Determining an Optimal Reaction Co-Ordinate for the Alzheimer's Aβ-42 Peptide

Submitted in part fulfilment of the requirements of the degree of

Bachelor of science in

Biotechnology with Enterprise

Faculty of Biological Sciences Undergraduate School

University of Leeds

Leeds

LS2 9JT

Declaration of Academic Integrity

Determining an Optimal Reaction Co-Ordinate for the Alzheimer's Aβ-42 Peptide

Submitted in accordance with the requirements for the degree of

Bachelor of Science

University of Leeds

Faculty of Biological Sciences Undergraduate School

The candidate confirms that the work is submitted in accordance with the Declaration of Academic Integrity signed by the candidate at the start of the academic year.

Signed: 

Abstract

The study and modelling of intrinsically disordered proteins (IDPs) is challenging. It is not possible to define these proteins by a single structure, they are functional proteins which present high conformational and structural flexibility. Instead they can be defined in terms of multiple structural ensembles and their corresponding free energy landscapes. Defining these free energy landscapes as a function of a reaction coordinate (RC), or as diffusion along the free energy landscape, gives a good description of the system dynamics. For this description of the dynamics to be accurate, the RC is be selected in a way so that the dynamic properties can be computed exactly. In this paper the committor for a structural ensemble of the Alzheimer's associated Aβ-42 peptide is determined by numerically improving the root mean squared displacement (rmsd), through an iterative approach. The committor is an ideal RC which can compute the dynamic properties between two points along it exactly. In this study one example of such a committor is determined, along with the diffusive dynamics along the RC. The determined committor passes strict validation criteria used to test RC optimality. Five structural ensembles are constructed from points of interest along the committor. These ensembles represent the shift in structure along the coordinate and give a visual guide to the proteins folding.

Acknowledgments

- Thanks to Dr Sergei Krivov for the provision of the notebooks and feedback which made this project possible.
- Thanks to Löhr et al. (2021) for making their structural ensembles and trajectory data open access.

List of Abbreviations

Abbreviation	Meaning
Aβ-42	Amyloid β – 42
CFEP	Cut Based Free Energy Profile
FEL	Free Energy Landscape
FEP	Free Energy Profile
HFEP	Histogram Based Free Energy Profile
IDP	Intrinsically Disordered Protein
MFPT	Mean First Passage Time
MTPT	Mean Transition Path Time
RC	Reaction Coordinate
rmsd	Root Mean Squared Deviation
MetSo	Methionine Oxidised
q	Committor
qn	Natural Committor

Table of Contents

Introduction

Attempting to understand how proteins function at an atomic level experimentally is highly challenging. Viewing protein structures function in real-time is difficult, while experimental processes such as purification and crystallisation can alter the final observed structure (Gupta, Akhtar and Bajpai. 2014). An attractive solution for studying complex biomolecules is to run them in computer-based simulations (Hollingsworth and Dror. 2018). Simulating complex proteins *in silico* grants a wider range of ways in which the protein can be studied, examples include ligand binding, temperature shifts and changes in water model and ion content. The availability of trajectory data obtained from molecular dynamics simulations is increasing, alongside more user-friendly software with which to produce them (Lindorf-Larsen et al., 2011). It is, however, how the data is interpreted and analysed which present challenges, both the effective computation of complex molecular mechanisms and the vast data these simulations produce require relative experts in their respective fields to understand and utilise them effectively (Jung, Covino and Hummer. 2019).

This study aimed to model and describe the kinetics of the Aβ-42 peptide, a member of the Amyloid β family and a major Alzheimer's associated protein. Alzheimer's occurs due to the incorrect processing of the amyloid precursor protein by α and γ-secretases (Murphy and LeVine. 2010). This incorrect processing results in the formation of amyloid-β peptides, which in turn aggregate to form toxic oligomers, and subsequently senile plaques (Sengupta, Nilson and Kayed. 2016).

Aβ-42 is classified as an intrinsically disordered protein (IDP). IDPs make-up approximately 1/3rd of our proteome and cannot be classified in the same way as conformational proteins (Lohr et al., 2021). IDPs do not have a specified 3D structure that one would expect to find at the minima of the "folding funnel", instead they are often represented by many folding funnels with multiple free-energy minima across a shallow landscape. IDPs confound multiple rules defining protein behaviour, making them difficult to study and define in terms of popular working theories which fall under the "structure-function paradigm" (Uversky. 2019). In addition, they are far more challenging to study experimentally than "ordered" proteins, commonly used techniques such as small-angle X-ray scattering and nuclear magnetic resonance (NMR) give structural and dimensional data but give an overall incomplete picture of dynamics. IDPs instead of being defined by one native ensemble can be defined by many ensembles, which in turn brings the challenge of exhaustively mapping all functionally relevant forms using experimental data (Bhattacharya and Lin. 2019). Additionally, the unique metastable states are often impossible to define by changes in 3D structure, as

regular proteins are, additionally the transitions between these states are often rapid (Löhr et al., 2021).

For IDPs the determination of the rates of transition between them and the kinetics of the transition are of equal importance as the determination of 3D structure. One common challenge in the simulation of protein folding and kinetics is achieving an experimentally relevant timescale while also being able to analyse the large quantity of produced data effectively. A commonplace method for dealing with this is the use of Markov State Models (MSM) (Pande, Beauchamp and Bowman. 2010). This approach allows the simulation of a large number of short trajectories and overcomes the sampling problems in biomolecular simulations, such as an inability to access the whole configuration space (Zuckerman. 2011). There are however limitations to the MSM approach, such as the assumption of sufficient sampling. Hence, we took an alternative approach.

Alternative to an MSM based approach, which uses a large number of short trajectories, the dynamics of the system have instead been described as a function of a reaction coordinate (RC). Defining the kinetics of a peptide by its diffusion along a free energy landscape (FEL) or as a function of a reaction coordinate is convenient. However, for the description of the dynamics to be accurate, the RC has to be chosen in a way that ensures optimality (Banushkina and Krivov. 2016).

Conventional RCs such as the root mean squared deviation (rmsd) or fraction of native contacts, 'Q', provide an informative description of the reaction kinetics (Best, Hummer and Eaton. 2013). However, in the case of IDPs, the rmsd and fraction of free contacts are insufficient to accurately define diffusion along the FEP (Krivov. 2021). They are likely to be suboptimal as RCs, so the predictions made about the dynamics will be inaccurate. These inaccuracies can include an oversimplified free energy landscape where the free energy barriers are often smaller which in turn results in accelerated estimated kinetics for the diffusive model (Banushkina and Krivov. 2016).

One optimal form of RC is the committor. The committor can be used to compute the dynamics of a system between any two points along it exactly. The projected dynamic properties include the number of transitions between the metastable states (NAB), the equilibrium flux, the mean first passage time (mfpt) and mean transition path time (mtpt) (Krivov. 2018). Both the mtpt and mfpt are important dynamic quantities, the mfpt is the average time for a selected stochastic process to occur from an initial state to a separate secondary state, this description of the dynamics can be used for a range of biochemical analysis such as in protein folding and charge hopping (Polizzi, Therien and Beratan. 2016). The transition path is the cross-over of the molecule when a free energy barrier between

stable/semi-stable conformations of a model is passed and the conformation of the protein shifts (Chung and Eaton. 2018).

This study aims to solve the committor using trajectory and structural data obtained from Lohr et al. (2021). The 5,119 trajectories obtained contained a total of 1,259,172 frames for the regular (reduced) model. For the methionine-oxidised (MetSo) model of Aβ-42, a total of 3,071 trajectories with 1,268,139 total frames were obtained. The methods used in this paper involved the construction of a neural network in the VAMPNet approach and the subsequent development of an MSM. These approaches firstly require extensive knowledge of the target system, secondly, they require large amounts of computing capacity, finally, they also require a great degree of experience in protein modelling. The approach taken in this study reduces the computational overhead required by a significant amount.

In this study, the RC was first described as the rmsd between all heavy atoms. The rmsd was determined using the individual structural ensembles and the combined molecular dynamics trajectories obtained from Lohr et al. (2020). In practice computing the committor should be done using a long equilibrium trajectory rather than from the diffusion coefficient (Krivov. 2018). To achieve optimality and to compute the dynamic properties between stable states exactly, the projected rmsds were optimised in an adaptive non-parametric approach described in Krivov. 2021. The approach numerically optimises the rmsd as an RC by iteratively taking the lowest value for the root mean squared displacement across many iterations. The final optimised committor accurately describes the dynamics of a structural ensemble of Aβ-42 to an acceptable degree of uncertainty.

Methods

The project was carried out using a Microsoft Surface Pro 6, with 8 Intel i7-8650U CPU cores, 16 GB of RAM and 500GB of diskspace. The concatenation of trajectories and rmsd analysis was performed using Wordom (Seeber et al,. 2011). The determination of FEPs, dynamic properties, iterative determination of the committor and its FEP was performed using Jupyter Notebook and Google Colab.

Plotting Trajectory Along the Reaction Coordinate Through the Wordom Command Line

The initial analysis was performed using trajectory data and structure files kindly made available by Löhr et al. (2021). Seven pdb structure files were analysed, containing different conformations of the Aβ-42 peptide, two of these structures were methionine oxidated (MetSo) so as to give context to the findings and show the approach to be transferable. The other five will be referred to as reduced to clearly distinguish the two sets. Trajectories derived from the pdb structures were obtained in xtc format from Löhr et al. (2021), and subsequently converted to dcd format through the Wordom command line. The dcd files for the trajectories were combined into two concatenated dcd files for analysis: a large oxidated trajectory and large reduced trajectory. The combined trajectory files were used for root mean squared deviation (rmsd) analysis for the seven structures, the two MetSo structures (MetSo 3-0 an 4-0) and the five reduced structures (Structures 2-0 to 6-0). The command used for rmsd analysis was:

“wordom -ia rmsd --TITLE rmsd1 --SELE "/*/*/*" -imol file.pdb -itrj file.dcd > rmsd”

Each structure file was used as a reference, to produce a corresponding rmsd. Rmsds are an intuitive RC, which can be further used to determine the cut based free energy profiles (Krivov, 2013).

Plotting the Histogram Based Free Energy Profile

Following the generation of the rmsds, they were used to produce a histogram based free energy (HFEP) profile. The HFEP can be used to determine the position of observable free energy minima on the free energy landscape. The FEP here is described as:

$$\frac{F(x)}{kT} = -lnZ_H(x)$$

kT is Boltzman's constant times the system temperature (278K), F(x) is the free energy at a given point along the RC while $Z_H(x)$ is the partition function of the HFEP. Once observed, the free energy minima serve as approximations of metastable states that the protein takes on, they can define boundary states (Krivov. 2018).

Determining the Cut Based Free Energy Profile and Diffusion Coefficient

The cut based free energy profile was computed directly from the trajectory, this cut based free energy profile was then used to determine the diffusion coefficient. The below equations describe this:

$$Z_{C,0}(x, \Delta t) = Z_H(x)\sqrt{D(x)\Delta t/\pi}$$
$$Z_{C,1}(x, \Delta t) = Z_H(x)D(x)\Delta t$$
$$Z_{C,-1}(x, \Delta t) = Z_H(x)/2$$

These are each computed using the Comp_Zca from the cfeplib library, a library file in the CFEP notebook series (Krivov. 2020). The Z_c profiles can then be used to determine the free energy profile (F_H/kT) as a function of the reaction coordinate:

$$\frac{F_H}{kT} = -lnZ_H(x) = -\ln 2\, Z_{C,-1}(x)$$

While the diffusion coefficient is determined as follows:

$$D(x) = \frac{Z_{c,-1}(x)}{2\Delta t\, Z_{C,-1}(x)}$$

The comp_Zca function, with parameters a=-1, was used to generate the FEP as a function of the rmsd. From this profile, the free energy minima and the barrier separating them can be observed.

Determining the Equilibrium Flux, Mean First Passage Time and Mean Transition Path Time

The FEP along with the Diffusion Coefficient, D(x), describe the diffusive model of the dynamics of a given system. The definition of the diffusive model describes the dynamics exactly when projected onto the committor, the optimal RC. The HFEP is computed in the following way:

$$F_H(r) = -kTlnZ_H(r)$$
$$Z_H(r) = N_r/\Delta r$$

Using the Z_H one can plot the HFEP and observe the free energy minima along the RC.

$$Z_{c,1}(q, \Delta t) = N_{AB}$$
$$D(q) = \frac{N_{AB}}{\Delta t Z_H(q)}$$

The partition function for the CFEP ($Z_{C,1}$) and the partition function for the HFEP (Z_H) were computed. $Z_{c,1}$ and Z_H can be used to compute the equilibrium flux, the diffusion coefficient D(x), mfpt and mtpt. The D(x) is not denoted as D(q) until the RC passes the validation tests for committor approximation described below. D(x) and the FEP alone are not sufficient to determine the committor for complex systems (Krivov. 2018).

Dynamic properties can be computed exactly with the use of the committor function. The committor is the probability for a trajectory to reach one boundary state (defined by the free energy minima from the histogram plot) from a given position (Banushkina & Krivov. 2016). The committor is defined as an optimal RC, or *pfold* in the study of protein folding dynamics.

Applying the Committor Criterion

The RC is used to describe the transition between the two free energy minima along it. In order to check if the RC can indeed be defined as the committor (q) or 'optimal RC', an optimality criterion needs to be applied. The criterion determined the RC's proximity to the committor. Transition path summation was used to determine the optimality. For this the functions comp_ekn_tp and comp_Zca_ekn available from the CFEP notebook series. These functions transform the rmsd into the committor (rmsd→q(rmsd)), comp_ekn_tp computes

the total number of transitions between A and B by decomposing the trajectories on transition paths and determining the equilibrium network. Comp_Zca_ekn determines the cut energy profiles using the equilibrium kinetic network as determined by comp_ekn_tp (Krivov. 2020).

Once the trajectory projected onto the committor, the value for $Z_{c,1}$ should be near constant, bar some fluctuation bar an expected breakdown at the borders of the diffusive approximation. An optimal RC or committor will accurately fit into $Z_{c,1}(q, \Delta t) = N_{AB.}$ This can be used as a committor validation test following non-parametric optimisation of the RC. The initially determined RC is however unlikely to be the committor, as Aβ-42 is an IDP the rmsd and fraction of native contacts may serve poorly as RCs (Krivov. 2020). Therefore, if the initial putative RC fails the validation criterion, it will need to be numerically optimised by the method described below.

Non-Parametric Determination of the Committor from Non-Equilibrium Sampling

The final aim for optimisation of the RC (r) is for it to approximate the committor as closely as is possible. Once $\Delta r^2/2 \approx N_{AB,}$ the putative RC ≈ q, the RC timeseries is roughly equivalent to the committor time series; therefore $r(i\Delta t_o) \approx q(i\Delta t_o)$. The nonparametric optimisation process is as follows:

A single long trajectory is taken. This long trajectory ($X(i\Delta t_o)$ matches frames to their respective trajectory numbers using the structural dcd files as a reference. The optimisation equation in terms of the putative RC is:

$$\min r \sum_t [r(t + \Delta t) - r(t)]^2 \; It(t)$$

$$r(t) = 0, X(t) \in A;\; r(t) = 1, X(t) \in B$$

Putative RC timeseries r(t) is 0 at boundary A and 1 at boundary B; simplified this means r(A) = 0 and r(B) = 1 are defined as optimisation constraints for the RC. Or more explicitly q=0 for when $x \in \partial A$ and q=1 when $x \in \partial B$ where A is the first boundary of interest and B is the second (Krivov. 2018)

$I_t(t)$ is defined as an indicator function in place of $n(i|j,\Delta t_0)$ – the number of transitions from state i to j (a->b) in a given time. $I_t(t) = 1$ when itraj(t) = itraj(t + Δt_0), when this condition is not met $I_t(t) = 0$ and any terms from other trajectories are ignored (Krivov. 2021).

Optimal coefficients denoted as α*. $\alpha^* = \arg min_\alpha I(r + \delta r)$. They update the target RC so that r(t) ← r(t) + δr*(t), growing the RC towards the committor as δr*(t) is the optimal variation. To achieve the optimal variation the system is iterated until the above parameters for committor are achieved. The optimal coefficients, in a single variation, give the closest approximation to the committor in that single variation.
The α* function is used as an alternative to a system using Steepest Decent Algorithm, an algorithm used to minimise the functional to its lowest possible value for every iteration.
The optimal value of α (α*) is found through the following equations (Krivov. 2021):

$$\sum_j A_{kj}\,\alpha_j^* = b_k$$

$$A_{kj} = \sum_t f_k(t) f_j(t) I_b(t) I_t(t)$$

$$b_k = \sum_t [r(t+\Delta t) - r(t)] f_k(t) I_b(t) I_t(t)$$

The iterations are recorded, alongside $|r_{new}\text{-}r_{old}|$ to measure the magnitude of change between a number of iterations. Where n = 1000, as results are recorded every 1000 iterations, the equation is:

$$\left|\left|r - r_{-1000}\right|\right| = \sqrt{\sum_t [r(t) - r_{-1000}(t)]^2}$$

The value of recording $\left|\left|r - r_{-1000}\right|\right|$ is that one can observe when the changes between iterations is negligible in relation to the magnitude of r*(t) and can therefore define it as q(t), the committor.

The committor criterion that was used to check the initial RCs optimality should then be applied to the committor as $Z_{c,1}$ will be constant across the entirety of the sampling interval (across all Δt).

Adapting the Committor as the Natural Committor

The diffusion coefficient D(q) alternates along the RC. So, manipulating the coordinate will give a more detailed and accurate picture of the FEP. The committor is altered to produce the natural coordinate (**qn**), where D(qn) is constant at 1. The resultant change is denoted by the equation in Krivov. (2018):

$$\frac{dqn}{dq} = D(q)$$

The determined FEP as a function of the natural committor will closely approximate the FEP as a function of the natural Eigenvector, $F(u\sim_1) \approx F(qn)$, another optimal RC which can be used to exactly compute the dynamics between two points on the RC (Krivov, 2020).

Results

Frame Plots of the rmsd time series

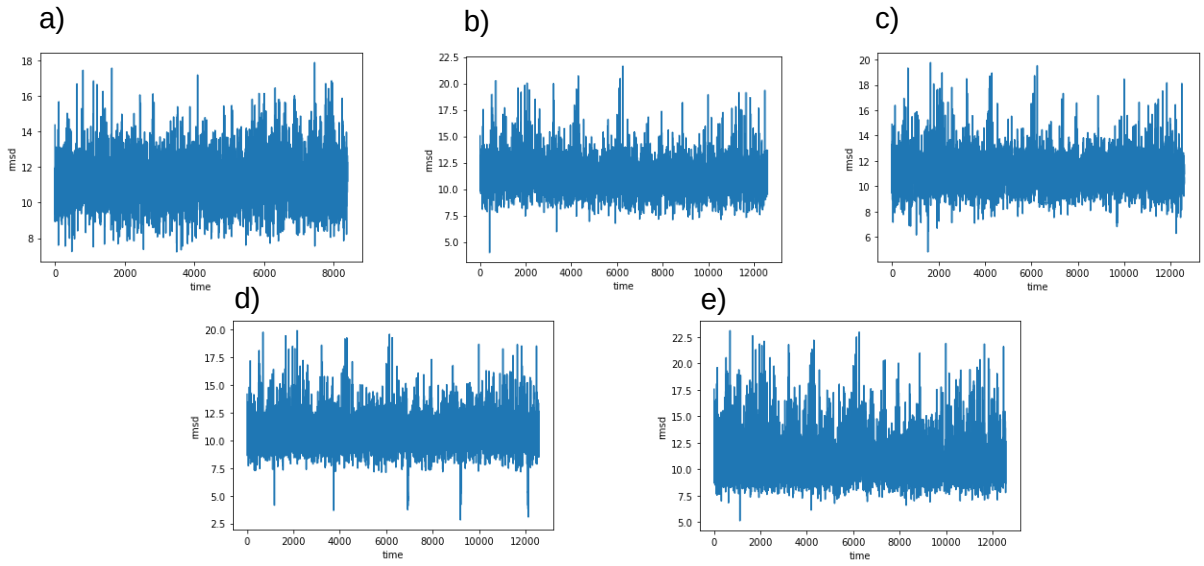


Figure 1: rmsd timeseries from the reduced trajectories: a-e show the rmsd time series and their fluctuations for the reduced trajectories. The plot gives an idea of the RCs behaviour, plotted every 100th frame so the plot is less dense. **a)** is rmsd 2-0, **b)** is rmsd 3-0, **c)** is rmsd 4-0, **d)** is rmsd 5-0 and **e)** is rmsd 6-0. The rmsds were titled so they correspond to the structure number they were derived from.

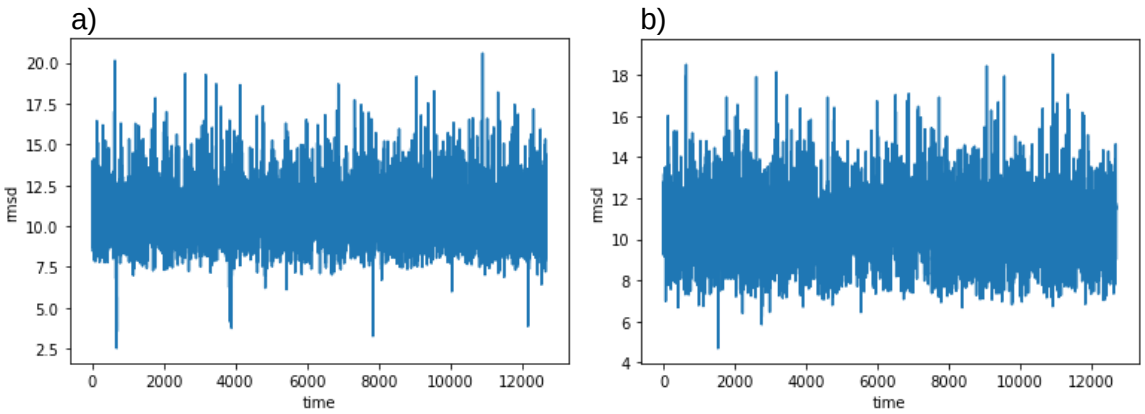


Figure 2: rmsd timeseries from the oxidative trajectories: show the rmsd time series for the rmsds derived from the oxidative trajectory and methionine oxidated (MetSo) structures. **a)** oxidative rmsd derived from MetSo structure 3-0. **b)** Is the oxidative rmsd derived from MetSo structure 4-0.

Figure 1 and 2 show the time series of the computed rmsds for the reduced (figure 1) and oxidative trajectories (figure 2). The time series, plotted for every 100th frame to make the plot less dense, give an indication of the structure transitioning between states. Figure 1.d and figure 2.a showed very clear transitions to a lower energy state, which was of particular interest.

The Free Energy Profiles Along the rmsd

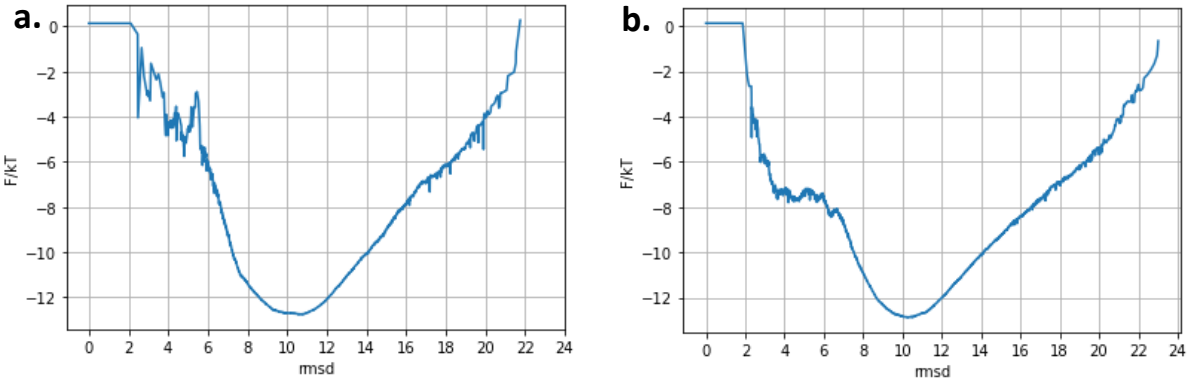


Figure 3: FEPs along the MetSo rmsds: a and b show the free energy landscapes along the oxidated rmsds. These rmsds were computed using the MetSo structural ensembles as a reference and the large combined oxidative trajectory.

Figure 3.a shows the FEP along the rmsd, which was computed using the MetSo rmsd 3-0 as a reference. It showed one small free energy minima at ~5 Å and a large free energy minima at 10-11 Å. The first minima showed a small area and depth within the folding funnel, which indicated the metastable state had poor energetic stabilisation. Between the two funnels there was a clear free energy barrier at ~5.5 Å which was 2 kcal/mol high, indicating some unfolding would be necessary to transition between states.

Figure 3.b, which represents the rmsd of methionine oxidated structure 4-0, showed little to no primary small free energy minima as seen in figure 3.a, instead there is a single minimum with only small energy barriers down the folding funnel. It is not possible to characterise multiple metastable states along this FEP.

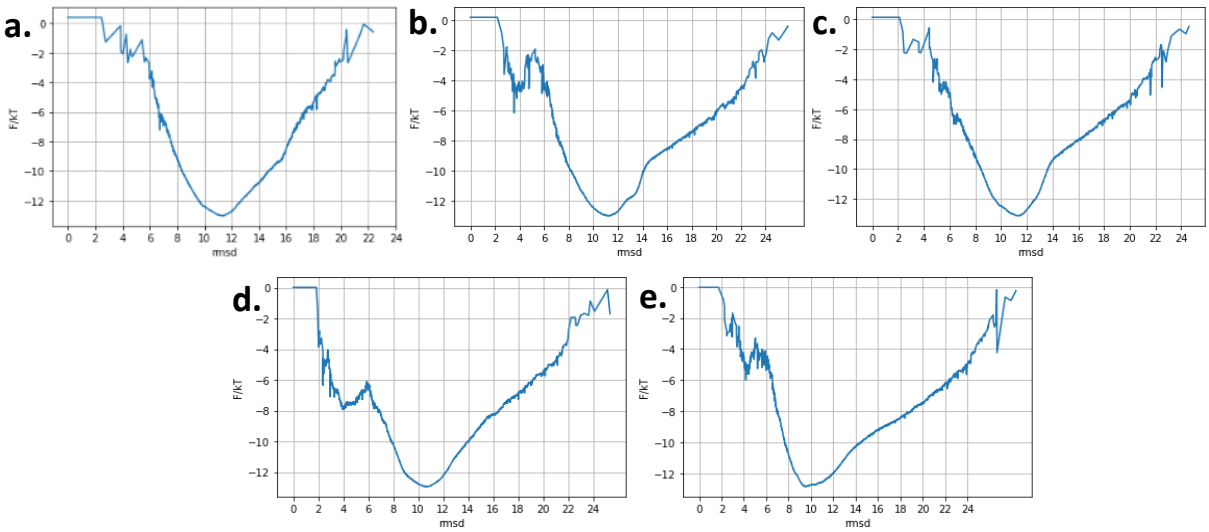


Figure 4: FEPs along the reduced rmsds: **a-e** are free energy landscapes along the reduced rmsds. These rmsds were computed using the reduced structural ensembles as references and the large combined reduced trajectory.

Figure 4 shows the FEPs along the rmsd for the reduced structures and trajectory. Figure 4 a and c show relatively smooth folding funnels, with only one metastable state visible. Figure 4 b, d and e show multiple clear metastable states along the FEP. Figure 4b has a small minima at ~4 Å, followed by a deeper minima at 11 Å. The free energy barrier between these states, at 5 Å, is between 3 and 4 kcal/mol high. Figure 4e shows a similar picture to 4b, it possesses a comparatively small first minima at 4 Å with a smaller barrier between the metastable states. Figure 4d on the other hand has a larger first free energy minima than the other rmsd plots at 4 Å, the larger area within the minima indicates a greater energetic stabilisation than the first minima on other plots.

Testing Optimality Using the Committor Validation Criteria

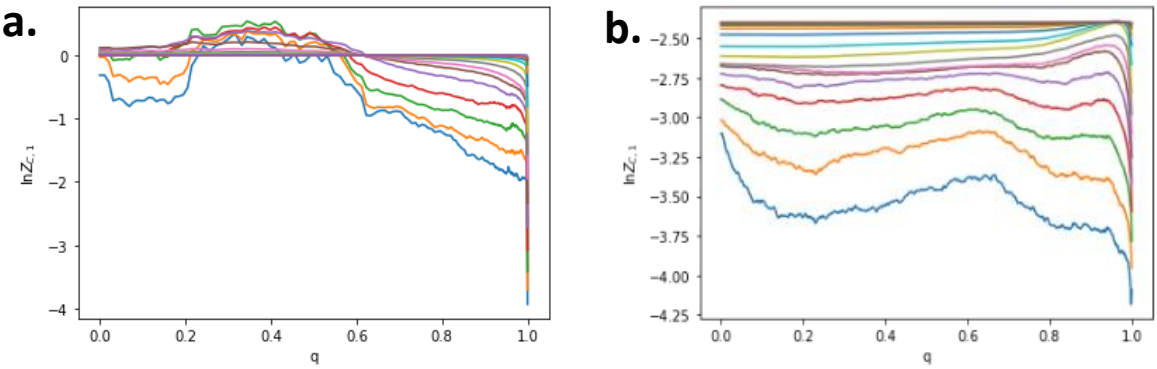


Figure 5: $Z_{c,1}$ Profiles for reduced trajectories: Profiles showing the $Z_{c,1}$ profiles along the RC derived from the reduced trajectories. 5a and b show the $Z_{c,1}$ profiles of the RCs shown in figure 4 b and d respectively. The blue line shows $\Delta t=1$, orange is $\Delta t=2$ going up to 2^{15}.

Figure 5.a and b show the $Z_{c,1}$ profiles at different sampling intervals of Δt. When the chosen RC is optimal, the $Z_{c,1}$ of the cut free energy profile ($F_{c,1}$) should remain roughly consistent, irrespective of the position of its sampling interval. Once the free cut profile can be said to be constant irrespective of Δt, then the putative RC can be considered to approximate the committor (Krivov, 2013). The above profiles indicate the selected RC, the rmsd projected onto the committor (rmsd→q(rmsd)), is suboptimal. Figure 5.b shows that when $\Delta t=1$ (blue line) the profile is not constant. This could be due to a breakdown in the diffusive approximation. The deviation from the committor in 5.b is approximately 4.5, which while not a large deviation, is still significant. Figure 5.a however fluctuates on both sides of 0, the profile deviates at its greatest point to around ~7.4 deviation. This is a significant deviation and shows the coordinate to be far from optimal. The rmsd as an RC for the reduced trajectories and structures was not an ideal coordinate, therefore the RC will have to be improved to approximate the committor.

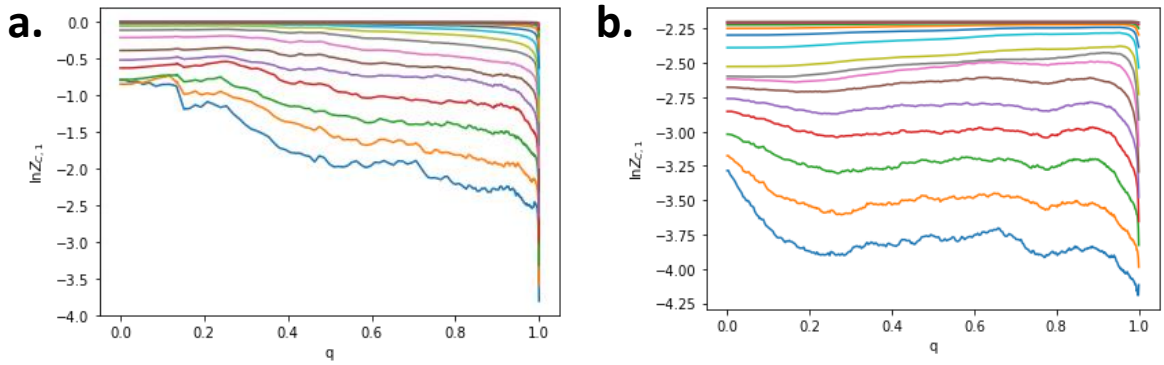


Figure 6: $Z_{c,1}$ Profiles for oxidative trajectories: Profiles showing the $Z_{c,1}$ profile along the RC. a) The $Z_{c,1}$ profile derived from structure oxidated structure 3-0, the rmsd of which is shown in figure 3.a. b) the $Z_{c,1}$ profile of structure 4-0, the rmsd of which is seen in figure 3.b.

The transformed rmsd profiles to the committor as a function of the rmsd (rmsd→q(rmsd)), were also subjected to the committor validation criteria. Figure 6.a shows that the $Z_{c,1}$ profile is not consistent across all of the sampling interval. At the furthest point the profile deviates by exp(2.5), or ~12.2. This profile is very suboptimal. Figure 6.b is the $Z_{c,1}$ profile of the methionine oxidated structure 4-0. It shows a deviation of roughly exp(1.75), or 5.75. This is not a huge deviation, it is however significant, indicating the need for optimisation of the RC in this instance.

Determination of Kinetic Properties

Key properties of the dynamics were then computed directly from the cut free energy profiles ($Z_{c,1}$ profiles) shown above. The properties computed for the cut profiles were the equilibrium flux, mean first path times and the mean transition path times. The dynamics were calculated between the observed free energy minima described in figures 3 and 4.

	Structure 3-0 reduced RC		Structure 5-0 reduced RC	
	Trajectory	Diffusive model	Trajectory	Diffusive model
NAB	1.00	2.82	11.00	43.63
Mfpt-AB	118	84.81	526.36	141.65
Mftp-BA	40999	573634	109830	28718
Mtpt	100.5	47.43	153.91	71.14

Figure 7: Dynamic properties calculated between points along the putative RCs, reduced q(rmsd) 3-0 and reduced q(rmsd) 5-0, for the reduced trajectory.

Figure 7 shows important dynamic properties between two points along the putative RCs. The boundary points (A & B) selected for structure 3-0 (rmsd shown in figure 4b) were 4 and 11. These boundaries are located in the free energy minima, meaning the dynamic properties are those for the transition between metastable sates. Structure 5-0 (rmsd shown in figure 4d) had its boundary states defined as 4 and 11. The number of times Aβ-42

transitioned from state A to B (N_{AB}) was estimated to be 11 when computed directly from the trajectory, the diffusive model gave an estimate four-fold higher. The resultant predictions of the dynamics were much faster in the diffusive model, potentially indicating an oversimplification of the kinetics. Figure 7 for structure 3-0 gives a similar story, where the dynamics differ by a factor of ≤ 3. Taken with the $Z_{c,1}$ profiles in figure 5, the difference between the trajectory dynamics and those based on the diffusive model indicate that the putative RC is definitely suboptimal. The diffusive model of the dynamics is therefore sub-diffusive and cannot be described to be Markovian as the dynamics are projected onto a suboptimal RC (Krivov. 2013).

	Structure 3-0 methionine oxidised RC		Structure 4-0 methionine oxidised RC	
	Trajectory	Diffusive Model	Trajectory	Diffusive model
NAB	1.00	2.82	9.00	54.04
Mfpt-AB	172	84.81	727	117.5
Mftp-BA	152654	450090	134489	23361
Mtpt	42	62.11	147.78	59.51

Figure 8: Dynamic properties calculated between points along the putative oxidative RCs, MetSo q(rmsd) 3-0 and Metso q(rmsd) 4-0, directly from the trajectory and from the diffusive model.

Figure 8 is similar to figure 7, but for the MetSo structure 3-0 and 4-0, the rmsds of which are shown in figure 3 a and b. Similar to figure 7, the results showed the diffusive model dynamics to be many times faster than that of the dynamics computed directly from the trajectory. The MetSo structure 3-0 shows nearly identical dynamics to that of the reduced structure 3-0, indicating the methionine oxidation had not affected the dynamics hugely. For MetSo structure 4-0, the diffusive model has an N_{AB} 6-fold larger. The predicted dynamics were also faster. Taken with the $Z_{c,1}$ profile shown in figure 4.b, the diffusive model was apparently sub-diffusive as it was projected onto a suboptimal RC.

FEP Along the Committor and validation by the Committor criterion

The initial $Z_{c,1}$ profiles and computed dynamics between minima indicate the RC to be suboptimal, predicting an oversimplified FEP and faster dynamics by a significant factor. The RC was optimised through a non-parametric determination approach described in (Krivov. 2018 and Krivov. 2021).

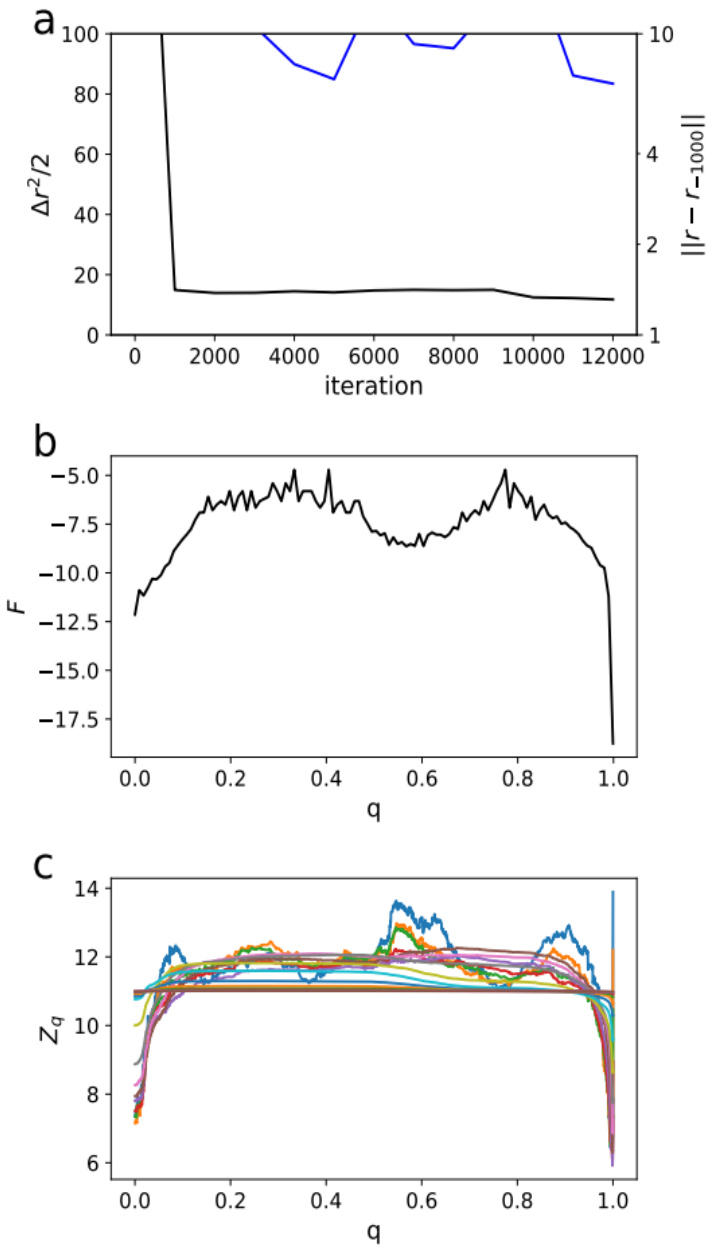


Figure 9: a) Value of $\Delta r^2/2$ per every 1000 iterations, the right-hand x axis shows the increment size between every 1000 iterations $||r\text{-}r_{\text{-}1000}||$. **b)** The FEP as a function of the newly determined committor. **c)** The Z_q profile along the committor in non-logarithmic form for multiple time series.

The approach for determining the committor in an equilibrium state gave the results seen in Figure 9. Figure 9.a shows the minimum value of the root mean squared for every 1000 iterations. It also shows the increment size between iterations. The figure indicated that past 1000 iterations the $\Delta r^2/2$ drops dramatically, then stabilises. The $||r-r_{-1000}||$ value was seen to decrease, whilst its decline was not constant the increment size did decrease.

Figure 9.b shows the FEP along the committor, the profile indicated a more complex landscape than the rmsd profile did. However, this profile shows boundary effects close to point q=1. New details emerge and the transition state is made clear, with two peaks which show more noise than the rmsd profile indicated. While the committor gave a very good image of the FEP, the accuracy of the image is limited by a high dx (a determinant of sampling size), meaning we have lost some resolution (Krivov. 2018). To obtain a more detailed profile, the committor was transformed into the natural committor (qn). The FEL along the natural committor can be seen in **figure 10.b**.

Figure 9.c is the Z_q profile of the putative committor, it shows very stable levels across the coordinate in comparison to the equivalent $Z_{c,1}$ profiles shown in figures 3 and 4. This stability lasts up to the boundaries, where some breakdown is to be expected, indicating the determined partition function of the putative committor, Z_q, is steady for all Δt. The committor therefore passed the validation criterion and the RC closely approximates the committor. The Z_q criterion is roughly for $Z_q(q,\Delta t_0)$, 2 higher than $N_{AB} = 11$. The dynamics is accurate within a factor of ~1.2 for when the time scale $\Delta t_0 = 0.2$, or accurate within 20%.

Conversion of the committor to the Natural Coordinate

The FEP in figure 9.b. is not a the most convenient way in which to look at the coordinate; the diffusion coefficient is likely to fluctuate along the RC. Therefore, adapting the committor to its natural counterpart gives a more detailed view of the FEP (Krivov. 2018).

Figure 10.b shows the FEP along the natural committor. The FEP of the natural RC, qn, showed a more detailed FEP where the diffusion coefficient is constant at D(qn)=1. The dx was also turned down. The main points of observation were the two free energy barriers either side of the transition state. These two barriers, located at qn=10 and qn=26 respectively, are noisy. This possibly indicates a degree of folding and unfolding as the peptide enters and withdraws from the transition state. Both of these barriers are higher than observed in the rmsd by 1/3, which is shown in figure 10.b. The rmsd gives a simpler view of the dynamics, it showed only one smaller free energy barrier and did not give show a clear transition state.

After finding the new transition state, the time series of qn was plotted. The F/kT values were recorded for the aforementioned key states and frames with similar values of F/kT were located to give representative structures at the indicated energy level. The frames identified key points along the committor were plotted. Representative values were taken from the points along the committor, and frames showing the equivalent F/kT values along a timeplot of qn were selected and the structures were viewed. These key points are shown in figure 11 as 3D structures.

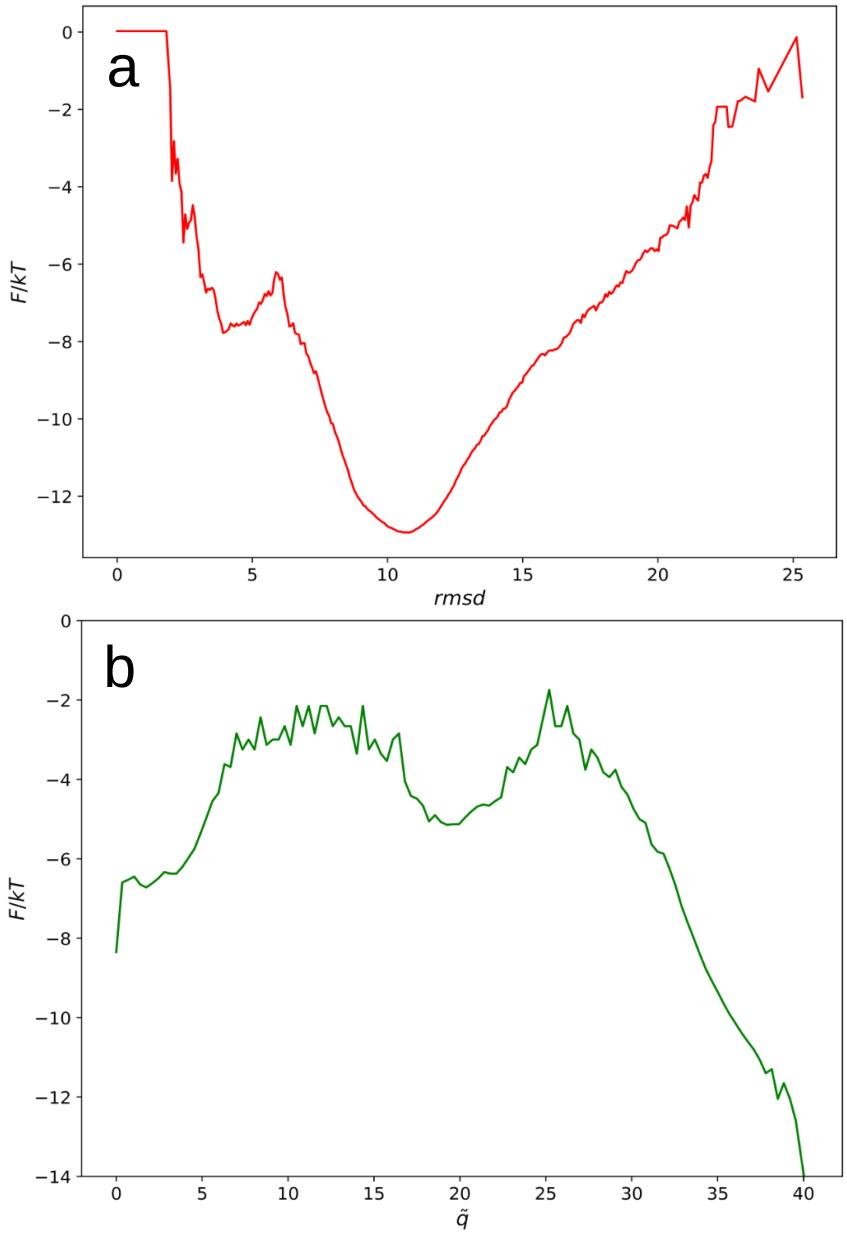


Figure 10) a) The FEP of Aβ-42 as a function of the rmsd. **b**) The FEP of Aβ-42 as a function of the Natural RC: it shows the FEP where D(qn)=1, giving a good description of the dynamics.

Using the Natural Coordinate Timeseries to Identify and Render Key Structures Along the Committor

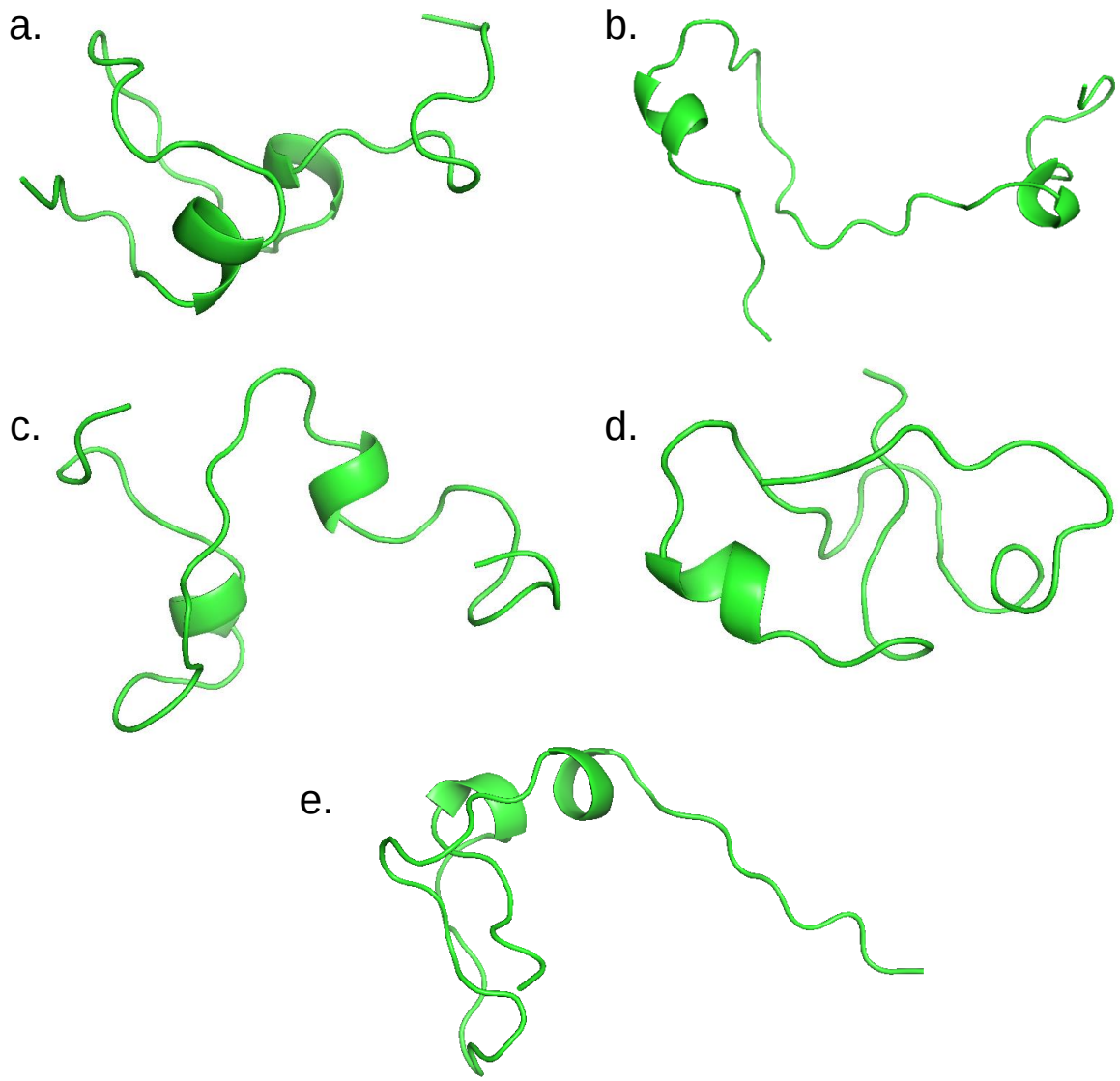


Figure 11) The 3D structures of Aβ-42 at key points along the committor. a) Structure taken from qn=1 along the committor. b) Structure taken from qn=10 along the committor. c) Structure taken from the transition state, where qn=20. d) Structure from the second energy barrier, where qn=26. e) Final structure taken from point qn=38-40.

Figure 11 shows representative structures at key points along the committor, namely the initial state (a), the first free energy barrier state (b), the transition state (c), the second free energy barrier state (d) and the final state (e). It is hard to assume much from each of these states, the only observable secondary structures forming are the beginnings of an α helix.

Dynamics Computed Directly from Trajectory and Updated Diffusive Model

	Natural Committor for Structure 5-0	
	Trajectory	Diffusive Model
NAB	12.00	13.64
Mfpt-AB	458.25	2294.14
Mftp-BA	100700	90034.8
Mtpt	98.25	215.612

Figure 12: Key dynamic properties between points along the natural committor.

Figure 12 was calculated between points qn=3 and qn=37 along the qn, these boundary values were selected to avoid boundary distortion. The figure indicates the diffusive dynamics are in better agreement with those calculated from the trajectory, than the rmsd RCs showed. The difference in N_{AB} is less than 10%, indicating the diffusive model deviates within an acceptable range. The N_{AB} values indicate the dynamics are projected onto an optimal RC, however, the other values differ.

Discussion & Conclusion

This study has outlined and successfully used methods to determine and validate the committor for the IDP, Aβ-42. The dynamics of a transition and the corresponding, relevant 3D structures along the transition have been mapped along the selected committor. To determine the committor from the suboptimal putative RC, a non-parametric approach for the determination of the committor from an equilibrium trajectory was used for Krivov. (2018). This was however not a perfect approach to the problem, as an equilibrium analysis has been applied to a non-equilibrium simulation/trajectory. The agreement between values for the dynamic properties along the committor suggested the non-equilibrium character of the dynamics has not affected the diffusive model much. Aside from the dynamics, the determined committor has passed the validation criterion for non-equilibrium situations. The Z_q profile in figure 9.c indicates the Zc,1 profile (partition function) of the committor is constant across multiple sampling intervals within 20%.

Löhr et al. (2021) produced a complete model for the mfpt between the different populations of Aβ-42 conformations. Each of the transition times between states was accompanied by the probability of the transition occurring. The kinetic model provides a good picture of the model's transitions between an array of states, defining a complete picture for the peptides shifts. The models had a high probability to transition to a form of "middle" state. An accurate comparison to the model made by Löhr et al. (2021) was unfortunately not possible as my results were not as comprehensive as I had hoped. The committor was only validated with certainty for a single structural ensemble, along with the predicted diffusive dynamics. Values such as the mfpt between states couldn’t be compared, as their model differed from ours and no marked transition similar to ours could be identified, partly as structural similarities with my results could not be seen. Having only one transition model is a limitation to this study, as Aβ-42 is an IDP with a large number of relevant conformations. For a complete picture of Aβ-42s transitions and kinetics, one must predict the dynamics between a large number of structural ensembles (Bhattacharya and Lin. 2019). In addition to the above, Löhr et al. (2021) measured the inter residue distances using α Carbon atoms, meanwhile our analysis utilised rmsds based on the nearest neighbour heavy atom distances.

This study did not manage to determine the committor for all of the structural ensembles obtained. This is probably due to a few of the rmsd profiles being hard to analyse via an equilibrium approach, namely those shown in figures 3.b and 4.a, c and e. These profiles were difficult to fit into the equilibrium analysis, due to their lack of multiple obvious metastable states. For profiles with very small free energy minima, the optimisation did not

work well. If only one metastable state could be identified, or one of the metastable had near negligible energetic stabilisation compared to the other, the equilibrium approach determined committors which poorly described the FEL. Figures showing attempts determining the committor in these situations can be seen in the appendix in S1, S2 and S3.

In conclusion, the results shown here show that the committor determination approaches and validation techniques outlined by Krivov. (2021), are sufficient for determining the dynamics for a given IDP structure and trajectory. The results here however give a less complete view of the protein dynamics between all the ensembles, than the kinetic ensembles developed by Lohr et al. (2021).

The next steps of the study should be to approach the putative RCs which could not be improved via an equilibrium-based approach by using one suited to determining the committor for non-equilibrium systems. Alongside this the determination of the eigenvectors would be useful for the further validation of the determined committors and differences between them would be instructive. The eigenvectors, however, could present similar issues I faced when determining the committor; the blind determination approach for determining the eigenvector in Krivov. (2020) is used for equilibrium scenarios, also inherent instability in the eigenvector optimisation might be an issue (Krivov. 2021). Additionally, a trajectory set determined at a higher temperature, i.e body temperature, would be useful firstly for comparison with other papers and may be more relevant to clinical studies of the peptide. This is as the trajectories from Lohr et al. (2021) were produced at 274K.

Declared Word Count: 5,219

References

Banushkina, P.V. and Krivov, S.V. 2015. Nonparametric variational optimization of reaction coordinates. *The Journal of Chemical Physics*. **143**(18),p.184108.

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

Bhattacharya, S. and Lin, X. 2019. Recent Advances in Computational Protocols Addressing Intrinsically Disordered Proteins. *Biomolecules*. **9**(4),p.146.

Chung, H.S. and Eaton, W.A. 2018. Protein folding transition path times from single molecule FRET. *Current Opinion in Structural Biology*. **48**,pp.30–39.

Jung, H. Covino, R. and Hummer, G. 2019. *Artificial Intelligence Assists Discovery of Reaction Coordinates and Mechanisms from Molecular Dynamics Simulations*. Unpublished. [Online][Accessed: April 15th]. Available from: https://arxiv.org/abs/1901.04595

Krivov, S. 2013. On Reaction Coordinate Optimality. *Journal of Chemical Theory and Computation. **9,** 135-146.*

Krivov, S. 2020. CFEP2, A Criterion For Committor. [Online] [Accessed March 25th – April 15th]. *Open Access via: GitHub*. Available from: https://github.com/krivovsv/CFEPs/blob/main/CFEP2.ipynb

Krivov, S. 2021. *Non-Parametric Analysis of Non-Equilibrium Simulation*s. [Online] [Accessed: March 28th] *Unpublished*. Available from: https://arxiv.org/abs/2102.03950

Krivov, S.V. 2018. Protein Folding Free Energy Landscape along the Committor - the Optimal Folding Coordinate. *Journal of Chemical Theory and Computation*. **14** (7),pp.3418–3427.

Krivov, S.V. 2021. Blind Analysis of Molecular Dynamics. *Journal of Chemical Theory and Computation*.1-37.

Lindorff-Larsen, K., Piana, S., Dror, R.O. and Shaw, D.E. 2011. How Fast-Folding Proteins Fold. *Science*.. **334**(6055),pp.517–520.

Löhr, T., Kohlhoff, K., Heller, G.T., Camilloni, C. and Vendruscolo, M. 2021. A kinetic ensemble of the Alzheimer's Aβ peptide. *Nature Computational Science*.. **1**(1),pp.71–78.

Murphy, M.P. and Levine, H. 2010. Alzheimer's Disease and the Amyloid-β Peptide. *Journal of Alzheimer's Disease*.. **19**(1),pp.311–323.

Pande, V., Beauchamp, K., Bowman, G. 2010. Everything you wanted to know about Markov State Models but were afraid to ask. *Methods* **52**(1): 99-105.

Pietrucci, F. 2017. Strategies for the exploration of free energy landscapes: Unity in diversity and challenges ahead. *Reviews in Physics*. **2**,pp.32–45.

Polizzi, N.F., Therien, M.J. and Beratan, D.N. 2016. Mean First-Passage Times in Biology. *Israel Journal of Chemistry*. **56**(9-10),pp.816–824.

Seeber, M., Cecchini, M., Rao, F., Settanni, G. and Caflisch, A. 2007. Wordom: a program for efficient analysis of molecular dynamics simulations. *Bioinformatics*. **23**(19),pp.2625–2627.

Seeber, M., Felline, A., Raimondi, F., Muff, S., Friedman, R., Rao, F., Caflisch, A. and Fanelli, F. 2011. Wordom: A user-friendly program for the analysis of molecular structures, trajectories, and free energy surfaces. *Journal of Computational Chemistry*. **32**(6),pp.1183–1194.

Sengupta, U., Nilson, A.N. and Kayed, R. 2016. The Role of Amyloid-β Oligomers in Toxicity, Propagation, and Immunotherapy. *EBioMedicine*. **6**,pp.42–49.

Sgourakis, N.G., Merced-Serrano, M., Boutsidis, C., Drineas, P., Du, Z., Wang, C. and Garcia, A.E. 2011. Atomic-Level Characterization of the Ensemble of the Aβ(1–42) Monomer in Water Using Unbiased Molecular Dynamics Simulations and Spectral Algorithms. *Journal of Molecular Biology*. **405** (2), pp.570–583.

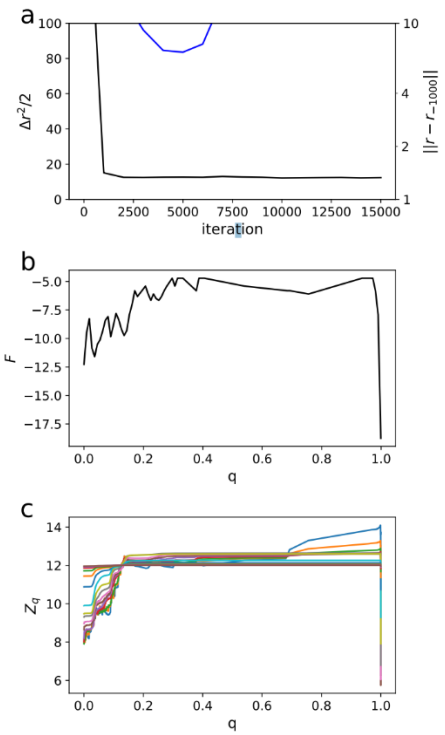
Turoverov, K.K., Kuznetsova, I.M. and Uversky, V.N. 2010. The protein kingdom extended: Ordered and intrinsically disordered proteins, their folding, supramolecular complex formation, and aggregation. *Progress in Biophysics and Molecular Biology*. **102**(2-3),pp.73–84.

Uversky, V.N. 2019. Intrinsically Disordered Proteins and Their “Mysterious” (Meta)Physics. *Frontiers in Physics*. **7**.

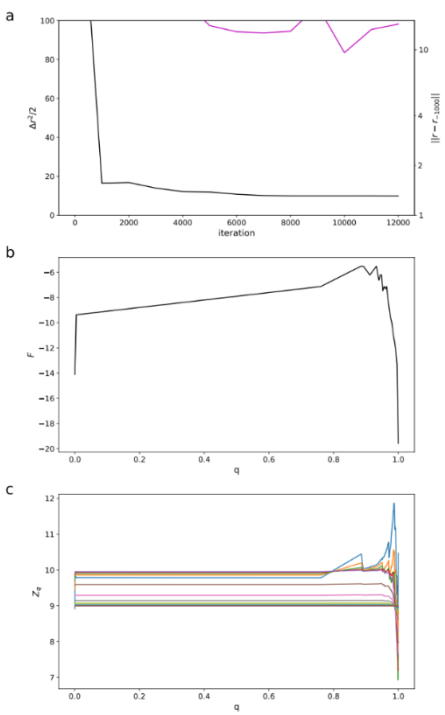
Zheng, W., Tsai, M.-Y. and Wolynes, P.G. 2017. Comparing the Aggregation Free Energy Landscapes of Amyloid Beta(1–42) and Amyloid Beta(1–40). *Journal of the American Chemical Society*. **139**(46),pp.16666–16676.

Zuckerman, D.M. 2011. Equilibrium Sampling in Biomolecular Simulations. *Annual Review of Biophysics*. **40**(1),pp.41–62.

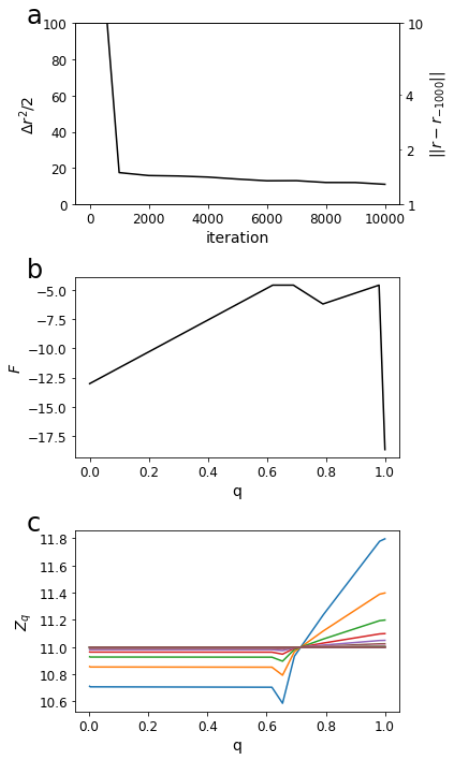
Appendix



S1: Attempted committor determination for the reduced trajectory using reduced structure 3-0 as a reference.



S2: Attempted committor determination for the oxidative trajectory using MetSo structure 3-0 as a reference.



S3: Attempted committor determination for the oxidative trajectory using MetSo structure 4-0 as a reference.