Molecular Dynamics Simulation Suggests Possible Interaction Patterns at Early Steps of β_2-Microglobulin Aggregation

Federico Fogolari,* Alessandra Corazza,* Paolo Viglino,* Pierfrancesco Zuccato,† Lidia Pieri,* Pietro Faccioli,†‡ Vittorio Bellotti,§ and Gennaro Esposito*

*Dipartimento di Scienze e Tecnologie Biomediche, Università di Udine, Udine, Italy; †European Centre for Theoretical Studies in Nuclear Physics and Related Areas (E.C.T.), Villazzano (Trento), Italy; ‡Dipartimento di Fisica and Istituto Nazionale di Fisica Nucleare, Università degli Studi di Trento, Povo (Trento), Italy; and §Department of Biochemistry, University of Pavia, Laboratori di Biotecnologie, Istituto di Ricovero e Cura a Carattere Scientifico Policlinico San Matteo, Pavia, Italy

ABSTRACT Early events in aggregation of proteins are not easily accessible by experiments. In this work, we perform a 5-ns molecular dynamics simulation of an ensemble of 27 copies of β_2-microglobulin in explicit solvent. During the simulation, the formation of intermolecular contacts is observed. The simulation highlights the importance of apical residues and, in particular, of those at the N-terminus end of the molecule. The most frequently found pattern of interaction involves a head-to-head contact arrangement of molecules. Hydrophobic contacts appear to be important for the establishment of long-lived (on the simulation timescale) contacts. Although early events on the pathway to aggregation and fibril formation are not directly related to the end-state of the process, which is reached on a much longer timescale, simulation results are consistent with experimental data and in general with a parallel arrangement of intermolecular β-strand pairs.

INTRODUCTION

β_2-microglobulin is the major component of amyloid fibrils in dialysis-related amyloidosis (1). The molecular mechanism by which β_2-microglobulin aggregates form and eventually lead to fibril formation is largely unknown, although many structural studies have provided clues about this issue (2–4).

Early events on the pathway to fibril formation are very difficult to study because of the intrinsic dynamic nature of forming oligomers (5). Molecular dynamics simulations allow a detailed view of intermolecular association, notwithstanding the current limitations in the treatable size of the system and in the length of the simulation.

A number of simulations of wild-type, mutants, and fragments of β_2-microglobulin have indeed been reported in the literature (6–9). Much attention has been focused on conformational changes that can be related to fibril formation. To observe conformational transitions to disordered states of the molecule, high-temperature simulations have been performed by Ma and Nussinov (6) and Daggett and co-workers (7). Ma and Nussinov were able to reproduce many experimental observations in their simulations, the most notable of which is the β-sheet to α-helix transition of strands C, C′, and D.

The unfolding simulations of Daggett and co-workers were performed at low pH for four of the best characterized fibrillogenic proteins (7). Low pH was simulated by protonating all acidic residues and histidines. Surprisingly, in all simulations, formation of the unusual α-pleated sheet was observed. This unusual structure, characterized by alternating right- and left-handed helical backbone conformations (α-sheet), has peptidic amides and carbonyls on opposite sides of the sheet, giving rise to a strong electrostatic dipole. In a subsequent article by the same group (9), the conformational dynamics of β_2-microglobulin was specifically analyzed by molecular dynamics simulations at low and neutral pH, at 310 K and 498 K. Very little α-sheet was observed in the neutral pH trajectories. At low pH, instead, the alternating α_R, α_L conformation proved significantly populated. A self-assembly mechanism was thus proposed that involves only α-sheet intermediates, i.e., those resulting from evolution of the early native-like unfolding intermediate.

Molecular dynamics simulations of β_2-microglobulin aggregation and fibril formation are much more difficult, because the time required for these processes is orders of magnitude longer than the time presently accessible by molecular dynamics simulations. The problem may be circumvented by exploring the configurational space of ensembles of molecules by docking techniques. Using protein-protein docking techniques, Nussinov and co-workers were able to propose a model for the protofibril that agrees with experimental observations and with the pattern of conservation of contacts in the structurally aligned protein family (10,11).

Simulations of ensembles of peptides to study their aggregation behavior have been performed before. It is worth noting that molecular dynamics simulations can only sample a time range of at most hundreds of nanoseconds and all-atom simulations have therefore sampled only early events in aggregation. To speed up the occurrence of interesting

Submitted October 2, 2006, and accepted for publication November 2, 2006.

Address reprint requests to Federico Fogolari, Dipartimento de Scienze e Tecnologie Biomediche, Università di Udine, Piazzale Kolbe 4, 33100 Udine, Italy. E-mail: ffogolari@mail.dstb.uniud.it.

Pierfrancesco Zuccato's present address is EXADRON, a division of EuroTech S.p.A., Via Jacopo Linussio, 1-33020 Amaro (UD), Italy. E-mail: p.zuccato@exadron.com.

Lidia Pieri's present address is Istituto Nazionale di Astrofisica, Astronomical Observatory of Padova, Vicolo dell'Osservatorio 5, I-35122 Padova, Italy. E-mail: lidia.pieri@oapd.inaf.it.

© 2007 by the Biophysical Society

0006-3495/07/03/1673/09 $2.00

doi: 10.1529/biophysj.106.098483

events, a number of nonstandard approaches to molecular dynamics simulations have been adopted. The most significant are:

1. designing initial conformations that could represent (12–16) or lead to (17,18) an ordered aggregate;
2. simulation of the most probable pathway between two boundary (initial and final) states (19) (related to item 1, and a very interesting, approach);
3. constraining part of the system into a preformed β-sheet or in extended conformation (19,20);
4. using (at least in the early stage of the simulation) simplified models of proteins (21–26) with smoother energy landscape (see Kolinski and Skolnick (21));
5. adopting an implicit solvent representation (27–29);
6. using, for thermodynamic analysis, molecular mechanics and implicit solvent models (13,19,24).

All models (including most detailed explicit solvent simulations) could generate artifacts, and therefore comparison with the available body of experimental data is fundamental for computational studies. Without going into the details of all these simulation studies, a few conclusions seem to hold generally:

1. many amyloidogenic peptides do aggregate spontaneously during the simulation (24–28,30);
2. the presence of β-sheet structures drives disruption of α-structure (or collapsed conformations) (22,31) and formation of β-sheet aggregates (20,24, 5,29). This effect could, however, arise from molecular crowding (see, e.g., Paci et al. (29));
3. although the cross-βstructure, involving backbone atoms, is the common feature of all amyloid aggregates, side-chain interactions play an important role in many studies (12,14–16,19,24,27,30).

Comparison of the features of the aggregates shows, however, a large diversity depending on the specific system studied. β-hairpin and coiled β-sheet aggregates (see e.g., Ding et al. (24)), and parallel and antiparallel arrangements of extended peptides have been found (see *e.g.* (29)).

In this work, we performed 5-ns molecular dynamics simulations of an ensemble of 27 β_2-microglobulin molecules, equally spaced in a cubic arrangement and randomly oriented. The system contained a half-million atoms and required extensive computational resources. The analysis of molecular dynamics trajectories suggests interaction patterns at early steps of β_2-microglobulin aggregation.

To avoid misunderstanding, it should be clarified that the pathway leading to amyloid formation involves much longer timescale processes, like those involving partial unfolding, or proline *cis-trans* isomerization, as recently suggested (32,33). This article does not address these processes, but rather the interaction between folded proteins in solution, which lead to the formation of dynamic oligomers observed in solution before fibril formation (5).

MATERIALS AND METHODS

Simulation system setup

The crystal structure of β_2-microglobulin was obtained from the Research Collaboratory for Structural Bioinformatics repository (PDB code 3HLA, chain B) (34). Protons were added using the pdb2gmx utility of the GROMACS simulation package (35). The program assigned the uncharged protonation state with proton on N_ϵ to His-13 and His-31, the uncharged protonation state with proton on N_δ to His-51, and the charged protonation state to His-84. This procedure was adopted because it is widely accepted, although it should be mentioned that the charged form of His-84 is not consistent with the computed pKa of 4.1 reported by Giorgetti et al. (36).

The protonated structure was then used to generate a topology and coordinate file using the psfgen utility of NAMD simulation software (37). The resulting system included 1624 atoms. The structure was used to compute the electrostatic potential around the molecule using the program UHBD (38). To neutralize the -1 protein charge, a sodium ion was placed in the most negative potential point on a surface enclosing the molecule at a distance of 7 Å from any heavy atom of the protein. This system (protein + ion) was rotated randomly in space using an algorithm developed by J. Arvo and available from the author's web page (http://www.cs.caltech.edu/~arvo/). Twenty-seven randomly rotated systems were arranged in a $3 \times 3 \times 3$ cubic lattice with 50 Å spacing to avoid steric hindrances. An alternative choice would be to arrange molecules in a smaller (say $2 \times 2 \times 2$) lattice and perform a longer molecular dynamics simulation. In this case, however, the cubic box containing a molecule would be in contact with both the box containing another molecule and its periodic image. To avoid this in a cubic arrangement we chose the minimal $3 \times 3 \times 3$ lattice.

The system was relaxed by 200 conjugate-gradient minimization steps. The dielectric was set to 10.0 to minimize the effect of missing solvent, and the cutoff was 12.0 Å. The minimized system was solvated using the module solvate in the VMD software package (39) in a box with margins at 5.0 Å distance from any solute atom. The system contained 484,407 atoms. The solute molecules, including ions, were fixed, and the system was energy-minimized by 300 conjugate-gradient steps using periodic boundary conditions and the particle mesh Ewald (PME) method for electrostatic interactions (40). PME employed a grid of $192 \times 192 \times 192$ points. The PME tolerance was set to 10^{-6}, which, together with the cutoff of 12 Å, resulted in an Ewald coefficient of 0.257952 $Å^{-1}$. The minimized system was further relaxed, keeping the solute molecules (including ions) fixed, by molecular dynamics simulation. The system was heated to 300 K in 2 ps and a further 18-ps simulation was run to let water molecules reorient, consistent with the average lifetime of a hydrogen bond in water (41). The system without restraints on solute molecules was energy-minimized by 300 conjugate-gradient minimization steps. The system was then heated to 300 K in 2 ps and a further 118-ps simulation was run to let the system equilibrate so that the production run could finally start.

In all molecular dynamics simulations, the temperature was kept constant through a simple velocity rescaling procedure, and the pressure was controlled through a Berendsen bath (42) using a relaxation time of 100 fs. The size of the box was fluctuating around $168 \times 168 \times 168$ Å within fractions of Å. The effective concentration of the protein in the simulation box was 9.5 mM, which is very high but still not unrealistic as far as in vitro studies are concerned. For instance, typical concentrations for NMR studies are in the range of 1 mM. The serum concentration in vivo (e.g., in patients undergoing hemodialysis) is much lower, in the range of 5 μM (1), although little is known about possible local concentration variations (see, e.g., the discussion in Relini et al. (43)).

RESULTS

Analysis of molecular dynamics runs

The analysis has been performed in two distinct phases: first, the features concerning isolated molecules were examined,

then intermolecular interactions were examined. We suggest that these two aspects are related to each other in that partial unfolding or mobility could be prerequisites, here and for many other proteins, for ordered aggregates to be formed.

The availability of a 5-ns molecular dynamics run for 27 replicas of the same system allows us to consider ensemble averages taken over different molecular dynamics snapshots for the same molecule or over different molecules taken from a single molecular dynamics snapshot.

Molecular dynamics snapshots have been superimposed on the deposited x-ray structure to assess the convergence of the molecular dynamics run. The average root mean-square deviation (RMSD) of the 27 molecules from the native structure along the 5-ns trajectory has been computed on C_α atoms. After a fast increase during 120 ps equilibration close to 1 Å, the RMSD from the native structure shows a steady increase that apparently reaches a plateau only in the last half-nanosecond of the run.

The RMSD versus time curve (Fig. 1) could be fitted (with a correlation coefficient of .99) by the equation, $\text{RMSD} = 1.0+1.9\times(1-\exp(-\frac{t}{5.5}))$ reaching, thus, the asymptotic RMSD of 2.9 Å. These values are qualitative because of the limited range of sampling, but nevertheless confirm that equilibration of a system by molecular dynamics can take much longer than usually assumed.

To assess the contributions of each region to the RMSD from the native structure, the ensemble of the 27 structures in the last snapshot was considered. At 5 ns, the average C_α RMSD from the starting structure, after global superposition, is displayed in Fig. 2 (*thick line*). Large RMSDs are found at the two terminals and at residues 15–18 and 85–88 corresponding to loops AB and FG, respectively. The pattern of mobility, deduced from the simulation by pairwise comparison of all molecules in the last snapshot (Fig. 2, *thin line*), shows a pattern largely overlapping with the average RMSD from the starting structure. However, small differences are observed in the region containing residues 85–88, with a deviation from native structure larger than the average RMSD. This could be a consequence of the protonation state chosen for His-84, which leads to a conformational rearrangement in this mobile loop. The largest difference between the crystallographic structure and the ensemble of conformations in this region is located at the ϕ angle of residue Q89 in the FG loop. The overall data, however, suggest that no major conformational transitions are observed during the molecular dynamics simulation and that differences from the starting structure are mainly related to the flexibility of the protein. The flexibility of the protein mostly does not match the B factor found in the crystallographic structure containing, however, also the other subunit of the major histocompatibility complex (MHC-I). The only region where large flexibility is found is at residues 15–18, although large differences are found at the N- and C-terminal regions, which prove to be flexible. This suggests that the latter two regions possess some propensity to undergo local changes that accompany the conformational transition into a fibril-competent conformation.

Torsion-angle order parameters (computed on the ensemble at the end of the simulation) are consistent with the flexibility resulting from RMSD analysis. In particular, order parameters are <0.9 for ϕ angles of residues Gly-18, Trp-60, Gly-43 and Lys-48, and for ψ angles of N- and C-terminal residues and for residues Ala-15 to Lys-19, Ile-46 to Lys-48, Ser-57, Asp-59, Trp-60, and Pro-72.

Secondary structure of the starting crystallographic structure is preserved in all 27 structures at 5 ns except for the two residues (35 and 84) at the end of their strands.

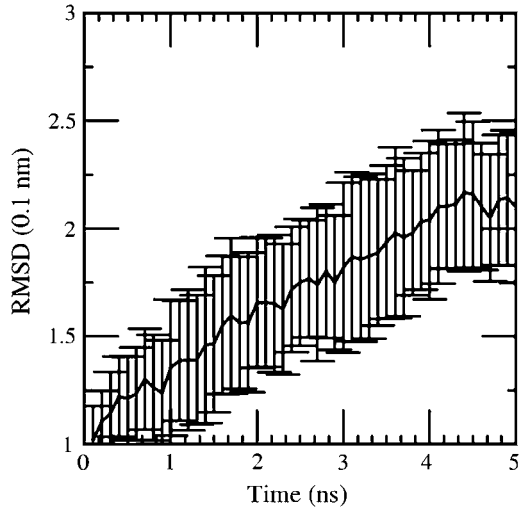

FIGURE 1 Root mean-square deviation of C_αs from the starting crystal structure (PDB code 3HLA, chain B) versus simulation time. The average and error bars are computed over the ensemble of 27 molecules.

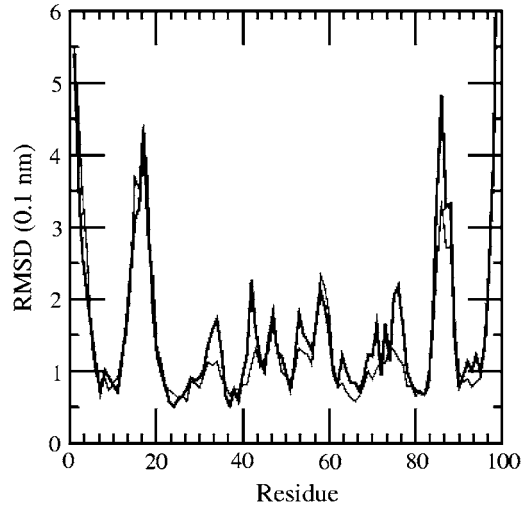

FIGURE 2 RMSD versus residue at the end of the simulation. The average deviation from the starting structure is shown by a thick line. The average deviation upon pairwise superposition of all 27 structures onto each other is shown by a thin line.

Diffusional properties

For a spherical particle, the translational (D) and rotational (D_r) diffusion coefficients are: $D = (kT/6\pi r\eta)$ and $D_r = (kT/8\pi r^3\eta)$, where k is the Boltzmann constant, T is the temperature, r is the hydrodynamic radius of the molecule, η is the viscosity (for water at room temperature, $\eta = 9 \times 10^{-4}$ Ns/m^2) (44). The translational diffusion of the molecules during simulation should take place according to the Einstein-Smoluchowski equation $\langle x^2\rangle = 6Dt$ where $\langle x^2\rangle$ is the average quadratic displacement of the molecule and t is the time (44).

The quadratic displacement of the center of mass of each molecule from the starting position (averaged over the ensemble of 27 molecules) versus time could be well fitted by a straight line with slope 127 Å^2 ns^{-1} (1.27×10^{-9} m^2 s^{-1}).

The rotational diffusion of molecules obeys the following equation (44): $\langle\cos(\theta)\rangle = \exp(-2D_r t)$, where θ is the angle between a vector rigidly rotating with the molecule with respect to its starting direction. If $R(t)$ is the rotation matrix for optimally superimposing the starting structure to the structure at time t, then the average (over all possible starting directions) of the scalar product $\vec{v} \cdot R\vec{v}$ is given by the trace of matrix R divided by 3, i.e., by $(R_{xx}+R_{yy}+R_{zz})/3$.

The plot of this quantity versus time could be well fitted by an exponential with a time constant of 9.6 ns.

The two diffusion constants allow one to estimate the hydrodynamic radius of the protein (17.5 Å) and the simulated viscosity of the solution (5.9×10^{-4} kg m^{-1} s^{-1}), which are in reasonable agreement with the size of the protein and the viscosity of pure water at room temperature, 9×10^{-4} kg m^{-1} s^{-1}.

These results support the overall accuracy, as far as ensemble dynamics is concerned, of the simulation.

Analysis of protein-protein contacts

All pairs of proteins were considered at the start of the production run and at 1-ns intervals to check contact formation. Formation and disruption of contacts are expected, based on the large movements the molecules experienced during the 5-ns simulation. A contact between two molecules is defined when the heavy atom of one molecule is within a cutoff distance from the heavy atom of the other molecule. The cutoff has been chosen as the sum of the van der Waals radii plus 1 Å. This definition of contact was found to be significant in protein-structure recognition tasks (45). At variance with the previous study (45), contacts involving backbone atoms are also considered here.

At the beginning of the production run, there are few contacts between two pairs of proteins involving just one residue pair per protein pair. After 1 ns, the number of pairs of molecules in contact increases to 3 and later up to 13 at 5 ns.

The number of residue pairs involved in contacts, however, steadily increases from 7 at 1 ns to 62 at 5 ns. This means that the average contacts per pair of molecules are steadily increasing, thus stabilizing the formed complex.

Contacts involve a large number of different residues, although only a few of them appear to be particularly important.

To increase statistical validity and to have stable complexes better represented, we took snapshots of molecular dynamics simulations at 1-ns intervals and pooled all pairs of molecules having at least one contact.

Overall, 39 pairs of contacting molecules were saved. Only two of these pairs, formed during the first nanosecond of simulation, are found at the end of the simulation. These two pairs are very interesting because together they form a network of four interacting molecules. Their properties will be discussed later.

The plot of the number of contacts (on a residue basis) versus residue reveals that residue Trp-60 plays the most important role together with nearby residues and N- and C-terminal residues (Fig. 3). Although Trp-60 is also the residue displaying the largest surface area exposed to the solvent, the number of contacts does not correlate in general directly with solvent exposure, but rather with both solvent exposure and the hydrophobic character of the residue.

When all those residues involved in more than five intermolecular residue-residue contacts are displayed in the crystal protein structure, their predominantly apical localization with respect to the β sandwich structure is strikingly apparent (Fig. 4).

The group of residues close to the N-terminus (i.e., residues 1, 3, 31, 34, 56, 58, 59, and 60) that is involved in the largest number of contacts clusters at one apex of the molecule (arbitrarily termed the head) and is a contact partner 122 times, whereas residues close to the C-terminus that are involved in the largest number of contacts (i.e., residues 16, 17, 18, 74, 75, 97, 98, and 99), at or close to the opposite apex of the molecule (termed tail), are contact partners 65 times. The overall number of contacts is 140. The location of the contacting residues in the molecular structure is shown in Fig. 4.

The preferential involvement of some regions in intermolecular contacts may be assayed by superimposing alternatively

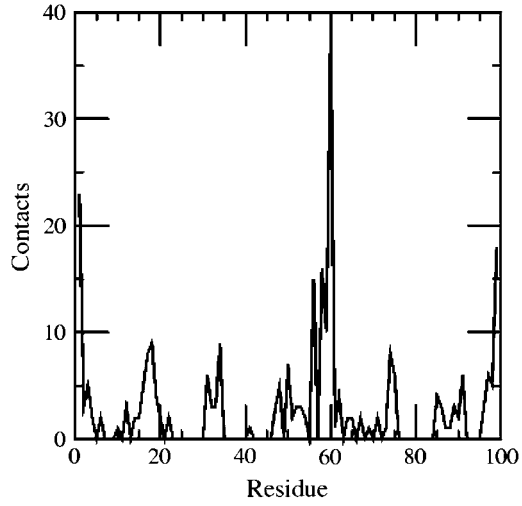

FIGURE 3 Number of contacts versus residue. For definitions of contact and molecular ensemble, see text.

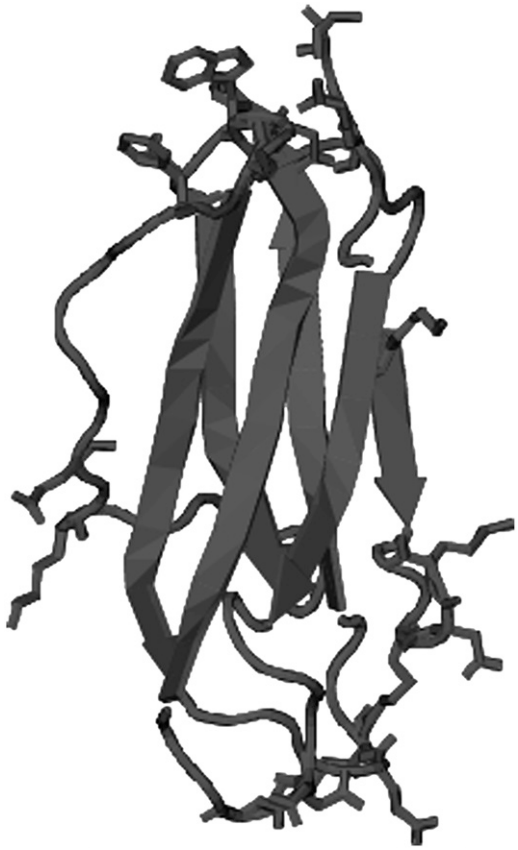

FIGURE 4 Apical location of the most important contacting residues. The side chain of residues 1, 3, 16, 17, 18, 31, 34, 48, 50, 56, 58, 59, 60, 74, 75, 91, 97, 98, 99 is shown. The N-terminal (head) apex of the molecule is displayed at the top of the figure. The C-terminal (tail) apex is displayed at the bottom of the figure.

one of the two molecules of the contacting pair onto the crystallographic structure. This procedure provides a plot of all pairs in contact (each pair is displayed twice) in a common frame of reference. A striking picture emerges with the part of the molecule close to the N-terminus most frequently involved in contacts, and the face corresponding to strands C, C′, and F having very few contacts.

The 78 pairs of molecules displayed in Fig. 5 were tentatively clustered according to distance-based methods (e.g., by hierarchical methods like the unweighted pair group method with averaging), with little effectiveness. Finally, a simple diagram pathway grouping proved more effective. In this method, the distance matrix is scanned and all those pairs whose RMSD is less than a given threshold value are grouped together. A small threshold value will result in all pairs grouped separately, whereas a large threshold value will group all pairs together. A threshold of 0.8 nm was used, which resulted in 27 clusters. Most of them were populated by just one to three pairs. The largest cluster, however, contains 22 pairs that display an N-terminus-to-N-terminus arrangement (Fig. 5). Due to the overall symmetry of the arrangement, the 22 pairs in the cluster are just 11 pairs with one or the other protein of the pair taken as a reference. It is interesting that these 11 pairs involve the two pairs that associate in a four-molecule cluster. The head-to-head arrangement is somewhat unexpected, because the dipole moment of the molecule is roughly oriented along the C-terminus-to-N-terminus direction. Salt bridge interactions between Lys-58 or the N-terminal amino group and Asp-34 can also form in this arrangement, but most relevant are hydrophobic interactions involving Trp-60, Phe-56, Ile-1, and Lys-58.

To understand how this reciprocal orientation between contacting molecules is observed in the ensemble and why it is so stable, the electrostatic potential at the surface of the molecule in its starting conformation and the isopotential curves around the molecule were computed (Fig. 6). The orientation in the figure is the same as in Fig. 4.

It is interesting to note that whereas the overall orientation of the polarity is as expected from the dipole orientation, there are both positive and negative potential regions in the N-terminal top of the molecule. This allows the molecules to establish electrostatic complementarity in head-to-head contacts.

The next three most populated clusters display tail-to-tail (Fig. 5) and antiparallel arrangements (Fig. 5). The two clusters showing tail-to-tail contacts involve exactly the same molecular pairs, the only difference being in the molecule used for superposition on the reference starting structure. These pairs are found at 3 ns and they are first stabilized by salt-bridge interactions, which are not found in subsequent snapshots. There is, however, no special trend during simulation for salt-bridge contacts, which are rather few compared to hydrophobic contacts.

The cluster showing an antiparallel arrangement of the molecule is interesting, because it would be closer to what could be expected as a starting stage for cross-β forming structure, i.e., with the β-strands running parallel or antiparallel to each other. The ensemble is, however, formed by just two pairs, which, for reasons of symmetry, are replicated by the choice of the reference molecule for superposition.

The next populated clusters show diverse orientations also involving head-to-tail contacts. It is worthy of note that only two backbone-to-backbone contacts are observed in the snapshots and they both involve the N-terminal group of Ile-1.

Dynamics of aggregation

All available evidence seems to point at hydrophobic residues exposed to solvent as the mechanism responsible for intermolecular interactions. The highest number of contacts is displayed by residues Trp-60, Ile-1, Met-99, Lys-58, and Phe-56. Except for Lys-58, these residues are mainly hydrophobic and exposed to solvent in the isolated β_2-microglobulin monomer, whereas they are involved in intermolecular contacts in the complex resolved by x-ray crystallography.

A plausible mechanism of aggregation of the protein therefore involves the ''sticky'' apical ends of β_2-microglobulin molecules, which lead to transient complexes where the β-strands of different molecules are antiparallel and in a linear arrangement.

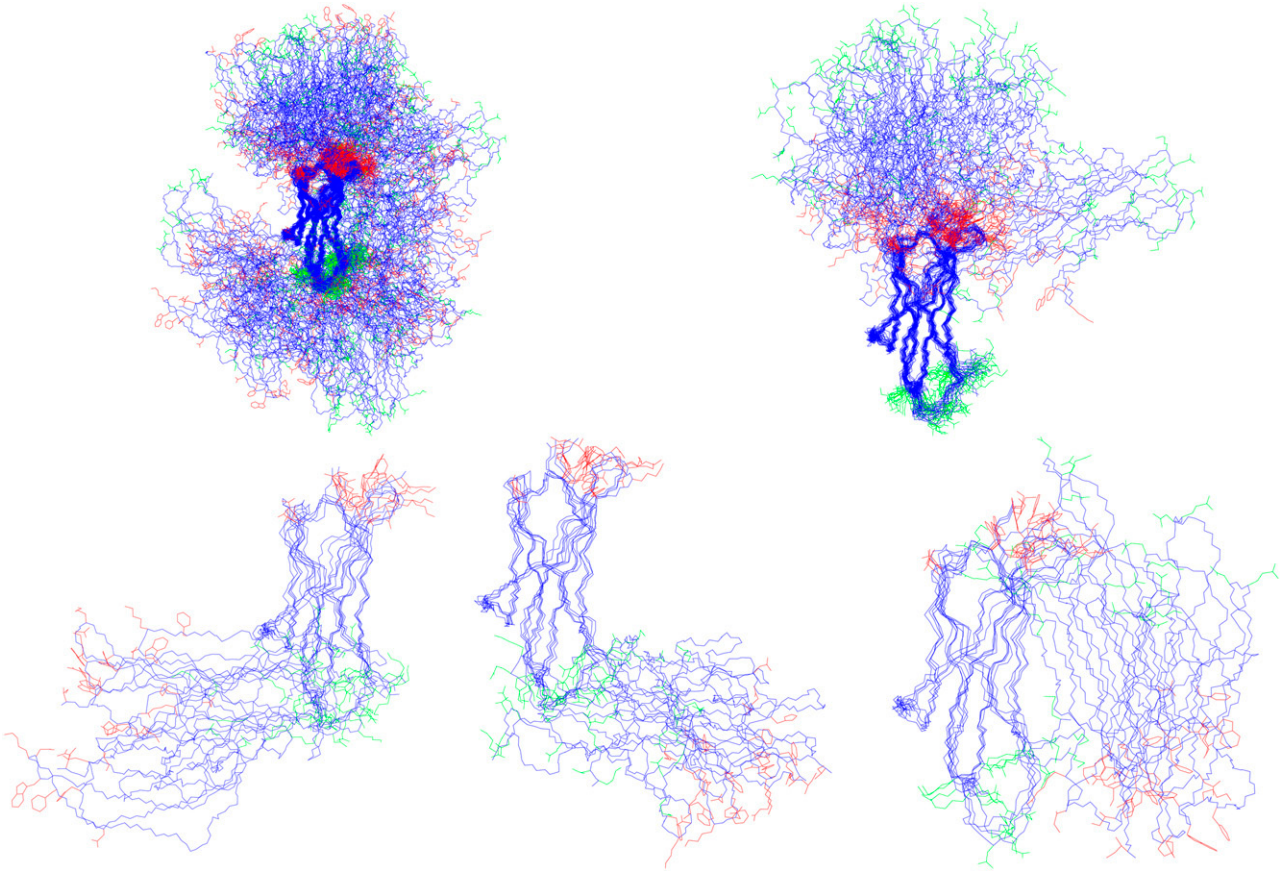

FIGURE 5 All the selected pairs of molecules are displayed in such a way that one of the two molecules is superimposed on the starting structure. The head residues are colored in red, the tail residues are colored in green (*upper left*). The most populated cluster displays a head-to-head arrangement (*upper right*). The second- and third- most populated clusters display a tail-to-tail arrangement (*lower left*). The fourth-most populated cluster displays an antiparallel arrangement (*lower right*).

At first glance, this is at variance with what could be expected based on the accepted cross-β structure of amyloid. However, formation of intramolecular β-strand pairing should take place at a second stage on a much longer timescale, as observed in the laboratory. The sample in the present simulation should concern only the early stages of aggregation, which may be distinct from fibrillogenesis.

Among the pairs of contacting molecules considered previously, there are two long-lived pairs, i.e., two pairs that form during the first nanosecond of simulation and survive throughout the simulation. These two pairs together form a cluster of four molecules. The arrangement of this cluster in the second half of the simulation is peculiar. Two molecules seem to compete for the same end of a third molecule. This leads to an arrangement of the two molecules parallel to each other (Fig. 7). Any attempt to explain this arrangement in terms of molecular dipoles fails due to shortcoming of the same concept of dipole for a macromolecular system. Indeed, the dipole-dipole energy shows no trend at all. However, electrostatic effects are expected to be important. Using the Poisson-Boltzmann equation, we could compute the electrostatic energy for assembling the cluster of four molecules from infinitely distant monomers. Within the accuracy limits of the approach itself, the electrostatic free energy starts from a small negative value (-0.2 kcal/mol) and reaches ~ -5.0 kcal/mol at the end of the simulation.

Although the latter value slightly depends on the choice of inner dielectric constant and calculation parameters, the value is negative and in the range of 5 kcal/mol. This result shows that electrostatic energy is decreased by the molecular arrangement.

Our results suggest that despite the common structure and properties shared by amyloid fibers formed by different peptides and proteins, the earliest stages of β_2-microglobulin aggregation seem to be rather specific to the molecular properties of β_2-microglobulin and the mechanism is not easily generalizable to other proteins.

The stable linear arrangement observed here for three molecules could, to some extent, drive the cluster toward a regularly oriented configuration, but if this happens, it takes a much longer time than that used for our simulation.

Comparison with experimental data

The results of the present simulation can be compared with the previous experimental data in several aspects. First, all solution studies (46–48) have shown by indirect evidence that

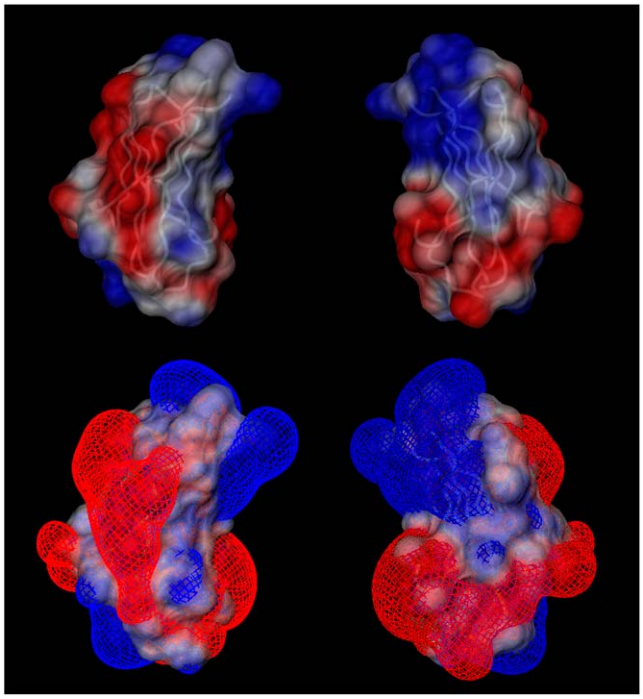

FIGURE 6 The electrostatic potential at the surface of β_2-microglobulin (*upper*) and isopotential surfaces (*lower*). Blue is saturated at 1.0 kcal mol^{-1} q^{-1} and red at -1.0 kcal mol^{-1} q^{-1}. The blue isopotential surface refers to 0.5 kcal mol^{-1} q^{-1} and red to -0.5 kcal mol^{-1} q^{-1}. The pictures at right refer to the molecule rotated by 180° about the vertical axis.

the N-terminal and C-terminal regions of β_2-microglobulin are not involved in fibril formation. The same studies provide evidence of possible involvement in intermolecular pairing of strand D, loop DE, and possibly strands F and C. Thus the flexibility of both N- and C-terminal regions that emerges from our simulations is consistent with experimental evidence, because flexible regions are unlikely to establish the intermolecular contacts that will prove determinant for fibrillogenesis. By a symmetric argument it could be argued that simulations successfully identify residues in the DE loop with some propensity for intermolecular contacts, according to experiments. It has already been pointed out that molecular dynamics simulations are able to sample only the early events that occur within the considered ensemble of molecules, whereas fibrillogenesis is a much slower process. For amyloidogenic proteins, however, several experimental results have led to the conclusion that fibrillogenesis takes place by a nucleation-dependent mechanism, with the nucleation step being the rate-determining step (49–53) A similar mechanism has been proposed also for β_2-microglobulin (5), where nucleation should be preceded by a lag phase during which dynamic oligomeric species are formed. Hence, the molecular dynamics simulations described here should reproduce those early assemblies, rather than the formation of nuclei or protofibrils. Conceivably, some features of the early assemblies may be similar or even conserved by the subsequent nucleated adducts. This could be the case with the contacts observed in the simulation for residues of the DE loop, which also appear to be involved in intermolecular contacts in experiments addressing certain later stages of the aggregative process (2,54).

Besides the information on the location of the intermolecular contacts, the simulations provide also some hints concerning the relative orientation of the interacting β_2-microglobulin molecules. These hints may convey again valuable inferences for fibril geometry, though no conclusion can be validated due to the lack of specific experimental evidence. Nevertheless, it is worth noting that the predominance of head-to-head arrangements within the clusterized pairs of interacting β_2-microglobulin molecules recalls the related findings that were recently reported for the NM domain of Sup35 and a peptide thereof (55,56). In particular, it was proposed that the NM domain of sup35 paired into amyloid structures through head-to-head and tail-to-tail interactions, the former mode occurring first to nucleate the whole assembly (55). The crystal structure of a small fragment of the sup35 NM domain (containing six to seven residues), on the other hand, exhibits all the features of the cross-β pairing, which alternate with complementary surface interactions between adjacent sheets (56).

CONCLUSIONS

Early events in aggregation of proteins proved elusive to experimental investigations. The dynamic formation of

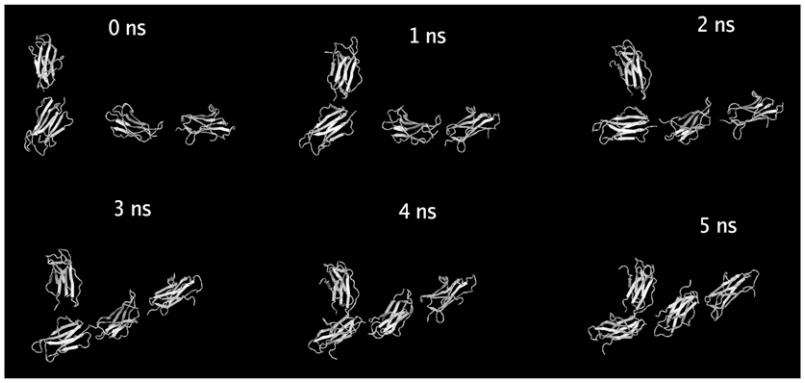

FIGURE 7 Snapshots taken at 1-ns intervals of four selected β_2-microglobulin molecules.

short-lived oligomers of β_2-microglobulin has been proved experimentally (5), but no structural characterization of these species has been possible yet. The currently available computers and programs allow the simulation of nanosecond molecular dynamics of large systems, including up to one million atoms. In this work we simulated 5-ns molecular dynamics of an ensemble of 27 β_2-microglobulin molecules densely, but not unrealistically, arranged in a cube.

During the simulation, the formation of intermolecular contacts is observed. The simulation highlights the importance of apical residues and, in particular, of those at the end containing the N-terminus of the molecule. Hydrophobic contacts appear to be important for the establishment of long-lived (on the simulation timescale) contacts. It should be obvious that these early events on the pathway to aggregation and fibril formation are not directly related to the end-state of the process, which is reached on a much longer timescale.

However, the simulation findings are in line with experimental data. In particular, loop DE and the adjacent β-strands have been implicated in intermolecular contacts by different studies (2,3,46). Here, the residue involved in the largest number of contacts is Trp-60, a residue in the DE loop, which is in close association with the α subunit in class I MHC. Curiously, the conformation at residues Asp-59 and Trp-60 is close to the repeating unit of the α-pleated sheet proposed by Daggett and co-workers to promote fibril formation (7).

How the long-lived (on the simulation timescale) dimers and oligomers involving Trp-60 eventually aggregate and form amyloid fibrils cannot be assessed by a 5-ns molecular dynamics simulation. One of the consequences of a long-lived head-to-head contact of the molecules would be the possibility of a hinge motion about the contacting residues that would facilitate parallel intermolecular β-sheet pairing. Interestingly, this would be consistent with the model proposed by Nussinov and co-workers (11) for assembly of the fibril following the detachment of strands A and G from the structure, a condition consistent in turn with the experimental observations by our group (3). Mutants are being prepared for testing and further detailing this picture.

It is difficult to state how generalizable to other proteins is the mechanism proposed here for early molecular encounters. It is interesting that for most amyloidogenic proteins partial unfolding (implying exposure of buried hydrophobic surfaces) has been hypothesized as a prerequisite for aggregation. For the system studied here, detachment from the MHC complex is sufficient for hydrophobic surface area exposure. The requirement for electrostatic complementarity would modulate this general mechanism.

We note that the specificity of the patterns of interaction found here, and which could foreseeably be found for other proteins, is in sharp contrast with the overall similar structure of fibrils formed by very different proteins. Here, only the early steps in aggregation are addressed by simulation. The timescale of these steps is in the range of nanoseconds, orders of magnitude faster than that of fibril formation. For the latter process, the details of the mechanism at the molecular level appear still elusive.

Simulations were performed using BEN, the Teraflop cluster facility at ECT* Supercomputing lab as part of the multidisciplinary activities of the Centre. The Supercomputing Lab @ECT* is a joint initiative of ECT*, the Istituto Trentino di Cultura, Exadron (the Eurotech Group), the Istituto Nazionale di Fisica Nucleare, and the Provincia Autonoma of Trento to test and develop innovative technologies in high-performance computing.

This research was funded by the Investment Fund for Basic Research (FIRB) grants RBNE03B8KK and RBNE03PX83, and Projects of National Interest grant PRIN2005 from the Italian Ministry for Education, University and Research.

REFERENCES

1. Yamamoto, S., and F. Gejyo. 2005. Historical background and clinical treatment of dialysis-related amyloidosis. *Biochim. Biophys. Acta.* 1753:4–10.
2. Radford, S. E., W. S. Gosal, and G. W. Platt. 2005. Towards an understanding of the structural molecular mechanism of β_2-microglobulin amyloid formation in vitro. *Biochim. Biophys. Acta.* 1753:51–63.
3. Esposito, G., A. Corazza, P. Viglino, G. Verdone, F. Pettirossi, F. Fogolari, A. Makek, S. Giorgetti, P. Mangione, M. Stoppini, and V. Bellotti. 2005. Solution structure of β_2-microglobulin and insights into fibrillogenesis. *Biochim. Biophys. Acta.* 1753:76–84.
4. Kardos, J., D. Okuno, T. Kawai, Y. Hagihara, N. Yumoto, T. Kitagawa, P. Zavodszky, H. Naiki, and Y. Goto. 2005. Structural studies reveal that the diverse morphology of β_2-microglobulin aggregates is a reflection of different molecular architectures. *Biochim. Biophys. Acta.* 1753:108–120.
5. Corazza, A., F. Pettirossi, P. Viglino, G. Verdone, J. Garcia, P. Dumy, S. Giorgetti, P. Mangione, S. Raimondi, M. Stoppini, V. Bellotti, and G. Esposito. 2004. Properties of some variants of human β_2-microglobulin and amyloidogenesis. *J. Biol. Chem.* 279:9176–9189.
6. Ma, B., and R. Nussinov. 2003. Molecular dynamics simulation of the unfolding of β_2-microglobulin and its variants. *Prot. Eng.* 16:2310–2321.
7. Armen, R. S., M. L. deMarco, D. O. V. Alonso, and V. Daggett. 2004. Pauling and Corey's α-pleated sheet structure may define the prefibrillar amyloidogenic intermediate in amyloid disease. *Proc. Natl. Acad. Sci. USA.* 101:11622–11627.
8. Nishino, M., Y. Sugita, T. Yoda, and Y. Okamoto. 2005. Structures of a peptide fragment of β_2-microglobulin studied by replica-exchange molecular dynamics simulations: towards the understanding of the mechanism of amyloid formation. *FEBS Lett.* 579:5425–5429.
9. Armen, R. S., and V. Daggett. 2005. Characterization of two distinct β_2-microglobulin unfolding intermediates that may lead to amyloid fibrils of different morphology. *Biochemistry.* 44:16098–16107.
10. Benyamini, H., K. Gunasekaran, H. Wolfson, and R. Nussinov. 2003. β_2-microglobulin amyloidosis: insights from conservation analysis and fibril modelling by protein docking techniques. *J. Mol. Biol.* 330:159–174.
11. Benyamini, H., K. Gunasekaran, H. Wolfson, and R. Nussinov. 2005. Fibril modelling by sequence and structure conservation analysis combined with protein docking techniques: β_2-microglobulin amyloidosis. *Biochim. Biophys. Acta.* 330:121–130.
12. Ma, B., and R. Nussinov. 2002. Stabilities and conformations of Alzheimer's β-amyloid peptide oligomers ($A\beta_{16-22}$, $A\beta_{16-35}$, and $A\beta_{10-35}$): sequence effects. *Proc. Natl. Acad. Sci. USA.* 99:14126–14131.
13. Tiana, G., F. Simona, R. A. Broglia, and G. Colombo. 2005. Thermodynamics of β-amyloid fibril formation. *J. Chem. Phys.* 120:8307–8317.
14. Lopez de la Paz, M., G. M. S. de Mori, L. Serrano, and G. Colombo. 2005. Sequence dependence of amyloid fibril formation: insights from molecular dynamics simulations. *J. Mol. Biol.* 349:583–596.

15. Zanuy, D., B. Ma, and R. Nussinov. 2003. Short peptide amyloid organization: stabilities and conformations of the islet amyloid peptide NFGAIL. *Biophys. J.* 84:1884–1894.

16. Zanuy, D., and R. Nussinov. 2003. The sequence dependence of fiber organization. A comparative molecular dynamics study of the islet amyloid polypeptide segments 22–27 and 22–29. *J. Mol. Biol.* 329: 565–584.

17. Wei, G., N. Mousseau, and P. Derreumaux. 2004. Exploring the early steps of aggregation of amyloid-forming peptide KFFE. *J. Phys. Condens. Matter.* 16:S5047–S5054.

18. Simona, F., G. Tiana, R. A. Broglia, and G. Colombo. 2004. Modeling the α-helix to β-hairpin transition mechanism and the formation of oligomeric aggregates of the fibrillogenic peptide Aβ(12–28): insights from all-atom molecular dynamics simulations. *J. Mol. Graph. Mod.* 23:263–273.

19. Lipfert, E., J. Franklin, F. Wu, and S. Doniach. 2005. Protein misfolding and amyloid formation for the peptide GNNQQNY from yeast prion protein sup35: simulation by reaction path annealing. *J. Mol. Biol.* 349:648–658.

20. Wu, C., L. Hongxing, and Y. Duan. 2005. Elongation of ordered peptide aggregate of an amyloidogenic hexapeptide NFGAIL observed in molecular dynamics simulations with explicit solvent. *J. Am. Chem. Soc.* 127:13530–13537.

21. Kolinski, A., and J. Skolnick. 2004. Reduced models of proteins and their applications. *Polym.* 45:511–524.

22. Malolepsza, E., M. Boniecki, A. Kolinski, and L. Piela. 2005. Theoretical model of prion propagation: a misfolded protein induces misfolding. *Proc. Natl. Acad. Sci. USA.* 102:7835–7840.

23. Urbanc, B., F. Ding, D. Sammond, S. Khare, S. V. Buldyrev, H. E. Stanley, and N. V. Dokholyan. 2004. Molecular dynamics simulation of amyloid β dimer formation. *Biophys. J.* 87:2310–2321.

24. Ding, F., J. J. LaRocque, and N. V. Dokholyan. 2005. Direct observation of protein folding, aggregation, and a prion-like conformational conversion. *J. Biol. Chem.* 280:40235–40240.

25. Nguyen, H. D., and C. K. Hall. 2004. Molecular dynamics simulations of spontaneous fibril formation by random-coil peptides. *Proc. Natl. Acad. Sci. USA.* 101:16180–16185.

26. Favrin, G., A. Irback, and S. Mohanty. 2004. Oligomerization of amyloid Aβ_{16-22} peptides using hydrogen bonds and hydrophobicity forces. *Biophys. J.* 87:3657–3664.

27. Gsponer, J., U. Habertur, and A. Caflisch. 2003. The role of side-chain interactions in the early steps of aggregation: molecular dynamics simulations of an amyloid-forming peptide from the yeast prion sup35. *Proc. Natl. Acad. Sci. USA.* 100:5154–5159.

28. Cecchini, M., F. Rao, M. Seeber, and A. Caflisch. 2004. Replica exchange molecular dynamics simulations of amyloid peptide aggregation. *J. Chem. Phys.* 121:10748–10756.

29. Paci, E., J. Gsponer, X. Salvatella, and M. Vendruscolo. 2005. Molecular dynamics studies of the process of amyloid aggregation of peptide fragments of transthyretin. *J. Mol. Biol.* 340:555–569.

30. Klimov, D. K., and D. Thirumalai. 2003. Dissecting the assembly of Aβ_{16-22} amyloid peptides into antiparallel β-sheets. *Structure.* 11:295–307.

31. Peng, Y., and U. H. E. Hansmann. 2003. Helix versus sheet formation in a small peptide. *Phys. Rev. E.* 68:041911.

32. Jahn, T. R., M. J. Parker, S. W. Homans, and S. E. Radford. 2006. Amyloid formation under physiological conditions proceeds via a native-like folding intermediate. *Nat. Struct. Biol.* 13:195–201.

33. Eakin, C. M., A. J. Berman, and A. D. Miranker. 2006. A native to amyloidogenic transition regulated by a backbone trigger. *Nat. Struct. Biol.* 13:202–208.

34. Saper, M. A., P. J. Bjorkman, and D. C. Wiley. 1991. Refined structure of the human histocompatibility antigen HLA-A2 at 2.6 Å resolution. *J. Mol. Biol.* 219:277–319.

35. Berendsen, H. J. C., D. van der Spoel, and R. van Drunen. 1995. GROMACS: A message-passing parallel molecular dynamics implementation. *Comput. Phys. Commun.* 91:43–56.

36. Giorgetti, S., A. Rossi, P. Mangione, S. Raimondi, S. Marini, M. Stoppini, A. Corazza, P. Viglino, G. Esposito, G. Cetta, G. Merlini, and V. Bellotti. 2005. β_2-microglobulin isoforms display an heterogeneous affinity for type I collagen. *Protein Sci.* 14:696–702.

37. Kale, L., R. Skeel, M. Bhandarkar, R. Brunner, A. Gursoy, N. Krawetz, J. Phillips, A. Shinozaki, K. Varadarajan, and K. Schulten. 1999. NAMD2: greater scalability for parallel molecular dynamics. *J. Comput. Phys.* 151:283–312.

38. Madura, J. D., J. M. Briggs, R. Wade, M. E. Davis, B. A. Luty, A. Ilin, J. A. Antosiewicz, M. K. Gilson, B. Bagheri, L. Ridgway Scott, and J. A. McCammon. 1995. Electrostatics and diffusion of molecules in solution: simulations with the University of Houston Brownian Dynamics program. *Comput. Phys. Commun.* 91:57–95.

39. Humphrey, W., A. Dalke, and K. Schulten. 1996. VMD: visual molecular dynamics. *J. Mol. Graph.* 14:33–38.

40. Darden, T., D. York, and L. Pedersen. 1993. Particle mesh Ewald. An N.log(N) method for Ewald sums in large systems. *J. Chem. Phys.* 98:10089–10092.

41. Keutsch, F. N., and R. J. Saykally. 2001. Water clusters: untangling the mysteries of the liquid, one molecule at a time. *Proc. Natl. Acad. Sci. USA.* 98:10533–10540.

42. Berendsen, H. J. C., J. P. M. Postma, W. F. van Gunsteren, A. Di Nola, and J. R. Haak. 1984. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* 81:3684–3690.

43. Relini, A., C. Canale, S. D. Stefano, R. Rolandi, S. Giorgetti, M. Stoppini, A. Rossi, F. Fogolari, A. Corazza, G. Esposito, A. Gliozzi, and V. Bellotti. 2006. Collagen plays an active role in the aggregation of β2-microglobulin under physiopathological conditions of dialysis-related amyloidosis. *J. Biol. Chem.* 281:16521–16529.

44. Hunter, R. J. 1987. Foundations of Colloid Science. Oxford University Press, Oxford, UK.

45. Berrera, M., H. Molinari, and F. Fogolari. 2003. Amino acid empirical contact energy definition for fold recognition in the space of contact maps. *BMC Bioinformatics.* 4:8.

46. Esposito, G., R. Michelutti, G. Verdone, P. Viglino, H. Hernandez, C. Robinson, A. Amoresano, F. Dal Piaz, M. Monti, P. Pucci, P. Mangione, M. Stoppini, G. Merlini, G. Ferri, and V. Bellotti. 2000. Removal of the N-terminal hexapeptide from human β_2-microglobulin facilitates protein aggregation and fibril formation. *Protein Sci.* 9:831–845.

47. Hoshino, M., H. Katou, K. Hagiwara, H. Naiki, and Y. Goto. 2002. Mapping the core of the β_2-microglobulin amyloid fibril by H/D exchange. *Nat. Struct. Biol.* 9:332–336.

48. McParland, V. J., A. P. Kalverda, S. W. Homans, and S. E. Radford. 2002. Structural properties of an amyloid precursor of β_2-microglobulin. *Nat. Struct. Biol.* 9:326–331.

49. Naiki, H., K. Higuchi, K. Nakakuki, and T. Takeda. 1991. Kinetic analysis of amyloid fibril polymerization in vitro. *Lab. Invest.* 65:104–110.

50. Naiki, H., N. Hashimoto, S. Suzuki, H. Kimura, K. Nakakuki, and F. Gejyo. 1997. Establishment of a kinetic model of dialysis-related amyloid fibril extension in vitro. *Amyloid Int. J. of Exp. Clin. Invest.* 4:223–232.

51. Jarrett, J., and P. J. Lansbury. 1992. Amyloid fibril formation requires a chemically discriminating nucleation event: studies of an amyloidogenic sequence from the bacterial protein OsmB. *Biochemistry.* 31:12345–12352.

52. Jarrett, J., and P. J. Lansbury. 1993. Seeding one-dimensional crystallization of amyloid: a pathogenic mechanism in Alzheimer's disease and scrapie? *Cell.* 73:1055–1058.

53. Lomakin, A., D. Teplow, D. Kirschner, and G. Benedek. 1997. Kinetic theory of fibrillogenesis of amyloid β-protein. *Proc. Natl. Acad. Sci. USA.* 94:7942–7947.

54. Chatani, E., and Y. Goto. 2005. Structural stability of amyloid fibrils of β_2-microglobulin in comparison with its native fold. *Biochim. Biophys. Acta.* 1753:51–63.

55. Krishnan, R., and S. L. Lindquist. 2005. Structural insights into a yeast prion illuminate nucleation and strain diversity. *Nature.* 435:765–772.

56. Nelson, R., M. R. Sawaya, M. Balbirnie, A. O. Madsen, C. Riekel, R. Grothe, and D. Eisenberg. 2005. Structure of the cross-β spine of amyloid-like fibrils. *Nature.* 435:773–778.