

A lattice-gas model for amyloid fibril aggregation

LIU HONG^{1,2}, XIANGHONG QI² and YANG ZHANG^{2(a)}

¹ Zhou Pei-Yuan Center for Applied Mathematics, Tsinghua University - Beijing, 100084, PRC

² Center for Computational Medicine and Bioinformatics, University of Michigan - 100 Washtenaw Ave., Ann Arbor, MI 48109, USA

received 10 March 2011; accepted in final form 9 May 2011

published online 8 June 2011

PACS 87.10.Hk – Lattice models

PACS 87.14.em – Fibrils (amyloids, collagen, etc.)

PACS 87.15.ak – Monte Carlo simulations

Abstract – A simple lattice-gas model, with two fundamental energy terms —elongation and nucleation effects, is proposed for understanding the mechanisms of amyloid fibril formation. Based on the analytical solution and Monte Carlo simulation of 1D system, we have thoroughly explored the dependence of mass concentration, number concentration of amyloid filaments and the lag-time on the initial protein concentration, the critical nucleus size, the strengths of nucleation and elongation effects, respectively. We also found that thickening process (self-association of filaments into multi-strand fibrils) is not essential for the modeling of amyloid filaments through simulations on 2D lattice. Compared with the kinetic model recently proposed by Knowles *et al.*, highly quantitative consistency of two models in the calculation of mass fraction of filaments is found. Moreover our model can generate a better prediction on the number fraction, which is closer to experimental values when the elongation strength gets stronger.

Copyright © EPLA, 2011

Amyloid proteins can spontaneously converge from isolated water-soluble monomers into large insoluble aggregated fibrils under appropriate conditions either *in vitro* or *in vivo*. If these amyloid fibrils abnormally accumulate in tissues and organs, the patient may suffer from amyloidosis; more seriously, if this happens in the brain, degeneration of neuronal processes and synaptic abnormalities may appear, such as Alzheimer's and Parkinson's diseases etc. [1–3]. Thus fully uncovering the mechanisms of amyloid fiber formation will lead a key step in improving the medical diagnosis and therapy of amyloid-related diseases [4,5].

In the past decade, many high-resolution amyloid fiber structures were resolved by X-ray crystallography and solid state NMR [6–10]; some common key steps in the formation of amyloid fibrils were identified [11,12]; more and more quantitative relationships between fibrillation kinetics and experimental conditions (such as protein concentration, temperature, pH value and etc.) were established [13–15].

Among them, one of the most striking findings would be that protein aggregation may be a generic property of polypeptide chains, which relies on their common peptide

backbone rather than specific amino acid sequences [16]. As we can see, despite large variations of amyloid proteins in secondary structure (dominated by α -helix, β -sheet, or random coil) and tertiary structure (either globular or intrinsically disordered) prior to fibrillation, all fibrils share a similar morphology. They are unbranched and rope-like, typically 10 nm in width and 0.1–10 μ m in length, which reflects a common continuous cross- β -sheet substructure, with β -strands perpendicular to and inter-strand hydrogen bonds parallel to the fibril axis [6].

The above statement serves as a basis of our present study, *i.e.* a molecular-level coarse-grained lattice-gas model, which is built on a n -dimensional discrete lattice system, with each site either empty (noted by “0”) or occupied (noted by “1”) to represent water molecules and amyloid proteins in cross- β -structure, respectively. And we define a filament as a set of occupied lattice points that are adjacent to each other in the same row or same column (its size should be larger than the critical nucleus size n_c); while a fibril as a bundle of connected filaments on 2D and 3D lattice (as illustrated in fig. 1).

As the intersheet interactions are always weaker than the intrasheet hydrogen bonding interactions in fibril, we can make a fair separation of these two types of interactions, and focus more on the energy contribution that

^(a)E-mail: zhng@umich.edu

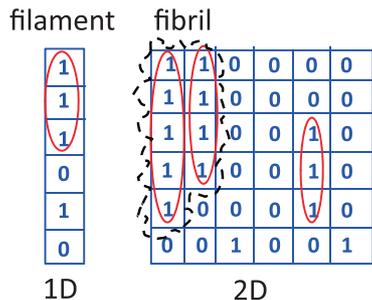


Fig. 1: (Colour on-line) An illustration of the 1D and 2D lattice-gas model.

stabilizes the fibril structure. Most amyloid filaments need first to form some kind of folding nucleus with critical size around 2–6 monomers before aggregation can further proceed [11]. These nuclei usually are not very stable and take a long time to form, which can be accounted by a large positive-energy barrier on the free-energy surface, although whether the major contribution to this barrier is entropy or enthalpy is still controversial [17,18]. After the nucleation stage, filaments grow linearly by subsequent monomer addition at both ends [12]. Fibril elongation often proceeds rapidly, with a negative-energy contribution at each step to stabilize the existing structure. Therefore we can introduce some empirical energy terms, that account for the above key processes, respectively, to describe the fibrillation of amyloid proteins. And the total energy of a fibril system can be constructed as

$$E = E_{nuc} + E_e = \epsilon_n N_f - \epsilon_e N_i, \quad (1)$$

if the interaction energies between water molecules and within amyloid peptides are neglected. E_{nuc} and E_{elg} represent the energy terms for nucleation and elongation separately; ϵ_n and ϵ_e denote the energy strength of nucleation and elongation effects; N_f and N_i denote the total number of single-strand filaments and aggregated peptide pairs in the system.

In fact, as natural amyloid fibrils are unbranched and grow mainly in one dimension, we can constrain our studies on 1D discrete lattice, in which the system energy can be further expressed as

$$E_{1d} = \epsilon_n \sum_i (1 - g_i) g_{i+1} \cdots g_{i+n_c} - \epsilon_e \sum_i g_i g_{i+1}, \quad (2)$$

where $n_c \geq 2$ stands for the critical nucleus size. g_i can take either +1 or 0 value, representing whether site i is occupied by the amyloid peptide or not. In the current case, the total protein concentration is a constant, with $\sum_i g_i = n$.

Although the above energy form looks quite simple, quantitative calculation is not an easy task, as there are enormous possible fibril structures needed to be considered when the system size is large. Here an exact combinatoric formula for the partition function of 1D system under

free boundary condition is given, which constitutes the theoretical foundation of following applications of the lattice-gas model:

$$\begin{aligned} Z(n < n_c) &= 1, \\ Z(n_c \leq n \leq N-1) &= \sum_{m=0}^n \sum_{d=0}^{\lfloor (n-m)/n_c \rfloor} C_{n-m-(n_c-1)d-1}^{d-1} \\ &\cdot C_{m+d}^m C_{N-n+1}^{d+m} e^{[(n-m-d)\epsilon_e - d\epsilon_n]/(k_B T)}, \\ Z(n = N) &= e^{[(N-1)\epsilon_e - \epsilon_n]/(k_B T)}, \end{aligned} \quad (3)$$

where $C_n^m = \frac{n!}{m!(n-m)!}$. m and d are two internal variables. N stands for the size of the lattice system; n for the total number of occupied sites; and thus $\phi = n/N$ is the fraction of occupied sites, corresponding to the protein concentration.

The derivation of eq. (3) is straightforward. Since the first and last formula are quite simple, we will focus on the second one — a case when only a part of the lattices is occupied. First, we remove m monomers from n occupied sites, and partition the remaining $n-m$ occupied sites into d regions, with the restriction that each region contains at least n_c sites. This gives the partition number $C_{n-m-(n_c-1)d-1}^{d-1}$. Second, we insert m monomers between the d regions. As all monomers are equal, there are C_{m+d}^m different ways. Next we divide $N-n$ empty sites into $d+m+1$ intervals between occupied regions. Each of the $d+m-1$ internal intervals requires at least one empty site, while two outer ones can be empty. Thus we add two virtual empty sites at both ends of $N-n$ successive empty sites, then partition them into $d+m+1$ pieces. All possible ways are C_{N-n+1}^{d+m} . Finally multiplying all factors, as well as the Boltzmann factor $e^{[(n-m-d)\epsilon_e - d\epsilon_n]/(k_B T)}$ together, we get the desired result.

Many interesting statistical properties of the fibril system could be learned from the partition function [19], among which two important ones are the mass concentration of filaments (defined as the total number of aggregated peptides in the filaments divided by system size N) and the number concentration of filaments (defined as the total number of single-strand filaments divided by system size N), *i.e.*

$$\langle M \rangle = -\frac{k_B T}{N} \left(\frac{\partial \ln Z}{\partial \epsilon_n} - \frac{\partial \ln Z}{\partial \epsilon_e} \right), \quad (4)$$

$$\langle P \rangle = -\frac{k_B T}{N} \frac{\partial \ln Z}{\partial \epsilon_n}. \quad (5)$$

Quantitative relationships of mass and number concentration of filaments on the initial protein concentration, critical nucleus size, and variable combinations of nucleation and elongation effects are explored in fig. 2. As we can see in fig. 2(A), the curves for mass concentration of filaments on the initial protein concentration gets steeper with the increase of system size, which means there exists some critical value of the protein concentration for amyloid

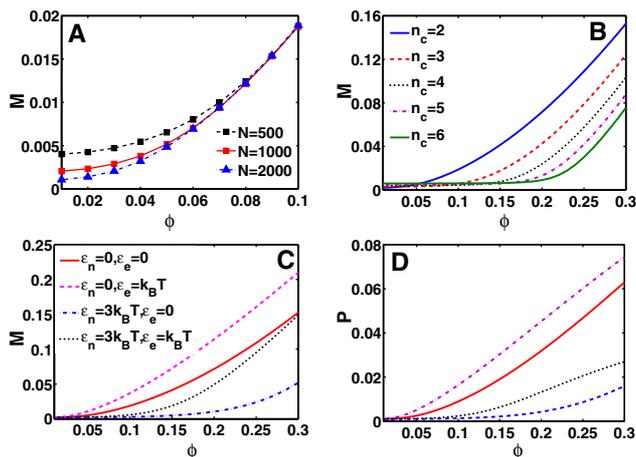


Fig. 2: (Colour on-line) Dependence of mass and number concentration of filaments on protein concentration obtained through eqs. (3)–(5). (A) Influence of system size, with $\epsilon_n = \epsilon_e = 0, n_c = 2$. (B) Influence of different nucleus size, with $\epsilon_n = \epsilon_e = 0, N = 1000$. (C,D) Different combinations of nucleation and elongation effects, with $N = 1000, n_c = 2$.

fibril formation (ϕ^*). If the fraction of occupied sites in the system is below this threshold ($\phi \leq \phi^*$) at the beginning no apparent filaments can be detected. A similar result on the critical protein concentration for amyloid fiber formation has also been reported by Lomakin *et al.* [20]. The critical protein concentration shows a strong positive correlation with the critical nucleus size, which however seems to have less effect on the slope of the curves for the mass concentration of filaments (see fig. 2(B)).

The effect of nucleation and elongation energy can be qualitatively read from a comparison in fig. 2(C) and (D). In general, the negative elongation energy promotes the formation of amyloid fibrils; while the positive nucleation energy requires a larger critical protein concentration. A combination of both energy terms produces a steeper curve for the mass concentration of filaments, and a less steep one for the number concentration of filaments by contrast, which means a higher cooperativity is achieved mainly through the formation of longer filaments.

The partition function mainly deals with the thermal equilibrium state. More prosperous dynamic properties about the fibrillation process could be learned from direct Monte Carlo simulations of 1D lattice-gas model, whose basic step is the state exchange of two neighboring sites. In fig. 3(A), (C), (E), typical MC trajectories under various protein concentrations, strengths for nucleation and elongation effect are illustrated. And their corresponding impacts on the lag-time (which here is defined as the time when mass concentration of filaments reaches 10% of the total protein concentration) are quantitatively summarized in fig. 3(B), (D), (F), *i.e.* $\tau_{lag} \propto \phi^{-5}$, $\tau_{lag} \propto e^{\epsilon_n}$ and $\tau_{lag} \propto e^{-\epsilon_e}$, respectively. The latter two exactly obey the Boltzmann relationship; while the first one gives a scaling exponent which is obviously larger than $(n_c + 1)/2$

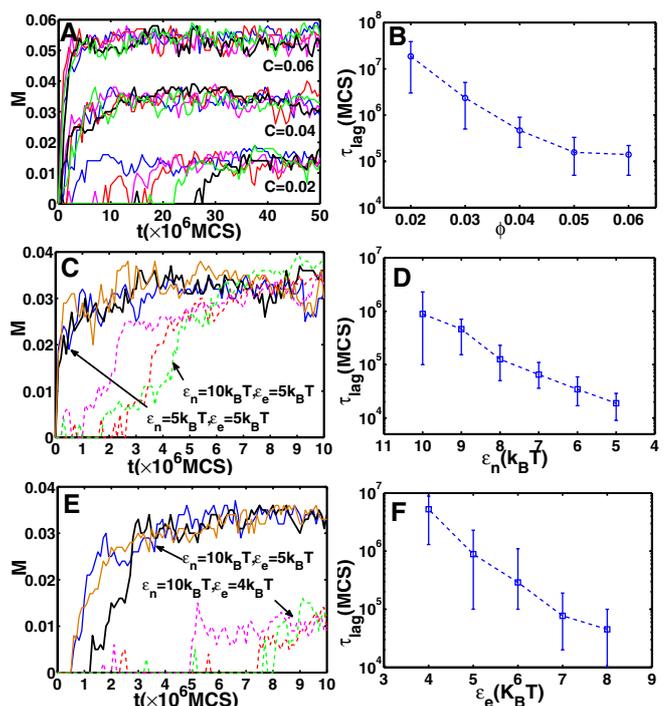


Fig. 3: (Colour on-line) Time evolution of the mass concentration of filaments and corresponding lag-time (τ_{lag}) as functions of the initial protein concentration (ϕ), nucleation effect (ϵ_n) and elongation effect (ϵ_e). (A) Fibril growth at various concentrations, with $N = 1000, \epsilon_n = 10k_B T, \epsilon_e = 5k_B T, n_c = 2$. Each curve represents an independent Monte Carlo simulation. (B) Lag-time *vs.* concentration. Each error bar represents 20 simulation runs. (C) Fibril growth at various nucleation effects, with $N = 1000, \phi = 0.04, n_c = 2$. (D) Lag-time *vs.* nucleation effect. (E) Fibril growth at various elongation effects, with $N = 1000, \phi = 0.04, n_c = 2$. (F) Lag-time *vs.* elongation effect.

as predicted by the classical nucleation theory [21]. This means the lag-time can be effectively shortened by an increase in protein concentration and elongation effect, and largely extended by the increase of the nucleation effect.

The predictions of the current model on the relationship between initial protein concentration and final mass concentration of filaments are further validated by the experimental data of three different amyloid proteins, *i.e.* Csg *B_{trunc}* [22], α -synuclein [23] and Apo C-II [24]. The good agreement as shown in fig. 4 confirms the reliability and usefulness of the lattice-gas model in quantitative exploration of the mechanisms of amyloid fiber formation.

Although real amyloid filaments grow in only one dimension, they do self-associate to form bundles. Thus whether this thickening process is essential for the modeling of amyloid fiber formation is still doubtful. In the last part, we will extend the lattice-gas model to the 2D (or 3D) system and make a quantitative evaluation. In this case, the total energy of a two-dimensional lattice system can be written as (here we choose the x -axis as the fiber

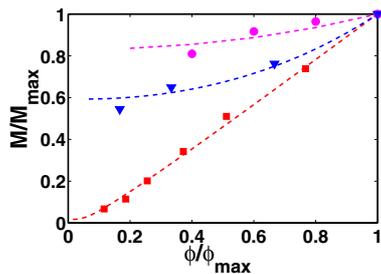


Fig. 4: (Colour on-line) Comparison of theoretical predictions with experimental data on the mass concentration of filaments. For a better illustration, each data set is normalized by the value under its maximum protein concentration. The red squares stand for the experimental data of the amyloid protein Csg *Btrunc*, blue triangles for α -synuclein, and pink circles for Apo C-II. Three dashed lines are calculated by the 1D lattice-gas model (eqs. (3) and (4)), with different best-fitting energy parameters (the red line with $\epsilon_n = 0, \epsilon_e = 4k_B T$; the blue line with $\epsilon_n = 0.5k_B T, \epsilon_e = 3.2k_B T$; and the pink line with $\epsilon_n = 0.5k_B T, \epsilon_e = 3.3k_B T$). The protein concentration is set the same as in experiments.

growing direction)

$$E_{2d} = \epsilon_n \sum_i (1 - g_{i,j}) g_{i,j+1} \cdots g_{i,j+n_c} - \epsilon_e^1 \sum_{i,j} g_{i,j} g_{i,j+1} - \epsilon_e^2 \sum_{i,j} g_{i,j} g_{i+1,j}, \quad (6)$$

where $g_{i,j}$ can take either +1 or 0 value, representing the status of being occupied of site (i, j) . ϵ_e^1 and ϵ_e^2 represent the interaction strength along two perpendicular axis, respectively. Compared to the energy form of the 1D lattice-gas model (eq. (2)), the major difference lies in that eq. (6) not only considers the energy contribution from protein-protein interaction within each filament, but also includes the interaction between different filaments, which is a key to understand how single-strand filaments aggregate into multi-strand fibrils [25].

In fig. 5, the results of Monte Carlo simulations on a two-dimensional lattice system with different strengths of the thickening process (expressed through the energy parameter ϵ_{elg}^2) are compared. We can see that in general the mass concentration of filaments gets larger, while the number concentration of filaments gets smaller with the increase of the interaction strength between different filaments. This means the thickening process helps the amyloid fiber formation mainly through the stabilization of long filaments. However the limited difference in M and P (less than 10%), considering relatively big changes in the interaction strength between different filaments, shows that the thickening process may not be as effective as elongation for the fibril formation, which coincides with the fact that the self-association of fibrils happens after the filaments grow.

Kinetic models based on chemical master equations have recently witnessed considerable success in describing

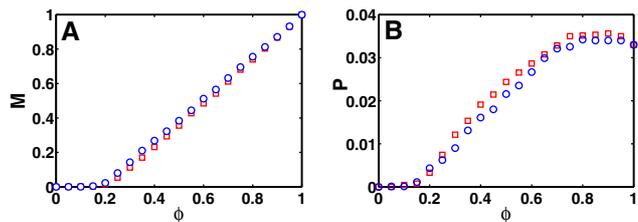


Fig. 5: (Colour on-line) (A) Mass concentration and (B) number concentration of filaments with and without thickening process. Blue circles represent the results without thickening process, $\epsilon_n = 5k_B T, \epsilon_e^1 = k_B T, \epsilon_e^2 = 0$; red rectangles for the results with thickening process, $\epsilon_n = 5k_B T, \epsilon_e^1 = k_B T, \epsilon_e^2 = k_B T$. Each data point on the plot corresponds to the statistical average of 100 Monte Carlo simulations of the 2D lattice-gas model on a $N = 30 \times 30$ square lattice, with each run taking 10^6 movements.

the fibrillation process of amyloid proteins [26,27]. An interesting question therefore raised is the relationship between these two models. As a special example, here we compare the static solutions of Knowles' model [27], a recently proposed kinetic model, with the equilibrium statistical results predicted by the lattice-gas model. Since both models are built on similar mechanisms — nucleation, elongation, and fragmentation (not so apparent in the lattice-gas model) in describing the formation process of amyloid fibrils, we argue that if the parameters in the two models satisfy the following Boltzmann's relationship:

$$k_+ m_{tot} / k_- \propto e^{\epsilon_e / (k_B T)}, \quad (7)$$

$$k_n m_{tot} / k_- \propto e^{-(\epsilon_n - \epsilon_e) / (k_B T)}, \quad (8)$$

where m_{tot} is the initial protein concentration; k_+ , k_- and k_n represent the reaction rates of nucleation, elongation and fragmentation in Knowles's model [27], respectively, we can expect the statistical results of the lattice-gas model will be in quantitative agreement with the static solution of the kinetic model, when the initial protein concentration is not very high. This conjecture is well validated through numerical comparison of both models on the prediction of mass fraction of filaments (mass concentration of filaments divided by initial protein concentration).

In fig. 6(A), we can see that under the same fibrillation conditions (with Boltzmann relationship exactly held), predictions of both models on the mass fraction of filaments are astonishingly the same. Even on the number fraction of filaments (see fig. 6(B)), both models give similar results in a large region of model parameters ($k_+ m_{tot} / k_- \ll 1$). However as the strength of elongation effect gets stronger, the difference between the two models becomes more pronounced. The kinetic model predicts that the number fraction of filaments grows at the same rate as the mass fraction of filaments, which consequently leads to an almost constant value for the average length of filaments (M/P); while the lattice-gas model predicts the

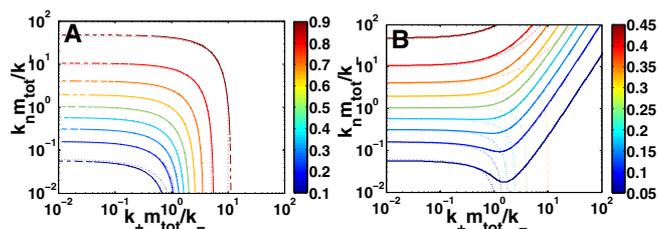


Fig. 6: (Colour on-line) Comparison of the lattice-gas model with Knowles' model on (A) mass fraction of filaments and (B) number fraction of filaments. The solid lines represent the dependence of mass (and number) fraction of filaments on two Boltzmann factors $e^{\epsilon_e/(k_B T)}$ and $e^{-(\epsilon_n - \epsilon_e)/(k_B T)}$ as predicted by the 1D lattice-gas model with $N = 500, n = 50$ (eq. (3)); the dashed lines show the dependence of static mass (and number) fraction of filaments on the dimensionless parameters $k_+ m_{tot}/k_-$ and $k_n m_{tot}/k_-$ as predicted by Knowles' model. Two contour plots are matched together with $k_+ m_{tot}/k_- = e^{\epsilon_e/(k_B T)}/8$ and $k_n m_{tot}/k_- = e^{-(\epsilon_n - \epsilon_e)/(k_B T)}/8$ (see eqs. (7) and (8)).

average length of filaments will increase with the growth of the elongation effect. Judging from the experimental data on amyloid fiber formation [28,29], we believe that the lattice-gas model offers a more reasonable result, since it is based on a truly thermal reversible system, while Knowles' model only includes forward reactions but no backward reactions, which will lead to an overestimation on the effect of fragmentation [30].

In conclusion, we have proposed a lattice-gas model for understanding the mechanisms of amyloid fibril formation, which possesses a most dramatic characteristic—simplicity. In fact in the current model, all atomic details of amyloid proteins are completely neglected. Even peptide interactions are highly simplified. Only the inter-peptide ones are kept. They are characterized by empirical energy terms, which are directly correlated with the aggregation process of amyloid fibrils. Despite the highly simplified nature of the current model in the description of structural, chemical and energetic details of amyloid proteins, many detailed dynamic properties of fibril formation and quantitative relationships between initial experimental condition and final fibril concentration are predicted with high precision. In fact, a quantitative comparison with a sophisticated kinetic model demonstrates the consistency of two different models in the calculation of the mass fraction of filaments. And the lattice-gas model can generate a better prediction on the number fraction, which is closer to experimental values when the elongation strength gets stronger.

Another advantage is generality. Due to its extremely simplified physical picture, the lattice-gas model can be easily extended to account for more complicated structures and interactions. A possible worthwhile extension is the inclusion of different energy forms. For brevity, here we have only considered two most fundamental

energy forms—nucleation and elongation, even though the real fibrillation process of amyloid proteins can be far more complex than what we have described. For instance, conformational transition [6,31], on and off-pathway competition [32], homo- or heterogeneous nucleation [20,33], secondary nucleation [34–36], autocatalytic surface growth [9], merging [9,25] and branching [37], may all play important roles. Thus a thorough exploring of these model systems should significantly enhance our current understandings on the mechanisms of amyloid fibril formation, and also shed light on the diagnosis and therapy of amyloid-related diseases.

This work is supported in part by the NSF Career Award 0746198, the National Institute of General Medical Sciences Grants GM083107 and GM084222, and the Tsinghua University Initiative Scientific Research Program (20101081751).

REFERENCES

- [1] MERLINI G. and BELLOTTI V., *N. Engl. J. Med.*, **349** (2003) 583.
- [2] CHITI F. and DOBSON C. M., *Annu. Rev. Biochem.*, **75** (2006) 333.
- [3] HARRISON R. S. *et al.*, *Rev. Physiol. Biochem. Pharmacol.*, **159** (2007) 1.
- [4] FENG B. Y. *et al.*, *Nat. Chem. Biol.*, **4** (2008) 197.
- [5] ROBERTS B. E. and SHORTER J., *Nat. Struct. Mol. Biol.*, **15** (2008) 544.
- [6] SUNDE M. and BLAKE C. C. F., *Q. Rev. Biophys.*, **31** (1998) 1.
- [7] NELSON R. *et al.*, *Nature*, **436** (2005) 773.
- [8] NELSON R. and EISENBERG D., *Adv. Protein Chem.*, **73** (2006) 235.
- [9] PARAVASTU A. K., LEAPMAN R. D., YAU W. M. and TYCKO R., *Proc. Natl. Acad. Sci. U.S.A.*, **105** (2008) 18349.
- [10] WASMER C. *et al.*, *Science*, **319** (2008) 1523.
- [11] MORRIS A. M. *et al.*, *Biochemistry*, **47** (2008) 2413.
- [12] COLLINS S. R. *et al.*, *PLoS Biol.*, **10** (2004) 1582.
- [13] FANDRICH M., *J. Mol. Biol.*, **365** (2007) 1266.
- [14] KODAKA M., *Biophys. Chem.*, **107** (2004) 243.
- [15] NIELSEN L., *Biochemistry*, **40** (2001) 6036.
- [16] DOBSON C. M., *Nature*, **426** (2003) 884.
- [17] CHENG S. Z. D. and LOTZ B., *Polymer*, **46** (2005) 8662.
- [18] RUSCHAK A. M. and MIRANKER A. D., *Proc. Natl. Acad. Sci. U.S.A.*, **104** (2007) 12341.
- [19] KERSON H., *Statistical Mechanics* (Wiley, New York) 1987.
- [20] LOMAKIN A. *et al.*, *Proc. Natl. Acad. Sci. U.S.A.*, **93** (1996) 1125.
- [21] OOSAWA F. and ASAKURA S., *Thermodynamics of the Polymerization of Protein* (Academic Press, New York) 1975.
- [22] HAMMER H. D. *et al.*, *Proc. Natl. Acad. Sci. U.S.A.*, **104** (2007) 12494.
- [23] UVERSKY V. N. *et al.*, *J. Biol. Chem.*, **276** (2001) 44284.

- [24] BINGER K. J. *et al.*, *J. Mol. Biol.*, **376** (2008) 1116.
- [25] PALLITTO M. M. and MURPHY M., *Biophys. J.*, **81** (2001) 1805.
- [26] KUNES K. C. *et al.*, *Phys. Rev. E*, **72** (2005) 051915.
- [27] KNOWLES T. P. J. *et al.*, *Science*, **326** (2009) 1533.
- [28] BAN T. *et al.*, *J. Biol. Chem.*, **278** (2003) 16426.
- [29] RAAIJ M. E., *Biophys. J.*, **95** (2008) 4871.
- [30] HILL T. L., *Biophys. J.*, **44** (1983) 285.
- [31] SERIO T. R. *et al.*, *Science*, **289** (2000) 1317.
- [32] POWERS E. T. and POWERS D. L., *Biophys. J.*, **94** (2008) 379.
- [33] HILLS R. H. jr. and BROOKS C. L. III, *J. Mol. Biol.*, **368** (2007) 894.
- [34] PADRICK S. B. and MIRANKER A. D., *Biochemistry*, **41** (2002) 4694.
- [35] HILL E. K. *et al.*, *Biomacromolecules*, **7** (2006) 10.
- [36] XUE W. F. *et al.*, *J. Biol. Chem.*, **284** (2009) 34272.
- [37] ANDERSEN C. B. *et al.*, *Biophys. J.*, **96** (2009) 1529.