Effects of Core-Packing on the Structure, Function, and Mechanics of a Four-Helix-Bundle Protein ROP

Marc A. Ceruso,* Alessandro Grottesi, and Alfredo Di Nola
Department of Chemistry, University of Rome "La Sapienza," Rome, Italy

ABSTRACT **The effects of core-packing on the structure, function and mechanics of the RNA-binding 4-helix-bundle Rop have been studied by molecular dynamics simulations. The structural, dynamical and geometrical properties of the Rop homodimer, (formed by the antiparallel juxtaposition of two helix-turn-helix motifs), have been compared with those of three protein variants described by Munson et al. (Protein Sci, 5:1584–1593, 1996), where the core of the native protein has been systematically repacked using a two-amino acid alphabet: Ala_2Leu_2-8, Ala_2Leu_2-8-rev, and Leu_2Ala_2-8. The results showed that it was possible to readily distinguish the inactive protein Leu_2Ala_2-8 from the other functionally active systems based on tertiary and quaternary structure criteria. Structural properties such as native secondary structure content did not correlate with biological activity. Biological activity was related in part to the relative arrangement of the residues within the binding site. But, more global aspects, related to the overall topology of the helical bundle, accounted for the small functional differences between Ala_2Leu_2-8 and Ala_2Leu_2-8-rev. Mechanically, the 4-helix-bundle absorbed core mutations by altering the local structure at the sequence termini and in the turns that join the two helices of each monomer, and by changing the overall orientation and separation of the extremely rigid helices. Proteins 1999;36:436–446.**

Key words: protein structure; molecular dynamics; Rom; hydrophobic; coiled-coil; alpha-helix; RNA-binding; structure-function relationships

INTRODUCTION

A remarkable feature of protein structure is the tight and regular packing of core residues.[1–4] Core residues contribute significantly to a protein's stability,[5] and the thermostability of a protein can be increased by engineering more efficiently packed cores.[6–10] In addition, the core of a protein participates in its biological function by providing a more or less flexible scaffold onto which the array of biologically relevant residues is disposed.[3,11] In the same manner, the packing of core residues controls the topology of the polypeptidic chain (the fold),[12–14] and small changes to core residues lead to dramatic changes in quaternary structure, as in the case of GCN4, a naturally dimeric coiled-coil, which reorganizes itself into trimeric or tetrameric bundles after modest amino acid mutations at core packing sites.[15,16] However, the role of packing in a protein's structure, function, and stability is not so clear cut.[17] Increased hydrophobic surface burial and filling of cavities within the core do not necessarily improve the stability of a protein.[18–20] Mutagenesis studies have shown that it is possible to mutate simultaneously the quasi-totality of the core residues of a protein while maintaining biological activity and presumably a native-like structure.[7,9,21–23] In many proteins, multiple core mutations are accommodated by small rearrangements of side chain and backbone elements without loss of function or fold topology.[22,24–30] And finally, the de novo design of proteins has demonstrated that side-chain complementarity (efficient packing) alone is not a sufficient principle for driving the formation of "native-like" proteins.[31,32]

In this work, we were interested in determining at an atomic level the effects of core mutations on the structure, function, and mechanics of a protein. To this end, we have chosen to study by molecular dynamics (MD) simulations the RNA-binding protein Rop (also known as Rom), and to compare the structural and mechanical properties of the wild-type protein with those of three recently reported[7] Rop variants, where the native protein core has been systematically repacked: Ala_2Leu_2-8, Ala_2Leu_2-8-rev, and Leu_2Ala_2-8.

Rop is an RNA-binding protein that regulates the replication of ColE1 and related plasmids in *Escherichia coli*.[33–36] Rop is an antiparallel 4-α-helical coiled-coil protein, formed by the noncovalent association of two identical monomers (63 amino acids each) consisting of two helices connected by sharp bend. Like coiled-coils, every helix in Rop is characterized by a seven-residue repeat, noted (*abcdefg*). Residues at position *a* and *d* are mostly apolar and form the core of the protein. The core residues are distributed in eight four-residue layers which pack on top of each other and lie perpendicular to the bundle's principal axis (Fig. 1). Each layer consists of two *a* residues and two *d* residues, with every monomer contributing one *a*,*d*-pair. In Rop most *a* residues are small (alanines) and most *d* residues are large (leucines). Regan and cowork-

Grant sponsor: EC; Grant number: ERBFMRX-CT96-0013; Grant sponsor: MURST (Progetto Nazionale: "Biologia Strutturale").
*Correspondence to: Marc Ceruso c/o Prof. Alfredo Di Nola, Department of Chemistry, University "La Sapienza," P.O. Box 34 Roma 62, P.le Aldo Moro 5, Rome I-00185, Italy. E-mail: mceruso@signac.chem.uniroma1.it
Received 4 January 1999; Accepted 8 April 1999

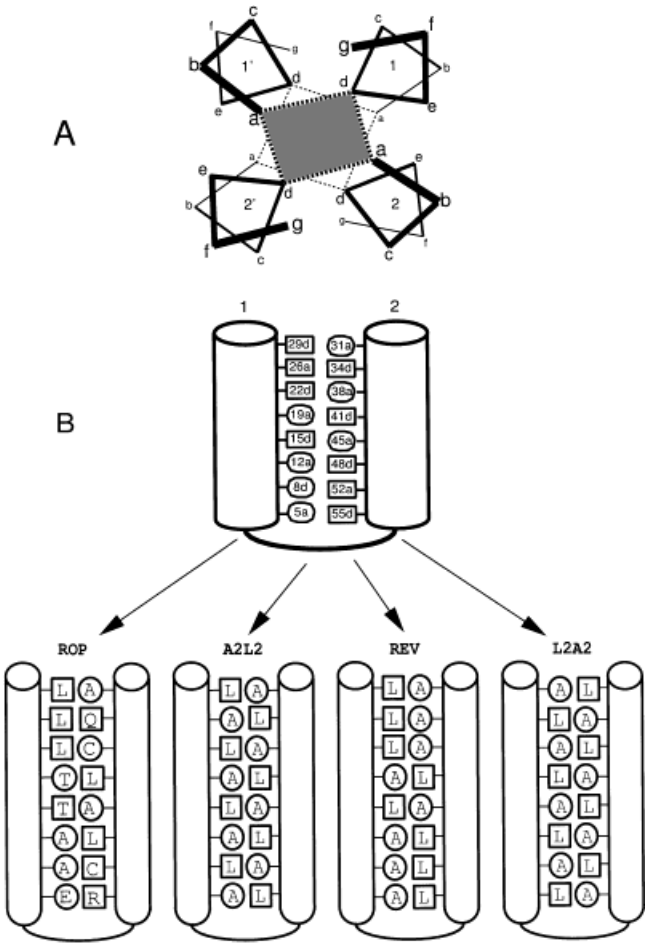

Fig. 1. Schematic representation of Rop's helical bundle. **A:** top view of the core of the bundle, showing two out of the eight planar layers formed by two "a,d"-pairs each. **B:** vertical view of one of the bundle's monomers with numbering of core residues: the first planar layer consist of residues 29, 31, 5′, and 55′, the second planar layer consists of residues 26, 34, 8′, and 52′ . . . Locations with "large" native residues are boxed, while those with "small" native residues are circled. Location and nature of amino acids within the core for wild-type and core-repacked mutants.

ers[7,9] have used single and two-amino acid alphabets at *a* and *d* positions to entirely repack the core of Rop. Mutants with single amino acid alphabets are either completely unstructured (Ala at *a* and *d* positions), or have a molten globule structure (Leu at *a* and *d* positions). But, by incorporating alternatively alanine and leucine residues at *a* and *d* positions, as in Ala_2Leu_2-8, Ala_2Leu_2-8-rev, and Leu_2Ala_2-8 (Fig. 1), a stable native-like protein is obtained. Significant differences in biological activity between the three proteins are found. While Ala_2Leu_2-8-rev is as active as Rop, Ala_2Leu_2-8 has an RNA-dissociation constant five-fold larger than that of the wild-type protein, and Leu_2Ala_2-8 is completely inactive. The core-repacked proteins have slight differences in secondary structure content as well, the mean residue ellipticity at 222 nm decreasing from Leu_2Ala_2-8 to Rop, to Ala_2Leu_2-8, and to Ala_2Leu_2-8-rev.

The aim of the present work was to uncover the physical changes that resulted from the various packing organizations in Rop and its variants, in order to explain the experimentally observed differences in secondary structure and biological function, and, more generally, to study the relationships between core packing patterns and the structural and mechanical organization of coiled-coil 4-helix-bundled proteins. MD simulations agreed well with available experimental data, reproducing the structural and dynamical properties of Rop's nuclear magnetic resonance (NMR) and X-ray determined structures, and affording secondary structure contents directly comparable with circular dichroism data. In terms of structure-function relationships, the various conformational and geometrical reporters allowed a clear distinction between the biologically active systems (Rop, Ala_2Leu_2-8, and Ala_2Leu_2-8-rev) and the inactive one (Leu_2Ala_2-8). No correlation was found between native secondary structure content and biological activity. Structurally, Ala_2Leu_2-8 was less "native-like" than Ala_2Leu_2-8-rev, but it maintained an orientation and an arrangement of residues within the binding site which were comparable to those of Ala_2Leu_2-8-rev, suggesting that the observed small differences in activity between the two systems were related to changes in mechanical properties rather than binding site configuration. The core-packing patterns in the various mutants gave rise to specific perturbations in the geometrical organization of the 4-helix-bundles. These perturbations were both local (introduction of strain, loss of secondary structure) and global (coiled-coil pitch length, helix packing orientation) in their nature, and were related to the symmetry of the packing patterns in the core of the proteins. The results indicated that MD simulations represent an efficient tool to predict the structural and biological behavior of a set of related proteins differing by a large number of mutations. The results highlighted some aspects of the relationship between a protein's structure and its biological activity, and suggested a mechanical decomposition that helped to understand how the 4-helix-bundle system responded to core mutations. This work made extensive use of algorithms and computational software (Gromacs, SHAKE, thermal-bath-coupling) that were originally developed in the group of Prof. H.J.C. Berendsen.

METHODS

Starting Configurations

The initial wild-type protein configuration was taken from the second NMR model in entry 1rpr[37] of the protein databank, because it is the closest to the average NMR structure. Mutants were constructed within Insight II 97.0 (MSI, San Diego, CA) using the above NMR coordinates as a template and replacing target residues with the desired amino acid. The following Rop variants, Ala_2Leu_2-8, Ala_2Leu_2-8-rev, and Leu_2Ala_2-8 were constructed and simulated by molecular dynamics (see Fig. 1 or Munson et al.[7] for definitions).

TABLE I. Global Structural Reporters†

System	R_G (nm)	SASA (nm^2)	$RMSD_{XRAY}$ (nm)	$RMSD_{NMR}$ (nm)	Alpha	Turn	Coil	Strain	HB5	HB4	HB3	HB2
ROPXray	1.33	62.22	—	0.178 (0.037)	100	4	8	0	2	92	4	2
ROPNMR	1.44 (0.01)	59.86 (2.30)	0.178 (0.037)	0.117 (0.024)	96.6 (2.6)	7.4 (2.6)	8 (0.0)	1.2 (0.9)	4.1 (1.1)	86.9 (2.6)	10.5 (2.7)	1.3 (1.0)
ROPMD	1.46 (0.01)	65.82 (1.02)	0.223 (0.015)	0.221 (0.020)	91.4 (3.6)	8.1 (3.3)	13.3 (1.2)	2.3 (1.0)	1.7 (1.3)	84.1 (3.0)	10.2 (2.6)	4.1 (1.4)
A2L2	1.45 (0.01)	65.19 (1.13)	0.282 (0.017)	0.258 (0.017)	89.6 (2.7)	7.5 (2.6)	13.6 (1.8)	1.2 (0.9)	1.0 (0.8)	80.9 (3.5)	12.6 (3.0)	3.0 (1.4)
REV	1.46 (0.01)	63.87 (1.16)	0.241 (0.013)	0.233 (0.013)	84.4 (2.4)	12.2 (2.6)	14.5 (1.5)	2.3 (1.2)	0.5 (0.6)	76.0 (2.0)	15.9 (2.7)	4.3 (1.95)
L2A2	1.48 (0.01)	67.00 (1.00)	0.403 (0.024)	0.402 (0.022)	92.1 (2.0)	7.7 (2.1)	12.4 (0.9)	2.3 (1.0)	1.6 (0.9)	84.2 (1.7)	11.5 (2.6)	2.6 (1.1)

†All reported values are average values over the xray-subset, numbers in parenthesis refer to standard deviations conformations were sampled every 10 ps. Secondary structure assignments were done using the program DSSP.[45] Alpha, Turn, and Coil designate respectively the total number of residues in alpha-helical, in turn and in random-coil conformation. Strain designates the total number of residues in unfavorable regions of the Ramachandran map.[49] HB5 (resp. HB4, HB3, HB2) designates the average number of hydrogen bonds of type $O_i \leftarrow HN_{i+5(resp.4,3,2)}$.

Molecular Dynamics

All simulations were performed with the GROMACS simulation package.[38] Each system was immersed in a pre-equilibrated box of simple point charge (SPC) water.[39] There were approximately ~3,230 water molecules for a total of ~10,900 atoms in each system. All simulations were at least 1.61 ns long, and the first 0.30 ns of each simulation were discarded in order to ensure equilibration. Molecular dynamics simulations were initiated as follows: using a restraining harmonic potential, all heavy atoms of the protein were constrained to their initial positions, while surrounding SPC water molecules were first minimized and then submitted to 5 ps of constant volume MD at 300 K. The resulting system was then minimized, without any constraints, before starting constant temperature and constant volume MD. A nonbonded cutoff of 1.2 nm was used for both Lennard-Jones and Coulomb potentials (this is a typical value within the commonly used ranged of 1.0 to 1.4 nm). The pair lists were updated every ten steps. A constant temperature of 300 K was maintained by coupling to an external bath[40] using a coupling constant ($\tau = 0.002$ ps) equal to the integration time step. The SHAKE algorithm[41] was used to constrain bond lengths. No counter ions were used in the simulations, since in order to accomodate their large number (14 to achieve neutrality of the simulated box) the box dimensions should have been enlarged to a computationally prohibitive size, if undesirable ion/protein contacts were to be prevented. The comparison of the results with experimental data did not show any significant effect due to this choice.

RESULTS AND DISCUSSION

Definitions

The three-dimensional structure of Rop has been determined by X-ray crystallography[42] and solution NMR.[37,43] In this study, the coordinates of the second NMR model (see Methods) were used as a template for the starting configuration of all systems, since the last seven amino acids of each monomer (residues 57 to 63) and parts of the side chain of residues Lys3, Lys6, and Met11 are not present in the crystal coordinates. In what follows, ROPMD, A2L2, REV, and L2A2 designate the molecular dynamics simulations of the proteins Rop, Ala_2Leu_2-8, Ala_2Leu_2-8-rev, and Leu_2Ala_2-8, respectively. ROPNMR refers to the set of ten NMR models, and ROPXray designates the crystallographic structure of Rop. Finally, the set of residues corresponding to the first 56 amino acids of each monomer will be called the "xray-subset."

Average properties for ROPMD, A2L2, and REV systems were computed over the 300–1,610 ps time interval, while average properties for L2A2 were computed between 600 and 1,910 ps, where L2A2 showed a relatively more stable character (see below, and Fig. 2). Structural and dynamical properties were computed over the xray-subset, unless otherwise specified.

Comparison of ROPMD, ROPNMR, and ROPXray

The experimentally determined structures, ROPNMR and ROPXray, were similar but not identical (first two rows in Table I).The radius of gyration and solvent accessible surface area were clearly different, but the secondary structure and hydrogen-bonding patterns were more comparable, with ROPNMR having less alpha-helical structure, and more turns and bends. The average root-mean-square deviation (RMSD) between ROPNMR and ROPXray, calculated over C-alpha atoms only, was 0.178 nm. Eberle et al.[37] reported that differences (≥0.15 nm) between the NMR set of structures and the X-ray model were located principally around residues 32–36 (this region is past the turn that joins the two helices of each monomer), the N-terminus (residues 1–7) and the "C-terminus" (residues 51–56). These differences were attributed to crystal packing effects (N-terminus and turn region) and to the presence of slowly exchanging conformers at the C-terminus, around residue Cys52 in particular.[37] Conformational differences between ROPMD and ROPXray were also located at the N- and C-termini of each monomer, but the third location of deformation involved the turn residues 28–33 instead.

In ROPMD, as in ROPNMR and ROPXray, the last seven residues of each monomer were highly flexible. Plots of the RMSD with respect to the initial NMR configuration and to ROPXray (Fig. 2A and 2B), and plots of the radius of gyration (Fig. 2C) and of the solvent-accessible surface area as a function of time (Fig. 2D) indicated that ROPMD reached a conformational equilibrium within 200–300 ps of simulation.

ROPMD was equally distant from ROPNMR and ROPXray (~0.22 nm, Table I). The average radius of

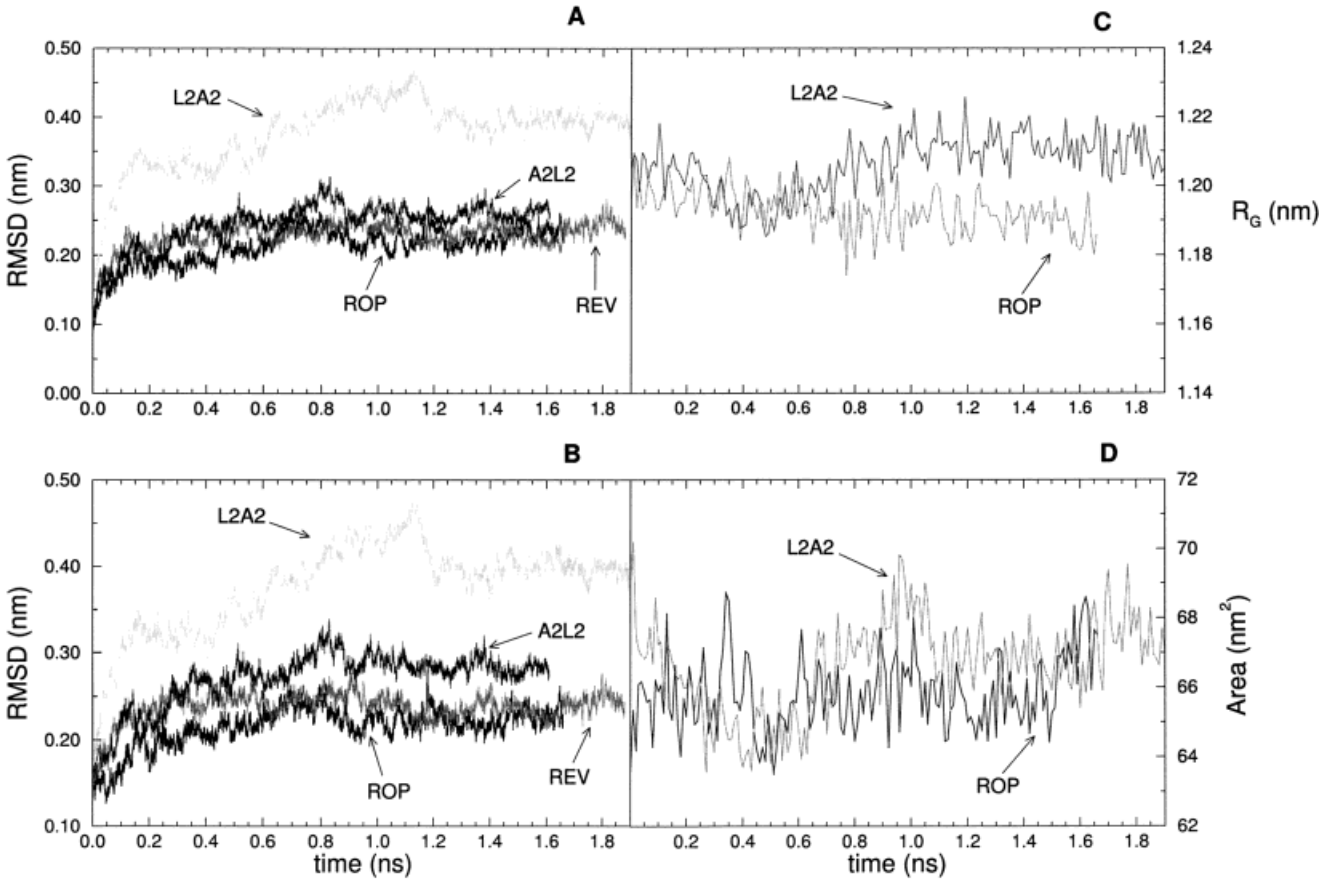

Fig. 2. Atomic root-mean-square deviation (nm), radius of gyration (nm), and solvent accessible surface area (nm^2) as a function of time. The RMSD was calculated over the backbone atoms of the xray-subset (see text) by fitting the corresponding C-alpha atoms onto those of the NMR model **(A)**, or those of the crystallographic structure **(B).** The radius of gyration **(C)** and the solvent accessible surface area **(D)** for ROP and L2A2 were calculated over the X-ray-subset as well. Individual protein systems are indicated within each frame.

gyration for ROPMD (1.46 nm) was comparable to that of the average of the ten NMR models (1.44 nm), but larger than that of the X-ray structure (1.33 nm). This was not surprizing since we had used an NMR-derived structure as a starting configuration. However, the average solvent accessible surface area (SASA) for ROPMD (65.82 nm^2) was larger than that of ROPNMR (59.86 nm^2), and was closer to the SASA of the X-ray structure (62.22 nm^2). The difference in SASA between ROPMD and the X-ray structure was explained by taking into account the six methylenes, two methyls and four amino groups which are missing from the X-ray model and that belong to the solvent facing residues Lys3, Lys6, and Met11. In an amino acid in extended conformation, the average SASA of a methyl, methylene, and amino moiety are approximately 0.40 nm^2, 0.30 nm^2, and 0.48 nm^2 respectively.[44] The difference in SASA between ROPMD and the average NMR model reflects probably a difference in protocols. NMR structures were refined using in vacuo simulations and this may have "collapsed" the side chains towards the core of the protein, while explicit solvent molecules were used in this study. We checked that the differences in SASA between ROPMD and the NMR models did not reflect a solvent exposure of core residues, but were due to a systematic increase in the solvent exposure of amino acids that face the solvent in the NMR models (Fig. 3).

Comparison of experimental and MD-derived B factors showed strong consistency in intensity and helical periodicity within the helical sections of Rop, but the N-terminal (residues 1–7) and C-terminal (54–56) extremities as well as the central loop of the protein (residues 25–32) were significantly more flexible in the simulation than in the solid-state structure (Fig. 4). The backbone root-mean-square fluctuation of individual residues in the MD simulation of Rop and the standard deviation of backbone atoms among the ten NMR-models had closely related tendencies across monomer sequences (Fig. 5). The MD structures had smaller fluctuations in general, but the loop region presented a more pronounced difference in mobility relative to the two helices in the MD simulation than in the NMR models.

Thus, ROPMD reached an equilibrium configuration with patterns of fluctuation closely related to those of ROPNMR and ROPXray, with remarkably "rigid" helical sections and with the extremities of the 4-helix bundle being the most flexible regions. The overall shape of

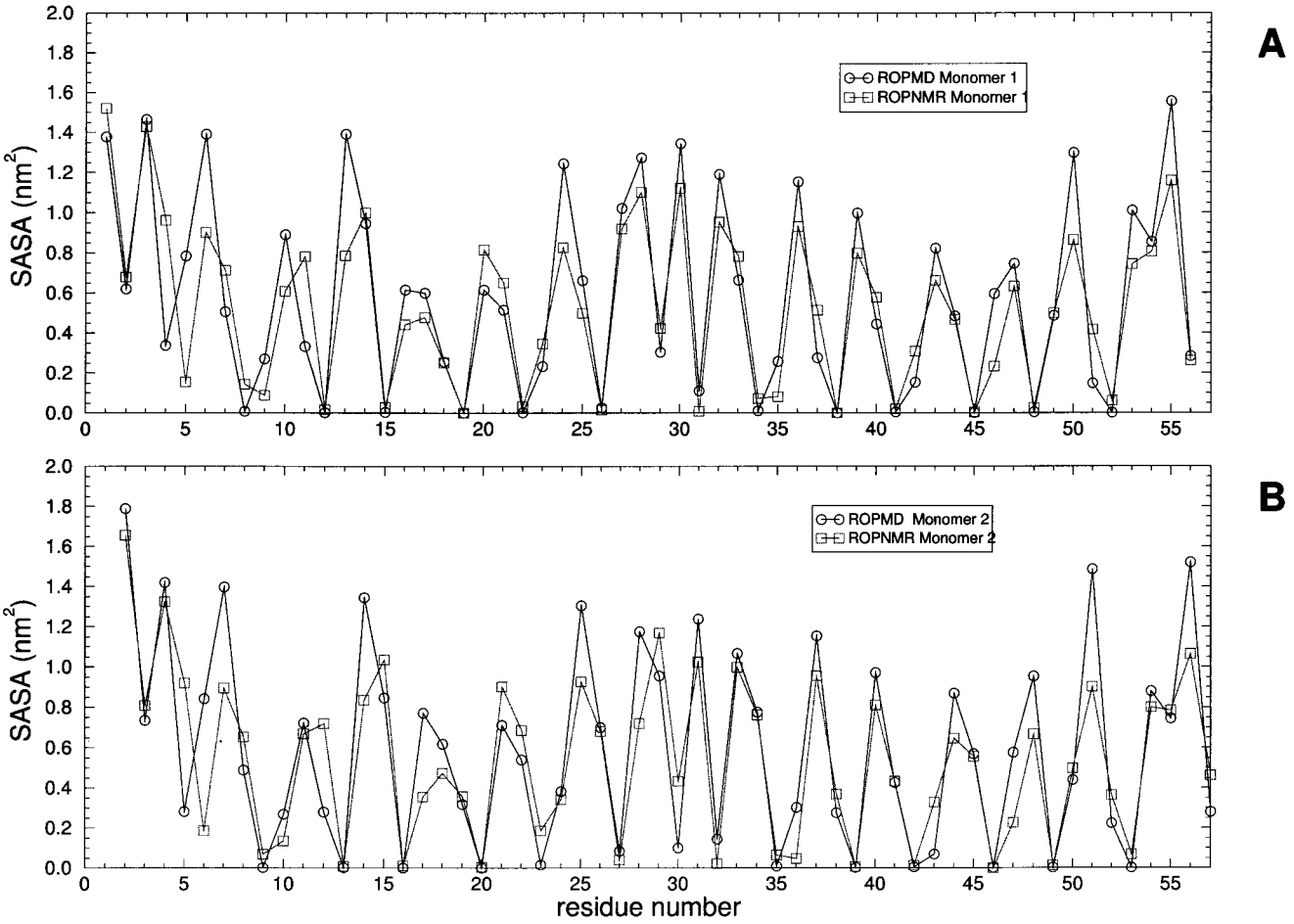

Fig. 3. Comparison of the average solvent accessible surface area (nm^2) per residue between ROP and NMR models. Averages were calculated between 0.30 ns and 1.61 ns every 10 ps for the MD trajectory of Rop (open circle), and over the ten NMR models (open squares). Data represent averages per residue and per monomer.

ROPMD was a compromise between ROPNMR and ROPXray, the differences between ROPMD, ROPNMR, and ROPXray reflecting the differences between the various methodologies.

Comparison of Simulated Systems

Global structural reporters and biological activity

L2A2, the only system which is not biologically active, was clearly unique in its behavior (Fig. 2, Table I). L2A2 equilibrated more slowly than the other systems, reaching relatively stable plateaux in RMSD, R_G, and SASA, only after ~600 ps. L2A2 underwent the largest conformational changes with respect to ROPNMR, ROPXray, and the other systems as well (Table I and Table II). L2A2 had also the largest R_G and SASA values. Nevertheless, and in agreement with experimental data, L2A2 conserved a degree of helicity comparable to that of ROPMD (vide infra).

The three biologically active systems, ROPMD, A2L2, and REV formed a closely related set of structures (Fig. 2, Table I, and Table II). But, some differences between A2L2 and REV were apparent (Table I and Table II). REV had smaller C-alpha RMSD values with respect to ROPMD, to ROPXray and to ROPNMR than A2L2, in correspondence with Ala_2Leu_2-8-rev being more active biologically than Ala_2Leu_2-8.[7]

Secondary structure and circular dichroism ellipticity

The evolution of the secondary structure of each system was monitored as a function of time. Secondary structure assignments were done using the program DSSP.[45] Conformations were sampled every 10 ps. Only ROPMD's and REV's results are shown Figure 6. Results concerning secondary structure and hydrogen-bonding statistics (xray-subset only) are summarized in Table I. In general, secondary structure was well conserved but there were various degrees of fraying at the N-terminal and C-terminal extremities and at the edges of the central loop: helical residues adopting 3_{10}, bend or turn-like conformations, HB4 decreasing in favor of HB3 and HB2 hydrogen-bonding. The more dramatic losses in alpha-helical (Al-

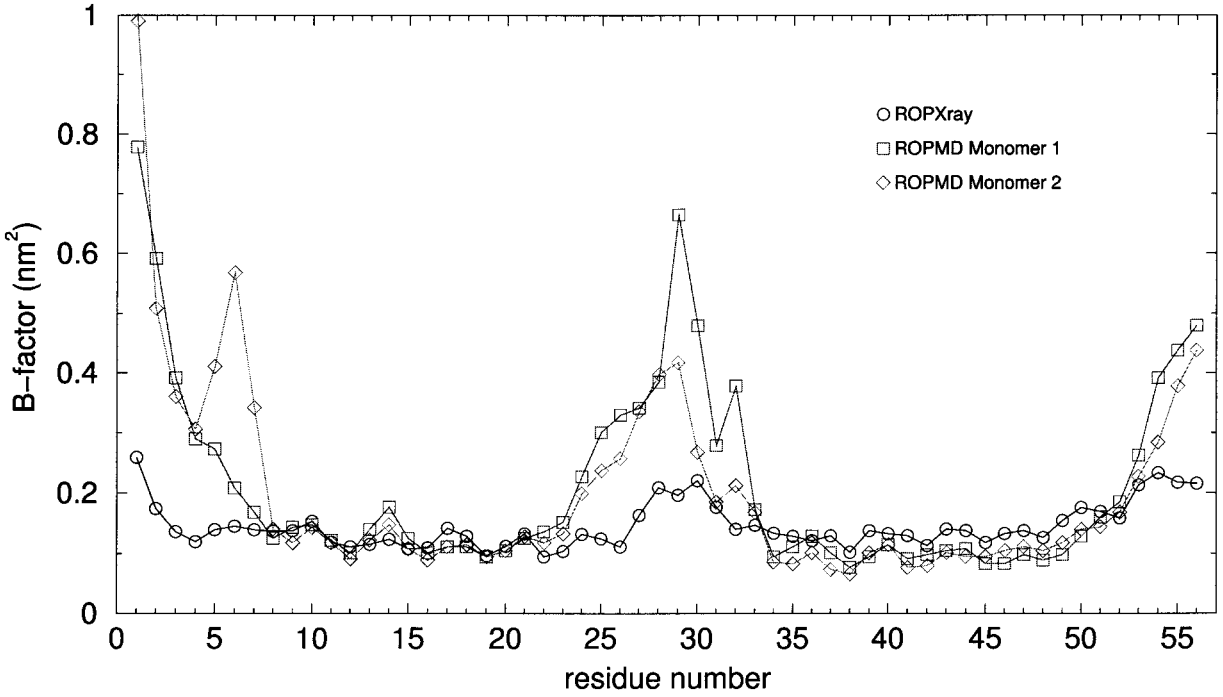

Fig. 4. Comparison of MD and experimental B factors (nm^2) for Rop protein. B factors of backbone heavy atoms, averaged within a residue (for residues 1–56), were evaluated from atomic root-mean-square fluctuations (RMSF) using[51] $B = 8\pi^2/3 < (\Delta r)^2 >$. Open circles crystallographic values, open diamonds for the first monomer in ROP and open squares for the second in ROP.

pha) secondary structure and alpha-helical hydrogen-bonding (HB4) took place in REV. The statistical significance of the differences was ascertained by comparing the means using a t-Student test (reported averages and standard deviations derive from an uncorrelated sample). The values of Alpha and HB4, within the xray-subset, decreased as L2A2 $\simeq$ ROPMD $>$ A2L2 $>$ REV, in correspondence with ellipticity data at 222 nm.[7] This observation remained valid when taking into account the full protein sequences, with "Alpha" (helical) counts decreasing from L2A2 (92.4 (1.9)) to ROPMD (92.1 (3.6)), to A2L2 (90.2 (3.6)) and to REV (84.5 (2.3)) and HB4 counts decreasing from ROPMD (85.3 (3.0)), to L2A2 (84.4 (1.7)), to A2L2 (82.4 (3.8)), and to REV (76.4 (2.0)). For comparison, the mean-residue ellipticity values at 222 nm are 30.5 10^3, 30.1 10^3, 28.5 10^3, and 27.5 10^3 degree.cm^2.dmol^{-1}, for Leu_2Ala_2-8, Rop, Ala_2Leu_2-8 and Ala_2Leu_2-8-rev, respectively.[7]

A closer look at structure-function relationships

The preceding results have shown that the MD simulations have reproduced with fidelity the available experimental data concerning the tertiary/quaternary structure and the fluctuations of Rop as well as the secondary structure content of the various proteins. From a functional activity (RNA-binding) point of view, the large structural differences between L2A2 and the rest of the systems are probably linked to the incapacity of Leu_2Ala_2-8 to bind RNA. However, it is interesting to note that the experimental or MD-derived secondary structure contents did not correlate with biological activity: Leu_2Ala_2-8, which conserves native secondary structure well, is inactive, while Ala_2Leu_2-8-rev, which undergoes a distinct loss in helical content is as active as Rop. In the same manner, the helical content and the solvent-accessible surface area were respectively more comparable within the A2L2-ROPMD pair than within the REV-ROPMD pair, although the RNA-dissociation constant of Ala_2Leu_2-8 is five-fold greater than that of Ala_2Leu_2-8-rev.[7]

The only structural evidence indicating that REV was more native-like than A2L2 (and hence more active than A2L2?) was obtained when comparing the average RMSD among the various MD systems (Fig. 2 and Table II) or with the experimentally determined structures (Table I). Note that these differences are statistically significant (Null hypothesis for the comparison of the means gave $P \ll 0.001$ or $0.01 < P < 0.001$ by t-test, possible correlation effects were taken into account and the standard deviations were corrected accordingly[46]). Thus, A2L2 seemed to present some global structural differences that could be related to its lower biological activity. What are those differences and how does Ala_2Leu_2-8, in spite of these differences, still bind RNA?

The residues responsible for RNA recognition in Rop are all located on the solvent exposed face of helices 1 and 1′, they form a symmetrically arranged set, consisting of residues Asn10, Phe14, Gln18, and Lys25 on each monomer.[33] To investigate the difference in activity between Ala_2Leu_2-8-rev and Ala_2Leu_2-8 further, we constructed matrices of distances (distance maps) among all the atoms

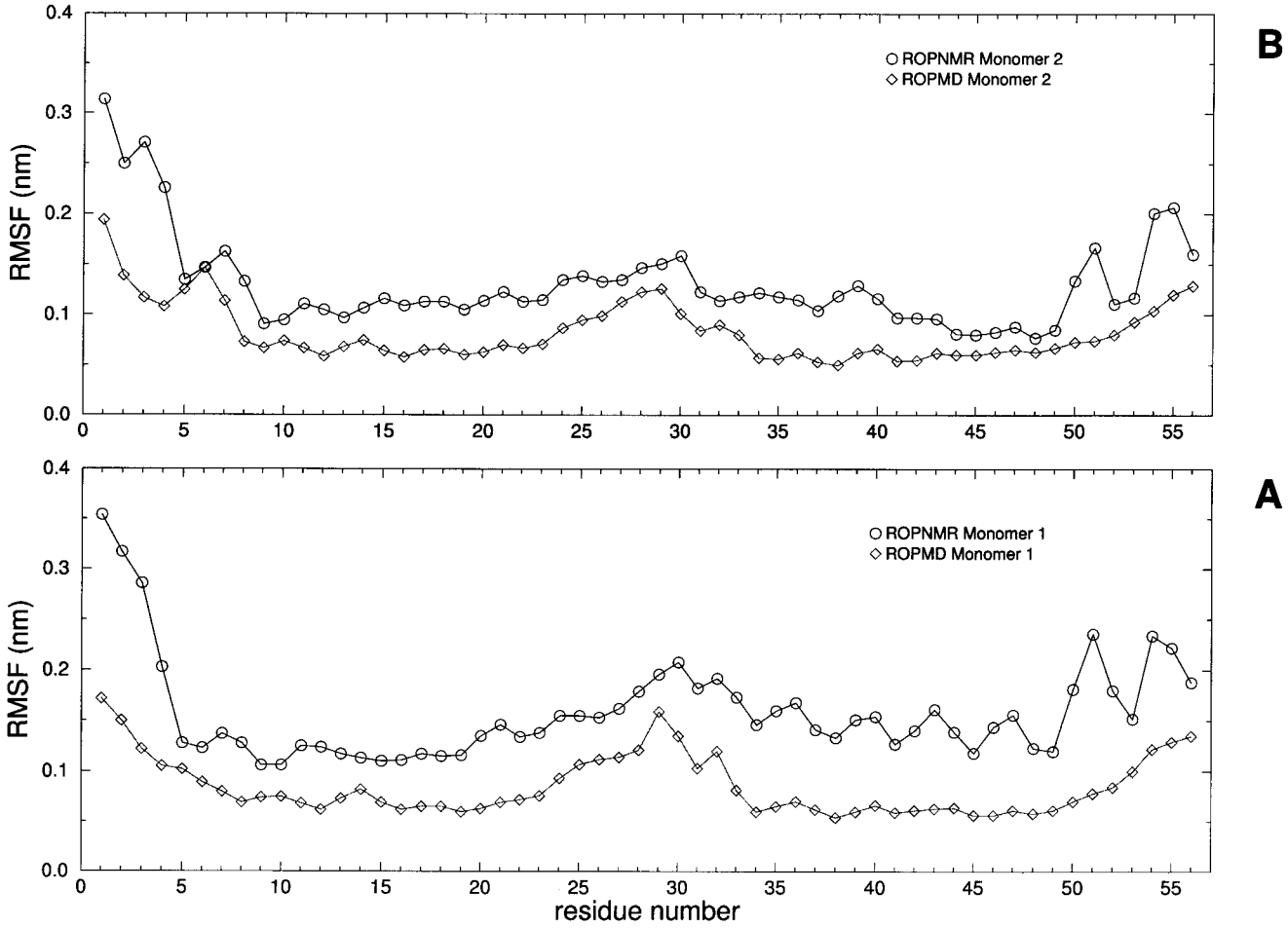

Fig. 5. Residue root-mean-square fluctuation (nm) in ROP (open diamonds) and NMR-models (open circles). Data were averaged over backbone heavy atoms. **A**: first monomer, **B**: second monomer.

TABLE II. Average RMSD (nm) Values Relative to Average MD Structures†

System	⟨ROPMD⟩	⟨A2L2⟩	⟨REV⟩	⟨L2A2⟩
ROPMD	0.090 (0.015)	0.224 (0.014)	0.208 (0.011)	0.327 (0.025)
A2L2	0.226 (0.018)	0.095 (0.017)	0.212 (0.015)	0.0316 (0.021)
REV	0.215 (0.014)	0.216 (0.030)	0.106 (0.018)	0.347 (0.013)
L2A2	0.327 (0.041)	0.316 (0.025)	0.343 (0.036)	0.093 (0.023)

†RMSD values were calculated over C-alpha atoms only within the X-ray-subset (see text). Standard deviations are reported in parenthesis. ⟨ ⟩ denotes the average MD structure computed over adequate time intervals and sampled every 10 ps (see text). A given row contains the average RMSD of the corresponding simulation with respect to the average structure of the title column.

of those residues directly involved in RNA-binding (e.g., 10, 14, 18, and 25 of each monomer). These distance maps represent a finger print of the binding site organization and allow a conformational comparison between the binding sites of two proteins free of any fitting scheme. The distance map computed from the time-average configuration of ROPMD was used as reference. The time average of the root-mean-square difference between the distance map of a given variant system and the reference map (that of ROPMD) is given Table III. Once again, L2A2 marked its difference. But, most remarkably, A2L2 presented an average binding site organization which was indistinguishable from that of REV, suggesting that the structural differences which were observed between REV and A2L2 at a more global level (vide supra) were the ones responsible for the differences in activity between these two proteins. Indeed, we are comparing binding site configurations in their unbound state, but a binding event, such as

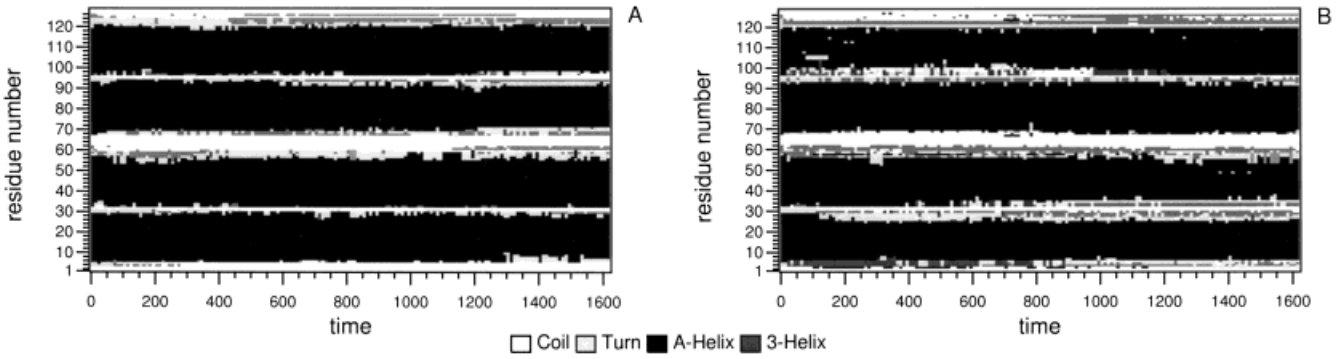

Fig. 6. Evolution of secondary structure in **(A)** ROPMD and in **(B)** REV. Secondary structure is color coded using a gray scale which becomes lighter in going from alpha-helix, to 3_{10}, to turn or bend and finally to random-coil.

TABLE III. Average Root Mean Square Difference (nm) Between Distance Maps[†]

	A2L2	REV	L2A2
$dmap_{ROPMD}$	0.200 (0.028)	0.192 (0.028)	0.296 (0.044)

[†]$dmap_{ROPMD}$ designates the time-average distance map computed from ROPMD. Rows contain the time average of the root-mean-square difference between the distance map of the corresponding system and $dmap_{ROPMD}$. Standard deviations are given in parenthesis.

RNA-binding, is not a static one, and the mechanical and dynamical properties of each protein play a crucial role. In other words, the difference in the overall shape between A2L2 and REV must give rise to differences in their mechanical and dynamical properties which, in turn, would affect RNA-binding. The results in the next section ("core-packing patterns and protein mechanics") confirmed this interpretation. Munson et al.[7] had suggested the existence of an "end effect" related to the "reversed" layers of Ala_2Leu_2-8-rev to account for the difference in activity between REV and L2A2. When looking at the distance maps in detail, it was indeed found that distances relating residues at the extremities of the binding site as well as inter-helical distances were more "ROPMD-like" in REV than in A2L2 (data not shown). But, A2L2 showed better conservation of the binding site configuration among intra-helical residues 10′, 14′ and 25′ in helix 1′, and that the contact between residue 25 and residue 10′ was also better conserved in A2L2 than in REV (data not shown).

Core-packing patterns and protein mechanics

Comparison of the backbone root-mean-square fluctuation among the various MD-simulated systems at the residue level showed that in the mutant systems, as in the native protein, the most flexible regions were the N- and C-termini as well as the region around the central turn of each monomer, while the remaining helical sections presented a remarkably low level of fluctuation (<0.1 nm) in all systems. Differences between the various systems concerned principally residues 25–35.

The average backbone RMS deviation from the X-ray structure was distributed through the sequence in a way that was specific to each of the simulated systems, suggesting that the various core-packing patterns had had specific structural effects. These effects were mostly global ones (e.g., orientation and separation of helices within the bundle) and were investigated by evaluating a series of geometrical properties, related to the intra-monomeric as well as the inter-monomeric organization (Table IV) of the helices in the 4-helix-bundle.

Among the intra-monomeric properties, the loss of symmetry in the behavior of the two monomer units was particularly striking (compare $\Theta_{1,2}$ vs $\Theta_{1',2'}$ for any MD system and the corresponding values for ROPXray). Such loss of the two-fold symmetry of the dimer was present in the starting NMR configuration.[37] ROPMD had the most symmetrical behavior of all systems, with comparable values for the torsion (1.4 degrees difference), the distance (0.007 nm difference), and the pitch (0.80 nm difference) in the two monomer units. The others systems, presented various degrees of symmetry in their properties: in REV, the helical torsions were comparable (differing by 1.5 degrees), but the inter-helical distances were not maintained (0.093 nm difference), and the pitches differed (1.7 nm) twice as much as in ROPMD; in A2L2 the relative orientation of the helices and the inter-helical distances were both different (8.6 degrees and 0.122 nm of respective difference), and the pitches were markedly different; L2A2 had an intermediate level of symmetry, with inter-helical distances and torsions being different, but to a lesser extent (4.4 degrees and 0.063 nm difference), the pitches had also an intermediate behavior differing by 2.4 nm. Thus the two-fold symmetry of the 4-helix-bundle was maintained with a decreasing capacity in going from ROPMD, to REV, to L2A2, and finally to A2L2. It is also interesting to note that both L2A2 and A2L2 (only for one of the monomers in the case of A2L2) were capable of producing shorter coiled-coils than the other two systems.

TABLE IV. Protein Geometry[†]

	Intra-monomer properties						Inter-monomer properties			
System	$\Theta_{1,2}$ (°)	$d_{1,2}$ (nm)	$\Theta_{1',2'}$ (°)	$d_{1',2'}$ (nm)	$Pitch_{1,2}$ (nm)	$Pitch_{1',2'}$ (nm)	$\Theta_{1,1'}$ (°)	$d_{1,1'}$ (nm)	$\Theta_{2,2'}$ (°)	$d_{2,2'}$ (nm)
ROPXray	18.0	0.870	18.0	0.870	19.2 (2.3)	19.2 (2.3)	22.5	1.020	17.4	0.987
ROPMD	19.9 (2.9)	0.870 (0.025)	21.3 (2.6)	0.877 (0.022)	16.5 (0.7)	15.8 (1.4)	24.2 (3.2)	1.040 (0.043)	27.7 (2.9)	0.996 (0.032)
A2L2	23.1 (1.7)	0.773 (0.033)	14.5 (3.2)	0.895 (0.033)	12.0 (1.1)	18.8 (2.1)	19.8 (3.9)	1.004 (0.034)	27.3 (2.6)	1.074 (0.032)
REV	20.7 (2.5)	0.908 (0.030)	19.2 (3.0)	0.815 (0.029)	17.2 (3.7)	15.5 (3.3)	19.3 (3.0)	0.976 (0.023)	27.8 (2.3)	0.990 (0.047)
L2A2	27.8 (2.7)	0.840 (0.033)	23.4 (2.0)	0.903 (0.025)	11.7 (0.8)	14.3 (3.4)	14.1 (1.9)	0.846 (0.019)	41.8 (7.2)	1.379 (0.080)

[†]$\Theta_{i,j}$ and $d_{i,j}$ denote respectively the torsion angle and the distance between helices i and j. Helices were taken to run from residues 6–28 and 32–54 for each monomer. The helix axis was determined by the geometric center, O_i, of the C-alpha trace, and the first principal axis, $\mathbf{u_i}$, (that with the largest eigenvalue) of helix i. The torsion between two helices was defined as the torsion between the pairs (O_i, $\mathbf{u_i}$) and (O_j, $\mathbf{u_j}$). A clockwise rotation was positive. The distance between two helices was defined as the distance between the corresponding geometric centers. The pitch between two helices was obtained first by calculating the local pitch as described by Seo and Cohen,[50] and then by averaging the local pitches over the length of the helix.

This capacity could be related to their regularly alternating pattern of small and large residues at *a* and *d* positions, while in REV and ROP where the patterns are interrupted (see introduction and Fig. 1), the super-helical pitch is larger, suggesting that the "reverse" layers of REV could play a role in pitch conservation, by preventing super-helical coiling, in a manner similar to how "stutters" and "stammers" lead to local underwinding and overwinding of alpha-helical coiled-coils.[47]

Inter-monomeric properties, showed that, with the exception of L2A2, the MD-systems had inter-helical distances ($d_{1,1'}$ and $d_{2,2'}$), which were comparable to those of ROPXray, but that the relative orientation of the helices in the MD-systems and ROPXray were different. Nevertheless, among the simulated systems, A2L2 and REV were again similar to each other. The torsion angle between helices 2 and 2′, in A2L2 and REV, was comparable with its value in ROPMD. However, in REV and A2L2, the helices 1 and 1′, which form the RNA-binding face of the bundle, adopted a different orientation than in ROPMD. But, the markedly different values of the inter-monomeric properties of L2A2 were particularly striking, and afforded further justification for the biological inactivity of this protein.

It is interesting to draw a parallel between the highly asymmetrical values of $\Theta_{1,1'}$ and $\Theta_{2,2'}$, in L2A2 and the asymmetry (albeit less marked) between $\Theta_{1,2}$ and $\Theta_{1',2'}$ in A2L2. Indeed, the packing pattern of these two proteins in the 4-helix-bundle are related to each other by a 90 degree rotation about the bundle's long axis, which would bring in correspondence helices 1′ and 1 of Leu_2Ala_2-8 with helices 1 and 2 of Ala_2Leu_2-8. This and the above data concerning coiled-coil pitch lengths suggest that similar packing patterns could lead to similar mechanical effects.

This correspondence between core sequence patterns and mechanical properties was also observed when comparing the location of strained residues in the various protein systems (see Table I for the definition of "strained"). Besides various deformations at the extremities of the bundle which were observed in all systems (data not shown), residues 34 and 34′, that belong to the second and seventh core layer, were "strained" only in the REV and L2A2 systems. Indeed, Ala_2Leu_2-8-rev and Leu_2Ala_2-8 share the same arrangement of small and large residues in these layers (Fig. 1), and these layers are the ones which distinguish the proteins Ala_2Leu_2-8 and Rop, from the proteins Ala_2Leu_2-8-rev, and Leu_2Ala_2-8. Note that the latter two systems do differ from Rop at this location as well: inspection of Rop's sequence shows that two large residues (Leu26 and Gln34) are positioned in the second layer (Leu26′ and Gln34′ for the seventh layer), instead of an alternation of small and large residues. The appearance of strain after a cavity creating mutation has been observed previously,[48] and is related to an increased conformational flexibility.

CONCLUSION

The present work has explored the relationships between a given core-packing pattern within a precise fold (the 4-helix-bundle of Rop) and the structure and function of the corresponding protein. The results from the molecular dynamics simulations of Rop and the three variant proteins were in full agreement with experimental data, reproducing the secondary structure content of the various protein systems (Table I), as well as the structural and dynamical properties of ROPMD with respect to ROPNMR and ROPXray (Fig. 2, 4, and 5).

Biological activity was correlated in part with the conservation of native tertiary/quaternary structure, but not with native secondary structure conservation. This was well illustrated by L2A2, which had native-like (ROPMD-like) levels of secondary structure, but had three-dimensional properties that showed a very different structural and mechanical behavior (Tables I, II, III, and IV). The biologically active systems (ROPMD, A2L2, and REV) were easily distinguished from the inactive one (L2A2). The difference in activity between Ala_2Leu_2-8 and Ala_2Leu_2-8-rev appeared to be related to a difference in global or mechanical (vide infra) organization of the helical bundle: three-dimensional parameters showed that A2L2 and REV did not behave similarly (Table I, II, and IV) with respect to ROPMD or to the experimentally determined structures, but that the RNA-binding face of A2L2 maintained a very native-like structure (Table III and IV).

Mechanically, each monomer in the various protein systems could be seen as the juxtaposition of five building-blocks, corresponding to the two termini (residues 1–7 and 54–56), the central loop-turn (residues 25–35) and the two helices (8–24 and 36–53). The systems differed from each other locally (at the termini and the central loop-turn), by their secondary structure (Fig. 6), their atomic fluctuations (Fig. 4 and 5) and their deviations from X-ray or NMR structures (see paragraph on "core-packing patterns and protein mechanics"). The systems differed also globally by the relative arrangement of their helical blocks (Table IV). The different packing patterns gave rise to various organization of the helical bundle, but the loss of the two-fold symmetry of the native Rop protein was particularly interesting as it decreased from ROPMD, to REV, to L2A2, and to A2L2 (note the lack of correlation with functional activity). Some specific relationships between local or global packing patterns and the quaternary structure of the helical bundle were suggested: the REV's reversed layer, which interrupts the regularly alternating pattern of small and large residues along the helix coil, seemed to maintain a native pitch value, while those proteins with regularly alternating patterns seemed to generate (at least in three out of four cases) super-coiled coils. In the same manner, the packing arrangement within the second and seventh layer, which is identical in REV and L2A2, lead to the creation of strain at residue 34 (resp. 34′ in the other monomer) in those two proteins only. However, establishing sure relationships between the packing patterns and the organization of the corresponding helical bundle would require further investigation.

The results, presented in this work, have shown that it is possible to use molecular dynamics to predict successfully differences in structural and biological behavior among proteins that differ from each other by a large number of mutations. In addition, biological activity was principally related to the conservation of the three-dimensional arrangement of residues within the binding site, while small differences in activity were related to the three-dimensional organization of the protein. Finally, a mechanical decomposition of the protein was achieved, where core-packing patterns within the 4-helix-bundle motif were accommodated mostly by global changes in the relative orientation and separation of the very rigid helices, and by local changes concentrated at the sequence termini and in the central loop region.

ACKNOWLEDGMENTS

We would like to thank CASPUR (Consorzio per le Applicazione di Supercalcolo per Università e Ricerca) for access to computational facilities.

REFERENCES

1. Chothia C. Structural invariants in protein folding. Nature 1975; 254:304–308.
2. Harpaz Y, Gerstein M, Chothia C. Volume changes on protein folding. Structure 1994;2:641–649.
3. Levitt M, Gerstein M, Huang E, Subbiah S, Tsai J. Protein folding: the endgame. Annu Rev Biochem 1997;66:549–579.
4. Ragunathan G, Jernigan RL. Ideal architecture of residue packing and its observation in protein structures. Protein Sci 1997;6: 2072–2083.
5. Dill KA. Dominant forces in protein folding. Biochemistry 1990;29: 7133–7155.
6. Malakauskas SM, Mayo SL. Design, structure and stability of a hyperthermophilic protein variant. Nat Struct Biol 1998;5:470–475.
7. Munson M, Balasubramanian S, Fleming KG, et al. What makes a protein a protein? Hydrophobic core designs that specify stability and structural properties. Protein Sci 1996;5:1584–1593.
8. Russell RJ, Taylor GL. Engineering thermostability: lessons from thermophilic proteins. Curr Opin Biotechnol 1995;6:360–374.
9. Munson M, O'Brien R, Sturtevant JM, Regan L. Redesigning the hydrophobic core of a four-helix-bundle protein. Protein Sci 1994; 3:2015–2022.
10. Ishikawa K, Nakamura H, Morikawa K, Kanaya S. Stabilization of Escherichia coli ribonuclease HI by cavity-filling mutations within a hydrophobic core. Biochemistry 1993;32:6171–6178.
11. Ponder JW, Richards FM. Tertiary templates for proteins: use of packing criteria in the enumeration of allowed sequences for different structural classes. J Mol Biol 1987;193:775–791.
12. Chothia C, Finkelstein AV. The classification and origins of protein folding patterns. Annu Rev Biochem 1990;59:1007–1039.
13. Chothia C. Principles that determine the structure of proteins. Annu Rev Biochem 1984;53:537–572.
14. Crick FHC. The packing of alpha-helices: simple coiled-coils. Acta Crystallogr 1953;6:689–697.
15. Harbury PB, Kim PS, Alber T. Crystal structure of an isoleucine-zipper trimer. Nature 1994;371:80–83.
16. Harbury PB, Zhang T, Kim PS, Alber T. A switch between two-, three-, and four-stranded coiled coils in GCN4 leucine zipper mutants. Science 1993;262:1401–1407.
17. Behe MJ, Lattman EE, Rose GD. The protein folding problem: the native fold determines packing, but does packing determine the native fold? Proc Natl Acad Sci USA 1991;88:4195–4199.
18. Takano K, Yamagat Y, Yutani K. A general rule for the relationship between hydrophobic effect and conformational stability of a protein: stability and structure of a series of hydrophobic mutants of human lysozyme. J Mol Biol 1998;280:749–761.
19. Ishikawa K, Kimura S, Kanaya S, Morikawa K, Nakamura H. Structural study of mutants of Escherichia coli ribonuclease HI with enhanced thermostability. Protein Eng 1993;6:85–91.
20. Jackson SE, Moracci M, elMasry N, Johnson CM, Fersht AR. Effect of cavity creating mutations in the hydrophobic core of chymotrypsin inhibitor 2. Biochemistry 1993;32:11259–11269.
21. Axe DD, Foster NW, Fersht AR. Active barnase variants with completely random hydrophobic cores. Proc Natl Acad Sci USA 1996;93:5590–5594.
22. Gassner NC, Baase WA, Matthews BW. A test of the "jigsaw puzzle" model for protein folding by multiple methionine substitutions within the core of T4 lysozyme. Proc Natl Acad Sci USA 1993;93:12155–12158.
23. Lim WA, Sauer RT. Alternative packing arrangements in the hydrophobic core of λ repressor. Nature 1989;339:31–36.
24. Závodszky P, Kardos J, Svingor Á, Petsko GA. Adjustment of conformational flexibility is a key event in the thermal adaptation of proteins. Proc Natl Acad Sci USA 1998;95:7406–7411.
25. Dahiyat BI, Mayo SL. Probing the role of packing specificity in protein design. Proc Natl Acad Sci USA 1997;94:10172–10177.
26. Su A, Mayo SL. Coupling backbone flexibility and amino acid sequence selection in protein design. Protein Sci 1997;6:1701–1707.
27. Dahiyat BI, Mayo SL. Protein design automation. Protein Sci 1996;5:895–903.
28. Harbury PB, Tidor B, Kim PS. Repacking protein cores with backbone freedom: structure prediction for coiled coils. Proc Natl Acad Sci USA 1995;92:8408–8412.
29. Baldwin EP, Matthews BW. Core-packing constraints hydrophobicity and protein design. Curr Opin Biotechnol 1994;5:396–402.
30. Baldwin EP, Hajiseyedjavadi O, Baase WA, Matthews BW. The role of backbone flexibility in the accomodation of variants that repack the core of T4 lysozyme. Science 1993;262:1715–1717.
31. Betz SF, Liebman PA, DeGrado WF. De novo design of native proteins: characterization of proteins intended to fold into antiparallel, Rop-like, four-helix bundles. Biochemistry 1997;36:2450–2458.

32. Betz SF, DeGrado WF. 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.
33. Predki PF, Nayak LM, Gottlieb MBC, Regan L. Dissecting RNA-protein interactions: RNA-RNA recognition by Rop. Cell 1995;80: 41–50.
34. Castagnoli L, Scarpa M, Kokkinidis M, Banner DW, Tsernoglou D, Cesarini G. Genetic and structural analysis of the ColE1 Rop (Rom) protein. EMBO J 1989;8:621–629.
35. Tomizawa J-I, Som T. Control of ColEl1 plamid replication: enhancement of binding of RNA I to the primer transcript by the ROM protein. Cell 1984;38:871–878.
36. Cesarini G, Muesing MA, Polisky B. Control of ColE1 DNA replication: the rop gene product negatively effects transcription from the replication primer promoter. Proc Natl Acad Sci USA 1982;79:6313–6317.
37. Eberle W, Pastore A, Sander C, Rosch P. The structure of ColE1 rop in solution. J Biomol NMR 1991;1:71–82.
38. Van der Spoel D, Berendsen HJC, Van Buuren AR, et al. Gromacs User Manual. Nijenborgh 4, 9747 AG Groningen, The Netherlands. Internet: http://rugmd0.chem.rug.nl/~gmx 1995.
39. Berendsen HJC, Postma JPM, Van Gunsteren WF, Hermans J. Interaction models for water in relation to protein hydration. In: Pullman B, editor. Intermolecular forces. Dordrecht: D. Reidel Publishing Company; 1981. p 331–342.
40. Berendsen HJC, Postma JPM, Di Nola A, Haak JR. Molecular dynamics with coupling to an external bath. J Chem Phys 1984;81:3684–3690.
41. Ryckaert JP, Ciccotti G, Berendsen HJC. Numerical integration of the cartesian equations of motion of a system with constraints; molecular dynamics of n-alkanes. J Comp Phys 1977;23:327–341.
42. Banner DW, Kokkinidis M, Tsernoglou D. Structure of the ColE1 Rop protein at 1.7 Å resolution. J Mol Biol 1987;196:657–675.
43. Eberle W, Klaus W, Cesareni G, Sander C, Röch P. Proton nuclear magnetic resonance assignments and secondary structure determination of the ColE1 rop (rom) protein. Biochemistry 1990;29: 7402–7407.
44. Creighton TE. Proteins: structures and molecular properties. 2nd ed. New York: W. H. Freeman and Company; 1993. 507 p.
45. Kabsch W, Sander C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers 1983;22:2577–2637.
46. Di Nola A, Brünger AT. Free energy calculations in globular proteins: methods to reduce errors. J Comp Chem 1998;19:1229–1240.
47. Brown JH, Cohen C, Parry DAD. Heptad breaks in α-helical coiled-coils: stutters and stammers. Proteins 1996;26:134–145.
48. Ceruso MA, Amadei A, Di Nola A. Mechanics and dynamics of B1 domain of protein G: role of packing and surface hydrophobic residues. Protein Sci 1999;8:147–160.
49. Swindells MB, MacArthur MW, Thornton JM. Intrinsic Φ, Ψ propensities of amino acids, derived from the coil regions of known structures. Nat Struct Biol 1995;2:596–603.
50. Seo J, Cohen C. Pitch diversity in α-helical coiled coils. Proteins 1993;15:223–234.
51. McCammon JA, Harvey SC. Dynamics of proteins and nucleic acids. New York: Cambridge University Press; 1987.