Proc. Natl. Acad. Sci. USA
Vol. 92, pp. 8408–8412, August 1995
Biochemistry

Repacking protein cores with backbone freedom: Structure prediction for coiled coils

PEHR B. HARBURY*†‡, BRUCE TIDOR†§, AND PETER S. KIM*†

*Howard Hughes Medical Institute, Department of Biology, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139; †Whitehead Institute for Biomedical Research, Nine Cambridge Center, Cambridge, MA 02142; and ‡Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA 02115

Communicated by Brian W. Matthews, Howard Hughes Medical Institute, Eugene, OR, June 5, 1995 (received for review March 14, 1995)

ABSTRACT **Progress in homology modeling and protein design has generated considerable interest in methods for predicting side-chain packing in the hydrophobic cores of proteins. Present techniques are not practically useful, however, because they are unable to model protein main-chain flexibility. Parameterization of backbone motions may represent a general and efficient method to incorporate backbone relaxation into such fixed main-chain models. To test this notion, we introduce a method for treating explicitly the backbone motions of α-helical bundles based on an algebraic parameterization proposed by Francis Crick in 1953 [Crick, F. H. C. (1953) *Acta Crystallogr.* 6, 685–689]. Given only the core amino acid sequence, a simple calculation can rapidly reproduce the crystallographic main-chain and core side-chain structures of three coiled coils (one dimer, one trimer, and one tetramer) to within 0.6-Å root-mean-square deviations. The speed of the predictive method [≈3 min per rotamer choice on a Silicon Graphics (Mountain View, CA) 4D/35 computer] permits it to be used as a design tool.**

Accurate prediction of side-chain packing and its influence on tertiary conformation represent a central goal in protein-structure prediction and design. Toward this end, several elegant computational methods (1–6) have recently been devised to "repack" side chains into models of proteins. These techniques have one serious limitation: they require precise knowledge of the protein backbone conformation. This constraint is unreasonable if one intends to truly "predict" protein structure (7–9). It is a more reasonable, but still very difficult, constraint when one wishes to use such techniques as "protein design" tools. Most importantly, x-ray crystallography studies show clearly that the backbone conformations of proteins often change in response to point mutations (9–14). This fact makes it extremely difficult to merge existing repacking algorithms with efforts in homology modeling. Backbone motions cannot be incorporated easily into packing calculations because the number of main-chain configurations available to a typical protein is astronomical. The computation time required to sample all backbone arrangements [by conventional molecular dynamics methods (15–18), for example] is prohibitive. Rapid sampling of backbone structures has only been achieved by deliberately ignoring the geometric details of packing [with lattice models (19, 20)].

The dilemma may be resolved by restricting the shape of a protein backbone to a family of parametric curves. Main-chain conformational freedom within the general backbone "fold" is achieved through variation of characteristic parameters that determine the curve shapes. By greatly reducing the number of possible main-chain arrangements, such a parameterization allows essentially every reasonable backbone structure to be sampled in a short period of time.

Coiled-coil proteins are particularly amenable to this type of analysis. Their sequences are made up of a repeated seven-amino acid pattern (the heptad repeat), conventionally denoted $(\mathbf{a–g})_n$, with hydrophobic residues generally present at the first (**a**) and fourth (**d**) positions. These sequences form amphipathic α-helices (Fig. 1*A*) that assemble into gently supercoiled bundles (Fig. 1*B*). Residues at positions **a** and **d** face inward to form a hydrophobic core. As observed by Crick (25), this structure can be parameterized with curves consisting of a circle of radius R_1 precessing with frequency ω_1 around the path of a superhelix of radius R_0, frequency ω_0, and unit path length d (Fig. 2*A*). A sixth parameter, ϕ, determines how points on the curve at integral values of the parametric variable τ (which correspond to the positions of C^α carbons in the coiled-coil backbone) are arrayed around the α-helix axis (Fig. 2*B*). If the α-helical secondary structure and bundle tertiary structure are assumed to be regular and symmetric (26, 27), the backbone freedom of parallel coiled coils is fully represented by variation of only three parameters: R_0 (the supercoil radius), ω_0 (the supercoil frequency), and ϕ (the **a**-position orientation angle) (Fig. 2*B*).

MATERIALS AND METHODS

The Crick Parameterization. The Crick parameterization (25) arises from a simple helix, $H(\tau)$, on the laboratory Z axis. At each point a primed coordinate system is embedded into the helix path, with X' on the radial vector and Z' tangent to the helix. The primed and laboratory frames are related by the Eulerian rotation matrix $E(\omega_0\tau, \alpha, 0)$. A circle $C'(\tau)$, rotating in the primed frame as it moves up the helix path, traces out the coiled-coil curve $CC(\tau)$. Thus: $CC(\tau) = E \cdot C'(\tau) + H(\tau)$, where $C'(\tau) = [R_1\cos(\omega_1\tau + \phi); R_1\sin(\omega_1\tau + \phi); 0]$ and $H(t) = \{R_0\cos(\omega_0\tau); R_0\sin(\omega_0\tau); [d\cos(\alpha)\tau]\}$. The superhelical crossing angle α is a dependent variable defined by the condition $\sin(\alpha) = R_0\omega_0/d$.

The degrees of freedom in Crick's curve are limited by structural regularity. (*i*) Equivalence of heptads requires that ω_1 be fixed to $4\pi/7$ radians per amino acid, so that seven residues complete exactly two full turns relative to the superhelix axis. This condition places every seventh residue in the same local environment (Fig. 2 *A* and *B*). (*ii*) The regularity of α-helical secondary structure limits the value of d to the rise per amino acid of a regular α-helix and the value of R_1 to the C^α radius of a regular α-helix. Accordingly, ω_1, d, and R_1 were fixed, respectively, at $4\pi/7$ radians per residue, 1.52 Å (26), and 2.26 Å (27) for all calculations. (*iii*) Symmetry around the superhelix axis implies that the monomers in an n-mer coiled coil should share an equivalent backbone geometry.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: SD, steepest descents; ABNR, adopted-basis Newton–Raphson; cR/aa, centiradian per amino acid.

§Present address: Department of Chemistry, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139.

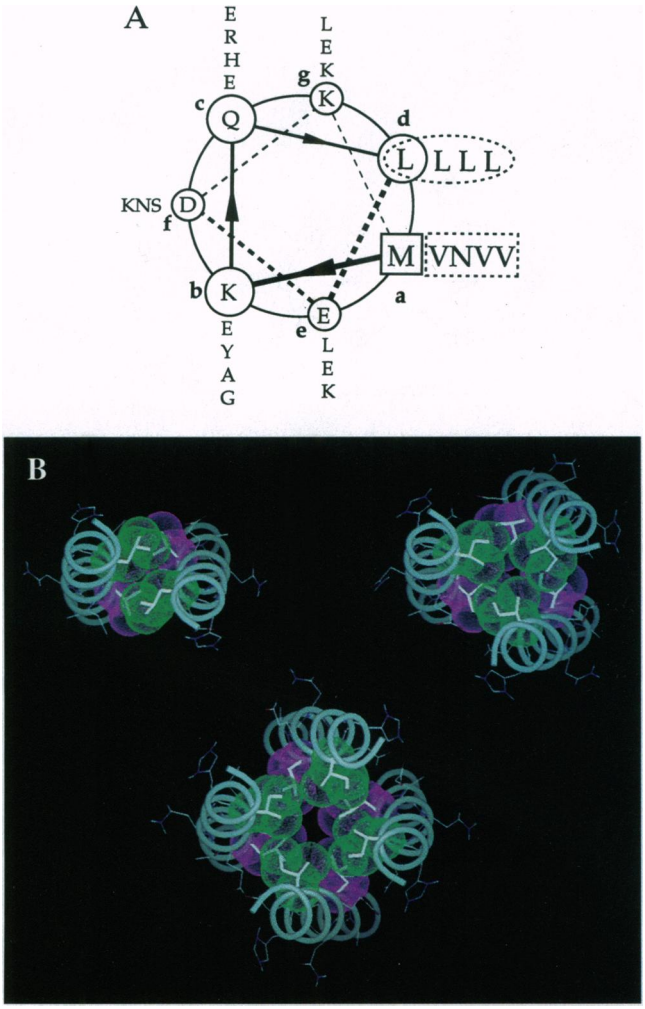

FIG. 1. The coiled-coil fold. (*A*) A helical projection of residues 2 (methionine)–32 (glutamate) of GCN4-p1, the wild-type GCN4 coiled-coil peptide sequence (21). Substitution of the residues in the dashed box (position **a**) and the dashed oval (position **d**) to differently shaped apolar amino acids changes the oligomerization order of the protein (22). Core mutants are named with the suffix p*XZ*, where *X* denotes the residue inserted at the four **a** positions and *Z* denotes the residue inserted at the four **d** positions. (*B*) Crystal structures of GCN4-p1 [dimer (23)], GCN4-pII [trimer (24)], and GCN4-pLI [tetramer (22)]. Purple van der Waals surfaces identify residues at **a** positions, and green van der Waals surfaces identify residues at **d** positions.

The physical parameters of the Crick framework were represented here by the coordinates of virtual atoms. The C^α carbons of the coiled-coil backbone were restrained onto the framework by a harmonic potential with a force constant of 12.5 kcal/(mol·Å^2) (1 cal = 4.184 J). Derivatives of the restraint potential with respect to each of the physical parameters were applied as forces to the virtual atoms. Two- and three-fold axial symmetry was applied to the dimer and trimer frameworks, while a pair of independent 2-fold axes were applied across the diagonals of the tetramer framework. A periodic boundary was created by using the IMAGE facility of CHARMM (28). Specifically, a single primary heptad was placed between two "image" heptads of identical conformation. The image heptads were rotated around and translated along the superhelix axis relative to the primary heptad according to the current values of the Crick variables. The primary and image heptads were covalently linked, but with the force constants for bonds, bond angles, and improper dihedrals reduced by factors of 100, 100, and 10, respectively, to accelerate the convergence of the backbone parameters.

Conformational Search. The calculations were done with CHARMM using the PARAM19 potential (28). All bonded potentials, the Lennard–Jones potential (switched off between 6.0 and 6.5 Å), and the explicit hydrogen-bond potential [in the polar EF2 configuration (29)] were used. Radial contraction of initially separated helices was accomplished by applying a 5 kcal/(mol·Å) force directed radially inward to all C^α carbons during 600 steps of steepest descents (SD) minimization. The χ_1 and χ_2 dihedral angles of candidate side chains were restricted with a 100 kcal/(mol·radian^2) force constant during radial contraction.

The radial force was removed after the initial 600 SD steps and remained off for the rest of the calculation. After its removal, the structure was allowed to relax during 150 additional SD steps. Subsequently, the dihedral restraints were zeroed, and the structure was subjected to 250 final steps of adopted-basis Newton–Raphson (ABNR) minimization. The side chains at positions **e** and **g** were then placed by restraining the χ_1 dihedral angles with a 10 kcal/(mol·radian^2) force constant into each of the four possible arrangements [(−) or (t) at **e** and **g**] followed by 100 SD steps, removal of the restraint, and 100 ABNR steps. The lowest potential conformation was retained. Finally, cycles over ω_0 [1 centi-radian per amino acid (cR/aa) increments from −8 to 0 cR/aa] consisted of 100 SD steps at a framework force constant of 50 kcal/(mol·Å^2) (to regularize the backbone at the new value of ω_0) followed by 100 SD steps and 150 ABNR steps at the original framework force constant of 12.5 kcal/(mol·Å^2). Side-chain rotamers were enumerated over the library of ref. 1. A single cycle for a particular rotamer choice required ≈3 min on a Silicon Graphics (Mountain View, CA) 4D/35 computer.

RESULTS

Calculating Coiled-Coil Structures. Our calculations allow all atoms of a high-resolution coiled-coil model to move under a standard molecular mechanics potential. The C^α carbons of the backbone are restrained to lie on a Crick curve (25), which acts as a structural framework (Fig. 2*A*). Forces on the main chain are transmitted to the framework, which evolves appropriately and, in turn, allows the main chain to find the regular coiled-coil conformation that minimizes the potential of the system. To additionally simplify the prediction we consider only core atoms, consisting of an alanine backbone, the full side chains at positions **a** and **d**, and side chains truncated after the C^γ at positions **e** and **g** (Fig. 1*A*). Finally, we analyze a single heptad under periodic boundary conditions (Fig. 2*C*). Several heptads can then be concatenated to generate a full-length coiled coil.

To calculate a three-dimensional structure, we begin with a choice of amino acid and rotamer angles for position **a** and independently for position **d**. The specified side chains are built onto the interior positions of straight and well separated helices, which are each seven residues long and arranged in a bundle of appropriate symmetry. The helices are subsequently "squashed" together by the application of a gentle radial force with minimization of the configurational potential. During radial contraction, the main chain is pushed into a shape that accommodates the side chains in their chosen rotamers. After radial collapse, the radial force is removed, and the structure is allowed to relax. The supercoiling parameter ω_0 is then actively varied from the straight helix limit to a severely overwound limit. This core-packing procedure is iterated over a canonical library of rotamer angles (1), allowing the conformation of minimum potential energy to be identified for each core sequence.

Comparison with Crystallographic Structures. In the ensuing discussion, p*XZ* refers to the core sequence with

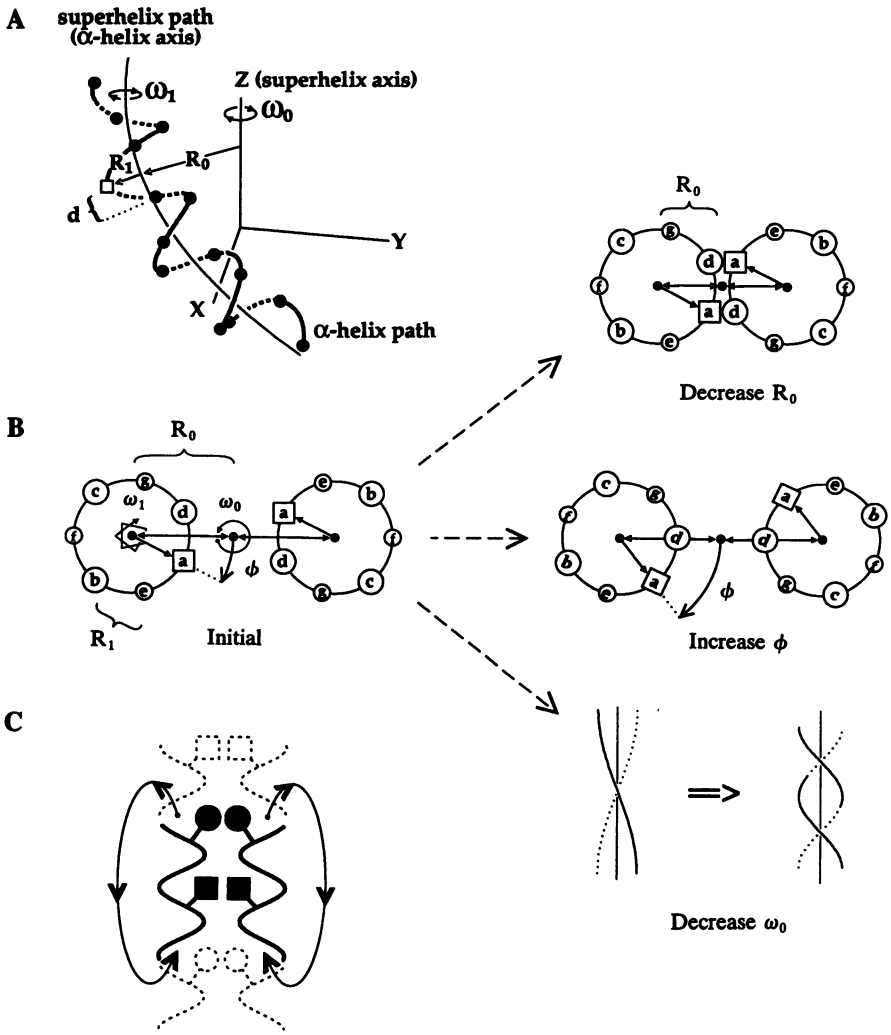

FIG. 2. Crick parameterization (25). (*A*) Two heptads of a coiled-coil framework. The roles of the variables *d* (rise per amino acid in the α-helix), ω_0 (frequency of the superhelix), ω_1 (frequency of the α-helix), R_0 (radius of the superhelix), and R_1 (radius of the α-helix) are depicted. Black beads correspond to C^α positions at points of integral τ. (*B*) Helical projections illustrate the three free backbone variables. Conformations with decreased superhelix radius (R_0), increased **a**-position orientation angle (ϕ), and decreased (more negative) superhelix frequency (ω_0) relative to an initial conformation are shown. (*C*) A single heptad with periodic boundary conditions is analyzed. A primary heptad (—) is sandwiched between two appropriately rotated and translated image heptads (- - - -) with the primary conformation. Any atom that passes out of the top of the primary heptad reenters at the bottom in the form of an image atom.

amino acid *X* at position **a** and amino acid *Z* at position **d** in the apolar background used for the packing calculation, while GCN4-p*XZ* denotes the experimentally characterized peptide with *X* at **a** positions and *Z* at **d** positions in the GCN4 coiled-coil background (Fig. 1). Exhaustive conformational searches were done with the sequences pVL, pII, and pLI, for comparison against the crystal structures of the corresponding GCN4 core variants, GCN4-p1 (23), GCN4-pII (24), and GCN4-pLI (22) [the pVL core was used to model GCN4-p1, which contains three valines and an asparagine at the **a** positions and four leucines at the **d** positions (Fig. 1)].

Starting from a range of extreme initial conditions (R_0 = 10 Å; ϕ = 0° and 40°; ω_0 = 0 and −8 cR/aa), a unique minimum-potential structure for each sequence was identified. The different initial conditions yielded identical results in all cases. Comparison of the empirically calculated conformations and crystallographically determined structures (Table 1) reveals an excellent agreement in the radius and ϕ parameters. The calculated and observed values of the supercoiling parameter ω_0 agree within 10% for the dimer and tetramer and within 25% for the trimer. For the dimer and tetramer conformations, a unique, energetically distinct, minimum potential side-chain geometry is identified by the calculation. In both cases, these calculated rotamers match exactly the side-chain rotamers observed in the four heptads of the corresponding crystal structure (Table 1). For the trimer conformation, the (+,*t*), (−,*t*) and (−,*t*), (−,*t*) isoleucine rotamer pairs are predicted to have roughly equal stabilities [(+,*t*), (−,*t*) is favored by 0.2 kcal/mol per helix per heptad]. In the trimer crystal structure, the (+,*t*), (−,*t*) conformation is observed in the most C-terminal heptad, while the three N-terminal heptads exhibit the (−,*t*), (−,*t*) conformation. Main-chain and core side-chain coordinates in the dimer, trimer, and tetramer models differ from the crystallographic coordinates by average root-mean-square deviations of 0.3, 0.4, and 0.6 Å, respectively (Table 1). These results are depicted graphically in Figs. 3 and 4.

DISCUSSION

The primary advantage of backbone parameterization is that it greatly simplifies and accelerates the conformational search required to find a minimum-potential protein conformation. However, the technique is based upon an underlying assumption: that the optimal structure for the protein sequence in

Table 1. Comparison of crystallographically observed and empirically calculated structures

GCN4 peptide variant, p*XZ*	R_0, Å		ϕ		ω_0, (cR/aa)		Rotamers at **a**, **d***			
	Obs.	Calc.	Obs.	Calc.	Obs.	Calc.	Obs.	Calc.	[ΔE]	rmsd†, Å
p1 [pVL]‡ (dimer)	4.9	4.8	22°	24°	−6.3	−6.1	(*t*), (−,*t*)	(*t*), (−,*t*)	[1.4]	0.24–0.38
pII§ (trimer)	6.7	6.5	20°	20°	−5.3	−6.2	(−,*t*), (−,*t*)	(+,*t*), (−,*t*)	[0.2]	0.37–0.47
		6.5		22°		−4.1	(+,*t*), (−,*t*)	(−,*t*), (−,*t*)	[0.7]	
pLI (tetramer)	7.6	7.5	20°	22°	−4.5	−4.2	(−,*t*), (−,*t*)	(−,*t*), (−,*t*)	[2.4]	0.57–0.64

Comparison of calculated (Calc.) and crystallographic structures (Obs.) of a dimer [GCN4-p1 (23)], a trimer [GCN4-pII (24)], and a tetramer [GCN4-pLI (22)] coiled coil (Fig. 1). Results are illustrated graphically in Fig. 3; the supercoil parameters are illustrated in Fig. 2. The observed parameters are averages over entire crystal structures.

*Rotamers follow the convention of ref. 1. ΔE represents the potential energy gap (kcal/mol per heptad per helix) between the indicated rotamer packing and the next best solution.

†Root-mean-square deviations between the coordinates of calculated and observed heptads for main-chain and core side-chain atoms are reported (Fig. 3). The maximum and minimum deviations resulting from superposition of the calculated heptad with each of the three central heptads in the appropriate crystal structure are indicated. For the pII trimer, the (−,*t*), (−,*t*) calculated coordinates were used.

‡GCN4-p1 contains three valines and an asparagine at the **a** positions and contains four leucines at the **d** positions. The pVL core sequence was used for calculation of the GCN4-p1 structure.

§For the pII trimer, the (+,*t*), (−,*t*) rotamer pair is evaluated as only slightly more stable than the (−,*t*), (−,*t*) rotamer pair (ΔE = 0.2 kcal/mol). In the crystal structure of GCN4-pII, the (+,*t*), (−,*t*) conformation appears in the COOH-terminal heptad, whereas the remaining three heptads exhibit the (−,*t*), (−,*t*) conformation.

question lies within the parametric family being searched. We demonstrate that this conceptual approach can be applied successfully to α-helical bundle proteins.

The Crick parameterization allows the core side-chain and backbone conformation of GCN4 coiled-coil variants to be determined with orders-of-magnitude greater speed than is possible by any other method applied to modeling coiled coils (30–33). Moreover, the result of the parametric search is more accurate than that produced by brute-force searches; no other computational approach has been able to correctly identify the optimum side-chain rotamers in each of the three parallel coiled-coil structures formed by the GCN4 core mutants. The calculation has no "adjustable" parameters and uses a general molecular mechanics potential. Because of its speed and accuracy, the method may be used to predict the specificity of untested core sequences for the different coiled-coil oligomeric states (41). The ability of the present calculation to reproduce crystal structures despite treating only core side chains suggests that interior packing dominates the backbone architecture of coiled coils [in accord with the general premise of Ponder and Richards (1)].

The technique may be extended in several ways. (*i*) The incorporation of a continuum solvent model (34, 35) (in place of the current *in vacuo* calculations) would allow for treatment of charged and polar residues at a minimum expense of computation time. We presently model only apolar side chains

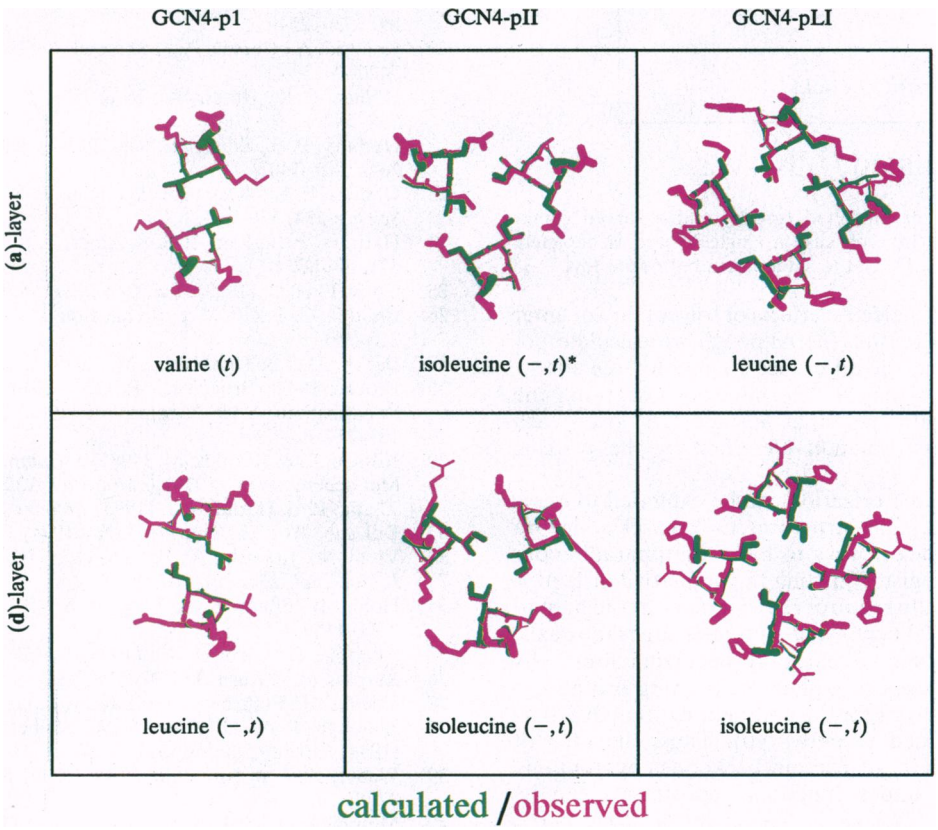

FIG. 3. Superposition of calculated and observed structures. (*A*) Comparison of the predicted (green) and observed (pink) side-chain packing at the **a** and **d** levels of the dimer [GCN4-p1 (23)], trimer [GCN4-pII (24)], and tetramer [GCN4-pLI (22)] parallel coiled-coil structures. *, The (−,*t*) rotamer solution for the **a**-layer of the trimer is depicted (see Table 1 legend).

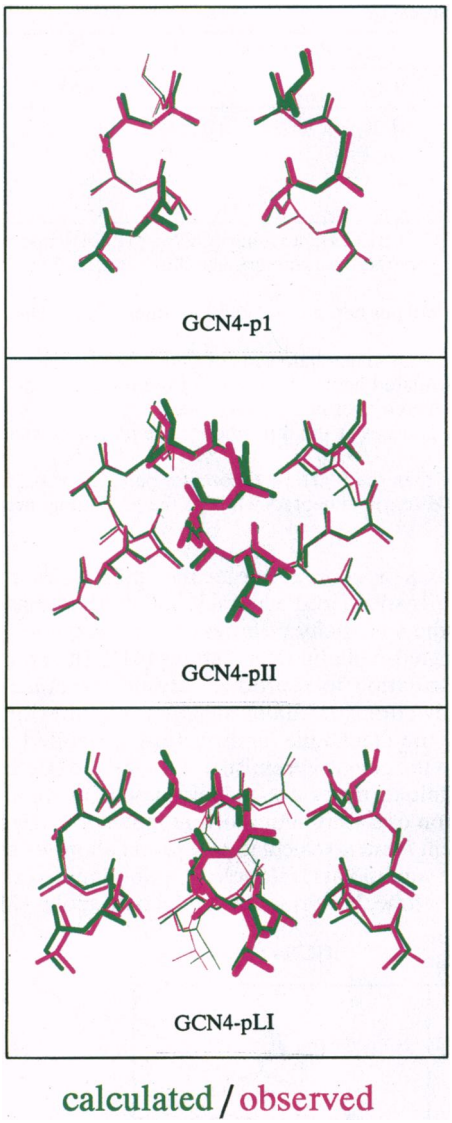

FIG. 4. Comparison of predicted (green) and observed (pink) main-chain coordinates. In each case a single heptad is depicted, running NH_2 terminal to COOH terminal from bottom to top.

so that the charge and dielectric effects of solvent do not enter into the computed potential. (*ii*) Although we calculate potential differences here, the difference Helmholtz free energy between conformations may be determined by integrating the approximately quadratic potential surface (36) over each low-lying rotamer conformation identified by the packing calculation.

(*iii*) Backbone parameterization can be extended to many additional and less regular structures (37). Crick's algebra may be modified in direct ways to handle antiparallel conformations, heptad register shifting [a cumulative drift of ϕ (Fig. 2*B*) that eventually requires reassignment of the heptad repeat (38)], and radial expansion along the superhelix axis, which is often observed in naturally occurring four-helix bundles (39). In the area of β-sheet-containing proteins, a mathematical groundwork for the parameterization of β-barrels has been described recently (40). Thus, the idea of reducing the complexity of protein backbones by coupling degrees of freedom under functional constraints can be applied broadly.

We gratefully acknowledge Z. Y. Peng, D. J. Lockhart, T. Alber, and members of the Kim laboratory for many insightful conversations.

This work was supported by the Howard Hughes Medical Institute (P.S.K.) and grants from the National Institutes of Health (GM47678 to B.T.), the Lucille P. Markey Charitable Trust (B.T.), and the W. M. Keck Foundation (B.T.).

1. Ponder, J. W. & Richards, F. M. (1987) *J. Mol. Biol.* **193,** 775–791.
2. Correa, P. E. (1990) *Proteins* **7,** 366–377.
3. Lee, C. & Subbiah, S. (1991) *J. Mol. Biol.* **217,** 373–388.
4. Holm, L. & Sander, C. (1991) *J. Mol. Biol.* **218,** 183–194.
5. Tuffery, P., Etchebest, C., Hazout, S. & Lavery, R. (1991) *J. Biomol. Struct. Dynam.* **8,** 1267–1289.
6. Desmet, J., Maeyer, M. D., Hazes, B. & Lasters, I. (1992) *Nature (London)* **356,** 539–542.
7. Richards, F. M. & Lim, W. A. (1993) *Q. Rev. Biophys.* **26,** 423–498.
8. Lee, C. & Levitt, M. (1991) *Nature (London)* **352,** 448–451.
9. Lim, W. A., Hodel, A., Sauer, R. T. & Richards, F. M. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 423–427.
10. Baldwin, E. P., Hajiseyedjavadi, O., Baase, W. A. & Matthews, B. W. (1993) *Science* **262,** 1715–1719.
11. Hurley, J. H., Baase, W. A. & Matthews, B. W. (1992) *J. Mol. Biol.* **224,** 1143–1159.
12. Varadarajan, R. & Richards, F. M. (1992) *Biochemistry* **31,** 12315–12327.
13. Eigenbrot, C., Randal, M. & Kossiakoff, A. A. (1990) *Protein Eng.* **3,** 591–598.
14. McRee, D. E., Redford, S. M., Getzoff, E. D., Lepock, J. R., Hallewell, R. A. & Tainer, J. A. (1990) *J. Biol. Chem.* **265,** 14234–14241.
15. McCammon, J. A., Gelin, B. R. & Karplus, M. (1977) *Nature (London)* **267,** 585–589.
16. Levitt, M. (1983) *J. Mol. Biol.* **170,** 723–764.
17. Weiner, S. J., Kollman, P. A., Case, D. A., Singh, U. C., Ghio, C., Alagona, G., Profeta, S. J. & Weiner, P. (1984) *J. Am. Chem. Soc.* **106,** 765–784.
18. Gunsteren, W. F. v. & Berendson, H. J. C. (1987) *Groningen Molecular Simulation (GROMOS) Library Manual* (Biomos, Groningen, The Netherlands).
19. Hinds, D. A. & Levitt, M. A. (1992) *Proc. Natl. Acad. Sci. USA* **89,** 2536–2540.
20. Kolinski, A., Godzik, A. & Skolnick, J. (1993) *J. Chem. Phys.* **98,** 7420–7433.
21. O'Shea, E. K., Rutkowski, R. & Kim, P. S. (1989) *Science* **243,** 538–542.
22. Harbury, P. B., Zhang, T., Kim, P. S. & Alber, T. (1993) *Science* **262,** 1401–1407.
23. O'Shea, E. K., Klemm, J. D., Kim, P. S. & Alber, T. (1991) *Science* **254,** 539–544.
24. Harbury, P. B., Kim, P. S. & Alber, T. (1994) *Nature (London)* **371,** 80–83.
25. Crick, F. H. C. (1953) *Acta Crystallogr.* **6,** 685–689.
26. Chothia, C., Levitt, M. & Richardson, D. (1981) *J. Mol. Biol.* **145,** 215–250.
27. Barlow, D. J. & Thornton, J. M. (1988) *J. Mol. Biol.* **201,** 601–619.
28. Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S. & Karplus, M. (1983) *J. Comput. Chem.* **4,** 187–217.
29. Nilsson, L. & Karplus, M. (1986) *J. Comput. Chem.* **7,** 591–616.
30. McLachlan, A. D. (1978) *J. Mol. Biol.* **122,** 493–506.
31. Zhang, L. & Hermans, J. (1993) *Proteins* **16,** 384–392.
32. DeLano, W. L. & Brunger, A. T. (1994) *Proteins* **20,** 105–123.
33. Vieth, M., Kolinski, A., Brooks, C. L., III & Skolnick, J. (1994) *J. Mol. Biol.* **237,** 361–367.
34. Honig, B., Sharp, K. & Yang, A. S. (1993) *J. Phys. Chem.* **97,** 1101–1109.
35. Abagyan, R. & Totrov, M. (1994) *J. Mol. Biol.* **235,** 983–1002.
36. Karplus, M. & Kushick, J. N. (1981) *Macromolecules* **14,** 325–332.
37. Sklenar, H., Etchebest, C. & Lavery, R. (1989) *Proteins* **6,** 46–60.
38. Bullough, P. A., Hughson, F. M., Skehel, J. J. & Wiley, D. C. (1994) *Nature (London)* **371,** 37–43.
39. Weber, P. C. & Salemme, F. R. (1980) *Nature (London)* **287,** 82–84.
40. Murzin, A. G., Lesk, A. M. & Chothia, C. (1994) *J. Mol. Biol.* **236,** 1369–1381.
41. Harbury, P. A. B. (1994) Thesis (Harvard Univ., Cambridge, MA).