Online journal ISSN - 1676-5680
GMR Genetics and Molecular Research
EVOLUTION AND TECHNOLOGY

# *Ab initio* 3-D structure prediction of an artificially designed three-α-helix bundle via all-atom molecular dynamics simulations

**A. Breda[1,2], D.S. Santos[2,3], L.A. Basso[2,3] and O. Norberto de Souza[1-4]**

[1]Laboratório de Bioinformática, Modelagem e Simulação de Biossistemas, Faculdade de Informática, PUCRS, Porto Alegre, RS, Brasil
[2]Programa de Pós-Graduação em Biologia Celular e Molecular, PUCRS, Porto Alegre, RS, Brasil
[3]Centro de Pesquisa em Biologia Molecular e Funcional, TECNOPUC, Porto Alegre, RS, Brasil
[4]Programa de Pós-Graduação em Ciência da Computação, PUCRS, Porto Alegre, RS, Brasil
Corresponding author: O. Norberto de Souza
E-mail: osmar.norberto@pucrs.br

Genet. Mol. Res. 6 (4): 901-910 (2007)
Received August 03, 2007
Accepted September 25, 2007
Published October 05, 2007

**ABSTRACT.** The rate at which knowledge about genomic sequences and their protein products is produced is increasing much faster than the rate of 3-dimensional protein structure determination by experimental methods, such as X-ray diffraction and nuclear magnetic resonance. One of the major challenges in structural bioinformatics is the conversion of genomic sequences into useful information, such as characterization of protein structure and function. Using molecular dynamics (MD) simulations, we predicted the 3-dimensional structure of an artificially designed three-α-helix bundle, called A3, from a fully extended initial conforma-

 

tion, based on its amino acid sequence. The MD protocol enabled us to obtain the secondary, in 1.0 ns, as well as the supersecondary and tertiary structures, in 4.0-10.0 ns, of A3, much faster than previously described for a similar protein system. The structure obtained at the end of the 10.0-ns MD simulation was topologically a three-α-helix bundle.

**Key words:** *Ab initio* prediction, Three-α-helix bundle, Molecular dynamics simulations, Protein 3-D structure

## INTRODUCTION

The large increase in both genomic and molecular biology data observed in the 1990's revealed that the number of unidentified genes, and their protein products, was greater than had been expected (Norin and Sundström, 2002). One of today's major challenges in the post-genomic era is to convert the data obtained from genome sequencing into useful information, such as the translation of nucleotide sequences into known genes and, ultimately, to proteins with known structure and function.

Once established by Anfinsen et al. (1961) that the linear sequence of amino acids alone contains all the information required for a protein to fold, computer models (including all atoms and their interactions) that reliably reproduce these characteristics can be used to study protein behavior by molecular dynamics (MD) simulations (Hansson et al., 2002; Karplus and McCammon, 2002).

Several research laboratories have been searching for the answer to Levinthal's paradox (Levinthal, 1969) concerning the principles that govern protein tertiary structure assignment (Bradley et al., 2005), but few of them have used *ab initio* MD simulation methods. Among those few, the focus has been on protein folding pathways (Berriz and Shakhnovich, 2001; Simmerling et al., 2002; Chowdhury et al., 2003).

We have been developing MD simulation protocols that will enable correct prediction of any polypeptide or protein three-dimensional (3-D) structure based on its amino acid sequence, which is generally called *ab initio* structure prediction (Sternberg et al., 1999; Bonneau et al., 2001; Hardin et al., 2002). The prediction should be made fast enough compared to the time scale in which the motions involved in protein 3-D structure formation occur *in vitro* and *in vivo* (Clarke et al., 1999). Initially, we are not interested in how the protein reaches its 3-D structure during the simulation, but only in its final, correct 3-D structure. If the prediction works, given the deterministic nature of the MD method (Karplus and McCammon, 2002), we can return to the beginning of the MD trajectory and perform further analysis in order to learn about the mechanisms by which that particular polypeptide or protein achieved its native structure. Knowledge concerning such mechanisms could have a key role in our understanding of sequence-structure-dynamics-function relationships, and would be particularly useful for obtaining the 3-D structure of proteins that have no homologues in protein families with known structures.

In this article, we report *ab initio* MD simulations started from a fully extended conformation of an artificially designed three-α-helix bundle (Johansson et al., 1998). The three-α-helix bundle is a common structural domain often found in many soluble proteins, including spectrin (Yan

et al., 1993) and the extramembraneous portion of *Staphylococcus aureus* protein A (Starovasnik et al., 1996).

## MATERIAL AND METHODS

### Model

We report an *ab initio* structure prediction by MD simulations of a 65-residue (Figure 1) three-α-helix bundle artificially designed by Johansson et al. (1998) to adopt this specific conformation. Based on the side-chain packing diagram of Figure 1 from this publication, we manually designed an ideal model of this three-α-helix bundle (Figure 2), named herein A3m, using the Swiss-PdbViewer program (Guex and Peitsch, 1997). A3m has all the proposed favored side-chain interactions (not shown for clarity) and was considered the reference structure to compare the results from the MD simulations, since there was no available experimental structure for this polypeptide. Furthermore, since a three-α-helix bundle can also adopt a counterclockwise arrangement (Johansson et al., 1998), we built a second A3m model, as described for the first,


**Figure 1.** The 65-amino acid sequence of A3m, from N-terminus (left) to C-terminus (right), and its secondary structure according to Johansson et al. (1998). Helices are represented as colored boxes (helix I in blue, helix II in green and helix III in red) and the glycine-rich turns as gray boxes.

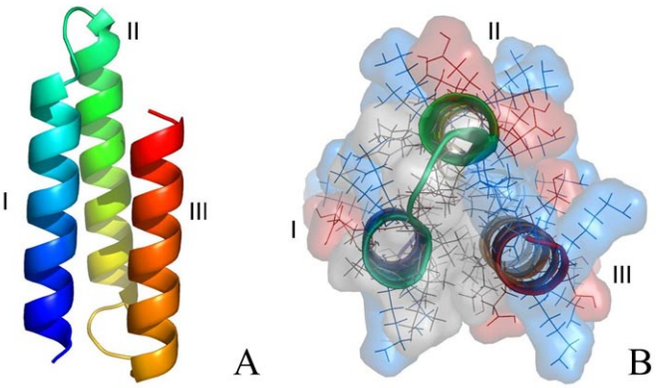


**Figure 2. A.** A manually designed ideal model for the three-α-helix bundle A3m based on the diagram in Figure 1 from Johansson et al. (1998), represented as ribbons and colored from blue (N-terminus) to red (C-terminus). **B.** A top down view, with the same orientation as Johansson's diagram. Helix I is in blue, helix II in green and helix III in red. Side chains and the molecular surface are colored blue for basic, red for acidic, and gray for non-polar amino acid side chains, respectively. The clockwise orientation of helix I through helix III can be easily observed. Helices I and III are parallel to each other and both are antiparallel to helix II.

but with helix I through helix III running in a counterclockwise orientation for further analyses. The neutral forms of the amino acids lysine and glutamate were used to reduce the total electric charge of the system.

### Molecular dynamics simulations

We started with a fully extended chain - named herein A3, with all φ (phi) and ψ (psi) dihedral angles equal to 180°. The A3 topology was built from its primary sequence with the Link, Edit and Parm modules of AMBER 4.1 (Pearlman et al., 1995). Energy minimization and MD simulations were performed with the SANDER module of Amber 6.0 (Case et al., 1999) using the all-atom force field (Cornell et al., 1995). The polypeptide was energy minimized for 500 steps in order to relax any possible strains introduced during model construction. After that, it was submitted to an initial 10-ns MD simulation, later extended to 50 ns, in an NT ensemble, at an average temperature of 298.16 K and a 6.0-10.0 Å cut-off radius for the evaluation of the long-range van der Waals and electrostatic interactions. Starting at 6.0 Å, the cut-off radius increased by 1.0 Å every 1.0 ns of MD simulation until it reached 10.0 Å at 5.0 ns, where it remained until the end of the simulation. The solvent was treated implicitly within the generalized born with surface area approximation (Bashford and Case, 2000). The SHAKE algorithm (Ryckaert et al., 1977) was used to restrain all hydrogen-heavy atom bond distances, allowing an integration time-step of 0.002 ps for the equations of motion. The simulation was performed on a PC cluster with 16 CPUs and the atomic positions were saved at every 500 steps (1.0 ps). A total of 50,000 snapshots were used for analysis.

### Structural analysis

The convergence of the simulation was monitored based on the RMSD (root mean-squared deviation) of the A3 MD trajectory with respect to the ideal three-α-helix bundle A3m. The RMSD is a measure of how similar or dissimilar two structures are and was calculated with the p-traj utility distributed with AMBER 6.0 (Case et al., 1999). Formation of secondary, supersecondary and tertiary structures were visualized with the Swiss-PdbViewer (Guex and Peitsch, 1997) and VMD (Humphrey et al., 1996) graphic packages. Stereochemical analysis and secondary structure calculations were performed with PROCHECK (Laskowski et al., 1993). The solvent accessible surface area (SASA) of residue tryptophan at position 32 (Trp32) was calculated with NACCESS (Hubbard and Thornton, 1993), using a probe radius of 1.4 Å for the water molecule. All structure illustrations were prepared with Pymol (DeLano, 2002).

## RESULTS AND DISCUSSION

The secondary, supersecondary and tertiary structural evolution of A3 during its dynamic trajectory is illustrated in Figure 3. After starting from its fully extended conformation at the initial simulation time (0 ns, not shown), helices I and III were basically formed at around 1.0 ns, with helix II only partially structured at its center. This is exceptionally fast, since α-helix formation *in vitro* has been reported to only occur in a millisecond time scale (Clarke et al., 1999). At 2.0 ns, helix II had extended along its length and all

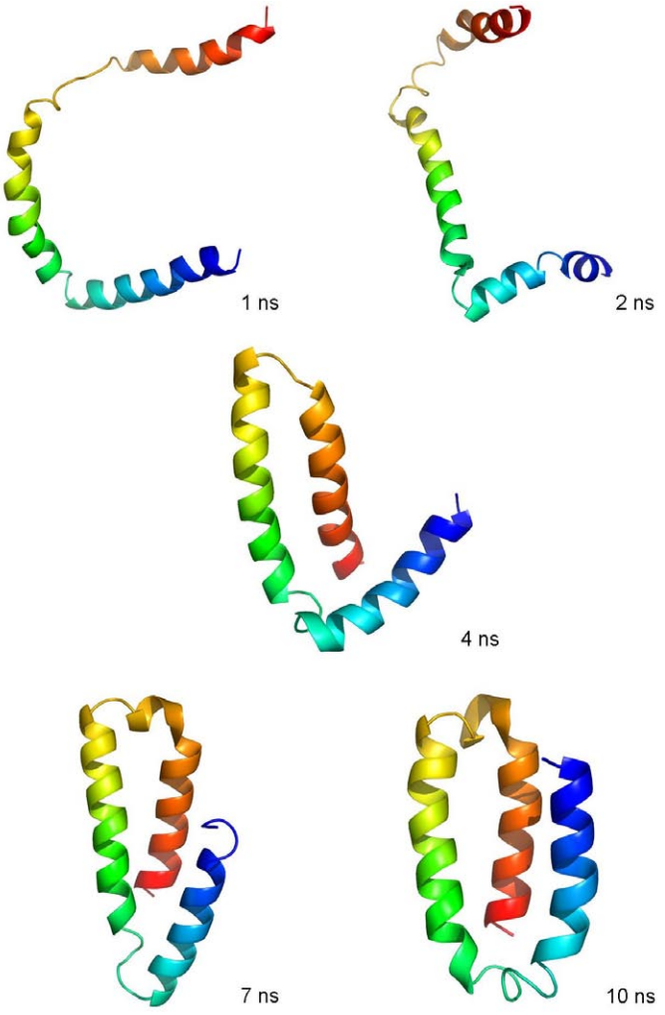


**Figure 3.** Snapshots from the initial 10.0-ns MD trajectory of A3 represented as a ribbon model. *1.0 ns* and *2.0 ns*, formation of secondary structure. Each of the three helices can already be distinguished, as well as the glycine-rich turns. *4.0 ns*, helices II (green) and III (red) begin to pack against each other. *7.0 ns*, once the contacts between helices II and III are formed, helix I packs against them. *10.0 ns*, at the end of the initial simulation, the tertiary structure of A3, characteristic of a three-α-helix-bundle, can be observed.

three helices were almost fully formed. Formation and rupture of helix units were mainly due to the inherent flexibility of the system imposed by the glycine-rich turns. This was more or less expected, as packing of side chains had not yet occurred at this time; hence the helices were free to move.

At 4.0 ns, helices II and III appeared packing against each other, with helix I following soon after at 7.0 and 10.0 ns (Figure 3). A similar behavior was described by Berriz and Shakhnovich (2001) and Bottomley et al. (1994) from analysis of the B domain of *S. aureus* protein A. They suggested a possible intermediate in the folding process of three-α-helix-bun-

dles, comprising helices II and III, with relatively high stability and functioning as a base for attachment of helix I. At the end of a 10.0-ns MD simulation, A3 adopted the topology of a three-α-helix bundle, with an RMSD value of approximately 7.5 Å (Figure 4) with respect to

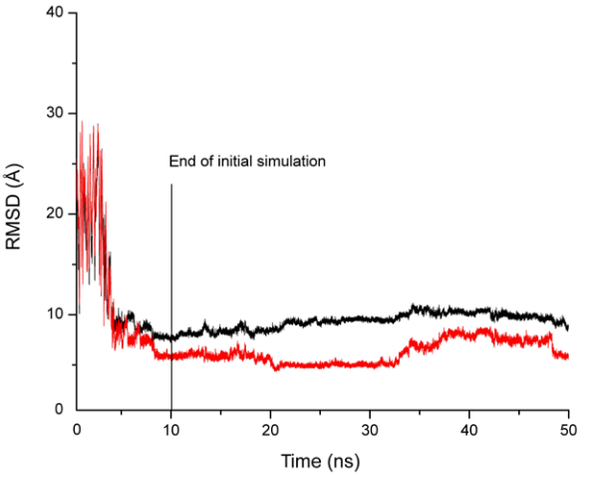


**Figure 4.** Root mean-squared deviation (RMSD) plot for the dynamic trajectory of A3 compared with A3m in the clockwise (black line) and counterclockwise (red line) orientations. At the end of the initial 10.0-ns simulation the RMSD values were 7.5 Å and 5.9 Å for the clockwise and counterclockwise model of A3m. The simulation was extended to 50.0 ns; despite the variations in RMSD, the final values did not differ significantly from those at 10.0 ns (9.0 Å and 5.8 Å, respectively).

the ideal, A3m conformation. Although we consider this a high RMSD value, it was close to the values obtained by participants of the last CASP encounters who used *ab initio* prediction methods, where the best prediction results of correct models had RMSD values of Cα atoms varying from 4.0 to 8.0 Å (Baker and Săli, 2001; Bradley et al., 2005; Moult, 2005).

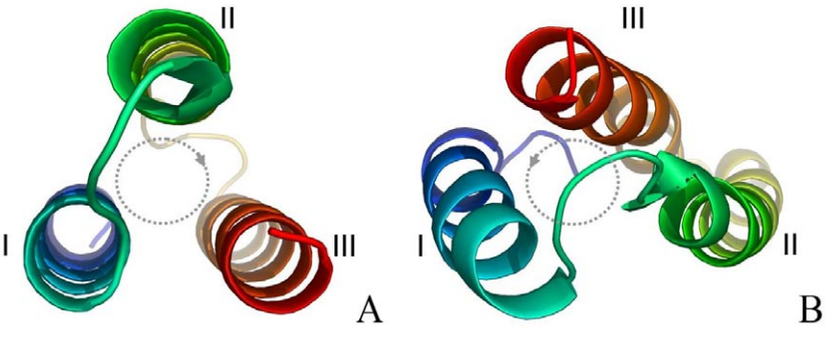


**Figure 5.** A top-down view of the three-α-helix bundle. **A.** The A3m model as in Figure 2. Starting from helix I (blue), passing through helix II (green) to helix III (red), this polypeptide has a clockwise orientation. **B.** A3, with the same view as in *A*, at the end of a 10.0-ns MD simulation; helix I through helix III runs in a counterclockwise orientation. The orientations in *A* and *B* are indicated by the broken arrowed circles.

Despite the promising results we describe above, the A3 orientation observed at the end of the 10.0-ns MD simulation was not the one expected based on the studies by Johansson et al. (1998), who concluded that the clockwise orientation was more favorable. On the contrary, at the end of our initial 10.0-ns MD simulation we obtained an A3 topology with a counterclockwise orientation (Figure 5).

Since a counterclockwise orientation is theoretically possible, as diagrammed in Figure 1b from Johansson et al. (1998), and was obtained by us in a 10.0-ns MD simulation, we developed the following hypotheses: one, the counterclockwise topology is simply an intermediate in the folding pathway of a clockwise A3, and two, A3 actually adopts a counterclockwise topology.

The intermediate hypothesis came from experimental observations by Ferreira and co-workers (Chapeaurouge et al., 2001) who showed that this polypeptide has two stable intermediates ($I_1$ and $I_2$). In order to test this hypothesis, we extended the simulation to 50.0 ns. As can be seen from the RMSD in Figure 4, the A3 topology did not change significantly and still maintained a counterclockwise orientation. Under these conditions and within the time scale in which the simulation was carried out, A3 adopted a counterclockwise orientation (Figure 5) as depicted in Figure 1b in Johansson et al. (1998).

Encouraged by these results, we further extended our analysis by comparing the A3 simulation trajectory with experimental data on the folding properties of three-α-helix bundles. The experimentally observed fluorescence emission spectra of Trp32 of A3 (Johansson et al., 1998) show a blue shift in a low dielectric environment. This might occur when Trp32 moves from the more hydrated (high dielectric) environment found in the extended conformation of the polypeptide to the less hydrated (low dielectric) interior of a polypeptide or protein. Since this

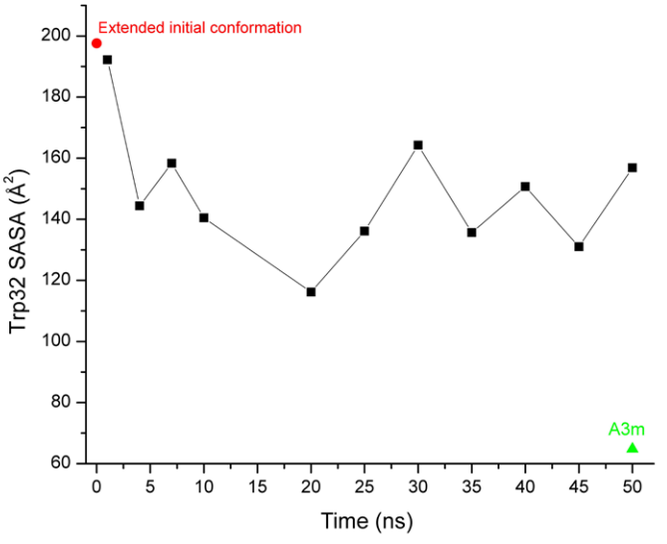


**Figure 6.** The solvent accessible surface area (SASA) of the Trp32 residue for 11 snapshots along the A3 MD simulation trajectory. The red-filled circle shows the standard Trp SASA (197.6 Å²) in the initial, extended conformation of the A3 polypeptide. The green-filled triangle shows the Trp32 SASA (64.8 Å²) for A3m, the ideal three-α-helix bundle model of A3.

change results in a solvent-protected Trp32, we were able to track it along the MD simulation trajectory as a change in the SASA of Trp32. We observed a decrease in the Trp32 SASA (Figure 6), from an initially extended conformation, as a function of simulation time. As expected, Trp32 SASA decreased as A3 formed a structure characteristic of the three-α-helix bundle observed in the simulation. This should be the behavior independent of the orientation of A3.

## CONCLUSION

We have presented the results of a 10.0-ns MD simulation, further extended to 50.0 ns, on the 65-amino acid polypeptide A3, supposed to adopt a clockwise three-α-helix bundle topology (Johansson et al., 1998; Chapeaurouge et al., 2001). Our objective was to develop MD simulation protocols for the prediction of protein and polypeptide 3-D structures based on their sequence of amino acids. Also, these predictions should be much faster compared to the time scale in which real motions occur in a protein *in vitro* or *in vivo* (Clarke et al., 1999).

Detailed analyses of the simulations showed how, since its onset, A3 rapidly developed its secondary structural units, namely α-helices I, II, and III and the two glycine-rich turns. This was obtained in a time frame of 1.0 ns and took less than 24 h of CPU time. The partial packing of these secondary structures into supersecondary structures occurred within the next 3.0 to 5.0 ns, much faster than the experimental rate (Clarke et al., 1999), as we expected our protocol to predict. Experimental data suppose a slow helix initiation due to a single rate-limiting nucleation event, followed by a fast helix elongation, and the transition of a polypeptide with high helix content into a helix-turn-helix by new nucleation points or by breakage of the long pre-existing helix. Both are events occurring in the millisecond timescale (Clarke et al., 1999). Our prediction protocol overcame these barriers in 10.0 ns of MD simulation, at the end of which the supersecondary and tertiary structures were stabilized into a 3-D topology characteristic of a three-α-helix-bundle motif, but with the helices arranged in a counterclockwise orientation. Extending the simulation to 50.0 ns did not change these features. In addition, the decrease in the solvent accessible surface area of Trp32 along the simulation correlates well with a denatured protein, where it is more exposed, folding into a more compact structure, where it is less exposed (Johansson et al., 1998).

Our initial MD simulation protocol for the prediction of polypeptide and protein 3-D structure from their amino acid sequence was moderately successful, allowing secondary structure formation at much faster rates than had been observed experimentally. This must be in part due to the usage of a varying cut-off radius, for the evaluation of the long-range interactions in the polypeptide, which in turn enabled simultaneous multiple nucleation points. However, although we did not obtain a clockwise three-α-helix bundle, a counterclockwise orientation should not be discarded, as it is also possible based on the design of Johansson et al. (1998). There is as yet no experimental 3-D structure for this α-helix bundle. More studies are necessary to find out why we obtained the counterclockwise instead of the preferred clockwise orientation in order to improve the prediction power of our MD protocol. Additional studies with different simulation protocols (temperature, simulation length, etc.) could reveal if the simulated A3 polypeptide got trapped at a local minimum.

## ACKNOWLEDGMENTS

We thank the reviewers for helpful comments on the article and Dr. César A.F. De Rose and the CPAD/PUCRS group for the administration of our PC cluster. Research supported by grants from CAPES, FAPERGS, and CNPq to O. Norberto de Souza and FINEP, Millennium Institute and CNPq/MCT to D.S. Santos and L.A. Basso. D.S. Santos, L.A. Basso and O. Norberto de Souza are recipients of CNPq fellowships. A. Breda was supported by a research-training scholarship from CNPq.

## REFERENCES

Anfinsen CB, Haber E, Sela M and White FH Jr (1961). The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain. *Proc. Natl. Acad. Sci. USA* 47: 1309-1314.

Baker D and Săli A (2001). Protein structure prediction and structural genomics. *Science* 294: 93-96.

Bashford D and Case DA (2000). Generalized born models of macromolecular solvation effects. *Annu. Rev. Phys. Chem.* 51: 129-152.

Berriz GF and Shakhnovich EI (2001). Characterization of the folding kinetics of a three-helix bundle protein via a minimalist Langevin model. *J. Mol. Biol.* 310: 673-685.

Bonneau R, Tsai J, Ruczinski I and Baker D (2001). Functional inferences from blind *ab initio* protein structure predictions. *J. Struct. Biol.* 134: 186-190.

Bottomley SP, Popplewell AG, Scawen M, Wan T, et al. (1994). The stability and unfolding of an IgG binding protein based upon the B domain of protein A from *Staphylococcus aureus* probed by tryptophan substitution and fluorescence spectroscopy. *Protein Eng.* 7: 1463-1470.

Bradley P, Misura KM and Baker D (2005). Toward high-resolution *de novo* structure prediction for small proteins. *Science* 309: 1868-1871.

Case DA, Pearlman DA, Caldwell JW, Cheathem TE III, et al. (1999). AMBER version 6.0. University of California, San Francisco.

Chapeaurouge A, Johansson JS and Ferreira ST (2001). Folding intermediates of a model three-helix bundle protein. Pressure and cold denaturation studies. *J. Biol. Chem.* 276: 14861-14866.

Chowdhury S, Lee MC, Xiong G and Duan Y (2003). *Ab initio* folding simulation of the Trp-cage mini-protein approaches NMR resolution. *J. Mol. Biol.* 327: 711-717.

Clarke DT, Doig AJ, Stapley BJ and Jones GR (1999). The alpha-helix folds on the millisecond time scale. *Proc. Natl. Acad. Sci. USA* 96: 7232-7237.

Cornell WD, Cieplak P, Bayly CI, Gould IR, et al. (1995). A second generation force field for the simulation of proteins. Nucleic acids, and organic molecules. *J. Am. Chem. Soc.* 117: 5179-5197.

DeLano WL (2002). PyMOL molecular graphics system. DeLano Scientific, San Carlos.

Guex N and Peitsch MC (1997). SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis* 18: 2714-2723.

Hansson T, Oostenbrink C and van Gunsteren WF (2002). Molecular dynamics simulations. *Curr. Opin. Struct. Biol.* 12: 190-196.

Hardin C, Pogorelov TV and Luthey-Schulten Z (2002). *Ab initio* protein structure prediction. *Curr. Opin. Struct. Biol.* 12: 176-181.

Hubbard SJ and Thornton JM (1993). 'NACCESS', Computer program. Department of Biochemistry and Molecular Biology, University College, London.

Humphrey W, Dalke A and Schulten K (1996). VMD: visual molecular dynamics. *J. Mol. Graph.* 14: 33-38.

Johansson JS, Gibney BR, Skalicky JJ, Wand AJ, et al. (1998). A native-like three-helix-bundle protein structure-based redesign: a novel maquette scaffold. *J. Am. Chem. Soc.* 120: 3881-3886.

Karplus M and McCammon JA (2002). Molecular dynamics simulations of biomolecules. *Nat. Struct. Biol.* 9: 646-652.

Laskowski RA, MacArthur MW, Moss DS and Thornton JM (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.* 26: 283-291.

Levinthal C (1969). Mössbauer spectroscopy in biological system. Proceedings of a Meeting held at Alperton House. University of Illinois Press, Monticello, 22-24.

Moult J (2005). A decade of CASP: progress, bottlenecks and prognosis in protein structure prediction. *Curr. Opin. Struct. Biol.* 15: 285-289.

Norin M and Sundström M (2002). Structural proteomics: lessons learnt from the early case studies. *Farmaco* 57: 947-951.

Pearlman DA, Case DA, Caldwell JW, Ross WS, et al. (1995). AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. *Comp. Phys. Commun.* 91: 1-41.

Ryckaert JP, Ciccotti G and Berendsen HJC (1977). Numerical integration of the Cartesian equation of motion of a system with constraints: molecular dynamics of N-alkanes. *J. Comp. Phys.* 23: 327-341.

Simmerling C, Strockbine B and Roitberg AE (2002). All-atom structure prediction and folding simulations of a stable protein. *J. Am. Chem. Soc.* 124: 11258-11259.

Starovasnik MA, Skelton NJ, O'Connell MP, Kelley RF, et al. (1996). Solution structure of the E-domain of staphylococcal protein A. *Biochemistry* 35: 15558-15569.

Sternberg MJ, Bates PA, Kelley LA and MacCallum RM (1999). Progress in protein structure prediction: assessment of CASP3. *Curr. Opin. Struct. Biol.* 9: 368-373.

Yan Y, Winograd E, Viel A, Cronin T, et al. (1993). Crystal structure of the repetitive segments of spectrin. *Science* 262: 2027-2030.