

## Hydrophobic interactions in the formation of secondary structures in small peptides

Cristiano L. Dias,<sup>1,2,\*</sup> Mikko Karttunen,<sup>3,4,†</sup> and Hue Sun Chan<sup>1,2,‡</sup>

<sup>1</sup>*Department of Biochemistry and Department of Molecular Genetics, University of Toronto, Toronto, Ontario, Canada M5S 1A8*

<sup>2</sup>*Department of Physics, University of Toronto, Toronto, Ontario, Canada M5S 1A7*

<sup>3</sup>*Department of Applied Mathematics, The University of Western Ontario, London, Ontario, Canada N6A 5B7*

<sup>4</sup>*Department of Chemistry, University of Waterloo, 200 University Avenue West, Waterloo, Ontario, Canada N2L 3G1*

(Received 13 July 2011; published 25 October 2011)

Effects of the attractive and repulsive parts of hydrophobic interactions on  $\alpha$  helices and  $\beta$  sheets in small peptides are investigated using a simple atomic potential. Typically, a physical spatial range of attraction tends to favor  $\beta$  sheets, but  $\alpha$  helices would be favored if the attractive range were more extended. We also found that desolvation barriers favor  $\beta$  sheets in collapsed conformations of polyalanine, polyvaline, polyleucine, and three fragments of amyloid peptides tested in this study. Our results provide insight into the multifaceted role of hydrophobicity in secondary structure formation, including the  $\alpha$  to  $\beta$  transitions in certain amyloid peptides.

DOI: [10.1103/PhysRevE.84.041931](https://doi.org/10.1103/PhysRevE.84.041931)

PACS number(s): 87.15.Cc, 87.14.ef, 87.15.bd

### I. INTRODUCTION

Hydrophobic interactions are important in protein folding, denaturation [1–4], and packing. They are a major driving force for protein tertiary structure [5–7]; they can lead to non-native interactions during folding [8]; and they can affect the formation of  $\alpha$  helices and  $\beta$  sheets [9] in proteins through packing [10] and local sterics [11]. The importance of hydrophobic interactions in the formation of secondary structures in globular proteins is highlighted by an experimental study of the mutations at the helical segments of *T4* lysozyme. The unfolding free energy of the mutated sequences was found to correlate with the surface area of the nine mutated amino acids, and the slope of this dependence was close to the one expected from hydrophobic stabilization (between 20 and 30 kcal mol<sup>-1</sup> Å<sup>-2</sup>) [12]. In the case of the arc repressor homodimer, interchanging the positions of a hydrophobic and a polar residue turns a two-strand  $\beta$  sheet into a pair of helices [13]. Such hydrophobicity-related conformational switches [14] have important ramifications for molecular evolution [15]. Indeed, in *de novo* protein design, an important step is to choose a hydrophobic-polar pattern that packs nonpolar residues in secondary structures well within a core [16]. For example, to design  $\alpha$  helices it is common to insert nonpolar residues at every three or four positions in a sequence. This accounts for a configuration where nonpolar sidechains point toward the same face of the helix which can then be stabilized through contacts within the nonpolar core of the protein [17,18]. These results are corroborated by computational studies [19–21]. Recently, simulations of a coarse grained model suggested that the structure of small globular proteins is determined predominantly by the hydrophobic-polar pattern regardless of the intrinsic secondary structure propensity of their amino acids [22].

However, even for small peptides that are not part of a hydrophobic core, secondary structures are sensitive to the effects of the hydrating medium. For example, fragments of

amyloid peptides with many hydrophobic residues form preferentially  $\beta$  or coil conformations in aqueous environments; but they form  $\alpha$ -helical structures in membrane-mimicking environments [23,24]. This indicates that, aside from protein core packing, hydrophobicity plays an important role in the conformation of small peptides. Understanding this role is useful, for example, to account for the emergence of enhanced  $\beta$  structures in proteins implicated in amyloid and prion diseases [25–27].

However, the effects of hydrophobic interactions on secondary structure preference in these situations are less clear. A key role for hydrophobicity in secondary structure formation was suggested by atomic simulations of the  $\alpha$  helix to  $\beta$  sheet transition of a short polyleucine peptide that exhibited non-Arrhenius behaviors [28] characteristic of the hydrophobic effect [29]. An early simulation of a 12mer polyalanine indicated that its ground state was an  $\alpha$  helix in vacuum and a  $\beta$  hairpin in aqueous solution [30]. However, subsequent simulations of aqueous short hydrophobic-polar [31] and polyalanine [32] peptides using different potential functions predicted an  $\alpha$ -helical ground state except for very high hydrophobic interaction strengths [32].

In this paper, we show that the spatial range of hydrophobic attraction determines whether the interaction favors  $\beta$  sheets or  $\alpha$  helices in a short polyalanine peptide. This finding offers a possible resolution for the divergent computational results noted above. Based on explicit-water simulations of a pair of methanes in water, we argue that a realistic range is  $\sim 5.0$  Å and show that this favors  $\beta$  sheets. We also considered desolvation barrier (db) effects as they are critical in protein energetics [33–43]. db effects were neglected in many implicit-water potential functions [31,32] despite studies showing its importance for protein structure prediction [44]. For polyalanine, we found that sidechains subjected to hydrophobic interactions in  $\alpha$  helices are constrained to the db region and thus are disfavored [45,46]. In contrast, alanine sidechains in  $\beta$  sheets are able to avoid db configurations and maintain favorable contacts. As a result,  $\beta$  sheets are favored over  $\alpha$  helices by both the attractive and repulsive parts of the hydrophobic interaction. We found further that this trend applies also to polyvaline, polyleucine, and three

\*diasc@biosimulations.org

†mkarttu@uwo.ca

‡chan@arrhenius.med.toronto.edu

fragments of amyloid proteins, as detailed below. Based on these findings, we explore their implications for the structure of amyloid peptides and proteins under denaturing conditions in the discussion below.

## II. MODEL

We used the model of Irback and Mohanty that has rationalized the folding thermodynamics of several natural and designed peptides [47]. This model treats hydrogen atoms explicitly. The model was chosen for its simplicity and physical transparency—it has no torsional terms, thus it avoids the numerical and physical uncertainties related to them. Bond lengths and angles are fixed at experimental values. The degrees of freedom are the  $\phi, \psi$  backbone dihedral angles and sidechain torsional angles. The potential energy of a conformation is given by

$$E = E_{\text{EV}} + E_{\text{loc}} + E_{\text{HB}} + E_{\text{HP}}, \quad (1)$$

where  $E_{\text{EV}}$  is the excluded-volume term,  $E_{\text{loc}}$  accounts for local electrostatics among backbone atoms,  $E_{\text{HB}}$  is a direction-dependent energy for hydrogen bonds between -NH and -CO groups, and  $E_{\text{HP}}$  is the effective (implicit-water) hydrophobic energy. Further details are provided in Ref. [47].

In the analysis below, we consider the original  $E_{\text{HP}}$  term as well as a modified form of  $E_{\text{HP}}$  with a db. The  $E_{\text{EV}}$ ,  $E_{\text{loc}}$ , and  $E_{\text{HB}}$  in our simulation are identical to those in Ref. [47]. The original  $E_{\text{HP}}$  is given by [48]

$$E_{\text{HP}} = \sum_{I < J} \frac{M_{IJ}}{N_I + N_J} \left[ \sum_{i \in A_I} f\left(\min_{j \in A_J} r_{ij}^2\right) + \sum_{j \in A_J} f\left(\min_{i \in A_I} r_{ij}^2\right) \right]. \quad (2)$$

Here  $I, J$  label the amino acids along the peptide sequence,  $M_{IJ}$  is an energy scale that depends on the amino acid types, and  $A_I$  and  $A_J$  are the sets of  $N_I$  and  $N_J$  sidechain heavy (nonhydrogen) atoms (labeled  $i$  and  $j$ ), respectively. Since the hydrophobic interaction is expected to be weaker for sidechains that are close in sequence,  $M_{I, I+1} = 0$  and  $M_{I, I+2}$  is reduced by a factor of 2. The formulation in Ref. [47] is equivalent to setting

$$f(x) = \begin{cases} -1 & \text{if } x \equiv r_{ij}^2 \leq A, \\ (x - R_{\text{cutoff}}^2)/(R_{\text{cutoff}}^2 - A) & \text{if } A < x < R_{\text{cutoff}}^2, \\ 0 & \text{if } x \geq R_{\text{cutoff}}^2, \end{cases} \quad (3)$$

where  $A = (3.5 \text{ \AA})^2$  and  $R_{\text{cutoff}} = 4.5 \text{ \AA}$ .

To test the robustness of our conclusions, we considered also an alternative potential,

$$E_{\text{HP}} = \sum_{I < J} \mathcal{M}_{IJ} \sum_{i \in A_I} \sum_{j \in A_J} f(r_{ij}^2), \quad (4)$$

which includes not only the minimal distances in Eq. (2) but all pairwise interactions between heavy atoms in different residues. Note that Eqs. (2) and (4) are different for sidechains with more than one heavy atom; but they are equivalent for polyalanine (with only one heavy  $C_\beta$  atom per sidechain) provided that  $M_{IJ} = \mathcal{M}_{IJ}$ .

In this work, we first focus on the 17mer polyalanine peptide Ace-ALA<sub>17</sub>-NH<sub>2</sub> before extending our analysis to other polypeptides. Parallel tempering was performed with the software SMMP [49] for the Irback and Mohanty model using eight temperatures between  $T = 275$  and  $362.5$  K. Statistics were gathered for  $10^7$  Monte Carlo sweeps.

## III. DEFINING SECONDARY STRUCTURES

We have used two methods to assign secondary structures to protein conformation to minimize possible biases by any one particular definition, so as to provide more robustness to our conclusions. The first method uses the ranges of dihedral angles (deg) for  $\alpha$  ( $-90 < \phi < -30, -77 < \psi < -17$ ) and for  $\beta$  ( $-150 < \phi < -90, 90 < \psi < 150$ ), as used by Irback and Mohanty [47,50]. The second method is based on the Define Secondary Structure of Proteins (DSSP) algorithm of Kabsch and Sander [51] in which residues are classified as  $\alpha$  or  $\beta$  depending on the main-chain amide-carbonyl H-bond pattern around them. Each of these patterns involves at least two ( $\alpha$ ) or four ( $\beta$ ) main-chain H bonds that define a smallest unit of the secondary structure. For an  $\alpha$  helix, main-chain H bonds are required between residues  $(I, I+4)$  and  $(I+1, I+5)$  in order for residues  $I+1, I+2, I+3$ , and  $I+4$  to belong to a helix. We follow the definition for which the first residue in the H-bond bracket, i.e.,  $I$  in  $(I, J)$  is the residue number of the H-bond acceptor CO group and  $J$  is the residue number of the donor NH group. For  $\beta$  sheets, at least two consecutive residues of the same type (parallel or antiparallel) within each strands are required. This condition serves to exclude ill-formed sheets. Two residues  $I$  and  $J$  situated on two different strands ( $|I - J| \geq 3$ ) are considered to be part of a parallel  $\beta$  sheet if a pair of H bonds exists between residues  $(I-1, J)$  and  $(J, I+1)$  or between residues  $(J-1, I)$  and  $(I, J+1)$ . They are considered to be part of an antiparallel  $\beta$  sheet if a pair of H bonds exists between residues  $(I, J)$  or between residues  $(I-1, J+1)$  and  $(J-1, I+1)$ . Here we use the H-bond energy in the DSSP algorithm as well as the cutoff energy it provides to define a H bond. As in the original DSSP formulation, when a residue is assigned to an  $\alpha$  helix and a  $\beta$  sheet simultaneously, the  $\alpha$  assignment is given priority [51].

In Fig. 1 we show the correlation between the number of  $\alpha$  amino acids in protein conformations computed using DSSP ( $n_{\text{DSSP}}$ ) versus using dihedral angles ( $n_{\text{dihedral}}$ ). The ensemble of conformations used to compute this figure is our baseline situation with no sidechain hydrophobic interaction for polyalanine. Because the minimum  $\alpha$ -helix length in the DSSP calculation is 4, there are no data for  $n_{\text{DSSP}} = 1, 2$ , and 3. On the margins of Fig. 1, we show characteristic structures for three sets of  $(n_{\text{DSSP}}, n_{\text{dihedral}})$  values. When  $n_{\text{dihedral}}$  is large ( $n_{\text{dihedral}} \gtrsim 10$ ), the structures resemble standard  $\alpha$  helices with their backbone wrapping around a main axis (top margin of Fig. 1). However, when  $n_{\text{dihedral}}$  is large and  $n_{\text{DSSP}}$  is small ( $n_{\text{DSSP}} \lesssim 10$ ), the structure lacks some of the H-bond connectivity between residues  $I$  and  $I+4$  required for  $\alpha$  helices (top left margin). At the bottom margin of Fig. 1, we show a structure that has large  $n_{\text{DSSP}}$  ( $n_{\text{DSSP}} \gtrsim 10$ ) but small  $n_{\text{dihedral}}$  ( $n_{\text{dihedral}} \lesssim 10$ ). This conformation shows a large amount of twists and its backbone does not wrap around a

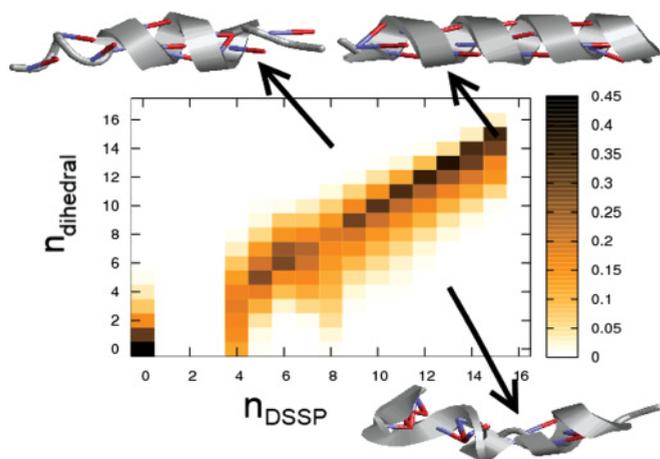


FIG. 1. (Color online) Correlation between the number of  $\alpha$  amino acid residues computed using the DSSP ( $n_{\text{DSSP}}$ ) and dihedral-angle ( $n_{\text{dihedral}}$ ) definitions. The shading scale provides the probability of structures characterized by both  $n_{\text{DSSP}}$  and  $n_{\text{dihedral}}$ , wherein the probabilities for values of  $n_{\text{dihedral}}$  for a given  $n_{\text{DSSP}}$  are normalized to that  $n_{\text{DSSP}}$ . Three representative structures (ribbon diagrams) are shown for three values of  $(n_{\text{DSSP}}, n_{\text{dihedral}})$  with sticks representing main-chain H bonds.

main axis. Overall, the distribution in Fig. 1 indicates that the concurrent presence of  $\alpha$  H-bond patterns and  $\alpha$  dihedral angles is a good indicator of  $\alpha$  helices (top right margin).

In Fig. 2 we show the correlation between the amount of  $\beta$  amino acids computed using the DSSP definition ( $n_{\text{DSSP}}$ ) and that using the range of dihedral angles ( $n_{\text{dihedral}}$ ). The ensemble of conformations used to compute this figure is the same as that in Fig. 1. Representative structures for three  $(n_{\text{DSSP}}, n_{\text{dihedral}})$  values are shown in a stick representation for backbone atoms. When  $n_{\text{dihedral}}$  is large ( $n_{\text{dihedral}} \gtrsim 8$ ), the backbone has the zigzaglike regularity characteristic of an ideal  $\beta$  geometry, as shown at the top right margin. However, when  $n_{\text{dihedral}}$  is large

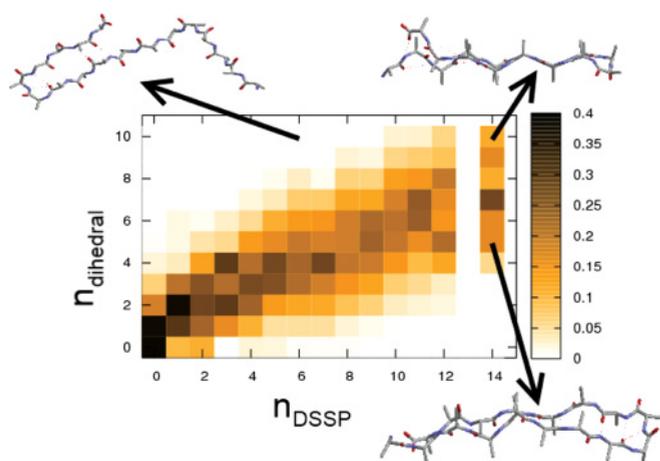


FIG. 2. (Color online) Correlation between the numbers of  $\beta$  amino acid residues computed using the DSSP definition ( $n_{\text{DSSP}}$ ) and using dihedral angles ( $n_{\text{dihedral}}$ ). The shading-coded probabilities are normalized as in Fig. 1. Three representative structures are shown for three sets of  $(n_{\text{DSSP}}, n_{\text{dihedral}})$  values. In the conformations shown, red, blue, and gray sticks denote oxygen, nitrogen, and carbon atoms, respectively.

but  $n_{\text{DSSP}}$  is small ( $n_{\text{DSSP}} \lesssim 10$ ), strands are not guaranteed to be H bonded to each other and some conformations can have isolated strands (structure at the top left margin). At the bottom margin we show a structure with a large  $n_{\text{DSSP}}$  ( $n_{\text{DSSP}} \gtrsim 10$ ) and small  $n_{\text{dihedral}}$  ( $n_{\text{dihedral}} \lesssim 8$ ). While a large  $n_{\text{DSSP}}$  leaves little freedom for isolated strands, a small  $n_{\text{dihedral}}$  does not enforce the geometrical zigzaglike nature of  $\beta$  sheets. As a result, these structures resemble distorted hairpinlike conformations.

Thus, the main limitation of using dihedral angles to identify secondary structures occurs for  $\beta$  sheets as this parameter is not able to distinguish between isolated strands and hairpinlike conformations. On the other hand, we found that the DSSP method can fail when assigning helical structures to protein conformations, as a H bond between residues  $I$  and  $I + 4$  is not always a signature of ideal helical structures but can also indicate significantly distorted helices (e.g., the structure at the bottom margin of Fig. 1). To minimize artifacts that may arise from these limitations, all analyses in this work were carried out with both DSSP and dihedral-angle definitions.

## IV. POLYALANINE

### A. Attractive range of the hydrophobic interaction

As the hydrophobic interaction between molecules depends on their distance to each other, we study here the distance distributions between  $C_{\beta}$  and  $C_{\beta}$  atoms in  $\alpha$  helices and  $\beta$  sheets. These secondary structures are determined from our baseline simulation, i.e., with no sidechain hydrophobic interaction for polyalanine, using both the DSSP and the dihedral-angle methodologies. To ensure that we are studying well behaved secondary structures, analyses were performed on conformations with a large amount of secondary structures. For  $\alpha$  helices, selected conformations have  $\geq 15$  residues assigned as  $\alpha$  using the DSSP or dihedral-angle method. For  $\beta$  sheets we analyze structures that have  $\geq 11$  residues assigned as  $\beta$  by DSSP and  $\geq 10$  residues assigned as  $\beta$  using the dihedral-angle definition.

Figure 3(a) shows that the distribution of  $C_{\beta}$ - $C_{\beta}$  distances for  $\alpha$  helices first peaks at  $\approx 6.2 \text{ \AA}$ , whereas  $\beta$  sheets do not peak until  $\approx 6.9 \text{ \AA}$  [52]. Notice that the dihedral-angle and DSSP methods yield very similar distributions, and thus these distributions are robust within reasonable variations in the definition of secondary structures. A consequence of these distributions is that, if only the attractive nature of the hydrophobic energy were considered ( $E_{\text{HP}} < 0$  within the range  $R_{\text{cutoff}}$  and  $E_{\text{HP}} = 0$  elsewhere),  $E_{\text{HP}}$  would be approximately proportional to the number of contacts with  $r_{ij} \leq R_{\text{cutoff}}$  and, thus, to the cumulative distribution at  $R_{\text{cutoff}}$ . The inset in Fig. 3(a) shows that the cumulative distributions for  $\alpha$  and for  $\beta$  intersect at  $\approx 5.5 \text{ \AA}$ . This implies that  $E_{\text{HP}}$  favors  $\alpha$  helices if  $R_{\text{cutoff}} \gtrsim 5.5 \text{ \AA}$ , but favors  $\beta$  sheets if  $R_{\text{cutoff}} \lesssim 5.5 \text{ \AA}$ .

To illustrate this increase in preference for  $\alpha$  helices when the spatial range of the attractive part of the hydrophobic interaction ( $R_{\text{cutoff}}$ ) is extended, we compute the ratio between the average number of amino acids in  $\alpha$  structures and the average number in  $\beta$  structures, i.e.,  $\langle N_{\alpha} \rangle / \langle N_{\beta} \rangle$ , at different

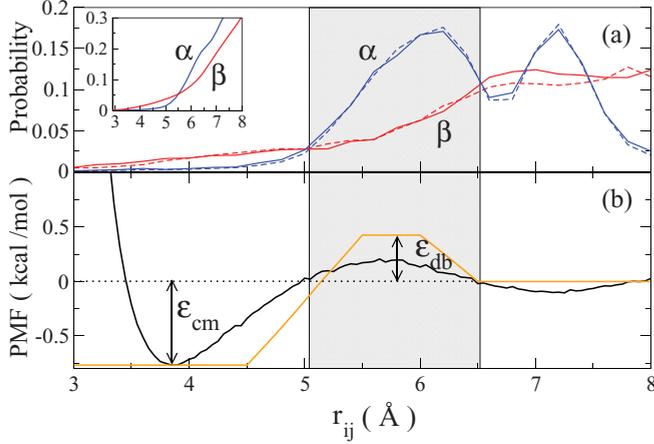


FIG. 3. (Color online) Interplay between secondary structure geometry and hydrophobic energetics. (a) Distribution of  $C_\beta$ - $C_\beta$  distance  $r_{ij}$  in  $\alpha$  and  $\beta$  structures in polyaniline for all amino acid residue pairs  $I < J - 1$ . Here the conformations were simulated using the formulation in Ref. [47], which sets  $M_{IJ} = 0$  and thus  $E_{HP} = 0$  for interactions between alanine sidechains. Both the dihedral-angle (dashed curves) and DSSP (solid curves) definitions were used to identify  $\alpha$  and  $\beta$  structures (as marked in the figure). Inset: the corresponding cumulative distributions using the DSSP definition. The corresponding cumulative distributions using the dihedral-angle definition (not shown) are essentially identical. (b) Potential of mean force of a pair of methane molecules (smooth line) in 396 TIP4P water molecules at 298.15 K computed using the test-particle insertion method [34]. Also shown is an example of our  $E_{HP}$  function in Eq. (5) (line with apparent rectilinear parts). The shaded box emphasizes that the first peak of the  $\alpha$ -helix distribution in (a) overlaps with the db region in (b).

interaction strengths. This ratio is shown in Fig. 4 for different  $R_{\text{cutoff}}$  values. For our purpose of model evaluation, and in view of the favorable contact interaction in the explicit-water methane-methane potential of mean force (PMF) [53] [see Fig. 3(b)], we used  $M_{IJ} = 0.5$  in Eq. (2) in the present simulation that varies  $R_{\text{cutoff}}$  (instead of  $M_{IJ} = 0$  for alanine-alanine interactions in the original Irbäck and Mohanty formulation [54]), and set  $A = (R_{\text{cutoff}} - 0.5 \text{ \AA})^2$  in Eq. (3). In Fig. 4, a relative increase in the amount of  $\alpha$  helices versus that of  $\beta$  sheets with increasing  $R_{\text{cutoff}}$  is evident for both the DSSP (a) and dihedral-angle (b) definitions of secondary structures. This observation indicates that choosing an appropriate range of attraction is critically important in the modeling of hydrophobic effects on secondary structure formation in polypeptides.

### B. Desolvation barrier

Whereas the  $E_{HP}$  interactions used in the model of Irbäck and Mohanty [47] contain only attractive (for  $M_{IJ} > 0$ ) and neutral parts, it has long been known that the PMF of two approaching nonpolar solutes does not decrease monotonically [53]. As illustrated by the simulated PMF in Fig. 3(b), the hydrophobic potential is expected to have a desolvation barrier (db) and a solvent-separated minimum (ssm), in addition to the contact minimum (cm) [34,35]. The difference between the cm and ssm positions  $\sim d_w$ , where  $d_w$  is the diameter of

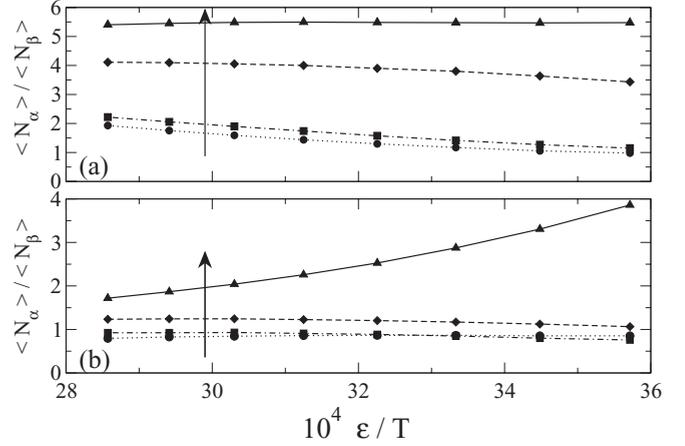


FIG. 4. Effect of the attractive range of hydrophobic interactions on the ratio of amino acid residues in  $\alpha$  versus  $\beta$  conformations.  $\langle N_\alpha \rangle / \langle N_\beta \rangle$  ratios were computed in accordance with the description in the text for  $R_{\text{cutoff}} = 4.50, 5.25, 6.25,$  and  $6.50 \text{ \AA}$ , and plotted using different symbols. The arrows indicate increasing values of  $R_{\text{cutoff}}$ . Secondary structures are determined using (a) the DSSP and (b) the dihedral-angle definitions. Here and in subsequent figures with  $\epsilon/T$ ,  $\epsilon = 1$  is used to set the energy scale for the interaction strength  $\epsilon/T$ .

a water molecule. Stipulating that hydrophobic properties of each of the sidechains'  $-\text{CH}_3$  and  $-\text{CH}_2-$  groups are similar to those of a methane molecule ( $\text{CH}_4$ ), the coincidence of the first peak of the  $\alpha$  distribution in Fig. 3(a) and the db region in Fig. 3(b) indicates that the db can be expected to disfavor the formation of helices in polyaniline. To include db effects in our simulations, we define  $f(x)$  in Eqs. (2) and (4) (with  $M_{IJ} = 1$ ) by approximating the cm and db positions of the PMF of the two methane molecules:

$$f(x) = \begin{cases} -\mathcal{E}_{\text{cm}} & \text{if } \sqrt{x} \leq r_{\text{cm}}, \\ \mathcal{E}_{\text{cm}}(x - D)/(D - C) & \text{if } r_{\text{cm}} < \sqrt{x} < r_{0,1}, \\ \mathcal{E}_{\text{db}}(x - D)/(E - D) & \text{if } r_{0,1} \leq \sqrt{x} < r_{\text{db},1}, \\ \mathcal{E}_{\text{db}} & \text{if } r_{\text{db},1} \leq \sqrt{x} \leq r_{\text{db},2}, \\ \mathcal{E}_{\text{db}}(G - x)/(G - F) & \text{if } r_{\text{db},2} < \sqrt{x} < r_{0,2}, \\ 0 & \text{if } \sqrt{x} \geq r_{0,2}, \end{cases} \quad (5)$$

where  $r_{\text{cm}} = 4.5 \text{ \AA}$ ,  $r_{\text{db},1} = 5.5 \text{ \AA}$ ,  $r_{\text{db},2} = 6.0 \text{ \AA}$ ,  $r_{0,1} = r_{\text{cm}} + \mathcal{E}_{\text{cm}}(r_{\text{db},1} - r_{\text{cm}})/(\mathcal{E}_{\text{cm}} + \mathcal{E}_{\text{db}})$ ,  $f((r_{0,1})^2) = 0$ ,  $r_{0,2} = 6.5 \text{ \AA}$ ,  $C \equiv (r_{\text{cm}})^2$ ,  $D \equiv (r_{0,1})^2$ ,  $E \equiv (r_{\text{db},1})^2$ ,  $F \equiv (r_{\text{db},2})^2$ , and  $G \equiv (r_{0,2})^2$ . An example of how  $f$  varies with interatom distance is given in Fig. 3(b). We do not include a ssm in Eq. (5) because test simulations indicated that it plays a much lesser role than the cm and db in the behaviors of our present interest.

We monitor the effects of the cm and db by studying the relative populations of  $\alpha$  and  $\beta$  conformations, i.e.,  $\langle N_\alpha \rangle / \langle N_\beta \rangle$ , as a function of the inverse temperature (interaction strength) in simulations with and without db's (Fig. 5). Secondary structures are defined using the dihedral-angle (a) and DSSP (b) methods. Using either definition, the relative populations of  $\alpha$  helices decrease in the presence of a db. This trend is consistent with the inference above from our  $C_\beta$ - $C_\beta$  distance distribution (Fig. 3). Thus, alanine sidechains in  $\beta$  conformations can better avoid the db penalty. They also form

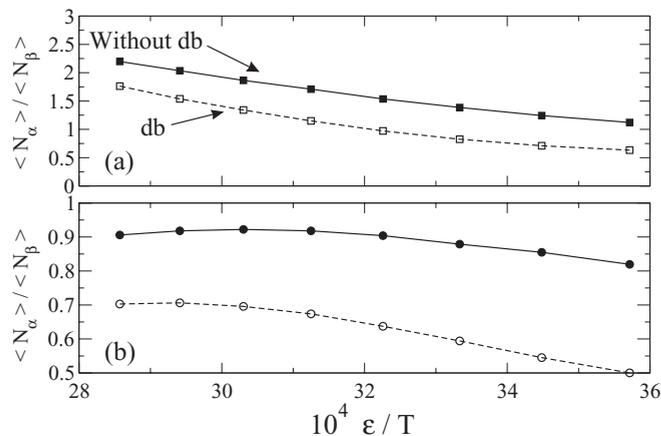


FIG. 5. Effect of the desolvation barrier (db) on the populations of  $\alpha$  and  $\beta$  conformations. Secondary structures were determined according to the dihedral-angle (a) and the DSSP (b) methods. Filled and open symbols are results from simulations without db ( $\mathcal{E}_{db} = 0$ ) and with db ( $\mathcal{E}_{db} = 5/9 \mathcal{E}_{cm}$ ), respectively, using Eqs. (4) and (5) with  $\mathcal{M}_{IJ} = 0.5$  and  $\mathcal{E}_{cm} = 1$ . Lines drawn through data points are merely guides for the eye.

more close contacts, as is evident from the higher probabilities for  $\beta$  than for  $\alpha$  in the  $r_{ij} \lesssim 5 \text{ \AA}$  region in Fig. 3(a). This is because two H-bonded  $\beta$  strands of a hairpin [Figs. 6(a) and 6(b)] have some freedom to rotate around their axis such that the distances between sidechains above and below the plane of the hairpin can be significantly different [Fig. 6(b)]. This configuration makes it possible for one pair of sidechains to contact closely ( $d_1$ ) while allowing another pair to have a  $d_2 \gtrsim d_1 + d_w$  to avoid the db.

These trends are even more apparent in the Ramachandran plots [55] for four different  $E_{HP}$ 's (Fig. 7). The  $\alpha$  and  $\beta$  regions are centered, respectively, at the lower right ( $\phi, \psi \approx -60, -45$ ) and at the upper left ( $\phi, \psi \approx -135, +135$ ). Our results indicate that  $\alpha$  is strongly favored when  $E_{HP} = 0$ , i.e., when secondary structures are stabilized only by H bonding [large dark spot at the lower right in Fig. 7(a)]. Individually, either a db or a cm reduces the favorability of  $\alpha$  vis-à-vis  $\beta$  [smaller and fainter spots at the lower right in Figs. 7(b) and 7(c)]. When both cm and db are operative,  $\alpha$  is seen to be even more disfavored [Fig. 7(d)].

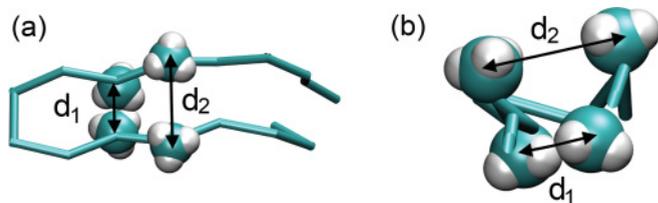


FIG. 6. (Color online) Desolvation barriers can be avoided by  $\beta$ -like conformations. Shown are the top (a) and front (b) views of an example  $\beta$ -hairpin conformation with significantly different separations  $d_1$  and  $d_2$  between pairs of sidechains above and below the plane of the sheet; see text for details.

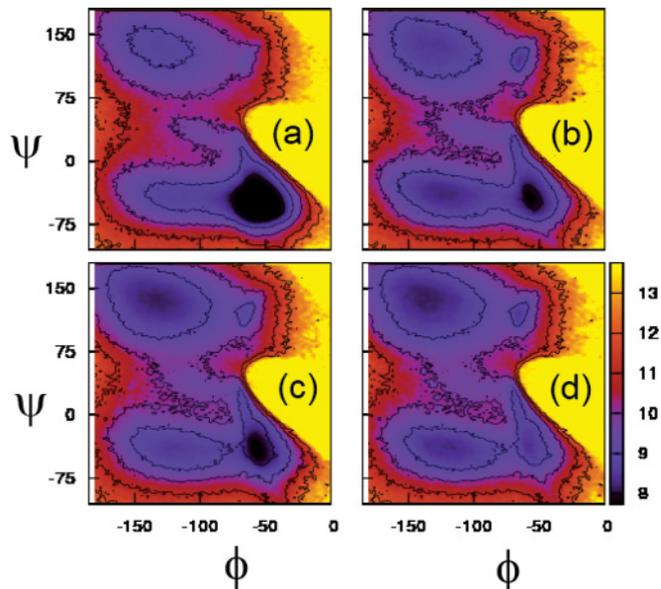


FIG. 7. (Color) Effects of db's on the Ramachandran plot. Contours and color coding show the distribution of  $-\ln[\Omega(\phi, \psi)] + \text{const}$ , where  $\Omega(\phi, \psi)$  is the probability of a residue with dihedral angles within a bin of  $2^\circ \times 2^\circ$  centered at  $(\phi, \psi)$ . Results were simulated with  $\mathcal{M}_{IJ} = 0.5$  in Eq. (4) and are shown for four different parametrizations of  $E_{HP}$  defined by Eq. (5): (a)  $E_{HP} = 0$  ( $\mathcal{E}_{cm} = \mathcal{E}_{db} = 0$ ); (b) no contact attraction ( $\mathcal{E}_{cm} = 0$ ) but with db ( $\mathcal{E}_{db} = 10/9$ ); (c) with contact attraction ( $\mathcal{E}_{cm} = 1$ ) but no db ( $\mathcal{E}_{db} = 0$ ); and (d) with both contact attraction and db ( $\mathcal{E}_{cm} = 1$ ,  $\mathcal{E}_{db} = 10/9$ ).

## V. AMINO ACIDS WITH LARGER SIDECHAINS (LEU AND VAL)

We further explore the effect of db on model peptides with larger sidechains [Ace-Leu<sub>17</sub>-NH<sub>2</sub> and Ace-Val<sub>17</sub>-NH<sub>2</sub>] by studying the relative populations of  $\alpha$  and  $\beta$  structures, i.e.,  $\langle N_\alpha \rangle / \langle N_\beta \rangle$ , as a function of the interaction strength (Figs. 8 and 9). The  $\alpha$  and  $\beta$  populations in the absence and presence of db were compared for each model. Different results were computed using the  $E_{HP}$  formulation in Eq. (2) that does not involve all heavy atom pairs (Fig. 8) and the alternate formulation that takes into account all geometric details of the sidechains through Eq. (4) (Fig. 9). Our goal is to ascertain the general trend of db effects by studying modified forms of  $E_{HP}$ , even though the balance among the energies in Eq. (1) in our alternate models is not identical to that in Ref. [47]. In both Figs. 8 and 9, it is clear that the db enhances  $\beta$  at the expense of  $\alpha$ .

## VI. SECONDARY STRUCTURES IN AMYLOIDS

To explore the generality of the above observations for homopolypeptides, we study further the effects of  $R_{\text{cutoff}}$  and db on short peptides with heterogeneous sequences. We consider three sequences extracted from amyloid peptides. They are all sufficiently flexible to assume both  $\alpha$  and  $\beta$  conformations. Each of these peptides contains several hydrophobic residues and the relative  $\alpha$  versus  $\beta$  populations are sensitive to solvation effects, suggesting that hydrophobicity might play an important role in the conformation of these peptides. We avoided sequences containing aromatic amino acids as our

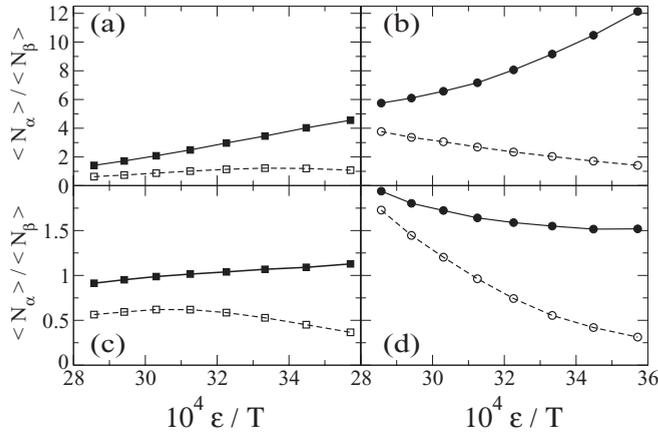


FIG. 8. Combined effects of db and hydrophobic interaction strength on secondary structure in model polyleucine (a),(b) and polyvaline (c),(d). The average numbers of  $N_\alpha$  and  $N_\beta$  residues are defined by dihedral angles on the left (a),(c) and by DSSP on the right (b),(d). Data for no-db ( $E_{db} = 0$ ) and with-db models ( $E_{db} > 0$ ) are plotted, respectively, with filled and open symbols. Lines are merely guides for the eye. Results were computed using Eqs. (2) and (5) with  $M_{IJ} = 0.9$  [54] and  $\mathcal{E}_{cm} = 1$ . Following Ref. [36],  $\mathcal{E}_{db} = (5/9)\mathcal{E}_{cm}$  is used for the with-db models.

hydrophobic model for  $-\text{CH}_2$  and  $-\text{CH}_3$  groups might not be suitable for them. We also avoided sequences that terminate in residues with large hydrophobic sidechains as the ends of  $\alpha$  or  $\beta$  structures can deviate from the ideal conformation, biasing our analyses.

The first sequence (GSNKGAIIGLM) we studied corresponds to residues 25–35 of amyloid- $\beta$  which is found in patients with Alzheimer’s disease [23,56–58]. The second sequence (KHMAGAAAAGA) corresponds to residues 110–120 of the amyloidogenic  $\beta$ -hairpin peptide of the Syrian hamster prion protein, H1 peptide [23]. The third sequence (GAVVTGVTAVA) corresponds to residues 68–78 in the middle of the hydrophobic domain of human  $\alpha$ -synuclein

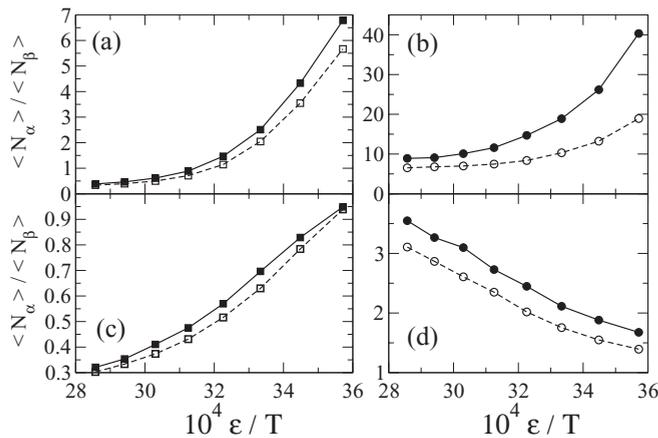


FIG. 9. Combined effects of db and hydrophobic interaction strength on secondary structure in the alternate model. As Fig. 8 except the results here were computed using the modified potential in Eq. (4) with  $\mathcal{M}_{IJ} = 0.1$  for polyleucine and  $\mathcal{M}_{IJ} = 0.05$  for polyvaline. As in Fig. 8, here we apply Eq. (5) with  $\mathcal{E}_{cm} = 1$ ,  $\mathcal{E}_{db} = 0$  for the no-db models and  $\mathcal{E}_{cm} = 1$ ,  $\mathcal{E}_{db} = 5/9$  for the with-db models.

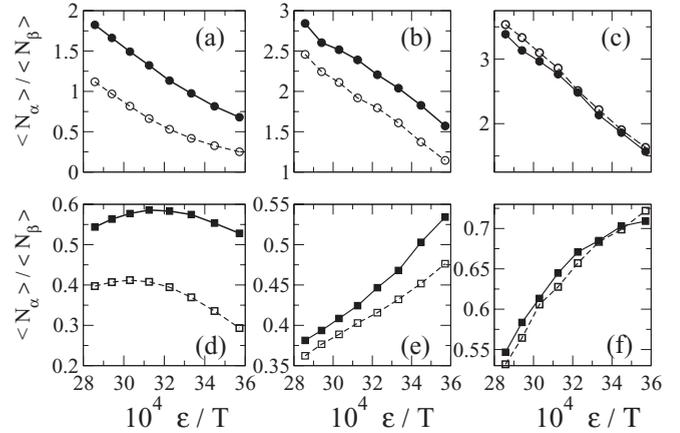


FIG. 10. Effect of  $R_{cutoff}$  on secondary structures in amyloid peptides, computed using variations of the Irbäck and Mohanty model, which does not consider desolvation barriers (see text).  $\alpha$  and  $\beta$  structures are defined by DSSP [(a)–(c), top panels] and by dihedral angles [(d)–(f), bottom panels]. Filled and open symbols are results for  $R_{cutoff} = 6.5 \text{ \AA}$  and  $5.0 \text{ \AA}$ , respectively. Results for the peptides from  $\alpha$ -synuclein, amyloid- $\beta$ , and protein H1 are shown, respectively, in the left (a),(d), middle (b),(e), and right (c),(f) panels.

[59,60] implicated in Parkinson’s disease. Explicit solvent simulations and experiments have shown that these sequences assume a large amount of  $\beta$  and coiled conformations in aqueous environments whereas the  $\alpha$  helix is their dominant conformational pattern in solvents that mimic the interior of membranes (e.g., hexafluoroisopropanol-water mixture) [23,24,56].

#### A. Effect of the attractive range of hydrophobic interactions

Figure 10 shows how  $R_{cutoff}$  affects secondary structure content. The results here were simulated using Eqs. (2) and (3), with the same  $M_{IJ}$  interaction parameters for the amino acid residues as those in the original Irbäck and Mohanty model [54] and  $\sqrt{A} = R_{cutoff} - 0.5 \text{ \AA}$  (as for Fig. 4). The results in Fig. 10 indicate that the effect of  $R_{cutoff}$  is sequence dependent. According to both the DSSP and dihedral-angle definitions for secondary structure, increasing the attractive range  $R_{cutoff}$  of hydrophobic interactions from  $5.0 \text{ \AA}$  to  $6.5 \text{ \AA}$  increases the population of  $\alpha$  relative to  $\beta$  (filled symbols taking higher values than open symbols) for the  $\alpha$ -synuclein and amyloid- $\beta$  peptides [Figs. 10(a), 10(b), 10(d), and 10(e)]. This trend is similar to that observed above for polyalanine, polyleucine, and polyvaline. However, the same change in  $R_{cutoff}$  has little effect on the relative  $\alpha, \beta$  population for the H1 peptide [Figs. 10(c) and 10(f)].

#### B. Effect of desolvation barriers

Figure 11 shows how a db affects secondary structure content. For this comparative study, the simulations were conducted using Eqs. (2) and (5), with the same values for  $r_{cm}$ ,  $r_{db,1}$ ,  $r_{db,2}$ , and  $r_{0,2}$  as described after Eq. (5). We set  $M_{IJ} = 0.1$  for alanine-alanine interactions; all other  $M_{IJ}$  values are identical to those provided by Irbäck and Mohanty [54]. We further set  $\mathcal{E}_{cm} = 0$  for alanine-alanine interactions so that the

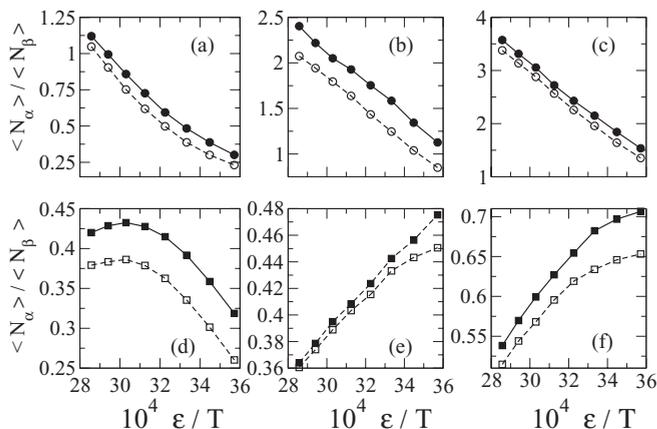


FIG. 11. Effect of db on secondary structure content in amyloid peptides. As Fig. 10 except here filled and open symbols are results from no-db ( $\mathcal{E}_{db} = 0$ ) and with-db ( $\mathcal{E}_{db} > 0$ ) models. See text for details.

no-db model is identical to that of Irbäck and Mohanty and we used  $\mathcal{E}_{db} = 1$  for the alanine-alanine with-db interaction to capture the effect of db's. For all other interactions, we set  $\mathcal{E}_{cm} = 1$  and  $\mathcal{E}_{db} = (5/9)\mathcal{E}_{cm}$  as in Figs. 5, 8, and 9. The results in Fig. 11 indicate that the population of  $\alpha$  relative to that of  $\beta$  is consistently lower with db. In other words, the effect of the db on these amyloid sequences is to penalize  $\alpha$  with respect to  $\beta$  structures (the open symbols always take lower  $\langle N_\alpha \rangle / \langle N_\beta \rangle$  values than the filled symbols). However, the magnitude of this effect is apparently sequence dependent and also sensitive to how the secondary structures are defined. The effect is more prominent in Figs. 11(b), 11(d), and 11(f) but less so in Figs. 11(a), 11(c), and 11(e).

## VII. DISCUSSION

Based on considerations of atomic structures of polypeptides, a main finding of our investigation is that physical hydrophobic interactions are unfavorable to the formation of  $\alpha$  helices in polyalanine. Geometrical constraints dictate that no water can occupy the space separating a pair of sidechains that are one turn apart along an ideal  $\alpha$  helix. As the exclusion of water in this small void leads to a higher energy, helical conformations in water are penalized relative to nonhelical conformations in polyalanine by this desolvation effect. A similar trend is observed for polyvaline and polyleucine. Because this effect hinges on the detailed packing of sidechains, we expect it to be sequence dependent. The dependence of the impact of this effect on sequence is illustrated by our results for three amyloid peptides. Taken together, our results underscore the importance of desolvation effects in the relative favorability of  $\alpha$  versus  $\beta$  conformations in polypeptides, and serve to emphasize that these effects should be taken into account in modeling [41,43] and in interpreting experimental data on conformational dependence of water-peptide interactions [61]. For example, as mentioned above, some amyloid peptides form  $\beta$  or coil conformations in water but they adopt  $\alpha$ -helical conformations in membrane mimicking environments [24]. In light of our model results, a possible scenario is that the desolvation effects are not

conducive to helix formation when these peptides are in an aqueous environment. The db effect, however, is absent in a membranelike environment, in which case the dominant interaction might be via backbone H bonds. These bonds can favor  $\alpha$  over  $\beta$  conformations because helices tend to have fewer unpaired backbone -NH and -CO groups than  $\beta$  sheets. This scenario deserves further examination.

The present model uses a pairwise additive potential. In this regard, it should be noted that hydrophobic interactions are in general not pairwise additive. Explicit-water simulations of aqueous solutions of methane molecules showed deviations from pairwise additivity in their interactions; but the deviations are mild [38,62,63]. For instance, for the interaction between a pair of polyalanine helices, the total potential based on an implicit-solvent model that assumes pairwise additivity [40] is similar to the corresponding PMF determined by explicit-water simulation [42]. Both potentials clearly show a db, although the db in the explicit-water PMF is somewhat lower than that predicted by the additive treatment (Fig. 1 of Ref. [40]). In view of these considerations, we do not expect the nonadditivity of hydrophobic interactions to alter the general db-related trend observed in the present study. Indeed, our results are consistent with the finding from an early explicit-water simulation of the interaction between water and an alanine-based  $\alpha$  helix ( $\text{CH}_3\text{-ALA}_{18}\text{-NH-CH}_3$ ) that water does not penetrate much between spatially adjacent  $C_\beta$  groups along the helix axis [64]. The same study showed that water rather tends to occupy the space above the NH-O hydrogen bond in a triangle-shaped region defined by three proximate  $C_\beta$  groups (see Fig. 3 of Ref. [64]). In light of the apparent favorability of this spot for water, this position may correspond to a three-body version of the two-body solvent-separated minimum [34]. It would be instructive to ascertain the extent to which this phenomenon can be rationalized by simple implicit-water models in future studies.

Another general area for which the concepts put forth in this work should be applicable is protein folding and misfolding. Secondary structure formation is context dependent [14,65]. It follows that a secondary structure that is favored in the protein core might not be favored under denaturing conditions. For instance, under denaturing conditions muscle myoglobin was shown to assume fibril-like conformations that resemble amyloid and prion aggregates consisting of  $\beta$  strands oriented perpendicular to the main fiber axis, confirming the idea that the sequences of amyloid and prion peptides do not need to be fundamentally different from those of other proteins [66]. If db's indeed favor  $\beta$  versus  $\alpha$  for a polypeptide that is fully exposed to an aqueous environment, db effects might be a contributing factor in fibril formation.

An obvious generalization of our approach is to extend the present analysis of db-related geometric constraints to all possible amino acid residue pairs. Such a systematic analysis should provide some degree of physical rationalization for the aforementioned sequence dependent effects on db-promoted  $\beta$ -structure formation among amyloid peptides. More generally, information from such a systematic analysis may offer insight into the nonrandom statistics of residue-residue correlation in  $\beta$  sheets and in  $\alpha$  as well as  $3_{10}$  helices [67]. As suggested above, extension of the present approach should also be relevant to understanding secondary

structure formation in transmembrane proteins, which can have predominantly  $\alpha$ -helical [68,69] or  $\beta$ -barrel [70,71] structures. It would be interesting to explore the role of db-related and steric constraints in the conformational transitions between water-soluble and membrane-bound forms of these proteins.

In summary, we trust that the general trends observed in our simulations are robust because they are based on simple geometrical considerations, although our model might not account for more subtle solvation effects that can arise from an explicit treatment of water [64,72] or a possible coupling between hydrophobic and other interactions [29,73]. Our results show that both a repulsive db and a physically realistic range of cm attraction of  $\sim 5.0$  Å between methyl or methylene groups tend to favor  $\beta$  over  $\alpha$  structures, at least for the polyalanine, polyvaline, and polyisoleucine chains we have studied. This finding may help to assess prior studies that used effective ranges of cm attraction as large as 6.5 Å [31,32]. db's arise from empty spaces created by water exclusion between hydrophobic groups [37–40,42] and are the hallmark of interactions between nonpolar groups [33,34,36–43] that are partially exposed to water. As such, db's are pertinent to the  $\alpha$ - $\beta$  transition in amyloidogenesis [26,27], which can

be affected by pressure [27,74–76]. Atomic PMF simulations indicated that temperature lowers both the cm and the db [34,37,38,42] whereas pressure raises them [77]. Inasmuch as a lower cm favors  $\beta$  (Fig. 7) but a lower db favors  $\alpha$  (Figs. 5–9), and vice versa, the impact of either temperature or pressure on the relative favorabilities of  $\alpha$  and  $\beta$  will hinge on the balance between the opposite effects on  $\alpha$  vs  $\beta$  propensities caused by temperature- or pressure-induced changes in the PMF at the cm and db. We are confident that the conceptual framework developed here will facilitate elucidation of these fundamental questions through further efforts in theory and experiment.

#### ACKNOWLEDGMENTS

We thank Dr. Markus Miettinen for helpful discussions. This work was supported in part by grants from the Natural Sciences and Engineering Research Council of Canada and Ontario Early Researcher Award Program (M.K.) and the Canadian Institutes of Health Research and the Canada Research Chairs Program (H.S.C.). We are also grateful for the computing resources provided by SciNet ([www.scinet.utoronto.ca](http://www.scinet.utoronto.ca)) and SHARCNET ([www.sharcnet.ca](http://www.sharcnet.ca)) of Compute Canada.

- 
- [1] C. L. Dias, T. Ala-Nissila, M. Karttunen, I. Vattulainen, and M. Grant, *Phys. Rev. Lett.* **100**, 118101 (2008).
- [2] C. L. Dias *et al.*, *Cryobiology* **60**, 91 (2010).
- [3] P. De Los Rios and G. Caldarelli, *Phys. Rev. E* **63**, 031802 (2001).
- [4] T. Petermann and P. De Los Rios, *Phys. Rev. E* **73**, 026114 (2006).
- [5] W. Kauzmann, *Adv. Protein Chem.* **14**, 1 (1959).
- [6] K. A. Dill, *Biochemistry* **29**, 7133 (1990).
- [7] H. Li, C. Tang, and N. S. Wingreen, *Phys. Rev. Lett.* **79**, 765 (1997); H. S. Chan, *Nat. Struct. Biol.* **6**, 994 (1999).
- [8] A. Zarrine-Afsar *et al.*, *Proc. Natl. Acad. Sci. USA* **105**, 9999 (2008); Z. Zhang and H. S. Chan, *ibid.* **107**, 2920 (2010).
- [9] L. Pauling, R. B. Corey, and H. R. Branson, *Proc. Natl. Acad. Sci. USA* **37**, 205 (1951); L. Pauling and R. B. Corey, *ibid.* **37**, 251 (1951).
- [10] H. S. Chan and K. A. Dill, *Proc. Natl. Acad. Sci. USA* **87**, 6388 (1990).
- [11] T. P. Creamer and G. D. Rose, *Proc. Natl. Acad. Sci. USA* **89**, 5937 (1992).
- [12] M. Blaber, X. J. Zhang, and B. W. Matthews, *Science* **260**, 1637 (1993).
- [13] M. H. J. Cordes, N. P. Walsh, C. J. McKnight, and R. T. Sauer, *Science* **284**, 325 (1999).
- [14] P. A. Alexander, Y. He, Y. Chen, J. Orban, and P. N. Bryan, *Proc. Natl. Acad. Sci. USA* **106**, 21149 (2009).
- [15] E. Bornberg-Bauer, *Biophys. J.* **73**, 2393 (1997); Y. Cui, W. H. Wong, E. Bornberg-Bauer, and H. S. Chan, *Proc. Natl. Acad. Sci. USA* **99**, 809 (2002); D. Shortle, *ibid.* **106**, 21011 (2009).
- [16] H. Xiang, B. L. Buckwalter, H.-M. Shieh, and M. H. Hecht, *Proc. Natl. Acad. Sci. USA* **92**, 6349 (1994).
- [17] M. H. Hecht, J. S. Richardson, D. C. Richardson, and R. C. Ogdén, *Science* **249**, 884 (1990).
- [18] L. Regan and W. F. DeGrado, *Science* **241**, 976 (1988).
- [19] K. A. Dill, S. Bromberg, K. Yue, K. M. Fiebig, D. P. Yee, P. D. Thomas, and H. S. Chan, *Protein Sci.* **4**, 561 (1995).
- [20] A. Rey and J. Skolnick, *Proteins* **16**, 8 (1993).
- [21] A. Kolinski, W. Galazka, and J. Skolnick, *J. Chem. Phys.* **103**, 10286 (1995).
- [22] G. Bellesia, A. I. Jewett, and J.-E. Shea, *Protein Sci.* **19**, 141 (2010).
- [23] I. Daidone *et al.*, *Proteins* **57**, 198 (2004).
- [24] J. Heller *et al.*, *Protein Sci.* **5**, 16551 (1996).
- [25] G. H. Wei, N. Mousseau, and P. Derreumaux, *Proteins* **56**, 464 (2004).
- [26] M. G. Krone *et al.*, *J. Am. Chem. Soc.* **130**, 11066 (2008).
- [27] C. Dirix *et al.*, *J. Mol. Biol.* **347**, 903 (2005).
- [28] O. Collet and C. Chipot, *J. Am. Chem. Soc.* **125**, 6573 (2003).
- [29] H. S. Chan, *Proteins* **40**, 543 (2000); H. Kaya and H. S. Chan, *ibid.* **52**, 510 (2003).
- [30] Y. Levy, J. Jortner, and O. M. Becker, *Proc. Natl. Acad. Sci. USA* **98**, 2188 (2001).
- [31] F. Ding *et al.*, *Proteins* **53**, 220 (2003); Y. Mu and Y. Q. Gao, *J. Chem. Phys.* **127**, 105102 (2007).
- [32] H. D. Nguyen, A. J. Marchut, and C. K. Hall, *Protein Sci.* **13**, 2909 (2004).
- [33] N. Hillson, J. N. Onuchic, and A. E. García, *Proc. Natl. Acad. Sci. USA* **96**, 14848 (1999).
- [34] S. Shimizu and H. S. Chan, *J. Chem. Phys.* **113**, 4683 (2000).
- [35] C. L. Dias *et al.*, *J. Chem. Phys.* **134**, 065106 (2011).
- [36] M. S. Cheung, A. E. García, and J. N. Onuchic, *Proc. Natl. Acad. Sci. USA* **99**, 685 (2002).
- [37] D. Paschek, *J. Chem. Phys.* **120**, 6674 (2004).
- [38] M. S. Moghaddam, S. Shimizu, and H. S. Chan, *J. Am. Chem. Soc.* **127**, 303 (2005).

- [39] Z. Liu and H. S. Chan, *J. Mol. Biol.* **349**, 872 (2005).
- [40] A. Ferguson, Z. Liu, and H. S. Chan, *J. Mol. Biol.* **389**, 619 (2009); **401**, 153(E) (2010).
- [41] Y. Levy and J. N. Onuchic, *Annu. Rev. Biophys. Biomol. Struct.* **35**, 389 (2006).
- [42] J. L. MacCallum *et al.*, *Proc. Natl. Acad. Sci. USA* **104**, 6206 (2007).
- [43] H. S. Chan, Z. Zhang, S. Wallin, and Z. Liu, *Annu. Rev. Phys. Chem.* **62**, 301 (2011).
- [44] M. S. Lin, N. L. Fawzi, and T. Head-Gordon, *Structure* **15**, 727 (2007).
- [45] G. Némethy, *Angew. Chem., Int. Ed. Engl.* **6**, 195 (1967).
- [46] G. Némethy and H. A. Scheraga, *J. Phys. Chem.* **66**, 1773 (1962).
- [47] A. Irbäck and S. Mohanty, *Biophys. J.* **88**, 1560 (2005)
- [48] Note that the overall minus sign in the expression for  $E_{HP}$  in Eq. (7) of Ref. [47] is absorbed into the definition of  $f(x)$  in Eq. (3) of the present paper.
- [49] J. H. Meinke *et al.*, *Comput. Phys. Commun.* **178**, 459 (2008).
- [50] A. Irbäck, *J. Phys.: Condens. Matter* **17**, S1553 (2005).
- [51] W. Kabsch and C. Sander, *Biopolymers* **22**, 2577 (1983).
- [52] In G. E. Schulz and R. H. Schirmer, *Principles of Protein Structure* (Springer-Verlag, New York, 1979), the listed ideal  $(\phi, \psi)$  values for right-handed  $\alpha$  helix and antiparallel, parallel, and twisted  $\beta$  sheets are, respectively,  $(-57, -47)$ ,  $(-139, +135)$ ,  $(-119, +113)$ , and  $\approx(-110, +140)$ . Thus, the corresponding  $C_{\beta}-C_{\beta}$  distances for the  $(I, I + 3)$  and  $(I, I + 4)$  positions in an ideal  $\alpha$  helix and for the  $(I, I + 2)$  positions in the three types of  $\beta$  sheet are, respectively, 5.84, 6.55, 6.93, 6.54, and  $\approx 7.05$  Å. Notably, the first peak in our simulated  $\alpha$ -helix distribution [Fig. 3(a)] is located in between the distances separating residue pairs  $I, I + 3$  and  $I, I + 4$  in an ideal  $\alpha$  helix, whereas the first peak in our simulated  $\beta$ -sheet distribution is located near the distances separating  $I, I + 2$  residue pairs in the three types of ideal  $\beta$  sheets.
- [53] L. R. Pratt and D. Chandler, *J. Chem. Phys.* **67**, 3683 (1977).
- [54] See Table 1 of Ref. [47].
- [55] G. N. Ramachandran, C. Ramakrishnan, and V. Sasisekharan, *J. Mol. Biol.* **7**, 95 (1963).
- [56] G. Wei and J.-E. Shea, *Biophys. J.* **91**, 1638 (2006).
- [57] A. Baumketner and J.-E. Shea, *J. Mol. Biol.* **362**, 567 (2006).
- [58] Y. Xu, J. Shen, X. Luo, W. Zhu, K. Chen, J. Ma, and H. Jiang, *Proc. Natl. Acad. Sci. USA* **102**, 5403 (2005).
- [59] W. S. Davidson, A. Jonas, D. F. Clayton, and J. M. George, *Int. J. Biol. Chem.* **273**, 9443 (1998).
- [60] A. J. Trexler and E. Rhoades, *Biochemistry* **48**, 2304 (2009).
- [61] M. Kohtani and M. F. Jarrold, *J. Am. Chem. Soc.* **124**, 11148 (2002).
- [62] J. A. Rank and D. Baker, *Protein Sci.* **6**, 347 (1997).
- [63] S. Shimizu and H. S. Chan, *Proteins: Struct., Funct., Genet.* **48**, 15 (2002).
- [64] A. E. García, G. Hummer, and D. M. Soumpasis, *Proteins: Struct., Funct., Genet.* **27**, 471 (1997).
- [65] D. L. Minor and P. S. Kim, *Nature (London)* **380**, 730 (1996).
- [66] M. Fändrich, M. A. Fletcher, and C. M. Dobson, *Nature (London)* **410**, 165 (2001).
- [67] G. von Heijne and C. Blomberg, *J. Mol. Biol.* **117**, 821 (1977); E. G. Hutchinson, R. B. Sessions, J. M. Thornton, and D. N. Woolfson, *Protein Sci.* **7**, 2287 (1998); B. Wathen and Z. Jia, *Int. J. Mol. Sci.* **10**, 1567 (2009); *Biochem. Cell Biol.* **88**, 325 (2010).
- [68] K. R. MacKenzie, *Chem. Rev.* **106**, 1931 (2006).
- [69] J. L. Popot and D. M. Engelman, *Biochemistry* **29**, 4031 (1990).
- [70] E. Gouaux, *Curr. Opin. Struct. Biol.* **7**, 566 (1997).
- [71] W. C. Wimley, K. Hristova, A. S. Ladokhin, L. Silvestro, P. H. Axelsen, and S. H. White, *J. Mol. Biol.* **277**, 1091 (1998).
- [72] A. E. García and K. Y. Sanbonmatsu, *Proc. Natl. Acad. Sci. USA* **99**, 2782 (2002).
- [73] H. Kaya and H. S. Chan, *Phys. Rev. Lett.* **85**, 4823 (2000).
- [74] J. L. Silva, Y. Cordeiro, and D. Foguel, *Biochim. Biophys. Acta – Proteins and Proteomics* **1764**, 443 (2006).
- [75] S. Grudzielanek, V. Smirnovas, and R. Winter, *J. Mol. Biol.* **356**, 497 (2006).
- [76] S. Grudzielanek, Y. Zhai, and R. Winter, *ChemPhysChem* **11**, 2060 (2010); **11**, 2060(E) (2010).
- [77] G. Hummer *et al.*, *Proc. Natl. Acad. Sci. USA* **95**, 1552 (1998).