
SHORT COMMUNICATION

Structural Polymorphism of a Marginally Stable 4- α -Helical Bundle. Images of a Trapped Molten Globule?

Nicholas M. Glykos and Michael Kokkindis*

IMBB, FORTH, Crete, Greece and Department of Biology, University of Crete, Crete, Greece

The general experience from structural studies of single amino-acid-substituted mutant proteins is that the effect of mutation is rather localized and minor. We have reported¹ an exception to this rule by showing that a single alanine to proline substitution at position 31 of the Rop protein is sufficient for changing the topology of this 4- α -helical bundle (Fig. 1), leading to drastic changes both in its surface properties and the packing of its hydrophobic core.

The original structure determination of this mutant was achieved through the analysis of an orthorhombic crystal form.² A few puzzling observations were made at that time concerning both the mutant structure itself and some of its functional properties. The first observation concerned the reduced number (and density) of hydrophobic core packing interactions. This was manifested and reinforced by (i) the destruction of the layered structure of the hydrophobic core as seen in the wild-type protein; and (ii) by the absence of interactions between hydrophobic residues obeying a certain sequence periodicity (known as heptad repeat) which characterizes associating α -helices. A related second observation concerned the large number of buried hydrophobic residues (approximately 30% of the total number of buried residues) for which the electron density maps suggested the presence of static or discrete disorder. The third was the presence of a large continuous cavity buried in the middle of the protein (and effectively separating the hydrophobic core in two discrete patches) with a volume of approximately 270 Å³. Lastly, the finding that this mutant retained some of the biological activity of the wild-type protein⁴ could not be reconciled with the large movements (of the order of approximately 4 Å) that would be required to bring some side-chains (like Phe14, known to be essential for RNA-binding^{4,5}) into a suitable (for RNA-binding) distance.

In the hope that we could address some of these questions, we pursued our attempts to determine the structure of a second (monoclinic) crystal form of A31P. It was only very recently that we succeeded in completing the crystallographic analysis of this second crystal form (and this only after we resorted to a computationally expensive 23-dimensional molecular replacement search³). As shown in Figure 1, the monoclinic A31P crystal structure retained the same topology and folding motif as its ortho-

rhombic counterpart, but everything else appears to be different: The relative juxtapositions of the helices deviate so much between the two structures, that when any one of them is used for superposition, none of the remaining three helices overlap. When the whole bundles are superimposed, the root mean square deviation (RMSD) between equivalent C $_{\alpha}$ atoms is approximately 3 Å. Not surprisingly, these differences are accompanied by changes in the mode of hydrophobic core packing and the bundles' surface properties (data not shown). This in spite of the fact that the only consistent difference between the crystallization conditions of the two crystal forms is an increase of the pH by half a unit for the monoclinic form.

The large differences between the two crystal forms (indicative of conformational polymorphism) together with (i) the presence of a mobile, loosely packed hydrophobic core; (ii) the presence of a large continuous (and probably solvent-exposed) hydrophobic cavity in the middle of the protein's interior; and (iii) the greatly reduced stability of this mutant⁶ (compared with the wild-type protein), led us to hypothesize that A31P at its equilibrium state may have molten-globule-like characteristics.^{7,8,9,10,11} Clearly, such a proposition can not be supported solely by evidence obtained through the analysis of two static images, especially when these images have been determined from crystal structures with completely different packing environments. For this reason, and in an attempt to obtain evidence concerning the behavior of this mutant in solution and to further characterize its equilibrium properties, a 3-ns molecular dynamics simulation of A31P in explicit water was performed as described in the legend to Figure 2. Furthermore, and in order to have a measure of comparison for the analyses of the molecular dynamics results, similar 3-ns simulations were performed on wild-type Rop

Nicholas M. Glykos's present address is Department of Molecular Biology & Genetics, Democritus University of Thrace, Dimitras 19, 68100 Alexandroupolis, Greece.

*Correspondence to: Michael Kokkindis, IMBB, FORTH, PO Box 1527, 71110 Heraklion, Crete, Greece or Department of Biology, University of Crete, P.O. Box 2208, 71409 Heraklion, Crete, Greece. E-mail: kokkinid@imbb.forth.gr

Received 18 November 2003; Accepted 19 February 2004

Published online 14 May 2004 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.20167

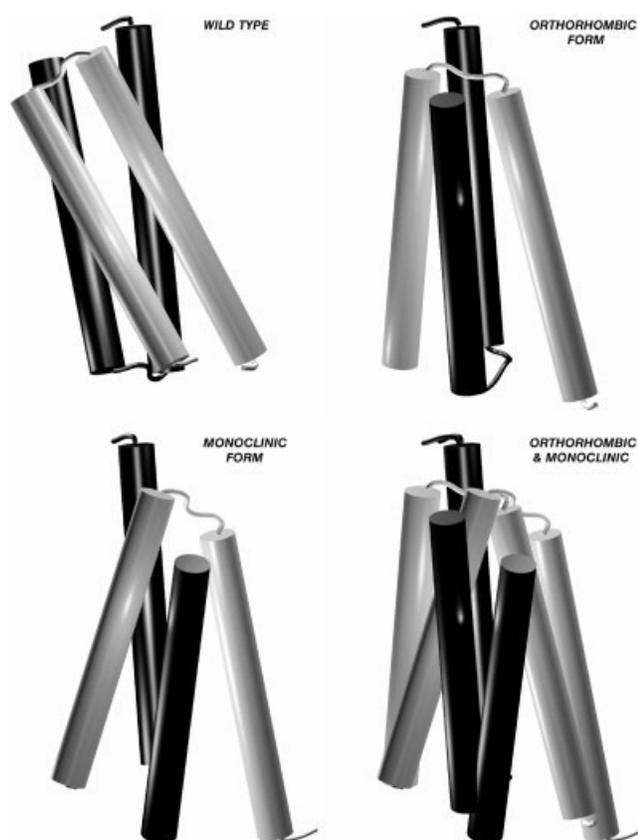


Fig. 1. Comparison between the crystal structures of native Rop (upper-left), orthorhombic A31P (upper-right), monoclinic A31P (lower-left) and a superposition of the two A31P structures (lower-left). In all diagrams α -helices are depicted as cylinders, the connecting loops as thin tubes, and the two monomers (of each bundle) are coded with different shades of gray. All structures shown in this figure have been oriented in such a way as to align (both in position and orientation) the helix lying furthest away from the viewer. The structures were aligned with the program LSQKAB²² from the CCP4 suite of programs,²³ and the figure prepared with the programs VMD²⁴ and Raster3D.²⁵

and on a hypothetical A31P structure which was constructed by keeping the wild-type Rop crystal structure fixed and artificially (computationally) replacing the alanine at position 31 with a proline. The simulation of this hypothetical structure would also allow us to confirm (or otherwise) the paradoxical finding—based mostly on homology-modeling studies—that a native-like structure for the A31P mutant could have been as good a structure as for any of the numerous Rop mutant structures previously characterised.^{12,13,14,15,16} In the following paragraphs we discuss and compare the results obtained from these three molecular dynamics simulations.

Probably the most notable characteristic of the A31P molecular dynamics trajectory (few representative instances of which are shown in Figure 2) concerns the malleability of the mutant structure. Although the simulation was initiated from the orthorhombic form crystal structure (having a C_{α} RMSD of 2.9 Å from the monoclinic form), it twice converged to structures rather similar to the monoclinic form crystal structure (see instances recorded at $t = 0.72$ and $t = 1.52$ ns, with corresponding C_{α} RMSD of

1.68 and 1.79 Å). This is shown more clearly in Figure 3(b) in which the evolution of the C_{α} RMSD between the monoclinic crystal structure and the simulation-recorded structures is depicted. The simulation also sampled conformations characteristically different from both crystal forms (e.g., instances at $t = 1.36$, 2.00, and 3.05 ns in Figure 2). Indeed, a 2D-RMSD-based cluster analysis¹⁷ of the A31P trajectory which was performed with a cutoff of 1.5 Å (sufficient to guarantee that both the wild-type and hypothetical A31P simulations would be represented by just one cluster) revealed five distinct clusters, only two of which were closely related to the crystallographically

Fig. 2. Snapshots recorded from the A31P molecular dynamics trajectory at the indicated times. The cartoon (cylinder) representations correspond to the monoclinic form crystal structure whose orientation and position is kept fixed throughout this figure. The trajectory-recorded structures are depicted as tubes (with different shades of gray used for each monomer) and are oriented in such a way as to align (both in position and orientation) the helix lying furthest away from the viewer with the equivalent helix from the monoclinic form crystal structure. The initial ($t = 0$) trajectory structure is the orthorhombic form A31P crystal structure. Superpositions performed with the program X-PLOR²⁶ and figure prepared with the programs VMD²⁴ and Raster3D.²⁵

METHODS. Orthorhombic A31P simulation: Starting from the crystallographically determined coordinates of the orthorhombic form of A31P (entry 1b6q.pdb), missing side-chain and hydrogen atoms were built with X-PLOR.²⁶ A31P was solvated in an orthogonal box of pre-equilibrated TIP3 water²⁷ with dimensions $75.4 \times 56.6 \times 56.6$ Å³. All water molecules lying closer than 2.0 Å from the protein surface were removed. The final system comprised 1814 protein and 22,245 water atoms, with a total net charge of $-1.32e$. All other calculations were performed with the program NAMD²⁸ using the CHARMM22 force field²⁹ as follows. The system was first energy minimized for 500 conjugate gradient steps, and then equilibrated for 40 ps under NpT conditions with a stepwise increase ($\Delta = 10$ K) of the temperature from 10 to 320 K over a period of 10 ps. The pressure was maintained at 1 atm by weak coupling to an external bath³⁰ with a pressure coupling relaxation time of 0.5 ps⁻¹. At the end of the equilibration the dimensions of the simulation box were $75.2 \times 56.5 \times 56.5$ Å³. A NpT molecular dynamics simulation was then performed for 3 ns using Berendsen's method³⁰ and the temperature reassignment algorithm to maintain the pressure and temperature at 1 atm and 320 K respectively. The time step was 2 fs, periodic boundary conditions were imposed, long-range full electrostatics interactions were evaluated every four timesteps with the particle mesh Ewald method,³¹ a cutoff for the van der Waals interactions was applied through a switching function acting between 10 and 12 Å, 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.

Hypothetical A31P simulation: Starting from the crystallographically determined coordinates of wild-type Rop (entry 1rop.pdb), the alanine at position 31 was replaced by a proline with the program Xfit from the XtalView suite of programs.³² Missing side-chain and hydrogens atoms were built with X-PLOR and the resulting structure solvated (using VMD) in an orthogonal box of TIP3 water with dimensions $78.6 \times 60.1 \times 56.8$ Å³. The final system comprised 1814 protein and 23,181 water atoms. All other calculations were performed as described above with the exception that the temperature and pressure were controlled using the Nosé-Hoover Langevin dynamics and Langevin piston barostat control methods as implemented by the NAMD program (and maintained, as previously, at 320 K and 1 atm).

Wild-type Rop simulation: Starting from the crystallographically determined coordinates of wild-type Rop (entry 1rop.pdb), missing side-chain and hydrogen atoms were built with X-PLOR and the resulting structure solvated (using VMD) in an orthogonal box of TIP3 water with dimensions $78.6 \times 60.2 \times 57.8$ Å³. The final system comprised 1810 protein and 24,021 water atoms. All other calculations were performed as described for the hypothetical A31P simulation.

Unless otherwise stated, all further analyses of the molecular dynamics data were performed with X-PLOR,²⁶ VMD²⁴ and locally written software (program *carma*, available via <http://origin.imbb.forth.gr/~glykos/>).

determined structures. To reinforce the notion of the mutant structure's mobility, we compare in Figure 3(a) and (c) the behavior of the A31P and wild-type simulations versus the starting structures and the trajectory-average

structures respectively. Wild-type Rop appears to be as stable and its structure as well-preserved as could be expected on the basis of its thermodynamic stability^{6,14} and the results from the reported NMR experiments.¹⁸ As

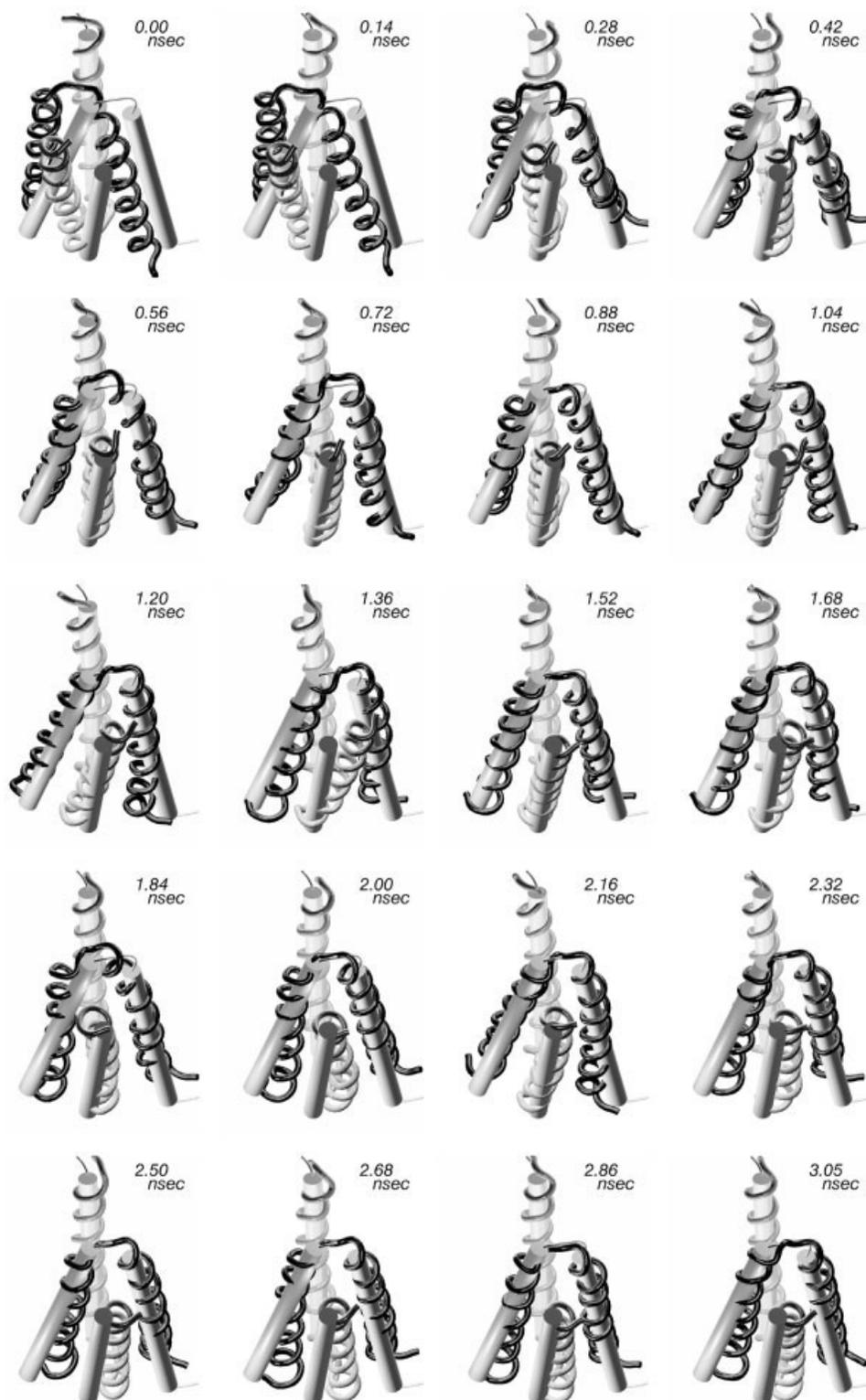


Figure 2.

shown in Figure 3(d), the stability of the wild-type structure is uniformly distributed throughout the molecule with (expected) deviations at the N- and C-termini of the two polypeptide chains. In sharp contrast, the A31P simulation deviates markedly from the starting structure, and shows no sign of convergence towards a—possibly different from the starting crystal structure—trajectory-average conformation. As indicated by Figure 3(d), the mobility of the A31P structure is not the result of localized hot-spots (for example at the termini) but is again distributed throughout the whole molecule. These results are illustrated more clearly in Figures 3(e) and 3(f) which compare the trajectory-average C_{α} - C_{α} distance maps and the corre-

sponding RMS deviations from them for the wild-type (above the diagonals) and A31P (below the diagonals) simulations. The dense, well-packed hydrophobic core of the wild-type structure [Fig. 3(e)], leads to a uniform distribution of low RMS deviations not only within the individual helices, but more importantly, between them [Fig. 3(f)]. The only clear signs of higher than average mobility correspond to the termini of the two polypeptide chains. This is clearly not the case with mutant whose only stable elements appear to be the individual helices per se (but not their relative juxtaposition in the A31P structure). The behavior of the A31P simulation suggested that the 3-ns interval may have been too short to adequately sample its conformational space. To confirm that this is indeed the case, a principal component analysis was performed on the two simulations as previously described.^{19,20,21} Figure 3(g) and (h) compares the projections of the C_{α} motion of the A31P and wild-type simulations on the planes of the three eigenvectors corresponding to the three largest eigenvalues: whereas for the wild-type simulation the 3-ns interval appears to have been sufficient to adequately sample its conformational space, this is definitely not the case with the A31P simulation. This finding suggests that a longer simulation length for A31P would have allowed us to observe an even greater structural polymorphism.

In Figure 4 we compare the behavior of the A31P and wild-type simulations versus a 3-ns simulation of a hypothetical (wild-type-like) A31P structure, which was constructed as described in the legend to Figure 2. The C_{α} RMS deviation of this hypothetical mutant from its starting structure [shown in Fig. 4(a)] shows an evolution that appears to be intermediate between the wild-type and the true A31P structure. But as Figure 4(b) indicates, the deviation from the starting structure is not due to the mobility or structural instability of this hypothetical molecule. Rather, it is due to a genuine structural transition toward a different (from the starting structure) trajectory-average conformation. This transition mainly corresponds

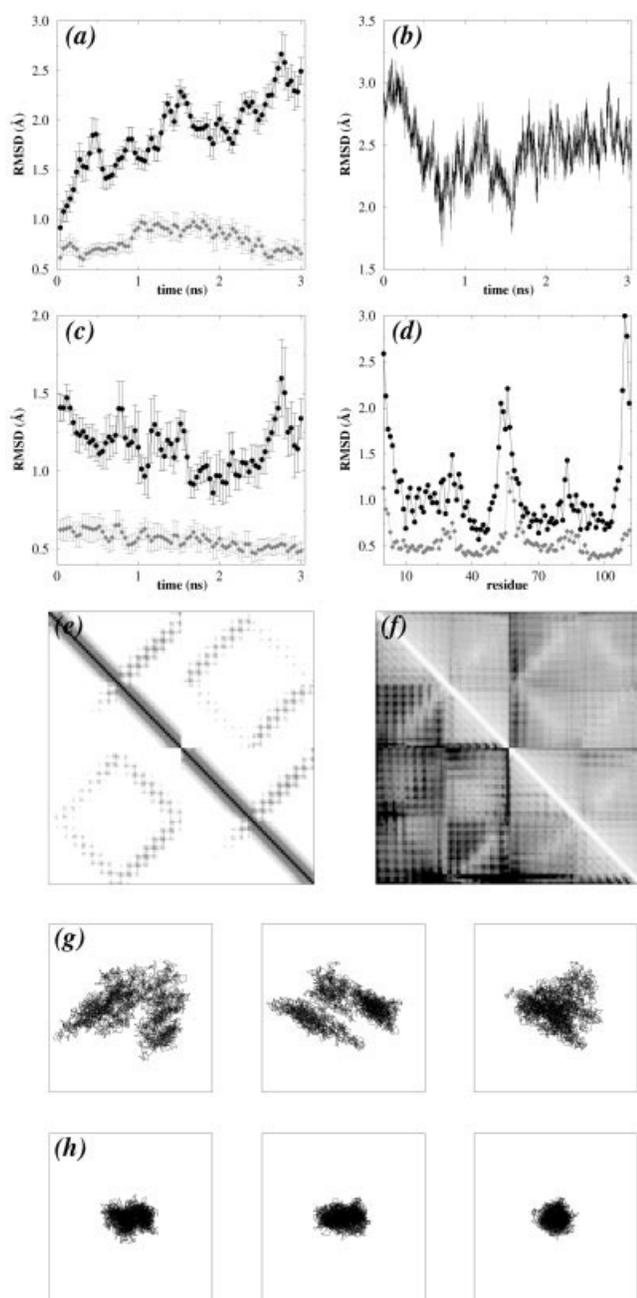


Fig. 3. **a:** C_{α} RMS deviations from the starting structures versus time for the A31P (upper curve) and wild-type Rop (lower curve) simulations. Individual points represent running averages and standard deviations in a 80-ps window. **b:** C_{α} RMSD from the monoclinic form crystal structure versus time for the A31P simulation. **c:** C_{α} RMS deviations from the average structures versus time for the A31P (upper curve) and wild-type Rop (lower curve) simulations. **d:** C_{α} RMS deviations from the average structures versus residue number (both chains included) for the A31P (upper curve) and wild-type Rop (lower curve). **e:** Trajectory-averaged C_{α} - C_{α} distance map for the wild-type Rop (above the diagonal) and A31P (below the diagonal). A linear gray scale gradient is used to depict average C_{α} - C_{α} distances between zero (black) and 12 Å (white). **f:** Corresponding RMS deviations (from the average) of the C_{α} - C_{α} distances for wild-type Rop (above the diagonal) and A31P (below the diagonal). A linear gray scale gradient is used to depict RMS deviations between zero (white) and 1.5 Å (black). **g:** Projection of the C_{α} motion of the A31P simulation on the planes of the three eigenvectors corresponding to the three largest eigenvalues. From left to right the 1-2, 1-3, and 2-3 eigenvector planes are shown. In all three diagrams the values on all axes range from -20 to 20 Å. **h:** Projection of the C_{α} motion of the wild-type Rop simulation on the planes of the three eigenvectors corresponding to the three largest eigenvalues. From left to right the 1-2, 1-3, and 2-3 eigenvector planes are shown. The graphs are on the same scale as for panel (g).

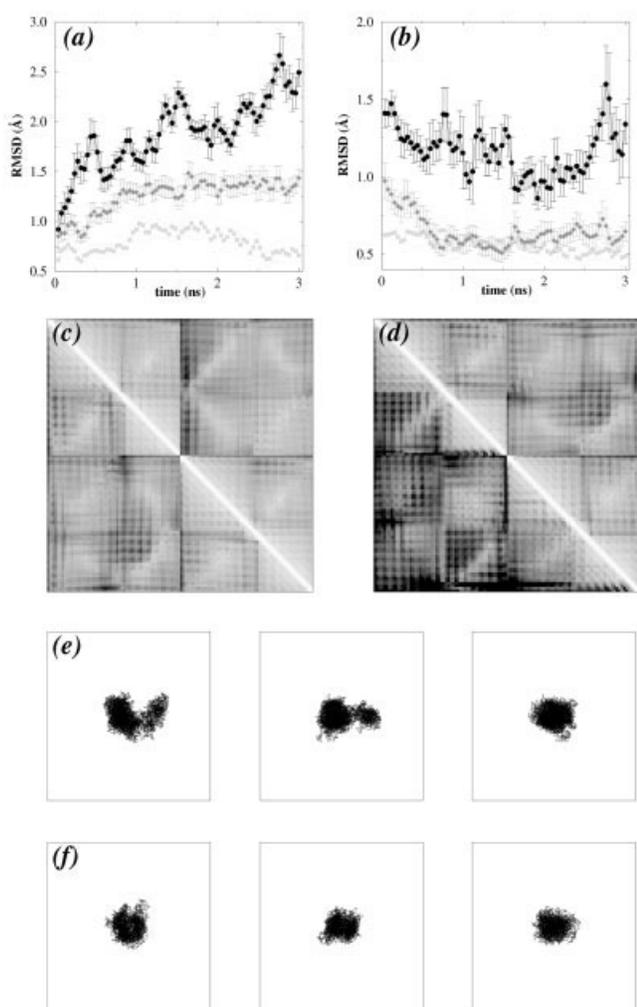


Fig. 4. **a:** C_{α} RMS deviations from the starting structures versus time for the A31P (upper curve), hypothetical A31P (middle curve) and wild-type Rop (lower curve) simulations. **b:** C_{α} RMS deviations from the average structures versus time for the A31P (upper curve) and hypothetical A31P (lower curve) simulations. The graph for the wild-type Rop is also included but mostly overlaps with the hypothetical A31P curve. **c:** RMS deviations of the C_{α} - C_{α} distances (from the trajectory-average) for wild-type Rop (above the diagonal) and hypothetical A31P (below the diagonal). **d:** RMS deviations of the C_{α} - C_{α} distances (from the trajectory-average) for hypothetical A31P (above the diagonal) and A31P (below the diagonal). **e:** Projection of the C_{α} motion of the hypothetical A31P simulation on the planes of the three eigenvectors corresponding to the three largest eigenvalues. From left to right the 1-2, 1-3, and 2-3 eigenvector planes are shown. **f:** Same as (e), but only using the last 2 ns of the hypothetical A31P simulation for the calculation. (c) through (f) are on the same scale as the equivalent diagrams from Figure 3.

to a reorganization (of limited extent) of the turn region to absorb the effects of the induced mutation. This is shown more clearly in Figure 4(c) which compares the RMS deviations of the C_{α} - C_{α} distances (from their trajectory-average) for the hypothetical A31P (below the diagonal) and wild-type (above the diagonal) simulations: the main differences between the two maps arise from (and correspond to) a movement of the turn regions with respect to the helices in the case of the hypothetical A31P simulation. A similar comparison [Fig. 4(d)] between the true A31P

(below the diagonal) and the hypothetical A31P (above the diagonal) simulations clearly shows that hypothetical A31P appears to be far more stable than the experimentally determined A31P (and nearly as stable as wild-type Rop). Figure 4(e) and (f) compares the results from a principal component analysis of the hypothetical A31P simulation which was performed using either the whole trajectory [Fig. 4(e)], or, only the last 2 ns of it [Fig. 4(f)]. The comparison shows that following the turn's structural rearrangement, which occurs within the first nanosecond of the simulation [Fig. 4(b)], the conformation of the hypothetical A31P structure remains relatively stable and its dynamics appear to have been adequately sampled.

The results presented above are corroborated and strengthened by an analysis of the average side-chain mobility in the hydrophobic cores of the three structures. For this analysis we have chosen 22 side-chains (mainly leucines and isoleucines) that take part in the formation of the hydrophobic core of all three structures and calculated the RMS deviation from their average positions for the whole length of the three dynamics simulations (excluding from the analysis all main-chain, C_{β} , and hydrogen atoms of the respective residues). For the wild-type simulation we found an average RMS deviation of 0.62 ± 0.14 Å, whereas for the A31P simulation we obtained an average deviation of 1.03 ± 0.16 Å. The hypothetical A31P structure gave a value of 0.70 ± 0.14 Å which is reduced to 0.61 ± 0.11 Å if only the last 2 ns of the simulation are considered.

In summary, the molecular dynamics studies outlined above together with the crystallographic evidence discussed in the introduction are consistent with the hypothesis that A31P is a mutation-induced molten globule. This proposition is also consistent with the thermodynamic studies of the mutant,⁶ and especially with its greatly reduced stability ($\Delta\Delta G = 29$ KJ/(mol dimer) at 25°C), the reduced ratio of the ellipticity values $[\Theta_{222nm}]/[\Theta_{208nm}]$, and its reduced transition enthalpy.⁶ Although the absence of aromatic side-chains from the hydrophobic core of Rop precludes a confident assignment of the molten globule state (through a comparison of the near and far UV spectra), the crystallographic evidence for the presence of disordered side-chains in the hydrophobic core and the behavior of its molecular dynamics simulation leaves little doubt that the mutant's hydrophobic core is molten-globule-like in character. The apparent stability of a hypothetical wild-type-like A31P structure indicates that the experimentally determined A31P structure may correspond to kinetically trapped molten globule. If this proposition is accepted, then the Rop A31P mutant is one of the best structurally characterized molten globules known to-date.

REFERENCES

1. Glykos NM, Cesareni G, Kokkinidis M. Protein plasticity to the extreme: changing the topology of a 4- α -helical bundle with a single amino acid substitution. *Structure* 1999;7:597-603.
2. Glykos NM, Kokkinidis M. Meaningful refinement of polyaniline models using rigid-body simulated annealing: application to the structure determination of the A31P Rop mutant. *Acta Crystallogr D Biol Crystallogr* 1999;55:1301-1308.
3. Glykos NM, Kokkinidis M. Structure determination of a small

- protein through a 23-dimensional molecular replacement search. *Acta Crystallogr D Biol Crystallogr* 2003; 59:709–718.
4. Castagnoli L, Scarpa M, Kokkinidis M, Banner DW, Tsernoglou D, Cesareni G. Genetic and structural analysis of the ColE1 Rop (Rom) protein. *EMBO J* 1989;8:621–629.
 5. Predki PF, Nayak LM, Gottlieb MBC, Regan L. Dissecting RNA-Protein Interactions: RNA–RNA Recognition by Rop. *Cell* 1995;80: 41–50.
 6. Peters K, Hinz H-J, Cesareni G. Introduction of a proline residue into position 31 of the loop of the dimeric 4- α -helical protein ROP causes a drastic destabilization. *Biol Chem* 1997;378:1141–1152.
 7. Ptitsyn OB. Molten globule and protein folding. *Adv Protein Chem* 1995;47:83–229.
 8. Chamberlain AK, Marqusee S. Comparison of equilibrium and kinetic approaches for determining protein folding mechanisms. *Adv Protein Chem* 2000;53:283–328.
 9. Arai M, Kuwajima K. Role of the molten globule state in protein folding. *Adv Protein Chem* 2000;53:283–328.
 10. Betz SF, Raleigh DP, DeGrado WF, Lovejoy B, Anderson D, Ogihara N, Eisenberg D. Crystallization of a designed peptide from a molten globule ensemble. *Fold Des* 1996;1:57–64.
 11. Hill CP, Anderson DH, Wesson L, DeGrado WF, Eisenberg D. Crystal structure of α_1 : implications for protein design. *Science* 1990;249:543–546.
 12. Vlasi M, Steif C, Weber P, Tsernoglou D, Wilson KS, Hinz HJ, Kokkinidis M. Restored heptad pattern continuity does not alter the folding of a four- α -helix bundle. *Nature Struct Biol* 1994;1:706–716.
 13. Castagnoli L, Vetriani C, Cesareni G. Genetic and structural analysis of the ColE1 Rop (Rom) protein. *J Mol Biol* 1994;237:378–387.
 14. Steif C, Hinz H-J, Cesareni G. Effects of cavity-creating mutations on conformational stability and structure of the dimeric 4- α -helical protein ROP: Thermal unfolding studies. *Proteins* 1995;23: 83–96.
 15. Predki PF, Agrawal V, Brünger AT, Regan L. Amino-acid substitutions in a surface turn modulate protein stability. *Nature Struct Biol* 1996;3:54–58.
 16. Munson M, Anderson KS, Regan L. Speeding up protein folding: mutations that increase the rate at which Rop folds and unfolds by over four orders of magnitude. *Fold Des* 1997;2:77–87.
 17. Simmerling C, Elber R, Zhang J. MOIL-View—a program for visualization of structure and dynamics of biomolecules and STO—a program for computing stochastic paths. In: Pullman A, et al., editors. *Modelling of biomolecular structure and mechanisms*. Netherlands: Kluwer; 1995. p 241–265.
 18. Eberle W, Pastore A, Sander C, Rosch P. The structure of ColE1 rop in solution. *J Biomol NMR* 1991;1:71–82.
 19. Ichiye T, Karplus M. Collective motions in proteins: a covariance analysis of atomic fluctuations in molecular dynamics and normal mode simulations. *Proteins* 1991;11:205–217.
 20. Amadei A, Linssen ABM, Berendsen HJC. Essential dynamics of proteins. *Proteins* 1993;17:412–425.
 21. Kitao A, Go N. Investigating protein dynamics in collective coordinate space. *Curr Opin Struct Biol* 1999;9:164–169.
 22. Kabsch W. A solution for the best rotation to relate two sets of vectors. *Acta Crystallogr A* 1976;32:922–923.
 23. Collaborative Computational Project, Number 4. The CCP4 Suite: Programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr* 1994;50:760–763.
 24. Humphrey W, Dalke A, Schulten K. VMD—visual molecular dynamics. *J Mol Graph* 1996;14:33–38.
 25. Merritt EA, Bacon DJ. Raster3D—photorealistic molecular graphics. *Methods Enzymol* 1997;277:505–524.
 26. Brünger AT. X-PLOR Version 3.1. A system for X-ray crystallography and NMR. Connecticut: Yale University Press; 1992.
 27. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein, ML. Comparison of simple potential functions for simulating liquid water. *J Chem Phys* 1983;79:926–935.
 28. Kale L, Skeel R, Bhandarkar M, Brunner R, Gursoy A, Krawetz N, Phillips J, Shinozaki A, Varadarajan K, Schulten K. NAMD2: Greater scalability for parallel molecular dynamics. *J Comput Phys* 1999;151:283–312.
 29. MacKerell AD, Bashford D, Bellott M, Dunbrack RL, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczera K, Lau FTK, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Reiher WE, Roux B, Schlenkrich M, Smith C, Stote R, Straub J, Watanabe M, Wiorkiewicz-Kuczera J, Yin D, Karplus M. All-atom empirical potential for molecular modeling and dynamics Studies of proteins. *J Phys Chem Series B* 1998;102:3586–3616.
 30. Berendsen HJC, Postma JPM, van Gunsteren WF, Dinola A, Haak JR. Molecular dynamics with coupling to an external bath. *J Chem Phys* 1984;81:3684–3690.
 31. Darden T, York D, Pedersen L. Particle mesh Ewald. An $N \log(N)$ method for Ewald sums in large systems. *J Chem Phys* 1993;98: 10089–10092.
 32. McRee DE. A visual protein crystallographic software system for X11/XView. *J Mol Graph* 1992;10:44–46.