A novel canonical dual computational approach for prion AGAAAAGA amyloid fibril molecular modeling

Jiapu Zhang *, David Y. Gao, John Yearwood

Centre in Informatics and Applied Optimization, Graduate School of Sciences, Informatics Technology and Engineering, University of Ballarat, Mount Helen, VIC 3353, Australia

1. Introduction

According to a recent comprehensive review (Chou, 2011), to develop a useful model for biological systems, the following things were usually needed to consider: (i) the material of benchmark used to develop and test the model, (ii) the formulation of modeling method, (iii) operating procedures during the modeling process, (iv) properly perform the cross-validation tests to objectively evaluate the anticipated accuracy of the model, and (v) web-server establishment. Below, let us elaborate some of these procedures. In this paper, the material used to develop the model is 3NHC.pdb and its 3D-crystal structures; the modeling method is the Mathematical Optimization methods of the canonical dual theory (CDT) (Gao, 2000; Gao et al., 2010; Gao and Wu, to appear) (procedure 1) and of the Amber 11 package’s steepest descent (SD) method (Case et al., 2010) and conjugate gradient (CG) method (Case et al., 2010; Sun and Zhang, 2001) (procedure 2); and the test to the accuracy of the model is performed by the RMSD (root-mean-square deviation) value of last snapshots between procedures 1 and 2.

Various computational molecular dynamics approaches have been used to study PrP (106–126) (Kuwata et al., 2003; Wagoner, 2010) but, to the best of our knowledge, to predict molecular structures of prion AGAAAAGA amyloid fibrils the computational approaches are few (Zhang, 2011; Zhang et al., 2011). Zhang (2011) successfully constructed three AGAAAAGA amyloid fibril models by the standard simulated annealing (SA) method and several traditional optimization methods within AMBER 10 package. In Zhang et al. (2011), the hybrid simulated annealing discrete gradient (SADG) method was successfully used for modeling two AGAAAAGA amyloid fibril models (instead of the Insight II (http://accelrys.com) package used in Zhang, 2011), and then the models were refined/optimized by the SDCG methods, SA method and SDCG methods again. In this paper, all the optimization approaches of Zhang (2011) and Zhang et al. (2011) will be replaced by the optimization theory of CDT. Numerical computational results show that the optimization approaches of CDT have a very perfect performance. It is even no need to do furthermore SDCG refinements by the AMBER package. We could not do comparisons (for example, the angstrom values between adjacent β-sheets and β-strands) for the models of Zhang (2011) and Zhang et al. (2011) and of this paper, because these models have different number of chains and different structural classes listed in Kuwata et al. (2003).

As we all know, the disease prions PrPSc are rich in β-sheets amyloid fibrils (about 43% β-sheet) (Griffith, 1967). There are some classical works on the β-sheets and β-barrels (Chou et al., 1983a, 1990a,b, 1991). X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy are two powerful tools to determine the protein 3D structure. However, not all proteins can be successfully crystallized, particularly for membrane proteins. Although NMR is indeed a very powerful tool in determining the
3D structures of membrane proteins (see, e.g. Schnell and Chou, 2008; Oxenoid and Chou, 2005; Call et al., 2010; Pielak and Chou, 2010; Pielak et al., 2009; Wang et al., 2009 and a recent review (Pielak and Chou, 2011)), it is also time-consuming and costly. To acquire the structural information in a timely manner, one has to resort to various structural bioinformatics tools (see, e.g. Chou, 2004b–d, 2005b) and a comprehensive review (Chou, 2004c). Particularly, computational approaches allow us to obtain a description of the protein 3D structure at a submicroscopic level. Under many circumstances, due to the unstable, noncrystalline and insoluble nature of the amyloid fibrils, it is very difficult to use traditional X-ray and NMR experimental methods to obtain atomic-resolution structures of amyloid fibrils (Tsai, 2005; Zheng et al., 2006). Although X-ray and NMR techniques cannot determine the 3D structures of some proteins and their binding interactions with ligands in a timely manner that are important for drug design and basic research, many structural bioinformatics tools can play a complementary role in this regard as demonstrated by a series of papers published recently (see, e.g. Cai et al., 2011; Chou, 2004a, 2005a; Chou et al., 2003; Du et al., 2007, 2010; Gong et al., 2009; Liao et al., 2011; Wang and Chou, 2010; Wang et al., 2009; Wei et al., 2009). This paper, in some sense, presents a structural bioinformatics tool in view of the CDT-based mathematical optimization theory.

The accuracy of the models presented in this paper is tested by the RMSD value. The last snapshot of procedure 2 will be superposed onto the last snapshot of procedure 1, and the RMSD value is zero after the alignment by VMD 1.8.7beta5 (Humphrey et al., 1996). This implies to us that the CDT strategy can accurately build the prion AGAAAAGA amyloid fibril models. To test the accuracy of their model, some examination validation methods are always used. In developing a prediction model or algorithm, the following three cross-validation methods are often used for examining its effectiveness in practical application: independent dataset test, subsampling (5-fold or 10-fold cross-validation) test, and jackknife test (Chou and Zhang, 1995). However, as demonstrated by Chou (2011, Eqs. (28)–(32)), among the three cross-validation methods, the jackknife test is deemed the least arbitrary that can always yield a unique result for a given benchmark dataset, and hence has been increasingly used and widely recognized by investigators to examine the accuracy of various models and predictors (see, e.g. Chen et al., 2009; Chou et al., 2011; Ding et al., 2009; Hayat and Khan, 2011; Kandaswamy et al., 2011; Lin and Ding, 2011; Mohabatkar, 2010; Zeng et al., 2009; Zhou et al., 2007; all these papers reflect the current trend of increasingly and widely using the jackknife test to examine varieties of models or predictors).

There is another criteria to evaluate the models. To avoid homology bias and remove the redundant sequences from the benchmark dataset, a cutoff threshold of 25% was recommended (Chou, 2011; Chou et al., 2011) to exclude those proteins from the benchmark datasets that have equal to or greater than 25% sequence identity to any other. However, in this study we did not use such a stringent criterion because the currently available benchmark dataset, a cutoff threshold of 25% was recommended (Chou, 2011; Chou et al., 2011) to exclude those proteins from the benchmark datasets that have equal to or greater than 25% sequence identity to any other. However, in this study we did not use such a stringent criterion because the currently available benchmark dataset, a cutoff threshold of 25% was recommended (Chou, 2011; Chou et al., 2011) to exclude those proteins from the benchmark datasets that have equal to or greater than 25% sequence identity to any other. However, in this study we did not use such a stringent criterion because the currently available benchmark dataset, a cutoff threshold of 25% was recommended (Chou, 2011; Chou et al., 2011) to exclude those proteins from the benchmark datasets that have equal to or greater than 25% sequence identity to any other. However, in this study we did not use such a stringent criterion because the currently available benchmark dataset, a cutoff threshold of 25% was recommended (Chou, 2011; Chou et al., 2011) to exclude those proteins from the benchmark datasets that have equal to or greater than 25% sequence identity to any other. However, in this study we did not use such a stringent criterion because the currently available benchmark dataset, a cutoff threshold of 25% was recommended (Chou, 2011; Chou et al., 2011) to exclude those proteins from the benchmark datasets that have equal to or greater than 25% sequence identity to any other. However, in this study we did not use such a stringent criterion because the currently available benchmark dataset, a cutoff threshold of 25% was recommended (Chou, 2011; Chou et al., 2011) to exclude those proteins from the benchmark datasets that have equal to or greater than 25% sequence identity to any other.

2. The canonical dual approach

We briefly introduce the CDT of Gao et al. (2010), Gao and Wu (to appear), and Gao (2000) specially for solving the following minimization problem of the sum of fourth-order polynomials:

\[
\min_{x} \left\{ P(x) = \sum_{i=1}^{m} W_i(x_i) + \frac{1}{2} \lambda^T Q x - x^T f : x \in \mathbb{R}^n \right\},
\]

where \( W_i(x_i) = \frac{1}{2} \sum_{j=1}^{d_i} x_i^j (x_i + c_i)^j \) and \( x, Q \in \mathbb{R}^{n \times n} \). This problem is a critical point of the fourth-order polynomial, \( \int \int \int \int \sum_{i=1}^{m} W_i(x_i) + \frac{1}{2} \lambda^T Q x - x^T f : x \in \mathbb{R}^n \). The dual problem of \( P(x) \) is

\[
\max_{\lambda, \gamma} \left\{ P^d(\lambda, \gamma) = \sum_{i=1}^{m} \left( \gamma_i - \frac{1}{2} \lambda_i^T x_i^2 \right) - \frac{1}{2} \lambda^T F(\lambda) G^T(\lambda)^T F(\lambda)^T : \gamma \in \mathbb{R}^n \right\},
\]

where \( F(\lambda) = f - \sum_{i=1}^{m} \gamma_i b_i \), \( G(\lambda) = Q + \sum_{i=1}^{m} \gamma_i A_i \), \( \gamma = \{ \gamma \in \mathbb{R}^n | F(\lambda) \in \text{Col}(G(\lambda)) \} \), and \( x^* \) denotes the Moore–Penrose generalized inverse of \( G \) and \( \text{Col}(G(\lambda)) \) is the column space of \( G(\lambda) \). The prime-dual Gao–Strang complementary function of \( P(\lambda, \gamma) \) is a critical point of \( P(\lambda, \gamma) \) over \( S^2 \), \( x^* = \text{Col}(G(\lambda)) \), and

\[
E(x, \gamma) = \sum_{i=1}^{m} \left( \frac{1}{2} x^T A_i x + b_i^T x + c_i \right) \gamma_i - \frac{1}{2} \gamma^T \gamma \right\}.
\]

(1)

For \( P(x) \) and \( P^d(\lambda, \gamma) \) we have the following CDT:

**Theorem 1** (Gao et al., 2010; Gao and Wu, to appear; Gao, 2000). The problem \( P(x) \) is canonically dual to \( P(\lambda, \gamma) \) in the sense that if \( x^* \) is a critical point of \( P(x) \), then \( x = G^T(\gamma)^T F(\gamma) \) is a critical point of \( P(x) \) and \( P(\lambda, \gamma) \). Moreover, if \( x \in S^2 \), then \( G^+ \) is a global minimizer of \( P^d(\lambda, \gamma) \) over \( S^2 \), and \( P(\lambda, \gamma) = \max_{\gamma \in \mathbb{R}^n} P^d(\lambda, \gamma) = P^d(\lambda) \).

It is easy to prove that the canonical dual function \( P^d(\lambda) \) is concave on the convex dual feasible space \( S^2 \). Therefore, Theorem 1 shows that the nonconvex primal problem \( P(x) \) is equivalent to a concave maximization problem \( P^d(\lambda) \) over a convex space \( S^2 \), which can be solved easily by well-developed methods. Over \( S^2 \), \( x^* = \text{Col}(G(\lambda)) \), and we have the following theorem:

**Theorem 2** (Gao and Wu, to appear). Suppose that \( \gamma \) is a critical point of \( P^d(\lambda) \) and the vector \( x \) is defined by \( \gamma = G^T(\gamma)^T F(\gamma) \). Then \( x \in S^2 \), then on a neighborhood \( x_0 \subset X \), \( x^* = \text{Col}(G(\lambda)) \), and we have either

\[
P(\lambda, \gamma) = \max_{\lambda, \gamma} \left\{ P(\lambda, \gamma) = \sum_{i=1}^{m} \left( \gamma_i - \frac{1}{2} \lambda_i^T x_i^2 \right) - \frac{1}{2} \lambda^T F(\lambda) G^T(\lambda)^T F(\lambda)^T : \gamma \in \mathbb{R}^n \right\},
\]

(2)
Fig. 1. The prime and dual Double Well Potential functions (Prime: blue, Dual: red). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

or

\[ P(x) = \max_{x \in X} P(x) = \mathbb{E}(x, \zeta) = \max_{\zeta \in S_\zeta} P^d(\zeta) = P^d(\zeta). \]  

(6)

By the fact that the canonical dual function is a d.c. function (difference of convex functions) on \(S_\zeta\), the double-min duality (5) can be used for finding the biggest local minimizer of \((P)\) and \((P^d)\), while the double-max duality (6) can be used for finding the biggest local maximizer of \((P)\) and \((P^d)\). In physics and material sciences, this pair of biggest local extremal points play important roles in phase transitions.

To illuminate that the CDT works, we minimize the well-known Double Well Potential (DWP) function \((\text{Gao, 2000})\) (blue colored in Fig. 1):

\[ P(x) = \frac{1}{2} x^2 - 2x^3 - x. \]

We can easily get \(\mathbb{E}(x, \zeta) = (\frac{1}{2} x^2 - 2x^3 - \frac{1}{2} x), \)

\[ P^d(\zeta) = -\frac{1}{2} \zeta^2 - 2\zeta \]  

(8)

(red colored in Fig. 1) and \(S_\zeta = \{ \zeta \in \mathbb{R}_+ \mid \zeta > 0 \} \). Let \(\mathbb{E}(x, \zeta) = 0\), we get three critical points of \(\mathbb{E}(x, \zeta): (x_1, \zeta^1) = (2.11491, 0.236417), (x_2, \zeta^2) = (-1.86081, -0.268701), (x_3, \zeta^3) = (-0.254102, -1.967722). \) By Theorem 1, we know \(x^1 = 2.11491\) is the global minimizer of \((7)\), \(x^1 = 0.236417\) is the global maximizer of \((7)\) over \(S_\zeta\), and \(P(x^1) = \mathbb{E}(x^1, \zeta^1) = P^d(\zeta^1) = -1.02951. \) By Theorem 2, we know that the local minimizers: \(x^2 = -1.86081, x^2 = -0.268701\) (over \(S_\zeta\)), \(P(x^2) = \mathbb{E}(x^2, \zeta^2) = P^d(\zeta^2) = 0.9665031\) and the local maximizers: \(x^3 = -0.254102, x^3 = -1.967722\) (over \(S_\zeta\)), \(P(x^3) = \mathbb{E}(x^3, \zeta^3) = P^d(\zeta^3) = 2.063. \) Thus, by Fig. 1 illuminating the application of CDT to the DWP problem, we may see that the canonical dual approach works. The powerful of canonical dual approach is preliminarily shown in Tables 1–3 of arXiv:1105.2270v3 (http://arxiv.org/PS_cache/arxiv/pdf/1105/1105.2270v3.pdf). In the next section, we will apply this successful canonical dual approach to the molecular model building and solving problem of prion AGAAAAGA amyloid fibrils.

3. Prion AGAAAAGA amyloid fibril molecular model building and solving

Many experimental studies such as Brown (2000, 2001), Brown et al. (1994), Holscher et al. (1998), Jobling et al. (2001, 1999), Kuwata et al. (2003), Norstrom and Mastriani (2005), and Wegner et al. (2002) have shown two points: (1) the hydrophobic region (113–120) AGAAAAGA of prion proteins is critical in the conversion from a soluble PrP\(^C\) into an insoluble PrP\(^Sc\) fibrillar form and (2) normal AGAAAAGA is an inhibitor of prion diseases. Various computational approaches were used to address the problems related to “amyloid fibril” (Carter and Chou, 1998; Chou, 2004b, 2004c; Chou and Howe, 2002; Wang et al., 2008; Wei et al., 2005; Zhang, 2011; Zhang et al., 2011; Zhang, 2009). By introducing novel mathematical canonical dual formulations and computational approaches, in this paper we may construct atomic-resolution molecular structures for prion (113–120) AGAAAAGA amyloid fibrils.

The atomic structures of all amyloid fibrils revealed steric zippers, with strong van der Waals interactions between β-sheets and hydrogen bonds to maintain the β-strands (Sawaya et al., 2007). About β-sheets and β-barrels, there are various interactions and motions, such as the interactions between β-strands (Chou et al. 1982a,b, 1983a,b), interaction between two β-sheets (Chou et al., 1986), as well as the low-frequency accordian-like motion in a β-sheet and breathing motion in a β-barrel (Chou, 1985) and their biological functions (Chou, 1988). The “amyloid fibril” problem can be looked as a molecular distance geometry problem (MDGP) (Grosso et al., 2009), which arises in the interpretation of NMR data and in the determination of protein structure [as an example to understand MDGP, the problem of locating sensors in telecommunication networks is a DGP. In such a case, the positions of some sensors are known (which are called anchors) and some of the distances between sensors (which can be anchors or not) are known: the DGP is to locate the positions of all the sensors. Here we look sensors as atoms and their telecommunication network as a molecule]. The three dimensional structure of a molecule with \(n\) atoms can be described by specifying the 3-dimensional coordinate positions \(x_1, x_2, \ldots, x_n \in \mathbb{R}^3\) of all its atoms. Given bond lengths \(d_i\) between a subset \(S\) of the atom pairs, the determination of the molecular structure is

\[ \{P_d\} \text{ to find } x_1, x_2, \ldots, x_n \text{ s.t. } \|x_i - x_j\| = d_{ij}, \quad (i, j) \in S, \]

where \(\| \cdot \|\) denotes a norm in a real vector space and it is calculated as the Euclidean distance 2-norm in this paper. Eq. (9) can be reformulated as a mathematical global optimization problem (GOP)

\[ \min P(X) = \sum_{(i, j) \in S} w_{ij} \|x_i - x_j\|^2 - d^2_{ij} \]

(10)

in the terms of finding the global minimum of the function \(P(X)\), where \(w_{ij}(i, j) \in S\) are positive weights, \(X = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^{n \times 3}\) (More and Wu, 1997) and usually \(S\) has many fewer than \(n^2/2\) elements due to the error in the theoretical or experimental data (Zou et al., 1997; Grosso et al., 2009). There may even not exist any solution \(x_1, x_2, \ldots, x_n\) to satisfy the distance constraints in (9), for example when data for atoms \(i, j, k \in S\) violate the triangle inequality; in this case, we may add a perturbation term \(-\epsilon X\) to \(P(X)\):

\[ \min P_e(X) = \sum_{(i, j) \in S} w_{ij} \|x_i - x_j\|^2 - d^2_{ij} - \epsilon^T X \]

(11)

where \(\epsilon \geq 0\). In some cases, instead exact values \(d_{ij}(i, j) \in S\) can be found, we can only specify lower and upper bounds on the distances: \(l_{ij} \leq \|x_i - x_j\| \leq u_{ij}(i, j) \in S\); in such cases we may penalize all the unsatisfied constraints into the objective function of \((P_e)\) by adding \(\sum_{(i, j) \in S} \max(l_{ij}^2 - \|x_i - x_j\|^2, 0^2) + \max(\|x_i - x_j\|^2 - u_{ij}^2, 0^2)\)

into \(P_e(X)\) (Zou et al., 1997; Grosso et al., 2009), where we may let \(d_{ij}\) be the interatomic distance (less than 6 Å) for the pair in successive residues of a protein and set \(l_{ij} = (1 - 0.05)d_{ij}\) and \(u_{ij} = (1 + 0.05)d_{ij}\) (Grosso et al., 2009). In this paper we will use
the canonical duality approach introduced in Section 2 (Gao et al., 2010; Gao and Wu, to appear; Gao, 2000) to solve (9)–(11). Because the canonical dual is a perfect dual with zero duality gap between prime and dual problems, we can get the accurate global optimal solutions of (9)–(11). Thus by canonical dual approach we may successfully construct the molecular structure of prion AGAAAAGA amyloid fibrils as follows.

If we look at the prion AGAAAAGA molecular modeling problem as a MDGP with two anchors and two sensors, we can easily construct the prion AGAAAAGA amyloid fibril models. In fact we may let the coordinates of these two anchors being variable. But, these two anchors belong to one body of Chains A and B, and the two sensors belong to another body of Chains G and H. This is a simple two-body problem model of theoretical physics, i.e. Einstein’s absolute relative theory. Hence, we may look the coordinates of two anchors being fixed. The constructions will be based on the most recently released experimental molecular structures of human M129 prion peptide 127–132 (PDB entry 3NHC released into Protein Data Bank (http://www.rcsb.org) on 04 August 2010) (in brief, this paper will use the PrP structured region 127–132 to do homology modeling for the PrP unstructured region 113–120). The atomic-resolution structure of this peptide is a steric zipper, with strong van der Waals (vdw) interactions between β-sheets and hydrogen bonds to maintain the β-strands (Fig. 2, where the dashed lines denote the hydrogen bonds).

In Fig. 2 we see that G (H) chains (i.e. β-sheet 2) of 3NHC.pdb can be obtained from A (B) chains (i.e. β-sheet 1) by

\[
G(H) = \begin{pmatrix}
1 & 0 & 0 \\
0 & -1 & 0 \\
0 & 0 & -1
\end{pmatrix} A(B) + \begin{pmatrix}
9.07500 \\
4.77650 \\
0.00000
\end{pmatrix}.
\] (12)

And other chains can be obtained by

\[
l(J) = G(H) + \begin{pmatrix}
0 \\
9.5530 \\
0
\end{pmatrix}, \quad k(L) = G(H) + \begin{pmatrix}
0 \\
-9.5530 \\
0
\end{pmatrix},
\]

\[
c(D) = A(B) + \begin{pmatrix}
0 \\
9.5530 \\
0
\end{pmatrix}, \quad e(F) = A(B) + \begin{pmatrix}
0 \\
-9.5530 \\
0
\end{pmatrix}.
\] (13)

Basing on the template 3NHC.pdb from the Protein Data Bank, three prion AGAAAAGA palindrome amyloid fibril models—an AGAAAA model (Model 1), a GAAAAG model (Model 2), and an AAAAGA model (Model 3)—will be successfully constructed in this paper. AB chains of Models 1–3 were, respectively, obtained from AB chains of 3NHC.pdb using the mutate module of the free package SwissPdbViewer (SPDBV Version 4.01) (http://spdbv.vital-it.ch). It is pleasant to see that almost all the hydrogen bonds are still kept after the mutations, where for the donor O (oxygen) atom and the acceptor H (hydrogen) atom if the distance cutoff is less than 3.00 Å and the angle cutoff is less than 120.00° then a hydrogen bond is maintained.

**Fig. 2.** Protein fibril structure of human M129 prion GYMLGS (127–132).

**Fig. 3.** Far vdw contacts of AG chains and BH chains of Model 1.
kept; thus we just need to consider the vdw contacts only. Making mutations for GH chains of 3NHC.pdb, we can get the GH chains of Models 1–3. However, the vdw contacts between A chain and G chain, between B chain and H chain are too far at this moment (Figs. 3–5) because the shortest distance of atoms between Chain A and Chain G, and between Chain B and Chain H, is still very larger than the double size of the vdw radius of CB carbon atom.

Seeing Figs. 3–5, we may know that for Models 1–3 at least two vdw interactions between A.ALA3.CB-G.ALA4.CB, B.ALA4.CB-H.ALA3.CB should be maintained. Fixing the coordinates of A.ALA3.CB and B.ALA4.CB (two anchors) ((6.014,5.917,0.065), (5.658,1.630,−0.797)), letting \( d = \) the twice of the vdw radius of Carbon atom (i.e. \( d = 3.4 \) Å), and letting the coordinates of G.ALA4.CB and H.ALA3.CB (two sensors) be variables, we may get a simple MDGP with six variables and its dual with two variables:

\[
P(x_1, x_2) = \frac{1}{2}(x_{11}−6.014)^2 + (x_{12}−5.917)^2 + (x_{13}−0.065)^2 − 3.4^2)^2 \\
+ \frac{1}{2}(x_{21}−5.658)^2 + (x_{22}−1.630)^2 + (x_{23}+0.797)^2 − 3.4^2)^2.
\]

Fig. 4. Far vdw contacts of AG chains and BH chains of Model 2.

We can get a local maximal solution (−11.56, −11.56) for \( P^d(\zeta_1, \zeta_2) \) and its corresponding local maximal solution to \( P(x_1,x_2) \). But we need the global maximal solution of \( P^d(\zeta_1, \zeta_2) \). Thus, by introducing perturbation parameters \( \epsilon = 0.05 \), we have to seek the global optimal solutions from the perturbed problems of \( P(x_1,x_2) \) and \( P^d(\zeta_1, \zeta_2) \):

\[
P^\epsilon(x_1,x_2) = \frac{1}{2}(x_{11}−6.014)^2 + (x_{12}−5.917)^2 + (x_{13}−0.065)^2 − 3.4^2)^2 \\
+ \frac{1}{2}(x_{21}−5.658)^2 + (x_{22}−1.630)^2 + (x_{23}+0.797)^2 − 3.4^2)^2 \\
−0.05x_{11} − 0.05x_{12} − 0.05x_{13} − 0.05x_{21} − 0.05x_{22} − 0.05x_{23},
\]

\[
P^\epsilon(\zeta_1,\zeta_2) = 59.6233\zeta_1−0.5\zeta_2^2 + 23.7451\zeta_2−0.5\zeta_2^2 \\
− \frac{1}{2} \left( \frac{(0.05 + 12.028\zeta_1)^2}{2\zeta_1} + \frac{(0.05 + 11.834\zeta_1)^2}{2\zeta_1} \right)
\]

Fig. 5. Far vdw contacts of AG chains and BH chains of Model 3.
We can easily get the global maximal solution \((0.0127287, 0.0127287)\) for \(P_d^i(\tau_1, \tau_2)\). Then, we get its corresponding solution for \(P_E(x_1, x_2)\):

\[
(x_1, x_2) = \left(7.97807, 7.88107, 2.02907, 7.62207, 3.59407, 1.16707\right).
\]

By Theorem 1 we know that \(x\) is a global minimal solution of \(P_E(x_1, x_2)\). We set \(x\) as the coordinates of G.ALA4.CB and H.ALA3.CB and taking the average value we get

\[
G(H) = \begin{pmatrix}
1 & 0 & 0 \\
0 & 1 & 0 \\
0 & 0 & -1
\end{pmatrix} A(B) + \begin{pmatrix}
1.96407 \\
9.51107 \\
1.23207
\end{pmatrix}.
\]

By (15) we can get very close vdw contacts between A chain and G chain, between B chain and H chain (Figs. 6–8).

Thus, we successfully constructed Models 1–3, and through further refinements by the Amber 11 package (Case et al., 2010) we at last get the optimal Models (Figs. 9–11). We find the RMSD (root mean square deviation) between Figs. 6–8 and Figs. 9–11 is 0 Å; this implies that the Amber 11 refinements are not necessary and the CDT is good enough to get the optimal Models 1–3 as illuminated in Figs. 6–8. The other CDIJ and EFKL chains can be obtained by parallelizing ABGH chains in the use of mathematical formulas (13) and (14).

As the end of this section, we give some remarks on the Models 1–3. (1) The canonical dual approach exactly makes the closest CB atoms between Chain A and Chain G, and between Chain B and Chain H, just being equal to the double size of the vdw radius of CB carbon atom (Figs. 6–8) and this is the perfect structure of the Models 1–3. Figs. 9–11 were obtained by the further refinements.
through the SDCG optimization methods of Amber 11 package. The zero RMSD value between Figs. 6–8 and Figs. 9–11 implies to us that the canonical dual approach of this paper works well. (2) The SDCG optimization methods of Amber 11 package automatically considered the bond angles and dihedral angles, and during the canonical dual molecular model building and optimization procedure, the perfect bond angles and dihedral angles automatically produced by the Swiss-PdbViewer package are still being kept. (3) The molecular modeling problem of this paper is in fact a very simple two-body problem of theoretical physics, i.e. Einstein’s absolute relative theory. In mathematics, it is a sensor network problem with two anchors and two sensors.

4. Conclusion

This paper presents an important method and provides useful information for treatments of prion diseases. X-ray crystallography is a powerful tool to determine the protein 3D structure. However, it is time-consuming and expensive, and not all proteins can be successfully crystallized, particularly for membrane proteins. Although NMR spectroscopy is indeed a very powerful tool in determining the 3D structures of membrane proteins, it is also time-consuming and costly. Due to the noncrystalline and insoluble nature of the neurodegenerative amyloid fibril or plaque, little structural data on the prion AGAAAAGA segment is available. Under these circumstances, the novel canonical dual computational approach introduced in this paper showed its power in the molecular modeling of prion AGAAAAGA amyloid fibrils. This indicated that computational approaches or introducing novel mathematical formulations and physical concepts into molecular biology can significantly stimulate the development of biological and medical science. The optimal atomic-resolution structures of prion AGAAAAGA amyloid fibrils presented in this paper are useful for the drive to find treatments for prion diseases in the field of medicinal chemistry.
Acknowledgments

This research is supported by US Air Force Office of Scientific Research Under the grant AFOSR FA9550-10-1-0487, and by a Victorian Life Sciences Computation Initiative (http://vlsci.org.au) grant number VR0063 on its Peak Computing Facility at the University of Melbourne. Last, but not the least, the authors appreciate the anonymous referees for their numerous insightful comments to improve this paper, and appreciate the great helps from Ms Janet Stein Journal Manager-JTB.

References


Fig. 10. Optimal structure of prion AGAAAAGA amyloid fibril Model 2.


Fig. 11. Optimal structure of prion AGAAAAGA amyloid fibril Model 3.


