Folding and association of a homotetrameric protein complex in an all-atom Go model

W. M. Berhanu, P. Jiang, and U. H. E. Hansmann
Department of Chemistry and Biochemistry, University of Oklahoma, Norman, Oklahoma 73019-5251, USA
(Received 14 September 2012; revised manuscript received 12 November 2012; published 11 January 2013)

The 84-residue homotetrameric BBAT1 is one of the smallest stable protein complexes and therefore is a good test system to study the self-assembly of multimeric proteins. We have researched for this protein the interplay between the folding of monomers and their assembly into tetramers. Replica exchange molecular dynamics simulations relying on a Go model are compared with earlier simulations that use the physics-based coarse-grained UNRES model.

DOI: 10.1103/PhysRevE.87.014701 PACS number(s): 87.15.Cc

I. INTRODUCTION

Experimental and computational studies have led to considerable insights into the principles of folding of small monomeric proteins [1]. However, the majority of proteins are built out of multiple subunits [2]. Despite their significance in controlling many biological processes [3,4], the folding and unfolding and association pathways for oligomeric proteins are much less understood than that of monomeric proteins. In this Brief Report we investigate the folding and association of the homotetrameric BBAT1 (Protein Data Base identifier 1SN9) [5] and how the observed mechanism depends on the energy functions.

BBAT1, shown in Fig. 1, is one of the simplest models of multisubunit proteins. The protein is built from four identical chains of 21 amino acids, each folding into a $\beta\beta\alpha$ motif. Even for such small protein complexes all-atom molecular dynamics simulations of folding and association are a challenge. This is because it requires long time scales (milliseconds to microseconds) to observe folding events, and the utilized force field must correctly describe the relative energies of a wide range of conformations during the folding process [7,8].

Enhanced sampling and coarse graining are two approaches to access the time scale of folding events [8,9]. Another possibility is the use of Go models [10], which introduce structural data in the energy functions by favoring contacts that appear in the native configuration [10]. This bias reduces the complexity of the resulting energy landscape to a perfect funnel with only residual energetic frustration caused by nonnative interactions [11,12]. While such structure-based models have helped us to understand the folding mechanisms of many proteins [13,14], there are cases when they are known to fail [15–17]. This is because in Go models intermediate states involving non-native contacts may be suppressed.

The question of whether Go models can describe the folding and association of the more complex oligomeric proteins then arises. In order to study this question we investigate in this Brief Report the folding and association pathway of BBAT1 using the all-atom Go model proposed by Whitford *et al.* [18] and compare our results with those of simulations [19] relying on the physics-based UNRES force field [20].

II. METHODS AND SIMULATION SETUP

Our simulations rely on the Structure-based Models in GROMACS (SMOG), developed by the Onuchic group [10]. Using the SMOG@ctbp Web tool located at http://smog.ucsd.edu, we prepared topology and coordinate files [18] of the BBAT1 tetramer that were then employed in molecular dynamics simulations with the GROMACS 4.5.5 software package [21]. Note that it is important to take into account the symmetry of the target structure by including all permutations of the native contacts. The simulation starts from a completely extended conformation to avoid bias toward the native state. The chains are placed at random and separated by at least the length of the extended chain in a cubic box with a box size of 100 Å and hard walls. Langevin dynamics with a time step of 0.5 fs is used, and configurations are saved every 0.5 ps. Note that the energy function is nonphysical, and therefore temperatures are given in reduced units u. In order to increase sampling efficiency, we rely on replica exchange molecular dynamics. Thirty-two replicas are used, with the temperatures spaced in a geometrical distribution between $170u$ and $200u$. Each replica is simulated for 160 ns, leading to a total simulation time of 5.1 μs. The simulation of the isolated BBAT monomer is done using the same protocol as in the tetramer above with 32 replicas and a temperature distribution between $90u$ and $220u$. All simulations start from a completely extended conformation, and each replica is simulated to 160 ns with a total simulation time of 5.1 μs. We display in Fig. 2 the random walk in the temperature space for one of the 32 replicas of the tetramer simulation. As in the other 31 replicas, it explores the full spectrum of temperatures, moving in the course of the simulation from the lowest to highest temperature and back, leading in this way to increased sampling of low-energy structures at the lowest temperature. The constant volume heat capacity as a function of temperature plotted in Fig. 3 was calculated for three different time intervals. As the three curves overlap within the errors (data not shown), we conclude that our simulation has converged.

III. RESULTS

The heat capacity as a function of temperature in Fig. 3 has a pronounced peak at $210u$. This peak corresponds to the temperature where the average root-mean-square deviation (RMSD) to the crystal structure and the radius of gyration, a measure for the compactness of configurations, steeply decrease; see Fig. 4. Hence, this temperature marks the folding temperature of BBAT1. It also corresponds to the association temperature where the fraction of tetramers with RMSD to a

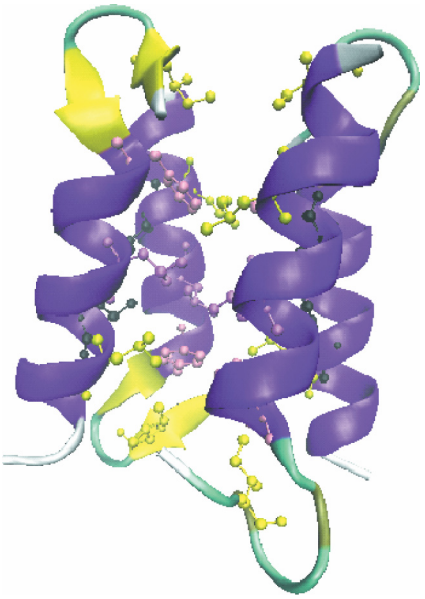

FIG. 1. (Color online) The crystal structure of BBAT1 (Protein Data Base identifier 1SN9), prepared using VMD [6].

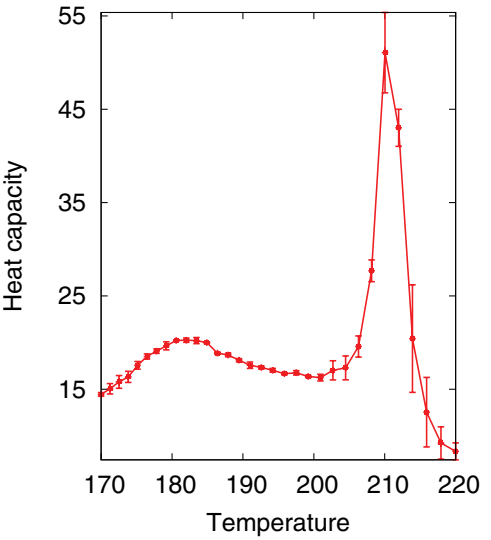

FIG. 3. (Color online) Heat capacity C_V as a function of temperature for the BBAT1 tetramer. The error bars are from a comparison between the averages taken over three time intervals (40–80, 80–120, and 120–160 ns).

crystal structure of less than 2 Å is approximately 0.5 (Fig. 5). The association of tetramers is accompanied by the steep decrease of the monomer population, whereas the populations of dimers and trimers are much lower (less than 0.1), indicating the strong cooperativity of the process.

Note that the heat capacity curve of Fig. 3 has a second broader and shallower peak in the temperature range $170u$–$195u$. This peak is related to the folding of the monomers. Within that temperature range intrachain native contacts grow from below 0.4 to 1, but only a small part of this increase (from 0.16 to 0.35 for helical contacts) is associated with the formation of tetramers (Fig. 6). In contrast, interchain native contacts grow by 0.5 within a small temperature change around $210u$ but only increase slowly afterwards. Therefore, the peak around $170u$ to $195u$ marks the secondary structure formation in the monomers, including their N-terminal hairpin and C-terminal helix.

Figures 6 and 7 reveal that the thermodynamics of monomer folding is a multiple-step process. Upon association of the monomers to tetramers (around $220u$), the number of non-native contacts decreases, and the end-to-end distance increases strongly with decreasing temperature. This indicates that association with a tetramer leads to a straightening of the monomer chains (resulting in a larger end-to-end distance and fewer non-native intrachain contacts). Only when the temperature is lowered further, at around $170u$–$195u$, do the monomers fold and assume their secondary structure (Fig. 6), which in turn leads to a shorter end-to-end distance (Fig. 7). This is in agreement with concentration and temperature-dependent CD spectroscopy data [22], which show a significant negative ellipticity around 222 nm at low temperature due to folding of

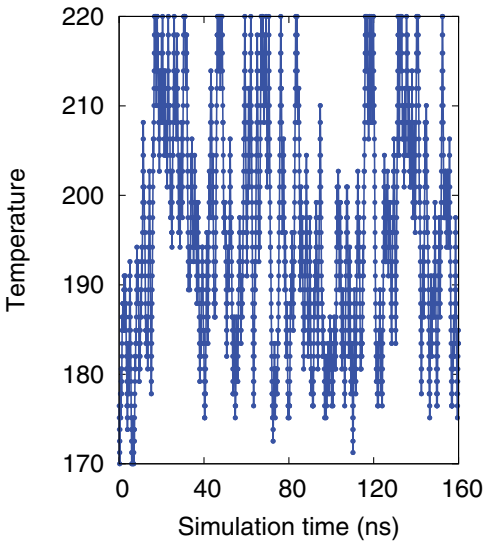

FIG. 2. (Color online) Walk of a single replica through temperature space for the BBAT1 tetramer.

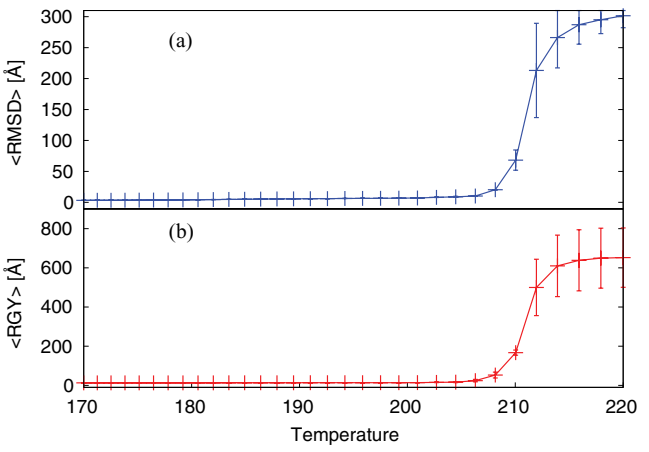

FIG. 4. (Color online) (a) Average root-mean-square deviation (RMSD) with respect to the native structure of the BBAT1 tetramer and (b) radius of gyration (RGY) as a function of temperature for the BBAT1 tetramer.

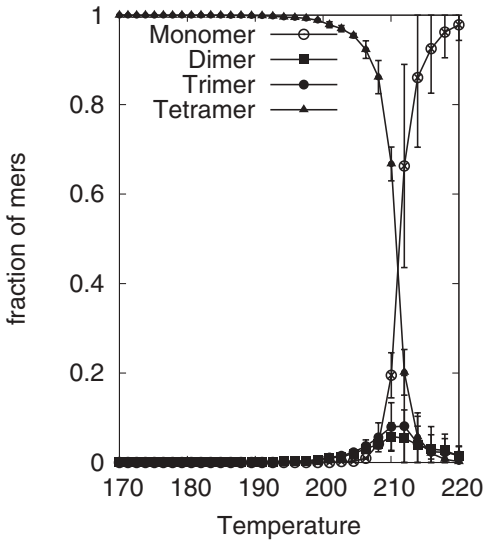

FIG. 5. Fraction of monomers, dimers, trimers, and tetramers as a function of temperature for BBAT1 (all four chains).

the peptide chain into an α-helical rich structure and a loss of α-helical content with increasing temperature [22].

The above folding and association process differs from what is observed in simulations [19] that rely on the physics-based coarse-grained UNRES force field [20]. In UNRES simulations [19], the folding of the monomers is also preceded and aided by association with a tetramer, but the mechanism by which the individual chains fold differs. Upon association, the monomers do not straighten out but at a lower temperature form instead a compact intermediate state with little secondary structure and a small end-to-end distance. The transition toward this intermediate state is marked by a peak in specific heat. A second transition from the intermediate state toward the native state at an even lower temperature leads also to a second peak in specific heat and is marked by an increase

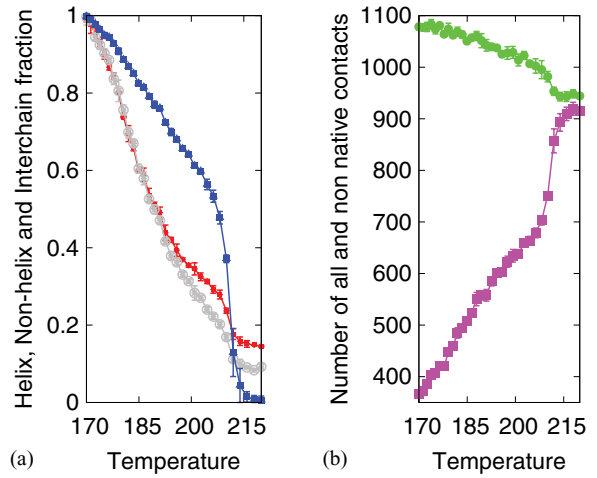

FIG. 6. (Color online) Fractional native contacts and number of all and non-native contacts as function of temperature for the BBAT1 tetramer. (a) Fraction of helical [red (medium gray) line], nonhelical (light gray line), and interchain contacts [blue (dark gray) line] vs temperature and (b) number of total contacts (top line) and only non-native contacts (bottom line) vs temperature.

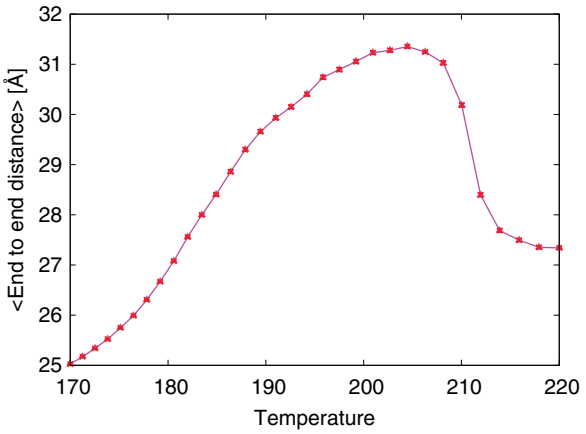

FIG. 7. (Color online) Temperature dependence of the average end-to-end distance of BBTA chains for the BBAT1 tetramer (in Å).

in end-to-end distance and secondary structure. Unlike in our Go-model simulations, one observes in UNRES simulations not two but three transitions: (1) association of the chains with a tetramer, (2) collapse of each chain into a compact intermediate, and (3) secondary structure formation and folding from this intermediate form into the native structure. Hence, a critical intermediate state is not sampled in Go-model simulations. One possible explanation for this difference is that the intermediate involves non-native contacts against which the Go-model energy function introduces a bias. This is a problem that has been observed earlier for monomeric proteins [15,16,23,24]. On the other hand, we like to emphasize that the essential elements of the folding and association process observed in the UNRES simulations are also seen in Go-model simulations: association and folding of their individual chains are separate processes, with the tetramer seeming to provide an environment that aids folding of the monomers.

Interestingly, the folding of the isolated monomer in our Go-model simulations resembles the mechanism observed in the UNRES simulations of both isolated monomers and tetramers. This can be deduced from Fig. 8, where we plot heat capacity, average RMSD, and average end-to-end distance as obtained in all-atom Go-model simulations of the isolated monomer. Note that in our Go model temperatures depend on the system, and therefore absolute temperature values cannot be compared between the isolated monomer and the tetramer. The folding of the monomer as indicated by a pronounced drop in root-mean-square deviation is marked by a peak in specific heat. This peak is also related to a minimum in end-to-end distance. This suggests that the monomer collapses first into a bendlike structure with small end-to-end distance. Upon further lowering the temperature, the isolated BBAT1 monomer stretches out of this compact intermediate state (characterized by small secondary structure and short end-to-end distance), increasing its end-to-end distance due to the formation of the helix and β turn (Fig. 8). The same process is observed in UNRES simulations of the isolated monomer, with the difference that the two steps lead there to two peaks in specific heat. This indicates that successful (all-atom) Go-model simulations of monomers do not necessarily guarantee that these models are also adequate

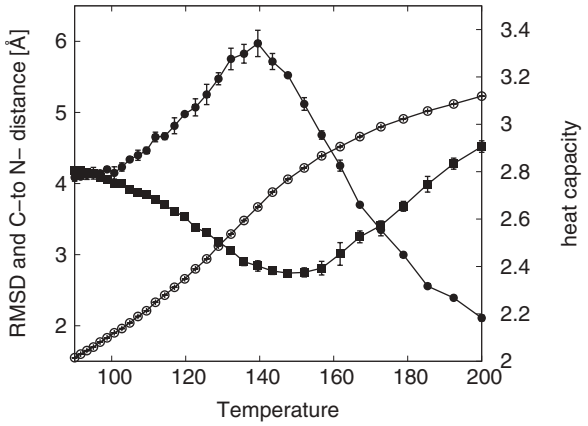

FIG. 8. End-to-end distance (squares), root-mean-square deviation (open circles), and heat capacity (solid circles) as a function of temperature for an isolated BBAT monomer. A constant factor of 18 Å is subtracted from the end-to-end distance (squares) in order to allow for a better comparison of all three curves.

for simulating the oligomeric system. Note that the monomer folding mechanism is consistent with experimental work and high-temperature unfolding molecular dynamics simulations in Refs. [25,26].

IV. DISCUSSIONS AND CONCLUSIONS

Go models, which favor the pairwise contacts in the native conformation, offer a way to circumvent the problem of slow sampling in protein folding simulations. While work so far has focused mostly on monomeric proteins, we have tested in the present Brief Report these models for the purpose of investigating folding and association of multimeric proteins. Our test case is the homotetrameric 84-residue BBAT1, chosen because of its small size. Comparing our results with those relying on earlier simulations [19] using the physics-based coarse-grained force field UNRES, we find that all-atom Go-model simulations reproduce the most important features of the folding mechanism observed also in the UNRES simulations. Both kinds of simulations predict that the association of the four chains with a tetramer precedes and aids the folding of the individual chains. The two models differ in the details. Both models lead, for the isolated monomer, to the same folding mechanism, which is consistent with experimental data. However, the Go model leads, for the tetrameric system only, to an abbreviated folding mechanism for the individual chains that omits an important intermediate state. Hence, the three-step process of folding and association observed in UNRES simulations is reduced to two steps. As a direct comparison with experimental results is lacking, it is not clear which of the two models describes correctly the folding and association process of BBAT1. However, the discrepancy between the two models shows that even if a Go model can reproduce the folding mechanism of a monomeric unit, it is not guaranteed that it also describes correctly the oligomer. In the present case, we conjecture that the differences in the folding mechanisms between the two models are due to non-native contacts appearing in the intermediate state that are suppressed in Go-model simulations. Similar differences in folding mechanisms have been observed earlier for monomeric proteins and can be explained with the same reasoning [15,16,23,24]. Hence, as with monomeric proteins, Go-model simulations are valuable tools but have to be used with care when exploring folding and the association mechanism of oligomeric proteins [27,28]. Our results suggest that to observe the details of the folding and association process of multimeric proteins and protein complexes in Go-model simulations, one may need to include sequence dependent non-native interactions. A statistics-based interresidue potential [29,30] may bridge the gap and help to sample such hydrophobicity-driven non-native contacts. Work is underway to study the usefulness of such an approach.

ACKNOWLEDGMENTS

We acknowledge support from the National Institutes of Health (Grant No. GM62838). This research used resources of the National Energy Research Scientific Computing Center, which is supported by the Office of Science of the US Department of Energy under Contract No. DE-AC02-05CH11231. We thank Adam K. Sieradzan for helpful discussions.

[1] J. A. O. Rumfeldt, C. Galvagnion, K. A. Vassall, and E. M. Meiering, Prog. Biophys. Mol. Biol. **98**, 61 (2008).
[2] M. J. Bennett, M. P. Schlunegger, and D. Eisenberg, Protein Sci. **4**, 2455 (1995).
[3] A. F. Mehl, K. Okada, S. M. Dehn, and S. Kurian, Biochem. Bioph. Res. Commun. **420**, 635 (2012).
[4] A. R. Mezo, R. P. Cheng, and B. Imperiali, J. Am. Chem. Soc. **123**, 3885 (2001).
[5] M. H. Ali, E. Peisach, K. N. Allen, and B. Imperiali, Proc. Natl. Acad. Sci. USA **101**, 12183 (2004).
[6] W. Humphery, A. Dalk, and K. Schulten, J. Mol. Graphics **114**, 33 (1996).
[7] R. O. Dror, R. M. Dirks, J. P. Grossman, H. Xu, and D. E. Shaw, Annu. Rev. Biophys. **41**, 429 (2012).
[8] P. L. Freddolino, C. B. Harrison, Y. X. Liu, and K. Schulten, Nat. Phys. **6**, 751 (2010).
[9] U. H. E. Hansmann, Chem. Phys. Lett. **281**, 140 (1997).
[10] J. K. Noel, P. C. Whitford, K. Y. Sanbonmatsu, and J. N. Onuchic, Nucleic Acids Res. **38**, W657 (2010).
[11] P. C. Whitford, K. Y. Sanbonmatsu, and J. N. Onuchic, Rep. Prog. Phys. **75**, 076601 (2012).
[12] J. D. Bryngelson, J. N. Onuchic, N. D. Socci, and P. G. Wolynes, Proteins **21**, 167 (1995).
[13] C. Clementi, H. Nymeyer, and J. N. Onuchic, J. Mol. Biol. **298**, 937 (2000).
[14] G. H. Zuo, J. Wang, and W. Wang, Proteins **63**, 165 (2006).
[15] Z. Zhang and H. S. Chan, Proc. Natl. Acad. Sci. USA **107**, 2920 (2010).

[16] N. Koga and S. Takada, J. Mol. Biol. **313**, 171 (2001).
[17] C. T. Friel, G. S. Beddard, and S. E. Radford, J. Mol. Biol. **342**, 261 (2004).
[18] P. C. Whitford, J. K. Noel, S. Gosavi, A. Schug, K. Y. Sanbonmatsu, and J. N. Onuchic, Proteins **75**, 430 (2009).
[19] A. K. Sieradzan, A. Liwo, and U. H. E. Hansmann, J. Chem. Theory Comput. **8**, 3416 (2012).
[20] A. Liwo, M. Khalili, and H. A. Scheraga, Proc. Natl. Acad. Sci. USA **102**, 2362 (2005).
[21] D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, and H. J. Berendsen, J. Comput. Chem. **26**, 1701 (2005).
[22] K. A. McDonnell and B. Imperiali, J. Am. Chem. Soc. **124**, 428 (2002).
[23] S. S. Plotkin, Proteins **45**, 337 (2001).
[24] W. L. Treptow, M. A. Barbosa, L. G. Garcia, and A. F. Pereira de Araújo, Proteins **49**, 167 (2002).
[25] M. H. Ali, C. M. Taylor, G. Grigoryan, K. N. Allen, B. Imperiali, and A. E. Keating, Structure **13**, 225 (2005).
[26] L. Wang, Y. Duan, R. Shortle, B. Imperiali, and P. A. Kollman, Protein Sci. **8**, 1292 (1999).
[27] Y. Levy, A. Caflisch, J. N. Onuchic, and P. G. Wolynes, J. Mol. Biol. **340**, 67 (2004).
[28] S. Yang, S. S. Cho, Y. Levy, M. S. Cheung, H. Levine, P. G. Wolynes, and J. N. Onuchic, Proc. Natl. Acad. Sci. USA **101**, 13786 (2004).
[29] S. Miyazawa and L. J. Robert, Macromolecules **18**, 534 (1985).
[30] Y. C. Kim and G. Hummer, J. Mol. Biol. **375**, 1416 (2008).