INSTITUTE OF PHYSICS PUBLISHING PHYSICAL BIOLOGY
Phys. Biol. **1** (2004) 89–99 PII: S1478-3967(04)81452-9

Side chain interactions determine the amyloid organization: a single layer β-sheet molecular structure of the calcitonin peptide segment 15–19

David Zanuy[1,5], Nurit Haspel[2,5], Hui-Hsu (Gavin) Tsai[3], Buyong Ma[3], Kannan Gunasekaran[3], Haim J Wolfson[2] and Ruth Nussinov[3,4]

[1] Laboratory of Experimental and Computational Biology, NCI-Frederick, Bldg 469, Rm 151, Frederick, MD 21702, USA
[2] School of Computer Science, Faculty of Exact Sciences, Tel Aviv University, Tel Aviv 69978, Israel
[3] Basic Research Program, SAIC-Frederick Inc., Laboratory of Experimental and Computational Biology, NCI-Frederick, Bldg 469, Rm 151, Frederick, MD 21702, USA
[4] Sackler Institute of Molecular Medicine, Department of Human Genetics, Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv 69978, Israel

E-mail: zanuyd@ncifcrf.gov

Received 24 February 2004
Accepted for publication 3 June 2004
Published 22 June 2004
Online at stacks.iop.org/PhysBio/1/89
doi: 10.1088/1478-3967/1/2/005

Abstract
In this paper we present a detailed atomic model for a protofilament, the most basic organization level, of the amyloid fibre formed by the peptide DFNKF. This pentapeptide is a segment derived from the human calcitonin, a natural amyloidogenic protein. Our model, which represents the outcome of extensive explicit solvent molecular dynamics (MD) simulations of different strand/sheet organizations, is a single β-sheet filament largely without a hydrophobic core. Nevertheless, this structure is capable of reproducing the main features of the characteristic amyloid fibril organization and provides clues to the molecular basis of its experimental aggregation behaviour. Our results show that the side chains' chemical diversity induces the formation of a complex network of interactions that finally determine the microscopic arrangement of the strands at the protofilament level. This network of interactions, consisting of both side chain–side chain and backbone–side chain interactions, confers on the final single β-sheet arrangement an unexpected stability, both by enhancing the association of related chemical groups and, at the same time, by shielding the hydrophobic segments from the polar solvent. The chemical physical characterization of this protofilament provides hints to the possible thermodynamical basis of the supra molecular organization that allows the formation of the filaments by lateral association of the preformed protofibrils. Its regular, highly polarized structure shows how other protofilaments can assemble. In terms of structural biology, our results clearly indicate that an amyloid organization implies a degree of complexity far beyond a simple nonspecific association of peptide strands via amide hydrogen bonds.

[5] These authors contributed equally to this work.

Introduction

Amyloid formation is a problem of intense interest. Many groups have attempted to derive the laws that drive globular, coherently folded proteins towards partially ordered, insoluble and potentially toxic fibres. From a physiological point of view, amyloid deposits appear to be the common link between many unrelated pathologies, such as Alzheimer's, prion related encephalopathies, type II diabetes and Huntington's disease. *In vivo*, its presence is related to the considerable damage in the tissues where it is deposited [1]. Interestingly, most of the proteins that characterize these pathologies do not appear to be similar, either in sequence or in function. Furthermore, currently, there is no direct evidence on the mechanisms involved in cellular death. Among the hypothetical mechanisms, the disruption of the cellular membrane by the oligomeric precursors prior to the full fibre precipitation [2] appears to be the most widely accepted.

From a structural chemistry standpoint, amyloid formation is still a major challenge and understanding the rules of formation of partially ordered molecular complexes, i.e., the fibres, should provide clues to the design of drugs to prevent or efficiently block the progression of these diseases. Despite the lack of correlation among the proteins that form amyloid fibres *in vivo*, the physical process of formation and the final organization share some general features, such as a common fibre morphology as observed by electron microscopy (EM) [3], a kinetic pattern very similar to that observed in nucleated polymerization [4], an inherent sequence-dependent capability to form fibres under the right physico-chemical conditions [5] and the common presence of certain types of amino acids in those sequences that are characterized by having the highest rates of amyloid fibril formation, i.e., the fastest kinetics of aggregation [6]. This recognition has led to a search for generalized models to discover the chemical principles that govern the pathological (local or global) unfolding and thereby to overcome its development [7, 8]. It is increasingly recognized, though, that amyloid self-organization implies some degree of structural complexity involving a process that goes beyond a nonspecific aggregation of a preformed β-sheet [9]. Yet, there are still unexplained points relating to the kinetics of amyloid aggregation and its elongation. A simplistic approach cannot explain either the lag phase or the processes that take place during amyloid growth. The lag phase itself is far from being a static black box: for example, the fibre formation of the amyloid beta peptide (Aβ) presents a dynamic formation and reorganization of complex secondary structure patterns, in which the formation of β-sheets represents the last step [10]. The islet amyloid polypeptide (IAPP) provides yet another example of multiple kinetic events in amyloid aggregation [11, 12]. A related unsolved problem emerges from studies of fusion proteins as models for amyloidogenic enzymes [13]. There, only some segments of the new fusion proteins actively participate in the filament assembly. These largely correspond to domains previously shown to present a significantly high tendency to form amyloid fibres. Thus, instead of inducing a global destabilization and propagation of fibrillogenesis over all the chimera sequences, only those segments that already presented a higher amyloidogenicity actively participate in the fibre organization of the fusion proteins. Therefore, even if there is no specificity in side chain interactions, there is likely to be some degree of propensity linked to the particular chemical features of the side chains and/or specific side chain clusters.

Here, we focus on yet another unresolved problem, the assembly of the molecular complexes, which is an indirect reflection of the thermodynamics of fibre formation. *In vitro*, the ordered aggregation is highly dependent on the peptide concentration, precipitating only at the proper conditions and once achieved, is very difficult to reverse. The equilibrium of the fibrillogenesis reaction changes under physiological conditions, affecting the shape of the energy landscape and invalidating a classical scenario that describes an equilibrium between the folded free conformers and a formation of an assembly towards a seed micro-fibre. *In vivo*, amyloid formation and stability overwhelm the cellular regeneration capability, drags the stationary state of protein synthesis into an unbalanced point that feeds the *de novo* fibre formation. At the end, the new structure becomes an invasive tissue pathogen [14]. Finally, the consensus fibre organization derives more from the lack of resolution of the experimental information than from univocal evidence. The low degree of crystallinity presented by the fibres, especially when obtained from complete protein sequences, makes the determination of the three-dimensional organization by x-ray diffraction data impossible. The extremely low solubility of the samples prohibits solution NMR. Thus, the only available information relates to the minimal geometrical parameters that characterize the association of peptide segments into supra-molecular assemblies of β-sheets, named cross-β structures. This leads to a pseudo super-helical organization, where the inter-strand hydrogen bonds are in the helical axis direction. This simple structural model seemed to explain the common morphology of the amyloid fibres, which present symmetric grooves on their surfaces as observed by EM. These grooves have classically been interpreted as the outcome of the twist between strands of the same sheet, without accounting for the few orders of magnitude difference between a sheet and a complete fibre. Furthermore, experimental evidence points to some higher degree of microscopic polymorphism within a similar mesoscopic organization. Solid state NMR studies on Aβ peptide segments pointed to molecular clusters closer to a β-helix of some complex proteins than to simple lateral and longitudinal strand association [15]. In other cases, high quality x-ray diffractograms of highly oriented fibres constituted by polar peptides showed an unexpected flat, lateral association [16].

As a first step towards the understanding of the physico-chemical rules that govern this self-assembly process, we attempt to characterize the molecular assemblies that compose the amyloid aggregates and the chemical implications of their patterns of interactions. In order to achieve this goal, we use a new approach based on the actual knowledge of the fibril formation process: experimentally, it is well established that fibril growth initiates from a minimal pre-formed, soluble self-assembled seed-complex, that must retain its organization

in water. This is essential in order for the seed to assume the role of a structural template for the final growth process. Since we always study peptides that are experimentally known to be fibre formers, we overcome the formation process. We assume that a seed complex must exist in solution and must preserve some degree of organization. On the basis of this assumption, we explore final possible arrangements of molecular complexes and evaluate their relative stabilities. By these means we can avoid sampling many micro-states of aggregation of a very small number of molecular assemblies, which is extremely time consuming. We do not attempt to directly assess the process of amyloid formation; instead, we model the structure of the minimal seed organization. Further, due to computational limitations, we do not study the final shape and size of this so-called seed complex, but a number of molecular species that could mimic the minimal organization requirements. Once we have built different combinations, through high temperature molecular dynamics (MD) simulations we establish the relative stabilities of the models. Those models that are able to retain some degree of self-organization for the time simulated would be good candidates to study the structural properties of the amyloid fibres formed by a given sequence.

In this paper we study the pentapeptide fragment DFNKF, which is a short segment extracted from the human calcitonin (hCT) (residues 15–19 of hCT) [17]. The whole protein is a 32 amino acid hormone, produced by the thyroid C-cells. Its function is related to calcium homeostasis. Amyloid fibril formation of hCT is associated with medullary carcinoma of the thyroid [18, 19]. It was shown that hCT has little secondary structure at room temperature. However, hCT fibrils were found to be highly ordered, consisting of both α-helix and β-sheet elements, experimentally shown to form fibrils similar to those of the entire hormone sequence [20, 21].

This particular pentapeptide represents a unique opportunity for studying the microscopic arrangement of a fibril complex: its short length makes it suitable for all atom simulations within a reasonable time frame. Its composition is particularly attractive, since it presents an alternation of highly polar (or charged) side chains and highly hydrophobic side chains (Phe). Studying how this chemical nature influences the organization of its assembly should provide insights into the cohesive interactions in the fibril. Finally, the reported fast rates of aggregation of this peptide [20] hint that its microscopic structure is related to this behaviour. As we will show, our molecular template reflects the complexity of its chemical nature and provides an apparent simple organization that is perfectly compatible with the observed fast kinetics. It presents the minimal degree of fibre organization (the protofibril) that is based on a single sheet, organized by parallel oriented strands. The main stabilizing non-covalent interactions are due to side chain–side chain and side chain–main chain interactions. This complex network of interactions provides regularity to the assembly of strands and enhances the protection of the nonpolar residues from the water solvent. In this paper, we describe the underlying chemistry of the interactions stabilizing the organization of this oligomer. The details of all molecular simulations leading to this model will be published elsewhere (Haspel *et al*, manuscript in preparation).

While carrying out this investigation, solid state NMR data on the DFNKF peptide have been published [22]. On the basis of the isotope labelling of two Phe positions, the authors have suggested that the DFNKF forms antiparallel architecture at neutral pH. However, our simulations carried out so far suggested the parallel arrangement of strands. We are in the process of examining additional possible anti-parallel β-sheets, varying the number of strands and in particular the side chain conformations. Furthermore, it is interesting to note that the full length calcitonin peptide forms a mixture of anti-parallel and parallel β-sheets when the pH is varied [23]. Thus, we may well have a case of structural polymorphism with both parallel and antiparallel arrangements. We note that currently there is no direct evidence that the spin couplings observed experimentally reflect the β-strands association rather than supra-molecular organization.

Methods

Molecular templates

The simulated models were built with the InsightII [32] molecular package by automatic backbone superposition of basic preformed structures. Due to the lack of experimental three dimension information, these preformed structural patterns were fixed in order to reproduce the minimal geometrical requirement of the cross β-helix motif. Thus, for any given arrangement, the inter-strand distance within sheets was fixed at 5.00 Å and between sheets at 10.00 Å. The intra-sheet twist was initially avoided, since it is highly dependent on the chemical constitution (i.e. peptide sequence). Nevertheless, our previous works have shown that if the system is correctly relaxed, the correct twist is achieved after sufficient steps of energy minimization. Each built model was simulated for 4 ns at 350 K.

A summary of all the simulations carried out can be found in table 1. We initially built models which consisted of two four-stranded sheet (named octamers), since their small size made them suitable for atomic studies. Further, in an organization based on a two-sheet structure, some strands were hidden from water and others exposed. This allowed studies of the effect of the different surroundings on the association process. The initial arrangements that we tested differed in the strand orientation within the sheets (parallel or anti-parallel) and of the sheets with respect to one another (again, parallel or anti-parallel). For each arrangement, more than one side chain conformation was chosen. First, an initial combination of side chain conformations is energetically minimized. Then, depending on the steric hindrance that this first template presents, different new combinations are set (in the third column of table 1, for each letter combination there is more than one template). Since none of the templates maintained an acceptable degree of regularity, we enlarged the number of associated sheets. For size considerations we reduced the

Table 1. Summary of all the simulations carried out in order to find a suitable organizational model for the sequence *DFNKF*. The relative stabilities between structural templates were established by comparing the ability of retaining a certain degree of order after 4 ns of simulation (at high temperature, 350 K). The size of the fibril-like complex, the number of arrangements used as structural probe, the details of each model, the size of the simulation box used and the total number of molecules for each particular case are specified.

System size and supramolecular organization	Number of different tested arrangements	Arrangement details[a]	Simulation box size (Å)	Number of TIP3 waters
Two sheets, four strand per sheet (*octamers*)	13	A-a (3 models)[b] A-p (3 models) P-p (3 models) P-a (4 models)	45 45 45	2720
Three sheets, three sheets per strand (*nonamers*)	7	P-p (1 model) P-a (2 models) A-p (2 models) A-a (2 models)	50 50 50	3830
Single sheet nine strands	1	P	84 45 45	5316

[a] Each organization varies depending on the orientation of the pentapeptide strands within and between sheets. The code used in the table above is: 'P', parallel orientation, 'A' antiparallel orientation, capital letters within sheet, lower case letters between sheets.

[b] When more than one model is tested to determine the strands' orientation pattern, that includes changing side chains conformation and possible shifting along the lateral association axis, i.e., the different arranged sheets would be out of register.

number of strands per sheet. This led to arrangements of three sheets with three stands in each (named nonamers). We used the relatively more ordered combinations of side chain conformers and strand orientations from the first series of simulations. We built seven new complexes. Again, none maintained a minimal degree of assembled integrity as a fibril template. Thus, we probed a final protofibril model based on a single sheet (last model of table 1).

Molecular dynamics simulations

All calculations were performed using the Charmm program [24]. All atoms of the system were considered explicitly and the potential energy of the system was computed using the Charmm 22 force field [25]. All simulations were performed using an NVT ensemble in a orthorhombic simulation box and periodic boundary conditions were applied using the nearest image convention. The box size was adjusted to the complex length in order to maintain infinite dilution conditions within the cut-off applied, i.e., in all cases the distance between any given atom of a peptide chain to any edge of the simulation box was at least 1.5 times the cut-off distance. Thus, we assume that any atom belonging to a peptide chain sees only the water images rather than its own. We used explicit representation of the solvent molecules, using the TIP3 water model [26]. This combination of parameters has proved to be conformationally and energetically accurate for studying polypeptides and proteins with high degree of secondary structure [27–31]. All initial amyloid models were solvated with water molecules taken from a Monte Carlo equilibrium simulation until an approximate total density of 1 g cc^{-1} was achieved.

Before running each molecular dynamics simulation, the potential energy of the system was minimized for a total of 10 000 minimization steps in order to relax bond geometries and overcome possible atomic overlaps. We applied a combination of the steepest-descent and Newton–Raphson methods, performing at first 6500 steps of the former and ending the procedure by 3500 of the latter. Each system was heated for 20 ps until the simulation temperature was reached and 20 ps of thermal equilibration were run before starting to monitor the molecular evolution. This protocol had been used successfully in our previous investigations. All simulations were carried out at 350 K, which imposed thermal stress that is slightly higher than physiological conditions. This strategy has allowed us to increase the rate of any structural rearrangement, i.e., to obtain some kinetic perspective. Residue based cut-offs were applied at 12 Å, i.e. if two residues or a residue and a water molecule have any atoms within 12 Å, the interaction between the entire pair is evaluated. The *SHAKE* algorithm [33] was applied to fix the bond lengths and a numerical integration time step of 2 fs was used for all simulations. The non-bonded pair list was updated every 25 steps. The MD trajectories were saved every 500 steps (1-ps interval) for subsequent analysis.

All the trajectories' analyses were performed with software developed by us and that had successfully been tested as in our previous work [17, 18]. The codes include the geometrical characterization of amyloidogenic ensembles and the calculation of statistical magnitudes such as root-mean square deviation (rmsd) or radial function distribution ($g(r)$).

Molecular electrostatic potential (MEP). Molecular electrostatic potentials (MEP) in a water solution were calculated with a classical description of the solute molecule using single point charges centred in every atom centre of mass, and solving the Poisson–Boltzmann equation using the finite difference method [34]. All calculations were performed with the Delphi module of the InsightII molecular package [31], using the Charmm 22 force field [25]. All studied structures were centred in a grid box of 30 Å edge size. A point grid space of 1 Å was used for a total $1.601\,613 \times 10^6$ grid points. Dielectric constants of water and solute were set at 80.0 and 1.0 respectively, using the vacuum dielectric constant for the latter since the reorientation of charges was already taken into account with the molecular dynamics simulations (explicit

TIP3 waters were used). The Poisson–Boltzmann equation was solved analytically until conversion was achieved. The conversion criterion was accomplished when the difference in total electrostatic potential between two consecutive iterations was lower than $1 \times 10^{-6}\,\kappa T/e$.

Results and discussion

In this work we present the model obtained following an extensive exploration of possible molecular arrangements that may permit the *DFNKF* peptide to form amyloid deposits. This peptide presented very attractive properties, such as short length, which made it suitable for atomistic simulations, variability of chemical groups in its side chains and, despite its length, experimentally it showed an unexpected high rate of fibre formation [20]. We followed the strategy we used in other sequences [28–31]. We performed molecular dynamics simulations of small molecular assemblies, described at a full atomistic level (both solvent and complex assemblies). For each peptide sequence, we modelled all possible strand arrangements in which β-sheets can associate (parallel or antiparallel orientation of backbones, either within or between sheets) and tested their reliability in terms of relative stabilities. For the few amyloid fibres that have been solved ultrastructurally, this strategy has proved successful [35–37]. In this case, based on the small peptide size and the presence of nonpolar residues, we assumed that the protofilament structure would consist of more than one sheet. This kind of organization would shield the aromatic groups from the polar solvent and favour inter-strand interactions between charged group clusters. We performed up to 21 different trajectories, with systems constituted by two or three sheets, for a total of 84 ns (see table 1). We observed a recurring tendency to lose inter-sheet association while the single sheets remained stable. Surprisingly, in aqueous surroundings, the four-stranded sheets were more stable when they lost their association. Hence, we inferred a potential molecular assembly based on a single sheet protofilament, an assumption first confirmed when testing the relative stability of a single sheet consisting of nine parallel strands (figure 1). Subsequent parallel tempering runs of four strands over a temperature range of 300–600 K have further confirmed that our model has the lowest energy among all explored conformers (manuscript in preparation).

The first remarkable aspect of our model is the fact that we obtained a stable organization based on a single β-sheet without a significant hydrophobic core. This is interesting since, in naturally occurring folded protein systems, β-sheets are almost always organized in a hydrophobic core. However, the current system does not fold around this type of chemical group. Rather, it organizes entirely by means of close range contacts between neighbouring chemical groups, either belonging to the peptide backbone or to the residue side chains of the charged/polar *DFNKF*, with electrostatic interactions dominating the system, assisted by aromatic and some hydrophobic interactions (see below). Intriguingly, our results are in agreement with previous experimental reports that showed formation of a single layer stable β-sheet conformation either via *de novo* peptide design [38]

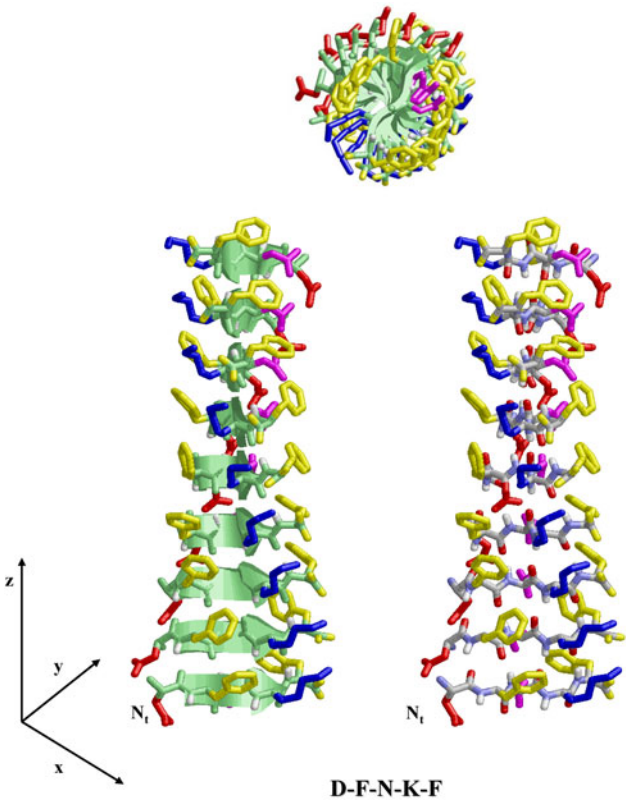

Figure 1. Atomistic representation of the protofilament model for the amyloid fibre constituted by the peptide DFNKF. On the top, the equatorial projection is depicted. On the bottom two different axial projections are represented. On the left side the main chain atoms are represented as secondary structure arrows, while side chain heavy atoms are depicted by sticks. On the right side all heavy atoms are depicted. Each side chain type is differently coloured: Asp in red, Phe in yellow, Asn in magenta and Lys in blue.

or through a genetic redesign of existing natural proteins [39]. Our results not only confirm the Koide *et al* proposition that such a conformation may also exist in amyloids, but extend that hypothesis to assemblies of small molecules that have no globular domains. Further, while in their case the strands are covalently connected to form hairpins, here the unlinked strands are unconstrained, optimizing their interactions.

The second interesting point in our model is the strand orientation within the filament sheet. At first sight, the combination of the *DFNKF* peptide sequence and the atoms' spatial distribution imposed by the main chain conformation (characterized by a backbone arrangement of approximately 2/1 symmetry) leads to an alternating distribution of both polar and aromatic side chains along the peptide sequence. Thus, on its own, the backbone conformation may favour a parallel strand arrangement, since it allows stacking of all aromatic rings of one strand with neighbouring chains. However, this simple hypothesis is insufficient to guarantee the high degree of organization *per se*. First, in our case experiments suggest that aromaticity may not be sufficient to trigger fibre growth. Amyloid formation experiments on tetrapeptides derived from *DFNKF*, truncating either the N or C residue termini, showed that *FNKF*, which contained both aromatic

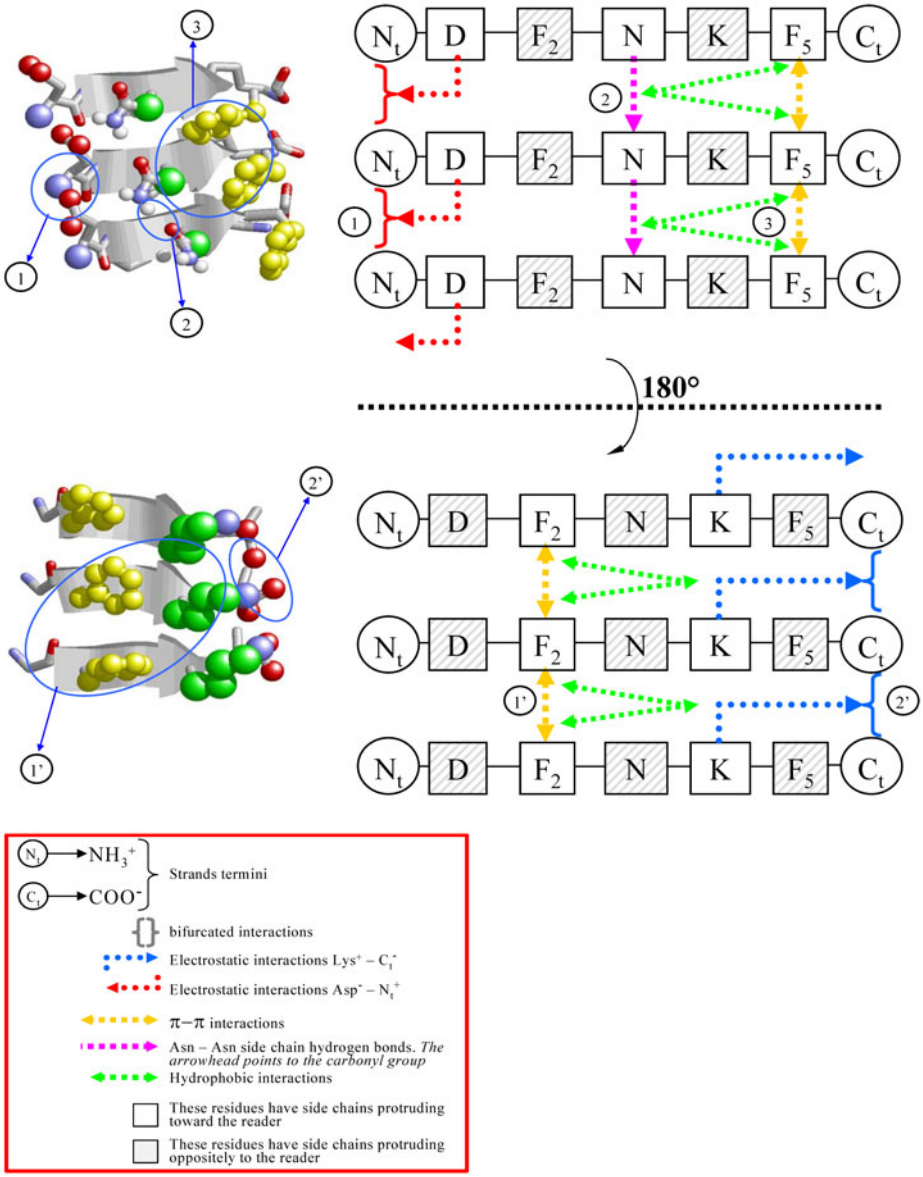

Figure 2. Schematic representation of the side chain interactions network described in our molecular model. The numerical determination of each interaction is presented in supplementary material. The backbone inter-strand hydrogen bonds are not depicted. Each side of the sheet is presented separately. On the left side the corresponding molecular arrangements are depicted with molecular representations of the each side chains involved in any given interaction. The colour code is the same as in figure 1. All hydrogen atoms have been omitted except those involved in the Asn side chains hydrogen bonds spine. Methylene and ammonium groups, and carboxylic oxygen atoms are depicted by hard spheres.

rings, did not form fibres but amorphous aggregates, while *DFNK* formed visible fibres, although their organization was not canonical [20]. Second, if aromatic rings drove the self-assembly, there should be some degree of lateral association, in order to reduce the contact with water. Yet, our attempts to obtain a regular lateral assembly of sheets always led to disordered arrangements. Thus, hydrophilic residues have some important role in the structure of the assembly. This assumption seems to be also supported when studying different chemical modifications, which also suggested a high degree of dependence between the total ionization state of the peptide and the possible arrangement that the structural seed could display [40].

Despite the apparent simplicity of the molecular model (figure 1), a high degree of complexity must permit the microfilament to present such an unexpected stability. A systematic analysis (details in supplementary material) indicated that this particular single sheet arrangement stabilizes the assembly solely *via* intra-sheet interactions: the characteristic hydrogen bonds of a β-sheet organization and a complex network of non-covalent interactions involving side chains (figure 1).

Figure 2 depicts the complexity of this network. First, short aliphatic segments are partially protected from the solvent by interacting with aromatic side chains. Second, the charged edges of the peptide strands are partially shielded from

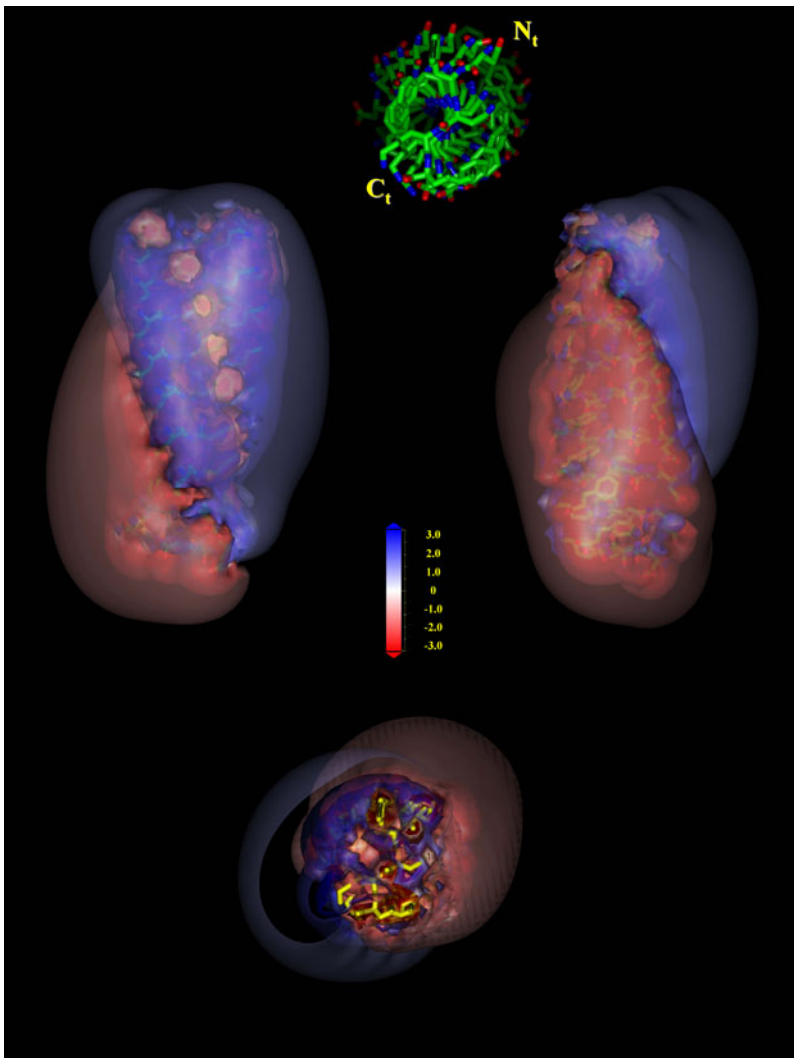

Figure 3. Molecular electrostatic potential distribution in water for the DFNKF fibre. Six different isoenergetic surfaces are depicted: all the points with an electrostatic potential of $+/-$ 1, 2 and 3 kcal mol^{-1} e^{-1}, respectively are shown as solid semi-transparent surfaces. The colour intensity of each surface is proportional to the potential value (see scale). Three different projections are provided, two lateral sections (centre) and one equatorial section (bottom). A molecular equatorial projection is also provided for clarification.

the solvent forming bifurcated salt bridges with the side chain charged groups. N termini interact with Asp side chains of the neighbouring strand, since its length allows the side chain to adopt a global conformation that resembles a C_5 interaction geometry. The C termini charged groups are stabilized by the interaction with Lys, whose long side chain folds and fits perfectly between the terminal carboxylic groups. These two interactions favour the helical twist that protects the aliphatic groups from exposure to the polar environment. Finally, the combination of both the inner super-helical twist and the conformation of the aliphatic segment of the Lys side chain allows their methyl groups to interact with the aromatic spine of this side of the sheet (Phe_2). Another set of interactions is provided by Asn through an extra set of side chain hydrogen bonds, increasing the thermal stability of the complex and anchoring the super-helical rotation axis. This interaction allows some shielding for the methyl group of the Asn side chain, by interacting with the other aromatic group (Phe_5). Finally, the perfect alignment of the aromatic groups, due to the parallel arrangement of the strands induces a perfect pi stacking of the phenyl groups. However, as both experimental and simulation results showed, this type of interactions gives extra global stability to the fibre but does not completely justify the formation of this organization.

Hence, this 3D arrangement protects the aromatic rings and the aliphatic segments from the solvent. This point is confirmed when calculating the approximate molecular electrostatic potential in different representative snapshots (figure 3). The high charge concentration at both sheet edges induces high polarization along the super-helical axis, mainly a consequence of the contact between charged side chains and the high dielectric solvent. However, since these chemical groups form stable salt bridges, the capability of their charged groups to reorient towards electromagnetic fields is compromised, generating a high electric dipole along the sheet structure. This scenario induces a partial isolation of the

hydrophobic groups from the solvent, since the low mobility of the charged residues does not allow a rapid redistribution of the water molecules around the fibre, and diminishes the effective dielectric surrounding the confronting apolar groups. This is confirmed when plotting the radial distribution function of the water molecules around the associated strands for the time simulated (see supplementary material). Finally, the distribution of oppositely charged chemical potentials explains how this protofilament can be formed in water without promoting lateral aggregation *via* aromatic residues. It also shows a possible path through which two protofilaments can associate, *via* lateral electrostatic attraction.

In summary, we have been able to model a detailed atomic complex at the lowest organization level of an amyloid fibre consisting of the peptide DFNKF. Remarkably, this protofilament which is constructed based on a single sheet with strands arranged in parallel, depicts a high degree of complexity mainly due to the side chains conformational organization. Thus, a complex network of non-covalent interactions is responsible for the high stability of the assembled complex in solution, *via* a notable optimization of all possible interaction geometries. It provides directionality (hydrogen bonds and stacking interactions) and shielding the charge repulsions by multicentric side chain–main chain interactions in both strand edges. Thus, even if the role of the side chains in the overall fibrillogenesis process is non-specific, they nevertheless shelter the final microscopic structure of the protofilaments. The non-covalent interactions provide the precise chemical environment that allows the lateral association of the filaments, that is, the final structure of the fibre. Moreover, this network of interactions indicates how a single β-sheet can exist despite the absence of a significant hydrophobic core.

Conclusions

(1) We have modelled a detailed molecular template for the amyloid protofibril formed by the peptide DFNKF. Its high stability in a water environment as well as its apparent arrangement in a single polarized sheet is in complete agreement with the experimental evidence: long ordered fibres formed through very fast kinetics.

(2) Our molecular model presents a high degree of molecular complexity due to the regular distribution of non-covalent interactions mediated by the side chains of each residue. This provides a favourable environment for the apolar side chains, allowing the formation of a molecular ensemble in a highly polar solvent.

(3) The overall distribution of chemical groups in the protofilament induces a highly polarized arrangement. This arrangement provides hints to the paths taken by the associating protofibrils into mature filaments, through electrostatic attraction. The high dielectric constant of the water environment further promotes the association of these polarized complexes.

Outlook

The specific role of the amino acid side chains in the amyloid organization is still being debated. Nevertheless, beyond the discussion of what is the final effect of the sequence variability, either through specific chemical interactions or changes in the energy landscape profile, i.e., the kinetics of the process [8], our observations illustrate that side chains explicitly influence the organization of the assembled strands, at least at the first protofilament level of fibre organization. The sequence composition of any peptide must determine the atomistic organization of its protofilament skeleton. Actually, this point has already been explored by other groups with molecular dynamics simulations. Caflisch and co-workers showed that the side chain interactions mediate the selection of the preferred orientation within sheets for the prion peptide 7–13 segment of the protein Sup35 [41]. Klimov and Thirumalai [42] have shown similar behaviour when studying β-sheet formation of the 16–22 segment of the Aβ peptide. For the A$\beta_{16–22}$ peptide, we observed that side chain interactions not only mediated in the strand arrangement within sheets, but also in their lateral association [31]. Hence, a peptide with the same length as *DFNKF* presents a completely different organization at the protofilament level. The *DFNKF* segment shows a preference towards a single sheet, while the A$\beta_{16–22}$ has at least two. However, from a chemical standpoint they differed solely in their side chains composition, i.e., their sequence. This point appears to be confirmed when revisiting other amyloidogenic sequences and their respective structural models: A$\beta_{11–25}$ [37], the Syrian hamster prion protein segment, PrP [28] and the IAPP segments 22–27 and 22–29 [29–30]: all presented a minimum assembly complex with more than one sheet. Yet, in these peptides, the side chains chemical character respectively varied from highly to moderately hydrophobic. Thus, an association of more than one sheet was needed to achieve an acceptable degree of stability, making a single sheet filament an unviable organization. In the current *DFNKF* segment, the highly charged edges, with a central polar side chain, and the potential stacking interactions of the aromatic side chains allow the strands to achieve a cylindrical projection, which enhances the contact surface of polar groups with the solvent and shields the low polar groups from interacting with water (as shown in figure 1). Thus, the organization of an amyloid fibre at the molecular level must depend on the balance of forces of all the chemical groups that constitute the peptide sequence, going beyond the simplistic approach of hydrogen bonds as the driving force or even the hydrophobic collapse.

The previous examples were limited to relatively short peptides. However, when more complex peptides are studied, similar observations can be obtained. We have previously shown that if there is a hydrophobic cluster, the inherent tendency of the strands is to self-assemble in more than one direction (laterally), since the hydrophobic side chains provided an untenable degree of destabilization, incompatible with any attempt to seed the self-assembly based on a single sheet filament. In longer peptides, such as A$\beta_{16–35}$, A$\beta_{10–35}$ and A$\beta_{1–40}$ [31, 35–36], the scenario appears to be different,

but the ultra structure of the assembly shows an analogous behaviour. In these cases, a single strand is long enough to lead to a hydrophobic collapse, driving the backbone towards the formation of a non-canonical turn (also referred to as a *bend*) that connects two long segments in a β-sheet conformation. This allows burying part of the hydrophobic side chains between the two segments, which not only provides stable organization but makes it more kinetically accessible. If we do not consider the bent zone, we would ultimately obtain a lateral association of two sheets similar to that described for shorter peptides. We believe that all these indications support the idea that the chemical composition of the side chains determines the way individual strands assemble into ordered structural patterns, maximizing the energy of cohesion either within and/or between sheets, depending on the kinetic accessibility of any potential organization.

This scenario of amyloid formation merges two opposing views: the first focuses on the simplicity of the final organization, which always leads to partially ordered fibres, while the second centres on the complexity of molecular arrangements that permit the association of structural modules that form these fibres. In this communication, we have shown how a simple peptide of five residues can capture an unexpected degree of complexity when self-assembled into β-sheet rich aggregates. Furthermore, our work solely focused on the smallest degree of organization, the protofilament level (i.e., the most basic scaffold of the fibre architecture). This microscopic complexity, along with the structural diversity associated with different sequences and lengths, leads to new challenges: are the partially folded states of entire proteins structured enough to serve as template for ordered aggregation [43], or is there a conformational selection, analogous to the protein folding process, to achieve the correct conformation that would lead to fibril formation? We believe that this last scenario describes better the real process of amyloid formation. However, at this stage of the investigation, one needs to collect more data to derive the general rules of this complex and challenging question. Our work, still based on small protein segments, constantly gives us new ideas to work with: the role of side chain interactions, as lateral association implies, the geometrical restraints imposed by the complex assembly of the structural modules and the electrostatic polarization of the protofilament in water that might guide the lateral association of the fibril structures. We truly believe that the final answer is going beyond β-sheet organization and leading us to take up supra-molecular chemistry.

Acknowledgments

We thank Dr Ehud Gazit for bringing to our attention this remarkably stable and interesting peptide and for several insightful discussions. We thank Dr C-J Tsai for discussions. In particular, we thank Dr Jacob V Maizel for encouragement. The computation times are provided by the National Cancer Institute's Frederick Advanced Biomedical Supercomputing Center and by the NIH Beowulf. The research of R Nussinov in Israel has been supported in part by the 'Center of Excellence in Geometric Computing and its Applications' funded by the Israel Science Foundation (administered by the *Israel Academy of Sciences*), and by the Adams Brain Center. This project has been funded in whole or in part with Federal funds from the National Cancer Institute, National Institutes of Health, under contract number NO1-CO-12400. The content of this publication does not necessarily reflect the view or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organization imply endorsement by the US Government.

Glossary of technical vocabulary

Amyloid and amyloid structures. The term 'amyloid' has been from the beginning surrounded by certain amount of uncertainty. It was originally used to name the inclusion bodies observed in brain tissue of Alzheimer's patients. Its etymology is related to the mistaken nature of those inclusions, since at the beginning of the 20th century they were believed to be constituted by complex carbohydrate polymers. After being identified as the main components of these aggregates, the name was kept for their major chemical component, the amyloid beta peptide (Aβ). This inherited terminology has resulted in a wide usage of the term amyloid to name any fibrous inclusion of proteins that relates with a known pathology. Actually, most recently, this term is being used to describe any insoluble fibre formed after incubating a solution with a dissolved protein. In order to avoid easy misunderstandings many authors have recently tended to use the terminology 'amyloid structure' when talking about the aforementioned fibre aggregates, in order to avoid confusion with the specific case of Aβ protein.

Fibril, protofilament, filament and fibre. This group of terms includes different related concepts, which describe the different degrees of organization biological fibres. This terminology is not solely used in amyloid structures but in other natural fibres such as silk and collagen. In general, the micro structure of a fibre is the sum of different assembled scaffold pieces. Thus, a '*fibril*' would be the lowest degree of organization, which in the context of amyloid proteins refers to the coherent organization between dozens or hundreds of protein chains assembled through β-sheets. The fibril term would include a range of sizes from assemblies that are still soluble (before precipitation) to the first insoluble particles. It is worth noting that molecular biology literature commonly uses the term 'fibril' when the degree of complexity of a protein fibre is unknown. Furthermore, the little amount of atomic resolution information available is always related to this organization level (i.e., solid state NMR). Once the growth of the fibrils reaches a stable size, they tend to collide and assemble laterally, like the different cords of a rope. These assembled fibrils constitute the '*protofilament*'. Analogously, the lateral association of a few protofilaments constitutes the '*filament*', which in more general terms is known as the 'fibre'. It is then the structures that can be observed by EM.

Molecular mechanics and molecular dynamics. An accurate description of the energy of any chemical system

can only be achieved by solving its electron wavefunction through quantum mechanics. This option becomes entirely unreliable for molecular systems of even a medium size, which are much smaller than biomacromolecules. However, when the properties of interest do not depend on changes in the electronic constitution of the studied chemical system, it is possible to obtain an accurate description of its internal energy through classical physics. Thus, it is possible to compute the molecular structure of any given macromolecule by representing its atoms as rigid spheres, with a determined partial charge and van der Waals radii, and its chemical bonds as springs. The atomic coordinates would be centred in the atom nuclei, which are also the centres of mass of each sphere. This approach is known as molecular mechanics and allows the computational chemists and biologists to explore the conformational minima of macromolecules by minimizing their potential energy. Furthermore, it allows the mathematical solution of Newton's equation of motion: the gradient of the total potential energy gives the forces that act over each atom of the system at any give moment. These forces are directly proportional to the acceleration that each atom presents. Through Newton's law, the double integral of the acceleration with respect to time step yields the new position of the atoms after a selected time frame. The automatic implementation of this procedure by mathematical algorithms that control temperature and pressure is the computational technique known as molecular dynamics.

Nucleation, nucleus-seed formation and lag phase. In macromolecular chemistry this concept describes the path followed by some chemical reactions of polymer formation and by some soluble molecular species when forming organized insoluble structures (such as the formation of salt crystals). Mainly the process is kinetically controlled by the size of the growing new species (either the polymer chain or the assembly of species). These new species are known as oligomers, which indeed are formed by the association of few monomers. While the oligomer size is not large enough the chemical process is very slow. Actually, from a mechanistic stand point there is a fast equilibrium exchange between oligomers and monomers. This period of time is known as the 'lag phase'. In the amyloid chemistry context, during this period there are not visible fibrils when studying the process *in vitro*. Once the ologimer reaches a minimal size (which is specific for each case) the reaction speeds up and the final structure is formed very promptly. In our particular case the fibril forms. This minimal size of the forming structure that becomes the threshold of the kinetic behaviour change is known as '*the nucleus of polymerization*' or '*nucleus-seed*'. This concept leads to the *seeding process*: in this nucleated chemical process, it is possible to bypass the lag phase by adding some pre-formed oligmers to the reaction media. When added, they act as reaction nuclei and increase dramatically the reaction velocity.

Stability of a molecular model. In the context of our research this concept relates to the capabilities of a molecular ensemble to maintain a particular organization. Thus, in general, one might describe as a 'stable model' any molecular conformation that is preserved under the simulation conditions. In the particular case of amyloid structures this term refers to the ability of a particular strands association to maintain a certain degree of ordered organization. Specifically, stability refers to regularity within sheets, in particular through the hydrogen bond direction axis and regularity in the lateral association of sheets (if it exists).

System. In general, in the context of molecular simulations, the term 'system' is used to account for all the molecular species included in a molecular simulation. Thus, when studying the conformational behaviour of macromolecules, it would include both the solvent molecules (if used) and the target molecules. However, in this communication this term might be used to name the peptide assemblies, since an amyloid aggregate is constituted by several chains that get to associate to form fibrils.

Self-assembly. In structural chemistry this term describes the chemical process of spontaneous association of molecules through non-covalent interactions, which leads to the formation of both thermodynamically and kinetically stable molecular complexes. The few classical examples of biological structures that form by this process include the DNA double helix association and membrane lipids bilayer organization.

References

[1] Kisilevsky R 2000 Amyloidogenesis: unquestioned answers and unanswered questions *J. Struct. Biol.* **130** 99–108

[2] Kagan B L, Hirakura Y, Azimov R, Azimova R and Lin M C 2001 The channel hypothesis of Alzheimer's disease: current status *Peptides* **23** 1311–5

[3] Sunde M and Blake C F 1998 From globular to the fibrous state: protein structure and structural conversion in amyloid formation *Q. Rev. Biophys.* **31** 1–39

[4] Rocher J C and Lansbury P T Jr 2000 Amylod fibrillogenesis: themes and variations *Curr. Opin. Struct. Biol.* **10** 60–8

[5] Fandrich M and Dobson C 2002 The behaviour of polyamino acids reveals an inverse side chain effect in amyloid structure formation *EMBO J.* **21** 5682–90

[6] Reches M, Porat Y and Gazit E 2002 Amyloid fibril formation by pentapeptide and tetrapeptide segmented of human calcitonin *J. Biol. Chem.* **277** 35475–80

[7] Gazi E 2002 Possible role of pi-stacking in self-assembly of amyloid fibrils *FASEB J.* **16** 77–83

[8] Chiti F, Stefani M, Taddei N, Ramponi G and Dobson C M 2003 Rationalization of the effects of mutations on peptide and protein aggregation rates *Nature* **424** 805–8

[9] Lopez de la Paz M, Goldie K, Zurdo J, Lacroix E, Dobson C M, Hoenger A and Serrano L 2002 De novo designed peptide-based amyloid fibrils *Proc. Natl Acad. Sci. USA* **99** 16052–7

[10] Kikirtadze M D, Condrom M M and Teplow D B 2001 Identification and characterization of the key kinetic intermediates in Amyloid b-protein fibrillogenesis *J. Mol. Biol.* **312** 1103–19

[11] Padrick S B and Miranker A D 2001 Islet Amyloid polypeptide: identification of long range contacts and local order on the fibrillogenesis pathway *J. Mol. Biol.* **308** 783–94

[12] Padrick S B and Miranker A D 2002 Islet amyloid: phase partitioning and secondary nucleation are central to the mechanism of fibrillogenesis *Biochemistry* **41** 4694–703

[13] Baxa U, Speransky V, Alasdair C S and Wickner R B 2002 Mechanism of inactivation on prion conversion of *Saccharomysces cerevisiae* Ure2 protein *Proc. Natl Acad. Sci. USA* **99** 5253–60
[14] Dobso C 2003 Protein folding and disease: a view from the first horizon symposium *Nat. Rev. Drug. Discov.* **2** 154–60
[15] Petkova A T, Ishi Y, Balbach J J, Antzutkin O N, Leapman R D, Delaglio F and Tycko R 2002 A structural model for Alzheimer's β-amyloid fibrils based on experimental constrains from solid state NMR *Proc. Natl Acad. Sci. USA* **99** 16742–7
[16] Balbirnie M, Grothe R and Eisenberg D S 2001 An amyloid forming peptide from the yeast prion sup35 reveals a dehydrated β-sheet structure for amyloid *Proc. Natl Acad. Sci. USA* **98** 2375–80
[17] Zaidi M, Inzerillo A M, Moonga B S, Bevis P J and Huang C L 2002 Forty years of calcitonin—where are we now? A tribute to the work of Iain Macintyre, FRS *Bone* **30** 655–63
[18] Butler M and Khan S 1986 Immunoreactive calcitonin in amyloid fibrils of medullary carcinoma of the thyroid gland. An immunogold staining technique *Arch. Pathol. Lab. Med.* **110** 647–99
[19] Berger G, Berger N, Guillaud M H, Trouillas J and Vauzelle J L 1988 Calcitonin-like immunoreactivity of amyloid fibrils in medullary thyroid carcinomas. An immunoelectron microscope study *Arch. A Pathol. Anat. Histopathol.* **412** 543–51
[20] Reches M, Porat Y and Gazit E 2002 Amyloid fibril formation by pentapeptide and tetrapeptide segmentd of human calcitonin *J. Biol. Chem.* **277** 35475–80
[21] Arvinte T, Cudd A and Drake A F 1993 The structure and mechanism of formation of human calcitonin fibrils *J. Biol. Chem.* **268** 6415–22
[22] Naito A, Kamihira M, Inoue R and Saito H 2004 Structural diversity of amyloid fibril formed in human calcitonin as revealed by site-directed 13C solid-state NMR spectroscopy *Magn. Reson. Chem.* **42** 247–57
[23] Kamihira M, Oshiro Y, Tuzi S, Nosaka Y A, Saito H and Naito A 2003 Effect of electrostatic interaction on fibril formation of human calcitonin as studied by high resolution solid state ^{13}C NMR *J. Biol. Chem.* **278** 2859–65
[24] Brooks R B, Bruccoleri R E, Olafson B D, Sate D J, Swaminathan S and Karplus M 1983 CHARMM: a program for macromolecular energy minimization and dynamics calculations *J. Comput. Chem.* **4** 187–217
[25] MacKerell J A D *et al* 1998 All-hydrogen empirical potential for molecular modeling and dynamics studies of proteins using the CHARMM22 force field *J. Phys. Chem.* B **102** 3586–616
[26] Jorgensen W L, Chandrasekhar J, Madura J D, Impey R W and Klein M L 1982 Comparison of simple potential functions for simulating liquid water *J. Chem. Phys.* **79** 926–35
[27] Bursulaya B D and Brooks C L III 1999 Folding free energy surface of three-stranded β-sheet protein *J. Am. Chem. Soc.* **121** 9946–51
[28] Ma B and Nussinov R 2002 Molecular dynamics simulations of alanine rich b-sheet oligomers: insight into amyloid formation *Protein. Sci.* **11** 2335–50
[29] Zanuy D, Ma B and Nussinov R 2003 Short peptide amyloid organization: stabilities and conformations of the islet amyloid peptide NFGAIL *Biophys. J.* **84** 1–11
[30] Zanuy D and Nussinov R 2003 The sequence dependence of fibre organization: a comparative molecular dynamics study of the islet amyloid polypeptide segments 22–27 and 22–29 *J. Mol. Biol.* **329** 565–84
[31] Ma B and Nussinov R 2002 Stabilities and conformations of Alzheimer's β-amyloid peptide oligomers ($A\beta_{16-22}$, $A\beta_{16-35}$, $A\beta_{10-35}$): sequence effects *Proc. Natl Acad. Sci. USA* **99** 14126–31
[32] INSIGHTII, Accelrys (2000) San Diego, California
[33] Ryckaert J P, Ciccoti G and Berendsen H J C 1977 Numerical integration of the Cartesian equations of motion of a system with constraints: molecular dynamics of *n*-Alkanes *J. Comput. Phys.* **23** 327–41
[34] Nicholls A and Honig B 1991 A rapid finite difference algorithm utilizing successive over-relaxation to solve Poisson–Boltzmann equations *J. Comput. Chem.* **12** 435
[35] Antzutkin O N, Leapman R D, Balbach J J and Tycko R 2002 Supramolecular structural constraints on Alzheimer's β-amyloid fibrils from electron microscopy and solid-state nuclear magnetic resonance *Biochemistry* **41** 15436–50
[36] Petkova A T, Ishii Y, Balach J J, Antzutkin O N, Leapman R D, Delaglio F and Tycko R 2002 A structural model for Alzheimer's β-Amyloid fibrils based on experimental constraints from solid state NMR *Proc. Natl Acad. Sci. USA* **99** 16742–7
[37] Silorski P, Atkins E D T and Serpell L S 2003 Structure and texture of fibrous crystals formed by Alzheimer's Aβ (11–25) peptide fragment *Structure* **11** 915–26
[38] Serran L 2000 The relationship between sequence and structure in elementary folding units *Adv. Protein Chem.* **53** 49–85
[39] Koide S, Huang X, Link K, Koide A, Bu Z and Engelman D M 2000 Design of single-layer β-sheets without a hydrophobic core *Nature* **403** 456–60
[40] Tsai H H (G), Zanuy D, Haspel N, Gunasekaran K, Ma B, Tsai C and Nussinov R 2004 The stability and dynamics of the human Calcitonin amyloid peptide DFNKF *Biophys. J.* at press
[41] Gsponer J, Haberthür U and Caflisch A 2003 The role of side-chain interactions in the early steps of aggregation: molecular dynamics simulations of an amyloid-forming peptide from the yeast prion Sup 35 *Proc. Natl Acad. Sci. USA* **100** 5154–9
[42] Klimov D and Thirumalai D 2003 Dissecting the assembly of $A\beta_{16-22}$ Amyloid Peptides into antiparallel β sheets *Structure* **11** 295–307
[43] DeMarco M L and Daggett V 2004 From conversion to aggregation: protofibril formation of the prion protein *Proc. Natl Acad. Sci.* **101** 2293–8