Final Report

Biological Datasets for Computational Physics

Università degli Studi di Padova

BDfCP: Project

Poccianti Gabriele

Last update: February 18, 2025

Contents

1 | Introduction

1.1 | Arc and Tau

In this work, we studied the interactions between two proteins, both playing important—though not yet fully understood—roles in the progression of Alzheimer's disease (AD): Arc[1] and TAU[2].

The activity-regulated cytoskeleton-associated protein (Arc) is an immediate early gene that plays a key role in synaptic plasticity, learning, and memory [2]. This function is likely related to its ability to form virus-like capsids [3, 4], facilitating the intercellular transfer of mRNA between adjacent cells.

On the other hand, TAU primarily stabilizes microtubules, but several studies indicate that its misfolded aggregates (known as neurofibrillary tangles) correlate with neurodegenerative diseases such as Alzheimer's disease (AD) and Parkinson's disease [5]. Some studies question whether neurofibrillary tangles are an essential feature of these illnesses [6, 7], but this issue is beyond the scope of our study. The exact mechanisms of ARC-TAU interaction remain unknown, despite strong evidence suggesting its involvement in neurodegeneration [8]. Similarly, the mechanisms of TAU release and transmission are not yet fully understood. Some studies propose the involvement of extracellular vesicles (EVs) in this process [9].

Our goal is to gain insight into how hTau interacts with Arc and, in particular, to determine whether Arc can maintain a stable monomeric structure or form a viral capsid after binding.

Since no experimental structures were available for our proteins, we generated these using AlphaFold [10] and then performed a conformational analysis, particularly comparing them to the structures obtained for isolated Arc. Finally, we conducted an all-atom molecular dynamics (MD) simulation in explicit solvent and repeated the analysis for the final configuration.

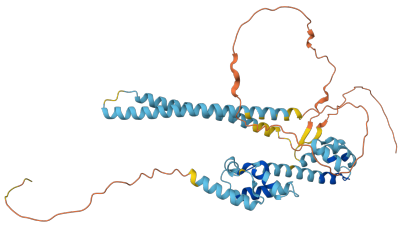

Figure 1.1: AF-Q7LC44 (Arc Human) - AlphaFold Protein Structure Database. Color represents the level of confidence: from very high (blue) to very low (orange).

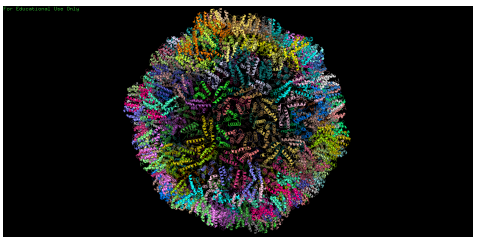

Figure 1.2: Experimental resolution of 6TAP - biological assembly of the dArc1 capsid, Drosophila melanogaster.

[1] Human, Uniprot: Q7LC44

[2] We considered the 2N4R isoform (human). Uniprot: P10636-8. AD P-tau preferentially interacts with this isoform [1]

2 | Methods

2.1 | Dataset Assembly

We used AlphaFold to generate the structure of the ARC-TAU complex, focusing on the four repeats of TAU (residues 244–369), as these binding domains are more prone to aggregation.

To conduct a population dynamics analysis, we generated the structures using 106 different random seeds.[1].

To enhance sampling, we considered the five different models generated by AlphaFold for each structure. Additionally, we generated ARC structures using the same set of random seeds for comparison.

For both ARC and ARC-TAU complex structures, we analyzed the distribution of angles between the N-terminal and the capsid domain. This was achieved by calculating the angle between the First Principal Component (FPC) of these regions. The angles were compiled into a dataset categorized by seed and model.

We applied k-means clustering (determining the optimal number of clusters via the elbow method) to both the ARC and ARC-TAU complex populations. The resulting data were then fitted with Gaussian distributions, with cluster amplitudes scaled according to their relative population fractions.

A second dataset (`interaction_df`) was created to track, for each seed, model, and TAU repeat domain (R1–R4), the fraction of residues interacting with either the N-terminal or the capsid domain. The dataset includes columns for `seed`, `model`, `tau_terminal (R1-R4)`, `N_terminal_fraction`, and `Capsid_fraction`.

Contact maps were generated for given PDB structures by creating a grid of TAU and ARC residues and highlighting pairs with distances below a predefined threshold (6Å, unless otherwise specified).

2.1.1 Sequence Alignment

We performed sequence alignment between human ARC (Uniprot: Q7LC44) and its homologs dArc1 and dArc2 in the brachyceran fly lineage using Clustal Omega [11]. Given the asymmetric unit of the viral domains (PDBs: 6TAP and 6TAQ), we com-

[1] 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 18323, 375994, 790690, 90638, 169881, 657436, 773018, 317349, 52779, 213773, 830834, 693800, 255822, 349839, 872026, 268123, 268123, 914923, 547515, 964412, 896980, 717051, 603037, 413174, 528017, 309838, 304693, 288233, 38706, 803721, 319530, 303268, 706246, 112014, 6231, 415984, 669627, 374627, 816328, 373294, 272166, 258814, 983841, 868525, 197085, 62426, 435229, 723114, 455106, 748535, 237310, 482750, 984735, 445807, 414227, 416917, 390067, 490618, 512984, 910069, 998565, 134145, 649600, 839941, 957586, 754773, 573537, 74062, 260587, 802990, 918271, 643021, 11957, 580611, 873809, 773396, 316612, 741413, 226462, 608486, 868314, 320419, 950211, 682301, 621333, 313213, 989736

puted the RMSD between mapped residues of each chain and our ARC model. The chain with the lowest RMSD was selected for further analysis.

2.1.2 Simulations

The simulations were conducted primarily using the Python library *OpenMM*.

The main steps included:

- Cleaning the input file (adding missing residues, atoms, and hydrogens).
- Setting up a **truncated octahedron** simulation box and aligning the principal axes of inertia along a diagonal to minimize the number of solvent molecules required.
- Equilibrating the system using *Packmol*[2], with an ionic concentration of 0.150 M and ensuring a **minimum distance of 10 Å** between the ions and the protein.
- Using the **AMBER14-all** force field for proteins and **AMBER14/TIP3P** for water.
- Treating the ARC N-terminal and viral capsid as rigid bodies by importing a rigid body script from the `OpenMM Code Repository`[3].
- Running an (N,P,T) simulation at $T = 300K$ using the `LangevinMiddleIntegrator` with a 2 fs timestep, coupled with a **Monte Carlo Barostat** set at 1 atm.

Unless otherwise specified, simulations ran for 20 nanoseconds.[4]

2.1.3 Statistical Analysis

Unless specified otherwise, a significance level of $\alpha = 0.05$ was used.

TOST Test

We employed the two one-sided t-test (TOST) to determine whether the difference between two means was within a predefined equivalence margin θ:

$$\begin{aligned} H_0 &: \mu_2 - \mu_1 \leq -\theta \quad \text{or} \quad \mu_2 - \mu_1 \geq \theta \\ H_1 &: -\theta < \mu_2 - \mu_1 < \theta \end{aligned} \tag{2.1}$$

ANOVA and Tukey's HSD Test

We performed a one-way ANOVA to determine whether the fraction of interactions with the N-terminal ($N_{\text{terminal\_fraction}}$) significantly differed across the four TAU domains ($R1$, $R2$, $R3$, $R4$). If significant, we applied Tukey's Honestly Significant Difference (HSD) test for pairwise comparisons.

[2] https://m3g.github.io/packmol/

[3] https://simtk.org/plugins/moinmoin/openmm/VirtualRepository

[4] The following starting configurations were used: 416917 (model 4), 316612 (model 0), 802990 (model 0), and 260587 (model 3). These are referred to as Sim 1–4, respectively. Sim 1 was run for 40 nanoseconds.

2.1.4 Resources

Simulations were initially run on Google Colab and ColabPro, but exceeded runtime limits in both cases. Surprisingly, even ColabPro's GPU time limit was reached. As a result, simulations were conducted locally on a Windows 11 machine with an NVIDIA 4060 Ti GPU (16 GB RAM).

2.1.5 Code

All code was written in Python. Key libraries included:

- **sklearn** for PCA
- **Bio** and **MDAnalysis** for PDB structure manipulation
- **statsmodels** for statistical analysis

3 | Results

3.1 | Preliminary Population Analysis

We now present the results of the population analysis. Looking at the histograms (3.1-3.2), we observe that for ARC (Fig. 3.1), the data is well-fitted by a single curve, indicating a well-defined average conformation. No distinct physical meaning for the clusters was identified in this case.

On the other hand, when the structure is perturbed by TAU (Fig. 3.2), two peaks emerge, each corresponding to a separate cluster. This suggests the presence of two alternative conformations.

To test this hypothesis, we performed a two-sample equivalence test, which revealed that the difference between the mean angles of each cluster is equivalent to 83° with a margin of ±4° (p-value = 0.0259).

In Fig. 3.6, we present two representative structures.

When analyzing the contact patterns, we employed an ANOVA test and found that the fraction of interacting residues varies significantly with the domain (p-value = 1.7×10^{-9}).

Further inspection using Tukey's test showed that, for the N-terminal, the interacting fractions of $R2$ with $R3$ and $R4$ cannot be distinguished. However, the other pairwise quantities are statistically different (Fig. 3.3).

In contrast, only $R3$ shows statistically distinct behavior when interacting with the viral domain (Fig. 3.4).

Finally, since the experimental structure of ARC's capsid is available for Drosophila, we determined that dARC1 (biological assembly, chain E) is the most compatible structure, with an average RMSD of (13.0 ± 0.6) Å.

3.2 | After the Dynamics

We begin by noting that, due to the limited time available, we were unable to collect sufficient data for a comprehensive dynamical statistical analysis. As such, many of the conclusions presented here will be qualitative.

First, we analyzed how the angles between the N-terminal and viral domain evolved over time. Examining the difference between the initial and final configurations (3.10) and the time evolution (3.15) does not provide conclusive results, as there is no clear trend. In particular, structures belonging to the same cluster (as shown in Fig. 3.2) exhibit varying behavior. This is consistent with the "fuzzy" nature of the interaction between ARC and TAU [9, 12].

Furthermore, we refined our analysis by focusing on one trajectory to examine whether the evolution varied smoothly over time or not[1].

Moreover, we investigated how the interacting fractions changed during the dynamics (Fig. 3.7-3.8). We observe that the fractions tend to vary predominantly in one domain, which may indicate that the interactions between the two proteins are localized. Specifically, the interaction between TAU and the N-terminal appears more stable than that with the Gag domain (in agreement with [9]).

Finally, the compatibility with the capsid domain at the end of the simulation increases, as shown in Fig. 3.10 - 3.12. This result may contradict our expectations, as one theory suggests that tau overexpression may lead to the extracellular release of ARC capsids [13]. However, [8] reports that "Arc1 capsid formation is increased in the context of tauopathy." Additionally, [9] suggests that ARC and TAU are packaged into EVs, making a further stabilization of the viral structure compatible with this scenario.

[1] We considered seed 416917, model 4.

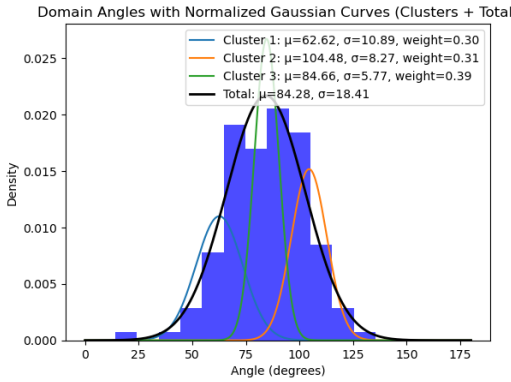

Figure 3.1: Histogram of the distribution of the angles between n-terminal and capsid domain across different models (Single Arc structure). In this case the best fit is achieved by a single gaussian, which suggests a unique structure. The smaller gaussians fits are relative to different clusters

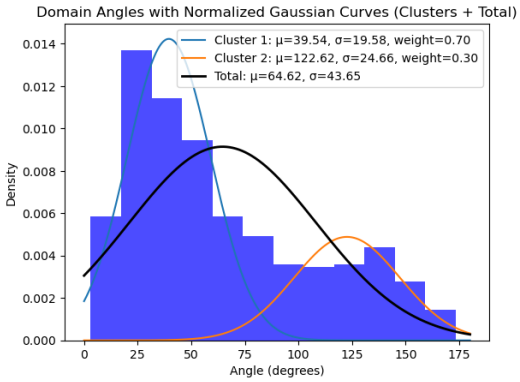

Figure 3.2: Histogram of the distribution of the angles between n-terminal and capsid domain across different models (Arc + Tau). In this case the two clusters seem to better describe the system, indicating the presence of new favorable conformations

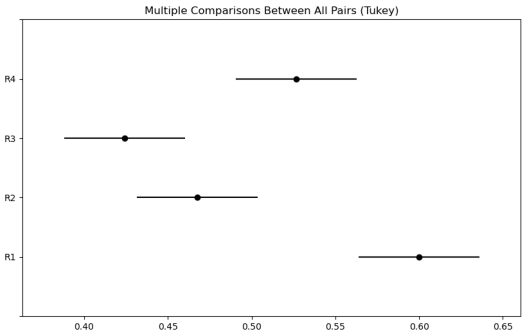

Figure 3.3: Tukey test for N-terminal fractions ($\alpha = 0.05$), showing significant differences in interactions.

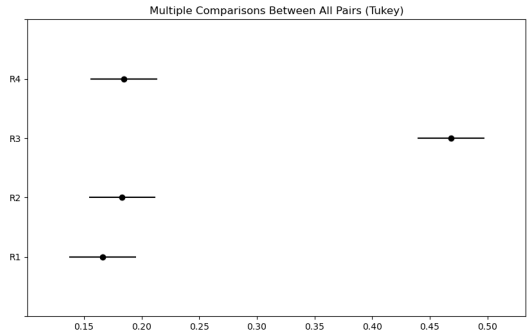

Figure 3.4: Tukey test for the viral domain.

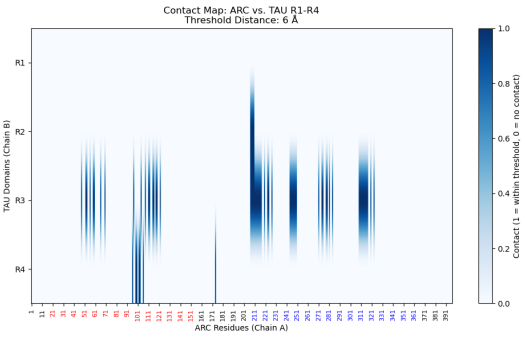

Figure 3.5: Contact map for the ARC and Tau complex (seed 8, model 4). Threshold distance: 6 Å.

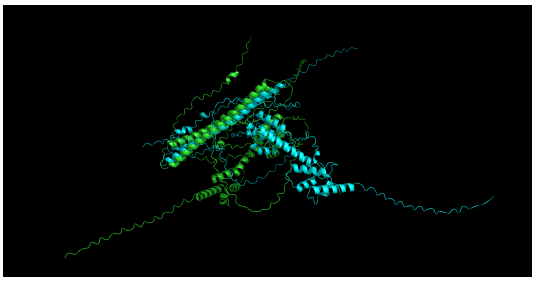

Figure 3.6: Representative structures: seed 2, model 0 (green, cluster 1) and seed 6, model 0 (blue, cluster 2).

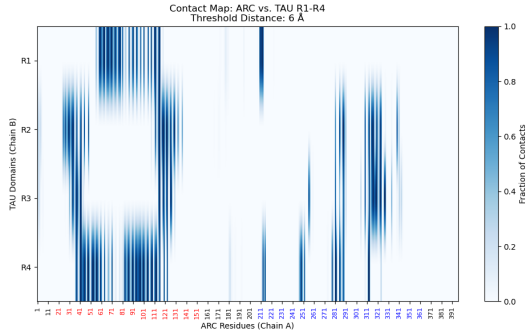

Figure 3.7: Dynamic contact map between Arc and the R1-4 repeats of Tau for simulation 1. The treshold is set as 6 Å and the colormap represents the fraction of interacting frames.

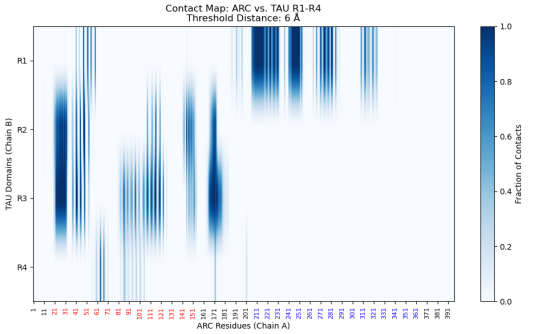

Figure 3.8: Dynamic contact map between Arc and the R1-4 repeats of Tau for simulation 3. The treshold is set as 6 Å and the colormap represents the fraction of interacting frames.

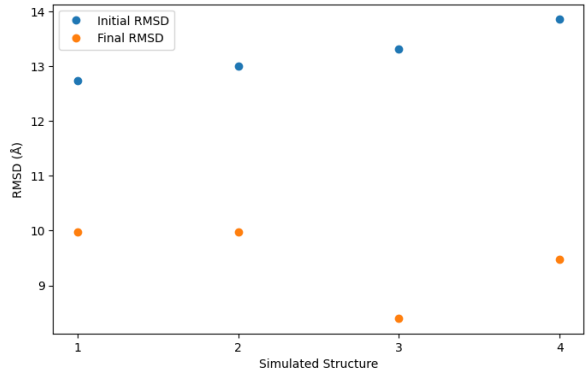

Figure 3.9: RMSD with dArc1 and different structures (at the beginning and end of the simulation). We can see that the ARC tends to become more compatible with a viral structure

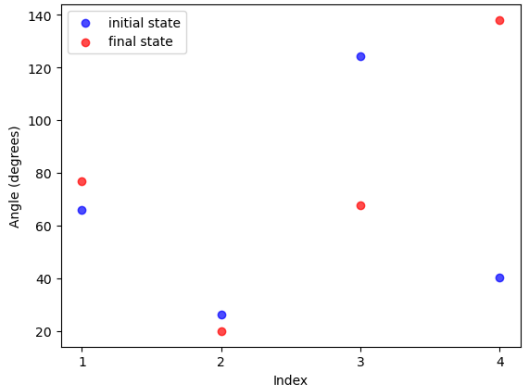

Figure 3.10: Angles between Arc n-terminal and viral domain before and after the simulations. We do not observe a clear tendency, a possible sign of a fuzzy interaction

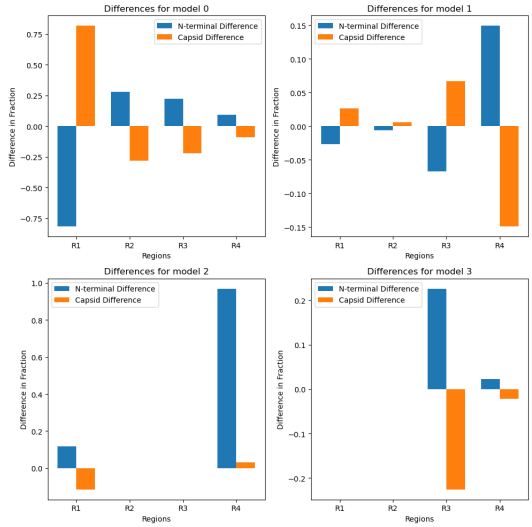

Figure 3.11: Difference between the final and initial fraction of interacting residues for each TAU domain. We can observe that the change is mostly localized in one region.

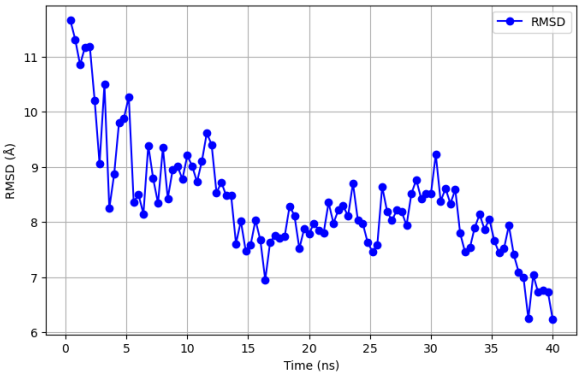

Figure 3.12: RMSD with the Drosophila structure (dArc1) for sim 1. Even if we don't reach equilibrium, we can see that the RMSD has decreased wrt the initial structure

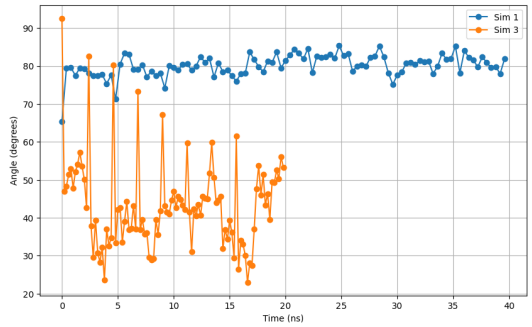

Figure 3.13: Angles between n-terminal and viral domain of Arc for the simulations one and three. We can see that the behaviors are very different: In the first case the angle increases and seems to stabilize, where in the second it decreases, with spikes, probably due to a very different configuration.

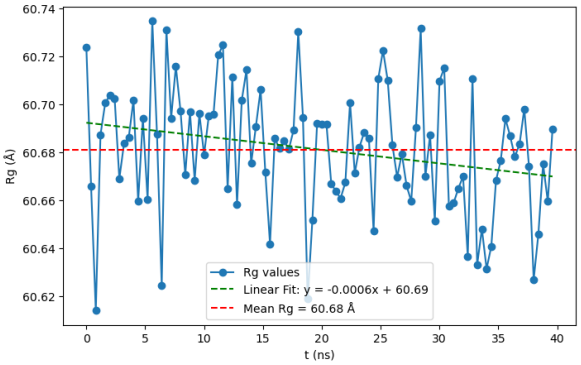

Figure 3.14: Radius of Gyration of the trajectory of sim 1 over time. We can observe almost no trend

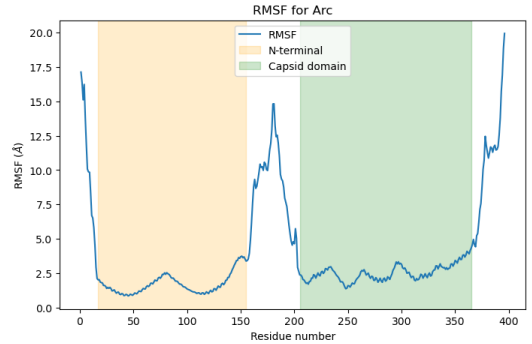

Figure 3.15: RMSF for ARC (sim 1). Unsurprisingly the rigid chains do not display major conformational changes, whereas most of the movement is focused on the middle part and in the chains

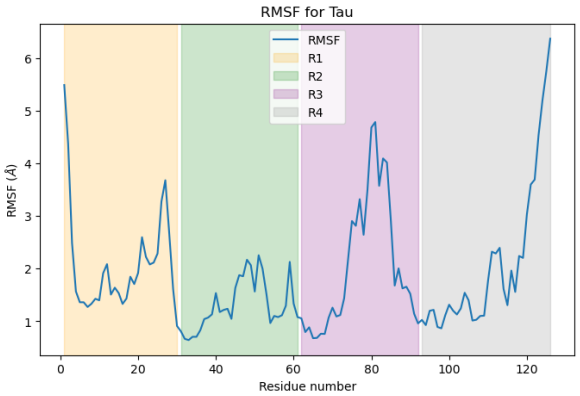

Figure 3.16: RMSF for TAU (sim 1). Here the most mobile sections are the chains, which tend to collapse during the dynamics

4 | Conclusions

We can note that, given the limited simulation time available, the system can't be considered at equilibrium. In proof of this, we can look at figure 4.1, where we clearly observe that the RMSD (frame at time t vs at time 0) of the complex in sim 1 has not converged. In fact the downspikes suggest, together with our previous observations about domain angles and contact patterns, that the complex is fuzzy, not capable of keeping a stable monomeric structure. This scenario may seem, at first glance, in contradiction with the improved ability of Arc, after binding with Tau, to form a capsid. However these aspects, even if related, may refer to different phenomena: the first may be linked to the Arc physiological function of regulating synaptic plasticity [14]. On the other side the formation of capsid may facilitate the encapsulation of Arc and Tau in EVs, an important part of the spreading of Tau pathology.

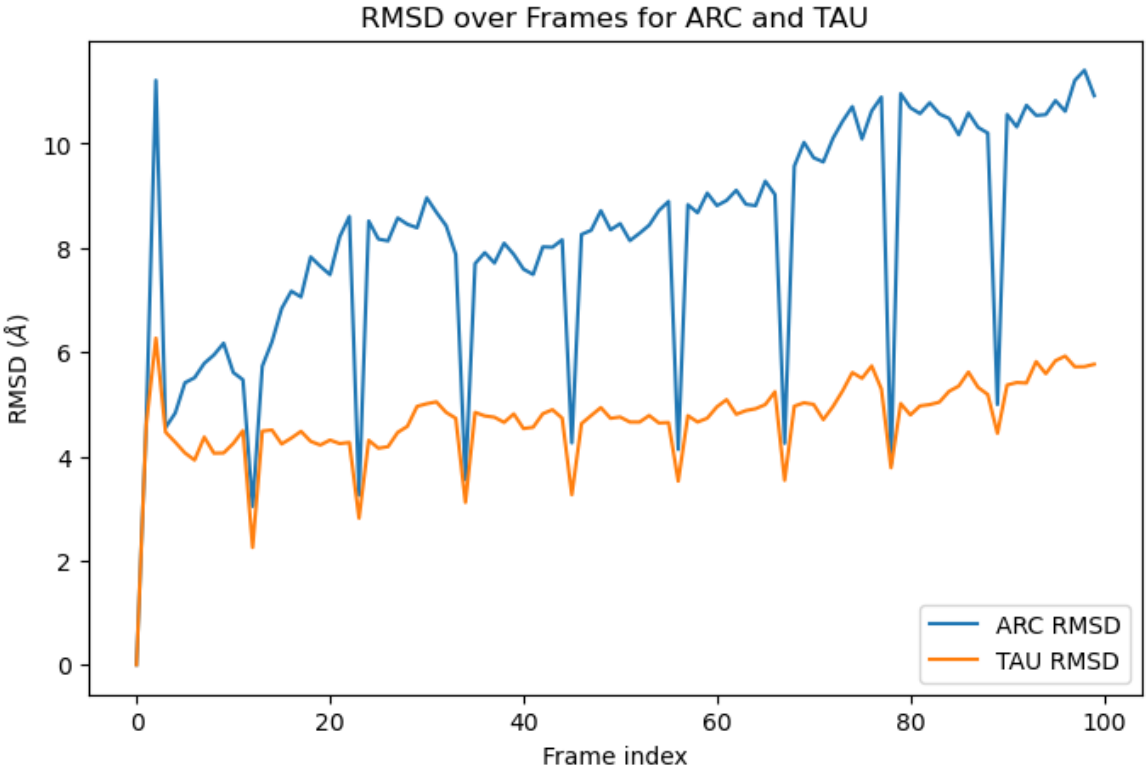

Figure 4.1: RMSD (structure at time t vs at time 0) as a function of time for ARC and TAU. After a transient we can see that the RMSD is still trending, indicating that the system hasn't reached equilibrium yet. Moreover we can note the presence of downspikes, a sign that the system is assuming different conformations

4.1 | Potential New Directions

One possible improvement to our work could be to consider beta propensities. We know that Tau, in particular adopts β-sheet conformations during aggregation[15], leading

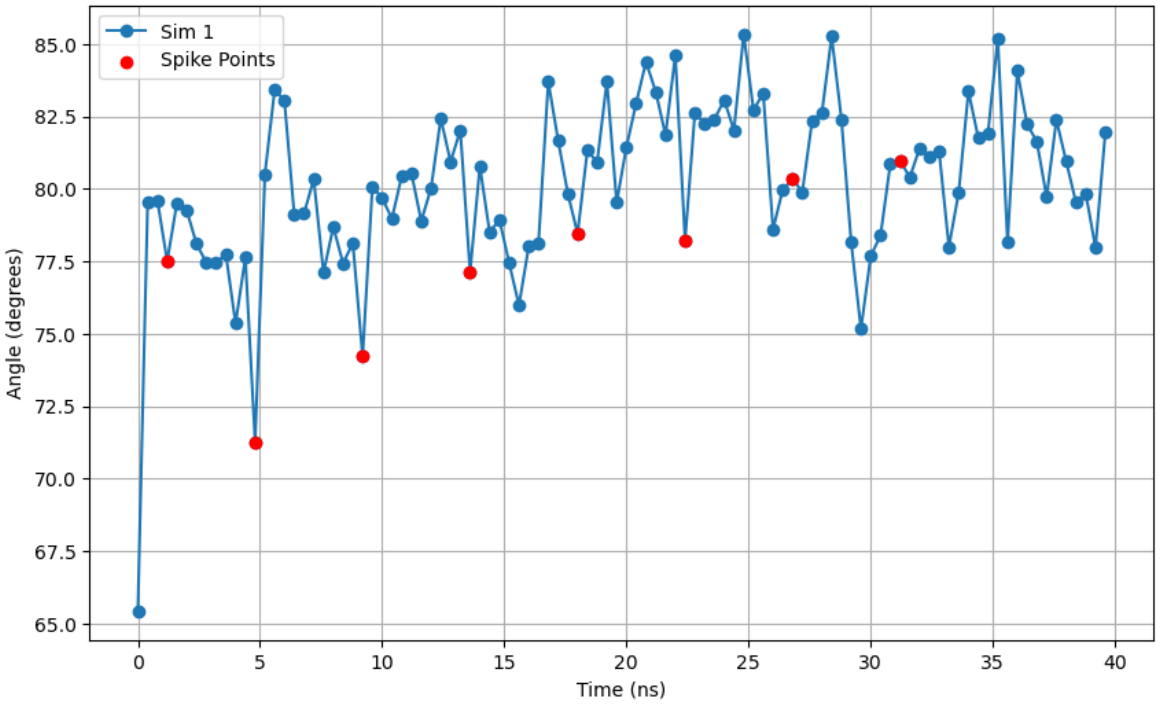

Figure 4.2: We can observe that most of the spikes correspond to local minima of the angles between Arc and Tau, indicating that they are related to structure quite different from the others

to the formation of neurofibrillary tangles associated with neurodegenerative diseases. By analyzing β-sheet propensities, we can gain insights into how Arc influences Tau's conformational transitions and aggregation behavior. It would be of special interest which residues of Arc are more prone to interact with the tangles.

Another crucial improvement to our work could be to run more simulations for longer time periods (for instancearound $200ns$) and observe how the quantities of interest vary wrt our previous analysis. This would allow for a relevant statistical analysis.

Finally, with more computational resources, we could relax the rigid bodies constrain; allowing for a more realistic threatement of the Arc's domains.

5 | Bibliography

[1] Alejandra del C. Alonso, Tanweer Zaidi, Michal Novak, Hector S. Barra, Inge Grundke-Iqbal, and Khalid Iqbal. Interaction of tau isoforms with alzheimer's disease abnormally hyperphosphorylated tau and in vitro phosphorylation into the disease-like protein. *Journal of Biological Chemistry*, 276(41):37967–37973, 2001.

[2] Edward D. Pastuzyn, Carolyn E. Day, Ryan B. Kearns, Melanie Kyrke-Smith, Andrew V. Taibi, James McCormick, Nathan Yoder, David M. Belnap, Sigurður Erlendsson, Dustin R. Morado, John A. G. Briggs, Cédric Feschotte, and Jason D. Shepherd. The neuronal gene arc encodes a repurposed retrotransposon gag protein that mediates intercellular rna transfer. *Cell*, 172(1-2):275–288.e18, 2018. Erratum in: Cell. 2018 Mar 22;173(1):275. doi: 10.1016/j.cell.2018.03.024.

[3] Edward D. Pastuzyn, C. E. Day, R. B. Kearns, M. Kyrke-Smith, A. V. Taibi, J. McCormick, N. Yoder, D. M. Belnap, S. Erlendsson, D. R. Morado, J. A. G. Briggs, C. Feschotte, and J. D. Shepherd. The neuronal gene arc encodes a repurposed retrotransposon gag protein that mediates intercellular rna transfer. *Cell*, 172(1-2):275–288.e18, 2018. Erratum in: Cell. 2018 Mar 22;173(1):275. doi: 10.1016/j.cell.2018.03.024.

[4] L. D. Nielsen, C. P. Pedersen, S. Erlendsson, and K. Teilum. The capsid domain of arc changes its oligomerization propensity through direct interaction with the nmda receptor. *Structure*, 27(7):1071–1081.e5, Jul 2019.

[5] Wim Annaert and Bart De Strooper. A cell biological perspective on alzheimer's disease. *Annual Review of Cell and Developmental Biology*, 18:25–51, 2002. Epub 2002 Apr 2.

[6] L. M. Fox, C. M. William, D. H. Adamowicz, R. Pitstick, G. A. Carlson, T. L. Spires-Jones, and B. T. Hyman. Soluble tau species, not neurofibrillary aggregates, disrupt neural system integration in a tau transgenic model. *J Neuropathol Exp Neurol*, 70(7):588–595, Jul 2011.

[7] N. Rudinskiy, J. M. Hawkes, S. Wegmann, K. V. Kuchibhotla, A. Muzikansky, R. A. Betensky, T. L. Spires-Jones, and B. T. Hyman. Tau pathology does not affect experience-driven single-neuron and network-wide arc/arg3.1 responses. *Acta Neuropathol Commun*, 2:63, Jun 2014.

[8] L. Schulz, P. Ramirez, A. Lemieux, E. Gonzalez, T. Thomson, and B. Frost. Tau-induced elevation of the activity-regulated cytoskeleton associated protein arc1 causally mediates neurodegeneration in the adult drosophila brain. *Neuroscience*, 518:101–111, May 2023. Epub 2022 Apr 27.

[9] Mitali Tyagi, Radhika Chadha, Eric de Hoog, Kaelan R. Sullivan, Alicia C. Walker, Ava Northrop, Balazs Fabian, Monika Fuxreiter, Bradley T. Hyman, and

Jason D. Shepherd. Arc mediates intercellular tau transmission via extracellular vesicles. *bioRxiv*, October 2024.

[10] J. Abramson, J. Adler, J. Dunger, R. Evans, T. Green, A. Pritzel, O. Ronneberger, L. Willmore, A. J. Ballard, J. Bambrick, S. W. Bodenstein, D. A. Evans, C. C. Hung, M. O'Neill, D. Reiman, K. Tunyasuvunakool, Z. Wu, A. Žemgulytė, E. Arvaniti, C. Beattie, O. Bertolli, A. Bridgland, A. Cherepanov, M. Congreve, A. I. Cowen-Rivers, A. Cowie, M. Figurnov, F. B. Fuchs, H. Gladman, R. Jain, Y. A. Khan, C. M. R. Low, K. Perlin, A. Potapenko, P. Savy, S. Singh, A. Stecula, A. Thillaisundaram, C. Tong, S. Yakneen, E. D. Zhong, M. Zielinski, A. Žídek, V. Bapst, P. Kohli, M. Jaderberg, D. Hassabis, and J. M. Jumper. Addendum: Accurate structure prediction of biomolecular interactions with alphafold 3. *Nature*, 636(8042):E4, Dec 2024. Erratum for: doi: 10.1038/s41586-024-07487-w.

[11] Fabian Sievers and Desmond G. Higgins. Clustal omega for making accurate alignments of many protein sequences. *Protein Science*, 27(1):135–145, Jan 2018. Epub 2017 Oct 30.

[12] M. Fuxreiter. Context-dependent, fuzzy protein interactions: Towards sequence-based insights. *Current Opinion in Structural Biology*, 87:102834, August 2024. Epub 2024 May 16, PMID: 38759297.

[13] D. W. Yakout, A. Shroff, W. Wei, V. Thaker, Z. D. Allen, M. Sajish, T. Y. Nazarko, and A. M. Mabb. Tau regulates arc stability in neuronal dendrites via a proteasome-sensitive but ubiquitin-independent pathway. *Journal of Biological Chemistry*, 300(5):107237, 2024. Epub 2024 Mar 27.

[14] O. Nikolaienko, S. Patil, M. S. Eriksen, and C. R. Bramham. Arc protein: a flexible hub for synaptic plasticity and cognition. *Semin Cell Dev Biol*, 77:33–42, May 2018.

[15] Sofia Lövestam, David Li, Jane L. Wagstaff, Abhay Kotecha, Daniel Kimanius, and S. H. W. Scheres. Disease-specific tau filaments assemble via polymorphic intermediates. *Nature*, 625(7993):119–125, January 2024.