

Simple model for the simulation of peptide folding and aggregation with different sequences

Marta Enciso and Antonio Rey

Citation: *J. Chem. Phys.* **136**, 215103 (2012); doi: 10.1063/1.4725883

View online: <http://dx.doi.org/10.1063/1.4725883>

View Table of Contents: <http://jcp.aip.org/resource/1/JCPSA6/v136/i21>

Published by the [American Institute of Physics](#).

Additional information on J. Chem. Phys.


Journal Homepage: <http://jcp.aip.org/>

Journal Information: http://jcp.aip.org/about/about_the_journal

Top downloads: http://jcp.aip.org/features/most_downloaded

Information for Authors: <http://jcp.aip.org/authors>

ADVERTISEMENT



AIPAdvances

Special Topic Section:
PHYSICS OF CANCER

Why cancer? Why physics? [View Articles Now](#)

Simple model for the simulation of peptide folding and aggregation with different sequences

Marta Enciso^{a)} and Antonio Rey

Departamento de Química Física I, Facultad de Ciencias Químicas. Universidad Complutense, E-28040 Madrid, Spain

(Received 27 March 2012; accepted 19 May 2012; published online 7 June 2012)

We present a coarse-grained interaction potential that, using just one single interaction bead per amino acid and only realistic interactions, can reproduce the most representative features of peptide folding. We combine a simple hydrogen bond potential, recently developed in our group, with a reduced alphabet for the amino acid sequence, which takes into account hydrophobic interactions. The sequence does not pose any additional influence in the torsional properties of the chain, as it often appears in previously published work. Our model is studied in equilibrium simulations at different temperatures and concentrations. At low concentrations the effect of hydrophobic interactions is determinant, as α -helices (isolated or in bundles) or β -sheets are the most populated conformations, depending on the simulated sequence. On the other hand, an increase in concentration translates into a higher influence of the hydrogen bond interactions, which mostly favor the formation of β -type aggregates, in agreement with experimental observations. These aggregates, however, still keep some distinct characteristics for different sequences. © 2012 American Institute of Physics. [<http://dx.doi.org/10.1063/1.4725883>]

I. INTRODUCTION

Coarse-grained models are widely used in protein and peptide studies. This intermediate-resolution approach replaces the individual details of the atomic forces by effective bead-bead interactions, speeding up the sampling process.^{1,2} They average out many degrees of freedom while trying to grasp the most relevant aspects of the studied process. The obtention of realistic results lies, then, on the usage of purpose-based potentials. Many minimalist potentials can be found in the literature (see below), often differing in the kind of information they want to reproduce and in the extent of simplification of the system description. Along this work, we will focus on those potentials that aim to describe peptide folding and aggregation, using just one bead per amino acid.

In relation to these single-bead minimalist models, many strategies have been proposed in the last decades^{3,4} to study both peptide and protein folding. The use of just one bead per amino acid derives in a drastic simplification of the system, whose implications must be weighted beforehand. As a result, this kind of models and their associated interactions often needs to rely on a reference structure, as it happens in $G\phi$ or structure-based potentials, where all the system interactions are built according to the native structure of the system.⁵ This necessity is partially removed with so-called Sorenson-like models,^{6–8} that introduce realistic driving forces, mainly hydrogen bonds and/or hydrophobic interactions. Although they rely on physics-based principles, they also need an *a priori* knowledge of a reference secondary structure for the parametrization of the local geometry of the

chain, which helps to avoid an unspecific hydrophobic collapse of the chains.

In other cases, the information from a reference structure is completely removed and just physics-based forces apply in the system.^{9–12} However, the large simplification of single-bead descriptions often limits the scope of these studies, that are blind to some characteristic features of peptides, such as sequence-dependent effects. “Average” peptides with generic attractive interactions have been considered instead,^{10,12} leading to interesting conclusions which are, however, sequence-insensitive. Some heteropolymers have been modeled using a similar background, but using sequences without a biological meaning.¹³

As a more complex alternative, other authors have used models in which several interaction centers per residue are used to describe the geometry of the peptide chain.¹⁴ These models allow for a more accurate description of some of the present interactions, but result in a higher computational cost. We precisely try to find results comparable to these models, without such a detailed description of the polypeptide chain.

In this work, we present a single-bead minimalist potential that includes the effect of sequence without any bias towards reference structures. We have included three physics-based interaction terms: backbone hydrogen bonds,¹⁵ hydrophobic interactions between different types of residues, and a sequence-independent dihedral term that controls the chain stiffness without imposing any secondary structure. We have based the two latter on a widely used and accepted Sorenson-like potential,⁶ where we have removed any information about reference states or *a priori* constraints related to secondary structures. This simple yet realistic potential has been developed thanks to a thorough analysis of geometric protein properties.

^{a)}Present address: Institut für Mathematik, Freie Universität, 14195 Berlin, Germany.

Peptides have traditionally been a very relevant tool for the understanding of biological systems,^{16,17} as they share the same building blocks as proteins (the individual amino acids) and can hold some keys about the behavior of full proteins in aspects such as folding, enzymatic activity, etc. Besides, peptides play an important role for testing interaction potentials.^{18,19} Their small size allows extensive analysis at a moderate computational cost, constituting therefore valuable model systems for further (and larger) protein studies.

Currently, peptides have also become the object of many computational and experimental efforts from the scientific community, due to their relevance in aggregation studies and their connection to neurodegenerative diseases.²⁰ In fact, evaluating the behavior of amyloid-forming peptides sheds light on some key factors of aggregation, such as the role of concentration and sequence in this process.^{21,22}

Our goal in this article includes the design of the potential mentioned above and its use for the evaluation of sequence-dependent effects in peptide systems, using for that purpose simple sequences (i.e., with very regular distributions of hydrophobic and hydrophilic residues). We have simulated two different sequences (either α -prone or β -prone) under different temperature and concentration conditions, observing a proper sensitivity of our model towards changes in these variables. We have analyzed how sequence modifies the energetic and structural landscape on these peptide systems, building schematic “structural phase diagrams” in terms of those variables.

The study of these peptide systems provides, in addition, a valuable tool to understand how hydrophobic and hydrogen bonds balance in low concentration conditions (where a native-like behavior is expected) and high concentration ones (where aggregates are said to be favored).²³ In fact, recent studies relate the competition between folding and aggregation to an inversion on the relationship between these two driving forces.²⁴

We have to remark that the potential introduced in this work is not directly aimed towards the solution of the protein folding problem. This is a too complex problem for the simple approach we are using here. What we try to do is to find the simplest geometric model (i.e., one single center of interaction per amino acid) able to reproduce the most distinctive features of peptide behavior in solution, including sequence-dependent folding and aggregation above a certain concentration. Moreover, we try to do this by using only realistic driving forces, namely, backbone hydrogen bonds and hydrophobic interactions. As far as we know, this is the first time all these requirements are simultaneously fulfilled since, as previously described, other models either include more centers of interaction per amino acid or do not consider the effect of sequence, or even include additional forces based on a single desired structure.

II. METHODS

Peptides are described in our model through a single bead per amino acid, placed at the α -carbon position. Beads within a same peptide chain are connected through virtual bond vectors, with a fixed length of 3.8 Å. The chain flexibility has

been modeled according to the chemistry of the real bonds: the virtual bond angle associated with three consecutive beads is allowed to range from 65° to 150°.

Using this simplified representation, we aim to describe peptide behavior in realistic terms, i.e., the model has to be able to provide reasonable native-like geometries and responses towards changes in sequence, concentration, and temperature. For this purpose, we have only used the two most relevant physics-based interactions that are present in peptide systems: hydrogen bonds and hydrophobics.

The election of these potentials has been very careful, as most single-bead potentials tend to stabilize non-natural structures (which are distorted versions of the secondary structure elements) and/or need to rely on information from a reference state in order to define the secondary structure of the different chains. Our global expression for the interaction energy of the system is the following:

$$E = \omega^{hb} E^{hb} + \omega^{hp} E^{hp} + \omega^{stiff} E^{stiff}. \quad (1)$$

The first term corresponds to a hydrogen bond contribution that has been described in full detail recently,¹⁵ for this reason, we only mention here a few relevant aspects about it. It is applied between any pair of residues i and j (where $j > i + 2$ and $j \neq i + 4$). Its energy calculation consists of two-steps: first, we check three geometrical restrictions (namely, the length of the tentative hydrogen bond between beads, R1; the orientation between the auxiliary vectors, R2; and the relative orientation between those auxiliary vectors and the tentative hydrogen bond, R3). Second, a step-like potential applies if the values of the former restrictions fall within certain limits, expressed in Table I. Acceptable ranges and potential strength differ depending on the kind of hydrogen bond (either local/helical or non-local/ β -type) and are described elsewhere.¹⁵

The two latter terms in Eq. (1) describe the hydrophobic interactions that can be formed among the system beads and control the stiffness of the polypeptide chain, respectively. They have been based on a Sorenson-like minimalist potential designed by Head-Gordon and co-workers.⁶ The original definition, inspired by a lattice representation,²⁵ merged a Lennard-Jones potential for the hydrophobic contribution (based on a three-letter alphabet where residues can be either neutral N , polar L , or hydrophobic B) with a severe control of the local geometry of the chain. It results in a pre-determined definition of the secondary structure of each given fragment as a function of its sequence. As we discussed in the Introduction, this is the kind of feature we want to avoid in our potential.

Therefore, we have performed two major modifications in relation to the original potential. Regarding the

TABLE I. Optimal ranges for the three geometrical restrictions chosen in our model for backbone hydrogen bonds.

Restriction	Local range (Å)	Non-local range (Å)
R1	$4.7 \leq R1 \leq 5.6$	$4.0 \leq R1 \leq 5.6$
R2	$0.74 \leq R2 \leq 0.93$	$0.75 \leq R2 \leq 1.00$
R3	$0.92 \leq R3 \leq 1.00$	$0.94 \leq R3 \leq 1.00$

TABLE II. Values of the parameters for the sequence-dependent terms of E^{hp} . The right hand side contains the values that σ_{int} , S_1 , and S_2 take depending on the kind of the interaction (note that the two latter quantities are adimensional).

Type of interaction	σ_{int} (Å)	S_1	S_2
$B - B$	4.00	15.45	1
$L - L$ or $L - B$	3.00	15.45	-1
$N - L$, $N - B$, or $N - N$	3.00	15.45	0

hydrophobic interaction itself, we have slightly changed its shape to avoid a too strong non-specific collapse of the chains and to adapt it to real native-like geometries, in agreement with the hydrogen bond term.¹⁵ The resulting expression is defined as follows:

$$E^{hp} = \sum_i^{N-4} \sum_{j>i+4}^N E_{i,j}^{hp};$$

$$E_{i,j}^{hp} = S_1 \left[\left(\frac{\sigma_{int}}{r_{i,j}} \right)^{12} - S_2 \left(\frac{\sigma_{int}}{r_{i,j}} \right)^{10} \right]. \quad (2)$$

In this equation, $r_{i,j}$ is the distance between the interacting beads and S_1 , S_2 , and σ_{int} are parameters that depend on the nature of the interaction, as shown in Table II. If two hydrophobic residues interact ($B - B$), the positive value of S_2 indicates an attractive minimum in the Lennard-Jones potential; if hydrophilic residues interact among them ($L - L$) or with hydrophobic beads ($L - B$), S_2 becomes negative and the interaction is repulsive; if a neutral residue is involved, there is also some short-range repulsion among beads. As an example, we show in Figure 1(a) the $B - B$ interaction, compared to the original definition by Brown *et al.*⁶ Our version shifts the minimum towards slightly larger distances, but at the same time makes the attractive well narrower, to avoid too sticky situations.

Regarding the chain stiffness (third term in Eq. (1) and also included in the potential by Brown and co-workers), we have observed that this term, defined through the virtual torsional angles (ϕ), is also crucial to avoid an excessive “stickiness” of the potential when long range hydrophobic interactions are included.⁶ In the original definition this term imposes, too, the local geometry of the fragment. The authors define from the beginning three types of residues (helical, extended, or turn) and stabilize different secondary structures depending on them. This *a priori* choice of secondary structure needs to be removed in our model, as we pursue a different aim: to reproduce the natural trend of proteins, where a particular combination of hydrophobic and polar residues favors the formation of α -helices or β -sheets (also stabilized by the formation of hydrogen bonds) without imposing the local geometry in any additional way. Moreover, we aim to study the interconversion of folded structures (that may have different kinds of local geometries) into aggregated ones (that mostly exhibit a β -type structure²⁶) by the effect of concentration. Changes in this property of our system must then be able to modify the balance between the two types of local arrangements and, therefore, the observed stable

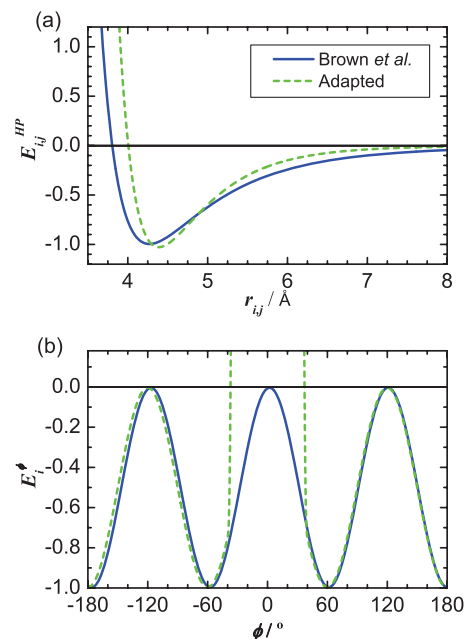


FIG. 1. Representation of some of the energetic contributions of our adapted hydrophobic interaction (green), compared to the original work by Brown and co-workers (blue).⁶ (a) Sequence-dependent long-range term for the hydrophobic-hydrophobic interactions. (b) Stiffness term associated to the torsional virtual angle between four consecutive α -carbons, compared to the turn-type interaction of the former work.⁶

structure. Obviously, if the local geometry is selected from the beginning, this structural change is strongly hampered, or even becomes impossible.

To avoid these drawbacks, we have used the same functional form for the stiffness contribution for all the virtual torsional angles of our chains, regardless the type of amino acid. Its functional shape is similar to the turn-type one of the Brown potential.⁶ It equally favors α -helices and β -sheets. In addition, we have strongly penalized local conformations whose torsional angle lies near 0° , that would lead to unnatural and too compact conformations. Therefore, the final functional form, represented in Figure 1(b), is as follows:

$$E^{stiff} = \sum_i^N E_i^\phi;$$

$$E_i^\phi = \begin{cases} 0.5 (1 + \cos 3\phi_i) - 1 & \text{if } |\phi_i| \geq 40^\circ \\ 10.0 & \text{otherwise.} \end{cases} \quad (3)$$

The relative values of the three weighting factors, ω^{hb} , ω^{hp} , and ω^ϕ in Eq. (1), have been optimized through extensive simulations of peptide systems and complete proteins under different conditions. We have analyzed different sets of parameters looking for optimal structural features and thermodynamic properties, such as, a proper obtention of helical and β -type structures in terms of concentration, the acquirement of reasonable folded structures for protein sequences (not shown in this article), etc. We have found that the most suitable weights are $\omega^{hb} = 9.5$, $\omega^\phi = 7.0$, and $\omega^{HP} = 6.5$, which have been therefore used to get all the results presented here.

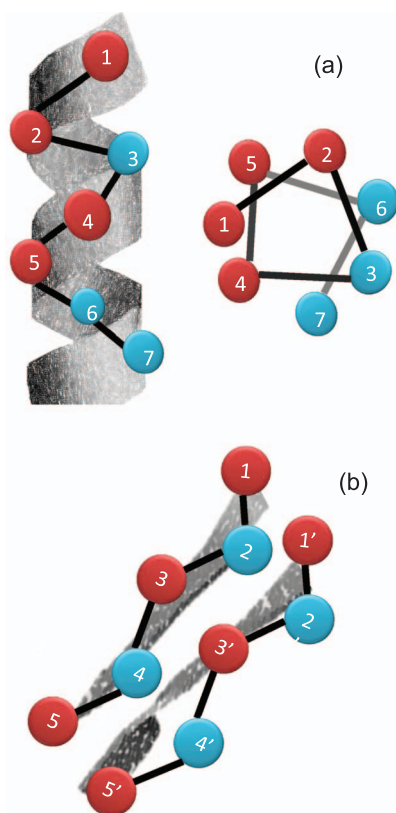


FIG. 2. Schematic description of the relative arrangement of hydrophobic (blue) and hydrophilic (red) residues for (a) a helical fragment of sequence $L_2BL_2B_2$, and (b) a β -sheet where each chain's sequence is $(LB)_2L$.

Once we have set up the potential definition, we need to define proper sequences to model helical and β -forming peptides, as particular series of amino acids favor those particular arrangements in native proteins. Note that the desired structures will be achieved by a suitable combination of residues “that tend to be buried in the structure” (hydrophobic, B) and residues “that repel each other” (polar, L), without any additional help. In our case, the simplicity of our potential definition needs to keep our sequences simple, too. Inspired by previous efforts found in the literature for the Brown *et al.* potential,⁸ we have used very regular sequences. We have also adapted the peptide lengths according to the experimental average lengths of α -helices (close to 13 residues) and β -strands (9 residues). In this way, our “helical sequence” is $L_2B(BL_2)_2B_2L_2$ and the “ β -type” one is $(LB)_4L$, as we have shown (in shorter versions) in Figure 2. As seen in this figure, the first sequence, when the chain adopts a helical conformation, shows a hydrophobic and a hydrophilic face, which could result in a proper association of the chains into a helical bundle. On the other hand, the second sequence creates two faces when every chain takes an extended conformation, favoring the association of the different chains into β -sandwich structures. As we shall show, these situations are found in some, but not all, the environmental conditions considered here.

In order to study the characteristics of the whole energetic and structural landscape of the simulated peptides according to our model, we have used a replica exchange Monte Carlo

simulation algorithm (REMC), as previously described.²⁷ We have carried out multichain numerical experiments using periodic boundary conditions, where each of our simulations presents 24–40 temperatures. Our simulations start from a completely extended conformation for each chain and consists of 5×10^6 Monte Carlo cycles at every temperature after 3×10^6 equilibration cycles. In each cycle, every bead of the system is subjected to a trial Monte Carlo move.

We have computed different concentrations for a six-chain system, modifying for this purpose the size of the simulation box. In this article we shall show only the four most representative ones, ranging from 0.2 to 2.0 chain moles/L (with intermediate concentrations at 0.5 and 1.0 chain moles/L). Note that the numerical values of the simulated concentrations exemplify the variation analyzed in this work, but they do not try to quantitatively reflect a real experimental concentration.

The results presented here correspond to statistical averages over the sampling at every temperature and over different independent runs. For each system, three or five independent runs have been carried out. The structural analysis has been performed in terms of the number and type of hydrogen bond interactions, the radius of gyration of the independent chains and the distances between their centers of mass, as found in the recorded simulation results. These quantities, chosen after an extensive visual analysis of the conformations appearing in the computed trajectories, have allowed us to characterize the most relevant (highly populated) structural arrangements resulting from our model in every set of conditions. Our approach to this calculation starts then by visually exploring the computed trajectories. Given the relative simplicity of the peptide systems studied, we have been able to readily detect distinct relevant structural situations, which are then properly differentiated in an automatic analysis of the output files through the combination of the quantities mentioned above. This way, we can obtain their relative populations, as well as individual snapshots which are considered as representative configurations of the system and are presented below. In addition, the energetic properties of the systems have been obtained by the calculation of the heat capacity curves versus temperature, computed from the thermal fluctuations of the systems. Both energy and temperature, as well as derived quantities such as the heat capacity, are measured in standard reduced units.

III. RESULTS

Along this work, we are introducing a single-bead coarse-grained model that includes sequence-dependent hydrophobic interactions and hydrogen bonds, without any reference towards native conformations or *a priori* definitions of secondary structure regions through the use of torsional preferences. We aim to describe the behavior of peptides with different sequences under changes in concentration, as this variable is considered to play a key role in peptide aggregation.²³

As mentioned in Sec. II, we have used a three-letter alphabet to describe the hydrophobicity of each amino acid, and we have designed peptides whose sequences are meant to stabilize, based on the hydrophobicity patterns alone,

either helical or strand-like configurations. In this section, we present some relevant properties of our α -prone and β -prone systems. We pay a particular attention to the impact of the sequence itself, as well as to the effect of concentration. These two variables constitute the most relevant features that a realistic coarse-grained model should consider. Besides, a realistic reproduction of the effects of these properties helps in the understanding of the forces that drive peptide folding and aggregation, from a microscopic point of view.

Sequence determines the stable tridimensional layout of a certain peptide or protein in a given environment. In this way, simulating different sequences should lead to different stable structures. Therefore, we start the analysis of each of our systems by presenting all the possible relevant structures that we have found in our simulations, compatible with their particular sequence.

In order to build up a comprehensive view in structural and energetic terms, we have computed the relative population of each of these structures, in terms of the different environmental conditions (temperature and concentration, in this particular study). We will show the results of these analyses for two representative concentrations, for each type of sequence.

A. α -prone systems

We show in Figures 3(a)–3(c) example snapshots of the most representative structural arrangements that we have found in helical-prone systems, aside from unstructured configurations. As we can see, all of these structures keep the “helical print” imposed by their sequence. We find helical bundles, such as the one in Figure 3(a), that are mainly stabilized by the hydrophobic interaction among three helical chains into a single associated unit; this helicity, favored both by local hydrogen bonds and by the hydrophobic intrachain interactions of the chosen sequence, has been checked by the detection of $(i, i + 3)$ hydrogen bonds, corresponding in our hydrogen bond interaction to local bonds.¹⁵ In some temperature/concentration conditions, we have also found β -type structures such as Figure 3(b), where most of the peptides acquire a strand-like arrangement, but some of them (typically, one or two) keep a helical configuration. If we look carefully at the β -sheet in that figure, we observe that the interchain hydrogen bonds are correctly formed between the different strands. The helical-prone sequence of these peptides creates a certain “accumulation” of exposed hydrophobic residues, that become somewhat protected through their interaction with the hydrophobic face of one α -helix, in this case. In Figure 3(c) we show a snapshot of independent helices, also found in this type of simulations, that are stabilized by intrachain hydrogen bonds and a small number of local hydrophobic interactions.

These structures have been identified along the computed trajectories using their geometric properties and the type and distribution of individual interactions, as we have explained in the Methods section. Using these analyses, we have evaluated the population of each one of these structures at the different concentrations and temperatures we have simulated. This

α -prone sequences β -prone sequences

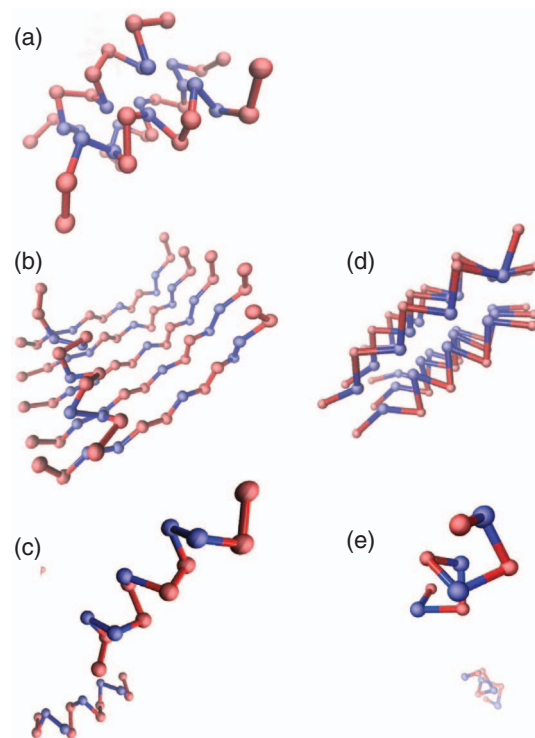


FIG. 3. Cartoon representations of the most representative configurations found in our simulations of α -prone systems (left, structures (a), (b), and (c)) and β -prone systems (right, structures (d) and (e)). (a) Three-helix bundle. (b) β -sheet interacting with one helix. (c) Isolated α -helices (from helical sequence). (d) β -sandwich. (e) Isolated α -helices (from β -prone sequences).

information is summarized in Figure 4, where we show the results of two representative concentrations (the most diluted and concentrated ones in our simulations). Figure 4(a) shows their heat capacity curves versus temperature; a peak in this plot indicates an energetic transition, whose temperature has been marked with a vertical line for clarity. In the lower part of Figure 4 we show the thermal evolution of the population of the different structures we have presented in Figure 3, for the diluted system, in Figure 4(b), and for the concentrated one, in Figure 4(c).

Starting with the most diluted system, we observe in the black curve of Figure 4(a) that there is one peak, indicating in principle the presence of a single transition. As we can see in the population analysis of Figure 4(b), the stable structure below the thermal transition in the three-helix bundle. At the transition temperature bundles are lost, partially populating an intermediate state where the isolated helices remain relatively stable in a narrow temperature range. At higher temperatures, the system is unstructured, with the individual peptide chains randomly changing their conformations.

In the concentrated system, the red curve in Figure 4(a) has two peaks, that indicate two different transitions whose nature is described in Figure 4(c). The transition at lower temperatures shows that the energetic minimum for this system (i.e., the most stable structure at very low temperatures) corresponds again to the three-helix bundle. In highly concentrated environments such as this one, however, the system

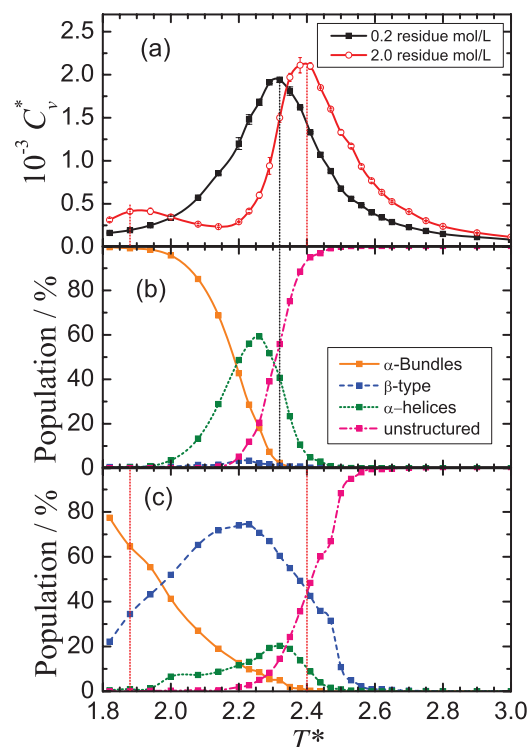


FIG. 4. Thermodynamic and structural data of the helical-prone system for two different concentrations, a “diluted” (0.2 residue mol/L) and a “concentrated” one (2.0 residue mol/L). (a) Heat capacity curves versus temperature for the two concentrations. (b) Thermal evolution of the population of the different structures (helical bundles, β -type, independent helices, and unstructured configurations) found in the diluted system. (c) Thermal evolution of the population of the observed structures in the concentrated system.

acquires, still at relatively low temperatures, a predominant β -type arrangement, that falls apart at the higher transition temperature. Similarly to the diluted case, we detect isolated helices in the surroundings of that second transition temperature, although in a proportion much smaller than in the diluted conditions.

Therefore, the effect of concentration in helical-prone sequences induces a change in the kind of structural and energetic transitions that take place in the system. In this way, in diluted systems we observe the unfolding of helical bundles when the temperature rises, while high concentration conditions promote the formation of β -type structures, which become stable in a relatively wide low temperature range.

B. β -prone systems

The structures obtained from β -prone systems reflect, again, the specific properties of the used sequence. In this way, we have mainly found two different types of configurations. The most abundant one is the β -sandwich. We show an example of this structural type in Figure 3(d), where each sheet of the sandwich is formed by three chains, mainly stabilized by interchain hydrogen bonds, and the two sheets remain together due to the interactions between their hydrophobic faces. In addition, we have found some independent α -helices (see Figure 3(e)), stabilized by intrachain hydrogen bonds. They reflect the trend of isolated peptides to get at least

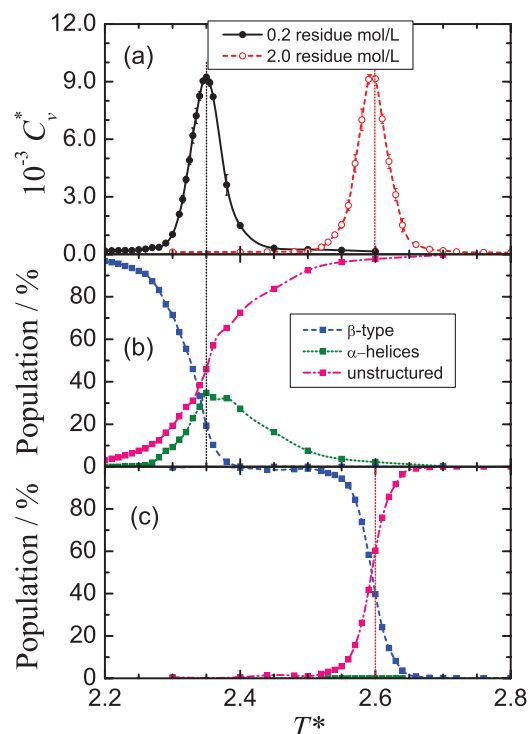


FIG. 5. Thermodynamic and structural data of the β -prone system at a “diluted concentration” (0.2 residue mol/L) and a “concentrated” one (2.0 residue mol/L). (a) Heat capacity curves versus temperature. (b) Thermal evolution of the population of the different structures (β -sandwich, independent helices, and unstructured configurations) found in the diluted system. (c) Thermal evolution of the population of the observed structures in the concentrated system.

partially stabilized by the formation of local hydrogen bonds, as it also happened in sequenceless peptides¹⁵ and has been reported from different models as well.¹²

We show in Figure 5 the evolution of β -type systems with temperature, for two different concentrations (0.2 residue mol/L and 2.0 residue mol/L). Starting with the heat capacity curves in Figure 5(a), we observe that both present now a single peak of similar size and shape whose position along the temperature axis is highly sensitive to changes in concentration. If we analyze the structural evolution of the system along temperature, in Figures 5(b) and 5(c) for the diluted and concentrated systems, respectively, we observe in both cases a similar situation, where the β -sandwich unfolds at the transition temperature. In the most diluted case, we can appreciate some helical configurations at the transition temperature, that are minority at 0.2 residue mol/L and almost negligible (i.e., below 5%) at higher concentrations.

This indicates that, in general terms, β -prone systems present the same kind of structural transition at every concentration. An increase in the transition temperature with concentration indicates a raise in the stability of the sandwich. Although this may in part reflect the different characteristics of the denatured state as a function of concentration (mainly entropic), it can be also related to the tendency to aggregate (i.e., to form a multichain structure) in high concentration conditions.²³

IV. DISCUSSION

We have applied the coarse-grained model we defined in the Methods section to simulate peptides using different sequences, temperatures, and concentration conditions. As we have explained in Sec. III, our model presents a convenient sensitivity towards both parameters.

In the case of helical-prone sequences, we have shown that an increase in the concentration conditions changes the kind of structural transition that takes place upon thermal folding/denaturation. In this way, three-helix bundles are the most stable arrangement at low temperatures in mildly concentrated environments. The formation/disappearance of these bundles takes place through an intermediate composed by independent α -helices, whose population is only relevant in a narrow temperature range. At high concentrations, the most stable structure at intermediate and relative low temperatures is a β -type one, mainly stabilized by interchain hydrogen bonds. This structure is lost above the transition temperature without noticeable intermediates. At intermediate concentrations, we observe the competition between both phenomena.

This information has allowed us to map the energetic landscape of our systems, building a “structural phase diagram” which can be seen as a summary of our results, shown in Figure 6(a) for the case of helical systems. The straight boundaries among the different regions in the diagram just reflect the lack of a detailed knowledge, as it would appear in a real phase diagram.¹² Here, we are just showing the configurations with a significant population for a system in the region centered at a simulated temperature and concentration. The transition temperature leading to the unstructured region remains nearly constant at all the concentrations. Therefore, we do not observe a large change in the stability of the system, if this is to be measured by an important shift in the transition temperature, as detected in a heat capacity experiment. However, we have observed a substantial rearrangement of the interactions that stabilize the system below this transition temperature, depending on the concentration.

This fact can be related to the intrinsic propensity of peptides towards aggregation, regardless their particular sequence,²³ something that is especially relevant in the case of sequences which are prone to form helical structures at low concentrations. We have showed that a very simple model like ours is able to reproduce this fact, shifting the kind of stabilizing interactions of the system, from hydrophobic to hydrogen bonds as the concentration increases, as it has recently been proposed.²⁴

In the case of β -type peptides, changes in concentration modify the stability of the system as defined above, but keep a similar structural landscape, as we can observe in the “structural phase diagram” of Figure 6(b). In all cases, we observe that the β -sandwich structure is the preponderant configuration below the transition temperature. A decrease in the concentration lowers the stability of the β -sandwich, shifting the related transition temperature towards smaller values. The kind of β -type structure obtained for this sequence, which already includes many hydrogen bonds at

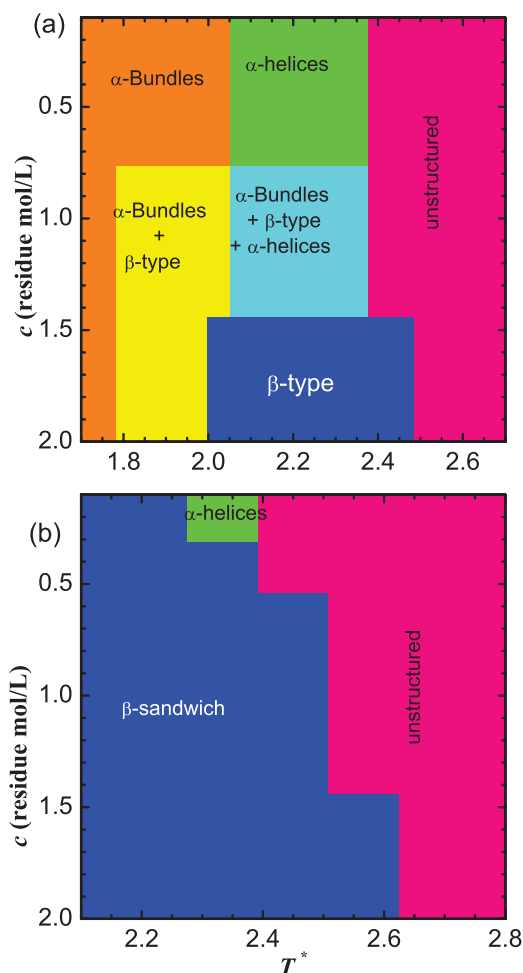


FIG. 6. Schematic structural phase diagram for multichain peptides according to our simulation model. (a) α -forming peptides. (b) β -prone peptides.

low concentrations, creates a situation which is, structurally, less heterogeneous than in the sequences favoring helical structures.

Regarding our interest in peptide aggregation, we may claim that our observations match the experimental observations that state that high concentration conditions promote the formation of aggregates for peptides and proteins with different sequences.²³ The change in the structure of helical bundles at high concentrations is a clear indication of this fact. Besides, our simulations of β -type systems provide an additional evidence, as they link a concentration rise with an increase in the stability of the multichain structure.

The model presented is therefore adequate for the study of peptides with simple sequences. We are at this moment extending the work to proteins where different secondary structure elements can be found together, linked into the same chain and properly packed in solution (work in progress). The results we have got so far (data not shown) indicate that, with an adequate design of the full regular sequence, simple proteins can be also folded with this model in dilute conditions, and the effect of concentration taken into account.

V. CONCLUSIONS

We have designed a physics-based potential that, using just one center of interaction per amino acid, is able to reproduce the most relevant features of peptide behavior in solution under different temperatures and concentrations. In addition, it is able to distinguish between different sequences without any additional information, something which was missing in other recent models.¹² Although results like these are not new, we believe we have been able to get them with a model which significantly reduces the complexity of the system, in comparison with other previously reported,¹⁴ and which is not biased towards a given structure, being then more realistic than models of similar complexity.⁶

We have carried out REMC simulations at equilibrium conditions to investigate the thermal folding/denaturation of two different simplified sequences (either helical or β -prone), each of them at four different concentrations.

Helical sequences stabilize three-helix bundles in mild and moderate concentration conditions; in highly concentrated environments they can form alternative multichain structures with a β -type nature, mainly stabilized by interchain hydrogen bonds. This observation exemplifies the shift in the interaction balance from native-like (in this case, bundles) to aggregated-like (in this case, β -type structures) under the effect of concentration,²⁴ involving a change in the main determinants of the structure: from hydrophobic interactions to backbone hydrogen bonds. In the case of β -prone systems, a decrease in concentration lowers the stability of the β -sandwich, which is the most stable structure at all the simulated concentrations.

As a result of our analysis, we have obtained structural phase diagrams for both types of peptides, that illustrate the general behavior of our systems under different conditions. These diagrams are not as accurate from a thermodynamic point of view as others previously reported,¹² but they are able to include the effects of the peptides sequence in the type and stability of the structures detected. The comparison of the schematic diagrams in Figure 6 with those for sequenceless peptides with our model for hydrogen bond interactions (see Figure 9 in Ref. 15) better shows this effect of the peptide sequences. The hydrophobic potential contribute to a more realistic description of aggregates, avoiding the stabilization of the unnatural structures that usually appear within these (and also similar) simple models (see Ref. 15 and the discussion therein). In addition, the presence of stable three-helix bundles for the corresponding sequences reduces the presence of independent helices, stabilized by hydrogen bonds alone, as the only representative for α -type structures. This creates a richer structural diagram for α -prone sequences than for sequenceless peptides. On the other hand, with β -prone sequences and the corresponding hydrophobic interactions the comparison yields the opposite results, with a simpler structural diagram in which the presence of isolated peptides with α -helical structure is marginal when the sequence is considered, and framed to a very small temperature and concentration range.

We would expect the situation to be obviously more complex with less regular sequences, as those appearing in real

peptides and proteins. The phase diagrams could be more complicated, especially at low temperatures and high concentrations, where hydrophobic interactions play a more important role in the behavior of the system, as we have shown for our model. However, if the proportion of hydrophobic and hydrophilic residues in the sequences is "reasonable," we do not expect that the main gross characteristics of the results presented here, mainly those related to the effect of concentration in interchain aggregation through hydrogen bonds, and the influence of the sequence in the structures of the peptides in dilute systems, would be so different.

Finally, we want to stress out that all these realistic results have been obtained using a very simple and computationally efficient potential and model, able to sample tens of different temperatures for a given system in a reasonably short computation time. Other recent works have required the use of more sophisticated geometrical descriptions and/or interaction potentials for the consideration of the concentration or the sequence in coarse-grained peptide models,^{14,28} avoiding in some cases the structural changes upon aggregation. The results presented in this work show that a minimalist approach, if carefully designed, may be valid for the realistic simulation of simple peptide systems, at least at the same level as other more sophisticated (and computationally expensive) models previously reported.

ACKNOWLEDGMENTS

This work was partially supported by the Spanish Ministerio de Ciencia e Innovación (Grant No. FIS2009-13364-C02-02), by Comunidad Autónoma de Madrid (Grant No. S2009/PPQ-1551) and by Universidad Complutense de Madrid/Banco Santander (Grant No. GR35/10-A-910068). M.E. acknowledges a scholarship from Spanish Ministerio de Educación (FPU program).

¹C. Hyeon and D. Thirumalai, *Nat. Commun.* **2**, 487 (2011).

²C. Wu and J. E. Shea, *Curr. Opin. Struct. Biol.* **21**, 209 (2011).

³V. Tozzini, *Curr. Opin. Struct. Biol.* **15**, 144 (2005).

⁴C. Clementi, *Curr. Opin. Struct. Biol.* **18**, 10 (2008).

⁵R. D. Hills and C. L. Brooks, *Int. J. Mol. Sci.* **10**, 890 (2009).

⁶S. Brown, N. J. Fawzi, and T. Head-Gordon, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 10712 (2003).

⁷E.-H. Yap, N. L. Fawzi, and T. Head-Gordon, *Proteins* **70**, 626 (2008).

⁸J. W. Mullinax and W. G. Noid, *Proc. Natl. Acad. Sci. U.S.A.* **107**, 19867 (2010).

⁹T. X. Hoang, A. Trovato, F. Seno, J. R. Banavar, and A. Maritan, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 7960 (2004).

¹⁰S. Auer, C. M. Dobson, and M. Vendruscolo, *HFSP J.* **1**, 137 (2007).

¹¹S. Kaschiev and S. Auer, *J. Chem. Phys.* **132**, 215101 (2010).

¹²S. Auer and D. Kashchiev, *Phys. Rev. Lett.* **104**, 168105 (2010).

¹³T. Vogel, T. Neuhaus, M. Bachmann, and W. Janke, *Phys. Rev. E* **80**, 011802 (2009).

¹⁴H. D. Nguyen and C. K. Hall, *Biophys. J.* **87**, 4122 (2004).

¹⁵M. Enciso and A. Rey, *J. Chem. Phys.* **132**, 235102 (2010).

¹⁶H. Dyson and P. E. Wright, *Curr. Opin. Struct. Biol.* **3**, 60 (1993).

¹⁷B. E. Kemp and R. B. Pearson, *Trends Biochem. Sci.* **15**, 342 (1990).

¹⁸M. S. Shell, R. Ritterson, and K. A. Dill, *J. Phys. Chem. B* **112**, 6878 (2008).

¹⁹Y. Pak, S. Jang, and S. Shin, *J. Chem. Phys.* **116**, 6831 (2002).

²⁰R. M. Murphy, *Annu. Rev. Biomed. Eng.* **4**, 155 (2002).

²¹A.-M. Fernandez-Escamilla, F. Rousseau, J. Schymkowitz, and L. Serrano, *Nat. Biotechnol.* **22**, 1302 (2004).

- ²²W. B. Stine, K. N. Dahlgren, G. A. Krafft, and M. J. LaDu, *J. Biol. Chem.* **278**, 11612 (2003).
- ²³D. Thirumalai, D. K. Klimov, and R. I. Dima, *Curr. Opin. Struct. Biol.* **13**, 146 (2003).
- ²⁴A. W. Fitzpatrick, T. P. J. Knowles, C. A. Waudby, M. Vendruscolo, and C. M. Dobson, *PLoS Comput Biol* **7**, e1002169 (2011).
- ²⁵F. H. Stillinger, T. Head-Gordon, and C. L. Hirshfeld, *Phys. Rev. E* **48**, 1469 (1993).
- ²⁶R. Nelson, M. R. Sawaya, M. Balbirnie, A. Ø. Madsen, C. Riekel, R. Grothe, and D. Eisenberg, *Nature (London)* **435**, 773 (2005).
- ²⁷L. Prieto, D. de Sancho, and A. Rey, *J. Chem. Phys.* **123**, 154903 (2005).
- ²⁸D. Alemani, F. Collu, M. Cascella, and M. D. Peraro, *J. Chem. Theory Comput.* **69**, 315 (2010).