

Searching low-energy conformations of two elastin sequences

H. Arkin^a

Department of Physics, Hacettepe University, Beytepe 06532, Ankara, Turkey

Received 6 May 2003 / Received in final form 12 December 2003

Published online 15 March 2004 – © EDP Sciences, Società Italiana di Fisica, Springer-Verlag 2004

Abstract. The three-dimensional structures of two common repeat motifs Val¹-Pro²-Gly³-Val⁴-Gly⁵ and Val¹-Gly²-Val³-Pro⁴-Gly⁵-Val⁶-Gly⁷-Val⁸-Pro⁹ of tropoelastin are investigated by using the multicanonical simulation procedure. By minimizing the energy structures along the trajectory the thermodynamically most stable low-energy microstates of the molecule are determined. The structural predictions are in good agreement with X-ray diffraction experiments.

PACS. 02.70.Lq 02.70.Lq Monte Carlo and statistical methods – 05.50.+q Lattice theory and statistics (Ising, Potts, etc.) – 82.20.Wt Computational modeling; simulation

Determination of the folded structure of biological macromolecules such as polypeptides and proteins is an important goal in structural biology. Because the three-dimensional structure gives their biological activity [1]. The atomic interactions of a protein are commonly modeled by an empirical potential energy function, which typically leads to a complex energy landscape consisting of a tremendous number of local minima.

The major difficulty in conventional protein simulations such as the Metropolis Monte Carlo method or molecular dynamics lies in the fact that simulations are not effective at temperatures of experimental interest because the system becomes trapped in one of a huge number of energy local minima. The development of novel global optimization algorithms for the protein folding problem is still an active area of research. One way to overcome this problem is to perform simulation in a generalized-ensemble [2,3] where each state is weighted by non-Boltzmann probability weight factor, so that a flat histogram in potential energy space may be realized. This allows the simulation to escape from any energy barrier and to sample much wider conformational space than conventional methods. One of the well-known, powerful generalized-ensemble methods is the Multicanonical algorithm (MUCA) [4]. The trapping problem can be alleviated by the multicanonical simulation (MUCA) method. In one simulation all the temperature range is determined and all thermodynamic quantities are calculated.

The problem of protein folding entails the study of a nontrivial dynamics along pathways embedded in a rugged landscape. Therefore, development of efficient methods for conformational search is of central importance. Methods for searching energy landscape and low-energy conforma-

tions are proposed [5–7], energy landscape perspectives are investigated [8]. Such a goal can be achieved within the multicanonical ensemble approach.

While the MUCA ensemble is based on a probability function in which the different energies are equally probable:

$$P^{\text{MU}}(E) \sim n(E)w(E) = \text{const.} \quad (1)$$

where $w(E)$'s are multicanonical weight factors. Hence, a simulation with this weight factor, which has no temperature dependence, generates a one-dimensional random walk in the energy space, allowing itself to escape from getting trapped in any energy local minimum.

Re-weighting techniques (see Ref. [9]) enable one to obtain Boltzmann averages of various thermodynamic properties over a large range of temperatures. The advantage of this algorithm lies in the fact that it not only alleviates the multiple-minima problem but also allows the calculation of various thermodynamic quantities as functions of temperature from one simulation run. This demonstrates the superiority of the method.

To verify the coverage of the low-energy region by the MUCA sample, I also minimized the energy of conformations of the trajectory and indeed recovered the global energy minimized structure and other low energy minimized structures [10]. This suggests that MUCA can also serve as a useful conformational search technique for identifying the most stable wide microstates of a peptide. Further for comparison the effectiveness of searching low-energy conformations, a conformational search was obtained with the Monte Carlo minimization (MCM) method of Li and Scheraga [11].

The simulated molecule, tropoelastin, is soluble precursor protein of fibrous elastin and contains in its sequence several stretches of repeating oligopeptides.

^a e-mail: handan@hacettepe.edu.tr

The main three repeating oligopeptides are a tetrapeptide Val¹-Pro²-Gly³-Gly⁴, a pentapeptide Val¹-Pro²-Gly³-Val⁴-Gly⁵ and a nanopeptide Val¹-Gly²-Val³-Pro⁴-Gly⁵-Val⁶-Gly⁷-Val⁸-Pro⁹. In previous works, X-ray diffraction data were used to determine, at room temperature, the crystal structure of a repeat pentapeptide of elastin [12]. Nuclear magnetic resonance (NMR) studies have shown that the repeated-oligopeptide segments of elastin are composed of subunits that are conformationally equivalent within the NMR time scale. Several secondary structural elements have been proposed as features of one or more of these repeated peptide segments [13]. Alternative approaches such as computer molecular modeling starting from amino acid sequences can contribute to a better understanding of the three-dimensional structures of these repeating oligopeptides. In this work, I determine *all* the thermodynamically stable conformations populated by the molecule, not only the stable structure at room temperature.

The tropoelastin sequences namely the pentapeptide Val¹-Pro²-Gly³-Val⁴-Gly⁵ and a nanopeptide Val¹-Gly²-Val³-Pro⁴-Gly⁵-Val⁶-Gly⁷-Val⁸-Pro⁹ are modeled here by the well-known potential energy function ECEPP/2 [14], which is given by the sum of the electrostatic term, 12-6 Lennard-Jones term and the hydrogen bond term for all pairs of atoms in the peptide together with the torsion term for all torsion angles. The peptide bond angles ω are kept fixed at 180°, to their common value, and therefore a conformation of Val¹-Gly²-Val³-Pro⁴-Gly⁵-Val⁶-Gly⁷-Gly⁸-Gly⁹ sequence is defined solely by 28 degrees of freedom, the 16 backbone dihedral angles ϕ and ψ (the ϕ angles for the proline residues are fixed) and the 12 side chain dihedral angles χ . We use the standard dielectric constant $\epsilon = 2$ of ECEPP. This force field is implemented into the software package FANTOM [15].

First, I carried out two canonical (i.e., constant T) MC simulations at 50 K and 500 K, and multicanonical test runs which enabled us to determine the required energy ranges. At each MUCA update step a trial conformation was obtained by changing one dihedral angle at random within the range $[-180^\circ; 180^\circ]$, followed by the Metropolis test. The dihedral angles were always visited in a pre-defined order, going from first to last residue; a cycle of N MC steps ($N = 28$) is called a sweep.

The MUCA weights were built recursively during a long *single* simulation where the multicanonical parameters were re-calculated every 5000 sweeps and 200 times which adds up to 1 000 000 sweeps total. Several such simulations were carried out and the final MUCA weights of the best simulation were used in the following MUCA production run of another 1 000 000 sweeps. In all cases, each multicanonical simulation started from completely *random* initial conformation. No a priori information about the groundstate conformation is used in simulations.

As pointed out in the Introduction, for peptides it is not only of interest to obtain thermodynamic averages and fluctuations at different temperatures, but also to find the most stable regions in the conformational space, which allows to identify the most stable wide microstates. Proteins

Table 1. Number of Energy-Minimized Structures in Energy Bins of 0.5 kcal/mol above $E = -17.94$ kcal/mol as obtained by the MUCA and the MCM methods for the Val¹-Gly²-Val³-Pro⁴-Gly⁵-Val⁶-Gly⁷-Val⁸-Pro⁹ sequence. The results of only 10^5 sweeps are presented in the table.

Bin (kcal/mol)	Energy (kcal/mol)	MUCA	MCM ($T = 500$ K)
0.0 – 05	-17.95 to -17.45	1096	1774
0.5 – 1.0	-17.45 to -16.95	2374	3462
1.0 – 1.5	-16.95 to -15.95	2805	4089
1.5 – 2.0	-15.95 to -15.45	5038	5902
2.0 – 2.5	-15.45 to -13.95	4767	5646
2.5 – 3.0	-13.95 to -13.45	5117	4764
3.0 – 3.5	-13.45 to -12.95	5512	3609
3.5 – 4.0	-12.95 to -12.45	5116	2213
4.0 – 4.5	-12.45 to -11.95	4376	2169
4.5 – 5.0	-11.95 to -11.45	3632	1275
5.0 – 5.5	-11.45 to -10.95	2797	1147
5.5 – 6.0	-10.95 to -10.45	2470	611

are expected to populate low energy wide microstates even at room temperature, while peptides might also populate relatively higher energy microstates. Therefore, it is of interest to investigate the conformational coverage provided by MUCA, in particular in the low energy region [10]. In order to classify the microstates according to the potential wells they belong around thermodynamically stable different structures, each conformation of the simulation data was subjected to energy minimization.

Following the methods proposed by Meirovitch et al. [16], the configurations generated in 10^6 sweeps of the MUCA production run are minimized and the minimized structures were sorted according to a variance criterion where two structures are considered to be different if at least one dihedral angle differs by more than 2°. The lowest energy found (our suspected GEM) is

$$E = -17.94 \text{ kcal/mol} \quad (2)$$

and its conformation is depicted in Figure 1. The number of structures found in energy bins of 0.5 kcal/mol above $E = -17.94$ kcal/mol appear in Table 1. Only results of 10^5 sweeps are presented in Table 1. Here we compare the conformational searching of the low-energy region to that obtained with MCM [11]. For bins 7–12 the number for MCM is smaller than MUCA and for bins 1–4 the results of the two methods are close which means that a very good coverage of the lowest energy bin is provided by MUCA. From the table it is obvious that MUCA covers a large range of energies in an approximately homogeneous way.

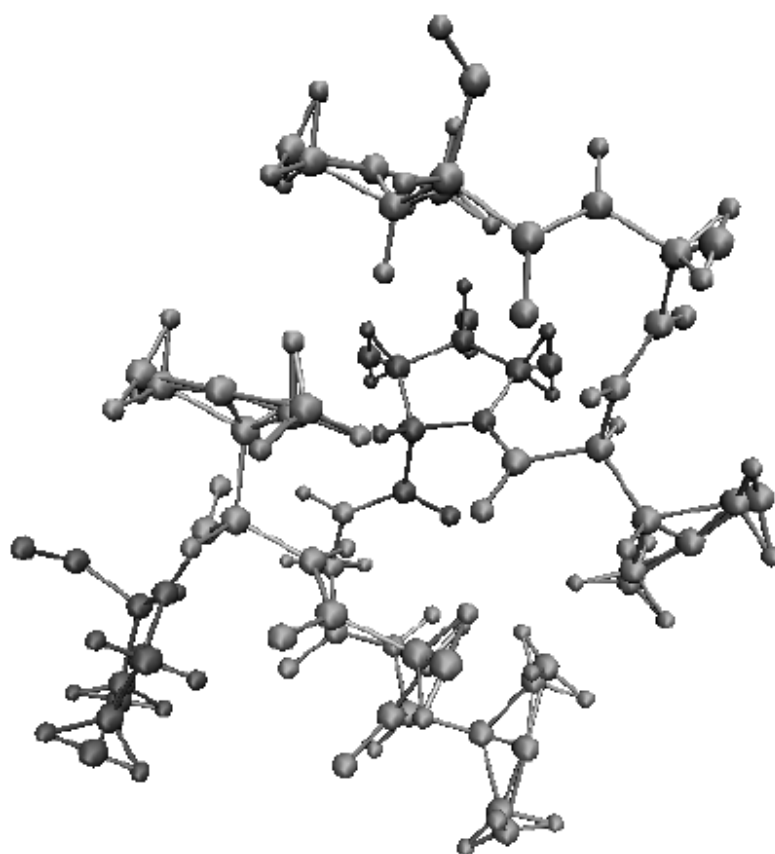


Fig. 1. Structure of the conjectured GEM of the peptide sequence Val¹-Gly²-Val³-Pro⁴-Gly⁵-Val⁶-Gly⁷-Val⁸-Pro⁹ at energy $E = -17.94$ kcal/mol.

The superiority of the multicanonical approach lies in the fact that MUCA provides the sampling of conformations at all temperatures from one simulation run, therefore enables one to study thermodynamics of the system under consideration. The distribution of backbone dihedral angles were analyzed and the Ramachandran plots were prepared for each residue of conformations. In Figure 2 the Ramachandran plots for the nanopeptide sequence (except proline residues) of typical structures for three temperature ranges are presented. The first set of seven plots (for the seven amino acid residues except prolines.) is for GEM structure and structures pertaining to the lowest energy bin above the GEM (bin 1 in Tab. 1). The second and third sets shows the most probable structures for the two temperature regions, 130–140 K and 290–300 K, respectively. One can easily see that the sample points in the set of plots at higher temperatures (the right column) are more scattered, but each plot contains the corresponding low temperature partner (the plot at the left column) as a subset. This means that these two groups are in the same well-defined valley in the energy space. Therefore it becomes possible to track a chosen conformation at room temperature goes to which one of the lowest energy states as the temperature lowered. Among the 1096 conformations (first bin in Tab. 1), there are

some typical different backbone features to those of the global minimum fold. The backbones of the typical conformations with the lowest energies, are presented in Figure 3. One can see this difference from the Ramachandran plots as well. The backbone differences occur mainly in the terminal parts that are expected to be the most flexible. Other differences comes from the side chain dihedral angles. The scattering of points in Ramachandran plots increase significantly for $T = 290\text{--}300$ K (third column in Fig. 2), suggesting that the sequences modeled with the ECEPP potential will populate at least several wide microstates. Analyzing the Ramachandran plots, the structures at $T = 290\text{--}300$ K consists of β -turns in Val¹-Gly²-Val³, Pro⁴-Gly⁵ and Val⁶-Gly⁷-Val⁸ bridges. These results are in agreement with previous studies, both experimental [17] and simulation [18] approaches, which suggest that the main secondary structure in elastin is short β -turns. Circular dichroism (CD) and NMR measurements gave evidence of flexible β -turns as the dominant structural feature [19,20].

In addition, the plots of Val³-Pro⁴-Gly⁵-Val⁶-Gly⁷ part of the nanopeptide are very similar to the Ramachandran plots of the pentapeptide (Val¹-Pro²-Gly³-Val⁴-Gly⁵), which is also simulated independently (plots are not presented).

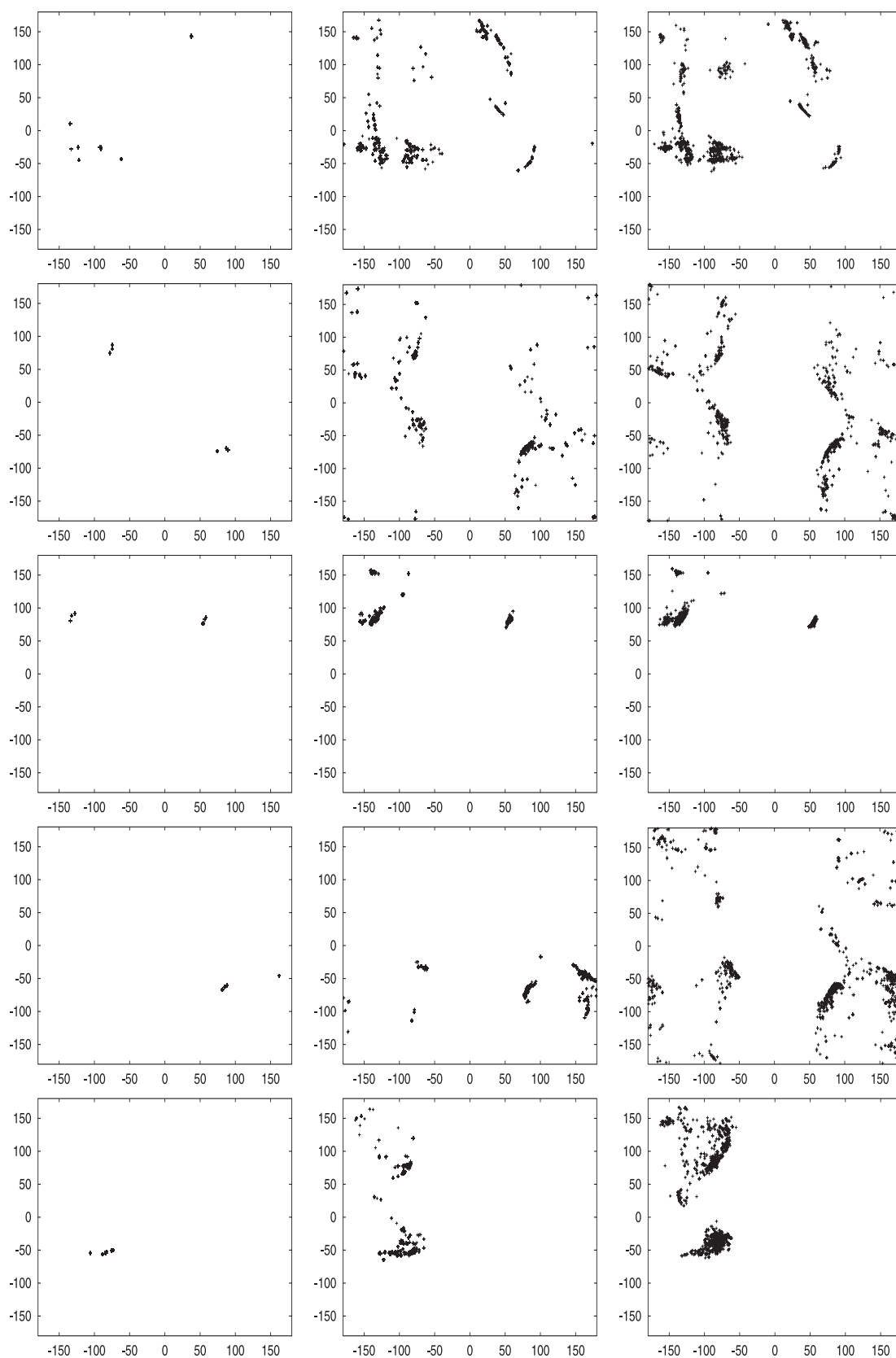


Fig. 2. Ramachandran plots of each residue (from top to bottom) Val¹-Gly²-Val³-Pro⁴-Gly⁵-Val⁶-Gly⁷-Val⁸-Pro⁹ (except proline residues). The abscissa is the angle ϕ and the ordinate is ψ . The angles are in the range of $[-180^\circ; 180^\circ]$. The first column shows the GEM and the conformations of the lowest energy bin above the GEM, the middle column shows conformations in the temperature range (130–140) K and the last column shows conformations in the temperatures range (290–300) K.

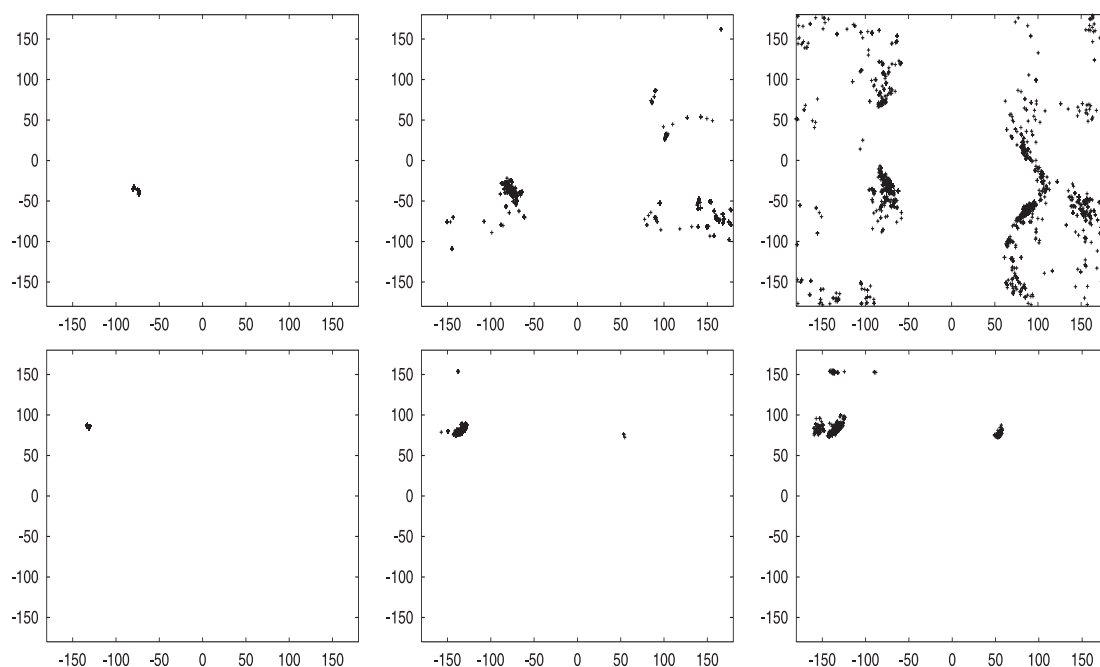


Fig. 2. Continued.

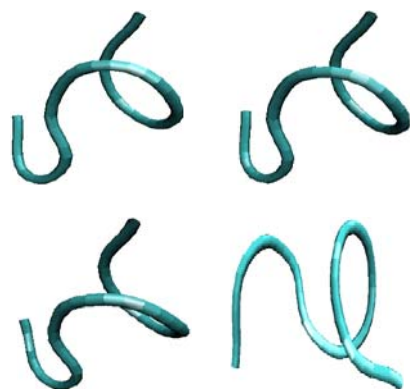
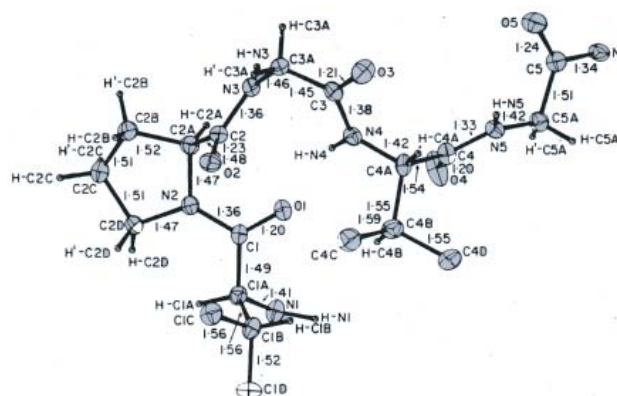
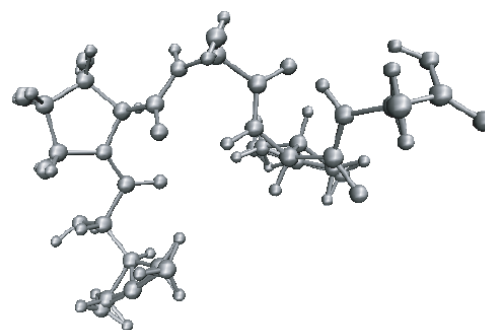


Fig. 3. Some typical low-energy conformations with energies below $E = -17.45$ kcal/mol (first bin in Tab. 1) of sequence Val¹-Gly²-Val³-Pro⁴-Gly⁵-Val⁶-Gly⁷-Val⁸-Pro⁹.

To compare the experimental findings and the multicanonical simulation results, the stable conformation of VPGVG (in one letter code) sequence at room temperature and the X-ray data are both shown in Figure 4. The crystal structure obtained with X-ray diffraction experiments are very close to the thermodynamically stable conformation at room temperature which is found by multicanonical simulation. The overlap parameter value (see Ref. [8] for the definition of the overlap parameter) between the two structures are found 0.94 which means that the two conformations are very close to each other. In Figure 5 the global minimum energy conformation of VPGVG sequence is given. By inspecting the Ramachandran plots, one can see that the global minimum conformation of Figure 5 and the conformation at room temperature shown in Figure 4, are in the same energy



(a)



(b)

Fig. 4. a) The conformation of pentapeptide sequence obtained with X-ray diffraction experiments. The data is from the paper: W.J. Cook et al., *J. Am. Chem. Soc.* **102**, 5502 (1980). b) The conformation of the pentapeptide obtained at room temperature by multicanonical simulation.

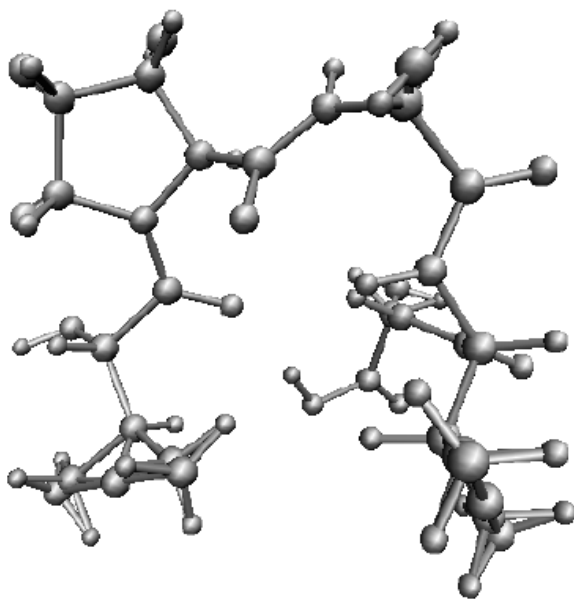


Fig. 5. Structure of the conjectured GEM of the pentapeptide sequence Val¹-Pro²-Gly³-Val⁴-Gly⁵ with energy $E = -4.40$ kcal/mol.

valley and the molecule at room temperature is folded almost to its native structure at low temperature.

To summarize, two common repeat motifs of tropoelastin have been simulated by using the multicanonical approach and demonstrated that a very good coverage of the conformational space especially in the low energy region is achieved. Sampling of conformations pertaining different microstates over wide range of temperatures, the most stable low energy microstates and the global energy minimum are obtained by multicanonical simulations. The results confirm the experimental findings of X-ray studies [12] and CD studies [19]. Considering computer time, MUCA simulation required a 9 hour production run of 1 million sweep for nano peptide on a DEC-Alpha 433 workstation.

References

1. M. Vásquez, G. Némethy, H.A. Scheraga, *Chem. Rev.* **94**, 2183 (1994)
2. U.H.E. Hansmann, Y. Okamoto, *Ann. Rev. Comp. Phys.* **5**, 129 (1999)
3. A. Mitsutake, Y. Sugita, Y. Okamoto, *Biopolymers (Peptide Science)* **60**, 96 (2001)
4. B.A. Berg, T. Çelik, *Phys. Rev. Lett.* **69**, 2292 (1992); B.A. Berg, *Fields Inst. Commun.* **28**, 1 (2000)
5. P.G. Wolynes, J. Onuchic, D. Thirumalai, *Science* **267**, 1619 (1995)
6. L.T. Wille, U.H.E. Hansmann, *Phys. Rev. Lett.* **88**, 068105 (2002)
7. H. Arkm, T. Çelik, *Int. J. Mod. Phys. C* **14**, 985 (2003)
8. H. Arkm, T. Çelik, *Int. J. Mod. Phys. C* **14**, 113 (2003); H. Arkm, T. Çelik, *Int. J. Mod. Phys. C* **14**, 567 (2003)
9. A.M. Ferrenberg, R.H. Swendsen, *Phys. Rev. Lett.* **61**, 2635 (1988); A.M. Ferrenberg, R.H. Swendsen, *Phys. Rev. Lett.* **63**, 1658 (1989)
10. H. Arkm, T. Çelik, *Eur. Phys. J. B* **30**, 577 (2002)
11. Z. Li, H.A. Scheraga, *Proc. Natl. Acad. Sci. USA* **84**, 6611 (1987); Z. Li, H.A. Scheraga, *J. Mol. Struct.* **179**, 333 (1988)
12. W.J. Cook, H. Einspahr, T.L. Trapane, D.W. Urry, C.E. Bugg, *J. Am. Chem. Soc.* **102**, 5502 (1980)
13. D. Urry, M.M. Long, *CRC Crit. Rev. Biochem.* **4**, 1 (1976)
14. F.A. Momany, R.F. McGuire, A.W. Burgess, H.A. Scheraga, *J. Phys. Chem.* **79**, 2361 (1975); M.J. Sippl, G. Némethy, H.A. Scheraga, *J. Phys. Chem.* **88**, 6231 (1984)
15. B. von Freyberg, T. Schaumann, W. Braun, *FANTOM User's Manual and Instructions: ETH Zürich, Zürich, 1993*; B. von Freyberg, W. Braun, *J. Comput. Chem.* **14**, 510 (1993)
16. H. Meirovitch, E. Meirovitch, *J. Comput. Chem.* **18**, 240 (1997)
17. B. Li, D.O.V. Alonso, V. Daggett, *J. Mol. Biol.* **305**, 581 (2001)
18. V. Villani, A.M. Tamburro, *J. Mol. Struct. (Theochem)* **431**, 205 (1998)
19. A.M. Tamburro, V. Guantieri, L. Pandolfo, A. Scopa, *Biopolymers*, **29**, 855 (1990)
20. D.W. Urry, *J. Phys. Chem. B* **101**, 11007 (1997)