FEBS Letters 580 (2006) 6199–6205

Flexible docking of an amyloid-forming peptide from β_2-microglobulin

Daron M. Standley[a,b,*], Yasushige Yonezawa[a], Yuji Goto[a], Haruki Nakamura[a]

[a] *Institute for Protein Research, Osaka University, 3-2 Yamadaoka, Suita, Osaka 565-0871, Japan*
[b] *Japan Science and Technology Agency, Institute for Bioinformatics Research and Development (BIRD), Japan*

Received 11 September 2006; revised 2 October 2006; accepted 9 October 2006

Available online 18 October 2006

Edited by Miguel De la Rosa

Abstract **Using an all-atom, molecular dynamics-based, flexible docking method, the tertiary and quaternary structures of protofilaments of the "K3" fragment from β_2-microglobulin (residues Ser20-Lys41) were predicted at low pH in a continuous mixture of water and 2,2,2-trifluoroethanol (TFE). Tetramers with energies very close to the global minimum were produced with C_α root-mean square deviation values under 4 Å over 88 residues compared to a recently solved SSNMR structure. The most accurate model distinguishes itself from other low-energy solutions in that it shows high structural similarity to another known fold, the parallel β-helix, in agreement with models proposed previously by several other groups. The method achieves efficiency without loss of generality or atomic detail by enforcing local symmetry on the individual peptides, rewarding intermolecular contacts, and iteratively building up the protofilaments by successively doubling the number of chains. Solvent effects were included in the model by treating the dielectric constant and surface tension as functions of the TFE concentration. In order to understand the physical basis for the stabilizing effects of TFE, the TFE concentration was varied from 0% to 50% (v/v) and a peak in stability was observed at 16%, where the polar and hydrophobic terms cancel out and close to the experimentally determined value of 20%.**

Keywords: Amyloid; Flexible docking; Symmetric restraints; Restrained molecular dynamics; β_2-microglobulin; Alcohol effects

1. Introduction

Protein misfolding, in particular aggregation into elongated, unbranched structures, known as amyloid fibrils, is associated with a number of serious diseases such as Alzheimer's, Parkinson's, type II diabetes, and dialysis-related amyloidosis. Experimental studies of amyloid fibril formation are complicated by the fact that the fibrils generally appear as non-crystalline aggregates, which are not amenable to standard solution nuclear magnetic resonance (NMR) or X-ray crystallography structure determination methods [1,2], and by the fact that it is often necessary to prepare fibrils in vitro under non-physiological conditions [3,4]. Nevertheless, a general consensus has emerged, based on a wide body of experimental data, that fibrils formed from various protein sequences are quite similar [5]. In particular, the "cross-β" motif, in which β-strands align perpendicular to the fibril axis, characterizes all known amyloid fibrils and thus represents a significant structural constraint [1,5]. The difficulties inherent in experimental structure determination, along with the shared structural features of fibrils arising from diverse amino acid sequences, make computational approaches a potentially valuable means of investigating their structure and stability.

Traditional protein folding and docking methods are, however, complicated in the case of amyloid fibrils for several reasons. One reason is that the number of states grows exponentially with the number of chains if the chains are treated independently. This problem is exacerbated by the fact that fibril formation is a rare event so the state of interest is generally not the most populated. Furthermore, it is not obvious how non-physiological conditions, such as alcohol effects or low pH should be incorporated into the physical model for such simulations. To date, much of the theoretical work concerning amyloid fibril formation has been based on atomic-level molecular dynamics simulations of monomers [6,7] or dimers [8], or on rigid or coarse-grained docking simulations [9]. All-atom simulations using physics-based potential energy functions and equations of motion are obviously the most realistic. However, protein aggregation, by definition, involves interactions between multiple protein chains, so it is not certain how much mechanistic information can be drawn from the simulation of only 1 or 2 chains. On the other hand, while coarse-grained models have the advantage of efficiently treating interactions between numerous chains, the mechanism of amyloid formation is not yet understood, so it is not clear what simplifications in the model are appropriate in a molecular-level simulation. Clearly, the experimental and theoretical difficulties are coupled: without unambiguous structural constraints, models of amyloid fibrils derived from folding or docking simulations cannot be verified, making improvement of simulation methods in this important area difficult.

β_2-Microglobulin (bm2), the light chain of the major histocompatibility complex class I antigens, is a case in point. Bm2 is a natural product of the immune system and is normally filtered from the blood in the kidney but forms fibrils in patients undergoing prolonged hemodialysis treatment. A large number of computational studies of bm2 have been performed from either the full-length protein or shorter fragments, resulting in a diverse set of proposed fibril architectures, including the antiparallel β-helix [10], zipper-spine [11], extended antiparallel sheet [8], and extended β-hairpin [7]. Each

*Corresponding author. Fax: +81 6 6879 8636.
E-mail address: standley@protein.osaka-u.ac.jp (D.M. Standley).

Abbreviations: bm2, β_2-microglobulin; TFE, 2,2,2-trifluoroethanol; RMSD, root-mean-square deviation; MD, molecular dynamics; NMR, nuclear magnetic resonance; SSNMR, solid state NMR

0014-5793/$32.00
doi:10.1016/j.febslet.2006.10.022

of these models is at least partially consistent with known experimental data, but without knowing the tertiary and quaternary structure of the amyloid-form of bm2 in detail, we cannot say which is the most accurate or biologically relevant.

The "K3" fragment of bm2, consisting of residues Ser20-Lys41, forms fibrils readily under acidic conditions and moderate amounts of alcohol [12]. The morphology, solid-state NMR (SSNMR) and fiber diffraction spectra of the K3 fibrils resemble those of full-length bm2 fibrils formed in vivo, suggesting that K3 may be a minimal or essential amyloidogenic sequence [13]. Recently, an SSNMR-based structural study of K3 fibrils was undertaken [14], making this peptide an ideal subject for testing the proposed docking method.

In the present work, all calculations were carried out using standard molecular mechanics potential energy functions without the use of experimentally derived restraints or adjustment of parameters to fit the experimental data. Several general assumptions about the fibril state were made that allow the system to be treated efficiently without loss of detail. Since the method does not depend on external restraints, it is general enough to be applicable to other amyloidogenic protein sequences. Knowledge of the SSNMR model allows the accuracy of the predicted structures to be discussed and suggestions for future improvements to be made. We also describe the similarity of the predicted structures to other known folds.

Fibril formation of the K3 peptide is accelerated by the addition of alcohol under acidic conditions. The optimal amount of alcohol has been shown to depend on the hydrophobicity of the alcohol species. These effects have been attributed to a balance between polar interactions, which increase in strength with solvent hydrophobicity, and hydrophobic protein-protein interactions, which decrease in strength with the solvent hydrophobicity [15]. In order to include such effects, the hydrophobic and electrostatic properties of the water-alcohol mixture were explicitly included in the solvation energy. In the GBSA solvation model, the polar and hydrophobic forces can be considered simple functions of the dielectric constant [16] and surface tension [17], respectively. By adjusting the dielectric constant and surface tension, we can examine the stability of fibrils over a wide range of 2,2,2-trifluoroethanol (TFE) concentrations.

2. Materials and methods

We make two assumptions about fibril structure that allow us to reduce the complexity of a multi-chain system dramatically. The first is that in the fibril form, all of the monomers adopt nearly the same local conformation. This assumption cannot be justified a priori; however, the authors are not aware of any proposed fibril model in which it does not hold. In a system of chains with only one effective local conformation, the number of degrees of freedom is equal to that of a single chain, plus the relative rotational and translational degrees of freedom of the individual chains. Thus, the complexity of the system scales linearly with the number of chains. Symmetry is enforced through intramolecular distance restraints, in analogy to the method used by Nilges for the NMR structure determination of homodimers [18]. This type of local symmetry has been employed in other computational studies of protein aggregation [19,20], the primary difference in the present work being the focus on structure prediction, rather than on the dynamics of a multichain system.

The rotational and translational degrees of freedom are further reduced by a contact potential that rewards intermolecular contacts between consecutive chains. The contact potential is a Gaussian function of the same form as that used previously for determining the number of equivalent residues in a structure alignment [21]. The second assumption is that fibrils can be formed by a build-up procedure where first the monomers are docked to form dimers, then dimers are docked to form tetramers, and so on. The above two assumptions are artificial, and are not meant to accurately reflect the actual dynamics of fibril formation. The reduction in the conformational space is sufficient, however, for an accurate atomic-level treatment of a multi-chain system using conventional molecular dynamics (MD) methods. Moreover, the restraints are not considered part of the physical model and thus are not used in the final energy evaluation.

Each docking run consisted of a 15-ps annealed MD simulation, where the temperature was lowered linearly from 300 to 50 K. The timestep used was 1.5 fs. Since the distance constraints (see below) force the molecule to move along what is essentially a collective coordinate trajectory, a large time-step is justified [22]. All calculations were performed using the *cosgene* program, part of the myPresto system, which is available from the web [23] and is free for academic use [24]. Consistent with acidic experimental conditions, acidic residues were protonated, as was the N-terminal Serine and Histidine 31.

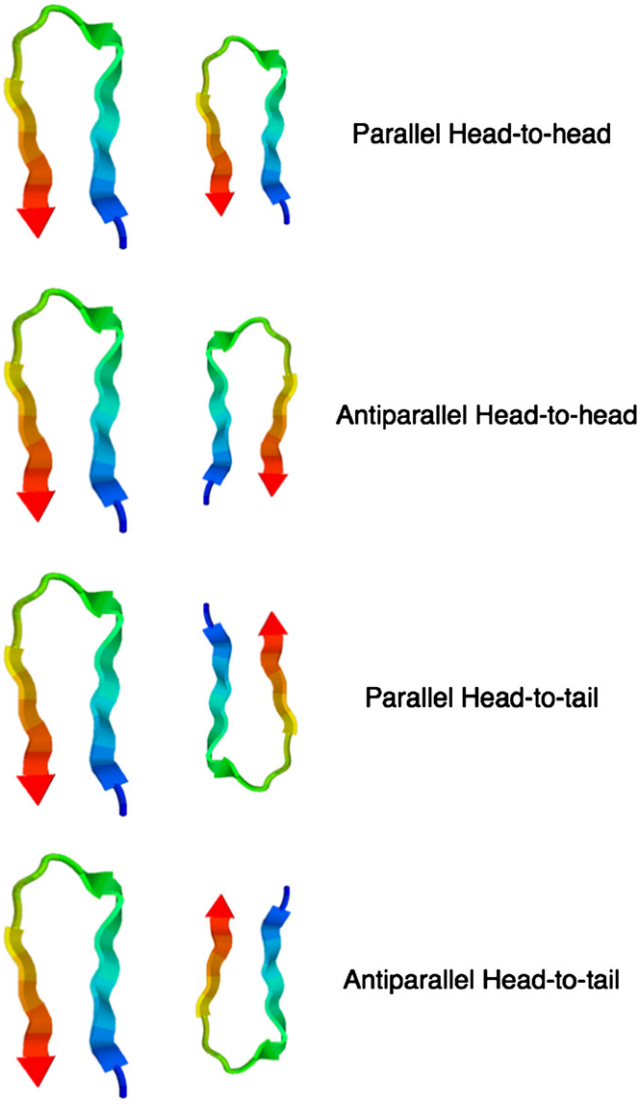

Fig. 1. The dimers were divided into four groups: parallel head-to-head, antiparallel head-to-head, parallel head-to-tail, and antiparallel head-to-tail. The four states were chosen with equal probability. Before each MD run, one of the monomers was translated by a random displacement vector (not shown).