Data in Brief 49 (2023) 109359

Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Data Article

# Early stages in A$\beta$1-42 spontaneous aggregation: An unbiased dataset from coarse-grained molecular dynamics simulations

Exequiel E. Barrera [a,b,*], Sergio Pantano [b]

[a] Instituto de Histología y Embriología (IHEM) - Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), CC56, Universidad Nacional de Cuyo (UNCuyo), Mendoza, Argentina
[b] Biomolecular Simulations Group, Institut Pasteur de Montevideo, Mataojo 2020, CP 11400 Montevideo, Uruguay

ARTICLE INFO

*Article history:*
Received 18 January 2023
Revised 7 June 2023
Accepted 26 June 2023
Available online 2 July 2023

Dataset link: Early stages in Ab1-42 spontaneous aggregation: an unbiased dataset from coarse-grained molecular dynamics simulations. (Original data)

*Keywords:*
SIRAH
Coarse-grained simulation
Soluble oligomer
Alzheimer
Aggregation
Oligomerization
Molecular dynamics

ABSTRACT

The small soluble aggregates of A$\beta_{1\text{-}42}$ are broadly documented as potential targets for the development of new compounds with the capacity to inhibit the early stages of Alzheimerś disease. Nevertheless, A$\beta_{1\text{-}42}$ peptides show an intrinsically disordered character with a high propensity for aggregation, which complicates the identification of conserved structural patterns. Because of this, experimental techniques find substantial difficulties in the characterization of such soluble oligomers. Theoretical techniques, such as molecular dynamics (MD) simulations, provide a possible workaround for this problem. However, the computational cost associated with comprehensively sampling the vast conformational space accessible to these peptides might become prohibitive. In this sense, coarse-grained (CG) simulations can effectively overcome that hurdle at a fraction of the computational cost.

In this dataset, we furnish an extensive collection of A$\beta_{1\text{-}42}$ peptides in dimeric conformation generated with the SIRAH force field for CG MD simulation. It comprises 25 independent trajectories in .xtc (gromacs) format of A$\beta_{1\text{-}42}$ couples of peptides that evolve towards dimeric states along eleven µs-

DOI of original article: 10.1016/j.csbj.2021.02.014
* Corresponding author.
*E-mail address:* ebarrera@mendoza-conicet.gob.ar (E.E. Barrera).

https://doi.org/10.1016/j.dib.2023.109359

long unbiased simulations. Thanks to the backmapping capabilities of our force field, pseudo atomistic coordinates can be straightforwardly recovered from MD trajectories reported here and analyzed with popular molecular editing programs. This set of simulations performed at room conditions and physiological salt concentrations may furnish a complete collection of inter-peptide interfaces that can be used in high-throughput docking or as new starting states for peptide oligomerization seeding of A$\beta_{1\text{-}42}$ dimerization.



## Specifications Table

| | |
|---|---|
| Subject | Biological Sciences. |
| Specific subject area | Protein Biophysics. Molecular dynamics simulations of A$\beta_{1\text{-}42}$ peptides. |
| Type of data | Filtered Data from molecular dynamics simulations. Coarse-grained molecular dynamics trajectories of A$\beta_{1\text{-}42}$ dimerization process. |
| How data were acquired | Hardware: CPU (Intel Xeon Gold 6138, 2.00 GHz) accelerated with an Nvidia Tesla P100 GPU. Software: Gromacs 2018.4 using the SIRAH 2.0 force-field for performing MD simulations and SIRAH Tools, along with AmberTools 2018 and Amber14SB force-field implemented in VMD 1.9.3 for backmapping. |
| Data format | Filtered Gromacs .xtc trajectories. |
| Description of data collection | Coarse-grained molecular dynamics trajectories of dimerization events of A$\beta_{1\text{-}42}$ peptides. Simulation frames of 25 independent replicas are reported every 0.1 ns of simulation. |
| Data source location | Primary Data was collected at the Uruguayan Centre for Supercomputing (ClusterUY)<br>Montevideo<br>Uruguay |
| Data accessibility | Repository name: Mendeley Data<br>Direct URL to data:<br>https://data.mendeley.com/datasets/h8y867fkry/2<br>Instructions for accessing these data: Data is freely accessible from the web address above. |
| Related research article | The primary data were obtained following the same protocol ported by us in "Dissecting the role of glutamine in seeding peptide aggregation" by E. E. Barrera, F. Zonta, and S. Pantano, Computational and Structural Biotechnology Journal, 2021, DOI:10.1016/j.csbj.2021.02.014 |

## Value of the Data

- Datasets of spontaneous and unbiased aggregation are scarce. The availability of ready-to-use data in standard format obviates the computational hurdle of producing the dataset.
- Data of interest to aggregation studies and screening of compounds/peptides with the potential to inhibit initial stages of A$\beta_{1\text{-}42}$ aggregation.
- Pre-calculated molecular dynamics trajectories are ready to be analyzed.

## 1. Objective

The rationales behind data generation were two-fold. On one side, data were generated to characterize the initial steps in A$\beta_{1\text{-}42}$ aggregation; on the other, we aimed to provide a corpus of conformers of A$\beta_{1\text{-}42}$ dimers that can be used as starting points for studying interactions with small compounds able to inhibit further aggregation processes.

## 2. Data Description

Soluble oligomers of A$\beta_{1\text{-}42}$ are widely recognized as crucial targets for the design of inhibitors for potential therapeutic intervention against Alzheimerś disease. However, the intrinsically disordered character of this polypeptide poses severe difficulties for the experimental determination of conserved structural motifs. Indeed, initial aggregation steps are extremely challenging for state-of-the-art experimental techniques. Although molecular dynamics simulations harbor the potential to capture such initial association events, unbiased exploration of the conformational landscape available to unstructured dimers implies a significant computational cost. Here, we provide a dataset of configurations of A$\beta_{1\text{-}42}$ dimers obtained by coarse-grained molecular dynamics (MD) simulations using the SIRAH force field. Trajectories are supplied in standard gromacs format and can be converted to fully atomistic representations for visualization and analysis using molecular visualization/analysis software. The dataset contains MD trajectories of A$\beta_{1\text{-}42}$ that undergo spontaneous and unbiased dimerization. We provide the time series of 25 replicates simulated for 11 microseconds under room conditions and physiological salt concentration. These multiple aggregation events provide valuable information not only on new binding pockets formed by the dimeric interface but also monomeric hot spots that small molecules can target on high-throughput docking campaigns. Alternatively, A$\beta_{1\text{-}42}$ dimers could be used as aggregation seeds in studies of A$\beta$ secondary nucleation.

The raw data were filtered so that only the protein coordinates were maintained, and simulation frames of 25 independent replicas were reported every 0.1 ns of simulation.

The set of coarse-grained (CG) molecular dynamics trajectories is deposited on Mendeley data with the DOI:10.17632/h8y867fkry.2

To visualize the CG trajectories:

1. Download the files from https://data.mendeley.com/datasets/h8y867fkry/2.
2. From a terminal, open the trajectory on VMD using the command line:

```
> vmd Abeta_1-42_dimer.psf Abeta_1-42_dimer_R1.xtc -e
sirah_vmdtk.tcl
```

Note that using the file sirah_vmdtk.tcl file, you are uploading a series of macros and definitions that allow using all the normal VMD drawing methods as vdw, licorice, etc., and coloring by restype, element, name, etc.

## 3. Experimental Design, Materials, and Methods

### 3.1. Primary data

A thorough description of the protocol used to produce the primary data is detailed in the associated paper [1]. Briefly, the coordinates from X-ray PDB structure: 1IYT were utilized as starting 3D model for A$\beta_{1\text{-}42}$ monomers. Fully atomistic coordinates were converted to the CG representation employing SIRAH Tools [2]. Subsequently, the CG peptides were inserted in a computational box, and NaCl ions were used to represent a salt concentration of 150 mM. MD simulations were performed in the NPT thermodynamical ensemble coupling the system at 1 atm. Randomized conformations were generated by heating the system to 373 K and then cooling it down to 300 K in consecutive steps. The version 2.0 of the SIRAH force field for CG and multiscale simulations [3] was used with and GROMACS 2018.4 MD package as a simulation engine [4].

We chose the three most populated conformational families using GROMACS utility gmx cluster, setting a cutoff of 0.1 nm for the gromos clustering algorithm. For simulating A$\beta_{1\text{-}42}$ dimerization, we placed two monomers in random reciprocal orientations imposing a separation of separating the by 5 nm between their centers of geometry. We prepared 54 independent systems and ran CG simulations at 300 K for 1 µs. Among those, 25 systems underwent dimerization within the simulated time window. The simulation of those systems was continued for 10 µs more.

### 3.2. Secondary data

The coarse-grained coordinates provided as primary data can be transformed into fully atomistic trajectories using SIRAHTools [2] to be used as secondary data. In order to do that, once you loaded the trajectory on VMD:

1. Open the Tk console from the *Extensions* pulldown in the main menu VMD window.
2. Typing `sirah_backmap` will backmap all the CG frames in the trajectory provided that you have Ambertools 18 or higher installed and the AMBERHOME environment variable properly set.

Typing `sirah_backmap help` will display all available options to backmap as, for instance, select only a few frames instead of the entire trajectories. The backmapping will create a new, fully atomistic molecule in VMD, which can be saved or analyzed as any normal all-atoms structure. It is important to notice that the backmapping procedure includes a minimization using the sander module of Amber, which does not conserve the chain identifiers in the newly created, fully atomistic, pdb files.

No other software than VMD and Ambertools and the sirah_vmdtk.tcl script (included in the dataset) are needed to analyze or backmap the CG trajectories.

## Ethics Statement

Not applicable.

## Credit author statement

- Conception and design of the study: E.E. Barrera
- Acquisition of data: E.E. Barrera
- Analysis and interpretation of data: E.E. Barrera
- Drafting the manuscript: E.E. Barrera, S. Pantano
- Revising the manuscript critically for important intellectual content: E.E. Barrera, S. Pantano.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships which have or could be perceived to have influenced the work reported in this article.

## Data availability

Early stages in Ab1-42 spontaneous aggregation: an unbiased dataset from coarse-grained molecular dynamics simulations. (Original data) (Mendeley Data).

## Acknowledgments

This work was partially funded by FOCEM (MERCOSUR Structural Convergence Fund), COF 03/11, and the Uruguayan Ministry of Education and Culture, II/FVF/2019/051. EEB and SP belong to the SNI program of ANII. Simulations were performed on the National Uruguayan Center for Supercomputing, ClusterUY.

## References

[1] E.E. Barrera, F. Zonta, S. Pantano, Dissecting the role of glutamine in seeding peptide aggregation, Comput. Struct. Biotechnol. J. (2021) In Press.
[2] M. Machado, S. Pantano, Structural Bioinformatics SIRAH Tools : mapping, backmapping and visualization of coarse-grained models, Bioinformatics 32 (2–3) (2016), doi:10.1093/bioinformatics/btw020.
[3] M.R. Machado, E.E. Barrera, F. Klein, M. Sóñora, S. Silva, S. Pantano, The SIRAH 2.0 Force Field: Altius, Fortius, Citius, J. Chem. Theory Comput. 15 (2019) 2719–2733, doi:10.1021/acs.jctc.9b00006.
[4] M.J. Abraham, T. Murtola, R. Schulz, S. Páll, J.C. Smith, B. Hess, et al., Gromacs: HIGH performance molecular simulations through multi-level parallelism from laptops to supercomputers, SoftwareX 1–2 (2015) 19–25, doi:10.1016/j.softx.2015.06.001.