Improving an all-atom force field

Sandipan Mohanty[*]

John von Neumann Institut für Computing, Forschungszentrum Jülich, Jülich, Germany

U. H. E. Hansmann[†]

Department of Physics, Michigan Technological University, Houghton, MI 49931, U.S.A
and John von Neumann Institut für Computing, Forschungszentrum Jülich, Jülich, Germany

Experimentally well-characterized proteins that are small enough to be computationally tractable provide useful information for refining existing all-atom force fields. This is used by us for reparametrizing a recently developed all-atom force field. Relying on high statistics parallel tempering simulations of a designed 20 residue β-sheet peptide, we propose incremental changes that improve the force field's range of applicability.

I. INTRODUCTION

Better algorithms and improved force fields have helped to better our understanding of protein folding. Both coarse-grained models and phenomenologically derived force fields have proved successful in predicting the structure of the folded state from the sequence. On the other hand, folding simulations of even short proteins in full atomistic detail have had only a comparatively limited success. This is in part due to the difficulties in describing precisely the delicate balance of interactions within the atoms of a protein, and between the protein and the surrounding solvent. Hence, an accurate parametrization of the force field takes on critical importance.

This study concerns one such force field, developed by Irbäck *et al.* [1–3], which has previously been shown to fold several small proteins with helical, β-sheet, and mixed secondary structure [1,4]. For small peptides, this Lund force field gives very reasonable estimates of the native population as a function of temperature [1]. The force field has also been used to study oligomerization of systems of small peptides [5], and mechanical as well as thermal unfolding of ubiquitin [6,7]. An implementation of the force field is available to academic users as a C++ program package with the name PROFASI [8]. The force field has also been implemented in an upcoming release of the (PYTHON/FORTRAN) program package SMMP [9].

In this paper, we use simulations of the artificial 20 residue protein beta3s [10] peptide to propose and test incremental changes for the Lund force field. The potential as presented in Ref. [1], fails to identify the native state of this protein as the global free energy minimum, whereas the minor modifications given here enable it to correctly fold beta3s. Note that a change to the energy function is an improvement only when it increases its scope of applicability, i.e., does not lose its ability to fold previously studied proteins. As the potential presented in Ref. [1] is already capable of folding a variety of small proteins of different folds, it is rather well constrained. The modifications proposed here are therefore small, and are consistent with the original ideas behind the potential.

II. MODEL AND METHODS

The model used for this study belongs in the class of all-atom protein models with fixed bond lengths and bond angles, and implicit water. With these constraints, the degrees of freedom of the protein molecules are the backbone Ramachandran ϕ and ψ angles and a number of side chain torsional angles. While these features are shared by many concurrent protein models, the model discussed here uses a phenomenological potential in which every term is short ranged. The potential has been presented in detail and discussed before in Refs. [1,3]. Here, it will be presented very briefly and qualitatively.

The potential is sequence based; that is, it does not make use of any information about the structure of the native state. It consists of four simple terms, which represent excluded volume effects, a local electrostatic term, a hydrogen bond term and an effective hydrophobic attraction:

$$E = E_{\text{ev}} + E_{\text{loc}} + E_{\text{hb}} + E_{\text{hp}}. \tag{1}$$

The excluded volume effects are represented by the E_{ev} term as a strong r^{12} repulsion between the atoms. This term takes special care of pairs of atoms separated by three covalent bonds. For such pairs, the effective radius of exclusion of atoms is taken to be larger than other pairs. This pure repulsion between atoms separated by three covalent bonds is found to help in ensuring reasonable Ramachandran angle distributions as well as side-chain angular distributions, without any explicit energy term that depends on the angles. The second term in Eq. (1), E_{loc}, represents a very limited electrostatic term. Only the partial charges on the backbone N (and attached H) and C (and attached O) are considered. Atoms in one amino acid only interact electrostatically with the other atoms of the same amino acid. These atoms are never very far away from each other and there is little room for effective screening by solvent molecules between these partial charges. For all other cases, partial charges of the atoms are ignored.

*Corresponding author. s.mohanty@fz-juelich.de
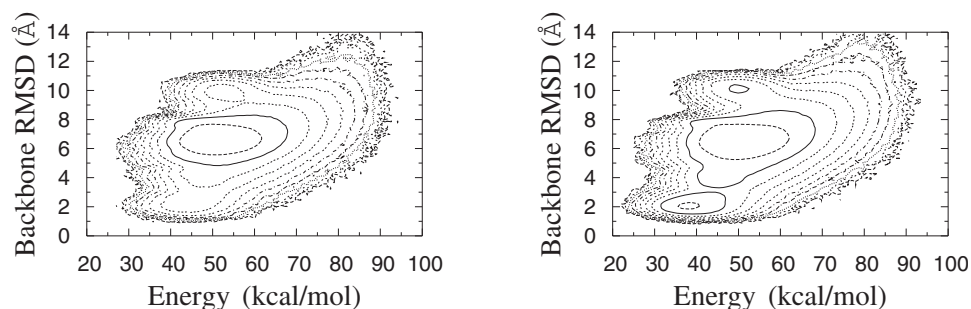
†u.hansmann@fz-juelich.de, hansmann@mtu.edu

FIG. 1. Comparison of the free-energy as a function of energy and backbone RMSD for beta3s in the (left) old version of the force field (obtained with default PROFASI version 1.0.1) and (right) that obtained after our corrections. The changes make a small shift in the weights in favor of the native minimum. Contour lines are separated by 1 k_BT. Both plots show free energy landscapes at $T \approx 283$ K in energy and temperature scales callibrated with respect to their own set of parameters.

The hydrogen bonds are represented by the E_{hb} term. This term has a (10,12)-Lennard-Jones type distance dependence and an orientation dependence that weakly favors the alignment of the four atoms N-H-O-C, in that order. Hence, the hydrogen bonds are regarded as bonds between electric dipoles: an NH dipole in an amino acid makes a hydrogen bond with a CO dipole in another.

Stable protein conformations tend to protect certain amino acids from exposure to water. This hydrophobic effect is often modeled by defining a solvent accessible hydrophobic surface, and using an energy term proportional to the area of the exposed hydrophobic surface. The hydrophobicity term E_{hp} in the Lund force field of Eq. (1) does not calculate the full hydrophobic surface area, but only uses a very rough estimate of the portion of that area that avoids contact with water by means of contact with other hydrophobic atoms. Computationally this is cheap when compared to a full surface area calculation.

The conformational space of the proteins is explored using two kinds of updates. Single angle updates for both backbone and side chain degrees of freedom set a randomly chosen torsional angle (that is free to rotate) to any value between 0 and 2π. A single backbone move of this kind can cause a very large change in the structure of the molecule but has also a high probability to be rejected. As a second kind of conformational update, a semilocal move [11] of the backbone degrees of freedom is used, which makes a local deformation of the chain leaving the atoms outside the chosen local region largely unmoved.

To study the folding and thermodynamic properties of the model, we used parallel tempering simulations, with the program package PROFASI [8], with certain modifications to the force field terms as explained below. Initial exploratory runs were done with eight temperatures in the approximate range from 273 and 373 K, distributed as a simple geometric series. For the more promising changes of the force field, longer simulations, with $1-2 \times 10^9$ elementary Monte Carlo updates per replica, with 24 temperatures (replicas) in the same range were performed.

III. RESULTS AND DISCUSSION

Starting point of our force field optimization was the observation that the designed three-stranded β-sheet protein

beta3s [10] does not fold under the form of the force field given in Ref. [1]. The free energy landscape of the protein as obtained with the original force field at $T \approx 283$ K is shown as function of backbone root mean square deviation (RMSD) to the native structure and of the potential energy in Fig. 1 (left). Nativelike states (defined by us as such with a backbone root mean square deviation, RMSD smaller than 2 Å) are observed in the simulation but do not form a free-energy minimum. A clue for the refinement of the force field was found when we observed that a modified molecule, having the same amino acid sequence but also containing Acetyl and NH2 capping groups at the N- and C-termini, respectively, led to a stable ground state that closely resembles the experimentally determined structure of beta3s. The terms of the force field which are affected by the extra capping groups were identified, and changes which have the same effect on the original molecule were listed.

An obvious change to try is to strengthen the contacts seen in the native states, such as those between the Isoleucine residues at position 3 and 18 with the Tyrosine and Tryptophan residues at positions 10 and 11, respectively. But it turns out that this can not be done indefinitely to strengthen the native state. A small increase in the hydrophobicity of Isoleucine stabilizes the native state, but a subsequent increase only results in enhancement of a set decoy states, with all the hydrophobic residues packed together. This is an illustration of a possible error that should be avoided while optimizing force fields. Optimization based on native states often results in force fields that describe the native minima of proteins well, but also stabilize decoys not considered during optimization. Since the Lund force field aims to describe the folding process, the global properties of the energy function, rather than its behavior around native folds, must be given greater importance.

Another possibility for optimization is to make some of the interactions which are stronger in presence of Acetyl and NH2 capping groups in general slightly stronger. The NH and CO dipoles in the first and the last amino acids of a protein chain are regarded as on average more exposed to water than those of the other amino acids. Their strengths were attenuated by a factor of 0.5 in the original Lund force field while in presence of capping groups the force field assigns full strengths to those dipoles. Since the observed native structure of beta3s contains one hydrogen bond involv-

TABLE I. Updated hydrophobicity table M_{IJ} of Ref. [1]. Hydrophobic amino acids are now divided into four categories. The matrix M_{IJ} represents the size of hydrophobicity interaction when an amino acid of type I is in contact with an amino acid of type J.

	I	II	III	IV
I Ala	0.0	0.1	0.1	0.1
II Met, Pro, Val		0.9	1.0	2.8
III Ile, Leu			1.1	3.0
IV Phe, Trp, Tyr				3.2

ing a CO dipole at the C terminus, an asymmetric attenuation is considered. As in the case of hydrophobicity, making too great a change in these factors turns out to be unacceptable, as it changes unacceptably the folding properties of the C-terminal hairpin of protein G, a previously studied peptide in the model.

The task of tuning the force field is further complicated by the fact that with a change of force field parameters, the energy and temperature scales of the model have to be recalibrated. The model uses an internal temperature scale. The ratio of any pair of melting temperatures calculated in the model is a prediction rather than an adjustable parameter. However, to interpret them in a commonly used unit of temperature, the calculated melting temperature for any one protein has to be identified with the corresponding experimental value. All other temperatures and energies are then determined through their ratios with that one temperature. In Ref. [1], the 20 residue helical protein Trp cage was chosen for this purpose. The melting temperature of Trp cage (PDB id: 1L2Y) in the model, using the force field of Ref. [1], is 0.47 model units. This was identified with the experimentally reported melting temperature of 315 K [12] for the peptide. When the force field is modified, the effects on the melting temperature of this peptide should be calculated, and a new scale factor between model and common temperature units determined. The new scale factor should be used for all simulations done with the modified force field. A candidate modified force field, that aparently succeeds in a simulation of the target molecule may have to be rejected if it changes the temperature scale in such a way that the aparently successful folding only occurs below the freezing point of water in the new temperature scale.

A combination of minor modifications that enhances beta3s folding at the right temperatures, without changing the folding behavior of previous sequences has been determined through repeated folding simulations. The updated hydrophobicity table is given in Table I. The suppression factors now used for the hydrogen bond dipole strengths at the

ends of the chain are N-terminus: NH3 (0.25), CO (0.7); C-terminus: NH (0.7), CO2 (0.5). These values and the data in Table I, along with a slight increase in strength of the term E_{loc}, so that κ_{loc} in Ref. [1] is changed to 125, augment the force field parameters listed in Ref. [3]. It turns out that the effect of our combined changes on the temperature scale is very small, though versions that were tried in the intermediate phases of the modification procedure affected the temperature scale by one percent or more. The melting temperature of Trp cage in the internal units of the model is now found to be 0.470, essentially the same given in the version of the force field presented in Ref. [1]. The model temperatures are therefore translated to K by equating 0.47 model units to 315 K.

The resulting free-energy landscape ($T \approx 283$ K) as obtained with the improved Lund force field is also displayed in Fig. 1 (right). A well-formed free-energy-minimum is now visible for configurations with a rmsd smaller than 2 Å. Note that in a 2D projection of the high-dimensional free energy landscape, several distinct minima might be projected onto the same location, as is the case with the minimum seen at RMSD of 6 Å that indeed includes various distinct structures. Several conformational characteristics of the molecule, such as the radius of gyration vary quite a lot among structures contributing to that minimum. One common characteristics is, however, that they all form a tight hydrophobic core with all the tryptophan, isoleucine and tyrosine groups packed close together. This is a known side effect of the pair-wise additive form of the hydrophobicity term in the potential, which fails to take into account multibody effects which become important for such conformations. Since this molecule has a large number of highly hydrophobic residues, the potential has a built in weakness for the study of this molecule. Nevertheless, the simulations succeed in escaping that minimum and finding the native state as the second most significant minimum.

In Table II, we indicate the backward compatibility of our force field modifications. The melting temperature for the helical proteins (Trp Cage and Fs) is estimated using a two state fit of the helix content observed in the simulations. For β-sheet proteins, a similar fit to the average hydrophobicity energy term is used, as in Ref. [1]. Using precisely the same analysis on both versions of the force field, we find that except for case of LLM, the thermodynamics of these small peptides remains largely unaffected by our changes. The changed force field also folds beta3s, which could not be folded with the older force field.

The molecule beta3s has a small probability in the model to be in the experimentally reported native state at $T \approx 283$ K. About 12% of the configurations have RMSDs smaller than 3 Å, i.e., can be considered to be similar to the

TABLE II. Estimated melting temperatures for a set of previously studied proteins of both helical and β-sheet secondary structure shows that our force field corrections are backward compatible.

	Trp Cg	Fs	Gb1p	Gb1m2	Gb1m3	betanova	LLM	beta3s
Old	315	304±1	297±1	321±1	322±1	314±1	314±1	
New	315	299±1	297±1	322±1	324±1	313±1	302±1	297±1

native state. This compares with an experimentally observed propensity of 13–31 % at 284 K for this structure [10]. A more detailed discussion of the folding behavior of beta3s as observed with the modified Lund force field, as well as certain other small proteins not previously studied with this force field, can be found in Ref. [4].

IV. CONCLUSION

Motivated by simulations of the small beta3s peptide we have proposed a modification of the recently introduced Lund force field that increase its range of applications. The modifications are not only backward compatible, but we

were also able to find the native structure in a number of other proteins that have not been studied with this force field previously. While further fine-tuning may be required, we believe that our results indicate a certain universality of the Lund force field, at least for small proteins.

ACKNOWLEDGMENTS

[1] Anders Irbäck and Sandipan Mohanty, Biophys. J. **88**, 1560 (2005).

[2] Anders Irbäck, Björn Samuelsson, Fredrik Sjunnesson, and Stefan Wallin, Biophys. J. **85**, 1466 (2003).

[3] Anders Irbäck and Fredrik Sjunnesson, Proteins **56**, 110 (2004).

[4] Sandipan Mohanty and U. H. E. Hansmann, Biophys. J. **92**, 3573 (2006).

[5] Giorgio Favrin, Anders Irbäck, and Sandipan Mohanty, Biophys. J. **87**, 3657 (2004).

[6] A. Irbäck and S. Mitternacht, Proteins: Struct., Funct., Bioinf. **65**, 759 (2006).

[7] Anders Irbäck, Simon Mitternacht, and Sandipan Mohanty, Proc. Natl. Acad. Sci. U.S.A. **102**, 13427 (2005).

[8] Anders Irbäck and Sandipan Mohanty, J. Comput. Chem. **27**, 1548 (2006).

[9] F. Eisenmenger, U. H. E. Hansmann, S. Hayryan, and C. K. Hu, Comput. Phys. Commun. **138**, 192 (2007).

[10] Eva de Alba, Jorge Santoro, Manuel Rico, and M. Angeles Jiménez, Protein Sci. **8**, 854 (1999).

[11] Giorgio Favrin, Anders Irbäck, and Fredrik Sjunnesson, J. Chem. Phys. **114**, 8154 (2001).

[12] J. W. Neidigh, R. M. Fesinmeyer, and N. H. Andersen, Nat. Struct. Biol. **9**, 425 (2002).