Side-chain repacking calculations for predicting structures and stabilities of heterodimeric coiled coils

Amy E. Keating[†‡§], Vladimir N. Malashkevich[†], Bruce Tidor[‡], and Peter S. Kim[†¶]

[†]Whitehead Institute for Biomedical Research, Howard Hughes Medical Institute, Department of Biology, Massachusetts Institute of Technology, Nine Cambridge Center, Cambridge, MA 02142; and [‡]Department of Chemistry, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139

Contributed by Peter S. Kim, October 22, 2001

An important goal in biology is to predict from sequence data the high-resolution structures of proteins and the interactions that occur between them. In this paper, we describe a computational approach that can make these types of predictions for a series of coiled-coil dimers. Our method comprises a dual strategy that augments extensive conformational sampling with molecular mechanics minimization. To test the performance of the method, we designed six heterodimeric coiled coils with a range of stabilities and solved x-ray crystal structures for three of them. The stabilities and structures predicted by the calculations agree very well with experimental data: the average error in unfolding free energies is <1 kcal/mol, and nonhydrogen atoms in the predicted structures superimpose onto the experimental structures with rms deviations <0.7 Å. We have also tested the method on a series of homodimers derived from vitellogenin-binding protein. The predicted relative stabilities of the homodimers show excellent agreement with previously published experimental measurements. A critical step in our procedure is to use energy minimization to relax side-chain geometries initially selected from a rotamer library. Our results show that computational methods can predict interaction specificities that are in good agreement with experimental data.

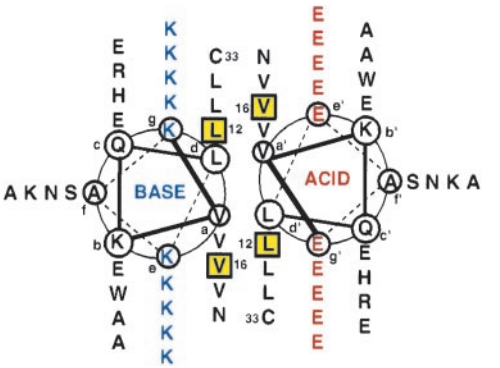

Fig. 1. Helical wheel diagram of the heterodimeric coiled coil GABH. Substitutions of Val, Ile, and Leu were made at positions d12 and a16 (yellow boxes) to give six peptides: $A_{LI}B_{LL}$, $A_{LL}B_{LL}$, $A_{IV}B_{LL}$, $A_{LL}B_{IV}$, $A_{LI}B_{IV}$, and $A_{IV}B_{IV}$ (notation: *A*cid$_{d12a16}$*B*ase$_{d12a16}$). The linear sequence is: Ac-(E/K)VKQL(E/K)A(E/K)VEEd12(E/K)S(E/K)a16WHL(E/K)N(E/K)VARL(E/K)K(E/K)NAEC(E/K)A-NH_2; the ACID peptide has E and the BASE peptide has K at positions in parentheses.

The goal of predicting protein interactions is a challenging task, even in cases where high-quality structural data are available. Making such predictions from sequence alone is beyond the reach of existing computational methods. We approach this problem by studying protein–protein interactions in a simple motif: the coiled coil. Here we describe our work predicting the structure and stability of a series of coiled-coil peptides by using a computational method that combines a fast conformational search procedure with molecular mechanics minimization.

The coiled coil is among the most common of all protein–protein interaction motifs. Computational analysis suggests that it is found in ≈5% of proteins in eukaryotic genomes (1). Many biological roles have been identified for coiled coils in diverse processes, such as transcription, membrane fusion, and chromosome segregation (see, for example, refs. 1–4). In most known examples, the coiled coil is involved in protein–protein interactions.

Coiled coils consist of two or more α helices that wrap around each other with a slight superhelical twist. Their regular structure results from a characteristic repeating heptad of amino acids, denoted $(\mathbf{abcdefg})_n$, in which the **a** and **d** positions are predominantly hydrophobic, and the **e** and **g** positions are typically charged or polar. The **a** and **d** positions make up the core of the coiled-coil interface, whereas the **e** and **g** positions occupy the region between the core and the solvent. The repeating heptad allows the motif to be predicted accurately from sequence data (5, 6). The structural regularity also provides a benefit in molecular modeling. With these advantages, coiled coils offer an opportunity to test methods for predicting interaction specificities and to analyze the computational challenges that confront such a task.

Several different types of interactions that influence partner selection among coiled coils have previously been exploited for prediction or design. Simple scoring rules based on counting the number of putative attractions and repulsions between charged **g**- and **e**-position residues have some predictive value (7–9). Singh and Kim recently introduced a new computational framework for addressing the partnering problem in coiled coils that includes interactions at additional positions (10). In this paper, we address the role of hydrophobic residues at the **a** and **d** positions in determining interaction specificity. We show that computational modeling of coiled-coil structures can be used to predict interaction energy differences that agree quantitatively with experimental results.

There are several obstacles to predicting protein structure accurately enough to describe interactions. The conformational space accessible to even a small protein is large and impossible to sample exhaustively. Further, protein structures are very plastic. Even when an approximate structure is known, small sequence-specific adjustments in side-chain and backbone conformation are difficult to

Abbreviations: VBP, vitellogenin-binding protein; DEE, Dead End Elimination; CD, circular dichroism; Gdm·HCl, guanidinium chloride; Apa, aminopentanoic acid; MIN-SC, fixed backbone side-chain minimization; MIN-FULL, full backbone minimization; GABH, *G*CN4 *A*cid/*B*ase *H*eterodimer.

Data deposition: The atomic coordinates reported in this paper have been deposited in the Protein Data Bank, www.rcsb.org (PDB ID codes 1KD8, 1KD9, and 1KDD).

§To whom reprint requests should be addressed at: Department of Biology, MIT 68-622, 77 Massachusetts Avenue, Cambridge, MA 02139. E-mail: keating@mit.edu.

¶Present address: Merck Research Laboratories, Merck & Co., Inc., 770 Sumneytown Pike, West Point, PA 19486.

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.

BIOCHEMISTRY

Table 1. Relative unfolding free energies in the absence of Gdm·HCl (ΔG^u, kcal/mol), the dependence of ΔG^u on [Gdm·HCl] (m, kcal/mol·M) and the concentration of Gdm·HCl where the peptide is half unfolded ($[Gdm·HCl]_{1/2}$, M) for all GABH peptides

Peptide	ΔG^u	m	$[Gdm·HCl]_{1/2}$
$A_{LI}B_{LL}$	11.3 (0.4)	1.9	5.8
$A_{LL}B_{LL}$	9.8 (0.3)	1.8	5.4
$A_{IV}B_{LL}$	7.4 (0.2)	1.7	4.5
$A_{LL}B_{IV}$	7.0 (0.1)	1.6	4.3
$A_{LI}B_{IV}$	6.1 (0.1)	1.6	3.9
$A_{IV}B_{IV}$	4.4 (0.1)	1.5	2.8

Standard deviations are given in parentheses.

incorporate into models. In our work on coiled coils, we address these issues by using three strategies. First, following Harbury *et al.* (11, 12), we use Crick's parameterization of the coiled coil to limit the conformational space accessible to the protein backbone (13). Second, we use an efficient search algorithm to sample the large side-chain conformational space of coiled-coil heterodimers. Third, we introduce a final minimization step to fine-tune the backbone and side-chain geometry of our predicted structure. The first two strategies address the problem of a large search space, whereas the third is critical for achieving quantitative estimates of interaction energies.

Materials and Methods

Peptide Synthesis and Characterization. Six peptides with sequences given in Fig. 1 were synthesized and purified by reverse-phase HPLC, as previously described (14). Purity was >95%, and masses were verified by matrix-assisted laser desorption ionization time-of-flight mass spectrometry (PerSeptive Biosystems, Framingham, MA).

Circular Dichroism (CD). CD spectra were collected at 25°C in PBS (50 mM sodium phosphate/150 mM sodium chloride) at pH 7.0 by using an Aviv (Lakewood, NJ) 62A/DS CD spectrometer. Denaturation of peptides by guanidinium chloride (Gdm·HCl) was monitored at 222 nm. Concentrations were determined by the method of Edelhoch (15). Unfolding curves were fit, assuming a two-state transition, to give Gdm·HCl concentrations at half-denaturation ($[Gdm·HCl]_{1/2}$), m values (the dependence of ΔG on [Gdm·HCl]), and free energies of unfolding in the absence of Gdm·HCl (ΔG^u) (Table 1). Because upper baselines were not always well defined, these were assigned a slope of zero; the average deviation for fits performed with and without this constraint was 0.16 kcal/mol.

Sedimentation Equilibrium. Sedimentation equilibrium experiments at 5, 10, and 25 μM were performed after dialyzing against PBS (pH 7.0) using a Beckman XLA-90 analytical ultracentrifuge (Beckman Coulter). Data were collected after equilibration was verified (18–24 h) at 25,000, 30,000, and 35,000 rpm at 25°C. Data were fit to a single ideal species. Partial specific volumes and solvent density were calculated from the amino acid sequence and solvent composition, respectively (16).

Crystallography. Initial crystallization conditions for peptides $A_{IV}B_{LL}$, $A_{LI}B_{LL}$, and $A_{LL}B_{LL}$ were identified by using Crystal Kits I and II (Hampton Research, Riverside, CA). Protein stocks at 10 mg/ml in water were mixed 1 μl:1 μl with buffer and equilibrated against the well reservoir using the hanging drop method. For all three peptides, crystals could be grown from 18–20% polyethylene glycol (PEG) 4000/0.1 M Na Hepes (pH 7.7)/10% 2-propanol. Crystals were flash-frozen in a nitrogen stream after soaking briefly in a cryoprotectant solution of the well buffer plus 15% PEG 400. All three peptides crystallized with the same packing in the space group $P4_12_12$.

Diffraction data were collected at beamline X4A at the National Synchrotron Light Source (Brookhaven National Laboratory, Upton, NY) with a Quantum IV detector (Area Detector Systems Corporation, Poway, CA). Data were integrated and scaled with the programs DENZO and SCALEPACK (17); statistics of the data collection and refinement are given in Table 2.

Initial phases were determined by using the program AMORE (18). The model for the molecular replacement procedure was generated by using Stage 1 of our computational procedure (see below). Models constructed from 12 different backbones and various side-chain rotamer possibilities were tested (the **a**- and **d**-position residues of peptide $A_{LI}B_{LL}$ were used for all data sets), which yielded initial phases for peptides $A_{IV}B_{LL}$ and $A_{LL}B_{LL}$. Peptide $A_{LI}B_{LL}$ was refined with initial phases from $A_{IV}B_{LL}$. Model building was done with the program O (19), and refinement calculations were done with the program CNS (20).

We took precautions to eliminate phase bias from the refinement procedure: all initial models were subjected to high-temperature (5,000 K) simulated annealing, all model building was done against composite omit maps, and side-chains at positions **d**12 and **a**16 were initially refined as Ala and rebuilt only when their density was clear in the maps.

Computation of Structure and Energies for Coiled Coils. To facilitate modeling of the peptides in Fig. 1, we represented each of them as an infinitely repeating segment consisting of residues **g**8–**f**28. The difference between the actual structure and the assumption of an infinite coil is minimal at positions **d**12 and **a**16. For calculations on vitellogenin-binding protein (VBP), the coiled coil was represented by a sequence consisting of residues 71–94 of the 96-residue constructs in ref. 21. Residue 72 (an **a**-position Val) was constrained to have the gauche (−) rotamer, which is favored in this position in coiled coils, to provide a boundary condition.

To address the difficulty of modeling solvent-exposed charged residues, we truncated the **e** and **g** residues beyond C_δ, substituting these residues with aminopentanoic acid (here denoted Apa). All **b**, **c,** and **f** positions were Ala in the calculations. Apa residues in the **e** and **g** positions were assigned conformations consistent with the formation of salt bridges, where these sites comprised Lys/Glu or Arg/Glu pairs. The only other inputs to the calculation were the sequence, oligomerization state and register assignment of the

Table 2. Data collection and refinement statistics

Peptide	λ, Å	Completeness, %	R_{merge},* %	Data resolution, Å	Refinement resolution, Å	Nonhydrogen atoms	Waters	Reflections	R_{cryst}/R_{free},† %	rms deviations (bonds, Å/angles, °)
$A_{IV}B_{LL}$	0.9763	98.9	6.0	35–1.85	30–1.9	1,705	288	23,529	23.2/27.6	0.018/0.99
$A_{LL}B_{LL}$	0.9763	98.3	4.0	35–2.00	10–2.1	1,650	174	17,771	24.2/29.6	0.013/1.40
$A_{LI}B_{LL}$	1.0093	97.5	6.0	35–2.05	20–2.14	1,595	113	16,580	25.2/30.2	0.004/0.83

*$R_{merge} = \Sigma\Sigma_j|I_j(hkl) - \langle I(hkl)\rangle|/\Sigma\Sigma_j|I_j(hkl)|$, where I is intensity.

†R_{cryst} (R_{free}) $= \Sigma||F_{obs}(hkl)| - |F_{calc}(hkl)||/\Sigma|F_{obs}(hkl)|$, where F_{obs} and F_{calc} are observed and calculated structure factors. Ten percent of reflections were excluded to compute R_{free}.

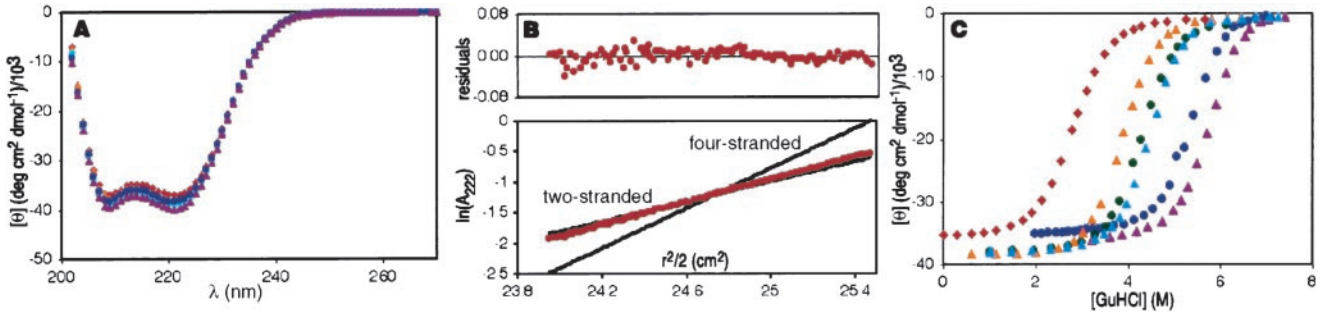

Fig. 2. Characterization of GABH peptides. (*A*) CD spectra for the six peptides show that all are highly helical and exhibit minima of similar magnitude at 208 and 222 nm, which is characteristic of coiled coils. (*B*) Analytical ultracentrifugation data for the least stable peptide, $A_{IV}B_{IV}$ at 25 μM and 25°C. The data for all six peptides can be fit to give the expected mass for a two-stranded species, with random residuals. (*C*) The [Gdm·HCl] dependence of the CD signal at 222 nm for peptides at 10 μM and 25°C. $A_{IV}B_{IV}$ = red, $A_{LI}B_{IV}$ = orange, $A_{LL}B_{IV}$ = green, $A_{IV}B_{LL}$ = light blue, $A_{LL}B_{LL}$ = dark blue, $A_{LI}B_{LL}$ = purple.

coiled coil of interest. This type of information is available from coiled-coil prediction programs such as MULTICOIL (6).

Stage 1 = Discrete Search by Using Dead-End Elimination (DEE)/A* (NOMIN). For each backbone in a predetermined library, low-energy side-chain configurations were identified by using a combination of DEE (22, 23) and an A* search algorithm (24). The standard DEE elimination criterion was modified by introducing a cut-off energy of 30 kcal/mol. This guaranteed that all solutions within this energy of the global minimum were retained in the search. The A* search procedure was used to sort solutions that survived DEE and return up to 10,000 of them per backbone in order of increasing energy. We implemented the A* search with bounds as described in ref. 24. Details are published as supporting information on the PNAS web site, www.pnas.org.

Side-Chain and Backbone Libraries. Default bond lengths and angles from the CHARMM PARAM19 parameter set were used for all rotamers (25). Side-chain torsion values were taken from Dunbrack's August 1999 library for backbones with $\phi = -60$ and $\psi = -50$ (26). The rotamers used are given in the supporting information on the PNAS web site. The number of rotamers per residue was six (Leu), six (Ile), three (Val), and six (Apa). Apa was assigned one of two rotamers ($\chi_1 = -70$, $\chi_2 = 180$; $\chi_1 = 180$, $\chi_2 = 180$) when assumed to be in a salt-bridging geometry. Subrotamers were constructed for all residues by including $+/-10°$ from the parent χ_1 and χ_2.

A library of 48 idealized coiled-coil backbones described by the parameters $R_0 = \{4.8, 4.9, 5.0$ Å$\}$, $\omega_0 = \{0.055, 0.06, 0.065, 0.07$ radians/residue$\}$, and $\phi = \{0.3, 0.35, 0.4$ radians$\}$ was generated as previously described (11). These values bracket those that describe the GCN4-p1 dimer structure.

Side-Chain Rotamer Energy Terms for DEE/A* Calculations. Interaction energies between side chains and the template (backbone plus **b**-, **c**-, and **f**-position Ala) and among side chains were calculated by using the subrotamer method of Mendes *et al.* (27), which replaces the energy between rotamers with a weighted

Table 3. Rotamer combinations observed in the crystal structures and how they were ranked by different calculations

		X-ray				Computation		
		Rotamer*						
Peptide	Copy no. in asymmetric unit	A_{d12}	B_{d12}	A_{a12}	B_{a16}	Method	Relative energy	Rank
A_{IV}/B_{LL}	1, 2	I_1	L_2	V_1	L_1	MIN-FULL	0	1
	3	I_1	L_2	V_1	L_2		0.3	5
	1, 2	I_1	L_2	V_1	L_1	MIN-SC	1.6	3
	3	I_1	L_2	V_1	L_2		0	1
	1, 2	I_1	L_2	V_1	L_1	NOMIN	50.4	488
	3	I_1	L_2	V_1	L_2		0	1
A_{LL}/B_{LL}	1	L_2	L_2	L_2	L_2	MIN-FULL	0	1
	2	L_2	L_2	L_1	L_2†		0.9	2
	3	L_2	L_2	L_1‡	L_1‡		2.5	38
	1	L_2	L_2	L_2	L_2	MIN-SC	8.9	142
	2	L_2	L_2	L_1	L_2†		2.7	5
	3	L_2	L_2	L_1‡	L_1‡		10.2	180
	1	L_2	L_2	L_2	L_2	NOMIN	0	1
	2	L_2	L_2	L_1	L_2†		23.6	54
	3	L_2	L_2	L_1‡	L_1‡		57.7	284
A_{LI}/B_{LL}	1, 2, 3	L_2	L_2	I_1	L_1	MIN-FULL	0.5	3
	1, 2, 3	L_2	L_2	I_1	L_1	MIN-SC	3.0	5
	1, 2, 3	L_2	L_2	I_1	L_1	NOMIN	30.0	156

The three copies of the dimer in the asymmetric unit are numbered 1–3. Calculated energies (kcal/mol) and ranks are given relative to the lowest energy predicted structure.

*The rotamer notation is Residue# = (χ_1, χ_2) with I_1 = (−, trans); L_1 = (trans, +); L_2 = (−, trans); V_1 = (trans).

†B factor >65 Å^2 for one Cδ.

‡Distorted χ_1 angles at these sites: $\chi_1 = -142°, -158°$ were assigned as trans.

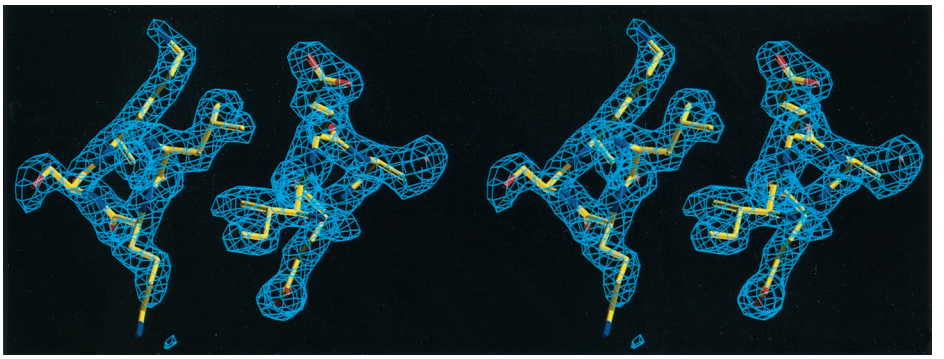

Fig. 3. Stereo view of a composite omit map for $A_{IV}B_{LL}$ showing the packing of Ile d12 against Leu d12 in the core (at 1.6σ). The view is from the N terminus and includes residues 12–15.

average of the energies between a set of subrotamers. The equations used are given in the supporting information on the PNAS web site. This procedure compensates for the inflexibility of library rotamers. Side-chain subrotamer energies were computed by using the program CHARMM and the PARAM19 force field (25) with 90% van der Waals radii. No electrostatic terms were included. Energies between side chains on residues with >10-Å separation between Cβ atoms were set to 0.0 kcal/mol, and pairwise energies between subrotamers >100 kcal/mol were set to a uniform large value (10^6 kcal/mol).

Stage 2 = Fixed Backbone Side-Chain Minimization (MIN-SC) or Full Backbone Minimization (MIN-FULL): Geometry Minimization. Up to 3,500 unique side-chain configurations identified by the A* search were minimized using the PARAM19 potential with 100% van der Waals radii, no electrostatics, and the addition of an explicit hydrogen-bonding term, as described in refs. 11 and 28 and in the supporting information on the PNAS web site. Minimization allowing compensating changes in side-chain and backbone geometry was carried out as described refs. 11 and 12.

Energy Corrections. Stages 1 and 2 were carried out in the gas phase. To estimate energies in aqueous solution, an empirical solvation correction was applied. We used the solvation parameters of Wesson and Eisenberg (29); however, we chose an atomic solvation coefficient of zero for Met sulfur. Solvent accessible surface areas were computed with CHARMM by using the method of Lee and Richards and a probe radius of 1.4 Å (30).

Molecular mechanics energies calculated for different sequences and structures were converted to relative unfolding free energies by using the thermodynamic cycle of Harbury *et al.* (12), with experimental helix propensities taken from ref. 31. This is described in more detail in the supporting information on the PNAS web site.

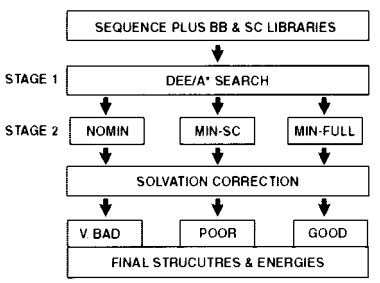

Fig. 4. Flow chart describing the computational methods. BB, backbone; SC, side chain.

Results

Design of a Model Heterodimeric Coiled Coil. Our computational method was designed to predict structures and relative unfolding free energies for parallel two-stranded coiled coils with different core sequences. Experimental testing of the calculations required a series of peptides that (*i*) span a range of stabilities and (*ii*) remain dimeric and parallel across this range. We chose to study a set of heterodimeric complexes to address factors that influence coiled-coil partner selection.

To meet these criteria, we designed a series of heterodimeric coiled coils that we denote GABH, for *G*CN4 *A*cid/*B*ase *H*eterodimer (Fig. 1). The stability of members of the family can be modulated through the substitution of different hydrophobic residues (Leu, Ile, Val) at positions **d**12 and **a**16. Heterodimerization is enforced by oppositely charged residues in the **e** and **g** positions of partnering strands (14). With this design, three ACID peptides with Glu at all **e** and **g** positions (A_{LL}, A_{IV}, A_{LI}) and two BASE peptides with Lys at all **e** and **g** positions (B_{LL} and B_{IV}) can be combined to generate six different heterodimers.

Ensuring the specific formation of dimers, rather than of trimers or tetramers, presents a special challenge in designed coiled coils. Destabilizing mutations in the core of coiled coils frequently lead to a change in oligomerization state (32). The GABH peptides contain several features that promote dimer specificity. First, the opposite charges on the ACID and BASE strands favor heterodimers over homo- and heterotrimers, in which similar-charged residues would be opposed at a minimum of one interface. Second, Asn is included at position **a**30 and Cys [which can form a disulfide bond with favorable geometry in a dimer (33)] at position **d**33 to promote dimer specificity (34). When oxidized, Cys at **d**33 also enforces a parallel orientation of the two strands and allows easy purification of covalently linked heterodimers.

The **b**, **c**, and **f** positions of GABH were taken from the sequence of GCN4-p1, with the exception of position **b**17, which was changed from Tyr to Trp. Many peptides incorporating surface residues from GCN4-p1 crystallize readily (see ref. 35 and refs. therein).

Characterization of Heterodimer Variants. The six GABH heterodimers are ≈100% helical and have CD spectra characteristic of coiled coils (Fig. 2*A*). Analytical ultracentrifugation data for all six peptides fit well to equations describing a single molecular species in solution (Fig. 2*B*). The observed molecular weights are in good agreement with those expected for the disulfide-linked dimer (for each peptide, the average mass ratio observed/calculated = 1.1/1). The relative stabilities of the six two-stranded coiled coils were determined by chemically induced unfolding with Gdm·HCl (Fig. 2*C*). Unfolding was fully reversible. The heterodimer stabilities span a range of 6.6 kcal/mol (Table 1).

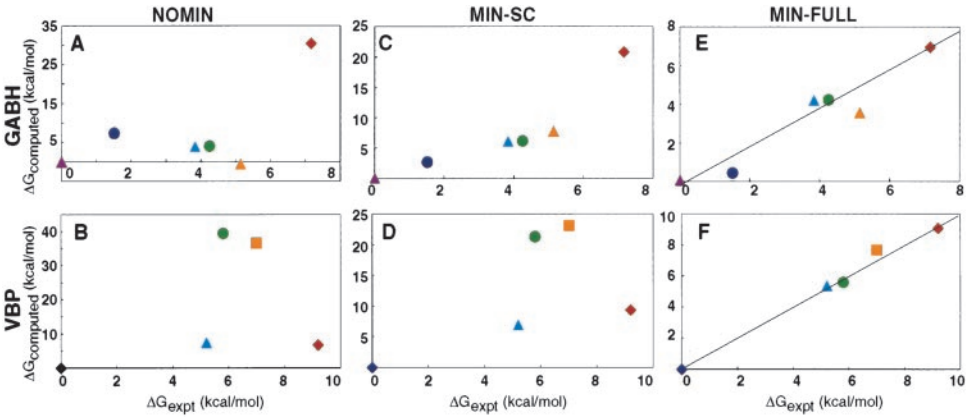

Fig. 5. Comparison of experimental and calculated relative free energies of unfolding. All data are reported in kcal/mol relative to the most stable peptide. Note the different scales for different plots. *A*, *C*, and *E* give results for GABH peptides, and *B*, *D*, and *F* give results for VBP derivatives: (*A* and *B*) NOMIN calculations followed by a solvation correction; (*C* and *D*) MIN-SC calculations plus a solvation correction; (*E* and *F*) MIN-FULL calculation. Colors are as in Fig. 2 for GABH. For VBP d position mutants, L = dark blue, M = light blue, I = green, V = orange, A = red.

Crystallography. We solved the crystal structures of $A_{LL}B_{LL}$, $A_{LI}B_{LL}$, and $A_{IV}B_{LL}$. All three peptides crystallized under similar conditions with the same packing of three copies of the dimer per asymmetric unit. The $A_{IV}B_{LL}$ structure is of excellent quality (Table 2). The $A_{LL}B_{LL}$ and $A_{LI}B_{LL}$ structures have significantly higher average B factors (45 Å^2 for $A_{LL}B_{LL}$ and 43 Å^2 for $A_{LI}B_{LL}$, compared with 28 Å^2 for $A_{IV}B_{LL}$), and poorer refinement statistics.

The peptides all adopt the expected structure of a two-stranded parallel heterodimeric coiled coil. The C terminus is ordered in some copies of the structures and shows how a disulfide bond can be accommodated at the **d** position of a coiled coil with both Cys residues occupying a χ_1=180° rotamer. Most, although not all, of the **e**- and **g**-position residues form salt-bridging interactions across the hydrophobic interface, as anticipated in the design. In several cases, the density for these side chains is missing, consistent with a dynamic structure observed by NMR for **e**- and **g**-position residues in a similar peptide (36). For $A_{IV}B_{LL}$ and $A_{LL}B_{LL}$, variations in core packing are found among the three copies of the dimer in the asymmetric unit (Table 3). Fig. 3*B* shows one arrangement of residues found at position **d**12 in the core of $A_{IV}B_{LL}$. In $A_{LI}B_{LL}$ and $A_{LL}B_{LL}$ a few core sites have high B factors (Table 3).

Comparison of Computed and Experimental Structures and Energies. Computational predictions of the structure and unfolding energy (relative to that of $A_{LI}B_{LL}$) were carried out for each of the experimentally characterized peptides. In addition, we performed calculations on a series of homodimeric coiled coils derived from VBP that have been characterized by Moitra *et al.* (21).

The calculations consist of two stages designed first to limit and subsequently to refine the protein conformational space (Fig. 4). Stage 1 (NOMIN) involves an intensive search through a discrete space comprising idealized two-stranded coiled-coil backbones and canonical side-chain rotamers (26). The DEE and A* algorithms are used to reduce the set of 5×10^9 possible conformations that can be constructed from these choices to a subset of up to 3,500 candidates to be further evaluated (24). In Stage 2, a molecular mechanics minimization procedure is added to refine the Stage 1 candidate structures. We tested two different approaches. In the first, only the side chains are minimized on a fixed backbone (MIN-SC). In the second, flexibility is introduced for both the side-chain and backbone atoms (MIN-FULL) (11).

As described below, structures and relative stabilities predicted by using the NOMIN or MIN-SC methods differ significantly from experimental data. The MIN-FULL calculation, however, yields relative stabilities that agree very well with experiment and structures that superimpose on the x-ray structures to within 0.7 Å.

Fig. 5 shows predicted relative stabilities of the different GABH and VBP structures and compares these with experimental values. In Fig. 5 *A* and *B*, the NOMIN energies do not agree with experiment—the stability of $A_{IV}B_{IV}$ is greatly underestimated, and that of $A_{LI}B_{IV}$ overestimated. The stability of the **d** = Ile and **d** = Val mutants of VBP are also significantly underestimated. MIN-SC improves the agreement of computed and measured stabilities, particularly for the GABH peptides. In Fig. 5*C*, the overall ordering of stabilities is now correct, although that of $A_{IV}B_{IV}$ is still underestimated by more than 13 kcal/mol. The VBP energies in Fig. 5*D* are not correctly ordered and show large errors.

Our best predictions come from calculations that use full-backbone and side-chain minimization (MIN-FULL). The predicted unfolding free energy differences for MIN-FULL calculations are compared with experiment in Fig. 5 *E* and *F*. Quantitative agreement is excellent, with an average error of less than 0.7 kcal/mol for the GABH series and 0.5 kcal/mol for the VBP mutants.

Side-chain rotamers observed in the crystal structures of $A_{LI}B_{LL}$, $A_{LL}B_{LL}$, and $A_{IV}B_{LL}$ are compared with rotamers predicted by each type of calculation in Table 3. The NOMIN calculation ranks many of the experimentally observed conformations poorly (see the last column of Table 3), and predicts them to be high in energy relative to other structures (second to last column). The MIN-SC calculation performs better, ranking some of the experimentally observed rotamer choices among the top five predictions. The

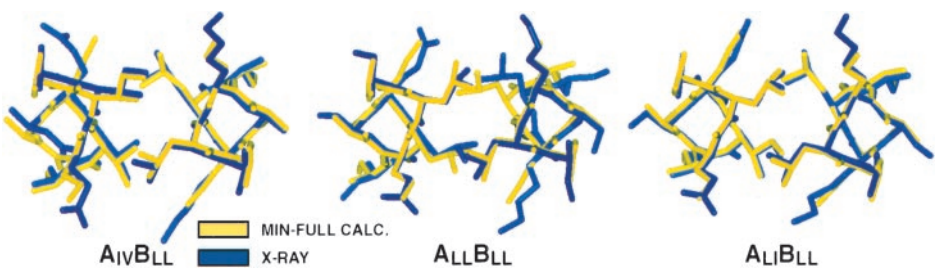

Fig. 6. Superposition of the x-ray and MIN-FULL-calculated structures. The structures shown correspond to the first entries in Table 3 for each peptide. The view is from the C terminus of the peptide and includes positions a16 and d12.

energy gaps between the best predicted structures and several of the experimentally observed rotamer choices are a few kcal/mol using the MIN-SC method. The MIN-FULL calculation shows the best performance. This method ranks all but one of the experimentally observed rotamer combinations within the top five structures predicted by the calculation. The one exception is the third copy of $A_{LL}B_{LL}$, which adopts a strained conformation for χ_1 of position **d**12 and is ranked poorly in all calculations (see Table 3). The energy differences between the rest of the experimentally observed conformations and the best structure predicted by MIN-FULL are less than 1 kcal/mol.

Thus, within a 0.5–1.0 kcal/mol error, the MIN-FULL calculation identifies the correct rotamers. Furthermore, the calculations suggest there are multiple low-energy packing states for some of these peptides, which is confirmed by the observation of alternate packing arrangements in different copies of the experimental structures.

Fig. 6 shows structures from the MIN-FULL calculation superimposed with experimental structures for $A_{LL}B_{LL}$, $A_{IV}B_{LL}$, and $A_{LI}B_{LL}$. The most frequently observed or best-defined rotamer configuration from the crystal structure is shown, along with the calculated structure that matches it (see Table 3). The agreement is very good for all peptides —the average rms deviation for all nonhydrogen atoms included in the model is <0.7 Å.

Discussion

The coiled coil is a common and important protein interaction motif that is involved in a wide variety of biological processes. As a step toward our goal of predicting coiled-coil interactions from sequence data, we have developed a computational method that can predict the influence of hydrophobic core mutations on interaction selectivity. As seen in Fig. 5 *E* and *F*, our computational method makes quantitatively correct predictions about the relative stabilities of different coiled-coil sequences. This is accomplished with a simple energy function that considers only strain energy, van der Waals packing energy, a solvation term, and amino acid helix propensities.

The calculations can be used to predict coiled-coil partnering preferences that arise from core packing. Among the GABH peptides, for example, the interaction of B_{LL} with A_{LI} is predicted to be $\approx$4 kcal/mol more favorable than the interaction with A_{IV}, and the interaction of B_{IV} with A_{LL} is predicted to be $\approx$3 kcal/mol more favorable than the interaction with A_{IV}, consistent with experimental measurements. Residues that are known to favor trimer or tetramer formation over dimerization (such as Ile or Val at the **d** position) are found, as anticipated, to destabilize the heterodimers studied here (32).

The calculated stability of $A_{LI}B_{IV}$, which is 1.5 kcal/mol greater than that observed experimentally, stands out as the largest error in the MIN-FULL calculations. We were unable to crystallize $A_{LI}B_{IV}$ for comparison with the computed structure. However, the predicted choice of rotamers includes the selection of a statistically rare conformation for Ile at position **a**16 (χ_1 = +; χ_2 = trans) (26). For calculations modeling an isolated Ile on the surface of a coiled coil, this rare rotamer is predicted to be more favorable (by 0.9 kcal/mol) than the most common Ile rotamer. Interestingly, the relative stability of the best predicted structure that does not contain this rare rotamer agrees very well with experiment (error = 0.6 kcal/mol). It is possible that the stability of the χ_1 = +; χ_2 = trans rotamer is overestimated by the CHARMM PARAM19 force field.

In conclusion, we consider some differences between protein structure prediction and protein design. In their protein design work, Dahiyat and Mayo have shown a good correlation of predicted stability with experimental melting temperature for variants of the coiled coil GCN4-p1, without using any minimization (37). For the GABH and VBP proteins, however, we find that minimization is essential to good performance. The two results differ because the Dahiyat side-chain identities and conformations were *selected* to be compatible with the GCN4-p1 backbone structure in the design process. Consequently, the side-chain and backbone geometries are unlikely to be severely strained, and relaxation is expected to be much less important. Many of the amino acid substitutions modeled in this work are not easily accommodated in a fixed-backbone framework, and relaxation is critical. The need to accommodate strain in structure prediction underscores a major difference between prediction and design.

Predicting experimental energy differences to high resolution remains a daunting challenge. Our work on coiled coils demonstrates the importance of including conformational flexibility, here introduced via minimization, in models constructed for this purpose. Future work will have to extend specialized methods for addressing backbone flexibility in coiled coils to other more general protein folds.

We thank Michael Burgess and James Pang (Whitehead Institute) for peptide syntheses, Justin Caravella and Ian Chan (Massachusetts Institute of Technology) for programming contributions, Leslie Gaffney for help with the manuscript, and Craig Ogata and the staff at beamline X4A at the National Synchrotron Light Source for their assistance. We thank members of the Kim and Tidor labs, especially Justin Caravella, Mona Singh, David Akey, Sam Sia, and John Newman, for scientific input at many levels. This work used the W. M. Keck Foundation X-Ray Crystallography Suite at the Whitehead Institute and was funded by National Institutes of Health Grant GM 44162. A.E.K. was supported by the Helen Hay Whitney Foundation and by a fellowship from the Merck/Massachusetts Institute of Technology collaboration program.

1. Newman, J. R., Wolf, E. & Kim, P. S. (2000) *Proc. Natl. Acad. Sci. USA* **97,** 13203–13208.
2. Burkhard, P., Strelkov, S. V. & Stetefeld, J. (2001) *Trends Cell Biol.* **11,** 82–88.
3. Hurst, H. C. (1995) *Protein Profile* **2,** 101–168.
4. Eckert, D. M. & Kim, P. S. (2001) *Annu. Rev. Biochem.* **70,** 777–810.
5. Berger, B., Wilson, D. B., Wolf, E., Tonchev, T., Milla, M. & Kim, P. S. (1995) *Proc. Natl. Acad. Sci. USA* **92,** 8259–8263.
6. Wolf, E., Kim, P. S. & Berger, B. (1997) *Protein Sci.* **6,** 1179–1189.
7. Beck, K., Dixon, T. W., Engel, J. & Parry, D. A. (1993) *J. Mol. Biol.* **231,** 311–323.
8. Vinson, C. R., Hai, T. & Boyd, S. M. (1993) *Genes Dev.* **7,** 1047–1058.
9. Moll, J. R., Olive, M. & Vinson, C. (2000) *J. Biol. Chem.* **275,** 34826–34832.
10. Singh, M. & Kim, P. S. (2001) in *5th Annual International Conference on Computational Molecular Biology* (Association for Computing Machinery, Montreal), pp. 279–286.
11. Harbury, P. B., Tidor, B. & Kim, P. S. (1995) *Proc. Natl. Acad. Sci. USA* **92,** 8408–8412.
12. Harbury, P. B., Plecs, J. J., Tidor, B., Alber, T. & Kim, P. S. (1998) *Science* **282,** 1462–1467.
13. Crick, F. H. (1953) *Acta Crystallogr.* **6,** 685–689.
14. O'Shea, E. K., Lumb, K. J. & Kim, P. S. (1993) *Curr. Biol.* **3,** 658–667.
15. Edelhoch, H. (1967) *Biochemistry* **6,** 1948–1954.
16. Laue, T. M., Shah, B. D., Ridgeway, T. M. & Pelletier, S. L. (1992) in *Analytical Ultracentrifugation in Biochemistry and Polymer Science* (R. Soc. Chem., Cambridge, U.K.), pp. 90–125.
17. Otwinowski, Z. (1993) in *Data Collection and Processing*, eds. Sawer, L., Isaacs, N. & Bailey, S. (Science and Engineering Research Council, Daresbury Laboratory, Warrington, U.K.), pp. 55–62.
18. Collaborative Computational Project, no. 4 (1994) *Acta Crysttallogr. D* **50,** 760–763.
19. Jones, T. A., Zou, J. W., Cowan, S. & Kjeldgaard, M. (1991) *Acta Crystallogr. A* **47,** 110–119.
20. Brunger, A. T., Adams, P. D., Clore, G. M., De, L. W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., *et al.* (1998) *Acta Crystallogr. D* **54,** 905–921.
21. Moitra, J., Szilak, L., Krylov, D. & Vinson, C. (1997) *Biochemistry* **36,** 12567–12573.
22. Desmet, J., De Maeyer, M., Hazes, B. & Lasters, I. (1992) *Nature (London)* **356,** 539–542.
23. Goldstein, R. F. (1994) *Biophys. J.* **66,** 1335–1340.
24. Leach, A. R. & Lemon, A. P. (1998) *Proteins* **33,** 227–239.
25. Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S. & Karplus, M. (1983) *J. Comp. Chem.* **4,** 187–217.
26. Dunbrack, R. L., Jr. & Karplus, M. (1993) *J. Mol. Biol.* **230,** 543–574.
27. Mendes, J., Baptista, A. M., Carrondo, M. A. & Soares, C. M. (1999) *Proteins* **37,** 530–543.
28. Nilsson, L. & Karplus, M. (1986) *J. Comput. Chem.* **7,** 591–616.
29. Wesson, L. & Eisenberg, D. (1992) *Protein Sci.* **1,** 227–235.
30. Lee, B. & Richards, F. M. (1971) *J. Mol. Biol.* **55,** 379–400.
31. O'Neil, K. T. & DeGrado, W. F. (1990) *Science* **250,** 646–651.
32. Harbury, P. B., Zhang, T., Kim, P. S. & Alber, T. (1993) *Science* **262,** 1401–1407.
33. Zhou, N. E., Kay, C. M. & Hodges, R. S. (1993) *Biochemistry* **32,** 3178–3187.
34. Lumb, K. J. & Kim, P. S. (1995) *Biochemistry* **34,** 8642–8648.
35. Akey, D. L., Malashkevich, V. N. & Kim, P. (2001) *Biochemistry* **40,** 6352–6360.
36. Marti, D. N., Jelesarov, I. & Bosshard, H. R. (2000) *Biochemistry* **39,** 12804–12818.
37. Dahiyat, B. I. & Mayo, S. L. (1996) *Protein Sci.* **5,** 895–903.