**Bachelor Thesis**

# Molecular dynamic simulations of a truncated measles $N_{TAIL}$-domain and predicted effects on its Photoinduced Electron Transfer decay time

# Moleküldynamik-Simulationen eines gekürzten Masern-$N_{TAIL}$-Proteins und prognostizierte Effekte auf seine Zerfallszeit des photoinduzierten Elektrontransfers

by
**Aaron Christopher Globisch**
from Bremen at the
Max Planck Institute of Multidisciplinary Sciences
Theoretical and Computational Biophysics Department
in Göttingen, Germany

Preparation time: 14.01.2022 – 19.04.2022
Supervisor: Dr. Gabor Nagy
First assessor: Prof. Dr. Helmut Grubmüller
Second assessor: Prof. Dr. Jörg Enderlein

# Contents

# 1. Introduction

Measles is a highly contagious airborne disease caused by the measles virus (MeV). Due to extensive vaccination within the Global Vaccination Action Plan of the World Health Organisation, death rates of Measles infections dropped by 62% from 2000 to 2019. [1] Despite these efforts, 207,500 people died of Measles in 2019 and to this point there is no cure for the disease. [1]

To develop a cure it is helpful to understand the biochemical pathway of MeV and its dynamics in detail. MeV is a negative sense single stranded RNA virus, composed of the RNA, six proteins and a lipid membrane.

Its nucleoprotein (MeV-N) binds to RNA and is composed of the $N_{CORE}$-domain from residues 1 to 405 and the $N_{TAIL}$-domain from residues 406–525. The $N_{CORE}$-domain forms the nucleocapsid and encapsulates the viral RNA. In contrast to the $N_{CORE}$, which is mostly $\alpha$-helical, the $N_{TAIL}$ is mostly disordered including a transient $\alpha$-helical Molecular Recognition Element ($\alpha$-MoRE) from residues 485–502. [2]

It has been proposed that $N_{TAIL}$ plays a role in transcription initiation and therefore is crucial for the virus replication. [2] In the absence of RNA, MeV-N forms heterodimers with the measles phospoprotein (MeV-P). In the presence of RNA, MeV-N forms the nucleocapsid while MeV-P forms homotetramers that bind the large RNA polymerase (MeV-L). It is likely that the XDomain of MeV-P (PXD) then binds the $\alpha$-MoRE region of $N_{TAIL}$ to recruit MeV-L and initiate the transcription. [3,4]

In order to show that the PXD-$N_{TAIL}$binding takes place, it is helpful to capture it on an atomic scale. However, the $N_{TAIL}$ protein is an Intrinsically Disordered Protein (IDP) and the binding processes of IDPs are hard to visualise on such scale. This is due to current structure revealing methods, such as Small-angle X-ray scattering, resolve the size and shape of monodisperse macromolecules, but IDPs alter their conformations highly frequently and therefore frequently vary in size. Hence, the highly variable dynamics of IDPs present considerable challenges to obtain information on an atomic scale.

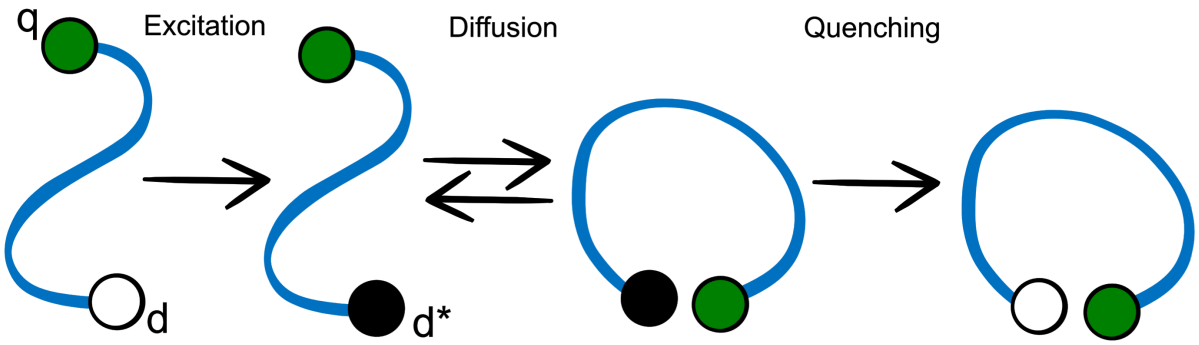


**Figure 1.1:** Scheme for the quenching process between a dye (d) and a quencher (q) of a random peptide. After laser pulses excite the dye (d $\rightarrow$ d*), the peptide randomly diffuses between extended and dq contact conformations. Upon electron transfer between the excited dye and the quencher, the dye gets quenched. The scheme is taken and modified from Buscaglia *et al.* [5]

To obtain contact ratios in between two atom(-groups), the group of Sara Vaiana (Arizona State University, Tempe) recently carried out Photoinduced Electron Transfer (PET) experiments. In PET a sample containing a predetermined concentration of the atom(-groups) is exposed to laser pulses. The wavelength of the laser is chosen such to match the excitation frequency of a desired atom(-group) (dye), as sketched in fig. 1.1.

Upon excitation, the dye undergoes spontaneous fluorescence decay while it diffuses freely. Upon contact with the other atom(-group) (quencher), green filled circle in fig. 1.1, the excited electron on the dye can be transferred to the quencher, which accelerates the observed decay noticeably. In proteins, tryptophan residues are suitable dyes due to their fluorescence, which can be quenched by cysteines (and histidines to lesser extent).

Using tryphophan-cysteine pairs as intrinsic the dye and quencher for PET is advantageous because we did not attach bulky fluorescence dyes, which likely affect the protein dynamics. Unfortunately, the $N_{TAIL}$ wild-type do not contain any tryptophan or cysteine residues, but they can be introduced into $N_{TAIL}$ easily, as its tyrosines can be substituted with tryptophans and its serines with cysteines.

These substitutions leave the overall chemical structure of the mutated $N_{TAIL}$ similar to the wild-type and hence, we assume its dynamics to be largely unperturbed. Therefore, PET provides intra-molecular contact dynamics of predetermined locations about $N_{TAIL}$, but it unfortunately does not provide its full atomic dynamics.

Here, we have used Molecular Dynamics (MD) simulations to complement PET experiments with full atomic details. MD simulations numerically approximate the solution of the differential equations of Newtonian mechanics of a known conformation to obtain a new conformation. Iterative computations of each new conformation then reveal the trajectory of the system.

Unfortunately, due to the applied approximations, the obtained dynamics have limited accuracy, which makes it necessary to verify the simulations against independent measurements.

Andrea Vaiana *et al.* [6] proposed that PET decay times depend on the distance between the dye and quencher. Hence, the full atomic detail in MD simulations enables us to calculate the dye to quencher (dq) distances directly from the obtained trajectory, which in turn allow the calculation of the expected PET decay rates from simulations.

Thus, PET experiments and MD simulations complement each other: The comparison of experimental and calculated decay times can verify the simulation accuracy. In turn, the validated simulation results provide an insight into the underlying dynamics.

Researches of a collaborative research project between the groups of Sara Vaiana and Helmut Grubmüller (Max Planck Institute of Multidisciplinary Sciences, Göttingen) applied this approach successfully to study the $N_{TAIL}$ dynamics. [7] They used MD simulations of numerous $N_{TAIL}$ mutants to calculate the PET decay curves of several dq pairs. Their results were verified using multiple PET experiments as well as Circular Dichroism and Short Angle X-Ray Scattering measurements. [7,8]

Because $N_{TAIL}$ is an IDP and behaves similar to a random coil, and in

turn random coils can be described by homopolymer models, they further compared their observed PET decay times to those obtained from homopolymer models. Simple homopolymer models predict that in a random coil the PET decay time $\tau_{\mathrm{dq}}(N) \propto N^{\gamma}$ is proportional to a function of the number of residues $N$ in the sequence between the dye and quencher. [8,9] For simple homopolymer models $\gamma \approx 3/2$ can be expected. [5] Interestingly, the PET decay times for two mutants differed significantly from what was expected from the homopolymer model calculations.

Due to the PET experiments, both mutants contained two point mutations near the $\alpha$-MoRE region: a tyrosine to tryptophan at residue 518 and a serine to cystine mutations at either residue 482 or at residue 488 (subsequently denoted as dq mutants Q482 and Q488).

Based on those homopolymer models, the decay times for Q482 and Q488 were expected to be 3.7 µs and 2.8 µs, respectively. However, Vaiana *et al.* [8] obtained nearly twice the decay time 5.94 µs for Q482 and a similar decay time 2.42 µs for Q488 from PET experiments compared to the homopolymer model estimates. Gabor Nagy (Max-Planck-Institute of Multidisciplinary Sciences, Göttingen), from the group of Grubmüller, found further indications that $\mathrm{N_{TAIL}}$ does not always behave like a random coil-like peptide, in ongoing research from subsequently performed MD simulations of $\mathrm{N_{TAIL}}$. These indications led to the hypothesis of competing effects between local secondary structure and non-local interactions.

We truncated a $\mathrm{N_{TAIL}}$ mutant from residue 482 to 518 to test if electrostatic long-range interactions from outside the truncated sequence alter the predicted PET decay times of the mutant pairs Q482 and Q488.

We then performed MD simulations of truncated $\mathrm{N_{TAIL}}$ mutants and computed the predicted PET decay times of Q482 and Q488. These were then compared to the full length $\mathrm{N_{TAIL}}$ decay times obtained by the groups of Sara Vaiana and Helmut Grubmüller under similar conditions. Because their used homopolymer model was optimised for peptide lengths similar to $\mathrm{N_{TAIL}}$, we improved the quantitative decay time estimate of a random

coil-like sequence to match the length of truncated $N_{TAIL}$ by performing additional heteropolymer $(AGQ)_n$ MD simulations. [9]

From the truncated $N_{TAIL}$ variants and $(AGQ)_n$ peptides, we calculated PET decay curves from the dq distances observed in the course of the simulations. This enabled us to examine whether the decay times will be more similar to the decay times of full length $N_{TAIL}$ or the random coil-like peptide.

We further extracted the secondary structure frequencies of the truncated $N_{TAIL}$ and $(AGQ)_n$ peptide simulations, which allowed us to separate the effect of secondary structure from random diffusion on PET decay times more precisely. The structural analysis also revealed secondary structures which affect the decay time of Q488.

Ultimately, the results of the study will contribute to clarify the complex dynamics of the $N_{TAIL}$ domain of MeV-N, which may lead to a better understanding of the measles virus replication.

## 2. Methods

### 2.1. Molecular dynamics simulations

We performed 23 MD simulations of a truncated $N_{TAIL}$ variant using the GROMACS 2019 package, the Amber99SB-disp (a99SB-disp) force field and a modified TIP4 water model for a total length of 130 µs simulation time.[10,11]

30 µs of the total simulation time were covered by three initial 10 µs simulations, which were started from differing starting conformations. However, because 30 µs simulation time in total were not sufficient to evaluate the dynamics of the truncated $N_{TAIL}$ variants with certainty, we prepared 20 additional 5 µs simulations.

In order to investigate the most frequently occurring dynamics but also cover a wide range in conformational space, we started the additional simulations from conformations which were selected based on a Principal

Component Analysis (PCA) of the initial three simulations. [12]

We performed PCA, from which we extracted the first five principal components (PCs), which preserved 83.2 % of the dynamics in the initial simulations and projected the second to fifth PC onto the first PC. Every projection displayed each adopted conformation from the initial simulations in dependence of the used PCs. To select the new starting conformations, we chose one conformation from the global energy minimum and three from different local minima of each of the four projections, which yielded the first 16 new starting conformations. To potentially reveal yet undiscovered energy minima, the last four new stating conformations were randomly selected from a sparsely populated area in each projection.

To ensure maximal similarity between the truncated $N_{TAIL}$ and the $N_{TAIL}$ peptides used in the PET experiments of Vaiana *et al.*, we introduced an acetyl group (ACE) to the N-terminus and an amine group ($NH_2$) to the C-terminus of the initial truncated $N_{TAIL}$ starting conformations. Next, we solvated each peptide conformation in 150 mmol NaCl solution in a truncated dodecahedral box, applied periodic boundary conditions to the box and set a wall distance from the centre of the peptide such that the extended peptide avoided contact through the periodic boundaries.

All simulations were propagated using a leapfrog integrator in 4 fs intervals. [13] Holonomic constraints using the LINCS algorithm, with a LINCS order of 6, and virtual atom sites for apolar hydrogen were used to ensure that molecular vibrations in systems allow for this time interval. [14,15]

We set the cutoff distances for grid cells neighbour searching as well as short-range electrostatics and van der Waals interactions to 1 nm and used EnerPres as a dispersion correction for the van der Waals cut off. [10] The long-range electrostatics were computed by a Particle Mesh Ewald (PME) scheme with an PME order of 4 at a Fourier grid spacing of 0.13 nm. [10,16,17]

All simulations were subjected to constant temperature and pressure. The temperature was maintained using velocity-rescaling, which is a modified Berendsen thermostat. [15,18] We chose protein and non-protein as tem-

perature coupling groups as well as a temperature time constant and a reference temperature for both groups to be 0.1 ps and 298 K, respectively.

The pressure was maintained by an isotropic Parrinello-Rahman barostat. [19] Here, we chose the pressure coupling constant and the reference pressure to be 0.1 ps and 1 bar, respectively, as well as the isothermal compressibility of water to be $4.5 \cdot 10^{-5}\,\mathrm{bar}^{-1}$ as a compressibility constant.

To obtain initial conditions for the peptides, we first performed a steepest descent energy minimisation on the starting conformations. [10]

Second, we ran a 100 ps long MD simulation under constant volume, a temperature of 300 K and positional restraints on the peptide atoms, using the function type 1 and an isotropic force constant of $1\,\mathrm{GJ\,mol^{-1}\,nm^{-2}}$. [10] The initial velocities were assigned based on a Maxwell-Boltzmann distribution such that the system reached a temperature of 300 K on average.

Third, a 100 ps constant pressure and temperature MD simulation was performed using a Berendsen barostat. [15] Here, a pressure of 1 bar was maintained using the isothermal compressibility of water ($4.5 \cdot 10^{-5}\,\mathrm{bar}^{-1}$) and a pressure coupling time constant of 0.5 ps. The position restraints on the protein were maintained with the function type 1 and an isotropic force constant of $1\,\mathrm{GJ\,mol^{-1}\,nm^{-2}}$. [10]

Finally, we performed a 500 ps MD simulation with constant temperature and pressure but without the positional restraints on the protein. Here, a constant pressure was maintained by a Parrinello-Rahman barostat with the same compressibility and reference pressure as before, but a coupling time constant of 2 ps. [19,20] The last conformations of these simulations were used to start all production simulation trajectories.

## 2.2. Decay time calculations

Zerze *et al.* [21] have shown that predicted PET decay times can be calculated from MD simulation trajectories, enabling a direct comparison between the dq contact dynamics of the simulated and experimentally used peptides. Moreover, previous studies have shown that there is no preferred orientation between an excited dye relative to the quencher for

quenching, but it requires van der Waals contacts.[6,22,23] Hence, decay times are calculated from the dq distances, where the dq distance is defined as the minimum distances

$$r_{\mathrm{dq}} = \min|\mathbf{r}_{\mathrm{d}} - \mathbf{r}_{\mathrm{q}}| \tag{2.1}$$

between any tryptophan indole ring system atom $\mathbf{r}_{\mathrm{d}}$ and any cysteine side-chain atom $\mathbf{r}_{\mathrm{q}}$. In literature, dye and quencher are considered to be in contact, if the distance between any $\mathbf{r}_{\mathrm{d}}$ and $\mathbf{r}_{\mathrm{q}}$ is smaller than a cutoff distance $r_{\mathrm{cut}} = 0.4\,\mathrm{nm}$ at any given time step.[6,21,22]

However, because constant instantaneous quenching below a cutoff distance is a strong assumption, we used a more realistic approach computing the quenching rate. It has been proposed that observed PET decay rates are a superposition of diffusion-limited contributions $k_{\mathrm{D+}}^{-1}$ and reaction-limited contributions $k_{\mathrm{R}}^{-1}$.[21–24]

In the diffusion-limited scenario, diffusion is slow compared to instantaneous quenching upon contact. Due to the slow diffusive dynamics, an excited dye might have undergone fluorescent decay before quenching could have occurred, which makes it necessary to quantify the survival probability[21]

$$S(t) = \langle \Theta[t_{\mathrm{c}}(t_0) - t - t_0] \rangle_{t_0} \tag{2.2}$$

of the dye excited triplet state. The probability is obtained by using the Heaviside step function $\Theta[t]$ for the excited triplet state survival probability of a time $t$ averaged over all $t_0$ time steps in the simulation, in which dq are in contact. Thereby the frame $t_{\mathrm{c}}(t_0)$ is the first dq contact time after a selected $t_0$.

To then obtain the diffusion-limited rate $k_{\mathrm{D+}}$, we calculated the survival probabilities for 30 times $t \in [1, 1000]$ and approximated a mono-exponential function $\exp[k_{\mathrm{D+}}t] = S(t)$ for the calculated survival probabilities for all trajectories individually.

In contrast, in the reaction-limited scenario, diffusion is fast compared

to a quenching event, such dye and quencher have to come into contact many times on average before a quenching event occurs. The quenching rate [22]

$$q(r_{\mathrm{dq}}) = q_{\mathrm{c}} \exp\left[-\zeta(r_{\mathrm{dq}} - r_{\mathrm{cut}})\right] \tag{2.3}$$

was computed with the constant quenching rate $q_{\mathrm{c}} = 1 \cdot 10^8\,\mathrm{s}^{-1}$ and a distance dependent contact probability estimate, containing the free parameter $\zeta = 33.3\,\mathrm{nm}^{-1}$. [22] The continuous probability estimate takes into account, that quenching events can or cannot happen on both dq distances below and above $r_{\mathrm{cut}}$.

Then, the reaction-limited rate [21]

$$k_{\mathrm{R}} = \int_0^\infty q(r_{\mathrm{dq}})\rho(r_{\mathrm{dq}})\mathrm{d}r_{\mathrm{dq}} \tag{2.4}$$

was calculated over the quenching rate and the normalised dq distance distribution $\rho(r_{\mathrm{dq}})$, which was discretised in bins of 0.1 nm from 0 nm to the maximum dq distance.

Having diffusion- and reaction-limited rate obtained, we were able to calculate the observed decay rate $k_{\mathrm{obs}}^{-1} = k_{\mathrm{D+}}^{-1} + k_{\mathrm{R}}^{-1}$. However, in our research we used decay times $\tau$ instead of the decay rates. Decay times are defined as time required for the number of excited state molecules reduced to $1/e$ of its original population and can be calculated as $\tau = k_{\mathrm{obs}}^{-1}$.

It should be noted that the two mutants Q482 and Q488 of the simulated truncated $\mathrm{N_{TAIL}}$ peptide differed only by a single oxygen and sulphur atom and we therefore assumed that the dynamics of both mutants were similar. Accordingly, we have used our simulations to calculate the decay times for both Q482 and Q488 from the same trajectory, which allowed us to use our simulation trajectories more efficiently.

Having 23 trajectories, we obtained each 23 decay times for Q482 and Q488. Because the resulting decay times did not follow a clear distribution, we chose bootstrapping to estimate the mean and uncertainty,

drawing a sample of 23 decay times with replacement repeated 10 000 times. The sample size of 23 was chosen, because it satisfies the condition that every drawn sample will contain 63.2% of the observations on average. [25]

## 2.3. Radius of gyration calculations

In order to quantify the spatial expansion of the protein during a simulation we calculated the radius of gyration [10]

$$R_{\mathrm{gyr}} = \left( \frac{\sum_i m_i ||\mathbf{r}_i||^2}{\sum_i m_i} \right)^{\frac{1}{2}} \tag{2.5}$$

as the mass-weighted root-mean-square distance $\mathbf{r}_i$ of every atom $i$ to the peptide centre of mass, where $m_i$ is the mass of the $i$-th atom. [26] Larger $R_{\mathrm{gyr}}$ values indicate a more expanded protein, whereas smaller $R_{\mathrm{gyr}}$ indicate a more compact one. Larger variances in $R_{\mathrm{gyr}}$ as a function of time point to more fluctuations in terms of protein expansion and smaller variations in $R_{\mathrm{gyr}}$ corresponds to more stable conformations.

## 2.4. Secondary structure calculations

To characterise internal structural changes of the protein during the simulations, we calculated the secondary structure of the protein every 1 ns using the DSSP algorithm of Kabsch and Sanders provided in SESCA. [27–29]

DSSP classifies the hydrogen bond patterns along the protein backbone for hydrogen bond donor (NH) and acceptor (C=O) groups. It first identifies hydrogen bonds within the backbone by calculating the electrostatic interactions between two neighbouring H-bonding donor-acceptor groups. In the next step H-bonds up to 5 residues away are grouped into 3-, 4-, or 5-turns ($n$-turns) and H-bonds above 5 residues away are identified as $\beta$-bridges. Subsequently consecutive $n$-turns are classified as $n$-helices and consecutive $\beta$-bridges between shared residues as $\beta$-sheets. Further, unclassified residues with a curvature above 70° are classified as bends.

## 2.5. Protein contact map calculations

To reveal the relative orientation of secondary structure elements of our simulated peptides, we used protein contact maps. Protein contact maps represent the contact probability between selected atoms pairs of residues in a protein of $N$ residues using a two-dimensional $N \times N$ matrix.

In order to obtain the contact map, we calculated the contact probabilites of backbone $C_\alpha$ atoms of each residue pair. We considered residues in contact when their $C_\alpha$ to $C_\alpha$ distance is below 0.6 nm and used the HBSS algorithm provided in SESCA. [29] Because neighbouring $C_\alpha$ atoms lie within their 0.6 nm cut off distance, they are displayed with a 100% contact probability.

We also computed salt-bridge formation probabilities. To this aim, we selected the positively charged nitrogen atom from arginines as well as lysienes and the negatively charged oxygen of aspartic acid as well as glutamic acid and used the HBSS algorithm provided in SESCA to calculate their contact probabilites, using a cut off distance of 0.6 nm.

## 2.6. Mutual information calculations

For the purpose of quantifying how $C_\alpha$ and salt-bridge interactions within the peptide correlate with dq contacts, we used the normalised pointwise mutual information. [30] The normalised pointwise mutual information (NPMI) [30]

$$I_{\mathrm{mut}}(X_i, Y_j) = -(\ln P(X_i, Y_j))^{-1} \ln\left(\frac{P(X_i, Y_j)}{P(X_i)P(Y_j)}\right) \tag{2.6}$$

quantifies the amount of information obtained about the state of variable $X$ by observing another variable $Y$, where $P(X_i)$ and $P(Y_j)$ are the marginal probabilities of $X$ being in state $i$ and $Y$ being in state $j$ and $P(X_i, Y_j)$ is the joint probability of $X$ being in states $i$ and $Y$ being in state $j$ at the same time.

In order to quantify the correlations, we first set variable $X_i$ as the

dq distances of either Q482 or Q488 and variable $Y_j$ as the distances of either $C_\alpha$ or charged residues. Next, we used a concatenated trajectory of all 23 simulations to determine the marginal and joint probabilities of contacts between the selected variables, using 0.4 nm as the dq contact cutoff distance and 0.6 nm as the $C_\alpha$ and salt-bridge contact cutoff. Lastly, with the probabilites obtained, we calculated the NPMI as shown in (2.6).

Because NPMI measures dependence between two variables, increasing values from 0 to 1 show that the occurrence of e.g. dq contacts increasingly promotes the occurrence of $C_\alpha$ contacts and vice versa, whereas decreasing values from 0 to $-1$ show an increasing hindering of simultaneous occurrence between the two variables.[30] Thus if, and only if the two variables are independent from each other the NPMI is zero.

## 2.7. $AGQ_n$ simulations

Alanine (A), glycine (G) and glutamine (Q) repeat peptides have been used extensively in literature for models of random coil-like behaviour. The zero net charge and high glycine content of the AGQ repeat sequences lead to mostly unstructured, random coil-like behaviour in PET experiments.[5,21,31–33]

To enable us to calculate decay rates of a random coil-like peptide, we simulated two sets of peptides in which tryptophan (W) and cysteine (C) are used as the dye and quencher at the termini, see table 2.1.

**Table 2.1:** Overview of the simulated (AGQ) repeat peptide sequences and their lengths $n$ and $m$. The amino acids Alanine (A), glycine (G) and glutamine (Q), as well as tryptophan (W) and cysteine (C) were used.

| Peptide sequence | $n$ | $m$ |
|---|---|---|
| C-$(AGQ)_n$-W | $1, 2, 3, 5$ | - |
| C-$(AGQ)_5$-S-$(AGQ)_m$-W | - | $1, 3, 5$ |

Whereas the first set of peptides C-$(AGQ)_n$-W with $n = 1, 2, 3, 5$ solely imitated a random coil-like sequence, the second set C-$(AGQ)_5$-S-$(AGQ)_m$-W with $m = 1, 3, 5$ included an additional serine/cystine after five AGQ

repeats. Because we assumed that the serine/cystine substitutions leave the dynamics of peptides unchanged, the second set allowed us to calculate the PET decay times between random coils of similar lengths to the truncated $N_{TAIL}$ peptide, as well as to estimate the PET decay times of peptides with a sequence S-$(AGQ)_5$-C-$(AGQ)_m$-W. These peptides probe the volume exclusion effect of a dangling long peptide chain attached to the quenching residue.

The volume exclusion effect describes the phenomenon, that occupied space of one part of a sequence cannot be occupied by another part of the sequence. [34] Therefore the decay times of dq pairs among a peptide sequence differed from decay times of the dq pairs which are located at the termini of a peptide.

We prepared 3 starting conformations of every AGQ repeat peptide and performed MD simulations under the same conditions as described in chapter 2.1, for the truncated $N_{TAIL}$ simulations, for 5 µs. Hence, we obtained 15 µs of simulations of each peptide length in both sets.

# 3. Results and discussion

## 3.1. PET decay time analysis

In order to investigate how the truncation of $N_{TAIL}$ affects its PET decay times, we first computed the predicted decay times, as described in chapter 2.2, for both of the dq pairs from the performed MD simulations of the truncated $N_{TAIL}$ peptide.

We obtained decay times from 0.21 µs to 3.02 µs and from 0.18 µs to 5.39 µs for Q482 and Q488, respectively, from 3 truncated $N_{TAIL}$ simulations of 10 µs stimulation time and 20 truncated $N_{TAIL}$ simulations of 5 µs simulation time using the a99 force field. The average decay times are $(0.68 \pm 0.22)$ µs for Q482 and $(0.6 \pm 0.4)$ µs for Q488 and hence, are indistinguishable within the error of one another, but it has to be noted that the uncertainties are rather large, see red diamonds in fig. 3.1.

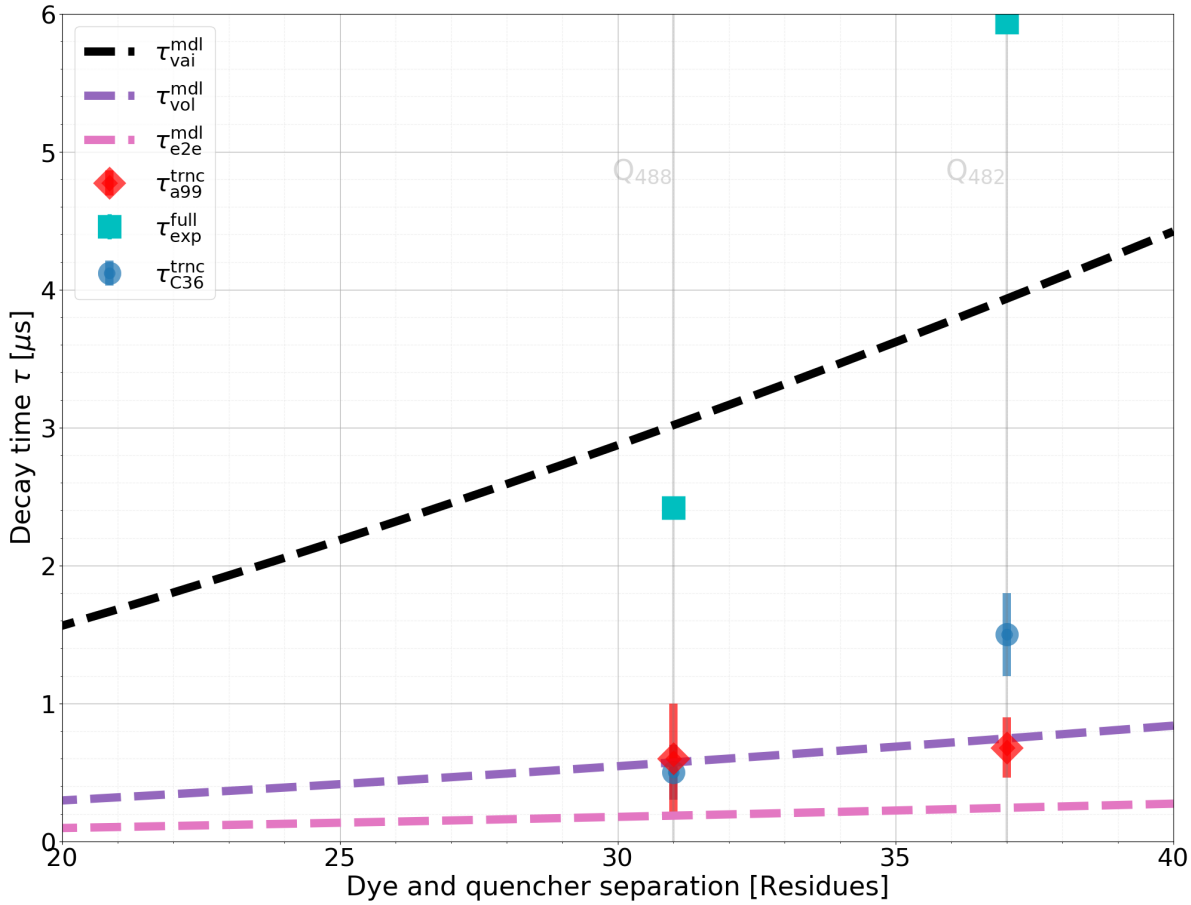


**Figure 3.1:** Overview of predicted and estimated PET decay times as a function of dq residue separation. The decay times for Q482 and Q488 are represented by signs on the left at residue 31 and right at residue 37, indicated by the vertical grey lines. The decay time estimated $\tau^{mdl}_{vai}$ of the homopolymer model used by Vaiana *et al.*[8] is shown as a dashed black line. Their obtained full-length $N_{TAIL}$ PET decay times $\tau^{full}_{exp}$ for Q482 and Q488 are shown as cyan squares.[8] Decay times of the truncated $N_{TAIL}$ $\tau^{trnc}_{c36}$, simulated in the Charmm36M/OPC force field, obtained and provided by Gabor Nagy, are displayed as blue circles. Our calculated decay times $\tau^{trnc}_{a99}$ of truncated $N_{TAIL}$ using the a99SB-disp force field are displayed as the red diamonds. The purple and pink dashed lines show the decay time estimate $\tau^{mdl}_{e2e}$ of the homopolymer model adapted to pure random coil-like peptides and the decay time estimate $\tau^{mdl}_{vol}$ of the homopolymer model adapted to random coil-like peptides subjected to an excluded volume, respectively.

However, the decay times obtained from the truncated peptide were considerably smaller than the decay times of the full-length $N_{TAIL}$ for Q482 3.7 µs and Q488 2.8 µs obtained by Vaiana *et al.*[8] in PET experiments,

see cyan squares in figure 3.1. A comparison between the decay times obtained from the truncated and full-length $N_{TAIL}$ variants clearly showed two things:

Firstly, the decay times of both mutants of the truncated peptide were roughly five-fold smaller than those of the full-length peptide. Secondly, the decay times of both mutants obtained from the truncated $N_{TAIL}$ were similar to each other, whereas the decay times obtained from the full-length $N_{TAIL}$ differ by a factor of 2.48.

In the following, we want to discuss why these differences occurred. First of all, we investigated if the different decay times obtained from simulations and experiments were caused by the used a99SB-disp force field. For this purpose we analysed the decay times $(1.5 \pm 0.3)\,\mu s$ for Q482 and $(0.50 \pm 0.20)\,\mu s$ for Q488 of truncated $N_{TAIL}$ peptides simulated in the Charmm36M/OPC force field, see blue circles in figure 3.1, which were provided by Gabor Nagy (Max Planck Institute of Multidisciplinary Sciences, Göttingen) in ongoing research. Both decay times were calculated from the same trajectories using the same method that we used. Because we used the same method, potential differences of the decay times in the different quenching positions can solely be reduced on the different peptide dynamics caused by the force fields.

Two inconsistent effects were observed. First, the decay times based on Charmm36M/OPC were significantly faster than those obtained from PET experiments, matching the trend of the decay times based on a99SB-disp. This indicates that the truncated $N_{TAIL}$ based on Charmm36M/OPC behaved similar to the truncated $N_{TAIL}$ based on a99SB-disp. Second, the decay times based on Charmm36M/OPC differ by a factor of 3, matching the trend of the decay times obtained from PET experiments. This indicated that the dq contact dynamics of the truncated $N_{TAIL}$ based on Charmm36M/OPC were more similar to the full-length peptide in the PET experiments.

However, it must be noted that the decay times based on Charmm36M/OPC were calculated from three simulations resulting in a total of only 31 µs simulation time. Due to the short simulation time,

it cannot be ruled out that the decay times were unrepresentative for truncated $N_{TAIL}$ simulations using Charmm36M/OPC. Therefore, it is difficult to tell from the decay times based on Charmm36M/OPC alone, if the different decay times obtained from simulations and experiments were caused by the use of the a99SB-disp force field.

To rule out if the different decay times obtained from the simulations and experiments were caused by the used a99SB-disp force field, we compared decay times computed from full-length $N_{TAIL}$ variants, simulated using a99SB-disp, with decay times computed from full-length $N_{TAIL}$ variants, simulated using Charmm36M/OPC.

The predicted full-length $N_{TAIL}$ decay times based on a99SB-disp were $(28 \pm 6)$ µs for Q482 and $(15 \pm 6)$ µs for Q488, see tab. A.1, and were computed from six simulations resulting in a total of 105 µs simulation time. Those based on Charmm36M/OPC were $(21 \pm 7)$ µs for Q482 and $(13 \pm 3)$ µs for Q488, see tab. A.1, and were computed from six simulations resulting in a total of 88 µs simulation time. All four decay times were provided by Gabor Nagy and calculated from the same trajectories using the same method that we used.

The predicted decay times of equal dq mutants based on a99SB-disp and Charmm36M/OPC fields were similar to each other, indicating that the dynamics of full-length $N_{TAIL}$ peptides behaved similar across both force fields. In addition to that, similar to the PET experiments, the full-length $N_{TAIL}$ simulations predict a significantly slower decay time for Q482 than for Q488 using both force fields, indicating the force fields captured the different behaviour of dq mutants observed in the PET experiments.

In summary, both force fields captured the different dynamics between both dq mutants of full-length $N_{TAIL}$ peptides reliably, resulting in consistently differing decay times between both mutants. Therefore, a reason for the different decay times can still be the inaccuracy of the used a99SB-disp and Charmm36M/OPC force fields.

Hence, we investigated if the different decay times obtained from the simulations and experiments were caused by inaccuracies of the simula-

tion. For this, we tested the simulation accuracy by comparing the observed decay times of full-length $N_{TAIL}$ variants obtained from the PET experiments with those computed from full-length $N_{TAIL}$ simulations.

Surprisingly, all predicted decay times calculated from full-length $N_{TAIL}$ simulations were significantly greater than those based on the PET experiments, indicating that the simulations might be inaccurate.

However, a reason why the decay times calculated from full-length $N_{TAIL}$ simulations were greater than those obtained from PET experiments, might lie in the accuracy of the decay curve fitting to obtain the decay times. Currently, the decay curve fitting is done differently between the experiments and simulations: The quenching process has multiple pathways and fitting a single exponential decay to multi-exponential processes may cause an increased decay time.

The potential inaccuracy of the decay curve fitting combined with the consistently observed similar dynamics of full-length $N_{TAIL}$ variants across different force fields, led us to the assumption that the simulations were accurate.

Therefore, the different decay times based on the simulations and experiments were unlikely caused by the used a99SB-disp force field.

Next, with the simulation accuracy verified, we were able to investigate if the truncation cuts down the decay times of $N_{TAIL}$. For this, we compared the decay times obtained from the truncated $N_{TAIL}$ simulations with those obtained from the full-length $N_{TAIL}$ simulations.

As expected, the decay times obtained from truncated $N_{TAIL}$ variants were significantly smaller than those obtained from the full-length $N_{TAIL}$ variants. This is likely due to the volume exclusion effect. The volume exclusion effect describes the effect, that occupied space of one part of a sequence cannot be occupied by another part of the sequence. [34] As a result, the additional 76 and 6 residues attached at the N- and C-terminus, respectively, of the truncated $N_{TAIL}$ occasionally blocked the path between the dq variants of the full-length $N_{TAIL}$, causing greater decay times compared to those of the truncated $N_{TAIL}$.

Consequently, the consistently increased decay times obtained from the full-length $N_{TAIL}$ peptides compared to those of the truncated $N_{TAIL}$ peptides suggested that the truncation cuts down the decay times of $N_{TAIL}$ significantly.

Lastly, to find out if the truncation of $N_{TAIL}$ also caused the more similar decay times for both dq mutants of the truncated $N_{TAIL}$ peptides, we examined the decay times ratios of Q482 and Q488 of the introduced peptides. The ratios of the full-length $N_{TAIL}$ using a99SB-disp and Charmm36M/OPC were similar being 1.87 and 1.61, respectively, but the ratios based on the truncated $N_{TAIL}$ simulations using a99SB-disp and Charmm36M/OPC differed being 1.13 and 3, respectively.

However, because the ratios obtained from the full-length $N_{TAIL}$ simulations were similar across both force fields on extensive sampling, we expected that the ratio of the truncated $N_{TAIL}$ using Charmm36M/OPC will be similar to the ratio of the truncated $N_{TAIL}$ using a99SB-disp on extensive sampling. Hence, more similar decay times likely result from the truncation of $N_{TAIL}$.

Therefore, the truncation of $N_{TAIL}$ likely not only caused the smaller but also the more similar decay times of the truncated $N_{TAIL}$ compared to those of the full-length $N_{TAIL}$.

This led to the question why the decay times of truncated $N_{TAIL}$ variants were more similar. To investigate this question, we compared the decay time ratios of the truncated and full-length $N_{TAIL}$ variants to those of a random coil. Vaiana *et al.* [8] estimated decay times $\tau(N) = aN^{2/3}$ of random coils using simple homopolymer models, where $N$ is the number of residues in the coil between the dye and quencher and $a$ is a free parameter. [5,8,9]

Surprisingly, the ratios of the full-length $N_{TAIL}$ simulations only matched the trend of the ratio 2.45 based on the PET experiments, but were more similar to the ratio 1.32 obtained from the homopolymer model used by Vaiana *et al.* [8], see dashed black line in fig. 3.1.

The greater similarities between the ratios obtained from full-length $N_{TAIL}$ variants with those obtained from random coil-like peptides indicated that the contact dynamics of simulated full-length $N_{TAIL}$ peptides were more similar to those of a random coil, than to those of the $N_{TAIL}$ peptides in the PET experiments.

Interestingly, the ratios of the truncated $N_{TAIL}$ showed even greater similarities to the ratio of the homopolymer model than the full-length $N_{TAIL}$ ratios, indicating that the truncated $N_{TAIL}$ behaved more similar to a random coil than the full-length $N_{TAIL}$. Unfortunately, the free parameter $a = 17.48 \cdot 10^{-9}$ of the decay time estimate used by Vaiana *et al.* was adapted to the length of the full-length $N_{TAIL}$ and therefore were not applicable to the truncated $N_{TAIL}$.

Hence, to compare the decay time ratio of the truncated $N_{TAIL}$ to the ratio of a random coil-like peptide, we compared the decay time ratios of the truncated $N_{TAIL}$ to those of the homopolymer models adapted to the length of the truncated $N_{TAIL}$.

For this purpose, we first carried out extensive simulations of two sets of homopolymers adapted to the length of the truncated $N_{TAIL}$, see ch. 2.7. In both sets tryptophan (W) was used as the dye and cysteine (C) as the quencher quencher at the termini of the sequence to ensure the comparability to the $N_{TAIL}$ simulations. In between, $n$ or $m$ repeats of alanine (A), glycine (G) and glutamine (Q) are used to mimic random coil-like behaviour. The first set of (AGQ) simulations (C-$(AGQ)_n$-W) solely mimic random coil-like sequences of different end-to-end dye to quencher distances. In the second set (C-$(AGQ)_5$-S-$(AGQ)_m$-W) we inserted an additional serine (S) after five AGQ repeats, which we utilised as a quencher in order to probe the influence of excluded volume in a random coil. Next, we calculated the decay times of all homopolymers from the simulations as described in ch. 2.2, and lastly, we fitted the decay time estimate $\tau_i(N) = a_i N^{3/2}$ to the obtained decay times of the $i$-th set individually.

The fitting yielded $a_1 = 1.08 \cdot 10^{-9}$ and $a_2 = 3.32 \cdot 10^{-9}$ to the estimated decay times of the random coil-like peptides as denoted by the purple

and pink dashed lines in fig 3.1, respectively. As expected, the estimate of the homopolymer model which was adapted to random coil-like peptides subjected to an excluded volume was greater than the estimate of the homopolymer model based on random coil-like peptides which were not subjected to an excluded volume.

To clarify if our random coil-like peptides behaved similar to those Vaiana *et al.* used to derive their estimate, we compared our fits to the fit of Vaiana *et al.* [8]

As expected, due to the use of the same fitting method, the ratio 1.26 obtained from the homopolymer model adapted to pure random coils and the ratio 1.30 obtained from the model adapted to random coils subjected to an excluded volume, matched to the ratio 1.32 obtained from the estimate used by Vaiana *et al.*

However, both of our decay time estimates fell drastically below the estimate used by Vaiana *et al.*, see black dashed line in fig 3.1. Hence, their model was not able to correctly estimate the decay times for a random coil-like peptide of similar lengths to the truncated $N_{TAIL}$. This highlighted the necessity to obtain appropriate decay time estimates which yielded decay times that were similar to those of the truncated $N_{TAIL}$ in absolute values.

To find out to what extend the contact dynamics of the truncated $N_{TAIL}$ agreed with the dynamics of a random coil of similar length to the truncated $N_{TAIL}$, we compared the decay times of the truncated $N_{TAIL}$ to those obtained from both of the homopolymer models.

The estimated decay times for similar lengths to Q482 and Q488, based on the pure random coil-like peptides, were 0.19 µs and 0.24 µs, respectively. Both decay times were smaller, but within the $2\sigma$-interval of the truncated $N_{TAIL}$ decay time times, agreeing in magnitude.

In contrast, the Q482 decay time estimated by the homopolymer model used by Vaiana *et al.* lay within the $15\sigma$-interval of the Q482 decay time obtained from the truncated $N_{TAIL}$ peptides and the Q488 decay time lay within the $7\sigma$-interval of the Q482 decay time obtained from the trun-

cated $N_{TAIL}$ peptides. Due to the high the estimates obtained from our homopolymer models seem more appropriate.

The Q482 and Q488 decay time estimates based on peptides subjected to an excluded volume were 0.71 µs and 0.54 µs, respectively. Surprisingly, both decay times match those of the truncated $N_{TAIL}$, indicating that it showed random coil-like contact dynamics.

However, we observed two inconsistencies. First, Q488 was affected by an excluded volume, but Q488 large uncertainty showed that its decay time also might agreed better with the estimate based on pure random coils. Second, Q482 was not affected by an excluded volume but only matched the decay time estimate based on random coil-like peptides subjected to an excluded volume.

Hence, these contradictions suggested phenomena within the sequence influencing the decay times, which the comparison between the decay times obtained from truncated $N_{TAIL}$ peptides and those obtained from homopolymer models cannot reveal.

In summary, the comparison of decay times calculated from truncated and full-length $N_{TAIL}$ peptides suggested that the truncation cuts down the decay times of $N_{TAIL}$ significantly, and made them more similar to each other. Additionally, the comparison of decay times calculated from truncated $N_{TAIL}$ peptides and from homopolymer models indicated decay time influencing processes within the truncated $N_{TAIL}$ sequence, which the decay time analysis cannot reveal.

## 3.2. Conformational state analysis

To better understand the conformational dynamics that give rise to the predicted PET decays times, we projected all truncated $N_{TAIL}$ simulations onto a free-energy landscape along the two dq distances shown in fig. 3.2. The free-energy landscape showed a diagonally extended energy well with a single free-energy minimum shown by the yellow 0.5 kT area.

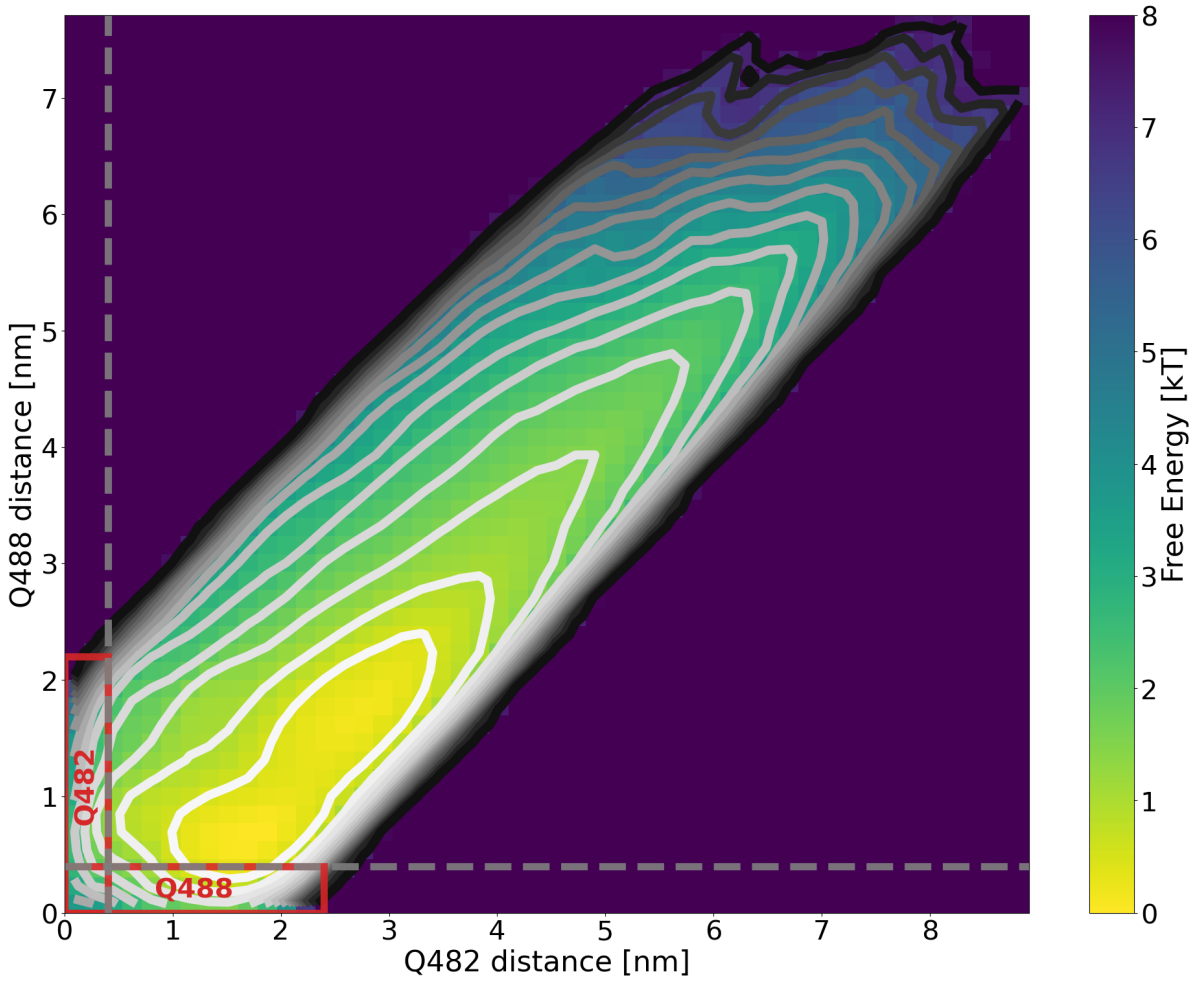


**Figure 3.2:** The free energy landscape of the truncated $N_{TAIL}$ variants as a projection of the simulation trajectories onto the dq distances. The colours from yellow to purple colour code the density of conformations from densest, the energy minimum, to sparsest, the energy maxima. Grey contour lines show a rise in free energy of 0.5 kT, black dashed lines show 0.4 nm dq distance cutoff and the red boxes show the selected contact frames of Q482 and Q488.

To begin with, we focused on how the most adopted conformational states in the free-energy minimum potentially affect the decay times. Here, we observed three interesting things:

As expected, the free-energy minimum is shifted towards smaller Q488 distances, slightly down from the principal diagonal. This is likely due to Q488 having had a six residue shorter dq separation compared to Q482,

clearly favouring shorter dq distances for Q488 in general.

Secondly, the location of the free-energy minimum revealed that the most probable conformational states were partially collapsed, corresponding to dq distances between 0.3 nm and 3.5 nm. Surprisingly, because the lower boarder of the free-energy minimum fell below the dq contact cut-off distance of 0.4 nm, see grey dashed lines, the free-energy minimum already contained Q488 contact conformations, but no Q482 contact conformations.

Lastly, because we simulated the truncated $N_{TAIL}$ variants in a temperature of 300 K, transitions between conformations of energies of up to 1 kT were rather freely available. Therefore, an even greater amount of Q488 contact conformations became available around the free-energy minimum, but still no Q482 contact conformations were as freely available as Q488 contact formations.

Combining these three observations shows that Q488 contact conformations were heavily favoured over Q482 contact conformations, which suggested that Q488 decay times were likely smaller than decay times of Q482. However, the averaged Q488 decay time were indistinguishable from the averaged Q482 decay time. This is likely because small decay times do not depend solely on high probabilites for the dye and quencher to be in a contact conformation, but rather depend on high transition rates between non-contact and contact conformations. Nonetheless, the larger uncertainty of the Q488 decay time compared to the uncertainty of the Q482 decay time suggested the possibility of faster decay times for Q488 compared to Q482.

## 3.3. Contact map analysis

To investigate how the internal structure affected the predicted decay times of the truncated $N_{TAIL}$, we focused on the adopted internal structure in its contact conformations. For this, we first collected all contact conformations marked inside the red rectangles in fig. 3.2 for Q482 and Q488 separately, and then analysed which internal structures were adopted in

these collections.

To reveal secondary structure we calculated $C_\alpha$ backbone contact maps, and to reveal salt-bridges we calculated contact maps between opposite charged residues, as described in ch. 2.5. Contact maps provided us a visualisation of contact probabilities as a $N \times N$ matrix of a $N$ residue long peptide, revealing more prominently adopted contacts.

In contact maps, lines of increased $C_\alpha$ contact probability 3, 4 and 5 residues adjacent to principal diagonal (close-range) are indicative for a 3-, 4- and 5-turn, respectively. Consecutive 3-, 4- and 5-turns can therefore be identified as helix formation. (Anti-)parallel lines of increased $C_\alpha$ contact probability more than five residues from the principal diagonal away (mid- and long-range) are indicative of (anti-)parallel $\beta$-sheets. Therefore, we were able to identify secondary structure from the $C_\alpha$ contact maps.

The contact maps for the Q482 and Q488 collections are shown in fig. 3.3 left and fig. 3.4l left, respectively. In each map, the $C_\alpha$ contact map is displayed in the upper diagonal and the salt-bridge contact map in the lower diagonal.

We knew locations of internal structures but not how they affected the predicted decay times. In order to quantify the impact of internal structures on dq distances, we computed normalised pointwise mutual information (NPMI) between $C_\alpha$ contacts and dq contacts as well as between salt-bridge contacts and dq contacts, using the concatenated trajectory of all truncated $N_{TAIL}$ simulations as described in ch. 2.6. Here, NPMI quantified the amount of information obtained about whether $C_\alpha$ or salt-bridges were in contact or not by observing whether the dq were in contact or not and vice versa. As a result, we obtained a correlation valued from -1 (strong hindering influence) over 0 (no influence) to 1 (strong promoting influence) for each contact pair.

Lastly, we visualised the correlations for the Q482 and Q488 collections as $N \times N$ matrices in fig. 3.3 right and fig. 3.4 right, respectively. In each matrix, the correlations between $C_\alpha$ contacts and dq contacts are displayed in the upper diagonal and the correlations between salt-bridge contacts and dq contacts in the lower diagonal.

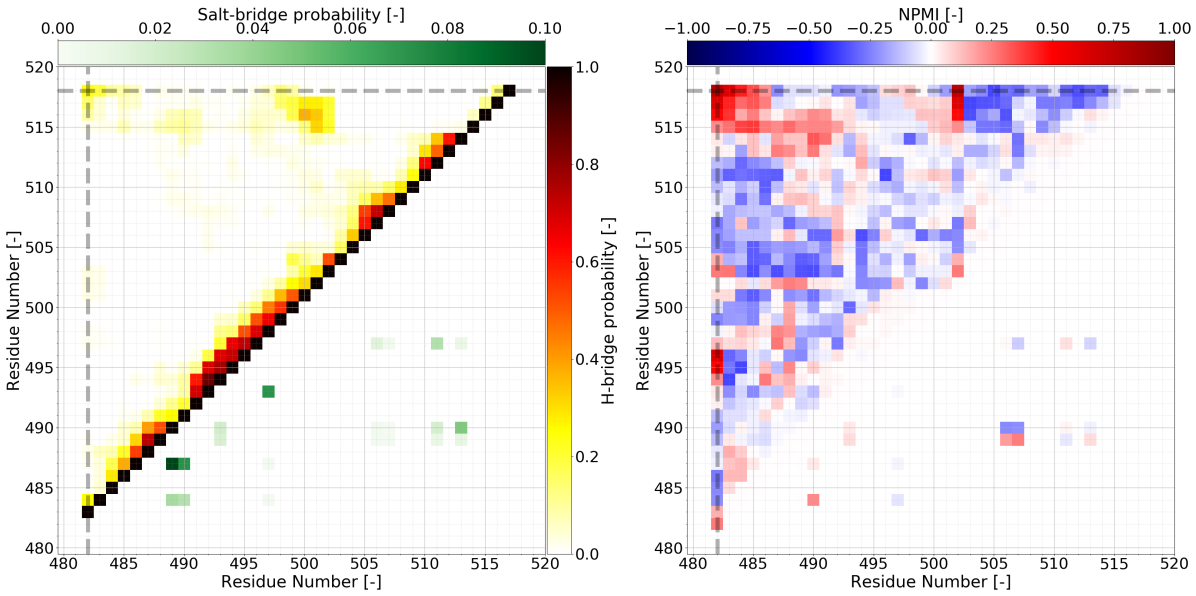


**Figure 3.3:** The left plot shows a contact map of all Q482 contact conformations of the truncated $N_{TAIL}$, in dependence of the residue number. The pixels show probabilites of backbone $C_\alpha$ interactions from low to high, colour coded from white to black, in the upper diagonal. In the lower diagonal the pixels show probabilites of salt-bridge interactions from low to high, colour coded from white to dark green. Both contact cutoff distances were set to 0.6 nm. The right plot shows the NPMI between backbone $C_\alpha$ contact (upper diagonal) and salt-bridge formation (lower diagonal) with Q482 contact conformations, in dependence of the residue number. The NPMI is colour coded from dark blue over white to dark red, where dark blue show a strong negative correlation, red a strong positive correlation and white no correlation whatsoever. In both plots the grey dashed lines show the location of the dye and the quencher. Because individual contacts occur pairwise the plots can be mirrored at their principal diagonal.

With the contact maps and correlation matrices obtained, we were able to analyse internal structure and their correlations with dq contact conformations.

For this, we first analysed the short-range interactions of Q482, see fig. 3.3. The black diagonal lines in fig. 3.3 left showed that the $C_\alpha$ of each amino acid in the truncated $N_{TAIL}$ were within the range of 0.6 nm to its neighbours in all selected frames. The helical structures along the principal diagonal were expected due to the transiently helical $\alpha$-MoRE region, and turn structures may have been necessary to bring the dye and

the quencher into close proximity. However, in the case of Q482 contact conformations, neither the helices, nor the three prominent $\beta$-turns were correlated strongly with the presence of dq contacts.

Nonetheless, there were a few $C_\alpha$ interactions accompanied by sparse salt-bridge occurrences at the N-terminus which promoted Q482 contact conformations, as their locations indicated a bend of the quencher towards the dye.

Next, we analysed the mid- and long-range interactions of Q482. In the vicinity of the Q482 contact area, where the dashed lines cross, the contact map of fig. 3.3 showed that nearby $C_\alpha$ contact only occurred in up to 20 %. The low frequency of interactions between the $C_\alpha$ of the dye and the quencher showed that their backbones did not necessarily have to come in contact when their side chains did. The NPMI of this area showed a high positive correlation up to 0.8 between the occurrence of $C_\alpha$ contacts of nearby residue pairs and Q482 contact conformations. Hence, it was likely that once the $C_\alpha$ of nearby residues interact, the dye and quencher were in close proximity in the first place.

The raised correlations up to 0.4 between the $C_\alpha$ interactions around the residue pair 490/515, and the diagonal running group of $C_\alpha$ interactions between the residues 487–492 and 505–513, which distinguished from the residual area having small positive correlations up to 0.2, which likely occurred for the same reason. Because these $C_\alpha$ interactions resembled a diagonal line directly between the dq contact area and the middle of the sequence, these H-bonds were more likely to be a consequence of a hairpin-like conformation than a cause of Q482 contacts.

Similar seemed to be valid for the $C_\alpha$ interactions between residue pairs adjacent to the residue pair 482/495, which reached correlations up to 0.8. Because these $C_\alpha$ interactions were not noticeable in the contact map, it is more likely they occurred randomly when Q482 was in contact in the first place.

Moreover, fig. 3.3 left shows a region of $C_\alpha$ interactions with an increased probability of up to 40% above the residue pair 500/515, reaching correlations up to 0.8. This region indicated a more prominently occurring

Q482 contact conformation in which the C-terminus was folded in half.

Interestingly, salt-bridges between residues 489–490 and residues 511, 513 as well as 506–507 showed either moderately strong contrary correlations up to ±0.2 but occurred in less than 1%, or slight contrary correlations of less than ±0.05 and occurred more frequent up to 5%, indicating they did not have much impact on Q482 contact formation.

Then, we analysed close-range $C_\alpha$ and salt-bridge interactions and their correlations observed for Q488 contact conformations in fig. 3.4.

The $C_\alpha$ and salt-bridge interaction probabilities as well as their correlations with Q488 contacts along the principal diagonal were mostly in accordance with those of Q482 in location and magnitude. This accordance is in line with the findings of Vaiana *et al.*, as it entails neither an advantage for Q482 nor for Q488 contact formations and therefore is not the reason for the different observed Q482 and Q488 decay times in the PET experiment. [8]

Lastly, we analysed the long-range interactions in the collection of Q488 contact conformations. The Q488 contact area, see crossed dashed lines in fig. 3.4 left, showed $C_\alpha$ interaction probabilites of up to 40% and correlations up to 0.8. Similar to the contact area of Q482, the presence of Q488 contacts likely entailed the close proximity of the $C_\alpha$ of residues in the vicinity of the dye and quencher and vice versa.

Additionally, there were $C_\alpha$ interactions with probabilities up to 20% and correlations up to 0.6 starting adjacent to the right of the Q488 contact area running diagonally down to the principal diagonal. This course of $C_\alpha$ interactions was indicative for a set of anti-parallel $C_\alpha$ contacts, possibly indicating the presence of an anti-parallel $\beta$-sheet. Depending on the location of the anti-parallel $C_\alpha$ contacts, they could not only folded the C-terminus in half, but also folded the C-terminus onto the N-terminal half, reducing the dye and quencher distance drastically.

Because the course of the assumed anti-parallel $\beta$-sheet in the Q482 NPMI map, show mainly none to negative correlations for Q482 contact

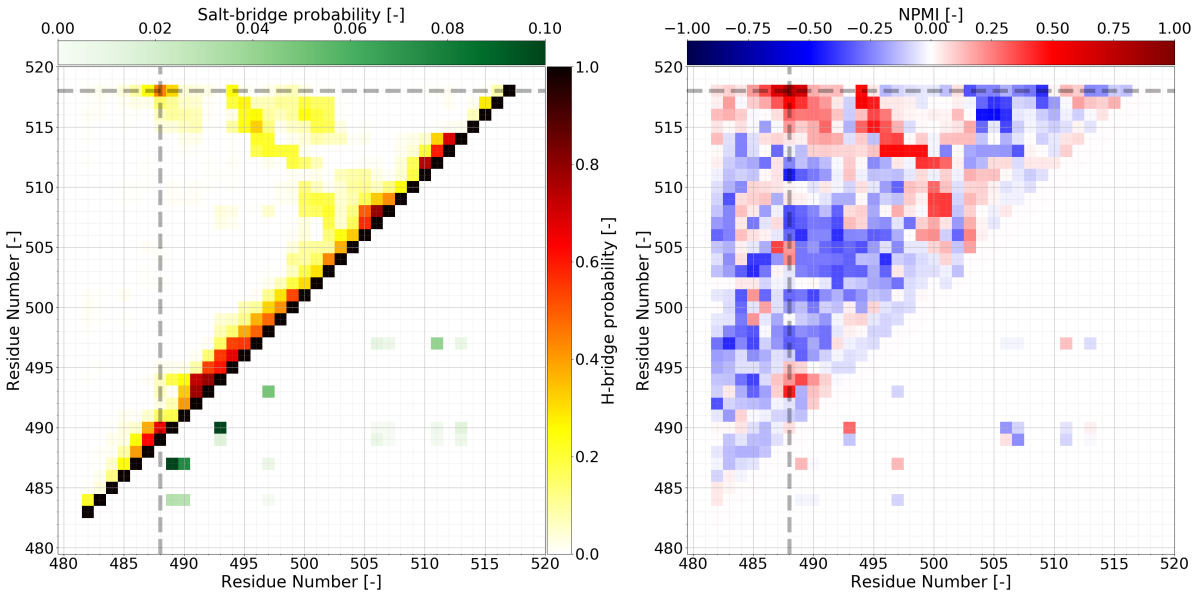


**Figure 3.4:** The left plot shows a contact map of all Q488 contact conformations of the truncated $N_{TAIL}$, in dependence of the residue number. The pixels display probabilites of backbone $C_\alpha$ interactions from low to high, colour coded from white to black, in the upper diagonal. In the lower diagonal the pixels represent probabilites of salt-bridge interactions from low to high, colour coded from white to dark green. Both contact cutoff distances were set to 0.6 nm. The right plot displays the NPMI between backbone $C_\alpha$ contact (upper diagonal) and salt-bridge formation (lower diagonal) with Q488 contact conformations, in dependence of the residue number. The NPMI is colour coded from dark blue over white to dark red, where dark blue indicate a strong negative correlation, red a strong positive correlation and white no correlation whatsoever. In both plots the grey dashed lines indicate the location of the dye and the quencher. Because individual contacts occur pairwise the plots can be mirrored at their principal diagonal.

conformations, the comparison reveals an Q488 contact conformation promoting conformational state.

The anti-parallel $C_\alpha$ contacts were accompanied by salt-bridges between the residue 499 and residues 511 and 513 with probabilities up to 4%, which show slight positive and negative correlations, respectively.

The probability of salt-bridges between residues 489–490 and residues 511, 513 as well as 506–507 decreased compared to Q482, and further showed mainly negative correlations, disclosing that the presence of these salt-bridges was more likely hindering Q488 contact formations.

In summary, the contact analysis revealed unexpectedly high correlations between two $C_\alpha$ contact regions and dq contacts. Firstly, the $C_\alpha$ contact region around residues 500 and 515 which correlated with Q482 contacts and secondly, the $C_\alpha$ contact region along the residues 494–504 and 505–518 which correlated with Q488 contacts. Because both $C_\alpha$ contact regions also occurred frequently, they seem to played an important role in quenching.

The diagonal course of the $C_\alpha$ contacts between residues 494–504 and 505–518 towards the principal diagonal was suggestive for an anti-parallel $\beta$-sheet.

Additionally, salt-bridges occurred near the quencher positions and were positively correlated with dye-quencher contact formation. But because salt-bridges occurred at rather low probability overall, they seem to played a less important role for quenching than the two mentioned $C_\alpha$ contact regions.

Lastly, the contact analysis showed that either minor or no correlations between close-range $C_\alpha$ interactions and dq contacts conformations along the $\alpha$-MoRE region. Hence, close-range $C_\alpha$ interactions within the $\alpha$-MoRE region insignificantly influenced quenching and therefore the transient $\alpha$-MoRE helix has insignificant influence on the decay times of Q482 and Q488. This insignificant influence on both decay times is in line with the hypothesis of Vaiana *et al.* which suggested that the transient $\alpha$-MoRE helix does not cause the different Q482 and Q488 decay times.

## 3.4. Secondary structure analysis

The contact analysis showed high correlations between the $C_\alpha$ contact region around residues 500 and 515 with Q482 contacts and between the $C_\alpha$ contact region along the residues 494–504 and 505–518 with Q488 contacts, suggesting that both regions may play an important role in Q482 and Q488 quenching, respectively. However, the contact analysis could not determine the influence of these regions on the decay times of the

truncated $N_{TAIL}$, as the analysis only provided a limited and static insight into the secondary structure of dq contact conformations.

In order to investigate influence of regions with high correlations on quenching dynamics, we inspected the time course of secondary structure formation and dq distances in our three longest truncated $N_{TAIL}$ simulations.

For this, we computed the secondary structure for all conformations during the course of the three simulations using the DSSP algorithm by Kabsch and Sanders provided in SESCA by Gabor Nagy, as described in ch. 2.4, and used the minimum distances we obtained during the decay time calculation, as described in ch. 2.2. [27,29] In addition, to get an overview over the spatial expansion of the observed truncated $N_{TAIL}$ peptide, we calculated the radius of gyration for all conformations during the course of the three simulations, as described in ch. 2.3.

Figure 3.5 illustrates the observed time correlations for the three longest simulation trajectories. In the first column of fig. 3.5 we show the analyses of the first simulation, which was started from a fully unfolded conformation. As seen in fig. 3.5 a) the truncated $N_{TAIL}$ peptide adopted neither stable helices (red) nor stable $\beta$-structures (blue) during the simulation. Due to the lack of both structures, the peptide was able to visit a wide range of collapsed and expanded conformations resulting in greater fluctuations of dq distance and radii of gyration, see fig. 3.5 b) and c), respectively.

Highlighted in yellow are two sections with continuously low dq distances. In the first section, between 1 µs and 2.5 µs, the peptide adopted a moderately compact conformation with two short $\beta$-sheets on the C-terminus (495/515) and (502/505). It allowed frequent Q488 contacts with only a few contacts for Q482.

The second section, between 4 µs and 6 µs, included one beta sheet (495/515) and two shorter helices (500–505 and 485–490). This conformation was even more compact and it reduced Q482 distances further, although allowed a larger variation of Q488 distances.

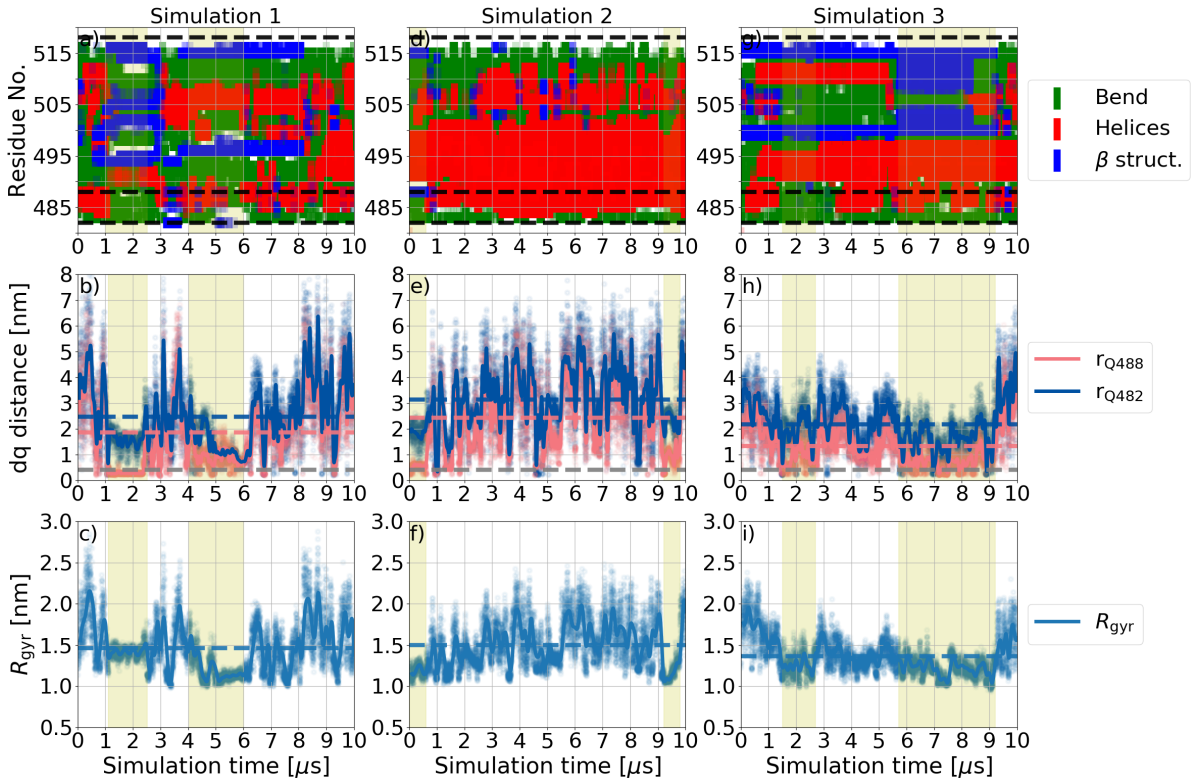


**Figure 3.5:** Comparison of the secondary structure (top), radius of gyration (middle) and dq distance (bottom) of our three longest truncated $N_{TAIL}$ simulations in each column. In the first row on the secondary structure of each residue is shown as a function of simulation time. Here, the black dashed horizontal lines show the positions of the quenchers (Q482 bottom and Q488 middle) and the dye (top). Helical structures (3-, 4-, 5-helices) are shown in red, $\beta$-structures ($\beta$-strands and -sheets) in blue and bends and turns in green. In the middle row the time series of the dq distances is shown in marine for Q482 and in magenta for Q488, and their corresponding average is shown as a like coloured horizontal dashed line. A horizontal black dashed line shows the contact cutoff distance $r_{cut} = 0.4\,\text{nm}$. The bottom row shows the time series of the radii of gyration in blue. Again, the averages are shown as a horizontal dashed like coloured lines. Along the columns regions with decreased dq distances are highlighted in yellow.

In the middle column of fig. 3.5 we show the analyses of the second simulation. Throughout the first 0.6 µs, highlighted in yellow, the peptide deviated little from its starting conformation which contained a $\beta$-structure (488/515) and a helix (491–496), causing the Q488 distances to border the dq cutoff distance.

Subsequently, the peptide formed a helix (485–500) but the C-terminus

only had a transient helix structure and was often unfolded. This state allowed a lot of conformational freedom for the termini, leading to a large variation in dq distances and radii of gyration. The large variation still allowed occasional dq contacts for both mutants, but these contacts were short-lived and only occurred rarely.

However, despite the presence of this stable helix (485–500) the peptide was able to adopt consecutive small dq distances between 9.1 µs and 9.9 µs, highlighted in yellow.

The right column of fig. 3.5 shows the analyses of the last simulation. The secondary structure analysis revealed that the peptide maintained a $\beta$-structure (500/515) and a relatively stable N-terminal helix (490–498) nearly throughout the entire simulation. The increased secondary structure content occurred simultaneously with considerably low values and low fluctuations of the dq distances and radii of gyration.

Most contacts occurred during the sections marked in yellow. The first section started 1.5 µs and ended at 2.8 µs when the initially folded shorter transient helices (485–488 and 504–506) unfolded. The second section, between 5.8 µs and 9.2 µs, was after a relatively stable shorter helix (510–513) unfolded, which allowed $\beta$-structure (500–504/503–515) to elongate, while the stable N-terminal helix simultaneously elongated transiently to residue 485. These elongations resulted in a compact conformational state which allowed frequent Q488 and occasional Q482 contacts.

In summary, the analysis of the time course of secondary structure formation showed that stable ($\geq 0.5$ µs) Q488 contact distances were mostly accompanied by $\beta$-structure formation along the residues 494–504 and 505–518, confirming that the observed $C_\alpha$ interactions are indeed anti-parallel $\beta$-bridges and -sheets, observed in fig. 3.4.

In stable Q488 contact distances, the anti-parallel $\beta$-bridges between the residues 494–504 and 505–518 increased Q488 contact probabilities considerably, causing numerous Q488 quenching contacts on nanosecond scale. This promoting behaviour is in line with the positive correlations of $C_\alpha$ interactions between the $\beta$-bridges and Q488 contact formations,

seen in fig. 3.4 right.

To check if the truncated $N_{TAIL}$ is more likely to adopt a stable secondary structure than a random coil-like peptide, we investigated potential differences in secondary structure frequencies between the two. For this we calculated the average secondary structure probabilities of all truncated $N_{TAIL}$ simulations and compared them to the average probabilities of the C-$(AGQ)_5$-S-$(AGQ)_5$-W ($AGQ_{10}$) simulations.

In this comparison we only used $AGQ_{10}$ because it has a length of 33 residues and resembled the length of 38 residues of the truncated $N_{TAIL}$ of all $(AGQ)_n$ simulations best. Additionally, by choosing a random coil-like sequence simulated in the exact same force field as the truncated $N_{TAIL}$, we ensured all resulting differences could be traced back to the sequence and following its behaviour during the simulation.

**Table 3.1:** Overview of the probability to form helices, $\beta$-structures, bends or unclassified secondary structure of the truncated $N_{TAIL}$ and $AGQ_{10}$. The secondary structure is calculated by DSSP implemented in SESCA. [29]

| Peptide/Model | Helices | $\beta$-structures | Bends | Unclassified |
|---|---|---|---|---|
| $N_{TAIL}$ | $(20 \pm 1)\,\%$ | $(3 \pm 1)\,\%$ | $(26 \pm 1)\,\%$ | $(51 \pm 1)\,\%$ |
| $AGQ_{10}$ | $(14 \pm 1)\,\%$ | $(5 \pm 1)\,\%$ | $(36 \pm 1)\,\%$ | $(45 \pm 1)\,\%$ |

As shown in tab. 3.1, the overall well-defined secondary structure probability computed to $(23 \pm 2)\,\%$ for the truncated $N_{TAIL}$ and to $(18 \pm 2)\,\%$ for $AGQ_{10}$, indicating that $AGQ_{10}$ and truncated $N_{TAIL}$ peptides adopted well-defined secondary structure fairly similar.

Interestingly, the truncated $N_{TAIL}$ had a significantly increased probability to adopt helices $(20 \pm 1)\,\%$ compared to $AGQ_{10}$ $(13 \pm 1)\,\%$, which was in line with the considerable helix contents, seen for the truncated $N_{TAIL}$ in fig. 3.5. This was expected, because the truncated $N_{TAIL}$ peptide includes the $\alpha$-MoRE region with an increased helix propensity. [8]

Furthermore, the truncated $N_{TAIL}$ formed $\beta$-structures in only $(3 \pm 1)\,\%$ in general, hence, the promoting behaviour between anti-parallel $\beta$-bridges

and Q488 contacts found in the secondary structure analysis is expected to be small.

In summary, the secondary structure probability analysis showed a significantly increased frequency of helices for the truncated $N_{TAIL}$ compared to $AGQ_{10}$. Additionally, the truncated $N_{TAIL}$ formed $\beta$-structures in only $(3 \pm 1)\,\%$, entailing that the promoting behaviour between $\beta$-bridges along the residues 494–504 and 505–518 and Q488 contacts take effect rather rarely.

## 4. Conclusion

To better understand the complex dynamics of the measles nucleoprotein, it is important to know if the absence of electrostatic long-range interactions alter the predicted PET decay times of the truncated $N_{TAIL}$ mutant pairs Q482 and Q488.

Our predicted PET decay times indicated that the truncation of $N_{TAIL}$ significantly reduces predicted PET decay times and diminishes the considerable difference between Q482 and Q488 decay times, observed in experiments and simulations of the full-length $N_{TAIL}$, below the computational uncertainty. Because the considerable difference between both decay times diminished after the truncation, long-range interactions from outside the sequence likely caused the difference of both decay times in the full-length $N_{TAIL}$.

Interestingly, the uncertainty of the Q488 decay time was nearly doubled compared to Q482 in the truncated $N_{TAIL}$ simulations indicating different contact dynamics between the two. The conformational state analysis showed that Q488 contact conformations were significantly more frequently adopted than the Q482 conformations, underlining different contact dynamics for Q488 compared to Q482.

Moreover, the contact map analysis showed that the conformational

states with anti-parallel $\beta$-bridges along the residues 494–504 and 505–515 significantly promoted Q488 contact conformations and vice versa. The secondary structure analysis indicated that these anti-parallel $\beta$-bridges were mostly accompanied by stable periods of low Q488 dq distances on a microsecond scale. In turn, these stable periods of low Q488 dq distances were accompanied by numerous quenching events on a nanosecond scale which is indicative for small Q488 decay times.

The secondary structure frequency analysis showed that the truncated $N_{TAIL}$ peptides adopted $\beta$-structures in only $(3 \pm 1)\,\%$, overall, entailing that the promoting behaviour between $\beta$-bridges along the residues 494–504 and 505–518 and Q488 contacts took effect rather rarely. Nonetheless, because the anti-parallel $\beta$-bridges between the residues 494–504 and 505–518 increased Q488 contact probabilities considerably, the microsecond scale folding dynamics of the $\beta$-structures may explain the larger uncertainty of the predicted Q488 decay times between our MD simulations.

Either way, because the $\beta$-structures promote Q488 contact conformations, the considerable difference between Q482 and Q488 decay times observed in the PET experiments cannot solely be traced back to long-range interactions from outside the truncated sequence. Hence, it is likely that also the promoting anti-parallel $\beta$-bridges along the residues 494–504 and 505–518 play an important role in causing this difference.

The contact map analysis showed no significant correlations between helix formation and dq contacts. The missing correlations were in line with the hypothesis that the transient $\alpha$-MoRE helix does not cause the considerably different Q482 and Q488 decay times observed in PET experiments of full-length $N_{TAIL}$ peptides, by Vaiana *et al.*

However, the secondary structure analysis indicated truncated $N_{TAIL}$ peptides with high helical content established dq contacts less frequently than those with less helical content. Additionally, the secondary structure frequency analysis showed that the truncated $N_{TAIL}$ peptides adopted helices significantly more often than random coils. Hence, it is likely that the increased adoption of helical structures increases the rigidity of the truncated $N_{TAIL}$, entailing fewer loop formations between the dye and

quenchers, which in turn cause greater decay times.

Interestingly, the predicted PET decay times obtained from $\mathrm{AGQ}_n$ simulations showed that the decay times of truncated $\mathrm{N_{TAIL}}$ peptides were greater than the decay time estimates obtained from a homopolymer model which was adapted to random coil-like peptides. In fact, both decay times of the truncated $\mathrm{N_{TAIL}}$ match the decay time estimates of a homopolymer model which was adapted to random coil-like peptides subjected to volume exclusion.

The increased decay times of truncated $\mathrm{N_{TAIL}}$ peptides compared to random coil-like peptides support the hypothesis that the increased frequency to adopt helices of truncated $\mathrm{N_{TAIL}}$ peptides likely increased both decay times compared to random coil-like peptides.

All together, Vaiana *et al.* observed that the Q482 decay time of full-length $\mathrm{N_{TAIL}}$ peptides was greater than estimated based on a random coil-like peptide and that the Q488 decay time of full-length $\mathrm{N_{TAIL}}$ peptides was smaller than estimated based on a random coil-like peptide.

Our study indicated that Q482 and Q488 decay times of truncated $\mathrm{N_{TAIL}}$ peptides were increased such they match the decay time of a random coil-like peptide subjected to volume exclusion, due to the truncated $\mathrm{N_{TAIL}}$ propensity to adopt helical structures. Additionally our study indicated that the Q488 decay time of truncated $\mathrm{N_{TAIL}}$ peptides got promoted by anti-parallel $\beta$-bridges along the residues 494–504 and 505–518 and vice versa, potentially decreasing the Q488 decay time to match the decay time of a random coil-like peptide.

Therefore, to match the results observed by Vaiana *et al.* further Q488 decay time decreasing impact is needed. Here, a missing link could be the long-range interactions which we excluded by truncating $\mathrm{N_{TAIL}}$.

# A. Decay time comparison

Table A.1 provides an overview of all decay times of Q482 and Q488 as well as Q482/Q488 ratios mentioned in this work.

As in fig. 3.1, $\tau_{\mathrm{a99}}^{\mathrm{full}}$ and $\tau_{\mathrm{C36}}^{\mathrm{full}}$ represent the decay time of full-length $\mathrm{N_{TAIL}}$ peptides simulated in the a99SB-disp and Charmm36M/OPC force fields, respectively, obtained and provided by Gabor Nagy from ongoing research, $\tau_{\mathrm{exp}}^{\mathrm{full}}$ represent the observed decay rate obtained in PET experiments from and $\tau_{\mathrm{exp}}^{\mathrm{mdl}}$ the used homopolymer model by Vaiana *et al.* [8]

The decay times of truncated $\mathrm{N_{TAIL}}$ variants simulated in the a99SB-disp and Charmm36M/OPC force fields are represented by $\tau_{\mathrm{a99}}^{\mathrm{trnc}}$ and $\tau_{\mathrm{C36}}^{\mathrm{trnc}}$, respectively, where the latter again was obtained and provided by Gabor Nagy from ongoing research.

Lastly, $\tau_{\mathrm{e2e}}^{\mathrm{mdl}}$ and $\tau_{\mathrm{vol}}^{\mathrm{mdl}}$ represent the decay time estimates of homopolymer models simulated in the a99SB-disp force field without and with the influence of volume exclusion. respectively.

**Table A.1:** Overview of the decay times of Q482 and Q488 of all introduced peptides and models.

| Decay time | Q482 [µs] | Q488 [µs] | Ratio [µs] |
|---|---|---|---|
| $\tau_{\mathrm{a99}}^{\mathrm{full}}$ | $28 \pm 6$ | $15 \pm 6$ | 1.87 |
| $\tau_{\mathrm{C36}}^{\mathrm{full}}$ | $21 \pm 7$ | $13 \pm 3$ | 1.61 |
| $\tau_{\mathrm{exp}}^{\mathrm{full}}$ [8] | 5.94 | 2.42 | 2.45 |
| $\tau_{\mathrm{exp}}^{\mathrm{mdl}}$ [8] | 3.7 | 2.8 | 1.32 |
| $\tau_{\mathrm{C36}}^{\mathrm{trnc}}$ | $1.5 \pm 0.3$ | $0.50 \pm 0.20$ | 3 |
| $\tau_{\mathrm{vol}}^{\mathrm{mdl}}$ | 0.71 | 0.54 | 1.31 |
| $\tau_{\mathrm{a99}}^{\mathrm{trnc}}$ | $0.68 \pm 0.22$ | $0.6 \pm 0.4$ | 1.04 |
| $\tau_{\mathrm{e2e}}^{\mathrm{mdl}}$ | 0.23 | 0.18 | 1.28 |

# References


1. Patel, M. K. Progress Toward Regional Measles Elimination — Worldwide, 2000 – 2019. *MMWR. Morbidity and Mortality Weekly Report* **69.** ISSN: 0149-21951545-861X. https://www.cdc.gov/mmwr/volumes/69/wr/mm6945a6.htm (2021) (2020).
2. Jensen, M. R. *et al.* Intrinsic disorder in measles virus nucleocapsids. *Proceedings of the National Academy of Sciences of the United States of America* **108,** 9839–9844. ISSN: 0027-8424. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3116414/ (2021) (June 14, 2011).
3. Longhi, S. *et al.* The C-terminal domain of the measles virus nucleoprotein is intrinsically disordered and folds upon binding to the C-terminal moiety of the phosphoprotein. *The Journal of Biological Chemistry* **278,** 18638–18648. ISSN: 0021-9258 (May 16, 2003).
4. Bourhis, J.-M. *et al.* The C-terminal domain of measles virus nucleoprotein belongs to the class of intrinsically disordered proteins that fold upon binding to their physiological partner. *Virus Research* **99,** 157–167. ISSN: 0168-1702 (Feb. 2004).
5. Buscaglia, M., Lapidus, L. J., Eaton, W. A. & Hofrichter, J. Effects of denaturants on the dynamics of loop formation in polypeptides. *Biophysical Journal* **91,** 276–288. ISSN: 0006-3495 (July 1, 2006).
6. Vaiana, A. C. *et al.* Fluorescence Quenching of Dyes by Tryptophan: Interactions at Atomic Detail from Combination of Experiment and Computer Simulation. *Journal of the American Chemical Society* **125.** Publisher: American Chemical Society, 14564–14572. ISSN: 0002-7863. https://doi.org/10.1021/ja036082j (2021) (Nov. 1, 2003).
7. Kunkel, J. *et al.* Dynamical Heterogeneity in the Measles Virus IDP N-TAIL in its Free and Bound States. *Biophysical Journal* **120.** Place: Cambridge Publisher: Cell Press WOS:000629601401284, 213A–213A. ISSN: 0006-3495. https://www.webofscience.com/wos/woscc/full-record/WOS:000629601401284 (2021) (Feb. 12, 2021).
8. Vaiana, A. C. *et al.* Non-local Interactions within the Disordered C-Terminal Domain of the Measles Virus Nucleoprotein. presented at the COMPUTER SIMULATION AND THEORY OF MACROMOLECULES, Huenfeld, Germany, Apr. 23, 2021 (Apr. 15, 2021).
9. Zerze, G. H., Zheng, W., Best, R. B. & Mittal, J. Evolution of All-Atom Protein Force Fields to Improve Local and Global Properties. *The Journal of Physical Chemistry Letters* **10,** 2227–2234. ISSN: 1948-7185 (May 2, 2019).
10. Lindahl, Abraham, Hess & Spoel, v. d. GROMACS 2019.6 Manual. Publisher: Zenodo. https://zenodo.org/record/3685925 (2021) (Feb. 28, 2020).

11. Robustelli, P., Piana, S. & Shaw, D. E. Developing a molecular dynamics force field for both folded and disordered protein states. *Proceedings of the National Academy of Sciences of the United States of America* **115,** E4758–E4766. ISSN: 0027-8424. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6003505/ (2021) (May 22, 2018).

12. in. *Principal Component Analysis* (ed Jolliffe, I. T.) 167–198 (Springer, New York, NY, 2002). ISBN: 978-0-387-22440-4. https://doi.org/10.1007/0-387-22440-8_8 (2022).

13. Hockney, R. W., Goel, S. P. & Eastwood, J. W. Quiet high-resolution computer models of a plasma. *Journal of Computational Physics* **14,** 148–158. ISSN: 0021-9991. (2022) (Feb. 1, 1974).

14. Hess, B., Bekker, H., Berendsen, H. J. C. & Fraaije, J. G. E. M. LINCS: A linear constraint solver for molecular simulations. *Journal of Computational Chemistry* **18,** 1463–1472. ISSN: 1096-987X. (2022) (1997).

15. Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A. & Haak, J. R. Molecular dynamics with coupling to an external bath. *The Journal of Chemical Physics* **81.** Publisher: American Institute of Physics, 3684–3690. ISSN: 0021-9606. https://aip.scitation.org/doi/10.1063/1.448118 (2021) (Oct. 15, 1984).

16. Darden, T., York, D. & Pedersen, L. Particle mesh Ewald: An N log(N) method for Ewald sums in large systems. *The Journal of Chemical Physics* **98.** Publisher: American Institute of Physics, 10089–10092. ISSN: 0021-9606. https://aip.scitation.org/doi/10.1063/1.464397 (2022) (June 15, 1993).

17. Essmann, U. *et al.* A smooth particle mesh Ewald method. *The Journal of Chemical Physics* **103.** Publisher: American Institute of Physics, 8577–8593. ISSN: 0021-9606. https://aip.scitation.org/doi/10.1063/1.470117 (2022) (Nov. 15, 1995).

18. Bussi, G., Donadio, D. & Parrinello, M. Canonical sampling through velocity rescaling. *The Journal of Chemical Physics* **126.** Publisher: American Institute of Physics, 014101. ISSN: 0021-9606. https://aip.scitation.org/doi/10.1063/1.2408420 (2021) (Jan. 7, 2007).

19. Parrinello, M. & Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied Physics* **52.** Publisher: American Institute of Physics, 7182–7190. ISSN: 0021-8979. https://aip.scitation.org/doi/10.1063/1.328693 (2021) (Dec. 1, 1981).

20. Nosé, S. & Klein, M. Constant pressure molecular dynamics for molecular systems. *Molecular Physics* **50.** Publisher: Taylor & Francis, 1055–1076. ISSN: 0026-8976. https://doi.org/10.1080/00268978300102851 (2022) (Dec. 10, 1983).

21. Zerze, G. H., Mittal, J. & Best, R. B. Diffusive Dynamics of Contact Formation in Disordered Polypeptides. *Physical Review Letters* **116,** 068102. ISSN: 1079-7114 (Feb. 12, 2016).

22. Zerze, G., Mittal, J. & Best, R. B. Diffusive dynamics of contact formation in disordered polypeptides. *Physical review letters* **116,** 068102. ISSN: 0031-9007. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5578460/ (2021) (Feb. 12, 2016).

23. Lapidus, L. J., Steinbach, P. J., Eaton, W. A., Szabo, A. & Hofrichter, J. Effects of Chain Stiffness on the Dynamics of Loop Formation in Polypeptides. Appendix: Testing a 1-Dimensional Diffusion Model for Peptide Dynamics. *The Journal of Physical Chemistry B* **106,** 11628–11640. ISSN: 1520-6106, 1520-5207. https://pubs.acs.org/doi/10.1021/jp020829v (2021) (Nov. 2002).

24. Lapidus, L. J., Steinbach, P. J., Eaton, W. A., Szabo, A. & Hofrichter, J. Effects of Chain Stiffness on the Dynamics of Loop Formation in Polypeptides. Appendix: Testing a 1-Dimensional Diffusion Model for Peptide Dynamics. *The Journal of Physical Chemistry B* **106.** Publisher: American Chemical Society, 11628–11640. ISSN: 1520-6106. https://doi.org/10.1021/jp020829v (2021) (Nov. 1, 2002).

25. Efron, B. & Tibshirani, R. Improvements on Cross-Validation: The 632+ Bootstrap Method. *Journal of the American Statistical Association* **92,** 548–560. ISSN: 0162-1459, 1537-274X. http://www.tandfonline.com/doi/abs/10.1080/01621459.1997.10474007 (2022) (June 1997).

26. Iedema, P. D. & Hoefsloot, H. C. J. Computing Radius of Gyration Distributions for Reactor Populations of Highly Random-Branched Polymers. *Macromolecular Theory and Simulations* **10,** 870–880. ISSN: 1521-3919. https://onlinelibrary.wiley.com/doi/abs/10.1002/1521-3919%2820011101%2910%3A9%3C870%3A%3AAID-MATS870%3E3.0.CO%3B2-M (2021) (2001).

27. Kabsch, W. & Sander, C. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22.** _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/bip.360221211, 2577–2637. ISSN: 1097-0282. https://onlinelibrary.wiley.com/doi/abs/10.1002/bip.360221211 (2021) (1983).

28. Nagy, G. & Grubmüller, H. Implementation of a Bayesian Secondary Structure Estimation Method for the SESCA Circular Dichroism Analysis Package. *bioRxiv.* Publisher: Cold Spring Harbor Laboratory Section: New Results, 2020.12.02.408302. https://www.biorxiv.org/content/10.1101/2020.12.02.408302v1 (2021) (Dec. 3, 2020).

29. Nagy, G., Igaev, M., Jones, N. C., Hoffmann, S. V. & Grubmüller, H. SESCA: Predicting Circular Dichroism Spectra from Protein Molecular Structures. *Journal of Chemical Theory and Computation* **15.** Publisher: American Chemical Society,

5087–5102. ISSN: 1549-9618. https://doi.org/10.1021/acs.jctc.9b00203 (2021) (Sept. 10, 2019).

30. Bouma, G. Normalized (pointwise) mutual information in collocation extraction. https://www.semanticscholar.org/paper/Normalized--mutual-information-in-Bouma/15218d9c029cbb903ae7c729b2c644c24994c201 (2009).

31. Ping, G., Dastidar, S. G. & Duan, Y. Statistical properties and kinetics of end-end contact formation of unfolded polypeptides: a systematic molecular dynamics study. *The Journal of Chemical Physics* **126,** 045108. ISSN: 0021-9606 (Jan. 28, 2007).

32. Soranno, A., Longhi, R., Bellini, T. & Buscaglia, M. Kinetics of Contact Formation and End-to-End Distance Distributions of Swollen Disordered Peptides. *Biophysical Journal* **96,** 1515–1528. ISSN: 0006-3495. https://www.sciencedirect.com/science/article/pii/S0006349508010485 (2021) (Feb. 18, 2009).

33. Norouzy, A., Assaf, K. I., Zhang, S., Jacob, M. H. & Nau, W. M. Coulomb Repulsion in Short Polypeptides. *The Journal of Physical Chemistry B* **119.** Publisher: American Chemical Society, 33–43. ISSN: 1520-6106. https://doi.org/10.1021/jp508263a (2021) (Jan. 8, 2015).

34. Hill, T. L. *An Introduction to Statistical Thermodynamics* (Dover Books, New York, US, 1987).

A year full of learning, improving and repeating come to an end.
You invested your time and effort so that I could take my first steps in science.

Thank you Gabor.

Best,
Aaron