

- Williams, A., Portelius, E., Kheterpal, I., Guo, J.-T., Cook, K., Xu, Y., and Wetzel, R. (2004). Mapping Abeta amyloid fibril secondary structure using scanning proline mutagenesis. *J. Mol. Biol.* **335**, 833–842.
- Xu, Y., and Xu, D. (2000). Protein threading using PROSPECT: Design and evaluation. *Proteins* **40**, 343–354.
- Yankner, B. A. (1996). Mechanism of neuronal degeneration in Alzheimer's disease. *Neuron* **16**, 921–932.
- Zanuy, D., Gunasekaran, K., Ma, B., Tsai, H. H., Tsai, C. J., and Nussinov, R. (2004). Insights into amyloid structural formation and assembly through computational approaches. *Amyloid* **11**, 143–161.

## [19] *Ab initio* Discrete Molecular Dynamics Approach to Protein Folding and Aggregation

By BRIGITA URBANC, JOSE M. BORREGUERO,  
LUIS CRUZ, and H. EUGENE STANLEY

### Abstract

Understanding the toxicity of amyloidogenic protein aggregates and designing therapeutic approaches require the knowledge of their structure at atomic resolution. Although solid-state NMR, X-ray diffraction, and other experimental techniques are capable of discerning the protein fibrillar structure, determining the structures of early aggregates, called oligomers, is a challenging experimental task. Computational studies by all-atom molecular dynamics, which provides a complete description of a protein in the solvent, are typically limited to study folding of smaller protein or aggregation of a small number of short protein fragments.

We review an efficient *ab initio* computer simulation approach to protein folding and aggregation using discrete molecular dynamics (DMD) in combination with several coarse-grained protein models and implicit solvent. This approach involves different complexity levels in both the protein model and the interparticle interactions. Starting from the simplest protein model with minimal interactions, and gradually increasing its complexity, while guided by *in vitro* findings, we can systematically select the key features of the protein model and interactions that drive protein folding and aggregation. Because the method used in this DMD approach does not require any knowledge of the native or any other state of the protein, it can be applied to study degenerative disorders associated with protein misfolding and aberrant protein aggregation.

The choice of the coarse-grained model depends on the complexity of the protein and specific questions to be addressed, which are mostly suggested by *in vitro* findings. Thus, we illustrate our approach on amyloid  $\beta$ -protein ( $A\beta$ )

associated with Alzheimer's disease (AD). Despite the simplifications introduced in the DMD approach, the predicted  $A\beta$  conformations are in agreement with existing experimental data. The *in silico* findings also provide further insights into the structure and dynamics of  $A\beta$  folding and oligomer formation that are amenable to *in vitro* testing.

## Introduction

An increasing number of neuropathological disorders, such as Alzheimer's, Parkinson's, Creutzfeldt-Jakob, motor neuron diseases, and polyglutamine disorders, are known to be associated with protein misfolding, followed by deposition of toxic protein aggregates in tissue (Dobson, 2004; Koo *et al.*, 1999). Alzheimer's disease (AD) is a progressive, neurodegenerative disorder, pathologically characterized by senile plaques and neurofibrillary tangles (Selkoe, 2001). The primary component of senile plaques is amyloid  $\beta$ -protein ( $A\beta$ ), which has been strongly linked to the etiology and pathogenesis of AD.  $A\beta$  aggregates into small assemblies (oligomers), protofibrils, and fibrils rich in  $\beta$ -sheet content. In the past decade, compelling evidence has emerged indicating that soluble oligomeric assemblies and protofibrillar intermediates that form before senile plaque deposition may be determinant pathogenetic factors (Klein *et al.*, 2004).

Determination of oligomer conformation at the atomic level and tracking pathways of assembly from monomers to oligomers requires efficient computational approaches. With the dramatic increase of computer power in recent decades, it has become possible to study the behavior of large biological molecular systems by computer simulations (Ash *et al.*, 2004; Feig and Brooks, 2004; Fersht and Daggett, 2002; Karplus and McCammon, 2002). Traditional, all-atom molecular dynamics (MD) with atomic-detail force fields in a physiological solution (which would be ideal for studying  $A\beta$  oligomerization) is not computationally accessible with current technology. An aggregation process amenable to study by all-atom MD should occur on time scales of  $\leq 10^{-7}$  sec and would require the use of advanced technologies such as worldwide distributed computing (Snow *et al.*, 2002; Zagrovic and Pande, 2003; Zagrovic *et al.*, 2002). However, *in vivo* and *in vitro* studies suggest that the initial stages of oligomerization occur on time scale of hours (Bitan *et al.*, 2003a; Kaye *et al.*, 2003).

The idea of applying a fast and efficient discrete molecular dynamics method (DMD) (Rapaport, 1997) to study protein folding was proposed in 1996 (Zhou *et al.*, 1996). Soon after, the method was combined with a one-bead protein Gō model to study folding of a model three-helix bundle protein (Dokholyan *et al.*, 1998, 2000; Zhou and Karplus, 1997, 1999; Zhou *et al.*, 1997). The interparticle interactions in the Gō model are assigned on the basis

of the knowledge of a native state of a protein (Taketomi *et al.*, 1975). Thus, Gō models are not *ab initio*, because they require the knowledge of a native state of a protein. Despite this drawback, they are the simplest thermodynamic models that yield a unique native state of a protein and describe folding reminiscent of a first-order phase transition (Dokholyan *et al.*, 2003). Recently, a two-bead Gō model was applied by Peng *et al.* to study aggregation of an ensemble of 28 A $\beta$ 40 peptides into a fibrillar structure (Peng *et al.*, 2004).

In 2001, a four-bead protein model in combination with the DMD method was first introduced by Smith and Hall (Smith and Hall, 2001a) inspired by earlier studies (Takada *et al.*, 1999). This model accounts for a rather accurate backbone description and is able to describe a cooperative transition of a polyaniline chain into an  $\alpha$ -helical conformation without any *a priori* knowledge of the native state (Smith and Hall, 2001b). With the four-bead model with hydrogen bond interactions on a single 16-residue polyaniline chain, Ding *et al.* (2003) demonstrated a temperature-induced conformational change from the  $\alpha$ -helix to the  $\beta$ -hairpin conformation. Because of these properties, the four-bead model with hydrogen bond interactions represents a base on which the *ab initio* modeling can be realized. The *ab initio* DMD computational approach introduces simplifications to the protein description, interparticle interactions, and treatment of the solvent. These simplifications make the DMD approach at least six orders of magnitude faster than all-atom MD with explicit solvent. To ensure biological relevance of the approach that targets different aspects of A $\beta$  folding, oligomer, and fibril formation, up-to-date experimental findings need to be integrated into the development of the protein model and interactions, creating a much-needed partnership between computation and experiment as recognized by others [e.g., the review by Ma and Nussinov (2004)].

This review is organized in two main sections. In the first section, we describe the applications of the DMD approach to model A $\beta$  folding and aggregation. The goal of this section is to give an idea of the kind of information we can obtain using the proposed DMD approach. In addition, we present the main hypotheses on the structure and dynamics of folding and assembly that emerged from these applications. The second section introduces in detail the implementation of the DMD method, coarse-grained protein models, interparticle interactions, and limitations that originate in simplifications associated with the approach.

### Applications to A $\beta$ Folding and Aggregation

We describe the applications of the DMD approach with the four-bead and the united-atom model to A $\beta$  folding and aggregation to demonstrate the variety of information that can be obtained. We also review selected

*in vitro* findings that shed light on different structural aspects of A $\beta$  folding and aggregation and help guide the development of the DMD approach.

### In Vitro Findings

The sequence DAEFRHDSGYEVHHQKLVFFAEDVGSNKGAIIGLMVGGVVIA defines the primary structure of A $\beta$ 42. A $\beta$ 40 lacks the last two amino acids, I41 and A42. The secondary structure of A $\beta$  monomer conformations depends strongly on the environment. In an apolar membrane-like environment, A $\beta$ 40 and A $\beta$ 42 monomers adopt predominantly an  $\alpha$ -helical conformation (Coles *et al.*, 1998; Crescenzi *et al.*, 2002), whereas in an aqueous solution A $\beta$  prefers a collapsed coil monomer structure with a bend in the V24-K28 region (Zhang *et al.*, 2000).

Recent limited proteolysis experiments on A $\beta$ 40 and A $\beta$ 42 have shown that the region A21–A30 is highly resistant to proteolytic attack under conditions favoring oligomerization, suggesting the presence of a folded structure (Lazo *et al.*, 2005). Similar results were observed for the A $\beta$  fragment A $\beta$ (21–30) in monomeric solution. Lazo *et al.* postulated that this decapeptide adopts a structure that nucleates the intramolecular folding of the full-length A $\beta$  monomer. The solution dynamics of A $\beta$ (21–30), as determined by NMR studies, yielded two families of folded A $\beta$ (21–30) structures both containing a turn-like motif centered at G25–S26 (Lazo *et al.*, 2005). These *in vitro* results raise questions that can be addressed *in silico*: (1) what is the driving force of folding, and (2) how does the folded structure affect the pathway of A $\beta$  assembly?

A $\beta$ 40 and A $\beta$ 42 both have high tendencies to aggregate into fibrils, which makes studies of oligomeric intermediates difficult. To study A $\beta$  oligomerization *in vitro*, the technique photo-induced cross-linking of unmodified proteins (PICUP) has been applied to covalently stabilize oligomers (see Chapter 12 by Bitan, 2006 in Volume 413). Using PICUP coupled with size-exclusion chromatography, Bitan *et al.* (2003a) showed that A $\beta$ 40 and A $\beta$ 42 display distinct oligomer size distributions. Solutions of A $\beta$ 40 display a rapid equilibrium among monomers, dimer, trimers, and tetramers, whereas A $\beta$ 42 preferentially forms pentamer/hexamer units (paranuclei), which further assemble into beaded superstructures similar to early protofibrils (Bitan *et al.*, 2003a). Further studies of primary structure elements controlling early oligomerization demonstrate that I41 is critical for paranucleus formation in solutions of A $\beta$ 42 and that A42 is necessary for further assembly of A $\beta$ 42 into larger oligomers (Bitan *et al.*, 2003b). In addition, oxidation of M35 blocks paranucleus formation in A $\beta$ 42 but does not alter the A $\beta$ 40 oligomer size distribution (Bitan *et al.*, 2003c). Mass spectroscopy and ion mobility measurements of A $\beta$ 42, which was subjected

to filtration to remove large assemblies and immediately electrosprayed, indicated the presence of dimers, tetramers, paranuclei, and pairs of paranuclei in agreement with PICUP results (Bernstein *et al.*, 2005).

Solid-state NMR studies yielded high-resolution information on the A $\beta$ 40 fibrillar structure, in which each individual peptide displays a bend, stabilized by a salt-bridge between D23 and K28 (Petkova *et al.*, 2002; see Chapter 6 by Tycko, 2006 in Volume 413). The kinetics of A $\beta$ 40 fibril formation is typically preceded by a lag phase that is not present in a recently synthesized A $\beta$ 40-lactam (D23/K28) that contains a lactam bridge between D23 and K28 (Sciarretta *et al.*, 2005). This experimental finding explains the importance of the bend in the V24-K28 region and the associated salt-bridge D23-K28 in A $\beta$ 40 fibrillogenesis and suggests that A $\beta$ 40-lactam (D23/K28) bypasses an unfavorable folding step, leading to  $\sim$ 1000-fold greater rate of fibril formation (Sciarretta *et al.*, 2005). The role of the salt-bridge D23-K28 formation at different stages of A $\beta$  folding and assembly can be addressed in the DMD approach by systematically varying the effective electrostatic potential.

#### *Four-Bead Model with Hydrogen Bonding: Planar $\beta$ -sheet Assemblies and the Role of Glycines*

Urbanc *et al.* (2004a) applied a four-bead protein model with backbone hydrogen bond interactions to study A $\beta$ 40 versus A $\beta$ 42 dimer formation. The A $\beta$ 42 sequence was simplified to a polyalanine chain with glycines at positions 9, 25, 29, 33, 37, and 38. This model exhibited conformational changes with increasing temperature. The monomer adopted an  $\alpha$ -helical conformation at low temperatures, several types of  $\beta$ -strand conformations including  $\beta$ -hairpin conformation at intermediate temperatures, and random coil-like conformation at high temperatures. A turn between G25 and G29 was consistently observed at intermediate temperatures and was shown to be induced by the presence of glycines, in particular G25. The importance of glycines was recently confirmed by an all-atom MD study of A $\beta$ 42 folding in explicit aqueous solution, which demonstrated that glycines induced local turns in the peptide and consequently caused the  $\alpha$ -helical to  $\beta$ -strand conformational change (Xu *et al.*, 2005).

The turn between G25 and G29 occurred in the same protein region as the bend in the model of A $\beta$  fibrils by Petkova *et al.* (2002). The local structure of a typical peptide within the fibril is quite different from the four-bead model prediction. Hydrogen bonds in the fibril are oriented along the fibrillar axis and link neighboring peptides with no significant intramolecular hydrogen bonding. In the simplified four-bead model, intramolecular hydrogen bonds first give rise to  $\beta$ -hairpin monomer

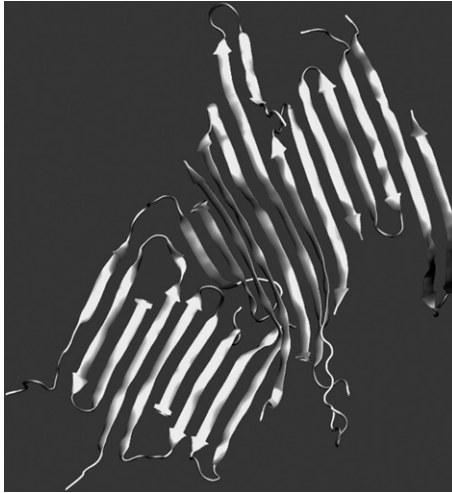


FIG. 1. An A $\beta$ 42 octamer as found within the four-bead model with hydrogen bond, but no amino acid-specific interactions. The octamer is an extended planar  $\beta$ -sheet with several domains that are slightly rotated with respect to one another.

conformations, which then further assemble into extended planar  $\beta$ -sheets. These planar  $\beta$ -sheet aggregates are held together exclusively by intramolecular and intermolecular hydrogen bonding (Fig. 1). A critical observation was the lack of stacking among the  $\beta$ -sheets, in contradiction with the model of A $\beta$  fibril formation (Petkova *et al.*, 2002). This result suggests that amino acid-specific interactions between pairs of side-chains are responsible for a correct description of the stacked  $\beta$ -sheet structure.

*Four-Bead Model with Amino Acid-Specific Hydrophobic Interactions:  
A $\beta$ 40 versus A $\beta$ 42 Oligomer Formation*

Ding *et al.* (2003) studied the effect of hydrophobic side chain interactions on the  $\alpha$ -helix and  $\beta$ -hairpin monomer conformations in a 16-residue polyaniline. They found that above a certain strength of effective hydrophobic interactions ( $E_{HP}/E_{HB} > 0.20$ ), the  $\beta$ -hairpin monomer conformation disappears, and it is replaced by a globular monomer conformation (Ding *et al.*, 2003). Nguyen and Hall demonstrated that the presence of a weak effective hydrophobic attraction ( $E_{HP}/E_{HB} < 1/6$ ) between the side chains of 16-residue polyaniline peptides leads to formation of a stacked  $\beta$ -sheet structure, consistent with the basic structural features of the fibril formation (Nguyen and Hall, 2004a,b, 2005; see Chapter 20 by Hall and Wagoner, 2006 in this volume).

Urbanc *et al.* introduced a four-bead  $A\beta$  model with hydrogen bond interactions and effective hydrophobic and hydrophilic interactions that were amino acid-specific (Urbanc *et al.*, 2004b). They showed that such a model with strong amino acid-specific hydrophobic and hydrophilic interactions ( $E_{HP}/E_{HB} = 0.3$ ) leads to the formation of globular oligomer structures (Urbanc *et al.*, 2004b). Urbanc *et al.* (2004b) demonstrated that this model is able to capture significant oligomerization differences between  $A\beta_{40}$  and  $A\beta_{42}$  that are consistent with *in vitro* results (Bitan *et al.*, 2003a,b,c). The effective hydrophobic attraction, as well as the effective hydrophilic repulsion, are critical features of the model that yields a steady-state distribution of  $A\beta$  oligomers of different sizes. If only the hydrophobic attraction was present in the model, the steady state would be a single globular oligomer because of the lack of forces opposing aggregation. In the presence of both hydrophobic attraction and hydrophilic repulsion, globular oligomers of various sizes coexist in a quasi-steady state. A typical globular oligomer consists of a core containing the hydrophobic parts of  $A\beta$  and a surface containing the hydrophilic N-terminal residues (Fig. 2). In agreement with experimental findings (Bitan *et al.*, 2003a), Urbanc *et al.* found that  $A\beta_{42}$  had an increased tendency to form pentamers, whereas dimers dominated in  $A\beta_{40}$ . Detailed structural analysis of these *in silico* results provided new structural insights and offered a plausible explanation of the role of M35 in  $A\beta_{40}$  versus  $A\beta_{42}$  oligomerization, indicating that oxidation of M35 disrupts  $A\beta_{42}$  paranuclei formation but does not affect  $A\beta_{40}$  oligomerization (Bitan *et al.*, 2003c). Statistical analysis of the tertiary structure of *in silico* pentamers showed important differences between the two alloforms in terms of contact formation involving C-terminal residues. In  $A\beta_{42}$ , the intramolecular contacts between V39-A42 on one side and I31, I32, L34, M35, and V36 on the other side dominated, whereas in  $A\beta_{40}$ , the C-terminal fragment V39-A40 did not form any significant intramolecular contacts (Urbanc *et al.*, 2004b). Thus, on the basis of this structural information, Urbanc *et al.* suggested that disrupting the hydrophobic nature of M35 by oxidation would cause a disruption of important hydrophobic contacts between oxidized M35 and C-terminal fragment in the  $A\beta_{42}$  pentamer. Because these contacts were not present in  $A\beta_{40}$  pentamers, oxidation of M35 would not make much of a difference in  $A\beta_{40}$  oligomer formation.

#### *United-Atom Model: $A\beta(21-30)$ Folding Initiated by a Hydrophobic Packing between V24 and K28*

Borreguero *et al.* (2005) developed a united-atom protein model in which all atoms except hydrogens are explicitly present. They applied the DMD approach with the united-atom model to study folding and unfolding transitions of  $A\beta(21-30)$  under different electrostatic interaction strengths

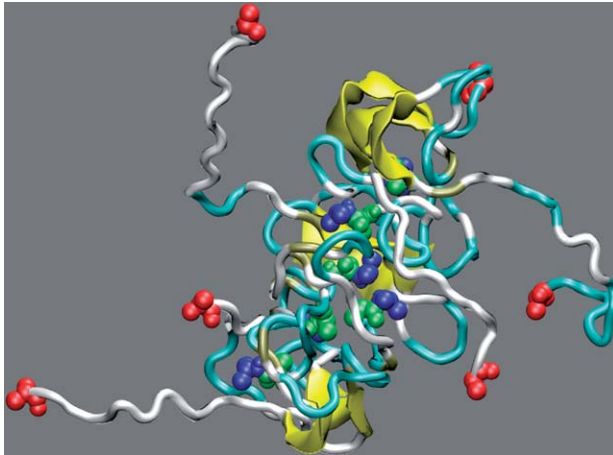


FIG. 2. Globular structure of A $\beta$ 42 hexamer as found within the four-bead peptide model with amino acid-specific interactions caused by hydrophathy. D1 is represented by four red spheres to illustrate the hydrophilic N-termini at the surface of the hexamer. I41 (four green spheres) and A42 (four blue spheres), as part of the C-terminal region, are at the hydrophobic core of the hexamer. Yellow ribbons represent a  $\beta$ -strand, cyan tube a turn, and silver tube a random coil-like secondary structure. The image was generated within the VMD software package (Humphrey *et al.*, 1996), which includes the STRIDE algorithm for calculating the secondary structure-propensity per residue (Heinig and Frishman, 2004).

(EIS) (Borreguero *et al.*, 2005). Hydrophobicity was shown to be the driving force of folding in A $\beta$ (21–30), inducing packing between V24 and the butyl portion of K28 (Fig. 3). In addition to hydrophobicity, intermediate EIS ( $\sim 1.5$  kcal/mol) predominantly between E22 and K28 contributed to an optimal stability of the folded structure (Borreguero *et al.*, 2005). At higher EIS ( $\sim 2.5$  kcal/mol)—typically occurring in the interior of proteins—A $\beta$ (21–30) was found to be partially unfolded because of a salt-bridge between D23 and K28. This observed prevalence of the D23-K28 interaction at highest EIS is in agreement with molecular models of protofibrils formed by full-length A $\beta$  (Petkova *et al.*, 2002) and A $\beta$ (16–35) (Ma and Nussinov, 2002) that show stabilization through D23-K28 salt-bridge and no E22-K28 interaction or V24-K28 packing. The study of Borreguero *et al.* exposed the binary nature of salt-bridge interactions between K28 and E22/D23 and provided a mechanistic explanation for the linkage of amino acid substitutions at E22 with AD and cerebral amyloid angiopathy (Borreguero *et al.*, 2005). Recently, Cruz *et al.* studied A $\beta$ (21–30) and its Dutch mutant (E22Q) by all-atom MD in water, reduced-density water, and in water with salt ions. They confirmed that in water A $\beta$ (21–30) folding is driven by hydrophobic



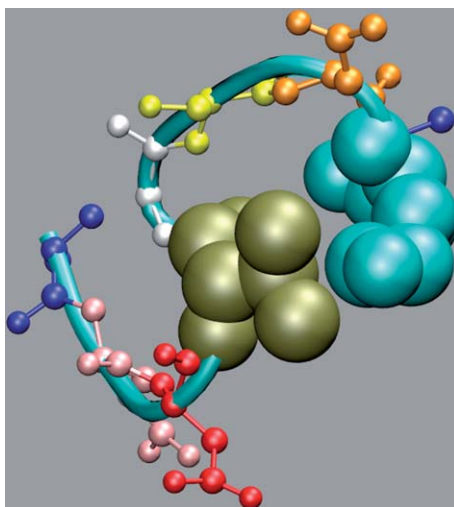


FIG. 3. Folded  $A\beta(21-30)$  decapeptide conformation as found within the united-atom model with amino acid-specific interactions caused by hydrophathy and charge. All atoms except hydrogens are drawn as small spheres: A21 and A30 (blue), E22 (pink), D23 (red), V24 (tan), G25 and G29 (white), S26 (yellow), N27 (orange), and K28 (cyan). V24 and K28 are presented by large opaque spheres to illustrate their packing, a critical event in the decapeptide folding. The image was generated within the VMD software package (Humphrey *et al.*, 1996).

forces involving V24 and K28. In addition, Cruz *et al.* showed that the  $A\beta(21-30)$  folded structure is very sensitive to changes in environment and that in the Dutch mutant folding events are rare (Cruz *et al.*, 2005).

### *Structural Hypotheses Derived from the DMD Studies*

Next we summarize the main hypotheses regarding the structure and dynamics of oligomer and fibril formation that are derived from the results of the DMD studies using either the four-bead model with amino acid-specific hydrophathies (Urbanc *et al.*, 2004b) or the united-atom model with atomic hydrophathies and effective electrostatic interactions (Borreguero *et al.*, 2005). These hypotheses are amenable to both *in silico* and *in vitro* testing.

1. Full-length monomers of  $A\beta_{40}$  and  $A\beta_{42}$  fold from the C-terminus toward the N-terminus. First intramolecular contacts during  $A\beta$  monomer folding are formed between V36 and V39 and their neighbors (Urbanc *et al.*, 2004b).

2.  $A\beta_{40}$  and  $A\beta_{42}$  monomers fold in different ways. The  $A\beta_{42}$  monomer folding is associated with a turnlike element centered at G37–G38, which is not present in the  $A\beta_{40}$  monomer. The  $A\beta_{40}$  monomer has an additional parallel  $\beta$ -strand between A2-F4 and the central hydrophobic cluster (L17-A21), not present in the  $A\beta_{42}$  monomer (Urbanc *et al.*, 2004b). The prediction of the turn at G37-G38 in  $A\beta_{42}$  (but not in  $A\beta_{40}$ ) is consistent with *in vitro* results of limited proteolysis, which shows that the region V39-A42 in  $A\beta_{42}$  is protease resistant, whereas the region V39-A40 in  $A\beta_{40}$  is not (Lazo *et al.*, 2005).

3. In  $A\beta_{40}$  oligomers, the most significant intermolecular contacts exist between pairs of central hydrophobic clusters (L17-V18-F19-F20-A21), whereas in  $A\beta_{42}$  oligomers, contacts between pairs of C-terminal regions (V39-A42) are the most important (Urbanc *et al.*, 2004b).

4. Despite similar globular structure with hydrophobic C-terminal residues in the core and hydrophilic N-terminal residues at the surface, the structure of  $A\beta_{40}$  and  $A\beta_{42}$  pentamers differs. The parallel  $\beta$ -strand structure at the N-termini of  $A\beta_{40}$  (as described in 2) persists in all assembly states and is completely absent from  $A\beta_{42}$  oligomers. Consequently, the N-termini of  $A\beta_{42}$  are spatially less restricted and can be found on average further away from the core of the oligomer. This difference in the N-termini properties might contribute to a more exposed hydrophobic core of  $A\beta_{42}$  oligomers, rendering  $A\beta_{42}$  more prone to further aggregate (Urbanc *et al.*, 2004b).

5. Hydrophobic attraction between V24 and the butyl portion of K28 drives the folding of  $A\beta(21-30)$ , whereas the salt-bridge E22-K28 contributes to the stability of the folded structure (Borreguero *et al.*, 2005).

6. Because experiments show that the same region V24-G25-S26-N27-K28 is protease resistant in the full-length  $A\beta$  (Lazo *et al.*, 2005), fibril formation should be preceded by an event that disrupts the folded loop V24-G25-S26-N27-K28. Because the E22-K28 salt-bridge contributes to the loop stability, a substitution of E22 by a non-negatively charged amino acid should enhance the fibril formation through: (1) decrease of the loop stability, and (2) increase of the rate of D23-K28 salt-bridge formation due to the absence of competition between E22 and D23 (Borreguero *et al.*, 2005).

Based on the above predictions, one can introduce selected amino acid substitutions in  $A\beta_{40}$  and  $A\beta_{42}$  that would hypothetically disrupt or change monomer and oligomer conformations. Should *in vitro* and other experimental findings that target the structure of  $A\beta$  folded monomers and oligomers determine that any of the above hypotheses is not valid, the DMD approach can be refined in two ways: (a) by introducing more detail into the protein model; and (b) by refining the interactions between the side-chain atoms and possibly introducing locally modified interactions.

## Methods

### *Discrete Molecular Dynamics Method*

MD is a computer simulation method in which particles move according to specific interparticle forces on the basis of classical dynamics. Newton's equations of motion must be numerically integrated at each time step for all particles to update instantaneous velocities and positions. DMD is a simplified version of MD and is applicable whenever the interparticle potentials can be represented by one or more square wells (Rapaport, 1997). Within each well, the potential is constant, because the force between the two particles is zero, and thus the particles move with constant velocities until they reach a distance at which the potential is discontinuous. At that moment, an elastic or inelastic collision occurs, and the two particles change their velocities instantaneously while conserving the total energy, momentum, and angular momentum. No numerical integration is needed. The only events are two-particle collisions, and the main challenge is to keep track of collision times. Consequently, DMD simulations are considerably faster than continuous MD simulations.

During a DMD simulation, the number of particles, volume, and temperature are held constant. Periodic boundary conditions are implemented to avoid interactions with the walls of the simulation box. The size of the box is chosen to be larger than the stretched protein under study. We implement temperature control in our model using the method proposed by Berendsen *et al.* (1984). In this method, a heating rate coefficient,  $\alpha$ , is introduced. The temperature is rescaled at regular intervals  $\Delta t$ :  $T(t + \Delta t) = T(t) + \alpha \Delta t [T_\infty - T(t)]$  where  $T(t)$  is the instantaneous temperature,  $T(t + \Delta t)$  is the rescaled temperature, and  $T_\infty$  is the target temperature of the heat bath;  $\alpha^{-1}$  is a characteristic time, in which the temperature equilibrates. The time interval  $\Delta t$  corresponds to about  $N$  collisions, where  $N$  is the number of particles. Temperature is defined by the total kinetic energy of particles as follows  $\frac{3}{2} \kappa_B T = \frac{1}{N} \sum_{i=1}^N \frac{m_i v_i^2}{2}$ , where  $\kappa_B$  is the Boltzmann constant,  $v_i$  are the velocities of each of the  $N$  particles, and  $m_i$  their masses. Rescaling temperature requires rescaling all  $N$  velocities  $v_i$  by a factor  $\sqrt{T(t + \Delta t)/T(t)}$ , which is followed by recalculation of the collision times. To avoid the time-consuming task of recalculating collision times, we introduce a rescaled time variable and rescaled potentials, keeping the velocities and collision times intact. This transformation does not alter the trajectory in any way. We keep track of the original simulation time by keeping track of the rescaling factors, so that results are expressed in original units. A more detailed description has been given elsewhere (Borreguero, 2004).

### Four-Bead Model Implementation

We use the four-bead protein model introduced by [Ding \*et al.\* \(2003\)](#). In the four-bead model, the backbone is represented by three beads, corresponding to the amide ( $N$ ), the  $\alpha$ -carbon ( $C_\alpha$ ), and the carbonyl ( $C$ ) groups. Each side-chain (except G, which lacks the side-chain group) is represented by one side-chain bead ( $C_\beta$ ). Each bead (atom) is characterized by its mass and hard-core radius. In the simplest version of the model, all atoms have equal mass and their hard-core radii are set to their van der Waals radii ([Creighton, 1993](#)). Each side-chain atom is characterized by a type, which determines its interactions with other atoms. Any two atoms can only be at a distance  $d > d_{min}$ , where  $d_{min}$  is the sum of their hard-core radii. Thus, the potential is set to an “infinitely” large value for  $d < d_{min}$ . Pairs of atoms can be linked by a covalent bond or an angular constraint to account for the protein geometry as shown in [Fig. 4A](#). If two atoms are linked in this way, there is a distance  $d_{max}$  such that for  $d > d_{max}$  the potential is infinite to prevent the two atoms from breaking the bond. The lengths of bonds and angular constraints are determined phenomenologically by calculating their distributions using known folded protein structures of  $\sim 7700$  proteins from the Protein Data Bank (PDB) (<http://www.rcsb.org/pdb>). The values of the lengths of covalent bonds and angular constraints, which are allowed to vary around their average values by 2%, were reported elsewhere ([Ding \*et al.\*, 2003](#)).

**Backbone Hydrogen Bond.** In proteins, the most ubiquitous hydrogen bond interaction involves the carbonyl oxygen and the amide hydrogen of two amino acids. In the four-bead model, because the carbonyl oxygen and the amide hydrogen are not explicitly present, an effective backbone

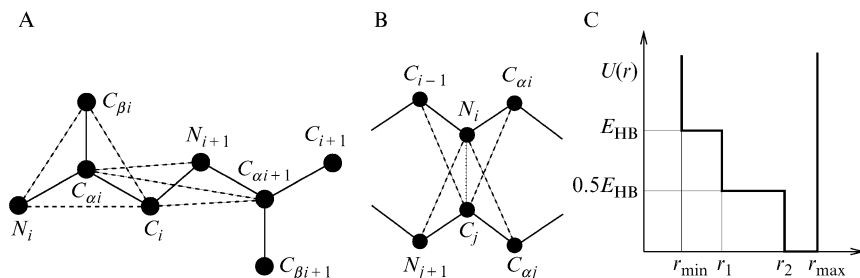


FIG. 4. (A) Covalent bonds (solid lines) and constraints (dashed lines) that define the four-bead peptide model. (B) Hydrogen bond between  $C_j$  and  $N_i$  (dotted line) and the corresponding auxiliary bonds (dashed lines) that define the geometry of hydrogen bonding in the four-bead model. (C) Interparticle potential  $U(r)$  for the four auxiliary bonds shown in (B).

hydrogen bond is introduced between the nitrogen atom  $N_i$  of the  $i$ -th amino acid and the carbon atom  $C_j$  of the  $j$ -th amino acid (Ding *et al.*, 2003) (Fig. 4B–C). On formation of the hydrogen bond, atoms  $N_i$  and  $C_j$  in the model change types to prevent their involvement in additional hydrogen bond formation. When amino acids  $i$  and  $j$  belong to the same protein, we allow them to form a hydrogen bond only if they are at least three amino acids apart,  $|i - j| \geq 4$ . This constraint is a consequence of angular restrictions of the hydrogen bond that requires that the  $CO$  and  $NH$  bonds are approximately collinear. These same angular restrictions are enforced by introducing four auxiliary bonds involving the left and the right neighboring beads of  $N_i$  and  $C_j$  (Fig. 4B). The hydrogen bond between  $N_i$  and  $C_j$  forms only if all six beads are at energetically favorable distances. Each of the four auxiliary interactions is modeled by a double-step potential (Fig. 4C), and the particular values of the hydrogen bond parameters are chosen phenomenologically to best match the distribution of the corresponding distance in real proteins (Ding *et al.*, 2003). The additional auxiliary interactions take place only in the presence of the hydrogen bond interaction. During the hydrogen bond formation or deletion, the other interactions involving the  $N_i$  and  $C_j$  beads remain intact. When a new hydrogen bond is formed, the new hard-core collision distance between  $N_i$  and  $C_j$  is assigned to be 4.0 Å, such that at the lowest energy state of a hydrogen bond, the optimal distances of the four auxiliary pairs allow for approximately linear alignment of the  $CO$  and the  $NH$  bonds.

*Amino Acid-Specific Interactions Caused by Hydrophathy.* Because the solvent is not explicitly present in our DMD approach, effective interactions between the side-chain atoms are introduced to mimic the solvent effects. We introduce hydrophobic attraction and hydrophilic repulsion between pairs of side-chains, depending on the hydrophatic nature of individual side-chains. In our model, the potential energy decreases when two hydrophobic residues interact, thus minimizing their solvent accessible surface area (SASA). Conversely, the potential energy increases when two hydrophilic residues interact. This potential thus favors noninteracting hydrophilic residues, which maximizes their SASA.

There are different ways of implementing amino acid-specific hydrophatic interactions. We chose the empirical amino acid hydrophathy scale derived by Kyte and Doolittle (1982) as previously described (Urbanc *et al.*, 2004b). We consider the following amino acids: I, V, L, F, C, M, and A hydrophobic; N, Q, and H non-charged hydrophilic; and D, E, K, and R charged hydrophilic. The remaining amino acids with absolute values of hydrophathies below threshold values are considered neutral, so that two neutral side-chain atoms only interact through their hard-core interaction. The hydrophatic interactions are of two types: (1) an attractive interaction between two hydrophobic side-chains; and (2) a repulsive interaction between two noncharged hydrophilic

or a charged hydrophilic and a noncharged hydrophilic side-chain. Interactions are implemented using a square-well potential between the pairs of side-chain beads  $C_{\beta,i}$  and  $C_{\beta,j}$ , so that they interact if the distance between their centers is less than the interaction range distance  $7.5 \text{ \AA}$  (Fig. 5A,B). The potential energy of the effective attractive hydrophobic interaction  $E_{HP}$  is proportional to the mean of the relative hydrophobic strengths, I (-1.0), V (-0.93), L (-0.84), F (-0.62), M (-0.42), and A (-0.40), where the negative sign reflects the attractive nature of the interaction. The potential energy of the effective repulsive hydrophilic interaction is proportional to the mean of the relative hydrophilic strengths, R (1.0), K (0.87), E (0.78), D (0.78), N (0.78), Q (0.78), and H (0.71), where the positive sign reflects the repulsive nature of the interaction. H, with a  $\text{pK}_a$  value  $\sim 6.0$ , is considered a noncharged hydrophilic amino acid because at physiological conditions ( $\text{pH} = 7.4$ ) only about 4% of H is charged.

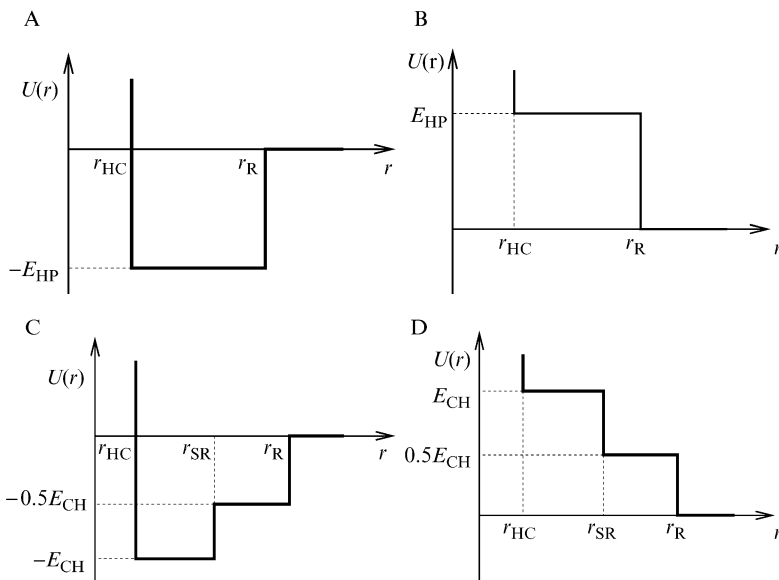


FIG. 5. (A) Effective hydrophobic interaction with the potential energy  $-E_{HP}$  between two hydrophobic side-chain atoms as implemented in the four-bead model. (B) Effective hydrophilic interaction with the potential energy  $E_{HP}$  between two hydrophilic side-chain atoms as implemented in the four-bead model. The range of the hydrophobic and hydrophilic interactions is set to  $r_R = 7.5 \text{ \AA}$ . (C) Effective electrostatic interaction with the electrostatic potential energy  $-E_{CH}$  between two oppositely charged atoms as implemented in the four-bead model. (D) Effective electrostatic interaction with the electrostatic potential energy  $E_{CH}$  between two atoms of the same charge as implemented in the four-bead model. The range of the interaction is  $r_R = 7.5 \text{ \AA}$  and the soft range is  $r_{SR} = 6 \text{ \AA}$ . The hard-core repulsion distance is denoted by  $r_{HC}$ .

*Amino Acid-Specific Interactions Caused by Charge.* The effective electrostatic interaction between two charged side-chain atoms is implemented using a double attractive/repulsive square well potential with the interaction range  $r_R = 7.5 \text{ \AA}$  and a “soft” interaction range  $r_{SR} = 6 \text{ \AA}$  (Fig. 5C,D). The energy of the effective electrostatic interaction,  $E_{CH}$ , is tunable and is typically set in the range  $E_{CH}/E_{HB} \in [0, 1]$ , with different  $E_{CH}$  values corresponding to different solvent conditions.

### *United-Atom Model Implementation*

In the united-atom model, all protein atoms except hydrogens are explicitly represented. The backbone of the protein is represented by four atoms, corresponding to the amide group ( $N$ ), the  $\alpha$  carbon ( $C_\alpha$ ), the prime carbon ( $C$ ), and the oxygen ( $O$ ). On formation of the backbone hydrogen bond,  $N$  and  $O$  change their types. A special atom type is introduced for the amide group ( $N$ ) of P to describe its characteristic covalent bond to the P  $C_\delta$  side-chain atom. We assign a different type for each side-chain atom of the 19 amino acids. For each atom, we assign an individual atomic mass, a phenomenologically estimated radius (Tsai *et al.*, 1999) and a nominal charge of (+1) to the amino groups of K and R, and a (-1) charge to the carboxy groups of D and E.

To achieve the correct description of the flexibility of the protein, we assign three types of bonding between protein atoms to account for the backbone and side-chain geometries: (1) covalent bonding; (2) angular constraints; and (3) rotameric constraints (Fig. 6). The rotameric constraints were first introduced in the context of the six-bead model (Ding *et al.*, 2005) and later expanded to account for the united-atom model (Borreguero, 2004). These lengths and their variances are derived from the library of potentials using a statistical analysis of a specific database of protein structures (PDB40), which is a subset of the Structural Classification of Proteins (SCOP) database (<http://scop.berkeley.edu>) of protein structures (Chandonia *et al.*, 2004). The exact values were given by Borreguero (2004). Typically, the covalent bonds are allowed to vary by  $\sim 4\%$ , angular constraints by  $\sim 6\%$ , and the rotameric constraints by 4–28%.

We implement covalent and angular constraints as square-well potentials with two “infinite” walls representing the limits of typical interatomic distances for the particular bond or constraint under consideration (Fig. 8A). As an example, we describe a rotameric constraint by implementing the  $\chi_1$  rotamer angle (e.g., for Val).  $\chi_1$  is the angle between the two planes generated by atoms  $N-C_\alpha-C_\beta$  and  $C_\alpha-C_\beta-C_{\gamma_2}$ , respectively (Fig. 7A). Thus,  $\chi_1$  is determined by the positions of the four atoms  $N-C_\alpha-C_\beta-C_{\gamma_2}$ . Other four-atom sets are equally valid, namely  $N-C_\alpha-C_\beta-C_{\gamma_1}$ ,  $C-C_\alpha-C_\beta-C_{\gamma_1}$ ,

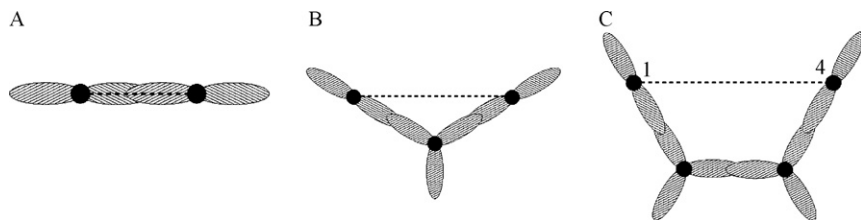


FIG. 6. Bond types between two atoms. (A) Covalent bond. (B) Angular constraint to model the central hybridized molecular orbital. (C) Rotameric constraint 1–4 to reproduce the statistically observed preference of atom 4 to orient itself with respect to the position of atom 1.

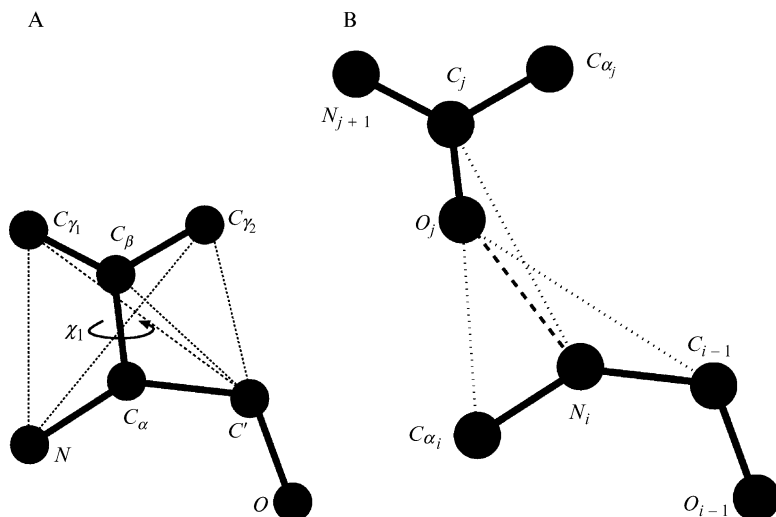


FIG. 7. (A) Schematic diagram of the  $\chi_1$  rotamer of valine. Dashed lines represent distances between atoms involved in rotameric-constraint interactions. Any two of these interactions uniquely determine  $\chi_1$ . (B) Schematic diagram of the backbone hydrogen bond. The dashed line represents the  $N-O$  bond, and the dotted lines represent auxiliary constraints maintaining the correct  $N-O$  orientation.

or  $C-C_\alpha-C_\beta-C_{\gamma_2}$ .  $\chi_1$  determines the distance between the first and last atom for each set (i.e.,  $\overline{NC_{\gamma_2}}$  (Fig. 8B–C),  $\overline{NC_{\gamma_1}}$ ,  $\overline{CC_{\gamma_1}}$ , and  $\overline{CC_{\gamma_2}}$ ). Conversely, any two of these four distances are sufficient to uniquely determine  $\chi_1$ . The distance distributions for each of the four constraints defines a respective potential energy function (Fig. 8D).



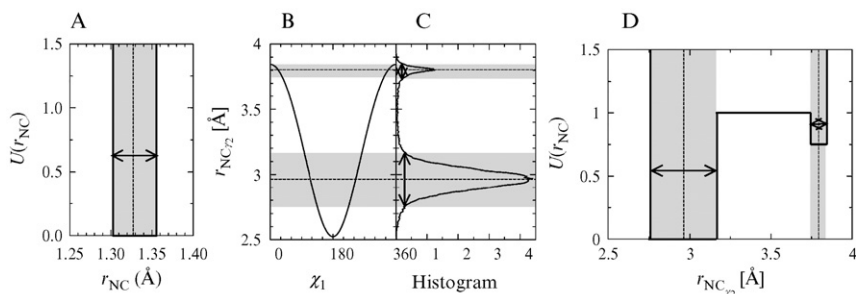


FIG. 8. (A) Potential energy associated with the bond between N and C groups in the protein backbone. The most probable distance (dashed line) and a typical range of distances (arrow) derived from the PDB40 structural database. (B) The distance  $r_{NC_2}$  as a function of the rotamer  $\chi_1$ . (C) Histogram of distances  $r_{NC_2}$  as derived from the PDB40 dataset. (D) Associated rotameric potential between  $N$  and  $C_{\gamma_2}$ . We estimate the width and the depth of the two potential wells as the width and the area of the two respective probability peaks. We normalize the potential units in this plot with respect to the deepest well.

*Soft-Core Interactions.* Two atoms can approach to distances smaller than the sum of their van der Waals radii. However, these distances are not energetically favorable, and the two atoms have to overcome an energy barrier (Tsai *et al.*, 1999). From the occurrence probability of these events at room temperature, we estimate the energy barrier to be threefold the thermal energy. We set the lower limit of the soft-core region to be 85% of the sum of their respective van der Waals radii (Fig. 9A). Distances smaller than this lower limit, which correspond to less than 1% of the observed distances in protein structures, are not allowed.

*Backbone Hydrogen Bond.* We adopt the hydrogen bonding first implemented into the four-bead model (between the amide  $N$  and carbonyl  $C$  groups) (Ding *et al.*, 2003) and further developed to account for the explicit backbone  $O$  group (Ding *et al.*, 2005). The hydrogen bond between  $N$  and  $O$  atoms is implemented into the model as a “reaction”-type interaction. On formation of the backbone hydrogen bond,  $N$  and  $O$  change their types into new “bonded” types to prevent additional hydrogen bonding to the third atom. The geometry of the backbone hydrogen bond is modeled by a four-body interaction. In addition to the  $N$ – $O$  bond, three additional constraints between neighboring  $N$  and  $O$  atoms (Fig. 7B) reproduce the orientation of the hydrogen bond. Details of this backbone hydrogen bond model were described elsewhere (Borreguero, 2004; Ding *et al.*, 2005).

*Amino Acid-Specific Interactions Caused by Hydropathy.* As in the four-bead model, we introduce an implicit treatment of the solvent with an attractive/repulsive potential energy between pairs of side-chain atoms

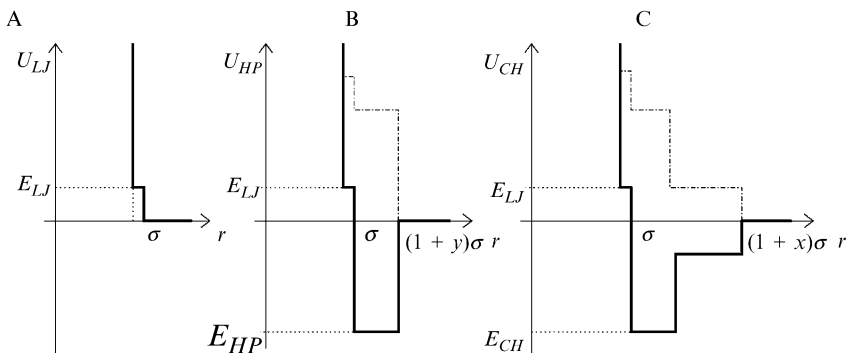


FIG. 9. Three types of potentials,  $U_{LJ}$ ,  $U_{HP}$ , and  $U_{CH}$ , in dependence on the interparticle distance  $r$ . (A) Lenard-Jones potential  $U_{LJ}$ .  $\sigma$  is the sum of the Van der Waals radii of the two interacting atoms, and  $E_{LJ} = 3\kappa_B T$  is a finite repulsive potential energy of two atoms at distances smaller than  $\sigma$ . (B) Effective hydrophobic potential  $U_{HP}$  with  $y = 0.2$  between two noncharged atoms with the potential energy  $E_{HP}$ . The dotted-dashed line represents the effective repulsion between a charged and a noncharged atom. (C) Effective electrostatic potential  $U_{CH}$  with  $x = 1.3$ . The first potential well at distances  $\sigma < r < 1.4\sigma$  corresponds to the potential energy  $E_{CH}$ . The second potential well at distances  $1.4\sigma < r < 2.3\sigma$  corresponds to the potential energy  $0.3E_{CH}$ .

(Fig. 9B). We define an atomic hydrophathy scale on the basis of an experimental estimation of the gain/loss of the free energy on transferring a particular amino acid from an aqueous solution to a gas phase (Wesson and Eisenberg, 1992). Knowing the gain/loss of the free energy for each amino acid, we then estimate the atomic solvation energies (Zhang *et al.*, 1997). When the distance between two atoms becomes smaller than 120% of the sum of their Van der Waals radii, we consider the interfacial volume as solvent excluded, and the two atoms interact with a potential energy equal to the sum of their hydrophathy values. For a particular atom of type  $t$ , we define hydrophathy values  $HP_t = -\sigma_t SASA_t/n_t$ , where  $\sigma_t$  is the atomic solvation parameter (free energy gain/loss per unit of solvent-exposed area of the atom type  $t$ ),  $SASA_t$  is the solvent accessible surface area (Wesson and Eisenberg, 1992), and  $n_t$  is an estimated number of the neighboring nonbonded atoms (usually  $1 \leq n_t \leq 3$ ).

*Amino Acid-Specific Interactions Caused by Charge.* We implement an effective electrostatic interaction between two charged atoms using a double attractive/repulsive square well potential (Fig. 9C). The cutoff between two charged atoms is set to a value 2.33-fold larger than the sum of their Van der Waals radii ( $\sim 7 \text{ \AA}$ ). A tunable potential energy of the charged interaction within the range 0–2.5 kcal/mol allows us to perform simulations in a wide range of solvent conditions. This energy range is within and

above the experimentally measured values for the free energy gain on salt-bridge formation on the surface of proteins, 0.24–1.26 kcal/mol (Horovitz and Fersht, 1992; Searle *et al.*, 1999). A more detailed description was given elsewhere (Borreguero *et al.*, 2005).

### *Limitations of the DMD Approach*

As described previously, the peptide model parameters (the peptide bonds and constraints) and the parameters of the hydrogen bonding are defined phenomenologically using the known crystalline structure of proteins from the Protein Data Bank. Such a phenomenological approach to modeling (Ding *et al.*, 2002) has been discussed extensively by Zhang *et al.* (2004). The phenomenologically derived force-field was shown to be essential for successful folding of the Trp-cage protein (Ding *et al.*, 2005). As the PDB expands, however, these parameters could change. Different protein databases ( $\alpha$ -protein database,  $\beta$ -protein database,  $\alpha/\beta$ -protein database, etc.) could yield different parameters as well. It has been shown that knowledge-based potentials yield different results when trained on either NMR or X-ray resolved structures (Godzik *et al.*, 1995). We considered the filtered protein structural database PDB40, which contains representatives of all known protein folds. The filter ensures that no two proteins have more than 40% sequence identity, preventing any bias in the statistical analysis toward overrepresented homologous sequences. For our purposes, the peptide model parameters obtained from this database are “fixed” and represent the definition of the model peptide.

In our approach, the hydrogen bond interaction is not amino acid-specific. The hydrogen-bond potential energy,  $E_{HB}$ , is the energy unit. This choice does not imply that we treat the hydrogen bond interaction as a fixed interaction independent of the environment. In fact, the free energy cost of breaking a hydrogen bond strongly depends on the local environment (Honig and Yang, 1995). This cost may be small in aqueous solutions but is typically large in organic solvents. In aqueous solution, the effective hydrogen bond is strong at the hydrophobic core and weak at the surface of a protein or a protein assembly. These local variations of the hydrogen bonding caused by variations in the dielectric constant within a solvent are neglected in our approach. However, our simulation approach allows for different environments by assigning different nonvariable dielectric constants to different solvent conditions and correspondingly renormalizing interactions in the model to the reference energy given by the hydrogen bond energy,  $E_{HB}$ .

Amino acid-specific interactions caused by hydrophathy and charge also depend on the particular environment under study. The strength of the

effective hydrophobic interactions depends on the solvent and other global variables, such as temperature, molar concentration, and pH. Our model accounts for these global variables by varying the amino acid-specific hydrophobicities and charge, temperature, number of peptides, and the volume of the simulation box. In an explicit solvent, however, the hydrophobic/hydrophilic effect and electrostatic interactions depend strongly on the local variations of SASA and the dielectric constant. As in the case of the hydrogen bond, in water the hydrophobic/hydrophilic effect is strongest at the surface of the protein or protein assembly. The opposite is true for the electrostatic interactions. The free energy change associated with breaking a salt-bridge within the hydrophobic core of a protein is much higher than at its surface, where polar water molecules shield the charged atoms and thus effectively weaken the electrostatic interactions between the side-chains.

The DMD approach described in this review neglects local variations in the dielectric constant and in the SASA of each side-chain. The effective hydrophobic/hydrophilic interactions in coarse-grained models are based on hydrophathy scales. A number of different hydrophathy scales exist, some of them phenomenological (for example, [Kyte and Doolittle \[1982\]](#)); others are based on the *in vitro* gain/loss of the free energy when a particular atom is transferred from an aqueous solvent to a gas phase (for example, [Wesson and Eisenberg \[1992\]](#)). A question that needs to be addressed in the future is how robust the results of the DMD approach are with respect to different hydrophathy scales.

The effective electrostatic interaction is modeled by a two-step square-well potential. We neglect the long-range nature of the Coulombic interaction between two charged particles (and/or two dipoles). Implementing a true long-range electrostatic interaction would require a considerable computational effort with a potential approximated by a multistep square-well of an “infinite” range. The interaction range problem is addressed in all-atom MD either by using Ewald sums in combination with periodic boundary conditions, multipole expansions, or a field-reaction method ([Rapaport, 1997](#)). However, even these sophisticated algorithms neglect the electrostatic forces above a certain cutoff distance. In the DMD approach, the solvent is implicit. When implementing the effective interaction between charged atoms, one needs to consider effects of the aqueous solution. The charged groups of a peptide, surrounded by water molecules, are effectively shielded because of the polar nature of water molecules. Because of this shielding, we can approximate the effective electrostatic potential by a two-step square-well potential with a finite distance range as a first-order approximation. In addition, one can assign an effective charge to a given side-chain. As the charge of a particular side-chain depends on pH, we can model different pH environments by reassigning the charge of a particular amino acid. For example, H, which is

considered neutral at  $\text{pH} = 7$ , would be considered positively charged at low  $\text{pH}$ . Again, our approach neglects the fact that in the core of a peptide assembly the electrostatic interactions are stronger than at the surface of the assembly.

In principle, it is possible to model local variations of interparticle interactions that depend on SASA of individual side-chain atoms in the DMD approach by keeping track of the neighborhood of each atom. All interparticle interactions except the hard-core and soft-core interactions would need to be rescaled at regular simulation time intervals to account for atom-to-atom variability of SASA. However, this would require computational effort and could significantly slow down the simulations.

## Conclusion

In this chapter, we described in detail the DMD method, coarse-grained protein models, and interparticle interactions that were developed to study  $A\beta$  folding and aggregation. We also addressed weaknesses of the DMD approach that originate in simplifications of the peptide description, interparticle interactions, and the absence of explicit solvent. The strengths of the DMD approach are: (1) its *ab initio* nature that does not require any experimental parameters specific to  $A\beta$  as input parameters; (2) its efficiency that allows for study of not only folding but also of oligomerization and fibril formation of full-length  $A\beta$ ; and (3) its biological relevance that can be achieved through structural *in vitro*  $\leftrightarrow$  *in silico* feedback-guided development of the model and interactions. Because of these advantages, this approach is also applicable to studies of proteins associated with other neurodegenerative diseases.

Soluble oligomers are a common feature of amyloid assembly, but their significance in the pathway of fibril assembly is not clear (Glabe, 2004). Are oligomers obligatory intermediates on a single pathway from misfolded monomers to fibrils or are oligomers reversible off-pathway intermediates that only buffer the concentration of misfolded monomers? Another fundamental question is how the assembly state of  $A\beta$  correlates with its function and toxicity (Klein *et al.*, 2004). The DMD approach can provide detailed structural information on different assembly states and explain the pathways of  $A\beta$  oligomer and fibril formation. Assuming that the assembly structure is directly correlated to toxicity, *in silico* findings could yield mechanistic hypotheses about why certain assemblies are more toxic than others and thus provide directions for further *in vitro* testing and identifying drug targets that would disrupt formation of these assemblies.

## Acknowledgments

We thank S. V. Buldyrev, F. Ding, and N. V. Dokholyan for helpful discussions on the implementation of coarse-grained models and for continued intellectual support. We greatly appreciate many valuable discussions with G. Bitan, N. D. Lazo, and D. B. Teplow that guided us in our approach. We are thankful to the NIH for support on the grant AG023661 and Alzheimer's Association for the Zenith Fellows award. We are grateful to Stephen Bechtel, Jr. for a private donation.

## References

- Ash, W. L., Zlomislic, M. R., Oloo, E. O., and Tieleman, D. P. (2004). Computer simulations of membrane proteins. *Biochim. Biophys. Acta* **1666**, 158–189.
- Berendsen, H. J. C., Postma, J. P. M., Vangunsteren, W. F., Dlinola, A., and Haak, J. R. (1984). Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **81**, 3684–3690.
- Bernstein, S. L., Wytttenbach, T., Baumketner, A., Shea, J.-E., Bitan, G., Teplow, D. B., and Bowers, M. T. (2005). Amyloid  $\beta$ -protein: Monomer structure and early aggregation states of A $\beta$ 42 and its Pro<sup>19</sup> alloform. *J. Am. Chem. Soc.* **127**, 2075–2084.
- Bitan, G. (2006). Structural study of metastable amyloidogenic protein oligomers by photo-induced cross-linking of unmodified proteins. *Methods Enzymol.* **413**, in press.
- Bitan, G., Kirkitadze, M. D., Lomakin, A., Vollers, S. S., Benedek, G. B., and Teplow, D. B. (2003a). Amyloid  $\beta$ -protein (A $\beta$ ) assembly: A $\beta$ -40 and A $\beta$ -42 oligomerize through distinct pathways. *Proc. Natl. Acad. Sci. USA* **100**, 330–335.
- Bitan, G., Vollers, S. S., and Teplow, D. B. (2003b). Elucidation of primary structure elements controlling early amyloid  $\beta$ -protein oligomerization. *J. Biol. Chem.* **278**, 34882–34889.
- Bitan, G., Tarus, B., Vollers, S. S., Lashuel, H. A., Condron, M. M., Straub, J. E., and Teplow, D. B. (2003c). A molecular switch in amyloid assembly: Met<sup>35</sup> and amyloid  $\beta$ -protein oligomerization. *J. Am. Chem. Soc.* **125**, 15359–15365.
- Borreguero, J. M. (2004). Computational studies of protein stability and folding kinetics. PhD. Thesis, Boston University.
- Borreguero, J. M., Urbanc, B., Lazo, N. D., Buldyrev, S. V., Teplow, D. B., and Stanley, H. E. (2005). Discrete molecular dynamics study of the amyloid  $\beta$ -protein decapeptide A $\beta$ (21–30). *Proc. Natl. Acad. Sci. USA* **102**, 6015–6020.
- Chandonia, J. M., Hon, G., Walker, N. S., Lo Conte, L., Koehl, P., Levitt, M., and Brenner, S. E. (2004). The ASTRAL compendium in 2004. *Nucleic Acids Res.* **32**, D189–D192.
- Coles, M., Bicknell, W., Watson, A. A., Fairlie, D. P., and Craik, D. J. (1998). Solution structure of amyloid- $\beta$  peptide (1–40) in a water-micelle environment. Is the membrane spanning domain where we think it is? *Biochemistry* **37**, 11064–11077.
- Creighton, T. E. (1993). “Proteins: Structures and Molecular Properties,” 2nd Ed. Freeman and Company.
- Crescenzi, O., Tomaselli, S., Guerrini, R., Salvatori, S., D’Ursi, A. M., Temussi, P. A., and Picone, D. (2002). Solution structure of the Alzheimer amyloid  $\beta$ -peptide (1–42) in an apolar microenvironment. Similarity with a virus fusion domain. *Eur. J. Biochem.* **269**, 5642–5648.
- Cruz, L., Urbanc, B., Borreguero, J. M., Lazo, N. D., Teplow, D. B., and Stanley, H. E. (2005). Solvent and mutation effects on the nucleation of amyloid  $\beta$ -protein folding. *Proc. Natl. Acad. Sci. USA* **102**, 18258–18263.
- Ding, F., Dokholyan, N. V., Buldyrev, S. V., Stanley, H. E., and Shakhnovich, E. I. (2002). Direct molecular dynamics observation of protein folding transition state ensemble. *Biophys. J.* **83**, 3525–3532.

- Ding, F., Borreguero, J. M., Buldyrev, S. V., Stanley, H. E., and Dokholyan, N. V. (2003). A mechanism for the  $\alpha$ -helix to  $\beta$ -hairpin transition. *Proteins: Struct. Func. Genet.* **53**, 220–228.
- Ding, F., Buldyrev, S. V., and Dokholyan, N. V. (2005). Folding Trp-cage to NMR resolution native structure using a coarse-grained protein model. *Biophys. J.* **88**, 147–155.
- Dobson, C. M. (2004). Principles of protein folding, misfolding, and aggregation. *Cell Dev. Biol.* **15**, 3–16.
- Dokholyan, N. V., Buldyrev, S. V., Stanley, H. E., and Shakhnovich, E. I. (1998). Discrete molecular dynamics studies of folding of a protein-like model. *Folding Design* **3**, 577–587.
- Dokholyan, N. V., Buldyrev, S. V., Stanley, H. E., and Shakhnovich, E. I. (2000). Identifying the protein folding nucleus using molecular dynamics. *J. Mol. Biol.* **296**, 1183–1188.
- Dokholyan, N. V., Borreguero, J. M., Buldyrev, S. V., Ding, F., Stanley, H. E., and Shakhnovich, E. I. (2003). Identifying importance of amino acids for protein folding from crystal structures. *Methods Enzymol.* **374**, 616–638.
- Feig, M., and Brooks, C. L., III (2004). Recent advances in the development and application of implicit solvent models in biomolecule simulations. *Curr. Opin. Struct. Biol.* **14**, 217–224.
- Fersht, A. R., and Daggett, V. (2002). Protein folding and unfolding at atomic resolution. *Cell* **108**, 573–582.
- Glabe, C. G. (2004). Conformation-dependent antibodies target diseases of protein misfolding. *Trends Biochem. Sci.* **29**, 542–547.
- Godzik, A., Kolinski, A., and Skolnick, J. (1995). Are proteins ideal mixtures of amino acids? Analysis of energy parameter sets. *Prot. Sci.* **4**, 2107–2117.
- Hall, C. K., and Wagoner, V. A. (2006). Computational approaches to fibril structure and formation. *Methods Enzymol.* **412**, 338–365.
- Heinig, M., and Frishman, D. (2004). STRIDE: A web server for secondary structure assignment from known atomic coordinates of proteins. *Nucl. Aci. Res.* **32**, W500–W502.
- Honig, B., and Yang, A. S. (1995). Free energy balance in protein folding. *Adv. Protein Chem.* **46**, 27–58.
- Horovitz, A., and Fersht, A. R. (1992). Co-operative interactions during protein folding. *J. Mol. Biol.* **224**, 733–740.
- Humphrey, W., Dalke, A., and Schulten, K. (1996). VMD: Visual molecular dynamics. *J. Mol. Graphics* **14**, 33–38.
- Karplus, M., and McCammon, J. A. (2002). Molecular dynamics simulations of biomolecules. *Nat. Struct. Biol.* **9**, 646–652.
- Kayed, R., Head, E., Thompson, J. L., McIntire, T. M., Milton, S. C., Cotman, C. W., and Glabe, C. G. (2003). Common structure of soluble amyloid oligomers implies common mechanisms of pathogenesis. *Science* **300**, 486–489.
- Klein, W. L., Stine, W. B., Jr., and Teplow, D. B. (2004). Small assemblies of unmodified amyloid  $\beta$ -protein are the proximate neurotoxin in Alzheimer's disease. *Neurobiol. Aging* **25**, 569–580.
- Koo, E. H., Lansbury, P. T., Jr., and Kelly, J. W. (1999). Amyloid diseases: Abnormal protein aggregation in neurodegeneration. *Proc. Natl. Acad. Sci. USA* **96**, 9989–9990.
- Kyte, J., and Doolittle, R. F. (1982). A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **157**, 105–132.
- Lazo, N. D., Grant, M. A., Condrón, M. C., Rigby, A. C., and Teplow, D. B. (2005). On nucleation of amyloid  $\beta$ -protein monomer folding. *Prot. Sci.* **14**, 1581–1596.
- Ma, B., and Nussinov, R. (2002). Stabilities and conformations Alzheimer's  $\beta$ -amyloid peptide oligomers ( $A\beta_{16-22}$ ,  $A\beta_{16-35}$ , and  $A\beta_{10-35}$ ): Sequence effect. *Proc. Natl. Sci. USA* **99**, 14126–14131.
- Ma, B., and Nussinov, R. (2004). From computational quantum chemistry to computational biology: Experiments and computations are (full) partners. *Phys. Biol.* **1**, P23–P26.

- Nguyen, H. D., and Hall, C. K. (2004a). Molecular dynamics simulations of spontaneous fibril formation by random-coil peptides. *Proc. Natl. Acad. Sci. USA* **101**, 16180–16185.
- Nguyen, H. D., and Hall, C. K. (2004b). Phase diagrams describing fibrillization by polyaniline peptides. *Biophys. J.* **87**, 4122–4134.
- Nguyen, H. D., and Hall, C. K. (2005). Kinetics of fibril formation by polyaniline peptides. *J. Biol. Chem.* **280**, 9074–9082.
- Peng, S., Ding, F., Urbanc, B., Buldyrev, S. V., Cruz, L., Stanley, H. E., and Dokholyan, N. V. (2004). Discrete molecular dynamics simulations of peptide aggregation. *Phys. Rev. E* **69**, 041908.
- Petkova, A., Ishii, T. Y., Balbach, J. J., Antzutkin, O. N., Leapman, R. D., Delaglio, F., and Tycko, R. (2002). A structural model for Alzheimer's  $\beta$ -amyloid fibrils based on experimental constraints from solid state NMR. *Proc. Natl. Acad. Sci. USA* **99**, 16742–16747.
- Rapaport, D. C. (1997). "The Art of Molecular Dynamics Simulation." Cambridge University Press.
- Sciarretta, K. L., Gordon, D. J., Petkova, A. T., Tycko, R., and Meredith, S. C. (2005). A $\beta$ 40-Lactam (D23/K28) models a conformation highly favorable for nucleation of amyloid. *Biochemistry* **44**, 6003–6014.
- Searle, M. S., Griffiths-Jones, S. R., and Skinner-Smith, H. (1999). Energetics of weak interactions in a  $\beta$ -hairpin peptide: Electrostatic and hydrophobic contributions to stability from lysine salt bridges. *J. Acad. Chem. Soc.* **121**, 11615–11620.
- Selkoe, D. J. (2001). Alzheimer's disease: Genes, proteins, and therapy. *Physiol. Rev.* **81**, 741–766.
- Smith, A. V., and Hall, C. K. (2001a).  $\alpha$ -Helix formation: Discontinuous molecular dynamics on an intermediate-resolution protein model. *Proteins* **44**, 344–360.
- Smith, A. V., and Hall, C. K. (2001b). Assembly of a tetrameric  $\alpha$ -helical bundle: Computer simulations on an intermediate-resolution protein model. *Proteins* **44**, 376–391.
- Snow, C. D., Nguyen, N., Pande, V. S., and Gruebele, M. (2002). Absolute comparison of simulated and experimental protein-folding dynamics. *Nature* **420**, 102–106.
- Takada, S., Luthey-Schulten, Z., and Wolynes, P. G. (1999). Folding dynamics with nonadditive forces: A simulation study of a designed helical protein and a random heteropolymer. *J. Chem. Phys.* **110**, 11616–11629.
- Taketomi, H., Ueda, Y., and Gō, N. (1975). Studies on protein folding, unfolding and fluctuations by computer simulations. *Int. J. Peptide Protein Res.* **7**, 445–459.
- Tsai, J., Taylor, R., Chothia, C., and Gerstein, M. (1999). The packing density in proteins: Standard radii and volumes. *J. Mol. Biol.* **290**, 253–266.
- Tycko, R. (2006). Characterization of amyloid structures at the molecular level by solid state nuclear magnetic resonance spectroscopy. *Methods Enzymol.* **413**, in press.
- Urbanc, B., Cruz, L., Ding, F., Sammond, D., Khare, S., Buldyrev, S. V., Stanley, H. E., and Dokholyan, N. V. (2004a). Molecular dynamics simulation of amyloid- $\beta$  dimer formation. *Biophys. J.* **87**, 2310–2321.
- Urbanc, B., Cruz, L., Yun, S., Buldyrev, S. V., Bitan, G., Teplow, D. B., and Stanley, H. E. (2004b). *In silico* study of amyloid  $\beta$ -protein (A $\beta$ ) folding and oligomerization. *Proc. Natl. Acad. Sci. USA* **101**, 17345–17350.
- Wesson, L., and Eisenberg, D. (1992). Atomic solvation parameters applied to molecular dynamics of proteins in solution. *Protein Sci.* **1**, 227–235.
- Xu, Y., Shen, J., Luo, X., Zhu, W., Chen, K., Ma, J., and Jiang, H. (2005). Conformational transition of amyloid  $\beta$ -peptide. *Proc. Natl. Acad. Sci. USA* **102**, 5403–5407.
- Zagrovic, B., Snow, C. D., Shirts, M. R., and Pande, V. S. (2002). Simulation of folding of a small  $\alpha$ -helical protein in atomistic detail using worldwide-distributed computing. *J. Mol. Biol.* **323**, 927–937.
- Zagrovic, B., and Pande, V. S. (2003). Solvent viscosity dependence of the folding rate of a small protein: Distributed computing study. *J. Comput. Chem.* **24**, 1432–1436.



- Zhang, C., Vasmatzis, G., Cornette, J. L., and DeLisi, C. (1997). Determination of atomic desolvation energies from the structures of crystallized proteins. *J. Mol. Biol.* **267**, 707–726.
- Zhang, C., Liu, S., Zhou, H., and Zhou, Y. (2004). The dependence of all-atom statistical potentials on structural training database. *Biophys. J.* **86**, 3349–3358.
- Zhang, S., Iwata, K., Lachenmann, M. J., Peng, J. W., Li, S., Stimson, E. R., Lu, Y., Felix, A. M., Maggio, J. E., and Lee, J. P. (2000). The Alzheimer's peptide A $\beta$  adopts a collapsed coil structure in water. *J. Struct. Biol.* **130**, 130–141.
- Zhou, Y., Hall, C. K., and Karplus, M. (1996). First-order disorder-to-order transition in an isolated homopolymer model. *Phys. Rev. Lett.* **77**, 2822–2825.
- Zhou, Y., Karplus, M., Wichert, J. M., and Hall, C. K. (1997). Equilibrium thermodynamics of homopolymers and clusters: Molecular dynamics and Monte-Carlo simulations of system with square-well interactions. *J. Chem. Phys.* **107**, 10691–10708.
- Zhou, Y., and Karplus, M. (1997). Folding thermodynamics of a three-helix-bundle protein. *Proc. Natl. Acad. Sci. USA* **94**, 14429–14432.
- Zhou, Y., and Karplus, M. (1999). Folding of a model three-helix bundle protein: A thermodynamic and kinetic analysis. *J. Mol. Biol.* **293**, 917–951.

## [20] Computational Approaches to Fibril Structure and Formation

By CAROL K. HALL and VICTORIA A. WAGONER

### Abstract

Assembly of normally soluble proteins into amyloid fibrils is a cause or associated symptom of numerous human disorders. Although some progress toward understanding the molecular-level details of fibril structure has been made through *in vitro* experiments, the insoluble nature of fibrils make them difficult to study experimentally. We describe two computational approaches used to investigate fibril formation and structure: intermediate-resolution discontinuous molecular dynamics simulations and atomistic molecular dynamics simulations. Each method has its strengths and weaknesses, but taken together the two approaches provide a useful molecular-level picture of fibril structure and formation.

### Introduction

The hallmark of many neurodegenerative diseases, including Alzheimer's disease, is the accumulation and deposition of protein plaques in specific tissues within various organs in the body (Koo, 2002). These plaques are composed of ordered protein aggregates known as amyloid fibrils that form when normally soluble disease-specific proteins undergo a conformational change that leads to their aberrant assembly. Many of the 24 known so-called