All-atom folding of the three-helix HIV accessory protein with an adaptive parallel tempering method

ARTICLE *in* PROTEINS STRUCTURE FUNCTION AND BIOINFORMATICS · DECEMBER 2004

Impact Factor: 2.63 · DOI: 10.1002/prot.20290 · Source: PubMed

3 AUTHORS, INCLUDING:

Alexander Schug
Karlsruhe Institute of Technology
65 PUBLICATIONS **1,016** CITATIONS
SEE PROFILE

Wolfgang Wenzel
Karlsruhe Institute of Technology
311 PUBLICATIONS **3,622** CITATIONS
SEE PROFILE

All-atom folding of the three-helix HIV accessory protein with an adaptive parallel tempering method

A. Schug, T. Herges and W. Wenzel
Forschungszentrum Karlsruhe, Institut für Nanotechnologie,
76021 Karlsruhe,Germany
email: wenzel@int.fzk.de
http://www.fzk.de/biostruct

July 22, 2004

Abstract

All-atom protein structure prediction from the amino acid sequence alone remains an important goal of biophysical chemistry. Recent progress in forcefield development and validation suggests that the PFF01 free-energy forcefield correctly predicts the native conformation of various helical proteins as the global optimum of its free energy surface. Reproducible protein structure prediction requires the availability of efficient optimization methods to locate the global minima of such complex potentials. Here we investigate an adapted version of the parallel tempering method as an efficient parallel stochastic optimization method for protein structure prediction. Using this approach we report the reproducible all-atom folding of the three-helix 40 amino-acid HIV accessory protein from random conformations to within 2.4Å backbone RMS deviation to the experimental structure with modest computational resources.

1 Introduction

Protein structure prediction on the basis of the amino acid sequence alone remains one of the major outstanding challenges of theoretical biophysics (1; 2; 3; 4; 5). In the post-genomic era, sequence information for proteins abounds, while structural and mechanistic information remains scarce. Theoretical methods for protein structure prediction may help to close the gap between the available sequences and structures and elucidate mechanisms of proteins that are difficult to handle experimentally (e.g. transmembrane proteins). With the development of reliable forcefields (5; 6) and robust simulation techniques (7; 4), protein structure prediction may assist in the understanding and quantitative analysis of protein-protein or protein-ligand association (8; 9) at an atomistic level.

While homology based methods have demonstrated steady progress in the past decade (10), the assessment of atomistic de-novo prediction strategies has been less favorable (2; 11; 3). Atomistic simulations of the folding process remain confined to small peptides due to their large computational cost (12; 13; 14). Free energy based methods exploit the thermodynamics hypothesis (15) that proteins in their native conformation are in thermodynamic equilibrium with their environment. Following this hypothesis, the prediction of protein tertiary structure may be possible with significantly reduced computational cost, because the native structure can be determined with global optimization methods (16; 17; 18) without recourse to the folding dynamics.

We have recently reported the rational development of a transferable all-atom free-energy forcefield (PFF01) (18) that correctly predicts the native structure of several proteins with 20-60 amino acids, as the global minimum of the free-energy surface (FES). Reproducible folding could be demonstrated for the 20-amino acid trp-cage protein using the stochastic tunneling method, the 36 amino-acid headgroup of villin (pdb-code:1VII) (12; 7) and the HIV accessory protein with a modified basin hopping technique (5). An alternate free energy forcefield was recently used to predict the structure of the B domain of staphylococcal protein A from first principles (6).

Despite these advances in the modeling of the free energy surface, little is presently known about the efficiency of various optimization strategies at the all atom level to reliably determine the global optimum of the FES. For a number of other, larger proteins there is evidence to suggest that the FES correctly predicts the native structure, but reproducible folding from random

configurations to the native structures has yet to be demonstrated. The lack of efficient optimization techniques to treat larger molecules thus emerges as one central bottleneck in all atom protein structure prediction. In the search for efficient optimization methods particular emphasis must be focused on the development of large-scale distributed optimization strategies, as presently the largest computational resources are provided in grid-based architectures.

In this investigation we used an adapted version of the parallel tempering method (PT) to predict the native structure of the HIV accessory protein (19) (pdb code 1F4I) to near experimental resolution. Our results provide additional evidence that the free-energy forcefield PFF01 stabilizes the native conformation of the HIV accessory protein as its native conformation. The development of the adapted PT technique as a global optimization method is a useful step in the development of stochastic computational strategies that are capable to exploit the characteristics of large-scale distributed architectures.

2 Methods

2.1 Free energy forcefield

We have recently developed an all-atom (with the exception of apolar CH_n groups) free-energy protein forcefield (PFF01) that models the low-energy conformations of proteins with minimal computational demand (20; 5). In the folding process at physiological conditions the degrees of freedom of a peptide are confined to rotations about single bonds. The forcefield is parameterized with the following non-bonded interactions:

$$\begin{aligned} V\left(\{\vec{r}_i\}\right) &= \sum_{ij} V_{ij}\left[\left(\frac{R_{ij}}{r_{ij}}\right)^{12} - \left(\frac{2R_{ij}}{r_{ij}}\right)^{6}\right] \\ &+ \sum_{ij}\frac{q_i q_j}{\epsilon_{g(i)g(j)} r_{ij}} + \sum_i \sigma_i A_i + \sum_{\text{hbonds}} V_{hb}. \end{aligned} \tag{1}$$

Here r_{ij} denotes the distance between atoms i and j and g(i) the type of the amino acid i. The Lennard Jones parameters (V_{ij}, R_{ij} for potential depths and equilibrium distance) depend on the type of the atom pair and were adjusted to satisfy constraints derived from as a set of 138 proteins of the PDB database (21; 20; 22). The non-trivial electrostatic interactions in proteins

are represented via group-specific dielectric constants ($\epsilon_{g(i),g(j)}$ depending on the amino-acid to which atom i belongs). The partial charges q_i and ϵ_i were previously derived in a potential-of-mean-force approach (23). Interactions with the solvent were first fit in a minimal solvent accessible surface model (24) parameterized by free energies per unit area σ_i to reproduce the enthalpies of solvation of the Gly-X-Gly family of peptides (25). A_i corresponds to the area of atom i that is in contact with a ficticious solvent. The σ_i were adjusted to stabilize the native state of the 36-amino acid headgroup of villin (pdb-code 1VII) as the global minimum of the forcefield (26). Hydrogen bonds are described via dipole-dipole interactions included in the electrostatic terms and an additional short range term for backbone-backbone hydrogen bonding (CO to NH) which depends on the OH distance, the angle between N,H and O along the bond and the angle between the CO and NH axis (18).

2.2 Optimization Technique

The parallel (or simulated) tempering technique was introduced to overcome difficulties in the evaluation of thermodynamic observables for models with very rugged potential energy surfaces (27; 28) and applied previously in several protein folding studies (29; 30; 31). In such systems, simulations at low temperatures are trapped for long times in similar metastable conformations because the energy barriers to structurally very different conformations of similar energy are very high. The idea of the PT method is to perform several concurrent simulations of different replicas of the same system at different temperatures and to exchange replicas (or temperatures) between the simulations i and j with probability:

$$p = min(1, \exp\left(-(\beta_j - \beta_i)(E_i - E_j)\right)), \tag{2}$$

where $\beta_i = 1/k_B T_i$ and E_i are the inverse temperatures and energies of the conformations respectively. The choice of the exchange probability in equation (2) ensures that all simulations remain in thermodynamic equilibrium at their respective temperatures. The PT method thus offers the opportunity to simultaneously evaluate thermodynamic expectation values over a wide range of temperatures. The exchange mechanism improves the conformational averaging of the low-temperature simulations, because the exchange with high-temperature simulations provides a mechanism to overcome the

high energy barriers between low-lying metastable conformations. Without loss of generality one may confine the exchange mechanism to simulations which are adjacent in temperature.

Applied as an optimization technique, however, only the simulation associated with the lowest temperature will ultimately yield the estimate for the global optimum, while all others may be required to generate different conformations. The computational effort rises linearly with the number of temperatures, thus the computational efficiency of a PT based optimizer thus decreases when more than the minimally required number of simulations is used. The temperature scale for the highest and lowest temperatures is set by the requirement to efficiently explore the conformational space and to accurately resolve local minima, respectively. For the proteins in questions the temperatures must thus fall in a bracket between approximately 2-600 K. Note that this temperature scale refers to the fictitious temperature scale of the simulation, the physical temperature scale is always set to 300 K by the parameterization of the implicit solvent model. The energy distribution is chosen often either geometric, which results in either geometric or constant differences of subsequent inverse temperatures.

3 Results

Given an minimal and maximal temperature and a temperature distribution scheme, the number of processors uniquely determines the values of the individual temperatures. We performed several exploratory simulations with 10-32 processors with up to 10^7 steps per temperature and found *no signs of convergence* of the energy associated with the lowest temperature towards the known energy of the optimal NMR decoy in the forcefield. An analysis of this data revealed, that even with a large number of temperatures, the exchange rates from lower to higher temperatures remained very small. As a result, one of the initial conformations tended to collapse to a low-lying metastable state. According to the exchange formula above, the acceptance for exchange is unity if the energy of the high-temperature simulation is below that of the low-temperature simulation. This conformation thus dropped to the lowest temperature of the previously chosen set and stayed there for the rest of the simulation. Since it is impossible on the low-temperature scale to escape the present local optimum, this conformation hardly ever escaped the basin of attraction of the metastable conformation. Simulations at higher tempera-

tures, on the other hand, are unlikely to explore the low-energy portion of their nearest local minimum — as a result they never attain energies sufficiently small to exchange to a lower energy, where such local minimization would be possible.

We overcame this obvious shortcoming of the original parallel tempering method with the introduction of an *adaptive temperature control* for all simulations: Starting with an initial, ordered set of geometrically distributed temperatures we monitored the exchange rate between adjacent temperatures. If the exchange rate between temperature i and i+1 was below 0.5%, then all temperatures above t_i were lowered by 10% of $t_{i+1} - t_i$. If the exchange rate was above 2%, then all temperatures above t_i were increased by the same difference. The highest temperature was maintained between 500-1000 K, otherwise all temperatures were rescaled by 10% in the appropriate direction.

Using this adjustment scheme, we ran a simulation comprising 6×10^6 steps per temperature on 20 processors of an INTEL XEON PC cluster. The low energy window of the energies of these simulations is shown in the upper panel, the associated temperatures in the lower panel of Figure (1). All simulations were started with random conformations at high temperatures to allow for rapid, unbiased relaxation of the structures and the temperature distribution. After an initial period of equilibration, the temperatures gradually relaxed into brackets, but kept fluctuating during the course of the simulation. This fluctuation appears to be required to maintain active exchange of conformations during the simulations, i.e. the adaptive temperature control should not be switched off during the simulation even though the ability to measure thermodynamic observables is lost. The lowest temperature converged to the desired minimal value of 5K. Nevertheless, the energies appeared to converge only very slowly, if at all, to the energy of the lowest energy NMR decoy (approx -119 kcal/mol).

Monitoring of the temperatures associated with individual conformations, we noted that the adaptive temperature control lead to increased temperature exchange between adjacent conformation, but the temperature windows associated with a given simulation remained confined to finite temperature band: the best conformations never visited high temperatures and vice versa. Adaptive temperature control thus appears as an important ingredient, but not as the cure for the transition state problem in protein folding. In addition, one notes that individual simulations of a traditional PT simulation do not exchange structural information. As a result, provided that all low

temperature simulations should find near native conformations, they would have to do so individually. The simulations at high temperature, on the other hand, produce essentially randomized structures by design.

To improve the computational efficiency of PT we thus introduced a *replication step*, in which the best conformation replaces the conformation at the highest temperature every 250,000 simulation steps. This mechanism results in a rapid, large scale exploration of the folding funnel around the best conformation found near the presently best conformation. The energy convergence of the continued simulation including the replication step is much accelerated, as shown in Figure (2). The figure also demonstrates that the simulations have now achieved the large exchange rates required for efficient protein folding as evidenced by the touching bands of the energy distribution. As a final step the simulation was quenched by fixing the lowest temperature to 5K for 10^7 steps. Thus, in total 30×10^6 energy evaluations were invested in each configuration, which corresponds to approximately 500 CPU hours on an 2.4 GHz INTEL XEON processor.

At this point the final conformation with the lowest energy/temperature had converged to within 1.23 / 2.46 Å backbone root mean square (RMSB) deviation to the best known decoy / NMR structure of the HIV accessory protein. The overlay of the experimental and the converged structure (see Figure (3)) demonstrates the good agreement between the conformations, the difference in NOE constraints demonstrates that not only short range, but also long range distances are correctly predicted. Considering the ensemble of final conformations, we find many structures closely resembling the native conformation. The RMSB deviations of the next four lowest conformations (all within 1.5 kcal/mol of the minimal energy) have RMSB deviations of 3.14/2.23/3.78/3.00 Å respectively to the native decoy. Figure (4) illustrates the distribution of the final conformation in energy/RMSB space and demonstrates the close correlation between energy and structural similarity.

The replication step introduced above led to a significant improvement of the relaxation rate of the overall simulation and introduced an element of true parallelism into the optimization scheme (which in the original PT method is mediated only indirectly by the temperature exchange). The use of such a replication step, while focusing the computational effort on the best available structure, has the potential to narrow the search to a small part of the conformations space. Figure (5) illustrates that this has not occurred in the present simulation, where the high temperature simulations continue to generate a large degree of structural diversity in the sampling space.

4 Conclusions

In this investigation we have introduced an efficient, modified version of the parallel tempering method for all atom protein structure prediction. Using this method we could demonstrate the unbiased folding of the 40-amino acid HIV accessory protein from random initial condition to within experimental resolution. Parallel tempering or replica exchange methods (32) have long been demonstrated to significantly speed relaxation processes on complex rugged potential energy surfaces, also in the context of all-atom protein folding (29; 31).

The approach presented here complements investigations into the protein folding problem (14; 13; 33) which aim to elucidate the molecular mechanisms of the protein folding process from the unfolded to the native state. In a recent application of the replica exchange method to the folding process of protein A (34), the folding path and transitions state ensemble could be characterized in an explicit water model.

In this work we have used a modified replica exchange approach as an optimization method to locate the global minimum of the HIV accessory protein in an all-atom model. Little is presently known about the suitability of different forcefields and/or optimization methods for all-atom protein structure prediction (35). In our approach, the free energy of the solvent and the complex electrostatic interactions of proteins in solution are represented by simple solvent-accessible surface area and group dependent dielectric constants respectively. We have previously been able to fold the trp-cage protein (4) and recently the villin headpiece (36) with this forcefield. It was also shown to stabilize various helical proteins with up to 60 amino acids in length at the all-atom level. The results presented here validate the applicability of the PFF01 forcefield to correctly predict the native state of the HIV accessory protein.

The use of optimization methods for protein structure prediction mandates the use of implicit solvent models which paramterize the free energy contributions of the solvent and of the side-chains. The accuracy of predictions of such models is subject to intense debate (37) and explicit water forcefields have been shown to yield improved structures (14) in some cases where direct comparison is possible. The results presented here demonstrate that at least for some proteins implicit solvent models can be parameterized to describe the native state of helical proteins.

The computational advantage of the optimization approach to *protein*

structure prediction stems from the fact that the native state of the protein can be obtained as the global optimum of the free energy surface without recourse to the folding dynamics. The use of optimization methods permits the determination of the native state — provided that it is correctly described by the forcefield — faster than the direct simulation of the dynamics because kinetic barriers and metastable conformations can be escaped in unphysical timescales.

In this work, this acceleration was achieved by the use of a modified version of the parallel tempering approach. We were able to show that temperature adjustments that dynamically optimize the exchange rate increase the efficiency of the search process. The optimization approach offers a rational criterion for unbiased *protein structure prediction*, whenever a particular structure occurs reproducibly in the low-energy spectrum of the simulations. The price paid for this relative certainty is the loss of direct insight into the kinetics of the folding process, its advantage a significant speedup of the folding simulation. The total computational effort invested in this study is comparable to a $5\mu s$ MD simulation, which (while substantial) is clearly insufficient to reproducibly fold a protein of this size.

The development of reliable free energy forcefields (18) and efficient optimization techniques offers an increasingly viable route for protein structure prediction at the all-atom level. Presently, the availability of efficient optimization methods, rather than inaccuracies of the forcefield, appear to be the bottleneck towards the treatment of larger proteins, in particular for beta-sheet proteins, which are computationally much more demanding. The three-helix HIV accessory protein is among the largest proteins predictably folded to date, the introduction of the modified parallel tempering method is thus an important step towards the development of parallel optimization techniques that can be efficiently implemented on massively parallel, distributed computational architectures that dominate the high-end of presently available computational resources.

We are grateful for many stimulating discussions with S. Gregurick and J. Moult, who contributed part of the code with which these simulations were performed. WW acknowledges support from the German National Science Foundation (We 1863/10-2), the BMWF and the KIST supercomputational materials lab, in particular Dr. K.-H. Lee, where some of the simulations reported here were performed.

References

[1] D. Baker and A. Sali. Protein structure prediction and structural genomics. *Science*, 294:93, 2001.

[2] J. Pillardy, C. Czaplewski, A. Liwo, J. Lee andD. R. Ripoll, R. Kamierkiewicz, Stanislaw Oldziej andW. J. Wedemeyer, K. D. Gibson, Y. A. Arnautova, J. Saunders, Y.-J. Ye, and Harold A. Scheraga. Recent improvements in prediction of protein structure by global optimization of a potential energy function. *Proc. Natl. Acad. Sci.(USA)*, 98:2329, 2001.

[3] J. Schonbrunn, W. J. Wedemeyer, and D. Baker. Protein structure prediction in 2002. *Curr. Op. Struc. Biol.*, 12:348–352, 2002.

[4] A. Schug, T. Herges, and W. Wenzel. Reproducible protein folding with the stochastisc tunneling method. *Phys. Rev. Letters*, 91:158102, 2003.

[5] T. Herges and W. Wenzel. Reproducible in-silico folding of a three-helix protein in a transferable all-atom forcefield. http://www.arXiv.org: physics/0310146, 2004.

[6] J.A. Vila, D.R. Ripoll, and H.A. Scheraga. Atomically detailed folding simulation of the b domain of staphylococcal protein a from random structures. *Proc. Natl. Acad. Sci.(USA)*, 100:14812–14816, 2004.

[7] U. H. E. Hansmann. Global optimization by energy landscpae paving. *Phys. Rev. Letters*, 88:068105, 2002.

[8] J. Drews. Drug discovery: a historical perspective. *Science*, 287:1960, 2000.

[9] H. Merlitz, B. Burghardt, and W. Wenzel. Application of the stochastic tunneling method to high throughput screening. *Chem. Phys. Lett.*, 370:68, 2003.

[10] E.E. Lattman. Casp4. *Proteins*, 44:399, 2001.

[11] R. Bonneau, J. Tsui, I. Ruczinski, D. Chivian, C. M. E. Strauss, and D. Baker. Rosetta in casp4: progress in ab-initio protein structure prediction. *Proteins*, 45:119–126, 2001.

[12] Y. Duan and P. A. Kollman. Pathways to a protein folding intermediate observed in a 1-microsecond simulation in aqueous solution. *Science*, 282:740, 1998.

[13] C. D. Snow, H. Nguyen, V. S. Panda, and M. Gruebele. Absolute comparison of simulated and experimental protein folding dynamics. *Nature*, 420:102–106, 2002.

[14] C. Simmerling, B. Strockbine, and A. Roitberg. All-atom strucutre prediction and folding simulations of a stable protein. *J. Am. Chem. Soc.*, 124:11258, 2002.

[15] C. B. Anfinsen. Principles that govern the folding of protein chains. *Science*, 181:223–230, 1973.

[16] Z. Li and H.A. Scheraga. Monte carlo minimization approach to the multiple minima problem in protein folding. *Proc. Nat. Acad. Sci. U.S.A.*, 84:6611, 1987.

[17] A. Liwo, P. Arlukowicz, C. Czaplewski, S. Oldizeij, J. Pillardy, and H.A. Scheraga. A method for optimising potential energy functions by a hierarchichal design of the potential energy landscape. *Proc. Natl. Acad. Sci.(USA)*, 99:1937–1942, 2002.

[18] T. Herges and W. Wenzel. Development of an all-atom forcefield for teriatry structure prediction of helical proteins. (submitted to Proteins), 2004.

[19] E. S. Withers-Ward, T.D. Mueller, I.S. Chen, and J. Feigon. Biochemical and structural analysis of theinteraction between the uba(2) domain of the dna repair protein hhr23a and hiv-1 vpr. *Biochemistry*, 39:14103, 2000. 1F4I.

[20] T. Herges, H. Merlitz, and W. Wenzel. Stochastic optimisation methods for biomolecular structure prediction. *J. Ass. Lab. Autom.*, 7:98–104, 2002.

[21] R. Abagyan and M. Totrov. Biased probability monte carlo conformation searches and electrostatic calculations for peptides and proteins. *J. Molec. Biol.*, 235:983–1002, 1994.

[22] T. Herges, A. Schug, B. Burghardt, and W. Wenzel. Low energy conformations of a three helix peptide in an all-atom biomolecular forcefield. accepted in Intl. J. Quantum Chem., 2004.

[23] F. Avbelj and J. Moult. Role of eloctrostatic screening in determining protein main chain conformational preferences. *Biochemistry*, 34:755–764, 1995.

[24] D. Eisenberg and A. D. McLachlan. Solvation energy in protein folding and binding. *Nature*, 319:199–203, 1986.

[25] K. A. Sharp, A. Nicholls, R. Friedman, and B. Honig. Extracting hydrophobic free energies from experimental data:relationship to protein folding and theoretical models. *Biochemistry*, 30:9686–9697, 1991.

[26] T. Herges, A. Schug, H. Merlitz, and W. Wenzel. Stochastic optimization methods for structure prediction of biomolecular nanoscale systems. *Nanotechnology*, 14:1161–1167, 2003.

[27] E. Marinari and G. Parisi. *Europ. Phys. Letters*, 451:19, 1992.

[28] A. P. Lyubartsev, A. A. Martinovski, S. V. Shevkunov, and P.N. Vorontsov-Velyaminov. *J. Chem. Phys.*, 96:1776, 1992.

[29] U.H.E. Hansmann and Y. Okamoto. Numerical comparison of three recently proposed algorithms in the protein folding problem. *J. Comput. Chem*, 18:920, 1997.

[30] U.H.E. Hansmann. *Eur. Phys. J. B*, 12:607, 1999.

[31] C.Y. Lin, C.K. Hu, and U.H. Hansmann. Parallel tempering simulations of hp-36. *Proteins*, 53:436–445, 2003.

[32] Y. Sugita and Y. Okamoto. Ab initio replica-exchange monte carlo method for cluster studies. *Chem. Phys. Lett*, 314:141–151, 1999.

[33] U. Mayor, N. R. Guydosh, C. M. Johnson, J. G. Grossmann, S. Sato, G. S. Jas, S. M. V. Freund, D. O. V. Alonso, V. Daggett, and A. R. Fersht. The complete folding pathway of a protein from nanoseconds to micorseconds. *Nature*, 421:863–867, 2003.

[34] A. E. Garcia and N. Onuchic. Folding a protein in a computer: An atomic description of the folding/unfolding of protein a. *Proc. Natl. Acad. Sci.(USA)*, 100:13898–13903, 2003.

[35] T. Yoda, Y. Sugita, and Y. Okamoto. Comparisons of force fields for proteins by generalized-ensemble simulations. *Chem. Phys. Lett*, 386:460–467, 2004.

[36] T. Herges, A. Schug, and W. Wenzel. Protein structure prediction with stochastic optimization methods: Folding and misfolding the villin headpiece. (accepted for publication in Lecture Notes in Compuational Science,(Springer, New York)), 2004.

[37] A. M. Fernandez-Escamilla, M. S. Cheung, M. C. Vega, M. Wilmanns, J. N. Onuchic, and L. Serrano. Solvation in protein folding analysis: Combination of theoretical and experimental approaches. *Proc. Natl. Acad. Sci.(USA)*, 101:2834–2839, 2004.

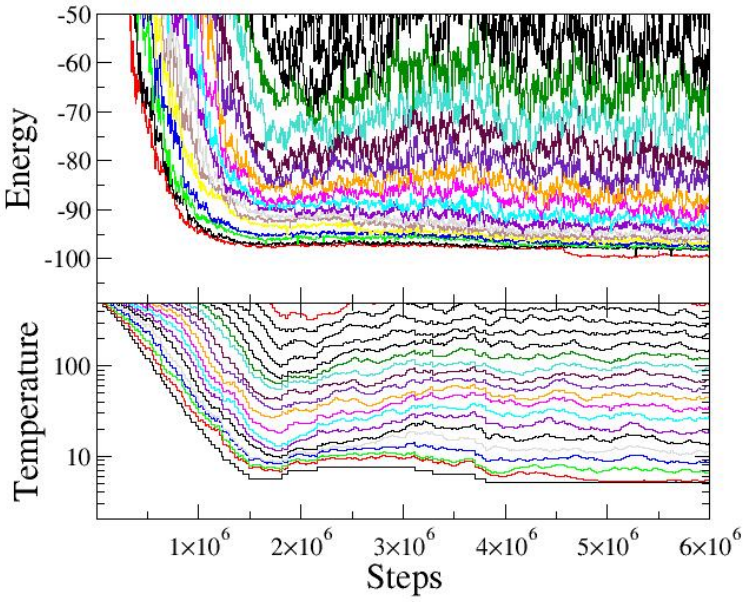

Figure 1: Energies and self-adjusting temperatures in the first 6 million PT steps for the simulation of the HIV accessory protein. The top panel shows the evolution of the energies with time. Note how prolonged periods without exchange (non-touching curves) lead to the adjustment of the temperatures (bottom panel) to regulate the exchange rate.

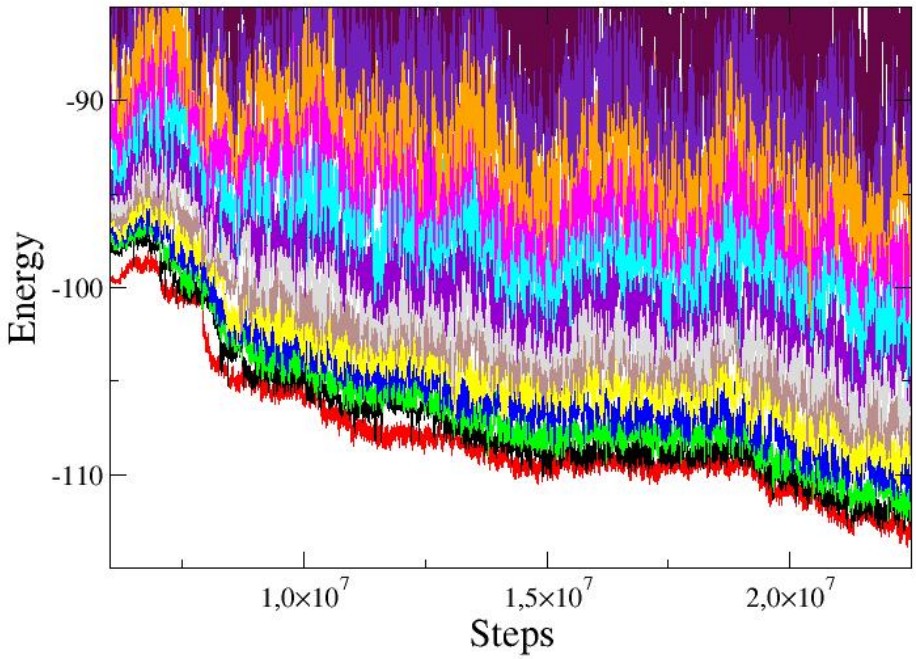

Figure 2: Energies of steps 6-22 $\times 10^6$ in the simulation of the HIV accessory protein with the adaptive parallel tempering method. The associated temperatures (not shown) have nearly equilibrated. The simulation was subsequently quenched and continued another 8 $\times 10^6$ steps to ensure that no other minima would be found (see text).

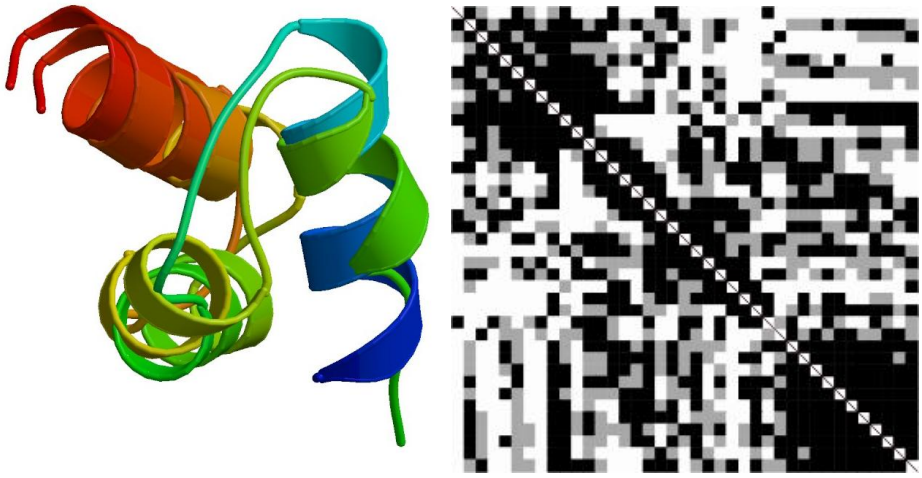

Figure 3: Overlay (left) and $C_\beta - C_\beta$ (right) distance map of the folded and the NMR structure and the predicted structure of the HIV accessory protein. A pixel in row i and colum j of the color coded distance map indicates the difference in the C_β-C_β distances of the native and the folded structure. Black (grey) squares indicate that the C_β-C_β distances of the native and the other structure differ by less than 1.5 (2.25) Å respectively. White squares indicate larger deviations.

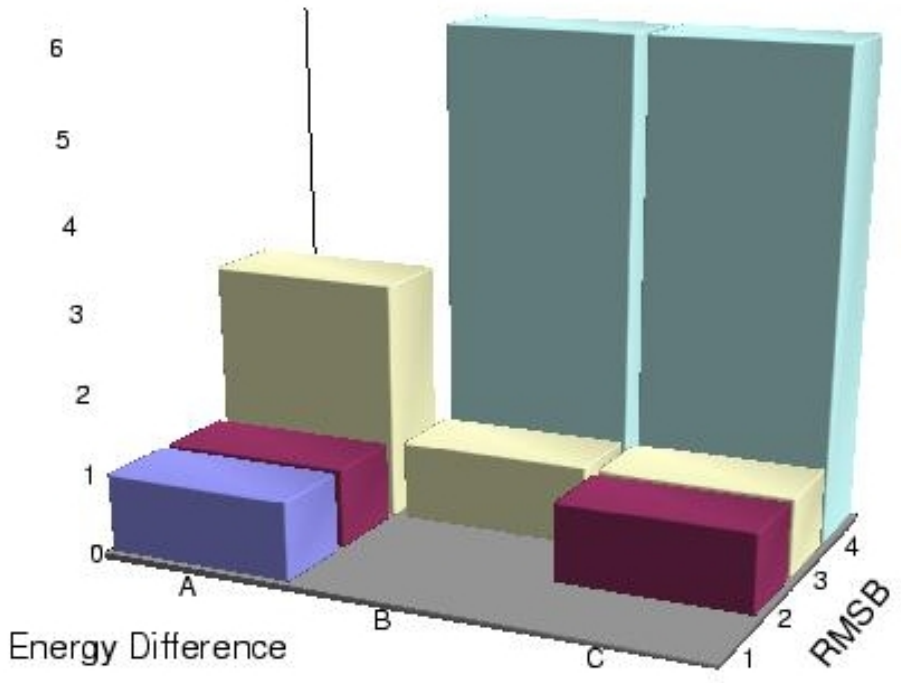

Figure 4: Histogram characterizing the distribution of energies and RMSB deviations (in Å, with respect to the best NRM decoy) of the final conformations of the modified PT simulation of the HIV accessory proteins. Classes A,B,C characterize structures within 1, 5 more than 5 kcal/mol energy difference to the best conformation found.

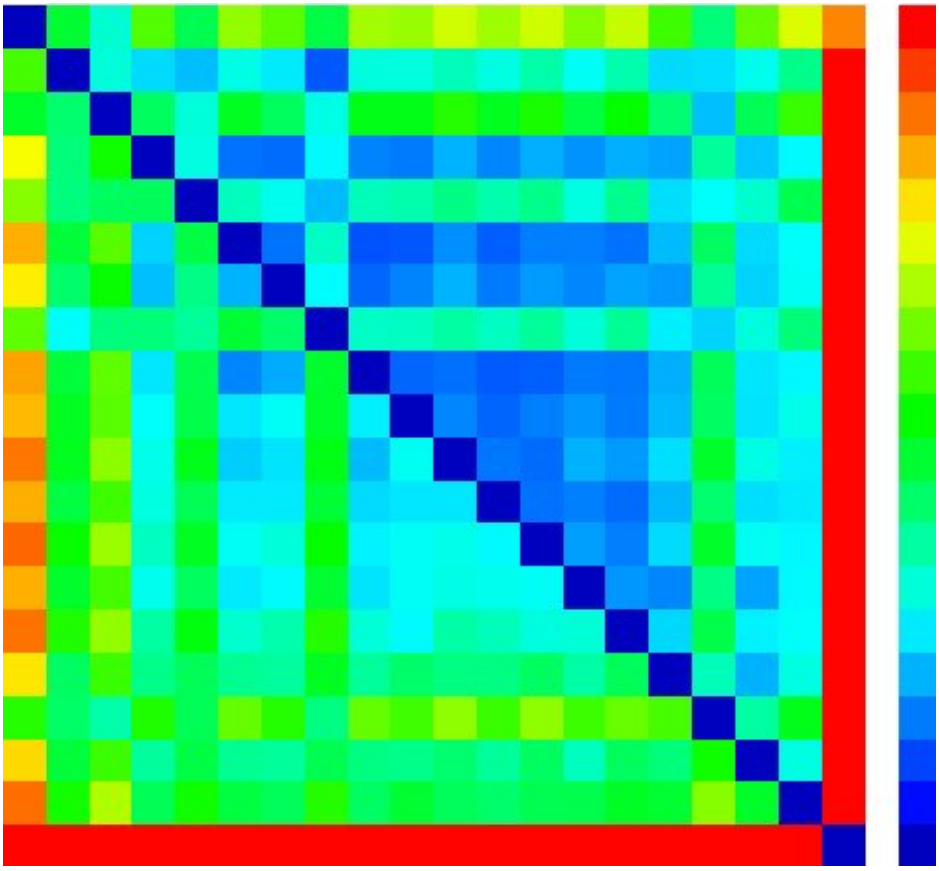

Figure 5: Illustration of the degree of similarity between the terminal configurations of the modified PT simulation of the HIV accessory protein. Each row represents a different structure (ordered by increasing energy from the top). Color codes (scale on the left, ranging from zero (blue) to more than 4 Å (red)) indicate the similarity to the other structures. The upper triangle measures backbone RMS, the lower triangle indicates the heavy atom RMSD. The red bars right and bottom indicate that the simulation at the highest temperature is very different from all other conformations. At intermediate temperatures there is a set of simulations which are similar among themselves, but still different from the native conformation.