doi:10.1016/j.jmb.2012.02.004 J. Mol. Biol. (2012) **xx**, xxx–xxx

Contents lists available at www.sciencedirect.com

Journal of Molecular Biology

journal homepage: http://ees.elsevier.com.jmb

Driving Forces and Structural Determinants of Steric Zipper Peptide Oligomer Formation Elucidated by Atomistic Simulations

Dirk Matthes†, Vytautas Gapsys† and Bert L. de Groot*

Computational Biomolecular Dynamics Group, Department of Theoretical and Computational Biophysics, Max-Planck-Institute for Biophysical Chemistry Am Fassberg 11, 37077 Göttingen, Germany

Received 17 October 2011; received in revised form 27 January 2012; accepted 1 February 2012

Edited by D. Case

Keywords:
peptide aggregation;
amyloid;
nucleated conformational conversion;
molecular dynamics;
collective coordinates

Understanding the structural and energetic requirements of non-fibrillar oligomer formation harbors the potential to decipher an important yet still elusive part of amyloidogenic peptide and protein aggregation. Low-molecular-weight oligomers are described to be transient and polymorphic intermediates in the nucleated self-assembly process to highly ordered amyloid fibers and were additionally found to exhibit a profound cytotoxicity. However, detailed structural information on the oligomeric species involved in the nucleation cannot be readily inferred from experiments.

Here, we study the spontaneous assembly of steric zipper peptides from the tau protein, insulin and α-synuclein with atomistic molecular dynamics simulations on the microsecond timescale. Detailed analysis of the forces driving the oligomerization reveals a common two-step process akin to a general condensation-ordering mechanism and thus provides a rational understanding of the molecular basis of peptide self-assembly. Our results suggest that the initial formation of partially ordered peptide oligomers is governed by the solvation free energy, whereas the dynamical ordering and emergence of β-sheets are mainly driven by optimized inter-peptide interactions in the collapsed state.

A novel mapping technique based on collective coordinates is employed to highlight similarities and differences in the conformational ensemble of small oligomer structures. Elucidating the dynamical and polymorphic β-sheet oligomer conformations at atomistic detail furthermore suggests complementary sheet packing characteristics similar to steric zipper structures, but with a larger heterogeneity in the strand alignment pattern and sheet-to-sheet arrangements compared to the cross-β motif found in the fibrillar or crystalline states.

Introduction

The deposition and accumulation of stable, filamentous aggregates of a specific protein or peptide in a variety of tissues are known to be associated with a number of human pathologies.[1] These aggregates are generally termed amyloid fibrils[2] and formed by many natural polypeptides.[3,4] In addition, truncated parts of such protein sequences[5,6] and de novo designed peptides[7,8] were

*Corresponding author. E-mail address: bgroot@gwdg.de.

† D.M. and V.G. contributed equally to this work.

Abbreviations used: MD, molecular dynamics; PCA, principal component analysis; k-NN, k-nearest neighbor; FMA, functional mode analysis.

shown to assemble into amyloid-like fibrils *in vitro*. Thus, it has been proposed that the intermolecular interactions of the invariant polypeptide backbone result in the amyloid structure as an alternative generic conformational state with the kinetics of the process being dependent on the side chains and external factors, such as pH, temperature or ionic strength.[4,8–12] Regardless of the sequence or native fold, the commonly formed amyloid fibrils are defined as self-assembled, elongated and unbranched (fibrillar) polypeptide aggregates with a cross-β conformation.[13] The cross-β architecture, as revealed by X-ray fiber diffraction, is described as stacked β-strands that run perpendicular to the fiber axis with extensive hydrogen bonding along the length of the fiber.[14] Moreover, a growing number of crystal structures of short model peptide sequences revealed a common steric zipper motif.[6,15–18] The atomic structures of the crystalline conformers show pairs of elongated β-sheets with parallel or antiparallel strand alignment. The opposing sheets are interdigitated such that a highly complementary packing of the side chains is achieved, yielding a tight and dry interface. Despite their fundamental similarity, the structures vary in their basic steric zipper motif, a feature that rationalizes the observed polymorphism of the self-propagating amyloid structures on a molecular basis.[16] It has been shown that crystalline and fibrillar amyloid polymorphs share structural characteristics such as the cross-β diffraction pattern and therefore offer a plausible paradigm for the general spine organization of amyloid fibrils.[4,15,17,18] Nevertheless, the degree of order in the crystal structure may not fully represent the one in the fibrillar form, as indicated by solid-state NMR experiments on various crystals and fibrils.[17,18]

Exposure of hydrophobic epitopes and regions of unstructured polypeptide backbone, such as found in partially folded or misfolded states, is among the accepted causes of amyloidogenic aggregation,[10,19,20] which is the irreversible formation of the β-sheet-rich amyloid structures.[4,8,20] Consensus aggregation-prone sequence patterns of amyloidogenic proteins have been identified, and they demonstrate that protein unfolding is necessary but not sufficient to promote aggregation.[11,21,22] In fact, experimental evidence is accumulating that short amyloidogenic signatures in natural protein sequences can facilitate self-assembly.[22–24]

The multi-staged aggregation process is canonically described as the conversion of isolated peptide monomers in solution to soluble oligomeric assemblies and the final, fibrillar aggregates via a nucleated growth process.[25,26]

It is likely that even more intermediate states have to be considered and that the conformational transitions between all of them are associated with different barrier heights.[27–29] In order to obtain and understand the full picture, detailed knowledge of the molecular structures of the involved species is indispensable.

Despite the considerable progress in characterizing the fibrillar end-states, it is still difficult to gain the biochemical and precise structural information for the oligomeric species in experiments. Heterogeneous oligomeric aggregates of different sizes are usually observed during the incubation of amyloidogenic peptide solutions[30] and discussed as either on-[31,32] or off-pathway[33,34] intermediates to the fibrils. Pre-fibrillar and fibrillar oligomers as well as annular protofibrils have been described among a variety of morphologies.[33,35–37] However, thorough investigations are hindered by either the transient or polymorphic and non-crystalline behavior of the oligomers.[28,35,38] Critical observations regarding the structural properties of oligomeric aggregates have been derived recently from experiments on short amyloidogenic peptides.[7,39–42]

Although some key aspects are not entirely understood, several studies report on the general consent that: (a) Given their qualitatively different morphologies from the characteristic appearance in transmission electron and atomic force microscopic images, oligomeric precursor states and amyloid fibrils are surprisingly similar in molecular conformation and supramolecular structure.[33,43,44] (b) Specific binding to antibodies alludes to common structural features shared by oligomers from different amyloidogenic proteins.[37,45] (c) The oligomeric aggregates assume β-sheet-rich conformations.[29,46] (d) Soluble amyloid intermediates are established as the primary pathogenic agents in several types of neurological amyloid diseases.[20,33,35–37,47]

In addition, computational studies have facilitated the current understanding of molecular determinants and events in the early stages of amyloidogenic peptide aggregation. Molecular dynamics (MD) simulations appear to be particularly suited to probe the formation of oligomeric species in atomistic detail, unraveling the transition pathways on timescales not amenable to experiments. The properties of small multimeric aggregates (dimers to decamers) of various amyloidogenic peptide sequences have been studied by atomistic simulations and described as partially ordered, nematic structures, which are subject to rapid fluctuations and large conformational rearrangements.[41,42,48–56] The obtained oligomer ensembles are described as distinct from the monomeric form[48,52,57] due to the conformational changes associated with emerging β-sheet structure.[49,56] These structural transitions were found to be accompanied by a loss of intra-peptide interactions and conformational entropy.[51,52] The desolvation of nonpolar surface and formation of inter-peptide backbone hydrogen bonds was found concomitantly.[52,54,58,59] Dynamical reorganization via sliding, reptation or flipping of individual strands as opposed to repeated dissociation and annealing has been observed in simulations[48,50,54,60] and has found experimental validation.[61] Furthermore, the initial stages of assembly are reported to be likely under kinetic control, and a multiplicity of association and interconversion pathways gives rise to polymorphic aggregate structures.[50,59,62,63] In the context of the clearly very complex underlying free-energy landscape, the prominent and crucial role of water in the aggregation process has been highlighted.[42,57–59,64,65]

Here, we perform unbiased, atomistic simulations of steric zipper peptide oligomerization in explicit solvent as model systems for amyloidogenic aggregation. The studied peptides are short segments identified from the fibril-forming proteins tau (^{306}VQIVYK^{311}, referred to as PHF6),[15,24] insulin (^{12}VEALYL^{17}, referred to as IB12)[5,15] and α-synuclein (^{51}GVATVA^{56}, referred to as AS51).[15] They have been found to be essential in fibril formation of the full-length proteins and adopt β-strand conformations in the fibrillar aggregate structures, respectively.[3,15,24,66,67] Moreover, the hexa-peptides PHF6, IB12 and AS51 were shown to spontaneously polymerize in solution, yielding amyloid-like fibrils, microcrystals or both.[15] Fibrillization assays of these peptides monitored with, for example, ThT fluorescence measurements, show a common lag time and imply a typical nucleation-dependent growth process.[15,66,68,69]

The aim of this work is to facilitate the understanding of the early events in spontaneous peptide oligomerization on a molecular level. To directly assess the multifaceted conformational ensemble of oligomers and the sampled structural transitions, we presented a novel mapping technique. Key structural elements of the spontaneously assembled oligomeric states are compared with the known X-ray crystallography structures of various amyloidogenic peptides, thereby contributing to the structural and dynamical characterization of low-molecular-weight peptide oligomers. Furthermore, addressed are the following questions: What drives the initially monomeric peptides to form oligomeric assemblies? What determines their subsequent structural ordering and stabilizes the formed aggregates? The comparison of aggregation pathways and structures of different peptide sequences and concentrations allows for insight into possible common mechanistic steps in amyloidogenic peptide aggregation.

Results

Spontaneous aggregation of steric zipper peptides results in β-sheet-rich decamers

Multiple unbiased MD simulations for the PHF6, IB12 and AS51 peptide systems were initiated from randomized and fully dispersed monomeric conformations (M10, see Table 3). Visual inspection of the PHF6, IB12 and AS51 simulation trajectories suggested common characteristics in the aggregation process. Regardless of sequence, a successive association of all 10 peptide chains as well as a gradual increase in β-sheet structure was found. While the onset of β-sheet formation was fast, usually within 10 ns, fluctuations in the content of β-sheet conformations of the decameric peptide aggregates were observed on the microsecond timescale (Fig. 1). In contrast to the PHF6 and AS51 simulations, a near-monotonical increase of β-sheet conformation was found for all IB12 trajectories (Fig. 1b). The most abundant IB12 peptide aggregate conformers were found to have a β-sheet content of around 65%, as seen from a combined histogram of all simulations (Fig. 1d). The overall fraction of such β-sheet-rich IB12 structures was more than four times higher compared to PHF6 and three times as high as for the AS51 aggregates (they sampled mostly a content of 40–50%). The set of independent PHF6 simulations showed large variations with respect to the level of β-sheet content (Fig. 1a). Notably, a significant proportion of PHF6 peptides were found with a low or without any β-sheet content in several simulations. Multiple AS51 simulations displayed a substantial, repeated loss and recovery of β-sheet structure, resulting in a high standard deviation for the β-sheet content (Fig. 1c).

The PHF6, IB12 and AS51 peptide chains associated in general to fairly ordered structures, as the increase in β-sheet structure with simulation time indicates. Nevertheless, as shown below, the 10 chains populated a multitude of different aggregate configuration types and heterogeneous topologies throughout the simulations, respectively.

Mapping peptide aggregates onto collective coordinates yields direct insight into common association pathways and diversity of structures

It is not straightforward to characterize the process of peptide oligomerization comprehensively using just one observable (e.g., the β-sheet content). In fact, multiple metrics are necessary to discriminate the aggregate conformation ensembles in a meaningful way or to examine specific structural properties. Although every chosen observable might provide its own information content, it is often not possible to compare them simultaneously. Yet another complication is the need to define or select the most suitable observable in the first place.

Here, we present a collective coordinate approach to describe the oligomeric structures and their transitions sampled in the different aggregation simulations (see Fig. 2 and Methods for a detailed description). In order to probe amyloidogenic β-aggregation adequately, we chose 25 observables as

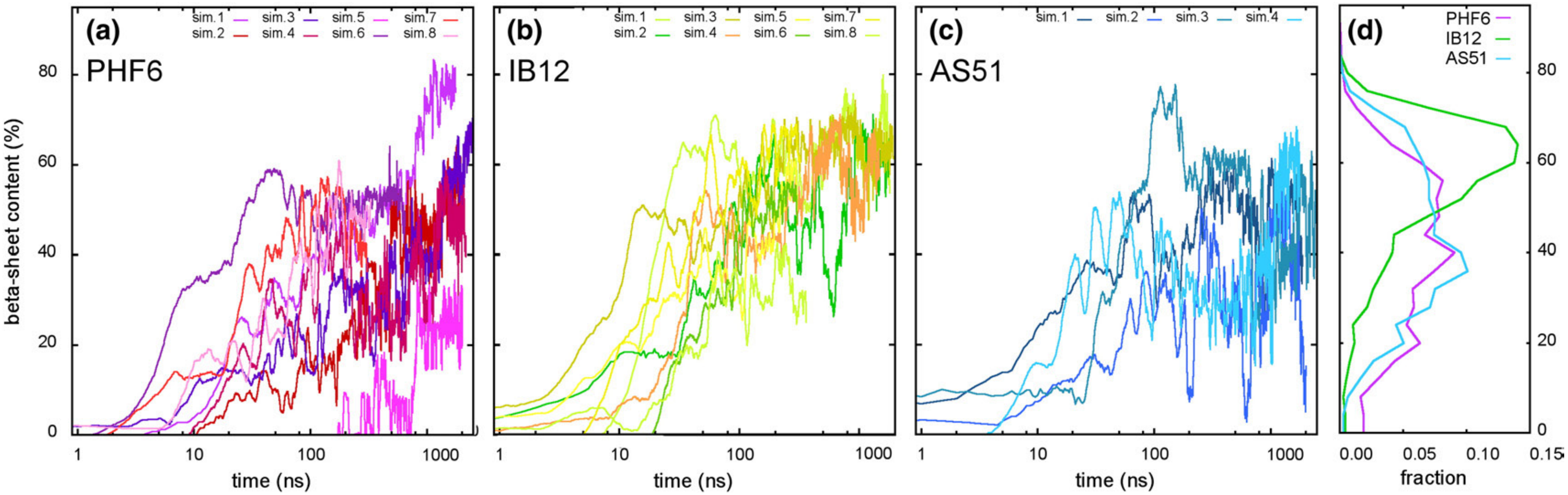

Fig. 1. Time dependence and probability of β-sheet structure content. The change in β-sheet conformation with simulation time is shown on a log scale for the (a) PHF6, (b) IB12 and (c) AS51 trajectories. In (d), the normalized abundance of a certain β-sheet content is given as histogram. The colors match the shades of the individual time traces for PHF6 (purple), IB12 (green) and AS51 (blue).

topological, structural and energetic descriptors of the sampled configurations. Through the use of this set of measures instead of the Cartesian coordinates, it is possible to apply a dimensionality reduction step using principal component analysis (PCA). The first three eigenvectors of the covariance matrix constructed from the observable data then represent a newly identified basis for the subsequent analysis. Thereafter, all the configurations from the simulations were projected onto these collective coordinates to obtain a low-dimensional representation of the sampled phase space and hence to identify underlying collective trends.

The projections of the combined simulation data for the PHF6, IB12 and AS51 peptide systems are shown in Fig. 3, where each sphere represents a simulation configuration snapshot. The mapping procedure discriminates structures with different features, therefore allowing the direct assessment of the multimeric aggregates found along the aggregation pathway. From each projection map, the common structural and energetic properties can be

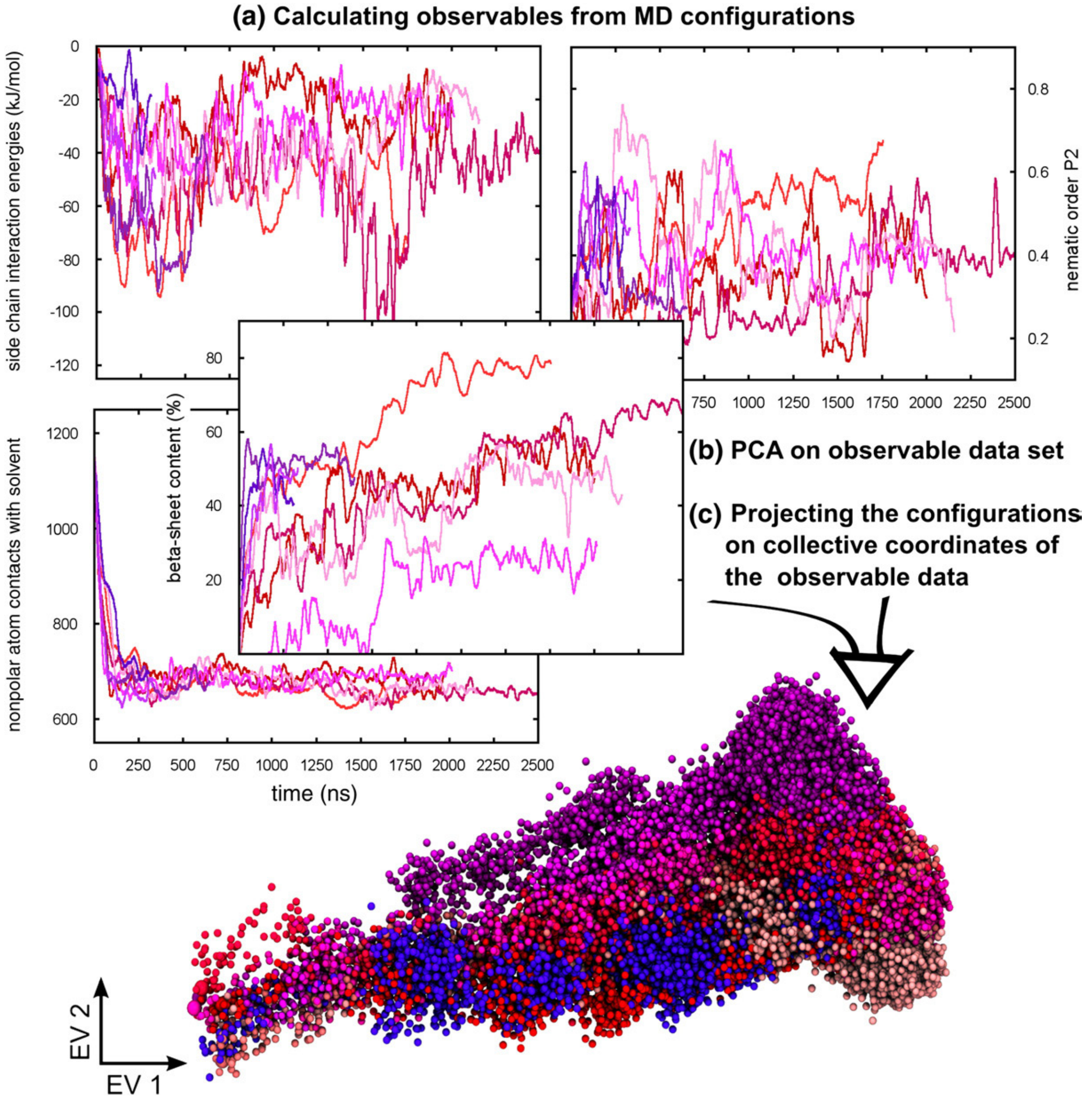

Fig. 2. Schematic representation of the collective coordinate mapping procedure. First, from multiple independent simulations, a set of observables (topological, structural and energetic descriptors) is calculated. Subsequently, a matrix is built, where each row represents the different observable values for a given MD conformation. A PCA on this multi-dimensional data set yields collective coordinates (first three eigenvectors of the covariance matrix with the largest eigenvalues) onto which the original data set can be projected. The resulting low-dimensional representation of the full data set allows a direct and intuitive mapping of the conformational ensembles, that is, the aggregation configuration space.

Please cite this article as: Matthes, D. *et al.*, Driving Forces and Structural Determinants of Steric Zipper Peptide Oligomer Formation Elucidated by Atomistic Simulations, *J. Mol. Biol.* (2012), doi:10.1016/j.jmb.2012.02.004

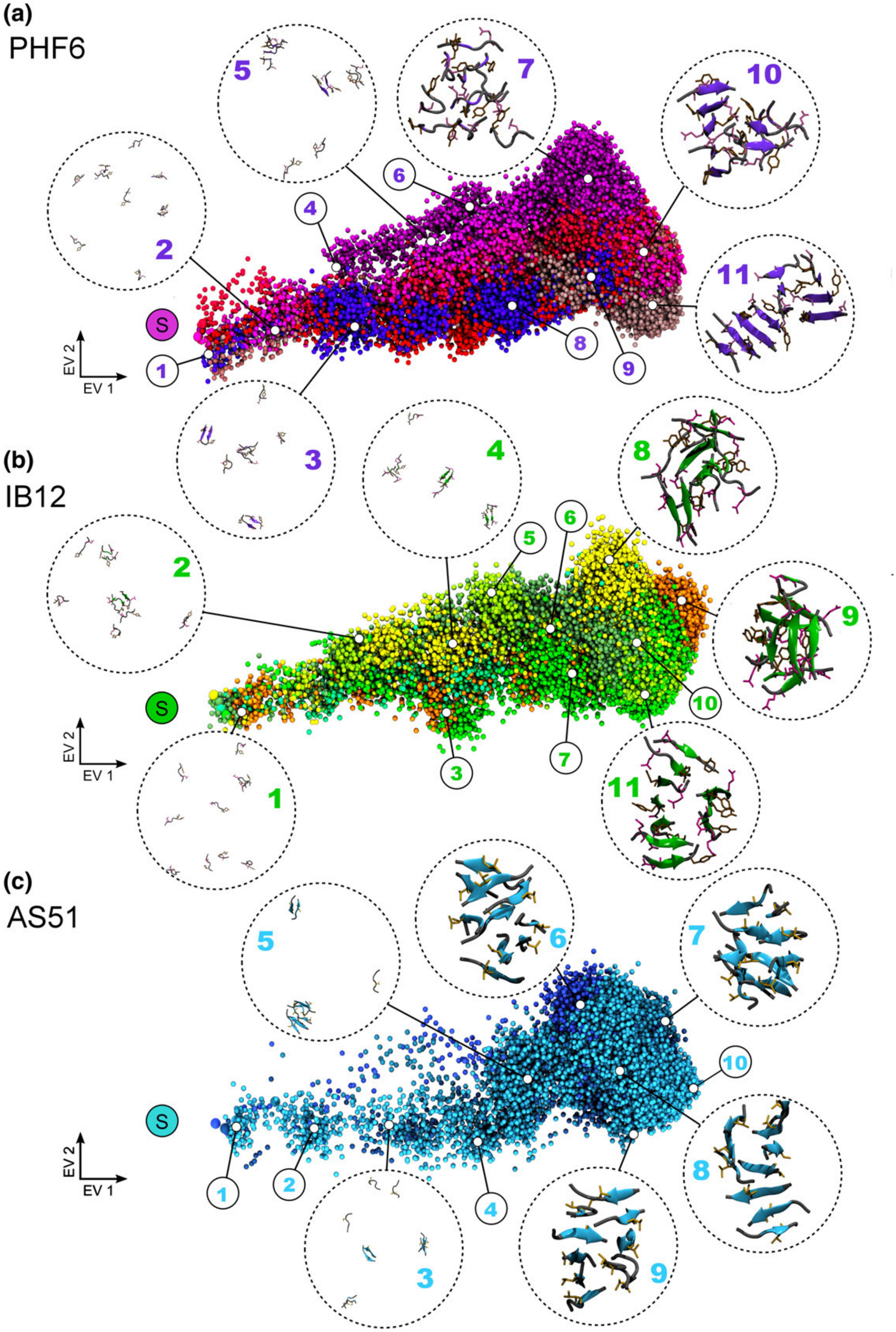

Fig. 3 (*legend on next page*)

read directly together with the associated oligomer structures and configurations. A visual inspection of the representative structures in the various regions of the projections indicates that, for all three peptide systems, the first collective coordinate (EV1) maps the conformational conversion from monomers to oligomeric aggregates (i.e., the general association state). Starting from initially dispersed peptides ("S" in Fig. 3), we observed a rapid and concerted change in several observables as the 10 peptide chains began to collide and aggregate with one another. The corresponding sampling along EV1 of the projections in the collective coordinate space as a function of simulation time is shown in Fig. S1, indicating a convergence within 100 ns. The inter-peptide atom contacts as well as the molecular surface burial increased, while the number of solvent molecules in contact with the peptides decreased. Seen from an energetic point of view, the first collective coordinate coincides with favorable coulombic inter-peptide main chain interactions, whereas those with solvent get less favorable. The same holds for the van der Waals interactions of the side chains between the peptides (increase) and interactions with water (decrease). The change in the other observables was explained to a smaller extent by the first collective coordinate.

The commonly observed general association for the PHF6, IB12 and AS51 peptides resulted in decameric, fully assembled oligomers as the most abundant general aggregate configuration type in all of the individual trajectories. The formed decamers were found to be stable assemblies as dissociation events were negligible (AS51) or not observed (PHF6, IB12). Note that due to the finite system size, decameric oligomers could not grow further in the present setup.

En route to the decamer, intermediate aggregate sizes ranging from dimers to nonamers were transiently formed, as previously reported.[50] The successive assembly of early and intermediate configurations proceeded via monomer addition, as well as condensation of primarily dimeric and trimeric precursor states.[50] Overall, the same association tendency was found for the three different peptide systems. The simulations suggest that assembly toward the final oligomeric state occurred the fastest for AS51. Similar to PHF6, in AS51 simulations, all the aggregate sizes have been sampled; however, the AS51 aggregates of intermediate size ($n = 3$–6) had on average a shorter lifetime, that is, were consumed more rapidly into larger oligomers. In addition, AS51 simulations did explore only a small portion of all the different possible association types. The temporal evolution of the aggregate sizes can be appreciated from Fig. S2, where the average size distributions of PHF6, IB12 and AS51 are shown for two time windows (0–300 ns; 300 ns, end of sim).

The sequence of events in the oligomerization process can be understood by looking at the representative structures and conformations mapped close to them in Fig. 3. Multiple conformations for oligomers of intermediate size were present, such as two- to four-stranded β-sheets in extended and untwisted conformations (Fig. 3a, structures 8 and 9; Fig. 3b, 2, 3 and 6; Fig. 3c, 4 and 5), while others appear bent and collapsed to compact structures (Fig. 3a, 4–6; Fig. 3b, 1, 2 and 5; Fig. 3c, 3). No apparent systematic pattern in strand alignment and registry was found for the diverse mixture of extended and collapsed structures of the different peptide systems.

The individual sampling routes and distribution of states along the first two collective coordinates in the projections of the independent PHF6, IB12 and AS51 simulations can be directly inferred from Fig. 3 and Fig. S1. For example, the PHF6 simulations explored common regions in the projections but also populated distinct configurations as seen for the purple-colored snapshots, which represent mostly collapsed aggregates without significant β-sheet content. These particular conformations diverged very early from the common sampling routes, suggesting an alternative aggregation pathway for PHF6. In multiple simulations, the stepwise assembly of IB12 oligomers was found to proceed as a single β-sheet up to the pentamer, while larger aggregate sizes exhibited lateral growth at the already established sheet surface and eventually β-sandwich structure formation (Fig. 3b, 5, 6 and 11).

As outlined before, the assembly of stable, decameric oligomers was observed in all simulations; however, the β-sheet aggregates sampled by each sequence, as well as in the independent trajectories for each peptide system, were different

Fig. 3. Projection of conformational ensembles and representative structures. Shown for (a) PHF6, (b) IB12 and (c) AS51 are all the sampled conformations projected in a three-dimensional collective coordinate space. Each sphere in the projection represents one trajectory configuration mapped into this space. The color shades encode the different independent simulations for each of the peptide sequences. The larger spheres marked with "S" indicate the location of the equilibrated initial configurations for each simulation. Enumerated representative structures for each map are shown and enclosed by circles with broken outline. Numbers enclosed by small circles point out structures reported in Fig. S3 because of space considerations. Each region of the projection corresponds to a distinct conformational state, as indicated by the representative structures. Spheres in close spatial proximity in the projection represent configurations with high structural similarity (see the text for detailed description). The chosen structures are shown in a cartoon drawing. A subset of side chains is shown in stick representation: PHF6 (Gln, Tyr), IB12 (Glu, Tyr) and AS51 (Thr).

in topology and abundance. PHF6 and AS51 aggregates were found with very similar types of β-sheet configurations within the fully assembled, decameric state. Especially, two- and three-stranded β-sheets in conjunction with a significant amount of disordered chains (PHF6>AS51) occurred with a high probability. The most common sheet topologies were two β-sheet dimers and six disordered chains (here written in the following notation: [2×2+6×1]), as well as [3×2+4×1] and [3+2+5×1]. In contrast, the IB12 decamers were found frequently to be composed of larger four- and five-stranded β-sheets and configuration types such as [5+4+1] and [4+3+2+1]. Interestingly, the total number of observed β-sheet configuration types was roughly the same for all the peptide systems (PHF6, 24; IB12, 24; AS51, 26). The wealth of distinct conformational basins in the decameric state for PHF6, IB12 and AS51 ranged from amorphous to ordered β-sheet assemblies and can be readily identified within each projection (Fig. 3a–c). Specifically, the second collective coordinate (EV2) describes the variance

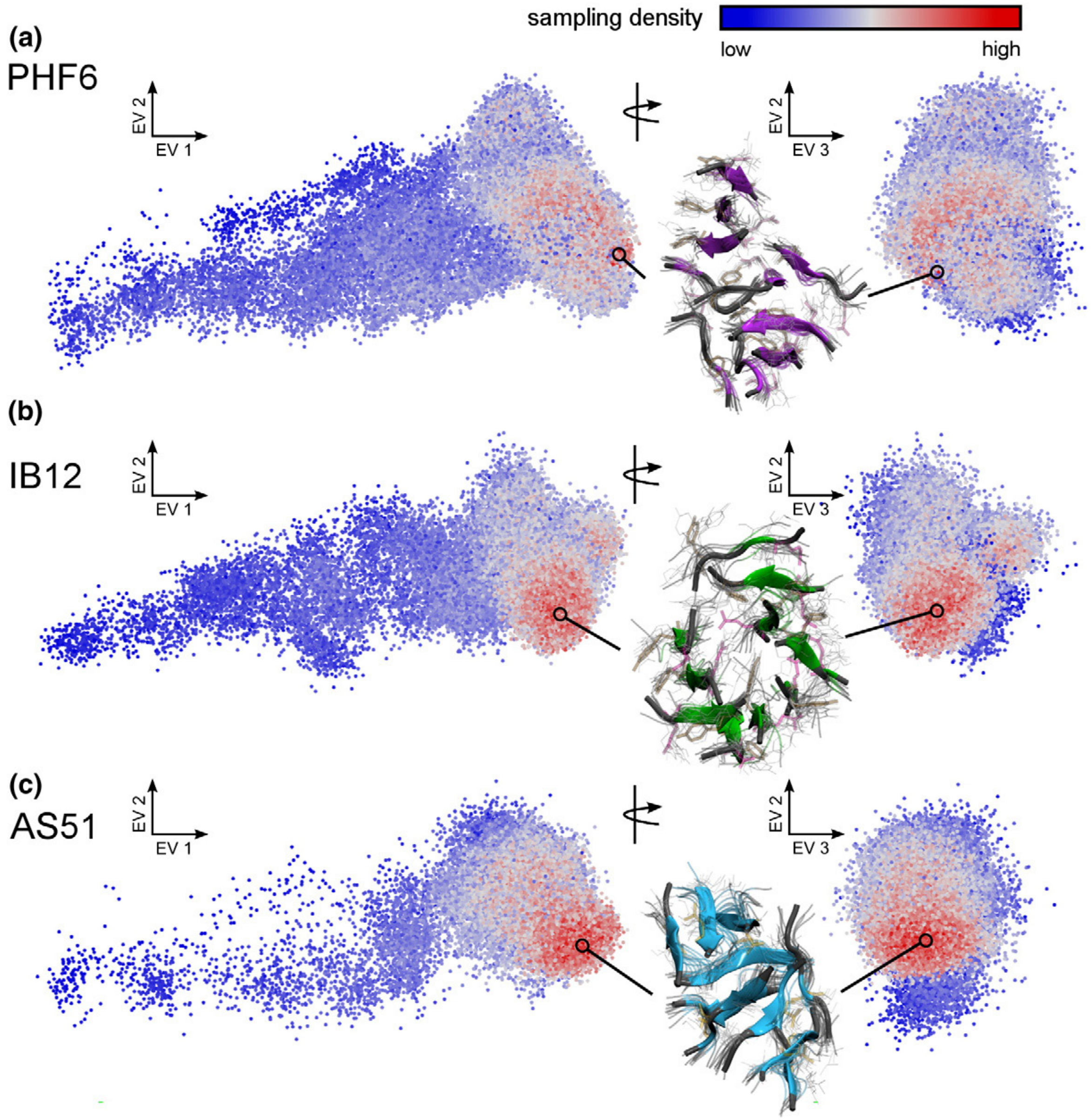

Fig. 4. Projection of conformational ensembles and their sampling density. The sampled conformations in the (a) PHF6, (b) IB12 and (c) AS51 trajectories are projected in the same collective coordinate space as in Fig. 3, where each configuration is represented by a point. The projection is viewed from two different angles, and the sampling density for all simulations is indicated by a color gradient (red, high; blue, low density). A superposition of the structure ensemble (10 oligomer conformations) corresponding to the respective region of highest sampling density is shown for each peptide system.

in β-sheet content (small or large number of β-sheets formed) and β-sheet aggregate configuration types (small or large intact sheets). Most of the decameric oligomers were found to be either β-sandwich-like structures (Fig. 3a, 10; Fig. 3c, 9) or single larger β-sheets facing smaller sheets (Fig. 3a, 11; Fig. 3b, 11) with multiple edge strands exposed to the solvent. In particular for IB12 and AS51 oligomers, oval β-sheet as well as bent barrel-like structures (Fig. 3b, 10; Fig. 3c, 7, 8 and 10) and orthogonal sheets were frequently sampled (Fig. 3b, 9). Aggregates with multiple smaller sheets positioned on top of each other (Fig. 3b, 8; Fig. 3c, 6), globular structures and amorphous assemblies with no specific packing order (Fig. 3a, 7) were sampled as well.

In addition to the supramolecular organization of the individual aggregates, the position and extent of regular packing of the side chains were a prominent feature observed for the different oligomer structures. A coarse but apparent classification of the side chain packing distribution in the PHF6 and IB12 oligomers could be derived by mapping the side chain solvent accessibility for all the aggregate configurations. From this, a preferential packing of the bulky Tyr residues to the interior (Fig. 3a, 7; Fig. 3b, 9), as well as the accumulation of Tyr residue side chains on the outside of the oligomeric aggregates (Fig. 3a, 9 and 10; Fig. 3b, 7, 10 and 11), was seen. Specifically, for all the decameric PHF6 conformations located in the upper and rightmost area of the projection in Fig. 3a (purple spheres, around structure 7), most of the Tyr residues were found to be strongly desolvated, whereas the Gln residues were uniformly oriented to the solvent. In the projection of the IB12 aggregates, a similar region could be identified (Fig. 3b; orange spheres, around structure 9), where the interior of the oligomers was found to be occupied mostly by Tyr residues, whereas the protonated Glu side chains were exposed on the aggregate surface. Interestingly, IB12 aggregate configurations that cluster around structure 11 of the projection in Fig. 3b (green spheres) showed the exact opposite solvent exposure characteristics. The respective oligomer conformations were in all cases stabilized by a transient hydrogen bonding network between the hydrophilic side chains.

Sampling density identifies prominently visited aggregate conformations

In order to investigate the prominently visited structures in each of the aggregate configuration ensembles, we determined the sampling density in the full-dimensional observable space using a k-nearest neighbor (k-NN) approach (see Methods). Despite the varying sampling routes in the independent simulations, the highest densities lie without exception in the ordered region of the decamers, although the highest density appears more localized for the IB12 and AS51 peptide systems (Fig. 4). In the densest sampled regions of the IB12 and AS51 configuration space, oligomers with an established sheet-to-sheet interface were identified. For AS51, an almost closed, flat β-barrel-like structure was found, while IB12 oligomers displayed a buckled β-sandwich aggregate architecture (Fig. 4). The ensemble of PHF6 oligomers extracted from the highest density region was associated with a lower β-sheet content, compared to the IB12 and AS51 aggregates. Interestingly, the rather distorted and less compact arrangement of smaller and twisted

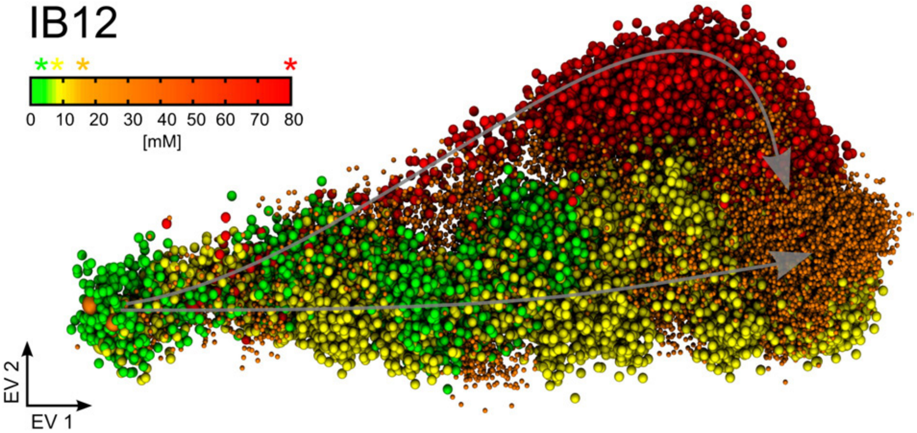

Fig. 5. Projection of IB12 conformational ensembles simulated at different concentrations. Shown are the IB12 conformations sampled at 3.3, 8.3, 16.6 and 83 mM initial concentration and projected in a three-dimensional collective coordinate space. Each sphere in the projection represents one trajectory configuration mapped into this space. All conformations from simulations with the same concentration are colored according to the shown scale. For comparison, the IB12 aggregate conformations from Fig. 3 (16.6 mM) are shown as smaller orange spheres. The gray arrows indicate sampling along the two prominent aggregation pathways observed: via disordered intermediates (upper arrow; red spheres) or ordered intermediates (lower arrow; green, yellow and orange spheres).

β-sheets featured numerous inter-peptide Gln side chain-to-main chain contacts (Fig. 4a).

Concentration dependence of the oligomerization process

To address the concentration dependence of the oligomer growth process, we have carried out additional M10 simulations of the IB12 peptide system in a range of concentrations above (83 mM) and below (3.3 mM, 8.3 mM) the initially simulated 16.6 mM (see Table 3). As before, the results are presented in the form of a collective coordinate mapping (Fig. 5). The projections clearly show that the systems with lower IB12 monomer concentration follow the same sampling routes toward the oligomeric states of higher order, as discussed above (16.6 mM, M10 simulations).

The lag time for the formation of decameric aggregates was found to be prolonged significantly due to the reduced probability of diffusional encounter of the peptide aggregates in comparison to the higher concentrations. No apparent concentration dependence on the stability of the formed aggregates was observed. In contrast, the highly saturated simulation systems (83 mM) showed a prominent excursion via initially isotropic and amorphous aggregates caused by the near-instantaneous collapse of the peptides, resulting from their small initial separation. For these disordered oligomer aggregates, a subsequent conformational transition toward β-strand structure was observed (Fig. 5).

Overall, the assembly pathways taken toward the decameric state were found to depend on the IB12 monomer concentration as sketched by the gray arrows in Fig. 5. However, the final structures of each of the simulations approached ordered aggregate conformations with consensus steric-zipper-like structural features, irrespective of the initial concentration.

Spontaneously formed aggregates and steric zipper oligomers converge to similar structures

Some of the observed oligomers have topological resemblance to a cross-β sheet motif, the basic structural element of peptides found in the fibrillar[4,17] and crystalline states,[15,18] which is characterized by a tightly packed pair of sheets, an ordered arrangement of strands and interdigitated side chains. To assess if and to what extent the spontaneously formed oligomers relate or even converge to aggregate structures with such an ideal sheet packing motif, both in terms of structural and energetic properties, we set up additional simulations for the PHF6 and IB12 peptide systems (see Methods). The two starting configurations were modeled based on the available crystal structure coordinates: a β-sheet composed of 10 strands (SH10) with extensive backbone hydrogen bonding and a planar β-sandwich of two facing β-sheets with 5 strands each (SH5-5), making up a steric zipper interface. Compared to the single β-sheet, the SH5-5 configuration featured four instead of two edge strands and less solvent exposed side chain surface due to the different packing. Note that the strands were organized in parallel within the PHF6 aggregates, whereas the strands were aligned in antiparallel fashion in the case of IB12. The trajectories that used the crystallographic conformations as initial structures will be referred to as reference simulations throughout. The configuration ensemble sampled in the reference simulations was investigated based on the same observables as before and evaluated together with the data set from the spontaneously formed oligomers. The low-dimensional projections of the individual structures on the collective coordinates of the concatenated ensemble (spontaneous and reference) are shown in Fig. 6a and b. The collective coordinates (and their components) did not change significantly compared to Fig. 3, such that the shape of the projected spontaneous structure configurations of PHF6 and IB12 was only slightly altered. The starting structures ("R" in Fig. 6a and b) and most of the sampled configurations of the reference simulations appeared in regions not accessed by the M10 simulations. As expected, the reference configurations initially clustered around the regions of the projection corresponding to decameric oligomers with very high β-sheet content as well as ordered, intact and large sheets.

In the collective coordinate representation, it can be seen that the PHF6 and IB12 reference simulations (SH10 and SH5-5) explored the space mostly along the second and third (perpendicular to the paper plane) PCA vectors toward the configurations spontaneously formed in the M10 simulations (Fig. 6a and b). The PHF6 reference structures lost their initially high β-sheet content and diverged quickly from aggregate topologies with large sheets and extended peptide chains. Furthermore, directly readable from the collective coordinates is a gradual reduction in nonpolar PHF6 atom contacts with the solvent. In comparison, through examination of the sampling of the IB12 reference structures in the collective coordinate map, in particular a loss in favorable coulombic interactions between the peptide main chains is found, while van der Waals interactions among the peptide main chain and side chain atoms are increased (also partially described by a higher number of nonpolar atom contacts). The projections in Fig. 6 and the shown representative structures clearly illustrate that the two sets of reference structures relaxed differently in the explicit solvent environment as well as when comparing the different independent PHF6 and IB12 reference simulations. In particular, the single

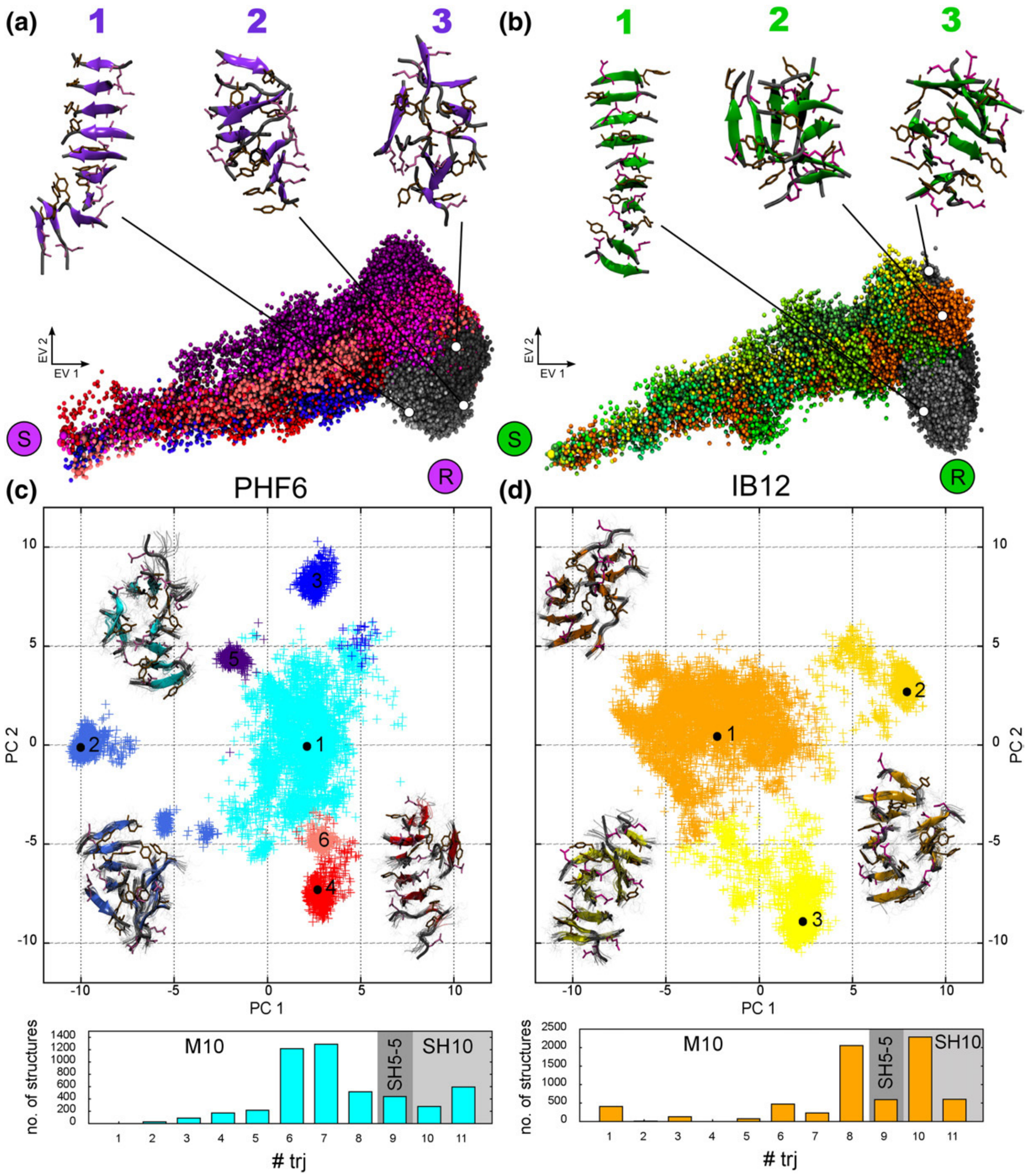

Fig. 6. Projection and clustering of spontaneously assembled and reference oligomer ensembles. The sampled configurations of various trajectories are shown as projections on the collective observable coordinates in (a) for PHF6 and for (b) IB12. Simulations starting from initially monomeric as well as the modeled, reference conformations are shown together in this projection. The simulations of the single β-sheets are shown in light gray and silver, and β-sandwich simulations, in dark-gray spheres. The larger spheres marked with "R" indicate the location of the initial reference configurations. Representative structures of the reference simulations are shown together with index numbers for each map as cartoon drawing. Commonly sampled oligomer conformations from spontaneous and reference simulations of (c) PHF6 and (d) IB12 peptides are shown in a projection onto the first two principal components obtained by PCA in Cartesian coordinate space. In (c), the six conformational clusters identified for PHF6 oligomer structures are shown in different colors, containing mixed (1, cyan) and conformations from spontaneous (4) and reference (2, 3, 5, 6) simulations only. In the same way, in (d), the three clusters identified for IB12 oligomers, containing mixed (1, orange) and conformations from spontaneous (2) and reference (3) simulations only, are shown. The location of the center structures of the three largest clusters is indicated by a black dot. For each of these clusters, a superposition of the center structures and the nine closest oligomer conformations are shown in cartoon representation. As before, Gln, Glu and Tyr side chains are shown in stick representation.

10-stranded PHF6 and IB12 β-sheets evolved toward different supramolecular organizations as also seen from the time evolution of the radius of gyration (R_g) in Fig. S4. The IB12 SH10 state reproducibly showed a prominent twisting and compactation on a timescale of 100–300 ns (Fig. 6b, 1–3). Interestingly, both IB12 SH10 simulations showed a break up roughly in the middle of the single β-sheet, after which the oligomer reorganized to more compact conformations very similar to a β-sandwich, transiently sampling orthogonal packed β-sheets (Fig. 6b, 2). In contrast, the observed conformational rearrangements in the single PHF6 β-sheet were mostly the result of strand reorientations and an overall β-sheet twisting (Fig. 6a, 1). The initially parallel PHF6 strands flipped mainly on the open edges but also within the β-sheet, after a partial break up and loss of β-sheet structure occurred. A stable β-sheet conformation was reformed in one PHF6 simulation, while in the other, the elongated sheet structure was only partially recovered and stabilized by two adjacent chains (Fig. 6a, 1).

Only a small decrease in R_g was observed for the PHF6 and IB12 simulations starting from the SH5-5 configurations. Given the comparable but minor extent of compactation seen for both PHF6 and IB12 SH5-5 reference simulations, the structural integrity of the respective oligomeric states differed substantially. The IB12 β-sandwich oligomer twisted from the planar starting configuration, and a partial loosening as well as a migration and sliding of edge strands was observed. However, the tight and complementary β-sheet interface was essentially preserved on the microsecond timescale. In the case of PHF6, a complete disarray of the SH5-5 configuration to a more globular organization of the oligomer was found (Fig. 6a, 2 and 3), where smaller sheets are packed together and with the β-strands rearranged again toward a predominant antiparallel alignment.

Identification of commonly sampled conformations and extent of overlap

The projections in Fig. 6a and b reveal a number of commonly sampled conformational states in the spontaneous aggregation and in the reference simulations for PHF6 and IB12 oligomers, respectively. Here, we examine whether conformations from the M10 simulations access the same configurational states as the SH10 or SH5-5 simulations and which features these aggregates share. The ensemble of similar structures from both pools of simulations was investigated by using a classification scheme, searching for k-NNs in the space of observable configurations (see Methods). The number of identified overlapping structures was found to be 8194 oligomer structures for PHF6 (3.2% of all configurations) and 10,626 in the case of IB12 (6.8%). To analyze the obtained configurations with high structural detail, we performed a PCA based on the corresponding Cartesian structure coordinates (see Methods). Afterwards, similar conformations were grouped by applying the *k*-means clustering algorithm.

The results are shown in Fig. 6c and d. For both PHF6 and IB12 oligomers, the largest identified conformational cluster (no. 1; PHF6, cyan and IB12, orange) contains structures from the M10, SH5-5 and SH10 simulations, confirming structural overlap in Cartesian space. Interestingly, seven out of eight PHF6 and IB12 simulations from the initially monomeric state sampled this overlap region (cluster 1). Transitions toward these particular conformations occurred rapidly, within 150 ns for PHF6 and within 120 ns for IB12. In Fig. S5, a histogram is shown, reporting the probability to find PHF6 and IB12 conformations in cluster 1 at a given time.

The other clusters are identified as dense and distinctly separated states in the two-dimensional PCA projection and contain structures from either the spontaneous or the reference conformational ensembles. In comparison to IB12, more than one conformational cluster with only reference structures was found for PHF6. This can be explained by the larger structural diversity observed for the different PHF6 reference simulations, although in total, fewer overlap conformations were identified than for IB12.

As one can see from the respective center structures shown in Fig. 6c and d, the main difference between clusters concerns the packing arrangement and size of the β-sheets. All identified PHF6 and IB12 oligomers were β-sheet-rich structures with extended peptide conformations. In detail, for cluster 1 (mixed) of PHF6, a fairly irregular and disordered packing was found, with small and twisted sheets positioned side by side and on top, while cluster 2 (reference structures only) showed an orthogonal β-sandwich structure (Fig. 6c). Finally, cluster 4 (spontaneous structures only) consists of a large twisted and bent sheet of eight strands with mixed alignment, facing two disordered peptide chains. The topology of the latter PHF6 structures is comparable to the respective spontaneously assembled IB12 oligomers (cluster 2), showing a similar elongated, twisted sheet with exposed edge strands and a smaller stabilizing sheet on the side (Fig. 6d). A preferential burial of Tyr and simultaneous exposure of Gln and Glu residue side chains were found for the PHF6 and IB12 oligomers in cluster 2 (reference structures only) and cluster 1 (mixed). The contrary scenario, where the Tyr residues point to the solvent, was found for the IB12 clusters 2 (spontaneous) and 3 (reference). These findings furthermore suggest that topologically similar oligomer populations can show quite different solvent accessibilities on the residue side-chain level, similar to what has been reported in Fig. 3a and b.

In summary, it could be deduced that the highest sampling density of the M10 simulations lies near the overlap region (Fig. 6). Moreover, it was shown that spontaneous and reference simulations of PHF6 and IB12 peptides sampled an ensemble of structures with the common characteristics of an established sheet-to-sheet interface. Although spontaneous assembly toward these particular structures was fast, the formed oligomers as well as the modeled reference states reorganized considerably throughout the simulations; therefore, complete convergence to highly ordered cross-β structures with steric zipper was not reached.

Elucidating the driving forces of the oligomerization process

The collective coordinate mapping illustrates that the interactions of peptide and solvent molecules are among the most important driving forces underlying the oligomerization process as well as an important determinant for the conformational dynamics in the assembled state. In order to quantify the peptide–solvent interactions in the spontaneous aggregation of the PHF6, IB12 and AS51 peptides, we estimated the free-energy difference of transfer from the initially monomeric to the assembled oligomeric state. Similar to protein folding processes, where the solvation free energies are lower for folded than for unfolded conformations,[70] a prominent reduction is expected for the aggregation process. Indeed, the computed atomic solvation free energies[70] decreased by more than half compared to the initial value for all the simulations starting from monomers.

The relation between the solvation free energy and the density of sampling in the collective coordinate space shows an interesting nonlinear behavior (Fig. 7) for all three peptide systems. The different degree of correlation apparently arises from the two different aggregation phases, which in turn directly reflect the observed two principal stages of the oligomerization process: the assembly from monomers to decamers, where a high correlation of the solvation free energy with the general assembly was found, and the subsequent structural transitions within the aggregated state, which show no apparent correlation.

This suggests that the solvation free energy drives the peptides together in the initial aggregation phase but does not play a prominent role in the further development of the oligomers. A functional mode analysis[71] (FMA, see Methods) was carried out to examine additional structural determinants and main forces driving the peptide aggregation. In general, the FMA attempts to correlate a suitable variable with a linear combination of a set of observables. Here, we select the sampling density in aggregate configuration space, that is, the degree of aggregation, as an indicator for the aggregation progress. The idea of the FMA approach is then to construct a linear multiple regression model for the aggregation process based on the sampling density, as exemplified above in one dimension for the solvation free energy.

For a model with adequate predictive power, this allows an assessment of the most prominent factors that govern the peptide oligomerization on a molecular level. Motivated by the observed difference in correlation to the solvation free energy and corresponding to the fast association to stable decamers as the predominant species, as well as convergence along EV1 in the collective coordinate projections (Fig. 3 and Fig. S1), we chose to split the analysis. Thus, the peptide association events and the decameric phase were investigated separately, as they apparently follow two different mechanisms.

It is important to note that although the analysis is based on the collective trends in the data of multiple trajectories, they may not be fully converged. Therefore, extending the simulations could change the observed densities, especially in the decameric phase.

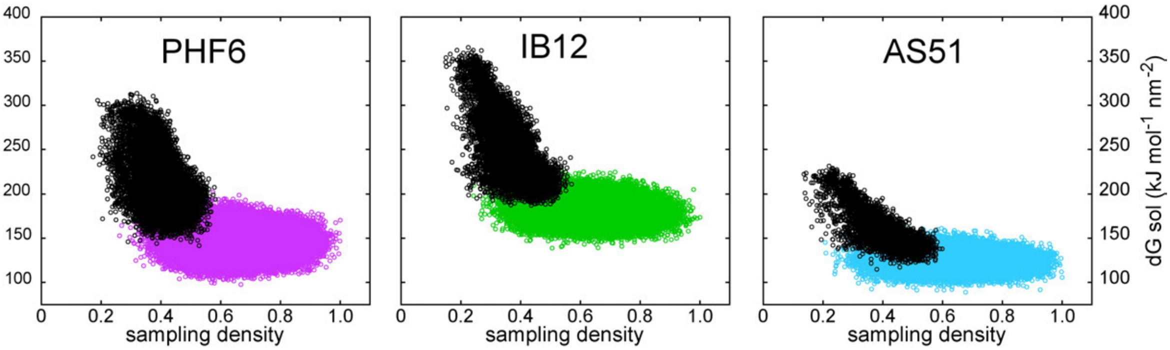

Fig. 7. Correlation of the solvation free energy to the sampling density. The correlation of the sampling density with the solvation free energy is shown for each peptide system. Notably, the sampling density in the initial association phase (black) and the subsequent structural ordering in the decameric state (colored) are correlated to a different extent.

Please cite this article as: Matthes, D. *et al.*, Driving Forces and Structural Determinants of Steric Zipper Peptide Oligomer Formation Elucidated by Atomistic Simulations, *J. Mol. Biol.* (2012), doi:10.1016/j.jmb.2012.02.004

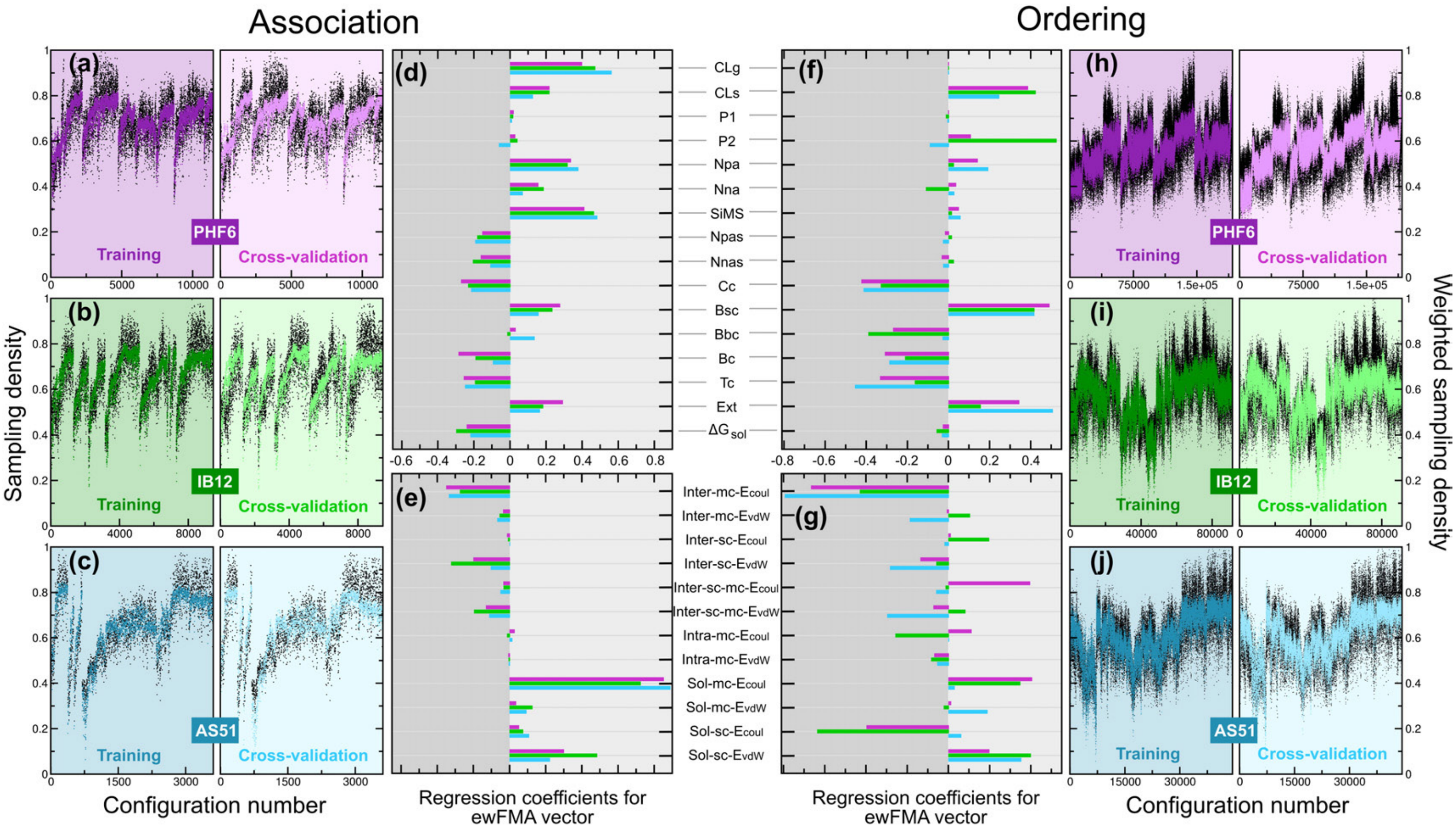

Fig. 8. FMA elucidates the aggregation driving forces. FMA models were built to predict the sampling densities in the aggregate configuration space of the (a and h) PHF6, (b and i) IB12 and (c and j) AS51 peptide systems, for both the peptide association (left) and the structural ordering in the decameric phase (right), respectively. The respective models were built on the full simulation data available, whereas in a second step, an iterative procedure was applied to test and cross-validate them against a completely independent set of data (see Methods). The contribution of each observable parameter to the changes in sampling density was analyzed using an ensemble-weighted FMA[71] model. The observables in terms of aggregate descriptors (d and f) and energetic components (e and g) were evaluated separately and had either positive or negative regression coefficient with the increase in the sampling density. Abbreviations and description of the observables are given in Methods.

To distinguish mere descriptors for the aggregation process (group 1) from putative true driving forces (group 2), we split the PHF6, IB12 and AS51 FMA models into two separate groups of observables. In the second group, the pairwise additive force field energy terms are included, which model the principal interactions between the molecules (hydrogen bonds, hydrophobic effect, etc.).

Figure 8a–c depict the FMA models for the three peptide systems in the association stage. The predictive power of the FMA models was assessed by cross-validation, using an independent data set not used for model training. The cross-validation correlation coefficients for the FMA descriptor models are shown in Table 1. For the association phase, the diagonal elements in Table 1 indicate that the created models are capable of predicting the progress of peptide assembly for the system used to train the algorithm (correlation coefficients > 0.7). Interestingly, the off-diagonal correlation coefficients reveal that the models trained on the data of one peptide system were able to mutually predict the densities of the other peptide systems.

After validating the FMA models for the peptide association phase, we investigated the individual model components to probe the contributions of specific interactions to the aggregation mechanism. In the subsequent analysis, positive coefficient values indicate that an observable was positively correlated with the configurational space density of the peptide aggregates (i.e., the conversion from monomeric to decameric structures). For the energy terms, negative regression coefficient values indicate favorable interactions driving the assembly, that is, favorable changes in an energy term were negatively correlated with the density.

In the following, the main findings for the association process are summarized (Fig. 8d and e). For example the AS51 FMA model for the association phase correctly indicates the aforementioned slightly faster aggregation to higher-order oligomers, as can be seen from the stronger influence of CLg (the assembly to general aggregates) compared to PHF6 and IB12. The oligomerization of all peptide systems is accompanied by a reduced solvent-accessible surface area and the preferential burial of hydrophobic groups. This is consistent with the finding that water molecules were excluded from the peptide aggregate interior upon assembly from the initially fully solvated monomers and can be seen from the negative correlation for the number of polar and nonpolar peptide contacts to water molecules (Npas, Nnas). While the exchange of solvent molecules with the peptide surface was on a picosecond timescale, the total extent of water release from the peptide hydration layer into the bulk was in general about 40% in the association phase and directly correlated with the decrease in the solvation free energy. The importance of the interaction with the solvent molecules for the aggregation process is illustrated furthermore by the strong positive correlation between the sampling density and the formation of large, solvent-inaccessible interfaces between the peptides (SiMS). The secondary structure conversion from predominantly coil and turn to β-sheet conformations as well as to more extended peptide chains was found to be correlated with the peptide association to a similar degree. Interestingly, neither the polar (P1) nor the nematic order (P2) of the peptide aggregates was a necessary component to predict the sampling density in the association phase, as they have relatively low correlation for either of the peptide systems.

All the energetic properties concerning the interactions of the peptides with the solvent have positive regression coefficients, that is, the aggregation process is accompanied by an increasing energy between the peptides and water. Here, the coulombic interactions of the water molecules with the peptide main chain groups (amides and carbonyls) showed the most prominent loss. The van der Waals interactions between the water and peptide side chains were reduced significantly as well. A concurrent gain in inter-peptide interactions was found mainly from contributions of coulombic main chain and van der Waals side chain interactions. These two mostly account for the hydrogen bond formation and hydrophobic collapse as the peptides rapidly oligomerize in the early association phase. In addition, mixed inter-peptide main chain and side chain interactions were observed frequently as the peptides formed oligomeric aggregates with a substantial degree of nonspecific contacts and disorder. Although not included in the FMA model analysis because of the large-scale fluctuations, the increase in solvent–solvent interactions again stresses the important role of the solvent in the aggregation process.

Table 1. Correlation coefficients for the PHF6, IB12 and AS51 FMA models of association and ordering in the decameric phase

	PHF6		IB12		AS51	
	Train	Test	Train	Test	Train	Test
Association						
PHF6	0.74	0.70		0.79		0.77
IB12		0.71	0.82	0.77		0.75
AS51		0.60		0.77	0.87	0.82
Decamer (weighted)						
PHF6	0.74	0.70		0.74		0.76
IB12		0.69	0.81	0.78		0.76
AS51		0.70		0.66	0.81	0.80

Row-wise: peptide system used for FMA model building. Column-wise: peptide system used for FMA model cross-validation.

Please cite this article as: Matthes, D. *et al.*, Driving Forces and Structural Determinants of Steric Zipper Peptide Oligomer Formation Elucidated by Atomistic Simulations, *J. Mol. Biol.* (2012), doi:10.1016/j.jmb.2012.02.004

The FMA model cross-validation and component analysis show that common driving forces and structural determinants are at play in the initial association stage of all three peptide systems. This suggests a largely sequence independent mechanism based on the generic physicochemical properties of the peptide molecules. In comparison, the correlation coefficients of the FMA models for the decameric phase were in general substantially smaller, while the trends in the predictive power of the models are preserved (AS51 > IB12 > PHF6, see Table S1). In particular, the PHF6 sampling density in the decamer appears to be the most difficult to predict for any of the three models. The rather poor predictive power of the PHF6 FMA model suggests that the sampled aggregate configurations and structural transitions are not described well by just a single model, and this is probably due to the less pronounced gradient in the sampling density, as also seen from Fig. 4a.

In order to probe the formation of particular ordered aggregates in the decameric phase featuring cross-β characteristics, we built an additional FMA model on the basis of a weighted sampling density. Oligomeric aggregates with high nematic order (P2) and β-sheet content, which are presumably similar to fibrillar β-sandwich aggregates, therefore received a higher weight prior to the FMA [weighting factor: exp(P2·Bsc)]. From the newly obtained weighted sampling density (Fig. S6), a striking result is the more localized density for PHF6 and IB12, while the AS51 sampling density remains virtually unchanged. This is due to the very small fraction of AS51 structures that fulfill both of the structural order criteria. Common for all the three peptide systems, the structural ordering in the initially collapsed decameric peptide assemblies was found to be positively correlated with the sheet configuration type of the aggregates (CLs, β-sheet aggregation to large sheets) (Fig. 8h–j). The common features of the most probable ordered oligomeric states were found to be the favorable interactions between coulombic inter-peptide main chain atoms (extensive hydrogen bonding), as well as van der Waals interactions from tight inter-peptide side chain packing. In detail, the ordering in the decameric phase was accompanied by an increase in β-sheet structure (Bsc) and peptide chain extension (Ext), while formation of extended β-strands occurs at the expense of any other secondary structure elements (coil, bend, turn). The conformational dynamics in the assembled state are governed by subtle, sequence-dependent differences as already seen in the different mapped structure ensembles. For example, the formation of β-sheets was found to be less prominent in the AS51 oligomers, whereas the sampling density in the configurational space was found to be positively correlated with the β-bridge content (Bbc). This is also expressed in the smaller correlation of the sampling density to the CLs observable and explained by the prominent occurence of closed, barrel-like sheet structures with low nematic order (P2). Moreover, there were notable contributions from interactions between main chain and side chain atoms, which stabilized the AS51 aggregates, opposed to a decrease in these quantities for PHF6 (coulombic) and IB12 (van der Waals) aggregates. The observed decrease in coulombic interactions between the peptide main chain and side chains in the PHF6 decamers is due to the reduced number of Gln side chain hydrogen bonds to the backbone, thus offering an explanation for the structural stability of oligomer structures with only small β-sheet content and their slow conversion to more ordered ones. The loss of these nonspecific interactions allowed for the growth of double-layer sheets with interdigitated side chains. The structural ordering in IB12 oligomers occurred mainly by orientational reorganizations to parallel pairs of sheets, hence explaining the high correlation with P2. Therefore, a more efficient packing of adjacent β-sheets was achieved, leading to a larger interface burial (SiMS) and a slightly higher solvent contact. These sheet rearrangements are indicative of transitions to IB12 oligomer structures with a dry and wet sheet interface, as well as solvent-exposed edge strands. While the loss in van der Waals peptide side chain-to-solvent interactions was commonly found to decrease further for the decameric states of PHF6, IB12 and AS51, this was not the case for the coulombic part, which was favorable for the aggregate ordering of the more hydrophilic PHF6 and IB12 peptides.

Overall, a picture emerges where a hydrophobic solvent effect appears to drive the initial clustering of the peptides into collapsed and partially ordered aggregates. This is followed by a phase of structural ordering in the fully assembled state, primarily characterized by a further buildup of β-sheets and, in particular, by the formation of dry inter-sheet interfaces. The FMA results corroborate the obtained structural information on the respective oligomer structures and rationalize the energetic contributions and sequence-dependent differences on a molecular level.

Effect of mutations on oligomer topology and aggregate order

Based on the analysis on structural determinants and driving forces underlying the primary aggregation steps, we hypothesize that a mutational study might give additional insight into the aggregation mechanism of amyloidogenic peptide model systems. Several mutations were chosen to probe the residue-specific effect on the interactions driving the formation of ordered, β-sheet-rich aggregates, similar to what has been done previously in hexapeptide fibrillization assays.[42,72] In particular, we tested the impact of side chain mutations on the aggregation kinetics and sheet-to-sheet packing organization in steric zipper peptide oligomers. The AS51 peptide (GVATVA) was selected for the mutations since it aggregates into oligomers with the least stabilization through β-sandwich structure-like interfaces and therefore is considered a suitable candidate for the predictions to directly test our simulation hypothesis.

In this context, exchanging Val, Ala and Thr residues in neighboring positions in the peptides should not affect the β-strand formation propensity strongly, whereas shifting Ala and Val residue side chains to the same side of the strand might lead to different sheet interfaces upon assembly (GVATAV, GVAVTA). Introducing an additional site with hydrogen bonding capability (Ala3Thr) should increase disordered inter-strand main chain-to-side chain interactions (GVTTVA). The presence of additional Ala residues is expected to reduce the inter-strand side chain stabilizing contacts, on the one hand leading to more random-coil conformations and on the other, weaker inter-sheet packing interactions (GAATAA). In contrast, a mutation to Phe at position 5 is predicted to have a high β-sheet propensity and the potential to form a large number of inter-peptide interactions. However, replacing Val5 with the bulky and strongly hydrophobic Phe residue might alter the initial assembly of the oligomers as well as the otherwise sterically less demanding sheet interface to a large extent (GVATFA).

The spontaneous oligomerization was studied for each peptide mutant with an additional 1. 25-μs-long simulation. Indeed, a number of predicted effects take place upon mutation, illustrated by the set of preliminary results summarized in Table 2 and the representative oligomer structures shown in Fig. S7. As expected for the GVTTVA peptide, ordering transitions were impeded by strong main chain-to-side chain interactions, such that the oligomer structures were mainly composed of dimer subunits packed together. The GAATAA peptide aggregates were indeed found to be mostly disordered, with a low β-sheet content and additionally no apparent interface formation. In stark contrast, the interior of the β-sheet-rich GVATFA decamers was found to be occluded by the introduced Phe side chains. The sheet topology in these oligomers was rather wedge shaped and did not resemble a β-sandwich. The interface characteristics in GVATAV and GVAVTA, in comparison to the GVATVA peptide aggregates, emphasize the subtle effects of side chain complementarity in small β-sheet oligomers. While the GVAVTA oligomers formed two orthogonal and stably packed β-sheets, the GVATAV decamers had in addition to a high β-sheet content also a high nematic order. This was found to be due to the parallel arrangement of the sheets with defined inter-sheet interface, leading to weaker main chain-to-side chain contacts and a high average number of exposed edge strands. We conclude that these mutations therefore also offer an attractive means to probe the aggregation kinetics of such peptides experimentally.

Discussion

The application of a collective coordinate mapping allowed for the detailed characterization of the diverse conformational ensembles for the simulated oligomerization process of three steric zipper peptides. This mapping procedure enabled the visualization of the structural inter-relationships of the sampled multimeric aggregates and thus provided a comprehensive overview of the global trends in the large-scale molecular changes accompanying the studied assembly process. At the same time, the discriminative power of the mapping allows a direct identification of appropriate reaction coordinates for the projection of high-dimensional configuration spaces.[73,74]

The investigation of the oligomerization process and the underlying driving forces revealed two principal phases and illustrates the delicate balance between inter-peptide and peptide–solvent as well as solvent–solvent interactions. The first step, the

Table 2. Summary of structural and energetic properties for the AS51 peptide mutant simulations

System	β-Sheet	P2	$E_{\mathrm{coul+vdW}}$ mc	$E_{\mathrm{coul+vdW}}$ sc	$E_{\mathrm{coul+vdW}}$ mc-sc
GVATVA	0.44 ± 0.11	0.26 ± 0.07	−806 ± 80	−208 ± 17	−349 ± 54
GVAT**AV**	0.46 ± 0.11	0.49 ± 0.11	−859 ± 69	−192 ± 19	−269 ± 27
GVA**VT**A	0.55 ± 0.09	0.22 ± 0.06	−817 ± 65	−207 ± 14	−360 ± 40
GV**T**TVA	0.30 ± 0.09	0.22 ± 0.07	−920 ± 66	−237 ± 26	−402 ± 51
GVAT**F**A	0.53 ± 0.13	0.41 ± 0.09	−798 ± 66	−330 ± 22	−361 ± 33
G**A**AT**A**A	0.30 ± 0.11	0.34 ± 0.11	−916 ± 87	−118 ± 13	−296 ± 33

Mutations in the AS51 peptide are indicated in boldface. The average and standard deviation were calculated for various observables over a 500-ns interval for each simulation (0.75–1.25 μs): β-sheet content, orientational order P2 and the sum of several coulombic and van der Waals potential energy terms ($E_{\mathrm{coul+vdW}}$) describing the interactions. Abbreviations and description of the observables are given in Methods.

Please cite this article as: Matthes, D. *et al.*, Driving Forces and Structural Determinants of Steric Zipper Peptide Oligomer Formation Elucidated by Atomistic Simulations, *J. Mol. Biol.* (2012), doi:10.1016/j.jmb.2012.02.004

association from monomers to oligomers, was commonly found to be correlated with the reduction in solvation free energy, leading to a hydrophobic collapse of the peptides and minimized peptide surface area in the aggregates. The primary cause of the loss in peptide–solvent interactions was determined to be the poor solvation properties of water for the polypeptide backbone and a hydrophobic dewetting of the peptide side chains. A similar preference of peptide–peptide over peptide–solvent interfaces has been discussed for the aggregation of strongly hydrophilic polyglutamine chains.[7,75] In addition to a gain in water–water hydrogen bonds, the release of water molecules into bulk solvent is postulated as a universal driving force to further facilitate the aggregation process by an increase in solvent entropy,[51,65,76] opposing the loss of conformational entropy in the peptide aggregates.

The observed concurrent increase in inter-peptide interactions is in accordance with previous simulations, where the partitioning of the residue side chains from the aqueous phase preceded the inter-peptide hydrogen bond formation.[62] Here, the emerging ensemble of early steric zipper peptide oligomers was characterized by stable and compact but only partially ordered aggregate structures, due to frequently formed inter-peptide main chain-to-side chain contacts. For the PHF6 peptide system, this scenario was particularly predominant, as the Gln side chains function as a good solvent for the peptide backbone and nonspecific hydrogen bond formation led to stronger and persisting aggregate disorder. Moreover, the subsequent conformational rearrangements in the decameric structures were marked by the development of more directional, nonbonded inter-peptide interactions in form of extensive hydrogen bonds between peptide strands, resulting in a significant amount of β-sheet structure. In line with previous findings,[75] we observe that the condensed peptide aggregates with dry interior, where peptides are primarily solvated by other chains, promote the backbone-driven β-sheet structure formation and contribute to the overall stability of the aggregates.

The formed decameric structures were subject to fluctuations on the microsecond timescale, and despite the rich structural diversity, a notable finding was the common formation of a dry inter-sheet interface. The ordered peptide arrangements originated from stronger inter-strand interactions in the β-sheets and further optimized inter-sheet packing through interdigitating side chains. Although multiple sampling routes, involving the presences of various intermediate conformations,[50] were taken on the way toward the decamers, the difference between the ordered PHF6, IB12 and AS51 oligomers in the decameric phase was mostly related to the sheet topology (β-sheet size and packing). For the hydrophobic and less bulky AS51 peptide, the largest fluctuation in β-sheet content and a more continuous hydrogen bonding pattern were found. Hence, the conformational characteristics of the most abundant AS51 aggregates were mostly β-barrel-like structures with a dry interior. In contrast to AS51, the PHF6 and IB12 peptides spontaneously formed frayed and twisted β-sandwich oligomers with a significant amount of solvent-exposed edge strands.

Some of the observed oligomers achieved a close structural resemblance to the basic cross-β element proposed for the fibrils.[15,17] However, they were not fully ordered compared to the steric zipper motif in the crystalline state, which displays two planar β-sheets positioned face to face.[15] In particular, the high polar and nematic β-strand order remains an important difference to the established interfaces of the small and polymorphic oligomers found here. Rather, additional degrees of freedom, such as sheet-to-sheet angle and surface curvature, strand alignment and residual coil and bend conformations characterized the oligomeric states. Compelling evidence for a larger heterogeneity in sheet pairing arrangements in non-fibrillar aggregates comes from recent crystal structures of a minimal, tetrameric assembly unit of macrocyclic peptides.[77] There, β-sheet dimers with steric zipper analogous interfaces were found.

This furthermore supports the notion that the formation of complementary and dry quaternary inter-sheet contacts is a common feature of oligomers and fibrils,[40,77] thereby extending the view on the structural organization of oligomers at a pre-nucleation stage. Further growth of pre-fibrillar intermediates beyond a critical size[25–27] is proposed to proceed by replicating the principal zipper structure via addition of β-strands.[49,56,57,77] In our simulations, the reference IB12 cross-β oligomers with the crystal packing were at most metastable for the studied size of 10 chains and relaxed to conformations similar to the spontaneously formed ones. Interestingly, the SH5-5 construct of the PHF6 peptide with a purely hydrophobic crystal structure interface was found not to be stable in our simulations and by others, where even larger steric zipper assemblies were probed.[78] Thus, the presence of polar and aromatic residues likely provides a stronger contribution than hydrophobic ones to the interface stability of opposing β-sheets in small oligomeric aggregates.[55,58,78] This was confirmed by our simulations of the AS51 peptide mutants, which showed an overall enhanced propensity to sample stably packed β-sheets when a Phe residue side chain was introduced into the sequence. The preferential antiparallel strand alignment as often observed for short peptides[42,50,59] offers an additional explanation for the destabilization and distortion of the all-parallel PHF6 steric zipper motif due to the less favorable intra-sheet interactions.

The observed aggregation kinetics and obtained conformational ensembles are inherently dependent on the accuracy of the applied force field; therefore, a force field effect cannot be fully excluded. We chose the GROMOS96 43A1 force field for the current study for the following reasons: (1) previous simulations on model peptide folding revealed a balanced secondary structure propensity with no particular bias toward β-sheet formation.[79] (2) This force field is known to spontaneously sample amyloid-like β-strand structures,[49,50,80] consistent with experimental findings,[81] and to stabilize preformed pairs of cross-β sheets.[56,82] (3) The GROMOS96 force fields are primarily calibrated against thermodynamic experimental data and therefore achieve an adequate accuracy for the free energy of solvation for the neutral analogs of the nonpolar (Ala, Val, Leu, Ile) and aromatic (Tyr) amino acids,[83] which mainly constitute the hydrophobic peptides (VQIVYK, VEALYL, GVATVA) investigated here. In addition, we observed in the present study that spontaneous and reference simulations converge to similar conformational ensembles compatible with the relaxed crystallographic steric zipper structures.

The detailed structure of the amyloid state is not entirely determined by sequence[84] and thus hinders the straightforward identification of the sequence-specific driving forces.[9,21,22,27] The partitioning into multiple aggregation pathways, where the formed oligomers exhibit different extents of β-sheet secondary structure, is proposed to be kinetic in nature,[50,62,63] as experimentally determined for, for example, Aβ[37,45,85] and PrP oligomerization.[29] The promiscuous polymorphism found for amyloid aggregates[17,18,84] may also be explained by the thermodynamic selection of the most stable steric zipper motif in the nucleus.[27,86] The observed initial hydrophobic collapse of the peptides is an early event in the oligomerization and therefore not rate limiting for the nucleation, as deduced from simulations[51] and experiments.[26,27] In line with the present findings, this might suggest a mechanistic continuum for amyloidogenic aggregation with the principal characteristics of a condensation-ordering mechanism. The determining conformational conversion step occurs in the oligomeric state directed by specific structural predilections of sheet packing interactions and is largely independent of the monomer concentration.[26,27,85,87,88]

The overall registry and alignment of the individual strands were found to be of less importance in the initial association phase for the small steric zipper peptides. This lack of homogeneous strand patterning in the β-sheet structure of early oligomers may be interesting to probe by solid-state NMR experiments, in particular since a regular strand and sheet packing is probably one of the key requirements for the nucleation and selection of a particular zipper structure and therefore also dependent on the sequence. In this context, it is expected that the formation of ordered pre-nucleation intermediates with intra- and inter-sheet β-structure elements similar to the fibrillar state critically depends on the steric complementarity of the sheet interfaces in the oligomeric species, that is, the fine detail of matching and properly packed side chains.[12,59,77]

The observed structural ordering and conformational sampling in the ensemble of collapsed peptide aggregates were not found to correlate well with the solvation free energy. It has been suggested along these lines that the surface geometries at the molecular level influence solvent-mediated forces, such as dewetting transitions,[76,89] and thus might govern the formation of the high-energy nucleus. The oligomer topologies with sheet pairing angles very different from parallel β-sandwich structures might be indicative of off-pathway oligomers.[77] For these aggregates, the expected height of the nucleation barrier to a cross-β transition state is large. This would be consistent with the observed lag time in amyloid formation and the observation that prefibrillar oligomers can acquire significant β-sheet structure as determined by experiments.[68,69,85] Similar effects could origin from the found positional preference of certain side chains, inside or outside relative to the oligomer surface. With reference to a free-energy landscape point of view, it has been argued that small toxic oligomers must exist as local free-energy minima with significant barriers to amyloid formation, although single amino acid mutations that increase the rate of disease progression often increase the rate of amyloid formation.[28]

Finally, the oligomerization of larger peptides and full-length proteins might obey similar principles but usually needs to be accompanied by a substantial change or loss of intra-peptide interactions. Moreover, aggregation is facilitated by transient misfolding or unfolding events of particular aggregation-prone segments, which populate amyloid competent conformers, such as hairpin motifs or intra-molecular β-strand stacks in natively disordered peptides[41,45,54,90] or accessible self-complementary stretches on protein surfaces.[22,23,67] Recent findings furthermore suggest that while oligomer formation might be facilitated by small segments and their local properties, the eventual maturation toward fibrillar structure will involve a remodeling stage to further incorporate and accommodate residues into the β-strands.[66,91]

Conclusion

In the present work, the oligomerization of three different amyloidogenic peptide sequences was studied with atomistic MD simulations in explicit solvent environment. The initial stages of the aggregation process were characterized by two distinct phases, and the conformational conversions from random configurations to β-sheet oligomers resembled the previously proposed condensation-ordering mechanism. As a first step, the peptides assembled via various intermediates to partially ordered aggregates, thereby creating desolvated interfaces between the chains. In a second step, sequence-dependent conformational reorganizations toward β-sheet-rich structures took place in the collapsed oligomeric state. The kinetics and stability of aggregates with β-sandwich structure motifs were found to exhibit a profound dependence on the hydrophobic character, steric constraints and positioning of the side chains participating in the sheet-to-sheet interfaces. This study furthermore provides qualitative evidence that early steric zipper peptide oligomers feature similar self-complementary sheet packing characteristics as it is proposed for the fibrillar aggregation end-states.

The elucidation of the energetic and structural determinants of amyloidogenic aggregation poses an essential challenge to biophysical studies and still needs to rely on the study of simplified model systems. While structural models of short peptides become more readily available, a thorough understanding of the dynamical transitions between the canonically defined states along the amyloidogenic aggregation pathway is still missing. Here, we found that solvent-mediated interactions, such as the prominent reduction in solvation free energy, drive the primary peptide oligomerization steps. After the initial collapse, the onset of an ordering process was observed, mainly driven by extensive backbone hydrogen bond formation and β-sheet lamination.

The overall dominant solvent effects and the observed conformational changes in the metastable oligomer species hint at a causal relation, also crucial for other molecular recognition processes.[92] Therefore, the view is emphasized that biomolecular aggregation of peptides and proteins is governed by the fine chemical details of peptide–solvent interactions and water structure at various stages of the self-assembly process.[51,52,54,57,65,76,89]

Methods

MD simulations

All MD simulations were carried out using the GROMACS software package (version 4.0).[93] The Berendsen coupling algorithm[94] was applied to keep the pressure constant by coupling the system to a pressure bath of 1 bar ($\tau = 1$ ps). Velocity rescale[95] was applied for temperature coupling to a temperature bath of 310 K. Initial velocities were sampled from a Maxwellian distribution at 310 K. All protein bonds were constrained with the P-Lincs algorithm.[96] All the hydrogens were replaced by virtual interaction sites, and therefore, all internal vibrational degrees of freedom of the hydrogen atoms were removed.[97,98] An integration time step of 5 fs was used. Neighbor lists for nonbonded interactions were updated every 5 steps. For production runs, the GROMOS96 43A1[99] force field and the SPC water model[100] were used. Water molecules were constrained using SETTLE.[101] The short-ranged van der Waals and electrostatic interactions were cut off at 1.4 nm and 0.9 nm, respectively. All simulations were carried out using periodic boundary conditions and the Particle Mesh Ewald[102,103] method. The electrostatic interactions with Particle Mesh Ewald were calculated at every step with a grid spacing of 0.12 nm. The relative tolerance at the cutoff was set to 10^{-6}.

Simulation setup and procedure

An overview of the simulated peptide systems, simulation lengths and sampling intervals (subscript) is given

Table 3. Summary of performed simulations and initial conformations

Sequence	System ID	Starting peptide configuration	Trajectory number	Simulation time (μs) and sampling interval (ps, subscript)
VQIVYK (PHF6)	M10	Random, monomeric configuration	8	2.50_{50}, 2.15_{50}, 2.01_{50}, 2.00_{50}, 1.75_{400}, 0.65_{50}, 0.33_{50}, 0.31_{50}
	SH10	β-Sheet, parallel strands	2	0.70_{50}, 0.50_{50}
	SH5-5	β-Sandwich, parallel strands	1	1.00_{50}
VEALYL (IB12)	M10	Random, monomeric configuration	8	1.92_{50}, 1.80_{500}, 1.80_{500}, 1.48_{50}, 0.36_{50}, 0.32_{50}, 0.30_{50}, 0.30_{50}
	M10 (3.3 mM)		1	0.17_{50}
	M10 (8.3 mM)		2	0.39_{50}, 0.37_{50}
	M10 (83 mM)		3	1.00_{50}, 0.50_{50}, 0.50_{50}
	SH10	β-Sheet, antiparallel strands	2	1.00_{50}, 0.75_{50}
	SH5-5	β-Sandwich, antiparallel strands	1	1.03_{50}
GVATVA (AS51)	M10	Random, monomeric configuration	4	2.44_{500}, 2.00_{500}, 1.87_{500}, 1.77_{50}
GVATAV	M10	Random, monomeric configuration	1	1.25_{50}
GVAVTA	M10		1	1.25_{50}
GVTTVA	M10		1	1.25_{50}
GVATFA	M10		1	1.25_{50}
GAATAA	M10		1	1.25_{50}

in Table 3. The simulations were labeled according to the name (sequence) of the peptide and the starting configuration. The total simulation time of all trajectories adds up to 42 μs for this study. The individual simulations (M10) with 10 initially monomeric PHF6 (VQIVYK), IB12 (VEALYL), AS51 (GVATVA) and mutant peptides (GVATVA, GVAVTA, GVTTVA, GVATFA, GAATAA) were set up according to the protocol described in Ref. 50: 1000 distinct peptide conformations were pre-generated with CONCOORD.[104] From this structure ensemble, 10 peptide structures were chosen at random and placed in random position and orientation in a cubic box (1000 nm^3) to result in a concentration of 16.6 mM if not stated otherwise. This procedure was applied to ensure a fully monomeric configuration and to reduce the bias of similar initial starting structures for each of the conducted simulation runs. Additional M10 simulations with different peptide concentrations (3.3, 8.3 and 83 mM) were realized by adjusting the box size (5000, 2000, 200 nm^3).

For the initial configuration of the reference simulations, the atomic coordinates of the crystal structures Protein Data Bank ID: 2ON9 (PHF6: VQIVYK) and Protein Data Bank ID: 2OMQ (IB12: VEALYL)[15] were used to model the ordered steric zipper aggregate conformations (SH5-5: two 5-stranded β-sheets in a sandwich and SH10: single, 10-stranded β-sheet). The peptide concentration and box size matched the ones of the M10 aggregation setup. There were no crystalline coordinates available for the GVATVA peptide.[15] Subsequently, all systems were solvated with explicit water molecules. The protonation state of the peptides was according to the one in solution at a pH of 7 for all simulations of the PHF6, as well as the AS51 peptides. For the IB12 peptide simulations and according to the crystallization conditions at pH 2.5, the C-terminus and glutamate side chains were assumed to be protonated.[5,15] Counter-ions (Na^+, Cl^-) were added to yield an appropriate ionic strength (150 mM) and to neutralize the net system charge. The simulation systems comprised roughly 100,000 atoms. After the system preparation, an energy minimization using steepest descent was performed.

Analysis

From the individual simulation trajectories, samples were taken for analysis every 50, 400 or 500 ps (Table 3). After pooling all independent simulations, the resulting total number of collected configurations was for PHF6: M10, 203,351; SH5-5 and SH10, 45,699; for IB12: M10, 100,851; SH5-5 and SH10, 55,604; M10 (3.3 mM), 3401; M10 (8.3 mM), 15202; M10 (83 mM), 40,003; and for AS51: M10, 47,960.

Mapping the aggregation dynamics by projections onto collective coordinates

The used mapping procedure enables us to describe an initially high dimensional problem in a reduced (low dimensional) collective coordinate space. The reexpression in terms of collective coordinates (principal components) preserves the trends and variance of the full-dimensional data as well as highlights similarities and differences. In this formulation, one is able to map the multimeric aggregates found along the aggregation pathway and discriminate structures with different properties. In addition, the projection onto collective coordinates allows for a direct visualization of common sampling routes and probabilities.

Calculation of observables

A total of 25 different observables were calculated from the Cartesian coordinates of the configurations sampled from the individual trajectories. The observables encompass diverse metrics that adequately capture the topological (i), structural (ii) and energetic (iii) properties of amyloidogenic β-aggregation. Some of them are established descriptors for simulated peptide assembly and were used in previous studies.[53,55] A detailed description is provided below.

(i) Topological properties. To describe the peptide assembly state at any given time and to ensure that all assemblies with different multichain topologies and configuration types are discriminated, we used a specific notation. The aggregation state was expressed in terms of a connectivity length (*cl*).[55,105] The *cl* was defined to be the sum over the square roots of the determined individual aggregate sizes, with *k* as the total number of aggregates and *n* as the number of peptides in aggregate *x*. A normalized connectivity length (*CL*) was used since this can be read more intuitively, with *N*=10 for all simulations, which is also the largest possible aggregate size.

$$cl = \sum_{x=1}^{k} \sqrt{n_x} \quad CL = \left(1 - \frac{cl - \sqrt{N}}{N - \sqrt{N}}\right)$$

The CL values for the aggregation configuration types range from 0 for fully monomeric to 1 for the state where all peptides were found in the same aggregate. A hierarchical categorization of the peptide aggregate sizes was pursued to monitor the assembly state of the peptides. General peptide assemblies or aggregates were defined as follows: pairwise contact analysis is used to identify the individual aggregates. Peptides that share an inter-chain residue contact are then counted to be within the same aggregate. For any two residues *i*, *j*, an inter-chain contact is considered formed if any heavy atom of residue *i* is within a cutoff of 0.45 nm from any heavy atom of residue *j*. Assemblies of peptides aggregated into intermolecular β-sheets were defined as follows: two peptides that share two consecutive inter-chain β-sheet contacts as defined by the DSSP definition[106] are counted within the same aggregate. The identified β-sheet aggregates are a subpopulation of the above considered general aggregates. These two metrics comprise the first two observables that are considered. (1) Assembly state of general aggregates according to van der Waals contacts (CLg); (2) assembly state of β-sheet aggregates according to β-sheet contacts (CLs). Instructive descriptors for different orientational properties of the system are the polar (P1) and the nematic (P2) order parameters, commonly used to analyze properties of anisotropic fluids such as liquid crystals. Several studies highlight the useful and complementary information of these order parameters in the context of peptide aggregation.[49,53] A suitable molecular vector ($\hat{\mathbf{z}}_i$) was defined, here the unit vector linking the C^α

atoms of residues 2 and 5 of each peptide. The choice was based on the observation that the terminal residues were not found to have high β-sheet propensities. The polarity of the system is described by P1, distinguishing between parallel, antiparallel or mixed strand (molecular vector) orientations.

$$\overline{P}_1 = \frac{1}{N}\sum_{i=1}^{N}\frac{3}{2}\hat{\mathbf{z}}_i\cdot\hat{\mathbf{d}} \quad \overline{P}_2 = \frac{1}{N}\sum_{i=1}^{N}\frac{3}{2}\left(\hat{\mathbf{z}}_i\cdot\hat{\mathbf{d}}\right)^2-\frac{1}{2}$$

The orientational or nematic order parameter P2 of the system discriminates between ordered and disordered peptide conformations: the extent of alignment and relative orientation of the individual peptides described by their molecular vector. The order parameter P1 and P2 were calculated using the Wordom program package.[107] (3) Polar order (P1); (4) nematic order (P2).

(ii) Structural properties. A heavy-atom contact analysis was used to calculate the total number of inter-strand contacts present between either polar or nonpolar peptide atoms. An atom was considered nonpolar (or hydrophobic) if the absolute value of the partial charge was below 0.2. Otherwise, this atom was defined as polar. A contact was considered formed if any heavy-atom was within a cutoff of 0.45 nm from any other heavy-atom of the same type (polar/polar) of any other strand. (5) Total number of polar atom contacts (Npa); (6) total number of nonpolar atom contacts (Nna). (7) Solvent-inaccessible molecular surface (SiMS). The total SiMS area was calculated using a solvent probe radius of 1.0 nm. Similar to the above described atom contact analysis, the sum of all peptide heavy-atom of either polar or nonpolar type in contact with solvent molecules was calculated. The cutoff criterion was chosen to be 0.35 nm, reflecting the approximate contact distance of atoms on the peptide surface to the water molecules of the first solvation shell. (8) Total number of polar atom contacts to the solvent (Npas), that is, the number of water molecules in contact with polar peptide atoms; (9) total number of nonpolar atom contacts to the solvent (Nnas). For an assessment of secondary structure type and content, the DSSP definition introduced by Kabsch and Sander[106] was used. The populations of various secondary structure elements were calculated (random coil, β-sheet, β-bridge, bend, turn) for the individual simulations as a function of simulation time. (10) Random-coil content (Cc); (11) β-sheet content (Bsc); (12) β-bridge content (Bbc); (13) bend content (Bc); (14) turn content (Tc). (15) Average peptide chain extension (Ext). The peptide extension was calculated by averaging the cumulative intra-peptide distances between the N- and C-terminal C^{α} atoms of each of the 10 chains.

(iii) Energetic properties. The sum of the potential energy terms of different groups of atoms was calculated to identify their contributions to the assembly process. The short- and long-range van der Waals, as well as short-range electrostatic interaction energies of the force field, were considered separately. In total, 10 potential energy terms were analyzed. (16) Sum of coulombic inter-main chain interactions (Inter-mc-Ecoul); (17) sum of van der Waals inter-main chain interactions (Inter-mc-EvdW); (18) sum of coulombic inter-side chain interactions (Inter-sc-Ecoul); (19) sum of van der Waals inter-side chain interactions (Inter-sc-EvdW); (20) sum of coulombic intra-main chain interactions (Intra-mc-Ecoul); (21) sum of van der Waals intra-main chain interactions (Intra-mc-EvdW); (22) sum of coulombic main-chain interactions with the solvent (Sol-mc-Ecoul); (23) sum of van der Waals main-chain interactions with the solvent (Sol-mc-EvdW); (24) sum of coulombic side chain interactions with the solvent (Sol-sc-Ecoul); (25) sum of van der Waals side chain interactions with the solvent (Sol-sc-EvdW).

Dimensionality reduction and projection of the observable data

The obtained observable data sets were concatenated (see below for details) and normalized (*z*-scored) prior to further analysis. Each set of a given observable (*x*) was normalized separately, such that the mean (μ) was set to 0 and the standard deviation (σ), to unity.

$$z = \frac{x-\mu}{\sigma}$$

This preprocessing step was carried out to circumvent issues when comparing observables with different units and dimensions, which might otherwise affect the outcome of the subsequent PCA. The PCA involved the diagonalization of the covariance matrix of the data set composed of the 25 described observables. The dimensionality reduction of the observable data was done by projecting the sets onto its respective principal components (the first three eigenvectors of the covariance matrix). PCA was performed for the 25-dimensional observable data sets of the PHF6, IB12 and AS51 peptide simulations separately if not stated otherwise, using the statistical software package R.[108] The subspace of the first three eigenvectors typically accounts for more than 55% of the variance (see Fig. S8).

Estimating the extent of overlap between mapped ensembles and identification of conformations from the overlap region

To examine whether the M10 and SH5-5 as well as SH10 simulations access the same configurational states as described by the various observables, we chose a stepwise approach. Firstly, the extent of overlap in the multi-dimensional observable space was determined. Secondly, the characteristics of the overlapping states were analyzed.

Identification of the overlap region

The observable data sets obtained from M10, SH5-5 and SH10 simulations were concatenated for the PHF6 and IB12 systems, respectively. On the concatenated data sets, PCA was performed, and subsequently, a projection onto the first three eigenvectors of the covariance matrix was carried out.

Determining the extent of overlap between the point clouds of spontaneous (M10) and reference (SH5-5 and SH10) simulations in the multi-dimensional observable configuration space was approached as a classification problem, whose solution also allowed the direct identification of the structures from the ensemble overlap regions. For each point in the observable data set, k-NNs were identified employing the k-NN algorithm[109,110] implemented in R.[108] Hereby the number of points for the classification is chosen such that N (number of neighbors) equals the square root of the number of samples in the data set (N_{PHF6}: 499 and N_{IB12}: 395). Each point was assigned a label of that ensemble (spontaneous or reference), the more members of which there were among its k-NNs. All those points of the data set that were assigned to the ensemble different from which they originally belonged to were considered to be in the overlap region of the two ensemble of points. For every point identified in this way, the closest point from the other ensemble not yet in the list of structures from the overlap region was added to the list. With this approach, 8194 nearest-neighbor conformations from the spontaneous and reference point clouds of PHF6 simulations and 10,626 for IB12 were found.

PCA and clustering of conformations from the overlap region

The identified data points from the overlapping regions of the ensembles were mapped back to their Cartesian coordinates. Since all the points correspond to decameric oligomers, a relabeling procedure described in Ref. 50 was performed in order to reduce permutation redundancy of the aggregates. The redundancy arises due to permutation of the labels (chain identifiers) in structures with equivalent conformations. The reference structure for relabeling was chosen to be the one closest to the center of the overlapping region in the observable space. In every step of the relabeling procedure 10! label permutations were tested. For each permutation, a fitting of the main chain and C^{β} atoms onto the reference structure was performed. The labeling that yielded the smallest RMSD to the reference structure was assigned to the aggregate.

Subsequently, a conformational clustering was carried out. PCA was performed using the Cartesian coordinates of the relabeled and superimposed structures, after a least-squares fit onto the main chain and C^{β} atoms. Projections onto the first 100 principal components were used for the *k*-means clustering, following the results of Ding and He showing that principal components are the solutions for the *k*-means clustering problem.[111] *k*-Means clustering was performed using the Hartigan–Wong algorithm[112] as implemented in R.[108] The Krzanowski–Lai criterion[113] was used to choose the number of conformational clusters, and the global *k*-means algorithm[114] was used to determine the respective cluster centers.

FMA analysis with partial least squares

Calculation of sampling density as external variable for the aggregation progress

For the density estimate in the full 25-dimensional observable configuration space, again the k-NNs method was used. For each data point, the distance to k-NNs was calculated.[115] Distances for all k-NN were summed up and inverted to yield a measure for the density for each data point. Densities for the PHF6, IB12 and AS51 M10 simulations were calculated from concatenated and commonly *z*-scored observable data set. The number of nearest-neighbors to be considered for the k-NN distance calculations was selected to be 593, which was the square root of the total number of data points in the concatenated set of PHF6, IB12 and AS51 configurations.

FMA model building and cross-validation

In order to elucidate the main forces driving the aggregation process, we employed the FMA[71] using the partial least squares algorithm.[116,117]

FMA is a technique to construct a linear multiple regression model, which maximizes the correlation between the observables and an external variable while at the same time maximizing the variances in both the variable and the observables. Here, 28 independent observables grouped into descriptors (1) and energetic properties (2) were used to predict the sampling density in the configuration space. In addition to the 25 described observables, the solvation free energy as defined by Eisenberg and McLachlan[70] (dGSol) and the sum of coulombic as well as van der Waals main chain-to-side chain interactions (Inter-mc-sc-Ecoul and Inter-mc-sc-EvdW) were used.

For each peptide system, PHF6, IB12 and AS51, a separate FMA model was built for the association and the decameric phase. In order to probe the formation of ordered aggregates in the decameric phase, we built an additional model where the sampling density was weighted by the factor exp(P2·Bsc) prior to the FMA. A cross-validation procedure was applied to estimate the predictive power of each FMA model. Every data set was divided into eight equal parts. One part at a time was left aside for the cross-validation, while an FMA model was built on the rest of the data. This process was repeated such that each part of the data set was used for the cross-validation once. The predicted FMA models were concatenated, yielding a completely cross-validated data set. The final FMA vectors, as well as the training sets in Fig. 8, were built from the full data sets, respectively. To estimate the quality of the built FMA models, we calculated the correlation coefficients for both the training and testing parts for the FMA models based on the descriptor observables. In order to avoid overfitting, we selected the number of components for the FMA model construction such that, with a further increase in the number of components (8), only minor changes in the correlation coefficient for the testing data set were observed.

Regression coefficients of the ensemble-weighted FMA model

The influence of the observables on the changes in sampling density was analyzed using an ensemble-weighted FMA[71] model, where the contribution of each observable parameter was weighted according to its variance in the data set at hand. This approach allowed us to investigate which of the observed aggregate features positively or negatively contributed to the sampling density in the observable configuration space, that is, the peptide aggregation.

Acknowledgements

We thank Ulrike Gerischer and Hadas Leonov for carefully reading the manuscript. This work was funded by the Graduate School Spectroscopy and Dynamics of Molecular Coils and Aggregates (grant no. GRK 782) and the Max-Planck Society. The

International Max-Planck Research School for Physics of Biological and Complex Systems is acknowledged for financial support to V.G.

Supplementary Data

Supplementary materials related to this article can be found online at doi:10.1016/j.jmb.2012.02.004

References

1. Sipe, J. D. (1992). Amyloidosis. *Annu. Rev. Biochem.* **61**, 947–975.
2. Collinge, J. (2005). Molecular neurology of prion disease. *J. Neurol., Neurosurg. Psychiatry*, **76**, 906–919.
3. Vilar, M., Chou, H. T., Lührs, T. S. K., Maji, D., Riek-Loher, R., Verel, G. *et al.* (2008). The fold of alpha-synuclein fibrils. *Proc. Natl Acad. Sci. USA*, **105**, 8637–8642.
4. Greenwald, J. & Riek, R. (2010). Biology of amyloid: structure, function, and regulation. *Structure*, **18**, 1244.
5. Ivanova, M. I., Thompson, M. J. & Eisenberg, D. (2006). A systematic screen of beta2 -microglobulin and insulin for amyloid-like segments. *Proc. Natl Acad. Sci. USA*, **103**, 4079–4082 .
6. Nelson, R., Sawaya, M. R., Balbirnie, M., Madsen, A. O., Riekel, C., Grothe, R. & Eisenberg, D. (2005). Structure of the cross-beta spine of amyloid-like fibrils. *Nature*, **435**, 773–778.
7. Murphy, R. M. (2002). Peptide aggregation in neurodegenerative disease. *Annu. Rev. Biomed. Eng.* **4**, 155–174.
8. Pastor, M. T., Kümmerer, N., Schubert, V., Esteras-Chopo, A., Dotti, C. G., de la Paz, M. & Serrano, L. (2008). Amyloid toxicity is independent of polypeptide sequence, length and chirality. *J. Mol. Biol.* **375**, 695–707.
9. Fändrich, M. & Dobson, C. M. (2002). The behaviour of polyamino acids reveals an inverse side chain effect in amyloid structure formation. *EMBO J.* **21**, 5682–5690.
10. Hartl, F. U. & Hayer-Hartl, M. (2009). Converging concepts of protein folding *in vitro* and *in vivo*. *Nat. Struct. Mol. Biol.* **16**, 574–581.
11. Gian Gaetano Tartaglia, A. P. P., Campioni, S., Dobson, C. M., Chiti, F. & Vendruscolo, M. (2008). Prediction of aggregation-prone regions in structured proteins. *J. Mol. Biol.* **380**, 425–436.
12. Ruschak, A. & Miranker, A. (2009). The role of prefibrillar structures in the assembly of a peptide amyloid. *J. Mol. Biol.* **393**, 214–226.
13. Nelson, R. & Eisenberg, D. (2006). Recent atomic models of amyloid fibril structure. *Curr. Opin. Struct. Biol.* **16**, 260–265.
14. Sunde, M. & Blake, C. (1998). From the globular to the fibrous state: protein structure and structural conversion in amyloid formation. *Q. Rev. Biophys.* **31**, 1–39.
15. Sawaya, M. R., Sambashivan, S., Nelson, R., Ivanova, M. I., Sievers, S. A., Apostol, M. I. *et al.* (2007). Atomic structures of amyloid cross-beta spines reveal varied steric zippers. *Nature*, **447**, 453–457.
16. Wiltzius, J. J. W., Landau, M., Nelson, R., Sawaya, M. R., Apostol, M. I., Goldschmidt, L. *et al.* (2009). Molecular mechanisms for protein-encoded inheritance. *Nat. Struct. Mol. Biol.* **16**, 973–978.
17. van der Wel, P. C., Lewandowski, J. R. & Griffin, R. G. (2010). Structural characterization of gnnqqny amyloid fibrils by magic angle spinning NMR. *Biochemistry*, **49**, 9457–9469.
18. Marshall, K. & Serpell, L. (2010). Fibres, crystals and polymorphism: the structural promiscuity of amyloidogenic peptides.. *Soft Matter*, **6**, 2110–2114.
19. Uversky, V. N. & Fink, A. L. (2004). Conformational constraints for amyloid fibrillation: the importance of being unfolded. *Biochim. Biophys.* **1698**, 131–153.
20. Dobson, C. M. (2003). Protein folding and misfolding. *Nature*, **426**, 884–890.
21. Pastor, M. T., de la Paz, M. L., Lacroix, E., Serrano, L. & Perez-Paya, E. (2002). Combinatorial approaches: a new tool to search for highly structured beta-hairpin peptides. *Proc. Natl Acad. Sci. USA*, **99**, 614–619.
22. Goldschmidt, L., Teng, P. K., Riek, R. & Eisenberg, D. (2010). Identifying the amylome, proteins capable of forming amyloid-like fibrils. *Proc. Natl Acad. Sci. USA*, **107**, 3487–3492.
23. Ventura, S., Zurdo, J., Narayanan, S., Parreno, M., Mangues, R., Reif, B. *et al.* (2004). Short amino acid stretches can mediate amyloid formation in globular proteins: the src homology 3 (sh3) case. *Proc. Natl Acad. Sci. USA*, **101**, 7258–7263.
24. von Bergen, M., Friedhoff, P., Biernat, J., Heberle, J., Mandelkow, E. M. & Mandelkow, E. (2000). Assembly of tau protein into Alzheimer paired helical filaments depends on a local sequence motif (306vqivyk311) forming beta structure. *Proc. Natl Acad. Sci. USA*, **97**, 5129–5134.
25. Lomakin, A., Chung, D. S., Benedek, G. B., Kirschner, D. A. & Teplow, D. B. (1996). On the nucleation and growth of amyloid beta-protein fibrils: detection of nuclei and quantitation of rate constants. *Proc. Natl Acad. Sci. USA*, **93**, 1125–1129.
26. Serio, T. R., Cashikar, A. G., Kowal, A. S., Sawicki, G. J., Moslehi, J. J., Serpell, L. *et al.* (2000). Nucleated conformational conversion and the replication of conformational information by a prion determinant. *Science*, **289**, 1317–1321.
27. Schmit, J. D., Ghosh, K. & Dill, K. (2011). Why do amyloid molecules assemble into oligomers and fibrils? *Biophys. J.* **100**, 450–458.
28. He, X., Giurleo, J. & Talaga, D. (2010). Role of small oligomers on the amylodogenic aggregation free-energy landscape. *J. Mol. Biol.* **395**, 134–154.
29. Cho, K., Huang, Y., Yu, S., Yin, S., Plomp, M., Qiu, S. *et al.* (2011). A multistage pathway for human prion protein aggregation *in vitro*: from multimeric seeds to beta-oligomers and nonfibrillar structures. *J. Am. Chem. Soc.* **133**, 8586–8593.
30. Bernstein, S. L., Dupuis, N. F., Lazo, N. D., Wyttenbach, T., Condron, M. M., Bitan, G. *et al.* (2009). Amyloid-beta protein oligomerization and the importance of tetramers and dodecamers in the aetiology of Alzheimer's disease. *Nat. Chem.* **1**, 326–331.

Please cite this article as: Matthes, D. *et al.*, Driving Forces and Structural Determinants of Steric Zipper Peptide Oligomer Formation Elucidated by Atomistic Simulations, *J. Mol. Biol.* (2012), doi:10.1016/j.jmb.2012.02.004

31. Ahmad, A., Uversky, V. N., Hong, D. & Fink, A. L. (2005). Early events in the fibrillation of monomeric insulin. *J. Biol. Chem.* **280**, 42669–42675.
32. Frare, E., Mossuto, M. F., de Laureto, P. P., Tolin, S., Menzer, L., Dumoulin, M. *et al.* (2009). Characterization of oligomeric species on the aggregation pathway of human lysozyme. *J. Mol. Biol.* **387**, 17–27.
33. Glabe, C. G. (2008). Structural classification of toxic amyloid oligomers. *J. Biol. Chem.* **283**, 29639–29643.
34. Kaylor, J., Bodner, N., Edridge, S., Yamin, G., Hong, D. & Fink, A. (2005). Characterization of oligomeric intermediates in alpha-synuclein fibrillation: FRET studies of y125w/y133f/y136f alpha-synuclein. *J. Mol. Biol.* **353**, 357–372.
35. Kodali, R. & Wetzel, R. (2007). Polymorphism in the intermediates and products of amyloid assembly. *Curr. Opin. Struct. Biol.* **17**, 48–57.
36. Bucciantini, M., Giannoni, E., Chiti, F., Baroni, F., Formigli, L., Zurdo, J. *et al.* (2002). Inherent toxicity of aggregates implies a common mechanism for protein misfolding diseases. *Nature*, **416**, 507–511.
37. Lasagna-Reeves, C., Glabe, C. & Kayed, R. (2011). Amyloid-beta annular protofibrils evade fibrillar fate in Alzheimer disease brain. *J. Biol. Chem.* **286**, 22122–22130.
38. Smith, D., Radford, S. & Ashcroft, A. (2010). Elongated oligomers in beta 2-microglobulin amyloid assembly revealed by ion mobility spectrometry mass spectrometry. *Proc. Natl Acad. Sci. USA*, **107**, 6794–6798.
39. Liang, Y., Lynn, D. G. & Berland, K. M. (2010). Direct observation of nucleation and growth in amyloid self-assembly. *J. Am. Chem. Soc.* **132**, 6306–6308.
40. Walsh, P., Neudecker, P. & Sharpe, S. (2010). Structural properties and dynamic behavior of nonfibrillar oligomers formed by prp(106–126). *J. Am. Chem. Soc.* **132**, 7684–7695.
41. Grabenauer, M., Wu, C., Soto, P., Shea, J. E. & Bowers, M. T. (2010). Oligomers of the prion protein fragment 106–126 are likely assembled from beta-hairpins in solution, and methionine oxidation inhibits assembly without altering the peptides monomeric conformation. *J. Am. Chem. Soc.* **132**, 532–539.
42. Lopez de la Paz, M., de Mori, G., Serrano, L. & Colombo, G. (2005). Sequence dependence of amyloid fibril formation: insights from molecular dynamics simulations. *J. Mol. Biol.* **349**, 583–596.
43. Chiti, F. & Dobson, C. M. (2006). Protein misfolding, functional amyloid, and human disease. *Annu. Rev. Biochem.* **75**, 333–366.
44. Ahmed, M., Davis, J., A.D., Sato, T., Ahuja, S., Aimoto, S., Elliott, J. *et al.* (2010). Structural conversion of neurotoxic amyloid-beta 1–42 oligomers to fibrils. *Nat. Struct. Mol. Biol.* **17**, 561–567.
45. Sandberg, A., Luheshi, L., So¨, S., llvander, T., Pereira de Barros, B., Macao, T. *et al.* (2010). Ha¨rd, Stabilization of neurotoxic Alzheimer amyloid-beta oligomers by protein engineering. *Proc. Natl Acad. Sci. USA*, **107**, 15595–15600.
46. Chimon, S., Shaibat, M., Jones, C., Calero, D., Aizezi, B. & Ishii, Y. (2007). Evidence of fibril-like bold beta-sheet structures in a neurotoxic amyloid intermediate of Alzheimer's bold beta-amyloid. *Nat. Struct. Mol. Biol.* **14**, 1157–1164.
47. Lashuel, H. & Lansbury, P. T. (2006). Are amyloid diseases caused by protein aggregates that mimic bacterial pore-forming toxins? *Q. Rev. Biophys.* **39**, 167–201.
48. Strodel, B., Whittleston, C. & Wales, D. (2007). Thermodynamics and kinetics of aggregation for the gnnqqny peptide. *J. Am. Chem. Soc.* **129**, 16005–16014.
49. Nguyen, P. H., Li, M. S., Stock, G., Straub, J. E. & Thirumalai, D. (2007). Monomer adds to preformed structured oligomers of abeta-peptides by a two-stage dock–lock mechanism. *Proc. Natl Acad. Sci. USA*, **104**, 111–116.
50. Matthes, D., Gapsys, V., Daebel, V. & de Groot, B. L. (2011). Mapping the conformational dynamics and pathways of spontaneous steric zipper peptide oligomerization. *PLoS ONE*, **6**, e19129.
51. Klimov, D. K. & Thirumalai, D. (2003). Dissecting the assembly of abeta16–22 amyloid peptides into antiparallel beta sheets. *Structure*, **11**, 295–307.
52. Wu, C., Lei, H. & Duan, Y. (2004). Formation of partially ordered oligomers of amyloidogenic hexapeptide (nfgail) in aqueous solution observed in molecular dynamics simulations. *Biophys. J.* **87**, 3000–3009.
53. Cecchini, M., Rao, F., Seeber, M. & Caflisch, A. (2004). Replica exchange molecular dynamics simulations of amyloid peptide aggregation. *J. Chem. Phys.* 10748–10756.
54. Straub, J. E. & Thirumalai, D. (2010). Principles governing oligomer formation in amyloidogenic peptides. *Curr. Opin. Struct. Biol.* **20**, 1–9.
55. Lu, Y., Derreumaux, P., Guo, Z., Mousseau, N. & Wei, G. (2009). Thermodynamics and dynamics of amyloid peptide oligomerization are sequence dependent. *Proteins: Struct., Funct., Bioinformatics*, **75**, 954–963.
56. Wallace, J. & Shen, J. (2010). Probing the strand orientation and registry alignment in the propagation of amyloid fibrils. *Biochemistry*, **49**, 5290–5298.
57. Reddy, G., Straub, J. E. & Thirumalai, D. (2009). Dynamics of locking of peptides onto growing amyloid fibrils. *Proc. Natl Acad. Sci. USA*, **106**, 11948–11953.
58. Gsponer, J. & Vendruscolo, M. (2006). Theoretical approaches to protein aggregation. *Protein Pept. Lett.* **13**, 287–293.
59. Nasica-Labouze, J., Meli, M., Derreumaux, P., Colombo, G. & Mousseau, N. (2011). A multiscale approach to characterize the early aggregation steps of the amyloid-forming peptide gnnqqny from the yeast prion sup-35. *PLoS Comput. Biol.* **7**, e1002051.
60. Santini, S., Wei, G., Mousseau, N. & Derreumaux, P. (2004). Pathway complexity of Alzheimers beta-amyloid abeta16–22 peptide assembly. *Structure*, **12**, 1245–1255.
61. Petty, S. A. & Decatur, S. M. (2005). Intersheet rearrangement of polypeptides during nucleation of beta-sheet aggregates. *Proc. Natl Acad. Sci. USA*, **102**, 14272–14277.
62. Hwang, W., Zhang, S., Kamm, R. D. & Karplus, M. (2004). Kinetic control of dimer structure formation in amyloid fibrillogenesis. *Proc. Natl Acad. Sci. USA*, **101**, 12916–12921.

Please cite this article as: Matthes, D. *et al.*, Driving Forces and Structural Determinants of Steric Zipper Peptide Oligomer Formation Elucidated by Atomistic Simulations, *J. Mol. Biol.* (2012), doi:10.1016/j.jmb.2012.02.004

63. Pellarin, R., Schuetz, P., Guarnera, E. & Caflisch, A. (2010). Amyloid fibril polymorphism is under kinetic control. *J. Am. Chem. Soc.* **132**, 14960–14970.
64. Klimov, D., Straub, J. & Thirumalai, D. (2004). Aqueous urea solution destabilizes abeta16–22 oligomers. *Proc. Natl Acad. Sci. USA*, **101**, 14760–14765.
65. Reddy, G., Straub, J. & Thirumalai, D. (2010). Dry amyloid fibril assembly in a yeast prion peptide is mediated by long-lived structures containing water wires. *Proc. Natl Acad. Sci. USA*, **107**, 21459–21464.
66. Ivanova, M. I., Sievers, S. A., Sawaya, M. R., Wall, J. S. & Eisenberg, D. (2009). Molecular basis for insulin fibril assembly. *Proc. Natl Acad. Sci. USA*, **106**, 18990–18995.
67. Teng, P. K. & Eisenberg, D. (2009). Short protein segments can drive a non-fibrillizing protein into the amyloid state. *Protein Eng. Des. Sel.* **22**, 531–536.
68. Quijano, F., Morrow, D., Wise, B., Brancia, F. & Goux, W. (2006). Prediction of nucleating sequences from amyloidogenic propensities of tau-related peptides. *Biochemistry*, **45**, 4638–4652.
69. Inoue, M., Hirata, A., Tainaka, K., Morii, T. & Konno, T. (2008). Charge-pairing mechanism of phosphorylation effect upon amyloid fibrillation of human tau core peptide. *Biochemistry*, **47**, 11847–11857.
70. Eisenberg, D. & McLachlan, A. D. (1986). Solvation energy in protein folding and binding. *Nature*, **319**, 199–203.
71. Hub, J. & de Groot, B. (2009). Detection of functional modes in protein dynamics. *PLoS Comput. Biol.* **5**, e1000480.
72. Lopez de la Paz, M. & Serrano, L. (2004). Sequence determinants of amyloid fibril formation. *Proc. Natl Acad. Sci. USA*, **101**, 87–92.
73. Gfeller, D., De Los Rios, P., Caflisch, A. & Rao, F. (2007). Complex network analysis of free-energy landscape. *Proc. Natl Acad. Sci. USA*, **104**, 1817–1822.
74. Stamati, H., Clementi, C. & Kavraki, L. (2010). Application of nonlinear dimentionality reduction to characterize the conformational landscape of small peptides. *Proteins: Struct., Funct., Bioinformatics*, **78**, 223–235.
75. Vitalis, A., Lyle, N. & Pappu, R. (2009). Thermodynamics of beta-sheet formation in polyglutamine. *Biophys. J.* **97**, 303–311.
76. Krone, M., Hua, L., Soto, P., Zhou, R., Berne, B. & Shea, J. (2008). Role of water in mediating the assembly of Alzheimer amyloid-beta abeta16–22 protofilaments. *J. Am. Chem. Soc.* **130**, 11066–11072.
77. Liu, M., Sawaya, P., Cheng, J., Zheng, J. & Nowick, D. (2011). Eisenberg, characteristics of amyloid-related oligomers revealed by crystal structures of macrocyclic beta-sheet mimics. *J. Am. Chem. Soc.* **133**, 6736–6744.
78. Vitagliano, L., Stanzione, F., De Simone, A. & Esposito, L. (2009). Dynamics and stability of amyloid-like steric zipper assemblies with hydrophobic dry interfaces. *Biopolymers*, **91**, 1161–1171.
79. Matthes, D. & de Groot, B. (2009). Secondary structure propensities in peptide folding simulations: a systematic comparison of molecular mechanics interaction schemes. *Biophys. J.* **97**, 599–608.
80. Nguyen, P., Lid, M. & Derreumaux, P. (2011). Effects of all-atom force fields on amyloid oligomerization: replica exchange molecular dynamics simulations of the ab16–22 dimer and trimer. *Phys. Chem. Chem. Phys.* **13**, 9778–9788.
81. Sgourakis, N., Yan, Y., McCallum, S., Wang, C. & Garcia, A. (2007). The Alzheimer's peptides abeta40 and 42 adopt distinct conformations in water: a combined MD/NMR study. *J. Mol. Biol.* **368**, 1448–1457.
82. Periole, X., Rampioni, A., Vendruscolo, M. & Mark, A. (2009). Factors that affect the degree of twist in beta-sheet structures: a molecular dynamics simulation study of a cross-beta filament of the gnnqqny peptide.. *J. Phys. Chem. B*, **113**, 1728–1737.
83. Villa, A. & Mark, A. E. (2002). Calculation of the free energy of solvation for neutral analogs of amino acid side chains. *J. Comput. Chem.* **23**, 548–553.
84. Petkova, A. T., Leapman, R. D., Guo, Z., Yau, W. M., Mattson, M. P. & Tycko, R. (2005). Self-propagating, molecular-level polymorphism in Alzheimer's beta-amyloid fibrils. *Science*, **307**, 262–265.
85. Lee, J., Culyba, E. K., Powers, E. T. & Kelly, J. W. (2011). Amyloid-beta forms fibrils by nucleated conformational conversion of oligomers. *Nat. Chem. Biol.* **7**, 602–609.
86. Park, J., Kahn, B. & Hwang, W. (2009). Thermodynamic selection of steric zipper patterns in the amyloid cross-beta spine. *PLoS Comput. Biol.* **5**, e1000492.
87. Bleiholder, C., Dupuis, N., Wyttenbach, T. & Bowers, M. (2011). Ion mobility–mass spectrometry reveals a conformational conversion from random assembly to beta-sheet in amyloid fibril formation. *Nat. Chem.* **3**, 172–177.
88. Auer, S., Meersman, F., Dobson, C. & Vendruscolo, M. (2008). A generic mechanism of emergence of amyloid protofilaments from disordered oligomeric aggregates. *PLoS Comput. Biol.* **11**, e1000222.
89. Ravikumar, K. & Hwang, W. (2011). Role of hydration force in the self-assembly of collagens and amyloid steric zipper filaments. *J. Am. Chem. Soc.* **133**, 11766–11773.
90. Suk, J. E., Lokappa, S. B. & Ulmer, T. S. (2010). The clustering and spatial arrangement of beta-sheet sequence, but not order, govern alpha-synuclein fibrillogenesis. *Biochemistry*, **49**, 1533–1540.
91. Scheidt, H., Morgado, I., Rothemund, S., Huster, D. & Fändrich, M. (2011). Solid-state NMR spectroscopic investigation of abeta protofibrils: implication of a beta-sheet remodeling upon maturation into terminal amyloid fibrils. *Angew. Chem., Int. Ed.* **50**, 2837–2840.
92. Hummer, G. (2010). Under water's influence. *Nat. Chem.* **2**, 906–907.
93. Hess, B., Kutzner, C., Van Der Spoel, D. & Lindahl, E. (2008). GROMACS 4.0: algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.* **4**, 435–447.
94. Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A. & Haak, J. R. (1984). Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **81**, 3684–3690.
95. Bussi, G., Donadio, D. & Parrinello, M. (2007). Canonical sampling through velocity rescaling. *J. Chem. Phys.* **126**, 014101.

96. Hess, B. (2008). P-lincs: a parallel linear constraint solver for molecular simulation. *J. Chem. Theory Comput.* **4**, 116–122.
97. Feenstra, K. A., Hess, B. & Berendsen, H. J. C. (1999). Improving efficiency of large time-scale molecular dynamics simulations of hydrogen-rich systems. *J. Comput. Chem.* **20**, 786–798.
98. van der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A. E. & Berendsen, H. J. (2005). GROMACS: fast, flexible, and free. *J. Comput. Chem.* **26**, 1701–1718.
99. van Gunsteren, W. F., Billeter, S. R., Eising, A. A., Hünenberger, P. H., Krüger, P., Mark, A. E. *et al.* (1996). *Biomolecular Simulation: The GROMOS96 Manual and User Guide.* Hochschulverlag AG an der ETH Zürich, Zürich, Switzerland.
100. Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F. & Hermans, J. (1981). Interaction models for water in relation to protein hydration. In *Intermolecular Forces*, pp. 331–342, D. Reidel Publishing Company, Dordrecht, Holland.
101. Miyamoto, S. & Kollman, P. A. (1992). Settle: an analytical version of the shake and rattle algorithm for rigid water models. *J. Comput. Chem.* **13**, 952–962.
102. Darden, T., York, D. & Pedersen, L. (1993). Particle mesh Ewald: an *n*-log(*n*) method for Ewald sums in large systems. *J. Chem. Phys.* **98**, 10089–10092.
103. Essmann, U., Perera, L., Berkowitz, M. L., Darden, T., Lee, H. & Pedersen, L. G. (1995). A smooth particle mesh Ewald method. *J. Chem. Phys.* **103**, 8577–8593.
104. de Groot, B., van Aalten, D., Scheek, R., Amadei, A., Vriend, G. & Berendsen, H. (1997). Prediction of protein conformational freedom from distance constraints. *Proteins: Struct., Funct., Bioinformatics*, **29**, 140–251.
105. Vekhter, B. & Berry, R. S. (1999). Modeling self-assembling of proteins: assembled structures, relaxation dynamics, and phase coexistence. *J. Chem. Phys.* **110**, 2195–2201.
106. Kabsch, W. & Sander, C. (1983). Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, **12**, 2577–2637.
107. Seeber, M., Cecchini, M., Rao, F., Settanni, G. & Caflisch, A. (2007). Wordom: a program for efficient analysis of molecular dynamics simulations. *Bioinformatics*, **23**, 2625–2627.
108. R Development Core Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0 (2010).
109. Ripley, B. (1996). *Pattern Recognition and Neural Networks.* Cambridge University Press, Cambridge, UK.
110. Venables, W. & Ripley, B. (2002). *Modern Applied Statistics with S.* Springer Verlag, New York, NY.
111. Ding, C. & He, X. (2004). *K*-means clustering via principal component analysis. pp. 29, ACM, New York, NY.
112. Hartigan, J. & Wong, M. (1979). Algorithm AS 136: a *K*-means clustering algorithm. *Appl. Stat.* **28**, 100–108.
113. Krzanowski, W. & Lai, Y. (1988). A criterion for determining the number of groups in a data set using sum-of-squares clustering. *Biometrics*, **44**, 23–34.
114. Likas, A., Vlassis, N. & Verbeek, J. J. (2003). The global *k*-means clustering algorithm. *Pattern Recognition*, **36**, 451–461.
115. Arya, S., Mount, D., Netanyahu, N., Silverman, R. & Wu, A. (1998). An optimal algorithm for approximate nearest neighbor searching fixed dimensions. *J. ACM*, **45**, 891–923.
116. Wold, H. (1966). Estimation of principal components and related models by iterative least squares. *Multivariate Anal.* **1**, 391–420.
117. Wold, H. (1975). Path models with latent variables: the NIPALS approach. In *Quantitative Sociology: International Perspectives on Mathematical and Statistical Model Building*. pp. 307–357.

Please cite this article as: Matthes, D. *et al.*, Driving Forces and Structural Determinants of Steric Zipper Peptide Oligomer Formation Elucidated by Atomistic Simulations, *J. Mol. Biol.* (2012), doi:10.1016/j.jmb.2012.02.004