ELSEVIER

SymROP: ROP protein with identical helices redesigned by all-atom contact analysis and molecular dynamics

Daniel Grell,* Jane S. Richardson,† David C. Richardson,† and Manfred Mutter*

**Institute of Organic Chemistry, University of Lausanne, Lausanne, Switzerland*
†Biochemistry Department, Duke University, Durham, North Carolina, USA

Experience has shown that protein redesigns (using the backbone from a known protein structure) are far more likely to produce well-ordered, native-like structures than are true de novo designs. Therefore, to design a four-helix bundle made of identical short helices, we here proceed by an extensive redesign of the ROP protein. A fully symmetrical SymROP sequence derived from ROP was chosen by modeling ideal-geometry side chains, including hydrogens, while maintaining the "goodness-of-fit" of side-chain packing by calculating all-atom contact surfaces with the Reduce and Probe programs. To estimate the probable extent of backbone movement and side-chain mobility, restrained molecular dynamics simulations were compared for candidate sequences and controls, including substitution of Abu for all or half the core Ala residues. The resulting 17-residue designed sequence is 41% identical to the relevant regions in ROP. SymROP is intended for construction by the Template Assembled Synthetic Proteins approach, to control the bundle topology, to use short helices, and to allow blocked termini and unnatural amino acids. ROP protein has been a valuable system for studying helical protein structure because of its simplicity and regularity within a structure large enough to have a real hydrophobic core. The SymROP design carries that simplicity and regularity even further.

Keywords: contact surfaces, four-helix bundle, protein design, side chain packing, Template Assembled Synthetic Proteins

Abbreviations: Abu, α-aminobutyric acid; MD, molecular dynamics; ROP, repressor of primer; TASP, Template Assembled Synthetic Proteins; U, α-aminobutyric acid.

Color plates for this article are on pages 309–310.

Corresponding author: Manfred Mutter, Institute of Organic Chemistry, University of Lausanne, BCH-Dorigny, CH-1015 Lausanne, Switzerland. Tel.: +41-21-6924011; fax: +41-21-6924015.

E-mail address: Manfred.Mutter@ico.unil.ch (M. Mutter)

INTRODUCTION

Protein de novo design aims to construct polypeptide chains that fold into a predetermined three-dimensional structure, without using the same backbone or a sequence homologous to any particular natural protein.[1–20] Although de novo designs have taught us a great deal about protein folding and structure, all of them, with only a very few exceptions,[21,22] have turned out to adopt a molten–globular-like ensemble rather than the unique, well-ordered conformational state characteristic of native proteins.[23,24] Therefore, most recent work in the design field has concentrated on what we will call protein redesigns, which involve designing changes, often very extensive, to the side chains of a known protein structure. Redesigns have taken good advantage of fast computational repacking algorithms,[25–29] they lend themselves to the development of modified or novel functions,[26,30–32] and a large fraction of them show well-ordered structures, especially if designed using specific hydrophobic interactions with all hydrogen atoms present.[33–35]

We recently developed a method for analyzing both visually and quantitatively the geometrical "goodness-of-fit" within or between molecules using small-probe contact dots,[36] calculated with all H atoms added and optimized.[37] This all-atom steric analysis can discriminate good *vs* bad packing, counting atomic clashes or sparse regions negatively but H-bonds and favorable van der Waals contacts positively. The method is very suitable for protein redesign and is used here to optimize the complementarity of packing in the core, the interface, and the surface of a shortened and symmetrized redesign of Repressor of Primer (ROP) protein. ROP is one of the most regular, and probably the most thoroughly studied and modified, of the naturally occurring four-helix bundle proteins. Its crystal structure has been determined at high resolution for wild-type (1rop;[38]) and for two different loop variants (1rpo[39]; 1gto). It is

Journal of Molecular Graphics and Modelling 18, 290–298, 2000

655 Avenue of the Americas, New York, NY 10010

1093-3263/00/$–see front matter
PII S1093-3263(00)00049-8

very stable and tolerant towards mutations,[39,40] although its stability,[41,42] folding rate,[43] and even oligomerization[44,45] have been altered. ROP has had its core redesigned,[46] has been reconnected into a monomer,[47] and a completely unrelated protein has even been transformed to the ROP fold with less than 50% sequence change.[48]

In contrast to most four-helical bundles in globular proteins, ROP is a dimer of two identical, 63-residue, helix hairpin subunits. The helices of the first monomer unit are here referred to as helix AI and AII and those of the second monomer as BI and BII. Helix I and helix II have the same hydrophobic periodicity but different sequences. The central hydrophobic core of ROP occurs in six well-defined, regularly repeated layers (Figure 1), each with two small (Ala, Cys, or Thr) and two large (Leu or Ile) side chains, one from each helix. This arrangement can be described as a four-helix anti-parallel coiled-coil, where the small core side chains are in position "a" of the heptad repeat and the large ones are in position "d." The helices coil slightly around the common axis to maintain a similar geometry and cross-section at each layer. The helices diverge slightly near their ends, producing one more core layers at each end with larger side chains (Phe, Ala, Leu, Glu) where the Phe comes from an "e" rather than a "d" heptad position and the Ala is in a different orientation and forms the N-cap of helix II. The ROP monomers pack together anti-parallel, head to tail, so that the single two-fold axis is perpendicular to the helix-bundle axis, running left to right through the center of Figure 1 between the monomers.

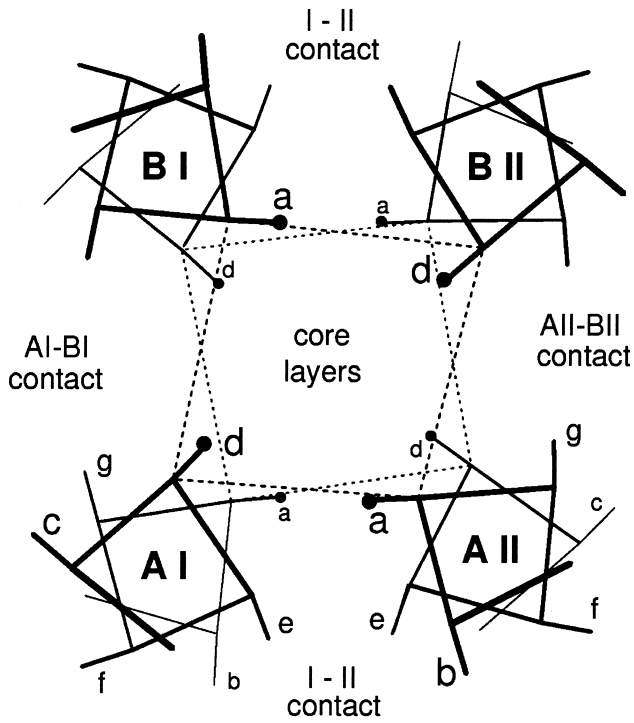


Figure 1. Idealized representation of a cross-section through the ROP structure showing the helix packing and arrangement of heptad positions (a–g) in each helix. The circumference C of each of the two core layers shown is defined as the sum of the four distances between Cα atoms, as outlined by the dashed lines. In ROP, the two monomer contacts (I-II) are tight and identical, whereas the AI-BI and the AII-BII contacts are wider and are different because the sequences of helix I and helix II differ.

An alternative way to describe the ROP tertiary structure is as an unusually tight (8.5 Å between helix axes) two-helix anti-parallel coiled-coil for each subunit as shown in Figure 2a, with the small Ala side chain on each helix touching the backbone of its partner helix, as described in Gernert et al.,[49] while the two monomers are related by wider (9.5–10 Å), more normal helix-helix contacts using the Leu residues, as shown in Figure 2b. As illustrated in the cross-section of Figure 1, the tight contacts within the two monomers are of course identical, while there are two similar but different helix-helix contacts that form the dimer, one between AI and BI (the I-I contact) and one between AII and BII (the II-II contact).

The symmetry and relative simplicity of the ROP structure have made it a popular and productive candidate for protein redesign. The possibility of exploiting an even higher level of latent symmetry makes it a suitable choice for our current attempt to use redesign rather than de novo design in creating a small anti-parallel bundle of four identical helices. To help further in predicting their suitability, the proposed bundles of identical helices are tested by energy minimization, restrained molecular dynamics (MD) simulations, secondary structure prediction, homology searches, and surface and volume calculations. The "SymROP" design should have the advantages of a redesign, because its backbone and interfaces are taken from regions of the ROP structure. But it also should share some theoretical and experimental advantages of simple de novo designs, because it consists of a single 17-residue sequence. Because helices of this length normally arrange themselves only poorly,[50] the four helices of SymROP are attached in a final step to a small template molecule, using the Template Assembled Synthetic Proteins (TASP) approach we developed

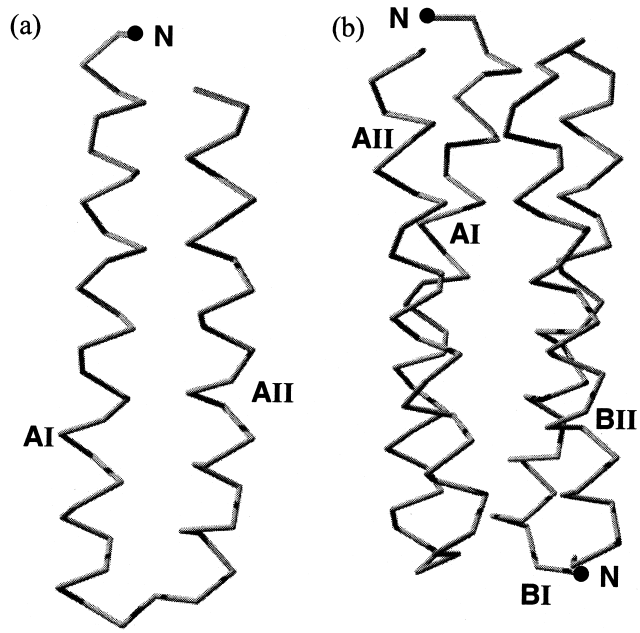


Figure 2. Cα trace of ROP viewed from the side. (a) Monomeric helix hairpin unit AI/AII. (b) Dimeric ROP molecule made up of two helix hairpin units.

to direct helical or β-sheet peptide blocks to characteristic folding units.[51] Producing SymROP by peptide synthesis requires the design criteria to include peptidosynthetical requirements such as the terminal Lys for template attachment, but it allows the use of non-natural amino acids and blocked chain termini.

RESULTS AND DISCUSSION

Initial Modeling

Starting from the structure of ROP protein, as described in the Introduction, our aim was to design a variant with full 222 symmetry and four identical helices, but to preserve the same types and geometries of helix contacts and about half of the same side-chain interactions. The changes essential for this process fall into two categories: (a) shortening the helix bundle to its regular central section, and (b) choosing between the two approximately equivalent realizations of each contact type that occur in the natural ROP protein and ensuring compatibility at the interfaces between changed and unchanged contacts.

The two endmost core layers of ROP are not suitable for a symmetrized design because they are broader, lack the four-residue symmetry of the central layers, and involve inherently asymmetrical interactions with the chain termini and the hairpin connection. It would be possible to imagine using either the central four layers or the central six layers. The outer layers of the central six involve an anomalous and asymmetrical Gln in one of the "a" core positions. Although it is known that a standard Ala-Leu sequence works in those layers,[46] the resulting structure is not known. It would necessarily involve slight backbone shifts and more substantial surface shifts to compensate for the Gln→Ala change, so that redesign surrounding it could not proceed from a known configuration. To avoid those complications and to provide a short sequence for convenient synthesis, the initial SymROP design uses a five-turn helix of 17 residues, including the central four core layers of ROP plus three residues on each end. That length exceeds the critical sequence length for α-helices[52] and can be expected to form and associate when kept in proximity by linkage to a suitable template. The equivalent portion of ROP consists of residues A9-A25 and B9-B25 of helices AI and BI, plus residues A35-A51 and B35-B51 of helices AII and BII, as shown in Figure 3. The ROP molecule as trimmed back to just those four regions, but with the natural sequence, is called ROPsh, for ROP shortened. The initial model of ROPsh was built by adding all H atoms to the 1rop X-ray structure (see Methods) and truncating. Both full-length ROP and ROPsh were subjected to energy minimization, and their all-atom contact dots were calculated and displayed before and after (see Methods). Although side chains moved on the surface, the interior was almost unchanged for either form. Color Plate 1 shows small-probe contact dots that document the excellent packing of side chains in the four central core layers of ROPsh after energy minimization. These are the core packing interactions that we must equal or surpass in the symmetrized design.

The second design stage is to achieve higher symmetry by picking one of the versions of each type of helix contact and duplicating it symmetrically. The first of three steps has already been done for us, with the successful ROP core redesign that symmetrizes each four-residue layer to contain Ala in both of the "a" positions and Leu in both "d" positions,[46] thus making the inner core sequence identical for all four helices. That set of changes (between ROPsh and ROPsh A_4L_4 in Figure 3) is incorporated in the choice for SymROP to follow Ala12 of ROP helix I and Leu41 and Ala45 of helix II (rather than Cys38, Ile15, or Thr19). The second step involves choosing between the I-I and the II-II dimer contacts, which are cleanly separated, consist of "g" and "c" positions, and each includes a residue at the center that makes self-contact around the ROP two-fold axis. The I-I contact, if duplicated, would produce four exposed Met (which could be changed) and four exposed Phe (which could not)—even Tyr does not work for

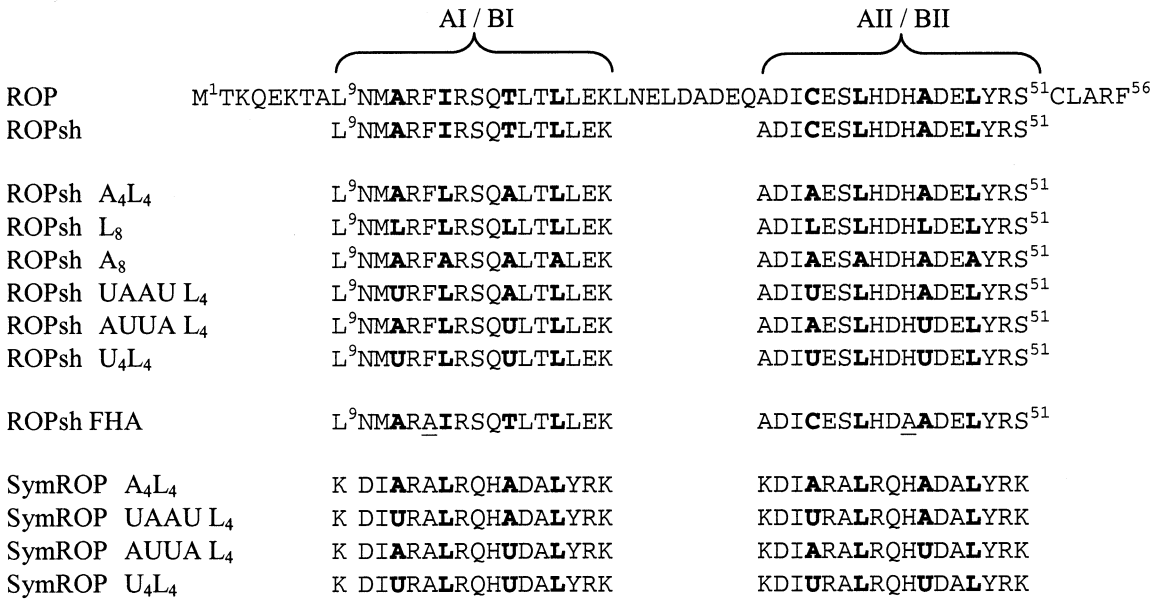


Figure 3. Sequence alignment of ROP with variants of ROP and SymROP. U = Abu. Core amino acid residues are indicated in bold; changes in ROPsh FHA are underlined.

their tight self-interaction. SymROP uses the II-II interface, which exposes mostly polar residues including His at the central two-fold, thus determining residues Ile37 and His44 in the "g" positions of SymROP. The "c" positions do not actually make helix-helix contacts in ROP, and they have been replaced with Ala40 and Ala47 to improve helix propensity. Unfortunately, the choice of using the II-II contact means there is no chance SymROP would show any RNA-binding activity, because that binding site is at the I-I interface.[53] The third step is to symmetrize the I-II contact by choosing one half of the ROP monomer contact to duplicate, thus determining SymROP heptad positions "e" and "b." Either half-contact must work internally, because they both occur in ROP, but at the central changeover point a new two-fold axis is introduced, which in our modeling worked better if SymROP used the "top" half of the I-II contact that includes the chain termini. That choice determines Arg13, Arg16, Asp46 (= Asp20), and Tyr49 (= Tyr23). The remaining heptad position "f" is completely exposed; SymROP uses either the residue from helix II where there is maximum homology (Asp36, Arg50) or else Gln (43) for a polar residue with high helix propensity. Asp36 (= Asp9) also serves as the helix N-cap.[54]

The amino acid residues at the outside of the helix-helix interfaces of coiled coils can be used to help stabilize the fold into the desired topology.[55] We think that these residues at positions "g" and "e" of a four-helix bundle make a major contribution to the stability and specificity of the bundle structure. On the surface of the natural dimeric ROP protein, several nonpolar amino acid residues can be found, including Leu20 and Ile37. This is surprising because of the risk of misfolding of the helices. However, these side chains are well packed with the neighboring helix. The arrangement of these relatively rigid residues at the helix interface (especially the Ile) might be able to hinder a turning motion of one helix alone—locking the cogwheels. Therefore, SymROP has kept Ile37, although it was judged that Leu20 could be replaced by an Asp.

After the decision to use the II-II contact, an even more symmetrical model was constructed by superimposing the helix II backbone onto the helix I location; this produces only very small shifts. All stages of the symmetrization were mutated on both models. They were checked for suitable charge distribution relative to the helix dipoles and to possible salt links. Favorable all-atom contacts were ensured by adjusting torsion angles (within favorable rotamer ranges) to give good Probe contact dots. The resulting basic sequence of SymROP is shown in Figure 3. It is 41% homologous to the relevant regions in wild-type ROP and 44% to the Ala-Leu core redesign sequence.[46]

Tests of Backbone and Side-Chain Mobility

An especially vulnerable point in this process, shared with almost all protein redesigns, is the assumption of unchanged backbone conformation from the native-protein model. To help judge the extent and direction of such changes and to help choose sequences that would minimize such movements, a series of restrained MD simulations were performed on full-length ROP, ROPsh, and SymROP, including various alternative sequences. To reduce the calculation time we did not use explicit solvent molecules; therefore, the results can only be treated in relation to structures subjected to the same MD protocols. We restrained the secondary structure of the protein to its initial helicity by applying a strong force constant to the hydrogen bonds of the helices, because our interest was in the arrangement of the helices—the tertiary structure. Figure 4a compares the time course of MD simulation for ROP and for ROPsh, following the root mean square deviation (RMSD) change from the X-ray coordinates. ROPsh is stable but changes more than ROP does, by a slight unwinding and separation at the ends of the helices as well as movement of some surface side chains. ROPsh is necessarily less stable than ROP, because ten residues in each helix plus the connecting loop are missing.

To follow changes in size and arrangement of the helix-bundle structure, we measured the circumference C of the core layers (as defined in Figure 1). The average circumference gives a single value that is sensitive to expansion or contraction of the helix contacts. As listed in Table 1, the circumference of ROP and ROPsh stayed very similar through the MD simulations, indicating that the basic ROP helix bundle can be maintained even with such drastic size reduction.

Special attention was given to the interface residues of ROP that interact with themselves on either end of the two-fold axis: His44 and Phe14. These two side-chain pairs sit in the center of the surface of the AI/BI or AII/BII contacts. The HisA44 and HisB44 rings form a hydrogen bond (Color Plate 2a), whereas PheA14 and PheB14 form extensive van der Waals contacts (Color Plate 2b). Both interactions give additional stability at these central interface positions.

Mutation studies on the complete ROP protein[53] showed that a point mutation F14A does not disturb the overall stability of ROP. In our shortened ROP form, however, these amino acids could have a greater effect on stability. Therefore, for restrained MD simulations in the context of our ROPsh model, we tried exchanging Phe 14 and His 44 with Ala, giving the sequence called ROPsh FHA in Figure 3. Figure 4b includes the time course of a MD simulation in which the overall structure of ROPsh FHA *vs* ROPsh is significantly changed, resulting in a less stabilized bundle. Especially at the ends of the structure the helices twist away from each other, as indicated by the increasing circumference of the outer layers of ROPsh FHA plotted in Figure 4c. The conclusion is that we should not omit such interactions in a shortened ROP, and indeed the SymROP design includes two copies of the His-His interaction.

Core Sequence Variants

The complementary arrangement of the side chains in the interior of a protein is essential for stability and specificity of structure. Mutations in the core of ROP protein have been extensively studied,[33,40,41,43,46,55,56] but the extreme stability of natural ROP protein might hide deleterious effects, which could do more harm in our shortened ROP structure. Therefore, we reinvestigated possible changes in terms of their ability to help or to hurt the quality of core packing.

Color Plate 1 shows all-atom contact dots for the sliced layers of ROPsh after addition of hydrogen atoms and energy minimization. All the residues in the core at positions "a" and "d" are very well packed, as shown by their contact dot surfaces, except for Ala45, which leaves a cavity at one side. However, the most stable and still active mutant of ROP so far is built up of an all Ala-Leu core.[33] This mutant, with an increased stability of $\Delta\Delta G = +4.9$ kcal mol^{-1} in comparison

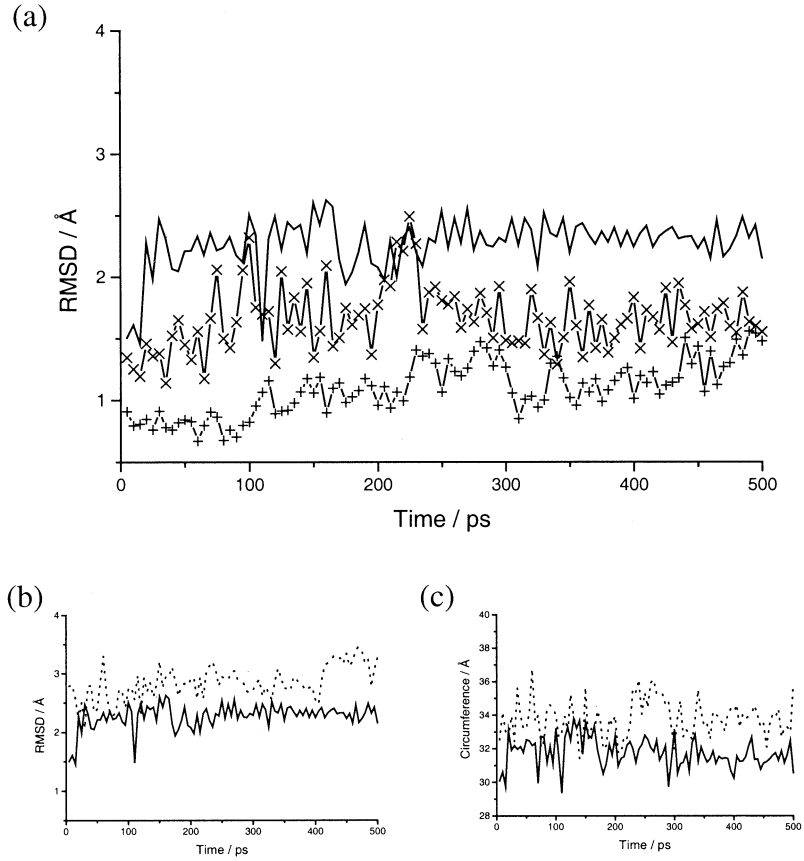


Figure 4. Time course of restrained MD simulation at 300K for ROP (−+−+−), ROPsh (——), SymROP A_4L_4 (−×−×−), and ROPsh FHA (- - -). Development of (a) RMSD of ROP, ROPsh, and SymROP A_4L_4 vs X-ray coordinates (1rop); (b) RMSD of ROPsh and ROPsh FHA vs 1rop; (c) layer circumference (average of layers 1 and 4) for ROPsh and ROPsh FHA.

to the wild-type protein, has a repeated alternation of the small residue Ala at the "a" position with the large residue Leu at the "d" position of all core layers. This same arrangement is used for the four layers of ROPsh A_4L_4, but the potential cavities are a concern.

Energy minimization of the ROPsh A_4L_4 showed no decrease in the size of the holes in the layers, as shown in Color Plate 3a. Cavities like these might promote disordered structure in a shortened ROP version. Therefore, non-natural amino acids, which have a similar volume as those in the original ROP sequence, were chosen (e.g., Abu = U = α-aminobutyric acid) as candidates to prevent cavities caused by core mutations. The van der Waals volumes are calculated according to F.M. Richards.[57] In particular, the exchange of Ala45 and Thr19 ("a" positions of middle layers) with Abu45 and Abu19 is ideal: $\Delta V < 4$ Å^3. Color Plate 3b shows the almost perfect packing in those central layers of the ROPsh U_4L_4 mutant. All Abu exchanges have an increased packing score value (Table 1) very similar to that of native-sequence ROPsh.

As controls for comparison of large size changes, two other

Table 1. Summary of geometric properties of full-length and shortened ROP proteins

Model	Core packing score	Volume (Å^3)	RMSD *vs* 1rop (Å)	Layer circumference C (Å)	RMS fluctuation of core side chains (Å)
ROP	78.8	12140	1.1	29.8	0.7
ROPsh	71.4	7562	2.3	30.6	0.9
ROPsh A_4L_4	69.3	7478	2.3	29.8	1.2
ROPsh L_8	80.1	7908	3.3	39.3	1.6
ROPsh A_8	48.9	7053	1.6	24.4	—
ROPsh U_4L_4	69.5	7581	2.0	30.9	1.2
ROPsh UAAU L_4	71.0	7527	2.2	29.7	1.6
ROPsh AUUA L_4	71.6	7521	1.5	29.8	0.8

Packing score for core side chains (normalized by area) and total volume were calculated on the energy minimized structure. RMSD was averaged over all restrained MD structures, measured *vs* 1rop X-ray coordinates for all backbone atoms. Cirucmference C (defined in Figure 1) was averaged over all restrained MD structures for all four layers. RMS fluctuation of nonhydrogen side-chain atoms was averaged over last 250 ps of MD.

Table 2. Summary of geometric properties of SymROP models

	Core packing score	Volume (Å)^3	RMSD *vs* 1rop (Å)	Layer circumference C (Å)
SymROP A_4L_4	79.2	7810	1.5	29.0
SymROP U_4L_4	78.4	7927	1.9	30.6
SymROP UAAU L_4	72.1	7888	1.9	29.6
SymROP AUUA L_4	77.6	7870	1.7	30.2

Packing score for core side chains (normalized by area) and total volume were calculated on the energy-minized structure. RMSD was averaged over all restrained MD structures, measured *vs* 1rop X-ray coordinates for all backbone atoms. Circumference C was averaged over all restrained MD structures for all four layers.

core model proteins were included in the MD studies. One was built of eight Ala residues in the core, ROPsh A_8, and the other with eight Leu residues, ROPsh L_8. Table 1 summarizes the overall structural similarity (RMSD *vs* X-ray coordinates of ROP), the size of the bundle (layer circumference C), and the mobility of the side chains in the core (RMS fluctuation of nonhydrogen atoms). The calculated volumes of the energy-minimized structures of the bundles (Table 1) show that only mutants containing Abu achieve the same size as ROPsh, whereas all the others change volume and circumference significantly according to the size of the employed core residues. ROPsh AUUA L_4 shows the highest degree of similarity and lowest mobility of all the core models, indicating that this is the most favorable sequence. Although the ROPsh L_8 core model results in the best packing score because of the large surface of the 16 total Leu residues, the helix-bundle structure is disturbed significantly (Table 1) and expanded. Despite its structural similarity, ROPsh A_8 has shrunk too much and it has more overall mobility. As expected in the context of ROP organization, the alternation of small and large residues not only inside the layers but on the helices as well is essential for a stable, well-defined bundled arrangement. In contrast, the various sequences using Abu maintain a close similarity to the ROP geometry while improving internal packing. The combination of Abu and Leu in just the two middle layers seems overall to be the most promising, with A_4L_4 next best.

To test these conclusions further, MD simulations were performed on the Abu variants in the context of the SymROP structure. The results are summarized in Table 2. All forms are highly similar to ROP, and the structure is very well maintained. The variant with Abu residues in the middle layers (SymROP AUUA L_4) again seems the most promising, because of its good packing and its structural similarity to ROP. Overall, the MD tests suggest that both the basic SymROP sequence (A_4L_4) and the SymROP AUUA L_4 version would be good candidates for experimental test.

Homology Tests and Secondary-Structure Prediction

The SymROP sequence shows a 30% homology to ROP helix AI/BI and 50% to AII/BII. Homology searches performed on the SWISSPROT database targeted only members of the ROP family as reasonable closest matches. Secondary structure predictions of such short sequences must be treated with caution. However, SymROP A_4L_4 is predicted to have an increased helicity in comparison to ROPsh according to GOR4, whereas other methods predict in general a high degree of helicity for all the sequences (Table 3). The AGADIR method predicts a 13% helix content for SymROP, in contrast to only 4% for the ROPsh helix I sequence and 1% for helix II. This demonstrates that a sequence with high helix-forming propensity was chosen for SymROP.

Attachment to the Template

We need to attach the four helices to a matrix—a template molecule—as shown in Color Plate 4, because of the low tendency of short helices to form bundled structures.[50] For this purpose, we chose a cyclic decapeptide cyclo(-Pro-Gly-Cys-Ala-Cys-Pro-Gly-Cys-Ala-Cys-) previously shown to produce preorganized structures of predetermined anti-parallel topologies, lowering the amount of entropy needed for association.[51] In this case, two different regioselectively addressable helix attachment sites are needed to provide separate synthetic connectivities for the up and the down helices. Orthogonal protec-

Table 3. Secondary structure predictions by GORIV,[63] LEV,[64] SOPM,[65] HNN, and percent helicity in solution based on AGADIR[62] for SymROP A_4L_4, ROPsh AI/BI and ROPsh AII/BII

	SymROP A_4L_4	ROPsh helix I	ROPsh helix II
GOR4	ccccccchhhhhheeeec	ccccccccceeeeeeeec	cccccccccccceeeec
LEV4	hhhhhhhhhhhhhhhhh	hchhhhhhhhhhehhht	hhhhhhhhhhhhhhhht
SOPM	hhhhhhhhhhhhhheec	hhhhhhhhhhhhhhhhh	hhhhhhhhhhhhhhhhh
HNN	chhhhhhhhhhhhhhcc	cchhhhhhhhhhhhhcc	ccchhhhhhhhhhhhcc
AGADIR	13%	4%	1%

c = random; h = helix; e = sheet; t = turn.

tion of Cys residues on the template allows stepwise condensation of the helices. Lys is needed at either end of the SymROP sequence, because the helices are covalently attached by a thioether linkage of the Lys side chains to the template [Cys(CH_2CO-N^6Lys)], two from the N-terminal and two from the C-terminal end. Additionally, we tested a second elongated linker system by inserting two Gly residues connected via amide bonds to the Lys side chains as spacers [Cys(CH_2CO-Gly-Gly-N^6-Lys)]. Although the linker systems were of different lengths, restrained MD simulations showed both a slight increase in the structural similarity to ROP (RMSD -0.2 to -0.4 Å) and a decrease of the mobility of the core side chains in comparison to simulations of the bundle alone. The averaged values of the circumferences C were comparable to the values of the bundle alone. The linkage to the template therefore does not disturb the overall structure and perhaps even makes a further stabilizing contribution.

The template organization and the synthesis as isolated helices make possible other future variations on SymROP. Versions could be made using all six ROP core layers or utilizing the I-I contact to try for RNA binding. A second template could be linked to the other end of the bundle structure to make an even more stable assembly.

CONCLUSION

The SymROP design makes explicit a previously unrecognized higher symmetry that is approximately present in the natural ROP molecule. ROP has been extremely useful as a model system, mostly because of its regularity and simplicity. SymROP takes those advantages even further, with a single 17-residue sequence expected to assemble into a highly symmetrical four-helix bundle with a real hydrophobic core despite its small size. All-atom contact analysis was essential to understanding the detailed choreography of the various ROP helix contacts; MD helped in analyzing the effects of particular substitutions in avoiding undesired structural changes; and the TASP approach for synthesis should allow rapid but controlled assembly of the pieces. SymROP and its variants can be used to elucidate the general factors governing helix assembly, topology, and packing, and they may serve as versatile scaffolds on which to build functional properties. Although many choices in SymROP design are highly constrained, others are not, so that it represents an entire family of possible molecules all based on a common symmetrical organization but differing significantly in size and surface properties.

METHODS

Intitial Modeling

Preliminary evaluation of the symmetry possibilities and the choice of contacts available were done on the 1rop and 1rpo X-ray structures of ROP protein, as displayed in Mage.[58] Hydrogen atoms were added and then optimized, including combinatorial analysis of local H-bond networks, by the program Reduce (version 1.13[37]); the side-chain amide of Gln 34 was flipped by 180° to improve H-bonds and reduce clashes. A shortened version was made by restriction to 17 residues per helix, centered on the central four core layers. A somewhat symmetrized backbone was constructed by superimposing helix II onto helix I. Side chains were mutated in Prekin (version 5.3[58]) for trials of alternative sequences. For both the original ROP structures and the SymROP trials, small-probe contact dots and packing scores were calculated with the program Probe (version 1.23[36]) using a probe sphere of 0.25 Å radius and were displayed in Mage along with the model.

Model Analyses, Energy Minimization, and Secondary-Structure Prediction

Energy minimization and MD simulations were done on a Silicon Graphics Octane using the program CVFF/FDiscover (version 97; Molecular Simulations, 1997), starting from the 1rop structure, with H atoms added by Reduce. Substitution of single residues was done either by default routines of the program InsightII (version 97; Molecular Simulations, 1997) or with Swiss-PdbViewer (version 3.1[59]), maintaining backbone geometry. Side-chain torsion angles were adjusted if possible to those of ROP. At each C-terminal end, NH_2 was added while the N-terminal ends were acetylated, to avoid unfavorable charges near the helix termini. In the calculations of TASPs, the coordinates of the template were fixed to their initial values, taken from its solution nuclear magnetic resonance structure.[60] Minimization procedures consisted of 500 steps of steepest descent, followed by a conjugate gradient minimization until the RMS gradient was <0.01 kcal/molÅ, not taking charges into account. For these minimizations, a dielectric term ($\epsilon = 1.00$) and a cut-off distance of 12 Å were chosen.

Volumes of the minimized models were calculated by the program VOIDOO (version 2.4[61]). Secondary-structure predictions were performed with the help of the Web server at Lyon, France (http://pbil.ibcp.fr/cgi-bin/NPSA) and with the software AGADIR1s-2[62] at http://www.embl-heidelberg.de/Services/serrano/agadir/agadir-start.html.

Restrained MD Simulations

The MD runs were performed using standard FDiscover protocols at a constant temperature of 300K, with a distance-dependent dielectric term ($\epsilon = 4.00$*r) approximating aqueous surroundings, including charges and a cut-off at 12 Å. During the initialize stages, constant temperature and pressure are accomplished by direct velocity scaling and temperature bath coupling. To constrain the helix secondary structure, a force constant of 1000 kcal/molÅ was applied to the backbone H-bonds. The H-atoms of the starting structure were adjusted to standard ionization states at pH 7. The structure was minimized for 5000 steps of steepest descent. After 1 ps of heating and 10 ps of initialization time, the system was subjected to a 500 ps simulation time with 1 fs time steps. Structures were saved every 5 ps. The average structure for the last 50 ps was generated by InsightII and once more energy minimized.

REFERENCES

1 Richardson, J.S., and Richardson, D.C. In: *Protein engineering*, Liss, A.R., Ed., Oxender & Fox, New York, 1987, pp 149–163

2 DeGrado, W.F., Wasserman, Z.R., and Lear, J.D. Protein design, a minimalist approach. *Science* 1989, **243,** 622–628

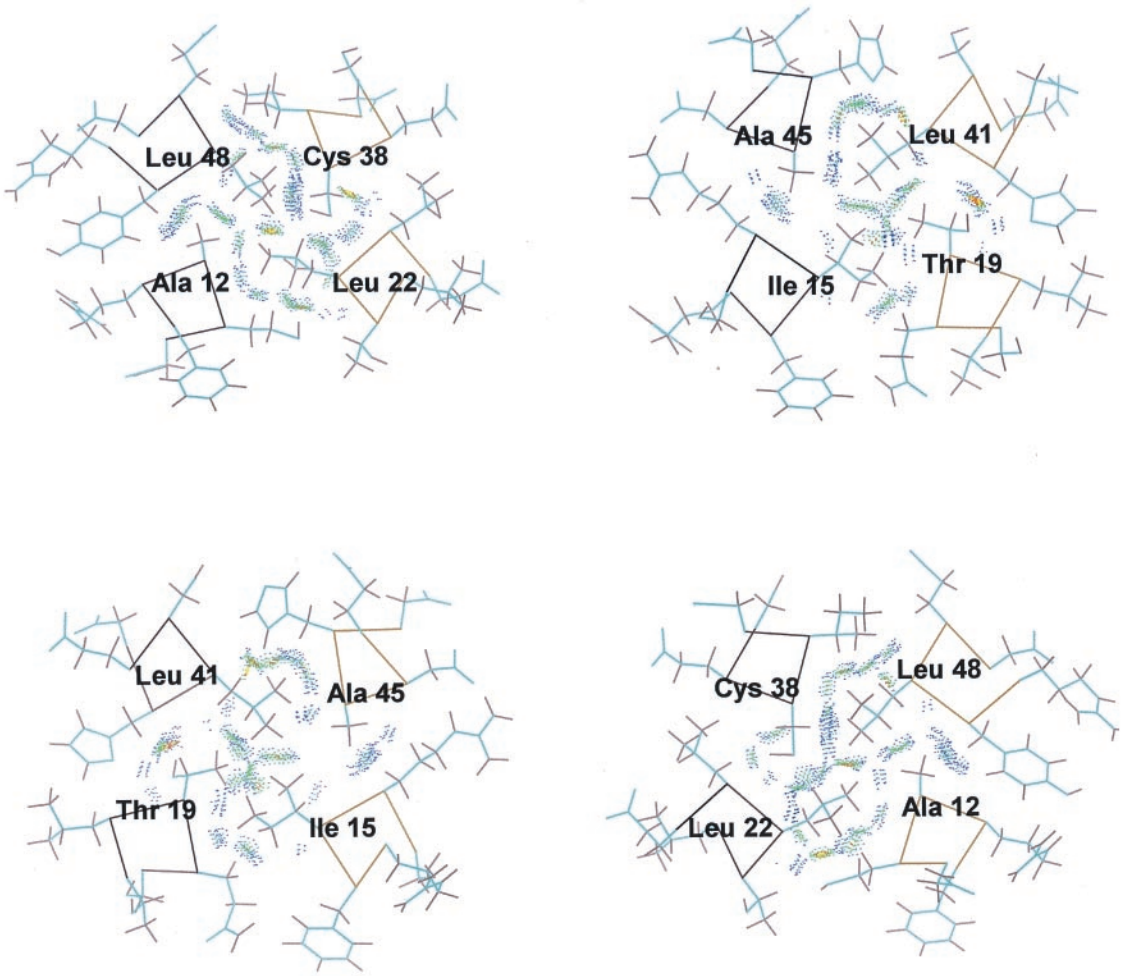
3 Fedorov, A.N., Dolgikh, D.A., Chemeris, V.V., Chernov, B.K., Finkelstein, A.V., Schulga, A.A., Alakhov

Yu, B., Kirpichnikov, M.P., and Ptitsyn, O.B. De novo design, synthesis and study of albebetin, a polypeptide with a predetermined three-dimensional structure. Probing the structure at the nanogram level. *J. Mol. Biol.* 1992, **225,** 927–931
4 Hecht, M.H., Richardson, J.S., Richardson, D.C., and Ogden, R.C. De novo design, expression, and characterization of Felix: A four-helix bundle protein of native-like sequence. *Science* 1990, **249,** 884–891
5 Mutter, M., Tuchscherer, G.G., Miller, C., Altmann, K.-H., Carey, R.I., Wyss, D.F., Labhardt, A.M., and Rivier, J.E. Template-Assembled Synthetic Proteins with four-helix-bundle topology. Total chemical synthesis and conformational studies. *J. Am. Chem. Soc.* 1992, **114,** 1463–1470
6 Kamtekar, S., Schiffer, J.M., Xiong, H., Babik, J.M., and Hecht, M.H. Protein design by binary patterning of polar and nonpolar amino acids. *Science* 1993, **262,** 1680–1685
7 Fezoui, Y., Weaver, D.L., and Osterhout, J.J. De novo design and structural characterization of an alpha-helical hairpin peptide: A model system for the study of protein folding intermediates. *Proc. Natl. Acad. Sci. U. S. A.* 1994, **91,** 3675–3679
8 Brenner, S.E., and Berry, A. Protein design by optimization of a sequence-structure quality function. *ISMB* 1994, **2,** 44–52
9 Eroshkin, A.M., Karginova, E.A., Gileva, I.P., Lomakin, A.S., Lebedev, L.R., Kamyinina, T.P., Pereboev, A.V., and Ignat'ev, G.M. Design of four-helix bundle protein as a candidate for HIV vaccine. *Prot. Eng.* 1995, **8,** 167–173
10 Bryson, J.W., Betz, S.F., Lu, H.S., Suich, D.J., Zhou, H.X.X., Oneil, K.T., and DeGrado, W.F. Protein design —A hierarchic approach. *Science* 1995, **270,** 935–941
11 Houbrechts, A., Moreau, B., Abagyan, R., Mainfroid, V., Preaux, G., Lamproye, A., Poncin, A., Goormaghtigh, E., Ruysschaert, J.M., and Martial, J.A. Second-generation octarellins: Two new de novo (beta/alpha) polypeptides designed for investigating the influence of beta-residue packing on the alpha/beta-barrel structure stability. *Prot. Eng.* 1995, **8,** 249–259
12 Smith, D.D., Pratt, K.A., Sumner, I.G., and Henneke, C.M. Greek key jellyroll protein motif design: Expression and characterization of a first-generation molecule. *Prot. Eng.* 1995, **8,** 13–20
13 Parthasarathy, R., Chaturvedi, S., and Go, K. Design of alpha-helical peptides: Their role in protein folding and molecular biology. *Prog. Biophys. Mol. Biol.* 1995, **64,** 1–54
14 Hodges, R.S. Boehringer Mannheim award lecture 1995. La conference Boehringer Mannheim 1995. De novo design of alpha-helical proteins: Basic research to medical applications. *Biochem. Cell. Biol.* 1996, **74,** 133–154
15 Mutter, M., and Tuchscherer, G. Non-native architectures in protein design and mimicry. *Cell. Mol. Life Sci.* 1997, **53,** 851–863
16 Beasley, J.R., and Hecht, M.H. Protein design—The choice of de novo sequences. *J. Biol. Chem.* 1997, **272,** 2031–2034
17 DeGrado, W.F. Proteins from scratch. *Science* 1997, **278,** 80–81
18 Smith, C.K., and Regan, L. Construction and design of beta-sheets. *Acc. Chem. Res.* 1997, **30,** 153–161
19 Rau, H.K., DeJonge, N., and Haehnel, W. Modular synthesis of de novo designed metalloproteins for light-induced electron transfer. *Proc. Natl. Acad. Sci. U. S. A.* 1998, **95,** 11526–11531
20 Schafmeister, C.E., and Stroud, R.M. Helical protein design. *Curr. Opin. Biotechnol.* 1998, **9,** 350–353
21 Betz, S.F., Bryson, J.W., and DeGrado, W.F. Native-like and structurally characterized designed alpha-helical bundles. *Curr. Opin. Struct. Biol.* 1995, **5,** 457–463
22 Harbury, P.B., Plecs, J.J., Tidor, B., Alber, T., and Kim, P.S. High-resolution protein design with backbone freedom. *Science* 1998, **282,** 1462–1467
23 Richardson, J.S., Richardson, D.C., Tweedy, N.B., Gernert, K.M., Quinn, T.P., Hecht, M.H., Erickson, B.W., Yan, Y., McClain, R.D., Donlan, M.E., et al. Looking at proteins: Representations, folding, packing, and design. Biophysical Society National Lecture, 1992. *Biophys. J.* 1992, **63,** 1185–1209
24 Betz, S.F., Raleigh, D.P., and DeGrado, W.F. De novo protein design: From molten globule to native-like states. *Curr. Biol.* 1993, **3,** 601–610
25 Harbury, P.B., Tidor, B., and Kim, P.S. Repacking protein cores with backbone freedom—structure prediction for coiled coils. *Proc. Natl. Acad. Sci. U. S. A.* 1995, **92,** 8408–8412
26 Hellinga, H.W. Rational protein design —Combining theory and experiment. *Proc. Natl. Acad. Sci. U. S. A.* 1997, **94,** 10015–10017
27 Dahiyat, B.I., and Mayo, S.L. De novo protein design—Fully automated sequence selection. *Science* 1997, **278,** 82–87
28 Desjarlais, J.R., and Clarke, N.D. Computer search algorithms in protein modification and design. *Curr. Opin. Struct. Biol.* 1998, **8,** 471–475
29 Regan, L. Proteins to order? *Structure* 1998, **6,** 1–4
30 Axe, D.D., Foster, N.W., and Fersht, A.R. Active barnase variants with completely random hydrophobic cores. *Proc. Natl. Acad. Sci. U. S. A.* 1996, **93,** 5590–5594
31 Baltzer, L. Functionalization of designed folded polypeptides. *Curr. Opin. Struct. Biol.* 1998, **8,** 466–470
32 Liu, D.R., and Schultz, P.G. Generating new molecular function: A lesson from nature. *Angew. Chem. Int. Ed.* 1999, **38,** 36–54
33 Munson, M., Balasubramanian, S., Fleming, K.G., Nagi, A.D., O'Brien, R., Sturtevant, J.M., and Regan, L. What makes a protein a protein? Hydrophobic core designs that specify stability and structural properties. *Prot. Sci.* 1996, **5,** 1584–1593
34 Dahiyat, B.I., and Mayo, S.L. Probing the role of packing specificity in protein design. *Proc. Natl. Acad. Sci. U. S. A.* 1997, **94,** 10172–10177
35 Davis, A.M., and Teague, S.J. Hydrogen bonding, hydrophobic interactions, and failure of the rigid receptor hypothesis. *Angew. Chem. Int. Ed.* 1999, **38,** 737–749
36 Word, J.M., Lovell, S.C., LaBean, T.H., Taylor, H.C., Zalis, M.E., Presley, B.K., Richardson, J.S., and Richardson, D.C. Visualizing and quantifying molecular goodness-of-fit: Small-probe contact dots with explicit hydrogen atoms. *J. Mol. Biol.* 1999, **285,** 1711–1733
37 Word, J.M., Lovell, S.C., Richardson, J.S., and Richardson, D.C. Asparagine and glutamine: Using hydrogen

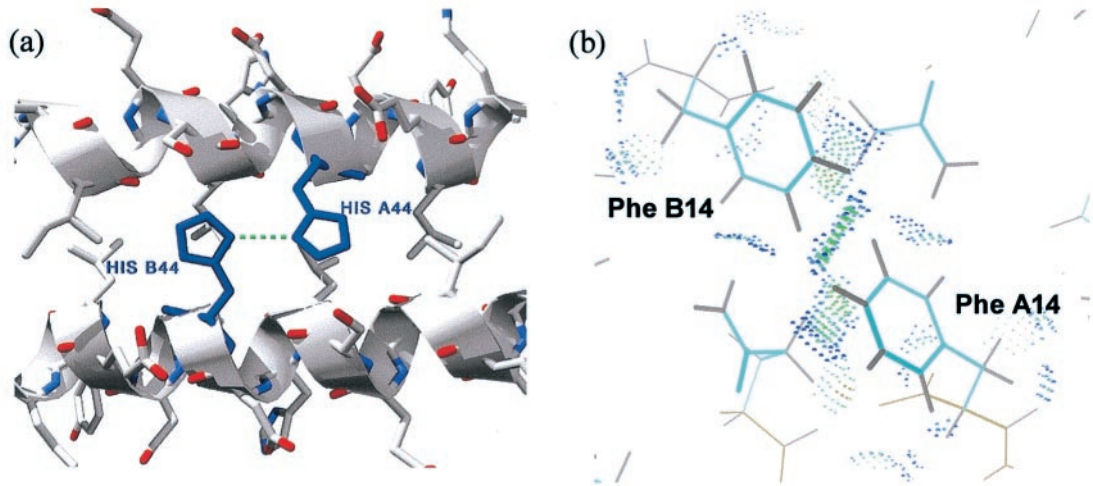
atom contacts in the choice of side-chain amide orientation. *J. Mol. Biol.* 1999, **285,** 1735–1747
38 Banner, D.W., Kokkinidis, M., and Tsernoglou, D. Structure of the ColE1 rop protein at 1.7 Å resolution. *J. Mol. Biol.* 1987, **196,** 657–675
39 Vlassi, M., Steif, C., Weber, P., Tsernoglou, D., Wilson, K.S., Hinz, H.J., and Kokkinidis, M. Restored heptad pattern continuity does not alter the folding of a four-alpha-helix bundle. *Nat. Struct. Biol.* 1994, **1,** 706–716
40 Kokkinidis, M., Vlassi, M., Papanikolaou, Y., Kotsifaki, D., Kingswell, A., Tsernoglou, D., and Hinz, H.J. Correlation between protein stability and crystal properties of designed ROP variants. *Prot. Struct., Funct., Genet.* 1993, **16,** 214–216
41 Steif, C., Weber, P., Hinz, H.J., Flossdorf, J., Cesareni, G., and Kokkinidis, M. Subunit interactions provide a significant contribution to the stability of the dimeric four-alpha-helical-bundle protein ROP. *Biochemistry* 1993, **32,** 3867–3876
42 Nagi, A.D., and Regan, L. An inverse correlation between loop length and stability in a four-helix-bundle protein. *Folding Des.* 1997, **2,** 67–75
43 Munson, M., Anderson, K.S., and Regan, L. Speeding up protein folding: Mutations that increase the rate at which Rop folds and unfolds by over four orders of magnitude. *Folding Des.* 1997, **2,** 77–87
44 Lassalle, M.W., Hinz, H.J., Wenzel, H., Vlassi, M., Kokkinidis, M., and Cesareni, G. Dimer-to-tetramer transformation: Loop excision dramatically alters structure and stability of the ROP four alpha-helix bundle protein. *J. Mol. Biol.* 1998, **279,** 987–1000
45 Glykos, N.M., Cesareni, G., and Kokkinidis, M. Protein plasticity to the extreme: Changing the topology of a 4-alpha-helical bundle with a single amino acid substitution. *Structure* 1999, **7,** 597–603
46 Munson, M., O'Brien, R., Sturtevant, J.M., and Regan, L. Redesigning the hydrophobic core of a four-helix-bundle protein. *Prot. Sci.* 1994, **3,** 2015–2022
47 Predki, P.F., and Regan, L. Redesigning the topology of a four-helix-bundle protein: Monomeric Rop. *Biochemistry* 1995, **34,** 9834–9839
48 Dalal, S., Balasubramanian, S., and Regan, L. Protein alchemy—Changing beta-sheet into alpha-helix. *Nat. Struct. Biol.* 1997, **4,** 548–552
49 Gernert, K.M., Surles, M.C., Labean, T.H., Richardson, J.S., and Richardson, D.C. The alacoil—A very tight, antiparallel coiled-coil of helices. *Prot. Sci.* 1995, **4,** 2252–2260
50 Jaenicke, R. Folding and association of proteins. *Prog. Biophys. Mol. Biol.* 1987, **49,** 117–237
51 Mutter, M., and Vuilleumier, S. A chemical approach to protein design: Template Assembled Synthetic Proteins (TASP). *Angew. Chem. Int. Ed.* 1989, **28,** 535–554
52 Mutter, M. Nature's rules and chemist's tools: A way for creating novel proteins. *TIBS* 1988, **13,** 260–265
53 Predki, P.F., Nayak, L.M., Gottlieb, M.B., and Regan, L. Dissecting RNA-protein interactions: RNA-RNA recognition by Rop. *Cell* 1995, **80,** 41–50
54 Richardson, J.S., and Richardson, D.C. Amino acid preferences for specific locations at the ends of alpha helices. *Science* 1988, **240,** 1648–1652
55 Betz, S.F., and DeGrado, W.F. Controlling topology and native-like behavior of de novo-designed peptides: Design and characterization of antiparallel four-stranded coiled coils. *Biochemistry* 1996, **35,** 6955–6962
56 Steif, C., Hinz, H.J., and Cesareni, G. Effects of cavity-creating mutations on conformational stability and structure of the dimeric 4-alpha-helical protein ROP: Thermal unfolding studies. *Prot. Struct., Funct., Genet.* 1995, **23,** 83–96
57 Richards, F.M. The interpretation of protein structures: Total volume, group volume distributions and packing density. *J. Mol. Biol.* 1974, **82,** 1–14
58 Richardson, D.C., and Richardson, J.S. Kinemages—Simple macromolecular graphics for interactive teaching and publication. *TIBS* 1994, **19,** 135–138
59 Guex, N., and Peitsch, M.C. Swiss-PdbViewer: A fast and easy-to-use PDB Viewer for Macintosh and PC. *Prot. Data Bank Q. Newslett.* 1996, **77,** 7
60 Dumy, P., Eggleston, I.M., Esposito, G., Nicula, S., and Mutter, M. Solution structure of regioselectively addressable functionalized templates: An NMR and restrained molecular dynamics investigation. *Biopolymers* 1996, **39,** 297–308
61 Kleywegt, G.J., and Jones, T.A. Detection, delineation, measurement and display of cavities in macromolecular structures. *Acta Crystallogr. D Biol. Crystallogr.* 1994, **50,** 178–185
62 Munoz, V., and Serrano, L. Development of the multiple sequence approximation within the AGADIR model of alpha-helix formation: Comparison with Zimm-Bragg and Lifson-Roig formalisms. *Biopolymers* 1997, **41,** 495–509
63 Garnier, J., Gibrat, J.F., and Robson, B. GOR method for predicting protein secondary structure from amino acid sequence. *Methods Enzymol.* 1996, **266,** 540–553
64 Levin, J.M., Robson, B., and Garnier, J. An algorithm for secondary structure determination in proteins based on sequence similarity. *FEBS Lett.* 1986, **205,** 303–308
65 Geourjon, C., and Deleage, G. SOPM: A self-optimized method for protein secondary structure prediction. *Prot. Eng.* 1994, **7,** 157–164

Daniel Grell, Jane S. Richardson, David C. Richardson, and Manfred Mutter

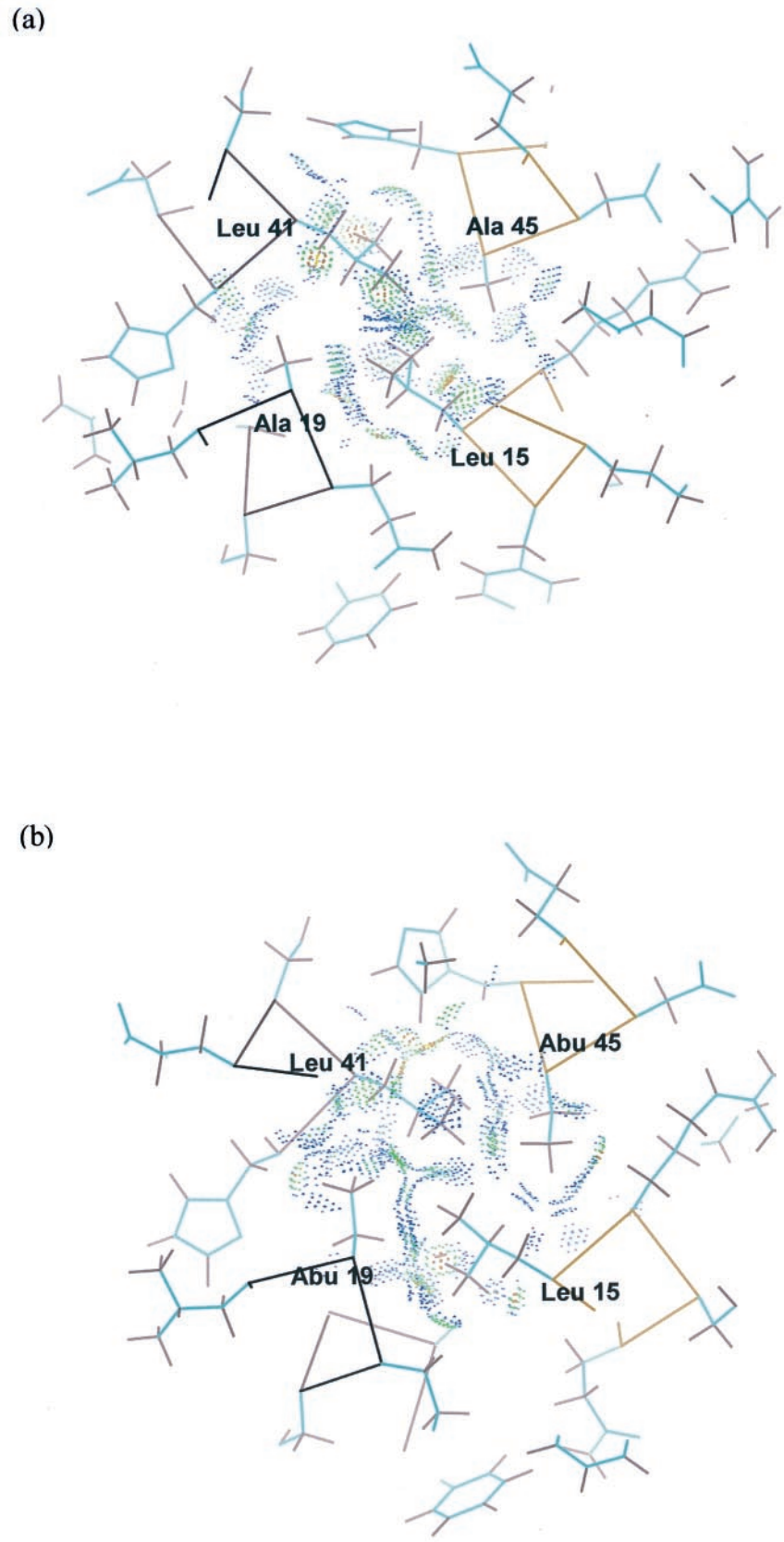
SymROP: ROP protein with identical helices redesigned by all-atom contact analysis and molecular dynamics



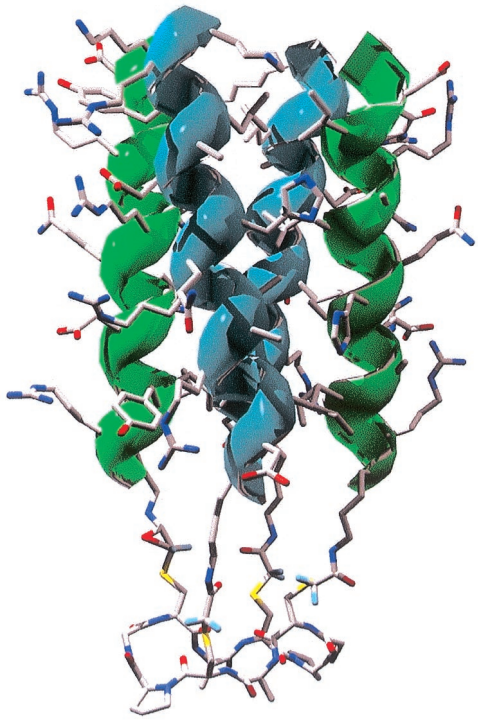
Color Plate 1. Representation of four-residue core layer slices after addition of hydrogen atoms and energy minimization of ROPsh. All-atom contact surfaces of the core amino acids are shown by Probe dots, to indicate the quality of packing. Dot color scheme: green or yellow = good contact (green for narrow gaps, yellow for slight overlaps <0.2 Å); pale green dots = H-bonds; blue dots = wider gaps (0.25 Å); orange or red spikes = unfavorable interpenetrations.



Color Plate 2. Close-up of the X-ray structure of the homodimer of ROP (1rop). (a) Ribbon drawing of HisA44 and HisB44 (black) and their H-bond (dashed line). (b) Surface contact dots representation of PheA14 and PheB14 van der Waals contacts. For dots color scheme see legend to Color Plate 1.



Color Plate 3. Representation of layer slice for residues 15, 19, 41, and 45 after energy minimization of (a) ROPsh A_4L_4 and (b) ROPsh U_4L_4. All-atom contact surfaces are shown by Probe dots. For dots color scheme, see legend to Color Plate 1.



Color Plate 4. Ribbon drawing of the template assembled four-helix bundle SymROP model. Helix backbones are shown by ribbons, joined to the template (bottom) by four linkers.