How much detail is sufficient?
A comparison of non-polarizable and polarizable force fields for protein folding

Anthony Hazel

July 12, 2018

NAMD Workshop

DICP, Dalian, CN

Folding free energy landscapes of β-sheets with non-polarizable and polarizable CHARMM force fields

Anthony J. Hazel,[1] Evan T. Walters,[2] Christopher N. Rowley,[2] and James C. Gumbart[1]
[1]*School of Physics, Georgia Institute of Technology, Atlanta, Georgia 30332, USA*
[2]*Department of Chemistry, Memorial University of Newfoundland, St. John's, Newfoundland and Labrador A1B3X7, Canada*

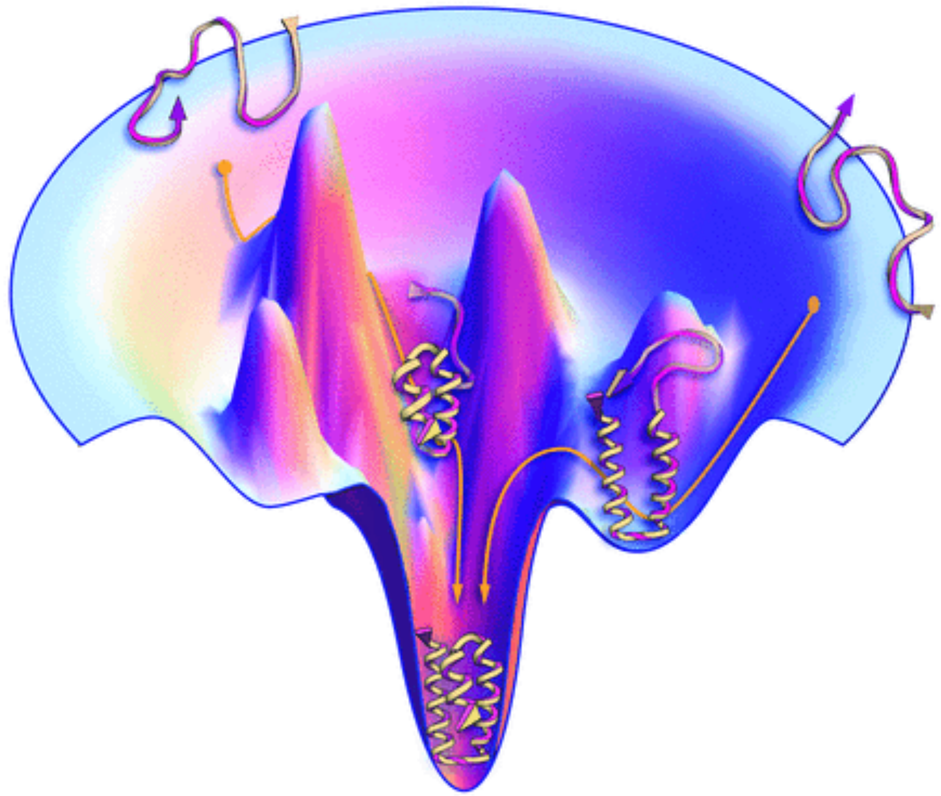
(Received 14 February 2018; accepted 22 May 2018; published online 8 June 2018)

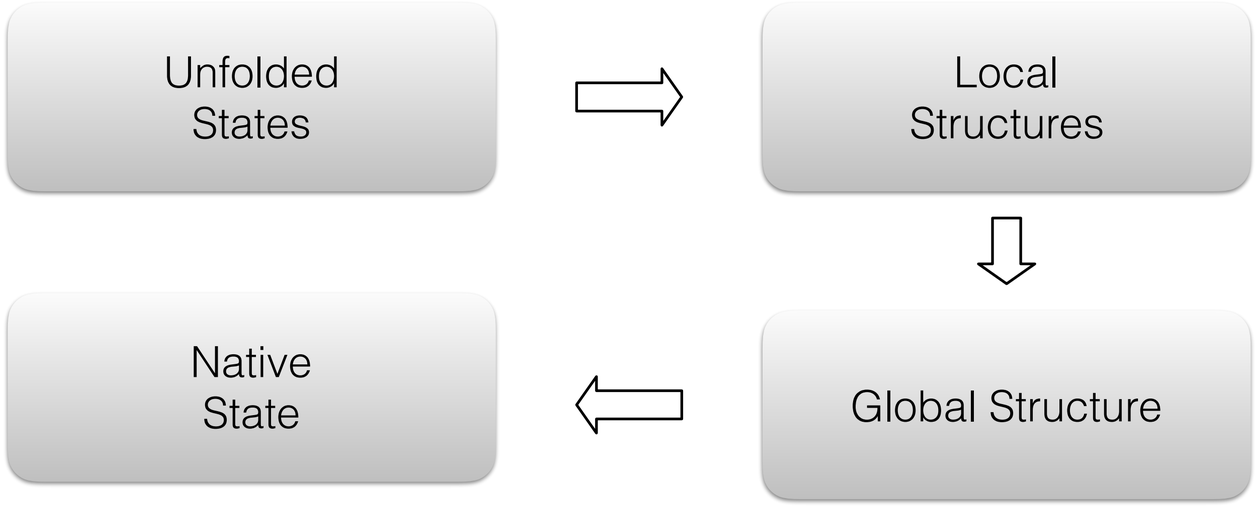
Molecular dynamics (MD) simulations of peptides and proteins offer atomic-level detail into many biological processes, although the degree of insight depends on the accuracy of the force fields used to represent them. Protein folding is a key example in which the accurate reproduction of folded-state conformations of proteins and kinetics of the folding processes in simulation is a longstanding goal. Although there have been a number of recent successes, challenges remain in capturing the full complexity of folding for even secondary-structure elements. In the present work, we have used all-atom MD simulations to study the folding properties of one such element, the C-terminal β-hairpin of the B1 domain of streptococcal protein G (GB1). Using replica-exchange umbrella sampling simulations, we examined the folding free energy of two fixed-charge CHARMM force fields, CHARMM36 and CHARMM22*, as well as a polarizable force field, the CHARMM Drude-2013 model, which has previously been shown to improve the folding properties of α-helical peptides. The CHARMM22* and Drude-2013 models are in rough agreement with experimental studies of GB1 folding, while CHARMM36 overstabilizes the β-hairpin. Additional free-energy calculations show that small adjustments to the atomic polarizabilities in the Drude-2013 model can improve both the backbone solubility and folding properties of GB1 without significantly affecting the model's ability to properly fold α-helices. We also identify a non-native salt bridge in the β-turn region that overstabilizes the β-hairpin in the C36 model. Finally, we demonstrate that tryptophan fluorescence is insufficient for capturing the full β-hairpin folding pathway. *Published by AIP Publishing.* https://doi.org/10.1063/1.5025951

Free Energy Landscapes

- Protein folding landscapes are narrower at the bottom; there are few low-energy, native-like conformations and many more open unfolded structures.
- Random steps that are mostly incrementally downhill in energy

Free energy: $G = H - TS$

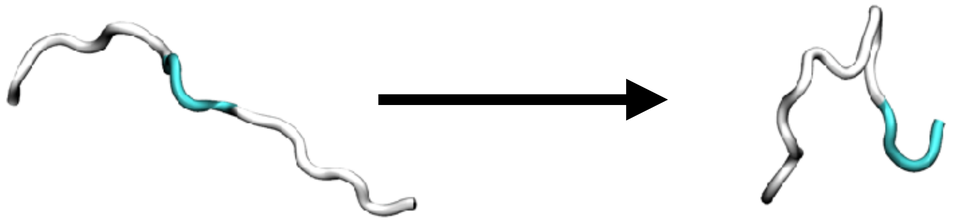




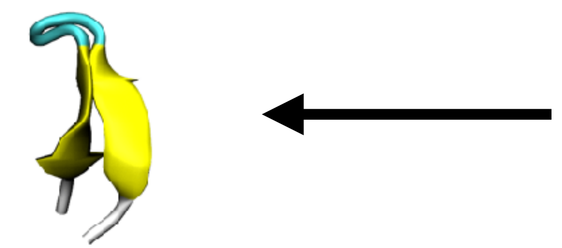
Wolynes, Onuchic, Thirumalai. Navigating the folding routes. *Science* 267: 1619-1620 (1995).

Umbrella Sampling

$k_1(\xi-\xi_1)^2$ $k_2(\xi-\xi_2)^2$

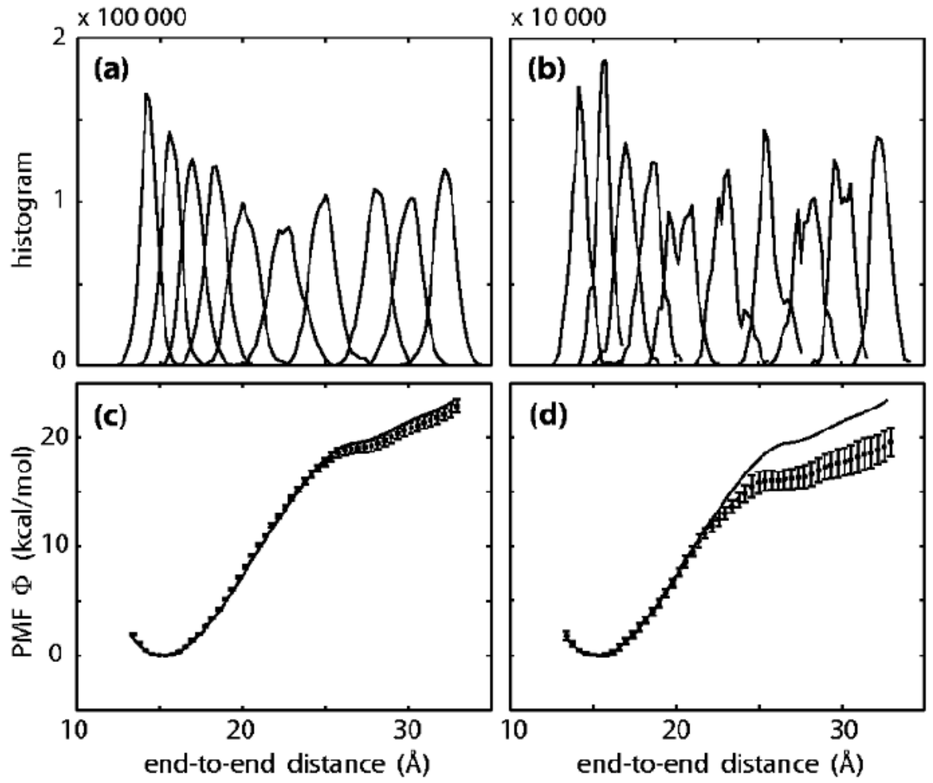


N independent simulations



$k_N(\xi-\xi_N)^2$ $k_{N-1}(\xi-\xi_{N-1})^2$

Potential of Mean Force (PMF)



Kumar et al. *J. Comput. Chem.* 13:1011 (1992).
Park et al. *J. Chem. Phys.* 119:3559 (2003).

The Weighted Histogram Analysis Method for Free-Energy Calculations on Biomolecules. I. The Method

Shankar Kumar,[1] Djamal Bouzida,[2] Robert H. Swendsen,[2] Peter A. Kollman,[3] and John M. Rosenberg[1]*

[1]*Department of Biological Sciences, University of Pittsburgh, Pittsburgh, Pennsylvania 15260*
[2]*Department of Physics, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213*
[3]*Department of Pharmaceutical Chemistry, University of California-San Francisco, San Francisco, California 94143*

Unbiased Hamiltonian

Bias Potential

$$\hat{H}_{\{\lambda\}}(x) = \hat{H}_0(x) + \sum_{i=1}^{L} \lambda_i \hat{V}_i(x) = \sum_{i=0}^{L} \lambda_i \hat{V}_i(x) \quad (1)$$

$$W_{\{0\},\beta}(\xi) = -\sum_{j=1}^{L} \lambda_j \hat{V}_j(\xi) + W_{\{\lambda\},\beta}(\xi) + C(\{\lambda\}, \beta) \quad (9)$$

Unbiased PMF

Subtract bias

Biased PMF

Arbitrary constant for each simulation

WHAM Equations

$$P_{\{\lambda\},\beta}(\{V\}, \xi) = \frac{\sum_{k=1}^{R} N_k(\{V\}, \xi) \exp\left(-\beta \sum_{j=0}^{L} \lambda_j V_j\right)}{\sum_{m=1}^{R} n_m \exp\left(f_m - \beta_m \sum_{j=0}^{L} \lambda_{j,m} V_j\right)} \quad (19)$$

and

$$\exp(-f_j) = \sum_{\{V\},\xi} P_{\{\lambda\}_j,\beta_j}(\{V\}, \xi) \quad (20)$$

... convergence of f_m's

Initial guess of constants ($f_m = 0$)

Journal of Computational Chemistry, Vol. 13, No. 8, 1011–1021 (1992)

Replica Exchange Molecular Dynamics (REMD)

Allow multiple, parallel simulations (replicas) to periodically exchange parameters (temperature, biasing potentials, etc.)

Hamiltonian exchange for US simulations (REUS)

Exchange probability given by Metropolis criterion:

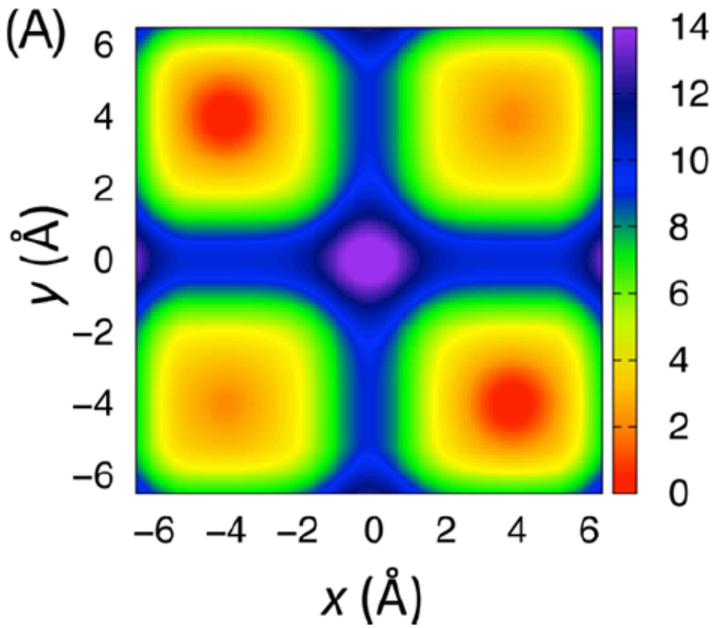
$$P(i \leftrightarrow j) = \min\left\{1, \exp(-\beta \Delta E)\right\} \qquad \beta = (k_B T)^{-1}$$

$$\Delta E = [E_i(q_i) + E_j(q_j)] - [E_j(q_i) + E_i(q_j)]$$

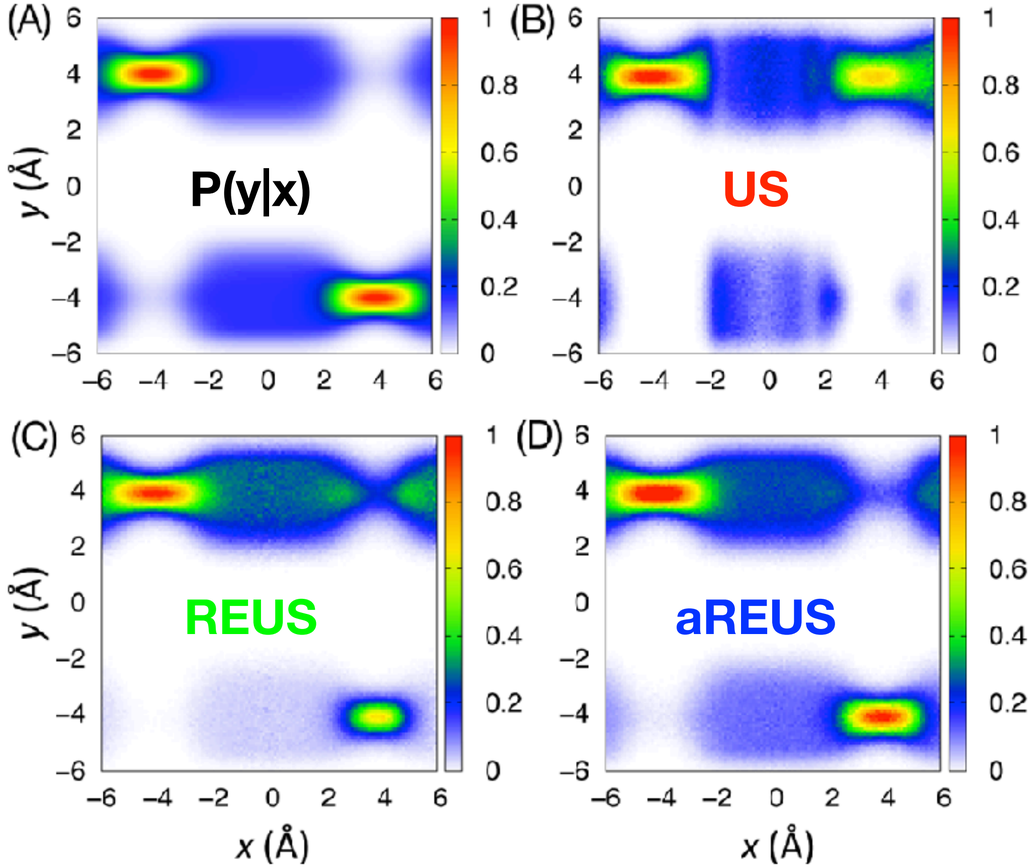
Sugita et al. *J. Chem. Phys.* 113:6042 (2000).

Replica Exchange Umbrella Sampling (REUS)

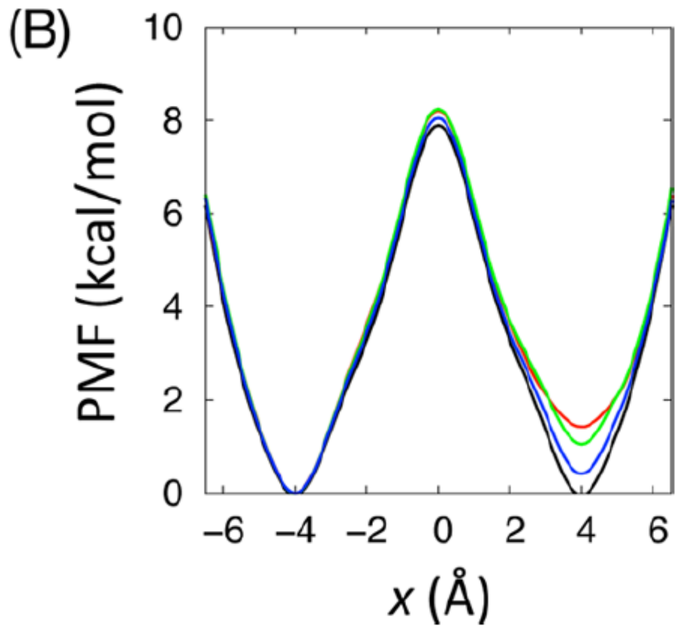
Toy Model: U(x,y)



Sampling

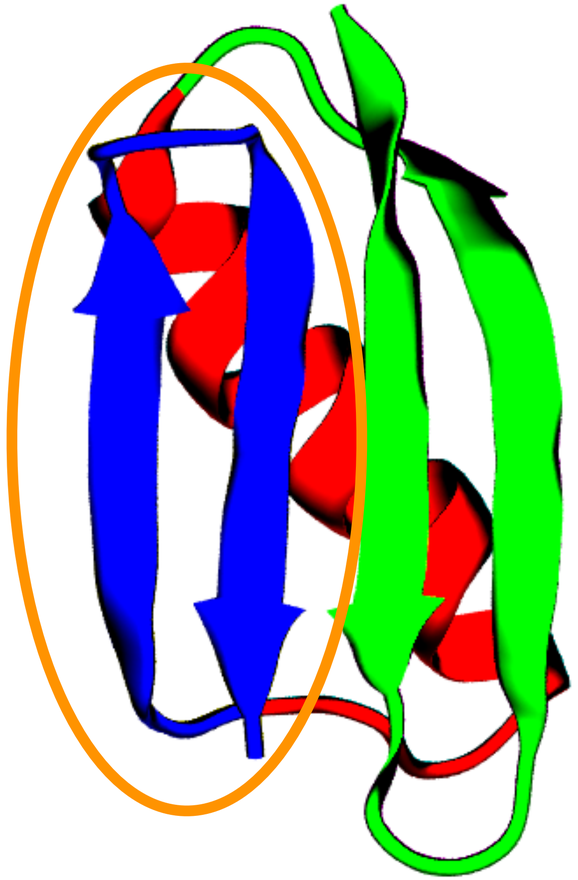


Free Energy

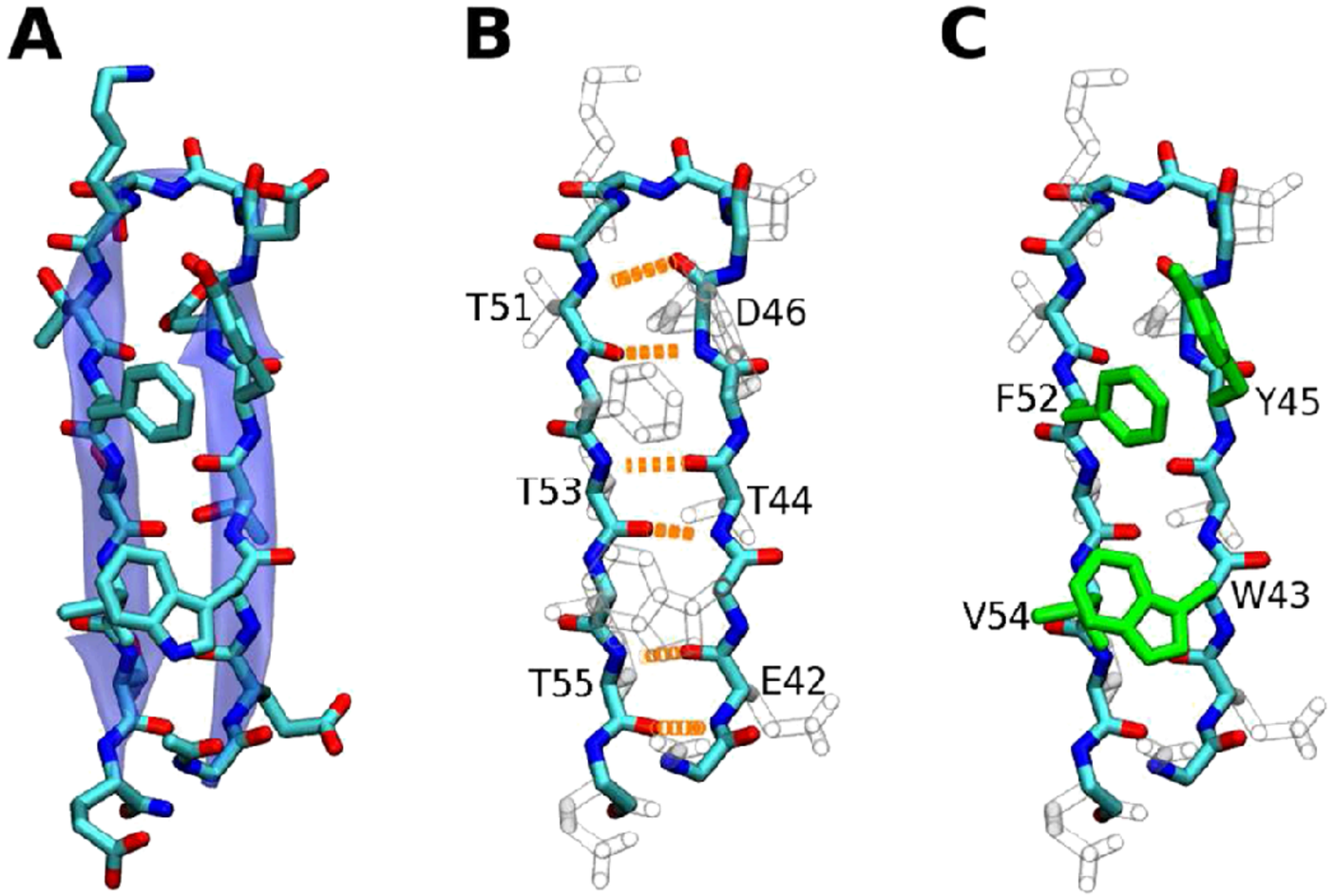


Park & Im. *J. Chem. Theory Comput.* 10:2719-2728 (2014).

B1 domain of streptococcal protein G (GB1)

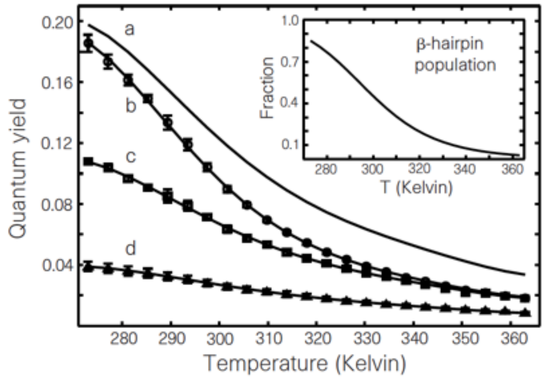


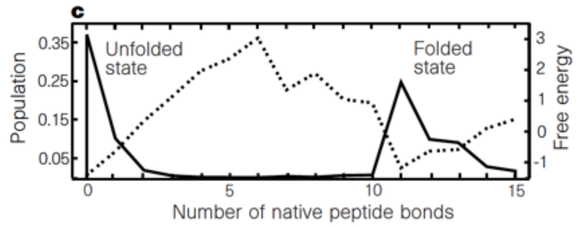
PDB: 1GB1
56 residues



C-terminal hairpin of GB1 (G41-E56)
A. All residues **B.** Hydrogen bonds **C.** Hydrophobic core

GB1 as a model for β-sheet folding





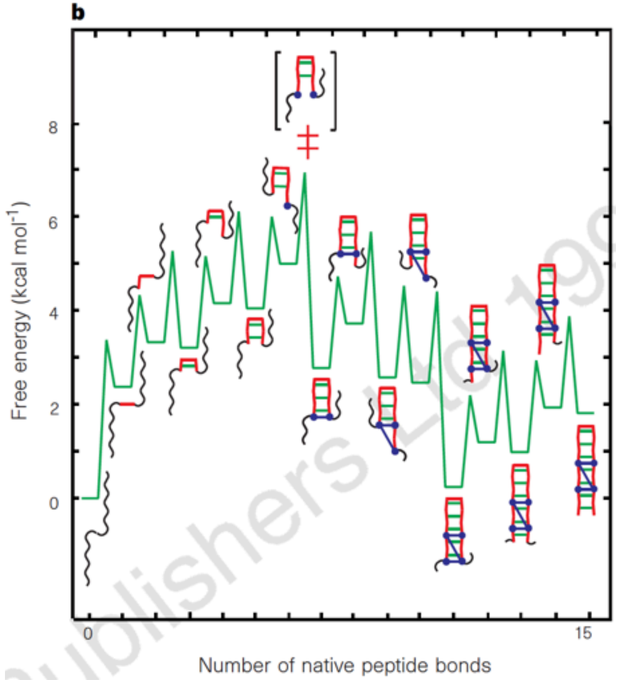


Table 1. Hairpin Sequences and Fold Populations

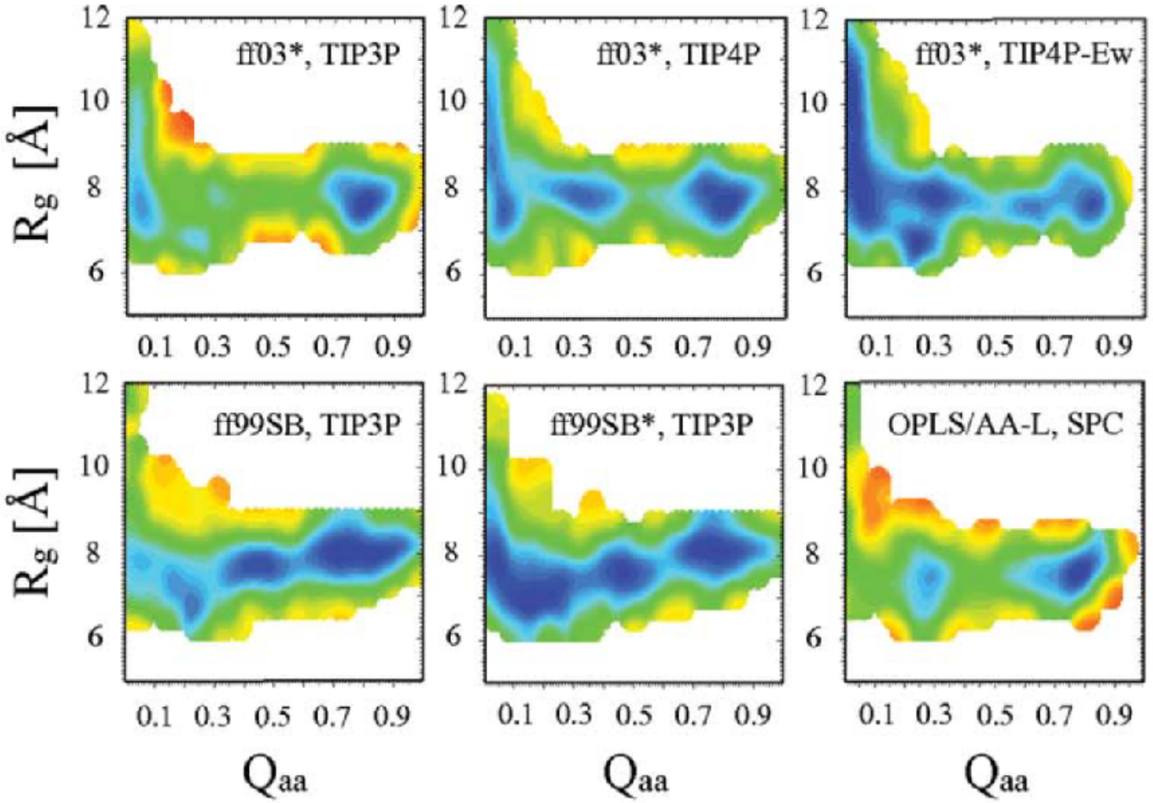
		% folded (298 K)	T_m (°C)
GB1p	GEWTYDDATKTFTVTE	ca. 30%	
GB1m1	GEWTYDDATKTATVTE	6 ± 6%	
GB1m2	GEWTYNPATGKFTVTE	74 ± 5%	47 ± 2
GB1m3	KKWTYNPATGKFTVQE	86 ± 3%	60 ± 2
trpzip4	GEWTWDDATKTWTWTE		70
HP5W4	KKWTWNPATGKWTWQE	>96%	85
HP5W	KKYTWNPATGKWTVQE	92 ± 2%	66
HP5F	KKYTWNPATGKFTVQE	82 ± 4%	53 ± 5
HP5A	KKYTWNPATGKATVQE	21 ± 10%	

30-50% folded @ 298/300K

Folds in ~6μs

Munoz et al. *Nature.* 149:072317 (1997).
Fesinmeyer et al. *J. Am. Chem. Soc.* 126:7238-7243 (2004).

Commonly used to calibrate force fields and enhanced sampling techniques



Best and Mittal. *Proteins.* 79:1318-1328 (2011).

CHARMM Drude Polarizable Force Field

Non-polarizable FFs tend to be overpolarized in order to mimic a solvent environment

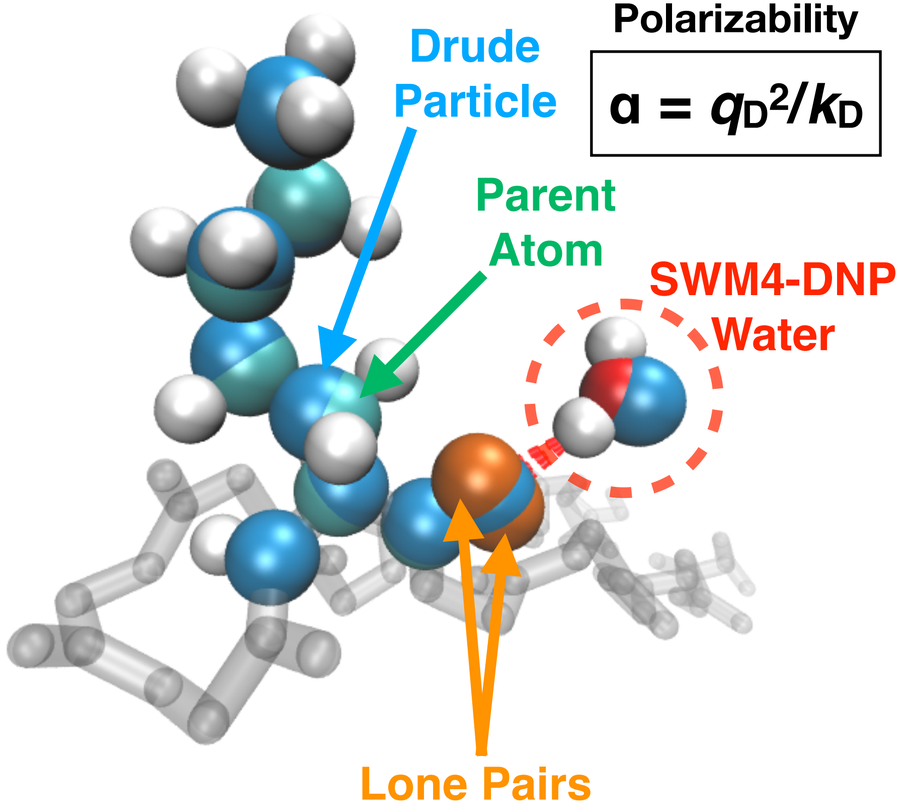
Drude oscillator polarizable FF splits each heavy atom into a (+) parent atom and a (—) Drude particle, connected by a stiff spring (k_D)

Drude model allows for polarization of molecule from environment, not just molecular geometry

1-fs timestep

```
drude on
drudeTemp 1
drudeDamping 20.0
drudeBondLen 0.25
drudeHardWall on
drudeNBTHOLEcut 5.0
LJcorrection yes
```

Drude Lysine



Lopes et al. *J. Chem. Theory Comput.* 9:5430–5449 (2013).
Huang and MacKerell, Jr. *Biophys. J.* 107:991–997 (2014).

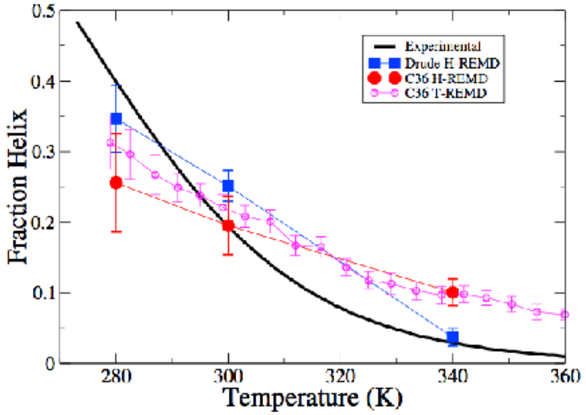


FIGURE 1 Fraction helix as a function of temperature determined from NMR experiments and replica exchange simulations with the Drude polarizable and the CHARMM36 additive force fields. The C36 T-REMD results were previously reported (11). To see this figure in color, go online.

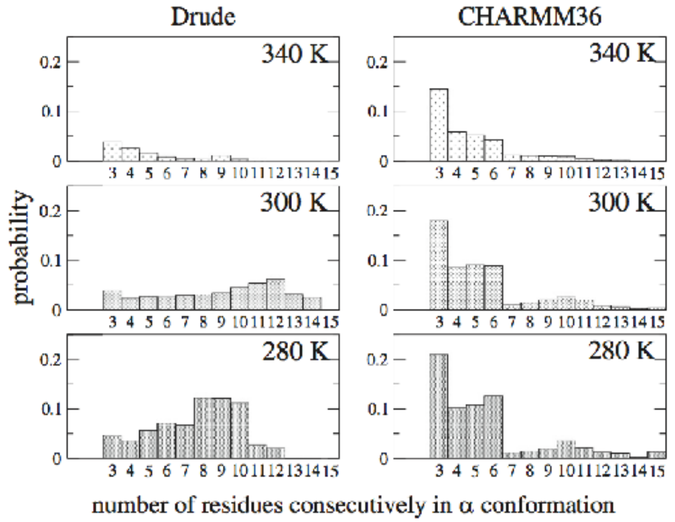
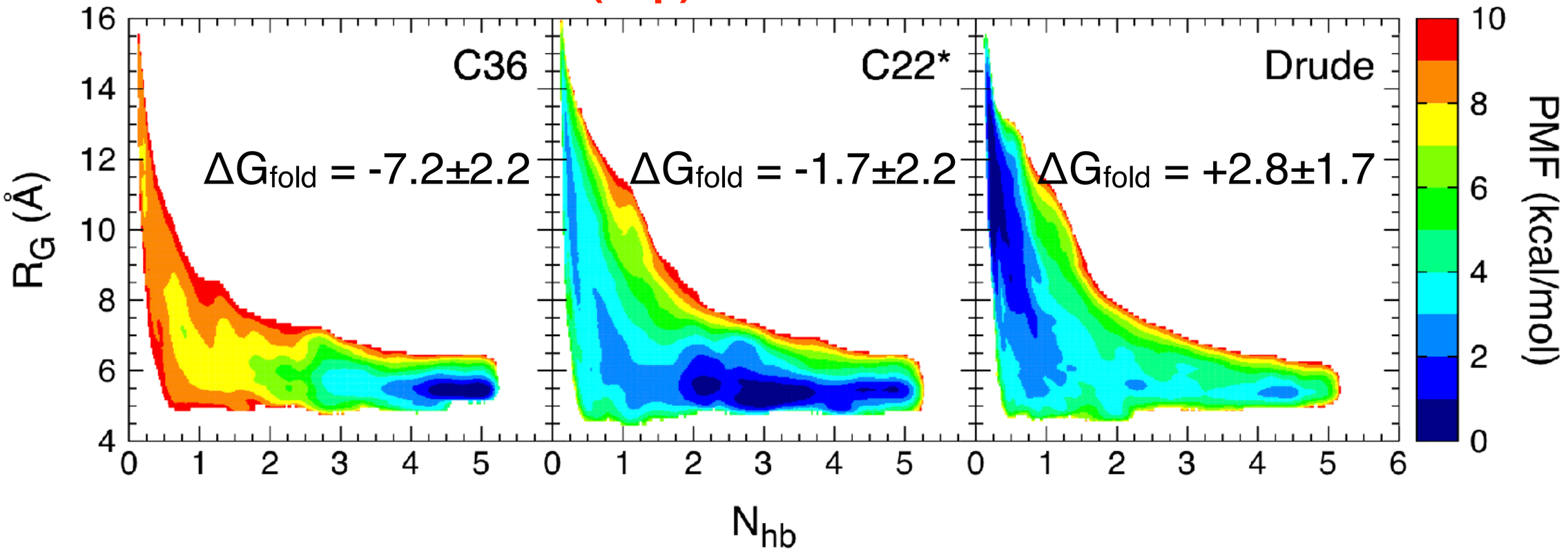


FIGURE 3 Probabilities of observing n consecutive residues in the α region (as a helix of length n) during the polarizable Drude and additive CHARMM36 simulations at 280, 300, and 340 K.

Folding PMFs of the GB1 β-hairpin

89-100 REUS windows 12-20ns/window

ΔG_{fold} (exp) = 0.0-0.5 kcal/mol



C36 greatly overestimates ΔG_{fold}

C22* slightly overestimates ΔG_{fold}

Drude slightly underestimates ΔG_{fold}

Side chain hydration free energies show room for improvement in the non-polarizable and polarizable models

Chris Rowley

molecule	residue	ΔG_{C36}	ΔG_{Drude}	ΔG_{expt}	molecule	residue	ΔG_{C36}	ΔG_{Drude}	ΔG_{expt}
n-butane	Ile	2.60	2.68	2.08	methylethylsulfide	Met	0.59	−1.12	−1.49
isobutane	Leu	2.56	2.50	2.28	3-methylindole	Trp	−5.40	−5.11	−5.91
methane	Ala	2.38	2.24	2.00	methylimidazole	His	−10.11	−11.67	−10.25
propane	Val	2.49	2.65	1.96	propionamide	Gln	−7.59	−8.06	−9.42
acetamide	Asn	−7.56	−9.69	−9.72	toluene	Phe	−0.30	−0.48	−0.76
p-cresol	Tyr	−4.92	−5.52	−6.13	acetate	Asp	−82.35	−84.92	−80.65
ethanol	Thr	−4.78	−3.74	−4.90	n-butylammonium	Lys	−66.00	−61.90	−69.24
methanethiol	Cys	−0.27	−1.04	−1.24	n-propylguanidinium	Arg	−60.21	−57.00	–
methanol	Ser	−4.94	−3.43	−5.08	propionate	Glu	−82.75	−80.76	−79.12
					n-methylacetamide	NMA	−7.65	−8.50	−10.10

40-stage Weeks–Chandler–Anderson (WCA)-decomposition free-energy perturbation (FEP) procedure

Y. Deng and B. Roux. *J. Phys. Chem. B.* 108:16567 (2004).

Y. Deng and B. Roux. *J. Phys. Chem. B.* 113:2234 (2009).

The Drude model underestimates hydroxyl, overestimates some charged, and improves sulfur-containing and amide hydration free energies. Improved backbone hydration needs to be compensated by improved N-H—O=C hydrogen bonding or peptide will unfold. Also, while backbone (NMA) hydration is improved in the Drude model over C36, further improvements could still be implemented.

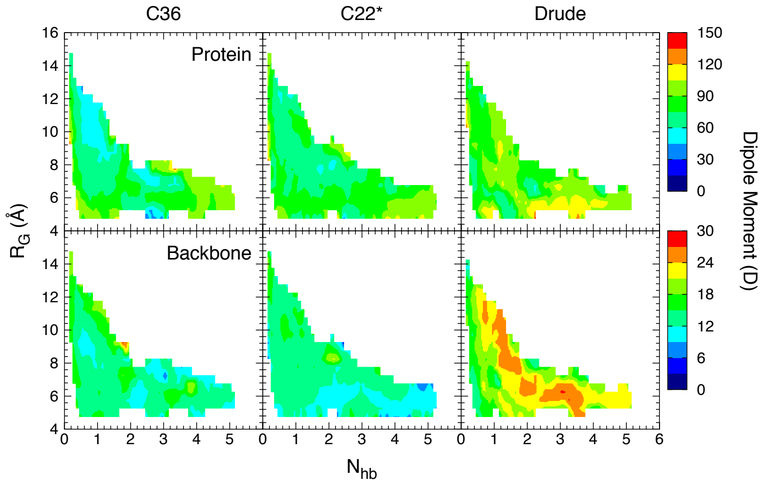
Examining backbone polarization in the Drude model

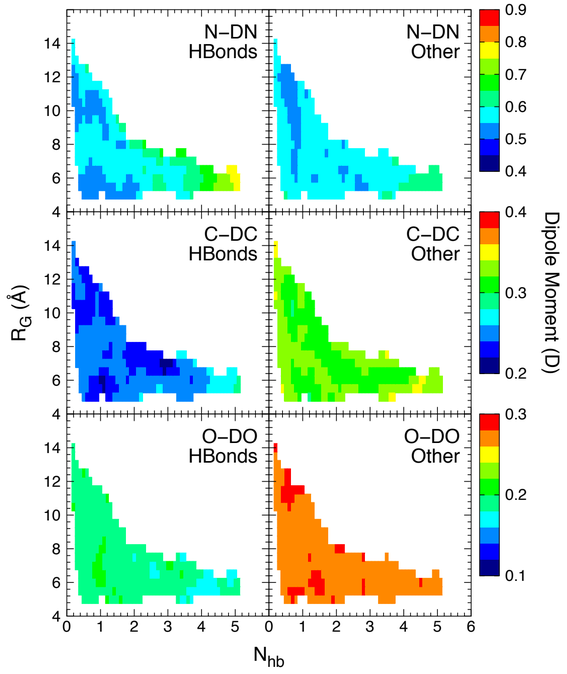
Dipole moments are enhanced in the Drude model

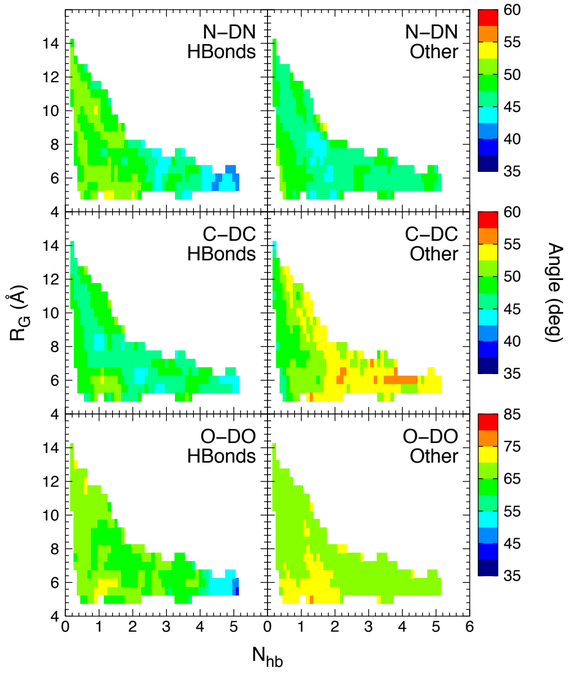
N-H bonds significantly polarize during intrapeptide hydrogen bonding, but not when hydrogen bonding with water

C=O bonds behave in the opposite manner, only polarizing significantly when hydrogen bonding with water

Unlike water-peptide hydrogen bonding, intrapeptide hydrogen bonding aligns parent-drude bond with chemical bond





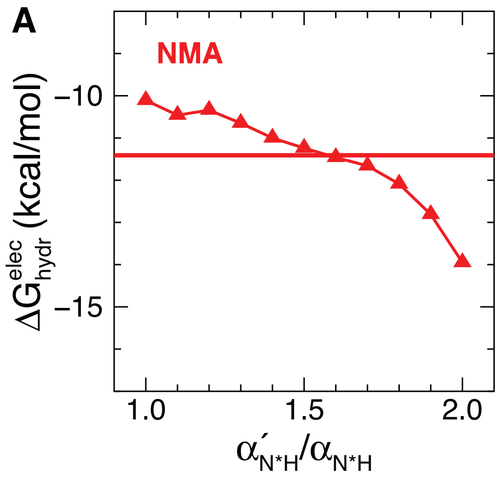


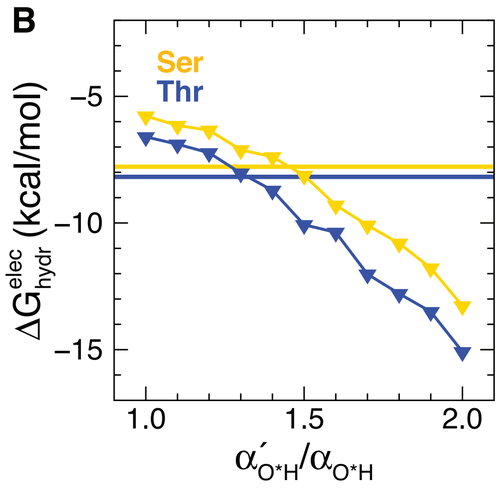
Altering backbone (and side chain) polarizabilities in the Drude model

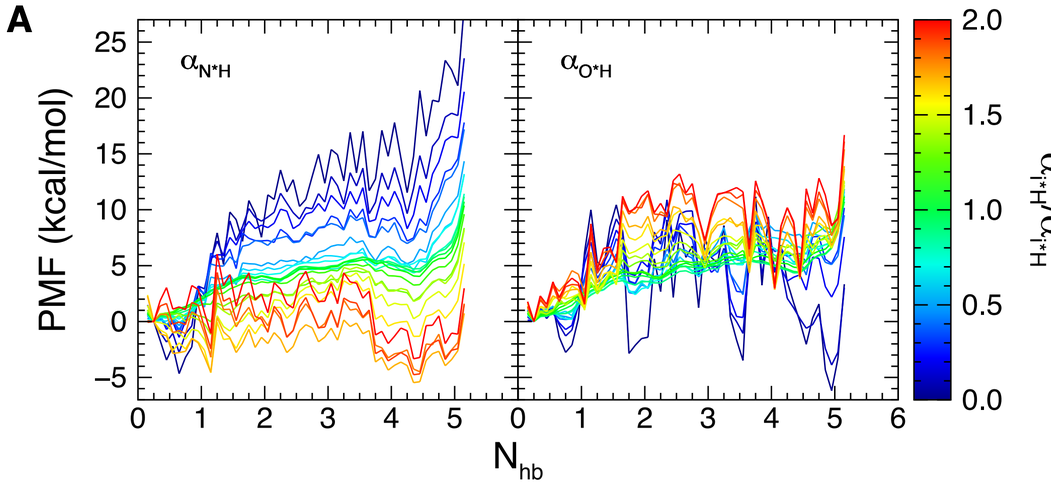
By rescaling the histograms in WHAM, we can recalculate the PMF with different parameters using the states already sampled by our REUS simulations:

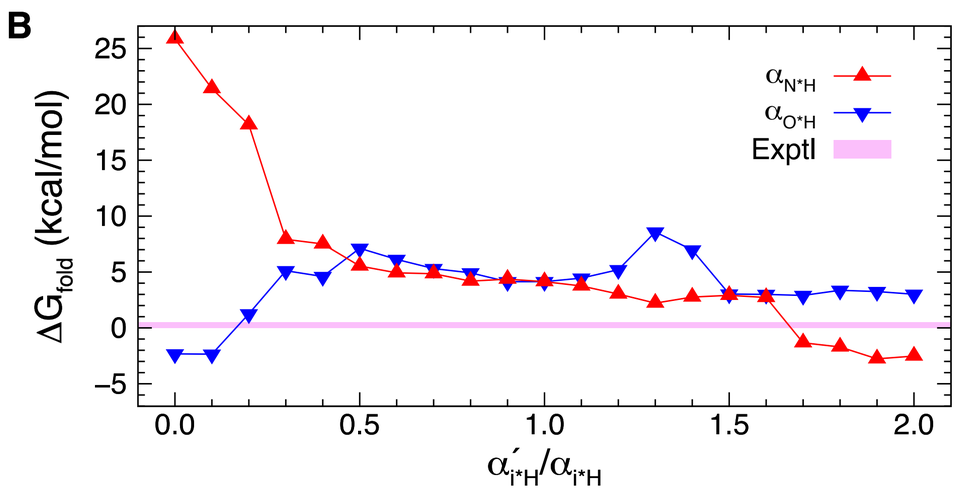
$$\mathrm{histogram}(r_{ij}) = \frac{\sum_t \exp\big(-\beta(U_{\mathrm{new}} - U_{\mathrm{old}})\big)\delta(r(t) - r_{ij})}{\sum_t \exp\big(-\beta(U_{\mathrm{new}} - U_{\mathrm{old}})\big)}$$

Increasing the backbone N polarizabilities by 60% showed the best improvement for both the backbone hydration and GB1 β-sheet folding

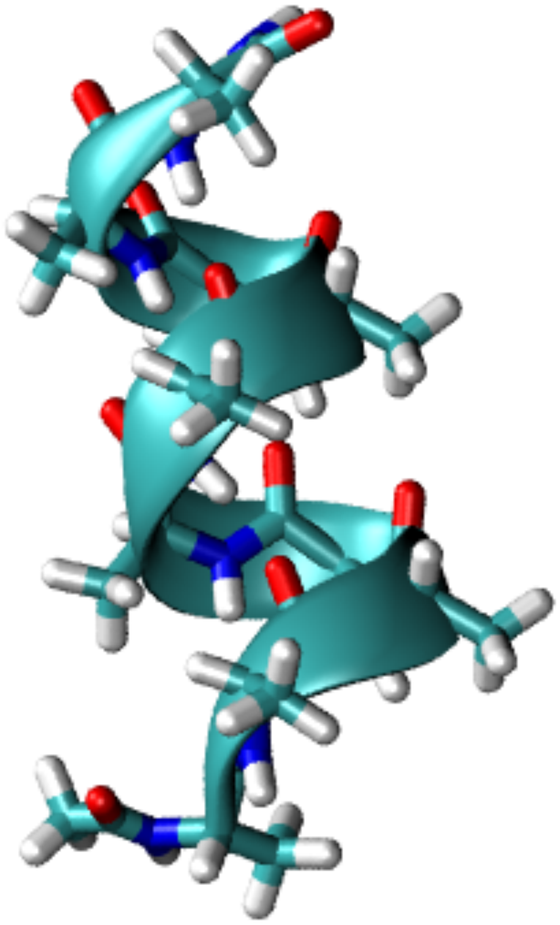


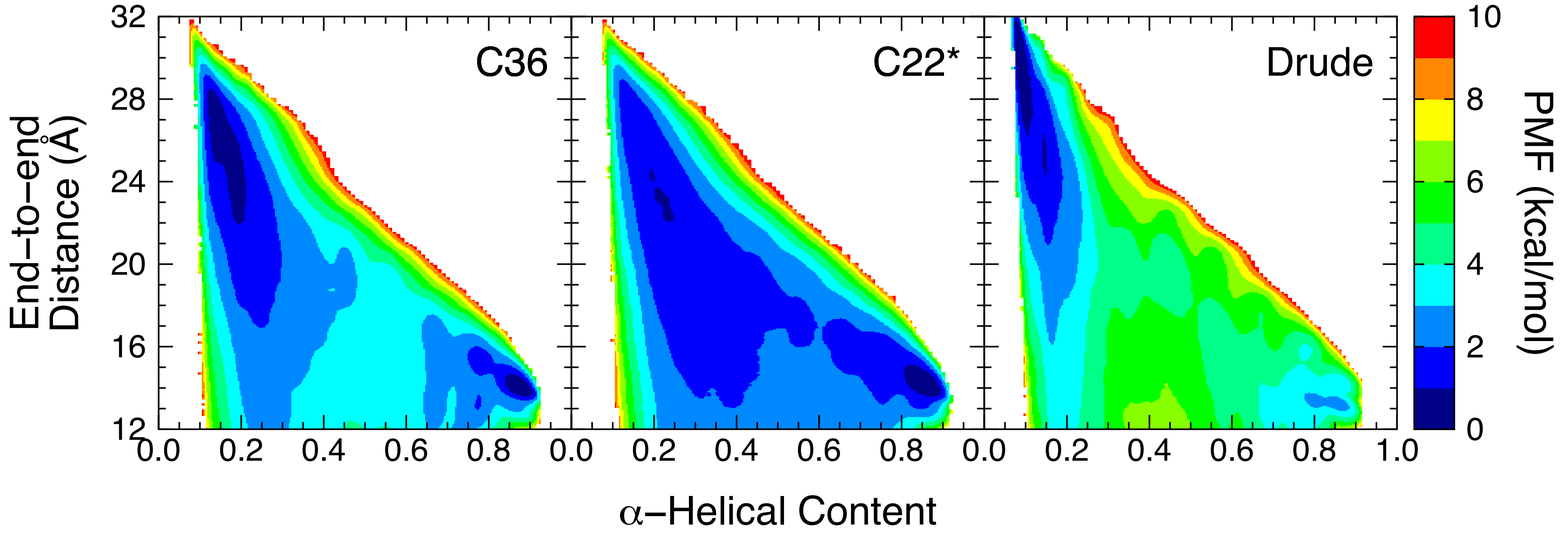


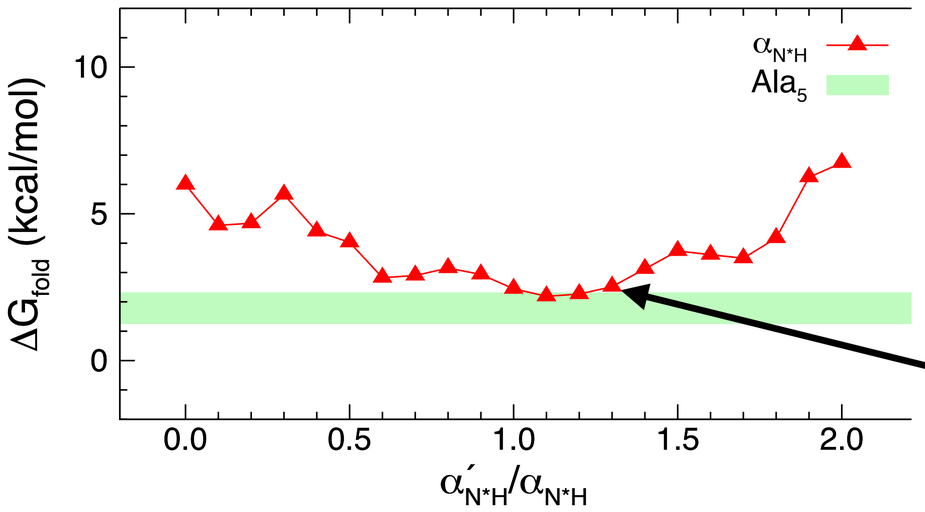




Checking the altered polarizabilities with an α-helical peptide, Ala_{10}







The Drude model is already well optimized for α-helices

Increasing in N polarizability by >30% would deteriorate quality of the model

30% increase is gives a good balance between α-helices and β-sheets

Conclusions

We used replica exchange umbrella sampling (REUS) to calculate the folding free energies of the model β-sheet peptide, GB1

The polarizable Drude force field is very well optimized for α-helical peptides

However, although it greatly outperforms the C36 force field, it performs slightly worse than the non-polarizable C22* force field in describing GB1 β-hairpin folding

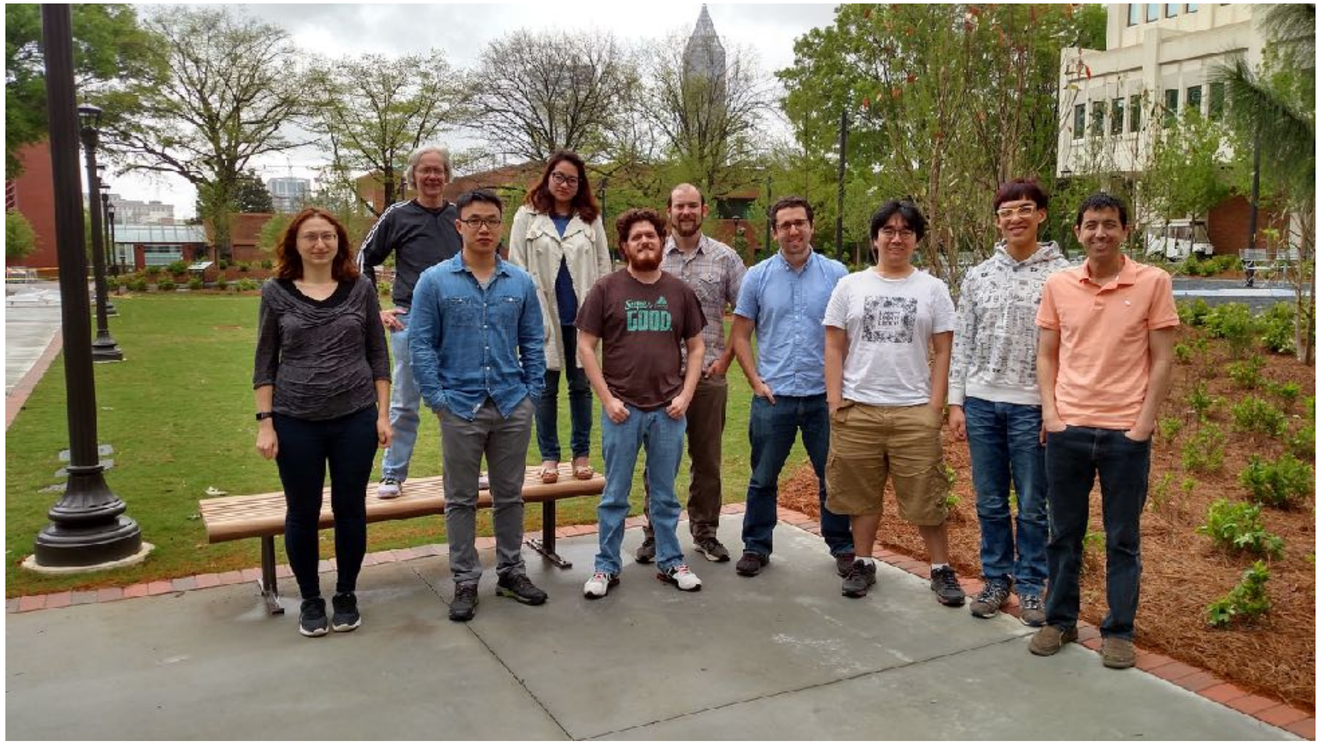
Using pertubations to the WHAM histograms, we showed that small enhancements to backbone polarizabilities improve both backbone hydration and β-hairpin folding while maintaining α-helical folding

While the Drude model is a relatively cheap method to introduce dynamic, inducible atomic polarization, the cheaper non-polarizable C22* force field is sufficient for modeling β-sheets

Acknowledgements

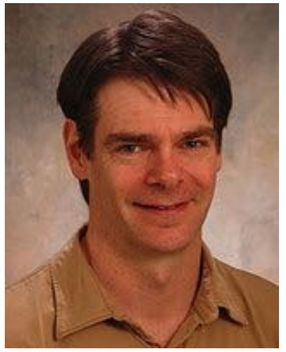
Funding

Computational resources



Gumbart lab

Chris Rowley



Benoit Roux

Questions?