Inclusion of Many-Body Effects in the Additive CHARMM Protein CMAP Potential Results in Enhanced Cooperativity of $\alpha$-Helix and $\beta$-Hairpin Formation

Robert B. Best,†* Jeetain Mittal,‡ Michael Feig,§ and Alexander D. MacKerell, Jr.¶*

†Department of Chemistry, University of Cambridge, Cambridge, United Kingdom; ‡Department of Chemical Engineering, Lehigh University, Bethlehem, Pennsylvania; §Department of Biochemistry and Molecular Biology and Department of Chemistry, Michigan State University, East Lansing, Michigan; and ¶Department of Pharmaceutical Sciences, School of Pharmacy, University of Maryland, Baltimore, Maryland

ABSTRACT Folding simulations on peptides and proteins using empirical force fields have demonstrated the sensitivity of the results to details of the backbone potential. A recently revised version of the additive CHARMM protein force field, which includes optimization of the backbone CMAP potential to achieve good balance between different types of secondary structure, correcting the $\alpha$-helical bias present in the former CHARMM22/CMAP energy function, is shown to result in improved cooperativity for the helix-coil transition. This is due to retention of the empirical corrections introduced in the original CMAP to reproduce folded protein structures—corrections that capture many-body effects missing from an energy surface fitted to gas phase calculations on dipeptides. The experimental temperature dependence of helix formation in (AAQAA)$_2$ and parameters for helix nucleation and elongation are in much better agreement with experiment than those obtained with other recent force fields. In contrast, CMAP parameters derived by fitting to a vacuum quantum mechanical surface for the alanine dipeptide do not reproduce the enhanced cooperativity, showing that the empirical backbone corrections, and not some other feature of the force field, are responsible. We also find that the cooperativity of $\beta$-hairpin formation is much improved relative to other force fields we have studied. Comparison with $(\phi,\psi)$ distributions from the Protein Data Bank further justifies the inclusion of many-body effects in the CMAP. These results suggest that the revised energy function will be suitable for both simulations of unfolded or intrinsically disordered proteins and for investigating protein-folding mechanisms.

INTRODUCTION

Computer simulations are now starting to access the time and length scales necessary for ab initio simulation of protein folding using molecular dynamics (MD) simulations. Several impressive MD studies have demonstrated folding of miniproteins in explicit solvent (1,2), providing an implicit validation of the atomistic energy functions or force fields that have largely been used for protein folding to date (3). Nonetheless, such ab initio folding simulations have also shown that some force fields tend to stabilize $\alpha$-helical structure, whereas others favor $\beta$-structure (4–6). This bias arises because a common treatment of the $\phi,\psi$ energy surface of the polypeptide backbone is usually used for all amino acids in each force field. Therefore, even a slight bias toward one type of secondary structure at the level of a single residue can represent a large error when summed over the whole protein. It has recently been shown that NMR data for weakly structured peptides in solution provides a highly sensitive probe of such secondary structure bias, which can be used to both validate and refine existing force fields (6–12); although experimental data for folded proteins are also useful, they are relatively insensitive to biases in the backbone potential (13,14). Indeed, the magnitude of the resulting corrections ($<0.5 k_B T$) is smaller than the typical difference between the $\phi,\psi$ potential energy surface of a force field and the quantum mechanical energy surfaces used to parametrize the force field (8). Force fields fine-tuned in this way have been shown to fold peptides and proteins with $\alpha$, $\beta$, and mixed $\alpha/\beta$ structure (9,11,15–17).

Nevertheless, a remaining challenge is that many-body effects are missing from currently used additive force fields, and these make a significant contribution to the total energy in the condensed phase (18). In the context of the helix-coil transition, many-body effects may be manifested by a stronger hydrogen bond formation in helices, due to electronic polarization (19,20), and possibly charge-transfer effects (21). The abovementioned efforts to reparameterize the backbone in additive force fields therefore face a frustration between reproducing the energy surface for unstructured or unfolded proteins, and that for folded peptides and proteins, as has been recognized in several recent studies (22,23). A related feature of additive force fields is that their reproduction of protein folding thermodynamics is generally quite inaccurate. Thus, although some replica-exchange studies of peptides and miniproteins in explicit solvent may obtain melting temperatures close to experiment (9,12,23–25), a common feature of most studies is that the unfolding transition is generally much less cooperative than in experiment (9,23,26–29), that is, it takes place over a much wider range of temperature relative to experiment. More precisely, this is a consequence of the entropy and enthalpy of folding being too small, typically around half of the experimental values (8,9,23,24). Earlier work on the helix-coil transition with Amber ff03 and ff99SB and modified variants showed that both the enthalpy and...
entropy of helix extension were too small (8). Although it may be possible to capture some of the missing many-body effects in a new generation of polarizable force fields, such force fields are currently not widely used and their use incurs a computational cost that exacerbates the sampling problem inherent in studies of protein folding.

In this work, we apply the revised backbone potential for the additive CHARMM (30) force field, referred to as CHARMM36, developed in conjunction with new side-chain parameters (13) to the helix coil equilibrium of model alanine-based peptides. The new model is able to match the experimental data at room temperature very well, as has been achieved previously for other force fields. More notably, however, it is shown that the use of the CHARMM CMAP (31,32) backbone potential allows some many-body effects to be included in the additive force field, resulting in improved cooperativity for the helix-coil transition and α-helix nucleation and extension parameters, and also for formation of β-hairpins. We show unequivocally that this effect arises directly from the CMAP and not from any other detail of the force field, by considering an alternative model in which we fit the CMAP to a gas phase quantum mechanical surface for the alanine dipeptide. Such improved cooperativity is expected to be beneficial for the reproduction of both the structure of unfolded proteins and protein folding mechanisms and dynamics (33–35).

RESULTS AND DISCUSSION

We focus here mainly on a simple prototype for protein folding: the helix-coil transition. As a model, we studied the experimentally well-characterized 15-residue helix (42), Ac-(AAQAA)3-NH2, for which temperature-dependent chemical shifts have been determined. Due to its low sequence complexity and the availability of comprehensive experimental data, this system formed part of the training set for parametrizing helix propensity at 300 K in the recent CHARMM36 force field (13). Equilibrium conformational distributions at different temperatures were obtained using a replica exchange MD algorithm. This has the advantage of both enhancing sampling at low temperatures and efficiently generating equilibrium data as a function of temperature.

The average fraction helix computed from the simulations with the CHARMM36 force field is shown in Fig. 1A along with an experimental estimate based on chemical shifts (a similar fraction helix was obtained from circular dichroism data (42,43)). As expected, given its inclusion in the force field parametrization, the fraction helix in the simulations matches that in experiment well at 300 K, significantly better than CHARMM22/CMAP (31,44) (Fig. 1, inset). Furthermore, the temperature dependence of the transition—which was not part of the force field optimization—is also in more satisfactory agreement with experiment. The transition is cooperative and occurs over a small temperature range, as in experiment, features that were not obtained with other force fields. Interestingly, analysis of the CHARMM22/CMAP results shows similar cooperative behavior, albeit upshifted by ~200 K; the shift of the melting curve results from the known bias of CHARMM22/CMAP toward helical structure. This contrasts with the lack of cooperativity obtained with other force fields that reasonably reproduce the equilibrium between helical and extended conformations: data for the AMBER ff03* force field are shown in Fig. 1, and are representative of the results of other additive force fields (8). We therefore suspected that this improvement came from the CMAP potential, which is specific to CHARMM. CMAP is a two-dimensional cubic spline V_{CMAP}(\phi,\psi) added to more accurately parameterize the \phi,\psi energy surface of the backbone. In earlier studies it was found that parametrizing the CMAP purely based on a quantum mechanical vacuum energy surface for the alanine dipeptide was unable to reproduce details of \phi,\psi distributions in the helical region, requiring empirical corrections in the CMAP to reproduce these distributions in folded proteins (31,32). This suggests that these empirical corrections capture some many-body effects not included in the rest of the force field. To test whether this, rather than some other feature of the CHARMM36 force field is responsible for the improved cooperativity, a variant of CHARMM36 was developed in which the CMAP backbone potential was optimized to reproduce a gas phase RIMP2/cc-pVTZ//MP2/6-31G* quantum mechanical surface for the
a model for the helix-coil transition. The Lifson-Roig model
parameters

Helix-coil parameters

This observation may be quantified by fitting the data to a model for the helix-coil transition. The Lifson-Roig model (45) characterizes helix formation in terms of parameters \( v \) and \( w \): \( v \) is effectively an equilibrium constant for conversion of a residue in a coil conformation to a residue in a helical conformation, without formation of helical hydrogen bonds (i.e., it primarily describes nucleation), whereas \( w \) is the equilibrium constant for extending the length of an existing helix by one residue (i.e., it includes the effect of forming one additional \( i,i+4 \) hydrogen bond). We fitted this model to the simulation data for Ac-(AAQAA)\(_3\)-NH\(_2\) using a previously described Bayesian method (8), with the resulting values shown in Fig. 1, B and C, together with some experimental estimates. As can be seen, both the CHARMM36 and CHARMM36-MP2 data match quite well the experimental estimates for the helix elongation parameter \( w \) at 300 K, but the temperature dependence of \( w \) is too weak for CHARMM36-MP2 and for AMBER ff03*. We characterize the temperature dependence of \( w \) by fitting a thermodynamic model; we assume a negligible \( \Delta C_p \), because the experimental data cover only a small range of temperature. The best-fit curves are shown in Fig. 1, with the parameters in Table 1. Given that \( w \) reflects the free energy change upon forming a helical hydrogen bond, the parameters in Table 1 reveal that (as was found for other force fields (8)), the gain in enthalpy and loss of entropy upon hydrogen bond formation for CHARMM36-MP2 or AMBER ff03* are too small relative to the experimental estimates; the fitted parameters for CHARMM36 reveal enthalpy and entropy differences in much better agreement with experiment. A possible explanation for the discrepancy is a deficiency in the description for hydrogen bonding in the force field, as we had previously proposed, i.e., the entropic cost and enthalpic gain for forming a hydrogen bond are each too small. For example, plane-wave density functional theory calculations on model helices have suggested that one contribution to stronger hydrogen bonding may be electron density redistribution within the helix.

### Table 1

<table>
<thead>
<tr>
<th>Lifson-Roig parameter</th>
<th>Force field</th>
<th>( \Delta H ) [kcal mol(^{-1})]</th>
<th>( \Delta S ) [cal mol(^{-1}) K(^{-1})]</th>
</tr>
</thead>
<tbody>
<tr>
<td>( w )</td>
<td>Experiment (Moreau et al.) (50)</td>
<td>-0.87</td>
<td>-2.32</td>
</tr>
<tr>
<td></td>
<td>Experiment (Rohl et al.) (51)</td>
<td>-1.28</td>
<td>-3.78</td>
</tr>
<tr>
<td>( w )</td>
<td>CHARMM36</td>
<td>-1.17</td>
<td>-3.41</td>
</tr>
<tr>
<td>( w )</td>
<td>AMBER ff03*</td>
<td>-0.61</td>
<td>-1.81</td>
</tr>
<tr>
<td>( w )</td>
<td>CHARMM36-MP2</td>
<td>-0.36</td>
<td>-0.65</td>
</tr>
<tr>
<td>( v )</td>
<td>Experiment (Rohl et al.) (51)</td>
<td>-</td>
<td>-6.6</td>
</tr>
<tr>
<td>( v )</td>
<td>CHARMM36</td>
<td>-0.21</td>
<td>-4.5</td>
</tr>
<tr>
<td>( v )</td>
<td>AMBER ff03*</td>
<td>-0.72</td>
<td>-0.94</td>
</tr>
<tr>
<td>( v )</td>
<td>CHARMM36-MP2</td>
<td>1.02</td>
<td>-0.57</td>
</tr>
</tbody>
</table>

In each case, a simple thermodynamic function \(-k_B T \ln(x) = \Delta H_T - T \Delta S_T\) was fitted to the data, where \( x = w/v \). In experiments, \( v \) is usually assumed temperature independent.
Reproduction of this effect represents an interesting challenge for future force field development, and might be realized with polarizable force fields. However, in the context of an additive force field our results show that these many-body effects may be implicitly included in the backbone CMAP energy (indeed a many-body term itself).

Naturally, there is a limit to the extent to which many-body effects can be included: creating a more focused \( \alpha \)-helical minimum can mimic the larger entropy and enthalpy changes associated with helix formation caused by having more orientationally specific hydrogen bonding (20), and strengthening of helical hydrogen bonding by electron density redistribution in the helical conformation (19). CMAP of course does not consider whether or not a residue is in a helix or forming a hydrogen bond, but it can be an effective correction because when the peptide is in the \( \alpha \)R region, the appropriate CMAP surface will direct the peptide toward helical hydrogen bonds, directly compensating for deficiencies in the nonbonded portions of the model that lead to the relative lack of cooperativity. Notably, this may be achieved without significantly affecting nonhelical states, another important aspect of CMAP because the approach allows for more accurate correction of the entire \( \phi,\psi \) surface in contrast to the traditional use of cosine-based torsional parameters alone. However, CMAP cannot capture additional differences between longer and shorter helices caused by long-range effects on the charge distribution. Ultimately, it may be that this type of cooperativity can be captured by polarizable force fields, although that still remains to be demonstrated.

**Comparison with chemical shifts**

Because comparison of a fraction helix inferred from experiment (via fitting a model) with one calculated from simulation (subject to variations of helix definition) is somewhat unsatisfactory, agreement of our results with experiment was confirmed by direct comparison with experimental observables. To do this, the state of the art chemical shift prediction package SPARTA+ (46) was used to calculate ensemble-average chemical shifts for each temperature, which we compare with experimental carbonyl chemical shifts for Ac-(AAQAA)\(_3\)-NH\(_2\) in Fig. 2. As can be seen, the agreement with experiment is excellent for CHARMM36, with only small deviations near the N-terminus; on the other hand, CHARMM36-MP2 or AMBER ff03* show a systematic deviation from the temperature dependence revealed in experiment.

**Origin of cooperativity**

As discussed previously, the origin of the improved cooperativity in the CHARMM force field may be attributed to the CMAP term in the energy function. To get more insight into the origin of this effect, we have analyzed the backbone \((\phi,\psi)\) distribution in the \( \alpha \)-helical region of the Ramachandran map obtained with the Amber ff03*, CHARMM36-MP2, and CHARMM36 force fields, and that observed in the Protein Data Bank (PDB). As a representative of nonhelical conformations, we use the \((\phi,\psi)\) distribution from residues with a coil secondary structure assignment in the PDB (using the TOP500 set of high quality crystal structures compiled by Lovell et al.) (47), and the \((\phi,\psi)\) distribution for conformations of Ac-(AAQAA)\(_3\)-NH\(_2\) in which no helix is formed in the simulations (Fig. 3, A, C, E, and G). We also have compiled \((\phi,\psi)\) distributions for residues in helices of 11–15 residues in length, in both the TOP500 PDB data set and the simulations, shown in Fig. 3, B, D, F, and H. If we look first at the residues in helices, simulations with Amber ff03* or CHARMM36-MP2 result in a broader free energy minimum and a slight shift (particularly toward lower \( \phi \)), relative to the PDB. A similar shift in average torsion angles for simulations of folded proteins was the motivation for empirical corrections in the original CMAP. As expected, the empirically corrected CMAP, results in the correct average helical torsion angles, Fig. 3 F. Notably, however, because the PDB distribution was used to guide the empirical correction, the CHARMM36 force field also results in a narrower distribution of torsion angles in the helical state—most likely responsible for the increased entropic cost of helical hydrogen bond formation. This empirical correction does of course slightly distort the \( \alpha \) free energy minimum in the nonhelical states, relative to the PDB. However, this is not expected to be a significant error due to the low average \( \alpha \) population in unfolded or unstructured peptides. Indeed, the CHARMM36 force field compares well with NMR data for disordered peptides (13).
The dependence of folding thermodynamics on parameters in the force field is complex, and adjusting torsion or CMAP potentials is of course only one factor affecting the helix-coil equilibrium. Another important influence is the interaction with the solvent. For example, it has been shown that using a more accurate water model than TIP3P may improve helix (10) or hairpin (12) folding thermodynamics in simulations. It should be pointed out, however, that using a solvent model that more accurately reproduces the properties of pure water does not guarantee better results for peptide and protein folding due to the need to properly balance the solute-solute, solute-solvent, and solvent-solvent interactions in the force field. Accordingly, use of an alternative water model in principle requires a reparametrization of the protein force field (10,22,48), and the CHARMM all-atom additive force fields have been optimized specifically in conjunction with the TIP3P water model.

**Folding of $\beta$-hairpin**

We have also examined the effect of the CMAP on the folding of the GB1 hairpin, by sampling the temperature-dependent equilibrium via REMD (as described in (13)). The resulting melting curve, shown in Fig. 4, also reveals a cooperative folding transition. Although the peptide is slightly too stable, with the folding midpoint upshifted from experiment, the width of the folding transition is in better agreement with experiment than in our previous studies (9), which yielded results for AMBER ff99SB* and other additive force fields similar to the data for AMBER ff03* shown in Fig. 4. We fit the global folding transition to a two-state thermodynamic model, because, unlike the helix, the folding of the peptide appears to be two-state using multiple probes (40,49). The fitted folding parameters

---

**Figure 3** Distribution of backbone angles from simulations (A–F) of (AAQAA)$_3$ and from the PDB (G and H), shown as $-\ln[P(\phi,\psi)]$ for probability density $P(\phi,\psi)$. Left-hand column is for residues not in helices, whereas the right-hand column is for helices of length 11–15 residues. (A and B) are from simulations with Amber ff03*, (C and D) with CHARMM36-MP2, (E and F) with CHARMM36, and (G and H) from the PDB. Broken magenta lines in all cases indicate the location of the helical minimum in the PDB (i.e., defined by the minimum in H). All apparent free energies are measured relative to the lowest value in the region of interest shown.

**Figure 4** Cooperativity of $\beta$-hairpin formation. The melting curve for the GB1 hairpin determined by REMD simulations with the CHARMM36 force field (solid circles: data; curve through data: fit to two-state thermodynamic model), AMBER ff03* (9) (up triangles: REMD from unfolded; down triangles: REMD from folded) compared with experiment (40) (thick curve). The empty circles and broken line show the CHARMM36 simulation data shifted down by 70 K. The folded structure of the hairpin is shown in the inset.
(Table 2) are in much better agreement with experimental estimates than those obtained with AMBER ff03* with a midpoint unfolding enthalpy $\Delta H_m = 10.6 \text{ kcal} \cdot \text{mol}^{-1}$ (experiment: $\sim 12 \text{ kcal} \cdot \text{mol}^{-1}$) and entropy $\Delta S_m = 28.2 \text{ cal} \cdot \text{mol}^{-1} \cdot \text{K}^{-1}$ (experiment: $\sim 40 \text{ cal} \cdot \text{mol}^{-1} \cdot \text{K}^{-1}$). The shift of the midpoint to higher temperatures may be attributed to the fact that the $\Delta S_m$, although improved, is still somewhat too small. The reason for the additional cooperativity is likely related to that for helix formation: the CMAP effect on the overall thermodynamics of the hairpin. The magnitude of the enthalpy and entropy of folding are too small. Ultimately, it is anticipated that a more accurate force field representation of the potential energy surface will allow thermodynamics to be more accurately captured: an important step in this direction are polarizable force fields currently under development. However, our work has shown that, at least to some extent, the missing cooperativity can be encoded in an additive force field, by means of the multibody CMAP potential. Whether such corrections will result in improved thermodynamics for other peptides remains to be tested, but we are currently engaged in testing the updated force field on larger all-$\alpha$ and all-$\beta$ proteins.

### CONCLUSIONS

A figure demonstrating convergence of the fraction helix in Ac-(AAQAA)$_3$-NH$_2$ in REMD simulations with CHARMM36, and a figure showing Ramachandran distributions of the GB1 hairpin in two force fields and the PDB, are available at http://www.biophysj.org/biophysj/supplemental/S0006-3495(12)00855-7.

R.B. is supported by a Royal Society University Research Fellowship, A.D.M. by National Institutes of Health (NIH) grants GM072558 and GM051501, and M.F. by NIH grants GM084953 and GM092949. This work used the Extreme Science and Engineering Discovery Environment, which is supported by National Science Foundation grant No. MCB-120014.

### SUPPORTING MATERIAL

A figure demonstrating convergence of the fraction helix in Ac-(AAQAA)$_3$-NH$_2$ in REMD simulations with CHARMM36, and a figure showing Ramachandran distributions of the GB1 hairpin in two force fields and the PDB, are available at http://www.biophysj.org/biophysj/supplemental/S0006-3495(12)00855-7.

### REFERENCES


### TABLE 2 Thermodynamic parameters for GB1 hairpin unfolding, obtained from a thermodynamic fit to the melting curve (Fig. 4), or from experimental estimates from spectroscopy (40) and calorimetry (49)

<table>
<thead>
<tr>
<th>Thermodynamic parameter</th>
<th>Experiment</th>
<th>CHARMM36</th>
<th>AMBER ff03*</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Delta H_m$ [kcal mol$^{-1}$]</td>
<td>11.6$^1$, 12.2$^2$</td>
<td>10.6 (0.1)</td>
<td>5.75 (0.18)</td>
</tr>
<tr>
<td>$\Delta S_m$ [cal mol$^{-1} \cdot K^{-1}$]</td>
<td>39.0$^1$, 41.4$^4$</td>
<td>28.2 (0.2)</td>
<td>18.2 (0.6)</td>
</tr>
<tr>
<td>$\Delta C_p$ [cal mol$^{-1} \cdot K^{-1}$]</td>
<td>0</td>
<td>85.3 (2.6)</td>
<td>24.0 (6.4)</td>
</tr>
<tr>
<td>$T_m$ [K]</td>
<td>297$^1$, 295.3$^3$</td>
<td>376 (5)</td>
<td>316 (14)</td>
</tr>
</tbody>
</table>

Thermodynamic model $\Delta G_{f \rightarrow u}(T) = \Delta H_m + T \Delta C_p(T - T_m) - T \Delta S_m - T C_p(T/T_m)$ was fitted to the data, with $T_m = \Delta H_m / \Delta S_m$. 

Biophysical Journal 103(5) 1045–1051
Cooperativity of $\alpha$-Helix and $\beta$-Hairpin Formation


