Supporting Information for "Optimized molecular dynamics force fields applied to the helix-coil transition of polypeptides"

Robert B. Best,[*,†] and Gerhard Hummer[‡]

Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, U.K., and Laboratory of Chemical Physics, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland 20892-0520, U.S.A.

E-mail: rbb24@cam.ac.uk

Alternative Definitions of "Helix"

We have tested the sensitivity of the "fraction helix" calculated from the simulations to different definitions of "helix". The following four definitions were considered:

1. (ϕ,ψ) cutoff I. This definition is the one used in the main text, requiring that at least three consecutive residues lie within the $\alpha_{\rm h}$ region, defined as[1] $|-65^\circ-\phi|<35^\circ$ and $|-37^\circ-\psi|<30^\circ$.

2. (ϕ,ψ) cutoff II. The same as I, but extending the definition of $\alpha_{\rm h}$ by 10°, i.e., $|-65^\circ-\phi|<45^\circ$ and $|-37^\circ-\psi|<40^\circ$.

[†]Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, U.K.

[‡]Laboratory of Chemical Physics, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland 20892-0520, U.S.A.

3. (ϕ, ψ) cutoff III. As for I and II, but α_h is defined as the maximally inclusive α_+ region, i.e., $-160^\circ \leq \phi \leq -20^\circ$ and $-120^\circ \leq \psi \leq 50^\circ$.

4. Hydrogen bond definition. We defined helical hydrogen bonds using the Kabsch-Sander hydrogen-bond energy from DSSP.[2] For this definition, partial charges of $+0.2$, -0.2, $+0.42$, and $-0.42e$ are assigned to the amide H and N, and the carbonyl C and O, respectively. A hydrogen bond is assigned if the electrostatic energy of interaction between the NH and CO dipoles is less than 0.5 kcal.mol^{-1}; all residues lying between a pair of (i,i+4) hydrogen-bonded residues are assigned as helical.

The effect of each of these four definitions on the per residue fraction helix and overall fraction helix is shown in Supporting Information (SI) Figure 2 and SI Figure 3, respectively. Most of the alternative definitions tend to increase the helix content, although only slightly, over that used in the main text. The temperature dependence of the helical fraction is, if anything, weaker with the alternative definitions, so the major conclusions of the paper are unaffected. The apparently larger helix fraction obtained for ff99SB using the hydrogen bond definition 4 has no dependence on temperature - these interactions are most likely transient contacts formed in the unfolded peptide.

Effect of Density and Pressure on Helix Formation

The results obtained with different box sizes in the main text can be used to construct a phase diagram for the fraction of helix as a function of density and temperature, as shown in SI Figure 7 (A). Although the pressure varies strongly with density (SI Figure 7 (B)), there is only a small effect of density or pressure on the overall fraction helix, consistent with the simulation study of Paschek et al.[3]

Alternative Lifson-Roig Models

In addition to the elementary Lifson-Roig (LR) model described in the main text, we have also considered two potential modifications, which improve the agreement between the model and simulation data:

1. Treating Ala, Gln independently (i.e., with separate w_A, v_A, w_Q, v_Q)

2. Inclusion of capping effects (e.g., the strong N-terminal capping effect of the acetyl blocking group).

Lifson-Roig Model Including N- and C-Capping

The modified LR model including capping effects is the same as the original, except that the non-helical residues adjacent to the N- and C-termini of a helical segment are weighted by n_i and c_i, respectively, to account for "capping" effects. The subscripts i indicate that each of these parameters is potentially dependent on the residue types concerned. The partition function for this model for a peptide with N amino acid residues can be written in a compact matrix form as:[4]

$$Z = \mathbf{V}^{\mathrm{T}} \prod_{i=0}^{N+1} \mathbf{M}_i \mathbf{V} \tag{1}$$

in which the matrix $\mathbf{M}_i$ is defined as:

$$\mathbf{M}_i = \begin{pmatrix} w_i & w_i c_{i+2} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & v_i & 0 & 0 \\ w_i n_{i-2} & w_i n_{i-2} c_{i+2} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & v_i & v_i & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{pmatrix} \tag{2}$$

and $\mathbf{V}^{\mathrm{T}}$ is the row vector $\begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$. The residues 0 and $N+1$ correspond to the acetyl and amide blocking groups, respectively. The average "fraction helicity" for a given residue i is defined as the population of conformations where it has weight w_i, given by $\langle h_i \rangle_{\mathrm{LR}} = \partial \ln Z / \partial \ln w_i$. The average fraction of helical residues, $\langle f_h \rangle_{\mathrm{LR}}$, is the average over these for the non-terminal residues $\langle f_h \rangle_{\mathrm{LR}} = 1/(N-2) \sum_{i=2}^{N-1} h_i$ (since terminal residues cannot have a w weight). The average number of helical segments $\langle n_s \rangle_{\mathrm{LR}}$ is given by $\langle n_s \rangle_{\mathrm{LR}} = \sum_{i=2}^{N-1} \partial^2 \ln Z / \partial \ln w_i \partial \ln c_{i+2}$, in which $\partial^2 \ln Z / \partial \ln w_i \partial \ln c_{i+2}$ identifies the fraction of the ensemble in which residue i is second to last in a helix.

As for the simpler model, we have fitted the simulation data to the model using a Bayesian formalism, assuming uniform priors for the parameters $\ln w_i$, $\ln v_i$, $\ln c_i$, and $\ln n_i$. The modified log-likelihood can be written as

$$\ln L = \sum_i N_{w,i} \ln w_i + \sum_i N_{v,i} \ln v_i + \sum_i N_{n,i} \ln n_i + \sum_i N_{c,i} \ln c_i - N_k \ln Z \tag{3}$$

where $N_{w,i}$ is the total number of times that residue i has weight w_i in the N_k conformations, and the other terms are defined analogously.

Results with Modified Models

In the following, we denote the LR model used in the main text as "LR", the LR model with independent parameters for Ala, Gln as "LR-res", the LR model with common Ala,Gln parameters but including capping as "LR-nc" and the model with both independent Ala, Gln parameters and capping as "LR-res-nc".

Inclusion of capping whilst still using the same parameters for Ala, Gln (model "LR-nc") gives a much better account of the distribution of helical residues within the peptide (SI Figure 8 (C)), as expected. Remarkably however, the helix-coil parameters w and v for the "LR-nc" model are almost identical to those obtained without capping, model "LR" (SI Figure 11). This motivated our decision not to include a description of capping in the main text, since w and v are the key

parameters of interest. A second observation that can be made is that the n parameter for the acetyl group (Ace) is much larger than Ala (defined to be 1), in agreement with experiment;[5] the relatively weaker effect of the C-terminal amide group (NH_2) relative to Ace is also in agreement with experiment.[5] Note that the correction terms to ψ do not affect the capping parameters (e.g., n and c are the same for ff03 and ff03*), as expected (data not shown); further, essentially the same capping parameters are obtained for Ace and NH_2 for both the "LR-nc" and "LR-res-nc" models, indicating that the capping parameters are independent of the treatment of the main LR parameters w and v.

Using separate parameters for Ala and Gln (models "LR-res", "LR-res-nc") clearly improves the agreement with the simulation (also indicated by a significantly increased log-likelihood): see SI Figure 9 and SI Figure 10. We find that w and v are both lower for Gln than for Ala. The finding of a lower w for Gln than Ala is also consistent with experimental estimates,[5] as is the fact that Ace is a better N-cap than Gln, and Gln is a better C-cap than NH_2.[5] Inclusion of both separate Ala and Gln parameters, and capping clearly improves the prediction of the number of helical segments by the model, as shown in SI Figure 10(B).

References

(1) García, A. E.; Sanbonmatsu, K. Y. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 2782–2787.

(2) Kabsch, W.; Sander, C. *Biopolymers* **1983**, *22*, 2577–2637.

(3) Paschek, D.; Gnanakaran, S.; García, A. E. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 6765–6770.

(4) Rohl, C. A.; Chakrabartty, A.; Baldwin, R. L. *Protein Sci.* **1996**, *5*, 2623–2637.

(5) Doig, A. J.; Baldwin, R. L. *Protein Sci.* **1995**, *4*, 1325–1335.

(6) Graf, J.; Nguyen, P. H.; Stock, G.; Schwalbe, H. *J. Am. Chem. Soc.* **2007**, *129*, 1179–1189.

(7) Lovell, S. C.; Davis, I. W.; III, W. B. A.; de Bakker, P. I. W.; Word, J. M.; Prisant, M. G.; Richardson, J. S.; Richardson, D. C. *Proteins* **2003**, *50*, 437–450.

(8) Mackerell, A. D.; Feig, M.; Brooks, C. L. *J. Comp. Chem.* **2004**, *25*, 1584–1604.

(9) Mackerell, A. D., Jr. *J. Comp. Chem.* **2004**, *25*, 1584–1604.

Tables

Table 1: Atom types and charges for the S-methylated cysteine (CME) residue in HEWL19.[6] Figure 1 shows the structure of the residue.

Atom	Amber Type	ff03 Charge	ff99SB Charge
N	N	-0.396165	-0.41570
H	H	0.295187	0.27190
CA	CT	-0.073501	0.02130
HA	H1	0.140510	0.11240
CB	CT	-0.221371	-0.12310
HB1	H1	0.146537	0.11120
HB2	H1	0.146537	0.11120
SG	S	-0.285182	-0.31190
CD	CT	-0.080726	-0.07670
HD1	H1	0.090000	0.09000
HD2	H1	0.090000	0.09000
HD3	H1	0.090000	0.09000
C	C	0.643035	0.59730
O	O	-0.584861	-0.56790

Table 2: Scalar couplings of Ala_5 calculated with original and optimized force fields.

	Expt.	σ	ff03	ff99SB	ff03*	ff99SB*
A2 $^1J_{NC\alpha}(\psi_2)$	11.36	0.59	11.28	11.33	11.45	11.38
A3 $^1J_{NC\alpha}(\psi_3)$	11.26	0.59	11.06	11.22	11.27	11.11
A4 $^1J_{NC\alpha}(\psi_4)$	11.25	0.59	11.05	11.27	11.18	11.09
A2 $^2J_{NC\alpha}(\psi_1)$	9.20	0.50	8.58	8.55	8.59	8.54
A3 $^2J_{NC\alpha}(\psi_2)$	8.55	0.50	8.10	8.18	8.39	8.22
A4 $^2J_{NC\alpha}(\psi_3)$	8.40	0.50	7.80	8.02	8.17	7.86
A5 $^2J_{NC\alpha}(\psi_4)$	8.27	0.50	7.71	8.13	8.02	7.80
A2 $^3J_{H\alpha C}(\phi_2)$	1.85	0.38	1.80	2.28	1.80	2.09
A3 $^3J_{H\alpha C}(\phi_3)$	1.86	0.38	1.86	2.49	1.86	2.14
A4 $^3J_{H\alpha C}(\phi_4)$	1.89	0.38	1.95	2.20	1.94	2.51
A5 $^3J_{H\alpha C}(\phi_5)$	2.19	0.38	1.86	2.56	1.87	2.80
A4 $^3J_{HNC}(\phi_4)$	1.15	0.59	0.76	0.83	0.81	0.98
A5 $^3J_{HNC}(\phi_5)$	1.16	0.59	1.12	1.14	1.09	1.46
A2 $^3J_{HNC\beta}(\phi_2)$	2.30	0.39	2.54	2.10	2.49	2.17
A3 $^3J_{HNC\beta}(\phi_3)$	2.24	0.39	2.62	2.05	2.54	2.34
A4 $^3J_{HNC\beta}(\phi_4)$	2.14	0.39	2.48	2.07	2.47	2.03
A5 $^3J_{HNC\beta}(\phi_5)$	1.96	0.39	2.29	1.30	2.33	1.63
A2 $^3J_{HNH\alpha}(\phi_2)$	5.59	0.91	6.69	7.39	6.71	7.45
A3 $^3J_{HNH\alpha}(\phi_3)$	5.74	0.91	6.89	7.54	6.91	7.47
A4 $^3J_{HNH\alpha}(\phi_4)$	5.98	0.91	7.17	7.85	7.07	7.65
A5 $^3J_{HNH\alpha}(\phi_5)$	6.54	0.91	6.73	8.68	6.74	7.50
A2 $^3J_{HNC\alpha}(\phi_2,\psi_1)$	0.67	0.10	0.68	0.70	0.69	0.70
A3 $^3J_{HNC\alpha}(\phi_3,\psi_2)$	0.68	0.10	0.58	0.63	0.64	0.62
A4 $^3J_{HNC\alpha}(\phi_4,\psi_3)$	0.69	0.10	0.55	0.63	0.62	0.59
A5 $^3J_{HNC\alpha}(\phi_5,\psi_4)$	0.73	0.10	0.56	0.70	0.61	0.60
A2 $^3J_{CC}(\phi_2)$	0.19	0.22	0.96	1.06	1.01	1.06
$\chi^2(J)$			1.2	1.7	1.1	1.7

Table 3: Scalar couplings of central Ala residues in the HEWL19 peptide.

		Expt.	σ	ff03	ff99SB	ff03*	ff99SB*
A9	$^1J_{NC\alpha}(\psi_9)$	10.54	0.59	10.13	10.84	10.68	10.53
A10	$^1J_{NC\alpha}(\psi_{10})$	10.58	0.59	10.04	10.84	10.51	10.27
A11	$^1J_{NC\alpha}(\psi_{11})$	10.57	0.59	9.98	10.64	10.38	10.21
A10	$^2J_{NC\alpha}(\psi_9)$	7.24	0.50	6.48	7.50	7.43	7.06
A11	$^2J_{NC\alpha}(\psi_{10})$	7.02	0.50	6.35	7.54	7.01	6.64
M12	$^2J_{NC\alpha}(\psi_{11})$	7.17	0.50	6.32	7.01	6.74	6.57
A9	$^3J_{H\alpha C}(\phi_9)$	2.06	0.38	1.22	2.47	1.50	1.76
A10	$^3J_{H\alpha C}(\phi_{10})$	1.72	0.38	1.40	2.55	1.55	1.82
A11	$^3J_{H\alpha C}(\phi_{11})$	2.20	0.38	1.48	2.87	1.75	2.16
A9	$^3J_{HNC}(\phi_9)$	1.39	0.59	0.81	1.05	0.96	1.01
A10	$^3J_{HNC}(\phi_{10})$	1.33	0.59	0.58	1.10	1.00	0.65
A11	$^3J_{HNC}(\phi_{11})$	1.09	0.59	0.56	1.22	0.59	0.77
A9	$^3J_{HNC\beta}(\phi_9)$	2.26	0.39	3.56	2.03	3.00	2.74
A10	$^3J_{HNC\beta}(\phi_{10})$	2.19	0.39	3.48	2.23	2.86	3.01
A11	$^3J_{HNC\beta}(\phi_{11})$	2.21	0.39	3.38	2.31	3.03	2.37
A9	$^3J_{HNH\alpha}(\phi_9)$	5.18	0.91	5.10	7.49	5.79	6.22
A10	$^3J_{HNH\alpha}(\phi_{10})$	5.10	0.91	5.72	7.08	5.95	6.46
A11	$^3J_{HNH\alpha}(\phi_{11})$	5.67	0.91	5.93	6.67	6.50	7.45
A10	$^3J_{HNC\alpha}(\phi_{10},\psi_9)$	0.46	0.10	0.24	0.52	0.46	0.39
A11	$^3J_{HNC\alpha}(\phi_{11},\psi_{10})$	0.43	0.10	0.22	0.49	0.36	0.38
M12	$^3J_{HNC\alpha}(\phi_{12},\psi_{11})$	0.43	0.10	0.24	0.49	0.39	0.38
$\chi^2(J)$				3.2	1.2	1.0	1.0

Figures

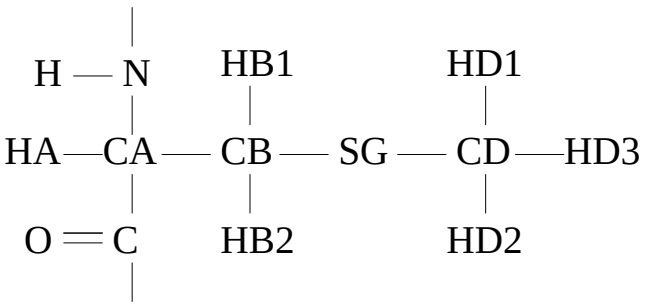

Figure 1: Covalent structure of the modified CME residue (see SI Table 1 for parameter values in ff03 and ff99SB).

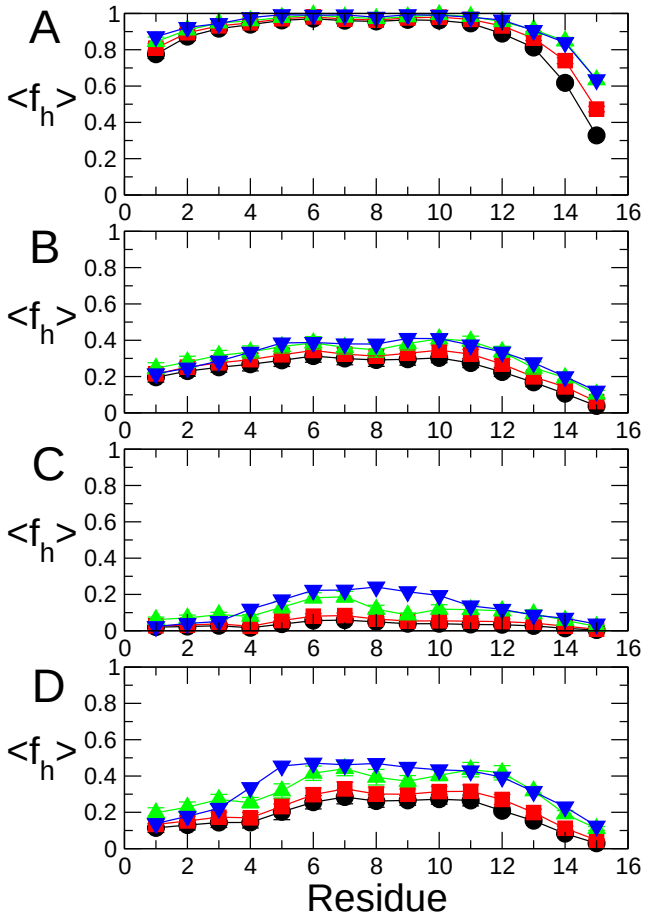

Figure 2: Dependence of per residue fraction helix on definition of helix: (A) ff03, (B) ff03*, (C) ff99SB, (D) ff99SB* at 303 K. Definitions 1-4 (see SI text) are given, respectively, by black circles, red squares, green “up” triangles and blue “down” triangles.

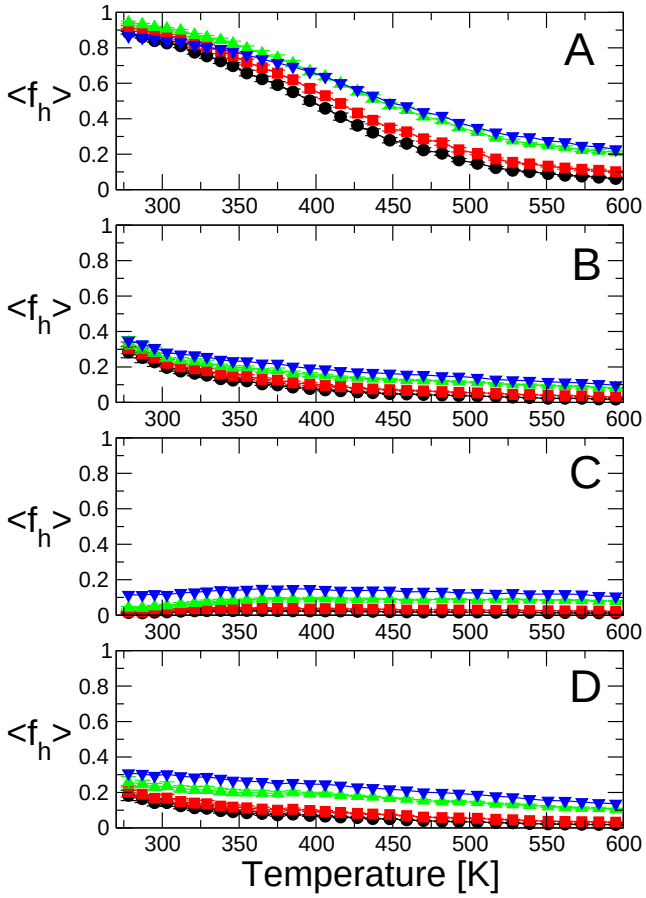

Figure 3: Dependence of overall fraction helix on definition of helix: (A) ff03, (B) ff03*, (C) ff99SB, (D) ff99SB*. Definitions 1-4 (see text) are given respectively by black circles, red squares, green "up" triangles and blue "down" triangles.

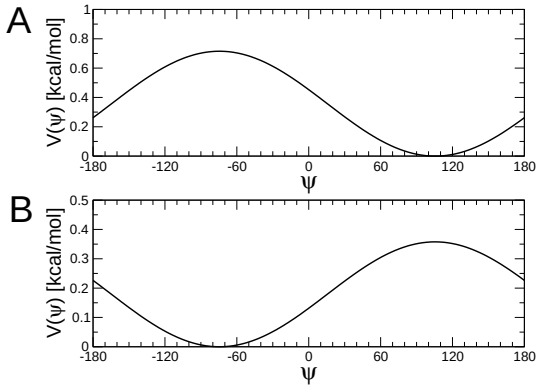

Figure 4: Torsional energy corrections $V(\psi)$ used in the (A) ff03* and (B) ff99SB* force fields.

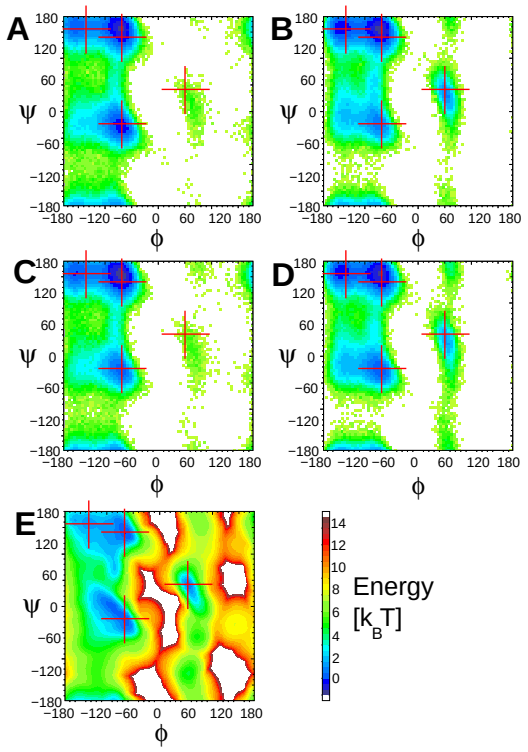

Figure 5: Ramachandran potentials of mean force $-\ln P(\phi,\psi)$ for the central residues of Ala_5 obtained from long equilibrium simulations with (A) ff03, (B) ff99SB, (C) ff03*, (D) ff99SB*. Panel (E) shows the corresponding potential of mean force computed from the protein data bank for residues not in secondary structure.[7] Red crosses in each panel indicate that the minima for each type of secondary structure are in similar positions on the map.

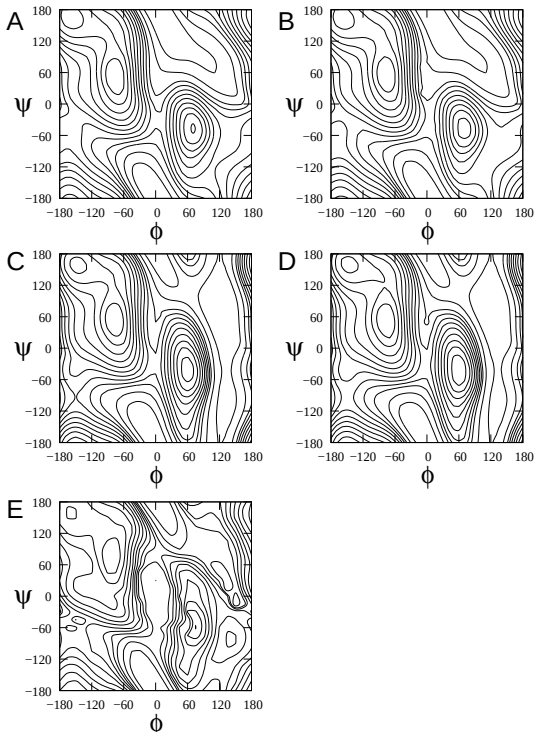

Figure 6: Gas phase (ϕ, ψ) potential energy surfaces for alanine dipeptide computed with (A) ff03, (B) ff03*, (C) ff99SB, and (D) ff99SB*, using infinite cut-offs. In panel (E) is shown the LMP2/cc-pVQZ//MP2/6-31g* surface reported by Mackerell *et al.* [8,9] Contours are given in 1 kcal/mol intervals from 0 to 10 kcal/mol above the minimum energy in each map, with two additional contours at 12 and 15 kcal/mol. Least square differences from the LMP2 energies, after adding an optimal constant offset, are (kcal/mol): 1.32 (ff03), 1.29 (ff03*), 1.69 (ff99SB) and 1.69 (ff99SB*), over regions with energies less than 7 kcal/mol on the LMP2 surface.

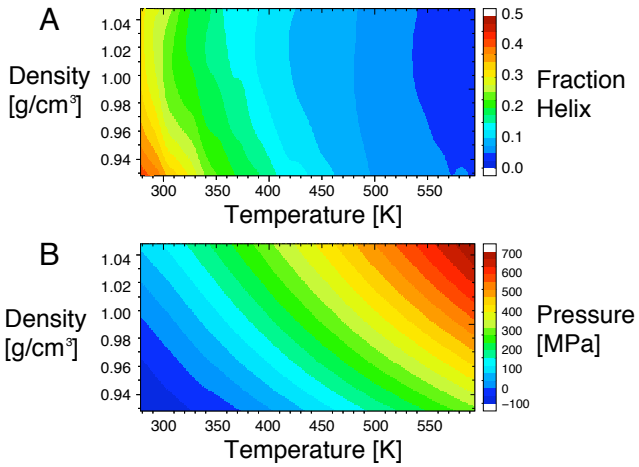

Figure 7: Effect of pressure on helix formation. (A) Total fraction helix as a function of system density and temperature, constructed from simulations at three different densities; (B) corresponding average pressure at each density and temperature. Smooth surfaces were constructed by interpolation using a two-dimensional cubic spline.

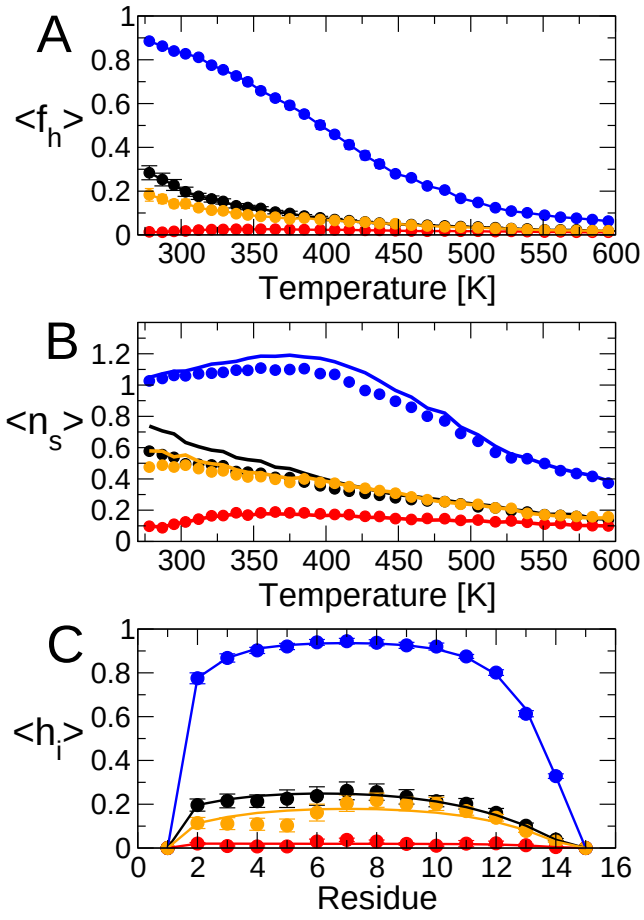

Figure 8: Quality of LR fit for model "LR-nc": capping included, Ala, Gln combined. (A) The average fraction helix $\langle f_h \rangle$, with simulation data as symbols and model predictions as solid lines (blue: ff03, red: ff99SB, black: ff03*, orange: ff99SB*). (B) The average number of "helical segments", $\langle n_s \rangle$, with simulation data as symbols and model predictions as solid lines; broken lines give the prediction using the expression $\langle n_s \rangle \approx \sum_i \partial \ln Z / \partial \ln v_{12,i}$. (C) Fraction helicity for each residue in the 303 K replica, symbols and lines as above.

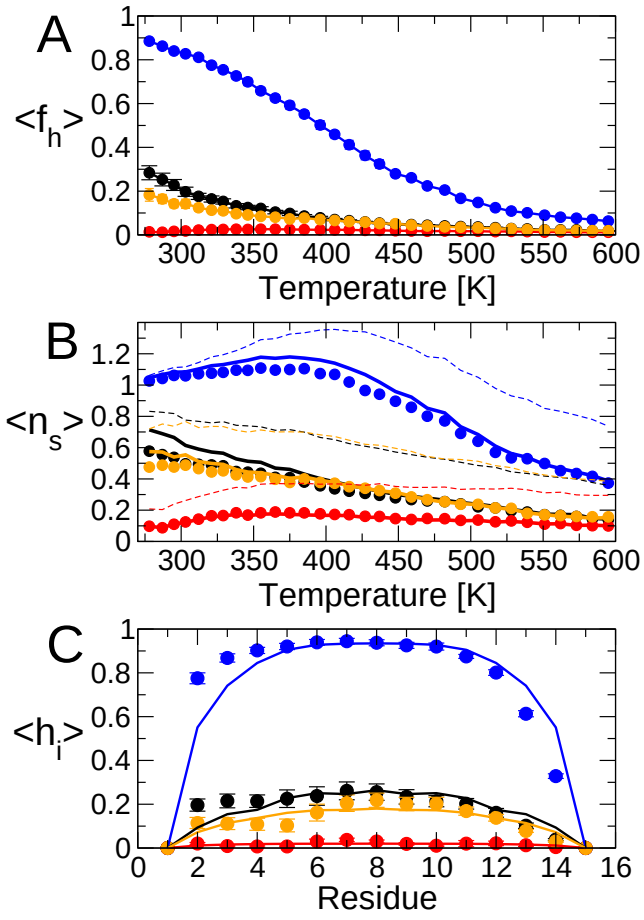

Figure 9: Quality of LR fit for model "LR-res": capping not included, Ala, Gln independent. (A) The average fraction helix $\langle f_h \rangle$, with simulation data as symbols and model predictions as solid lines (blue: ff03, red: ff99SB, black: ff03*, orange: ff99SB*). (B) The average number of "helical segments", $\langle n_s \rangle$, with simulation data as symbols and model predictions as solid lines; broken lines give the prediction using the expression $\langle n_s \rangle \approx \sum_i \partial \ln Z / \partial \ln \nu_{12,i}$. (C) Fraction helicity for each residue in the 303 K replica, symbols and lines as above.

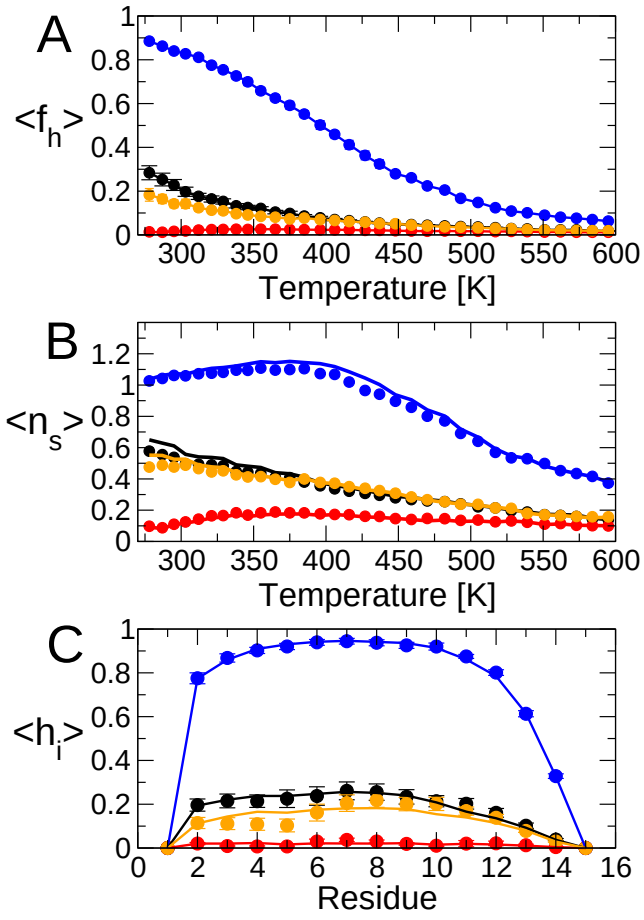

Figure 10: Quality of LR fit for model "LR-res-nc": capping included, Ala, Gln independent. (A) The average fraction helix $\langle f_h \rangle$, with simulation data as symbols and model predictions as solid lines (blue: ff03, red: ff99SB, black: ff03*, orange: ff99SB*). (B) The average number of "helical segments", $\langle n_s \rangle$, with simulation data as symbols and model predictions as solid lines; broken lines give the prediction using the expression $\langle n_s \rangle \approx \sum_i \partial \ln Z / \partial \ln v_{12,i}$. (C) Fraction helicity for each residue in the 303 K replica, symbols and lines as above.

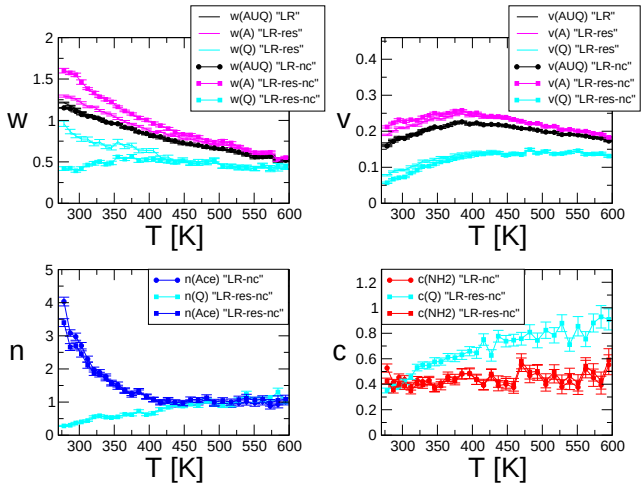

Figure 11: Comparison of fit parameters for different helix-coil models, for the ff03* force field: parameters are given in the key for each panel.

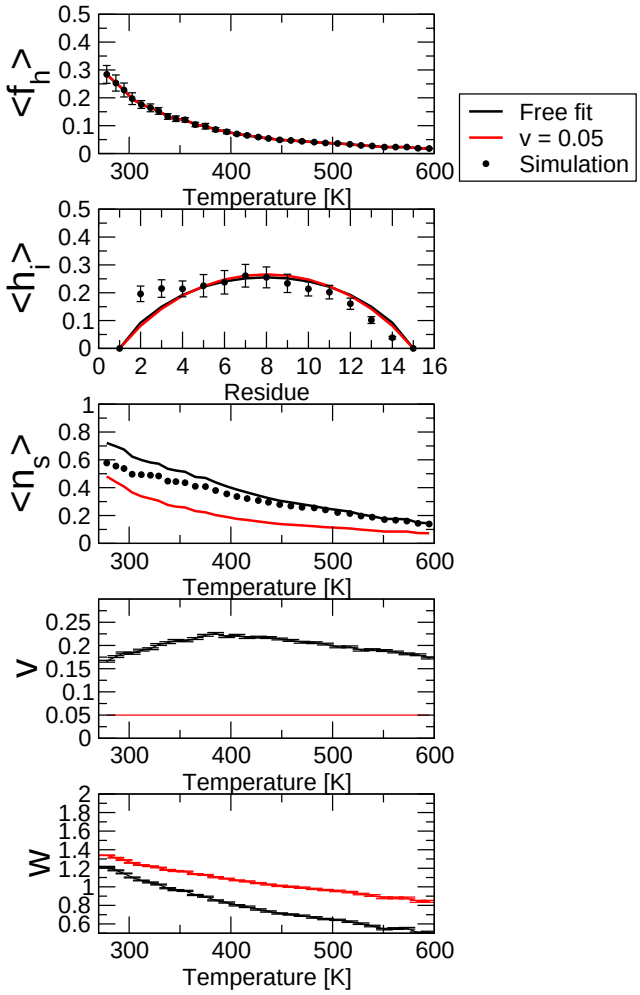

Figure 12: Bayesian fit of simple LR model to ff03* data with w,v both free (black curves) and with v fixed to 0.05 (red curves). Although the fit with fixed v can reproduce the overall fraction helix $\langle f_h \rangle$ and per residue fraction helix $\langle h_i \rangle$, it markedly underestimates the average number of helical segments $\langle n_s \rangle$.