The Journal of Chemical Physics

Folding and self-assembly of a small heterotetramer

Cite as: J. Chem. Phys. **140**, 105103 (2014); https://doi.org/10.1063/1.4868140
Submitted: 18 December 2013 . Accepted: 27 February 2014 . Published Online: 14 March 2014

Fatih Yaşar, Adam K. Sieradzan, and Ulrich H. E. Hansmann

View Online Export Citation CrossMark

ARTICLES YOU MAY BE INTERESTED IN

A general method for the derivation of the functional forms of the effective energy terms in coarse-grained energy functions of polymers. I. Backbone potentials of coarse-grained polypeptide chains
The Journal of Chemical Physics **146**, 124106 (2017); https://doi.org/10.1063/1.4978680

Velocity scaling for optimizing replica exchange molecular dynamics
The Journal of Chemical Physics **134**, 044124 (2011); https://doi.org/10.1063/1.3533236

Lock-in Amplifiers
Find out more today

Zurich Instruments

AIP Publishing

© 2014 AIP Publishing LLC.

Folding and self-assembly of a small heterotetramer

Fatih Yaşar,[1,a)] Adam K. Sieradzan,[2,b)] and Ulrich H. E. Hansmann[3,c)]
[1]*Department of Physics Engineering, Hacettepe University, Beytepe-Ankara 06800, Turkey*
[2]*Faculty of Chemistry, University of Gdańsk, Wita Stwosza 63, 80-952 Gdańsk, Poland*
[3]*Department of Chemistry and Biochemistry, University of Oklahoma, Norman, Oklahoma 73019-5251, USA*

(Received 18 December 2013; accepted 27 February 2014; published online 14 March 2014)

Designed miniproteins offer a possibility to study folding and association of protein complexes, both experimentally and *in silico*. Using replica exchange molecular dynamics and the coarse-grain UNRES force field, we have simulated the folding and self-assembly of the heterotetramer BBAThet1, comparing it with that of the homotetramer BBAT1 which has the same size and $\beta\beta\alpha$-fold. For both proteins, association of the tetramer precedes and facilitates folding of the individual chains. © *2014 AIP Publishing LLC*. [http://dx.doi.org/10.1063/1.4868140]

I. INTRODUCTION

For a long time, numerical studies of proteins have focused on monomeric proteins. This is in part due to the computational hurdles in modeling the folding of even small proteins (of order 100 residues), but also motivated by the long-prevalent picture that the native structure of a protein is determined solely by its sequence of amino acids. However, for a many proteins, the biologically active structure depends on the interaction with other proteins or biomolecules, and may not even be unique.[1] Hence, the formation of oligomeric proteins or protein complexes is not necessarily a hierarchic process where first the components fold and afterwards assemble, but rather an interplay of folding and association. For instance, we have recently[2,3] studied the folding and self-assembly of the 84-residue synthetic BBAT1,[4] build out of four identical chains that are 21-residue long and have a $\beta\beta\alpha$-fold. We found that association of the four chains precedes their folding; and distinct intermediates are observed before the BBAT1 assumes its final fold. While the folding of the individual chains follows the same mechanism as isolated monomers, the various transitions along the folding pathway are facilitated by the tetramer environment.

In the present paper, we extend this study to the more complex heterotetramer BBAThet1 (PDB-identifier: 1XOF)[5] shown in Fig. 1. Similar to BBAT1, the protein has a mixed α/β structure and is formed by four chains of 21 residues. However, BBAThet1 is not a homotetramer but build out of two kinds of peptides that differ in four residues. The sequences of both chains (called by us A and B) and that of the BBAT1 chain are listed in Table I, with the differences marked in boldface. All three peptides contain the artificial amino acid DapBz (Dbz) which consists of α, β-diaminopropionic acid derivatized with a benzoyl functionally on the side chain nitrogen. They differ in length and position of the helical segment: for BBAT1 it spans from Asp7 to Gly21, while for chain A of BBAThet1 it extends from Asp7 to Dbz20, and for chain B from Phe8 to Ala19. In all three peptides, β-sheet segments are made of Arg2-Ile3 and Tyr6-Asp7. The different compositions of the chains, leading to the same structure of the tetramer, allows us to investigate how the amino acid sequence modulates coupling between conformational changes of the monomers and their assembly into a tetramer.

In order to overcome the sampling difficulties involved in studies of association and folding of proteins of this size, we use the same combination of enhanced sampling techniques[6] and coarse-grained model[7] as in our previous BBAT1 study. Namely, we use replica exchange molecular dynamics simulations[8–10] with the UNRES force field[11–13] to determine the folding and association pathway of BBAThet1 in solution, and compare it with that of the homotetrameric BBAT1. Unlike many other coarse-grained models, UNRES is a physics-based energy function and has been derived as a potential of mean force of polypeptide chains immersed in water. It is characterized by temperature-dependent parameters[12,14] fitted to reproduce the heat capacity curves of many proteins.[15,16] The force field has been used successfully in the past to study folding of proteins, including oligomeric proteins.[17] For instance, use of this force field leads our previous work[2] to good correlation with experimental results for oligomeric proteins with $\beta\beta\alpha$-fold. For the above reasons, we have chosen in the present study this force field, albeit this choice is by no means the only possible one. Alternative approaches include the use of an all-atom Go-model as previously used by the Onuchic group[18,19] in the folding and association of Rop-dimers;[20] or all-atom physical force fields, relying either on "brute-force" molecular dynamics[21] or enhanced sampling techniques.[22,23] Similarities between physical all-atom simulations and such of coarse-grained Go-models have been recently studied in Ref. 24. Using our combination of UNRES energy function and replica exchange sampling technique, we find a mechanism of association mediated folding where the association of unfolded coil-like chains generates an environment that facilitates folding of the chains. The latter is a two-step process

a)Electronic mail: fatih@hacettepe.edu.tr
b)Electronic mail: adam86@wp.pl
c)Electronic mail: uhansmann@ou.edu

 © 2014 AIP Publishing LLC

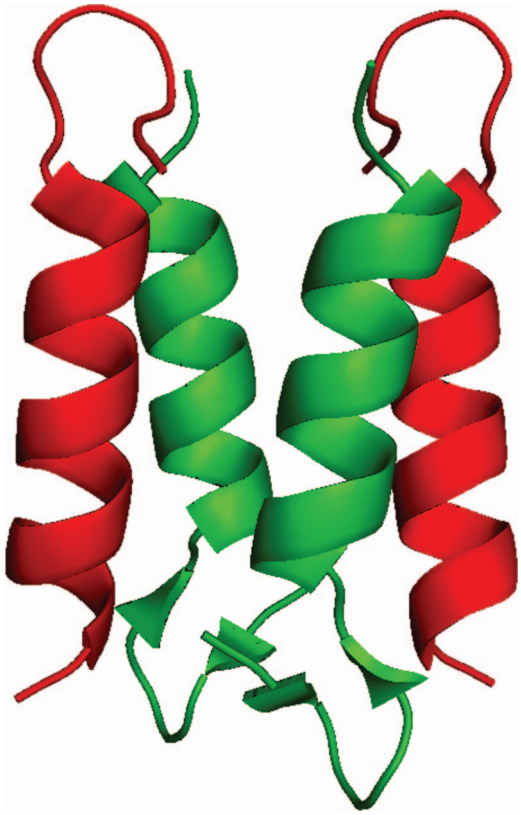

FIG. 1. Crystal structure of BBAThet1 (chain A is drawn in red, and chain B in green). The picture was prepared with PyMOL.[38]

involving a bend-like intermediate that lacks most of the native secondary structure.

II. METHODS

All simulations are performed with the latest version of the UNRES molecular modeling package[13,14] which implements the UNRES force field[11–13] and a multiplexed extension of the replica exchange molecular dynamics algorithm.[25] UNRES represents a polypeptide chain by a sequence of α-carbon atoms with united side chains attached, and peptide groups positioned halfway between two consecutive α-carbons. The effective energy function is defined as the free energy of the chain constrained to a given coarse-grained conformation plus the surrounding solvent. This potential of mean force of the virtual-bond chain contains temperature dependent prefactors and its explicit form can be found in Ref. 14. Additional parameters to describe the D-proline and the side chain of the benzyloxy diaminopropionic acid DapBz have been taken from Refs. 26 and 27.

Our replica exchange molecular dynamics simulations of the isolated monomers A and B, and of the heterotetramer, start from either fully extended or random configurations. All-virtual bonds-dihedral angles of four extended monomer chains are set to 180°, after which each extended chain is rotated by a random angle around the z, y, and x axis. The chains are subsequently translated by random vectors whose lengths are within the interval $(\sqrt{3}l_{max}, \sqrt{3}l_{max} + 10$ Å), where l_{max} is the maximum length of the extended chain B (which is larger than that of chain A). If the distance between the geometric center of any two chains is greater than the l_{max}, the configuration is discarded and another attempt is made.

As in previous work, 36 temperatures between 250 K and 600 K are used for both the isolated monomers and the tetramer. This flow-optimized temperature distribution[28] is given by 250 K, 256 K, 262 K , 268 K, 274 K, 278 K, 281 K, 284 K, 287 K, 290 K, 293 K, 297 K, 301 K, 305 K, 309 K, 313 K, 317 K, 321 K, 325 K, 329 K, 333 K, 338 K, 344 K, 351 K, 359 K, 368K , 379K , 391 K, 405 K, 421 K, 439 K, 450 K, 485 K, 520 K, 560 K, and 600 K. The tetramer system is simulated with 4 copies at each temperature, and isolated monomers with 2 copies per temperature. In our tetramer simulation, we use an energy penalty if the distance between chains growth beyond 60 Å. The total simulation time is 400 ns and 200 ns for the tetramer and each isolated monomer, respectively. The temperature is controlled by the Berendsen thermostat,[29] with angular momentum reset every 1000 steps. The maximum acceleration change is set at 4 and the equations of motion are solved by a multiple time step algorithm.[30] While the Berendsen thermostat does not guarantee sampling of a canonical distribution,[31] and with some force fields leads to artificial distributions of velocities,[32] there is extensive evidence that the deviations are small when using the Langevin thermostat with UNRES in MREMD simulations.[13]

The obtained data are analyzed using the weighted histogram analysis method to elucidate the thermodynamic quantities at any temperature.[33] Note that the calculation of the root-mean-square deviation (RMSD) between a given configuration and the reference structure requires accounting for the permutational symmetry of the tetramer. We define for this purpose $RMSD(\mathbf{X}) = \min_{\mathcal{P}}\{RMSD[\mathcal{P}(\mathrm{ABCD})]\}$ where X denotes a given conformation of the tetramer, and $\mathcal{P}$ marks a permutation of the four chains (A, B, C, and D).

III. RESULTS AND DISCUSSION

In order to understand the mechanism of folding and association of BBAThet1, we start with investigating the folding of the isolated monomers. In our previous work, we found that isolated monomers of the simpler BBAT1 fold in

TABLE I. Sequences of BBAT1[4] and BBAThet1[5] chains (BBAThet1 = 2*(Chain A) + 2*(Chain B), a = D-ala, p = D-pro, and Z = DapBz). Differing residues are in boldface.

Peptide	N-ter	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	C-ter
BBAT1	Ace	Y	R	I	p	S	Y	D	F	**G**	D	**E**	**L**	**A**	K	L	L	R	**Q**	A	Z	G	NH2
Chain A	Ace	Y	R	I	p	S	Y	D	F	**a**	D	**E**	**A**	**E**	K	L	L	R	**D**	A	Z	G	NH2
Chain B	Ace	Y	R	I	p	S	Y	D	F	**a**	D	**K**	**F**	**K**	K	L	L	R	**K**	A	Z	G	NH2

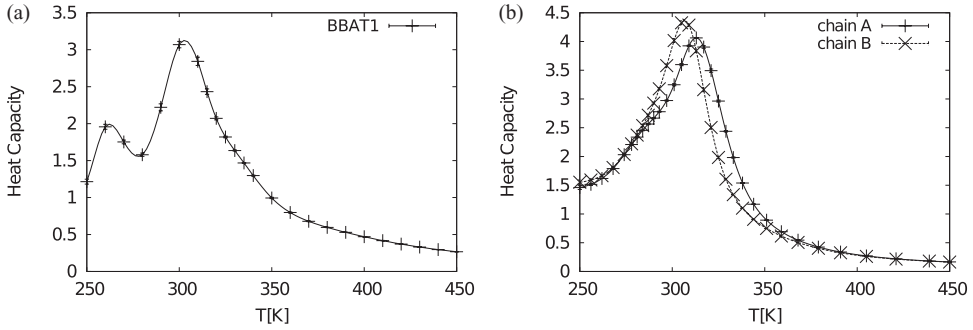

FIG. 2. Heat capacity as function of temperature for (a) an isolated BBAT1 monomer (data taken from Ref. 2). and (b) the isolated BBAThet1 monomers A and B.

a two-step process marked by two distinct peaks in the specific heat at temperatures $T_{cu} \approx 305$ K and $T_f \approx 270$ K, see Fig. 2(a). The corresponding specific heat curves for the BBAThet1 monomers are shown in Fig. 2(b). Unlike for BBAT1 monomers, we observe for both chain A and B of BBAThet1 only a single peak at $T_{cu}^A = 313 \pm 2$ K ($T_{cu}^B = 307 \pm 2$ K).

The high-temperature peak of the BBAT1 monomers ($T_{cu} \approx 305$ K) marks the collapse of the BBAT1 monomers into a compact form, where the C- and N-termini are closer together than in the folded structure. This collapse is also observed for both kind of BBAThet1 monomers; however, the corresponding minima in the end-to-end distance, shown in Fig. 3, are at a temperature $T \approx 330$ K well above the positions of the specific heat peaks. Instead, the peaks in specific heat mark the temperatures where the average radius of gyration as function of temperature has a kink, separating a compact phase at lower temperatures from a phase where the volume of the molecules increases quickly with temperature.

No low-temperature peak is found for the two BBAThet1 chains, but in the case of chain A we see a shoulder at 276 ± 18 K. No clear signal for such a shoulder is observed for chain B. The location of the chain A shoulder corresponds roughly to the low-temperature peak in the specific heat curve of BBAT1 ($T_f \approx 270$ K), which signals for BBAT1 monomers the formation of secondary structure. In order to see whether the shoulder in the specific heat curve of chain A is also related to secondary structure formation, we display in Fig. 4 the average percentage of random-coil configurations and such with a native-like helix as function of temperature. The later is defined as having 90% of the helical contacts found in the PDB-structure. With decreasing temperature, the percentage of random-coil structures decreases rapidly, with the mid-point of the transition around 325 K and 318 K. These temperatures correspond to the ones where the end-to-end distance curves in Fig. 3 have their minimum values. On the other hand, the percentage of configurations with the native-like helix increases rapidly below $T \approx 300$ K, with the mid-points at $T_f = 272 \pm 1$ K for chain a and $T_f = 267 \pm 1$ K for chain B. Here, we define a configuration as having native helicity if it has a helical segment reaching for chain A from Asp7 to Dbz20, and for chain B from Phe8 to Ala19. Note that the temperature values for chain A correspond with the position of the shoulder in the heat capacity curve for this peptide. We also note that the average distance between residues Arg2-Asp7, which is a measure for the formation of the β-hairpin and displayed in the inset of Fig. 3, has a kink at around $T \approx 270$ K, indicating appearance of the hairpin for lower temperatures.

Summarizing the above figures, we find that at high temperatures both peptides are elongated and unstructured. With decreasing temperature, the peptides become more compact, and around $T \approx 330$ K they start to assume a bend-like form. Visual inspection (see also Fig. 3) shows that the bend mostly appears between the 9th and 12th residue. As temperature is lowered further, small helical segments appear. Below $T \approx 320$ K, these segments start to grow from both terminals and finally merge, leading to an increase in end-to-end distance. This is similar to what has been observed earlier for the BBAT1 chains; however, here the process of secondary structure formation is not connected with a signal in the specific

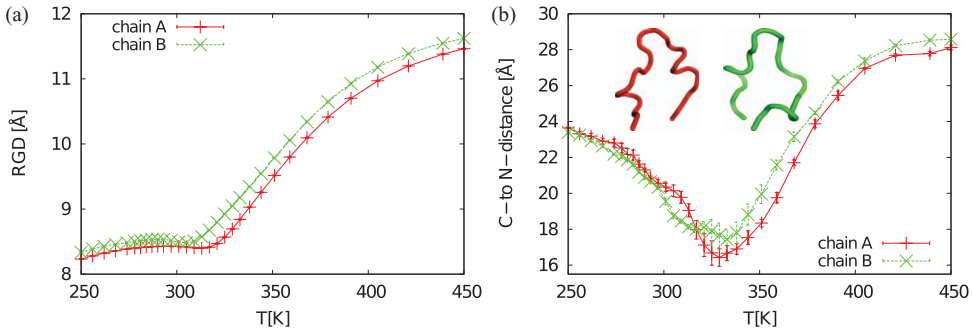

FIG. 3. (a) Average radius of gyration (RGD) and (b) end-to-end distance of isolated BBAThet1 monomers of type A (solid line) and B (dashed line) as function of temperature. Shown are also representative structures of the intermediate (type A, red; B, green).

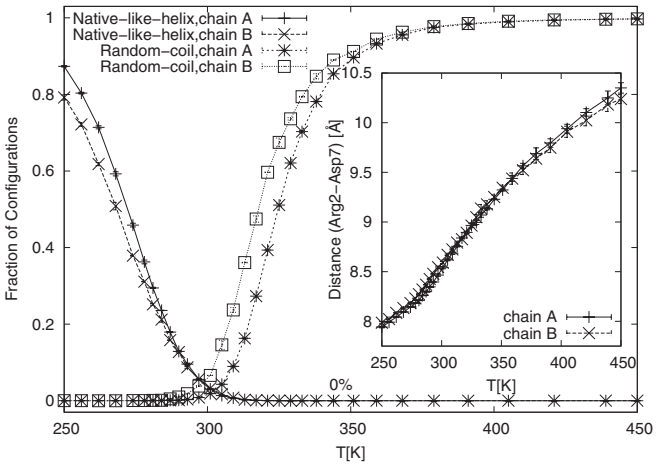

FIG. 4. Percentage of random-coil configurations and such with native-like helix as function of temperature comparing the isolated BBAThet1 monomers A (short dashed, solid line) and B (dotted, dashed line, respectively). The inset displays the average distance between residues Arg2-Asp7, a measure for formation of the N-terminal β-hairpin (type A, solid line; B, dashed line).

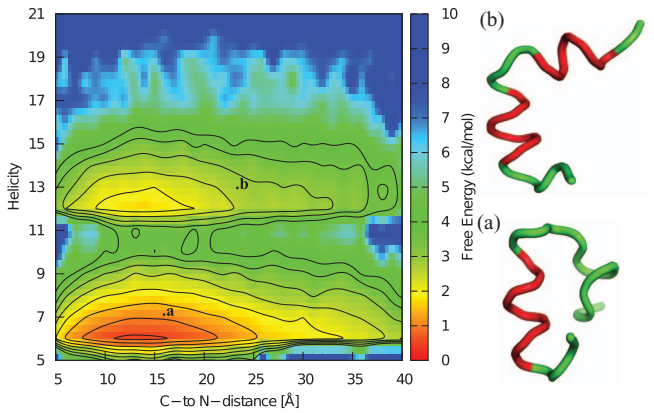

FIG. 5. Free energy landscape of BBAThet1 monomers (chain A) at 300 K, projected on the end-to-end distance and helicity as reaction coordinates. Helical segments are also shown (red) in their representative structures.

heat, i.e., not associated with a strong change in potential energy. Instead, we observed a peak in specific heat at a about 20 K lower temperature where the two chains become maximal compact. We conjecture that at this temperature there is competition between bended configurations with two separated helical segments and such with a single long helix. The percentage of configurations with such native-like elongated helix grows with decreasing temperature, with the mid-point of that transition around 270 K. At this temperature also appears a signal for formation of the N-terminal hairpin.

Assuming that thermal ordering corresponds to temporal ordering we conjecture the following folding mechanism for the isolated BBAThet1 chains. For both peptides, folding starts with chain collapse and formation of short helical segments at both termini leading to bend-like configurations. These helical segments in the intermediates grow together, stretching the bend-like configurations till a fully elongated helix is formed. In a final step, the N-terminal hairpin forms. This picture is also supported by Fig. 5 where we show the free energy landscape of the type A monomer at $T = 300$ K projected on the end-to-end distance and helicity as reaction coordinates. Such projections are often used to obtain a more direct picture of the folding mechanisms. In the present case, it leads to a landscape shaped like a laying H that is consistent with the above proposed folding mechanism.

This folding mechanism is essentially the same two-step mechanism as observed for the earlier studied BBAT1 monomers; however, the changes in energy appear to be weaker, and the two steps do not lead to separate peaks in the specific heat. Hence, while the two-step mechanism appears to be very similar for the various BBA peptides and likely results from the shared fold, the energy changes involved with the two transitions differ with sequence. For BBAT1, they lead to two separated peaks in the specific heat. In the case of chain A of BBAThet1, one of these peaks is reduced to a shoulder, and it disappears completely for chain B.

In a second set of simulations, we then study the folding and association of the BBAThet1 tetramer. Again, we start by monitoring the specific heat for possible transitions. This quantity as function of temperature is shown in Fig. 6 and exhibits one maximum and two shoulders. In order to ensure that the shoulder are not due to insufficient sampling, we have repeated the simulations with two different start configurations of the system (all four chains extended and on a single linear line, the four chains in random configurations and randomly distributed). Both runs lead to the same curve. The peak of the curve is at a temperature of $T_p \approx 323$ K, and the two shoulders at $T_h \approx 420$ K and $T_L \approx 290$ K.

The shoulder at the higher temperature T_h marks the association of the monomers. This can be seen from Fig. 7 where we display the fraction of monomers, dimers, trimers, and tetramers as function of temperature. Our definition of these entities depends on the distance between the center of each chain and mid-point of all centers. If the distance is less than 16 Å for all chains (the corresponding distances in the native structure are ≈10 Å), we

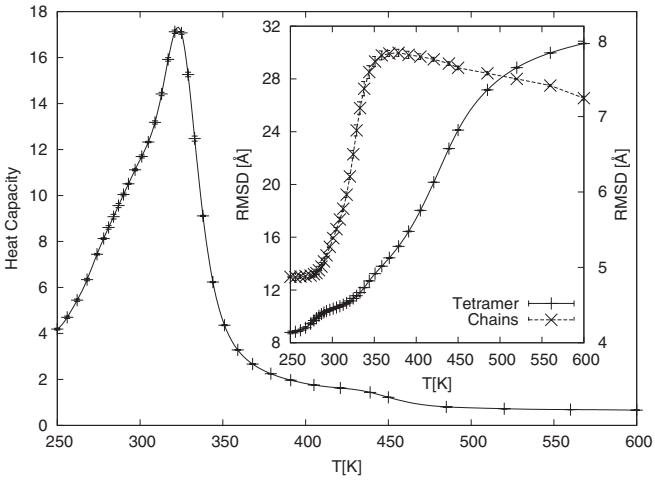

FIG. 6. Heat capacity of the BBAThet1 tetramer system as function of temperature. The inset displays the root-mean-square deviation (RMSD) of the tetramer with respect to the PDB structure (dashed line). The corresponding scale is on the left side. The second curce in the inset (dot dashed line) shows the averaged RMSD of the four component chains individually calculated with respect to their form in the PDB-structure. The corresponding scale is on the right side.

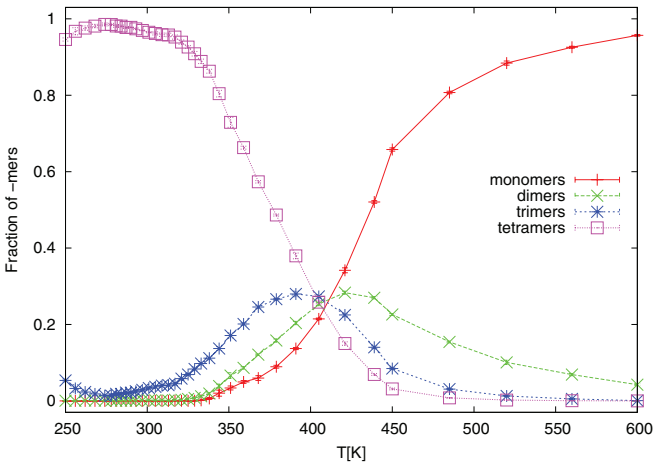

FIG. 7. Fraction of BBAThet1 monomers (solid line), dimers (dashed line), trimers (short dashed line), and tetramers (dotted line) as function of temperature.

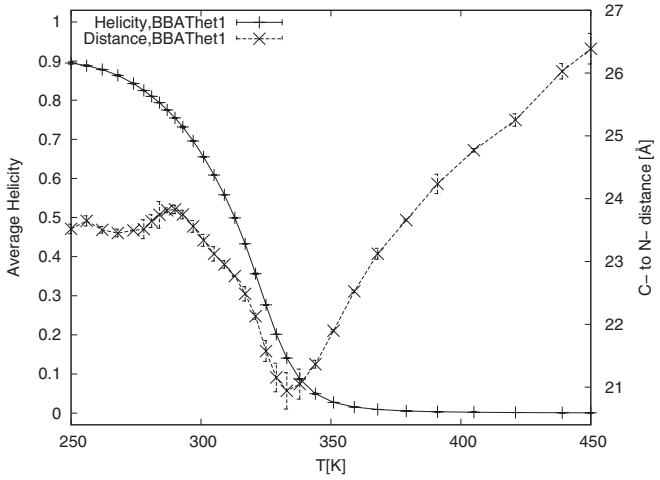

FIG. 8. Average end-to-end distance (dashed line) and average fraction of helical residues (solid line) of the four BBAThet1 chains as function of temperature.

define the configuration as a tetramer. If the distance is larger for one chain, it defines a trimer, if two a dimer. At high-temperatures, only monomers are observed. With decreasing temperatures, the fraction of monomers decreases and oligomers form. Dimers form first, and their rate reaches a maximum around 420 K. Trimers appear with at least 9% at 450 K, and reach a maximum value in the neighborhood of $T = 390$ K. On the other hand, tetramers appear only at around 450 K, but begin to dominate as temperature decreases. Hence, the shoulder in the specific heat between ≈ 400 and 480 K corresponds to the temperature region where association of BBAThet1 chains start. Note that this is also the temperature region where for BBAT1 association of the chains starts.

The fraction of BBAThet1 tetramers approaches a plateau at around 320 K, i.e., at temperatures where the specific heat has its peak. This suggests that once tetramers are formed, a second transition happens. This can be seen also from the inset of Fig. 6 where we show the average root-mean-square deviation of tetramer configurations with respect to the PDB-structure. Note that calculation of this value takes into account the symmetry of the structure with all its 24 possible permutations. The rapid decrease of this quantity in the region of the shoulder in specific heat mirrors the increase in the fraction of tetramers in Fig. 7, hence, this decrease in the "global" RMSD reflects the association of the chains to a tetramer. Once the tetramer fraction reaches its plateau, the decrease in RMSD slows down. A complementary picture appears into the average root-mean-square-deviation of the four chains (calculated for each chain separately), also shown in the inset of Fig. 6. From the change of this "local" RMSD, one sees that the four chains are unfolded when they associate, and start folding only when the tetramer is formed.

This is similar to the mechanism that we have observed in earlier work for BBAT1. For that protein folding of the four identical chains follows the same mechanism as observed for the isolated monomers. Figures 8 and 9 show that this is also the case for BBAThet1. The peak in specific heat at $T \approx 320$ K corresponds to the first step in the folding of the four chains. As in the case of the isolated monomers this is the chain collapse and formation of separated helical segments. This process leads to the increase in the average helicity of the chains and the minimum in average end-to-end distance seen in Fig. 8. Note that these figures represent averages over all four chains, i.e., neglect the differences between chains of type A and such of type B. We present separate curves for the behavior of the two peptides in the tetramer environment in Figs. S1 and S2 of the supplementary material, see Ref. 34. The minimum is located at around 330 K while the mid-point of the steep increase in helicity is around 320 K. Both values are around the position of the peak of the specific heat, indicating that this transition is mainly energy-driven. Lowering temperature further, the separated small helices in the bend-like configurations elongate and merge, leading to the observed increase in end-to-end distance. Note that this process starts earlier for the chains of type A than for the

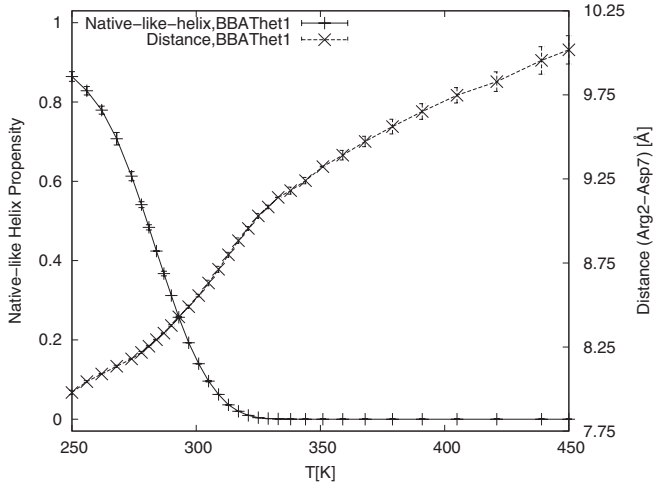

FIG. 9. Percentage of the chains in the BBAThet1 tetramer simulation that have native-like helix chains (solid line); and average distance between residues Arg2-Asp7 in these chains (dashed line). The later quantity is a measure for formation of the N-terminal β-hairpin. Both quantities are shown as function of temperature.

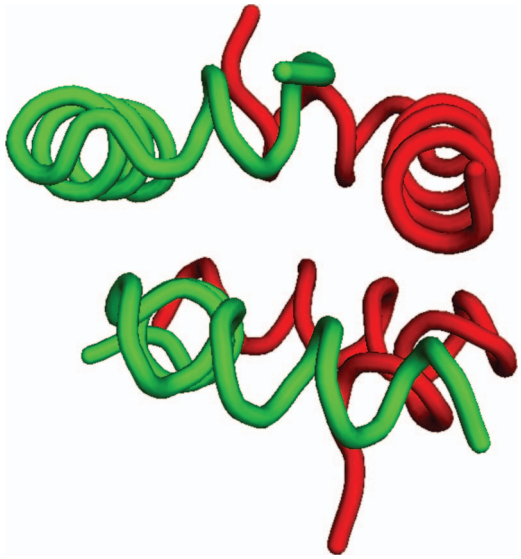

FIG. 10. Representative BBAThet1 tetramer structures as found in our simulation at $T = 250$ K. The two types of chain are drawn in different colors (red and green).

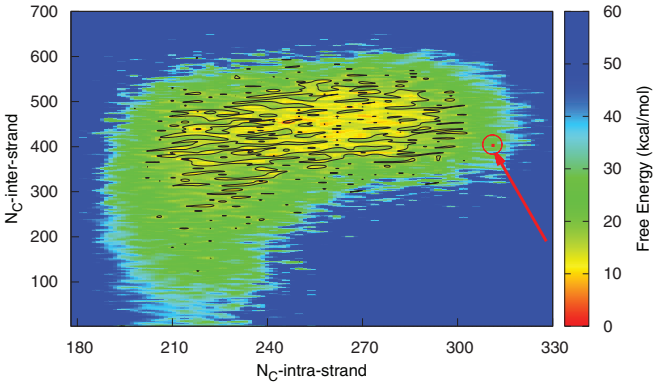

FIG. 11. Free energy landscape of the BBAThet1 tetramer at 300 K, projected on the number of intra-strand and inter-strand contacts as reaction coordinates. The position of the native structure (Fig. 1) is marked by a circle to which a red arrow points.

ones of type B (see Figs. S1 and S2 of the supplementary material[34]).

As a consequence, the percentage of configurations with native-like helix, shown in Fig. 9, increases rapidly around $T_L \approx 290$ K, i.e., in the temperature region where the specific heat has its low-temperature shoulder. The presented data are again averages over all four chains. Separate curves for chain A and chain B peptides are shown in Figs. S3 and S4 of the supplementary material, see Ref. 34. The average distance between residues Arg2 and Asp7 indicates that this temperature region also corresponds to the one where N-terminal β-hairpins form. A randomly drawn (i.e., representative) tetramer configuration with all four BBAThet1 chains fully folded, as observed at our lowest temperature $T = 250$ K, is shown in Fig. 10. This structure has a root-mean-square deviation of 8 Å to the native structure of Fig. 1 with the rmsd resulting from helix-bending by the D-Alanine at position 9 and exaggerating the differences: apart from this single mismatch are the secondary and tertiary structure of the protein correctly reproduced. The energy of this structure is with −557 kcal/mol comparable to that of the relaxed PDB-structure (−542 kcal/mol). These energies demonstrate that the force field needs further improvement to increase its resolution, for instance by refining the valence bending potentials, as was previously also noticed in Ref. 35.

The above thermal ordering suggests that BBAThet1 folds and associates in a three-step mechanism where association of unfolded chains precedes folding of the four individual chains. Folding of the chains starts after association by a collapse into a bend-like structure with nascent helical segments that in the final step grow together into the native helix while at the same time the N-terminal β-hairpin is formed. The association into the tetramer generates an environment that facilitates the folding of the four chains: the transition temperatures for collapse transition and secondary structure formation are about 20 K higher than observed for the isolated monomers. The relation between tetramers association and folding of the four component chains can be also seen directly in Fig. 11 where we project the free energy landscape of the tetramer at $T = 300$ K on the number of intra-strand and inter-strand contacts as reaction coordinates. The form of the landscape shows that first the number of contacts between strands grows unto to about 400 contacts, i.e., the chains associate to the tetramer, while the number of contacts with the strands changes little around a value of ≈220 contacts (note that this number is for all four chains together). Once the number of contacts between strands reaches this values of ≈400 contacts, does also the number of intra-chain contacts increases, indicating folding of the four chains. Lowering further the temperature, the global minimum in free energy moves accordingly to the right until it includes the marked position of the native structure (as shown in Fig. 1).

In an experimental setting, one usually has a large number of chains that interact and fold. Because of limited computational resources, we have studied in the present investigation only the association and folding of four chains into a tetramer, and therefore we cannot exclude the possibility that our results are distorted by the finite (and minimal) number of chains, a common problem in the simulation of multimeric proteins or protein aggregates. For instance, having a larger number of chains could lead in principle to oligomers with more than four chains, i.e., hexamers, octamers, and so on. However, formation of such oligomers is unlikely as the tetramer leaves few hydrophobic residues exposed. It is therefore not probable that larger oligomers are formed, or if formed that they would be stable. Indeed, in test runs we could see that hexamers did dissolve in a short time at $T = 300$ K. While such runs can only provide anecdotal evidence (as their outcome depends on the assumed form of the hexamer), it appears to be likely that simulations with a larger number of chains would not lead to formation of oligomers larger than tetramers. A second possibility is that given more than two chains of type A and two of type B, not only heterotetramer are formed but also homotetramers of type AAAA or BBBB. Again, we believe that this is an unlikely scenario. This is because we observe in our simulations a higher frequency of AB dimers over AA or BB dimers. At

450 K (the temperature where the abundance of dimers is highest, see Fig. 7), we find that 58% of dimers are of type AB, 7% of type AA, and 35% of type BB. Hence, our dimer frequencies indicate that the interaction between chains of different kinds is preferred over that of the same kind. Indeed, we needed to go to higher temperatures to dissolve the heterotetramer than needed to dissolve either of the two mono tetramers. Homotetramers of type BBBB decay already at $T \approx 400$ K, while we had to raise temperature to $T = 405$ K to dissolve homotetramers of type AAAA. On the other hand, 75% of the heterotetramers stayed stable during the 3 ns trajectories (which correspond to 3 μs in all-atom simulations[36]) at this temperature. These differences in "melting" temperature suggest that the heterotetramer is more stable than the homotetramer, and therefore more likely observed in experiments. Hence, our stability analysis indicates that our main simulation results are not qualitatively distorted by the small number of chains.

IV. CONCLUSIONS

While the Anfinsen experiments[37] suggest that biologically active proteins have a single defined structure that (under physiological conditions) depends solely on its sequence of amino acids, research over the last decades increasingly shows that the structure of proteins depends also on the interaction with other biomolecules. Hence, formation of protein complexes and oligomers is not necessarily a hierarchical process where the individual proteins fold first, and tertiary structure determines quartinary structure. Their complexity and time scales renders an experimental investigation of these association and folding processes difficult. A valid alternative are computer simulations of small designed model systems. In the present paper, we report results from coarse-grained replica exchange molecular dynamics simulations of the 84-residue heterotetramer BBAThet1, where each of the four chains takes a $\beta\beta\alpha$-fold. In order to understand the sequence-dependence of our results, we compare them with our earlier study of the homotetramer BBAT1 that has a similar structure. In both cases, we find that the protein complex takes its final form by a mechanism where assembly into a tetramer of unfolded chains generates an environment that facilitates the later folding of the four chains. Folding of the chains follows the same mechanism as for isolated monomers. Both BBAThet1 and BBAT1 peptides fold first into a bend-like intermediate with little secondary structure. Nascent helical segments are observed at both terminals. In a second step, these helical segments grow together, elongating the configurations, and the N-terminal β-hairpin forms. The details of the folding of the chains and the relative strength of the two steps is sequence dependent. For the BBAT1 chain, the steps are clearly separated and lead to pronounced peaks in the specific heat, for chain A of BBAThet1 the low-temperature peak is reduced to a shoulder and disappears for the chain B of BBAThet1. However, in all cases the transition temperatures are in the tetramer environment raised by about $\approx$20 K over that observed in folding of isolated monomers. Comparing homo and heterotetramer, we conjecture that folding promoted by association is a common mechanism in the formation of protein oligomers and complexes with $\beta\beta\alpha$-fold. This mechanism might be shared with other protein oligomers, and it may also be the mechanism by that intrinsically unfolded proteins assume their functional structure. We believe that our data give evidence for the need to generalize the common single funnel picture of protein folding by one where the folding landscape (with potentially more than one funnel) is modulated by environment and interaction with other proteins.

ACKNOWLEDGMENTS

This work was supported, in part, by research Grant No. GM62838 of the National Institutes of Health (USA), by Hacettepe University Scientific Research Fund under project number 012.D12.602.001. F.Y. thanks the Department of Chemistry and Biochemistry for kind hospitality during his sabbatical stay at the University of Oklahoma. A.K.S. is supported by the Government of the Republic of Poland from budget funds for science in years 2013-2014 within project 0558/IP3/2013/72 (in a part multichain development and testing in UNRES force field) and by Foundation for Polish Science (in part of introduction of non-standard amino acid).

[1] O. Keskin, A. Gursoy, B. Ma, and R. Nussinov, Chem. Rev. **108**, 1225 (2008).
[2] A. K. Sieradzan, A. Liwo, and U. H. E. Hansmann, J. Chem. Theory Comput. **8**, 3416 (2012).
[3] W. M. Berhanu, P. Jiang, and U. H. E. Hansmann, Phys. Rev. E **87**, 014701 (2013).
[4] M. H. Ali, E. Peisach, K. N. Allen, and B. Imperiali, Proc. Natl. Acad. Sci. U.S.A. **101**, 12183 (2004).
[5] M. H. Ali, C. M. Taylor, G. Grigoryan, K. N. Allen, B. Imperiali, and A. E. Keating, Structure **13**, 225 (2005).
[6] O. Zimmermann and U. H. E. Hansmann, Biochim. Biophys. Acta, Proteins Proteomics **1784**, 252 (2008).
[7] Z. Zhang, K. Y. Sanbonmatsu, and G. A. Voth, J. Am. Chem. Soc. **133**, 16828 (2011).
[8] C. J. Geyer and E. A. Thompson, J. Am. Stat. Assoc. **90**, 909 (1995).
[9] K. Hukushima and K. Nemoto, J. Phys. Soc. Jpn. **65**, 1604 (1996).
[10] U. H. E. Hansmann and Y. Okamoto, Phys. Rev. E **54**, 5863 (1996).
[11] A. Liwo, M. Khalili, and H. A. Scheraga, Proc. Natl. Acad. Sci. U.S.A. **102**, 2362 (2005).
[12] A. Liwo, M. Khalili, C. Czaplewski, S. Kalinowski, S. Ołdziej, K. Wachucik, and H. A. Scheraga, J. Phys. Chem. B **111**, 260 (2007).
[13] C. Czaplewski, S. Kalinowski, A. Liwo, and H. A. Scheraga, J. Chem. Theory Comput. **5**, 627 (2009).
[14] A. K. Sieradzan, U. H. E. Hansmann, H. A. Scheraga, and A. Liwo, J. Chem. Theory Comput. **8**, 4746 (2012).
[15] S. Ołdziej, A. Liwo, C. Czaplewski, J. Pillardy, and H. A. Scheraga, J. Phys. Chem. B **108**, 16934 (2004).
[16] S. Ołdziej, J. Łagiewka, A. Liwo, C. Czaplewski, M. Chinchio, M. Nanias, and H. A. Scheraga, J. Phys. Chem. B **108**, 16950 (2004).
[17] A. V. Rojas, A. Liwo, and H. A. Scheraga, J. Phys. Chem. B **111**, 293 (2007).
[18] P. C. Whitford, J. K. Noel, S. Gosavi, A. Schug, K. Y. Sanbonmatsu, and J. N. Onuchic, Proteins: Struct., Funct. Bioinform. **75**, 430 (2009).
[19] J. K. Noel, P. C. Whitford, K. Y. Sanbonmatsu, and J. N. Onuchic, Nucleic Acids Res. **38**, W657 (2010).
[20] A. Schug. P. C. Whitford, Y. Levy, and J. N. Onuchic, Proc. Natl. Acad. Sci. U.S.A. **104**, 17674 (2007).
[21] D. E. Shaw, P. Maragakis, K. Lindorff-Larsen, S. Piana, R. O. Dror, M. P. Eastwood, J. A. Bank, J. M. Jumper, K. K. Salmon, Y. Shan, and W. Wriggers, Science **330**, 341 (2010).
[22] A. Schug, W. Wenzel, and U. H. E. Hansmann, J. Chem. Phys. **122**, 194711 (2005).

[23]D. Paschek, H. Nymeyer, and A. E. Garcia, J. Struct. Biol. **157**, 524 (2007).
[24]R. B. Best, G. Hummer, and W. A. Eaton, Proc. Natl. Acad. Sci. U.S.A. **110**, 17874 (2013).
[25]Y. M. Rhee and V. S. Pande, Biophys. J. **84**, 775 (2003).
[26]A. R. Mezo, J. J. Ottesen, and B. Imperiali, J. Am. Chem. Soc. **123**, 1002 (2001).
[27]K. A. McDonnell and B. Imperiali, J. Am. Chem. Soc. **124**, 428 (2002).
[28]W. Nadler, J. H. Meinke, and U. H. E. Hansmann, Phys. Rev. E **78**, 061905 (2008).
[29]H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak, J. Chem. Phys. **81**, 3684 (1984).
[30]F. Rakowski, P. Grochowski, B. Lesyng, A. Liwo, and H. A. Scheraga, J. Chem. Phys. **125**, 204107 (2006).
[31]E. Rosta, N. V. Buchete, and G. Hummer, J. Chem. Theory Comput. **5**, 1393 (2009).
[32]A. Mor, G. Ziv, and Y. Levy, J. Comput. Chem. **29**, 1992 (2008).
[33]S. Kumar, D. Bouzida, R. H. Swendsen, P. A. Kollman, and J. M. Rosenberg, J. Comput. Chem. **13**, 1011 (1992).
[34]See supplementary material at http://dx.doi.org/10.1063/1.4868140 for supplemental figures of various quantities as function of temperature for the four chain in the tetramer simulation.
[35]Y. He, M. A. Mozolewska, P. Krupa, A. K. Sieradzan, T. K. Wirecki, A. Liwo, K. Kachlishvili, S. Rackovsky, D. Jagiela, R. Ślusarz, C. R. Czaplewski, A. Ołdziej, and H. A. Scheraga, Proc. Natl. Acad. Sci. U.S.A. **110**, 14936 (2013).
[36]M. Khalili, A. Liwo, A. Jagielska, and H. A. Scheraga, J. Phys. Chem. B **109**, 13798 (2005).
[37]C. B. Anfinsen, Science **181**, 223 (1973).
[38]PyMOL(TM) Molecular Graphics System, Version 1.2r2, DeLano Scientific LLC, Palo Alto, CA, 2009.