Relazione scientifica sui risultati dell'attività di ricerca

Programma Short Term Mobility – STM 2013

Computational biophysics of amyloid growth pathways

Proponente / fruitore:

Toni Giorgino

Istituto di Ingegneria Biomedica (ISIB-CNR)
Dipartimento di scienze biomediche
Corso Stati Uniti 4, 35127 Padova, Italy

Periodo STM:

16 settembre – 7 ottobre 2013

Istituzione ospitante:

Computational Biophysics Group – Universitat Pompeu Fabra
Research Programme on Biomedical Infomatics (GRIB-IMIM-UPF),
Barcelona Biomedical Research Parc (PRBB)
C/ Aiguader 88, E-08003 Barcelona, Spain

Abstract

Amyloid fibrillation is a ubiquitous and still poorly understood phenomenon in which proteins aggregate into ordered and often pathological structures. Molecular dynamics (MD) simulations, a very compute-intensive technique, provide an excellent opportunity to observe aggregation pathways and polymorphism. The short-term mobility (STM) funding was used to advance a joint research programme between ISIB-CNR and the Computational Biophysics Group of the Universitat Pompeu Fabra involving the analysis of a data set consisting in over half millisecond of all-atom MD trajectories of Sup35 aggregation-prone fragments, an amount of sampling far beyond the state of the art for the field. This report summarizes the most important results of the STM stay, which have been (a) the design of a combination of reaction coordinates that made the collected data amenable of analysis via a Markov-state model formalism. Additional results have been (b) extensive discussions held in order to frame the experimental protocol and pro and cons of alternative approaches; (c) testing of the metric with the high-throughput MD analysis software being developed at the host institution; (d) finalization of the MEMBPLUGIN software and the corresponding manuscript; and (e) a preliminary version of a paper to be submitted.

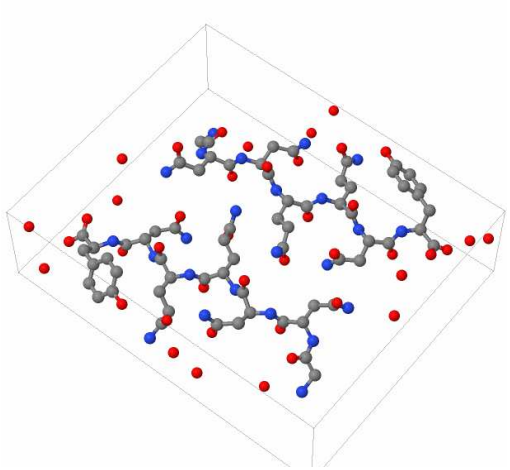

Figure 1 – Transversal cross-section of one strand of the Sup35 proto-fibril formed by the GNNQQNY residues (PDB:1YJP).

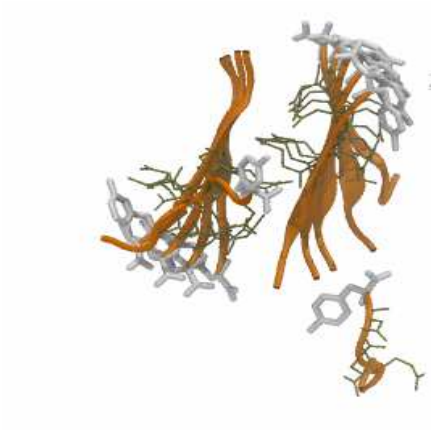

Figure 2 – A simulated configuration of the system containing the simulated Sup35 aggregation-prone peptides. Hydration is not shown for clarity.

Introduction

Amyloid fibrillation is a process related to "transmissible" protein misfolding occurring at the molecular scale; pathologic amyloid aggregates are found in numerous diseases, such as transmissible spongiform and related encephalopathies, and there are hints for their involvement in neurodegenerative diseases as dramatic as Alzheimer's and Parkinson's (1). The objective of the short-term mobility stay was to interpret the results obtained from over 0.5 millisecond of all-atom molecular dynamics (MD) trajectories of Sup35 fragments in order to provide an atomistic view of the amyloid growth polymorphism.

The work was performed in collaboration with a leading expert in the field of high-throughput MD, Prof. G. De Fabritiis, head of the hosting laboratory, the Computational Biophysics Group (GRIB-IMIM-UPF) located at the Barcelona Biomedical Research Parc (PRBB). Discussions also involved Prof. G. Tartaglia, group leader of the Gene Function and Evolution Laboratory at the Centre for Genomic Regulation (CRG) and expert in protein aggregation.

The project strengthened the collaboration between ISIB-CNR and the Computational Biophysics Group. The collaboration has a track record of experiments and co-authored papers based on the use of the host's resources, most notably the GPUGRID.net volunteer computing network (2). Previous joint research include one publication in the Proceedings of the National Academy of Sciences (3,4), which set forth the methodology – Markov-state based analysis – that has been applied in this work, consisting in the analysis of the new simulation data on a much more complex system than those tackled in the methodological papers.

Objectives

The stay realized the objectives set forth, namely:

- Analyze the results of a high-throughput computational experiment which uncovered events in the proto-fibrillation pathways of a model prion system.
- Rationalize the observed events in a narrative that highlights the pathway polymorphism in the context of current amyloidal growth theories.

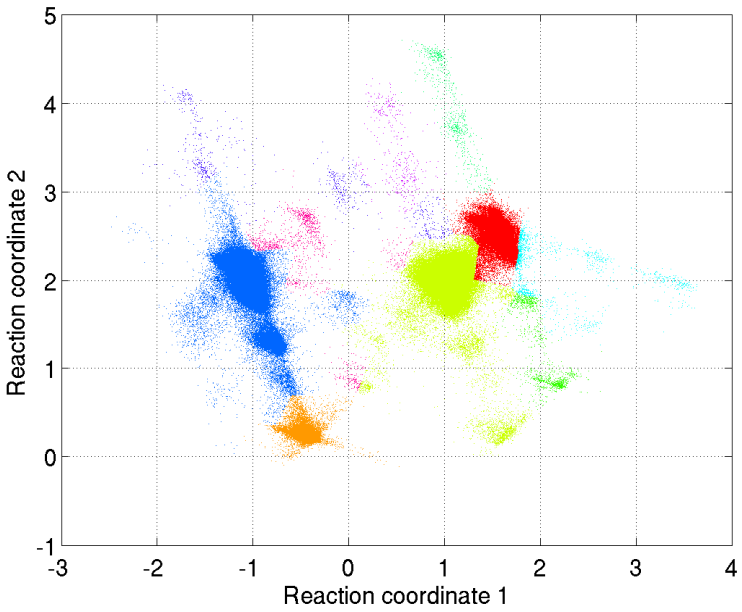

Figure 2 – Clustering of configurations in the space spanned by the reaction coordinate projection defined in this work.

- Outline a dissemination plan with co-authors at PRBB concerning the experimental results.
- Strengthen the collaboration between ISIB-CNR and start further collaborations with PRBB-hosted research groups.

The following sections will outline the scientific results in detail.

Results

To elucidate the pathways of binding of amyloid-like monomers, we setup large-scale simulations to sample different systems (Figure 1). Macromolecular dynamics, especially those involving large structural fluctuations such as disordered peptides, engage so many degrees of freedom that "brute force" simulation is insufficient to fully characterize the equilibrium properties (e.g. binding affinities) and kinetics of biomolecular systems. We tackled the problem using the vast amounts of simulation time provided by the GPUGRID.net computing network, supported by users distributed worldwide that volunteer GPUs for biomedical research (2). Volunteer computing is a computing model which involves users from all over the world to participate in scientific efforts and to "donate" the processing time of their computers and GPUs (5). This scheme allows scientists to control only the server, while all the hardware and running costs of the hardware are contributed to science by the volunteers. The infrastructure provided more than 600 µs of aggregated simulated time for this project.

The objective was to elucidate the spontaneous binding process (including structural and kinetics data) of monomers to a variety of geometries (Figure 2). We chose to analyze the GNNQQNNY fragment of the Sup35 protein (residues 7-13 of ERF3_YEAST, UniprotKB P05453), a well-known model system for fibrillar growth (6,7). The unbiased simulations provided events that have been categorized in order to provide a meaningful, unifying view of the process and pathways involved. We reviewed the data and discussed a forthcoming manuscript with local experts in MD and fibrillation. Technically, the analysis required the development of meaningful metrics (from the structural biology point of view) on the basis of the available trajectory data. The sheer amount of data (approximately 10 TB) required the analysis and visualization to be performed at the computational resources available at the facility where the data is hosted.

To this aim, during the stay we investigated the way to apply a Markov-based analysis methodology (3,8) to the modeled fibrils. We discussed and formalized a set of reaction coordinates that allow the distinction of the structural intermediates of interest. The definition of the collective coordinate system is an important step in the rationalization of the growth pathways because it represents a simplified (bidimensional) projection of an otherwise complex process.

The projection chosen has to preserve distinctness of structural configurations; this property was successfully verified, by the successful clustering obtained via a preliminary version of the High Throughput MD software package under development at the host institution (Figure 3). Another important test of the goodness of the sampling and the validity of the markovianity assumption is the convergence of implied timescales (ITS) (8,9). The ITS plot confirmed that the state decomposition and amount of sampling yielded a well converged model; the model is therefore amenable to analysis of the systems' dynamics and energetics (10). Such analysis is the objective of the next steps of the collaboration, and is currently ongoing.

Further collaborations, software and papers

STM 2013 funding also enabled the finalization of version 1.0 of a scientific code, *MEMBPLUGIN*, for the analysis of biological membranes, in collaboration with the group of Prof. Jana Selent (Computer-Assisted Drug Design Laboratory, GRIB-IMIM-UPF). MEMBPLUGIN is a collection of visual and command-line tools that can be run within the Visual Molecular Dynamics (VMD) environment to analyze biomolecular simulations of lipid bilayers, which is now released[1] with an open-source license. Furthermore, a manuscript describing the software (in shared senior authorship) has been finalized and submitted to a major journal during the STM period.

Conclusions

The STM-supported stay has been indispensible in the ongoing collaborative investigation of the structural determinants of amyloid growth set forth in this project. In-person discussions were essential to develop analysis scripts based on the high-throughput infrastructure of the host institution, which provided the computing power for the experiments, not readily available elsewhere. A scientific plan for the analysis of pathways and kinetics emerged from the discussions, together with corresponding drafts. Work is ongoing towards a paper suitable for a high-impact journal, highlighting the molecular and drug-design implications of the study (11).

Acknowledgments

The research stay was funded by the STM 2013 programme of the National Research Council of Italy, which is gratefully acknowledged. I am grateful to the hosting institution, the Universitat Pompeu Fabra, and to the people at the Research Programme on Biomedical Informatics (GRIB). Special acknowledgments go to Prof. G. De Fabritiis (Computational Biophysics Group, GRIB-IMIM-UPF) and members of his group; to Dr. M. J. Harvey (Accelera ltd.); to Prof. J. Selent (Computer-Assisted Drug Design Laboratory, GRIB-IMIM-UPF) and members of her group; and to Prof. G. Tartaglia (Gene Function and Evolution, Centre for Genomic Regulation).

[1] "MEMBPLUGIN: studying membrane complexity in VMD", at http://membplugin.sourceforge.net

Description of the hosting institution:

Computational Biophysics Group – Universitat Pompeu Fabra
Research Programme on Biomedical Infomatics (GRIB-IMIM-UPF)
Barcelona Biomedical Research Parc (PRBB)
C/ Aiguader 88, E-08003 Barcelona, Spain

Computational Biophysics Group – The host scientist, Dr. G. De Fabritiis, head of the Computational Biophysics Group, pioneered the use of accelerated processors for large-scale molecular dynamics (MD) simulations. The hosting group has innovated the field of MD developing (a) accelerated MD engines, using recent high-performance hardware such as graphical processing units (GPU); and (b) a distributed computing infrastructure, GPUGRID, which leverages the contribution of thousands of volunteers which donate GPU power for use by the group's scientists, thus providing a sampling capacity for all-atom MD simulations *on par* with the world's largest supercomputers (2).

Universitat Pompeu Fabra (UPF) – The UPF is one of the best known universities in Europe, with a prominent role in research and teaching and a noted position of excellence in biotechnology and economics. The UPF serves nowadays over 11,500 students, around 25% being international.

Barcelona Biomedical Research Park (PRBB) – PRBB is a new research infrastructure hosting more than 1,000 researchers with a unique mixture of experimental and computational laboratories. PRBB is part of a cluster of research centres based in Barcelona that includes the Barcelona Supercomputing Centre, hosting MareNostrum, one of the largest supercomputers in Europe.

References

1. Aguzzi A. Cell biology: Beyond the prion principle. Nature. 2009 Jun 18;459(7249):924–5.
2. Buch I, Harvey MJ, Giorgino T, Anderson DP, De Fabritiis G. High-Throughput All-Atom Molecular Dynamics Simulations Using Distributed Computing. J Chem Inf Model. 2010 Mar 22;50(3):397–403.
3. Buch I, Giorgino T, De Fabritiis G. Complete reconstruction of an enzyme-inhibitor binding process by molecular dynamics simulations. Proc Natl Acad Sci. 2011 Jun 6;108(25):10184–9.
4. Giorgino T, Buch I, De Fabritiis G. Visualizing the Induced Binding of SH2-Phosphopeptide. J Chem Theory Comput. 2012;8(4):1171–5.
5. Sansom C. The power of many. Nat Biotechnol. 2011;29(3):201–3.
6. Nelson R, Sawaya MR, Balbirnie M, Madsen AO, Riekel C, Grothe R, et al. Structure of the cross-β spine of amyloid-like fibrils. Nature. 2005 Giugno;435(7043):773–8.
7. Esposito L, Pedone C, Vitagliano L. Molecular dynamics analyses of cross-β-spine steric zipper models: β-Sheet twisting and aggregation. Proc Natl Acad Sci. 2006 Aug 1;103(31):11533–8.
8. Bowman GR, Beauchamp KA, Boxer G, Pande VS. Progress and challenges in the automated construction of Markov state models for full protein systems. J Chem Phys. 2009 Sep 28;131(12):124101.
9. Pérez-Hernández G, Paul F, Giorgino T, De Fabritiis G, Noé F. Identification of slow molecular order parameters for Markov model construction. J Chem Phys. 2013 Jul 3;139(1):015102–015102–13.

10. Noé F, Schütte C, Vanden-Eijnden E, Reich L, Weikl TR. Constructing the equilibrium ensemble of folding pathways from short off-equilibrium simulations. Proc Natl Acad Sci. 2009 Nov 10;106(45):19011–6.
11. Borhani DW, Shaw DE. The future of molecular dynamics simulations in drug discovery. J Comput Aided Mol Des. 2011 Dec 20;26(1):15–26.

Padova, 20 novembre 2013.