

Helix-Formation and Folding as Studied in Generalized-Ensemble Simulations

Nelson A. Alves

*Departamento de Física e Matemática, FFCLRP,
Universidade de São Paulo. Av. Bandeirantes 3900.
CEP 14040-901 Ribeirão Preto, SP, Brazil*

and Ulrich H. E. Hansmann

*Department of Physics, Michigan Technological University,
Houghton, MI 49931-1291, USA*

Received on 23 April, 2003

Studying simple artificial peptides, we show that recently developed simulation techniques enable efficient investigations of secondary structure formation and folding in small peptides.

I Introduction

Besides questions of sex and aging that capture not only the interest of Dietrich Stauffer [1], whose 60th birthday is celebrated by this issue of the *Brazilian Journal of Physics*, many other problems in biology and chemistry are now addressed by computer experiments. Probably the best known example is the protein-folding problem, i.e. the attempt to understand the mechanism by which a protein is driven into its unique biologically active structure.

While in principle it is possible to study the folding problem *in silico*, such simulations are extremely difficult for realistic protein models. This is because all-atom models lead to a rough energy landscape with a huge number of local minima separated by high energy barriers. Consequently, sampling of low-energy conformations becomes a hard computational task, and physical quantities cannot be calculated accurately from simple low-temperature molecular dynamics or Monte Carlo simulations.

A number of novel simulation techniques have been proposed for overcoming the multiple-minima problem (for a review, see Ref. [2]). One successful example is the *generalized-ensemble* approach [3], that was first applied to the protein-folding problem in Ref. [4]. Examples of this group of closely related techniques are multicanonical sampling [5], the broad histogram method [6], or simulated tempering [7]. In the following, we will present a short review of this technique followed by a discussion of its applications to one particularly important aspect of the protein-folding problem, namely the role of secondary structure for-

mation in the folding process. To be more specific, we investigate the relation between helix formation and folding considering simple artificial peptides.

II Generalized-ensemble techniques

A generalized-ensemble simulation [3] is characterized by the condition that a Monte Carlo or molecular dynamics simulation will lead to a uniform distribution of a pre-chosen physical quantity. For instance, in multicanonical sampling [5] the weight $w(E)$ is chosen such that the distribution of energies $P(E)$ is given by:

$$P(E) \propto n(E)w(E) = \text{const}, \quad (1)$$

where $n(E)$ is the spectral density. A free random walk in the energy space is performed that allows the simulation to escape from any local minimum. From this simulation one can calculate the thermodynamic average of any physical quantity A by re-weighting: [8]

$$\langle A \rangle_T = \frac{\int dx \mathcal{A}(x) w^{-1}(E(x)) e^{-E(x)/k_B T}}{\int dx w^{-1}(E(x)) e^{-E(x)/k_B T}}. \quad (2)$$

Here, x stands for configurations and k_B is the Boltzmann constant. The weights $w(E)$ are not *a priori* known in generalized ensembles and estimators have to be determined by an iterative procedure described in Refs. [5, 9, 10]. We remark that it is straightforward to define ensembles that lead to flat distributions in more than one variable [11].

Another way of enhancing sampling of low-energy configurations in protein simulations is parallel tempering [10,

12]. In its most common form, one considers in this technique an artificial system built up of N *non-interacting* copies of the molecule, each at a different temperature T_i . In addition to standard Monte Carlo or molecular dynamics moves that affect only one copy, parallel tempering introduces now a new *global* update [12]: the exchange of conformations between two copies i and $j = i + 1$ with probability

$$w(\mathbf{C}^{old} \rightarrow \mathbf{C}^{new}) = \min(1, \exp(-\beta_i E(C_j) - \beta_j E(C_i) + \beta_i E(C_i) + \beta_j E(C_j))) . \quad (3)$$

This exchange of conformations leads to a faster convergence of the Markov chain than in regular canonical simulations since the resulting random walk in temperatures allows the configurations to move out of local minima and cross energy barriers. Note that parallel tempering does not require Boltzmann weights. The method can be combined easily with other generalized-ensemble techniques as was demonstrated first in Ref. [10].

III Helix-formation and Folding

In the following we want to demonstrate that the above described generalized-ensemble techniques are well suited for protein-folding studies.

Our understanding of the folding process has increased considerably over the last few years. However, many questions remain open. For instance, it is still under debate at what step in the folding process secondary structure forms; and to what degree secondary structure formation and folding are determined by the intrinsic properties of the protein or by the interaction with the surrounding solvent.

In order to study these questions we have investigated two alanine based peptides: Ala₁₀-Gly₅-Ala₁₀ and (Ala-Asp)₅-Gly₅-(Ala-Asp)₅. These peptides are chosen because we know from earlier work that polyalanine (Ala_N) exhibits a helix-coil transition in gas phase and with certain implicit solvent models [13-17]. Hence, an investigation of these peptides allows to probe the role of helix formation in the folding process.

Our investigation of the two peptides is based on a detailed, all-atom representation with the interaction between the atoms described by a standard force field, ECEPP/2,[18]

(as implemented in the program package SMMP [19]):

$$E_{ECEPP/2} = E_C + E_{LJ} + E_{HB} + E_{tor}, \quad (4)$$

$$E_C = \sum_{(i,j)} \frac{332q_i q_j}{\epsilon r_{ij}}, \quad (5)$$

$$E_{LJ} = \sum_{(i,j)} \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} \right), \quad (6)$$

$$E_{HB} = \sum_{(i,j)} \left(\frac{C_{ij}}{r_{ij}^{12}} - \frac{D_{ij}}{r_{ij}^{10}} \right), \quad (7)$$

$$E_{tor} = \sum_l U_l (1 \pm \cos(n_l \chi_l)). \quad (8)$$

Here, r_{ij} (in Å) is the distance between the atoms i and j , and χ_l is the l -th torsion angle. The parameters ($q_i, A_{ij}, B_{ij}, C_{ij}, D_{ij}, U_l$ and n_l) are calculated from crystal structures of amino acids. Since the bond lengths and bond angles are set constant, the true degrees of freedom are rotations around these bonds characterized by dihedral angles ϕ, ψ, ω , and χ_i . The peptide bond angles were set to their common value $\omega = 180^\circ$. All energies are in kcal/mol, hence the factor '332' in the electrostatic energy term E_C where the permittivity is usually set to $\epsilon = 2$, its supposed value in the protein interior. E_{LJ} is a Lennard-Jones term, E_{HB} the hydrogen-bond energy and E_{tor} accounts for the torsion energy of the molecule.

The interactions between the peptide and the surrounding water are approximated by adding a solvent accessible surface term [20] to the energy function:

$$E = E_{ECEPP/2} + E_{solv} \quad \text{with} \quad E_{solv} = \sum_i \sigma_i A_i . \quad (9)$$

Here, E_{solv} is the solvation energy. In this approximation, one assumes that the free energy difference between atomic groups immersed in the protein interior and groups exposed to water is proportional to the solvent accessible surface area A_i of the i th atom with the parameters σ_i as experimentally determined proportionality factors.

Our results for each of the two peptides rely on a multicanonical simulation [5] of 4,000,000 sweeps following 10,000 sweeps for "thermalization". Our simulations start from completely random initial conformations (Hot Start) and one Monte Carlo sweep updates every torsion angle of the peptide once. The time series of energy for the simulations of both peptides is shown in Fig. 1a and 1b, respectively.

We have found in previous work [21] that the peptide Ala₁₀-Gly₅-Ala₁₀ folds in gas phase in a two-step process: first, two α -helices are formed that afterward arrange themselves into a U-like structure. In the present paper, we extend this research to the case where protein-solvent interactions are taken into account, and to a molecule where alternatingly an alanine residue is replaced by an asparagine. This

residue is chosen because unlike alanine (a helix-former) it is often found in β -sheets.

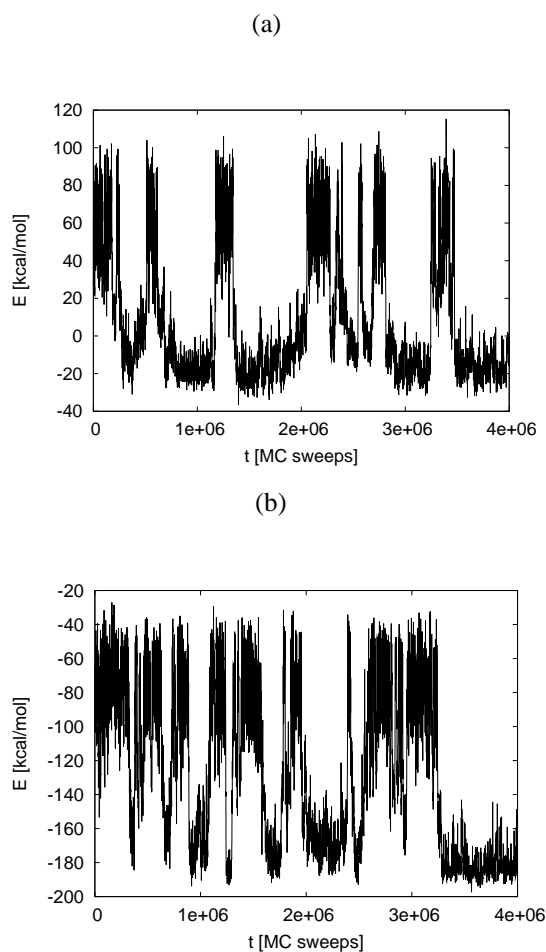


Figure 1. Time series of energy for (a) Ala₁₀-Gly₅-Ala₁₀ and (b) (Ala-Asp)₅-Gly₅-(Ala-Asp)₅ as obtained by multicanonical simulations.

We start our analysis with the peptide Ala₁₀-Gly₅-Ala₁₀ in an implicit solvent. Fig. 2a displays the lowest energy structure. As in the earlier investigated case of the peptide in gas phase [21] we find a hairpin formed by two α -helices. However, the folding process differs from the gas-phase case. For instance, we find only a single peak in the specific heat $C(T)$ (Fig. 3) at a temperature $T_c = 408 \pm 10$ K. At this temperature, the average number of helical residues $\langle n_H \rangle (T)$ (Fig. 4a) is increasing rapidly with decreasing temperature, while at the same temperature the end-to-end distance $\langle d_{e-e} \rangle (T)$ (a measure for the compactness of the peptide configuration, displayed in Fig. 4b) exhibits a sharp drop when the temperature is lowered. A more detailed investigation [21] reveals that this transition is either a weak first-order or a strong second order transition. Note that the folding scenario for the solvated Ala₁₀-Gly₅-Ala₁₀ molecule differs from the one in gas phase where helix formation and folding happen at separate temperatures, a helix-coil transition temperature $T_{hc} = 483 \pm 8$ K and a folding temperature $T_f = 265 \pm 7$ K. We conjecture that this different behavior is due to the hydrophobic nature of alanine

which favors energetically the close alignment of the two α -helices (built out of the alanine residues) over extended helical structures that are found in gas phase for $T_f \leq T \leq T_{hc}$. The resulting force is so strong that the helices fold immediately after their formation into the hairpin merging in this way the two transitions that are observed in gas phase into a single one.

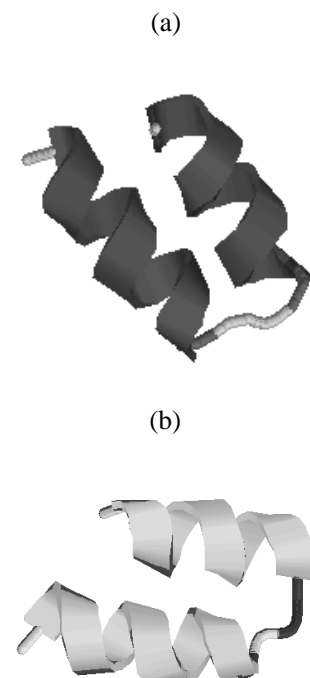


Figure 2. The lowest energy structure of (a) Ala₁₀-Gly₅-Ala₁₀ and (b) (Ala-Asp)₅-Gly₅-(Ala-Asp)₅ as obtained by multicanonical simulations.

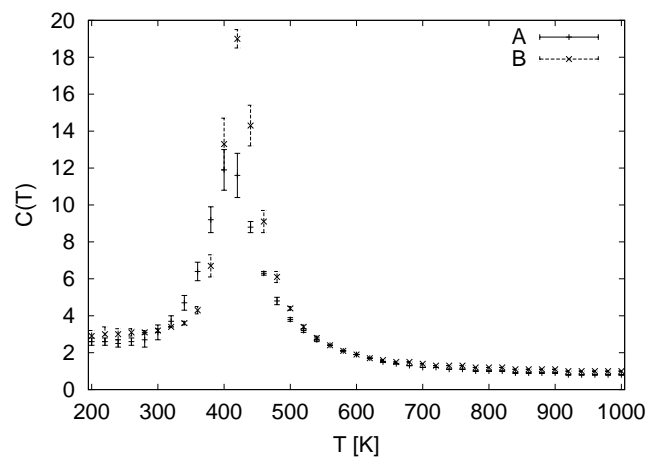


Figure 3. The specific heat $C(T)$ as a function of temperature for Ala₁₀-Gly₅-Ala₁₀ (A) and (Ala-Asp)₅-Gly₅-(Ala-Asp)₅ (B). The values are calculated from multicanonical simulations of the two peptides.

The above results indicate that Ala₁₀-Gly₅-Ala₁₀ folds in a two step process reminiscent of the well known framework [22, 23] and collision-diffusion model [24]. The first step is the formation of α -helices. It is connected with the large gain in the intramolecular ECEPP/2 energy that can be seen in Fig. 5a. Note that we have shifted the energy by a

constant (and therefore irrelevant) factor such that its value at $T = 1000$ K is zero. Such normalization allows for a better comparison of our two molecules. On the other hand, the gain in ECEPP/2 energy with helix-formation is partially compensated by a loss in potential energy (see Fig. 5b). This is because the OONS parameter set emphasizes the hydrophobic character of carbon atoms, therefore favoring coil structures over helical ones. However, the contribution by the solvation term is small compared with the ECEPP/2 term and overall, helix-formation is favored by a large gain in energy. The formation of α -helices then restricts the possible configuration space. Energetically most favorable is the folding of two α -helices (made out of the alanine residues) into a hairpin. Since the OONS parameter set emphasizes the hydrophobic character of the carbon atoms and decreases the hydrophilic character of uncharged oxygen and nitrogen atoms, the resulting force makes a close alignment of the two α -helices (build out of the alanine residues) energetically especially favorable, and the two helices fold immediately after formation into a helix hairpin [25].

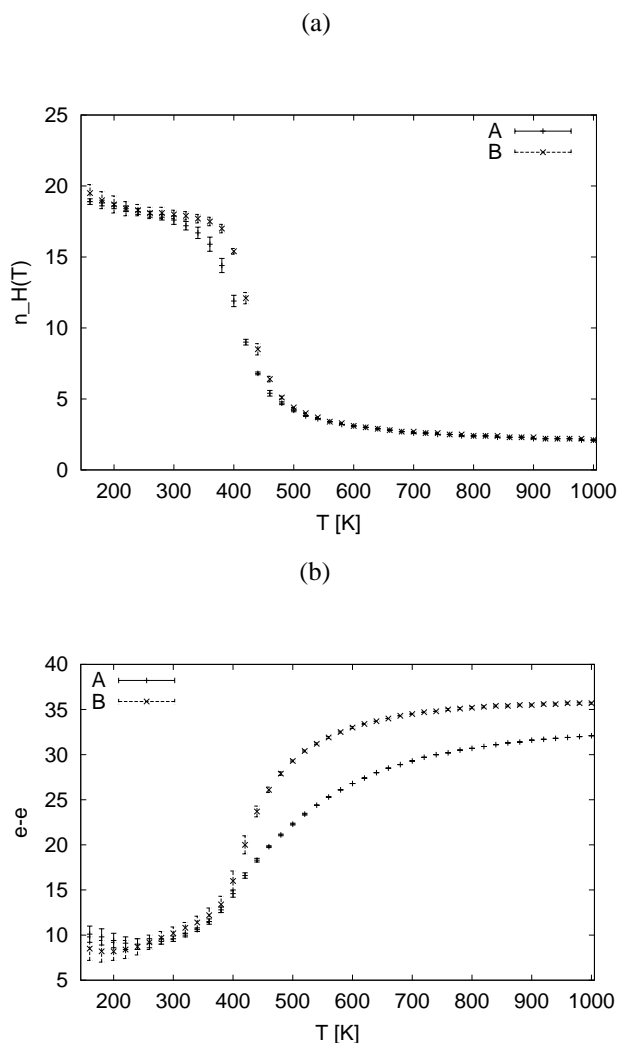


Figure 4. (a) The average helicity $\langle n_H \rangle(T)$ and (b) the average end-to-end distance $\langle d_{e-e} \rangle(T)$ as a function of temperature for Ala₁₀-Gly₅-Ala₁₀ (A) and (Ala-Asp)₅-Gly₅-(Ala-Asp)₅ (B) as obtained by multicanonical simulations.

How does this picture change if one replaces some of the helix-forming alanine residues by asparagine, an amino acid that unlike alanine has polar side groups? In order to study this question, we also consider a second peptide, (Ala-Asp)₅-Gly₅-(Ala-Asp)₅, where alternately alanine is replaced by asparagine. The OONS implicit solvent model [20] is again used to approximate the peptide-water interaction. We show in Fig. 3 also the specific heat as a function of temperature for this peptide. As in the case of (the solvated) Ala₁₀-Gly₅-Ala₁₀ molecule, only one peak is observed. However, the peak for (Ala-Asp)₅-Gly₅-(Ala-Asp)₅ is higher and narrower than the one for Ala₁₀-Gly₅-Ala₁₀, indicating a sharper transition. The corresponding transition temperature, $T_c = 422 \pm 5$, is only slightly higher than the one for Ala₁₀-Gly₅-Ala₁₀ ($T_c = 408 \pm 10$). Plots of the average number of helical residues $\langle n_H \rangle(T)$ (Fig. 4a) and the average end-to-end distance $\langle d_{e-e} \rangle(T)$ (Fig. 4b) indicate that this temperature marks both helix-coil transition and the folding into a helix hairpin (Fig. 2b). Both quantities seem again to indicate a sharper transition. This is surprising as asparagine is a polar amino acid. Hence, hydrophobic forces that would favor a helix hairpin over more extended structures are weaker for (Ala-Asp)₅-Gly₅-(Ala-Asp)₅ than for Ala₁₀-Gly₅-Ala₁₀, and one would therefore expect more extended structures. However, the corresponding larger values of $\langle d_{e-e} \rangle$ are only observed above T_c in the disordered phase. On the other hand, we do notice in Fig. 5b that the magnitude of the solvation energy term is smaller for the (Ala-Asp)₅-Gly₅-(Ala-Asp)₅ peptide than for Ala₁₀-Gly₅-Ala₁₀ while at the same time the corresponding ECEPP/2 energy (Fig. 5a) is larger in magnitude. As a consequence, helix-formation (that is opposed by the solvation term) is enhanced, leading to a slightly higher helix-coil transition temperature and increased average number of helical residues that is observed in Fig. 4a. Hence, formation of the two helices is enhanced. This increased helix formation may counteract the weaker hydrophobic forces as formation of the two helices again restricts the conformation space of the molecule and the energetically most favorable state is the folding into the helix hairpin.

Note, that for both molecules the final structure is determined solely by the intramolecular interactions while the details of the folding process depend also on the solvent contributions. Such a simple picture will be in general not true. However, our results point out that often the details of protein sequence are not important as long as they lead locally to the same secondary structure. Both Ala₁₀-Gly₅-Ala₁₀ and (Ala-Asp)₅-Gly₅-(Ala-Asp)₅ form two α -helices that are separated by the flexible chain of glycine residues. The formation of these two helices determines the unique final structure.

IV Conclusion

We gave a brief introduction into generalized-ensemble techniques and their applications to the protein folding prob-

lem. Using one of these techniques we have probed the relation between helix-formation and folding in two artificial peptides. Our results point out that the lowest energy structure does not depend on the details of the sequence but on the secondary structure that they encode. This is consistent with predictions of the framework [22, 23] and collision-diffusion model [24] of folding. These and our earlier results demonstrate that generalized-ensemble algorithms are well-suited for investigations of the thermodynamics of proteins and may lead to a deeper understanding of the protein-folding problem.

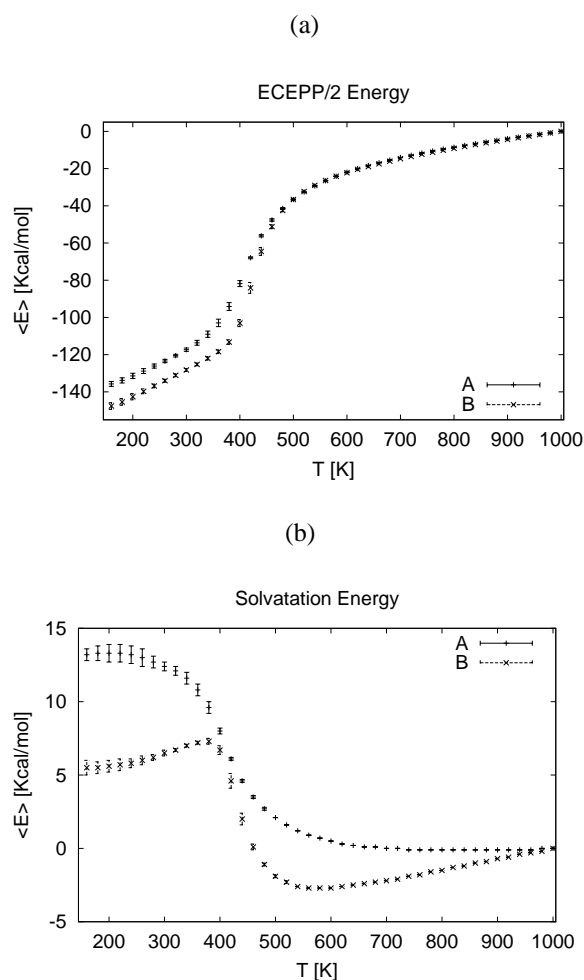


Figure 5. (a) The average ECEPP/2 energy of Ala₁₀-Gly₅-Ala₁₀ (A) and (Ala-Asp)₅-Gly₅-(Ala-Asp)₅ (B) as a function of temperature. The values were calculated from multicanonical simulations. The corresponding solvation energies E_{solv} for both peptides as a function of temperature are displayed in (b).

Acknowledgments:

This article developed out of the talk that U.H. presented at the workshop in Niteroi (Feb. 26-28, 2003) that celebrated Dietrich Stauffer's 60th birthday. Parts of the results presented in this article are also published in Refs. [21, 25]. N.A. Alves gratefully acknowledges support by CNPq (Brazil), and U.H. Hansmann support by a research grant from the National Science Foundation (CHE-9981874).

References

- [1] See, for instance: D. Stauffer, *Life, Love and Death: Models of Biological Reproduction and Aging*, proceedings of the international workshop "Monte Carlo and Structure Optimization Methods for Biology, Chemistry and Physics", Tallahassee, March 1999; available under: www.scri.fsu.edu/MCatSCRI/proceedings/stauffer/stauffer.pdf
- [2] U.H.E. Hansmann and Y. Okamoto, *Curr. Opin. Struc. Biol.* **9**, 177 (1999).
- [3] U.H.E. Hansmann and Y. Okamoto, In *Annual Reviews in Computational Physics VI*. Edited by Stauffer D. Singapore: World Scientific; 1999, 129-157.
- [4] U.H.E. Hansmann and Y. Okamoto, *J. Comp. Chem.* **14**, 1333 (1993).
- [5] B.A. Berg and T. Neuhaus, *Phys. Lett.* **B267**, 249 (1991); *Phys. Rev. Lett.* **68**, 9 (1992).
- [6] P.M.C. de Oliveira, T.J.P. Penna, and H.J. Herrmann, *Braz. J. Phys.* **26**, 677 (1996); P.M.C. de Oliveira, *Int. J. Mod. Phys. C* **9**, 497 (1998).
- [7] A.P. Lyubartsev, A.A. Martinovski, S.V. Shevkunov, and P.N. Vorontsov-Velyaminov, *J. Chem. Phys.*, **96**, 1776 (1992); E. Marinari, G. Parisi, *Europhysics Letters*, **19**, 451 (1992).
- [8] A.M. Ferrenberg and R.H. Swendsen, *Phys. Rev. Lett.* **61**, 2635 (1988); *Phys. Rev. Lett.* **63**, 1658(E) (1989), and references given in the erratum.
- [9] U.H.E. Hansmann and Y. Okamoto, *Physica A* **212**, 415 (1994).
- [10] U.H.E. Hansmann, *Chem. Phys. Lett.* **281**, 140 (1997).
- [11] S. Kumar, P.W. Payne, and M. Vásquez, *J. Comp. Chem.* **17**, 1269 (1996).
- [12] K. Hukushima and K. Nemoto, *J. Phys. Soc. (Jpn.)* **65**, 1604 (1996); G.J. Geyer, *Stat. Sci.* **7**, 437 (1992).
- [13] D. Poland and H.A. Scheraga, *Theory of Helix-Coil Transitions in Biopolymers* (Academic Press, New York, 1970).
- [14] U.H.E. Hansmann and Y. Okamoto, *J. Chem. Phys.* **110**, 1267 (1999); **111**, 1339(E) (1999).
- [15] N.A. Alves and U.H.E. Hansmann, *Phys. Rev. Lett.* **84**, 1836 (2000).
- [16] Y. Peng and U.H.E. Hansmann, *Biophys. J.* **82**, 3269 (2002).
- [17] Y. Peng, U.H.E. Hansmann, and N.A. Alves, *J. Chem. Phys.* **118**, 2374 (2003).
- [18] M.J. Sippl, G. Némethy, and H.A. Scheraga, *J. Phys. Chem.* **88**, 6231 (1984), and references therein.
- [19] F. Eisenmenger, U.H.E. Hansmann, Sh. Hayryan, and C.-K. Hu, *Comp. Phys. Comm.* **138**, 192 (2001).
- [20] T. Ooi, M. Obatake, G. Némethy, and H.A. Scheraga, *Proc. Natl. Acad. Sci. USA* **84**, 3086 (1987).
- [21] N.A. Alves and U.H.E. Hansmann, *J. Chem. Phys.* **117**, 2337 (2002).
- [22] O.B. Ptitsyn, *Protein Eng.* **7**, 593 (1994).
- [23] P.D. Kim and R.L. Baldwin, *Annu. Rev. Biochem.* **59**, 631 (1990).
- [24] M. Karplus and D.L. Weaver, *Protein Sci.* **3**, 650 (1994).
- [25] U.H.E. Hansmann, *Int. J. Quant. Chem.*, **90** 1515 (2002).