UNIVERSITY OF
LIMERICK
 OLLSCOIL LUIMNIGH

Bernal
 Institute

Predictive modelling of peptide assemblies

Prof Damien Thompson
 Department of Physics
 Bernal Institute
 University of Limerick, Ireland

ACS Spring Meeting
 Philadelphia, PA
 COMP: Drug Design
 26th March 2020

Today I'm going to talk you through work in my group that is seeking new drug targets from peptide assemblies. Ultimately through our collaborative work with the SSPC which is The Irish Research Centre for Pharmaceuticals, we also want to design peptide-based drug formulations but for today I'll stick with targetting naturally occurring assemblies, so that is of course amyloid self-assembly.

1

The self-assembly mechanism of these IDPs can be generally described as a nucleated polymerization, where through primary nucleation a nucleus of certain sized disordered oligomers is formed, which through conformational re-arrangements forms ordered low molecular weight oligomers, and protofibrils through seeded aggregation... these protofibrils undergo polymerization to form the mature fibril, which through further elongation gives rise to the amyloid deposits, so plaques and tangles in the brain.

Nevertheless studies show that these low molecular weight species are the most toxic and damaging species.

Reversible Amyloid Self-assembly?

Also the fibrils might directly fragment or reverse dissociate to refill that pool of toxic low molecular weight species.

IDPs Behave Differently from Folded Proteins

So just to remind you how these bad proteins, these disease causing IDPs, are different from good folded proteins that are functional in body... the video on the left shows that the structured folded proteins are able to retain their shapes as the sticky hydrophobic regions are on the inside, while the natively structureless IDPs on the right generally fluctuate randomly exposing their hydrophobic regions, which leads them to interact easily with other IDPs and aggregate or self-assemble.

These low molecular weight toxic oligomers are thought to form mainly beta sheets. But these are elusive to bulk experimental techniques in water due to their extremely transient nature experiments have applied constraints to covalently cross-link and obtain structures like this one but this does not solve the problem, because you still are not seeing them in the wild.

These low molecular weight beta sheet oligomers self-assemble from a predominantly disordered ensemble consisting of beta hairpin monomers and further give rise of cross-beta fibrils. Recent experimental evidences also suggest that alpha helical intermediates such as a partially folded helical monomer may initiate helical oligomerization through helix-helix associations to form toxic oligomers. Also fully folded helical monomers, were shown to be aggregation impeding however it is extremely challenging to characterize such subtle differences in helicities. Recent experiments also indicate that a helical tetramer of alpha synuclein may prevent toxic aggregation and may remain in dynamic equilibrium with a disordered monomer. Finally a preformed dodecamer of amyloid beta can seed fibril formation in 42 but not 40. Our research focusses on identifying conformational poses that are more prone to aggregation, so we model the helical monomers and oligomers, like these tetramers and dodecamers.

We use predictive modelling based on molecular dynamics simulations coupled with cross correlation networks, which can provide free energy maps for the assembly of these oligomers, and so provide leads for future experiments. As we're all horribly aware there is still no effective treatment. The National Institute on Aging has been funding research since 1974, the amyloid-beta and tau proteins were identified in the mid 1980s and there has been a stream of funding initiatives proposed drug candidates since then. In the past 5 years there has been an increased effort in computational chemistry in the past 5 years. Due largely to development of force fields that can actually handle dispersion interactions as the IDPs shift constantly in water; earlier methods biased towards folded states because of course they were parameterised to model globular not disordered proteins.

Research Questions

1. Could the population of stable α -helical amyloidogenic monomers be optimized to make them more resistant to toxic oligomerization and if so, what are the molecular signatures that govern such stabilities?
2. What causes the partially folded α -helical amyloidogenic monomers to be so aggregation-prone, compared to the helically folded or unfolded states?
3. What are the molecular determinants in the formation of non-aggregating α -synuclein helical tetramers over the aggregation-prone disordered monomers?
4. What are the molecular determinants of the difference in aggregation profiles between amyloid-b40 and amyloid-b42 from the perspective of a preformed, profibrillar oligomer?

So here are our four research questions

I. Optimizing the Helical Stabilities of A β and α S Monomers

Can the helical stabilities of A β and α S monomers be optimized to render them aggregation resistant to helical oligomers?

Bhattacharya, S., Xu, L., & Thompson, D. (2019). **Molecular simulations reveal terminal group mediated stabilization of helical conformers in both amyloid- β 42 and α -synuclein.** *ACS chemical neuroscience*, 10(6), 2830-2842.

What can we learn by studying both; what features if any do they share as IDPs.

For this let's look at the structures of amyloid beta 42 and alpha synuclein

It might seem counterintuitive to refer to these peptides as structures as we know they are intrinsically disordered in vivo but these are solved in a low dielectric organic solvent or in a membrane-bound environment where they present broken helices or helix-kink-helix motifs.

Ab42 is a 42 residue peptide and a-synuclein is 140 residues...the regions encompassing helix 1 and helix 2 are marked along with the loop... the red surface on the structures and on the sequence of both ab42 and a-syn represent the N-terminal, grey... the central hydrophobic domains, which is called central hydrophobic cluster in case of Ab42 and non-amyloid beta component in case of alpha synuclein, and green... the C-terminal ... additionally another region in blue known as the turn region exist for ab42..

Below are shown the residue types according to their hydropathy ... and we can see that amyloid beta is significantly more hydrophobic than alpha synuclein

We use these as starting structures to run long and extensive simulations of

unfolding, and we also try other things like solute-tempered metadynamics to capture more rare events and starting from random coils and watch evolution of transient structures so we sample in both directions, and these tricks can help us build up a very detailed map of the conformational landscapes of the peptides. Here I'm going to keep it simple and focus just on the brute force, following time's arrow, path from folding to unfolding.

Modelling Strategy

- Three degrees of helical stabilities: **fully folded** or *F*, **partially folded** or *P* and **Unfolding** or *U*)
- Differential helix propensities of **10 different** force field/water model combinations
- **36 μ s** of simulations

Folded

Charmm36 with mTIP3P
(C36/mTIP3)

Charmm36 with TIP4P
(C36/TIP4)

Charmm36 with TIP4P/2005
(C36/2005)

Charmm36 with TIP4P-Ew
(C36/Ew)

Partially folded

Charmm36 with TIP4P-D
(C36/D)

Amberff03 with TIP4P-D
(A03/D)

Unfolding

Charmm22* with TIP4P-D
(C22*/D)

Charmm36m with mTIP3P
(C36m/mTIP3)

Charmm36m with sTIP3P
(C36m/sTIP3)

Amberff03W with TIP4P/2005s
(A03Ws/2005)

Beauchamp, K. A., Lin, Y. S., Das, R., & Pande, V. S. (2012). Are protein force fields getting better? A systematic benchmark on 524 diverse NMR measurements. *Journal of chemical theory and computation*, 8(4), 1409-1414.
Piana, S., Lindorff-Larsen, K., & Shaw, D. E. (2011). How robust are protein folding simulations with respect to force field parameterization? *Biophysical Journal*, 100(9), L47-L49.

The modelling strategy that we use here exploits the differences in helix-coil converting propensities through 10 different force field and water model combinations. We generate the folded, partially folded and unfolding states. The force fields suitable for folded proteins generated the helically folded states, and the ones suitable for IDPs generate the partially folded helix and the unfolded states. So we use the heterogeneity of the available modelling parameter sets to scan a wide variety of poses.

Classification of Helical 'Foldedness'

Fraction of Native Contacts

$$Q(X) = \frac{1}{N} \sum_{(i,j)} \frac{1}{1 + \exp[\beta(r_{ij}(X) - \lambda r_{ij}^0)]}$$

1. Best, R. B., Hummer, G., & Eaton, W. A. (2013). Native contacts determine protein folding mechanisms in atomistic simulations. *Proceedings of the National Academy of Sciences*, 110(44), 17874-17879.

N = all pairs of heavy atoms (i, j) , which are in contact if the distance between them is less than 5 Å and separated by at least 3 residues.

$r_{ij}(X)$ = distance between i and j in conformation X .

r_{ij}^0 = distance in the native state (initial helical conformation).

β = smoothening parameter (5 Å⁻¹)

λ = describes fluctuations when the contact is formed.

We classify the helical foldedness according to the time evolution of fraction of native contacts, using the method of Best, Hummer and Eaton. States above Q value of 0.5 were regarded as being folded, states below 0.5 but not touching 0.1 as being partially folded and states touching 0.1 as unfolding.

We can use this to classify the different states as the temporal evolution of helical fraction followed the same trend as fraction of native contacts.

Termini frequently contacts with other regions in the helically folded states

Next we estimated the evolution of timed contact frequencies between heavy atoms within 5Å of residue pairs. We find that for Ab42 we lose contacts between the N-terminal and the central hydrophobic domain as we lose helicity.

Similarly the contact frequencies between C-terminal and the central hydrophobic domain in alpha synuclein is very prominent for the folded states but as we move from folded to partially folded the contact fades and finally disappears in the unfolding state.

In order to investigate the type of contacts, we looked into the interaction maps ... three kinds of interactions are shown here hydrophobic, hydrogen bond and salt bridge.. We see that the main contact leading to stabilizing the helices in folded states of Ab42 is hydrophobic interactions between the flexible N-terminus and the central hydrophobic domain followed by a secondary salt bridge between the terminus and other domains. Again these interactions are not as prominent for the partially folded states and disappears completely for the unfolding states...

For alpha synuclein a similar trend is observed but in this case between C-terminal and hydrophobic domains...

So we can deduce that hydrophobic interactions between the terminal regions and the central hydrophobic domains may help optimize the stability of helices

Terminal deleted peptides show significantly higher helix destabilizing tendency

$G = H - TS$ } where G = free energy, H = enthalpy, T = absolute temperature in K, S = entropy. Conformational entropy (S) was estimated by Normal mode analyses. All energies are normalized per residue.

Case, D. A. (1994). Normal mode analysis of protein dynamics. *Current Opinion in Structural Biology*, 4(2), 285-290.
 Pearlman, D. A., et al. (1995). AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. *Computer Physics Communications*, 92(1-3).

To test this we ran controls where we deleted the N-terminal of amyloid-beta and the C-terminal of alpha-synuclein. We found that these truncated proteins lost helicity much faster than their full length counterparts. Here are commonly formed conformational states. And here are the estimated energies, normalised per residue as we are comparing different sized proteins here.

For both the proteins the terminal deletion doesn't change the entropy, but gives far weaker enthalpy and so big free energy penalty, which supports the predicted helix stabilizing effect of the tails on the central hydrophobic domains.

II. Regulation of Partially Folded Helical States of Monomers

What are the molecular signatures in A β and α S monomers that causes the partially folded helical monomers to be so aggregation-prone?

Bhattacharya, S., Xu, L., & Thompson, D. (2020). **Long-range Regulation of Partially Folded Amyloidogenic Peptides.** *Scientific Reports* (Accepted).

In the previous work we determined the molecular determinants of aggregation-impeding helically folded states...here we delve deeper to try and identify the reason behind the partially folded helical states being so aggregation-prone, compared to the fully folded helical states or the unfolded states.

Modelling strategy

- MD simulations of A β 42 and α S helical monomer conformations from a subset (**four**) of physical models used in the previous study.

Folded

Charmm36 with mTIP3P
(C36/mTIP3)

Charmm36 with TIP4P
(C36/TIP4)

Partially folded and Unfolded

Charmm22* with TIP4P-D
(C22*/D)

Amber ff03W with TIP4P/2005s
(A03Ws/2005)

- Dynamic cross-correlation (DCC) network analyses:** map the intra-molecular coupled atomic fluctuations from MD data – C α -C α Pearson inner product cross-correlation coefficient is given by:

$$c_{ij} = \frac{\text{Tr}([\sigma_{ij}]^3)}{(\text{Tr}([\sigma_{ii}]^3))^{1/2} \cdot (\text{Tr}([\sigma_{jj}]^3))^{1/2}}$$

$\left. \begin{array}{l} [\sigma_{ij}]^3 = 3 \times 3 \text{ cross sub-matrix of the } 3 \times 3 \times 3 \text{ covariance} \\ \text{matrix for fluctuations between residues } i \text{ and } j \\ [\sigma_{ii}]^3 = 3 \times 3 \text{ diagonal sub-matrix for residue } i \\ [\sigma_{jj}]^3 = 3 \times 3 \text{ diagonal sub-matrix for residue } j \\ \text{Tr}() = \text{trace of the matrix} \end{array} \right\}$

Yao, X. Q., Malik, R. U., Griggs, N. W., Sijpe, L., Traynor, J. R., Sivaramakrishnan, S., & Grant, B. J. (2016). Dynamic coupling and allosteric networks in the α subunit of heterotrimeric G proteins. *Journal of Biological Chemistry*, 291(9), 4742-4753.
Grant, B. J., Rodrigues, A. P., Elisei, K. M., McCammon, J. A., & Cases, L. S. (2006). Bio3d: an R package for the comparative analysis of protein structures. *Bioinformatics*, 22(21), 2695-2696.

For understanding the potential aggregation-prone nature of partially folded helical states, we selected four subsets of force field and water model combinations, such that two of them generates folded models and the other two gives partially folded and unfolded models.

We used a method called dynamic cross-correlation network analyses that calculates correlated atomic fluctuations as cross-correlation coefficients defined by the trace of 3X3 covariance matrix of fluctuations between two different atoms over covariance matrix of fluctuations between the same atoms.

An example of the Consensus Network Analyses for the Partially Folded Helical A β 42

Next, the network edges are constructed based on the minimum Ca-Ca cross-correlation (c_{ij}) value between all residues. I'm showing here data obtained using a threshold of 0.4. We generated a stack of these maps and this one seemed to give the best signal to noise ratio but it is a somewhat arbitrary choice and certainly I would not expect that this 0.4 will not work automatically for all IDPs; we're currently testing that.

Partially Folded Helical states show Dynamic Coupling between Central Hydrophobic Domains and the Charged termini

We calculated about 500 different paths of disorder propagation. The propagation of disorder from the partially folded helical states

Helically Folded and Unfolded States do not Capture these Long-range Dynamic Coupling

18

Long-range (within 20 Å) Regulation of the Partially Folded Helical States

We calculated about 500 different paths of disorder propagation. The propagation of disorder from the partially folded helical states

Summary of Findings

- A common intra-peptide route to *folded helical stabilization* is found through **short-range interactions** between the flexible termini and the central hydrophobic domains in both A β 42 and α S monomers.
- **Long-range regulations** of the central hydrophobic domains by the charged termini may play a dominant role in promoting initial helical oligomerization of *partially folded helical states* of A β 42 and α S monomers.
- A common intra-peptide route to optimization of helical stability was determined by fine tuning the distances between charged terminal residues and central hydrophobic domains – this computational design strategy could be applied to other amyloidogenic peptides which present helical intermediates during fibril formation.
- Targeting the charged terminal residues in neurotoxic protein monomers with small molecules that facilitate close interactions with the central hydrophobes could be an important strategy to impede toxic oligomerization in these peptides.

III. Design and Characterization of helical α -synuclein tetramers

α -synuclein native structure : Monomer or tetramer?

Targeting α S tetramer provides a potential way for the development of treatment for PD.

Currently, we do not know exactly how α S monomers assemble into tetramers because of the lack of details of tetramer structures. Hence, we rationally designed thermodynamically stable α S tetramer and related multimers and validated these structural models.

We assume that protein-protein interactions should maximize the hydrophobic interacting surface, which is also one of dominant driving forces for protein assembly. And we found that current α S tetramer models do not satisfy this criteria, and we re-designed a new α S tetramers with optimal contact of their aggregation-prone region, that is their hydrophobic NAC region, by introducing symmetric operations.

The designed oligomer each has a symmetry in their hydrophobic region (residues 61-95) shown in Cartoon.

Free energy landscape of assembly of α -synuclein: Thermodynamic stability

The stability of helical α S
oligomers linearly increases
with the size of oligomers

The stability was evaluated by the conformational energy of each oligomer, which includes the solvation effects.

Free energy landscape of assembly of α -synuclein: Binding affinity

The binding affinity of helical α S oligomers linearly increases with the size of oligomers, suggesting that the formation of large oligomers is feasible.

$$BE = E(\text{complex}) - \sum_{i=1}^n E_i(\text{monomer})$$

The binding energy (BE) for assembly of α S oligomers is calculated according to the formula:

$BE = E(\text{complex}) - \sum_{i=1}^n E_i(\text{monomer})$ where $E(\text{complex})$ is the conformational energy of the multimer, and $E_i(\text{monomer})$ is the conformational energy of the monomer i within the multimer. The binding affinity increases linearly from dimer to octamer, suggesting that it is favourable to form α S oligomers.

We then estimated the activation energy (AE) that reflects the average energy barrier for assembly;

where $\langle \dots \rangle$ represents ensemble average, $E_i(\text{monomer})$ is the conformational energy of the monomer i within the multimer, and $E^0(\text{monomer})$ is the conformational energy of the monomer sampled in the free state.

The helical tetramer has the smallest activation energy among all the multimers investigated, which provides an explanation for why the tetramers are the most populated species observed experimentally.

Summary: Why α S tetramers?

Liang Xu, Shayon Bhattacharya and Damien Thompson. Chem. Commun., 2018, 54, 8080-8083.

Liang Xu, Shayon Bhattacharya and Damien Thompson. Phys.Chem.Chem.Phys., 2019, 21, 12036-12043.

In summary, we predicted that the assembly of helical α S oligomers is thermodynamically favorable, but only kinetically most favorable for tetramers, explaining the ubiquity of α S tetramers identified experimentally.

The orientations A–C are excluded according to the model selection criteria. N and C indicate the N- and C-terminal of α S tetramers with each chain coloured differently.

- (1) It is well-known that the N-terminal of α S interacts with membrane, so we exclude orientations with the C-terminal interacting with membrane, model A;
- (2) The modelled α S tetramers are homogeneous with an initial C4 symmetry, which further precludes many possible orientations. Models B and C are the same.

Such weak interactions of α S with membrane suggest that α S may tend to leave the membrane and stay relatively stable in solution. The present results here provide support the hypothesis of a dynamic equilibrium of α S between cytosol and membrane: α S helical tetramer assembled on membrane surface could release from membrane and resist aggregation in the cytosol. Taken together, our results imply that helical α S tetramer could be a potential target for the novel development of treatment for PD.

III. Molecular Signatures of the Differences in Aggregation Profile between A β 40 and A β 42: Comparison with AFM Experiments

What are the molecular determinants of the difference in aggregation profiles between amyloid- β 40 and amyloid- β 42 oligomers as seen from AFM experiments at the graphene-water interface?

Nirmalraj, P., N., List, J., Battacharya, S., Howe, G., Xu, L., Thompson, D., & Mayer, M. (2020). **Complete Aggregation Pathway of Amyloid- β (1-40) and (1-42) Resolved on an Atomically Clean Interface.** *Science Advances* (Accepted).

The first project that I am going to talk about focusses on identifying molecular signatures to optimize the helical stabilities of amyloid beta 42 and alpha synuclein...

Preferred Orientation of Oligomers (dodecamers or 12-mers) at the Graphene-Water Interface

- Cryo-EM solved structure of Aβ42 fibrils showing the LS-fold (PDB 5OQV) – dodecamers (two adjacent layers of hexamers) constructed for both Aβ42 and Aβ40 in the same fibrillar fold – models for profibrillar oligomers in self-assembly pathway of Aβ42.
- Starting structure aligned in three orientations – I (fibril axis perpendicular to graphene sheet), II (axis parallel to graphene), and III (axis parallel to graphene with the two layers adjacent to each other)
- MD in three orientations for 100 ns – **orientation III** most favourable according to binding energies between dodecamers and graphene – these potentially fibril seeding structures were sampled for another 500 ns

A β 42 Dodecamers are More Stable than A β 40: Comparison to AFM

- AFM: Oligomers and fibrils are very stable and densely packed for A β 42 but not for A β 40
- Conformational energy shows dodecamers of A β 42 are more energetically favourable and stable than A β 40, consistent with previous aggregation studies showing A β 42 dodecamers as prevalent oligomers

Account for the stability of A β 42 over A β 40 due to the two extra C-terminal residues: Intra-chain salt-bridge of Lys28-Ala42 (absent in A β 40) with sidechain NH₃⁺ in Lys28 and terminal COO⁻ in Ala42.

Economou, Nicholas J., et al. Amyloid β -protein assembly and Alzheimer's disease: dodecamers of A β 42, but not of A β 40, seed fibril formation. *JACS* 138.6 (2016): 1772-1775.
 Bernstein, S.L., Dupuis, N.F., et al. 2009. Amyloid- β protein oligomerization and the importance of tetramers and dodecamers in the aetiology of Alzheimer's disease. *Nature Chemistry*, 1(4), p.326.
 Xiao, Y., Ma, B., McElheng, D., Parthasarathy, S., Long, F., Hoshi, M., Nussinov, R. and Ishii, Y., 2015. A β (1-42) fibril structure illuminates self-recognition and replication of amyloid in Alzheimer's disease. *Nature Structural & Molecular Biology*, 22(6), pp. 499-505.

Maximum heights on surface of graphene – correlation with AFM experiments

- The computed size distribution (maximum dodecamer heights on top of graphene) from MD simulations shows a range of ~7–11 nm for Aβ42 with the maximum height peaking at ~8 nm – commensurate with the statistical distribution of oligomer diameter with a mean particle height of ~8.4 nm as recorded from high-resolution AFM experiments.
- For Aβ40, the most populated species has a diameter ~4.1 nm from AFM, way smaller than the predicted height of the 12-mer (~7 nm).

In-register Hydrogen Bonds Identify Unidirectional Growth for A β 40, but Bidirectional Growth for A β 42 Dodecamers

Sum of H-bonds in two hexamer layers: Number of in-register H-bonds of the dimer at one end (E1) is significantly less than that the dimer at the other end (E2) for both A β 40 and A β 42 – potential to undergo unidirectional growth?

Occupancy	E1 (%)			C (%)			E2 (%)		
	40-60	60-80	≥80	40-60	60-80	≥80	40-60	60-80	≥80
A β 40	2	4	11	3	12	32	1	9	30
A β 42	6	2	15	5	3	41	6	6	26

Hydrogen bond occupancies: A difference of 11% for the strongest (≥80%) H-bonds between E2 and E1 for A β 40 – a significantly larger difference of 19% for A β 40 – implications for A β 40 dodecamers being more open (almost twice more than for A β 42) on the fluctuating end – monomer addition on E1 end may proceed in a more ordered fashion for A β 42 than for A β 40 – potential for *unidirectional* growth for A β 40 and *bidirectional* growth for A β 42.

Summary of Findings

- We can predict the molecular shape of oligomers that persist in the AFM studies.

Acknowledgements

Group alumni

- ❖ Shane O'Mahony
- ❖ Jose Luis Dominguez
- ❖ Greg Gannon
- ❖ Adam Orlowski
- ❖ Raghav Singh
- ❖ Melissa Gunnoo
- ❖ Liang Xu
- ❖ Grzegorz Nawrocki
- ❖ Geoffrey Howe

Current group

- ❖ Pierre-Andre Cazade
- ❖ Sarah Guerin
- ❖ Shayon Bhattacharya
- ❖ Oguzhan Maraba
- ❖ Shubham Vishnoi

Collaborators

- ❖ Chris Nijhuis (NU Singapore)
- ❖ Peter Nirmalraj (EMPA, formerly IBM-Zurich)
- ❖ Marek Cieplak (Polish Academy of Science)
- ❖ Pascal Jonkheijm and Dorothee Wasserberg (U Twente)
- ❖ Ed Bayer (Weizmann)
- ❖ Caitriona O'Driscoll (UC Cork)
- ❖ Tom Simonson (Ecole Polytechnique France)
- ❖ Philip Moriarty (U Nottingham)
- ❖ Tofail Syed (U Limerick)

Science
Foundation
Ireland
For what's next

European
Commission

Horizon 2020
European Union funding
for Research & Innovation

Future MD simulations of multi-oligomer assembly could test our hypothesis by predicting the 3D structure of the extended protein films formed at the interface. Our study on the role of the liquid and imaging environment on the accurate determination of amyloid aggregate size and morphology demonstrates that the correct balance of measurement precision and biological relevance is required for the characterisation of amyloid aggregates implicated in the pathology of neurodegenerative diseases. The findings can also trigger future studies for experimentally quantifying the beta-sheet content of the amyloid aggregates and further mapping those sticky hot spots as future targets for drug design.