

## MECHANISM OF OLIGOMERIZATION OF SHORT PEPTIDES

*Mai Suan Li*

Institute of Physics, Polish Academy of Sciences, Warsaw

Nonfibrillar soluble oligomers, which are likely to be transient intermediates in the transitions from monomers to amyloid fibrils, may be the toxic species in Alzheimer's disease. For this reason it is very important to understand early events that direct assembly of amyloidogenic peptides. Using all-atom simulations with the GROMOS96 force field 43a1 in explicit water we have recently shown that the oligomerization of  $A\beta_{16-22}$  peptides obeys the dock-lock mechanism. We have also proposed a toy lattice model which allows us to ascertain this conclusion using a much larger number of monomers. In this contribution we review our all-atom as well as lattice simulation results on the dock-lock mechanism of short peptides which is probably a generic mechanism for fibril elongation of proteins and long peptides.

Нефибрилярные растворимые олигомеры, которые, вероятно, являются короткоживущими интермедиатами, могут быть токсичными выделениями при болезни Альцгеймера. По этой причине очень важно исследовать первичные явления, которые управляют процессом образования фибриллярного состояния для пептидов амилоида. Используя полноатомные симуляции в пакете GROMOS96 с силовым полем 43a1 в воде, мы недавно показали, что процесс олигомеризации пептидов  $A\beta_{16-22}$  подчиняется док-лок-механизму. Также мы предложили игрушечную решеточную модель, позволяющую изучать гораздо большее число мономеров и дающую уверенность в надежности этого механизма. В работе представлены результаты, полученные с помощью как полноатомных, так и решеточных моделей. Они подтверждают механизм док-лок для коротких пептидов, и, вероятно, этот механизм также применим к образованию фибриллов из белков и длинных пептидов.

PACS: 02.60.Cb; 87.15.-V

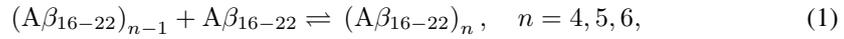
### INTRODUCTION

Protein folding and function take place in the environment crowded with biological macromolecules. As a result, proteins are exposed to intermolecular interactions that may lead to aggregation [1]. In many cases protein aggregates take the form of amyloid fibrils, which appear as unbranched rod-like nanostructures with the diameter of an order of 10 nm and varying length [2]. There is intense interest in determining the structures, kinetics, and growth mechanisms of amyloid fibrils because a large body of evidence suggests that amyloid fibrils and associated oligomeric intermediates are related to a number of diseases, including Alzheimer's, Parkinson's, Huntington's, and prion diseases [3–6]. For example, in the case of the Alzheimer's disease the memory decline may result from the accumulation of the amyloid  $\beta$ -protein ( $A\beta$ ) present in two forms:  $A\beta_{1-40}$  and  $A\beta_{1-42}$ .

Although details of the molecular structures of amyloid fibrils are becoming available [2,7] it is also important to understand the mechanisms of their formation starting from monomers. The kinetics of addition of soluble  $A\beta$  and Sup35 amyloid monomers to the preformed fibril

structures has been investigated experimentally [8–10]. These important studies showed that the association of monomers to the amyloid fibril occurs by a two-stage dock-lock mechanism. Furthermore, it is unlikely that the preformed fibrils themselves undergo substantial conformational changes as the templated-assembly may take place [8–10].

Previous all-atom and coarse-grained simulations have been focused on various aspects of fibril formation of short peptides such as the stability of oligomers [11, 12], the nature of intermediates [13, 14] and the role of mutation [13, 15]. However, the kinetics of adding monomers to the preformed template assembly has not been studied theoretically so far. Recently, we have made the first attempt in this direction [16]. In order to mimic the experiments [8–10] we investigated the oligomerization process of



using the GROMOS96 force field 43a1 [17] to perform extensive all-atom molecular dynamics simulations in explicit water. One of the main aims of the present paper is to review our results on kinetics of reaction (1). The main conclusion from our simulations is that, in agreement with the experiments [8–10], the adding of monomer to the preformed template follows the dock-lock mechanism, but the preformed subsystem fluctuates a lot. For illustration we present the results, obtained for the case with  $n = 4$ , which have not been described in detail in the previous work [16].

Since the all-atom simulations are restricted to a few peptides, we have developed a toy lattice model [18] which allows us for studying the fibril-like growth kinetics of large assemblies of monomers. In this model, a peptide consists of eight amino acids, each of which is represented by a single bead on the simple cubic lattice. The dynamics of the lattice model is defined by the standard Monte Carlo move set. We have shown that the simple lattice model can capture not only the dock-lock mechanism, but also other experimental observations including the activation dynamics of fibril growth at low temperatures [9] and the linear dependence of fibril formation time on the number of peptides [19].

## 1. ALL-ATOM SIMULATIONS

Our basic idea is illustrated in Fig. 1, *a*. We first obtain the antiparallel configuration by long enough simulations starting from random configurations of the trimer (the typical time of antiparallel arrangement of the  $A\beta_{16-22}$  trimer is  $\approx 250$  ns [16]). Then, one peptide is randomly added to the preformed three monomers and monitor the dynamics of assembly of the whole system. We have performed four trajectories of 224, 240, 400 and 600 ns. The volume of the simulation box is  $117 \text{ nm}^3$ , which corresponds to the peptide concentration of 57 mM.

We used the dihedral principal component analysis (PCA) to represent the conformational distribution of the  $3N$ -dimensional system [20]. It uniquely defines the distance in the space of periodic dihedral angles using the variables [20]  $q_k = \cos(\alpha_k)$ ,  $q_{k+1} = \sin(\alpha_k)$ . Here,  $\alpha_k \in \{\phi_k, \psi_k\}$  and  $k = 1, \dots, N - 1$ , with  $N$  being the number of backbone and side-chain dihedral angles. The correlated internal motions are probed using the covariance matrix  $\sigma_{ij} = \langle (q_i - \langle q_i \rangle)(q_j - \langle q_j \rangle) \rangle$ .

The free energy surface along the  $n$ -dimensional reaction coordinate  $V = (V_1, \dots, V_n)$  (obtained by diagonalizing matrix  $\sigma$ ) is given by  $\Delta G(V) = -k_B T [\ln P(V) - \ln P_{\max}]$ , where

$P(V)$  is the probability distribution obtained from a histogram of the MD data,  $P_{\max}$  is the maximum of the distribution, which is subtracted to ensure that  $\Delta G = 0$  for the lowest free energy minimum. We use dPCA to compute the free energy landscapes (FEL) using the first two eigenvectors  $V_1$  and  $V_2$ .

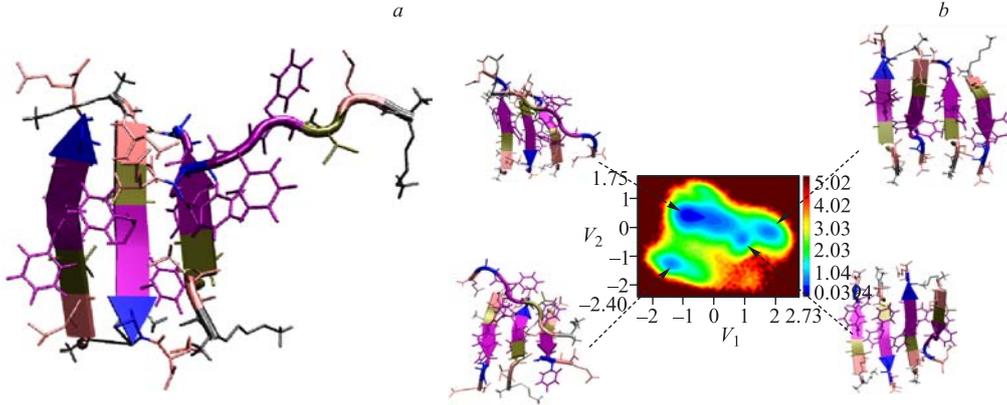


Fig. 1. *a*) The initial conformation for study of reaction (1) with  $n = 4$ . The preformed ordered three peptides have been obtained from our simulations for the trimer [16]. The conformation of monomeric  $A\beta_{16-22}$  was extracted from the structure of  $A\beta_{10-35}$  peptide available in the Protein Data Bank (ID: 1hz3). It is randomly added to the preformed subsystem; *b*) the FEL as a function of  $V_1$  and  $V_2$ . The typical conformations of four basins are shown

In order to characterize the fibril state of short peptides one uses the nematic order parameter  $P_2$  [21]. In terms of the unit vector  $\mathbf{u}_i$  linking  $N$ - and  $C$ -termini for the  $i$ th peptide, the order parameter  $P_2$  is

$$P_2 = \left[ \frac{1}{N} \sum_{i=1}^N \frac{|\mathbf{r}_{NC}^i|}{L_i} \right] P_2^0, \quad P_2^0 = \frac{1}{2N} \sum_{i=1}^N \frac{3}{2} (\mathbf{u}_i \cdot \mathbf{d})^2 - \frac{1}{2}, \quad (2)$$

where  $\mathbf{d}$  (the director) is a unit vector defining the preferred direction of alignment of the oligomer;  $N$  is the number of molecules;  $\mathbf{r}_{NC}^i$  is the end-to-end vector that connects two  $C_\alpha$  atoms from the termini of the  $i$ th peptide. The end-to-end distance in the fully stretched state  $L_i = (N_i - 1)a$  ( $a \approx 4 \text{ \AA}$ ), where  $N_i$  is a number of amino acids in the  $i$ th monomer, and  $a$  is the distance between two  $C_\alpha$  atoms. It follows from Eq. (2) that  $P_2^0 = 1$ , if all peptides are precisely parallel or antiparallel, even if they are not fully extended. In order to characterize the fibril conformations adequately, we define  $P_2$  as a product of  $P_2^0$  and the factor which is equal 1, if all peptides are stretched, and less than 1 otherwise. If  $P_2 > 0.5$ , the system has the propensity to be in an ordered state.

Figure 1, *b* shows the two-dimensional FEL as a function of  $V_1$  and  $V_2$ . The population of four dominant minima is 49, 8, 7 and 4% of sampled conformations, respectively. The low-populated basin 4 corresponds to the most structured conformations ( $\langle P_2 \rangle \approx 0.76$ ). Thus, the antiparallel fibril-like structures are energetically favorable, but they are marginally stable. This is also consistent with low free energy barriers ( $\approx 1 \text{ kJ/mol}$ ) separating different basins

(Fig. 1, *b*). By constructing the contact maps one can show that the antiparallel arrangement of peptides occurs due to interpeptide side chain-side chain interaction and formation of the salt bridge between oppositely charged amino acids at the termini [16]. The hydrogen bonds play the minor role.

We used  $P_2$  as a global order parameter to monitor the overall time-dependent fluctuations in the preformed and growing oligomer. Interestingly, both  $P_2$  of the entire system (Fig. 2, *a*) and of the preformed monomers (Fig. 2, *b*) fluctuate a lot. Although the initial value of  $P_2$  for the preformed trimer is about 0.8, we find that during the growth process it becomes as low as 0.25 (Fig. 2, *b*). For the typical trajectory shown in Fig. 2, on times between  $\approx 50$  and 125 ns the tetramer is ordered around  $P_2 \approx 0.8$ , but the orientational ordering is lost for about 25 ns (Fig. 2, *a*). The ordering is regained again at larger time scales. Thus, for the tetramer the order–disorder transition is reversible. Furthermore, the preformed subsystem fluctuates drastically to accommodate the nascent monomer. The similar behaviour has been also observed for the pentamer and hexamer [16]. We have quantified the dependence in the order parameter fluctuations on the oligomer size using  $\Delta P_2 = \sqrt{\langle P_2^2 \rangle - \langle P_2 \rangle^2}$ , where  $\langle \dots \rangle$  refers to the time average over all trajectories. We find that  $\Delta P_2$  is the largest for the trimer and generally decreases as the oligomer size increases. Since  $P_2(t)$  of the hexamer still fluctuates markedly, we suggest that the size of the critical nucleus of  $A\beta_{16-22}$  peptides exceeds six (the critical nucleus is defined as a minimal oligomer, which can serve as a weakly fluctuating template for addition of new monomers).

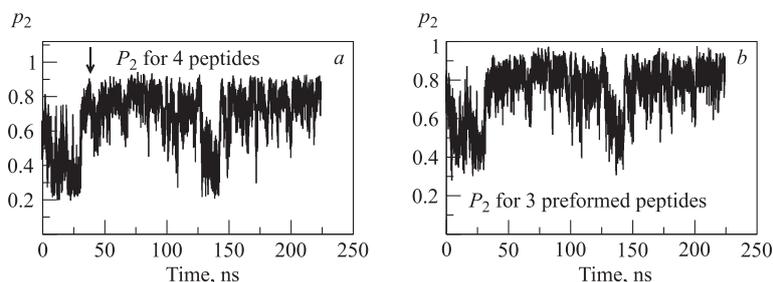


Fig. 2. *a*) The time dependence of the order parameter  $P_2$  for the entire system of four peptides; the arrow refers to  $P_2 = 0.9$ ; *b*) as in *a*, but for the preformed three peptides

In order to show that the added peptide joins the preformed monomers by the dock-lock mechanism, we monitor the time dependence of the  $\beta$  content,  $\beta(t)$  (in the presence of interpeptide interaction the content of helix and random coil is less than 20%). The secondary structure contents were computed using the definitions given in Ref. [13].

Initially,  $\beta(t)$  of the added peptide is  $\approx 0.3$  (results not shown) and for the time shorter than 0.5 ns it reaches the value of about 0.8 (Fig. 3, *a*). For  $t \leq 40$  ns tetramer there is a clear difference between the  $\beta$  contents of the preformed peptides and the nascent monomer. We attribute this period to the dock phase of oligomerization. This phase is followed by the lock stage, in which the added peptide joins the fluctuating template to form the antiparallel arrangement. Given the small number of peptides, the transition between two phases can be considered as reasonably sharp. In order to illustrate the dynamics of approach of the  $\beta$ -strand content of the added monomer to the value expected in the tetramer (roughly that of

the structured trimer), we have computed  $\bar{\beta}(t) = \frac{1}{N_T} \sum_{i=1}^{N_T} \frac{1}{t} \int_0^t \beta_i(s) ds$ , which is the running time average of the strand content averaged over  $N_T$  trajectories. If the added monomer is fully incorporated into the preformed oligomer,  $\bar{\beta}(t)$ , at long times, would approach the equilibrium value. From the dynamics of  $\bar{\beta}(t)$  (Fig. 3, b) we find that for  $t < 75$  ns (the docking stage) the strand content decreases from the value of 0.8, which has been reached during time  $< 0.5$  ns, to about  $\beta \approx 0.75$  and then grows up to 0.8. Thus, most of the fast conformational changes in the monomer occur in the initial phase. For  $t > 75$  ns, which corresponds to the lock phase, the strand content of the monomer increases albeit slowly. Indeed, if one applies the criterion that the ordered phase is formed if  $P_2 \geq 0.9$ , our results, obtained for four trajectories, show that the addition of a monomer to a trimer with well-formed initial  $\beta$  sheet is not complete even at  $t \approx 170$  ns. Thus, the time scale for the lock stage is considerably longer than for the dock phase. The large separation in the rates of the dock and lock phases is consistent with experimental findings that have probed the kinetics of addition of a monomer to the ends of a fibril [8,9]. It appears that in the growth of the fibrils and prenucleus oligomers the rate limiting step is the locking phase.

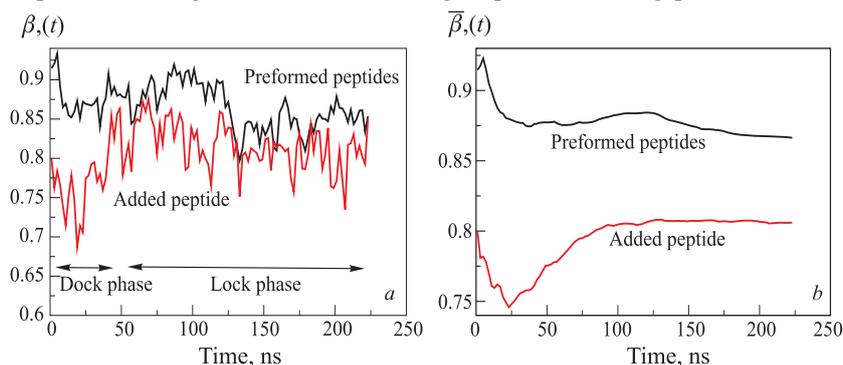


Fig. 3. a) Time dependence of  $\beta(t)$  for the tetramer from the same trajectory shown in Fig. 2. In order to filter the high frequency noise, the instantaneous values of  $\beta(s)$  are averaged over a time interval  $\Delta = 1$  ns; b) time dependence of  $\bar{\beta}(t)$ . The results are averaged over four trajectories

## 2. TOY LATTICE MODEL

Our toy lattice model [18] consists of identical peptides of  $N = 8$  residues each (Fig. 4). The sequence of a peptide is +HHPPHH−, where + and − denote charged residues, H and P represent hydrophobic and polar amino acids. The hydrophobic effect was taken into account by setting the contact energies between H residues  $e_{HH}$  to  $-1$  (in the units of the hydrogen bond energy  $\epsilon_H$ ). The propensity of polar (including charged) residues for solvation is achieved by setting the contact energy  $e_{P\alpha}$  to  $-0.2$ , where  $\alpha = P, +, \text{ or } -$ . Salt bridge formation between oppositely charged residues is given by the contact energy  $e_{+-} = -1.4$ . All other contact interactions are assumed to be repulsive. The generic value of  $e_{\alpha\beta}$  is 0.2, although for the residue pairs with the same charges the repulsion is stronger (0.7). Peptides were confined to the vertices of the three-dimensional lattice model with periodic boundary conditions.

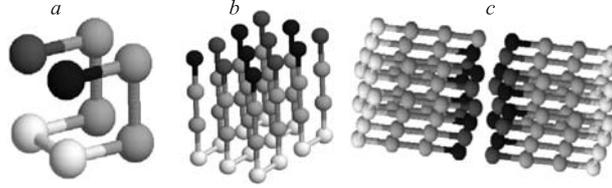


Fig. 4. *a*) Structure of the monomeric ground state. The charged residues are denoted by dark grey (+) and black (-). H and P residues are colored in medium grey and light grey, respectively; *b*) the ordered state for six monomers; *c*) the same as in *b*, but for 12 monomers

The interactions include excluded volume and contact (nearest neighbor) interactions. Excluded volume is imposed by the condition that an infinite energy is assigned to any conformation, in which a lattice site is occupied more than once. The energy of a unit cell is

$$E = \sum_{m=1}^M \sum_{i < j}^N e_{s(i)s(j)} \delta(r_{ij} - a) + \sum_{m < l}^M \sum_{i,j}^N e_{s(i)s(j)} \delta(r_{ij} - a), \quad (3)$$

where  $r_{ij}$  is the distance between residues  $i$  and  $j$ ;  $a$  is a lattice spacing;  $s(i)$  indicates the type of residue  $i$ , and  $\delta(0) = 1$  and zero, otherwise. The first and second terms in Eq. (3) represent intrapeptide and interpeptide interactions, respectively.

Since a monomer has only 8 residues, one can do exact enumeration of all possible conformations. The monomeric native conformation, shown in Fig. 4, *a*, is not degenerated and its energy  $E_N = -3.8$ . The single peptide has the folding temperature  $T_F = 0.5$ , which is a bit higher than the temperature of collapse  $T_\theta = 0.46$  associated with the maximum in energy fluctuations.

We use Monte Carlo (MC) algorithm to study the kinetics of amyloid formation. Typically, a local move is accomplished using standard MC move set, which consists of tail rotation, corner flip, and crankshaft rotation. Using MC simulated annealing simulations we determined the conformation of multipetide oligomer with the lowest energy (Fig. 4, *b*, *c*). In the fibril state the conformation of monomers is more extended compared to the monomeric native one (Fig. 4, *a*). By exact enumeration, one can show that this extended conformation corresponds to the first excited state of a single monomer ( $E_{\text{ex}} = -3.4 > E_N$ ). Thus, our simple model captures the fact that proteins do not fold to the compact native states (misfold) if they aggregate. It is worth to note that the fully ordered fibril conformations have single- (Fig. 4, *b*) and double-layer structures (Fig. 4, *c*) for  $M \leq 10$  and  $M > 10$ , respectively. However, the specific feature of the ordered conformations has been shown [18] to be irrelevant to the mechanism of fibril formation.

In order to demonstrate two-stage dock-lock mechanism of oligomerization we consider the system of  $M = 20$  peptides. The cubic cell is chosen as a cube with the side of 13 lattice sites. Therefore, the volume fraction occupied by the peptides is 0.073 and corresponds to the concentration of 230 mM. This concentration about three and one order of magnitude higher than typical experimental and all-atom simulation value, respectively. We have shown that changing the peptide concentration affects the fibril formation time, but not the mechanism itself. For  $M = 20$  the simulation has been carried out at the temperature of fastest fibril assembly  $T_s = 0.7$  using multiple MC trajectories. In all, 100 MC trajectories starting with

random initial conditions were obtained. The length of MC trajectories ( $10^8$  MCS) at  $T_s$  was sufficiently long to observe fibril assembly events in each trajectory. Consequently, MC trajectories were used to obtain both kinetic characteristics of fibril formation.

The kinetics of fibril assembly may be probed using the probability of being in the fibril state,  $f(t)$ , which is defined as the fraction of the fibril contacts. As seen from Fig. 5, the three-exponential fit ( $f(t) = f_0 - f_1 \exp(-t/\tau_1) - f_2 \exp(-t/\tau_2) - f_3 \exp(-t/\tau_3)$ ) is the best one (dashed line). Here, we have three different time scales  $\tau_1 \approx 0.01\tau_{\text{fib}}$ ,  $\tau_2 \approx 0.09\tau_{\text{fib}}$  and  $\tau_3 \approx 0.5\tau_{\text{fib}}$ , where the fibril formation time  $\tau_{\text{fib}} \approx 2 \cdot 10^7$  MCS. The partition of these phases is  $f_1 \approx 0.18$ ,  $f_2 \approx 0.47$  and  $f_3 \approx 0.11$ .  $\tau_1$  is a characteristic time scale for formation of the «burst phase». At this time scale only  $\approx 0.6\%$  of interpeptide fibril contacts are formed and the largest oligomer contains, on average, only nine peptides (results not shown). At time scales  $\tau_2$  about half of fibril contacts are formed (Fig. 5).  $\tau_3$  has the same order of magnitude as the fibril formation time.

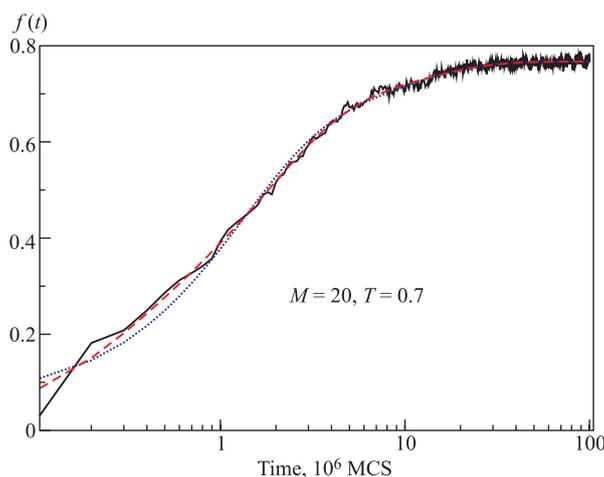


Fig. 5. The time dependence of  $f(t)$  for  $M = 20$  at  $T = 0.7$ . Here  $f(t) = \frac{Q_{\text{fc}}(t)}{Q_{\text{fc}}^{\text{total}}}$ ,  $Q_{\text{fc}}$  is the sum of intra- and inter-peptides fibril contacts,  $Q_{\text{fc}}^{\text{total}}$  is the number of fibril contacts in the ordered state. The results are averaged over 100 trajectories. The dotted and dashed curves refer to the bi-exponential and three-exponential fits, respectively

Using the bi-exponential fit (blue line), we obtain  $\tau_1 \approx 0.06\tau_{\text{fib}}$  ( $f_1 = 0.54$ ) and  $\tau_2 \approx 0.45\tau_{\text{fib}}$  ( $f_2 = 0.16$ ). Similar to the three-exponential case, this fit captures the overall behaviour of  $f(t)$ , except the fast «burst phase» at very short time scales. Thus, our results are in qualitative agreement with the experiments of Esler et al. [9], where the bi-exponential kinetics has been used to describe the template-dependent dock-lock mechanism of  $A\beta$  fibril assembly. We identify  $\tau_1$  and  $\tau_2$ , obtained from bi-exponential fit, as characteristic time scales for the dock and lock phases, respectively. Therefore, in agreement with the experiments and the all-atom simulations described above, the lock stage is much slower than the dock one.

## CONCLUSIONS

Our all-atom simulations in explicit water revealed that the oligomerization of A $\beta$ <sub>16–22</sub> peptides occurs in two steps. In the short dock step monomers undergo large conformational changes. The slow orientational ordering is accomplished in the lock phase. The transition between two phases may be probed by the time dependence of the beta content.

The dock-lock mechanism is also supported by the toy lattice model. As seen from Fig. 5, the lag phase, observed in many experiments, does not appear in our model. Presumably, this is related to the fact that the nucleation process is too fast in our model. However, we have shown [18] that this model can capture the activation dynamics of the fibril assembly of different peptides, but the typical activation energy is about two orders of magnitude lower than the experimental one [22,23]. It is found that  $\tau_{\text{fib}} \sim M$  [18] and this scaling behavior supports the Lifshitz–Slezov nucleation growth picture.

**Acknowledgements.** We are very grateful to D. K. Klimov, P. H. Nguyen, and D. Thirumalai for fruitful collaboration. The work was supported by Ministry of Science and Informatics in Poland, project No. 202-204234.

## REFERENCES

1. *Dobson C. M.* // *Nature*. 2003. V. 426. P. 884.
2. *Tycko R.* // *Curr. Opin. Struct. Biol.* 2004. V. 14. P. 96.
3. *Selkoe D. J.* // *Nature*. 2003. V. 426. P. 900.
4. *Dobson C. M.* // *Science*. 2004. V. 304. P. 1259.
5. *Ross C. A., Poirier M. A.* // *Nature Med.* 2004. V. 10. P. S10.
6. *Bossy-Wetzel E., Schwarzenbacher R., Lipton S. A.* // *Ibid.* P. S2.
7. *Nelson R., Eisenberg D.* // *Curr. Opin. Struct. Biol.* 2006. V. 16. P. 260.
8. *Cannon M. J. et al.* // *Anal. Biochem.* 2004. V. 328. P. 67.
9. *Esler W. P. et al.* // *Biochemistry*. 2000. V. 39. P. 6288.
10. *Collins S. R. et al.* // *PLOS Biol.* 2004. V. 2. P. 1582.
11. *Ma B., Nussinov R.* // *Proc. Nat. Acad. Sci. USA*. 2002. V. 99. P. 14126.
12. *Buchete N. V., Tycko R., Hummer G.* // *J. Mol. Biol.* 2005. V. 353. P. 804.
13. *Klimov D. K., Thirumalai D.* // *Structure*. 2003. V. 11. P. 295.
14. *Gnanakaran S., Nussinov R., Garcia A. E.* // *J. Am. Chem. Soc.* 2006. V. 128. P. 2158.
15. *Gsponer J., Haberthur U., Caflisch A.* // *Proc. Nat. Acad. Sci. USA*. 2003. V. 100. P. 5154.
16. *Nguyen P. H. et al.* // *Proc. Nat. Acad. Sci. USA* (submitted).
17. *Berendsen H. J. C., van der Spoel D., van Drunen R.* // *Comp. Phys. Commun.* 1995. V. 91. P. 43.

18. *Li M. S., Klimov D. K., Thirumalai D.* // *Polymer*. 2004. V. 45. P. 573.
19. *Kowalewski T., Holtzman D. H.* // *Proc. Nat. Acad. Sci. USA*. 1999. V. 96. P. 3688.
20. *Nguyen P. H. et al.* // *Proteins*. 2005. V. 61. P. 795.
21. *Cecchini M. et al.* // *J. Chem. Phys.* 2004. V. 121. P. 10748.
22. *Sabaté R., Gallardo M., Estelrich J.* // *Intern. J. Biol. Macromolecules*. 2005. V. 35. P. 9.
23. *Kusumoto Y. et al.* // *Proc. Nat. Acad. Sci. USA*. 1998. V. 95. P. 12277.