Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Molecular Graphics and Modelling 22 (2004) 397–403
Journal of Molecular Graphics and Modelling
www.elsevier.com/locate/JMGM

All-atom generalized-ensemble simulations of small proteins

Brian S. Kinnear [a], Martin F. Jarrold [b], Ulrich H.E. Hansmann [c,*]

[a] *Portland Technology Development, Intel Corp., Hilisboro, OR 97124, USA*
[b] *Chemistry Department, Indiana University, Bloomington, IN 47405, USA*
[c] *Department of Physics, Michigan Technological University, Houghton, MI 49931, USA*

Abstract

We give an overview of some generalized-ensemble techniques that have proven successful in all-atom simulations of proteins. We show that these techniques enable efficient investigations of secondary structure formation and folding in peptides and small proteins. Results are presented for various alanine-based artificial peptides and a small protein, the 36-residued villin headpiece subdomain (HP-36). Our results indicate that all-atom simulations of proteins may be more restricted by the accuracy of the present energy functions than by the efficiency of the search algorithms.

Keywords: Generalized-ensemble simulations; Proteins; HP-36

1. Introduction

The successful deciphering of whole genomes has aggravated an old challenge in protein science: for most of the resolved protein sequences one does not know the corresponding structures and functions. Since proteins are only functional if they fold into their specific shape, and misfolded proteins can even cause a variety of diseases, it is important to understand how the structure and function of proteins emerge from their sequence of amino acids. Computer experiments offer one way to gain such knowledge but are extremely difficult for realistic protein models. This is because all-atom models of proteins 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 developed that promise to alleviate the above-stated multiple-minima problem (for a review, see [1]). One successful method is the so-called *genera1ized-ensemb1e* approach [2] that was first applied to protein simulations in [3]. In the following we will present a short review of this approach and demonstrate its usefulness for protein simulations. We will focus in our examples on one particularly important aspect of the protein-folding problem, namely the role of secondary structure formation in the folding process.

2. Generalized-ensemble techniques

The complex form of the intramolecular forces and of the interaction with the solvent, containing both repulsive and attractive terms, leads for all-atom models of proteins to a very rough energy landscape with a huge number of local minima. Hence, a typical thermal energy of the order $k_B T$ is much less than the energy barriers that the protein has to overcome in the low-temperature region. This is because the probability to cross an energy barrier is proportional to $\propto \exp(-\Delta E/k_B T)$ where ΔE is the heights of the energy barrier. Hence, simple canonical Monte Carlo or molecular dynamics simulations will get trapped in a local minimum. The essence of all generalized-ensemble simulations is to replace the canonical weights by such that do allow escape out of local minima. In most cases the weights are chosen in such a way 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 [4] the weight $w(E)$ is chosen such that the distribution of energies $P(E)$ is given by

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

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

* Corresponding author.
E-mail address: hansmann@mtu.edu (U.H.E. Hansmann).

1093-3263/$ – see front matter © 2003 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmgm.2003.12.006

can calculate the thermodynamic average of any physical quantity A by re-weighting: [5]

$$\langle \mathcal{A} \rangle_{\mathrm{T}} = \frac{\int \mathrm{d}x \, \langle \mathcal{A} \rangle(x) w^{-1}(E)(x)) \mathrm{e}^{-E(x)/k_{\mathrm{B}}T}}{\int \mathrm{d}x \, w^{-1} E(x)) \mathrm{e}^{-E(x)/k_{\mathrm{B}}T}}.$$

Here, x stands for configurations and k_{B} is the Boltzmann constant. The weight $w(E)$ is not a priori known in generalized ensembles, and estimators have to be determined by an iterative procedure described in [4,6]. We remark that it is straightforward to define ensembles that lead to flat distributions in more than one variable [7].

Even within the generalized-ensemble approach simulations of proteins can still be hampered by large correlations between the sampled conformations. Improved updates are one way to overcome this obstacle. One example is parallel tempering. The method is also known as replica exchange or Multiple Markov Chain method [8], and was first introduced to protein folding in [9]. 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 effect only one copy, parallel tempering introduces now a new *global* update: the exchange of conformations between two copies i and $j = i+1$ with probability

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

The exchange of conformations will lead especially at low temperatures to a faster convergence of the Markov chain toward the stationary distribution than is observed in regular canonical simulations with only local moves. 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 [9].

The emphasis in todays computational protein studies is often on the *prediction* of the three-dimensional shape in which a protein is biologically active. Most proteins are at room temperature thermodynamically stable, i.e. the biologically active configuration is the global minimum in *free energy* at $T = 300\,\mathrm{K}$. Since this state is unique, one can identify the free-energy global minimum with the lowest *potential* energy conformation. Hence, structure prediction of proteins can be considered as a global optimization problem.

A general characteristic of successful optimization techniques is that they avoid entrapment in local minima and continue to explore the energy landscape for further solutions. Noticing that this also a characteristics of generalized-ensemble algorithms, Wille and Hansmann proposed recently a new optimization method, energy landscape paving (ELP), that proved very promising in protein studies [10].

In ELP, one performs low-temperature Monte Carlo simulations with a modified energy expression designed to steer the search away from regions that have been already explored:

$$w(\tilde{E}) = \mathrm{e}^{-\tilde{E}/k_{\mathrm{B}}T} \quad \text{with} \quad \tilde{E} = E + f(H(q,t)). \quad (4)$$

Here, T is a (low) temperature, $\tilde{E}$ serves as a replacement of the energy E and $f(H(q, t))$ is a function of the histogram $H(q, t)$ in a pre-chosen "order parameter" q. It follows that within ELP the weight of a local minimum state decreases with the time the system stays in that minimum. With time, ELP deforms the energy landscape locally till the local minimum is no longer favored, and the system will explore higher energies. It will then either fall in a new local minimum or walk through this high energy region till the corresponding histogram entries all have similar frequencies, and the system again has a bias toward low energies. Obviously, for $f(H(q,t)) = f(H(q))$ the method reduces to the various generalized-ensemble methods [2] (for instance for $f(H(q,t)) = \ln H(E)$ to multicanonical sampling).

The efficiency of the new global optimization method was tested and compared with simulated annealing and tabu search in simulations of small and medium-sized peptides [10]. ELP was also applied to the problem of crystal structure prediction from X-ray diffraction data [11].

3. Helix formation for non-polar amino acids

In the following, we want to demonstrate that generalized-ensemble techniques are well suited for protein research. While our understanding of the folding physics has increased considerably over the last few years, many questions remain unsolved. One example is the role of secondary structure formation in the folding process, and we focus here on the formation of α-helices.

It is long known that α-helices undergo a sharp transition toward a random coil state when the temperature is increased. The characteristics of this so-called helix–coil transition have been studied extensively [12]. Helix formation was also researched in [13] for polyalanine, polyvaline and polyglycine chains of chain length $N = 10$. each. The computational results of the latter work indicate that alanine has a strong propensity to form helices in gas phase while glycine shows no such tendency. Helix formation was observed for valine, but is much weaker than for alanine. However, recent experimental results of alanine and valine based peptides rather indicate that valine has in gas phase a higher helix propensity than alanine [14]. For this reason, we have repeated the simulations of [13] and extended them to homopolymer chains with up to 20 monomers.

Our investigation of helix formation for these three homopolymers is based on a detailed, all-atom representation of these molecules. The interactions between the atoms are

described by a standard force field, ECEPP/2 [15], as implemented in the program package SMMP [16]:

$$E_{\mathrm{ECEPP/2}} = \mathrm{E_C} + \mathrm{E_{LJ}} + \mathrm{E_{HB}} + E_{\mathrm{tor}}, \tag{5}$$

$$E_{\mathrm{C}} = \sum_{(i,j)} \frac{332 q_i q_j}{\varepsilon \tau_{ij}}, \tag{6}$$

$$E_{\mathrm{LJ}} = \sum_{i,j} \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} \right), \tag{7}$$

$$E_{\mathrm{HB}} = \sum_{(i,j)} \left(\frac{C_{ij}}{r_{ij}^{12}} - \frac{D_{ij}}{r_{ij}^{10}} \right), \tag{8}$$

$$E_{\mathrm{tor}} = \sum_{l} U_l (1 \pm \cos(n_l \chi_l)). \tag{9}$$

Here, r_{ij} (in Å) is the distance between the atoms i and j, and χ_l is the lth torsion angle. The peptide bond angles are set to their common value $\omega = 180^\circ$. We further assume that $\epsilon = 2$ in the protein interior.

Our results rely on multicanonical simulations of 2,000,000 sweeps for $\mathrm{Ala_0}$, $\mathrm{Gly_{10}}$ and $\mathrm{Val_{10}}$ and of 4,000,000 sweeps for $\mathrm{Ala_{20}}$, $\mathrm{Gly_{20}}$ and $\mathrm{Val_{20}}$. The simulations start with a random configurations and the first 10,000 sweeps are discarded to allow for equilibration. One of the quantities that we monitor in our simulations is $q_{\mathrm{H}} = \langle n_{\mathrm{H}}(T) \rangle / (N-2)$, i.e. the average number of helical residues divided by the number of residues that can be part of an α-helix. A residue is considered as "helical" if its backbone dihedral angles (ϕ, ψ)) take values in the range ($-70^\circ \pm 30^\circ$, $-37^\circ \pm 30^\circ$) [13]. The normalization factor $N-2$ is chosen instead of the number of residues N, because the terminal residues are flexible and usually not part of an α-helix. We remark that we have verified the usefulness of this definition by comparing it for some configurations with values calculated from the (i, $i+4$) backbone hydrogen bonding that is characteristic for α-helices.

We have displayed q_{H} in Fig. 1a–c for polyvaline, polyalanine and polyglycine, respectively. Data for chains of length $N = 10$ and 20 are plotted. The corresponding values for the specific heat

$$C(T) = \frac{[(\langle E^2 \rangle(T) - \langle E \rangle^2(T))/N]}{(K_{\mathrm{B}} T)^2} \tag{10}$$

are drawn in the inlets. In Fig. 1a and b, we observe for both polyvaline and polyalanine a helix–coil transition that separates a high-temperature phase with disordered coil configurations from a low-temperature phase that is dominated by helical configurations. This transition becomes more pronounced as the chain length increases from $N = 10$ to 20. The corresponding transition temperatures are determined from the location of the peaks in the specific heat shown in the inlets. For polyalanine the helix–coil transition temperature increases from $T_{\mathrm{c}} = 423(20)$ K for $N = 10$ to $T_{\mathrm{c}} = 510(10)$ K for $N = 20$. The corresponding temperatures for

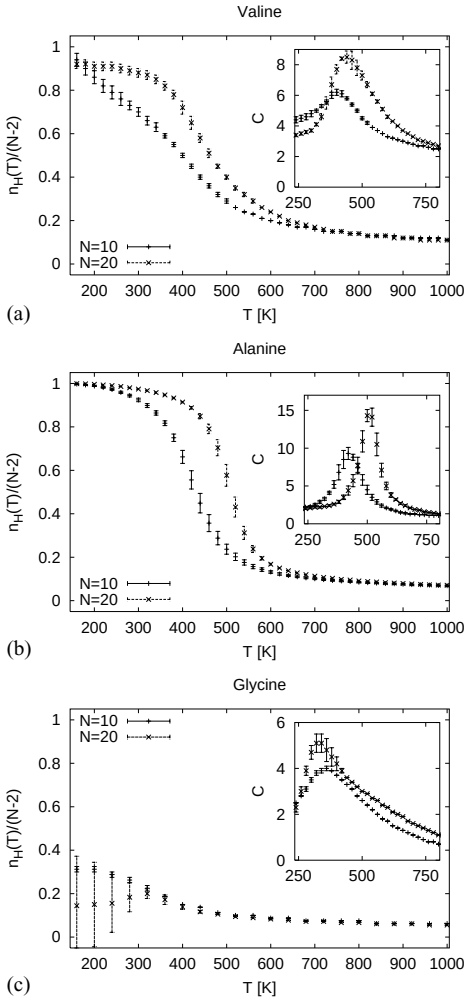

Fig. 1. Helicity q_{H} of (a) polyvaline, (b) polyalanine and (c) polyglycine as function of temperature. The corresponding values for the specific heat $C(T)$ are shown in the inlets.

polyvaline are lower: $T_{\mathrm{c}} = 405(25)$ K for $N = 10$ and $T_{\mathrm{c}} = 435(15)$ K for $N = 20$. On the other hand, no helix formation is observed for polyglycine. The peaks in the specific heat at $T_{\mathrm{c}} = 360(30)$ K ($N = 10$) and $T_{\mathrm{c}} = 330(30)$ K ($N = 20$) rather correspond to a collapse transition associated with a sharp decrease in the radius of gyration (data not shown). For comparison, we have drawn in Fig. 2 q_{H} for all three homopolymers, restricting ourselves to the case $N = 20$. The corresponding specific heat curves are shown again in the inlet. As one can see from both the curves in $q_{\mathrm{H}}(T)$ and the heights of the peaks in the specific heat $C(T)$, the helix–coil transition is stronger for polyalanine than for polyvaline, and missing for polyglycine. At $T = 300$ K, well below the helix–coil transition temperatures T_{c}, 88(2)% of the valine residues are in a helical state, while this number increases for polyalanine to 97(1)%. On the other hand only 33(3)% of the glycine residues have dihedral angles commonly found in α-helices.

Table 1 lists for the three homopolymers the differences in energy ΔE, free energy ΔG and entropy $T\Delta S$ between

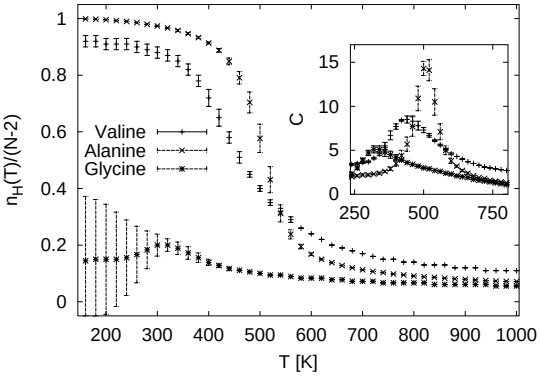

Fig. 2. Helicity q_H of Ala_{20}, Val_{20} and Gly_{20} as a function of temperature. The corresponding values for the specific heat $C(T)$ are shown in the inlets.

helical and coil configurations at this temperature. Here we define a coil configuration as one that has not more than two consecutive helical residues (i.e., residues where the pair of dihedral angles (ϕ, ψ) takes values in the range ($-70^\circ \pm 30^\circ$, $-37^\circ \pm 30^\circ$)). On the other hand, a helical configuration is defined as one where at least 60% of the residues are part of a single, extended helix. We are forced to relax this condition for polyglycine where we find no more than nine consecutive residues in a helix.

For $N = 20$, we find that for polyalanine helical configurations have an internal energy that is ≈30 kcal/mol lower than the one of coil configurations. This large energy difference can easily overcome the loss in entropy of $T\,\Delta S \approx -17$ kcal/mol when forming a helix, leading to a free-energy difference of $\Delta G \approx -13$ kcal/mol by that helical configurations are favored over coil configurations. The internal energy difference for polyvaline chains of the same length is with $\Delta E \approx -17$ kcal/mol considerably smaller than for polyalanine but still large enough to overcome the loss in entropy of $T\,\Delta S \approx -11$ kcal/mol when forming a helix. As a consequence, helical configurations are for polyvaline still favored by a free-energy difference of $\Delta G \approx -6$ kcal/mol over coil configurations. On the other hand, the energy gain by helix formation is for polyglycine chains of length $N = 20$ only $\Delta E \approx -3$ kcal/mol and not enough to overcome the associate loss in entropy of $T\,\Delta S \approx -5$ kcal/mol. Hence, coil configurations are favored over helical configurations, and the free-energy difference of $\Delta G \approx 2$ kcal/mol is positive.

We conclude that our data reaffirm qualitatively the previous results of [13] in that we find a strong helix forming tendency for polyalanine, no helix propensity for polyglycine and intermediate behavior for polyvaline. This ranking in helix propensities is different from the one reported in [14] where experimental results for homoploymers of the three amino acids are reported. Measurements of the three homopolymers indicate that in gas phase valine is a stronger helix former than alanine while again no helix formation is observed for glycine [14].

We remark that the ranking of helix propensities for the three types of homopolymers as calculated from our computer simulations is in agreement with the one measured in experiments on the solvated molecules. Hence, one possible explanation for the disagreement between our computational data and the results from the gas phase experiments in [14] is that the ECEPP/2 force field does not describe accurately molecules in gas phase. Since the force filed was developed to describe the structures of solvated proteins, it may be that its parameters are optimized for this environment and not a good description of molecules in gas phase. However, we have currently not sufficient data to test this conjecture. Another possible explanation for the discrepancy is that our computer simulations do not describe correctly the experiments that required partial charges on the ends of the homopolymers. Such charged polymers cannot be studied with the current version of SMMP, and the different charge distribution maybe responsible for the differences in our results.

4. Helix formation and folding

In order to understand in greater detail the relation between secondary structure formation and folding we have further studied the artificial peptide Ala_{10}–Gly_5–Ala_{10} in gas phase [17]. Since polyalanine has a pronounce helix–coil transition in gas phase, we expect formation of α-helices in our peptide, and we observe indeed a sharp transition between a coil phase at high temperatures and a helical phase at low temperatures in Fig. 3a, where the average number of helical residues $\langle n_H \rangle$ is displayed as a function of temperature. The helix–coil transition temperature can be determined from the corresponding peak in the specific heat (Fig. 3c) at $T = 480 \pm 10$ K. However, we find for this quantity a second, smaller peak at the lower temperature $T_f = 265 \pm 7$ K indicating yet another transition. That transition is related to

Table 1
Free-energy differences ΔG, energy differences ΔE, and entropy differences $T\,\Delta S$ (all in kcal/mol) between helix and coil states at $T = 300$ K

Amino acid	$N = 10$			$N = 20$		
	ΔG	ΔE	$T\,\Delta S$	ΔG	ΔE	$T\,\Delta S$
Alanine	−4.1 (2)	−11.8 (8)	−7.7 (8)	−13.0 (1.2)	−30.2 (2.8)	−17.2 (1.6)
Valine	−1.1 (7)	−6.4 (1.2)	−5.3 (1.5)	−5.6 (1.6)	−16.6 (5.1)	−11.0 (6.1)
Glycine	1.5 (1)	−2.6 (1)	−3.6 (2)	2.0 (7)	−3.3 (1.6)	−5.3 (2.0)

The numbers in parentheses represent errors.

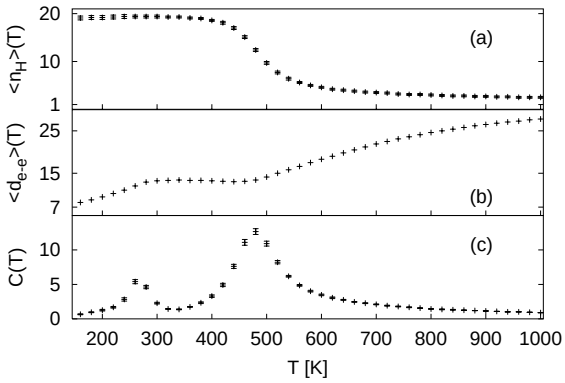

Fig. 3. (a) Average number of helical residues $\langle n_{\rm H}\rangle(T)$, (b) average end-to-end distance $\langle d_{\rm e-e}\rangle(T)$ and (c) specific heat $C(T)$ as function of temperature as calculated from simulations of Ala_{10}–Gly_5–Ala_{10} in gas phase.

a drop in the average end-to-end distance $\langle d_{\rm e-e}\rangle_{\rm T}$ as plotted in Fig. 3b. This quantity is a measure for the compactness of a protein conformation and decreases with lowering temperature. Below the helix–coil transition $T_{\rm hc}$ the decrease slows down and the curve becomes almost flat at a value of $\langle d_{\rm e-e}\rangle \approx 10$ Å indicating that there is little further change in the compactness of the molecule. However, at temperature $T_{\rm f}$ the end-to-end distance decreases again sharply toward a new value $\langle d_{\rm e-e}\rangle = 6.1$ Å. Hence, $T_{\rm f}$ marks the folding of the molecule into a defined compact structure with the two terminal ends of the peptide close together. This scenario is supported by Fig. 4 in which we display our lowest-energy configuration, a hairpin build out of two helices.

Our analysis suggests that in gas phase, Ala_{10}–Gly_5–Ala_{10} folds in a two step process. The first step is the formation of α-helices and can be characterized by a helix–coil transition temperature $T_{\rm hc} = 485 \pm 5$ K. 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. This second step can be characterized by a lower folding temperature $T_{\rm f} = 265 \pm 7$ K. We are currently investigating whether this behavior can be reproduced in gas-phase experiments [18].

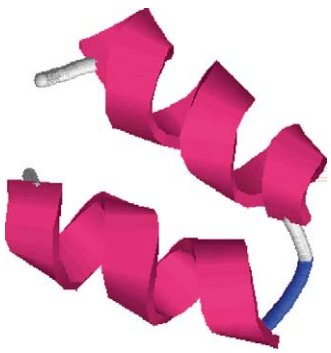

Fig. 4. Lowest-energy configuration of Ala_{10}–Gly_5–Ala_0 as obtained from multicanonical simulations in gas phase.

5. Structure prediction of small proteins

The importance of secondary structure formation in the folding process can be seen again in recent simulations of the villin headpiece subdomain, a 36-residue peptide (HP-36), whose experimental structure was determined by NMR analysis [19]. We have chosen this peptide here because it has raised in the past considerable interest in both theoretical [22] and experimental studies [23]. Its structure consists of three helices between residues 4–8, 15–18, and 23–32, respectively, which are connected by a loop and a turn. Using the above described energy landscape paving technique, an essentially correct structure was found in [10] as lowest-energy configuration when protein-solvent interactions where accounted for by a solvent accessible surface term [20]. Fig. 5a shows the experimental structure and Fig. 5b the obtained lowest-energy configuration. The computational determined structure has 95% of the native helical content and 60% of native contacts are formed. The root-mean-square deviation (RMSD) to the native structure is 5.8 Å when all backbone atoms where counted. On the other hand, if protein-solvent interactions are neglected and the peptide is simulated in gas phase, the native structure differs has $a \approx 30$ kcal/mol higher energy than the lowest-energy configuration (Fig. 5c) that was found in this simulation.

In order to understand more the differences between the gas-phase results and that with a solvent accessible surface term, parallel tempering simulations of HP-36 were performed on 20 nodes of a cluster of IBM four-ways 375 MHz SMP Thin Nodes with temperatures $T = 1000$, 900, 800, 700, 610, 560, 530, 510, 495, 485, 475, 465, 450, 420, 390, 360, 330, 300, 27S, 250 [21]. On each node, 150,000 MC sweeps were performed and a replica exchange move was attempted after each sweep. The acceptance rate was on average 30%. However, more important than the exchange rate is that replicas walk indeed across the whole range of temperatures. This was carefully monitored by us [21]. Both gas-phase simulations and such relying on a solvent-accessible surface term with the parameter set of [20] were performed.

From these parallel tempering simulations we have calculated the number of helical residues as function of and displayed in Fig. 6. Little difference is found at high temperatures. However, below the transition temperature $T \approx 490$ K the data for both simulations diverge. The helicity grows rapidly with decreasing temperature in the simulation of the solvated peptide while it stays small in gas phase. As one can see from Fig. 7, configurations in the two simulations differ also in their compactness. Here, we show the average radius of gyration $\langle r_{\rm gy}\rangle(T)$ as a function of temperature. The corresponding values for the total number of contacts $\langle n_{\rm TC}(T)\rangle$ are shown in the inlet. Both quantities indicate that configurations in gas phase are substantially more compact than the ones for the solvated molecule. For instance, at $T = 300$ K, we find $r_{\rm rgy} = 9.6(1)$ Å in gas phase compared to $r_{\rm rgy} =$

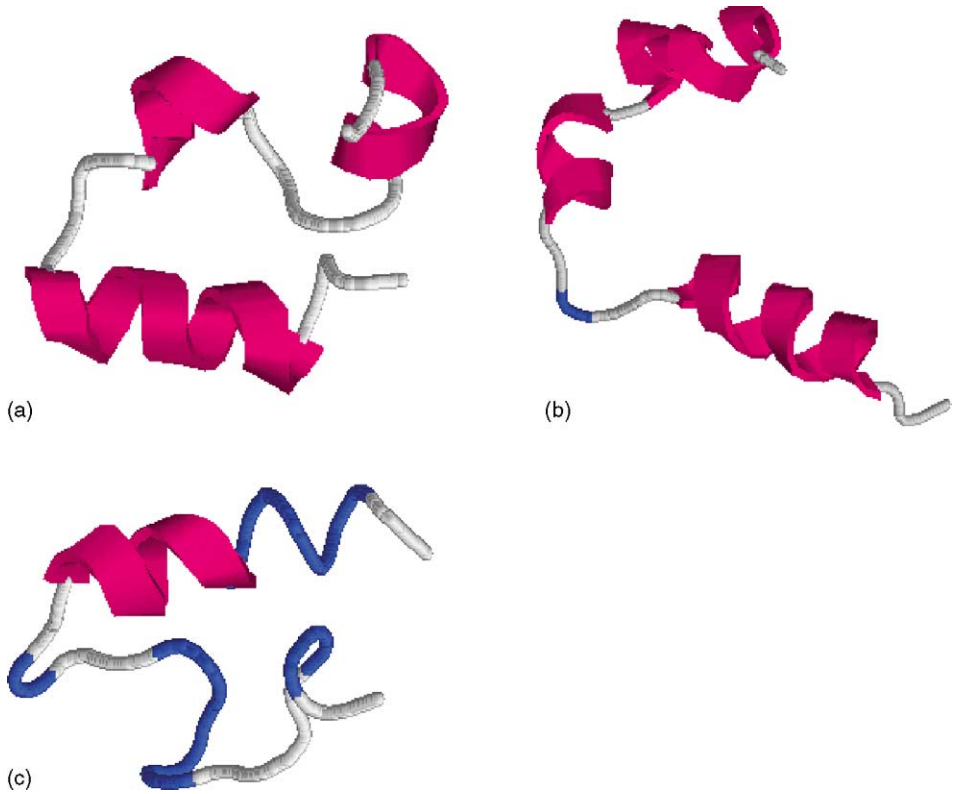

Fig. 5. (a) Experimentally determined structure of HP-36 as deposed in the PDB data bank, (b) lowest-energy configuration of HP-36 as obtained from an ELP simulation of the solvated peptide and (c) lowest-energy configuration of HP-36 as obtained from an ELP simulation of the peptide in gas phase. The figures are taken from [21].

12.5(1) Å in simulations of the solvated molecule. We conjecture that this bias toward compact configurations inhibits the formation of α-helices, and that the low-energy states of HP-36 in gas phase are characterized by large density and low helicity. We also remark that the high transition temperature of $T = 490\,\mathrm{K}$ agrees well with some recent findings in simulations of the C-terminus β-hairpin from protein G using the OPLSAA and AMBER force field [24].

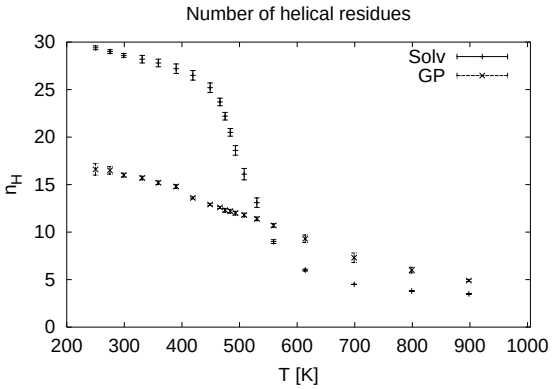

Fig. 6. Average number of helical residues $\langle n_{\mathrm{H}}\rangle(T)$ as function of temperature as obtained from simulations of HP-36 in gas phase and solvated. The figure is taken from [21].

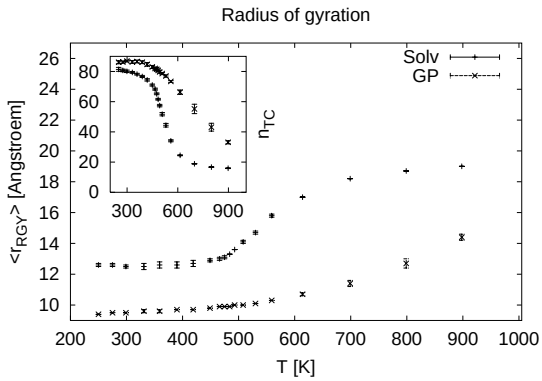

Fig. 7. Average radius of gyration $\langle r_{\mathrm{gy}}\rangle(T)$ of HP-36 as a function of temperature for both the solvated protein and in gas phase. The inlet shows the corresponding total numbers of contacts $\langle n_{\mathrm{c}}\rangle(T)$. The figure is taken from [21].

6. Conclusion

We gave a brief introduction into generalized-ensemble techniques and their applications to the protein folding problem. Our examples demonstrate that these techniques are well-suited for investigations into the physics of proteins. With the advent of these and other modern sampling techniques, all-atom simulations of proteins may now be more

restricted by the accuracy of the present energy functions than by the efficiency of the search algorithms.

Acknowledgements

Parts of the results presented in this article are also published in [17,21]. Financial supports from a research grant (CHE-9981874) of the National Science Foundation (USA) is gratefully acknowledged.

References

[1] U.H.E. Hansmann, Y. Okamoto, New Monte Carlo algorithms for protein folding, Curr. Opin. Struct. Biol. 9 (1999) 177–183.
[2] U.H.E. Hansmann, Y. Okamoto, The generalized-ensemble approach for protein folding simulations, in: D. Stauffer (Ed.), Annual Reviews of Computational Physics, vol. VI, World Scientific, Singapore, 1999, pp. 129–157.
[3] U.H.E. Hansmann, Y. Okamoto, Prediction of peptide conformation by multi-canonical algorithm: A new approach to the multiple-minima problem, J. Comp. Chem. 14 (1993) 1333–1338.
[4] B.A. Berg, T. Neubaus, Multicanonical algorithms for first order phase transitions, Phys. Lett. B267 (1991) 249–253.
[5] A.M. Ferrenberg, R.H. Swendsen, Optimized Monte Carlo data analysis, Phys. Rev. Lett. 61 (1988) 2635–2638.
[6] U.H.E. Hansmann, Y. Okamoto, Comparative study of multicanonical and simulated annealing algorithms in the protein folding problem, Physica A 212 (1994) 415–437.
[7] S. Kumar, P. Payne, M. Vásquez, Method for free-energy calculations using iterative techniques, J. Comp. Chem. 17 (1996) 1269–1275.
[8] K. Hukushima, K. Nemoto, Exchange Monte Carlo method and applications to spin glass simulations, J. Phys. Soc. (Jpn.) 65 (1996) 1604–1608;
G.J. Geyer, E.A. Thompson, Annealing Markov chain Monte Carlo with applications to ancestral inference, J. Am. Stat. Assn. 90 (431) (1995) 909–920.
[9] U.H.E. Hansmann, Parallel tempering algorithm for conformational studies of biological molecules, Chem. Phys. Lett. 281 (1997) 140–150.
[10] U.H.E. Hansmann, L. Wille, Global optimization by energy landscape paving, Phys. Rev. Lett. 88 (2002) 068105.
[11] H.P. Hsu, S.C. Lin, U.H.E. Hansmann, Energy landscape paving for X-ray structure prediction of macromolecules, Acta Cryst. A 58 (2002) 2337–2346.
[12] D. Poland, H.A. Scheraga, Theory of Helix–Coil Transitions in Biopolymers, Academic Press, New York, 1970.
[13] Y. Okamoto, U.H.E. Hansmann, Thermodynamics of helix–coil transitions studied by multicanonical algorithms, J. Phys. Chem. 99 (1995) 11276–11287.
[14] B.S. Kinnear, D.T. Kaleta, M. Kohtani, R.R. Hudgins, M.F. Jarrold, Conformations of unsolvated valine-based peptides, J. Am. Chem. Soc. 122 (2000) 9243–9256.
[15] (a) F.A. Momany, R.F. McGuire, A.W. Burgess, H.A. Scheraga, Energy parameters in polypeptides. VII. Geometric parameters, partial atomic charges, nonbonded interactions, hydrogen Bond interactions, and intrinsic torsional potentials for the naturally occurring amino acids, J. Phys. Chem. 79 (1975) 2361–2381;
(b) G. Nemethy, M.S. Pottle, H.A. Scheraga, Energy parameters in polypeptides. 9. Updating of geometrical parameters, nonbonded interactions, and hydrogen bond interactions for the naturally occurring amino acids, J. Phys. Chem. 87 (1983) 1883–1887;
(c) M.J. Sippl, G. Némethy, H.A. Scheraga, Intermolecular potentials from crystal data 6 determination of empirical potentials for O–H···O=C hydrogen bonds from packing configurations, J. Phys. Chem. 88 (1984) 6231–6233.
[16] F. Eisenmenger, U.H.E. Hansmann, S. Hayryan, C.-K. Hu, [SMMP] a modern package for simulation of proteins, Comput. Phys. Commun. 138 (2001) 192–212.
[17] N.A. Alves, U.H.E. Hansmann, Helix formation and folding in an artificial peptide, J. Chem. Phys. 117 (2002) 2337–2343.
[18] U.H.E. Hansmann, M.F. Jarrold, in preparation.
[19] C.J. McKnight, D.S. Doehring, P.T. Matsudaria, P.S. Kim, A thermostable 35-residue subdomain within villin headpiece, J. Mol. Biol. 260 (1996) 126–134.
[20] T. Ooi, M. Obatake, G. Nemethy, H.A. Scheraga, Accessible surface areas as a measure of the thermodynamic parameters of hydration of peptides, Proc. Natl. Acad. Sci. U.S.A. 84 (1987) 3086–3090.
[21] C.-Y. Lin, C.-K. Hu, U.H.E. Hansmann, Parallel tempering simulations of HP-36, Proteins Struct. Funct. Genet. 52 (2003) 436–445.
[22] M.-Y. Shen, K.F. Freed, All-atom fast protein folding: the villin headpiece, Proteins 49 (2002) 439–445.
[23] L. Vugmeyster, O. Trott, C.J. McKnight, D.P. Raleigh, A.G. Palmer, Temperature-dependent dynamics of the villin headpiece helical subdomain, an unusually small thermostable protein, J. Mol. Biol. 320 (2002) 841–854.
[24] R. Zhou, Free energy landscape of protein folding in water: explicit versus implicit solvent, Proteins 53 (2003) 148–161.