# New method for determining size of critical nucleus of fibril formation of polypeptide chains

Nguyen Truong Co and Mai Suan Li

Citation: J. Chem. Phys. **137**, 095101 (2012); doi: 10.1063/1.4749257
View online: http://dx.doi.org/10.1063/1.4749257
View Table of Contents: http://jcp.aip.org/resource/1/JCPSA6/v137/i9
Published by the American Institute of Physics.

# New method for determining size of critical nucleus of fibril formation of polypeptide chains

Nguyen Truong Co[1] and Mai Suan Li[2]
[1]*Saigon Institute for Computational Science and Technology, 6 Quarter, Linh Trung Ward, Thu Duc District, Ho Chi Minh City, Vietnam*
[2]*Institute of Physics, Polish Academy of Sciences, Al. Lotnikow 32/46, 02-668 Warsaw, Poland*

(Received 4 July 2012; accepted 17 August 2012; published online 6 September 2012)

A new method for determining the size of critical nucleus of fibril formation of polypeptide chains is proposed. Based on the hypothesis that the fibril grows by addition of a nascent peptide to the preformed template, the nucleus size $N_c$ is defined as the number of forming template peptides above which the time to add a new monomer becomes independent of the template size. Using lattice models one can show that our method and the standard method which is based on calculation of the free energy, provide the same result for $N_c$. © *2012 American Institute of Physics*. [http://dx.doi.org/10.1063/1.4749257]

## I. INTRODUCTION

The problem of protein aggregation is of paramount importance because it is associated with a number of diseases such as Alzheimer's, Parkinson's, Huntington's, type II diabetes, etc. This spurred many experimental[1] and theoretical studies[2] to understand mechanisms that drive oligomer and fibril formation. One of the central concepts in the aggregation theory is the critical nucleus size $N_c$ above which the aggregate is favorable to form. Below this size, the clusters will tend to dissolve rather than grow. In this paper we deal with fibril formation of polypeptide chains and $N_c$ is a number of peptides but in the traditional nucleation theory it is defined as the nuclei radius. Understanding amyloid nucleation remains a big challenge because the critical nucleus cannot be detected directly as it exists only transiently. However, its size may be inferred, e.g., from the concentration dependence of the lag time, i.e., the time required for nuclei formation. Having used this approach, $N_c$ was experimentally determined for the Ure2p yeast prion[3] and polyglutamine.[4] In prior simulations, $N_c$ is estimated either from the concentration dependence of the lag phase time[5] or from the dependence of the free energy on the number of monomers.[6] In the latter case, the critical nucleus size is defined as a turn-over point of the free energy.[7] With the help of this method and all-atom simulations, $N_c$ was obtained for peptide STVIYE[6] and NFGAIL at the physiological peptide concentration.[8] Using the coarse-grained model Fawzi *et al.* estimated the size of critical nucleus for $A\beta_{1-40}$ peptide.[9] The phenomenological approach has been developed to compute $N_c$ for beta amyloid peptides.[10]

Each approach for estimation of $N_c$ has advantages and disadvantages. The method, based on calculation of the free energy,[6,7] is accurate but it requires a lot of computational effort for good sampling. The atomic nucleation theory[10] is computationally less demanding in expense of lower accuracy. In view of importance of the critical nucleus concept in aggregation, here we propose a simple but efficient method for estimating $N_c$. Our method is based on the experimental[11–13] and theoretical[14–16] observation that the association of monomers to the preformed fibril obeys the dock-lock mechanism, i.e., a nascent monomer can dock and then undergo the structural arrangement to lock onto the template. As the number of template monomers exceeds $N_c$, the time for adding a new monomer, $\tau_{add}$, is expected to become independent of the template size. Based on this we propose to define $N_c$ the number of monomers of the preformed template above which $\tau_{add}$ becomes scale-invariant. Applying this idea to lattice models[17] we obtain $N_c$ for 8-bead sequences which weakly depends on monomer concentration. One can show that the standard approach, based on the dependence of the free energy variation on the number of monomers,[7] supports our approach as it yields the same result for $N_c$.

## II. MATERIALS AND METHODS

In order to illustrate our approach we use the simple lattice model,[17] where each chain consists of $M$ connected beads that are confined to the vertices of a cube. The simulations are done using $N$ identical chains with $M = 8$. The sequence of a chain is +HHPPHH−, where + and − are charged beads, while H and P refer to hydrophobic and polar amino acids, respectively. The assignment of chemical character and the nature of interactions between the beads should be viewed as a caricature of polypeptide chains, and are not realistic representation of amino acids. Despite such drastic simplification it has been shown that lattice models are useful in providing insights into protein folding[18] and aggregation[19] mechanisms.

The inter- and intra-chain potentials include excluded volume and contact (nearest neighbor) interactions. Excluded volume is imposed by the condition that a lattice site can be occupied by only one bead. The energy of $N$ chains is[17]

$$E = \sum_{l=1}^{N} \sum_{i<j}^{M} E_{sl(i)sl(j)} \delta(r_{ij} - a) + \sum_{m<l}^{N} \sum_{i,j}^{M} E_{sl(i)sm(j)} \delta(r_{ij} - a), \quad (1)$$

 

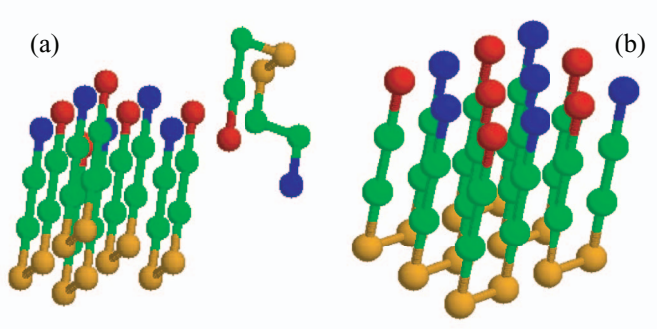


FIG. 1. (a) A typical initial conformation for the (5+1) system. Five monomers of the preformed template are antiparallel, while the conformation of a nascent monomer is randomly generated. (b) The final fibril conformation with the lowest energy $E = -60$.

where $r_{ij}$ is the distance between residues $i$ and $j$, $a$ is a lattice spacing, $sm(i)$ indicates the type of residue $i$ from $m$th peptide, and $\delta(0) = 1$ and zero, otherwise. The first and second terms in Eq. (1) represent intrapeptide and interpeptide interactions, respectively.

The contact energies between H beads $e_{HH}$ is $-1$ (in the unit of hydrogen bond energy $\epsilon_H$). The propensity of polar (including charged) residues to be "solvated" is mimicked using $e_{P\alpha} = -0.2$, where $\alpha = $ P, +, or −. "Salt-bridge" formation between oppositely charged beads is accounted for by a favorable contact energy $e_{+-} = -1.4$. All other contact interactions are repulsive. The generic value for repulsion $e_{\alpha\beta}$ is 0.2. For a pair of like-charged beads the repulsion is stronger, i.e., $e_{++} = e_{--} = 0.7$. It should be noted that energy parameters are chosen in such a way that they roughly describe the interaction energy between amino acids which are modeled as single beads.[17] The chains were confined to the vertices of the three-dimensional hypercube.

In this paper we focus on the peptide concentration $\rho = 5.8$ mM that corresponds the volume fraction occupied by peptides of $\approx 1.87 \times 10^{-4}$. This concentration is about two orders of magnitude denser than that used in typical experiments. To keep it fixed for all simulated systems we have chosen sizes of hypercubes equal $L = 25$, 29, 33, 35, 37, 37, 39, and $41a$ for $N = 4$, 6, 8, 10, 11, 12, 14, and 16, respectively.

The 8-bead sequence monomer in our model folds to the compact native state,[17] while the ground state of multi-sequence systems has the fibril-like structure where monomers are antiparallel (Fig. 1). The folding temperature of monomer $T_F = 0.55$ and it is identified as room temperature in our model.[17] All calculations will be performed at this temperature.

## III. RESULTS AND DISCUSSIONS

### A. New method for determination of $N_c$

.

The kinetics of association of an added monomer to the preformed template is monitored by studying the reaction $MR_{N-1} + MR \rightleftharpoons MR_N$, where MR refers to a 8-bead monomer. Simulations were performed by enclosing N chains in a box with periodic boundary conditions and Monte Carlo (MC) move sets described in Ref. 17. Initially the preformed template is generated as an ordered fibril conformation of $(N-1)$ monomers and a nascent disordered monomer is randomly added (Fig. 1(a)). Red and blue colors refer to positively and negatively charged residues, while hydrophobic (H) and polar (P) residues are green and golden, respectively.

The time to add a nascent molecule to the template, $\tau_{add}$, is defined as an average of first passage times needed to reach the fibril state, where $N$ monomers are antiparallel (Fig. 1(b)) starting from the fully ordered preformed template with an unstructured new monomer. For each system 50 independent MC trajectories were generated to compute $\tau_{add}$. We measure time in units of a MC step (MCS) which is a combination of local and global move.[17]

It should be noted that the starting configuration that consists of preset template and one randomly added monomer may cause some misunderstanding. To avoid this we have also calculated $\tau_{add}$ as follows. The simulation initiates from a random conformation of $N$ monomers and when the fibril of $N-1$ monomers (preset template) is formed we start to count the time to reach the fibril state of $N$ monomers. This time is also defined as adding time $\tau_{\rm add}$. We have checked that $\tau_{\rm add}$ estimated in this way coincides with that starting from preformed template and a randomly added monomer.

The dependence of $\tau_{add}$ on the number of chains that belong to the template, $N_{\rm template}$, is shown in Fig. 2(a). Values of $\tau_{add}$ have been collected at room temperature $T_F = 0.55$.[17] Within error bars, for $N_{\rm template} \geq 11$, the addition time ceases to depend on the template size. Therefore, according to our new approach, the size of critical nucleus is equal $N_c = 11$.

To check the validity of our method we estimate the size of critical nucleus by the another independent free energy approach. We adopt thermodynamics arguments developed by Ferrone[7] for nucleation-polymerization reactions relevant for aggregation kinetics. At equilibrium one can estimate the

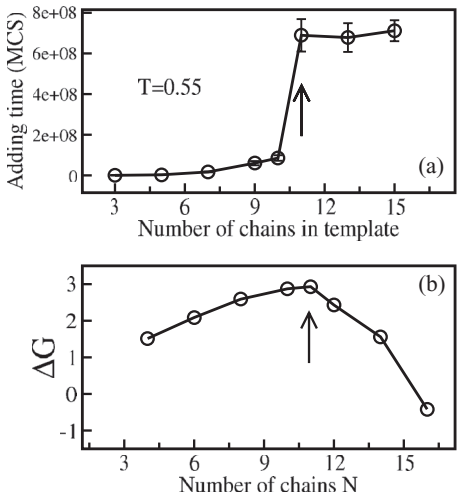


FIG. 2. (a) Dependence of the adding time $\tau_{add}$ on the number of monomers that belong to the preformed template. Results are averaged over 50 MC trajectories. Within error bars $\tau_{add}$ becomes independent of $N$ for $N_{\rm template} \geq 11$. The arrow refers to the size of critical nucleus $N_c = 11$. (b) Dependence of $\Delta G$ on $N$. This quantity displays maximum at $N = N_c$.

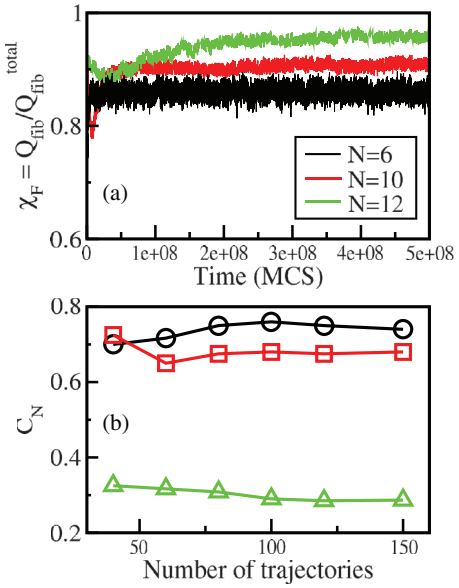


FIG. 3. (a) The time dependence of $\chi_F$ for N = 6, 10, and 12. (b) The dependence of $C_N$ on the number of MC trajectories. It saturates if the number of trajectories exceeds 100 for $N = 6$ and 10 and 120 for $N = 12$.

change in free energy, $\Delta G$, per unit monomer as

$$\frac{d\Delta G}{dN} = -k_B T \ln\left(\frac{C_{N-1}}{C_N}\right). \quad (2)$$

Here $C_N$ is a population of $N$-ordered protofibril or seed with intact end monomers and $C_{N-1}$ is a population with free monomer. To estimate $C_N$ for each system 150 independent trajectories are generated starting from random configurations. The MC simulations have been carried out at the room temperature $T_F = 0.55$.[17] The fraction of of fibril (interpeptide) contacts $\chi_F$, defined as the number of fibril contacts $Q_{\rm fib}$ divided by the total number contacts $Q_{\rm fib}^{\rm total}$ in the fully ordered state, $\chi_F = Q_{\rm fib}/Q_{\rm fib}^{\rm total}$. In equilibrium $\chi_F$ increases as $N$ grows (Fig. 3(a)). We define $C_N$ as the fraction of trajectories that have $\chi_F \geq 0.5$.[7,9] Since $C_N$ may be sensitive to the number of MC trajectories one should generate the number of trajectories large enough to obtain reliable results. Obviously, $C_N$ reaches saturation if the number of trajectories exceeds 100 for $N = 6$ and 10 and 120 for $N = 12$ (Fig. 3(b)). For all studied systems 150 MC trajectories are sufficient (Fig. 3(b)) and values of $C_N$ collected for this number of trajectories are shown in Table I. The results, obtained for $\frac{d\Delta G}{dN}$ using Eq. (2), are also displayed.

The free energy change due to addition of a monomer computed by formula $\Delta G = \int \frac{d\Delta G}{dN} dN$, where the integral is replaced by the sum. Choosing $\Delta G(N = 3) = 0$ and values of $\frac{d\Delta G}{dN}$ from Table I we obtain the dependence of $\Delta G$ on $N$ (Fig. 2(b)) which shows the maximum exactly at the critical point $N_c = 11$. This result agrees with that obtained by our new approach. Thus, two independent methods give the same answer implying that our approach is expected to provide reasonable estimations for the size of critical nucleus in fibril formation of polypeptide chains. The sudden jump in addition time at $N_{\rm template} \sim 10$ without any drastic change in $\Delta G$ around that region (Fig. 2) is presumably an artifact of the lattice model, but this problem requires further investigation. It should be noted that the new method is computationally more efficient than the free energy method as the latter requires three-fold more MC trajectories compared to the first one (Figs. 2 and 3).

TABLE I. Equilibrium populations of ordered fibrils $C_N$, populations with free monomer $C_{N-1}$, and $d\Delta G/dN$ calculated using Eq. (1). Results have been obtained at $T = 0.55$.

| N | $C_N$ | $C_{N-1}$ | $d\Delta G/dN$ |
|---|---|---|---|
| 4 | 0.9400 | 0.0600 | 1.5133 |
| 6 | 0.7400 | 0.2600 | 0.5753 |
| 8 | 0.7133 | 0.2867 | 0.5013 |
| 10 | 0.6600 | 0.3400 | 0.3648 |
| 11 | 0.5267 | 0.4733 | 0.0588 |
| 12 | 0.2867 | 0.7133 | − 0.5013 |
| 14 | 0.1702 | 0.8298 | − 0.8713 |
| 16 | 0.0267 | 0.9733 | − 1.7987 |

From the concentration dependence of lag phase times it was experimentally shown that $N_c$ is equal 6 for the Ure2p yeast prion[3] and even 1 for polyglutamine.[4] Identifying $N_c$ as a turn-over point of the free energy[7] plotted as a function of the number of monomers, Hills and Brooks[6] obtained $N_c = 5$ for peptide STVIYE. Using a slightly different method, Wu *et al.*[8] speculated that at the physiological peptide concentration of 1 nM the nucleus of NFGAIL is an octamer. With the help of the coarse-grained model[20] Fawzi *et al.* obtained $N_c = 10$ for full-length peptide A$\beta_{1-40}$.[9] Teplow and

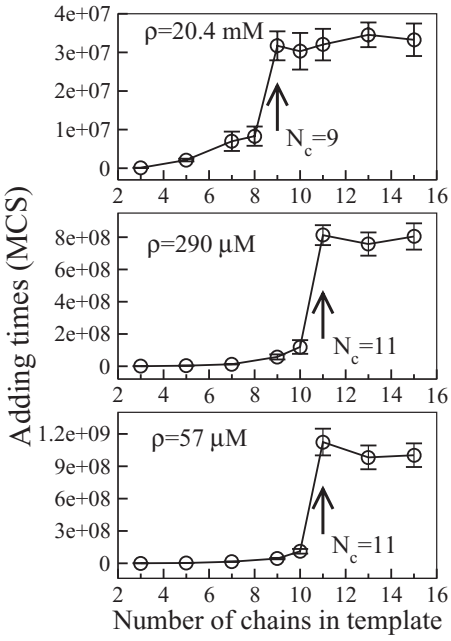


FIG. 4. Dependence of the adding time $\tau_{add}$ on the number of monomers that belong to the preformed template. Results are obtained at $T = 0.55$ and averaged over 50 MC trajectories. For concentrations $\rho = 20.4$ mM, 290 $\mu$M, and 57 $\mu$M, respectively, we set side of cubic box to 23, 95, and 163 for $N = 10$. Cubic sizes for other systems can be obtained from the condition that the monomer concentration is fixed. Within error bars $\tau_{add}$ becomes independent of $N$ for $N_{\rm template} \geq 9$ ($\rho = 20.4$ mM) and 11 ($\rho = 290$ and 57 $\mu$M). The arrow refers to the size of critical nucleus.

co-workers have also shown that kinetic models of amyloid formation fit time-course data when the number of peptides involved in the critical nucleus of A$\beta$ aggregation is set to ten peptides.[21] Thus, our estimate of the critical nucleus size using simple lattice models is in qualitative agreement with prior experimental as well as theoretical works in the sense that $N_c$ is of order of a few to tens chains.

### B. Dependence of critical nucleus size on monomer concentration

We have carried out the study on the dependence of $N_c$ on monomer concentration $\rho$ at the room temperature ($T = 0.55$). The dependence of adding times on the number of monomers in template for several concentrations is shown in Fig. 4. Clearly the dependence of $N_c$ on $\rho$ is rather weak because $N_c = 9$ for $\rho = 20.4$ mM and 11 for $\rho = 290$ and 57 $\mu$M. Such a weak dependence is probably an artifact of the lattice model but this model correctly captures the fact that the higher is the concentration the lower is the critical nucleus size.

## IV. CONCLUSION

In conclusion we have proposed the new approach for estimating the number of monomers that form the critical nucleus for fibril formation of polypeptide chains. In this approach $N_c$ is defined as a size of preformed template above which the time to add a nascent monomer becomes independent of system sizes. The validity of our method has been verified by the standard method where $N_c$ is defined from the dependence of free energy on the number of monomers using lattice models. It would be interesting to apply the new method to off-lattice models where more details on atom structures of sequences are taken into account. Within lattice models the new approach is more efficient than the free energy scaling method but more work should be done to clarify if this conclusion remains valid for other models.

## ACKNOWLEDGMENTS

This work was supported by the Ministry of Science and Informatics in Poland (Grant No. UMO-2011/01/B/NZ1/01622) and Department of Science and Technology at Ho Chi Minh city, Vietnam.

[1] F. Chiti and C. M. Dobson, Ann. Rev. Biochem. **75**, 333 (2006).
[2] J. E. Straub and D. Thirumalai, Curr. Opin. Struct. Biol. **20**, 187 (2010).
[3] N. Fay, Y. J. Inoue, L. Bousset, H. Taguchi, and R. Melki, J. Biol. Chem. **278**, 30199 (2003).
[4] A. M. Bhattacharyya, A. K. Thakur, and R. Wetzel, Proc. Natl. Acad. Sci. U.S.A. **102**, 15400 (2005).
[5] R. Pellarin and A. Caflisch, J. Mol. Biol. **360**, 882 (2006).
[6] R. D. Hills and C. L. Brooks, J. Mol. Biol. **368**, 894 (2007).
[7] F. Ferrone, Methods Enzymol. **309**, 256 (1999).
[8] C. Wu, H. X. Lei, and Y. Duan, Biophys. J. **87**, 3000 (2004).
[9] N. L. Fawzi, Y. Okabe, E. H. Yap, and T. Head-Gordon, J. Mol. Biol. **365**, 535 (2007).
[10] R. Cabriolu and A. Auer, J. Mol. Biol. **411**, 275 (2011).
[11] W. P. Esler, E. R. Stimson, J. M. Jennings, H. V. Vinters, J. R. Ghilardi, J. P. Lee, P. W. Mantyh, and J. E. Maggio, Biochemistry **39**, 6288 (2000).
[12] S. R. Collins, A. Douglass, R. D. Vale, and J. S. Weissman, PLoS Biol. **2**, 1582 (2004).
[13] T. Kowalewski and D. H. Holtzman, Proc. Natl. Acad. Sci. U.S.A. **96**, 3688 (1999).
[14] P. H. Nguyen, M. S. Li, G. Stock, J. E. Straub, and D. Thirumalai, Proc. Natl. Acad. Sci. U.S.A. **104**, 111 (2007).
[15] T. Takeda and D. K. Klimov, J. Mol. Biol. **368**, 1202 (2007).
[16] A. Rojas, A. Liwo, D. Browne, and H. A. Scheraga, J. Mol. Biol. **404**, 537 (2010).
[17] M. S. Li, D. K. Klimov, J. E. Straub, and D. Thirumalai, J. Chem. Phys. **129**, 175101 (2008).
[18] N. D. Socci and J. N. Onuchic, J. Chem. Phys. **101**, 1519 (1994).
[19] M. S. Li, N. T. Co, G. Reddy, C. K. Hu, J. E. Straub, and D. Thirumalai, Phys. Rev. Lett. **105**, 218101 (2010).
[20] S. Brown, N. J. Fawzi, and T. Head-Gordon, Proc. Natl. Acad. Sci. U.S.A. **100**, 10712 (2003).
[21] A. Lomakin, D. B. Teplow, D. A. Kirschner, and G. B. Benedek, Proc. Natl. Acad. Sci. U.S.A. **94**, 7942 (1997).