Loopless Rop: Structure and Dynamics of an Engineered Homotetrameric Variant of the Repressor of Primer Protein

Nicholas M. Glykos,[‡,§] Yannis Papanikolau,[§,‖] Metaxia Vlassi,[§,⊥] Dina Kotsifaki,[‖] Giovanni Cesareni,[#] and Michael Kokkinidis*,[‖]

Department of Molecular Biology and Genetics, Democritus University of Thrace, Dimitras 19, 68100 Alexandroupolis, Greece, Department of Biology, University of Crete, P.O. Box 2208, GR-71409 Heraklion, Crete, Greece, Institute of Molecular Biology and Biotechnology, P.O. Box 1527, GR-71110 Heraklion, Crete, Greece, Institute of Biology, NCSR "Demokritos", 15310 Ag. Paraskevi, P.O. Box 60228, Athens, Greece, and Department of Biology, University of Rome Tor Vergata, Via della Ricerca Scientifica, 00133 Rome, Italy

ABSTRACT: The repressor of primer (Rop) protein has become a steady source of surprises concerning the relationship between the sequences and the structures of several of its mutants and variants. Here we add another piece to the puzzle of Rop by showing that an engineered deletion mutant of the protein (corresponding to a deletion of residues 30−34 of the wild-type protein and designed to restore the heptad periodicity at the turn region) results in a complete reorganization of the bundle which is converted from a homodimer to a homotetramer. In contrast (and as previously shown), a two-residue insertion, which also restores the heptad periodicity, is essentially identical with wild-type Rop. The new deletion mutant structure is a canonical, left-handed, all-antiparallel bundle with a completely different hydrophobic core and distinct surface properties. The structure agrees and qualitatively explains the results from functional, thermodynamic, and kinetic studies which indicated that this deletion mutant is a biologically inactive hyperstable homotetramer. Additional insight into the stability and dynamics of the mutant structure has been obtained from extensive molecular dynamics simulations in explicit water and with full treatment of electrostatics.

The wild-type repressor of primer (Rop) protein is a RNA-binding protein involved in the regulation of the copy number of the ColE1 plasmid (*1*). The structure of Rop has been studied extensively using both X-ray crystallography and NMR and serves as the paradigm of a canonical homodimeric all-antiparallel four-α-helix bundle (*2*, *3*). The apparent structural simplicity of its folding motif led several groups to believe that Rop could be used as a model system to investigate the sequence−structure relationships in the folding and dynamics of four-α-helix bundles in general (*4*, *5*). The result is a large body of knowledge comprising functional (*6*−*8*), thermodynamic (*9*−*15*), kinetic (*16*−*18*), and structural (*19*−*25*) investigations of many Rop mutants and variants. Unfortunately, the structural simplicity of Rop proved to be deceiving; if there is a consistent lesson to be learned from these sequence−structure studies, it is our consistent failure to understand the dependency of the Rop fold on its sequence. Figure 1 illustrates this point. Four very similar amino acid sequences lead to four completely different structures, with different bundle topologies, different hydrophobic cores, and different surface properties. Although all structures do remain four-α-helix structures, they are so different in every other respect that, at least at the atomic level, it would make more sense to note their similarities rather than their differences.

Here we present the 2 Å crystal structure of Δ_{30-34}, an engineered, non-naturally occurring deletion mutant of Rop (we will hereafter refer to this variant as RM6). The design principle for RM6 has its roots in the heptad sequence periodicity (*4*) characterizing associating α-helices; interacting helices usually display a periodicity of seven residues in their primary sequence, with the first and fourth residue of the repeat, designated as *a* and *d* positions, respectively, being hydrophobic (and, through their interaction, responsible for the formation of the bundle's hydrophobic core). As shown in Figure 1, the pattern of successive heptad repeats in the wild-type Rop sequence [indicated with the repeating pattern of underlined (with green) amino acids marking the *a* and *d* positions] is disrupted only once at the position of the turn region (centered on Ala$_{31}$). It has been hypothesized that it is exactly this deviation from the heptad periodicity that leads to the formation of the turn connecting the two helices in the wild-type Rop monomer. To test this hypothesis, we have designed (*6*) two variants of wild-type Rop, both of which lead to a restoration of the heptad repeat throughout the protein sequence. The first variant is a two-

---

* To whom correspondence should be addressed: Department of Biology, University of Crete, P.O. Box 2208, GR-71409 Heraklion, Crete, Greece. Telephone and fax: +302810394351. E-mail: kokkinid@imbb.forth.gr.
  ‡ Democritus University of Thrace.
  § These authors contributed equally to this work.
  ‖ University of Crete and Institute of Molecular Biology and Biotechnology.
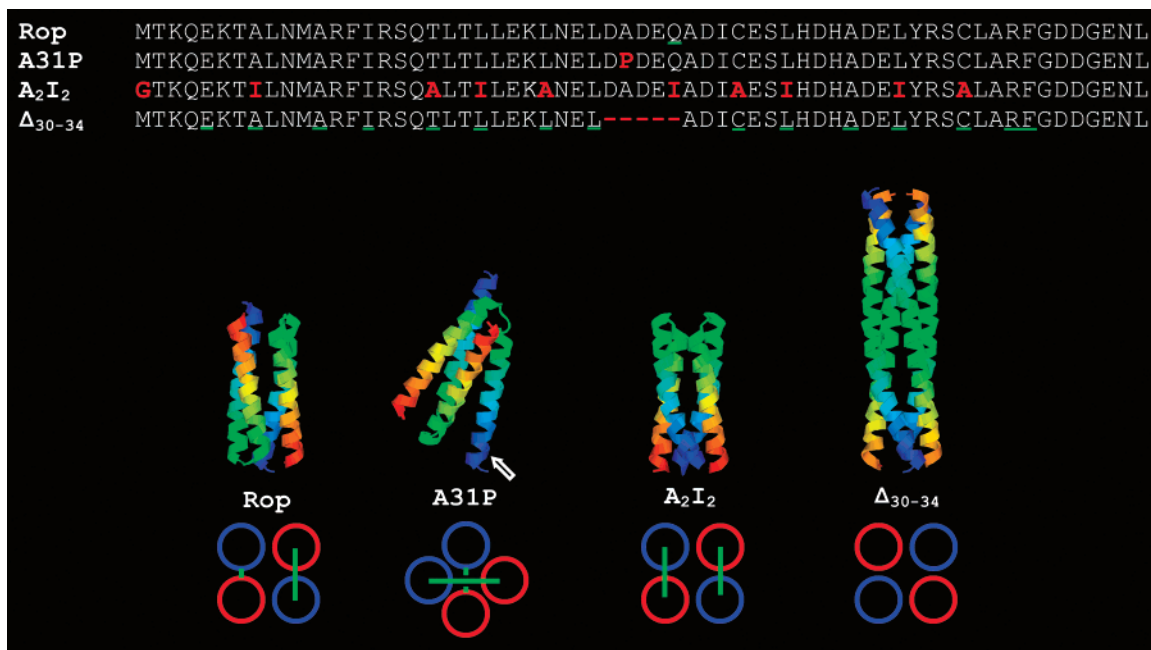 ⊥ NCSR "Demokritos".
  # University of Rome Tor Vergata.

FIGURE 1: Sequence–structure relationships in the Rop family. Comparison between the sequences, the structures, and the topological diagrams of wild-type Rop (PDB entry 1rop), the A31P mutant (PDB entry 1b6q), the (Ala$_2$Ile$_2$)$_6$ Rop variant (PDB entry 1f4n), and the RM6 deletion mutant described in this paper (PDB entry 1qx8). In the sequence alignment (top), differences between the sequences are highlighted using bold red characters. The underlined amino acids in the RM6 sequence (plus Gln$_{34}$ in the wild-type Rop sequence) correspond to residues occupying the *a* and *d* positions of the heptad repeat (see the text for details). Note how the RM6 deletion restores the heptad periodicity in the turn region. The schematic diagrams of the structures corresponding to these sequences (middle) are all on the same scale and have been oriented to align (in both orientation and position) the helix shown with an arrow in the A31P structure. The coloring scheme used for all structures is the same and corresponds to a color gradient for each monomer (polypeptide chain) ranging from blue for the N-terminus to red for the C-terminus. In the topological diagrams (bottom), each circle corresponds to a helix, with its color denoting the relative helix direction. Connecting loops are represented as green lines.

residue insertion, with two alanine residues inserted on either side of Asp$_{30}$ (this insertion mutant will be termed <2aa>). The second variant is RM6 which corresponds to the deletion of five residues (from Asp$_{30}$ up to and including Gln$_{34}$, DADEQ sequence; see Figure 1). Both sequences have an uninterrupted succession of heptad repeats, with <2aa> being seven residues longer than RM6. The high-resolution crystal structure determination of <2aa> (*24*) showed that restoration of the heptad pattern (through the aforementioned insertion) does not alter the fold of Rop. Here we show that restoring the heptad repeat through a deletion (i.e., the RM6 variant) results in a complete reorganization of the bundle structure.

RESULTS

Structure Description

Overall Fold and Topology. A summary of the structure determination procedure is given in Structure Determination and Analysis. The electron density maps shown in Figure 2 illustrate the quality of phase determination and of the final model. A schematic diagram of the structure is shown in the left panel of Figure 3 (with its topological diagram shown in Figure 1). RM6 is a homotetrameric, all-antiparallel, left-handed four-α-helix bundle. Its molecular surface (shown in the right panel of Figure 3) can be approximated by a slightly curved (see below) cylinder 78 Å in length and 28 Å in diameter. Each monomer (of the tetramer) corresponds to one Rop polypeptide chain and forms a continuous α-helix (instead of a helix–turn–helix hairpin as in wild-type Rop; see Figure 1). The helices twist around one another and form

a four-stranded, left-handed coiled coil. The homotetrameric state of the protein as determined from its crystal structure is in excellent agreement with previously published kinetic and thermodynamic data for this mutant (*17*). The crystallographically determined coordinates of the four monomers are not independent; the complete bundle is formed through the application of a crystallographic 2-fold symmetry axis. To simplify the discussion that follows, we will use different names for the four individual helices. Chains A and B (green and yellow helices in Figure 3, respectively) correspond to the contents of the crystallographic asymmetric unit and are chosen in such a way that they are on the same side of the intramolecular dyad axis. Their symmetry-related monomers will be denoted as chains C and D (with the A–C and B–D pairs related through the intramolecular symmetry axis). It is worth noting here that, at least in principle, it would have been possible for this homotetrameric protein to have a higher internal (molecular) symmetry with three exact and intersecting orthonormal 2-fold axes (point group 222) instead of just one exact 2-fold axis (point group 2). The fact that this is not observed agrees with the notion (also discussed below) that the RM6 structure may better be described as a "dimer of dimers".

Coiled-Coil and Helical Parameters. Figure 4 shows some pertinent parameters of the RM6 coiled coil (see Structure Determination and Analysis for a description of how these parameters were calculated). The bundle as a whole is slightly curved as can be inferred from the deviation of the local bending angles of the coiled-coil axis from the value of 180° (Figure 4A). Note also that the bending angles show a consistent trend, with higher deviations from linearity
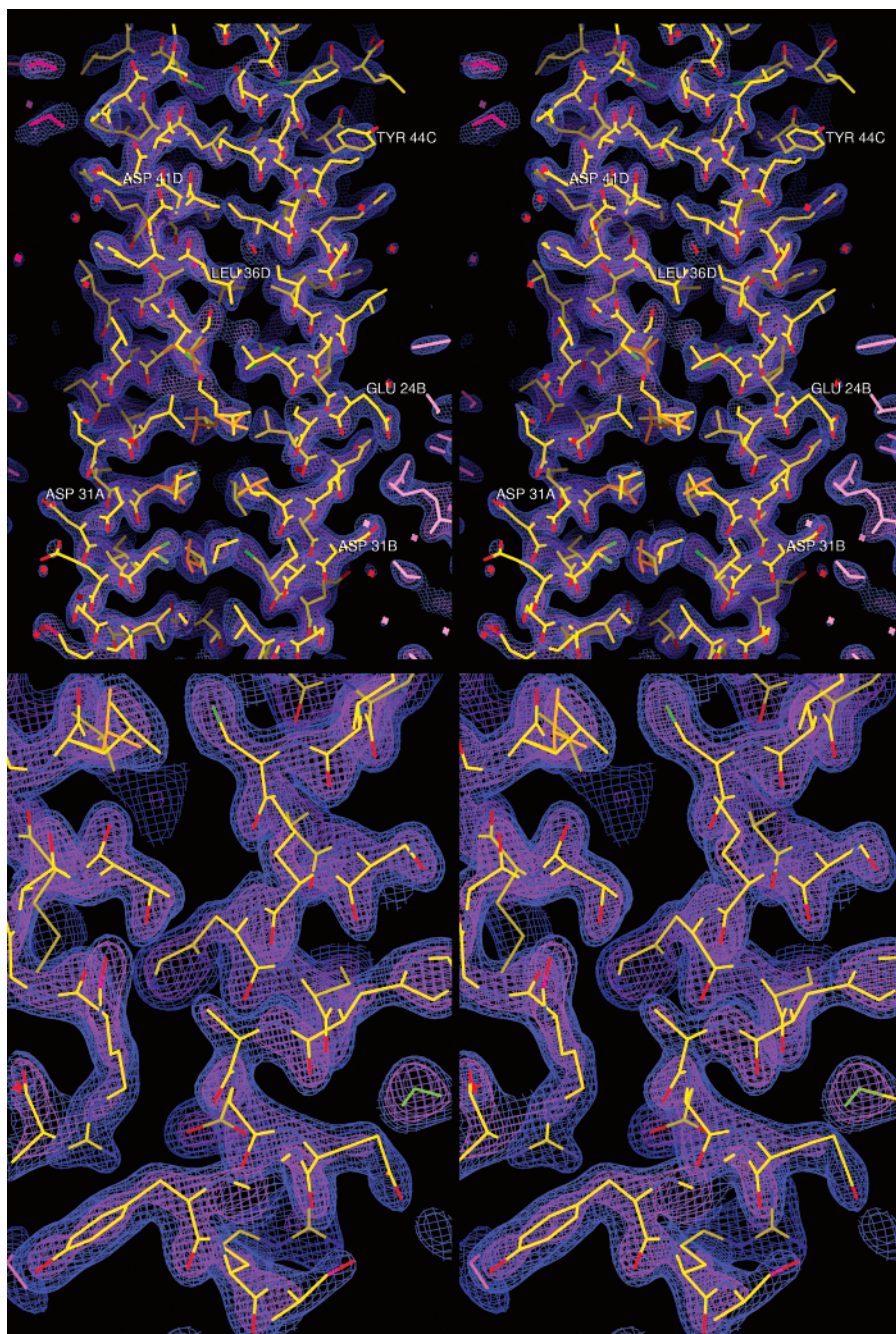
FIGURE 2: Stereodiagrams (wall-eyed) of the electron density distribution with the final model superimposed. The top panel shows a large volume containing portions of both helices contained in the crystallographic asymmetric unit. Residues Glu_{24}, Asp_{31}, Leu_{36}, Asp_{41}, and Tyr_{44} are labeled. The map is contoured using two isosurfaces 1.5σ and 3σ above mean, and all density features in the volume that is shown are drawn. The bottom panel illustrates the quality of phase determination by showing a detailed view of the distribution of electron density around part of the structure containing residues Tyr_{44} (bottom left corner) and His_{37} (center). The electron density in both panels corresponds to an approximation to the maximum-entropy estimate [calculated with *GraphEnt* (*50*)] of a σ_A-weighted difference map of the form $2mF_o - DF_c \exp(i\phi_c)$. All data between infinity and 2.02 Å (not counting the maximum-entropy extrapolated data) have been used for the calculation.

toward the protein's edges. As can be seen in Figure 3, individual helices are highly curved, twisting about the bundle axis. What is not immediately evident from the structure schematics is that this twist is not uniform across the bundle's length. As Figure 4B shows, the bundle's edges are significantly more twisted than its middle with an approximately linear relationship between the twist and the distance from the center of the protein. This observation concurs not only with the behavior of the local bending angles of the coil axis (Figure 4A) but also with the values of the helix−helix crossing angles as shown in Figure 4D.

These, again, start from a value of around 30°, decrease almost linearly to a value of ~10° at the middle of the protein, and then increase as we move toward the other end of the bundle. The pattern of helix−helix distances shown in Figure 4C indicates that although the protein is homotetrameric, the contacts between neighboring helices are not equivalent, suggesting that the RM6 structure may better be described as a dimer of dimers. Not only is the average distance between the A−B (or, equivalently, C−D) helices significantly smaller than that for the A−C (or, B−D) pairs, but the variation of their values along the bundle's length is
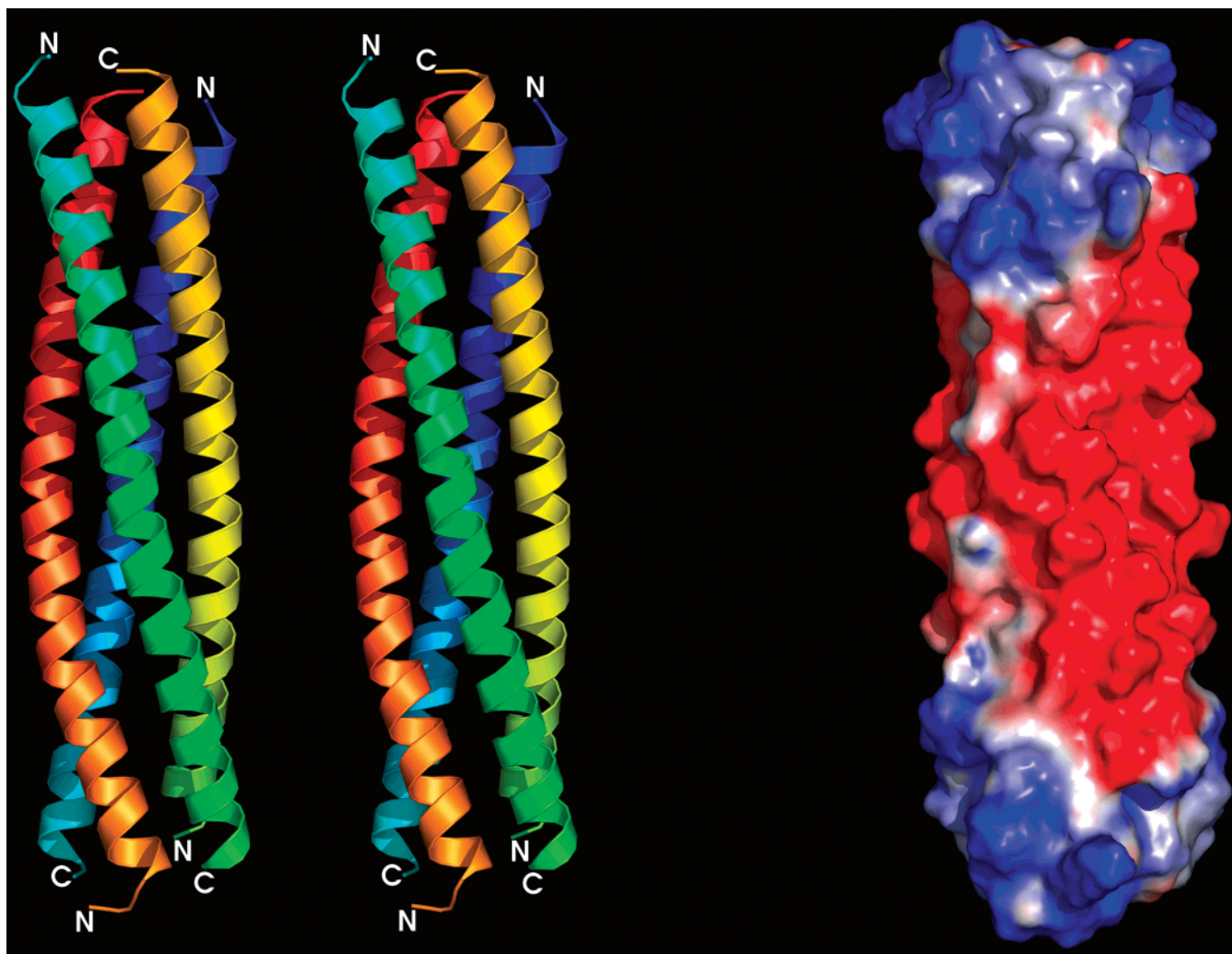
FIGURE 3: Overall structure and surface properties of RM6. The left panel is a stereodiagram (wall-eyed) of a schematic (cartoon) representation of the RM6 crystal structure. A different color is used for each monomer (with their N- and C-termini marked). Note that the structure shown is the complete bundle which is formed through the application of a crystallographic (exact) 2-fold axis. This intramolecular 2-fold axis relates the two monomers in the crystallographic asymmetric unit (green and yellow in this figure, called chains A and B in the text, respectively) with the other two (red and blue) and is approximately horizontal and passing through the center of the bundle. The right panel is a molecular surface representation of the structure (in the same orientation as the left panel) color-coded according to the electrostatic potential from −5 kT/e (red) to 5 kT/e (blue). Note the highly asymmetric charge distribution. The structure used for calculating the electrostatic potential had all missing side chains and hydrogen atoms built using PSFGEN (*44*).

systematically different with the A−B and C−D pairs following closely a sinusoidal curve with a period of half the bundle's length (Figure 4C). This behavior may be related to differences in the composition (and size) of the buried side chains at the corresponding hydrophobic core layers (see Hydrophobic Core and Surface Properties).

Turning our attention to individual helices, we note that their parameters are essentially identical to those expected from an undistorted α-helix. Averaging over all residues of the structure gives a mean of 3.62 residues per turn and a mean rise per residue of 1.51 Å. As panels A and B of Figure 5 show, both the number of residues per turn and the rise per residue remain fairly constant across the length of the helices with the larger deviations occurring at the highly mobile (and mostly nonhelical) termini. It is worth noting that individual helices are not uniformly curved. This can be seen in Figure 5C which shows a grayscale representation of a matrix whose elements are the angles between all possible pairs of segments from the local helix axis (*41*). The matrix indicates the presence of four relatively straight

helical segments in each helix, namely, residues 5−15, 16−30, 31−38, and 39−48. The limits of these segments are indicated by the colored bars drawn below Figure 5C. Note that there is a difference between the A and C helices and between the B and D helices concerning the limits of the two last segments: for the A−C helices the segments are residues 31−37 and 38−48, but for the B−D pair, their limits are residues 31−39 and 40−48. On the basis of this matrix alone, it would be difficult to decide whether these segments do indeed behave as relatively rigid bodies in solution and/ or whether some other classification scheme would have been more appropriate. For example, it could be argued that the matrix indicates the presence of three (residues 5−15, 16−30, and 31−48) or even only two segments (residues 5−30 and 31−48). As will be discussed in Molecular Dynamics Simulations, molecular dynamics simulations of RM6 indicate the presence of two relatively rigid segments (residues 5−30 and 31−48). It is worth noting here that the limit between these two segments coincides with the RM6 deletion point.

(A) Bundle axis bending angles

(B) Coiled−coil twist per residue

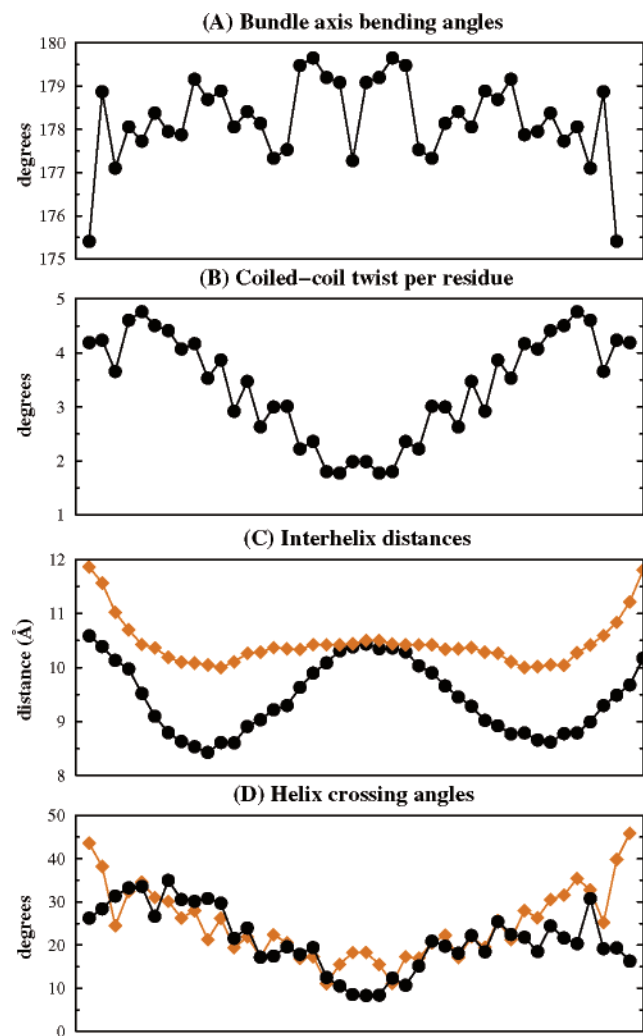**(C) Interhelix distances**

(D) Helix crossing angles

FIGURE 4: Coiled-coil parameters. The horizontal axes in all diagrams correspond to the (per residue) position along the whole bundle or individual helices. (A) Bending angles (in degrees) between successive triplets of points representing the local coiled-coil axis. Note the systematic deviation from 180° indicating that the bundle is slightly curved. (B) Twist (per residue) of the coiled coil. The reported value is in degrees, and the rotation is calculated with respect to the direction of the local coiled-coil axis. (C) Local helix−helix distances (per residue and in angstroms). The top curve (orange) corresponds to the distances between neighboring helices related through the crystallographic 2-fold axis (A−C and B−D pairs). The bottom curve shows the distances between the two non-symmetry-related helices (A−B and C−D pairs). (D) Local helix−helix crossing angles (per residue and in degrees). The orange curve corresponds to the angles between neighboring helices related through the crystallographic 2-fold axis (A−C and B−D pairs). The black curve shows the angles between the two non-symmetry-related helices (A−B and C−D pairs).

(A) Residues per helical turn

(B) Rise per residue
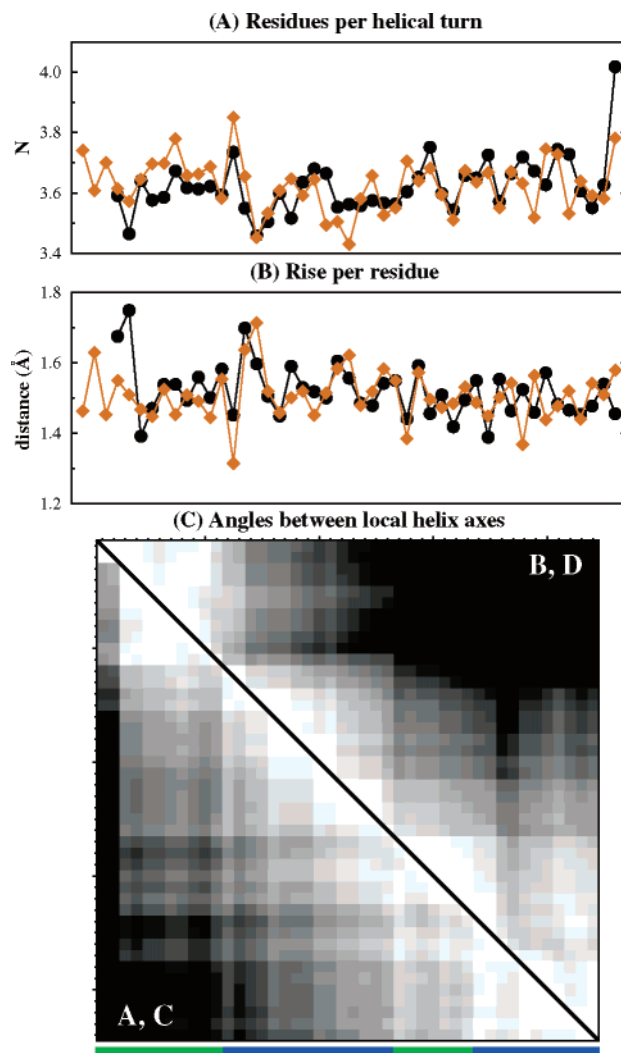
(C) Angles between local helix axes

FIGURE 5: Helical parameters. The horizontal axes in the top two diagrams correspond to residue numbers (of the respective helices). In the bottom diagram, both the horizontal and the vertical axes correspond to successive residues of the helices. (A) Number of residues per helix turn. The orange curve corresponds to helix B (or, equivalently, D) and the black curve to helix A (or C). (B) Rise (in angstroms) per residue. The color coding is the same as that for panel A. (C) Grayscale representation of a matrix whose elements (i and j) are equal to the angle (in degrees) between the local helix axis segments corresponding to residues i and j of the helix. The grayscale representation is such that an angle of 0° corresponds to white and any value greater than or equal to 25° is colored black. The top half (above the diagonal) shows the angles for helices B and D and the bottom half those for helices A and C. The colored bars below the matrix indicate the limits of relatively straight helical segments (see the text for details). The origin of the matrix is in the top left corner.

Hydrophobic Core and Surface Properties. As mentioned in the introductory section, associating α-helices display a heptad sequence periodicity with the first and fourth position of the repeat (denoted as positions a and d, respectively) being usually hydrophobic. The residues corresponding to these (a and d) positions are responsible for forming the protein's hydrophobic core. In wild-type Rop, the a and d residues from all four helices (underlined in Figure 1) interact at the protein's core in quartets: one a and one d residue from the first monomer interact with a d and an a residue from the second monomer. The result is that the hydrophobic core of native Rop (2) can be visualized as a stack of eight

hydrophobic layers, with each layer comprising two a and two d residues. Due to the presence of an intramolecular 2-fold axis, these eight hydrophobic layers are symmetry-related in pairs (the fourth layer is symmetry-related with the fifth, the third layer with the sixth, etc.). The composition of the layers, however, is not symmetric: when one moves from the center of the bundle toward its end, the four unique wild-type Rop layers consist of Ala_{45}−Leu_{41}−Thr_{19}−Ile_{15}, Ala_{12}−Leu_{22}−Cys_{38}−Leu_{48}, Cys_{52}−Gln_{34}−Leu_{26}−Ala_8, and Glu_5−Leu_{29}−Ala_{31}−Phe_{56} residues (in the order *adad*).

The RM6 hydrophobic core retains the organization in hydrophobic layers comprising *adad* residues, but as shown
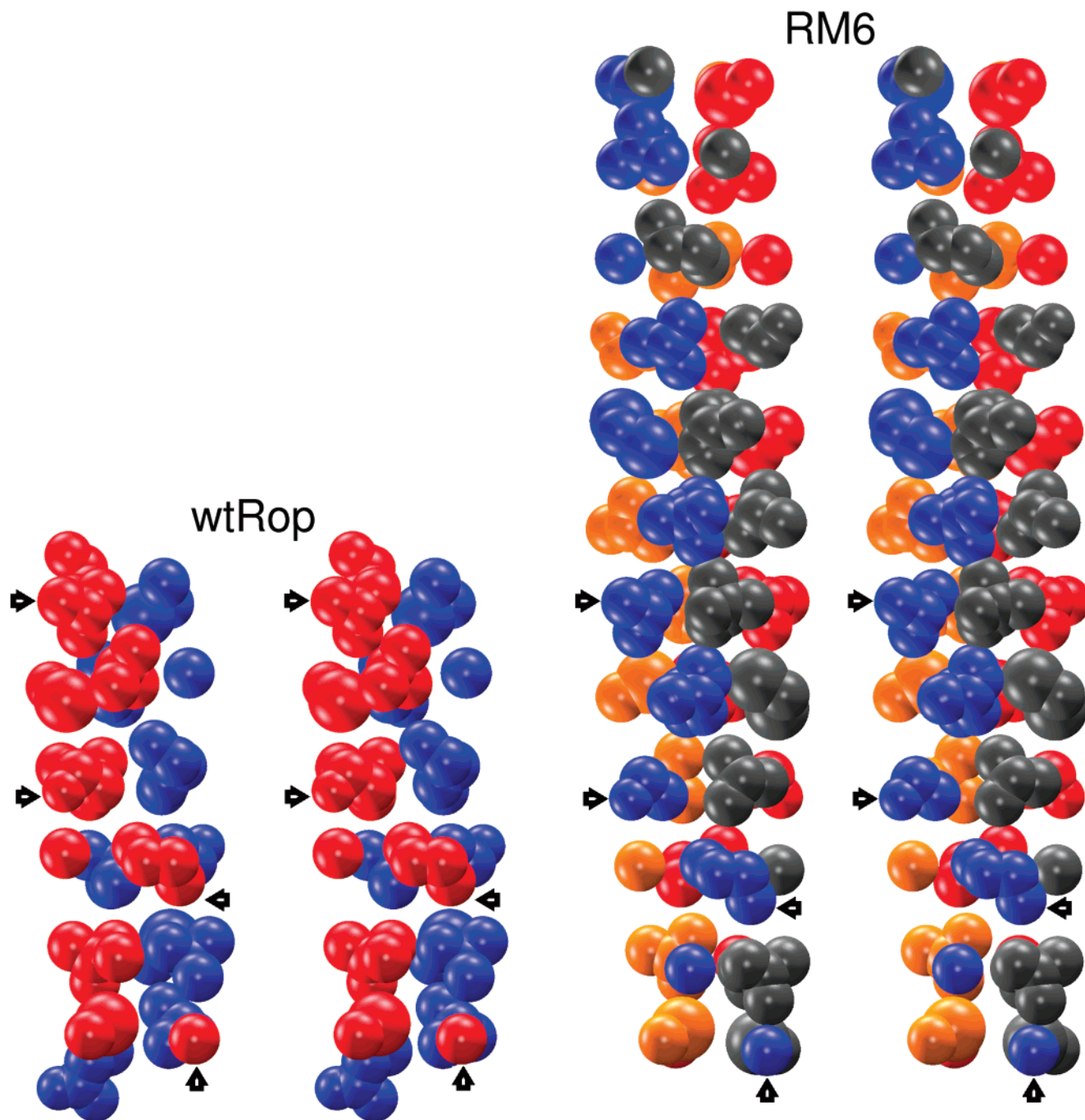
RM6

wtRop

FIGURE 6: Hydrophobic core comparison. Stereodiagrams (wall-eyed) of the side chains comprising the hydrophobic cores of wild-type Rop (left) and RM6 (right). Side chains are shown as space-filling models using their corresponding van der Waals radii and colored according to the polypeptide chain to which they belong (two chains for wild-type Rop and four for RM6). To aid the comparison between the two proteins, the side chains of residues 5−27 of one of their helices have been aligned in both position and orientation. The arrows point to equivalent sets of some of these side chains in the two proteins.

in Figure 6, this is as far as the similarities with wild-type Rop go. RM6 comprises 14 layers related in pairs through the application of the intramolecular 2-fold axis. The composition of the RM6 layers is internally symmetric with not a single layer sharing the same residues with native Rop. The RM6 layers are as follows (from the center of the bundle toward its edge and in the order *adad*): $Leu_{26}−Leu_{29}−Leu_{26}−Leu_{29}$, $Cys_{33}−Leu_{22}−Cys_{33}−Leu_{22}$, $Thr_{19}−Leu_{36}−Thr_{19}−Leu_{36}$, $Ala_{40}−Ile_{15}−Ala_{40}−Ile_{15}$, $Ala_{12}−Leu_{43}−Ala_{12}−Leu_{43}$, $Cys_{47}−Ala_8−Cys_{47}−Ala_8$, and, finally, $Glu_5−Phe_{51}−Arg_{50}−Arg_{50}$. Note that the last layer should not have been named as such, given that it is open, asymmetric, and

penetrated by water and its residues are highly mobile. As can be seen from Figure 6, the central layers of the RM6 hydrophobic core are rich in bulky side chains which causes a noticeable bulge at the middle of the protein. This asymmetric distribution of the large side chains may be correlated with the sinusoidal variation in the interhelix distances (Figure 4C) and the increased values of the coiled-coil twist near the bundle's edges (Figure 4C).

With such a drastically repacked hydrophobic core, it would be considered probable that some buried cavities may have been formed in RM6. In reality, the inverse happened: whereas wild-type Rop contains two symmetry-related cavi-

ties of volumes equal to approximately 26 Å³ (for a probe with a 1.4 Å radius), RM6 contains no buried cavities at all. This, together with the extent of the buried accessible area (discussed below) and the presence of eight interhelix salt bridges, may explain the apparent thermostability of this variant (*17*) (see also the Discussion). The total accessible surface area of RM6 (ignoring hydrogens and missing side chains) for a 1.4 Å probe is 10 439 Å³. Repeating the calculation for the four isolated chains gives an average total accessible surface area per monomer equal to 4930 Å³ (5010 and 4850 Å³ for A−C and B−D chains, respectively). The average buried accessible surface area of a monomer upon tetramer formation is thus 2320 Å³ (9281 Å³ for all four monomers) which gives an estimated solvation free energy gain (upon tetramer formation) of approximately −110 kcal/mol (*26*). Finally, and as shown in Figure 3, the distribution of the electrostatic potential of RM6 is highly polarized: the middle of the bundle is negatively charged, whereas its edges are prominently positive. This distribution of electrostatic potential is totally different from the one seen in wild-type Rop which is characterized by the presence of two distinct (front−back) surfaces, one negatively and the other positively charged [and used for RNA binding (*8*)].

Molecular Dynamics Simulations

Stability and Consistency. To gain an understanding of the equilibrium dynamics of RM6, we performed an extensive set of molecular dynamics simulations using periodic boundary conditions with explicit representation of water and a full treatment of the electrostatics. We followed the dynamics of six independent trajectories amounting to a total simulation time of 54.4 ns (see Molecular Dynamics Simulations for a detailed description of the simulation protocol).

Probably the most consistent finding from these simulations is the stability of the RM6 structure even at the relatively high temperature of 320 K (47 °C) that was used for all of our calculations. Figure 7 illustrates exactly this point by showing representative results obtained from the longest continuous trajectory (simulation *Hex2*): the rms deviation from the starting (experimentally determined) structure (Figure 7A) remains below 1 Å for almost the whole length of the trajectory, with an average deviation of 0.92 Å. Calculation of the rms deviation between the average structure (from the simulation) and each frame from the trajectory (Figure 7B) shows that the simulation converges to a stable average structure with a mean rms deviation (from this average structure) of 0.71 Å. Similarly, and as shown in Figure 7C, the value of the radius of gyration remains essentially identical with its starting value with a mean (over the length of the trajectory) of 19.84 Å and a standard deviation of only 0.08 Å. Note that the highly mobile terminal residues of the helices have been excluded from these calculations (see the following section).

As a further indication of the consistency between the experimental and simulation-derived structures, Table 1 compares the rms deviations and the average displacements between the crystal structure and the average structures obtained from the six trajectories. The mean value of the rms deviation between the crystal structure and the simulation-derived structures is only 0.58 Å. This is not very different from the expected coordinate error (approximately 0.3 Å) of the experimentally determined structure. Note also
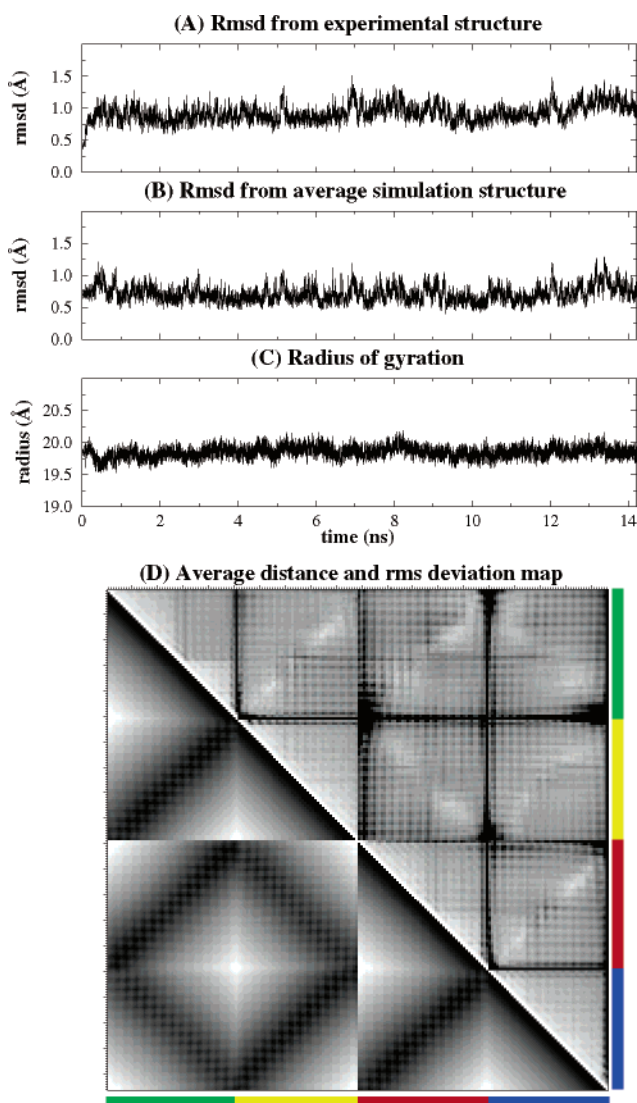

FIGURE 7: Stability of molecular dynamics simulations. The *Hex2* simulation was used for all calculations shown in this figure. For panels A−C, the flexible terminal residues have been excluded from the calculations. (A) Evolution of the rms deviation between the starting (crystal) structure and each of the structures recorded during the simulation. (B) Evolution of the rms deviation between the simulation-derived average structure (calculated from the *Hex2* trajectory) and each of the structures recorded during this same simulation. (C) Evolution of the value of the radius of gyration during the *Hex2* simulation. (D) Grayscale representation of the average $C_\alpha-C_\alpha$ distance map and the corresponding rms deviation from it. The half of this figure below the diagonal shows the average $C_\alpha-C_\alpha$ distances in the range from 0 Å (corresponding to black) to 78 Å (corresponding to white). The top half of the figure shows the rms deviations (from these average $C_\alpha-C_\alpha$ distances) during the length of the simulation. For the rmsd map, the contrast has been reversed and ranges from 0 Å (white) to 1 Å (black). The origin of the matrix is in the top left corner. The limits (on these maps) of the individual helices are indicated with the horizontal and vertical color bars (chain A corresponds to the green bar, chain B to yellow, chain C to red, and chain D to blue).

that some of the simulation-derived structures are closer to the crystal structure than to other simulations. Indeed, a dendrogram which was calculated by treating the rms deviations as distances (in a distance matrix) indicated the presence of three clusters of structures (at a rmsd cutoff of 0.59 Å): the first cluster comprised the experimental structure plus the *Hex1*-derived structure, a second cluster

Table 1: Comparison between the Experimentally Determined RM6 Structure and the Trajectory-Averaged Structures Obtained from Six Molecular Dynamics Simulations[a]

	crystal	*Hex1*	*Hex2*	*Hex3*	*LHex*	*Oct1*	*Oct2*
crystal	—	0.582	0.591	0.758	0.503	0.596	0.468
Hex1	0.517	—	0.653	0.908	0.572	0.741	0.598
Hex2	0.531	0.601	—	0.462	0.363	0.322	0.440
Hex3	0.681	0.831	0.377	—	0.590	0.431	0.600
LHex	0.463	0.512	0.328	0.524	—	0.419	0.354
Oct1	0.541	0.676	0.266	0.337	0.376	—	0.349
Oct2	0.433	0.534	0.367	0.492	0.304	0.307	—

[a] The experimental structure is denoted as crystal. All other columns correspond to the six molecular dynamics simulations that were performed (see Molecular Dynamics Simulations for details). The top half of the table shows the rms deviation (in angstroms) between the coordinates of the C_α atoms of residues 7−49 of the respective structures. The bottom half shows the corresponding average displacements (again in angstroms).

(most distant to all other) contained only the *Hex3* structure, whereas the third cluster contained the structures from simulations *Hex2*, *Oct1*, *Oct2*, and *LHex*. It is worth noting that even the simulation which appears to be the most distant from all others (i.e., *Hex3*) gives results that are fully consistent with those obtained from the other simulations [compare, for example, the upper half (obtained from *Hex2*) with the lower half (obtained from *Hex3*) of the cross-correlation matrix shown in Figure 9 (the linear correlation coefficient between the *Hex2*- and *Hex3*-derived values is 0.902)].

Mobility and Rigidity. Turning our attention to the dynamics of RM6, we start by examining the variability of the structure's C_α−C_α distances during the length of the simulations. A compact representation of these results is shown in Figure 7D. The lower (below the diagonal) half of this figure is a grayscale representation of the *Hex2* trajectory-average C_α−C_α distance map. The upper half of the same figure is the map corresponding to the rms deviations, during the length of the trajectory, from these average C_α−C_α distances. The average distance map has the characteristic appearance expected from an antiparallel four-α-helix bundle (with dark areas corresponding to short interatomic vectors and the inverse). The rmsd map, on the other hand, contains information about the relative stability of the protein (on a per residue basis) and clearly differentiates between the relatively rigid parts of the structure (white and light gray areas) and the highly mobile ones (black and dark gray areas). Its most pronounced feature is the set of black lines running horizontally and vertically, creating a checkerboard appearance. These highly mobile segments correspond to the residues belonging to the termini of the individual helices (residues 5, 6, 50, and 51 for helices A−C and residues 1−6, 50, and 51 for helices B−D). These residues do not participate in the formation of the bundle's hydrophobic core (see Hydrophobic Core and Surface Properties) and appear to be mostly disordered. The mobility of these residues (as indicated by the simulations) is in very good agreement with their crystallographically determined temperature factors (see below). With the exception of these tails, the RM6 structure appears to be well-preserved during the simulation. Nevertheless, some trends related to the dynamics of the bundle can be discerned directly from the rmsd map. The best preserved substructures correspond to the individual helices, whereas the helix−helix distances show higher variability,
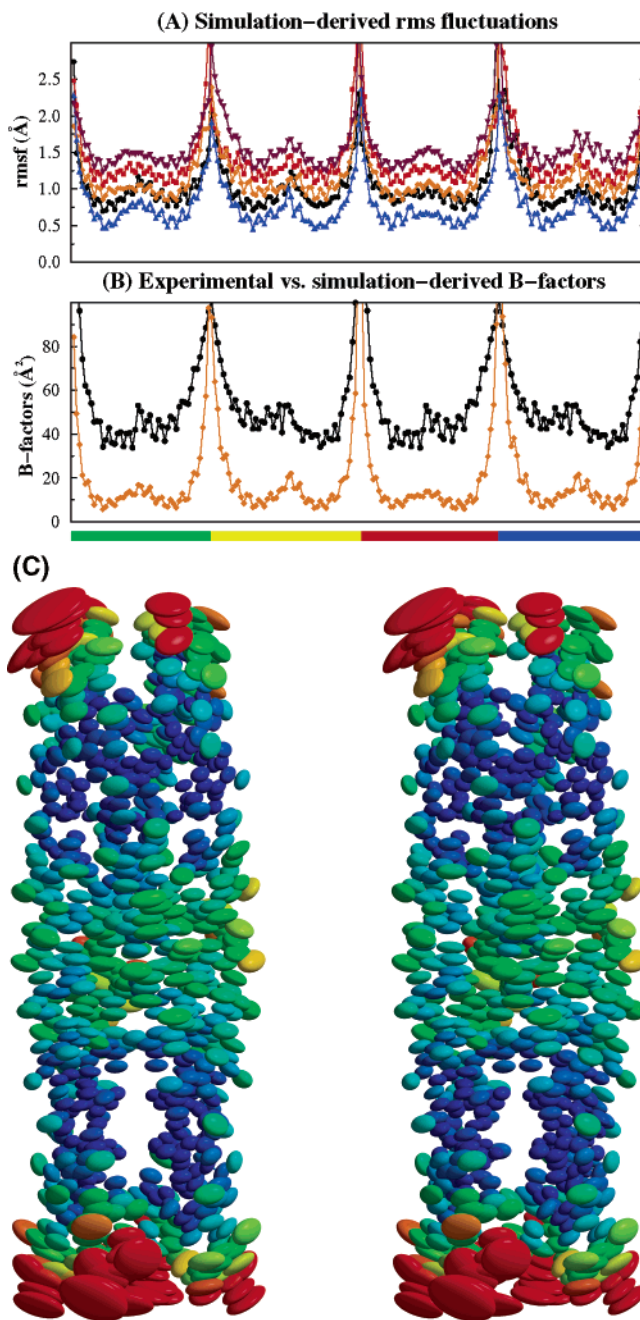

FIGURE 8: (A) Root-mean-square fluctuations of the C_α atoms for five independent trajectories. To aid comparison between the curves, four of them have been translated along the vertical axis by 0.25, 0.50, 0.75, and 1.0 unit, respectively. (B) Comparison between the experimental (top black curve) and average simulation-derived (bottom orange curve) atomic temperature factors of the C_α atoms. The limits (on these graphs) of the individual helices are indicated with the horizontal colored bars (chain A corresponds to the green bar, chain B to yellow, chain C to red, and chain D to blue). (C) Stereodiagram (wall-eyed) representation of the average structure of RM6 (from the *Hex2* trajectory) using the thermal anisotropic ellipsoids calculated from the motion of the backbone and C_β atoms during the simulation. The ellipsoids' isosurfaces are drawn at the 90% probability level and are colored according to the value of the equivalent isotropic temperature factor from 8 Å^2 (or less) corresponding to dark blue to 30 Å^2 (or more) corresponding to red. This figure was prepared with Raster3D (*51*).

especially for those helices that are lying on opposite sides of the intramolecular 2-fold axis (this agrees with the longer distances between these helices; see Coiled-Coil and Helical Parameters). It should also be noted that individual helices
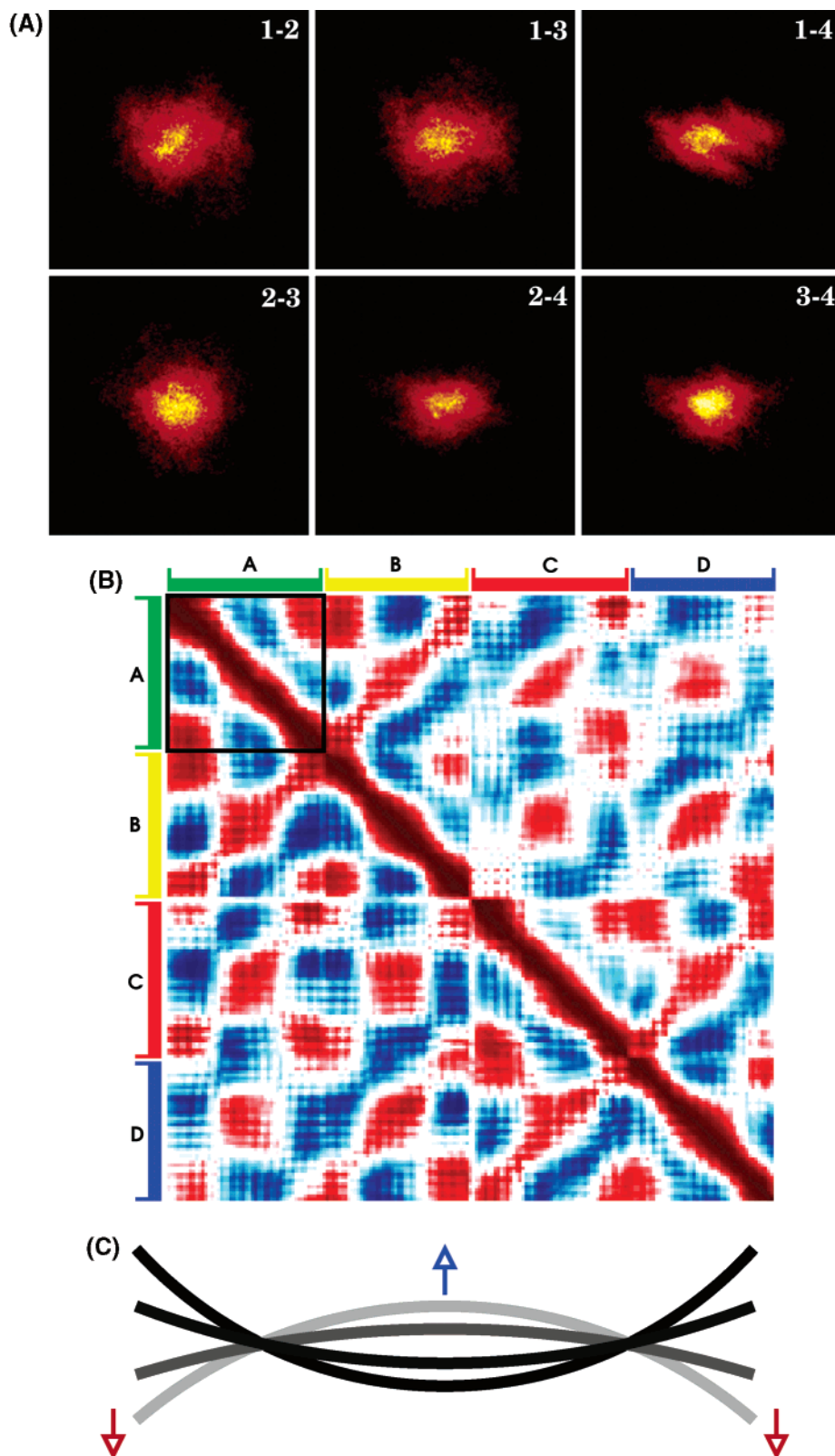
FIGURE 9: Sufficient sampling and cross-correlation matrices. (A) Pseudocolor representation of the density function corresponding to the projections of the fluctuations of the C_α motion (during the *Hex2* simulation) on the planes of the four eigenvectors associated with the four largest eigenvalues. The 1−2, 1−3, 1−4, 2−3, 2−4, and 3−4 eigenvector planes are shown. In all six diagrams, the values on all axes range from −20 to 20 Å. (B) Cross-correlation matrix calculated from the motion of the C_α atoms during the *Hex2* (above the diagonal) and *Hex3* (below the diagonal) simulations. The matrix values range from 1 (for fully correlated motions, depicted as dark red), through zero (uncorrelated, white), to −1 (fully anticorrelated, dark blue). The contrast of the cross-correlation matrix has been artificially increased through the application of a sigmoidal function. The limits of individual monomers are indicated by the marked bars shown above and on the left side of the matrix. The part of the matrix corresponding to the cross-correlations between the atoms of helix A is boxed. (C) Simplified schematic diagram of the type of motion of a helix that would result in the pattern of positive and negative correlations seen in the cross-correlation matrix. The curved lines correspond to successive snapshots of the helix axis motion (from black to light gray and the reverse). The arrows indicate the presence of correlated (red) and anticorrelated (blue) motion between different parts of the helix.

appear to behave not as a single rigid body but as two relatively rigid segments connected by a more flexible connection (this is seen most clearly in the part of the rmsd map that corresponds to vectors between the C_α atoms of helix A).

The indication that individual helices may behave as two rigid segments connected through a more flexible (but still helical) region is further supported by the distribution of the rms fluctuations of the C_α atoms during the length of the simulations as shown in Figure 8A for five trajectories. All simulations show the same trend with highly mobile termini and a higher-than-average mobility around the center of the bundle. In Figure 8B, we compare the crystallographically determined atomic temperature factors with those calculated from the molecular dynamics simulations. The experimentally determined and simulation-derived *B*-factors show exactly the same trend as before and are in very good agreement with each other with a linear correlation coefficient of 0.83. Still, and in terms of absolute values, the crystallographic temperature factors (at least at the core of the structure) are approximately 3 times higher than those derived by the simulation. We believe that this apparent disparity has its roots not in a simulation-specific deficiency but in the properties of the RM6 crystals which show very strong diffuse scattering even at relatively high angles. Finally, we note the presence of systematic differences between the simulation-derived temperature factors of the A−C pair of chains as opposed to the B−D pair (most easily seen from the higher average mobility near the center of the B−D helices). To quantify this observation, we calculated the value of the linear correlation coefficient between the simulation-derived *B*-factors of all possible pairs of the four chains. The A−C and B−D pairs gave a value of 0.98, significantly higher than the value of 0.82 obtained for the A−B and C−D pairs. The symmetry characterizing these pairs of helices is present not only in their average mobility but also in their average structures (data not shown). The fact that the symmetry between the A−C and B−D pairs is preserved even during the molecular dynamics simulations (and in the absence of crystallographic symmetry) further supports the view of RM6 as a dimer of dimers.

The previous discussion clearly showed the significant differences in the average mobility of the RM6 residues but gave no indication of whether the atomic fluctuations are isotropic or anisotropic. The availability of a trajectory of the atomic motions allows the calculation of anisotropic ellipsoids describing the trajectory-average probability density distribution of the atomic positions, very similar to the crystallographic anisotropic thermal ellipsoids. A representation of the RM6 structure using these anisotropic ellipsoids is shown in the form of a stereo diagram in Figure 8C (and colored according to *B*-factor values). This diagram explicitly reinforces a new view of the RM6 structure in terms of its dynamics: the protein can be viewed as two relatively stable four-α-helix bundles of half the RM6 length, joined end to end along their long axes, and with their connection coinciding with the RM6 deletion point. This view of the dynamics of the RM6 structure is further supported by (i) an analysis of the protein's correlated motions as observed in its molecular dynamics simulations and discussed in the next section and (ii) a reduction of the crystallographic *R* and R_{free} values by 0.5% when the crystallographic TLS

refinement (*27*) was repeated using two TLS bodies per helix [instead of the one TLS body per helix used during the original crystal structure determination procedure (Structure Determination and Analysis)].

Correlated Motions and Their Principal Components. Evident from Figure 8C is the fact that the directions and lengths of the major axes of the anisotropic ellipsoids show highly correlated preferences, most easily seen by their tangential alignment around the bundle's circumference. These localized preferences are suggestive of the presence of correlated protein motions. To analyze these, we performed a principal component analysis of the motion of the C_α atoms during the length of the molecular dynamics trajectories. We will focus our analysis on the *Hex2* simulation, the results from which are comparable with (and representative of) the results obtained from all other RM6 simulations. An important prerequisite for the validity of the (tentative) assignment of the principal components to protein motion modes is that sufficient sampling has been attained. The trajectories must be of adequate length to guarantee that all accessible protein conformations (for the given simulation's equilibrium conditions) have been sufficiently sampled. An indirect indicator of sufficient sampling is the distribution of the projection of the atomic fluctuations on the eigenvector planes corresponding to the largest eigenvalues. The diagrams in Figure 9A show these projections for all unique combinations the four largest eigenvectors of the *Hex2* simulation. Their closeness to two-dimensional Gaussians centered at the origin indicates that correlated protein motions have been sufficiently sampled.

Principal component analysis is based on the diagonalization of the variance−covariance matrix of the (x,y,z) components of the atomic fluctuations (*28*, *29*). A closely related matrix that is far easier to visualize is the dot product-based cross-correlation matrix (normalized variance−covariance matrix) shown in Figure 9B. Some aspects of this matrix are worth discussing. The strong diagonal band corresponds to the fully correlated motion between neighboring C_α atoms. There is also significant correlation between the motion of the neighboring antiparallel pairs of helices (A−B and C−D) creating the X-like patterns in the top left and bottom right quadrants. The correlation between helices lying on opposite sides of the intramolecular 2-fold axis (i.e., A−C, A−D, B−C, and B−D pairs) is significantly lower and inversely proportional to the distance between the corresponding helices. Running parallel to the bands of positive correlation are bands of negative matrix values indicating the presence of anticorrelated motion between parts of the RM6 structure. The interpretation of this anticorrelated motion is best understood in terms of the cross-correlation pattern of a single helix. Taking, for example, the part of the matrix that corresponds to helix A (marked in Figure 9B), we observe a tripartite organization: the two ends of the helix show highly correlated motion, whereas the center of the helix appears to move in a manner anticorrelated to its ends. A schematic interpretation of this pattern of movement is shown in Figure 9C: if the helix comprises two relatively rigid parts connected by a more flexible connection and it oscillates in the form shown in Figure 9C, then we will indeed observe a positive correlation between the motion of its ends and an anticorrelated movement of its central part. This interpretation of the cross-correlation
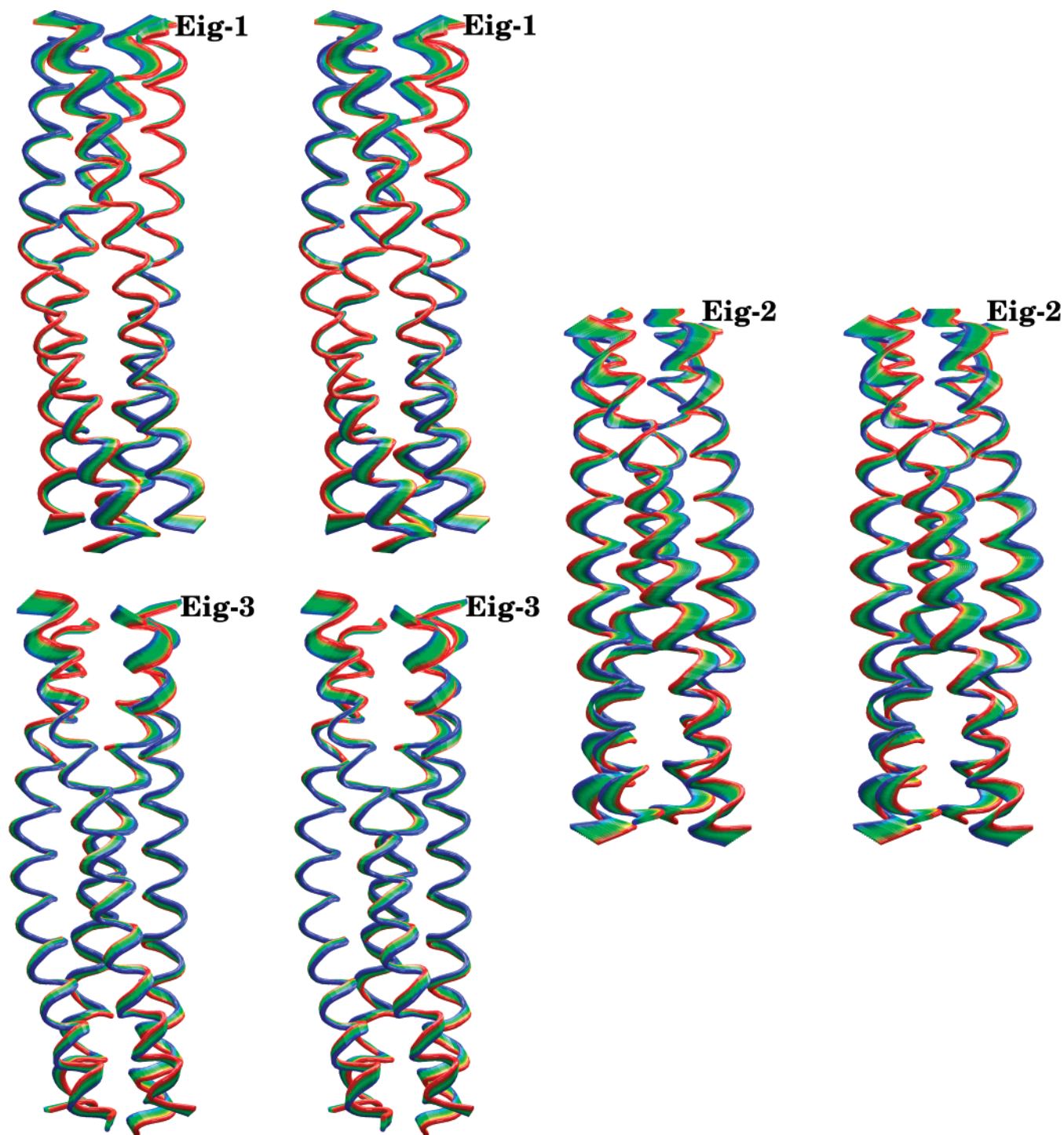
FIGURE 10: Principal component analysis. The three stereodiagrams (wall-eyed) depict the C_α motion corresponding to the three eigenvectors associated with the three largest eigenvalues. The movement of the chains is indicated through the superposition of 17 structures (for each diagram) which cover the whole range of conformations associated with each component. These 17 structures are colored from blue, through green, to red, corresponding to successive conformations along the given eigenvector motion.

matrix is in good agreement with the view of the RM6 dynamics presented in the previous section.

Although the hypothetical model presented in Figure 9C can qualitatively explain some of the features observed in the cross-correlation matrix, it does make the erroneous assumption that RM6 dynamics can be described by just one type of correlated motion. A quantitative decomposition of the RM6 correlated motion (as observed in the molecular dynamics trajectories) based on principal component analysis is described below.

If the flexible terminal residues are ignored, RM6 comprises 172 residues. With three (x,y,z) components per C_α atom, the variance−covariance matrix is a 516×516 real square symmetric matrix. Its diagonalization results in 516 eigenvectors (principal components) and their associated 516 eigenvalues (which are related to the amplitude with which the corresponding eigenvectors contribute to the correlated molecular motions). As is usually the case, we observe a markedly asymmetric distribution of eigenvalues with the three largest (corresponding to only 0.58% of the total

number of principal components) accounting for 41% of the total correlated motion [with 66 eigenvalues (12.8% of the total) required for explaining 90% of the correlated motion]. A schematic representation of the protein motion corresponding to each of the top three principal components is shown in Figure 10. This figure clearly exposes the oversimplification involved in the interpretation of the cross-correlation matrix in terms of one major motion. The component associated with the largest eigenvalue (equal to 16.3) corresponds to a twisting motion of the bundle's edges relative to its center, resulting in a periodic overwinding and underwinding of the coiled coil's edges. The second component (with an associated eigenvalue of 10.5) illustrates and exemplifies the rigid-body motion of the two, in terms of dynamics, halves of RM6 as previously discussed. The third eigenvector (with an associated eigenvalue of 9.9) corresponds to a bending (closing) movement of the whole bundle toward its intramolecular 2-fold axis (a finding consistent with the curvature of the coiled-coil axis as discussed in Coiled-Coil and Helical Parameters). The fourth and fifth components have associated eigenvalues significantly smaller than those already presented (4.9 and 3.6, respectively) and will not be discussed further. When the analysis is performed using the *Hex3* or *LHex* trajectories (instead of *Hex2*), very similar results are obtained (data not shown). It should be mentioned, however, that in the case of the *Hex3* trajectory the order of the top two principal components is reversed (with respect to *Hex2*).

DISCUSSION

We have shown that a five-residue deletion results in a complete reorganization of the Rop structure. The helix−turn−helix motif is abandoned, and a homotetrameric four-α-helix bundle of double the wild-type Rop length is formed. Along with this change of fold, everything else changes: the coiled-coil parameters, the hydrophobic core, the surface properties, and the dynamics. But even more surprising than these results is the fact that, statistically speaking, these results should not have been surprising: to our knowledge this is the third time (see Figure 1) that a relatively small change in the Rop sequence leads to dramatic reorganization of the Rop structure. Adding to these results cases such as that of the <2aa> variant (discussed below) only serves to accentuate the Rop sequence−structure puzzle. The fact that all these structures are of sufficient thermodynamic stability to be crystallized and studied only implies (tautologically) that they represent minima of their corresponding free energy landscapes, but this does little to help our understanding of how the sequence alters (or determines) the folding pathway toward one minimum or another. The silent assumption in the previous sentence is that for a given Rop sequence more than just one pronounced free energy minima may exist (i.e., that in terms of thermodynamic stability alone, a given sequence could have folded in at least two different ways). Computationally derived indications for the presence of such multiple free energy minima have been obtained for the A31P Rop mutant (*19*) and more recently for the $(Ala_2Ile_2)_6$ and $(Ala_2Leu_2)_6$ Rop variants (N. M. Glykos and M. Kokkinidis, unpublished results), but even if these computational indications are correct, knowing the existence of multiple free energy minima only serves to further complicate the question

of why only one of them is experimentally observed [allowing for the possibility that crystallization (even in different space groups) could be responsible for selecting specific conformers (*30*)].

Even if we take the complexity of the Rop sequence−structure relationships for granted, there is still an important question that remains unanswered: is this complexity characteristic of this protein alone, or can these results be generalized to other proteins? In the absence of similar experimental findings from other proteins, we would rather prefer to exercise caution by noting that it is indeed difficult to imagine such small sequence changes causing such drastic structural rearrangements in larger and/or nonsymmetric proteins.

Comparison of the RM6 structure with the <2aa> structure (see the introductory section for a description of this Rop variant) creates a new question: why restoration of the heptad repeat through an insertion has practically no effect on the Rop structure, whereas restoration through a deletion does. We believe that a qualitative and descriptive answer in this "to loop or not to loop" question can be obtained from an examination of the relevant sequences. The <2aa> variant is a full heptad repeat longer than RM6. The sequence of the seven residues of <2aa>, which are missing from RM6, is ADAADEQ. The first three residues (ADA) were designed to occupy positions *e*−*g* of one repeat and the last four (ADEQ) to occupy positions *a*−*d* of the next repeat. The closest match to the ADAADEQ sequence (from within Rop) is the HDHADEL sequence (starting at position 42 of the native protein). Our working hypothesis to explain why the ADAADEQ sequence failed to restore the helical continuity at the turn region is that glutamine may not be compatible with a *d* position in the context of an RM6-like structure (in wild-type Rop Q_{34} does occupy a *d* position, but this is a structurally special case since the side chain lies at the end of the bundle and its polar group is positioned toward the solvent and facing Arg_{55}). An attempt to overexpress a <2aa>-like variant of Rop but with a leucine replacing this glutamine is currently underway.

The comparison between the hydrophobic cores of native Rop and RM6 (shown in Figure 6) is worth mentioning again, but this time in a more generalized context. The current (metaphorical) view of protein hydrophobic cores is that of well-packed three-dimensional jigsaw puzzles. The implication of taking this metaphor at face value, however, is that we would consider it highly unlikely to find the same protein sequence being responsible for the formation of two different (but equally well-packed) cores. The RM6 structure clearly shows that even protein hydrophobic cores can be unexpectedly malleable. It should be admitted, however, that it is probably the symmetry of association together with the sequence periodicities of the Rop family that made such an exceptionally extensive repacking feasible.

The RM6 structure, together with the results from its molecular dynamics simulations, is in excellent agreement with the functional, thermodynamic, and kinetic data available for this variant (*6, 17, 52, 53*). The absence of any detectable biological activity is the direct consequence of the completely different fold and surface properties of the protein. The structure-based estimate of the solvation free energy gain upon tetramer formation, together with the very stable trajectories and average structures obtained from the

simulations, fully corroborates the experimentally determined thermostability of RM6. The computational (simulation-derived) predictions concerning the dynamics of RM6 are more difficult to validate experimentally, but indirect evidence has been obtained from the agreement with the crystallographic data. In this respect, it is probably noteworthy how the atomic mobility predicted from molecular dynamics agrees quantitatively better with the known stability of RM6 than with the mobility implied by (the rather high) crystallographic temperature factors.

To summarize, we have added a new member to the growing family of metamorphosed Rop structures, illustrating once more the plasticity of protein structures and the complexities of protein folding. In contrast to the other Rop variants previously studied, RM6 is more stable than wild-type Rop, offering a clear example of how evolution toward larger protein domains could have proceeded. In a more application-oriented note, we should mention that the stability and symmetry of RM6 could make it a promising starting material for the creation of artificial fibers with customizable surface properties.

EXPERIMENTAL AND COMPUTATIONAL PROCEDURES

Structure Determination and Analysis. The purification, crystallization, and data collection have been described previously (*31*). In summary, RM6 was purified from an overexpressing *Escherichia coli* strain (K38) using a single Q-Sepharose chromatography step by utilizing its differential solubility in solutions containing ethanol (20%, v/v) and varying amounts of sodium chloride. The protein was crystallized using a combination of mixed (organic solvent with salt) systems in microdialysis experiments complemented by successive micro- and macroseeding steps. The resulting crystals belong to space group *C*2 and contain the equivalent of half a bundle (two monomers) per asymmetric unit (with the whole bundle being formed through the application of a crystallographic 2-fold axis) with the following unit cell dimensions: $a = 54.5$ Å, $b = 42.5$ Å, $c = 51.7$ Å, and $\beta = 104.7°$. Crystallographic data were collected to 2 Å resolution on a MAR Research image plate detector mounted on a Rigaku RU-3HR rotating anode X-ray generator using the copper K_α radiation focused and mono-chromatized via a double nickel-coated mirror system. The rotation method was used with an oscillation range of 1°. Indexing and integration were performed with the HKL system (*32*) and resulted in data useful to 2.02 Å with an overall R_{symm} of 6.6% (31.2% in the last resolution shell), an overall completeness of 95.3% (76.0% in the last shell), and an average redundancy of 3.8 (2.6 in the last shell).

Molecular replacement calculations were performed with various programs (*33, 34*) and starting models and allowed the consistent determination of the approximate positions and orientations of two helices in the asymmetric unit. Initial difference maps indicated the absence of turns and the presence of continuous helical segments sufficiently elongated to account for the whole length of each monomer. An initial polyalanine model of the protein was constructed using helical backbone segments from other Rop structures. This initial polyalanine model accounted for 45 residues from the first monomer and 46 residues from the second. Rigid-body simulated annealing refinement (*34, 35*) of this model (using successively shorter rigid bodies, starting from one helix per body, and going down to two residues per rigid body) resulted in a polyalanine model giving an R of 0.397 and an R_{free} of 0.413 for all data between 8 and 2 Å resolution.

A $2mF_o - DF_c$ difference map (*36*) which was calculated using the phases from this refined polyalanine model allowed the unambiguous determination of the side chains' identities and orientations. Refinement was continued with successive rounds of torsion angle dynamics, simulated annealing, and addition of the most well-ordered water molecules using CNS (*37*) and Xfit [from the XtalView (*38*) suite of programs]. This procedure converged to a mostly complete model with an R of 0.235 and an R_{free} of 0.267 for all data. The refinement was completed with successive rounds of maximum likelihood refinement [using REFMAC (*39*)] interspersed with addition of further water molecules and fitting of discretely disordered residues. A round of TLS refinement (*27*) (using REFMAC and one set of TLS parameters per helix) led to the final model with an R of 0.189 and an R_{free} of 0.212. The model comprises residues 5−51 of the first monomer (chain A), residues 1−51 of the second (chain B), and 54 water molecules. All but one residue are in the core Ramachandran region (one in the allowed) with an overall G-factor [calculated with PROCHECK (*40*)] of 0.23. The crystallographically determined coordinates of RM6 have been deposited with the Protein Data Bank (PDB entry 1QX8).

Helical parameters (for individual helices) were calculated using HelAnal (*41*). To analyze the coiled-coil parameters, we used the coordinates of pseudoatoms representing the local helical axes of each helix. The local helical axes (with one pseudoatom for each residue and for each helix) were again calculated with HelAnal. The coordinates of the local coiled-coil axis were calculated as the average (x,y,z) coordinates of quadruplets of pseudoatoms belonging to the same layer (the classification in layers was based on the heptad sequence periodicity). An estimate of the curvature of the bundle axis was obtained by calculating the local bending angles of the coiled-coil axis. The calculation of the twist of the coiled coil per residue again used the local helix pseudoatoms and was based on the calculation of the polar rotation angles (ω, ϕ, and κ) needed to least-squares superimpose the four pseudoatoms of one layer onto the four pseudoatoms of the previous layer. The value of κ is the reported twist value. Helix−helix distances were calculated on a per residue basis using a distance matrix calculated from the coordinates of the local helix axis pseudoatoms. Similarly, to calculate the helix−helix crossing angles, we calculated the torsion angle among four pseudoatoms belonging to two successive layers of the (two) helices under examination.

All calculations reported here were performed using programs from the CCP4 suite of programs (*33*), as well as CNS (*37*), X-plor (*34*), APBS (*42*), Xfit (*38*), VMD (*43*), PyMol (DeLano Scientific), HELANAL (*41*), and locally written programs (for analyzing the helical and coiled-coil parameters).

Molecular Dynamics Simulations. Starting from the crystallographically determined coordinates of RM6, we built missing side chain and hydrogen atoms with PSFGEN from the NAMD distribution (*44*) and assuming an acidic pH (with the histidines fully protonated). Three different explicit

solvation systems were prepared, each with a successively larger volume (and, thus, number of water molecules). The reason for this multiplicity of hydration systems lies with the highly asymmetric electrostatic potential of RM6 (see Figure 3): when a hydration system with a tight bounding box is used, RM6 molecules can experience the electrostatic potential of neighboring molecules. This leads to a slow (but consistent) rotation of the molecules (during the molecular dynamics simulation) that allows them to interact electrostatically with their neighboring images. Having observed this behavior for all simulations performed with a tight bounding box, we decided to undertake a series of simulations using larger systems (in which a direct interaction between solute molecules is physically impossible). Comparison of these molecular dynamics trajectories allowed us to confirm that the electrostatic interaction between the periodic images of RM6 (arising from the periodic boundary conditions used for the simulations) does not significantly alter the dynamics of individual RM6 molecules.

The three hydration systems used in our simulations are (i) a relatively tight hexagonal cell comprising 4675 pre-equilibrated TIP3 water molecules (*45*) with unit cell basis vectors (projections along the orthogonal axes) of (100,0,0), (0,39.8,23), and (0,0,46) and a shortest (initial) solute−solute distance of 20 Å, (ii) a larger hexagonal cell comprising 7966 pre-equilibrated TIP3 water molecules with unit cell basis vectors (projections along the orthogonal axes) of (110,0,0), (0,47.6,27.5), and (0,0,55) and a shortest (initial) solute−solute distance of 30 Å, and (iii) a truncated octahedral cell with a shortest dimension of 89 Å (which is approximately 8 Å longer than the longest dimension of the RM6 bundle) comprising 10 321 pre-equilibrated TIP3 water molecules. For all three systems, the crystallographically determined waters were kept, while those water molecules less than 1.8 Å from the protein surface were removed. The final systems comprised 3216 protein atoms and 14 025, 23 898, or 30 963 water atoms depending on the periodic boundary cell that was used (see above). The net charge of the solute was neutralized through the addition of sodium and chloride ions to a final concentration of 100 mM (one sodium and three choride ions for the tight hexagonal cell, three and five for the large hexagonal cell, and five and seven for the octahedral cell).

The molecular dynamics simulations were performed with NAMD (*44*) using the CHARMM27 force field (*46*) as follows. The systems were first energy minimized for 2000 conjugate gradient steps with the positions of the backbone atoms fixed, and then for another 2000 steps without positional restraints. They were then slowly heated to a final temperature of 320 K (with a temperature step ΔT of 20 K) over a period of 66 ps with the positions of the C_α atoms harmonically restrained about their energy-minimized positions. Subsequently, the systems were equilibrated for 200 ps under NpT conditions without any restraints. This was followed by the production NpT runs with the temperature and pressure controlled using the Nosé-Hoover Langevin dynamics and Langevin piston barostat control methods as implemented in NAMD (and maintained at 320 K and 1 atm). For the tight hexagonal cell system, we performed three independent production runs giving a total simulation time of 34.2 ns (with each run lasting for 7.6, 14.2, and 12.5 ns). These three simulations are called in the text of the paper

simulations *Hex1*, *Hex2*, and *Hex3*, respectively. For the large hexagonal cell, we performed one production run which lasted 12.7 ns (called *LHex*). Finally, for the truncated octahedral cell, we performed two production runs with lengths of 4.6 and 2.8 ns, respectively (called *Oct1* and *Oct2*). All production runs were performed with the impulse Verlet-I multiple-time step integration algorithm as implemented by NAMD. The inner time step was 2 fs; short-range nonbonded interactions were calculated every two steps and long-range electrostatic interactions every four time steps using the particle mesh Ewald method (*47*). A cutoff for the van der Waals interactions was applied through a switching function, and SHAKE was used to restrain all bonds involving hydrogen atoms. Trajectories were obtained by saving the atomic coordinates of the whole system every 0.4 ps. All simulations reported here were performed on the departmental (Department of Molecular Biology and Genetics, Democritus University of Thrace) 18-node Beowulf class I cluster.

Calculation of the average trajectory structures, of the rms deviation from the average structures, of the radius of gyration, and of the atomic rms fluctuations was performed with X-PLOR (*34*). Calculation of the anisotropic fluctuations was performed with g_rmsf from the GROMACS suite of programs (*48*). Removal of global rotations and translations, calculation of rms deviations from the experimental structure, the $C_\alpha-C_\alpha$ distance map (and the corresponding rms deviation from it), the cross-correlation matrix, and the principal component analysis were performed with *carma* (*49*), available via http://www.mbg.duth.gr/~glykos/.

ACKNOWLEDGMENT

REFERENCES

1. Polisky, B. (1988) ColE1 replication control circuitry: Sense from antisense, *Cell 55*, 929−932.
2. Banner, D. W., Kokkinidis, M., and Tsernoglou, D. (1987) Structure of the ColE1 Rop Protein at 1.7 Å Resolution, *J. Mol. Biol. 196*, 657−675.
3. Eberle, W., Pastore, A., Sander, C., and Rosch, P. (1991) The structure of ColE1 rop in solution, *J. Biomol. NMR 1*, 71−82.
4. Cohen, C., and Parry, D. A. D. (1990) α-Helical Coiled Coils and Bundles: How to Design an α-Helical Protein, *Proteins 7*, 1−15.
5. Chou, K.-C., Maggiora, G. M., Némethy, G., and Scheraga, H. A. (1988) Energetics of the structure of the four-α-helix bundle in proteins, *Proc. Natl. Acad. Sci. U.S.A. 85*, 4295−4299.
6. Castagnoli, L., Scarpa, M., Kokkinidis, M., Banner, D. W., Tsernoglou, D., and Cesareni, G. (1989) Genetic and structural analysis of the ColE1 Rop (Rom) protein, *EMBO J. 8*, 621−629.
7. Castagnoli, L., Vetriani, C., and Cesareni, G. (1994) Linking an easily detectable phenotype to the folding of a common structural motif. Selection of rare turn mutations that prevent the folding of Rop, *J. Mol. Biol. 237*, 378−387.
8. Predki, P. F., Nayak, L. M., Gottlieb, M. B. C., and Regan, L. (1995) Dissecting RNA-Protein Interactions: RNA-RNA Recognition by Rop, *Cell 80*, 41−50.
9. Levy, Y., Cho, S. S., Shen, T., Onuchic, J. N., and Wolynes, P. G. (2005) Symmetry and frustration in protein energy landscapes: A near degeneracy resolves the Rop dimer-folding mystery, *Proc. Natl. Acad. Sci. U.S.A. 102*, 2373−2378.
10. Nagi, A. D., and Regan, L. (1997) An inverse correlation between loop length and stability in a four-helix-bundle protein, *Folding Des. 2*, 67−75.

11. Predki, P. F., Agrawal, V., Brünger, A. T., and Regan, L. (1996) Amino-acid substitutions in a surface turn modulate protein stability, *Nat. Struct. Biol. 3*, 54−58.

12. Steif, C., Hinz, H.-J., and Cesareni, G. (1995) Effects of Cavity-Creating Mutations on Conformational Stability and Structure of the Dimeric 4-α-Helical Protein ROP: Thermal Unfolding Studies, *Proteins 23*, 83−96.

13. Steif, C., Weber, P., Hinz, H. J., Flossdorf, J., Cesareni, G., and Kokkinidis, M. (1993) Subunit Interactions Provide a Significant Contribution to the Stability of the Dimeric Four-α-Helical-Bundle Protein ROP, *Biochemistry 32*, 3867−3876.

14. Kokkinidis, M., et al. (1993) Correlation Between Protein Stability and Crystal Properties of Designed ROP Variants, *Proteins 16*, 214−216.

15. Chou, K.-C., Maggiora, G. M., and Scheraga, H. A. (1992) Role of loop-helix interactions in stabilizing four-helix bundle proteins, *Proc. Natl. Acad. Sci. U.S.A. 89*, 7315−7319.

16. Rosengarth, A., Rosgen, J., and Hinz, H. J. (1999) Slow unfolding and refolding kinetics of the mesophilic Rop wild-type protein in the transition range, *Eur. J. Biochem. 264*, 989−995.

17. Lassalle, M. W., Hinz, H.-J., Wenzel, H., Vlassi, M., Kokkinidis, M., and Cesareni, G. (1998) Dimer-to-Tetramer Transformation: Loop Excision Dramatically Alters Structure and Stability of the ROP Four α-helix Bundle Protein, *J. Mol. Biol. 279*, 987−1000.

18. Munson, M., Anderson, K. S., and Regan, L. (1997) Speeding up protein folding: Mutations that increase the rate at which Rop folds and unfolds by over four orders of magnitude, *Folding Des. 2*, 77−87.

19. Glykos, N. M., and Kokkinidis, M. (2004) Structural Polymorphism of a Marginally Stable 4-α-Helical Bundle. Images of a Trapped Molten Globule? *Proteins 56*, 420−425.

20. Kresse, H. P., Czubayko, M., Nyakatura, G., Vriend, G., Sander, C., and Bloecker, H. (2001) Four-helix bundle topology re-engineered: Monomeric Rop protein variants with different loop arrangements, *Protein Eng. 14*, 897−901.

21. Willis, M. A., Bishop, B., Regan, L., and Brunger, A. T. (2000) Dramatic structural and thermodynamic consequences of repacking a protein's hydrophobic core, *Structure 8*, 1319−1328.

22. Glykos, N. M., Cesareni, G., and Kokkinidis, M. (1999) Protein plasticity to the extreme: Changing the topology of a 4-α-helical bundle with a single amino acid substitution, *Structure 7*, 597−603.

23. Ceruso, M. A., Grottesi, A., and Di Nola, A. (1999) Effects of core-packing on the structure, function, and mechanics of a four-helix-bundle protein ROP, *Proteins 36*, 436−446.

24. Vlassi, M., et al. (1994) Restored heptad pattern continuity does not alter the folding of a four-α-helix bundle, *Nat. Struct. Biol. 1*, 706−716.

25. Brunet, A. P., et al. (1993) The role of turns in the structure of an α-helical protein, *Nature 364*, 355−358.

26. Eisenberg, D., and McLachlan, A. D. (1986) Solvation energy in protein folding and binding, *Nature 319*, 199−203.

27. Schomaker, V., and Trueblood, K. N. (1998) Correlation of Internal Torsional Motion with Overall Molecular Motion in Crystals, *Acta Crystallogr. B54*, 507−514.

28. Amadei, A., Linssen, A. B. M., and Berendsen, H. J. C. (1993) Essential dynamics of proteins, *Proteins 17*, 412−425.

29. Ichiye, T., and Karplus, M. (1991) Collective Motions in Proteins: A Covariance Analysis of Atomic Fluctuations in Molecular Dynamics and Normal Mode Simulations, *Proteins 11*, 205−217.

30. Betz, S. F., Raleigh, D. P., DeGrado, W. F., Lovejoy, B., Anderson, D., Ogihara, N., and Eisenberg, D. (1996) Crystallization of a designed peptide from a molten globule ensemble, *Folding Des. 1*, 57−64.

31. Papanikolau, Y., Kotsifaki, D., Fadouloglou, V. E., Gazi, A. D., Glykos, N. M., Cesareni, G., and Kokkinidis, M. (2004) Ionic strength reducers: An efficient approach to protein purification and crystallization. Application to two Rop variants, *Acta Crystallogr. D60*, 1334−1337.

32. Otwinowski, Z., and Minor, W. (1997) Processing of X-ray Diffraction Data Collected in Oscillation Mode, *Methods Enzymol. 276*, 307−326.

33. Collaborative Computational Project, Number 4 (1994) The CCP4 suite: Programs for protein crystallography, *Acta Crystallogr. D50*, 760−763.

34. Brünger, A. T. (1992) *X-PLOR*, version 3.1, Yale University Press, New Haven, CT.

35. Glykos, N. M., and Kokkinidis, M. (1999) Meaningful refinement of poly-alanine models using rigid-body simulated annealing: Application to the structure determination of the A31P Rop mutant, *Acta Crystallogr. D55*, 1301−1308.

36. Read, R. J. (1986) Improved Fourier coefficients for maps using phases from partial structures with errors, *Acta Crystallogr. A42*, 140−149.

37. Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) Crystallography & NMR System: A New Software Suite for Macromolecular Structure Determination, *Acta Crystallogr. D54*, 905−921.

38. McRee, D. E. (1992) A visual protein crystallographic software system for X11/XView, *J. Mol. Graphics 10*, 44−46.

39. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) Refinement of Macromolecular Structures by the Maximum-Likelihood Method, *Acta Crystallogr. D53*, 240−255.

40. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) PROCHECK: A program to check the stereochemical quality of protein structures, *J. Appl. Crystallogr. 26*, 283−291.

41. Bansal, M., Kumar, S., and Velavan, R. (2000) HELANAL: A program to characterise helix geometry in proteins, *J. Biomol. Struct. Dyn. 17*, 811−819.

42. Baker, N. A., Sept, D., Joseph, S., Holst, M. J., and McCammon, J. A. (2001) Electrostatics of nanosystems: Application to microtubules and the ribosome, *Proc. Natl. Acad. Sci. U.S.A. 98*, 10037−10041.

43. Humphrey, W., Dalke, A., and Schulten, K. (1996) VMD: Visual Molecular Dynamics, *J. Mol. Graphics 14*, 33−38.

44. Kale, L., Skeel, R., Bhandarkar, M., Brunner, R., Gursoy, A., Krawetz, N., Phillips, J., Shinozaki, A., Varadarajan, K., and Schulten, K. (1999) NAMD2: Greater scalability for parallel molecular dynamics, *J. Comput. Phys. 151*, 283−312.

45. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983) Comparison of simple potential functions for simulating liquid water, *J. Chem. Phys. 79*, 926−935.

46. MacKerell, A. D., et al. (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins, *J. Phys. Chem. B 102*, 3586−3616.

47. Darden, T., York, D., and Pedersen, L. (1993) Particle mesh Ewald. An $N \log(N)$ method for Ewald sums in large systems, *J. Chem. Phys. 98*, 10089−10092.

48. Lindahl, E., Hess, B., and van der Spoel, D. (2001) GROMACS 3.0: A package for molecular simulation and trajectory analysis, *J. Mol. Model. 7*, 306−317.

49. Glykos, N. M. (2006) Carma: A molecular dynamics analysis program, *J. Comput. Chem.* (in press).

50. Glykos, N. M., and Kokkinidis, M. (2000) GraphEnt: A maximum entropy program with graphics capabilities, *J. Appl. Crystallogr. 33*, 982−985.

51. Merritt, E. A., and Bacon, D. J. (1997) Raster3D: Photorealistic molecular graphics, *Methods Enzymol. 277*, 505−524.

52. Lassalle, M. W., and Hinz, H. J. (1998) Unfolding of the tetrameric loop deletion mutant of ROP protein is a second-order reaction, *Biochemistry 37*, 8465−8472.

53. Lassalle, M. W., and Hinz, H. J. (1999) Refolding studies on the tetrameric loop deletion mutant RM6 of ROP protein, *Biol. Chem. 380*, 459−472.