Relative stability of *de novo* four-helix bundle proteins: Insights from coarse grained molecular simulations

Giovanni Bellesia,* Andrew I. Jewett, and Joan-Emma Shea

Department of Chemistry and Biochemistry and Department of Physics, University of California Santa Barbara, Santa Barbara, California 93106, U.S.A.

Received 27 October 2010; Revised 4 February 2011; Accepted 6 February 2011

DOI: 10.1002/pro.605

Published online 22 February 2011 proteinscience.org

Abstract: We use a recently developed coarse-grained computational model to investigate the relative stability of two different sets of *de novo* designed four-helix bundle proteins. Our simulations suggest a possible explanation for the experimentally observed increase in stability of the four-helix bundles with increasing sequence length. In details, we show that both short subsequences composed only by polar residues and additional nonpolar residues inserted, via different point mutations in ad hoc positions, seem to play a significant role in stabilizing the four-helix bundle conformation in the longer sequences. Finally, we propose an additional mutation that rescues a short amino acid sequence that would otherwise adopt a compact misfolded state. Our work suggests that simple computational models can be used as a complementary tool in the design process of *de novo* proteins.

Keywords: four-helix bundle; coarse-grained model; sequence length; folding properties; sequence optimization

Introduction

The interactions between polar, nonpolar amino acids, and the water solvent largely influence the structural properties and the thermodynamic stability of globular proteins.^{1–4} This observation has led to a strategy for the design of *de novo* proteins based on the choice of amino acid sequences whose pattern of polar and nonpolar amino acids insures that hydrophobic residues will be buried on folding into the target structure.^{3–12} Other considerations, such as secondary structure propensity, geometric packing, shape complementarity, salt-bridge specificity, and negative design can also play a role in further narrowing (optimizing) the sequences considered in the design process.^{10,13–18}

Hecht and collaborators used a combinatorial approach to design a set of *de novo* amino acid sequences with the four-helix bundle as the target structure.^{7,11} The helical regions of those sequences had a periodicity of polar and nonpolar amino acids matching the periodicity of an amphiphilic α -helical structure (3.6 residues per helical turn), that is, a nonpolar residue placed every three or four positions. The combinatorial design of the sequences was achieved by randomly choosing polar and nonpolar amino acids from the sets (Lys, His, Glu, Gln, Asp, Asn) and (Met, Leu, Ile, Val, Phe), respectively. Only the turn and capping regions were defined explicitly, that is, they were identical for all the sequences in the library.^{7,11} The amphiphilic pattern of the amino acids in the helical regions is shown in Figure 1 (sequence F14). This first “library” of 74-residue long sequences was partially successful in generating soluble globular structures with native-like properties such as resistance to proteolytic degradation, high α -helical content (probed by CD spectroscopy), and cooperative denaturation (as a function of both temperature and denaturant concentration).

Grant sponsor: NSF; Grant number: #0642086; Grant sponsor: David and Lucile Packard Foundation.

*Correspondence to: Giovanni Bellesia, T-6 and CNLS, Los Alamos National Laboratory, 87545 Los Alamos, NM.
E-mail: gbellesia@lanl.gov; giovanni.bellesia@gmail.com

Figure 1. Sequences considered in this study. Legend: black circles = nonpolar (N), white circles = polar (P), gray thick lines = turns (T). The number in the sequence name corresponds to the length of a single strand. The amphiphilic pattern in sequence F14 has been studied experimentally by Hecht and coauthors. The amphiphilic pattern in sequence F21 has been studied by Hecht and coauthors and is generated by adding seven additional amino acids to sequence F14. In our study, we also consider sequences F16 and F18, which have been generated by adding to F14 the first 2 and the first four amino acids of the seven additional amino acids in F21. The sequence F14B has been generated from F14 via a single “amino acid” permutation within each strand.

However, the majority of sequences in the library formed structures mostly resembling molten globules.^{7,19–22} The limited length of the helical strands in the 74-residue sequences (14 residues per helical strand) was identified as a possible reason for the partial success of the binary pattern-based design strategy. Hence, a second combinatorial library was considered where the helical regions were 50% longer (21 residues per helical strand). Several proteins from the new library showed a generally improved stability and a more native-like nature when compared with the folded structures obtained from the first library.^{11,23,24} The amphiphilic pattern of the amino acids in the helical regions for this second library is shown in Figure 1 (sequence F21).

In a previous study, we have used a coarse-grained computational model to explore the balance of local secondary structural preferences, and nonlocal tertiary hydrophobic interactions, in determining the fold of four-helix bundle proteins.^{3,25} The poly-

peptide chain is characterized as a linear heteropolymer consisting of mutually repulsive and mutually attractive beads, representing polar and nonpolar residues, respectively (See Methods section for details).

In the current study, we further test the predictive qualities of our model by using it to investigate the relative stability of those two combinatorial protein libraries.^{11,23,24} In particular, we try to answer the following questions: (1) Can we reproduce the experimental results? (improved stability and more native-like nature for the longer sequences when compared with the shorter ones). (2) Can we identify the features in the F21 sequence that contribute to its improved folding behavior? (3) Is there a possibility to further optimize sequence F14 in terms of its polar, nonpolar pattern?

To address question (1), we analyze both the F14 and F21 sequences together with two additional sequences of intermediate lengths (F16 and F18). We observe, in agreement with experimental results,

an improved stability of the four-helix bundle conformations in the longer sequences. To address questions (2) and (3) we designed a number of a sequence “mutations”: (i) three different mutations for the F21 sequence, which show a partial role of additional polar residues in stabilizing the four-helix fold as well as the possibility to further optimize F21 by inserting nonpolar residues in ad hoc positions and (ii) one mutation for the F14 sequence, which leads to an equilibrium structure population dominated by stable four-helix bundles.

Results and Discussion

We analyze the protein structure populations at $T = 0.22$ (lowest temperature) using a clustering technique based on the root mean squared deviation (RMSD) from the lowest potential energy structure at $T = 0.22 \text{ } \epsilon/k_B$. For each system considered in our simulations, we first calculate the RMSD probability distribution and consider the number of peaks N_p as an estimation for the number of different clusters (families) of structures. Finally, we use a k -means clustering algorithm²⁶ considering N_p as the number of clusters. The iterative algorithm efficiently assigns different structures to different clusters so that the structural variation within each cluster reaches its minimum value.

The results of the clustering analysis for the three most populated clusters for each sequence are shown in Table I. The data show, in agreement with protein design experiments, that extending the strand length from 14 to 21 amino acids significantly increases the population of well-folded four-helix bundles. Although sequences F14 and F16 are predominantly found in an ensemble of misfolded bundles, sequence F18 and (more clearly) sequence F21 show the emergence and the dominance of the four-helix bundle as the preferred fold at low temperatures.

Two representative structures for the dominant misfolded bundles in F14 and F16 are shown at the top of Figure 2. Both structures are compact and stabilized by multiple contacts between the N beads. Although these structures retain some degree of helical content (data not shown), it is apparent that the “hydrophobic” packing of the N beads does not resemble the one found in typical four-helix bundle structures. Conversely, in the middle line of Figure 3, we show two representative structures for the dominant four-helix bundles observed in F18 and F21. In more details, we found that the statistical weight (fraction) of the four-helix bundle population increases from <0.01 in F14, to ≈ 0.12 in F16, ≈ 0.55 in F18, and to ≈ 0.72 in F21.

The simulation results discussed so far clearly show a generally improved stability and a more native-like nature for sequence F21 when compared with sequence F14. These findings are in agreement

Table I. Structure Population for the Five Sequences in Figure 1 at $T = 0.22$ (Low Temperature)

System	Cluster #	Stat. weight (%)	Structure
F14	1	35.3	Misfolded bundle
	2	24.6	Misfolded bundle
	3	20.8	Misfolded bundle
F16	1	33.2	Misfolded bundle
	2	30.8	Misfolded bundle
	3	14.2	Misfolded bundle
F18	1	37.2	Misfolded bundle
	2	24.8	Helical bundle
	3	17.4	Helical bundle
F21	1	50.0	Helical bundle
	2	22.4	Helical bundle
	3	15.2	Shifted helical bundle
F14B	1	27.2	Helical bundle
	2	25.0	Helical bundle
	3	17.3	Helical bundle

For each sequence, only the statistical weights for the three largest cluster are shown. Misfolded bundle: a compact state with “non-helical” pairwise contacts between beads in the protein core. Helical bundle: a stable four-helix bundle. Shifted helical bundle: a four-helix bundle where one of the helical strands is shifted along the bundle main axis and forms only a fraction of the native “helical” contacts.

with protein design experimental results obtained by Hecht and coauthors in their study on two different combinatorial libraries based on the F14 and F21 sequences, respectively.^{11,23,24}

In addition to the four sequences F14, F16, F18, and F21, we also studied sequence F14B, which was obtained from sequence F14 via a single residue permutation (swap) within each strand (see Fig. 1). We found that F14B folds into a stable, four-helix bundle (statistical weight ~ 1.0) at $T = 0.22$ (see Table I and Fig. 3). The substantial differences in the equilibrium structure populations between F14 and F14B (see Table I and Fig. 3) are not captured in the two-dimensional view of the helical-wheel diagrams in Figure 3, which show, for both F14 and F14B, a potentially similar burying of the nonpolar residues in the four-helix bundle conformation. A more detailed (3D) analysis of the differences in the equilibrium structure populations relies on the calculation of the close contacts between nonpolar residues (the only attractive residues in our model). A contact is defined by a cutoff distance $c = 7\text{Å}$ ²⁷ and a probability $P \geq 0.5$.

First, we observe that F14 tends to form (on average) a larger number of pairwise contacts between nonpolar beads than F14B (+9%). We then consider the equilibrium configuration phase space of F14B as a representative ensemble for four-helix bundle structures with each helical strand 14 residues long. From this configuration phase space, we calculate the average number of pairwise contacts between nonpolar beads for both F14B and F14 sequences.

Figure 2. Red: polar beads (*P*). Yellow: nonpolar beads (*N*). Blue: neutral (turn) beads (*T*). Top: representative structures for the dominant misfolded bundles in F14 (left) and F16 (right). Both structures are compact and stabilized by multiple contacts between the *N* beads. Mid: representative structures for the dominant four-helix bundles observed in F18 (left) and F21 (right). Bottom: representative structure for the dominant four-helix bundle observed in F14B. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

We find that the sequence F14B forms (on average) ~15% more nonpolar contacts than F14. Finally, a similar calculation considering the equilibrium configuration phase space of F14 (misfolded bundles) and both F14 and F14B sequences shows that sequence F14 forms (on average) a larger number of nonpolar contacts respect to F14B. These results suggest that although both structures have the right polar–nonpolar pattern to potentially bury nonpolar residues in a four-helix bundle conformation, sequence F14B does a much better job than F14 in stabilizing the native contacts typical of a four-helix bundle.

To understand how the presence of additional beads in the longer sequences influence both the

structure and the internal dynamics of the entire chain, we first analyze the cross-correlation of the internal motion of the beads (position fluctuations) for the different sequences via the cross-correlation matrix c_{ij} . The cross-correlation matrix for the internal motion is defined as:

$$c_{ij} = \frac{\langle (\mathbf{r}_i - \langle \mathbf{r}_i \rangle) \cdot (\mathbf{r}_j - \langle \mathbf{r}_j \rangle) \rangle}{\sqrt{\langle (\mathbf{r}_i - \langle \mathbf{r}_i \rangle)^2 \rangle} \sqrt{\langle (\mathbf{r}_j - \langle \mathbf{r}_j \rangle)^2 \rangle}}, \quad (1)$$

where $\mathbf{r}_i$ is the position of the *i*th bead and $\langle \rangle$ denotes an ensemble average.²⁸ Rotation and translation motions have been removed by minimizing

Figure 3. Hypothetical 4-helix bundles as viewed from the bundle axis (turns omitted), showing only the C_α atom positions. Each helix is represented by an ideal “helix-wheel” containing 3.6 residues per turn. In these ideal bundles, we assume that the nonpolar residues (black and gray with white numbers) face inward, that all four helical axis are aligned with the bundle axis, and that the 4-helix bundles alternate direction, in and out of the page (U-bundle topology³¹). The C-terminus (residue number 1) is closest to the viewer and residues that are further away fade into the background.

the root mean square deviation from a reference structure.²⁸ The cross-correlation maps for the five sequences are shown in Figure 4. In F14, the correlation map does not show any robust pattern aside from a persistent positive correlation between proximal beads (diagonal region). In more details, the map shows that the four strands do not behave as separate structural domains and that there are no robust correlation patterns among beads that are far

away from each other in the sequence. In F16, three of the four strands (first, third, and fourth) appear to behave as separate domains, whereas the correlation map in the regions away from the diagonal show the emergence of some patterning in terms of both positive and negative correlation. The map for sequence F18 shows a well-defined intrastand (intradomain) correlated motion. However, the “communication” between different strands (interdomain

Figure 4. Color-based map of the cross-correlation matrix calculated from Eq. (1) for the five sequences analyzed in our study. The cross-correlation matrix as defined in Eq. (1) contains information regarding the correlation in the internal motion of the beads (position fluctuations). Colors corresponding to positive values define regions of positive correlation, while colors corresponding to negative values define regions of negative (anti) correlation. 0 = uncorrelated. Different areas of the map with high positive correlation typically correspond to separate secondary structure domains (i.e., the four-helical strands). These domains are well defined in sequences F21 and F14B and only partially defined in sequence F18. See Section II for further details. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

correlation) still shows some degree of disorder, which is related to the non-negligible statistical weight of misfolded conformations in F18 (see Table I). In the longest sequence (F21), the strand domains are well-defined (diagonal region – positive correlation) and so is the correlation between them. The first helical strand is negatively correlated with the second and the third strands (which are posi-

tively correlated to each other) and positively correlated with the fourth strand (which is also negatively correlated to strands 2 and 3). In other words, strands 1 and 4 act as a single structural domain, which is negatively correlated to another structural domain composed by strands 2 and 3. In addition, the correlation map for F14B shows that the internal motion within each of the four strands

is positively correlated (similarly to what we found in F18 and F21), whereas the correlation between the different helical strands is uniformly negative.

The correlation maps for both F21 and F14B (and partially for F18) show that to obtain well-folded helical bundles (i) the four strands have to behave as separate structural domains and (ii) the cross-correlation between the different domains has to be well defined. However, cross-correlation data still do not explain what are the reasons for the improvement in the folding behavior of the longer sequences.

As the fraction of polar *P* beads (with respect to the number of beads in the strands, i.e., excluding the turns) increases from 0.57 in F14 to 0.59 in F16, 0.61 in F18, and to 0.62 in F21, one possibility could be that this higher fraction of polar beads in the longer sequences contributes to stabilize the native helical-bundle state. The importance of polar residues in stabilizing the protein's native state has been discussed in both experimental and theoretical studies.^{2,15–17} In particular, Hill and DeGrado¹⁶ showed that polar residues can play an essential role in stabilizing the native state by increasing the free energy gap between the native state itself and competing non-native, misfolded states.

To investigate the role of the polar beads in stabilizing F21 helical native structure, we run additional REX-MD simulations on three different “mutants” of sequence F21 obtained by increasing the number of nonpolar residues while maintaining a “helical” polar-nonpolar pattern (a nonpolar residue placed every three or four positions). The “mutations” together with the F21 “wild-type” are shown in Figure 5. In the first mutant (*mut_A*), we substitute two polar beads and disrupt two of the four “polar stretches” composed each by three neighboring polar beads (see Fig. 1). We observe that in *mut_A* the four-helix bundle population in the conformational phase space at $T = 0.22$ (lowest temperature) decreases by about 18% with respect to F21 (from ≈ 0.72 to ≈ 0.59). The value of ~ 0.59 is very similar to the fractional value of the four helix bundle population in F18 (~ 0.55) where, as in *mut_A*, only two polar stretches are found. This result seems to confirm a partial role of the three beads polar stretches in stabilizing the four-helix bundle structures in F21. However, we observe a different behavior in *mut_B* and *mut_C* where we still disrupt the polar stretches, this time by substituting four polar beads so as to have four additional pairs of subsequent nonpolar beads. In *mut_B*, the additional attractive interactions between the four new nonpolar pairs (16–17, 29–30, 64–65, and 77–78 in Fig. 1) overcome the absence of the polar stretches and overstabilize the four-helix bundle population with respect to F21 (fraction population changes from ≈ 0.72 in F21 to ≈ 0.95). It is worth noticing that the new polar pairs

Figure 5. Legend: black circles = nonpolar (*N*), white circles = polar (*P*), gray thick lines = turns (*T*). F21 sequence (wild-type) and three mutants analyzed in our study. Larger black circles represent the additional nonpolar beads substituting polar beads in the respective mutated sequences. Double arrows indicate a simple change of position between two neighboring beads.

in *mut_B* have the same kind of structural “overlap” found in F14B (see Fig. 1). When the nonpolar pairs do not “overlap” completely as in *mut_C*, the absence of the polar stretches, which contribute to stabilizing the four-helix bundle population is almost exactly counterbalanced by the attractive interactions between the nonpolar pairs (19–20, 29–30, 67–68, and 77–78 in Fig. 1). Indeed, the fraction four-helix bundle population is ≈ 0.73 against ≈ 0.72 for F21.

Conclusions

Using computer simulations on a minimalistic C_α model, we were able to reproduce the length-dependent increase in the stability of the four-helix bundles observed in experiments. The net increase in stability of the folded state is reflected in the high correlation of motions between residues in the longer helix bundles, F18 and F21 (see Fig. 4). Our model also allowed us to explore the behavior of new sequence templates, which have not yet been tested in experiments. In details, we analyzed, first, a set of mutated sequences to investigate the role of both short subsequences of polar residues and interacting pairs of nonpolar residues in stabilizing the four-helix bundle conformation in the longer sequences studied by Hecht and coworkers (F21). In addition, we constructed a new sequence (F14B) obtained from the original four-helix bundle studied by Hecht and coworkers (F14) by a simple residue permutation (see Figure 1). In simulations, F14B folds into a stable four-helix bundle, whereas F14 is mainly found in an

ensemble of misfolded compact states. Our simulations suggest that simple coarse-grained models can be used in experimental protein design studies as a complementary tool for structure optimization.

Methods

Coarse-grained model

Each amino acid in the model protein is represented by a single bead of diameter σ that can be polar (P), nonpolar (N) or “neutral” (T) for the beads in the turn regions. Both P and T beads are mutually repulsive, and they differ only in the parametrization of their respective dihedral potentials.²⁵ Each bead has mass m (m is the mass unit) and is connected to the neighboring beads in the peptide chain by harmonic “springs.”^{25,29–32} The force field used in our study has the following functional form:

$$\sum_{i \neq j} 4\epsilon \left[\left(\frac{\sigma}{r_{ij}} \right)^{12} - \lambda_{ij} \left(\frac{\sigma}{r_{ij}} \right)^6 \right] + \sum_{\text{bonds}} \frac{K_b}{2} (r_{ij} - r_0)^2 + \sum_{\text{angles}} \frac{K_\theta}{2} (\theta_{ij} - \theta_0)^2 + \sum_{\text{dihe}} C [\cos(3\gamma) + \cos(\gamma + \delta)], \quad (2)$$

where $\lambda_{PP} = \lambda_{TT} = \lambda_{PT} = \lambda_{PN} = \lambda_{TN} = 0$, $\lambda_{NN} = 1$, $r_0 = 3.8 \text{ \AA}$, $\lambda_0 = 105^\circ$. Further details on the parametrization of the force field can be found in a previous publication.²⁵ Here, we only mention that the dihedral potential term constraining quadruplets comprising one or more T beads is:

$$0.2\epsilon \cos(3x), \quad (3)$$

whereas for quadruplets composed of N and P beads the dihedral potential is:

$$U_{\text{dihe}}^{\alpha+\beta} = C [\cos(3x) + \cos(x + \delta)], \quad (4)$$

where $C = 1.2\epsilon$ and $\delta = 65^\circ$. This potential term (labeled as α_2 in Ref. 25) has three minima: $\sim 60^\circ$, $\sim 180^\circ$, and $\sim -60^\circ$, correspondent to α helix, β -strand and left helix conformations, respectively. The deepest minimum is the one at $\sim 60^\circ$, that is, the dihedral potential favors the α helix conformation over the β sheet conformation. The energy contribution from the net intrinsic secondary structure propensity in our model can be defined as the relative energy difference between the α helix and the β -strand minima $\Delta U_{\text{dihe}} = 0.17 \epsilon$. An analogous energy contribution from the net intrinsic secondary structure propensity for the 20 natural amino acids has been calculated from a recent snapshot of the PDB repository.²⁵ It has been found that $\Delta U_{\text{dihe}} = 0.17 \epsilon$ is consistent with the results obtained from the PDB repository for values of ϵ up to $\sim 2.9 \text{ Kcal/mol}$.

Sequences

In this study, we analyze different “amino acid” sequences all composed by four identical “strands”

containing P and N beads and connected by short flexible regions (turns) made up by three beads of type T . In Figure 1, we show the five sequences studied in our simulations. Four of these sequences (F14, F16, F18, and F21) enable us to analyze how the peptide chain length affects the folding of denovo designed proteins. Sequences F14 and F21 have been designed and studied experimentally by Hecht and coworkers.^{7,11,19–21,23,24} (See Introduction section for the details of the design experiments). Sequence F21 has been designed adding seven amino acids to each strand of sequence F14.¹¹ In addition to F14 and F21, we study (i) two “intermediate” sequences (F16 and F18), which are generated adding to F14 the first 2 and the first four amino acids of the seven additional amino acids in F21, respectively (see dotted rectangles in Figure 1), and (ii) three different mutants of F21 (see Fig. 5).

Sequence F14B has been generated from F14 via a single “amino acid” permutation within each strand, see double arrows in Figure 1. We chose this “amino acid” permutation to favor the four-helix bundle “hydrophobic packing” by exploiting the polar, nonpolar “helical” pattern $PPNNPPNNPP$.²⁵ In Figure 2, the helix wheel diagrams show that all the sequences studied have the nonpolar residues (black and gray) “buried” from the external environment on folding into a four-helix bundle structure.

Molecular simulations

We performed Replica-exchange Langevin dynamics (REX-LD) simulations for studying the thermodynamics of folding of the model proteins in Figure 1. We ran the REX-LD simulations using 20 replicas with temperatures in the interval $[0.22, 0.55]$ (temperatures are expressed in units of ϵ/k_B). The time step was $ts \approx 0.008\tau$, where $\tau = (m \sigma^2/\epsilon)$ is the time unit. The cutoff for the nonbonded interactions was fixed at 4σ and the damping coefficient in the Langevin integrator was set to $b = 0.8 \tau^{-1}$, swaps between the different replicas were attempted every 2×10^4 time steps, the acceptance ratio varied between 0.4 and 0.6, and the total simulation time for each replica was $1.6 \times 10^6 \tau$. We considered the last 1/3 of the REX-LD “trajectories” as the production runs. The NAMD software³³ was used for all the simulations. The VMD software was used for part of the analysis.³⁴

Acknowledgments

The authors thank Michael Hecht for helpful and constructive suggestions.

References

1. Kauzmann W (1959) Some factors in the interpretation of protein denaturation. *Adv Protein Chem* 14:1–63.
2. Dill KA, Bromberg S, Yue K, Fiebi KM, Yee DP, Thomas PD, Chan HS (1995) Principles of protein

- folding—a perspective from simple exact models. *Protein Sci* 4:561–602.
3. Xiong H, Buckwalter B, Shieh H, Hecht M (1995) Periodicity of polar and nonpolar amino acids is the major determinant of secondary structure in self-assembling oligomeric peptides. *Proc Natl Acad Sci USA* 92: 6349–6353.
 4. West M, Hecht MHM (1995) Binary patterning of polar and nonpolar amino acids in the sequences and structures of native proteins. *Protein Sci* 4: 2032–2039.
 5. Yue K, Dill KA (1992) Inverse protein folding problem: designing polymer sequences. *Proc Natl Acad Sci USA* 89:4163–4167.
 6. Eisenberg D, Wilcox W, Eshital SM, Pryciak PM, Ho SP, DeGrado WF (1986) The design, synthesis, and crystallization of an alpha-helical peptide. *Proteins: Struct Funct Genet* 1:16–22.
 7. Kamtekar S, Schiffer J, Xiong H, Babik J, Hecht M (1993) Protein design by binary patterning of polar and nonpolar amino acids. *Science* 262:1680–1685.
 8. Ho SP, DeGrado WF (1987) Design of a 4-helix bundle protein: synthesis of peptides which self-associate into a helical protein. *J Am Chem Soc* 109:6751–6758.
 9. Regan L, DeGrado WF (1988) Characterization of a helical protein designed from first principles. *Science* 241:976–978.
 10. Kolinski A, Galazka W, Skolnick J (1995) Computer design of idealized beta-motifs. *J Chem Phys* 103: 10286–10297.
 11. Wei Y, Liu T, Sazinsky S, Moffet D, Pelczer I, Hecht M (2003) Stably folded de novo proteins from a designed combinatorial library. *Protein Sci* 12:92–102.
 12. Beasley J, Hecht M (1997) Protein design: the choice of de novo sequences. *J Biol Chem* 272:2031–2034.
 13. Raleigh DP, DeGrado WF (1992) A de novo designed protein shows a thermally induced transition from a native to a molten globule-like state. *J Am Chem Soc* 115:100079–10081.
 14. Raleigh D, Betz S, DeGrado WF (1995) A de novo designed protein mimics the native state of natural proteins. *J Am Chem Soc* 117:7558–7559.
 15. Hill R, Raleigh D, Lombardi A, Degrado W (2000) De novo design of helical bundles as models for understanding protein folding and function. *Acc Chem Res* 33:745–754.
 16. Hill RB, DeGrado WF (2000) A polar, solvent-exposed residue can be essential for native protein structure. *Structure* 8:471–9.
 17. Hill RB, DeGrado WF (1998) Solution structure of 2d, a nativelylike de novo designed protein. *J Am Chem Soc* 120:1138–1145.
 18. Handel T, DeGrado WF (1990) De novo design of a Zn^{2+} -binding protein. *J Am Chem Soc* 112:6710–6711.
 19. Roy S, Ratnaswamy G, Boice J, Fairman R, McLendon G, Hecht M (1997) A protein designed by binary patterning of polar and nonpolar amino acids displays native-like properties. *J Am Chem Soc* 119:5302–5306.
 20. Roy S, Helmer K, Hecht M (1992) Detecting native-like properties in combinatorial libraries of de novo proteins. *Folding Design* 2:89–92.
 21. Roy S, Hecht M (2000) Cooperative thermal denaturation of proteins designed by binary patterning of polar and nonpolar amino acids. *Biochemistry* 39:4603–4607.
 22. Kuwajima K (1989) The molten globule state as a clue for understanding the folding and cooperativity of globular-protein structure. *Proteins* 6:87–103.
 23. Wei Y, Kim S, Fela D, Baum J, Hecht M (2003) Solution structure of a de novo protein from a designed combinatorial library. *Proc Natl Acad Sci USA* 100:13270–13273.
 24. Go A, Kim S, Baum J, Hecht MH (2008) Structure and dynamics of de novo proteins from a designed superfamily of 4-helix bundles. *Protein Sci* 17:821–832.
 25. Bellesia G, Jewett A, Shea J-E (2010) Sequence periodicity and secondary structure propensity in model proteins. *Protein Sci* 19:141–154.
 26. Hartigan JA (1975) *Clustering Algorithms*. New York, NY: Wiley and Sons.
 27. da Silveira CH, Pires DEV, Minardi RC, Ribeiro C, Veloso CJM, Lopes JCD, Meira W, Jr, Neshich G, Ramos CHI, Habesch R, Santoro MM (2009) Protein cutoff scanning: a comparative analysis of cutoff dependent and cutoff free methods for prospecting contacts in proteins. *Proteins* 74:727–43.
 28. Zhou Y, Cook M, Karplus M (2000) Protein motion at zero-total angular momentum: the importance of long-range correlations. *Biophys J* 79:2902–2908.
 29. Honeycutt J, Thirumalai D (1990) Metastability of the folded states of globular proteins. *Proc Natl Acad Sci USA* 87:3526–3529.
 30. Honeycutt J, Thirumalai D (1992) The nature of folded states of globular proteins. *Biopolymers* 32:695–709.
 31. Guo Z, Thirumalai D (1996) Kinetics and thermodynamics of folding of a de novo designed four-helix bundle protein. *J Mol Biol* 263:323–343.
 32. Li MS, Klimov DK, Straub JE, Thirumalai D (2008) Probing the mechanisms of fibril formation using lattice models. *J Chem Phys* 129:175101.
 33. Phillips J, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel R, Kale L, Schulten K (2005) Scalable molecular dynamics with namd. *J Comput Chem* 26:1781–1802.
 34. Humphrey W, Dalke A, Schulten K (1996) Vmd: visual molecular dynamics. *J Mol Graphics* 14:33–38.