Outline of research directions

A) Overview

Over the last few decades, computational science has extended the range of phenomena that can be investigated within the framework of physics. An important example are complex systems such as spin glasses, neural networks, to name only two. Such systems are hampered by similar problems and can be studied with similar techniques.

Proteins and their interaction are another important example of complex systems, and one that in the next few years will have more and more a need for high performance computing. This is because a new challenge has emerged after the successful deciphering of whole genomes: for most sequences we do not know the function of the corresponding proteins. Hence, there is a need to understand how the structure and function of proteins emerge from their sequence of amino acids. Such knowledge could lead to a deeper understanding of various diseases that are caused by the mis-folding of proteins, to simulating disease pathways and predicting the best target for treatment, or to the *de novo* design of novel drugs with customized properties.

Computer experiments offer one way to gain such knowledge but are extremely difficult for realistic protein models [1]: 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. The difficulties become even more pronounced if the structure of a protein depends on its interaction with other bio-molecules. Overcoming these obstacles may be one of the defining challenges of high performance computing of the next few years. It will require new hardware and new simulation techniques.

My research is concerned with the later point, the development and test of algorithms for overcoming the protein-folding problem [2]. For instance, I have previously demonstrated the feasibility of *generalized-ensemble* simulations [3] for polypeptides of up to $30 - 40$ residues. Since stable domains in proteins usually consist of 50-200 amino acids, an important research goal will be to increase the size of molecules that can be studied by computer experiments.

These algorithmic investigations will focus on simulations of carefully chosen proteins of increasing size ranging from the 28-residue Fsd-Ey [4] over the LysM-domain [5] (48 residues) up to the 75-residue apocalbindin D9K [6]. The data gained in these simulations are further used to create new analysis techniques for protein studies, to probe the mechanism of folding in small proteins, and to examine the limitations set on protein simulations by the available energy functions and solvent models.

Protein-protein interactions are the topic of another line of research. Especially interesting are the conditions under which proteins mis-fold and aggregate. As abnormally folded and aggregated proteins are related to the outbreak of various diseases, such simulations may provide insight into the mechanism of their pathogenesis. In the next years I want in my group focus on polyglutamine repeats (Huntington and other inherited neurological diseases) and the 42-residue β-amyloid peptide (Alzheimer's disease).

Protein-ligand binding and protein interaction networks belong to the same research direction and would provide an interface for collaborations with bioinformatics groups.

Related to the above described research is the development and publication of new software for simulations of protein. These programs could be included in future updates of SMMP [7], the freely available program package that was developed by my group.

B) Background: Advanced Simulation Techniques

The key-idea behind the novel techniques employed by us in protein simulations is to replace the canonical weights, that suppress the crossing of an energy barrier of height ΔE by a factor $\propto \exp(-\Delta E/k_BT)$ (k_B is the Boltzmann constant and T the temperature of the system), with such weights that allow the system to escape out of local minima. Often 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 [8] the weight $w(E)$ leads to a distribution

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

with $n(E)$ the density of states. 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: [9]

$$< \mathcal{A} >_T \;=\; \frac{\int dx\ \mathcal{A}(x)\ w^{-1}(E(x))\ e^{-E(x)/k_BT}}{\int dx\ w^{-1}(E(x))\ e^{-E(x)/k_BT}}\ . \tag{2}$$

Here, x labels the configurations. The weights $w(E)$ are not *a priori* known and estimators have to be determined by an iterative procedure described in Refs. [8, 10].

Another way of enhancing the sampling of low-energy protein configurations is parallel tempering (also known as replica exchange) [11], a technique that was first introduced to protein folding in Ref. [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 a new *global* update [11]: the exchange of conformations between two copies i and $j = i + 1$ with probability

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

This exchange of conformations leads to a faster convergence of the Markov chain than is observed in regular canonical simulations. 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. [12].

All generalized-ensemble techniques are designed to explore low energy configurations but avoiding at the same time entrapment in local minima. In *energy landscape paving* (ELP), a new optimization method that proved to be very promising in protein studies [13], this is achieved by performing low-temperature Monte Carlo simulations with a modified energy expression that steers the search away from regions already explored:

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

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. Within ELP the weight of a local minimum state decreases with the time the system stays in that minimum till it is no longer favored. The system will then explore higher energies till it falls into a new local minimum. Obviously, for $f(H(q,t)) = f(H(q))$ the method reduces to the various generalized-ensemble methods [3] (for instance for $f(H(q,t)) = \ln H(E)$ to multicanonical sampling).

The generalized-ensemble approach was first applied to protein research in Ref. [14]. Work by my group is reviewed in Ref. [3]. Applications by other groups include studies of the coil-globule transitions of a model protein [15], of the molecular mechanism of cooperative folding [16], and of the conformational flexibility of small peptides [17]; researches on the sequence dependence and folding properties of minimal protein models [18, 19], and on folding intermediates in a lattice model [20], to name only a few examples. A critical evaluation of various generalized-ensemble techniques can be found in Ref. [21].

C) Research Design

1) Algorithms for Protein Simulations

Over the past few years, I have extended the size of molecules that can be studied with generalized-ensemble techniques from previously $\lesssim 10$ residues up to proteins of $30-40$ residues [22, 23]. The first line of research aims at further advancing this approach in order that simulation of stable domains in proteins (usually consisting of 50-200 amino acids) become possible.

The computational effort increases in multicanonical simulations with the number of residues as $\approx N^4$ [24]. While this is a much better numerical performance than in canonical simulations where one would expect a supercritical slowing down (i.e. the computer time would grow as $\propto e^{aN}$ with a an unknown constant), this scaling limits the size of systems that can be studied. In general, the computational effort in generalized-ensemble algorithms scales as $\propto X^2$ where X is the variable in which one wants a flat distribution. This is because these algorithms generate an unbiased $1D$ random walk in the ensemble coordinate. In the multicanonical algorithm the coordinate is the potential energy $X = E$. Since $E \propto N^2$ the scaling relation for multicanonical simulations

is recovered. Hence, a better scaling of the computer time can be obtained by choosing a more appropriate ensemble coordinate than the energy. We have demonstrated this recently for the 36 residue villin headpiece subdomain HP-36 [25]. This molecule consists of three helices held together by a loop and a turn [25], i.e. has only α-helices as secondary structure elements. Hence, the helicity is a natural "order parameter" for the molecule and leads as an ensemble coordinate to faster sampling than observed with other generalized-ensemble methods such as the one used in Ref. [22].

I plan now to extend this approach to proteins that are not limited to α-helices as secondary structure elements. The first test model is the 28 residue de novo protein Fsd-Ey that has a $\beta\beta\alpha$ -fold [4]. This synthetic protein is of similar size as HP-36, but has both an α-helix and a two-stranded β-sheet as secondary structure elements. The more complicated form of Fsd-Ey will be used to explore possible parameters for generalized-ensembles other than the energy or the helicity. One possible candidate is the simple scoring function of Chang et al. [26]:

$$S(c) = \sum_{i,m} n(i,m,c)\varepsilon(i,m,c) \tag{5}$$

where $n(i,m,c)$ is the number of amino acids of type i that in configuration c are in the environment m, and $\varepsilon(i,m,c)$ is the associated score. It is assumed that an amino acid is in one of nine environments characterized by secondary structure (i.e. whether the amino acid is part of an α-helix, β-sheet, or other) and exposed area. The latter quantity is defined as the ratio of solvent accessible area of the amino acid X in the present configuration divided by the corresponding one in the extended structure of $Gly - X - Gly$, and evaluated according whether this ratio is smaller than 10%, between $10-50\%$, or larger than 50%. The scores $\varepsilon(i,m)$ were obtained by threading tests over a training set 387 proteins with low sequence homology and diverse folds [26].

Another quantity whose usefulness we will evaluate is the so-called hydrophobic ratio of Silverman [27]. In this approach one studies the hydrophobicity distribution of all residues that are enclosed by an ellipsoidal surface characterized by a distance d. Silverman has shown that the ratio of the distances at which the second and zero-order moments of this distribution vanish (the hydrophobic ratio) is quasi-invariant for native protein structures, and is larger than that of decoys [27]. This result points to the possibility of defining an "order parameter" for folding and suggests constructing a generalized ensemble with the hydrophobic ratio as ensemble coordinate. The drawback is that this ratio is poorly defined for proteins smaller than ≈ 100 residues. In addition, its evaluation is time consuming (≈ 1 s). However, Nelson Alves (FFCLRP,USP, Brazil) and I have recently introduced a modification of this quantity that can be calculated quickly and accurately for proteins larger than ≈ 40 residues [28].

In the first year, the we will test the usefulness of ensembles that rely on such newly introduced coordinates through simulations of HP-36 and Fsd-Ey. As the techniques evolve, they will be applied to increasingly larger and more complicated proteins. For

instance, research in the second year will target the LysM Domain from E.Coli Mltd, a peptidoglycan-binding bacterial protein with 48 residues that has two helices packing onto the same side of a two-stranded anti-parallel β-sheet [5]. The loop between strand 1 and helix is supposed to be the ligand-binding site. If time permits, we will also include in our investigation the 64-residue Chymotrypsin Inhibitor 2 that consist of an α-helix and three strands forming a mixed parallel and anti-parallel β-sheet. While considerably larger than the LysM domain, this protein allows comparison with experimental results [29] and another computational technique, unfolding MD simulations [31].

The experience gained in simulations of the LysM domain shall be used in the third year to attempt structure prediction of the 75-residue apo calbindin D9K (PDB-code 1CLB) [6]. This protein is a four-helix bundle composed of two EF-hand motifs where a calcium cation can bind to each of the two loops. Apo calbindin D9K is considerably larger than HP-36 or the Lysm domain and may push the simulations to the limits of what can be investigated currently in all-atom simulations. However, this protein allows one to compare results with previous work [30] that were obtained by a different method relying on a coarse-grained protein representation, the UNRES force field [32].

2) Development of Software for Protein Simulations

I believe that it is important to make new algorithms and techniques quickly available to interested researchers. For this reason, we will implement the newly developed methods in updates of our free program SMMP (Simple Molecular Mechanics for Proteins) [7]. SMMP is a modern package for simulation of proteins and available from either the program library of *Computer Physics Communications* or directly from the authors. The present version allows only simulation of isolated molecules but we are now re-writing the program to allow simulation of more than one (interacting) macromolecules. The package is written in FORTRAN but we are currently working on a C++ version. Test, modification and/or optimization of SMMP for Grid-computing are also planned.

3) Physics of Folding

The data gained in the simulations of Fsd-Ey, the LysM domain and apocalbindin D9K will also be used for investigating the characteristics of the folding process. This line of research includes the search for better analysis techniques (that will also be implemented in upcoming versions of SMMP), and the investigation into the limitations set on protein simulations by the available solvent models.

Folding involves a number of transitions between the different thermodynamic states of a protein and there is a need to describe and characterize these transitions. The superior sampling properties of generalized-ensemble techniques can be utilized for developing better methods for analysis of "phase transitions" in proteins. For instance, my co-workers and I have pointed out previously the relation between the fractal dimen-

sion of the energy landscape of small peptides and their glass transition temperatures [45] and introduced partition function zeros analysis as a tool for characterizing transitions in biomolecules [33, 34, 35]. For the case of the helix-coil transition in polyalanine this allowed us to identify exponents that characterize “phase transitions” in this molecule[35]. We will continue development and test of these techniques for protein studies and use them to investigate the folding mechanism in small proteins.

It is now widely believed that the energy landscape of proteins (in contrast to random heteropolymers) resembles a partially rough funnel with a free energy gradient toward the native structure (for a review, see, for instance, Ref. [36]). Folding occurs by a multi-pathway kinetics and the particulars of the funnel landscape determine the transitions between the different thermodynamic states [37, 38]. An important example for such transitions is the formation of α-helices or β-sheets during folding. Hence, in order to understand the folding mechanism it is important to research the relation between secondary structure formation and folding.

Such an investigation is considerably simpler if the peptide or protein is build up out of only α-helices or β-sheets (besides turns and loops). For this reason, we have extensively researched for alanine-based artificial peptides the helix-coil transition and its relation to folding [34]. For instance, our data suggest that the peptide Ala_{10}-Gly_5-Ala_{10} folds in a two-step process (reminiscent of the framework of collision-diffusion model): in the first step two α-helices are formed in what amounts to a first order transition. Afterward these helices re-arrange themselves into a U-like structure. The second step has the characteristics of a second order transition. Another molecule that we currently study is the 20-residue peptide Beta3 [39] that is build up solely out of β-sheets. We are evaluating for this molecule the order in that strands form, the role of β-hairpin formation in the folding process and whether the folding is driven by side-chain contact formation or that of hydrogen bonds.

In the above described case we have used that these proteins are build up out of either only α-helices or β-sheets. This allows in a simple way definition of an ”order parameter” for the folding process. The situation is different for $\alpha\beta$-proteins such as Fsd-Ey, the LysM-domain and Chymotrypsin Inhibitor 2. These molecules have both α-helices and β-sheets as secondary structure elements and are therefore of higher complexity. While they allow a more general study of small proteins, the problem is that there is no obvious reaction coordinate describing folding. However, such coordinate can be extracted *a posteriori* from generalized-ensemble simulations using the fact that these techniques allow one to sample whole ensembles of low-energy structures and to construct the corresponding energy landscape.

Analyzing the data from simulations of the 28-residue protein Fsd-Ey and the 48-residue LysM domain with clustering techniques, we will sample the ensemble of local minima of both proteins. For each pair will be probed whether there is a path between them that does not require crossing a free energy barrier of pre-set height. In this way, one obtains a connectivity network for the protein energy landscape. While it is interest-

ing in itself to study the topology of these networks, the main emphasize is on identifying the "optimal" path(s) that lead from high energy configurations down to the native state. Using dimension reduction techniques it is planed to identify the true degrees of freedom in the protein motion along the optimal path in the connectivity network. While protein motion is in general non-linear, we will start the investigation with principal component analysis (PCA) albeit this is a globally linear method and leads to a higher dimensional than necessary sub-space. We hope that the combination of our sophisticated sampling techniques with PCA will help identifying the true degrees of freedom and reaction coordinates for describing the folding process. We will then use these techniques to test whether the energy landscape of Fsd-Ey, the LysM-domain, Chymotrypsin Inhibitor 2, and apo calbindin D9K can be described with the funnel concept, how the tertiary structure formation is related to collapse and secondary structure formation, whether there are nucleation sides, and whether entropic or energetic factors guide the path(s) toward the native structure. The relative stability of secondary structure elements is another question that we want to probe.

The above mentioned tools will also be used by us to research the effect of various solvent representations on protein simulations. Since simulations with explicit water molecules would exceed the available computer time, we have to use an implicit water model. In most of the simulations we will follow a common practice and approximate the protein-solvent interaction by adding a solvent-accessible surface term [40] to the energy function. Such an approach accounts in a heuristic way for the hydrophobic effect that is supposed to be the dominant force in the protein-solvent interaction. Another effect, dielectric screening, can be approximated by use of a distance-dependent dielectric permittivity [41, 42]. Since such an implicit solvent approach is a crude approximation of protein-solvent interaction, it is not certain that the folded structure (as determined by X-ray or NMR experiments) will correspond to the global minimum configuration. For this reason, I propose to use generalized-ensemble techniques for investigating the dependence of simulation results on the specific solvation model. We will use the data from simulations of Fsd-Ey and later the LysM-domain to study the energy landscape of these proteins as a function of the solvent representation. Especially interesting is how the distribution of low-energy states depends on the solvent model and how it differs from the gas phase model. In this way, we will study systematically the accuracy of the model, and explore potential avenues for their betterment. Separating the effects of intramolecular and hydration interactions, such research allows one also to study the extend that folding is determined by intrinsic properties of the protein. Later on, these results (relying on a solvent-accessible surface term) will be compared with those from other continuum solvent representations such as the generalized Born model [43] or the approach introduced in Ref. [44]. An extension of this line of research could be a critical evaluation of present "physical" force fields and their comparison with statistical energy functions.

4) Mis-folding and Aggregation

As another line of research I propose to study the interactions between proteins. For many proteins, the structure is not only determined by their chemical composition (i.e. the sequence of amino acids) but also depends on the interaction with other proteins. Such environment-dependent structural changes include ligand-binding and chaperone-assisted folding of proteins, but also the unfolding of proteins in burned tissue or the auto-catalysis and aggregation of mis-folded proteins. The later case is especially interesting as abnormal protein folding and aggregation appears to be involved as a general mechanism in a number of diseases such as Alzheimer's, Huntington's or spongiform encephalopathies (prion-mediated) [46]. The most common of these diseases is Alzheimer's. Associated with its neuropathology are amyloid deposits, composed mainly of the β-amyloid peptide (βA). It is found in body fluids in a soluble form that has partial α-helical structure. In Alzheimer's disease, βA undergoes a conformational change toward a β-sheet structure in which it is insoluble and assembles in fibrils 60-90 Å in diameter. Fibrillar amyloids form lesions 10-200 μm in diameter known as senile plaques. These plaques are surrounded by degenerating and swollen nerve terminals, and found in extracellular space of the brain. The neurotoxity of the βA-peptide is related to the degree of β-aggregation. A similar situation is observed in a family of inherited neurodegenerative diseases that includes Huntington disease [47]. These polyglutamine (polyQ) disorders are characterized by long (> 35) glutamine repeats in the affected proteins forming protein aggregates that show a fibrillar morphology similar to that observed in Alzheimer [48]. Hence, the analysis of the structural changes in polyQ molecules or the βA-peptide, and their subsequent aggregation, could contribute to a developing understanding of the biogenesis of the corresponding neurological disorders [46].

We will start our research with investigating the mechanism of β-sheet versus α-helix formation in polyQ peptides. Chains of increasing length will be simulated in order to compare our results with the observed pathogenic threshold of $\sim 35 - 40$ glutamines. We expect to find as local minima the soluble α-helix form and the insoluble β-sheet structure, but other structures may also exist at room temperature as local minima in the free energy landscape. The relative weight of the different structures as a function of chain length will be determined and the separating free energy barriers measured. This will allow us estimating the life times of these conformers and to identify possible pathways between these local free-energy minima. As described in the previous section, principal component analysis will be used to identify the true degrees of freedom describing the motion along these pathways.

The autocatalytic properties of βA or polyQ fibrils let us expect that surface effects play an important role in the formation of β-sheets and the aggregation of the β-sheet form. Hence, we will simulate the molecule in the presence of hydrophobic (which will model previously aggregated molecules) or hydrophilic surfaces. We are especially interested in observing how the free-energy landscape of the peptide is modified through the presence of the surface, and how this change in the energy landscape depends on

the characteristics (especially its hydrophobicity) of the surface. We expect that such a detailed investigation of the free energy landscape of βA and its change with environment will lead to a better understanding of the mechanism of β-formation in this peptide.

Simulations and analysis will then be repeated for the 42-residue β-A peptide, and the mechanism of mis-folding and aggregation compared for both molecules. Since one expects that for this molecule environmental effects are important in the process of β-sheet formation, the free-energy landscape investigations will be performed under different conditions. Recent experiments indicate that metals such as Zn(II) and Cu(II) inhibit amyloid β-aggregation by inducing non-β-sheet aggregation [49]. This may be due to the fact that metal ions may have a role in dimerization of βA. However, these experiments are not conclusive since the inhibitor effect of the metal ions seems to be sensitive to the pH. In order to test whether and under which conditions these metal ions act as compounds that interrupt β-aggregation we will simulate the βA-peptide in the presence of metal-ions. Our aim is to study how the presence of the metal ions changes the free-energy landscape of the peptide and whether there is dimerization of βA molecules in presence of the metal ions.

5) Possible Other Research Directions

The above described research is concerned with the physics of single and interacting proteins. Since the number of known three-dimensional structures has grown large enough to allow a meaningful statistical analysis of structural information, this "physical" view point should be compared with and complemented by one of structural bioinformatics. It is likely that a combination of both approaches will be most fruitful. For instance, the results of our all-atom simulations may allow to extract rules that may be useful in the context of machine learning for prediction purposes. On the other hand, statistical analysis of protein structures is a way for developing energy functions. Hence, both topics may also become part of our research if suitable collaborations can be forged. Protein-ligand binding and protein interaction networks are other topics that could provide an interface for collaborations with bioinformatics groups.

D) Summary

I have outlined a group of research projects that uses high performance computing to study proteins and their interactions. Their center piece is the continuing development of novel algorithms (the "generalized-ensemble" approach) toward the final goal of structure prediction of stable domains in proteins (usually of order 50-200 residues). Simulating with these techniques carefully chosen proteins of up to 75 residues it is planned to explore the mechanism of folding and structural transition in these molecules, and investigate the limitations set by the accuracy of energy function. Another line of research is concerned with the appearance and aggregation of mis-folded structures of polygluatamine molecules and the 42-residue β-amyloid peptide that are both related to the pathology of neurodegenerative diseases.

References

[1] U.H.E. Hansmann, *Comp. Sci. Eng.* **5** (2003) 64.

[2] U.H.E. Hansmann and Y. Okamoto, *Curr. Opin. Struc. Biol.* **9** (1999) 177.

[3] *Protein folding* in silico - *The Quest for Better Algorithms*, in: A. Hartmann and H. Rieger (eds), *New Optimization Algorithms in Physics*, VCH-Wiley (2004).

[4] C.A. Sarisky and S.L. Mayo, *J. Mol. Biol.* **307** (2001) 1411.

[5] A. Bateman and M. Bycroft, *J. Mol. Biol.* **299** (2000) 1113.

[6] N.J. Skelton, J. Kordel and W.J. Chazin, *J. Mol. Biol.* **249** (1995) 441.

[7] F. Eisenmenger, U.H.E. Hansmann, Sh. Hayryan, C.-K. Hu, *Comp. Phys. Comm.* **138** (2001) 192.

[8] B. Berg and T. Neuhaus, *Phys. Lett.* **B267** (1991) 249; B. Berg and T. Neuhaus, *Phys. Rev. Lett.* **68** (1992) 9.

[9] A.M. Ferrenberg and R.H. Swendsen, *Phys. Rev. Lett.* **61** 2635 (1988); **63** (1989) 1658(E), and references given in the erratum.

[10] U.H.E. Hansmann and Y. Okamoto, *Physica A* **212** (1994) 415.

[11] K. Hukushima and K. Nemoto, *J. Phys. Soc. (Japan)*, **65** (1996) 1604; G.J. Geyer, *Stat. Sci.* **7** (1992) 437.

[12] U.H.E. Hansmann, *Chem. Phys. Lett.* **281** (1997) 140.

[13] U.H.E. Hansmann and L.T. Wille, *Phys. Rev. Let.*, **88** (2002) 068105.

[14] U.H.E. Hansmann and Y. Okamoto, *J. Comp. Chem* **14** (1993) 1333.

[15] M.H. Hao and H.A. Scheraga, *J. Phys. Chem. J. Phys. Chem.* **98** (1994) 4940.

[16] M.-H. Hao and H.A. Scheraga, *J. Mol. Biol.* **277** (1998) 973.

[17] Chr. Barthels and M. Karplus, *J. Phys. Chem. B* **102** (1998) 865.

[18] A. Irbäck, C. Peterson, and F. Potthast, *Phys. Rev. E* **55** (1997) 860.

[19] A. Irbäck and F. Potthast, *J. Chem. Phys.* **103** (1995) 10298.

[20] A. Kolinski, W. Galazka and J. Skolnick, *Proteins Proteins* **26** (1996) 271.

[21] D. Gront, A. Kolinski, J. Skolnick, *J. Chem. Phys.* **113** (2000) 5065.

[22] C.-Y. Lin, C.-K. Hu and U.H.E. Hansmann, *Proteins: Structure, Function and Genetics*, **52** (2003) 436.

[23] U.H.E. Hansmann, *J. Chem. Phys.* **120** (2004) 417.

[24] U.H.E. Hansmann and Y. Okamoto, *J. Chem. Phys.* **110** (1999) 1267.

[25] C.J. McKnight, D.S. Doehring, P.T. Matsudaria and P.S. Kim, *J. Mol. Biol.* **260** (1996) 126.

[26] I. Chang, M. Cieplak, R.I. Dima, A. Maritan and J.R. Banavar, *Proc. Nat. Acad. Sci. USA* **98** (2001) 14350.

[27] B.D. Silverman, *Proc. Nat. Acad. Sci. USA* **98** (2001) 4996.

[28] N.A. Alves, V. Aleksenko and U.H.E. Hansmann, submitted for publication.

[29] L.S. Itzhaki, D.E. Otzen and A.R. Fersht, *J. Mol. Biol.* **254** (1995) 260.

[30] J. Lee, A. Liwo and H.A. Scheraga, *Proc. Natl. Acad. Sci. USA* **96** (1999) 2025.

[31] A. Li and V. Daggett, *J. Mol. Biol.* **257** (1996) 412.

[32] A. Liwo, S. Oldziej, M.R. Pincus, R.J. Wawak, S. Rackovski and H.A. Scheraga, *J. Comp. Chem.* **18** (1997) 849; A. Liwo, M.R. Pincus, R.J. Wawak, S. Rackovski, S. Oldziej and H.A. Scheraga, *J. Comp. Chem.* **18** (1997) 874.

[33] N.A. Alves and U.H.E. Hansmann, *Phys. Rev. Lett.* **84** (2000) 1836.

[34] N.A. Alves and U.H.E. Hansmann, *J. Chem. Phys.*, **117** (2002) 2337.

[35] N.A. Alves and U.H.E. Hansmann, *J. Phys. Chem. B*, **107** (2003) 10284.

[36] K.A. Dill and H.S. Chan, *Nature Structural Biology* **4** (1997) 10.

[37] J.D. Bryngelson and P.G. Wolynes, *Proc. Natl. Acad. Sci. (USA)* **84** (1987) 7524.

[38] J.N. Onuchic, Z. Luhey-Schulten and P.G. Wolynes, *Annu. Rev. Phys. Chem.* **48** (1997) 545.

[39] T. Kortemme, M. Ramirez-Alvardo and L. Serrano, *Science* **281** (1998) 253.

[40] T. Ooi, M. Obatake, G. Nemethy and H.A. Scheraga, *Proc. Nat. Aca. Sci. U.S.A.* **84** (1987) 3086.

[41] B. Hingerty, R.H. Richie, T.L. Ferrel, J. Turner, *Biopolymers* **24** (1985) 427.

[42] R. Lavery, H. Sklenar, K. Zakrzewska, B. Pullman, *J. Biomol. Struct. & Dynamics* **3** (1986) 989.

[43] D. Qiu, P.S. Shenkin, F.P. Hollinger and W.C. Still, *J. Phys. Chem.* **101** (1997) 3005.

[44] P. Koehl and M. Delarue, *Proteins:Function, Structure and Genetics* bf 20 (1994) 264.

[45] N.A. Alves and U.H.E. Hansmann, *Int. J. Mod. Phys. C* **11** (2000) 301.

[46] J-C. Rochet and P.T. Lansbury, *Curr Op Struc Biol* **10** (2000) 60.

[47] L. Masino and A. Pastore, *Brain Res. Bull.* **56** (2001) 183.

[48] P.A. Temussi, L. Masino, and A. Pastore, *EMBO J.* **22** (2003) 355.

[49] Y. Yoshiike, K. Tanemura, O. Murayama, T. Akagi, T. Akagi, M. Murayama, S. Sato, X. Sun, N. Tanaka and A. Takashima, *J Biol Chem* **276** (2001) 32293.