DOI 10.5012/bkcs.2009.30.12.3006

Computational Study of Human Calcitonin (hCT) Oligomer

Youngshang Pak, Jungho Shin,[†,a] and Soonmin Jang[†,*]

Department of Chemistry, Pusan National University, Busan 609-735, Korea
[†]Department of Chemistry, Sejong University, Seoul 143-747, Korea. [*]E-mail: sjang@sejong.ac.kr
Received January 29, 2008, Accepted October 21, 2009

We have performed long time REMD simulation on 15-19 residues of human calcitonin hormone (DFNKF) which is known to form highly ordered amyloid fibril. The simulation started from randomly oriented multiple DFNKF strand. Using all-atom level simulations with the generalized Born solvation (GB) model (param99MOD3), we observed spontaneous formation of β-sheet for tetramer. Interestingly, the current simulation gives anti-parallel sheet as a major conformation, consistent with experiments. The major interaction stabilizing the anti-parallel sheet seems to be the inter-strand hydrogen bond.

Key Words: Molecular dynamics simulation, Computational chemistry, Human calcitonin (hCT), Amyloid

Introduction

Over the last couple of decades there has been more than enough evidences showing that the formation of highly ordered protein (or peptide) aggregates are directly related to many pathologically distinct diseases, called amyloid disease, such as Alzheimer's disease, CJD (Creutzfeldt-Jacob disease), Parkinson's disease, and Huntington's disease.[1-3] Several amyloidogenic proteins are isolated and characterized.[4] Usually, these proteins aggregate to form insoluble fibril, known as amyloid fibril, and they have a characteristic crossed beta sheet structure perpendicular to the fibril axis, suggesting there could be common mechanistic pathways toward the diseases. Therefore, a detailed understanding of insoluble protein aggregates is of fundamental importance for therapeutic purposes. So far, however, little is known of its detailed structure and formation mechanism. This is mainly due to experimental difficulties in conventional structural determination methods such as NMR and X-ray because the aggregate is not only insoluble but also it hardly forms single crystal. It is noteworthy that recent advances in various alternative experimental techniques, such as X-ray with micro single crystal, by which the GNNQQNY structure of yeast prion protein Sup35 has been determined,[5] and solid state NMR,[6] shed light on the structural information of amyloid fibrils. Thus, some of the common features of fibril have begun to emerge.[7] At present, however, numerous structural and dynamic characteristics including the detailed structure of general amyloid fibril, its formation mechanism, and the minimum size of critical nucleus seed that may acts as a template for fibril growth largely remain unknown.

Unlike experiments, carefully designed computational studies may give direct and detailed information on protein aggregates and the aggregation process, thereby complementing experimental studies. This has prompted many researchers in computational biochemistry over the past several years to perform numerous calculations on properties of amyloid fibrils, including structure characterization, stabilities, and its formation pathways.[8,9] One of the central areas of focus has been the detailed structure of each fibril and the main driving force toward the self assembly of amyloidogenic proteins into the final stable beta sheet connected structure. It seems the zipper is a common motif for sheet-sheet recognition.[7]

The human calcitonin (hCT) hormone is one of the amyloidogenic proteins with 32-amino acids and its amyloid formation is related to medullary carcinoma of the thyroid.[10] Experimentally, peptide with residues 15-19 (DFNKF) also forms stable amyloid fibril.[11,12] However, the detailed structure with atomic level is not known to date despite the simplicity of its constituting peptide. Recently, a series of molecular dynamics simulation in Nussinov group tried to address the stability and important interactions for this amyloid fibril.[13-16] Using molecular dynamics simulations at high temperature with various possible arrangements, they found that the stable structure of aggregates is a parallel β-sheet even though the experimental data implicates the existence of a large portion of an anti-parallel structure.[12] They noticed the electrostatic interaction between C terminus and Lys and between N terminus and Asp plays important role along with hydrogen bond network between Asn side chains in their parallel β-sheet arrangement. Also, they noticed the Asn-Asn interaction plays major role in DFNKF stability and formation

Over the past several years, the focus of amyloidosis study has shifted from fibrilar form to oligomers since oligomers are not only the intermediates for fibril formation but in most cases also have higher cell toxicity than the fibril itself.[19,20] In terms of the computational requirements, the oligomerization of peptide is very slow process even with today's high performance computer. Thus, direct simulation of oligomerization is rarely feasible unless some efficient simulation strategy is employed. In this paper, we performed computer simulation of the aggregation of hCT oligomers (tetramer) through efficient sampling scheme, the replica exchange molecular dynamics (REMD) simulation.[17,18] In REMD, molecular dynamics simulation is performed with several independent trajectories at different temperatures simultaneously. Then the trajectory exchange

[a]Current address, Korea Institute of Science and Technology, Seoul, 136-791, Korea

Table 1. The REMD temperature distribution and the corresponding replica exchange acceptance. Acceptance at 300 K indicates acceptance between 288 K and 300 K.

Temp. (K)	288	300	315	332	352	374	398	423	450	478	508	540
Acceptance (%)		39.9	33.7	28.8	21.1	21.6	24.5	28.3	30.0	31.4	31.5	34.7

between neighboring temperature are allowed at every given time interval and this can greatly enhance the overall sampling efficiency since the high temperature trajectories tend to explore configuration space better than the low temperature trajectories. Unlike the previous computational studies by Tsai *et al*.,[14] who performed a short REMD simulation of hCT trimer starting from putative stable parallel β-sheet conformation, we performed self-assembly of the hCT tetramer starting from randomly oriented monomers. Considering the time scale of most amyloid fibril formation, computational simulation of formation of the ordered hCT aggregates is a daunting task but this approach can give indirect information on the formation mechanism and stability of hCT oligomers as well as possible aggregated oligomer structures.

Even though the stability of oligomeric hCT conformers are addressed by extensive stability test that start from various possible arrangements,[13,15] there could be other possible arrangements especially as the number of hCT strands increases. The reason we performed REMD simulation starting from constituting monomers is two fold. First, what are the possible stable conformers when hCT aggregates into oligomer and what is their relative stability? Second, are the resulting conformers consistent with previous simulations and experiments? Tetramer may be good candidate for this purpose since oligomers of this size can exhibit more variations in monomer arrangement to form β-sheet connected oligomer than simple dimer or trimer.

Methods

The initial conformation of each monomer is extended and the N terminus and C terminus of each monomer is patched with acetyl group (Ace) and N-methyl group (Nme) respectively. Multiple strands of extended DFNKF monomers are placed inside the simulation box with a distance of about 20 Å between each strand. Then, the system is heated to 600 K for 500 ps followed by local energy minimization. Please note that this will generate completely random initial configuration. The simulation box consists of an imaginary sphere with a radius of 25 Å from the center of mass. If the atoms fall out of this imaginary boundary, the atoms will experience an attractive harmonic force toward the center of mass of the system with a force constant of 1.0 kcal·mol^{-1}·Å^{-2}. This will effectively confine the whole system and prevent from the molecules flying apart from each other, thus mimicking a cellular environment.[21,22] Even with REMD, self assembly of small peptides with all atom level is computationally challenging, especially under the presence of explicit solvation. In an attempt to reduce the overall computational requirement while considering the reasonable solvation effect, we used the generalized Born (GB) solvation model with a modified version of all atom amber99 force field (param99MOD3) that we have developed previously.[23] Most all atom force fields tend to exhibit somewhat unbalanced α/β propensity with over-stabilized salt bridge effects within the GB solvation framework. The param99MOD3 modified force field is a result of extensive empirical tuning of the param99-MOD2 force field by Onufriev *et al*.[24] such that the global minimum of free energy landscape obtained from REMD simulation corresponds to the native state for several training sets, i.e. αββ mini-proteins (1FSD, 1PSV, and BBA5) which are structurally subtle and computationally challenging. It has been shown that the resulting force field can describe free energy profile of several structurally distinct peptides reasonably.[25]

The temperature range for the replica exchange was set to 288 K ~ 540 K with 12 replicas as listed in Table 1. The replica exchange interval was 0.16 ps. This provides an average replica exchange acceptance of about 29%. We implemented the REMD scheme into the *Tinker*[26] mol ecular dynamics simulation package using version 4.2 and used stochastic dynamics with friction coefficient 1.0 ps throughout this study. Note that the friction constant we used is smaller than the actual water friction constant (79 ps at 300 K) to enhance the sampling speed and this will not change the overall simulation results.[27,28] The simulation time step was 2.0 fs with fixed bond distances between hydrogen atoms and heavy atoms using RATTLE.[29] Considering the time scale of peptide aggregation, the self assembly simulation requires extensive computing even with very efficient sampling methods such as the REMD we used here. The total REMD simulation time for each replica was 136 ns, resulting in total 136 × 12 = 1632 ns. The cut distance of both the non-bonded interaction and the GB solvation was set to 24 Å. The trajectory was saved at every 400 ps and the last 32 ns of REMD simulation results were used for further analysis and production purposes.

Results and Discussion

First, we checked the overall performance of the REMD simulation. Frequently, the change of the replica index as a function of simulation time is used as a criteria for this purpose.[30,31] Following the history of all trajectories, the replica exchange index of one of the selected trajectories for tetramer simulation was obtained as a function of the REMD simulation time and it is shown in Figure 1. The trajectory visits all replicas from the lowest temperature (index 1) to the highest temperature (index 12), indicating the trajectory is stochastic enough with reasonable trajectory mixing within the overall REMD simulation time. In addition to the good trajectory mixing, reasonable convergence is required to observe the overall ensemble of trajectories at given temperature. We have performed backbone RMSD (root mean square deviation) based hierarchical cluster analysis[32] at every 8 ns of REMD simulation time (20000 conformers) during the whole simulation time (136 ns) and calcu-

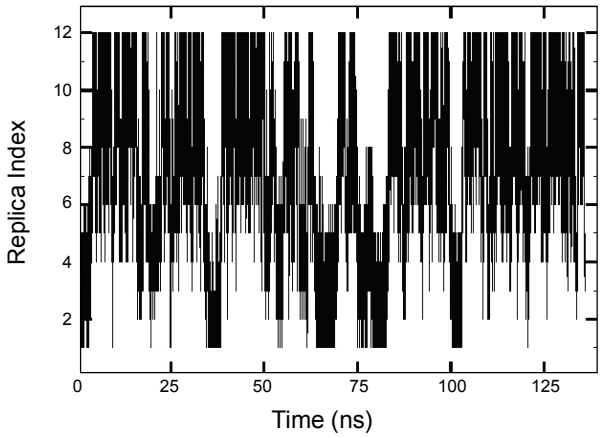

Figure 1. The replica index of one of the trajectories for the tetramer simulation as a function of REMD time. Other trajectories show similar behavior. The trajectory visits all replicas many times during the simulation, indicating the system is stochastic and reasonable trajectory mixing has been achieved.

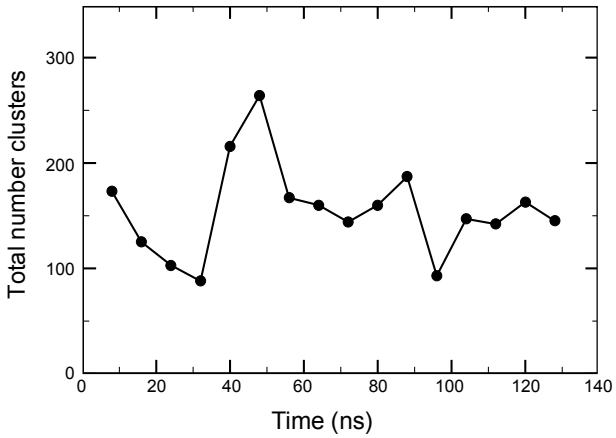

Figure 2. The total number of cluster at 300 K. The number of cluster is calculated at every 8 ns (20000 conformers) from last 32 ns of simulation time. As the RMSD criteria increases, the number of cluster decreases in general. After slow increase of total cluster number, it remains steady from around 104 ns even though there are some fluctuations. The figure indicates the REMD simulation has reached close to the equilibrium.

lated the total number of cluster to monitor the REMD convergence[33] at 300 K. In hierarchical clustering analysis, for each conformer, the RMSD with the rest of conformer is calculated and the number of conformers having RMSD value less than the given criteria (close neighbor) is counted. After that, the conformer with the highest number of this close neighbor (central conformer) is removed from the initial collection of structure, along with those close neighbors, and classified as single cluster. The same procedure is repeated with remaining conformers until no conformer is left. The number of total cluster at every 8 ns with RMSD criteria of 5.0 Å is plotted for last 32 ns of REMD simulation in Figure 2. The number of total cluster is fluctuating as simulation proceeds and it seems it remains steady after 104 ns, suggesting the REMD has reached to quasi-equilibrium. For the final structural analysis, we discarded initial 104 ns trajectory and performed hierarchical clustering analysis with trajectories from last 32 ns. One of the main results of present study comes from this clustering analysis at 300 K. When we used RMSD of 5.0 Å as clustering criteria, the relative populations of first seven clusters are 40.4%, 17.4%, 11.6%, and 3.7%. The top four highest populated clusters took 66.1% of total frames analyzed. The most highly populated cluster (40.4% of all total conformers) has a proto-fibril like single layer β-sheet with anti-parallel arrangement. This finding is different from previous reports on stability simulations, in which parallel β-sheet is the main stable structure.[15] The representative structure of most populated cluster for tetramer simulation is shown in Figure 3. The side chains are also shown as stick form. The representative structure is taken as the central conformer of the cluster during clustering analysis. The three strands forms stable β-sheet while the forth strand's sheet formation is rather loose. The structure shows a zigzag pattern by one residue in a sense that the end of each strand is not

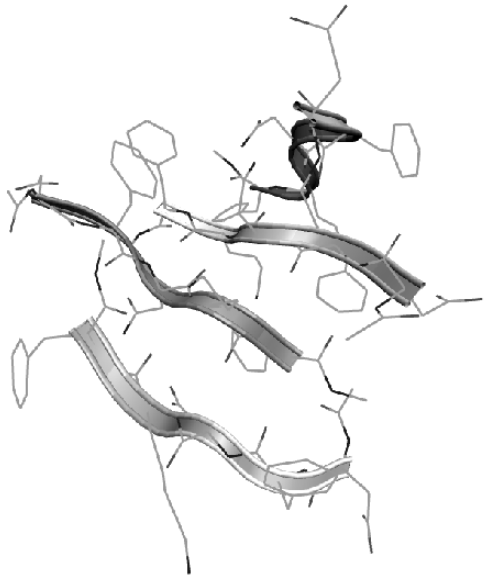

Figure 3. The representative structure of the most populated cluster. The hydrogen bonds are represented by dotted lines. The side chains are also shown in stick form for clarity.

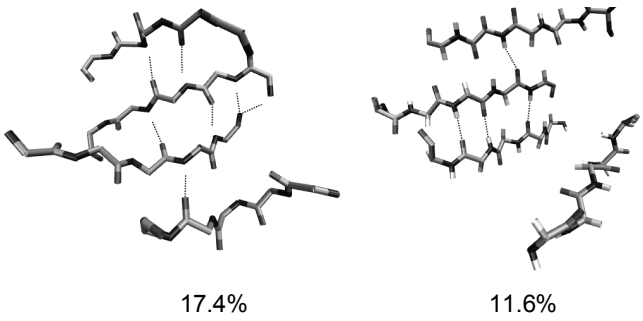

Figure 4. Representative structures of two clusters (cluster 2 ~ cluster 3) next to the highest populated cluster (cluster 1). The values indicate the population of each cluster. Note the highest populated cluster (cluster 1) represents 40.4% of the total population.

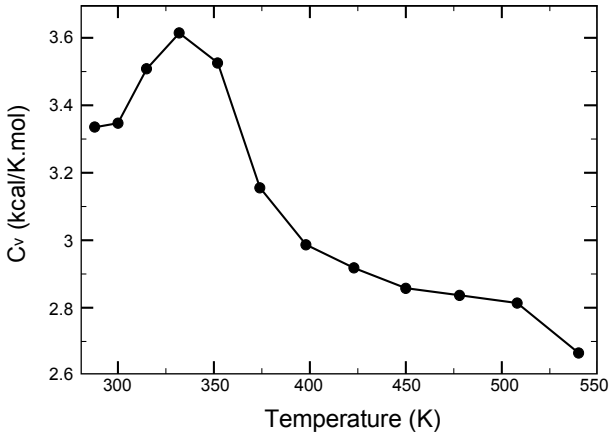

Figure 5. The constant volume heat capacities at all REMD replica temperature. The phase transition temperature is located at the heat capacity peak at around 330 K.

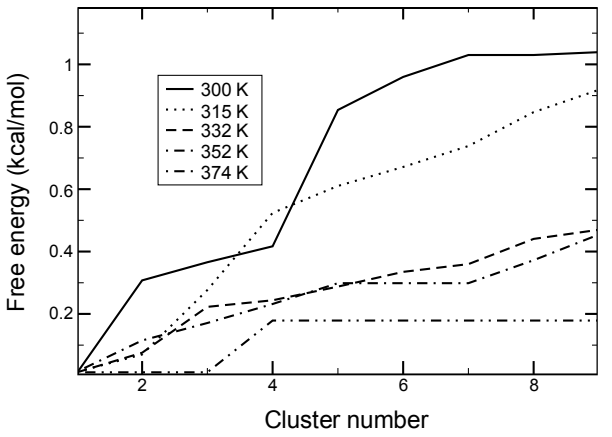

Figure 6. The free energy landscape as a function of discrete coordinate of cluster number at several temperatures. The free energy is relative energy.

linearly aligned. This is possibly due to inter-strand Phe-Phe π-π interaction with anti-parallel direction since the Phe residues are at second place from N-terminal and first place from C-terminal. As for this inter-strand π-π stacking the geometry shows it is not optimally placed, resulting in a weak π-π interaction. It seems the interstrand hydrogen bonds are a major factor in structural stability. Note that in the parallel β-sheet conformation of Tsai *et al*.,[15] π-π stacking also appears and its π-π stacking tendency is somewhat higher than current conformation. The interstrand backbone hydrogen bonds are between Lys and Lys, Phe-Phe, and Asn-Phe. In addition to the backbone hydrogen bonds, fluctuating hydrogen bonds between Asn side chains also appeared. The structure of the next two clusters, each having population of 17.4%, and 11.6% respectively, are basically composed of three strand β-sheet like + monomer, possibly intermediates for the formation of well aligned anti-parallel tetramer strand (Figure 4). The inter-stand hydrogen bonds are indicated with dotted lines. Here, we considered the hydrogen bond is "formed" if the bond distance is within 3.5 Å and the bond angle is larger than 150°. We did not observe any parallel strands in this tetramer simulation. The current simulation clearly shows anti-parallel β-sheet is the stable and dominant conformer for hCT oligomers, consistent with the experiment.[12]

We calculated the energy fluctuation, and thereby the constant volume heat capacities at all twelve REMD simulation temperatures. The heat capacity as a function of temperature in Figure 5 indicates that it exhibit phase transition at around 330 K, which corresponds to collapse of anti-parallel sheet. Thus, it is expected that this anti-parallel beta strand can survive up to around 330 K. The same tendency can be observed from the free energy landscape as a function of the discrete coordinate of cluster number in Figure 6. In this coordinate, cluster number 1 has the highest number of population. At 300 K, the free energy difference between the highest populated (anti-parallel β-sheet like structure) and the random structure (cluster number 9) is more than 1.0 kcal/mol. But from around 332 K, the free energy difference is less than 0.5 kcal/mol, suggesting that the anti-parallel β-sheet can disintegrates due to thermal fluctuation.

The trajectories obtained from REMD simulation represent ensemble of conformation at given temperature, lacking any direct dynamic information. Therefore, it is not the purpose of present work to probe the detailed assembly pathway of amyloid like hCT(15-19). Instead, the main focus of current study is finding the stable oligomer structure(s) connected with β-sheet that might be related to further aggregation into proto-fibril with no, or minimal simulation bias/constraint imposed on the system. Possibly, the largest Due to this limitation, the current simulation might not capture the water hydrogen bond network in the hCT fibril, which might play an important role in fibril stabilization,[7] since the simulation was performed under the GB implicit solvation environment. Therefore, in the future it needs to be studied further the stability and detailed water hydrogen bond network of the structure obtained from current simulation in the presence of explicit water using extensive simulations.

In terms of free energy, the aggregation of the hCT monomer into oligomer will increase free energy initially since the entropy decrease will dominate the enthalpy increase. But as the size of the oligomer increases, at some point, the enthalpy will dominate entropy and free energy will decrease as the oligomer size increases. In general, an oligomer of this size will serve as a nucleus seed for further growth into proto-fibril. Neither the size of this oligomer nor the structure is known at this time. As for the hCT, the size of this critical nucleus seed could be small.[15] It would be interesting to calculate and compare the free energy of oligomerization for hCT by changing the oligomer size systematically including the two oligomers we presented here even though each calculation requires considerable computational resources.

Conclusions

The detailed structure of the 15-19 segment of the hCT hormone protein, which forms amyloid fibril, is unknown with atomistic details. Recent computational studies show that the parallel beta-sheet strand might be the stable form. This con-

tracts to the experimental data which indicate that the anti-parallel beta-sheet is the major stable structure depending on the solution pH.[12] In an effort to understand this structural difference we performed the REMD simulation of acetyl and methylamine patched 15-19 segment of hCT hormone starting from fully extended structure with random orientation with all atom level using the GB implicit solvation model (param99-MOD3).

The simulation results indicate that the anti-parallel amyloid proto-fibril like β-sheet is the major conformer, stabilized mainly by interstrand hydrogen bonds. In general, the formation of amyloid fibril is very sensitive to the environments such as concentration, temperature, and pH. The solvation model we adapted is implicit solvent rather than explicit water molecules, which require a great amount of computational resources. This might be the main limitation for the simulation presented in this study. Clearly, the anti-parallel β-sheet structure from present simulation needs further study with more elaborated model including explicit solvent simulations. Furthermore, the mechanistic details such as the main driving force and pathway of this oligomer formation, possibly using hydrogen bond network analysis approach such as study by Tsai *et al*.,[14] may contribute to understanding the somewhat subtle structural and dynamic features of hCT (15-19).

Acknowledgments. SJ appreciates financial support from Sejong University.

References

1. Kelly, J. W. *Current Opinion in Structural Biology* **1998**, *8*, 101.
2. Hardy, J.; Selkoe, D. J. *Science* **2002**, *297*, 353.
3. Chiti, F.; Dobson, C. M. *Annual Review of Biochemistry* **2006**, *75*, 333.
4. Westermark, P. *Amyloidosis and Amyloid Proteins: Brief History and Definaitions*; WILEY-VHC: 2005; Vol. 1.
5. Nelson, R.; Sawaya, M. R.; Balbirnie, M.; Madsen, A. O.; Riekel, C.; Grothe, R.; Eisenberg, D. *Nature* **2005**, *435*, 773.
6. Tycko, R. *Biochemistry* **2003**, *42*, 3151.
7. Zheng, J.; Ma, B. Y.; Nussinov, R. *Physical Biology* **2006**, *3*, P1.
8. Caflisch, A. *Current Opinion in Chemical Biology* **2006**, *10*, 437.
9. Lee, S.; Kim, Y. *Bulletin of the Korean Chemical Society* **2004**, *25*, 838.
10. Zaidi, M.; Inzerillo, A. M.; Moonga, B. S.; Bevis, P. J. R.; Huang, C. L. H. *Bone* **2002**, *30*, 655.
11. Reches, M.; Porat, Y.; Gazit, E. *Journal of Biological Chemistry* **2002**, *277*, 35475.
12. Naito, A.; Kamihira, M.; Inoue, R.; Saito, H. *Magnetic Resonance in Chemistry* **2004**, *42*, 247.
13. Tsai, H. H. G.; Tsai, C. J.; Ma, B.; Gunasekaran, K.; Zanuy, D.; Nussinov, R. *Biophysical Journal* **2004**, *86*, 412A.
14. Tsai, H. H.; Reches, M.; Tsai, C. J.; Gunasekaran, K.; Gazit, E.; Nussinov, R. *Proceedings of the National Academy of Sciences of the United States of America* **2005**, *102*, 8174.
15. Tsai, H. H.; Zanuy, D.; Haspel, N.; Gunasekaran, K.; Ma, B. Y.; Tsai, C. J.; Nussinov, R. *Biophysical Journal* **2004**, *87*, 146.
16. Haspel, N.; Zanuy, D.; Ma, B. Y.; Wolfson, H.; Nussinov, R. *Journal of Molecular Biology* **2005**, *345*, 1213.
17. Sugita, Y.; Okamoto, Y. *Chem. Phys. Lett.* **1999**, *314*, 141.
18. Swendsen, R. H.; Wang, J. S. *Phys. Rev. Lett.* **1986**, *57*, 2607.
19. Walsh, D. M.; Klyubin, I.; Fadeeva, J. V.; Cullen, W. K.; Anwyl, R.; Wolfe, M. S.; Rowan, M. J.; Selkoe, D. J. *Nature* **2002**, *416*, 535.
20. Klein, W. L.; Krafft, G. A.; Finch, C. E. *Trends in Neurosciences* **2001**, *24*, 219.
21. Gsponer, J.; Haberthur, U.; Caflisch, A. *Proceedings of the National Academy of Sciences of the United States of America* **2003**, *100*, 5154.
22. Beglov, D.; Roux, B. *Journal of Chemical Physics* **1994**, *100*, 9050.
23. Jang, S.; Kim, E.; Pak, Y. *Proteins-Structure Function and Bioinformatics* **2006**, *62*, 663.
24. Onufriev, A.; Bashford, D.; Case, D. A. *Proteins-Structure Function and Bioinformatics* **2004**, *55*, 383.
25. Jang, S.; Kim, E.; Pak, Y. *Proteins-Structure Function and Bioinformatics* **2007**, *66*, 53.
26. Ponder, J. W. TINKER 4.2 : software tools for molecular design; Washington University, 2004.
27. Zagrovic, B.; Sorin, E. J.; Pande, V. *Journal of Molecular Biology* **2001**, *313*, 151.
28. Hornak, V.; Okur, A.; Rizzo, R. C.; Simmerling, C. *Proceedings of the National Academy of Sciences of the United States of America* **2006**, *103*, 915.
29. Palmer, B. J. *Journal of Computational Physics* **1993**, *104*, 470.
30. Rao, F.; Caflisch, A. *Journal of Chemical Physics* **2003**, *119*, 4035.
31. Periole, X.; Mark, A. E. *Journal of Chemical Physics* **2007**, *126*.
32. Daura, X.; Gademann, K.; Jaun, B.; Seebach, D.; van Gunsteren, W. F.; Mark, A. E. *Angewandte Chemie-International Edition* **1999**, *38*, 236.
33. Affentranger, R.; Tavernelli, I.; Di Iorio, E. E. *Journal of Chemical Theory and Computation* **2006**, *2*, 217.