, 6431 (1994).
15. D. S. Kemp, J. G. Boyd, C. C. Muendel, *Nature* **352**, 451 (1991).
16. K. T. O'Neil and W. F. DeGrado, *Science* **250**, 646 (1990).
17. A. Horovitz, J. M. Matthews, A. R. Fersht, *J. Mol. Biol.* **227**, 560 (1992).
18. M. Blaber *et al.*, *ibid.* **235**, 600 (1994).
19. A. Chakrabartty, T. Kortemme, R. L. Baldwin, *Protein Sci.* **3**, 843 (1994).
20. V. Muñoz and L. Serrano, *J. Mol. Biol.* **245**, 275 (1995).
21. M. Fukugita, T. Nakzawa, H. Kawai, Y. Okamoto, *Chem. Lett.* **1991**, 1279 (1991); R. H. Yun and J. Hermans, *Protein Eng.* **4**, 761 (1991); J. Hermans, A. G. Anderson, R. H. Yun, *Biochemistry* **31**, 5646 (1992); T. P. Creamer and G. D. Rose, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 5937 (1992); T. P. Creamer and G. D. Rose, *Proteins* **19**, 85 (1994).
22. M. Blaber, X.-j. Zhang, B. W. Matthews, *Science* **260**, 1637 (1993).
23. S. Padmanabhan and R. L. Baldwin, *Protein Sci.* **3**, 1992 (1994).
24. D. Eisenberg and A. D. McLachlan, *Nature* **319**, 199 (1986).
25. K. M. Armstrong and R. L. Baldwin, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 11337 (1993); H. Nicholson, W. J. Becktel, B. W. Matthews, *Nature* **336**, 651 (1988); H. Nicholson, D. E. Anderson, S. Dao-pin, B. W. Matthews, *Biochemistry* **30**, 9816 (1991); J. Sancho, L. Serrano, A. R. Fersht, *ibid.* **31**, 2253 (1992); D.-J. Lockhart and P. S. Kim, *Science* **260**, 198 (1993); C. R. Robinson and S. G. Sligar, *Protein Sci.* **2**, 826 (1993).
26. L. G. Presta and G. D. Rose, *Science* **240**, 1632 (1988); J. S. Richardson and D. C. Richardson, *ibid.*, p. 1648; D. F. Stickle, L. G. Presta, K. A. Dill, G. D. Rose, *J. Mol. Biol.* **226**, 1143 (1992); E. T. Harper and G. D. Rose, *Biochemistry* **32**, 7605 (1993); R. Aurora, R. Srinivasan, G. D. Rose, *Science* **264**, 1126 (1994).
27. P. Lyu, H. X. Zhou, N. Jelveh, D. E. Wemmer, N. R. Kallenbach, *J. Am. Chem. Soc.* **114**, 6560 (1992); P. C. Lyu, D. E. Wemmer, H. X. Zhou, R. J. Pinker, N. R. Kallenbach, *Biochemistry* **32**, 421 (1993); B. Forood, E. J. Feliciano, K. P. Nambiar, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 838 (1993); A. J. Doig, A. Chakrabartty, T. M. Klingler, R. L. Baldwin, *Biochemistry* **33**, 3396 (1994); A. J. Doig and R. L. Baldwin, *Protein Sci.* **4**, 1325 (1995).
28. L. Serrano and A. R. Fersht, *Nature* **342**, 296 (1989); M. D. Bruch, M. M. Dhingra, L. M. Gierasch, *Proteins* **10**, 130 (1991); J. T. J. Lecomte and C. D. Moore, *J. Am. Chem. Soc.* **113**, 9663 (1991); J. A. Bell, W. J. Becktel, U. Sauer, W. A. Baase, B. W. Matthews, *Biochemistry* **31**, 3590 (1992); E. A. Zhukovsky, M. G. Mulkerrin, L. G. Presta, *ibid.* **33**, 9856 (1994).
29. S. Y. M. Lau, A. K. Taneja, R. S. Hodges, *J. Biol. Chem.* **259**, 13253 (1984).
30. O. D. Monera, N. E. Zhou, C. M. Kay, R. S. Hodges, *ibid.* **268**, 19218 (1993).
31. E. K. O'Shea, K. J. Lumb, P. S. Kim, *Curr. Biol.* **3**, 658 (1993).
32. O. D. Monera, C. M. Kay, R. S. Hodges, *Biochemistry* **33**, 3862 (1994).
33. B. Lovejoy *et al.*, *Science* **259**, 1288 (1993).
34. S. Betz, R. Fairman, K. O'Neil, J. Lear, W. DeGrado, *Philos. Trans. R. Soc. London Ser. B.* **348**, 81 (1995).
35. W. F. DeGrado, R. Fairman, D. P. Raleigh, S. F. Betz, J. W. Bryson, unpublished data.
36. E. K. O'Shea, J. D. Klemm, P. S. Kim, T. Alber, *Science* **254**, 539 (1991).
37. P. B. Harbury, T. Zhang, P. S. Kim, T. Alber, *ibid.* **262**, 1401 (1993).
38. P. B. Harbury, P. S. Kim, T. Alber, *Nature* **371**, 80 (1994).
39. H. Wendt, C. Berger, A. Baici, R. M. Thomas, H. R. Bosshard, *Biochemistry* **34**, 4097 (1995).
40. N. E. Zhou, C. M. Kay, R. S. Hodges, *ibid.* **32**, 3178 (1993).
41. S. P. Ho and W. F. DeGrado, *J. Am. Chem. Soc.* **109**, 6751 (1987).
42. R. Lutgring and J. Chmielewski, *ibid.* **116**, 6451 (1994); E. Kitakuni, Y. Kuroda, M. Oobatake, T. Tanaka, H. Nakamura, *Protein Sci.* **3**, 831 (1994).
43. T.-M. Chin, K. D. Berndt, N. C. Yang, *J. Am. Chem. Soc.* **114**, 2279 (1992).
44. D. G. Myszka and I. M. Chaiken, *Biochemistry* **33**, 2363 (1994).
45. Y. Fezoui, D. L. Weaver, J. J. Osterhout, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 3675 (1994).
46. Y. Kuroda, T. Nakai, T. Ohkubo, *J. Mol. Biol.* **236**, 862 (1994).
47. S. R. Presnell and F. E. Cohen, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 6592 (1989); N. L. Harris, S. R. Presnell, F. E. Cohen, *J. Mol. Biol.* **236**, 13 (1994).
48. M. Mutter and S. Vuillevmier, *Angew. Chem. Int. Ed. Engl.* **28**, 535 (1989).
49. T. Sasaki and E. T. Kaiser, *J. Am. Chem. Soc.* **111**, 380 (1989).
50. K. Åkerfeldt *et al.*, *ibid.* **114**, 9656 (1992).
51. J. P. Schneider and J. W. Kelly, *Chem. Rev.*, in press.
52. M. Mutter *et al.*, *J. Am. Chem. Soc.* **114**, 1463 (1992).
53. A. Grove, J. M. Tomich, T. Iwamoto, M. Montal, *Protein Sci.* **2**, 1918 (1993); T. Iwamoto, A. Grove, M. O. Montal, M. Montal, J. M. Tomich, *Int. J. Pept. Protein Res.* **43**, 597 (1994).
54. C. T. Choma *et al.*, *Tetrahedron Lett.* **35**, 6191 (1994).
55. O. Nyanguile, M. Mutter, G. Tuchscherer, *Lett. Pept. Sci.* **1**, 9 (1994).
56. D. Eisenberg *et al.*, *Proteins* **1**, 16 (1986).
57. L. Regan and W. F. DeGrado, *Science* **241**, 976 (1988).
58. A. G. Murzin and A. V. Finkelstein, *J. Mol. Biol.* **204**, 749 (1988).
59. D. J. Ciesla, D. E. Gilbert, J. Feigon, *J. Am. Chem. Soc.* **113**, 3957 (1991).
60. C. P. Hill, D. H. Anderson, L. Wesson, W. F. DeGrado, D. Eisenberg, *Science* **249**, 543 (1990).
61. J. J. Osterhout *et al.*, *J. Am. Chem. Soc.* **114**, 331 (1992).
62. K. Kuwajima, *Proteins* **6**, 87 (1989); O. B. Ptitsyn, R. H. Pain, G. V. Semisotnov, E. Zerovnik, O. L. Razgulyaer, *FEBS Lett.* **262**, 20 (1990).
63. T. Sasaki and M. Lieberman, *Tetrahedron* **49**, 3677 (1993).
64. P. T. P. Kaumaya *et al.*, *Biochemistry* **29**, 13 (1990); H. Morii, K. Ichimura, H. Uedaira, *Proteins* **11**, 133 (1991); N. Nishino, H. Mihara, Y. Tanaka, T. Fujimoto, *Tetrahedron Lett.* **33**, 5767 (1992); S. Olofsson, G. Johansson, L. Baltzer, *J. Chem. Soc. Perkin Trans.* **2**, 1 (1995).
65. M. H. Hecht, J. S. Richardson, D. C. Richardson, R. C. Ogden, *Science* **249**, 884 (1990).
66. J. Skolnick and A. Kolinski, *J. Mol. Biol.* **221**, 499 (1991); E. Shakhnovich and A. M. Gutin, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 7195 (1993); *Protein Eng.* **6**,

793 (1993); K. A. Dill *et al.*, *Protein Sci.* **4**, 561 (1995); J. N. Onuchic, P. G. Wolynes, Z. Luthey-Schulten, N. D. Socci, *Proc. Natl. Acad. Sci. U.S.A.* **92**, 3626 (1995); K. Yue *et al.*, *ibid.*, p. 325.

67. S. Kamtekar, J. M. Schiffer, H. Xiong, J. M. Babik, M. H. Hecht, *Science* **262**, 1680 (1993).
68. L. Regan and N. D. Clarke, *Biochemistry* **29**, 10878 (1990).
69. T. Handel and W. F. DeGrado, *J. Am. Chem. Soc.* **112**, 6710 (1990).
70. T. M. Handel, S. A. Williams, W. F. DeGrado, *Science* **261**, 879 (1993).
71. D. P. Raleigh and W. F. DeGrado, *J. Am. Chem. Soc.* **114**, 10079 (1992).
72. D. P. Raleigh, S. F. Betz, W. F. DeGrado, *ibid.* **117**, 7558 (1995).
73. B. Honig and A.-S. Yang, *Adv. Protein Chem.* **46**, 27 (1995); K. Yue and K. A. Dill, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 4163 (1992).
74. M. Munson, R. O'Brien, J. M. Sturtevant, L. Regan, *Protein Sci.* **3**, 2015 (1994).
75. C. E. Schafmeister, L. J. W. Miercke, R. M. Stroud, *Science* **262**, 734 (1993).
76. D. S. Kemp and B. R. Bowen, *Tetrahedron Lett.* **29**, 5077 (1988); D. S. Kemp and B. R. Bowen, *ibid.*, p. 5081; K. Y. Tsang, H. Diaz, N. Graciani, J. W. Kelly, *J. Am. Chem. Soc.* **116**, 3988 (1994).
77. C. A. Kim and J. M. Berg, *Nature* **362**, 267 (1993); D. L. Minor Jr. and P. S. Kim, *ibid.* **367**, 660 (1994); *ibid.* **371**, 264 (1994); C. K. Smith, J. M. Withka, L. Regan, *Biochemistry* **33**, 5510 (1994); D. E. Otzen and A. R. Fersht, *ibid.* **34**, 5718 (1995).
78. J. Richardson *et al.*, *Biophys. J.* **63**, 1186 (1992).
79. T. P. Quinn, N. B. Tweedy, R. W. Williams, J. S. Richardson, D. C. Richardson, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 8747 (1994).
80. Y. Yan and B. W. Erickson, *Protein Sci.* **3**, 1069 (1994).
81. A. Pessi *et al.*, *Nature* **362**, 367 (1993); E. Bianchi, S. Venturini, A. Pessi, A. Tramontano, M. Sollazzo, *J. Mol. Biol.* **236**, 649 (1994).
82. A. N. Fedorov *et al.*, *J. Mol. Biol.* **225**, 927 (1992).
83. K. Goraj, A. Renard, J. A. Martial, *Protein Eng.* **3**, 259 (1990); M. Beauregard *et al.*, *ibid.* **4**, 745 (1991); A. Houbrechts *et al.*, *ibid.* **8**, 249 (1995); T. Tanaka, Y. Kuroda, H. Kimura, S. Kidokoro, H. Nakamura, *ibid.* **7**, 969 (1994); T. Tanaka, M. Hayashi, H. Kimura, M. Oobatake, H. Nakamura, *Biophys. Chem.* **50**, 47 (1994); T. Tanaka *et al.*, *Protein Sci.* **3**, 419 (1994).
84. R. S. Farid, C. C. Moser, P. L. Dutton, *Curr. Opin. Struct. Biol.* **3**, 225 (1993).
85. C. T. Choma *et al.*, *J. Am. Chem. Soc.* **116**, 856 (1994).
86. D. E. Robertson *et al.*, *Nature* **368**, 425 (1994).
87. D. L. Huffman, thesis, University of Illinois, Urbana-Champaign (1994).
88. H. Mihara *et al.*, *Chem. Lett.* **1992**, 1813 (1992); G. Basu, D. Anglos, A. Kuki, *Biochemistry* **32**, 3067 (1993); D. Anglos, V. Bindra, A. Kuki, *J. Chem. Soc. Chem. Commun.* **1994**, 213 (1994); D. G. McCafferty *et al.*, *Tetrahedron* **51**, 1093 (1995); D. R. Benson, B. R. Hart, X. Zhu, M. B. Doughty, *J. Am. Chem. Soc.* **117**, 8502 (1995).
89. M. Montal, *FASEB J.* **4**, 2623 (1990); M. R. Ghadiri, J. R. Granja, L. K. Buehler, *Nature* **369**, 301 (1994).
90. P. K. Kienker, W. F. DeGrado, J. D. Lear, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 4859 (1994).
91. B. Gutte, M. Däumingen, E. Wittschieber, *Nature* **281**, 650 (1979); W. F. DeGrado, *ibid.* **365**, 488 (1993).
92. K. Johnsson, R. K. Allemann, H. Widmer, S. A. Benner, *ibid.* **365**, p. 530.
93. S. E. Blondelle, E. Pérez-Payá, C. T. Dooley, C. Pinilla, R. A. Houghten, *Trends Anal. Chem.* **14**, 83 (1995).
94. K. W. Hahn, W. A. Klis, J. M. Stewart, *Science* **248**, 1544 (1990); M. J. Corey, E. Hallakova, K. Pugh, J. M. Stewart, *Appl. Biochem. Biotechnol.* **47**, 199 (1994).
95. K. T. O'Neil, R. H. Hoess, W. F. DeGrado, *Science* **249**, 774 (1990).
96. Abbreviations for the amino acid residues are: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.
97. T. E. Ellenberger, C. J. Brandl, K. Struhl, S. C. Harrison, *Cell* **71**, 1223 (1992).
98. D. E. Fisher, L. A. Parent, P. A. Sharp, *ibid.* **72**, 467 (1993).
99. A. R. Ferré-D'Amaré, P. Pognonec, R. G. Roeder, S. K. Burley, *EMBO J.* **13**, 180 (1994).
100. J. R. Desjarlais and J. M. Berg, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 2256 (1993).
101. E. J. Rebar and C. O. Pabo, *Science* **263**, 671 (1994); A. C. Jamieson, S.-H. Kim, J. A. Wells, *Biochemistry* **33**, 5689 (1994); Y. Choo and A. Klug, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 11168 (1994); *ibid.*, p. 11163; H. Wu, W.-P. Yang, C. F. Barbas III, *ibid.* **92**, 344 (1995).
102. Y. Choo, I. Sánchez-García, A. Klug, *Nature* **372**, 642 (1994).
103. J. U. Bowie and R. T. Sauer, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 2152 (1989); W. A. Lim and R. T. Sauer, *J. Mol. Biol.* **219**, 359 (1991); C. D. Waldburger, J. F. Schildbach, R. T. Sauer, *Nature Struct. Biol.* **2**, 122 (1995).
104. J. W. Ponder and F. M. Richards, *J. Mol. Biol.* **193**, 775 (1987); C. Lee and M. Levitt, *Nature* **352**, 448 (1991); H. W. Hellinga and F. M. Richards, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 5803 (1994); J. R. Desjarlais and T. M. Handel, *Protein Sci.*, in press.
105. V. Muñoz and L. Serrano, *Proteins* **20**, 301 (1994).
106. A. Wada, *Adv. Biophys.* **9**, 1 (1976); R. P. Sheridan and L. C. Allen, *Biophys. Chem.* **11**, 133 (1980); W. G. J. Hol, *Prog. Biophys. Mol. Biol.* **45**, 149 (1985).
107. P. J. Kraulis, *J. Appl. Crystallogr.* **24**, 946 (1991).

108. Supported by a grant from the Office of Naval Research.

Geometry of a Black Hole Collision

Richard A. Matzner,* H. E. Seidel, Stuart L. Shapiro, L. Smarr, W.-M. Suen, Saul A. Teukolsky, J. Winicour

The Binary Black Hole Alliance was formed to study the collision of black holes and the resulting gravitational radiation by computationally solving Einstein's equations for general relativity. The location of the black hole surface in a head-on collision has been determined in detail and is described here. The geometrical features that emerge are presented along with an analysis and explanation in terms of the spacetime curvature inherent in the strongly gravitating black hole region. This curvature plays a direct, important, and analytically explicable role in the formation and evolution of the event horizon associated with the surfaces of the black holes.

Black holes are small (a black hole of a million solar masses would be only as large as the sun), distant objects that have yet to be observed directly. But definite predictions about them come from detailed studies of solutions of Einstein's equations of general relativity. Fortunately, analytical methods (*1*) are powerful enough to solve these equations for a single, stationary black hole. However, this is not true for dynamically interacting black holes, believed to underlie some of the most dramatic phenomena in our universe.

The two-body problem in general relativity is still unsolved and is the subject of a National Science Foundation High-Performance Computing and Communications Grand Challenge project, termed the Binary Black Hole Alliance. The members of the Alliance include principal investigators at eight universities as well as a number of associates, collaborators, students, and postdoctoral fellows. The focus is to solve the problem as formulated topologically by Einstein and Rosen (*2*) in 1935 and as formulated numerically by DeWitt and Misner (*3*) in 1957. Solution of this 60-year-old problem will require the teraflop supercomputers of the late 1990s.

Supercomputers have advanced in speed by over 50,000 times since Hahn and Lindquist (*4*) made the first numerical attack on the problem 30 years ago. Nearly 20 years ago, Smarr and Eppley (*5*) obtained the first numerical solution of the head-on collision of two black holes of equal mass. They determined that the black holes did coalesce, radiating gravitational waves with energy of approximately 10^{-3} Mc^2, where M is the mass of the system and c is the speed of light. The gravitational waveform was similar to the damped vibrations ("ringing modes") familiar from perturbation calculations of black holes (*6*). However, numerical instabilities prevented those early calculations from being used to determine the details of the coalescence, which we report here.

R. A. Matzner is professor of physics and director of the Center for Relativity, University of Texas, Austin, TX 78712, USA. H. E. Seidel is senior research scientist at the National Center for Super Computing Applications (NCSA) and visiting associate professor of physics at the University of Illinois, Urbana, IL 61801, USA. S. L. Shapiro and S. A. Teukolsky are professors of astronomy and physics at Cornell University, Ithaca, NY 14853, USA. L. Smarr is director of NCSA and professor of astronomy and physics at the University of Illinois, Urbana, IL 61801, USA. W.-M. Suen is associate professor of physics at Washington University, St. Louis, MO 63130, USA. J. Winicour is professor of physics and astronomy at the University of Pittsburgh, Pittsburgh, PA 15260, USA.

*To whom correspondence should be addressed.