Replica exchange molecular dynamics simulations of amyloid peptide aggregation

M. Cecchini, F. Rao, M. Seeber, and A. Caflisch*

Department of Biochemistry, University of Zurich,
Winterthurerstrasse 190, CH-8057 Zurich, Switzerland
tel: +41 1 635 55 21, fax: +41 1 635 68 62, e-mail: caflisch@bioc.unizh.ch
(Dated: February 5, 2008)

The replica exchange molecular dynamics (REMD) approach is applied to four oligomeric peptide systems. At physiologically relevant temperature values REMD samples conformation space and aggregation transitions more efficiently than constant temperature molecular dynamics (CTMD). During the aggregation process the energetic and structural properties are essentially the same in REMD and CTMD. A condensation stage toward disordered aggregates precedes the β-sheet formation. Two order parameters, borrowed from anisotropic fluid analysis, are used to monitor the aggregation process. The order parameters do not depend on the peptide sequence and length and therefore allow to compare the amyloidogenic propensity of different peptides.

Keywords: molecular dynamics; replica exchange; peptide aggregation; amyloid; order transition; nematic order parameter

INTRODUCTION

A thorough sampling of conformational space is required to describe the thermodynamics of complex systems such as multiple peptide chains at finite concentrations. Constant temperature molecular dynamics (CTMD) techniques often fail to adequately sample conformational space of frustrated and minimally frustrated systems which are characterized by a rugged free-energy landscape where energy barriers between minima are higher than the thermal energy at physiological temperature. For this reason, a number of approaches to enhance sampling of phase space have been introduced [1, 2, 3, 4]. The parallel tempering technique (also known as replica exchange) was developed for dealing with the slow dynamics of disordered spin systems [5]. Sugita and Okamoto have extended the original formulation of replica exchange into an MD based version (REMD) and tested it on the pentapeptide Met-enkephalin in vacuo [6]. Although in the context of fragile liquids De Michele and Sciortino found that parallel tempering does not increase the speed of equilibration of the (slow) configurational degrees of freedom [7], in the case of atomistic simulations of proteins many different applications have shown the efficiency of the method. Sanbonmatsu and Garcia have used REMD to investigate the structure of Met-enkephalin in explicit water [8], and the α-helical stabilization by the arginine side-chain which was found to originate from the shielding of main chain hydrogen bonds [9]. REMD has also been applied to investigate the energy landscape of the C-terminal β-hairpin of protein G [10, 11] and a three-helix bundle protein [12]. REMD in implicit solvent has been used to investigate the thermodynamics of designed 20-residue structured peptides [13, 14], and recently to study folding of a helical transmembrane protein [15].

Highly ordered protein aggregates are associated with severe human disorders including Alzheimer's disease, type II diabetes, systemic amyloidosis, and transmissible spongiform encephalopathies [16, 17]. The soluble precursors of the ordered protein deposits do not share any sequence homology or common fold. However, X-ray diffraction data indicate a cross-β-structure for most fibrillar aggregates [18, 19]. These findings suggest that key steps in the aggregation process may be common to all amyloidogenic proteins. Despite the medical relevance of amyloidoses, many important questions about the formation of ordered aggregates remain unanswered. There is experimental evidence that cytotoxicity is more pronounced for the early aggregates than for highly organized fibrillar structures [20]. Moreover, some peptide fragments of amyloidogenic proteins display the same properties as the full-length protein, including cooperative kinetics of aggregation, fibril formation, binding of the dye Congo red, and the cross-β X-ray diffraction pattern [21]. Both findings are particularly interesting because current simulation approaches allow significant sampling only for oligomeric peptide systems.

There have been several lattice studies on aggregation in proteins. These simplified models have allowed to investigate the foldability and aggregation propensity [22, 23] and how interaction potentials affect the properties of aggregation-prone proteins [24]. Harrison *et al.* have shown that less stable proteins have a greater chance of assuming alternative native states as multimers [25]. MD simulations of aggregation have been performed by using a three-bead backbone and single-bead side chain model [26]. While this simplified model has allowed the simulation of the competition between folding and aggregation for two four-helix bundles it is probably not possible to extract detailed information on energetics and sequence dependence. Recently, a minimalist Go model of four peptide strands [27] has been investigated by MD simulations in a confining sphere and the aggregation process was shown to depend on both sequence and environment [28]. Atomic models of amyloidogenic

peptides have been simulated by MD with an implicit treatment of the solvent [29, 30, 31] and explicit water molecules [32, 33, 34, 35, 36].

Recently, a replica exchange Monte Carlo technique has been applied to a lattice Go model of a minimalist multichain system to study the interplay between folding and disordered aggregation [23] but atomic model REMD applications to ordered aggregation have not been reported yet.

In the present paper, REMD with implicit solvent [37] is used to investigate the thermodynamics of the early steps of peptide aggregation and comparison is made with CTMD. The present work was motivated by three questions: Is it possible to sample the early events of ordered peptide aggregation at physiologically relevant temperatures? Do the aggregation energetics sampled by REMD correspond to those observed in CTMD simulations? Are the nematic and polar order parameters, borrowed from liquid crystal theory, useful to describe aggregation? The simulation results indicate that all questions can be answered affirmatively. Moreover, the "liquid crystals" order parameters allow to discriminate amyloidogenic peptide sequences from those that form only disordered aggregates.

METHODS

Model

The MD simulations and part of the analysis of the trajectories were performed with the CHARMM program [38]. The oligomeric peptide systems were modeled by explicitly considering all heavy atoms and the hydrogen atoms bound to nitrogen or oxygen atoms (PARAM19 potential function [38, 39]). The remaining hydrogen atoms are considered as part of the carbon atoms to which they are covalently bound (extended atom approximation). The effective energy, whose negative gradient corresponds to the force used in the dynamics, is

$$E(\mathbf{r}) = E_{vacuo}(\mathbf{r}) + G_{solv}(\mathbf{r}) \qquad (1)$$

for a molecular system with atomic nuclei located at $\mathbf{r} = (\mathbf{r}_1, ..., \mathbf{r}_N)$. The PARAM19 vacuo energy function is

$$
\begin{aligned}
E_{vacuo}(\mathbf{r}) =\ & \frac{1}{2} \sum_{bonds} k_b (b - b_0)^2 + \frac{1}{2} \sum_{\substack{bond \\ angles}} k_\theta (\theta - \theta_0)^2 \\
& + \frac{1}{2} \sum_{\substack{dihedral \\ angles}} k_\phi [1 + \cos(n\phi - \delta)] \\
& + \frac{1}{2} \sum_{\substack{improper \\ dihedrals}} k_\omega (\omega - \omega_0)^2 \\
& + \sum_{i>j} \varepsilon_{ij}^{\min} \left[\left(\frac{d_{ij}^{\min}}{r_{ij}} \right)^{12} - 2 \left(\frac{d_{ij}^{\min}}{r_{ij}} \right)^6 \right] \\
& + \sum_{i>j} \frac{q_i q_j}{\epsilon(r_{ij}) r_{ij}}
\end{aligned}
$$

where b is a bond length, θ a bond angle, ϕ a dihedral angle, ω an improper dihedral, r_{ij} is the distance between atoms i and j, q_i and q_j are partial charges, and $d_{ij}^{\min}$ and $\varepsilon_{ij}^{\min}$ are the optimal van der Waals distance and energy, respectively. Parameters are given in Ref. [39].

An implicit model based on the solvent accessible surface was used to describe the main effects of the aqueous solvent on the solute [37]. In this approximation, the solvation free energy is given by:

$$G_{solv}(\mathbf{r}) = \sum_{i=1}^{N} \sigma_i A_i(\mathbf{r}) \qquad (2)$$

for a molecular system having N heavy atoms with Cartesian coordinates $\mathbf{r} = (\mathbf{r}_1, ..., \mathbf{r}_N)$. $A_i(\mathbf{r})$ is the solvent-accessible surface computed by an approximate analytical expression [40] and using a 1.4 Å probe radius. The solvation model contains only two σ parameters: one for carbon and sulfur atoms ($\sigma_{C,S} = 0.012$ kcal/mol Å^2), and one for nitrogen and oxygen atoms ($\sigma_{N,O} = -0.060$ kcal/mol Å^2) [37]. Hence, according to Eq. 2 hydrophobic side chains tend to be buried within the solute whereas hydrophilic side chains and the polar groups of the backbone prefer to be solvent accessible. Furthermore, ionic side chains were neutralized [41] and a linear distance-dependent screening function ($\epsilon(r_{ij}) = 2r_{ij}$) was used for the electrostatic interactions. The CHARMM PARAM19 default cutoffs for long range interactions were used, i.e., a shift function [38] was employed with a cutoff at 7.5 Å for both the electrostatic and van der Waals terms. This cutoff length was chosen to be consistent with the parameterization of the force-field and implicit solvation model. The model is not biased toward any particular secondary structure type. In fact, exactly the same force field and implicit solvent model have been used recently in MD simulations of aggregation [30, 31], folding of structured peptides (α-helices and β-sheets)

TABLE I: Simulations performed

Peptide sequence	Length (μs)	T (K)	Method	IP aggregation events	IA aggregation events
GNNQQNY	10×0.5	275	CTMD	0	6 (19.2) [a]
GNNQQNY	5×1.0	296	CTMD	3 (14.4)	5 (1.6)
GNNQQNY	10×3.4	330	CTMD	54 (7.6)	43 (1.4)
GNNQQNY	2×1.0	371	CTMD	0	0
GNNQQNY	6×2.0	275-400	REMD	14 (60.3)	15 (3.9)
QQQQQQQ	6×2.0	275-400	REMD	27 (54.8)	2 (9.4)
AAAAAAA	6×1.0	275-400	REMD	4 (0.8)	12 (0.9)
SQNGNQQRG	6×2.0	275-400	REMD	1 (1.6)	6 (1.0)

[a]The average time (ns) the three peptides remained aggregated in IP and IA is given in parentheses

ranging in size from 15 to 31 residues [42, 43, 44], and small proteins of about 60 residues [45, 46].

REMD simulations

The basic idea of REMD is to simulate different copies (*replicas*) of the system at the same time but at different temperatures values. Each replica evolves independently by MD and every t_{swap} states i, j with neighbor temperatures are swapped (by velocity rescaling) with a probability $w_{ij} = \exp(-\Delta)$, [6] where $\Delta = (\beta_i - \beta_j)(E_j - E_i)$, $\beta = 1/kT$ and E is the effective energy (potential and solvation energy, Eq. 1). A t_{swap} of 10000 MD steps (20 ps) was chosen in order to allow the kinetic and potential energy of the system to relax. High temperature simulation segments facilitate the crossing of the energy barriers while the low temperature ones explore in detail energy minima. The result of this swapping between different temperatures is that high temperature replicas help the low temperature ones to jump across the energy barriers of the system.

In this study six replicas were used with temperatures (in K): 275, 296, 319, 344, 371, 400. This range corresponds to a subset of values used in a previous study of reversible peptide folding with the same force-field and solvation model [14]. The acceptance ratios of exchange between neighbor temperatures ranged between 15% and 24%. Each trajectory has a length of 2 μs for a total of 12 μs of simulation time (see Table I).

Constant temperature MD simulations

A series of control runs were performed at constant temperature: (i) Ten simulations at 330 K (total of 34 μs) used as a comparison for the aggregation process between CTMD and REMD (see Table I); (ii) ten 0.5 μs simulations at 275 K and (iii) five 1 μs simulations at 296 K to compare CTMD and REMD sampling at physiologically relevant conditions; (iv) two 1 μs simulations at 371 K to study the system near the *condensation* temperature (see below).

For both REMD and CTMD, Langevin dynamics with a friction value of 0.15 ps^{-1} was used. This friction coefficient is much smaller than the one of water (43 ps^{-1} at 330 K computed as $3\pi \eta d/m$, [47] where η is the viscosity of water at 330 K, and d and m are the effective diameter, i.e., 2.8 Å , and mass of a water molecule, respectively) to allow for sufficient sampling within the μs time scale of the simulation. The small friction does not influence the thermodynamic properties of the system.

The SHAKE algorithm [48] was used to fix the length of the covalent bonds involving hydrogen atoms, which allows an integration time step of 2 fs. Furthermore, the nonbonded interactions were updated every 10 dynamics steps and coordinate frames were saved every 20 ps for a total of $5 \cdot 10^4$ conformations/μs. A 1 μs run requires approximately 2 weeks on a 1.4 GHz Athlon processor and the REMD simulations were run in parallel on a Linux Beowulf cluster.

Progress variables

Aggregation contacts. In-register parallel and antiparallel aggregation contacts were defined following the prescription given in Ref. [30]: a contact was considered to be present if the distance between two C_α atoms placed on different in-register strands was within 5.5 Å. The fraction of in-register parallel contacts Q_p and in-register antiparallel contacts Q_a were used to monitor the evolution of the aggregation process. In-register parallel and antiparallel aggregates, IP and IA respectively, were considered formed when Q_p and Q_a were larger than 0.75 ($Q_p, Q_a > 11/14$) whereas at values smaller than 0.25 ($Q_p, Q_a < 4/14$), the system was considered disordered. The aggregation time is defined as the temporal interval between the first time point where $Q_p, Q_a < 0.25$ and the following time point where $Q_p, Q_a > 0.75$.

Radius of gyration. The radius of gyration of the oligomeric system R_g was considered to monitor the degree of *condensation* and calculated using the minimum image convention. Large values of R_g indicate conformations with isolated and non interacting peptides (*uncondensed phase*). Small values of R_g indicate ordered as well as disordered aggregated conformations (*condensed phase*).

Orientational order parameters

The nematic and polar order parameters, $\overline{P_2}$ and $\overline{P_1}$ respectively, were considered in this study. These order parameters represent the first and second rank coefficients of the singlet orientational distribution expanded in a Wigner series [49, 50], i.e., a basis set of the Wigner rotation matrices. The nematic and polar order parameters are widely used for studying the properties of anisotropic

fluids such as liquid crystals [51, 52, 53, 54] and are defined as

$$\overline{P_2} = \frac{1}{N} \sum_{i=1}^{N} \frac{3}{2} (\hat{\mathbf{z}}_i \cdot \hat{\mathbf{d}})^2 - \frac{1}{2}, \qquad (3)$$

and

$$\overline{P_1} = \frac{1}{N} \sum_{i=1}^{N} \hat{\mathbf{z}}_i \cdot \hat{\mathbf{d}}, \qquad (4)$$

where $\hat{\mathbf{d}}$ (the director) is a unit vector defining the preferred direction of alignment, $\hat{\mathbf{z}}_i$ is a suitably defined molecular vector, and N is the number of molecules in the simulation box, i.e., three peptides in this study. The director is defined as the eigenvector of the ordering matrix [55] that corresponds to the largest eigenvalue. Here, the molecular vectors $\hat{\mathbf{z}}_i$ were defined as unit vectors linking the peptide's termini (from the N to the C terminus, Fig. 1). To optimally select the $\hat{\mathbf{z}}_i$ vectors, other choices were investigated: vectors linking the carbonyl C to the amide N of each residue ("amide" vectors) as well as vectors lying along the carbonyl bonds. Similar results were obtained with the three different choices of $\hat{\mathbf{z}}_i$. However, due to the atomic connectivity along the backbone the "amide" vectors are not fully independent; their orientations are strongly correlated and the description of the ordered macrostates results less precise. The same is true for the "carbonyl" vectors. Hence, vectors linking peptide's termini were preferred.
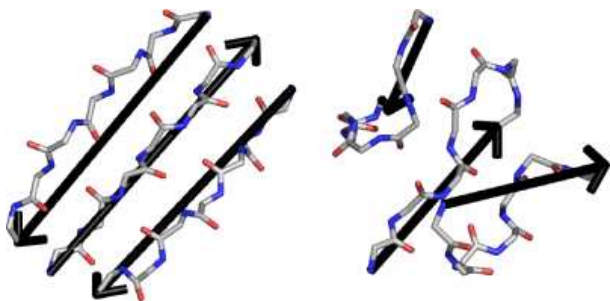

FIG. 1: Pictorial representation of the molecular vectors $\hat{\mathbf{z}}_i$ (black arrows) used to compute the order parameters $\overline{P_1}$ and $\overline{P_2}$. $\hat{\mathbf{z}}_i$ vectors are defined as full length peptide vectors (linking the peptide's termini) and allow to clearly discriminate between ordered (left, $\overline{P_2} = 0.87$) and disordered (right, $\overline{P_2} = 0.46$) conformations of the system. (The pictures were drawn using the program PyMOL [65]).

The order parameters (Eq. 3, 4) change value on going from one order macrostate to the other and should vanish when the transition to a fully isotropic state takes place. They describe different orientational properties of the system and yield useful and complementary information. The nematic $\overline{P_2}$ describes the orientational order of the system and discriminates between ordered and disordered conformations. The polar $\overline{P_1}$ describes the polarity

of the system, i.e., how much the molecular vectors ($\hat{\mathbf{z}}_i$) point in the same direction, and discriminates between parallel and antiparallel/mixed ordered aggregates.

Peptides

To evaluate the reliability of amyloidogenic propensity estimations, four oligomeric peptide systems were considered in this study: the amyloid-forming heptapeptide GNNQQNY and the soluble nonapeptide SQNGNQQRG both from the yeast prion Sup35 (residues 7-13 and 17-25 with the Gln/Arg mutation at position 24, respectively) [21], the amyloidogenic poly(L-glutamine) QQQQQQQ [56] and the non amyloidogenic poly(L-alanine) AAAAAAA [57]. To reproduce the experimental conditions [21, 56, 57], the peptide systems derived from the yeast prion Sup35 were modeled without blocking groups, while the Ala and Gln repeats were both N-acetylated and C-amidated.

All simulations were performed with three peptide replicas starting from random conformations, positions, and orientations. In the initial random positions there was no intermolecular contact, i.e., the peptides were separated in space. Each system was simulated in a cubic box of 75 Å per side yielding a sample concentration of 0.012 M. Since the oligomeric systems present different molecular weights, the above reported concentration corresponds to 3.4, 3.9, 5.4 and 3.4 mg/ml for GNNQQNY, SQNGNQQRG, QQQQQQQ and AAAAAAA, respectively.

Analysis tools

The aggregation contacts, radius of gyration and order parameters analysis was carried out with a GPL licensed program [58] developed in house to manipulate and analyze MD trajectories. The program is optimized for speed and ease of usage so that it allows extensive processing of large amounts of data and straightforward addition of new analysis tools. Compared to other available programs [38, 59], the analysis of MD trajectories is much faster.

RESULTS AND DISCUSSION

REMD diagnostics

The set of temperatures used in a REMD simulation is crucial for a correct and efficient sampling [8]. Since a simple "*a priori*" protocol for selecting the optimal temperature distribution has not been identified (yet), the choice often follows empirical considerations [8, 14, 23]: the highest temperature of the set has to be high enough
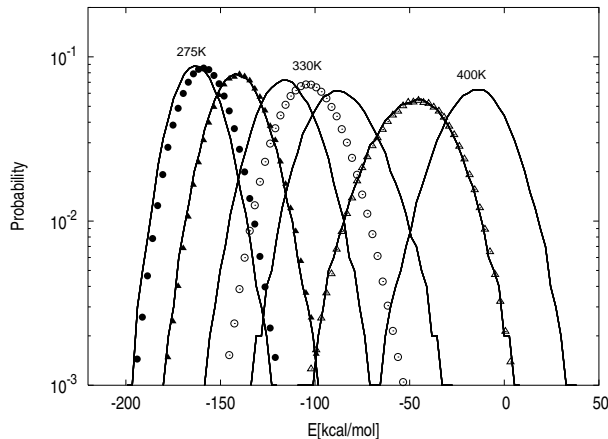
FIG. 2: Probability distribution of the effective energy for the REMD (solid lines) and the CTMD control simulations (filled circles, filled triangles, empty circles and triangles for 275, 296, 330 and 371 K, respectively). The REMD distributions correspond to the following temperatures (from left to right): 275, 296, 319, 344, 371, and 400 K. The asymmetry of the curves and the temperature dependence of the distributions indicate the presence of a phase transition around 371 K (see text).
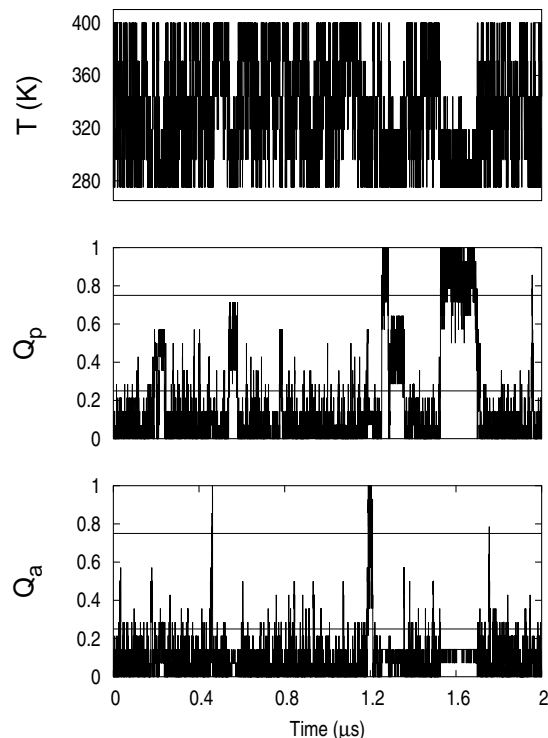

FIG. 3: Time series of (from top to bottom) the temperature T, the fraction of in-register parallel contacts Q_p, and the fraction of in-register antiparallel contacts Q_a for a REMD replica. Along the trajectory, replicas realize the desired free random walk in temperature space (top) so that an efficient sampling of the ordered aggregates is allowed (peaks in Q_p and Q_a plots). Horizontal lines in the time series of the fraction of aggregation contacts indicate the upper/lower thresholds used to define the ordered aggregation/disaggregation events.

to overcome energy barriers, while the lowest temperature has to allow the exploration of minima. However, given a fixed number of replicas the temperature range cannot be too wide. Temperature values need to be close enough to make the energy histograms overlap (see Fig. 2) in order to guarantee a high number of temperature swaps during a simulation run. In this study, a set of six temperature values ranging from 275 to 400 K has been selected (see Methods). The time series of temperature exchanges for one of the six replicas is shown in Fig. 3. During the simulation, each replica visits all the temperatures of the set several times realizing the desired free random walk in temperature space [6].

Symbols in Fig. 2 show the results from CTMD simulations carried out at 275 (filled circles), 296 (filled triangles), 330 (empty circles) and 371 K (empty triangles). At 330 K, the CTMD effective energy distribution is located between the REMD distributions extracted at 319 and 344 K and shows a consistent functional profile. At 371 K, CTMD and REMD effective energy distributions overlap. Therefore, the energetic properties of an aggregating system sampled by a REMD simulation at medium and high temperatures correspond to those observed in CTMD simulations. However, approaching the physiologically relevant conditions the CTMD distributions tend to shift toward less favorable energies (Fig 2, filled symbols). CTMD at low temperature can get trapped in local energy minima and REMD is superior in sampling conformational space [6, 14].

The time series of the fraction of in-register parallel contacts (Q_p) and in-register antiparallel contacts

(Q_a) have been monitored along the REMD trajectories (Fig. 3). A total of 14 IP and 15 IA aggregation events have been observed along the total simulation time of 12 μs (see Table I). The average aggregation time (see Methods) was 0.74 μs for IP and 0.75 μs for IA arrangements. The average aggregation time determined from the REMD simulation is similar to the values obtained from 34 μs CTMD simulations at 330 K. It is worth noting that in a preliminary REMD run with higher temperatures values ($6 \times 1\mu s$, 319-465 K; data not shown) only 3 IP and 4 IA aggregation events were sampled. The temperature range is crucial in REMD and it has to be carefully chosen in order to speed up the conformational search of relevant states [60], i.e., the *ordered states* when studying aggregation. To bias the search toward conditions where ordered states are more probable, the temperature was set to lower values (275 to 400 K as mentioned above) and the sampling of aggregation events turned out substantially improved.

Fig. 4 shows the projections of the free energy surface along Q_p and Q_a for both REMD and CTMD trajecto-
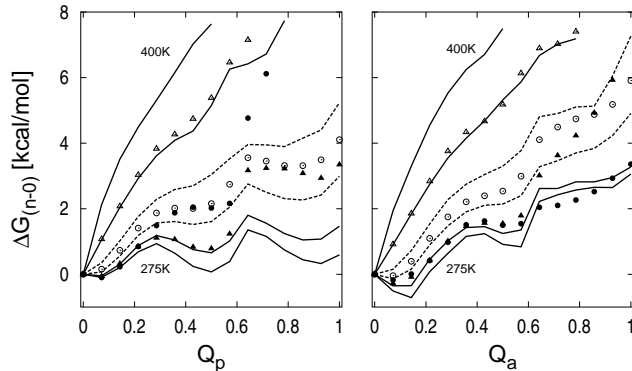
FIG. 4: Free-energy projections along the fraction of in-register parallel contacts Q_p (left) and in-register antiparallel contacts Q_a (right). Conformations with zero in-register contacts were chosen as reference states. $\Delta G_{(n-0)}$ was computed as $-k_B T \ln(N_n/N_0)$, where N_n indicates the number of conformations with n contacts and k_B is the Boltzmann constant. REMD data are shown in solid lines for all the temperature values except for 319 and 344 K which are in dashed lines. CTMD data are shown with symbols (filled circles, filled triangles, empty circles and triangles for 275, 296, 330 and 371 K, respectively).

ries. The profiles indicate that the structural properties of the aggregating system sampled by a REMD simulation correspond to those observed in CTMD simulations only at high and medium temperatures. At 371 K, CTMD and REMD free energy projections overlap. At 330 K, the CTMD free energy profiles (empty circles) are correctly placed between REMD projections at 319 and 344 K (dashed lines) and show patterns characterized by a well defined local minimum at $Q_p > 0.7$ and a monotonic uphill trend along Q_a, fully consistent with the profiles extracted from the REMD simulation. However, at low temperature (275 and 296 K) the free energy profiles extracted from CTMD and REMD trajectories are not consistent any more and the most "relevant" conformations, which correspond to in-register parallel and antiparallel arrangements ($Q_p, Q_a > 0.7$), are not correctly sampled by CTMD (Fig 4, filled symbols).

Temperature dependence of ordered amyloid peptide aggregation

Since the energetic and structural properties of the system are not artificially altered (see previous subsection), the REMD approach allows to evaluate thermodynamic quantities as a function of temperature in the chosen range [6]. From the REMD simulation performed for this study, the properties of interest have been extracted at any temperature of the set and the aggregation of

the amyloid-forming peptide GNNQQNY has been monitored in temperature space (275-400 K). This analysis gives interesting insights into the amyloid aggregation process.

The effective energy histograms shown in Fig. 2 are not symmetrically distributed around their mean value and their shape varies with temperature. The distributions, in fact, broaden toward higher energy values at low temperature (275-344 K) and toward lower energy values at high temperature (371-400 K). Moreover, by increasing temperature they progressively become lower and broader till the value of 371 K is reached. Mitsutake *et al.* have interpreted such a behavior as the evidence of a phase transition [61]. To characterize the transition, the radius of gyration R_g of the oligomeric system was considered and free energy projections along R_g were plotted (see Fig. 5). Conformations of the system producing non interacting peptides, namely conformations where all inter-peptide atomic distances are larger than the long-range interactions cutoffs (7.5 Å in this case), were used to determine R_g^C, i.e., the lowest detected radius of gyration for isolated peptides (see Fig. 5). The existence of two macrostates in equilibrium has been revealed: the first, named *uncondensed state*, includes high energy conformations with one or more isolated peptides ($R_g > R_g^C$); the second, named *condensed state*, consists of low energy conformations with aggregated peptides ($R_g < R_g^C$). For entropic reasons, the *uncondensed state* is preferred at high temperature. By cooling down, the *condensed state* is increasingly stabilized and around 371 K the fluctuations of R_g show a well defined peak highlighting the presence of the *condensation* transition (see Fig. 5). The equilibrium between the *condensed* and the *uncondensed* macrostates is clearly concentration dependent. If the concentration of amyloid-forming units increases, the equilibrium is moved toward the *condensed state* and the aggregation process is favored.

The free energy profiles along Q_p and Q_a at various temperatures help in understanding how the nucleation process evolves upon peptides *condensation*. At values of 400, 371 and 344 K both projections show steep uphill patterns with a single free energy minimum at $Q_p \approx Q_a \approx 0$ (see Fig. 4). This means that upon *condensation* the peptides are still more likely to form disordered aggregates characterized by non-specific interactions than amyloid-forming nuclei. In this range of temperatures, the enthalpic contribution due to in-register backbone or side-chain interactions does not dominate the entropic one and the growth of ordered nuclei is forbidden. However, when the temperature decreases the entropic contribution becomes less important and ordered in-register aggregates start forming. As shown in Fig. 4 in fact, below 330 K two and one additional free energy minima appear in the projection along Q_p and Q_a, respectively. The observed minima correspond to in-register parallel ($Q_p > 0.7$) and in-register mixed or

out-of-register ($0.3 \leq Q_p \leq 0.7$ and $0.4 \leq Q_a \leq 0.7$) arrangements and strongly suggest that the three-peptide system moves toward a higher degree of order when approaching the physiologically relevant conditions.

The simulation results indicate that in the early steps of amyloid aggregation a *condensation* stage toward disordered aggregates precedes the nucleation process and the disorder-order transition, in agreement with experimental evidence [62].
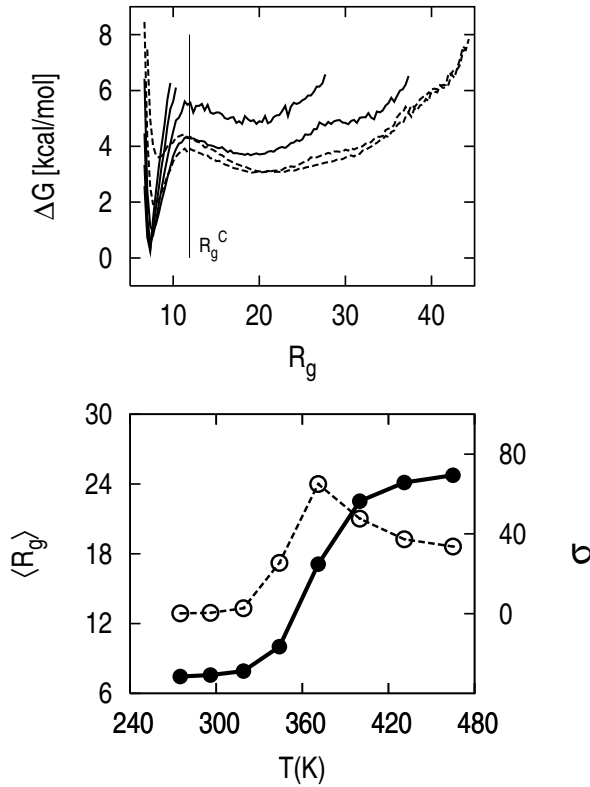

FIG. 5: (Top) Free-energy projections along the radius of gyration of the oligomeric system R_g computed from REMD trajectories. Solid lines correspond to temperature values below the *condensation* temperature (275-344 K); dashed lines correspond to temperature values above the transition temperature (371 and 400 K). The lowest radius of gyration for the *uncondensed* state is shown as a vertical line ($R_g{}^C = 11.9$ Å). (Bottom) Temperature dependence of the average radius of gyration $\langle R_g \rangle$ (filled circles) and its fluctuations σ (empty circles). The behavior of $\langle R_g \rangle$ and σ indicates the presence of a phase transition around 371 K between a *condensed* (low T) and an *uncondensed* phase (high T). Fluctuations of the radius of gyration σ are computed as $\langle R_g^2 \rangle - \langle R_g \rangle^2$. Data at 431 and 465 K were obtained from a preliminary REMD run carried out in a higher temperature range ($6 \times 1\mu s$, 319, 344, 371, 400, 431 and 465 K).

Disorder-order transition

In the early steps of aggregation, amyloidogenic peptides assemble into highly ordered β-sheet structures [21, 30]. During the assembly, the peptides tend to align adopting an extended β-strand conformation and a remarkable change in the local orientational order occurs. The aggregation of amyloid-forming peptides may then be interpreted as an order transition and orientational order parameters are suitable to monitor the time evolution of the process. Two orientational order parameters were employed and free energy projections are shown in Fig. 6. Along $\overline{P_2}$, the free energy profiles show a first broad minimum at $\overline{P_2} \approx 0.5$ for any temperature of the set and a second narrower one at $\overline{P_2} \approx 0.9$ for T values below 330 K. The first corresponds to a large free energy basin where orientational order is absent while the second corresponds to a smaller and well defined basin with a high orientational degree of order. Although the order parameters should vanish when order is absent, Fig. 6 shows that this is not the case when the number of vectors is small. Since only three peptides were simulated, a "background" order was always detected and the free energy minimum describing the *disordered state* is placed at $\overline{P_2} \approx 0.5$ which is consistent with the value of $\sqrt{81/40\pi\,N}$ expected for a completely randomly oriented array of N molecules [63]. The order parameter $\overline{P_2}$ shows the existence of two macrostates in equilibrium: the *disordered state* with a high entropy content, which corresponds to the global minimum of the free energy surface at high temperature, and the *ordered state* which becomes the global free energy minimum at low temperature. Interestingly, the free energy profiles along Q_p and Q_a do not lead to the same conclusion and the observed in-register arrangements correspond to local minima of the free energy surface (see Fig. 4).

Along $\overline{P_1}$, two narrow and well distinct minima corresponding to ordered macrostates at different polarity appear on the free energy projections (Fig. 6). The first, displayed at $\overline{P_1} \approx 0.35$, describes a free energy basin with a high-order and low-polarity content. Conversely, the second, displayed at $\overline{P_1} \approx 0.95$, corresponds to a basin with a high-order and high-polarity content. The order parameter $\overline{P_1}$ discriminates between parallel and antiparallel/mixed ordered conformations and provides complementary information since it allows to further characterize the *ordered state*.

Symbols in Fig. 6 show the free energy projections along the order parameters from CTMD simulations. Once again, the comparison with REMD profiles indicates that isothermal MD (filled symbols) does not sample the ordered aggregates with their correct statistical weight close to the physiological temperature range.

The REMD free energy profiles along $\overline{P_1}$ show that at low temperature (275 and 296 K) both polar macrostates

are highly populated. In the investigated temperature range, the system does not show an overall polar degree and frequent jumps between ordered states characterized by different polarity are observed. This suggests that below the order transition the equilibrium between polar macrostates might help amyloidogenic systems overcoming the entropy loss occurring during nucleation. In other words the growth of amyloid-forming nuclei might have an entropically favorable component due to the multiple ordered macrostates.

Sequence dependence of amyloidogenic propensity

Free energy projections along the nematic order parameter $\overline{P_2}$ show how the equilibrium between the *ordered* and *disordered state* changes in temperature space (Fig. 6). Upon cooling, the statistical weight of the *ordered state* increases and the mean of the $\overline{P_2}$ distribution moves toward higher values. The value of $\langle \overline{P_2} \rangle$, where $\langle \cdots \rangle$ indicates the average over the canonical ensemble, is then related to the thermodynamic stability of the *ordered state* and could be used to measure the amyloidogenic propensity of the system. $\langle \overline{P_2} \rangle$ values computed at different temperatures from REMD trajectories of the amyloid-forming peptide GNNQQNY are shown in Fig. 7 with filled circles. At high temperature, the $\langle \overline{P_2} \rangle$ values are close to 0.5 because no orientational order is present, and the system does not show amyloidogenicity. By decreasing temperature, the amyloidogenic
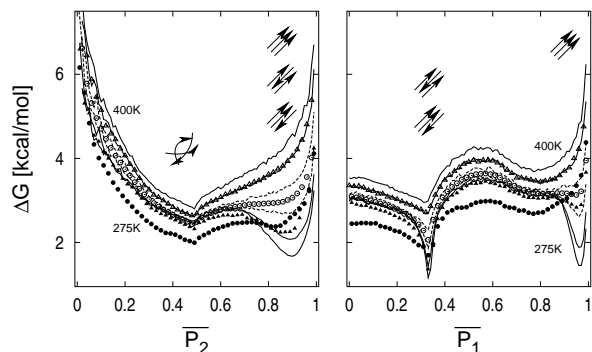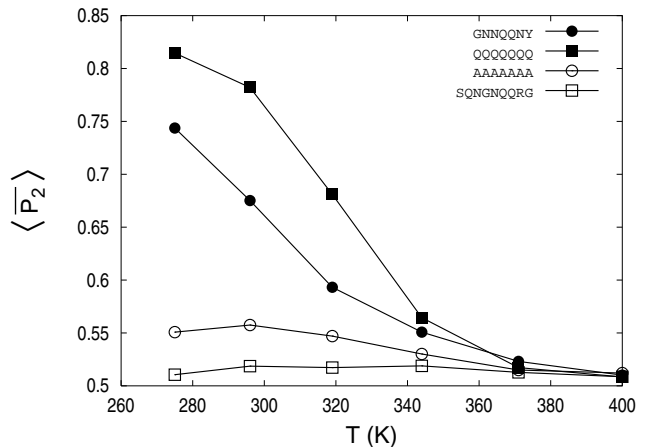

FIG. 7: Temperature dependence of the nematic order parameter $\langle \overline{P_2} \rangle$ averaged over the canonical ensembles sampled by REMD for four oligomeric peptide systems. $\langle \overline{P_2} \rangle$ estimates the amyloidogenic propensity of peptide systems and discriminates between amyloidogenic (GNNQQNY and QQQQQQQ) and non amyloidogenic (SQNGNQQRG and AAAAAAA) sequences in agreement with experimental data [21, 56, 57].

propensity grows and becomes increasingly larger until the order transition is completed. At physiologically relevant conditions, $\langle \overline{P_2} \rangle \approx 0.65$ and the system is highly amyloidogenic in agreement with experimental data [21].

Since the orientational order parameters do not depend on the peptide sequence and length, the reliability of the predictions could be further tested in sequence space. The REMD protocol was then applied to three additional oligomeric peptide systems (see "Methods") and $\langle \overline{P_2} \rangle$ values were evaluated to measure and compare amyloidogenic propensities. The testing set comprises a nonapeptide from the yeast prion Sup35 (SQNGN-QQRG) experimentally studied by Balbirnie *et al.* [21] and two heptapeptides (QQQQQQQ and AAAAAAA). Glutamine and alanine homopolymers flanked by basic residues to improve solubility have been investigated by Perutz *et al.* [56, 57].

Experimentally, the nonapeptide SQNGNQQRG shows solubility *in vivo* and *in vitro* and no formation of amyloid fibrils [21]. In agreement with these findings, $\langle \overline{P_2} \rangle$ is smaller than 0.55 in the whole temperature range (Fig. 7, empty squares) and the system is considered as non-amyloidogenic. The number of aggregation events and the average life time of aggregation extracted from REMD trajectories and reported in Table I. Remarkably, these quantities show that non-amyloidogenic sequences, i.e, SQNGNQQRG and AAAAAAA, do transiently assemble in a β-sheet conformation but still remain soluble because their ordered aggregates do not correspond to well defined free energy minima.

CD spectra, electron micrographs and X-ray diffraction photographs showed that poly(L-glutamine) peptides aggregate in solution at both pH 7.0 and 3.0 form-

FIG. 6: Free-energy projections along the nematic ($\overline{P_2}$, left) and the polar ($\overline{P_1}$, right) order parameters. REMD data are shown in solid lines for all the temperature values except for 319 and 344 K which are in dashed lines. CTMD data are shown with symbols (filled circles, filled triangles, empty circles and triangles for 275, 296, 330 and 371 K, respectively). Schematic representations of the aggregates (black arrows) are depicted to show that order parameters yield complementary information: $\overline{P_2}$ discriminates between ordered and disordered conformations while $\overline{P_1}$ discriminates between parallel and antiparallel/mixed ordered aggregates.
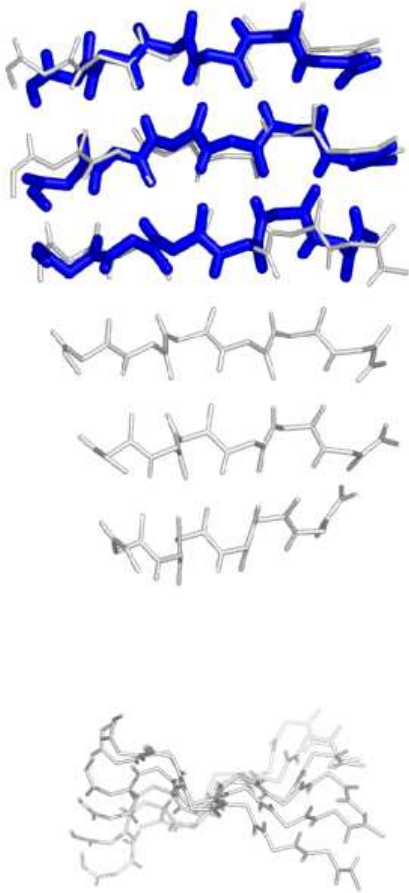
FIG. 8: (Top) Snapshots of ordered aggregates of three (thick sticks) and six (thin sticks) amyloidogenic SYVIIE peptides [64] extracted from CTMD simulations at 330 K. The simulations were performed at a sample concentration of 5 mg/ml. The overall conformation and twist of the three-stranded and six-stranded parallel β-sheets are indistinguishable. (Bottom) The six-stranded β-sheet upon 90° rotation to better visualize the twist. (The pictures were drawn using the program PyMOL [65]).

ing tightly linked β-sheets structures [56]. In particular, the X-ray diffraction picture exhibits a fiber diagram of the cross-β type distinctive of amyloid fibrils. On the other hand, poly(L-alanine) doesn't display amyloidogenicity and CD spectra showed α-helical structures at all pHs [57]. Again, the $\langle \overline{P_2} \rangle$ patterns shown in Fig. 7 (filled squares and empty circles) are consistent with experimental findings and correctly indicate amyloidogenicity only for QQQQQQQ.

Interestingly, Fig. 7 allows also to compare between amyloidogenic sequences. In fact, according to the $\langle \overline{P_2} \rangle$ patterns the glutamine repeat is more amyloidogenic than GNNQQNY at physiologically relevant conditions. To our knowledge, no experimental data is available to verify this finding. Testing of this prediction is a challenge for experimentalists.

CONCLUSIONS

The present study shows that atomistic REMD simulations with implicit solvent allow to sample the early steps of ordered aggregation of amyloidogenic peptides at physiologically relevant temperatures. The free energy profiles projected along structural and orientational progress variables are essentially the same in REMD and CTMD. The discrepancies at temperature values below 330 K are due to the limitations in sampling in CTMD simulations which indicates that REMD is a more efficient approach in the physiological range.

The early steps of amyloidosis can be interpreted as a condensation followed by an order transition. Therefore, the REMD simulation results were analyzed with two order parameters originally introduced to study liquid crystals. Interestingly, the nematic order parameter averaged over a canonical ensemble is able to discriminate amyloidogenic from soluble peptides in agreement with experimental data.

Although the present study was performed with three peptides for reasons of computational efficiency, the description of the ordered aggregates is likely to be independent of the size of system, i.e., the number of simulated peptide replicas. Very recent MD simulations of the amyloidogenic SYVIIE peptide (Cecchini *et al.*, unpublished results), which has been experimentally investigated by de la Paz and Serrano [64], have shown ordered aggregates of six peptides. Interestingly, the parallel β-sheet consisting of six peptides has the same overall conformation and twist as the three-peptide aggregate (Fig. 8).

* corresponding author, tel: +41 1 635 55 21, fax: +41 1 635 68 62, e-mail: caflisch@bioc.unizh.ch

[1] D. Frenkel and B. Smit, *Understanding Molecular Simulations* (Academic Press, San Diego, 2002).

[2] B. J. Berne and J. E. Straub, Curr. Opin. Struct. Biol. **7**, 181 (1997).

[3] A. Mitsutake, Y. Sugita, and Y. Okamoto, Biopolymers **60**, 96 (2001).

[4] N. Rathore, T. A. Knotts IV, and J. J. de Pablo, J. Chem. Phys. **118**, 4285 (2003).

[5] E. Marinari and G. Parisi, Europhys. Lett. **19**, 451 (1992).

[6] Y. Sugita and Y. Okamoto, Chemical Physics Letters **314**, 141 (1999).

[7] C. D. Michele and F. Sciortino, Phys. Rev. E **65**, 051202 (2002).

[8] K. Sanbonmatsu and A. Garcia, Proteins: Structure, Function and Genetics **46**, 225 (2002).

[9] A. E. Garcia and K. Sanbonmatsu, Proc. Natl. Acad. Sci. USA. **99**, 2782 (2002).

[10] A. E. Garcia and K. Sanbonmatsu, Proteins: Structure, Function and Genetics **42**, 345 (2001).

[11] R. Zhou, B. Berne, and R. Germain, Proc. Natl. Acad. Sci. USA. **98**, 14931 (2001).

[12] A. E. Garcia and J. N. Onuchic, Proc. Natl. Acad. Sci. USA. **100**, 13898 (2003).

[13] J. W. Pitera and W. Swope, Proc. Natl. Acad. Sci. USA. **100**, 7587 (2003).

[14] F. Rao and A. Caflisch, J. Chem. Phys. **119**, 4035 (2003).

[15] W. Im and C. L. Brooks III, J. Mol. Biol. **337**, 513 (2004).

[16] C. M. Dobson, Trends Biochem. Sci. **24**, 329 (1999).

[17] M. F. Perutz, Trends Biochem. Sci. **24**, 58 (1999).

[18] C. Blake and L. Serpell, Structure **4**, 989 (1996).

[19] S. B. Malinchik, H. Inouye, K. E. Szumowski, and D. A. Kirschner, Biophys. J. **74**, 537 (1998).

[20] M. Bucciantini, E. Giannoni, F. Chiti, F. Baroni, L. Formigli, J. Zurdo, N. Taddei, G. Ramponi, C. M. Dobson, and M. Stefani, Nature **416**, 507 (2002).

[21] M. Balbirnie, R. Grothe, and D. Eisenberg, Proc. Natl. Acad. Sci. USA. **98**, 2375 (2001).

[22] R. A. Broglia, G. Tiana, S. Pasquali, H. E. Roman, and E. Vigezzi, Proc. Natl. Acad. Sci. USA. **95**, 12930 (1998).

[23] D. Bratko and H. W. Blanch, J. Chem. Phys. **118**, 5185 (2003).

[24] G. Giugliarelli, C. Micheletti, J. R. Banavar, and A. Maritan, J. Chem. Phys. **113**, 5072 (2000).

[25] P. M. Harrison, H. S. Chan, S. B. Prusiner, and F. E. Cohen, J. Mol. Biol. **286**, 593 (1999).

[26] A. V. Smith and C. K. Hall, J. Mol. Biol. **312**, 187 (2001).

[27] B. Vekhter and R. S. Berry, J. Chem. Phys. **110**, 2195 (1999).

[28] M. Friedel and J. E. Shea, J. Chem. Phys. **120**, 5809 (2004).

[29] A. Fernandez and M. Boland, FEBS Lett. **529**, 298 (2002).

[30] J. Gsponer, U. Habertür, and A. Caflisch, Proc. Natl. Acad. Sci. USA. **100**, 5154 (2003).

[31] E. Paci, J. Gsponer, X. Salvatella, and M. Vendruscolo, J. Mol. Biol. (2004), in press.

[32] F. Massi, J. W. Peng, J. P. Lee, and J. E. Straub, Biophys. J. **80**, 31 (2001).

[33] B. Ma and R. Nussinov, Protein Science **11**, 2335 (2002).

[34] B. Ma and R. Nussinov, Proc. Natl. Acad. Sci. USA. **99**, 14126 (2002).

[35] D. Klimov and D. Thirumalai, Structure **11**, 295 (2003).

[36] G. Tiana, F. Simona, R. A. Broglia, and G. Colombo, J. Chem. Phys. **120**, 8307 (2004).

[37] P. Ferrara, J. Apostolakis, and A. Caflisch, Proteins: Structure, Function and Genetics **46**, 24 (2002).

[38] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus, J. Comput. Chem. **4**, 187 (1983).

[39] E. Neria, S. Fischer, and M. Karplus, J. Chem. Phys. **105**, 1902 (1996).

[40] W. Hasel, T. F. Hendrickson, and W. C. Still, Tetrahedron Comput. Methodol. **1**, 103 (1988).

[41] T. Lazaridis and M. Karplus, Proteins: Structure, Function and Genetics **35**, 133 (1999).

[42] A. Hiltpold, P. Ferrara, J. Gsponer, and A. Caflisch, J. Phys. Chem. B **104**, 10080 (2000).

[43] P. Ferrara and A. Caflisch, Proc. Natl. Acad. Sci. USA. **97**, 10780 (2000).

[44] P. Ferrara and A. Caflisch, J. Mol. Biol. **306**, 837 (2001).

[45] J. Gsponer and A. Caflisch, Proc. Natl. Acad. Sci. USA. **99**, 6719 (2002).

[46] J. Gsponer and A. Caflisch, J. Mol. Biol. **309**, 285 (2001).

[47] J. P. Hansen and I. R. McDonald, *Theory of simple liquids* (Academic Press, Oxford, 1990), 2nd ed.

[48] J. P. Ryckaert, G. Ciccotti, and H. J. C. Berendsen, J. Comp. Phys. **23**, 327 (1977).

[49] M. E. Rose, *Elementary Theory of Angular Momentum* (Wiley, 1957).

[50] C. Zannoni, *The Molecular Physics of Liquid Crystals* (Academic London, 1979), chap. 3.

[51] S. Chandrasekhar, *Liquid Crystals* (Cambridge University Press, Cambridge, England, 1992).

[52] P. G. de Gennes and J. Prost, *The Physics of Liquid Crystals* (Oxford University Press, Oxford, 1993), 2nd ed.

[53] C. Zannoni, J. Mater. Chem. **11**, 2637 (2001).

[54] R. Berardi, L. Muccioli, and C. Zannoni, Chem. Phys. Chem. **5**, 104 (2004).

[55] M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids* (Oxford Science Publications, Oxford, UK, 1987).

[56] M. F. Perutz, T. Johnson, M. Suzuky, and J. T. Finch, Proc. Natl. Acad. Sci. USA. **91**, 5355 (1994).

[57] M. F. Perutz, B. J. Pope, D. Owen, E. E. Wanker, and E. Scherzinger, Proc. Natl. Acad. Sci. USA. **99**, 5596 (2002).

[58] M. Seeber, M. Cecchini, F. Rao, G. Settanni, and A. Caflisch (2004), in preparation.

[59] W. Humphrey, A. Dalke, and K. Schulten, Journal of Molecular Graphics **14**, 33 (1996).

[60] M. K. Fenwick and F. A. Escobedo, J. Chem. Phys. **119**, 11998 (2003).

[61] A. Mitsutake, Y. Sugita, and Y. Okamoto, J. Chem. Phys. **118**, 6676 (2003).

[62] T. R. Serio, A. G. Cashikar, A. S. Kowal, G. J. Sawicki, J. J. Moslehi, L. Serpell, M. F. Arnsdorf, and S. L. Lindquist, Science **289**, 1317 (2000).

[63] T. P. Doerr, D. Herman, H. Mathur, and P. L. Taylor, Europhys. Lett. **59**, 398 (2002).

[64] M. Lopez de la Paz and L. Serrano, Proc. Natl. Acad. Sci. USA. **101**, 87 (2004).

[65] W. DeLano, *The PyMOL Molecular Graphics System* (DeLano Scientific, San Carlos, CA, USA, 2002).