Computational clarification of the molecular mechanism that allows determining the number of growing ends in polyglutamine aggregation

Aniruddha Bapat and Markus S. Miettinen*

Freie Universität Berlin Fachbereich Physik

(Dated: October 15, 2015)

Abstract. We use a combination of all-atom and coarse grained simulations to study the molecular details of an assay used for determining the number of growing ends in polyglutamine aggregation kinetics experiments. To this end we perform (all-atom) simulations of polyglutamine fibrils tagged with biotin-PEG-polyglutamine, as well as (coarse grained) simulations of the interaction of these fibrils with streptavidin. For the latter part we develop a model for biotin within the Martini framework.

* This work is progressed and discussed openly at `markussmiettinen.github.io/biotinPEGpolyQ`. Following the approach established in the NMRlipids project (`nmrlipids.blogspot.fi/2013/07/on-credits.html`), everyone is invited to join the discussion and make contributions. The manuscript will be eventually submitted to an appropriate peer reviewed journal, and everyone who has openly contributed, will be offered coauthorship. The final decision on authorship will be to the invited individuals themselves; they should base their decision on self-assessment of their scientific contribution to the project. We hope that all communications would be written such that they are assignable to a person and affiliation, as this will remove unnecessary ambiguity upon publishing. Finally, it is the responsibility of contributors to make sure that in case they write something in which other people have had a scientific contribution (such as supervisors writing comments based on discussions with their students), the identities of all contributors should be made known. These secondary contributors will also be offered coauthorship in the manuscript. As the project is part of A. Bapat's studies, he will be the first author; otherwise, the author names are alphabetically ordered with M. S. Miettinen as the corresponding author.

* This manuscript describes the current status of the project. We use four kinds of comments to point out issues: 1) general comments (green), 2) todo items (orange), 3) asking references to be added (blue), and 4) pointing out mistakes (red). The list of open issues is in the end of the manuscript.

`\comment`
`\todo`
`\reference`
`\wrong`

I. INTRODUCTION

Nine inherited neurodegenerative diseases, including Huntington's disease, are found to be caused by proteins having a polyglutamine (polyQ) repeat exceeding a threshold of (typically) 35+ consecutive glutamines [1, 2]. As the glutamines in the polyQ repeat are in DNA coded using exclusively the CAG nucleotide triplet, the *polyQ diseases* are also known as the *CAG triplet disorders*. The onset and severity of polyQ diseases correlate with the length of the polyQ sequence, and in the neuronal cells of the deceased patients fibrillar aggregates rich in polyQ are found. Furthermore, as the proteins causing the diseases lack common features other than the polyQ tract, it is thought that associated with this tract is some common molecular property that is critical for understanding all the diseases.

In our previous simulational work on polyQ we have focused on understanding the molecular details of how its aggregation is initiated. With all-atom explicit water molecular dynamics simulations and docking calculations we have estimated the relative feasibility of six different suggested conformations [3–8] of a Q_{40} peptide to be the aggregation-initiating conformer. Our simulations strongly support β-hairpin-containing conformers, and highlight the importance of side chain interdigitation between hairpins [9, 10].

A. Why is it important to know the number of growing ends?

Many molecular details of an amyloid formation process can be assessed through aggregation kinetics experiments, in other words, by measuring the concentration of aggregated peptides as a function of time, and analyzing these data using nucleated growth polymerization theory [11]. This applies not only to polyQ, but also to understanding the aggregation of peptides linked to other amyloid diseases, such as Alzheimer's or Parkinson's. A measure of particular interest here is the rate at which new peptides are being added to the growing fibril; for example, when combined with molecular dynamics simulations, this information can be used to assess which molecular structures can initiate the aggre-

---
* Author to whom correspondence may be addressed. `markus.miettinen@iki.fi`

gation process [10]. Quantitative determination of this rate, however, requires determining the number of growth sites (growing ends) in a kinetics experiment [11].

B. The assay used when experimentally determining the number of growing ends

To determine the number of growing ends in a polyQ aggregation experiment, biotinylated polyQ molecules (Fig. 1) are used [12, 13]. In the experiment one incubates a previously unknown number of preformed seeds (fibril fragments) of non-biotinyled polyQ in a bath of biotinylated polyQ. After incubation one adds Eu-streptavidin. The streptavidin binds to biotins and the europium allows concentration determination through fluorescence. The key feature of the assay is that it can be demonstrated [13] that, for some unknown reason, only one streptavidin appears to be bound per growing end: The Eu-fluorescence saturates orders of magnitude faster than the aggregation process itself, and the values for the concentration of growing ends are highly reproducible.
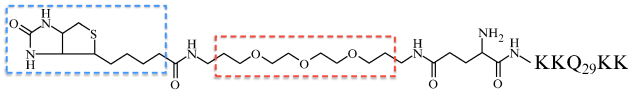

FIG. 1: Structure of the biotinylated polyQ, highlighting the biotin tag (blue), and the PEG linker (red).

C. Hypothesis on the molecular mechanism

Based on the experimental findings [12, 13], it appears that any biotins on the biotin-PEG-polyQ chains within the fibril are for some reason inaccessible to binding by streptavidin, in other words, only the biotin on the last bound molecule will bind a streptavidin and give signal; although there is some experimental support for this hypothesis, [14] the molecular mechanism remains unknown. [15]

D. Clarification of the molecular mechanism using all-atom simulations

To test the hypothesis, all-atom explicit water simulations with three different setups will be run: (1) single biotin-PEG-polyQ molecule in the growing end of a pure-polyQ fibril, (2) single biotin-PEG-polyQ molecule in the middle of a pure-polyQ fibril, and (3) fibril comprising only biotin-PEG-polyQ molecules.

Comparing the conformational distribution of the biotin-PEG tag between cases (1) and (2) will reveal if (and to what extent) the position of the tag within the fibril affects its availibility to streptavidin. A possible scenario is, for example, that when the tag is in the middle of the fibril, it will more strongly bind to the fibril surface than when being located at the growing end. Case (3) will reveal to what extent neighboring biotin-PEG tags interact with one another; do they, e.g., fold in such a way as to prevent them from being available for streptavidin binding.

E. Clarification of the molecular mechanism using coarse grained simulations.

Streptavidin is a tetrameric protein, whose each monomer comprises over 100 amino acids. Its sheer size thus prohibits direct all-atom simulations of its interaction with the biotinylated polyQ fibril. We, therefore, resort into coarse grained simulations using the Martini model [16] to study the interaction of the streptavidin tetramer with the same three fibril systems that we studied with the all-atom model. These simulations will allow us to confirm if the differences found in the biotin-PEG tag conformations using the atomistic model also translate into differences with regards to the access of biotin into the binding pocket of streptavidin.

II. METHODS

Building the all-atom model

> **Aniruddha**, start writing here a description on how you are building the fibril.

All-atom explicit water simulations with three different setups will be run: (1) single biotin-PEG-polyQ molecule in the growing end of a pure-polyQ fibril, (2) single biotin-PEG-polyQ molecule in the middle of a pure-polyQ fibril, and (3) fibril comprising only biotin-PEG-polyQ molecules. To guarantee the reliability of the results, 10 independent simulations for each setup will be performed.

There is no existing molecular dynamics force field for the biotin-PEG-polyQ. However, within the CHARMM force field family, models for both the biotin [17] and PEG [18] have been created. Thus a force field for the biotin-PEG-polyQ could, in principle, be built simply by joining

these existing building blocks of biotin and PEG to a polyglutamine chain described using the CHARMM36 [19] force field.

Having said that, the above mentioned works on biotin and PEG are rather old, and using the powerful modern tools for creating force fields for small molecules, such as CGenFF [20], to create a CHARMM36-compatible biotin-PEG from scratch, will possibly lead to a more reliable result.

> **Scaling plots (DL: Oct 25th).** To apply for computing resources we need to run scaling tests. For this we need a system that is as close to the actual system we will work on as possible. Naturally, we can not include biotins and PEGs yet, but pure polyQ will suffice here. To this end we must: **(1)** estimate the size of the fibril (number of Qs, number of waters) we need to simulate, **(2)** build a system of this size (does not need to be a fibril), **(3)** simulate with varying number of cores.

Building the coarse grained model

As in the all-atom case, there is exists no biotin-PEG-polyQ in Martini [16]. However, Dr. Luca Monticelli's group has recently developed a Martini model for PEG [21]. In addition, we have developed a model of glutamine that allows fibril formation by allowing for interdigitation of stacked hairpins—a feature that was prevented by the shape of the standard Martini [22]

glutamine, but which I have shown is highly important in correctly describing polyQ aggregation and fibril formation [10].

Thus the only part missing is the biotin, which we shall parametrize in such a way as to reproduce the conformational distributions in our all-atom simulations of the biotinylated polyQ fibrils.

Notably, we are not going to focus on carefully parametizing the interactions of biotin with the streptavidin binding pocket; we are only interested the ability of biotin to enter the pocket, not its actual binding affinity. (However, the biotin model we develop will probably be useful as a starting point for other researchers that wish to study the biotin–streptavidin interaction.)
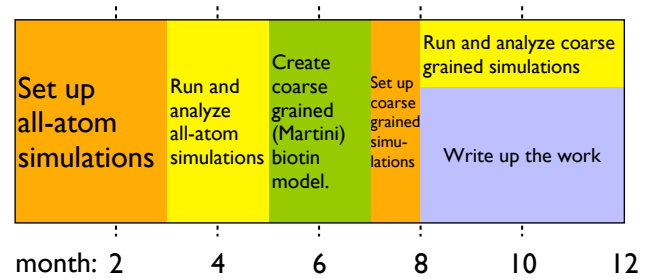
III. RESULTS

IV. DISCUSSION

> Should probably **discuss the recent paper by de Pablo [23]**, where they show how (badly) current MD FFs describe polyQ in solution. Not a crucial problem for us here (which we should point out), as our polyQ is in the fibril, but a worthwhile work nonetheless.

BIBLIOGRAPHY

[1] A. J. Williams and H. L. Paulson, Trends in Neurosciences **31**, 521 (2008).

[2] S. L. Hands and A. Wyttenbach, Acta Neuropathol. **120**, 419 (2010).

[3] M. F. Perutz, T. Johnson, M. Suzuki, and J. T. Finch, Proc. Natl. Acad. Sci. U.S.A. **91**, 5355 (1994).

[4] M. F. Perutz, J. T. Finch, J. Berriman, and A. Lesk, Proc. Natl. Acad. Sci. U.S.A. **99**, 5591 (2002).

[5] P. Sikorski and E. Atkins, Biomacromolecules **6**, 425 (2005).

[6] D. Zanuy, K. Gunasekaran, A. M. Lesk, and R. Nussinov, J. Mol. Biol. **358**, 330 (2006).

[7] F. Fiumara, L. Fioriti, E. R. Kandel, and W. A. Hendrickson, Cell **143**, 1121 (2010).

[8] K. Kar, M. Jayaraman, B. Sahoo, R. Kodali, and R. Wetzel, Nat. Struct. Mol. Biol. **18**, 328 (2011).

[9] M. S. Miettinen, V. Knecht, L. Monticelli, and Z. Ignatova, J. Phys. Chem. B **116**, 10259 (2012).

[10] M. S. Miettinen, L. Monticelli, P. Nedumpully-Govindan, V. Knecht, and Z. Ignatova, *Biophysical Journal*, Biophys. J. **106**, 1721 (2014).

[11] F. Ferrone, "Analysis of protein aggregation kinetics," (Academic Press, 1999) Chap. 17, pp. 256–274.

[12] A. M. Bhattacharyya, A. K. Thakur, and R. Wetzel, Proc. Natl. Acad. Sci. U.S.A. **102**, 15400 (2005).

[13] B. O'Nuallain, A. K. Thakur, A. D. Williams, A. M. Bhattacharyya, S. Chen, G. Thiagarajan, and R. Wetzel, Methods. Enzymol. **413**, 34 (2006).

[14] Private communication with professor Ronald Wetzel, University of Pittsburgh.

[15] Personal communication from professor Ronald Wetzel and from Dr. Elizabeth Landrum, University of Pittsburgh.

[16] S. J. Marrink, H. J. Risselada, S. Yefimov, D. P. Tieleman, and A. H. de Vries, J. Phys. Chem. B **111**, 7812 (2007).

[17] S. Izrailev, S. Stepaniants, M. Balsera, Y. Oono, and K. Schulten, Biophys. J. **72**, 1568 (1997).

[18] H. Lee, R. Venable, A. MacKerell, Jr., and R. Pastor, Biophys. J **95**, 1590 (2008).

[19] R. B. Best, X. Zhu, J. Shim, P. E. M. Lopes, J. Mittal, M. Feig, and A. D. MacKerell, Jr., J. Chem. Theory Comput. **8**, 3257 (2012).

[20] K. Vanommeslaeghe and A. D. MacKerell, Jr., J. Chem. Inf. Model. **52**, 3144 (2012).

[21] G. Rossi, P. Fuchs, J. Barnoud, and L. Monticelli, J. Phys. Chem. B **116**, 14353 (2012).

[22] L. Monticelli, S. K. Kandasamy, X. Periole, R. G. Larson, D. P. Tieleman, and S.-J. Marrink, Journal of Chemical Theory and Computation **4**, 819 (2008), http://pubs.acs.org/doi/pdf/10.1021/ct700324x.

[23] A. M. Fluitt and J. J. de Pablo, Biophys. J. **109**, 1009 (2015).

OPEN ISSUES

Rough work- and timetable for the planned 12 months duration of the project.