[Frontiers in Bioscience 13, 4039-4050, May 1, 2008]

Towards modeling of amyloid fibril structures

Jun-tao Guo[1], Ying Xu[2]

[1]Bioinformatics Research Center and Department of Computer Science, University of North Carolina at Charlotte, Charlotte, NC 28223, USA, [2]Department of Biochemistry and Molecular Biology and Institute of Bioinformatics, University of Georgia, Athens, GA 30602, USA

TABLE OF CONTENTS

1. ABSTRACT

Amyloid fibrils are associated with a number of debilitating diseases, including Alzheimer's disease and variant Creutzfeldt-Jakob disease. The elucidation of the structure of amyloid fibrils is an important step toward understanding the mechanism of amyloid formation and developing therapeutic agents for amyloid diseases. Despite great interests and substantial efforts from various research communities, deriving high-resolution structures of amyloid fibrils remains a challenging problem, due to the insolubility and non-crystalline nature of the fibrils. An array of experimental methods, such as electron microscopy, fiber diffraction, hydrogen-deuterium exchange, solid-state NMR, electron paramagnetic resonance spectroscopy and biochemical approaches, have been explored to study the problem, having yielded considerable amount of, though still partial, information about the fibril conformation. Computational modeling techniques can be used to predict and build structural models of amyloid fibrils, utilizing the available experimental data. Here, we describe a few computational methods for modeling of aggregate and fibril structures with a focus on protein threading-based approaches and discuss the challenging issues ahead.

2. INTRODUCTION

Amyloid fibrils are straight, unbranched, insoluble protein aggregates, formed through self-assembly of soluble proteins and peptides (1). As of today, over twenty proteins are known to be able to form amyloid fibrils in human. Although the exact role of amyloid fibrils in the pathology of amyloid diseases is still under debate, their association with an increasing number of diseases such as Alzheimer's disease (AD) and variant Creutzfeldt-Jakob disease (vCJD), and the apparent disruptive capability of the thread-like fibrils in brain tissues has generated strong interest in studying the amyloid fibril structure from different angles. Besides clinical implications, the formation of amyloid and amyloid-like fibrils also raised fundamental questions about protein folding as some proteins, not known to be associated with any diseases, can form amyloid-like fibrils *in vitro* under suitable conditions (2, 3). For example, myoglobin is a highly soluble and compact protein *in vivo*. Experiments from Dobson's lab revealed that apomyoglobin can form fibrillar structures with the typical characteristics of amyloid fibrils when the protein is incubated in solution with high pH, high temperature and increased salt concentration, suggesting the ability of proteins to form amyloid fibrils under the "right" conditions (3, 4).

Although amyloid forming peptides and proteins do not share any sequence homology, the structural characteristics of the amyloid fibrils derived from these peptides and proteins are remarkably similar. A dominating feature of amyloid or amyloid-like fibrils is the so-called "cross-beta pattern" revealed by X-ray diffraction, which was first reported from poached egg white in 1935 (5) and was later identified in isolated amyloid fibrils from infected tissues (6). In the cross-beta pattern, beta-strands run perpendicular to the principal axis of the fibrils while the hydrogen bonds between beta-strands are parallel to the axis (7).

Numerous fundamental and interesting questions have been raised about the amyloid fibrils regarding: i) the conversion mechanism from a natively folded protein structure to amyloid fibril, and ii) the detailed structural organizations of the amyloid fibrils (8). High resolution structures have been a subject of intense research because of its importance to our understanding of the functional roles of amyloid fibrils in amyloidoses and the potential bearing on the rational design of therapeutics. Although some success has been made in deriving the atomic-resolution microcrystal structures of non-physiological amyloid-like fibrils formed from short peptides of disease-related proteins (9, 10), many mysteries remain unsolved about amyloid fibril structures. One example is the polypeptide folding patterns in fibrils formed from longer peptides or full-length amyloid precursor proteins. What is available now is mostly low resolution structural data, collected using various experimental approaches, such as fiber diffraction, electron microscopy, hydrogen-deuterium exchange (HX), solid state NMR, small angle neutron scattering, limited proteolysis, and electron paramagnetic resonance spectroscopy (EPR), some of which have shed new lights on the structural details from various angles (11, 12).

Computational techniques offer a complementary approach to the study of amyloid fibril structures, which can be used: (a) to build structural models of the amyloid fibrils that are most consistent with the available experimental data; (b) to assess the stabilities of the modeled structures and the effects of mutations on the fibril structure; and (c) to study the fibril formation process (12, 13). To date, there are two general classes of computational approaches to modeling amyloid fibril structures. The first class of methods, which simulates the fibril formation process and the resulting structure, is in essence an "*ab initio*" approach. The advantage of this approach, at least theoretically, is that it can simulate the conversion process from soluble proteins to fibril, producing the fibril conformations at the same time. However, current atomic-level simulation methods are not yet quite up to the job due to the large system size and the long simulation time involved in fibril seed formation and in the fibrillation process, to provide very useful results. A more promising strategy, already making strides in the study of aggregate assembly and providing valuable insights, is to use low- to intermediate-resolution models, also called simplified folding models (12, 14). Instead of using atomic level force fields, low-resolution models rely on coarse-grained representation of peptides and proteins while ignoring the motion of the solvent molecules. Using a simplified representation of residues and a discontinuous molecular dynamics (DMD) simulation algorithm, Hall and colleagues were able to explore the protein aggregation process with a large system consisting of many polyalanine peptides, $KA_{14}K$. Their simulation results revealed a spontaneous formation of a protofilament or small fibril with a starting conformation of random coils (12, 14). The kinetics of aggregation can also studied using this simulation system.

The second class of methods predicts amyloid fibril structures through first generating fibril structural models, manually or automatically, and then assessing the model stability using molecular dynamics (MD) simulation techniques. This type of approach bypasses the fibril formation process and focuses on detailed end-product conformations including the chemical interactions that stabilize the fibril structure. The stability of such oligomer models can then be assessed using molecular dynamics simulations. It should be pointed out that this review is not intended to cover research works on simulating conformational changes of a single peptide. Rather we describe modeling techniques for oligomers and fibrils belonging to the second class, with a major focus on template-based modeling approaches. We refer readers to papers by Hall and Wagoner, Urbanc *et al.* (14, 15), and the papers by Sharma *et al.* and by Moroni *et al.* in this same issue for details of simulation and "*ab initio*" approaches to amyloid fibril modeling.

3. AN OVERVIEW OF AMYLOID FIBRIL STRUCTURAL MODELS

Most of the models proposed so far rely, more or less, on experimental observations. Other than all having the well-accepted common feature of amyloid or amyloid-like fibrils, the "cross-beta pattern", models differ from each other in two major aspects: 1) parallel or anti-parallel in strand arrangements within a beta-sheet; and 2) the number and locations of turns that separate the consecutive strands (Figure 1). Recent studies suggest that the arrangement of strands within a beta-sheet is length and sequence dependent. X-ray microcrystallography of amyloid fibrils formed from short disease-related peptides revealed both parallel and anti-parallel conformations (10). For example, the amyloid fibrils formed from two overlapping hexmers of amyloid-beta (peptide I: ^{35}MVGGVV^{40} and peptide II: ^{37}GGVVIA^{42}) belong to two different classes of the so-called "steric zippers". Peptide I forms anti-parallel sheets while peptide II forms parallel sheets (10). Compelling evidence from solid-state NMR and liquid suspension EPR studies on Abeta (1-40) and other amyloid fibrils formed from longer peptides suggests that the peptides in the fibril core is in-register, parallel arrangement (16-18) even though earlier models focused on anti-parallel beta-sheet conformation (19, 20). In the past several years, a number of parallel beta-sheet based models have been proposed for Abeta (1-40), but they don't agree with each other on the number and the locations of turns on each rung of the core structure, which range from zero turn

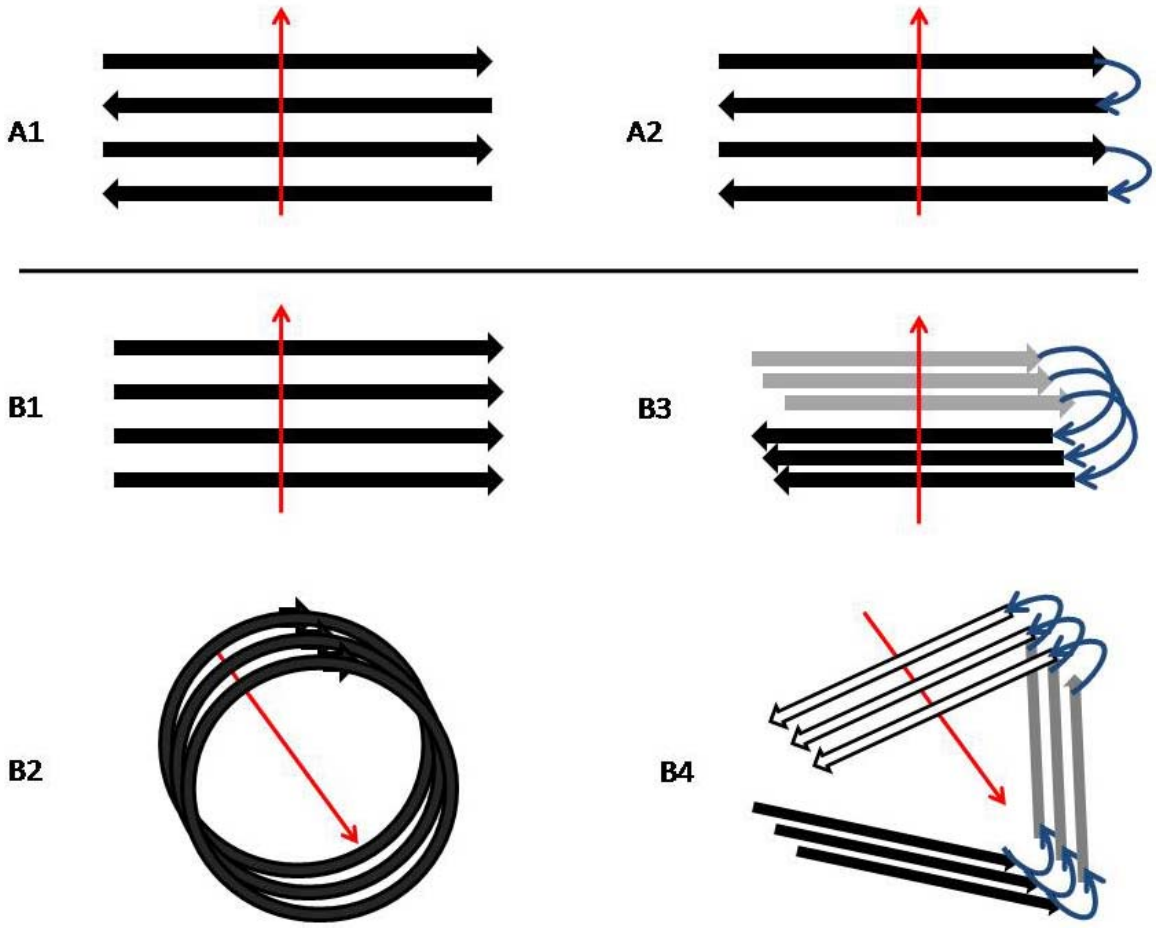

Figure 1. Possible arrangements of beta-strands in protofilaments of amyloid fibrils. A: anti-parallel arrangement. A1, inter-antiparallel; A2, intra- antiparallel. B: parallel arrangement. B1, extended parallel sheets; B2, circular parallel sheets; B3, one-turn parallel sheets; B4, two-turn parallel sheets; Red arrows represent the direction of the fibrils.

(21), one turn or bend (residues 25-29) (17, 22), to two turns (residues 22-23 and 29-30) (23, 24). Other than the defining feature, cross-beta motif, of amyloid fibrils, current experimental observations do not support any particular proposed amyloid fibril structural model (10, 11). In this review, we describe computational models of amyloid fibrils using both non-template and template-based techniques.

4. COMPUTATIONAL STUDY OF AMYLOID FIBRIL MODELS

4.1. Non-template-based modeling approaches

Non-template-based fibril modeling approaches are mainly used for studying fibril models formed from short peptides. Unlike modeling amyloid fibril structures formed by relatively long or full-length peptides, there is no need to concern about the number and the locations of turns in fibril models from short peptides. It is relatively easier to construct all the possible arrangements of such short peptides based on the typical cross-beta feature, about five angstroms between two hydrogen-bonded strands within a sheet and about ten angstroms between two sheets. Assisted with experimental data, fibrils formed from longer peptides can also be modeled using this approach. Basically, this approach consists of two main steps: 1) design and construct possible fibril models; and 2) assess the structural stability under various conditions and analyze the detailed interactions inside the model using molecular dynamics simulation. Below we briefly introduce research works done using this approach.

4.1.1. Amyloid fibrils formed from short peptides

Short peptides (4-8 residues) from disease-related proteins can form amyloid fibrils *in vitro* (10, 13). Though they only represent a portion of the disease peptides, the fibrils formed from these short peptides share similar fibril characteristics and are ideal for studying the underlying chemical principles of their spatial arrangement and atomic interactions using computational methods. Nussinov and colleagues have done some pioneering studies on oligomers from short amyloid peptides. These include an octomer (residues ^{113}AGAAAAGA^{120}) derived from the Syrian hamster prion protein (25), a short fragment from Abeta-peptide (residues ^{16}KLVFFAE^{22}) (22), a peptide from human calcitonin (residues ^{15}DFNKF^{19}) (26), and two

peptides from human islet amyloid polypeptide, hIAPP (residues ^{22}NFGAIL^{27} and ^{22}NFGAILSS^{29}) (27, 28). Through comparative studies, they found some stable amyloid structural conformations. For example, Abeta ^{16}KLVFFAE^{22} prefers an anti-parallel beta-sheet structure while the parallel beta-sheet arrangement is less stable, which is consistent with the solid-state NMR data (22, 29).

Not only can the above approach be used to probe the stability of the fibrils and to provide insights to the underlying chemical interaction patterns, it can also be applied to study the effects of different lengths and amino acid compositions (or mutations) on the stability of the structures. Their comparative study on two hIAPP peptides, ^{22}NFGAIL^{27} and ^{22}NFGAILSS^{29} revealed that the longer peptide with two additional serine residues at the C-terminal does not seem to have an effect on the overall structure, an anti-parallel conformation within sheets and parallel association between sheets. However subtle differences have been observed in the detailed structural arrangement (27, 28). They suggested that these subtle changes are a result of maintaining the number of hydrophobic interactions. The effects of mutations on the aggregate structure can also be investigated using simulation techniques. For example, systematic mutation experiments *in silico* provide an effective tool to study the contributions of single amino acids to the overall structural stability, which is convincingly demonstrated in the alanine scanning studies of hIAPP ^{22}NFGAIL^{27}. The results from the *in silico* alanine scanning of hIAPP ^{22}NFGAIL^{27} are in high agreement with the experimental observations (13). Though computational studies on short peptides described here provide us new insights about the chemical interactions and possible structural arrangements, there is a need to model amyloid fibril structures formed by longer or full-length peptides to provide insightful information about fibrils from disease tissues.

4.1.2. Amyloid fibrils formed from longer peptides

A similar approach to the one outlined above can also be used to model fibril structures formed from long and full-length amyloid peptides. Due to the existence of possible turns in the long peptide models and the bigger size of the systems, it is not practical to try all possible arrangements of strand association as in short peptide oligomer studies. Instead, modeling of fibrils structures of long peptides relies more on the experimental observations. For example, new evidence from solid state-NMR and liquid suspension EPR studies on full-length Abeta fibrils suggests peptides in the fibril core have an in-register, parallel arrangement (17, 18), even though anti-parallel strand arrangements are the dominant form in fibril models formed from short peptides as suggested in both computational and experimental studies. Accordingly, a number of models involving parallel beta-sheets have been proposed. Lakdawala *et al.* proposed a fully extended linear parallel model for Abeta(10-35) in 2002 (21). Through comparative studies, Ma and Nussinov found that the double-layered parallel hairpin model, in which a salt-bridge between D23 and K28 in the same strand, has higher stability while the fully extended linear parallel model cannot survive the high temperature, based on molecular dynamics simulation studies at 330K, suggesting that a bend with a salt-bridge D23-K28 can stabilize the fibril structure (22). The key features of the model, including the side-chain orientations of N-terminal strand and the D23-K28 salt-bridge, are consistent with experimental observations from solid-state NMR studies (17). The difference is in the C-terminal with I32- L34 facing inside in the Ma-Nussinov model and I32-L34 pointing outside in the model based on solid-state NMR experiment (17). Further experimental data on the quaternary structure of Abeta(1-40) indicate residue contacts between sidechains of L17 and F19 and sidechains of I32, L34, and V36, which is consistent with the Ma-Nussinov model (30). This model has been tentatively named the Ma-Nussinov-Tycko model (31). This example highlights the effectiveness and usefulness of computational studies in modeling amyloid fibril structures.

4.2. Template-based modeling approaches

The computational approach discussed above does not utilize any known protein structures as templates when constructing models. While computational structure predictions, especially template-based prediction methods, have been used extensively for soluble protein structure prediction, their application to misfolded structures and protein aggregates has been limited. This is partly due to the fact that in most cases the peptides or proteins adopt different conformations in soluble and fibril states. Given recent suggestions that fibrils are stabilized by forces common to all proteins, *i.e.*, hydrophobic interactions and hydrogen bonding, but not by forces particular to a specific sequence (32-34), template-based prediction methods, especially protein threading approach that will be introduced next, should be useful in modeling aggregate structures. In addition, the increasing number of solved protein structures has extended the coverage of protein folds. As of now, tens of thousands of protein structures have been solved experimentally and deposited into the Protein Data Bank (PDB) (35), providing a rich source of structures and sub-structures that could be used for structural modeling of amyloid fibrils. More importantly, proteins with cross-beta features might hold the key to understanding the folding pattern in amyloid fibrils (36, 37). For example, the parallel beta-helical fold fulfills the basic requirements for an underlying primordial structure of amyloid fibrils, such as the intrinsic cross-beta structure and the main chain hydrogen bonding. Protein threading seems to be a natural fit for amyloid fibril structure modeling. In this section, we introduce the protein threading method first and then review the applications of protein threading in modeling the structures of normally folded proteins as well as amyloid fibril structures.

4.2.1. Protein threading

Generally speaking, protein structure prediction techniques fall into three categories: *ab initio* prediction, protein threading (sometimes also referred as *fold recognition*), and homology modeling. *Ab initio* approach makes structure predictions based on the first principles of physics. It does not use any structural information of previously solved protein structures. Structure prediction by homology modeling is based on accurate sequence

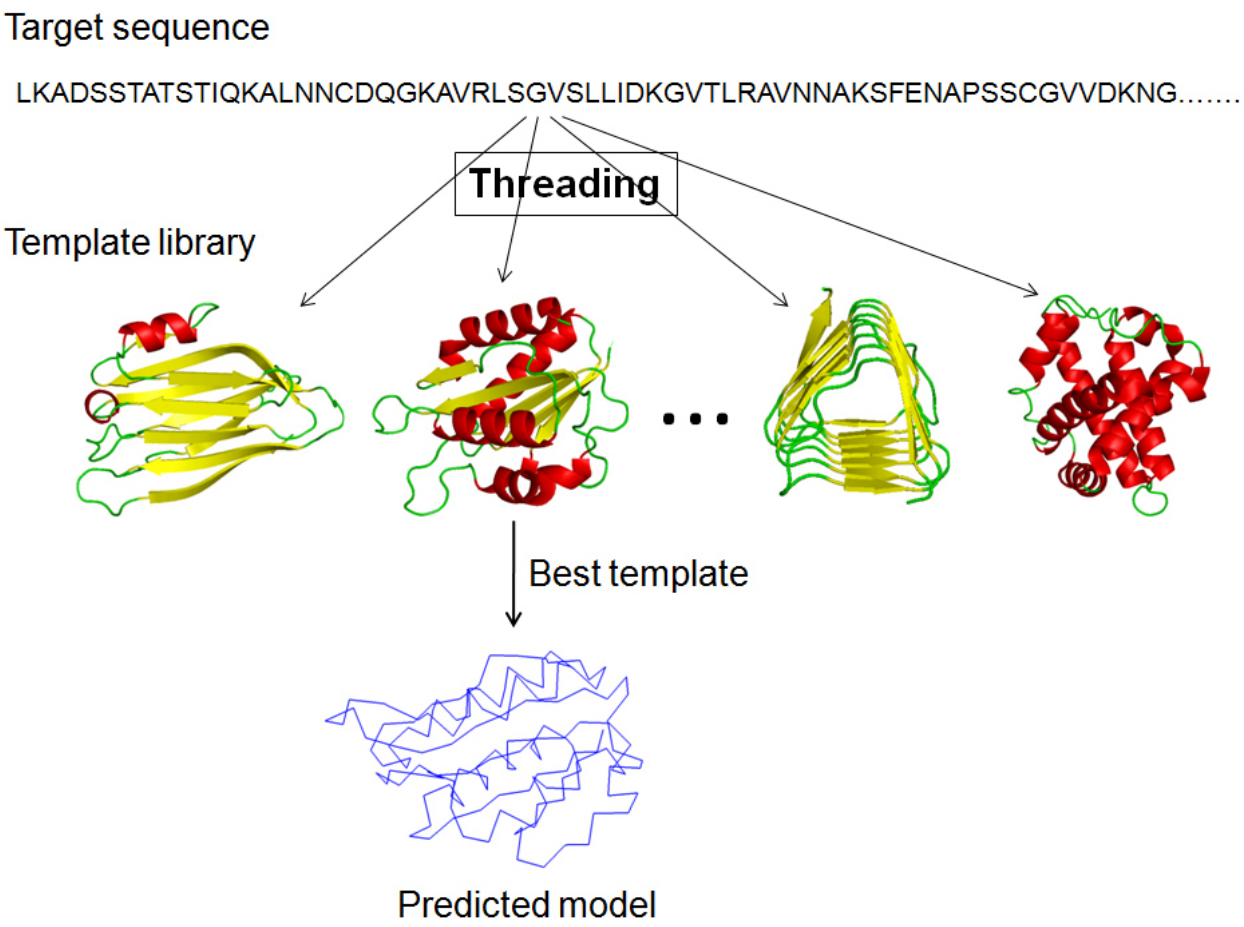

Figure 2. Schematic representation of protein threading approach. Target sequence is the sequence of a protein for modeling.

alignments between a target protein and a template protein structure; hence the prediction accuracy of this class of methods heavily depends on the sequence similarity between two proteins. Protein threading, introduced in the early 90's (38, 39), represents a more general class of prediction techniques than homology modeling as it uses both sequence similarity information if available and structural fitness information between the target protein and the template structure. In principle, protein threading does not require sequence similarity to find structural templates based on two assumptions: 1) many unrelated sequences fold into similar structures (40); and 2) some intrinsic interaction patterns between the residues of stable protein structures contribute to the specific folding pattern. The basic idea of protein threading is to "thread", literally, the amino acids of a target protein, preserving their sequential order and allowing for insertions and deletions, into structural positions of a template structure in such a way that the total energy defined by a scoring function is minimized. The quality of a sequence-structure alignment is typically assessed using a number of statistics-based energy terms or physical-based energies. This procedure is repeated for each structure in a template library and the "best" sequence-structure alignment of all the structural templates provides a prediction of the backbone atoms of the query protein (Figure 2). The assessment of protein threading programs is based on two important measures: 1) fold recognition, *i.e.* can a threading program find the correct template structure? and 2) sequence-structure alignment, *i.e.* given a correct template, can the target sequence be aligned to the template structure correctly?

A unique feature of threading approaches is that it applies structural information from the template proteins. Earlier threading programs use one-dimensional (1D) structural environments (secondary structure and solvent accessibility) to describe the template structures. For example, in their seminal paper, Bowie *et al.* defined eighteen different structural environments in terms of secondary structure and solvent accessibility, and represented a template structure as a 1D sequence of structural environments (38). Hence a threading problem is essentially to find an optimal alignment between a target sequence and a sequence of structural environments, which can be solved using an efficient dynamic programming approach. Despite the obvious computational advantages by representing a template 3D structure as a 1D string, it was shown that the threading performance can be improved by including residue-residue pair-wise interaction energy (41). A number of approximate and exact algorithms, such as, double dynamic programming, divide-and-conquer,

inter programming, and tree decomposition, have been developed to utilize the residue pair interactions (12).

Even though theoretically protein threading relies on interaction patterns and does not need sequence information for prediction, in practice, most threading programs utilize sequence information in a variety of ways as the prediction performance can be improved to some degree by incorporating protein sequence information. One of the most useful ways is to use evolutionary information. It has been observed that it improves the threading accuracy by using all homologous sequences of a query protein instead of using the query sequence alone during threading (42). One simple way to incorporate the evolutionary information is through using the sequence profile, derived from the aligned homologous sequences, hence generalizing the sequence-structure alignment problem to a sequence profile-structure alignment problem. It has been shown that the profile-profile alignment approach, which is done through comparing the sequence profiles of the query protein and the template protein, is very successful in detecting distantly related homologues (43). Govaerts *et al.* used a similar approach when modeling the PrP27-30 fibril structures (44).

4.2.2. Threading approach in non-aggregate protein structure modeling

The performance of protein threading programs has been assessed through the biennial, community-wide blind experiments, CASP (Critical Assessment of Structure Prediction), which started in 1994 (45). Threading-based prediction methods have performed well in the past several CASP contests (46, 47). In addition to successful applications in genome-scale predictions and in selecting targets for protein structural genomics (48-50), protein threading has also demonstrated its usefulness in modeling important and interesting proteins. For example, the first three-dimensional working model of human vitronectin, an adhesive protein that circulates in high concentrations in plasma, was predicted through a combination of computational methods, specifically protein threading and domain docking, and experimental observations (51). The predicted model is consistent with all known experimental observations, including positioning of the ligand binding sites, accessibility of protease cleavage sites (51), and data from small-angle scattering experiments (52). Using a combination of computational methods and biophysical analysis, Arluison *et al.* produced a model for Hfq of *Escherichia coli*., (please delete the "." after coli) a nucleic-binding protein, using threading techniques (53). Previous studies have shown that Hfq is involved in the regulation of many cellular proteins. Through threading analysis, the authors found that the Sm motif of small nuclear ribonucleoproteins is a good fit for Hfq. The derived model, which provide valuable information on possible RNA interaction residues, was validated through a series of biophysical and biochemical studies (53). Partial experimental data, especially distance related information, can be used as constraints in threading for a better model prediction. Young *et al.* demonstrated the effectiveness of using lysine cross-links as constraints on threading and its potential in high-throughput protein fold recognition (54). The distance constraints from using lysine-specific cross-linking agent, and time-of-flight mass spectrometry were used to assess the quality of threading models. In their study, they correctly identified FGF-2 as a member of the β-trefoil fold family.

Another application of protein threading has been demonstrated in a recent study in protein isoform structure modeling by Wang *et al.* (55). A large-scale analysis of known alternatively spliced variants at both protein sequence and structure levels reveals that the boundaries of alternative splicing events generally occur in exposed areas as well as in coil regions of secondary structures, and majority of the sequences involved in splicing are located on the surface of proteins. These findings suggest that threading, in general, provides a viable approach to modeling structures of alternatively spliced variants. Molecular dynamics simulation is then used to assess the fold stability of threading-based structure prediction (55).

Not only is the protein threading a popular method for soluble protein structure prediction, recently it has also been successfully applied to predict membrane protein structures (56, 57). Zhang *et al.* employed their threading assembly refinement program, TASSER, to predict the structures of over nine hundred putative G protein-coupled receptors (GPCRs) in the human genome. Based on their confidence measure derived from a benchmark set, the majority of the predicted models are considered to have the correct structural folds. To test the accuracy of predicted models, they selected several representative GPCRs and compared the models with experimental data from mutagenesis and affinity labeling. They found that the predicted models are consistent with the experimental observations (57). More importantly, they modeled the structure of bovine rhodopsin by excluding homologous template structures. TASSER produced a model with an RMSD (C-alpha) of 4.6 angstroms for the whole protein and 2.1 angstroms in transmembrane helix region when compared to the native structure and transmembrane helix region. The results highlight the power of protein threading and the usefulness of the computational models in studying important biological molecules.

4.2.3. Protein threading in amyloid fibril structure modeling

Despite protein threading is an effective approach in modeling soluble protein structures; its application to aggregate structure prediction has been limited. Depending on how a structural template is selected, there are two general threading-related approaches, implicit threading and explicit threading, in amyloid fibril structure modeling. In implicit threading, the peptide sequence can be mapped to a known structure that fits a proposed model. In 1999, Li *et al.* used an implicit threading approach to construct a twisted model of Abeta amyloid protofilaments based on limited experimental data that suggests Abeta may form an anti-parallel beta-sheet with a turn located around residues 25-28 in addition to the general cross-beta features. To construct the protofilament model, they selected the high-resolution structure of TTR (PDB ID: 2pab) (58) as the

template to construct the basic building block, a dimer of an anti-parallel beta-sheet with a turn located at residues 25-28 for Abeta(12-42) as the template fits most of the fibril structural features available at that time (20). In their model, 48 monomers (the basic building block) of Abeta(12-42) stack with four monomers per layer to form a twisted helical turn of beta-sheet. Molecular dynamics simulations of the protofilament model in explicit aqueous solution suggests that the observed twist in synchrotron x-ray studies might be the result of protofilament packing, rather than from the structure of individual protofilaments. In this study, the "threading"-sequence-structure alignment was done manually without using any threading energy or algorithms.

The first modeling of amyloid fibril structure using a threading algorithm (explicit threading) is the one generated by Chaney MO *et al.* (19). The threading program they used is TOPITS, which uses a 1D string of secondary structures and relative solvent accessibility to describe the structural templates (59). The structural environments of the residues in the target protein are predicted by a neural network approach. The predicted 1D string is then aligned to that of each protein template using an efficient dynamic programming algorithm. Using TOPITS, they identified three possible templates that share an anti-parallel beta-sheet structure for Abeta(1-42). The resulting model of Abeta(1-42) from threading studies has four anti-parallel beta-strands (1-6, 9-15, 18-24, and 29-36). In constructing the Abeta protofilament model, they argued that two Abeta molecules should form a dimer in order to shield unfavorable hydrophobic domains from the aqueous environment and proposed that the C-terminal domain (residues 30-42) of each dimer extend toward the center to form an anti-parallel beta-sheet with the other Abeta dimer. The fibril model consisting of three protofilaments was generated in an attempt to resolve the thermodynamically unfavorable twisted beta-sheet, which is highly hydrophobic but yet is exposed to an aqueous environment (19).

Though the aforementioned template-based fibril models are quite different, they do have one common feature, a core structure containing anti-parallel beta-sheets, either inter-molecular or intra-molecular. Since the introduction of the above anti-parallel models, more experimental data have been generated, painting a different picture of the fibril structure formed by long or full-length amyloid peptides/proteins from the ones proposed earlier. Firstly, strong evidence from solid-state NMR and liquid suspension EPR studies on amyloid fibrils formed from Abeta(1-40), a 37-residue islet amyloid peptide and α-synuclein suggests that the strands in fibril core are in-register and arranged in a parallel fashion (16-18, 60, 61). Secondly, a number of experimental studies using proline scanning mutagenesis, EPR, hydrogen-deuterium exchange, limited proteolysis, and solid-state NMR have suggested that not all of the residues are part of the highly structured core region, rather some of the N-terminal and C-terminal regions are less ordered (17, 18, 60-63). For example, recent proline scanning mutagenesis experiments, a technique used to search for regions involved in turns and disordered structure, have provided valuable information regarding the possible turn regions and non-core residues (23).

Based on newly generated experimental data, Guo *et al.* constructed a structural model for the Abeta amyloid fibril core structure using a threading technique and molecular dynamics simulation (24). Proline scanning mutagenesis experiments on Abeta(1-40) fibrils suggest that only part of Aβ(1-40), residues 16-35, is involved in the amyloid core formation. The experiment also indicated that there are three possible beta strands separated by two turns at residues 22-23 and 29-30 (23). These data suggested that the fibril structure might fit well with an existing folding pattern, the parallel beta-helical motif. The ideas of amyloid fibril being beta-helical like structure have been previously proposed for Abeta, insulin fibrils and polyglutamine fibrils (37, 64, 65). In constructing the fibril core structural model, Abeta(15-36) was threaded against a set of representative parallel beta-helical proteins, both left-handed and right-handed, and several non-beta-helical all-beta proteins as controls using an in-house threading program PROSPECT, which considers residue-residue pair-wise potential (24). The top ranked sequence-structure alignments using a threading program were consistent with the proline scanning mutagenesis data with respect to the locations of the turns in beta-strands. The non-beta-helical templates, including beta-hairpin structures, did not score as well as beta-helical proteins. Based on the threading alignments, both left-handed and right-handed 6-mer models were generated as the core of protofilaments and were subjected to molecular dynamics simulations to assess the stability of the models. The simulation results revealed that the left-handed model is more stable than the right-handed model (Figure 3). Most importantly, the total number of hydrogen bonds in the left-handed model during simulation is in agreement with the H-D exchange experiments (24). The left-handed beta-helical protofilament model with triangular cross section can explain many experimental observations except that it fails to account for the 10 angstroms equatorial reflections in X-ray diffraction data on Abet fibrils. A fibril model with a (please delete "a" from the sentence) 10 angstroms between sheets from neighboring protofilament has been proposed (24).

The threading approach has also been applied for the modeling of prion fibril structures (44). Govaerts *et al.* analyzed the position-specific residue preferences or profiles of 18 right-handed and 9 left-handed beta-helical proteins (48). Their analysis suggests that the sequence of PrP27-30 is compatible with a parallel left-handed beta-helical fold and the right-handed beta-helical fold is not a suitable template. In their approach to construct a model, residues 89-174 of PrP27-30 are threaded onto the structure of the beta-helical domain of uridyltransferase (PDB code: 1G97), resulting in four rungs of beta helices. The alpha-helical region of residues 177-227, solved by NMR spectroscopy, is arranged in the beta-helix fold in a similar way to other alpha-helices in known beta-helical protein structures. In addition, the density map data from the 2D crystals were used to optimize the exact position of the α-

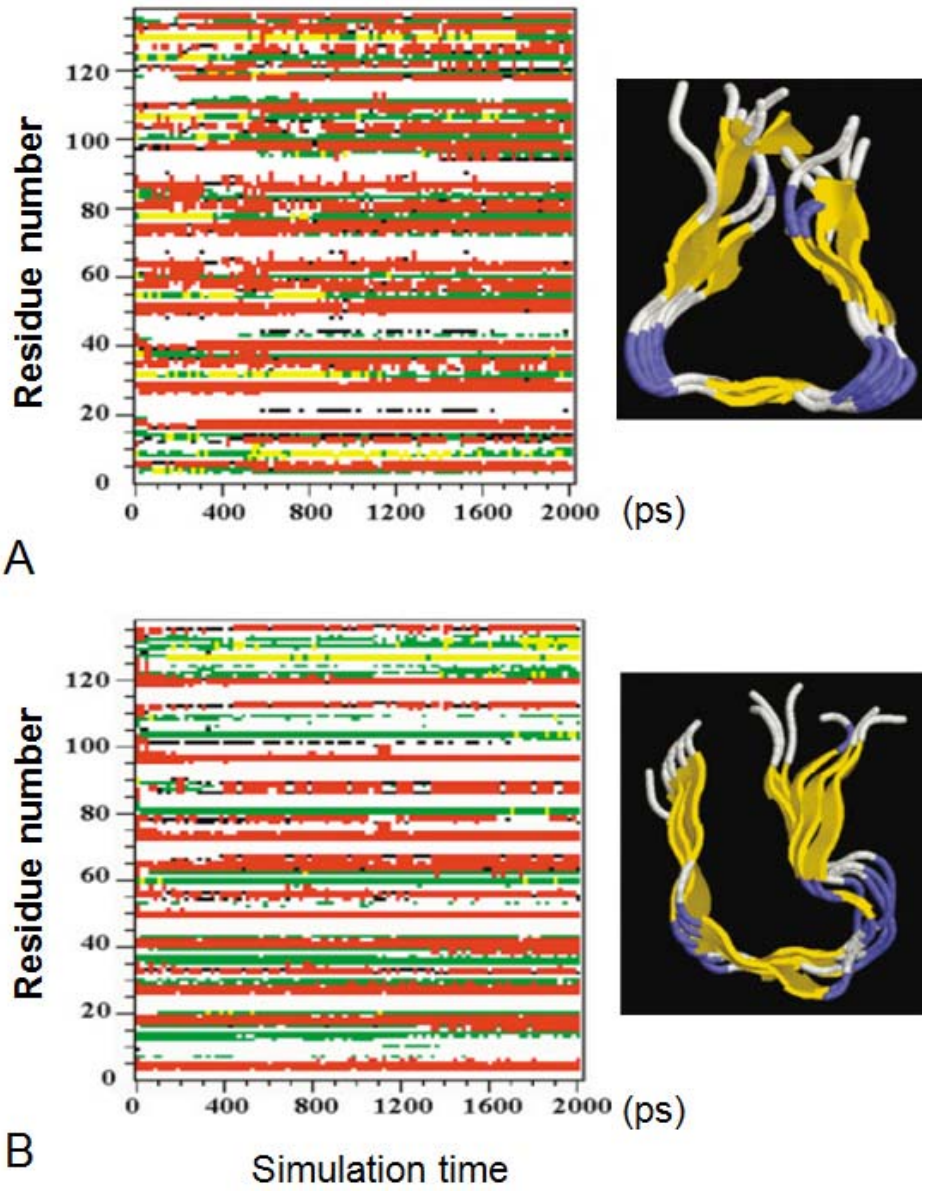

Figure 3. Secondary structure changes during molecular dynamics simulations of two different models. A. left-handed beta-helical model; B. Right-handed beta-helical model. Reproduced with permission from Wiley. (please remove this sentence since no permission is needed).

helices in the final structure. The trimeric model from the packing of three parallel left-handed beta-helical monomers that matches the structural constraints of the PrP27-30 crystals has also been proposed (44).

5. SUMMARY AND PERSPECTIVE

Two major computational techniques have been applied to studying aggregate and fibril structures, model construction and molecular dynamics simulation. The models can be generated either manually or through protein structure prediction programs. One of the limitations in using template-based approach is the incomplete coverage of protein structural folds. However, with the current rate of protein structure determination and the increasing number of structural folds in PDB, protein threading is set to play an even bigger role in modeling amyloid fibril structures. One challenging issue facing the protein structure prediction community is the accuracy of sequence-structure alignments, which in turn affects the accuracy of the model even though a correct fold is identified. Improved energy functions and efficient algorithms that can utilize more complicated energy functions will be keys to the future development of threading programs.

Structural polymorphism in amyloid fibrils presents a great challenge to computational modeling of

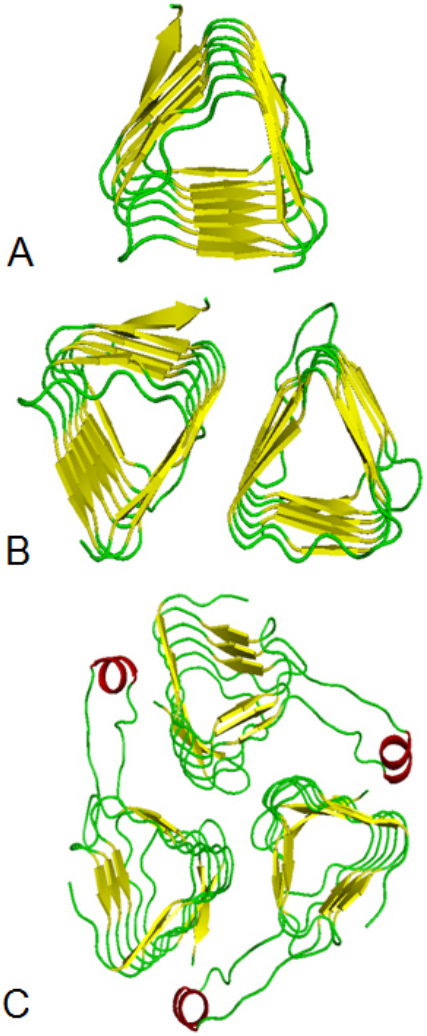

Figure 4. Examples of monomer, dimer, and trimer beta-helical template structures. A. Choristoneura Fumiferana Antifreeze Protein (PDB: 1M8N); B. A dimer structure of 1M8N; C. A trimer E. coli maltose-o-acetyltransferase (PDB: 1OCX). The images are generated using PyMOL(http://www.pymol.org).

amyloid fibrils (11). At the same time, it also presents an opportunity to computational structural biologists as computational approaches might be the best tools in providing new insights about structural variations of the amyloid fibrils. The structural polymorphism can be studied at two different levels, the protofilament structure and the packing of protofilaments. To date, threading method has been applied to predict protofilament structures using monomer structural templates. Fibril models are then proposed based on the protofilament structures and the possible arrangements from low resolution data, such as 2D crystal data. Recent advances in developing algorithms for quaternary structure prediction might be useful to fibril structure modeling using multimers as threading templates (Figure 4)(66). The development of energy functions that can accurately describe residue interactions within and among structural units will have a great bearing in quaternary structure threading using structural templates with at least two subunits (Figure 4). In addition, the interplay between experimental methods and computational approaches has clearly demonstrated the power of an integrated method in studying complex structural models like protein aggregation (31). More powerful prediction techniques and collaborations between experimental and the computational efforts will for sure provide more accurate structural models and hence help better understand the structural details of amyloid fibrils and the mechanism of amyloidoses.

6. ACKNOWLEDGEMENTS

This work is supported by the National Institutes of Health (AG #2813096) and in part by National Science Foundation (#NSF DBI-0354771, #NSF DBI-0542119, #NSF CCF-0621700, and NSF IIS-0407204), National Institutes of Health (NIH -1R01GM075331-01A2), and the startup fund to JTG from the University of North Carolina at Charlotte.

7. REFERENCES

1. P. Westermark: Aspects on human amyloid forms and their fibril polypeptides. *Febs J*, 272, 5942-9(2005)

2. M. Fandrich, C. M. Dobson: The behaviour of polyamino acids reveals an inverse side chain effect in amyloid structure formation. *Embo J*, 21, 5682-90(2002)

3. M. Fandrich, M. A. Fletcher, C. M. Dobson: Amyloid fibrils from muscle myoglobin. *Nature*, 410, 165-6(2001)

4. R. W. Pickersgill: A primordial structure underlying amyloid. *Structure (Camb)*, 11, 137-8(2003)

5. W. T. Astbury, S. Dickinson: The X-ray interpretation of denaturation and the structure of the seed globulins. *Biochem J*, 29, 2351-2360 1(1935)

6. E. D. Eanes, G. G. Glenner: X-ray diffraction studies on amyloid filaments. *J Histochem Cytochem*, 16, 673-7(1968)

7. M. Sunde, C. Blake: The structure of amyloid fibrils by electron microscopy and X-ray diffraction. *Adv Protein Chem*, 50, 123-59(1997)

8. D. Eisenberg, R. Nelson, M. R. Sawaya, M. Balbirnie, S. Sambashivan, M. I. Ivanova, A. O. Madsen, C. Riekel: The structural biology of protein aggregation diseases: Fundamental questions and some answers. *Acc Chem Res*, 39, 568-75(2006)

9. R. Nelson, M. R. Sawaya, M. Balbirnie, A. O. Madsen, C. Riekel, R. Grothe, D. Eisenberg: Structure of the cross-beta spine of amyloid-like fibrils. *Nature*, 435, 773-8(2005)

10. M. R. Sawaya, S. Sambashivan, R. Nelson, M. I. Ivanova, S. A. Sievers, M. I. Apostol, M. J. Thompson, M. Balbirnie, J. J. Wiltzius, H. T. Mcfarlane, A. O. Madsen, C. Riekel, D. Eisenberg: Atomic structures of amyloid cross-

beta spines reveal varied steric zippers. *Nature*, 447, 453-7(2007)

11. R. Tycko: Progress towards a molecular-level structural understanding of amyloid fibrils. *Curr Opin Struct Biol*, 14, 96-103(2004)

12. J. T. Guo, K. Ellrott, Y. Xu: A historical perspective of template-based protein structure prediction. In: Protein Structure Prediction, Methods and Protocols. Eds: M. Zaki&C. Bystroff. Humana Press, (2007)

13. D. Zanuy, K. Gunasekaran, B. Ma, H. H. Tsai, C. J. Tsai, R. Nussinov: Insights into amyloid structural formation and assembly through computational approaches. *Amyloid*, 11, 143-61(2004)

14. C. K. Hall, V. A. Wagoner: Computational approaches to fibril structure and formation. *Methods Enzymol*, 412, 338-65(2006)

15. B. Urbanc, J. M. Borreguero, L. Cruz, H. E. Stanley: Ab initio discrete molecular dynamics approach to protein folding and aggregation. *Methods Enzymol*, 412, 314-38(2006)

16. T. L. Benzinger, D. M. Gregory, T. S. Burkoth, H. Miller-Auer, D. G. Lynn, R. E. Botto, S. C. Meredith: Propagating structure of Alzheimer's beta-amyloid(10-35) is parallel beta-sheet with residues in exact register. *Proc Natl Acad Sci U S A*, 95, 13407-12(1998)

17. A. T. Petkova, Y. Ishii, J. J. Balbach, O. N. Antzutkin, R. D. Leapman, F. Delaglio, R. Tycko: A structural model for Alzheimer's beta -amyloid fibrils based on experimental constraints from solid state NMR. *Proc Natl Acad Sci U S A*, 99, 16742-7(2002)

18. M. Torok, S. Milton, R. Kayed, P. Wu, T. Mcintire, C. G. Glabe, R. Langen: Structural and dynamic features of Alzheimer's Abeta peptide in amyloid fibrils studied by site-directed spin labeling. *J Biol Chem*, 277, 40810-5(2002)

19. M. O. Chaney, S. D. Webster, Y. M. Kuo, A. E. Roher: Molecular modeling of the Abeta1-42 peptide from Alzheimer's disease. *Protein Eng*, 11, 761-7(1998)

20. L. Li, T. A. Darden, L. Bartolotti, D. Kominos, L. G. Pedersen: An atomic model for the pleated beta-sheet structure of Abeta amyloid protofilaments. *Biophys J*, 76, 2871-8(1999)

21. A. S. Lakdawala, D. M. Morgan, D. C. Liotta, D. G. Lynn, J. P. Snyder: Dynamics and fluidity of amyloid fibrils: a model of fibrous protein aggregates. *J Am Chem Soc*, 124, 15150-1(2002)

22. B. Ma, R. Nussinov: Stabilities and conformations of Alzheimer's beta -amyloid peptide oligomers (Abeta 16-22, Abeta 16-35, and Abeta 10-35): Sequence effects. *Proc Natl Acad Sci U S A*, 99, 14126-31(2002)

23. A. Williams, E. Portelius, I. Kheterpal, J-T. Guo, K. Cook, Y. Xu, R. Wetzel: Mapping abeta amyloid fibril secondary structure using scanning proline mutagenesis. *J Mol Biol*, 335, 833-842(2004)

24. J. T. Guo, R. Wetzel, Y. Xu: Molecular modeling of the core of Abeta amyloid fibrils. *Proteins*, 57, 357-64(2004)

25. B. Ma, R. Nussinov: Molecular dynamics simulations of alanine rich beta-sheet oligomers: Insight into amyloid formation. *Protein Sci*, 11, 2335-50(2002)

26. N. Haspel, D. Zanuy, B. Ma, H. Wolfson, R. Nussinov: A comparative study of amyloid fibril formation by residues 15-19 of the human calcitonin hormone: a single beta-sheet model with a small hydrophobic core. *J Mol Biol*, 345, 1213-27(2005)

27. D. Zanuy, B. Ma, R. Nussinov: Short peptide amyloid organization: stabilities and conformations of the islet amyloid peptide NFGAIL. *Biophys J*, 84, 1884-94(2003)

28. D. Zanuy, R. Nussinov: The sequence dependence of fiber organization. A comparative molecular dynamics study of the islet amyloid polypeptide segments 22-27 and 22-29. *J Mol Biol*, 329, 565-84(2003)

29. J. J. Balbach, Y. Ishii, O. N. Antzutkin, R. D. Leapman, N. W. Rizzo, F. Dyda, J. Reed, R. Tycko: Amyloid fibril formation by A beta 16-22, a seven-residue fragment of the Alzheimer's beta-amyloid peptide, and structural characterization by solid state NMR. *Biochemistry*, 39, 13748-59(2000)

30. A. T. Petkova, W. M. Yau, R. Tycko: Experimental constraints on quaternary structure in Alzheimer's beta-amyloid fibrils. *Biochemistry*, 45, 498-512(2006)

31. B. Ma, R. Nussinov: Simulations as analytical tools to understand protein aggregation and predict amyloid conformation. *Curr Opin Chem Biol*, 10, 445-52(2006)

32. M. Bucciantini, E. Giannoni, F. Chiti, F. Baroni, L. Formigli, J. Zurdo, N. Taddei, G. Ramponi, C. M. Dobson, M. Stefani: Inherent toxicity of aggregates implies a common mechanism for protein misfolding diseases. *Nature*, 416, 507-11(2002)

33. R. Kayed, E. Head, J. L. Thompson, T. M. Mcintire, S. C. Milton, C. W. Cotman, C. G. Glabe: Common structure of soluble amyloid oligomers implies common mechanism of pathogenesis. *Science*, 300, 486-9(2003)

34. A. D. Williams, M. Sega, M. Chen, I. Kheterpal, M. Geva, V. Berthelier, D. T. Kaleta, K. D. Cook, R. Wetzel: Structural properties of Abeta protofibrils stabilized by a small molecule. *Proc Natl Acad Sci U S A*, 102, 7115-20(2005)

35. H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, P. E. Bourne: The Protein Data Bank. *Nucleic Acids Res*, 28, 235-42(2000)

36. J. Jenkins, R. Pickersgill: The architecture of parallel beta-helices and related folds. *Prog Biophys Mol Biol*, 77, 111-75(2001)

37. R. Wetzel: Ideas of order for amyloid fibril structure. *Structure (Camb)*, 10, 1031-6(2002)

38. J. U. Bowie, R. Luthy, D. Eisenberg: A method to identify protein sequences that fold into a known three-dimensional structure. *Science*, 253, 164-70(1991)

39. D. T. Jones, W. R. Taylor, J. M. Thornton: A new approach to protein fold recognition. *Nature*, 358, 86-9(1992)

40. M. J. Sippl, H. Flockner: Threading thrills and threats. *Structure*, 4, 15-9(1996)

41. Y. Xu, D. Xu: Protein threading using PROSPECT: design and evaluation. *Proteins*, 40, 343-54(2000)

42. D. Kim, D. Xu, J. T. Guo, K. Ellrott, Y. Xu: PROSPECT II: protein structure prediction program for genome-scale applications. *Protein Eng*, 16, 641-50(2003)

43. K. Ginalski, N. V. Grishin, A. Godzik, L. Rychlewski: Practical lessons from protein structure prediction. *Nucleic Acids Res*, 33, 1874-91(2005)

44. C. Govaerts, H. Wille, S. B. Prusiner, F. E. Cohen: Evidence for assembly of prions with left-handed beta-helices into trimers. *Proc Natl Acad Sci U S A*, 101, 8342-7(2004)

45. J. Moult, J. T. Pedersen, R. Judson, K. Fidelis: A large-scale experiment to assess protein structure prediction methods. *Proteins*, 23, ii-v(1995)

46. J. Xu, M. Li: Assessment of RAPTOR's linear programming approach in CAFASP3. *Proteins*, 53 Suppl 6, 579-84(2003)

47. Y. Zhang: Template-based modeling and free modeling by I-TASSER in CASP7. *Proteins*, 69, 108-117(2007)

48. L. J. Mcguffin, D. T. Jones: Improvement of the GenTHREADER method for genomic fold recognition. *Bioinformatics*, 19, 874-881(2003)

49. D. Kihara, J. Skolnick: Microbial genomes have over 72% structure assignment by the threading algorithm PROSPECTOR_Q. *Proteins*, 55, 464-73(2004)

50. I. Friedberg, L. Jaroszewski, Y. Ye, A. Godzik: The interplay of fold recognition and experimental structure determination in structural genomics. *Curr Opin Struct Biol*, 14, 307-12(2004)

51. D. Xu, K. Baburaj, C. B. Peterson, Y. Xu: Model for the three-dimensional structure of vitronectin: predictions for the multi-domain protein from threading and docking. *Proteins*, 44, 312-20(2001)

52. G. W. Lynn, W. T. Heller, A. Mayasundari, K. H. Minor, C. B. Peterson: A model for the three-dimensional structure of human plasma vitronectin from small-angle scattering measurements. *Biochemistry*, 44, 565-74(2005)

53. V. Arluison, P. Derreumaux, F. Allemand, M. Folichon, E. Hajnsdorf, P. Regnier: Structural Modelling of the Sm-like Protein Hfq from Escherichia coli. *J Mol Biol*, 320, 705-12(2002)

54. M. M. Young, N. Tang, J. C. Hempel, C. M. Oshiro, E. W. Taylor, I. D. Kuntz, B. W. Gibson, G. Dollinger: High throughput protein fold identification by using experimental constraints derived from intramolecular cross-links and mass spectrometry. *Proc Natl Acad Sci U S A*, 97, 5802-6(2000)

55. P. Wang, B. Yan, J. T. Guo, C. Hicks, Y. Xu: Structural genomics analysis of alternative splicing and application to isoform structure modeling. *Proc Natl Acad Sci U S A*, 102, 18920-5(2005)

56. S. J. Fleishman, N. Ben-Tal: Progress in structure prediction of alpha-helical membrane proteins. *Curr Opin Struct Biol*, 16, 496-504(2006)

57. Y. Zhang, M. E. Devries, J. Skolnick: Structure modeling of all identified G protein-coupled receptors in the human genome. *PLoS Comput Biol*, 2, e13(2006)

58. C. C. Blake, M. J. Geisow, S. J. Oatley, B. Rerat, C. Rerat: Structure of prealbumin: secondary, tertiary and quaternary interactions determined by Fourier refinement at 1.8 A. *J Mol Biol*, 121, 339-56(1978)

59. B. Rost: TOPITS: threading one-dimensional predictions into three-dimensional structures. *Proc Int Conf Intell Syst Mol Biol*, 3, 314-21(1995)

60. S. A. Jayasinghe, R. Langen: Identifying structural features of fibrillar islet amyloid polypeptide using site-directed spin labeling. *J Biol Chem*, 279, 48420-5(2004)

61. M. Chen, M. Margittai, J. Chen, R. Langen: Investigation of alpha-synuclein fibril structure by site-directed spin labeling. *J Biol Chem*, 282, 24970-9(2007)

62. I. Kheterpal, S. Zhou, K. D. Cook, R. Wetzel: Abeta amyloid fibrils possess a core structure highly resistant to hydrogen exchange. *Proc Natl Acad Sci U S A*, 97, 13597-601(2000)

63. I. Kheterpal, A. Williams, C. Murphy, B. Bledsoe, R. Wetzel: Structural features of the Abeta amyloid fibril elucidated by limited proteolysis. *Biochemistry*, 40, 11757-67(2001)

64. J. L. Jimenez, E. J. Nettleton, M. Bouchard, C. V. Robinson, C. M. Dobson, H. R. Saibil: The

protofilament structure of insulin amyloid fibrils. *Proc Natl Acad Sci U S A*, 99, 9196-201(2002)

65. M. F. Perutz, J. T. Finch, J. Berriman, A. Lesk: Amyloid fibers are water-filled nanotubes. *Proc Natl Acad Sci U S A*, 99, 5591-5(2002)

66. H. Chen, J. Skolnick: M-TASSER: An Algorithm for Protein Quaternary Structure Prediction. *Biophys J*, in press, (2007)

Abbreviations: AD: Alzheimer's disease; HD: hydrogen-deuterium exchange; vCJD: variant Creutzfeldt-Jakob disease; MD: molecular dynamics; NMR: Nuclear magnetic resonance; EPR: paramagnetic resonance spectroscopy; PDB: protein data bank; 1D: one-dimensional; CASP: Critical Assessment of Structure Prediction; GPCR: G protein–coupled receptor; RMSD: root-mean-squared deviation; DMD: discontinuous molecular dynamics

Key Words: Amyloid fibril, Protein threading, Modeling, Beta-helical motif, Molecular dynamics simulation, Review

Send correspondence to: Jun-tao Guo, Bioinformatics Research Center, CARC 218, the University of North Carolina at Charlotte, 9201 University City Blvd., Charlotte, NC 28223, USA, Tel: 704-687-7492, Fax: 704-687-6610, E-mail: jguo4@uncc.edu

http://www.bioscience.org/current/vol13.htm