JCTC Journal of Chemical Theory and Computation
ARTICLE
pubs.acs.org/JCTC

50 Years of Lifson–Roig Models: Application to Molecular Simulation Data

Andreas Vitalis* and Amedeo Caflisch

Department of Biochemistry, University of Zurich, Winterthurerstrasse 190, CH-8057 Zurich, Switzerland

S *Supporting Information*

ABSTRACT: Simple helix–coil transition theories have been indispensable tools in the analysis of data reporting on the reversible folding of α-helical polypeptides. They provide a transferable means to not only characterize different systems but to also compare different techniques, viz., experimental probes monitoring helix–coil transitions in vitro or biomolecular force fields in silico. This article addresses several issues with the application of Lifson–Roig theory to helix–coil transition data. We use computer simulation to generate two sets of ensembles for the temperature-controlled, reversible folding of the 21-residue, alanine-rich FS peptide. Ensembles differ in the rigidity of backbone bond angles and are analyzed using two distinct descriptors of helicity. The analysis unmasks an underlying phase diagram that is surprisingly complex. The complexities give rise to fitted nucleation and propagation parameters that are difficult to interpret and that are inconsistent with the distribution of isolated residues in the α-helical basin. We show that enthalpies of helix formation are more robustly determined using van't Hoff analysis of simple measures of helicity rather than fitted propagation parameters. To overcome some of these issues, we design a simple variant of the Lifson–Roig model that recovers physical interpretability of the obtained parameters by allowing bundle formation to be described in simple fashion. The relevance of our results is discussed in relation to the applicability of Lifson–Roig models to both in silico and in vitro data.

INTRODUCTION

Elucidating the helix–coil transition microscopically has long been deemed to be of utmost importance for the understanding of protein folding, and the reader is referred to excellent review articles for further reading.[1,2] The process is of such elementary nature that it has also become an indispensable benchmark for the development of biomolecular force fields.[3−7]

Helix–coil transition data are often analyzed in an established statistical framework such as that of Zimm and Bragg,[8] Gibbs and DiMarzio,[9] or Lifson and Roig (LR).[10] In the latter, it is assumed that the potential energy function of the system can be mapped to terms written over the ϕ/ψ angles of individual polypeptide residues with the exception of an α-helical hydrogen-bonding term coupling residue i energetically to residues $i-1$ and $i+1$. This term is triggered as soon as three consecutive residues are in a helix-competent conformation, and the resultant favorable energy contribution is mapped exclusively onto residue i. In the absence of hydrogen bonds, the statistical weights of helix-competent vs helix-incompetent ("coil") states correspond to the respective, partial integrals over the Ramachandran map that due to the lack of residue–residue coupling can be formulated for each residue individually:

$$u_i' = \int_{c_i} e^{-\beta U(\varphi,\psi)} \mathrm{d}\varphi_i \mathrm{d}\psi_i \tag{1}$$

$$v_i' = \int_{h_i} e^{-\beta U(\varphi,\psi)} \mathrm{d}\varphi_i \mathrm{d}\psi_i \tag{2}$$

Here, c_i and h_i denote the helix-incompetent and helix-competent regions of ϕ/ψ space, respectively, while U is the (unknown) potential energy function. The LR model stipulates that whenever three consecutive residues are in helical conformation, stabilization occurs and another statistical weight, w'_i, is invoked. Recognizing the arbitrary absolute scale of the energy in the system, the statistical weights can be normalized by u'. Then, the low level of coupling allows the partition function to be expressed in matrix form:[10]

$$Z = \begin{pmatrix} 0 & 0 & 1 \end{pmatrix} \cdot \left[\Pi_{i=1}^{N_r} \begin{pmatrix} w_i & v_i & 0 \\ 0 & 0 & 1 \\ v_i & v_i & 1 \end{pmatrix} \right] \cdot \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} \tag{3}$$

Here, N_r is the number of residues with peptide bonds on both sides and is equivalent to the number of amino acids for capped polypeptides. If we ignore any sequence specificity (including end effects), the matrices become identical, i.e., all residue subscripts can be dropped, and it is possible to obtain global averages as follows:

$$\begin{aligned} \langle N_h \rangle &= \frac{\partial \ln Z}{\partial \ln w} \\ \langle N_s \rangle &= \frac{\partial \ln Z}{\partial \ln v_{12}} \end{aligned} \tag{4}$$

Here, N_h denotes the number of α-helical hydrogen bonds, and N_s the number of helical segments. Matrix element v_{12} refers to a single instance of v in the matrix. Note that N_s by definition includes segments of only two helical residues in a row with no hydrogen bonds formed. This is because the formalism in eq 3 only scans three consecutive residues, and v_{12} corresponds to states of the type "hhc" (helix, helix, coil) regardless of configuration of preceding residues. Further illustrations of LR theory using a simple example are provided in the Supporting Information.

Received: October 24, 2011
Published: December 14, 2011

Because w denotes the ratio of the statistical weights of hydrogen-bonded and coil states for an individual residue, it is often assumed to correspond directly to the stepwise equilibrium constant of helix elongation, i.e., $-\beta^{-1} \ln w \approx \Delta G_{\mathrm{hb}}$, where β is the inverse thermal energy, and ΔG_{hb} the free energy gain associated with the formation of a single hydrogen bond. It is possible to determine w from equilibrium experiments that are able to estimate helix content directly, such as temperature-dependent circular dichroism (CD) spectroscopy, by fitting the raw data to a two-state model that allows extraction of $\langle N_{\mathrm{h}} \rangle$ and subsequent fitting to yield w. This requires knowledge of v, which is often obtained independently or can be fit if data for multiple chain lengths are available.[11] Limitations of the LR model were established and analyzed soon after publication of the original model, and extensions were suggested.[12,13] Throughout the past two decades, specific modifications were proposed that incorporate helix capping,[14] short-range side chain interactions,[15] or extensions beyond the triplet model.[16]

Experimental analyses of the helix−coil transition designed to extract more than just helix content and to interpret the results in terms of a microscopic theory have had to utilize assumptions to avoid overfitting the data. Rohl et al.[17] used the kinetics of amide proton exchange to show that a single model with essentially three parameters can fit data at a single temperature for polypeptides of the series Ace-(AAKAA)$_m$Y-NH$_2$ reasonably well for values of m ranging from 1 to 10. Such a simple model was obtained by assuming a homopolymer and by assuming that exchange in the only considered hydrogen-bonded state (α-helix) is completely quenched. Then, the free parameters were the exchange rate in the coil-state and the aforementioned helix nucleation and propagation parameters.

Later, the same authors showed that, for a nearly identical series of peptides and over a limited range of temperatures, two types of fits with similar quality could be obtained, both using a T-independent helix nucleation parameter and again assuming homopolymeric behavior.[18] In the first fit, T-dependent propagation parameters were derived from CD, and the exchange rates in the coil-state were fitted, whereas in the second, the exchange rate was fixed to that observed for the shortest peptide, and helix propagation parameters were fit. These fits show small systematic deviations and yielded a slight inconsistency that was interpreted as stemming from the inapplicability of the exchange rate in the canonical coil state (shortest peptide) to the coil state seen for longer peptides capable of forming helices.

Thompson et al.[19] constructed a kinetic zipper model for a similar alanine-based peptide (termed FS-peptide)[20] to simultaneously interpret data from laser T-jump experiments and CD. They found that the equilibrium data could be equally well reproduced by different parameter sets but that relaxation rates were only consistent with values of the T-independent nucleation parameter that are significantly larger than those reported by Rohl and colleagues.[17] In their model, Thompson et al. were, however, restricted to the assumption of only a single helical segment being allowed to form. Examples, such as the three studies mentioned above, have led to the transferability of parameters derived from LR models being questioned.[21,22]

Based on the extensive literature on the subject, several assumptions inherent to the application of LR models to helix−coil transition data emerge as questionable:

(1) Even in the absence of helix formation, independence of the backbone angles of individual residues does not hold.[22,23]

(2) Helix stability does not just depend on hydrogen bonds but encompasses solvation and hydrophobic terms.[24−26]

(3) Scattering experiments and in silico studies have proposed that the single-sequence approximation is misleading even for relatively short peptides.[27,28] It appears quite likely that helix bundles form through stabilization by tertiary interactions that are not representable in LR models. Such interactions are also one possible explanation for the observed length-dependent propagation behavior of α-helices.[29−31]

(4) As a corollary to the previous point, it is worth mentioning that LR models predict that very long helices are extremely stable, which contrasts with the low prevalence of long helices in biological systems: The likelihood of observing helices longer than 15 residues in globular proteins decreases rapidly,[28] and even putative coiled-coil domains rarely exceed 150 residues despite the presence of stabilizing and specific tertiary interactions.[32] Of course, these data provide indirect evidence only as the impacts of evolutionary pressures vs physicochemical properties cannot be delineated.

In addition, applications of LR models to molecular simulation data have revealed that in almost all molecular force fields, the statistical likelihood of occupying the region of Ramachandran space compatible with α-helical hydrogen bonds is larger than proposed nucleation parameters suggest. Nucleation parameters are routinely overestimated when analyzing in silico data,[5,33,34] and this constitutes either a fundamental error in force fields or a disconnect between in vitro and in silico interpretations of helix nucleation.

In this contribution, we employ molecular dynamics simulations in an all-atom representation of the FS-peptide (acetyl-A$_5$(AAARA)$_3$A-*N*-methylamide) coupled to the recently developed ABSINTH implicit solvation model.[34] Our aim was to generate a diverse but statistically sound set of data that highlight limits of applicability and interpretability of LR fits and parameters to computational and atomistic sampling of the temperature-dependent helix−coil transition. We employ a wide range of simulation temperatures and compare models differing in the imposed rigidity of backbone bond angles to explore the thermodynamics of the transition in richer detail. The known impact of such constraints[35] is found to be large and is affecting qualitative features of the sampled ensembles as well. Using our simulation data, we show that LR fits yield results that are unsatisfactory either in terms of parameter interpretability or in terms of fit accuracy. We highlight the lack of transferability by showing that the temperature dependence of the fitted helix propagation parameter cannot be connected easily to the bulk behavior of the peptide. Finally, we suggest additional tests and alternative routes for analyzing in silico data, the most important one being the inclusion of the mean number of isolated residues in the α-helical basin, $\langle N_1 \rangle$, in the LR fitting.

■ METHODS

Simulation Design. The FS-peptide (acetyl-A$_5$(AAARA)$_3$A-*N*-methylamide)[20] was enclosed in a spherical droplet of 40 Å radius along with explicit sodium and chloride ions compensating the peptide's positive charge and adding a background electrolyte concentration of ∼150 mM. Starting configurations were random aside from satisfying excluded volume requirements. The effects of water were described by the ABSINTH implicit solvation model,[34] which is a group transfer-based model similar in spirit to the EEF1 model[36] and based in parts on the OPLS-AA/L force field[37] (see Methods in Supporting Information for further details). The simulations integrated Langevin equations of motion at constant volume with a time step of 2.5 fs and a universal atomic friction coefficient of 1.2 ps^{-1}.[38] With these settings, integration was stable, and net temperature artifacts due to integrator, cutoff, and other noise terms assumed maximal values of ~4K for the highest temperature (see below). The use of a Langevin integrator neglecting hydrodynamic interactions with artificially low friction in conjunction with an implicit solvent model means that the resultant conformational dynamics will not be physically realistic. The motivation for this setup lies in obtaining converged equilibrium data of the thermodynamics of the helix−coil transition as a function of temperature, which allow rigorous assessment of LR models.

To additionally enhance sampling, we employed the replica−exchange (RE) technique[39] and constructed two overlapping schedules each consisting of 16 temperatures. The low-temperature schedule used 220, 227, 234, 242, 250, 259, 268, 278, 288, 299, 310, 322, 334, 347, 360, and 374 K, and the high-temperature schedule used 260, 268, 276, 284, 292, 300, 310, 320, 330, 340, 350, 360, 375, 390, 410, and 440 K. Exchanges between neighbors were attempted every 25 ps in either case. The average acceptance probability for the swap moves generally exceeded 33% except for terminal replicas. The low exchange attempt frequency was intended, and the results show that sampling is robust regardless. Comparison of results from the two completely independent runs across the overlapping region allows a simple and rigorous assessment of sampling quality. It should be noted that implicit solvents do not exhibit phase transitions, thereby allowing the use of unusual temperatures. An exact mapping of simulation temperatures to realistic ones is typically not possible. Specifically for the ABSINTH continuum solvation model, temperatures between 280 and 350 K may be reasonably well-represented,[34] but the primary reason for using "unphysical" temperatures lies in our aims to create as diverse an ensemble of helical and coil states as possible and to optimize benefits from RE sampling to obtain statistically sound data.

Residue-based neighbor lists were recalculated every 5 steps, and interactions were generally truncated at 12 Å. Interactions between residues carrying a net charge were not truncated at all but instead computed in a monopole approximation if their distance exceeded 12 Å. The total simulation length of an individual temperature replica was always 250 ns, with the first 50 ns being discarded as equilibration. Two different sets of holonomic constraints were enforced during integration (see below). All simulations were performed using the homegrown CAMPARI software package.[40] The data for alanine dipeptide in Figure 7 were extracted from simulations of 125 ns in length. With the exception of the absence of any ions, these runs used identical conditions and settings and were performed independently for either set of constraints.

Constraints. We simulated the FS-peptide using two different sets of holonomic constraints enforced during integration of the equations of motion via the SHAKE algorithm.[41] The first set constrained the lengths of all covalent bonds. This corresponds to a standard setup in molecular dynamics applications. The second set specifically rigidified backbone bond angles by introducing additional distance constraints between C_α and O, C_α and HN, N and C_β, C and C_β, N and C, C_{i-1} and C_α, and C_α and N_{i+1}. Even though the coupling between constraints is increased, this set is still comfortably solvable by SHAKE. We used a relative tolerance of 10^{-4} and verified that the corresponding internal degrees of freedom were in fact constant throughout the simulations.

It should be noted that we ignored contributions to the equilibrium populations stemming from the mass-metric tensor determinant.[42] Given that fixed bond lengths are not typically considered as a source of bias error and that we introduce only a subset of possible bond angle constraints, we assume that the combination of a stochastic dynamics integrator and a structured energy landscape renders potential artifacts minor.[43,44] Support for this assumption is presented in Figure S1 in the Supporting Information, where we compare molecular dynamics to Monte Carlo data. The latter is based on an implementation[34] that rigidifies all bond angles (and some dihedral angles) and is inherently free of mass-metric tensor artifacts due to the absence of momenta. For the polypeptide backbone, it is therefore very similar to the case with rigidified backbone angles shown here. Consequently, quantitative similarity is expected and largely seen in Figure S1, Supporting Information. Formulations incorporating explicit corrections for mass-metric tensor artifacts exist but require dedicated integrators.[44,45]

Analysis of Simulation Data. Statistics for all data were collected every 25 ps. The α-helical region of Ramachandran space was defined identically to previous work.[34] Define secondary structure of proteins (DSSP) statistics were collected by assigning secondary structure based on hydrogen-bond patterns using the actual trajectory coordinates for amide hydrogen atoms. The default cutoff criteria employed by Kabsch and Sander[46] were used throughout, but numerical tests (not shown) revealed the sensitivity of altering the energetic cutoff for hydrogen bonds from −0.5 to −0.3 and −1.0 kcal/mol, respectively, to be insignificant compared to the differences between flexible and rigidified models or between measures of helicity in Figures 1 or S1, Supporting Information. Helical segments in DSSP require at least two, consecutive hydrogen bonds of $i \rightarrow i + 4$ registry. This means that three-residue segments are missed in the DSSP analysis, which in LR theory are assumed to possess one helical hydrogen bond. Furthermore, one- and two-residue segments are not accounted for at all. Both methods can theoretically yield false positives and false negatives, and this is partially intended: Torsional statistics are purely based on inference, and any three-residue segment assigned as helix may easily be in a conformation not amenable to hydrogen-bond formation. Conversely, α-helical hydrogen bonds may be formed even when not all three intervening residues are in the torsional basins due to compensatory effects. DSSP assignments on the other hand imply that not all hydrogen bonds throughout a helix may satisfy the significance cutoff, but that the residues are treated as a single helix nonetheless. Conversely, two consecutive hydrogen bonds may both be barely within the cutoff and may not correspond to a proper α-helical segment.

Length-dependent statistics for helical segments (continuous residues in helical conformation as determined by either DSSP or torsional occupancy) were collected and used to determine $\langle N_s \rangle$ and $\langle N_1 \rangle$. For each encountered segment, the contribution to $\langle N_h \rangle$ was inferred as $l_s - 2$, where l_s is the length of the corresponding segment. To be able to use DSSP-based statistics consistently, counts for three-residue segments contributing to $\langle N_h \rangle$, for two- and three-residue segments contributing to $\langle N_s \rangle$, and for one-residue segments constituting $\langle N_1 \rangle$ were taken from the torsional assignment instead (this gives rise to the "DSSP corr."

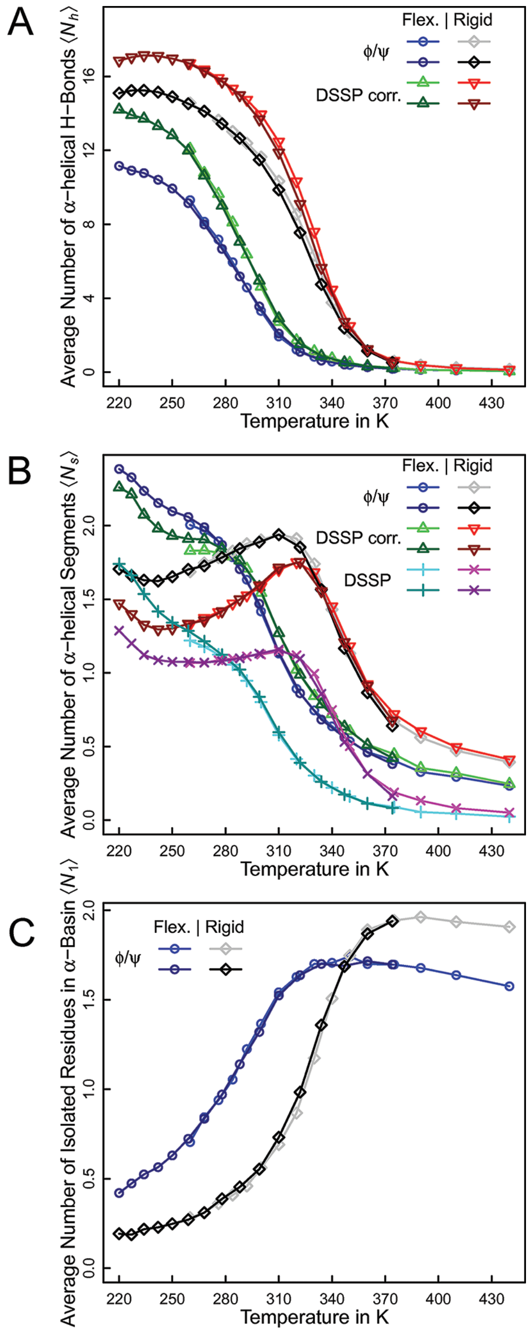

Figure 1. Quantification of helical content as a function of temperature for the FS-peptide using two different sets of holonomic constraints during the simulations. Panel A shows the mean number of α-helical hydrogen bonds, $\langle N_{\mathrm{h}}\rangle$, inferred from either torsional statistics ("ϕ/ψ") or DSSP assignments for the FS-peptide with corrections for three-residue segments (see Methods Section). Data for either set of constraints are indicated in the figure legend as "Rigid" (backbone bond angle constraints) and "Flex." (no bond angle constraints). Panel B plots the mean number of distinct segments with at least two consecutive residues in helical conformation, $\langle N_{\mathrm{s}}\rangle$. DSSP data not including the corrections are shown in addition to the rest. Panel C shows the average number of single residues in helical conformation surrounded by residues in nonhelical conformation, $\langle N_1\rangle$. By construction, only data based on torsional statistics are available. In all plots, darker colors correspond to the replica exchange molecular dynamics (REMD) run across a lower set of temperatures and lighter colors to the higher temperature run. Lines are drawn as a guide to the eye only.

data set in Figures 1, 3, and 6 and S1, S3, and S4, Supporting Information). We believe it is important to include two-residue segments in the counting in contrast to suggestions in the recent literature.[3] This is because otherwise $\langle N_{\mathrm{h}}\rangle$ and $\langle N_{\mathrm{s}}\rangle$ become more closely correlated, and less information than possible is being utilized.

Fitting Procedure. In all fits, the chosen model was fit to the data by a Monte Carlo procedure that allowed randomization over a reasonable interval (10% likelihood, 50% for the parameter f_3 that we introduce in eq 9 below) or stepwise perturbations (90% likelihood, 50% for f_3) of the fit parameters. All parameters were fit simultaneously, and a new set of values was accepted whenever the metric of goodness of fit was improved. The latter was defined as the normalized root-mean-square (rms) deviations of the two or three fitted quantities, viz., either $\langle N_{\mathrm{h}}\rangle$ and $\langle N_{\mathrm{s}}\rangle$ or $\langle N_{\mathrm{h}}\rangle$, $\langle N_{\mathrm{s}}\rangle$, and $\langle N_1\rangle$. The normalization values were 19, 2, and 2, for $\langle N_{\mathrm{h}}\rangle$, $\langle N_{\mathrm{s}}\rangle$, and $\langle N_1\rangle$, respectively. Normalized rms deviations were chosen to achieve a balanced impact of all three quantities irrespective of their value. The fits were generally highly reproducible and did not depend on the initial guess, indicating that a unique optimal solution given the metric of goodness exists. If this was not the case, it is noted in the text.

RESULTS AND DISCUSSION

In published computational work, connections to LR theory are usually made by parsing segment distributions for the peptide in question with respect to the α-basin which is defined by some heuristic.[3,5,33] From this, $\langle N_{\mathrm{s}}\rangle$ and $\langle N_{\mathrm{h}}\rangle$ are estimated by assuming that, just like the LR stipulation, three consecutive residues in helix conformation will yield a hydrogen bond. This is an indirect estimation, and we show in Figure 1 how such inference compares to more direct estimates based on DSSP hydrogen-bond energies.

Cooperative Helix Melting and the Influence of Rigid Backbone Bond Angles. Figure 1A shows results from two independent temperature RE runs each for two different sets of constraints using both DSSP and torsional estimates of the number of α-helical hydrogen bonds. The first noteworthy point is the excellent congruency between the two independent RE runs. Since this constitutes convincing evidence toward the statistical reliability of the data, error estimates from block averaging, which would inherently be less rigorous indicators, are omitted for reasons of clarity from this and all further plots.

What impact does backbone rigidity have on the helix−coil transition? For both sets of constraints, the peptide shows a well-defined melting transition with increasing temperature. The loss of α-helical hydrogen bonds appears cooperative in either case, but, as is observed experimentally,[19,47] occurs over a relatively broad temperature range. If bond angles along the backbone are rigidified, both the melting temperature and the limiting helical content in the helix phase experience a substantial upshift. This is true irrespective of whether hydrogen bonds are inferred by the DSSP algorithm or based on torsional segment statistics. The DSSP-based values are generally larger. This leaves at least two possibilities that are mutually compatible: On average the inference from torsional statistics misses hydrogen bonds (false negatives) and/or the DSSP inference overestimates numbers of hydrogen bonds (false positives, see Methods Section for details).

Panel B shows that there are qualitative differences between the two ensembles as well. With only bond lengths constrained, the average number of segments with at least two consecutive residues in helical conformation increases continuously with decreasing temperature. This suggests that the high flexibility of the chain favors conformations containing two or more shorter helices. Conversely, the introduction of angle constraints in the polypeptide backbone appears to stabilize conformations with just a single helix over a wider range of temperatures leading to an actual decrease in the number of helical segments when reducing the temperature from 300 to 250 K. The uncorrected DSSP-derived segment statistics are not applicable to LR theory since conformations with exactly two or exactly three residues in helical conformation are missed due to the lack of the two hydrogen bonds required according to DSSP (see Methods Section). They can, however, be used to quantify the actual number of well-defined helical segments. This confirms that the qualitative dissimilarity between the two ensembles is robust. If we add the counts from torsional statistics for those two- and three-residue segments to the DSSP counts ("DSSP corr." in the legend), the data for the case of flexible backbone angles are mutually consistent, i.e., the omissions of shorter segments implied by the DSSP algorithm are able to approximately explain the discrepancy in the data. This is *not* true for the case of rigidified backbone angles. Here, the discrepancy seems to stem mostly from a single, long helix being mistakenly broken into two or more pieces by the inaccuracy of the inference of hydrogen bonds based on torsional segment statistics. Later, we will therefore use both data sets.

Lastly, Figure 1C shows the number of single residues that are in helical conformation with both neighbors not being in helical conformation ("one-residue segments"). This is a complementary readout to the data in panel B and can, just like the other two, be directly estimated using the LR formalism:

$$\langle N_1 \rangle = \frac{\partial \ln Z}{\partial \ln v_{32}} \tag{5}$$

Note that v_{32} is the (only) element of the matrix corresponding to three-residue sequences "chc" (coil, helix, coil), i.e., an isolated, helical residue. We will use this readout, which we are unaware of having been employed in the recent literature, and its characteristic temperature dependence below as a weakly dependent test for fits obtained using eqs 4.

In summary, the data in Figure 1 show that bond angle constraints have a profound impact on the nature of helix-rich ensembles even though the melting transition itself may be robust. It is worth pointing out, however, that the differences observed here are still smaller than those seen when comparing different force fields to one another[3,5,48] or when comparing explicit to implicit solvent data.[33] It may be argued that increased local flexibility leads to access to larger parts of the Ramachandran map. Inspection of the corresponding data for alanine dipeptide (not shown) supports this statement and allows the tentative hypothesis that increased likelihood of helix nucleation leads to the shift in the melting transition upon rigidification of backbone bond angles. Importantly, the increased flexibility could also influence segment statistics in an artificial manner given the use of the same definition of the α-basin in either case. This is where the complementary DSSP analysis is important that shows consistent differences between flexible and rigid cases but should not be affected in a similarly straightforward manner by increased local flexibility. In fact, sensitivity analyses (not shown) emphasize the robustness of DSSP estimates with respect to changes in cutoff criteria.

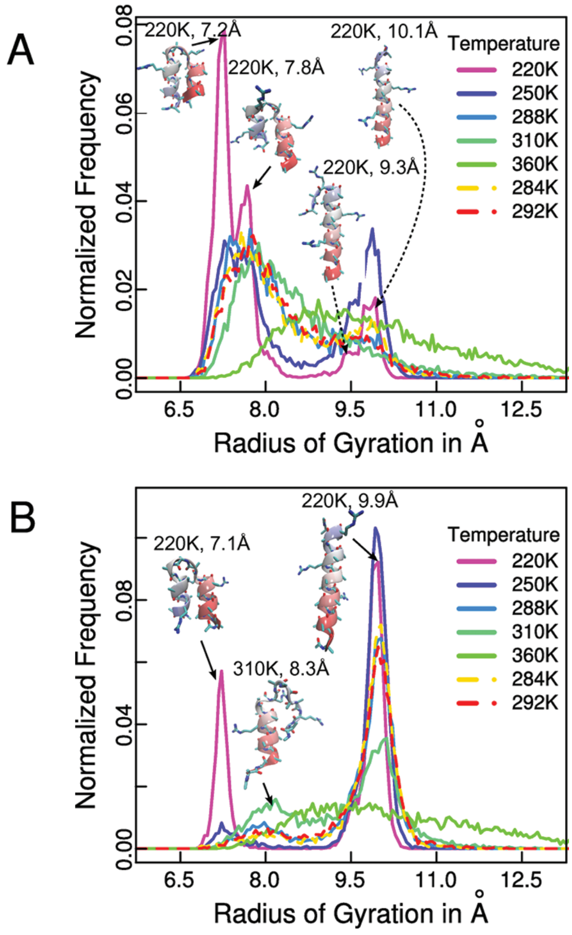

Figure 2. Histograms of radii of gyration (R_g) of the FS-peptide at different temperatures. Panel A shows the data for the case without any bond angle constraints, and panel B for the case with backbone bond angle constraints. The bin size for the construction of the histograms was 0.05 Å. All data drawn with solid lines are extracted from the low-temperature REMD run. To illustrate the statistical reliability of the data, we plot as dashed lines the two temperatures closest to 288 K found in the high-temperature REMD schedule (284 and 292 K). Clearly, the differences between substantially different temperatures vastly exceed the level of statistical noise in the data. To illustrate some of the dominant peaks, cartoon representations of individual structures along with their parent temperature and actual radius of gyration are given. These graphics were generated using VMD.[63]

As a corollary, we do not believe that it is possible to tune the cutoff parameters for the two analyses types to make the resultant estimates of helicity mutually consistent. In fact, qualitative differences should persist on account of the fundamentally different information utilized and DSSP's built-in fault tolerance. In this context, it should be stressed that the definition of the statistical weight w in LR theory does not require a specific interpretation in terms of dihedral angles that matches the one implied in the definition of v.

Single Helix or Collapsed Bundles? What is the nature of the qualitative differences observed between the two helix-rich ensembles? The data in Figure 1 suggest that with increased backbone flexibility, the peptide is more likely to form collapsed bundles of multiple helical segments, whereas the single helix is the dominant state with rigidified backbone angles. Figure 2 shows distributions of the radii of gyration for either case at a few different temperatures. In the coil regime (360 K), comparison of panels A and B, shows that the two distributions are broad and very similar indicating that extended and disordered structures are populated in either case. In the helical regime (≤288 K), however, substantial differences are found. The distributions are generally multimodal with the peak at about 9.8 Å corresponding to the single, extended α-helix and the sharp peak at ~7.2 Å corresponding to the symmetric two-helix bundle ("helix—turn—helix"). These states and their sizes are perfectly consistent with the work of Zhang et al.,[28] who report 10.2 Å for the straight helix and 7.2 Å for a helix—turn—helix conformer (symmetric bundle) based on implicit solvent molecular dynamics simulations using an AMBER force field.

In the presence of just bond length constraints (panel A), the straight helix is never populated in dominant fashion, and bundles are more prominent. Its population appears to increase with temperature before melting occurs (above 300 K) presumably on account of the lessened drive to collapse. Conversely, with a rigidified backbone, the dominant helical state is the single helix. Here, the probability of observing partially collapsed states with radii of gyration of 7—8 Å seems to increase with increasing temperature when compared to the data at 250 K. If the temperature is dropped even further, a secondary transition sets in, in which the single helix collapses to form the two-helix bundle. This transition is also apparent in panels A and B of Figure 1. Complex coupling of coil-to-globule and helix—coil transitions has been observed for simplified models.[49−51] One may ask whether the artificially low temperatures coupled with explicit representation of counterions influence these results, but an analysis of both ion—ion and peptide—ion pair correlation functions indicates that ions remain largely inert with very little direct binding at all temperatures (see Figure S2, Supporting Information).

LR Fitting. Next, we show that it is possible to fit a LR model to the data for just $\langle N_s \rangle$ and $\langle N_h \rangle$ by using eq 4 if no limits are placed on the values v and w can assume. In Figure 3A and B, we show most of the same data as in Figure 1 as solid lines along with the fitted values (symbols). There are two fits, one to the data obtained from torsional inferences and the other to the data obtained from DSSP inference. Obviously the quality of the fit is arbitrarily good in either case suggesting that the two LR parameters are sufficiently independent. However, panel C shows that the resultant values for the LR parameters are inconsistent with the observed propensity to form isolated residues in helical conformation (see eq 5). $\langle N_1 \rangle$ appears to be consistently overestimated when using the fitted values for v and w, more so for the torsional case than for the DSSP estimates. This indicates that the obtained nucleation parameters are generally too large.

In panels A and B of Figure 4, we plot the actual values for v and w resulting from the aforementioned fitting, respectively. We find the conjecture that large nucleation parameters cause an overestimation of $\langle N_1 \rangle$ to be qualitatively confirmed. The nucleation parameter traces the temperature dependence of the propagation parameter irrespective of the constraint set employed or the data set fit to. At low temperatures, it assumes values that are indeed nonsensically large if one considers that the nucleation parameter should be related to the likelihood of visiting the α-region of ϕ/ψ-space in the absence of any hydrogen bonds. Conversely, in the coil region, the assumed values appear reasonable and close to the estimation of Thompson et al.[19] of ~0.127 (page 9208, $\sigma_{ZB} \sim 0.01$, and $\sigma_{ZB} = v^2/(v + 1)^4$). Interestingly,

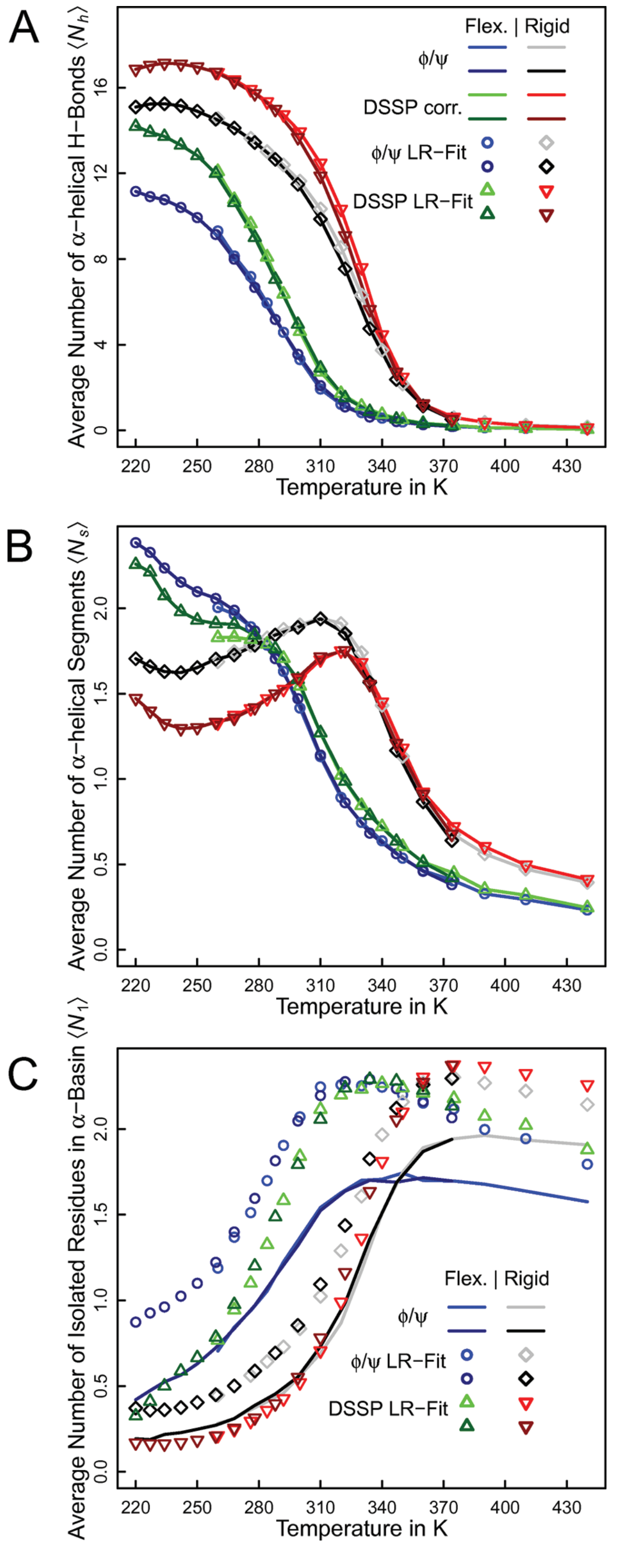

Figure 3. Quality of fits of LR theory to helical content data as a function of temperature for the FS-peptide using two different sets of holonomic constraints. The LR parameters obtained by these fits are plotted in Figure 4. To illustrate goodness of fits, solid lines are identical to those in Figure 1 and show the data measured directly from the simulations. Conversely, best-fitted values resulting from imposing the LR model are shown as symbols only (fits performed using eqs 4). Panel B uses the same legend as panel A.

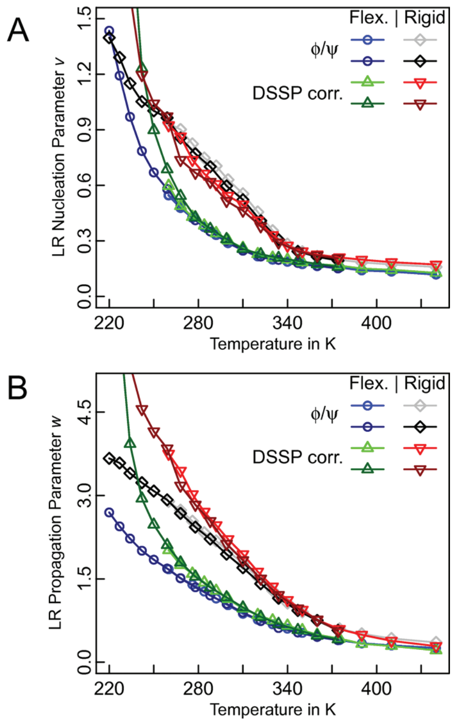

Figure 4. LR nucleation (A) and propagation (B) parameters as a function of temperature. Two types of fits are shown that use either DSSP-derived or torsional values for $\langle N_h \rangle$ and $\langle N_s \rangle$. The values shown give rise to the predictions shown as symbols in Figure 3 by using eq 4. The set of constraints enforced and the data set used are indicated in the legend similar to Figure 1. Note that low-temperature data for DSSP-based fits are cut off to allow visualization of all data in the same plot. They both continue to increase monotonously when further reducing the temperature.

given a set of constraints, the nucleation parameter seems to be mostly independent of the data set used ("DSSP corr." vs "ϕ/ψ") all the way down to temperatures of ~265 K, even though divergence of the fitted quantities occurs already at much higher temperatures in Figure 3A and B. Conversely, in Figure 4B, it is worth noting that the values of w and its dependency on temperature do appear to depend significantly on the data set used for fitting, suggesting that any enthalpy estimates using ln w will lack robustness (see below).

One may ask whether it is possible to fit to all three quantities ($\langle N_h \rangle$, $\langle N_s \rangle$, and $\langle N_1 \rangle$) with only one or two free variables. Figure S3, Supporting Information, shows that, when assuming a constant value for the nucleation parameter of 0.127, the quality of the fit drastically deteriorates. Essentially, it is impossible to predict correctly the values for $\langle N_s \rangle$ if only w is allowed to vary.[3] Even though the agreement for $\langle N_1 \rangle$ may be improved due to inclusion in the fitting procedure, it is overall very clear that LR predictions are unable to explain the data. As suggested by Figure 2, the largest discrepancy arises on account of the inability to represent the stabilization of helical bundles ($\langle N_s \rangle$ significantly larger than unity). An unconstrained fit in v/w-space masks this inapplicability by producing large values for v. This is almost certainly the reason why in silico data that match melting temperature and overall helicity well universally exhibit large values for v when analyzed with LR theory.[5,33,34] This inapplicability is masked of course if only data are analyzed that correspond to the transition and coil regimes but not to temperatures significantly below the observed melting temperature, or if the force field does in fact produce strictly LR-like results.[3,48]

Figure S4, Supporting Information, shows that the overall fit, as seen in Figure 3, can be improved when including $\langle N_1 \rangle$ as a fitted quantity. However, this may lead to a deterioration of fitting quality specifically for $\langle N_h \rangle$. Interestingly, both types of fits for either system now tend to agree more with the DSSP-derived hydrogen-bond counts. This is despite the fact that the values for $\langle N_1 \rangle$ are derived exclusively from torsional occupancies and indicates that the DSSP-derived statistics, which include torsional data for short segments (see Methods Section), may intrinsically be more consistent on account of the fact that they are much less prone to assign false breaks within long helices. In fact, overall fit quality is fairly good for the two DSSP-based fits. However, the values for the nucleation parameters remain large and exhibit even stronger dependencies on temperature. There are two ways to compensate the overestimation of single residues in α-conformation seen in Figure 3: making helices very stable (w large) or making the nucleation parameter so large that it is more likely to see two or more consecutive residues in helical conformation rather than one purely on account of v. Both paths are explored in Figure S4, Supporting Information; the former for DSSP and the latter for torsional statistics. This is an exacerbated demonstration of blind fitting yielding parameters that are impossible to interpret physically.

van't Hoff Analysis. The enthalpy change associated with the formation of a single hydrogen bond, ΔH_{hb}, is one of the parameters used most often to characterize the helix−coil transition experimentally. It is accessible from calorimetric experiments, and most recent estimates for alanine and alanine-like residues report a value of −0.9 kcal/mol[52] with earlier values being slightly larger (−1.3 kcal/mol).[53] For experiments that directly measure helix content (e.g., CD), it is common to extract ΔH_{hb} from a van't Hoff plot by assuming the following temperature dependence for ln w:[8,10,11,48,54]

$$-\beta \cdot \Delta G_{hb} = -\beta \cdot \Delta H_{hb} + \Delta S_{hb}/R = \ln k \approx \ln w \qquad (6)$$

Here, the subscript "hb" indicates that the process is interpreted to correspond to the addition of a single, α-helical hydrogen bond. We next critique the interpretation of ln w in eq 6 to arrive at a conclusion relevant to all LR-based analyses of helix−coil transition data.

If we consider a Schellman model[55] by assuming that only a single continuous helix is formed at any time and that no other residues, on average, reside in the helical basin at all, then the two-state equilibrium constant for the equilibrium between all-coil and all-helix states can be constructed as a product of stepwise constants:

$$K_{ch}^{cum} = \prod_{i=0}^{N_r-1} k_i = N_r v \cdot \frac{N_r - 1}{N_r} v \cdot \frac{N_r - 2}{N_r - 1} w \cdot \ldots = v^2 w^{N_r - 2}$$

$$\text{with } k_i = \frac{N_r - i}{N_r - i + 1} \cdot w \text{ and } K_i = (N_r - i) \cdot v^2 w^{i-1} \text{ for } i > 2 \qquad (7)$$

The statistical weight of a given sequence of N_r residues flanked by peptide bonds is the product of its residue weights, where the weight factor of a residue in the coil state, u, is set to 1 (normalization). Hence, a sequence "hhhcc" has total weight $v^2 w$ and sequence "hhhhc" has weight $v^2 w^2$. The combinatorial factors are simply related to the number of unique sequences that can accommodate a helical stretch of a given length. General

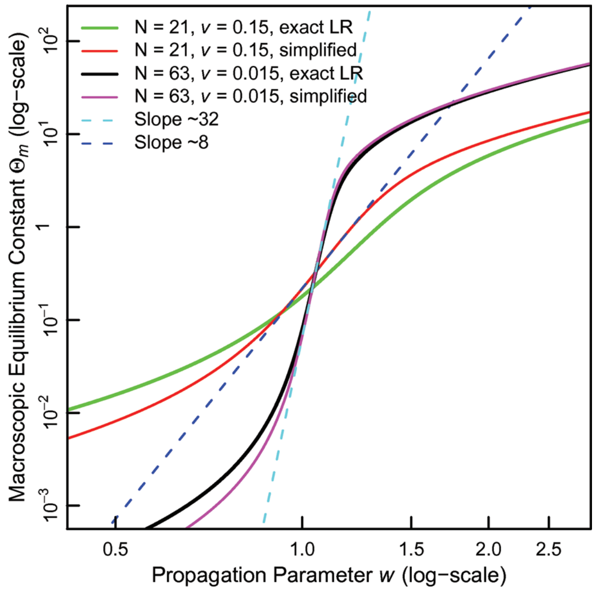

Figure 5. Comparison of the simplified single-sequence model (eq 7) to the exact LR formalism for two model systems. We show plots of Θ_m (see text) as a function of w when estimated using either the exact formula (eq 4) or the simplified version (eq 8). Agreement between the two depends on system parameters. Dashed lines indicate linear fits to the regions of maximal slope observed for the simplified model. The derivative can also be obtained analytically (see Supporting Information). Numerical tests show that the maximal slope does not approach $N_r - 2$ even when v is reduced by another 2 orders of magnitude for the case with $N_r = 63$.

forms for stepwise (k_i) and cumulative (K_i) equilibrium constants are provided in eq 7, the latter being referenced to the all-coil state. Clearly, the stepwise constants suggest the approximation in eq 6 to be applicable when considering an isolated growth step as long as the helix is nucleated and not yet close to its maximum length. The expected slope in a double logarithmic plot of K_{ch}^{cum} and w would indeed be $N_r - 2$ supporting the view that values obtained via eq 6 correspond to numbers per hydrogen bond. However, this equilibrium between the all-coil and all-helical states is monitored neither experimentally nor computationally; in both cases, ensemble averages are used to determine w. For $\langle N_h \rangle$, the simple model above yields

$$\langle N_h \rangle = \sum_{j=2}^{N_r - 1} \frac{K_j \cdot (j-1)}{Q} \text{ with } K_n = \prod_{i=0}^{n} k_i \text{ and}$$
$$Q = 1 + \sum_{i=0}^{N_r - 1} K_i \qquad (8)$$

We can thus construct a generalized equilibrium constant, Θ_m, for the helix−coil transition as $f_h/(1 - f_h)$, where $f_h = \langle N_h \rangle/(N_r - 2)$, i.e., the fractional helicity, and compare it in terms of its dependency on w to data extracted from exact application of LR theory (see eq 4). This is shown in Figure 5 for two cases: the first corresponds to a scenario where the single-sequence model above should be reasonably applicable (small v, larger N_r). Indeed, predictions from exact LR theory and from the simplified model agree very well. However, the relationship between the logarithms of Θ_m and w is complex. If we fit a line to the region of maximal variation (corresponding to states ranging from low to intermediate helicity), the resultant slope is only ~32, i.e., slightly more than half of the possible hydrogen bonds. It is therefore inaccurate to assume that application of eq 6 will yield values that can be interpreted as values per residue or per hydrogen bond (this would require a slope proportional to N_r). With parameters mimicking the system under study here, we find that the simple model becomes less applicable and that the slope for the full LR model is less than that found in the simplified model. Further numerical tests (see Figure S5, Supporting Information) clearly demonstrate that the maximum encountered slope has a nontrivial dependency on both N_r and v, that it is always larger in the simplified model, and that it will often lie close to to $(N_r - 2)/2$. This is an important point, as it means that results from fitting ln w in a van't Hoff-type plot should not be interpreted to be contributions per hydrogen bond.

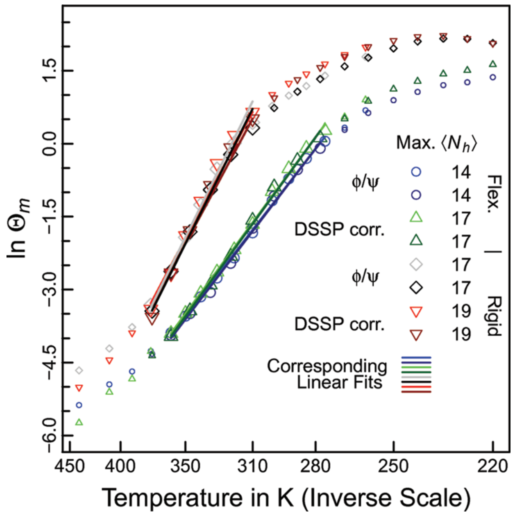

Figure 6. Van't Hoff determination of thermodynamic parameters of the helix−coil transition. Data are based on those in Figure 1A (see text). Linearity holds throughout the transition where both helix- and coil-rich states are populated to significant extent. With no angle constraints, the fitting region spanned from ~280−360 K, and with angle constraints we used ~310−375 K. These segments do in fact encompass the temperature regions exhibiting the largest change in Figure 1A. In the low-temperature region, secondary processes may prevent the van't Hoff assumption of temperature-independent enthalpy from being valid. The legend indicates the value assumed as the upper baseline for constructing f_h from $\langle N_h \rangle$ (see text). The obtained values are $\Delta S = -[34-35]$ cal·mol^{-1}·K^{-1} and $\Delta H = -[9.6-9.8]$ kcal·mol^{-1} for the case without bond angle constraints and $\Delta S = -[44-45]$ cal·mol^{-1}·K^{-1} and $\Delta H = -[13.8-14.5]$ kcal·mol^{-1} in the presence of angle constraints. The fits are no more dependent on the data set used than they are on the intrinsic accuracy of the data, which can be estimated by the differences obtained by independently fitting the low- and high-temperature REMD runs in each case. Lastly, values are not particularly sensitive to the definitions of upper baselines and temperature intervals. For example, the total variation is below 20% when including one additional temperature at each end or when changing the upper baseline from 14 to 19 for the case of flexible backbone and torsional data.

For simulation data, we therefore advocate to construct van't Hoff plots directly from measured equilibrium constants as described above, where the problem of identifying baselines is negligible. In vitro, van't Hoff fits of ln w usually require the definition of baselines implicitly that can often be determined with better accuracy using cosolute titrations. In Figure 6, van't Hoff plots of the values of Θ_m constructed from the data for $\langle N_h \rangle$ in Figure 1A are shown over temperature regimes where linearity holds. The lower baseline was always $\langle N_h \rangle = 0$, while the upper baselines we used are indicated in the legend. By this methodology, we obtain thermodynamic parameters for the entire process that are independent of whether DSSP or torsional statistics are used. The actual values agree well with literature estimates of $\Delta S = -36\ \text{cal}\cdot\text{mol}^{-1}\cdot\text{K}^{-1}$ and $\Delta H = -12\ \text{kcal}\cdot\text{mol}^{-1}$ [47] and $\Delta S = -51\ \text{cal}\cdot\text{mol}^{-1}\cdot\text{K}^{-1}$ and $\Delta H = -14.8\ \text{kcal}\cdot\text{mol}^{-1}$ that are obtained in similar fashion directly from spectroscopic data.[56] The agreement is congruent with the fact that the estimated melting temperatures from experiment (290−306 K) overlap with the interval defined by the apparent melting temperatures of the two simulated ensembles (see Figure 1A). The total enthalpy gives rise to estimated values for ΔH_{hb} of −0.5 and −0.75 kcal·mol^{-1} for flexible and rigidified backbones, respectively. These values are mutually consistent with the calorimetric estimate of −0.9 kcal·mol^{-1} that by definition has to be larger in magnitude given that it will include contributions from factors not related to hydrogen bonding (most prominently overall peptide swelling). They are also consistent with the values obtained for fits to ln *w*, which yield values between −1.0 and −1.3 kcal/mol experimentally,[11,18] if we consider that ΔH_{hb} in such a case should really correspond to the enthalpy associated with the formation of *more than one* hydrogen bond (see above). Of course, the agreement between the particular computational model in use and experimental data at the level of thermodynamics may be fortuitous. It is noteworthy that the force field in use here implies discarding most of the dihedral angle potential parameters[34] that continue to be optimized elsewhere.[3,5,57,58] Crucially, however, neither LR fits nor van't Hoff plots resolve potential discrepancies in mechanisms or dynamics of the helix−coil transition that could, for example, arise on account of the continuum solvation model lacking an appropriate description of water−peptide interfaces regarding wetting behavior, reorientation dynamics, etc.[59] It would therefore be ill-advised to arrive at conclusions on relative virtues of different computational models purely based on analyses like the ones presented here.

Modeling of Equilibrium between Single Helix and Multihelix Bundles. Lastly, is there a simple way to improve the original LR model, which specifically addresses issues identified here? For conceptual illustration, we test here a nongeneralizable modification to the fitting procedure that leaves the LR framework intact at the expense of an additional parameter. We focus on the statistics derived from torsional segments only since DSSP statistics need to be augmented by data on short segments derived from ϕ/ψ-values.

Following some of the ideas in the work of Ghosh and Dill,[60] we consider the system to be in equilibrium between a three-helix "bundle" and a single helix. Then, we may approximately treat the three-helix bundle as three independent sequences of one-third the length of the original peptide:

$$\langle N_h \rangle = 3f_3 \cdot \frac{\partial \ln Z_{N_r=7}}{\partial \ln w} + (1 - f_3) \cdot \frac{\partial \ln Z_{N_r=21}}{\partial \ln w} \qquad (9)$$

The averages $\langle N_1 \rangle$ and $\langle N_s \rangle$ are computed analogously (see eqs 4). The new parameter f_3 is simply the fractional occupancy of the three-helix bundle and setting it to zero recovers the original fitting functions as used in Figures 3 and 4. How are nonzero values of f_3 interpretable? Essentially, we stipulate that there are reasons external to LR theory that "stabilize" helix interruptions. In a thermodynamic sense, these can be tertiary interactions stabilizing compact bundles. However, in a statistical sense, they can also be errors in the counting of helical segments and their lengths. Evidence for both was presented above. Using eq 9 and treating f_3 as a free parameter, we obtain the fits and data in Figures S6, Supporting Information, and 7, respectively.

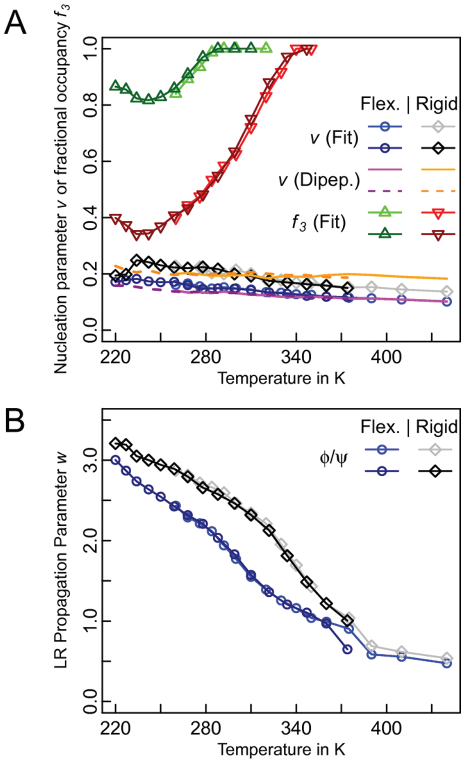

Figure 7. Fitted values *ν*, *w*, and f_3 as a function of temperature when employing eq 9. Only fits to the data based on torsional inference are considered. Panel A shows the values for the nucleation parameter and f_3 in the same plot. Only those values for f_3 are shown that proved reproducible through multiple independent Monte Carlo fits (see main text and Methods Section). Along with the fitted values for *ν*, we show the ratio of probabilities of occupying the helix and coil regions determined from simulations of alanine dipeptide, i.e., $p_h/p_c(T)$. The definition of the helical region was identical to the one used throughout to analyze data for the FS-peptide. Panel B shows the values for *w*. As in other figures, darker colors correspond to the low-temperature REMD run.

The first thing to note in Figure S6, Supporting Information, is that the overall fit quality is significantly improved over that shown in Figures S4, Supporting Information, which is of course expected due to the inclusion of an additional parameter. Nonetheless, f_3 is not able to explain all of the data consistently, as minor deviations are observed in the fitted values for $\langle N_h \rangle$ for the case with flexible bond angles. In Figure 7A, we show the obtained values for *ν* and f_3. Consistent with physical intuition, the nucleation parameter now assumes values in the interval from 0.1 to 0.25 and, for both systems, exhibits a very weak temperature dependence. This is despite the fact that no constraints were placed on the values *ν* can assume during the fitting. We show that these values are reasonable for the computational model in use by explicitly computing the ratio of weights of the helical vs coil regions for alanine dipeptide as a function of temperature. As can be gleaned from Figure 7A, the agreement is profound. Both the sign of the temperature derivative and the differences between flexible and rigidified backbones are mirrored in the dipeptide data. We can also infer that differences in local backbone conformational properties may in fact be able to explain the observed shift in the melting transition as was hypothesized above.

The values for f_3 are not particularly informative in the coil region since large differences have little impact on fit quality if long helical segments are generally unlikely to form. This means that the fits become ill-defined (not substantially dependent on f_3), and we omit those data points in Figure 7A. The model apparently suggests that the data are well-described by the three-helix bundle, in particular for the case of flexible bond angles. This is qualitatively consistent with Figure 2, in which the height of the peak at ~10 Å (single helix) strongly depends on the constraint set in use. The temperature dependence at low temperatures is consistent with Figure 2 as well in that bundled conformations are least likely at an intermediate temperature within the strongly helical region. In that sense, f_3 is physically interpretable. However, we wish to remind the reader that these fits are to quantities inferred from torsional statistics that are inherently prone to produce false negatives (see Methods Section and above). This may help to explain why in general the values for f_3 are large. Fitting this parameter may therefore simply represent a way to silently correct such faulty assignments. Unfortunately, the two effects are not easily deconvoluted. Along those lines, it may be interesting to ask whether a generalization of the model in eq 9 to arbitrary subsegment length distributions could produce even better results. The problem here is the limited data available for fitting a larger number of parameters. Figure S7, Supporting Information, shows a variant of eq 9, that can be fit unambiguously, producing inferior results. Lastly, it may be tempting to try to transform the data in Figure 2 into a direct and independent estimate for f_3 or related parameters, but such an effort would require the definition of a fair number of ad hoc structural criteria for clustering data.

■ CONCLUSIONS

This contribution makes a number of points that can be grouped into two categories. The first four all deal with the application of LR models to molecular simulation data and also with comparisons between in silico and in vitro results. Conclusions are as follows:

(1) Estimates of the LR nucleation and propagation parameters are not directly comparable to those extracted from experimental data if the processes for obtaining those are different (Figures 3 and 4 and S3 and S4, Supporting Information). For example, it is invalid to perform an unconstrained fit to $\langle N_h \rangle$ and $\langle N_s \rangle$, as in Figures 3 and 4 for a single chain length, and compare it to estimates such as those by Rohl and Baldwin[18] or Thompson et al.[19] that use a fundamentally different construct of assumptions. Moreover, values for v and w that agree with experiment at a specific temperature may mask inaccuracies, and we recommend reporting melting temperatures and van't Hoff enthalpies instead (Figures 1 and 6).

(2) Two checks are recommended: (i) mutual consistency of eligible helix−coil descriptors ($\langle N_h \rangle$ and $\langle N_s \rangle$) between torsional and DSSP inference and (ii) use of $\langle N_1 \rangle$ as either a weakly dependent test or an additional quantity to fit to (Figures 3 and S3, S4, and S6, Supporting Information). The robustness of estimation in particular of $\langle N_s \rangle$ will depend on the nature of the force field, and smaller deviations than those reported here may be found if the polypeptide backbone exhibits a larger amount of preorganization.[34]

(3) We show that it is misleading to interpret data from van't Hoff fits of ln w as quantities per residue or per hydrogen bond (Figures 5 and S5, Supporting Information). Of course, for similar procedures and identical systems, values obtained in such a way are still comparable to one another, but their physical meaning is not immediately obvious to us. In contrast, direct van't Hoff analyses of a generalized equilibrium constant, such as Θ_m, yield robust results that in this case also agree well with both IR and calorimetric estimates (Figure 6).[25,47,52,56]

(4) Lastly, we demonstrate that simple models can be found that preserve physical interpretability of fitted helix−coil parameters (Figure 7). It would be desirable to have a generalized framework for analyzing in silico data that satisfies the criteria spelled out above. One approach could be the ascending levels model of Lucas et al.[54] The problem thus far is that it is not routinely feasible to simulate reversible helix formation for many different peptides of differing lengths under a wide variety of conditions. Consequently, inconsistencies in the analysis are easily masked, and conclusions may be misleading.

The fifth and last point is more technical in nature:

(5) Bond angle constraints alter the free energy landscape substantially and give rise to quantitatively and qualitatively different ensembles (Figures 1 and 2). As noted,[35] force field reparametrization will often be required to add (or release) such constraints. Therefore, they should not be viewed as independent entities controlling computational efficiency only.[61] In contrast to backbone bond angle constraints, we did not observe strong changes of the kind seen in Figures 1 and 2 upon introduction of just bond length constraints (data not shown).

In summary, we suggest guidelines and checks for applying LR or similar theories to data obtained from atomistic simulations of helix-forming polypeptides. Ultimately, LR models may well be inapplicable to such data, and there is a clear need for a unified framework.[60,62] We also believe that this work helps to reconcile some of the discrepancies in interpreting helix−coil transition data using the LR or similar formalisms, for example, when comparing in vitro to in silico data and also when comparing different sets of in vitro data to each other.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information. Illustration of the LR model using a simple example. Methods describing the force field and solvation model in more detail. Methods and plots (S1) on control simulations using Monte Carlo sampling. Additional plots showing ion pair correlation functions (S2), and LR fits and fitted parameters using different assumptions (S3−S4). Analytical derivations and numerical exploration of the model defined by eqs 7 and 8 (S5). Details on the interpretation of the partition function underlying eq 9. Quality of fits associated with Figure 7 (S6) and exploration of a related model (S7). This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: a.vitalis@bioc.uzh.ch. Telephone: +41446355597.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENT

We thank Dr. Rohit V. Pappu for critical comments regarding this work. This work was supported in part by a grant from the Swiss National Science Foundation to A.C. and a personal grant by the UZH Forschungskredit to A.V. Most of the simulations were carried out on the Schrödinger supercomputer administered by the IT services of the University of Zurich.

■ REFERENCES

(1) Scholtz, J. M.; Baldwin, R. L. *Annu. Rev. Biophys. Biomol. Struct.* **1992**, *21*, 95–118.

(2) Makhatadze, G. I. *Adv. Protein Chem.* **2006**, *72*, 199–226.

(3) Best, R. B.; Hummer, G. *J. Phys. Chem. B* **2009**, *113*, 9004–9015.

(4) Song, K.; Stewart, J. M.; Fesinmeyer, R. M.; Andersen, N. H.; Simmerling, C. *Biopolymers* **2008**, *89*, 747–760.

(5) Sorin, E. J.; Pande, V. S. *Biophys. J.* **2005**, *88*, 2472–2493.

(6) Garcia, A. E.; Sanbonmatsu, K. Y. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *99*, 2782–2787.

(7) Ferrara, P.; Apostolakis, J.; Caflisch, A. *J. Phys. Chem. B* **2000**, *104*, 5000–5010.

(8) Zimm, B. H.; Bragg, J. K. *J. Chem. Phys.* **1959**, *31*, 526–535.

(9) Gibbs, J. H.; DiMarzio, E. A. *J. Chem. Phys.* **1959**, *30*, 271–282.

(10) Lifson, S.; Roig, A. *J. Chem. Phys.* **1961**, *34*, 1963–1974.

(11) Scholtz, J. M.; Hong, Q.; York, E. J.; Stewart, J. M.; Baldwin, R. L. *Biopolymers* **1991**, *31*, 1463–1470.

(12) Bixon, M.; Scheraga, H. A.; Lifson, S. *Biopolymers* **1963**, *1*, 419–423.

(13) Bixon, M.; Lifson, S. *Biopolymers* **1967**, *5*, 509–514.

(14) Doig, A. J.; Chakrabartty, A.; Klingler, T. M.; Baldwin, R. L. *Biochemistry* **1994**, *33*, 3396–3403.

(15) Shalongo, W.; Stellwagen, E. *Protein Sci.* **1995**, *4*, 1161–1166.

(16) Kemp, D. S. *Helv. Chim. Acta* **2002**, *85*, 4392–4423.

(17) Rohl, C. A.; Scholtz, J. M.; York, E. J.; Stewart, J. M.; Baldwin, R. L. *Biochemistry* **1992**, *31*, 1263–1269.

(18) Rohl, C. A.; Baldwin, R. L. *Biochemistry* **1997**, *36*, 8435–8442.

(19) Thompson, P. A.; Eaton, W. A.; Hofrichter, J. *Biochemistry* **1997**, *36*, 9200–9210.

(20) Lockhart, D. J.; Kim, P. S. *Science* **1992**, *257*, 947–951.

(21) Bierzynski, A.; Pawlowski, K. *Acta. Biochim. Pol.* **1997**, *44*, 423–432.

(22) Jacobs, D. J.; Wood, G. G. *Biopolymers* **2011**, *95*, 240–253.

(23) Pappu, R. V.; Srinivasan, R.; Rose, G. D. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 12565–12570.

(24) Shalongo, W.; Dugad, L. B.; Stellwagen, E. *J. Am. Chem. Soc.* **1994**, *116*, 2500–2507.

(25) Taylor, J. W.; Greenfield, N. J.; Wu, B.; Privalov, P. L. *J. Mol. Biol.* **1999**, *291*, 965–976.

(26) Ghosh, T.; Garde, S.; Garcia, A. E. *Biophys. J.* **2003**, *85*, 3187–3193.

(27) Zagrovic, B.; Jayachandran, G.; Millett, I. S.; Doniach, S.; Pande, V. S. *J. Mol. Biol.* **2005**, *353*, 232–241.

(28) Zhang, W.; Lei, H.; Chowdhury, S.; Duan, Y. *J. Phys. Chem. B* **2004**, *108*, 7479–7489.

(29) Kennedy, R. J.; Tsang, K. W.; Kemp, D. S. *J. Am. Chem. Soc.* **2002**, *124*, 934–944.

(30) Miller, J. S.; Kennedy, R. J.; Kemp, D. S. *J. Am. Chem. Soc.* **2002**, *124*, 945–962.

(31) Wang, T.; Zhu, Y. J.; Getahun, Z.; Du, D. G.; Huang, C. Y.; DeGrado, W. F.; Gai, F. *J. Phys. Chem. B* **2004**, *108*, 15301–15310.

(32) Rose, A.; Schraegle, S. J.; Stahlberg, E. J.; Meier, I. *BMC Evol. Biol.* **2005**, *5*, 66.

(33) Nymeyer, H.; Garcia, A. E. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 13934–13939.

(34) Vitalis, A.; Pappu, R. V. *J. Comput. Chem.* **2009**, *30*, 673–699.

(35) Chen, J.; Im, W.; Brooks, C. L., III *J. Comput. Chem.* **2005**, *26*, 1565–1578.

(36) Lazaridis, T.; Karplus, M. *Proteins: Struct., Funct., Bioinf.* **1999**, *35*, 133–152.

(37) Kaminski, G. A.; Friesner, R. A.; Tirado-Rives, J.; Jorgensen, W. L. *J. Phys. Chem. B* **2001**, *105*, 6474–6487.

(38) Skeel, R. D.; Izaguirre, J. A. *Mol. Phys.* **2002**, *100*, 3885–3891.

(39) Sugita, Y.; Okamoto, Y. *Chem. Phys. Lett.* **1999**, *314*, 141–151.

(40) Vitalis, A.; Steffen, A.; Lyle, N.; Mao, A. H.; Pappu, R. V. *CAMPARI*, v1.0; SourceForge/Geeknet, Inc.: Mountain View, CA, 2010; http://sourceforge.net/projects/campari (accessed December 13, 2011).

(41) Ryckaert, J. P.; Cicotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327–341.

(42) Fixman, M. *Proc. Natl. Acad. Sci. U.S.A.* **1974**, *71*, 3050–3053.

(43) Perchak, D.; Skolnick, J.; Yaris, R. *Macromolecules* **1985**, *18*, 519–525.

(44) Patriciu, A.; Chirikjian, G. S.; Pappu, R. V. *J. Chem. Phys.* **2004**, *121*, 12708–12720.

(45) Mülders, T.; Swegat, W. *Mol. Phys.* **1998**, *94*, 395–399.

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

(47) Williams, S.; Causgrove, T. P.; Gilmanshin, R.; Fang, K. S.; Callender, R. H.; Woodruff, W. H.; Dyer, R. B. *Biochemistry* **1996**, *35*, 691–697.

(48) Gnanakaran, S.; Garcia, A. E. *Proteins: Struct., Funct., Bioinf.* **2005**, *59*, 773–782.

(49) Sikorski, A.; Romiszowski, P. *Biopolymers* **2003**, *69*, 391–398.

(50) Varshney, V.; Carri, G. A. *Phys. Rev. Lett.* **2005**, *95*, 168304.

(51) Nowak, C.; Rostiashvili, V. G.; Vilgis, T. A. *J. Chem. Phys.* **2007**, *126*, 34902.

(52) Lopez, M. M.; Chin, D. H.; Baldwin, R. L.; Makhatadze, G. I. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 1298–1302.

(53) Scholtz, J. M.; Marqusee, S.; Baldwin, R. L.; York, E. J.; Stewart, J. M.; Santoro, M.; Bolen, D. W. *Proc. Natl. Acad. Sci. U.S.A.* **1991**, *88*, 2854–2858.

(54) Lucas, A.; Huang, L.; Joshi, A.; Dill, K. A. *J. Am. Chem. Soc.* **2007**, *129*, 4272–4281.

(55) Schellman, J. A. *J. Phys. Chem.* **1958**, *62*, 1485–1494.

(56) Ianoul, A.; Mikhonin, A.; Lednev, I. K.; Asher, S. A. *J. Phys. Chem. A* **2002**, *106*, 3621–3624.

(57) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. *Proteins: Struct., Funct., Bioinf.* **2006**, *65*, 712–725.

(58) Mackerell, A. D., Jr.; Feig, M.; Brooks, C. L., III *J. Comput. Chem.* **2004**, *25*, 1400–1415.

(59) Laage, D.; Stirnemann, G.; Sterpone, F.; Rey, R.; Hynes, J. T. *Annu. Rev. Phys. Chem.* **2011**, *62*, 395–416.

(60) Ghosh, K.; Dill, K. A. *J. Am. Chem. Soc.* **2009**, *131*, 2306–2312.

(61) Feenstra, K. A.; Hess, B.; Berendsen, H. J. C. *J. Comput. Chem.* **1999**, *20*, 786–798.

(62) Jacobs, D. J.; Dallakyan, S.; Wood, G. G.; Heckathorne, A. *Phys. Rev. E* **2003**, *68*, 061109.

(63) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graphics* **1996**, *14*, 33–38.