

Folding Simulations and Computer Redesign of Protein A Three-Helix Bundle Motifs

Krzysztof A. Olszewski,^{1,2} Andrzej Kolinski,^{1,2} and Jeffrey Skolnick¹

¹Department of Molecular Biology, The Scripps Research Institute, La Jolla, California 92037; ²Department of Chemistry, University of Warsaw, Pasteura 1, 02-093 Warsaw, Poland

ABSTRACT In solution, the B domain of protein A from *Staphylococcus aureus* (B domain) possesses a three-helix bundle structure. This simple motif has been previously reproduced by Kolinski and Skolnick (Proteins 18: 353–366, 1994) using a reduced representation lattice model of proteins with a statistical interaction scheme. In this paper, an improved version of the potential has been used, and the robustness of this result has been tested by folding from the random state a set of three-helix bundle proteins that are highly homologous to the B domain of protein A. Furthermore, an attempt to redesign the B domain native structure to its topological mirror image fold has been made by multiple mutations of the hydrophobic core and the turn region between helices I and II. A sieve method for scanning a large set of mutations to search for this desired property has been proposed. It has been shown that mutations of native B domain hydrophobic core do not introduce significant changes in the protein motif. Mutations in the turn region were also very conservative; nevertheless, a few mutants acquired the desired topological mirror image motif. A set of all atom models of the most probable mutant was reconstructed from the reduced models and refined using a molecular dynamics algorithm in the presence of water. The packing of all atom structures obtained corroborates the lattice model results. We conclude that the change in the handedness of the turn induced by the mutations, augmented by the repacking of hydrophobic core and the additional burial of the second helix N-cap side chain, are responsible for the predicted preferential adoption of the mirror image structure.

© 1996 Wiley-Liss, Inc.

Key words: lattice model of proteins, Monte Carlo method, mirror image fold, multiple mutations, hydrophobic core repacking, turn handedness modification

INTRODUCTION

In many cases, protein molecules in solution spontaneously adopt their unique three-dimensional na-

tive structure during the protein folding process. Since Anfinsen¹ demonstrated that the folding of proteins is reversible under appropriate conditions, it is generally believed that the native structure of proteins in solution corresponds to the thermodynamically stable conformation. Moreover, it is assumed that the primary structure of a protein (i.e., its amino acid sequence) contains all of the information necessary to determine its native structure.¹ However, in general, the prediction of the native conformation of a protein from its primary structure remains an unsolved problem, despite intensive research. Although the mechanism of protein folding is not yet thoroughly understood, many researchers have attempted to design de novo artificial proteins that will adopt a chosen structure in solution.^{2–6} This endeavor is an immensely complicated task. However, de novo design, if successful and followed by experimental verification, would provide insights into the relationship between protein sequence and structure and may especially help to single out those energy terms that are responsible for stabilizing the native state relative to folding intermediates^{2,4} and alternative topologies. This paper proceeds along this line and focuses on the more modest goal of redesigning a native sequence to modify the topology that the mutated protein will adopt. We intended to analyze the B domain of protein A, a simple protein with known structure that has already served as a successful test case for the lattice model of protein folding, which has been developed by Kolinski and Skolnick.⁷ The potential we are currently using is slightly rederived to improve reproducibility of the location of the helical termini. However, before we attempt to modify the native sequence of the B domain of protein A, we first scrutinize the intrinsic robustness of the new lattice model by examining a family of highly homologous domains of protein A that are also believed to fold to the same topology. We will then attempt to understand the relationship between sequence and both alternative forms of the three-helix bundle, i.e., the native to-

Received May 24, 1995; revision accepted December 15, 1995.

Address reprint requests to Jeffrey Skolnick, Department of Molecular Biology, The Scripps Research Institute, 10666 North Torrey Pines Road, MB1, La Jolla, CA 92037.

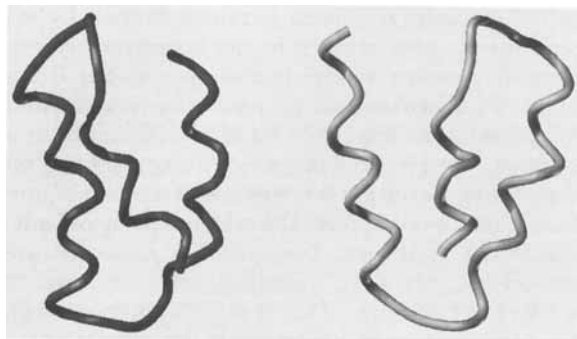


Fig. 1. The native structure (dark gray) and the topological mirror image (light gray) structure of the B domain of protein A. In the topological mirror image the first helix is located on the other side of the plane indicated by the hairpin formed by the helices II and III. Only a C α carbon ribbon tube representation is shown.

pology of the B domain and its topological mirror image (cf. Fig. 1) by studying what factors determine the native fold. The topological mirror image of the global three-helix bundle fold is a structure in which one helix is moved to the other side of the helical hairpin.

The extracellular part of protein A, a cell wall component of *Staphylococcus aureus* that binds to an F_c domain of several immunoglobulins, consists of five highly homologous domains designated E, D, A, B, and C (Table I). The X-ray crystallographic structure of the B domain complexed with the F_c portion of human polyclonal immunoglobulin G indicates the presence of two helices, from Gln10 to Leu18 (helix I) and from Glu26 to Asp37 (helix II), which are packed together in an antiparallel fashion.⁸ The three-dimensional solution structure of the B domain has also been determined by nuclear magnetic resonance (NMR).⁹ The B domain in water forms a stable three-helix bundle motif with helix I (Gln10–His19) tilted with respect to the antiparallel hairpin formed by packed helices II (Glu25–Asp37) and III (Ser42–Ala55). The N-terminal residues up to Glu9 and C-terminal from Gln56 to terminal lysine do not exhibit ordered structure. It has been proposed⁹ that the absence of the third helix in the crystal structure of the B domain complex with immunoglobulin is induced by the crystal contacts and that the native three-helix bundle topology is preserved in the solvated complex. Therefore, we assume that the B domain of *Staphylococcal* protein A has a stable, well-defined, simple native structure, which may serve as a good candidate for the purpose of redesign.

The structures of protein A domains other than the B domain have not been previously reported. However, since they are more than 80% homologous and also bind to immunoglobulin (which at least requires the proper packing of helices I and II), it is relatively safe to assume that the overall structure is conserved within the family of domains of protein

TABLE I. Sequences of Staphylococcal Protein A Domains and of Protein Z (Engineered B Domain of Protein A)*

	Domain					
	E	D	A	B	C	Z
10	Gln	Gln	Gln	Gln	Gln	Gln
11	Gln	Gln	Gln	Gln	Gln	Gln
12	Asn	Ser	Asn	Asn	Asn	Asn
13	Ala	Ala	Ala	Ala	Ala	Ala
14	Phe	Phe	Phe	Phe	Phe	Phe
15	Tyr	Tyr	Tyr	Tyr	Tyr	Tyr
16	Gln	Glu	Glu	Glu	Glu	Glu
17	Val	Ile	Ile	Ile	Ile	Ile
18	Leu	Leu	Leu	Leu	Leu	Leu
19	Asn	Asn	Asn	His	His	His
20	Met	Met	Met	Leu	Leu	Leu
21	Pro	Pro	Pro	Pro	Pro	Pro
22	Asn	Asn	Asn	Asn	Asn	Asn
23	Leu	Leu	Leu	Leu	Leu	Leu
24	Asn	Asn	Asn	Asn	Thr	Asn
25	Ala	Glu	Glu	Glu	Glu	Glu
26	Asp	Ala	Glu	Glu	Glu	Glu
27	Gln	Gln	Gln	Gln	Gln	Gln
28	Arg	Arg	Arg	Arg	Arg	Arg
29	Asn	Asn	Asn	Asn	Asn	Asn
30	Gly	Gly	Gly	Gly	Gly	Ala
31	Phe	Phe	Phe	Phe	Phe	Phe
32	Ile	Ile	Ile	Ile	Ile	Ile
33	Gln	Gln	Gln	Gln	Gln	Gln
34	Ser	Ser	Ser	Ser	Ser	Ser
35	Leu	Leu	Leu	Leu	Leu	Leu
36	Lys	Lys	Lys	Lys	Lys	Lys
37	Asp	Asp	Asp	Asp	Asp	Asp
38	Asp	Asp	Asp	Asp	Asp	Asp
39	Pro	Pro	Pro	Pro	Pro	Pro
40	Ser	Ser	Ser	Ser	Ser	Ser
41	Gln	Gln	Gln	Gln	Val	Gln
42	Ser	Ser	Ser	Ser	Ser	Ser
43	Ala	Thr	Ala	Ala	Lys	Ala
44	Asn	Asn	Asn	Asn	Glu	Asn
45	Val	Val	Leu	Leu	Ile	Leu
46	Leu	Leu	Leu	Leu	Leu	Leu
47	Gly	Gly	Ser	Ala	Ala	Ala
48	Glu	Glu	Glu	Glu	Glu	Glu
49	Ala	Ala	Ala	Ala	Ala	Ala
50	Gln	Lys	Lys	Lys	Lys	Lys
51	Lys	Lys	Lys	Lys	Lys	Lys
52	Leu	Leu	Leu	Leu	Leu	Leu
53	Asn	Asn	Asn	Asn	Asn	Asn

*Only amino acids with well-defined secondary structures are shown.

A. Moreover, to remove an Asn-Gly pair from the native sequence (to facilitate its purification by gene fusion¹⁰), the so-called protein Z has been proposed and subsequently expressed as a single point mutation of the B domain Gly30 to Ala30 (G30A).¹⁰ For protein Z and two of its single point mutants, i.e., (N29A) and (F31A), the unfolding free energies were reported.¹¹ In addition, a solution structure obtained from the NMR¹² reveals a three-helix bundle

topology of protein Z. Therefore, since single point mutations of protein Z are even more homologous to the B domain than domains A, C, D, or E, and are also known to bind to immunoglobulin, we assume that the three-helix bundle topology of the native structure for both protein Z mutants is also conserved.

Recent simulations of Kolinski and Skolnick⁷ reproduced reasonably well the experimental three-dimensional structure of the B domain of protein A with all well-defined secondary structure elements correctly represented. The average C α root mean square deviation (rms) of low energy structures, with respect to the experimental structure, varied from 2.55 Å to 3.42 Å for each of the different runs. Typically, exhaustive refinements at low temperatures produced structures as close as 2.25 Å rms from the native structure. It was also shown that during the folding simulations, the B domain may be temporarily trapped in a broad potential energy basin that corresponds to the topological mirror image of the native fold (also referred as the inverted structure), i.e., the first helix (Gln10–His19) lies on the other side of the plane marked by the second and third helices (Fig. 1). The second and the third helices do not change their positions. Rather, they swivel slightly, which exposes their hydrophobic faces to the other side and allows for the packing of the first helix on the other side of the C terminal hairpin. The resulting mirror image topology structure (the inverted structure) was kinetically stable, but could be discriminated from the correct fold on the basis of the average energy differences of the order of 25 k_BT. This finding suggests that, with suitable mutations, the mirror image topology might be favored.

In a typical protein, non-polar side chains are tightly packed in the interior to form a solvent-inaccessible hydrophobic core.¹³ The distribution of non-polar residues (hydrophobic pattern) along the protein chain is one of the most conservative determinants of the native structure.¹⁴ However, among sequences with similar hydrophobic patterns, the possibility of folding is restricted to the subset of sequences for which core packing is sterically allowed.¹⁵ The relative importance of turns as building blocks of protein structure is somewhat controversial. Sometimes they are regarded as key elements that rule the packing of supersecondary structure elements in a native protein structure.^{16–20} The other view of turns depicts them as passive hinges that adjust themselves to accommodate overall structural changes^{21–25}, hence the sequence dependence of the turn is almost unimportant.^{23–25} Below, in the context of a reduced protein model, we examine the relative importance of the turns versus hydrophobic packing in determining the unique topology of a protein.

This paper tests the robustness of the lattice

model interaction scheme parameterization by examining its applicability to the structural assessment of proteins closely homologous to the B domain. Furthermore, we propose a sieve algorithm for scanning an extensive set of possible mutations to screen for a desired property. We applied this algorithm to search for the mutated B domain of protein A that would prefer the stable topological mirror image structure. Two possible scenarios are explored. In the first, mutations are introduced in the hydrophobic core of the B domain; in the second, the native sequence is modified in the vicinity of the turn region between the first and the second helix. Finally, we examined the causes of the relative stability changes for the native three-helix bundle topology and its mirror image structure introduced by the turn mutation.

METHODS

Lattice Model

The lattice model used for the protein backbone representation²⁶ consists of 90 basis vectors, constructed by making all possible permutations of the components of the generic vectors (3,1,1), (3,1,0), (3,0,0), (2,2,1), and (2,2,0). These vectors serve as virtual bonds connecting consecutive C α s along the protein backbone. Only consecutive pairs of vectors with bond angles within the range of 72.5°–154° degrees were allowed. When the lattice unit vector length equals 1.22 Å, the lattice represents the positions of high-resolution library proteins C α carbons, with an rms of less than 0.7 Å.²⁷ No other explicit backbone atoms are used. The amino acid side chains are approximated by a library of single ball rotamers located at the side chain center of mass, which depends on the local backbone geometry.²⁶ The number of allowed rotamers varies from 1 rotamer (e.g., for alanine) to a maximum of 58 for arginine, in certain backbone configurations. The accuracy of this side chain representation is approximately 1.0 Å. To assess the total inherent accuracy of the lattice model with rotamers, we performed a series of simulations with a target side chain pair potential for the native structure of various proteins. Of course, for the folding experiments described in this paper, no such target side chain potential is used. For the 99 residue plastocyanin, the rms from the crystal structure is 2.0 Å; similarly, for the 138 residue flavodoxin, the rms is 1.8 Å, while myohemerythrin, a 118 residue protein, has an rms from the native structure of 1.8 Å. Therefore, after incorporation of the side chain representation into the model, the overall intrinsic geometrical accuracy of the model is about 2.0 Å.

The Monte Carlo dynamics of the protein on the lattice is simulated by performing moves accepted or rejected on the basis of the asymmetric Metropolis criterion.²⁸ The same set of moves as in the previous studies has been used here. These moves include:

predefined local two- and three-virtual-bond moves as well as large distance moves designed to enhance the search of the conformational space.²⁶ The latter moves are generated by concerted sequences of overlapping three-bond motions. In addition, random changes of rotamer positions are allowed to facilitate the packing of side chains.

The potential energy used in the lattice simulations consists of five terms, viz:

$$E = E_{\text{prop}} + E_{\text{hb}} + E_{\text{one}} + E_{\text{pair}} + E_{\text{NN}} \quad (1)$$

and was generated by analysis of the library of high-resolution Protein Data Bank (PDB) structures of globular proteins.²⁶

In particular, the local, sequence-dependent term

$$E_{\text{prop}} = \sum_{i=1}^{N-3} E_{i+2, i+3}^{\text{prop}}(\tilde{r}_{i, i+1, i+2}, \tilde{r}_{i+1, i+2, i+3}) \quad (2)$$

involves two consecutive C^α_i to C^α_{i+3} and C^α_{i+1} to C^α_{i+4} (the index i numbers consecutive C^α s) chiral distances, i.e., $\tilde{r}_{i, i+1, i+2}$ and $\tilde{r}_{i+1, i+2, i+3}$, respectively. The chiral distance is defined as follows: $\tilde{r}_{i, j, k} = \text{sign}(\mathbf{b}_i \times \mathbf{b}_j \cdot \mathbf{b}_k) \cdot \|\mathbf{b}_i + \mathbf{b}_j + \mathbf{b}_k\|$, where \mathbf{b}_i is a virtual bond that connects C^α_i to C^α_{i+1} . The subscripts on E_{ij}^{prop} in Equation 2 signify the amino acid sequence dependence of the term. Since two overlapping C^α - C^α chiral distances are involved, E_{prop} propagates protein-like elements of secondary structure along the protein backbone, acting like a generalized potential based on correlated, consecutive Φ , Ψ maps. Particularly in the case of protein A, the local propagator tends to favor formation of α -helices and defines the boundaries between secondary structure elements. This contribution to the potential was developed prior to the undertaking of the study, i.e., prior to the folding of the seven proteins homologous to the B domain of protein A and prior to the undertaking of the redesign of the B domain structure. While this potential is not yet demonstrated to be fully capable of folding an arbitrary protein sequence, we believe it is worthwhile to explore its predictive power on sequences that are foldable.

In addition to this local term, an effective interaction E^{H} between C^α_i and C^α_j that simulates formation of a hydrogen bond between backbone atoms of the i th and j th residues was introduced:

$$E_{\text{hb}} = \sum_{i < j}^N E^{\text{H}} \delta_{i, j} + \sum_{i < j}^N E^{\text{HH}} \delta_{i, j} \delta_{i \pm 1, j \pm 1} \quad (3)$$

where $\delta_{i, j} = 1$, when amino acids i and j form a hydrogen bonded pair and 0, otherwise. E^{H} and E^{HH} are equal to -0.5 kT. Amino acids i and j form a hydrogen bonded pair when $|i - j| \geq 3$, and when the following geometrical criteria are satisfied:

$$R_{\text{min}} \leq \|\mathbf{r}_{i, j}\| \leq R_{\text{max}} \quad (4a)$$

$$\|(\mathbf{b}_{i-1} - \mathbf{b}_i) \cdot \mathbf{r}_{i, j}\| \leq a_{\text{max}} \quad (4b)$$

$$\|(\mathbf{b}_{j-1} - \mathbf{b}_j) \cdot \mathbf{r}_{i, j}\| \leq a_{\text{max}} \quad (4c)$$

where $\mathbf{r}_{i, j}$ is a vector connecting C^α_i to C^α_j . $R_{\text{min}} = 4.6$ Å, $R_{\text{max}} = 7.3$ Å, and $a_{\text{max}} = 13.4$ Å.²⁶ This interaction component is sequence independent as well as non-directional, since the models do not specify the location of backbone atoms other than C^α . Proton donors and acceptors are not differentiated. Every amino acid but proline can form up to two hydrogen bonds, while proline can participate in only one hydrogen bond. The hydrogen bonding scheme accounts for the cooperativity of hydrogen bond formation by introducing an effective interaction between adjacent pairs of hydrogen bonds. It favors neither helices nor β -strand states.

A one body, centrosymmetric burial potential,

$$E_{\text{one}} = \sum_{i=1}^N E_i(r_i^0/s) \quad (5)$$

reflects the radial distribution of distances r_i^0 from amino acid side chain i to the center of mass of the protein (here, s is the expected radius of gyration, calculated for a closely packed protein²⁶). Note that this component is small for compact states and dominates denatured states; thus, it serves as a driving force in the initial stages of the folding process.²⁶

Detailed packing interactions are modeled by the combination of a pairwise soft core repulsion augmented by a square well potential, derived as a potential of mean force from the frequency of close contact occurrences between amino acids:

$$E_{\text{pair}} = \sum_{i < j}^N E_{ij}(r_{i, j}) \quad (6)$$

where i and j number interacting amino acids i and j and E_{ij} equals:

$$E_{ij} = \begin{cases} E_{\text{rep}}, & \text{for } r_{i, j} < R_{i, j}^{\text{rep}} \\ \varepsilon_{i, j}, & \text{for } R_{i, j}^{\text{rep}} \leq r_{i, j} < R_{i, j}, \text{ and } \varepsilon_{i, j} \geq 0 \\ f\varepsilon_{i, j}, & \text{for } R_{i, j}^{\text{rep}} \leq r_{i, j} < R_{i, j}, \text{ and } \varepsilon_{i, j} < 0 \\ 0, & \text{for } r_{i, j} \geq R_{i, j}. \end{cases} \quad (7)$$

The radius of repulsion $R_{i, j}^{\text{rep}}$, depth $\varepsilon_{i, j}$, and limits $R_{i, j}$ of the square well are publicly available by anonymous ftp.²⁹ Attractive interactions are modified by a factor f

$$f = 1 - (\cos^2(\angle(\mathbf{u}_i, \mathbf{u}_j)) - \cos^2(20^\circ))^2 \quad (8)$$

which is dependent on the angle between the vector $\mathbf{u}_i = \mathbf{r}_{i+2} - \mathbf{r}_{i-2}$ and the corresponding vector \mathbf{u}_j to induce proper supersecondary structure packing. The angle $\angle(\mathbf{u}_i, \mathbf{u}_j)$ represents the relative orientation of the secondary structure in the vicinity of the i th residue with respect to the supersecondary structure element surrounding the j th residue, and the 20° angle reflects the average packing angle of helices in helical bundles and strands in β -barrels.

Finally, there is a supplemental term E_{NN} that was designed to reproduce the occurrence of protein-like side chain contact maps of globular proteins. An artificial neural network with error back-propagation has been trained to recognize frequently occurring 7×7 fragments of side chain contact maps.³⁰ For each pair ij , if the 7×7 fragment of the side chain contact map centered at ij is recognized by the neural network as protein-like, then the pair interaction well depth ϵ_{ij} is modified in the following way

$$\epsilon_{ij}^{\text{new}} = 0.9 \epsilon_{ij} + 0.1 \bar{\epsilon}_{ij} \quad (9)$$

where

$$\bar{\epsilon}_{ij} = \frac{\sum_{7 \times 7} \epsilon_{kl} c_{kl}}{\sum_{7 \times 7} c_{kl}} \quad (10)$$

and the summation in Equation (10) is performed over the 7×7 fragment of an appropriate contact map. c_{kl} equals 1 if side chains k and l are in contact and 0, otherwise. Therefore, E_{NN} simulates the average effective many-body component of the potential energy responsible for the mutual packing of super secondary structure elements.

Previously, the same lattice model with a similar interaction scheme was successfully applied by Kolinski, Skolnick, and coworkers to the simulation of the folding process of small helical proteins,⁷ coiled coils,³¹ and crambin.⁷ The term E_{prop} used here is a generalization of the term $E_{\text{C}\alpha\text{-trace}}$ of Ref. 26 that has all the same features as the term $E_{\text{C}\alpha\text{-trace}}$, but the secondary structure interfaces are now better defined. The rotamer energy²⁶ has not been used and the many-body component E_{tem} of ref. 26 has been replaced by E_{NN} , the neural network packing regularizing term.³⁰ This extension of the original template term has been designed to include those residues that are not involved in the original template forming patterns (about 60% of all pairs). The changes in the potential described here do not qualitatively alter the results of those simulations. For the folding of protein A and for ROP, correct topologies and similar rms deviations were observed (Skolnick and Kolinski, unpublished results). Recently, the same version of the potential has also been used to study structural stability of the retro-sequence of the B domain of protein A³²; however, this potential version has not yet been tested on non-helical proteins.

Sieve Procedure for Mutation Screening

The following sieve procedure is proposed to screen for sequences that adopt the mirror image topology motif of the B domain of protein A (Fig. 2). First, the set of points in the sequence to be mutated must be chosen. Next, the mutation rules (i.e., what amino acids may be substituted at the chosen points)

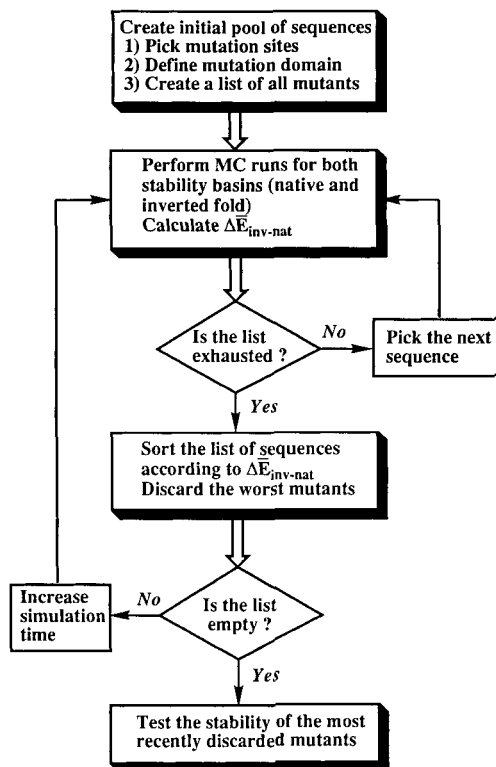


Fig. 2. The sieve method flowchart.

must be specified. Subsequently, for all possible mutations, two relatively short simulations were performed (ranging from 1,000 to 2,000 Monte Carlo cycles) that started from the native and the topological mirror image basins. The majority (around 80–90%) of the worst mutants (i.e., mutants for which the average energy difference indicates a preference for the native topology) are then rejected from further consideration. Next, we repeat the simulations for an order of magnitude longer time, and again, the worst mutants are rejected. Finally, long simulations (2 million Monte Carlo cycles) are performed. For all stages of the simulations, all mutants that cause protein unfolding are rejected. Certainly, such an algorithm may not lead to the most optimal solution, since the best mutant may be rejected even after the first set of simulations, since the sampling of the conformational space is sparse. On the other hand, the procedure for the exhaustive search of all possible mutations is computationally impractical. The prescreening runs are about a thousand times shorter than those of the final stage. Therefore, for the whole set of mutations, full screening would require about 10^3 longer cpu time. Moreover, in practice, we find that results obtained throughout the procedure present a consistent trend, since all accepted mutants must exhibit the desired property at all stages of the screening to be accepted.

RESULTS AND DISCUSSION

Stability and Folding of Protein A Domains

To establish the robustness of the structural prediction, we performed two pairs of consecutive long isothermal Monte Carlo simulations for the other four domains of protein A that are highly homologous to the B domain. In each pair of simulations, the starting point was either the native three-helix bundle structure or the topological mirror image structure previously identified⁷ as a possible alternative folding topology for the native sequence of the B domain. The simulations were conducted for 2 million Monte Carlo cycles at relatively low temperature ($T = 1$) to explore both potential energy basins related to the native and inverted three-helix bundle topologies. Throughout the simulations, we did not notice rearrangement of the protein fold; all structures accepted according to the Metropolis criterion were within 3.5 \AA C $^{\alpha}$ rms of the initial structure. The average energies \bar{E}_{nat} and \bar{E}_{inv} for those simulations are shown in Table II, where \bar{E}_{nat} (\bar{E}_{inv}) indicates the average energy in the native (the inverted) topology basin. For all naturally occurring domains of protein A, the native fold is preferred over its mirror image. In the cases of domains A, D, and E, the difference of the average energy between basins exceeds 10 kT, which may suggest a stable trend. The same tendency can also be observed for $E_{\text{nat}}^{\text{min}}$ and $E_{\text{inv}}^{\text{min}}$, the lowest energies for each sequence found during the course of the simulations for the native or mirror image basin, respectively (Table II). For the C domain, the average energy difference $\Delta\bar{E}_{\text{inv-nat}}$ is small; however, the difference between minimal energies visited throughout the simulations, $\Delta E_{\text{inv-nat}}^{\text{min}}$, also favors the native topology. Therefore, we assume that the tendency to preserve the native fold, although weaker, is still present. This difference probably occurs because the C domain is visibly different from its partners. It contains mutations in the turn regions between helices I and II (Asn24 to Thr24) and helices II and III (Gln41 to Val41), whereas, these regions are conserved in other domains. Also, the beginning of the third helix in C domain differs substantially from the other protein A domains (Table I).

Analogous stability simulations were performed for protein Z (an engineered version of the B domain of protein A) and two of its single point mutant sequences. Both mutations, as well as the mutation that leads to protein Z from the B domain of protein A, are situated close to the middle of the second helix, namely Asn29 to Ala29 (first mutant) and Phe31 to Ala31 (second mutant). The latter mutation has been reported as being the least stable in solution.¹¹ The average energies from our simulations tend to confirm this effect, provided that the free energies of the unfolded states for all closely homologous sequences are essentially the same (Table III). The

TABLE II. Relative Stability of Staphylococcal Protein A Domains*

	Domain				
	A	B	C	D	E
\bar{E}_{inv}	-166.1	-169.2	-166.1	-162.3	-166.9
\bar{E}_{nat}	-176.4	-194.5	-168.1	-174.4	-179.7
$\Delta\bar{E}_{\text{inv-nat}}$	10.3	25.3	2.0	12.1	12.8
$E_{\text{inv}}^{\text{min}}$	-206.6	-210.8	-200.0	-202.9	-215.1
$E_{\text{nat}}^{\text{min}}$	-213.5	-224.2	-205.6	-210.3	-217.1
$\Delta E_{\text{inv-nat}}^{\text{min}}$	5.9	13.4	5.6	7.4	2.0

* \bar{E}_{nat} and \bar{E}_{inv} are average energies calculated for native and mirror image structures, respectively. $E_{\text{nat}}^{\text{min}}$ and $E_{\text{inv}}^{\text{min}}$ are the lowest energies encountered during the simulations. All energies are given in kT units.

tendency to conserve the native fold of the B domain is also evident from Table III. The inversion of minimal energy difference found for protein Z (Table III) is probably due to the existence of a narrow local minimum that is virtually unpopulated, since the average energy difference favors the native topology structure.

Subsequently, we tested the ability of the homologous sequences to acquire the three-helix bundle topology from the random coil state. For each sequence of the A, C, D, and E domains of protein A, and for protein Z, five folding experiments starting from conformations that are far from the three-helix bundle topology were performed. The starting conformations were picked by a random choice of a PDB structure and by cutting a 44 residue fragment out of this structure. Then Monte Carlo annealing simulations were performed for 2 million Monte Carlo cycles. For all the above-mentioned homologous proteins, the three-helix bundle topology structure (native or inverted) was acquired for all folding runs. The rms of the final structures from the native or inverted structure of the B domain for all homologous protein tested lies within 3.5 \AA . Therefore, the stability simulation results are confirmed by the demonstration that the model system may fold into the lower energy topology basin.

Redesign of the Hydrophobic Core of the B Domain

The sieve procedure has been applied to search for the topological mirror image preferences among mutations of the hydrophobic core of the B domain. Due to limitations in computer resources, only three groups of six residue locations were selected. Only at selected sites were mutations performed. To single out the repacking effect of the hydrophobic core, the search was restricted to aliphatic and aromatic amino acid side chains only, i.e., the possible mutation set consisted of five amino acids: Ala, Val, Ile,

TABLE III. Relative Stability of Staphylococcal Protein A Mutants*

	Mutant		
	Z	Z(N29A)	Z(F31A)
\bar{E}_{inv}	-179.6	-175.7	-170.4
\bar{E}_{nat}	-190.7	-184.6	-174.6
$\Delta\bar{E}_{inv-nat}$	11.1	8.9	4.2
E_{inv}^{min}	-226.8	-205.3	-207.5
E_{nat}^{min}	-218.0	-220.4	-213.9
$\Delta E_{inv-nat}^{min}$	-8.8	15.1	6.4

* \bar{E}_{nat} and \bar{E}_{inv} are average energies calculated for native and mirror image structure, respectively. E_{nat}^{min} and E_{inv}^{min} are the lowest energies encountered during the simulations. All energies are given in kT units.

Leu, and Phe. Within each group of sites, all the mutation sites were allowed to be independently mutated by substituting for the native residue one from the above listed hydrophobic amino acids. Therefore, for a group of six sites and five possible substitutions with hydrophobic amino acids, the total number of possible mutations is $5^6 = 15,625$. The distribution of mutation sites within groups is as follows: In the first group of sites, two amino acid positions per each helix, corresponding to native Ala13, Phe14, Phe31, Ile32, Leu45 and Leu46, were mutated (Fig. 3). Within this group, all amino acids except Leu45 are conserved in all domains of protein A, while substitution of Leu45 by Val or Ile is seen in some protein A domains. Native amino acids sites comprised of Ile17, Leu18, Leu20, Phe31, Ala49, and Leu52 constitute the second group of mutation sites (Fig. 4) independently mutated; here, Ile17 and Leu20 are not totally conserved within the family of protein A domains. The third group of sites consists of residues originally involving Phe31, Ile32, Leu35, Leu45, Leu46, and Ala48 located on helices II and III (Fig. 5). Similar to the first group, the position involving Leu45 is the only position that is not conserved. Therefore, our choice reflects the (almost) invariant points in the native sequence of the protein A domains.

For each group of sites, in the first set of simulations, short runs (1,000 Monte Carlo cycles/run) were performed. In the second step, the Monte Carlo simulations were ten times longer. After completion of both initial steps, approximately 100 (from almost 16,000 possible) mutants underwent stability tests. It turned out that the native fold was preferred over its mirror image counterpart for all three sets of mutation sites tested. Therefore, in the case of the B domain of protein A, our results are consistent with the idea that the intrinsic details of hydrophobic core packing do not constitute the driving force for folding process but probably stabilize an already acquired motif.³³

Redesign of the Turn Region of the B Domain

To examine whether we could identify mutations in the turn region that favor the mirror image topology over the native one, a two-point mutation group in the turn connecting helix I and helix II was chosen. We have chosen this turn, since helices II and III seem to be relatively well packed.⁹ In addition to possible changes in the handedness of the turn induced by the mutation, we expected that changes in the turn region would destabilize the second helix. This may also affect the relative stabilities of the native and mirror image topology structures. All possible mutations (except for glycine, proline, and cysteine) of Asn22 and Asn24 were allowed. Glycine, proline, and cysteine were excluded, since their pair interaction and local propagator parameterization may be influenced by the specific roles they play in the formation of the structure of proteins. Asn22 preceded by proline is conserved in all protein A domains. Asparagine in position 24 constitutes the N-cap of the second helix and is relatively well conserved, since only in the C domain (which is itself slightly different from the other domains) is Asn24 substituted by Thr. We have chosen the Asn24 helix cap residue, since we want to destabilize the second helix to aid its swiveling motion, which may facilitate the topology change. The Asn22 choice has been made by elimination, since we wanted a mutation site within the turn; at the same time, we did not want to mutate hydrophobic residues such as Leu20 and Leu23, which also occur in the turn. Also, Pro21 is very typical as the second residue in the turn. Our choice of the turn mutation sites reflects the interplay between turn and helix that may be responsible for the three-helix bundle topology stabilization.²⁰

Applying the sieve procedure, this time we performed long Monte Carlo simulations (2 million Monte Carlo cycles) starting from both native and mirror image structures for all possible two-point mutations including asparagine (i.e., $17 \times 17 = 289$ mutants) at positions 22 and 24. All mutations that caused unfolding (i.e., did not stay within the 3.5 Å rms limit from the average native or mirror image structure) were rejected. As in the case of hydrophobic core mutation experiments, mutations in the turn region appeared to be extremely conservative, which is in accordance with experimental results for turn mutations in a four-helix bundle.²³⁻²⁵ A histogram of $\Delta\bar{E}_{inv-nat}$, the average energy difference (Fig. 6) for the 164 mutants that exhibit stable structures, shows that most mutations preserve the native fold. The same tendency may be observed in the histogram of the differences of lowest energies found throughout the simulations for a given mutant and for a given topology (Fig. 7). However, there are some mutations for which the average energy of the topological mirror image is lower than the respective native structure basin. Those mu-

FOLDING AND REDESIGN OF THREE-HELIX BUNDLE MOTIF

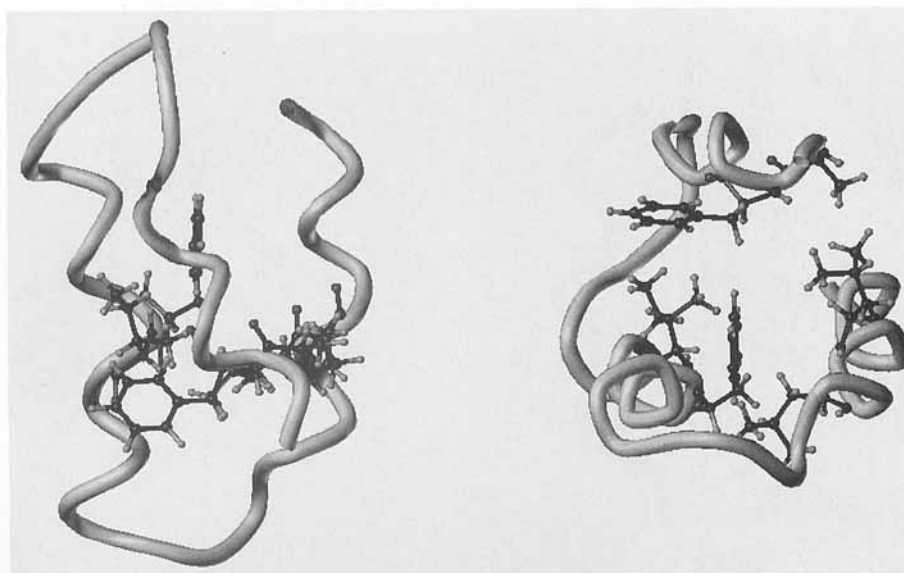


Fig. 3. First group of hydrophobic core mutations. Only native amino acids that underwent mutation (i.e., Ala13, Phe14, Phe31, Ile32, Leu45, and Leu46) are depicted in the two projections of the native three-helix bundle structure of the B domain.

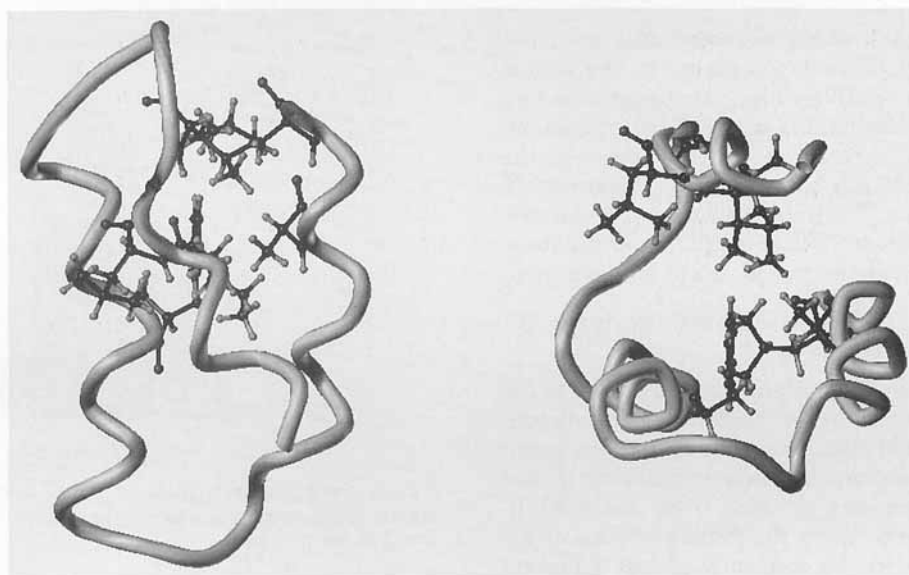


Fig. 4. Second group of hydrophobic core mutations. Only native amino acids that underwent mutation (i.e., Ile17, Leu18, Leu20, Phe31, Ala49, and Leu52) are depicted in the two projections of the native three-helix bundle structure of the B domain.

tants that constitute the far left portion of the $\Delta E_{\text{inv-nat}}$ histogram (Fig. 6) are presented in Table IV. The most promising mutants, however, do not have their counterparts in Table V (which presents the far left part of $\Delta E_{\text{inv-nat}}^{\text{min}}$ histogram) (Fig. 7). For the first mutant that exists in both tables, the EI (Asn22Glu and Asn24Ile) mutant, subsequent stability tests proved that the system was not completely equilibrated. Ultimately based on the results of eight consecutive runs, the average energy differ-

ence suggests a preference for the native topology for this mutant. The next possible mutant, RM (Asn22Arg and Asn24Met), was thoroughly tested, and, in this case, the average energy gap calculated for eight runs (2 million Monte Carlo cycles each) rose to 10.5 kT in favor of the mirror image structure. On the other hand, the difference between lowest minima found for both topologies after eight runs equals 3.9. However, the lowest local minimum encountered in the native topology basin was very

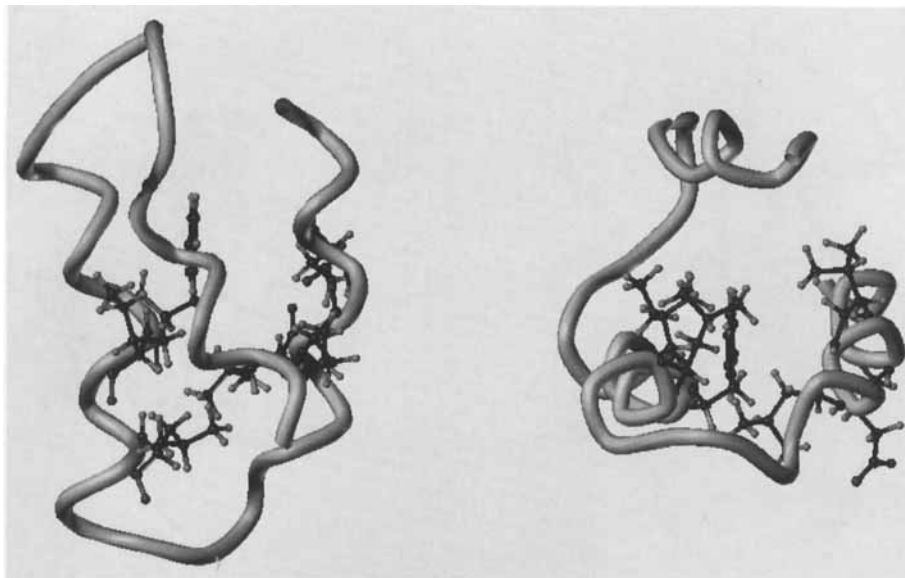


Fig. 5. Third group of hydrophobic core mutations. Only native amino acids that underwent mutation (i.e., Phe31, Ile32, Leu35, Leu45, Leu46, and Ala48) are depicted in the two projections of the native three-helix bundle structure of the B domain.

sparsely populated, which suggests that it is narrow. The probability of finding arginine in the second turn position is relatively high, compared with Lys, Glu, or Ala.³⁴ Methionine in the N-cap region, on the other hand, is rarely seen.³⁴ Nevertheless, the sequence LPRLM can be found in ATP synthase,³⁵ DNA polymerase,³⁶ nitrate reductase 1,³⁷ and protein sequences from SWISSPROT.³⁸ Unfortunately, the structures those sequences adopt are unknown.

Folding of the RM Mutant and Robustness of the Predicted Inversion

To confirm that RM mutants do, in fact, fold to the mirror image structure, we performed ten independent Monte Carlo simulations for this mutant with the starting structures chosen from randomly picked fragments of various proteins from the PDB library.³⁹ Similarly, as in the folding studies of homologous domains, we randomly picked a protein structure and then trimmed it to a 44 residue long fragment, which is equivalent to the random selection of the starting structure. Then, the Monte Carlo annealing procedure was used, i.e., we gradually lowered the temperature, thereby allowing protein to settle in the most probable potential energy basin. In six of ten simulations, the mirror image topology has been achieved as a stable structure with a mean rms of 1.21 Å from the average mirror image topology structure with a standard deviation 0.55 Å. The average mirror image topology structure is obtained as follows. For each simulation that yields the mirror image topology, once the topology is adopted the average positions of all C α s are calculated. The topology is considered adopted when the structures stays within 3.5 Å from either one of the generic

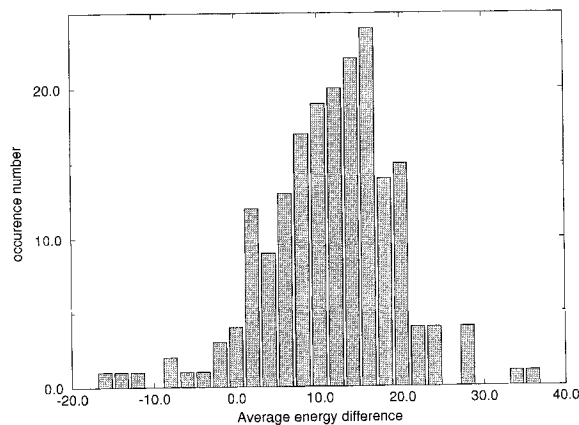


Fig. 6. The distribution of $\Delta E_{\text{inv-nat}} = E_{\text{inv}} - E_{\text{nat}}$ for 164 two-point mutations located in a turn region between helices I and II. Energies are in kT units.

three-helix bundle structures, i.e., the mirror image or the native structure, respectively. The average structure is obtained by averaging the positions of the C α s of these structures. The mean rms is the rms deviation from this averaged structure computed for all structures obtained from the Monte Carlo trajectory. In four runs, the B domain native structure was obtained. We did not notice, however, any stable structures alternative to the three-helix bundle topology.

The robustness of the RM mutation has been further tested by applying it to the A domain of protein A. In this case, we performed three consecutive isothermal simulation runs (2 million Monte Carlo cy-

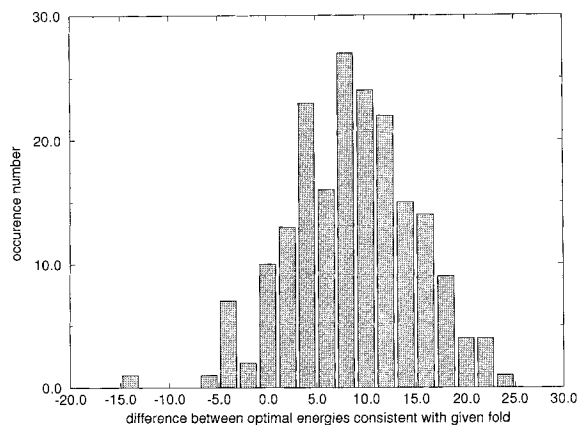


Fig. 7. The distribution of $\Delta E_{inv-nat}^{min} = E_{inv}^{min} - E_{nat}^{min}$ for 164 two-point mutations located in a turn region between helices I and II. Energies are in kT units.

TABLE IV. Average Energies of Mirror Image (Inverted) Structure Candidates for Mutations Located in a Turn Region, From First Set of Screening Runs*

Mutant	Energy		
	E_{inv}	E_{nat}	$\Delta E_{inv-nat}$
KY	-205.1	-189.8	-15.2
EI	-197.1	-184.6	-12.4
AV	-195.5	-184.6	-10.8
RR	-194.8	-187.7	-7.0
RM	-202.4	-196.0	-6.3
AI	-189.0	-184.0	-4.9
KA	-194.8	-191.7	-3.1
SE	-194.4	-192.8	-1.6
YK	-182.5	-181.2	-1.3
TW	-189.1	-188.7	-0.3

* E_{nat} and E_{inv} are average energies calculated for the native and mirror image structures, respectively. All energies are given in kT units.

cles/run) for each topology. The difference in the energy averaged over those runs $\Delta \bar{E}_{inv-nat} = 5.9$ suggests that the tendency of the RM mutation to enforce the mirror image packing is preserved. This tendency is also confirmed by the substantial difference in minimal energy found for both topologies, since $\Delta E_{inv-nat}^{min} = -22.7$. This may suggest that the RM mutant for the A domain is also a good candidate for topology inversion.

Origins of the Topology Inversion

Exploring the origins of topology inversion, we examined the long-lived (those present over 75% of the simulation time in stability simulations, i.e., entire trajectory remains in the native or mirror image topology) contacts for the native B domain sequence (Fig. 8) and for the RM mutant (Fig. 9). The native topology contacts are presented below the main di-

TABLE V. Minimal Energies of Inverted Structure Candidates for Mutations Located in a Turn Region, From First Set of Screening Runs*

Mutant	Energy		
	E_{inv}^{min}	E_{nat}^{min}	$\Delta E_{inv-nat}^{min}$
NL	-232.9	-219.7	-13.2
RD	-224.2	-220.0	-4.2
YF	-215.5	-211.8	-3.7
SF	-219.0	-215.4	-3.6
RM	-228.2	-225.0	-3.2
KH	-218.1	-215.4	-2.7
FA	-225.7	-223.2	-2.5
EI	-220.6	-218.2	-2.4
QL	-214.6	-212.5	-2.1
KA	-220.5	-220.0	-0.5
DI	-220.6	-220.4	-0.2
RR	-219.4	-219.4	0.0

E_{nat}^{min} and E_{inv}^{min} are the lowest energies encountered during the simulations for the native and mirror image structures, respectively. All energies are given in kT units.

agonal and the mirror image topology contacts above the diagonal. In the case of B domain, it is possible to judge at first glance which topology is preferred. In the native topology basin, there are 15 long-lived contacts. Thus, its side chain packing is well defined; it would be predicted to have native-like side chain packing. In contrast, in the mirror image topology there are only four long-lived contacts. In particular, there are no such contacts between helices I and III, which indicates a lack of well-defined packing interactions between those helices. Hence, the B domain, in its mirror image topology, resembles a molten globule rather than a well-defined structure. The main difference between the native topology of the B domain and the mirror image topology of the RM mutant also involves those contacts between helices I and III. For the B domain, the native contact set consists of three such contacts, namely, Phe14–Leu45, Phe14–Leu46, and Ile18–Ala49. For the RM mutant, in the mirror image topology (i.e., its putative native topology), the Phe14–Leu45 contact present in the native B domain disappears. Additional contacts involve Leu18–Ala49 and Leu18–Leu52. Leu46 is the side chain that experiences the most substantial change in contacts. In the B domain in the native topology, it has contact with Phe14. In the RM mutant in the mirror image topology Leu46 has five contacts with Phe14, Ile17, Phe31, Ile32, and Leu35. Concomitantly, Leu45 in the RM mutant mirror image topology loses all of its contacts (i.e., Phe14, Phe31, Ser34, and Leu35) compared with the native B domain structure. The repacking of hydrophobic core takes place because helix III swivels slightly to adjust the packing of the side chains in the mirror image topology. To a lesser extent, this effect can be seen for Leu18, which, in the B domain, has three

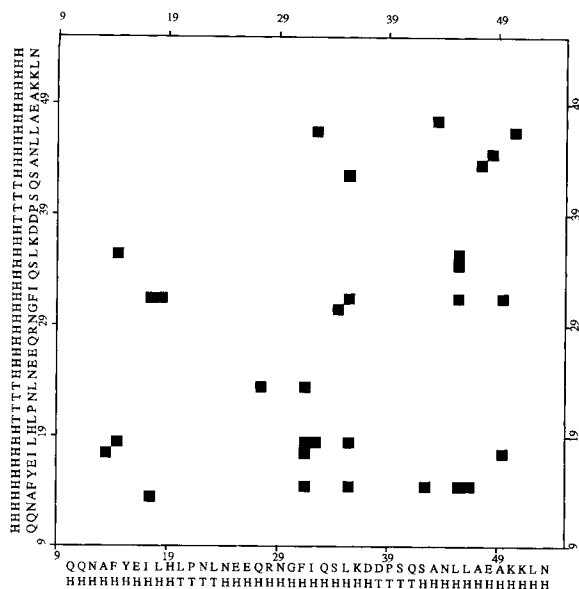


Fig. 8. Contacts with lifetimes over 75% of the total simulation time for the native sequence of the B domain of protein A are shown for the native topology (below diagonal) and for the mirror image topology (above diagonal).

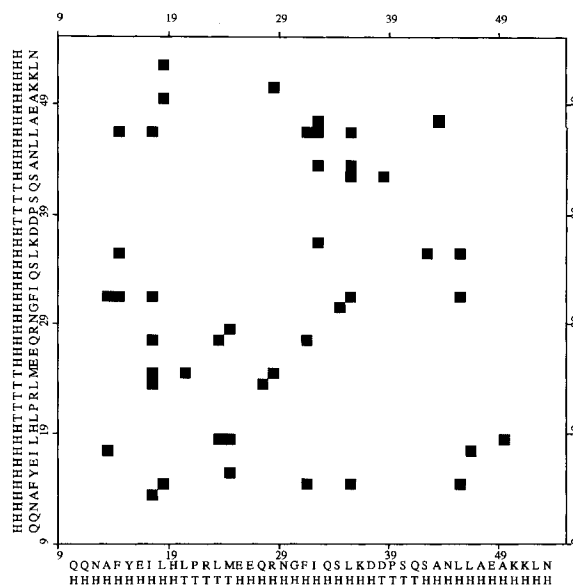


Fig. 9. Contacts with lifetimes over 75% of the total simulation time for the RM mutant sequence are shown for the native topology of the B domain (below diagonal) and for the mirror image topology (above diagonal).

contacts with helix II. For the RM mutant, all those contacts are lost in favor of contacts with Ala49 and Leu52 in helix III. Thus, we conjecture that the mirror image topology should have a well-defined pattern of side chain interactions, suggesting that it is native-like and not a molten globule-like state.

Next, we examined each contribution to the total energy as a function of the location of each amino acid in the sequence. Those components of the total energy related to E_{prop} and E_{pair} stabilize the mirror image structure. As expected, the RM mutation substantially changes the local secondary structure propensities in the turn region between helices I and II, but the second turn region is also affected (Fig. 10). The average \bar{E}_{prop} per amino acid is even lower for the RM mutant than for the native B domain, mainly because of the peak in the E_{prop} in the native topology located on the verge of the first turn for amino acids Leu23 and Arg24. It probably is a carryover effect of Leu20 and also Leu23, since the E_{prop} term spans a window of five residues and, in general, hydrophobic amino acids such as Leu are hardly seen in turn regions. For the RM mutant, on the other hand, the second helix is slightly longer and the turn sharper; therefore, the corresponding peak is narrower.

The pair interaction energy decomposition for the RM mutant in the mirror topology reveals an increased attractive interaction for amino acid residues not directly connected to the mutation region, namely Ile17, Phe31, Ala43, and Leu46 (Fig. 11) due to the repacking of the hydrophobic interactions that stabilize the mirror image structure. Addition-

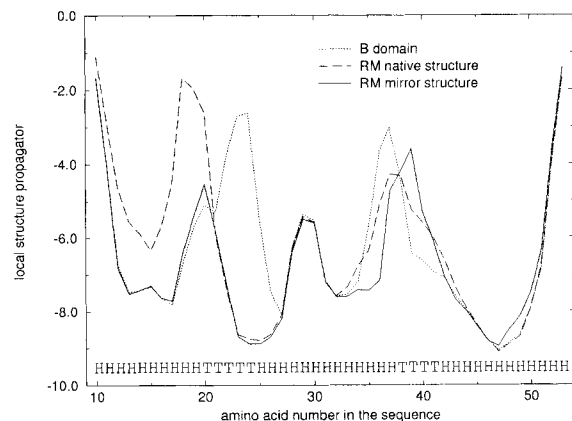


Fig. 10. Decomposition of the average local secondary structure propagator contribution E_{prop} for the native sequence and for the RM mutant. Energies are in kT units. H and T symbolize location of helices and turns, respectively, in the native structure of the B domain.

ally, there is also a substantial energy decrease for Arg28, which incidentally has a long-lived contact with Lys50. On the other hand, for the Lys50, there exists a large increase in the pair interaction component; therefore, the repulsive Arg28–Lys50 contact is stabilized by favorable pair interactions of Arg28 with Glu25 and Glu26. The pair interaction energy component related to the Asn22Arg mutation in the RM mutant does not vary significantly, and the single Asn22Arg mutation did not produce a stable inverted structure. On the other hand, the mutation of Asn24 to Met24 decreases the pair en-

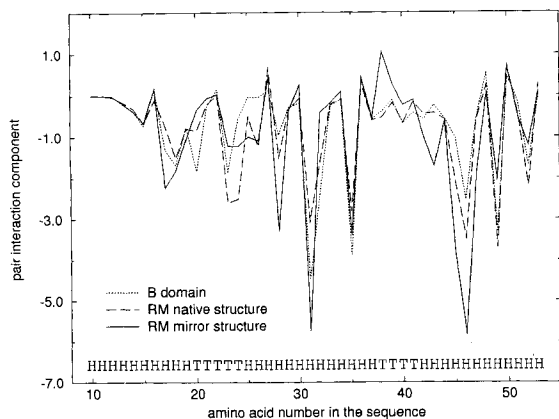


Fig. 11. Decomposition of the average pair energy contribution E_{pair} for the native sequence and for the RM mutant. Energies are in kT units. H and T symbolize the location of helices and turns, respectively, in the native structure of the B domain.

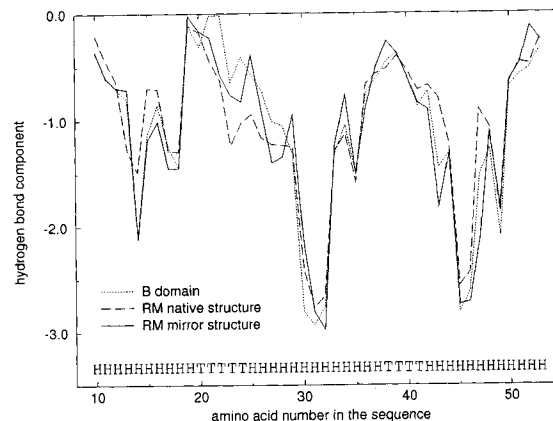


Fig. 12. Decomposition of the average hydrogen bond energy contribution E_{hb} for the native sequence and for the RM mutant. Energies are in kT units. H and T symbolize the location of helices and turns, respectively, in the native structure of the B domain.

ergy term for both mutant topologies, and therefore it cannot cause topology inversion by itself. However, in case of mirror image topology, the decrease is approximately 1.5 kT larger for Met24. The hydrogen bond energy (Fig. 12) is not very specific, and the packing regularizing neural network term indicates a slight preference toward the native structure over its mirror image (Fig. 13). Thus, based on Figure 10, we conclude that in the context of the lattice model, the RM mutation introduces the necessary preference for the change in the handedness of the turn, which induces a change in the fold topology that is stabilized further by appropriate packing. Therefore, our results are consistent with other work,²⁰ which suggests that the turn-helix interactions are responsible for the stabilization of the four-helix bundle topology. On the other hand, the difficulty in finding a proper mutation supports other results,^{23–25} which demonstrate, how promiscuous the four-helix bundle topology is with respect to mutations in the turn region.

All Atom Models of the RM Mutant

While lattice models have been demonstrated to be capable of reproducing the geometry of protein structures with high accuracy (about 0.7 Å rms for backbone), in principle it is possible that a given lattice structure cannot occur in reality. For example, this may be due to some unfavorable steric interactions that are absent in the reduced lattice model. This may be caused by the representation of the side chains in the reduced model as single balls and lack of backbone atoms other than C^α . This lack of full atomic detail may lead, upon all atom model construction of a given topology, to a structure that has a number of steric clashes that preclude the existence of that particular structure. Therefore, to establish if the lattice models are consistent with

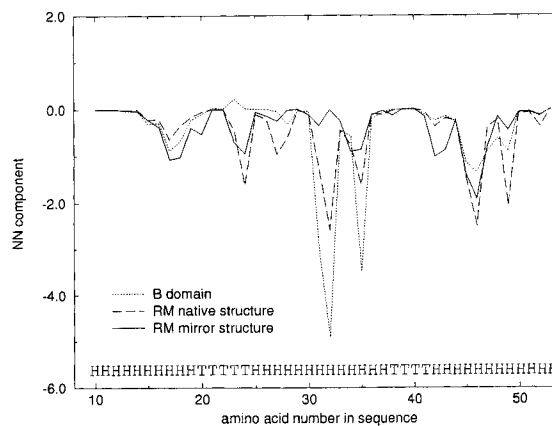


Fig. 13. Decomposition of the average side chain packing, neural network contribution E_{NN} for the native sequence and for the RM mutant. Energies are in kT units. H and T symbolize the location of helices and turns, respectively, in the native structure of the B domain.

atomic resolution models, for the RM mutant we constructed all atom models of the native and mirror image topology using SYBYL. Side chains built by SYBYL were in trans conformations and were initially relaxed in vacuo using the Kollman all-atom potential.⁴⁰ Both topologies were then relaxed in a water cube with periodic boundary conditions using the AMBER package⁴⁰ in TIP3P water⁴¹ at a constant temperature $T = 300$ K. We performed 50 ps molecular dynamics simulations intertwined with a short local minimization every 1 ps. The protein has been completely immersed in a sea of around 1,500 water molecules. It appears (Fig. 14) that the adoption of the inverted structure may be stabilized by the burial of the hydrophobic Met24 side chain (as opposed to the exposed Asn24 in both native and mirror image structures of the B domain native se-

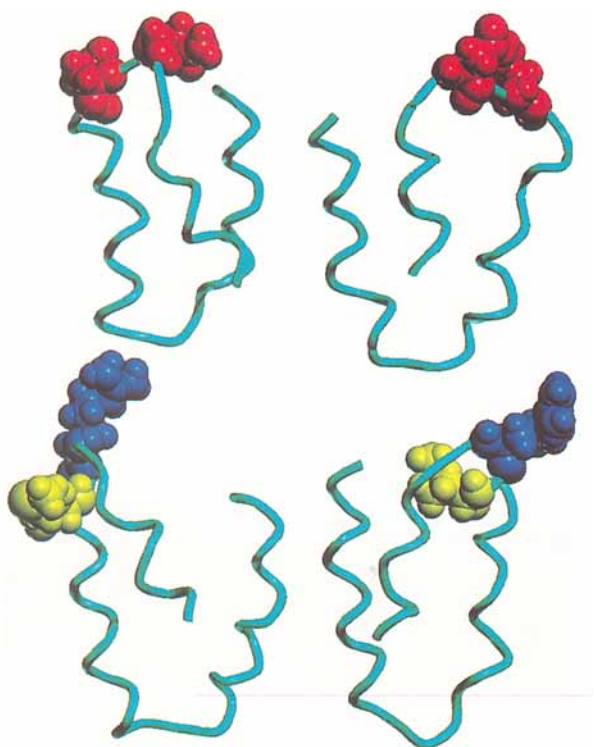


Fig. 14. Conformations of native Asn22 (red) and Asn24 (red) amino acids from the turn mutation site of the B domain of protein A (top row) are shown for the native structure from NMR (top left) and for the topological mirror image structure built from the lattice model (top right). Conformations of Arg22 (blue) and Met24 (yellow) from the mutation site of the RM mutant (bottom row) are shown for all atom models built from the lattice model. The native state of the B domain (bottom left) and its topological mirror image, the putative native state for the RM mutant (bottom right), are shown as a cyan ribbon tube representation depicted. The cyan ribbon tube represents the backbone trace of the all atom models.

quence; Fig. 14). Moreover, Met24 burial probably aids the swiveling motion of the second helix, which is consistent with pair interaction energy decreases for Met24 in the lattice model simulations. The results are in accordance with the lattice simulations, which report three contacts for Met24 in the RM mutant (namely, Ile17, Leu20, and Arg28), whereas Asn24 in the native sequence has no long-lived contacts. On the other hand, the native topology structure of the RM mutant has the methionine exposed to the solvent (Fig. 14). Simultaneous rearrangement of the second helix at its N-cap leads to a longer and slightly bent helix II in the mutated domain, an effect that is also present in the lattice simulations. Thus, the all-atom and lattice structures of the RM mutant are in complete qualitative agreement.

CONCLUSIONS

A previously developed lattice model of proteins with an improved potential energy function has been employed to fold a set of proteins homologous to the B domain of protein A. We have found that the

putative native fold of all three-helix bundle domains of protein A and protein Z and two of its mutants is predicted from the Monte Carlo folding simulations. In all cases, the differences in the average energy for simulations near the native structure and in the mirror image basin always indicate that the native topology is preferred. Thus, the potential energy used in these simulations responds in the expected manner for homologous changes in protein sequence, thereby providing some confidence that the potential recovers some aspects of protein behavior.

Application of the lattice model to the redesign of the three-helix bundle fold showed that changing the packing of the hydrophobic core is not enough to induce a global conformational change in the modeled structures. Additional analysis of the individual contributions to the total energy also confirms that the hydrophobic core packing interactions are not specific enough to differentiate between various structural alternatives. Mutations of the turn region were also found in the lattice model context to be very conservative, as experiments on four-helix bundles would suggest.^{23–25} However, in this case, as a result of intensive screening, it was possible to identify a mutant that adopts the topological mirror image fold. The acquisition of the mirror image topology is a concerted effect due to the change in the mutated turn secondary structure preferences, the burial of the N-cap of the second helix, and repacking of the hydrophobic core. Subsequently built all-atom models reinforce the conclusions of the lattice model. The resulting, mutated sequence should yield a native state with well-defined side chain packing as opposed to a molten globule state. Whether the conclusion about the possible importance of turns in defining the global topology holds in general or is just specific to the three-helix bundles analyzed here requires additional investigation. Nevertheless, we have found that the RM mutant has a well-defined, long-lived native state-like pattern of side chain contacts in the B domain mirror image topology. This constitutes a fairly strong prediction of the present lattice model of proteins. Future experimental work will be required to test whether or not this prediction is valid.

ACKNOWLEDGMENTS

We gratefully acknowledge NIH grant GM-37408 and the Joseph Drown Foundation for their partial support of this research. We thank Dr. Adam Godzik for useful comments.

REFERENCES

1. Anfinsen, C.B. Principles that govern the folding of protein chains. *Science* 181:223–230, 1973.
2. Hecht, M.H., Richardson, J.S., Richardson, D.C., Ogden, R.C. De novo design, expression and characterization of Felix: A four-helix bundle protein of native-like sequence. *Science* 249:884–891, 1990.

3. Fedorov, A.N., Dolgikh, D.A., Chemeris, V.V., Chernov, B.K., Finkelstein, A.V., Schulga, A.A., Alakhov, Y., Kirpichnikov, M.P., Ptitsyn, O.B. De novo design, synthesis and study of albebetin, a polypeptide with a predetermined three dimensional structure. *J. Mol. Biol.* 225:927-931, 1992.
4. Betz, S.F., Raleigh, D.P., DeGrado, W.F. De novo protein design: From molten globules to native-like states. *Curr. Opin. Struct. Biol.* 3:601-610, 1993.
5. Quinn, T.P., Tweedy, N.B., Williams, R.W., Richardson, J.S., Richardson, D.C. Betadoublet: De novo design, synthesis, and characterization of a β -sandwich protein. *Proc. Natl. Acad. Sci. USA* 91:8747-8751, 1994.
6. Godzik, A. In search of the ideal protein sequence. *Protein Eng.* 8:409-416, 1995.
7. Kolinski, A., Skolnick, J. Monte Carlo simulations of protein folding. II. Application to protein A, ROP, and crambin. *Proteins* 18:353-366, 1994.
8. Deisenhofer, J. Crystallographic refinement and atomic models of a human FC fragment and its complex with fragment B of protein A from *Staphylococcus aureus* at 2.9 Å resolution. *Biochemistry* 20:2361, 1981.
9. Gouda, H., Torigoe, H., Saito, A., Sato, M., Arata, Y., Schimada, I. Three-dimensional solution structure of the B-domain of staphylococcal protein A: Comparisons of the solution and crystal structures. *Biochemistry* 40:9665-9672, 1992.
10. Nilsson, B., Moks, T., Jansson, B., Abrahamsen, L., Elmblad, A., Holmgren, E., Henrichson, C., Jones, T.A., Uhlen, M. A synthetic IgG-binding domain based on staphylococcal protein A. *Protein Eng.* 1:107-113, 1987.
11. Cedergren, L., Andersson, R., Jansson, B., Uhlen, M., Nilsson, B. Mutational analysis of the interactions between staphylococcal protein A and human IgG1. *Protein Eng.* 6:441-448, 1993.
12. Lyons, B.A., Tashiro, M., Cedergren, L., Montelione, G.T. An improved strategy for determining resonance assignments for isotopically enriched proteins and its application to an engineered domain of staphylococcal protein A. *Biochemistry* 32:7839-7845, 1993.
13. Richards, F.M. Areas, volumes, packing and protein structure. *Annu. Rev. Biophys. Bioeng.* 6:151-176, 1977.
14. Bowie, J.U., Reidhaar, O.J.F., Lim, W.A., Sauer, R.T. Deciphering the message in protein sequences: Tolerance to amino acid substitutions. *Science* 247:1306-1310, 1990.
15. Rose, G.D., Wolfenden, R. Hydrogen bonding, hydrophobicity, packing, and protein folding. *Annu. Rev. Biophys. Biomol. Struct.* 22:381-415, 1993.
16. Wright, P.E., Dyson, H.J., Lerner, R.A. Conformation of peptide fragments of proteins in aqueous solution. *Biochemistry* 27:7167-7175, 1988.
17. Milburn, P.J., Meinwald, Y.C., Takahashi, S., Ooi, T., Scheraga, H.A. Chain reversal in model peptides: Studies of cysteine-containing cyclic peptides. *Int. J. Pept. Protein Res.* 31:311-321, 1988.
18. Skolnick, J., Kolinski, A. Dynamic Monte Carlo simulations of globular protein folding/unfolding pathways. I. Six member, Greek key β -barrels. *J. Mol. Biol.* 212:787-817, 1990.
19. Sikorski, A., Skolnick, J. Dynamic Monte Carlo simulations of globular protein folding/unfolding pathways. II. α -helical motifs. *J. Mol. Biol.* 212:819-836, 1990.
20. Chou, K.C., Maggiora, G.M., Scheraga, H.A. Role of loop-helix interactions in stabilizing four-helix bundle proteins. *Proc. Natl. Acad. Sci. USA* 89:7315-7319, 1992.
21. Kabsch, W., Sander, C. On the use of sequence homologies to predict protein structure: Identical pentapeptides can have completely different conformations. *Proc. Natl. Acad. Sci. USA* 81:1075-1078, 1984.
22. Argos, P. Analysis of sequence-similar pentapeptides in unrelated protein tertiary structures. *J. Mol. Biol.* 197:331-348, 1987.
23. Brunet, A.P., Huang, E.S., Huffine, M.E., Loeb, J.E., Weltman, R.J., Hecht, M.H. The role of turns in the structure of an α -helical protein. *Nature* 364:355-358, 1993.
24. Vlassi, M., Steif, C., Weber, P., Tsernoglou, D., Wilson, K.S., Hinz, H.J., Kokkinidis, M. Restored heptad pattern continuity does not alter the folding of a four-alpha-helix bundle. *Nature Struct. Biol.* 1:706-716, 1994.
25. Predki, P.F., Regan L., Redesigning the topology of a four-helix-bundle protein: Monomeric ROP. *Biochemistry* 34:9834-9839, 1995.
26. Kolinski, A., Skolnick, J. Monte Carlo simulations of protein folding. I. Lattice model and interaction scheme. *Proteins* 18:338-352, 1994.
27. Godzik, A., Kolinski, A., Skolnick, J. Lattice representation of globular proteins: How good are they? *J. Comp. Chem.* 14:1194-1202, 1993.
28. Metropolis, N.A., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., Teller, E. Equation of state calculations by fast computing machines. *J. Chem. Phys.* 21:1087-1092, 1953.
29. Skolnick, J. Lattice parameter set available via anonymous ftp from ftp.scripps.edu from /pub/skolnick/mutant directory. 1995.
30. Milik, M., Kolinski, A., Skolnick, J. Neural network system for the evaluation of side chain packing in protein structures. *Protein Eng.* 8:225-236, 1995.
31. Vieth, M., Kolinski, A., Brooks III, C.L., Skolnick, J. Prediction of the folding pathways and structure of the GCN4 leucine zipper. *J. Mol. Biol.* 1994:361-367, 1994.
32. Olszewski, K.A., Kolinski, A., Skolnick, J. Does a backwardly read protein sequence have a unique native state? *Protein Eng.* 9:5-14, 1996.
33. Behe, M.J., Lattman, E.E., Rose, G.D. The protein-folding problem: The native fold determines packing, but does packing determine the native fold. *Proc. Natl. Acad. Sci. USA* 88:4195-4199, 1991.
34. Wilmot, C.M., Thornton, J.M. Analysis and prediction of the different types of β -turn in proteins. *J. Mol. Biol.* 203:221-232, 1988.
35. Bergman, T., Erickson, K., Galyov, E., Persson, C., Wolf-Watz, H. The *lcrB* (*yscN-U*) gene cluster of *Yersinia pseudotuberculosis* is involved in Yop secretion and shows high homology to the *spa* gene clusters of *Chigella flexneri* and *Salmonella typhimurium*. *J. Bacteriol.* 176:2619-2626, 1994.
36. Zhang, J., Chung, D.W., C.K., T., Downey, K.M., Davie, E.W., So, A.G. Primary structure of the catalytic subunit of calf thymus DNA polymerase delta. Sequence similarities with other DNA polymerases. *Biochemistry* 30:11742-11750, 1991.
37. Hoff, T., Stummann, B.M., Henningsen, K.W. Cloning and expression of a gene encoding a root specific nitrate reductase in bean *Phaseolus vulgaris*. *Physiol. Plantarum* 82:197-204, 1991.
38. Bairoch, A., Boeckmann, B. The SWISS-PROT protein sequence data bank: Current status. *Nucleic Acids Res.* 22:3578-3580, 1994.
39. Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer Jr, E.F., Brice, M.D., Rodgers, J.R., Kennard, O., Simanouchi, T., Tasumi, M. The protein data bank: A computer-based archival file for macromolecular structures. *J. Mol. Biol.* 112:535-542, 1977.
40. Pearlman, D.A., Case, D.A., Caldwell, J.C., Seibel, G.L., Singh, U.C., Weiner, P., Kollman, P.A. AMBER. University of California, San Francisco, 1991, 4.0.
41. Jorgensen, W.L., Chandrasekhar, J., Madura, J.D., Impey, R.W., Klein, M.L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79:926-935, 1983.