

# Uncovering the Mechanical Coupling in SNARE Complex Disassembly

## First-Step Project (MSc)

Ida L. Neinhardt

*MSc in Molecular Life Science with a Bioinformatics Specialisation, University of Lausanne  
(UNIL), Lausanne, Switzerland*

DATE OF EXAMINATION  
Thursday, 30<sup>th</sup> May 2024 at 13:00

SUPERVISOR  
Aysima Hacisuleyman

*Department of Computational Biology, University of Lausanne (UNIL), Lausanne,  
Switzerland*

PROJECT DIRECTOR  
Dirk Fasshauer

*Department of Computational Biology, University of Lausanne (UNIL), Lausanne,  
Switzerland*

Lausanne 2024

# Abstract

Soluble N-ethylmaleimide-sensitive factor attachment protein receptors (SNAREs) are vital in eukaryotic cells where four SNARE proteins assemble into a four helical bundle called the SNARE complex and it catalyses the fusion of vesicles with lipid bilayers. After the SNARE complex has assembled and mediated membrane fusion, it is known that it disassembles for its constituents to become recycled, however, the mechanisms underlying this disassembly process remain unclear. This project utilises molecular dynamics and steered molecular dynamics to investigate stiffness of the residues of the SNARE proteins over time in 5 cognate SNARE complexes. The stiffness matrices acquired from the simulations were analysed with an autoencoder trained for anomaly detection. Stiffness propagation pathways of the SNARE complexes suggested that residues important for the disassembly process were located around the 0-layer of the complexes which is known as the most conserved layer in the SNARE protein complex. These results were compared with results from a replica of the 20s Supercomplex which showed similar results. All in all, the results from this project suggest that there is a specified pathway in which SNARE protein complexes tend to disassemble, however, how different complexes compare to one another, and why the specific path of residues is seen, remains to be elucidated.

# Introduction

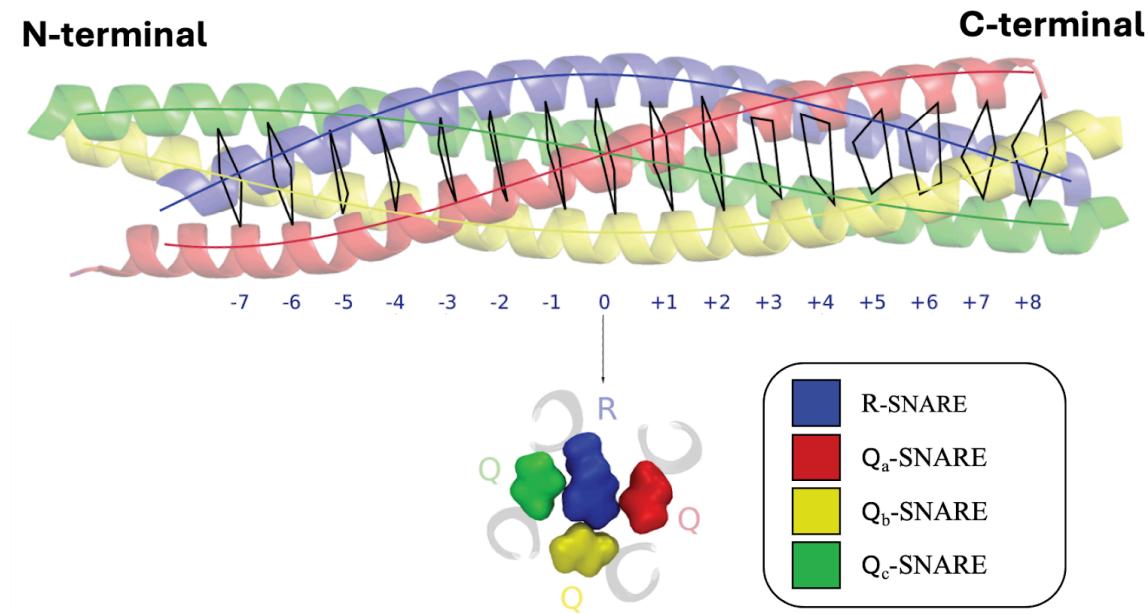
## Soluble N-ethylmaleimide-sensitive factor attachment protein receptors (SNAREs)

Soluble N-ethylmaleimide-sensitive factor attachment protein receptors (SNAREs) represent a group of small, evolutionarily conserved proteins in eukaryotes, from simple yeast to complex humans, that are vital for normal vesicular transport, the transport of molecules within a fluid-filled compartment surrounded by a lipid bilayer, often referred to simply as a vesicle, and membrane fusion among organelles and with the plasma membrane (1,2). Complementary SNAREs come together between membranes and with the help of accessory proteins they start zippering. This zippering physically pulls the membranes closer and mediates fusion. The SNARE mediated vesicular transport is important for many biological functions, including the release of neurotransmitters in synapses, protein transport from the endoplasmic reticulum and the Golgi apparatus, transportation of lysosomal catabolites among others (3).

## Structure and Naming of SNARE Proteins

Core SNARE domains have around 53 residues and are  $\alpha$ -helical when assembled into a complex. They can be grouped into two main categories based on their location, on the vesicle (v-SNAREs), which are primarily located in the secretory vesicle, or on the target membrane (t-SNAREs), which are located on the target membrane (4). Several SNAREs are located on both vesicles and target membranes, therefore further specialised classification can be done with consideration of their structural features. The SNARE protein complex is made

of a parallel 4-helix bundle composed of four different helices. Each SNARE domain can be divided into 16 mainly hydrophobic layers, where side chains of the proteins interact with each other in the core of the protein complex, many of which are conserved across species and subgroups (see *Figure 1*) (2,5). The most conserved layer is the 0-layer, an ionic layer in the centre of the complex, making up the sole hydrophilic region of the otherwise hydrophobic protein core layers (6). In almost all SNAREs, the 0-layer is composed of glutamine in the so-called Q-SNAREs, i.e., in the  $Q_a$ -,  $Q_b$ - and  $Q_c$ -SNAREs, and arginine in the R-SNARE. The other 15 layers are designated from -7, (N-terminal side) to +8 (C-terminal side) and are more or less conserved across SNARE subtypes.



*Figure 1: Layers of the SNARE complex coloured based on the corresponding SNARE protein, i.e., blue, red, green and yellow for the R-SNARE,  $Q_a$ -SNARE,  $Q_b$ -SNARE and  $Q_c$ -SNARE respectively. Locations of the N- and C-terminal regions as well as the 16 hydrophobic layers are marked. The figure is inspired by Yadav et al., (7).*

Abbreviations: SNARE, Soluble N-ethylmaleimide-sensitive factor attachment protein receptor

SNARE complexes can be divided into 4 groups, group I, II, III, and IV depending on their location and use (6). *Figure 2* shows how different groups are conserved and overall how SNARE domains are conserved. Working in the endoplasmic reticulum or Golgi apparatus catalogues the complexes in group I or II respectively. Group III is subdivided further into group III.a, which contains the protein complexes of the trans-Golgi network, and group III.b which comprises the SNAREs in the digestive endosomal compartments. The protein complexes involved in secretion are associated with group IV.

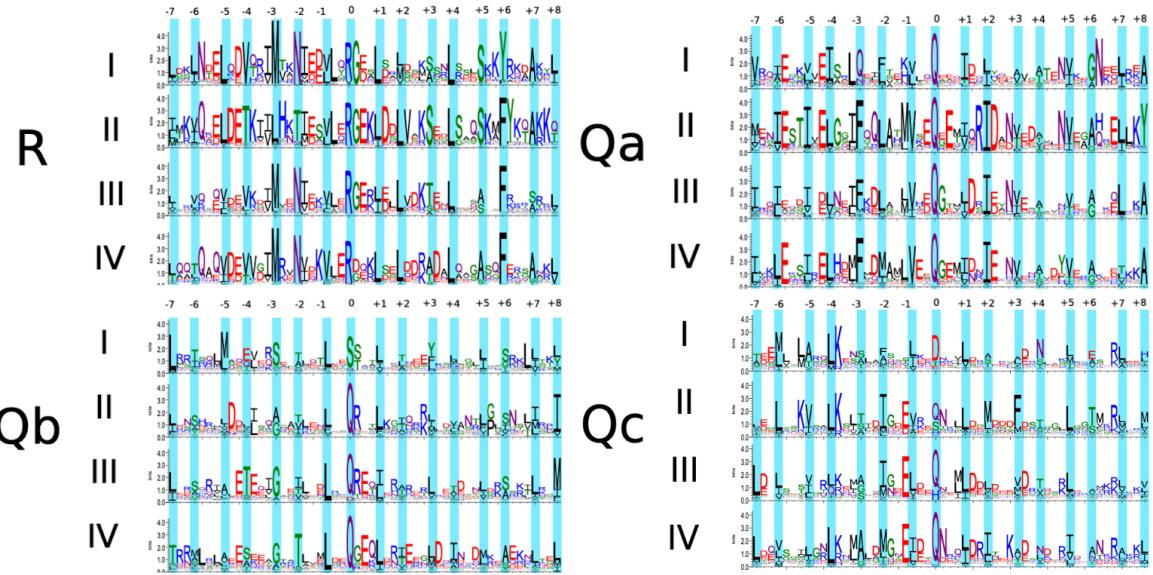


Figure 2: Weblogo representation of conserved layers of the sub-SNARE types. The SNARE proteins are written out with each of the 4 groups of SNAREs. The figure is adapted from Yadav et al., (7).  
Abbreviations: SNARE, Soluble N-ethylmaleimide-sensitive factor attachment protein receptor

## SNARE Complex Assembly

In the cell, individual SNARE proteins are kept in an inactive conformation to prevent premature SNARE zippering (8). The SNARE proteins assemble into the SNARE complex's standard four-helical bundle as they get into close contact with one another, creating a sort of bridge connecting the approaching vesicle with the membrane (9). There exist several steps prior to fusion; tethering, docking, priming, and fusion. The tethering of the SNARE proteins is facilitated by several different proteins called multisubunit tethering complexes (MTCs) (10). These tethering factors initiate the interaction between the vesicle and the target membrane and help place SNARE domains in close proximity. Tethering is followed by docking where SNARE domains come into contact and form the *trans*-SNARE complex and bring opposing vesicles and membranes together. Through the force generated by the SNARE assembly, the fusion occurs (9,11). Though there is still a lack of consensus on the order of SNARE protein assembly, one well-acknowledged hypothesis states that the Q<sub>a</sub>-protein first attaches to the Q<sub>b</sub>-protein before the Q<sub>c</sub>-protein and that these together then zipper with synaptobrevin, i.e., the R-protein, to complete the SNARE complex assembly (12). Alternatively, the Q<sub>a</sub>-protein may interact with the R-protein preceding the attachment to the Q<sub>b</sub>- and Q<sub>c</sub>-proteins (13).

## SNARE Complex Disassembly

Following a fusion event, the *trans*-SNARE complex resides on the membrane, and the resting complex is called zippered *cis*-SNARE (14). Individual SNARE proteins in the zippered *cis*-SNARE complex need to be recycled to be reused. While vesicle fusion and subsequent assembly of the SNARE complex takes a mere 200 microseconds, the disassembly for the reutilisation of materials occurs at a slower pace, more on the scale of around 5 seconds (11,15). Following the fusion of a vesicle with the cellular membrane, the

SNARE complex is attached by its C-terminal to the membrane (16). The unzipping for the disassembly then occurs in the N-to-C-terminal direction (8).

Two main proteins have been associated with the SNARE complex disassembly, *N*-Ethylmaleimide-sensitive Factor (NSF) and *N*-Ethylmaleimide-sensitive Factor Attachment Protein ( $\alpha$ -SNAP) (11,17). NSF is a homohexamer consisting of an N-terminal adapter domain and two ATPase associated with various cellular activities (AAA+) domains (NSF-D1 and NSF-D2). The N terminal domain is crucial for interacting with the SNARE complex and the  $\alpha$ -SNAP. The NSF-D1 domain is considered a catalytically active ATPase while the NSF-D2 domain functions as a structural scaffold, facilitating nucleotide-dependent hexamerisation. ATP hydrolysis in the NSF-D1 domain triggers significant conformational changes within NSF, providing the mechanical force required for unwinding (18).  $\alpha$ -SNAP comprises several distinct domains that collectively regulate membrane fusion processes in eukaryotic cells (18,19). Its N-terminal domain, also known as the NSF-binding domain, facilitates specific interaction with NSF, crucial for disassembling SNARE complexes. The central domain of  $\alpha$ -SNAP plays a pivotal role in SNARE protein interactions, particularly with the SNARE complex formed by the proteins syntaxin, SNAP-25, and synaptobrevin, promoting the disruption of SNARE complexes during membrane fusion events. Additionally, the C-terminal domain of  $\alpha$ -SNAP facilitates membrane binding, ensuring its localisation to sites of membrane fusion. Together, these domains orchestrate the intricate molecular mechanisms underlying vesicle trafficking and cellular membrane dynamics.

The NSF/ $\alpha$ -SNAP/SNARE (20s) supercomplex was solved by single-particle electron cryomicroscopy (18). ATP and ADP bound 20s supercomplex structures revealed that this complex displays asymmetry, transitioning from the six-fold symmetry of NSF ATPase domains to a pseudo-four-fold symmetry of the SNARE complex.  $\alpha$ -SNAPs interact with the SNARE complex in an opposing structural orientation, indicating a potential unwinding mechanism.

It has been shown that  $\alpha$ -SNAP can guide the NSF to the SNARE complex and help in disassembly in a processive unwinding manner (see *Figure 3A*), rather than a global unwinding (see *Figure 3B*). Whereas the global unwinding hypothesis suggests that there is a one-step disassembly of the SNARE complex, the progressive unwinding hypothesis states that there is a more gradual disassembly of the complex. In addition to these hypotheses, the distributive disassembly hypothesis which suggested that the NSF binds the complex in multiple cycles and this would then cause the protein complex to turn and disassemble (see *Figure 3C*) was suggested, however, it has since been disproven (5). It requires around 10 adenosine triphosphate (ATP) molecules for the disassembly of the SNARE protein complex (17). Though the instigating reasons for the disassembly are known, there is still a lot unknown regarding the protein itself and how the dynamics of different SNARE structures appear during disassembly.

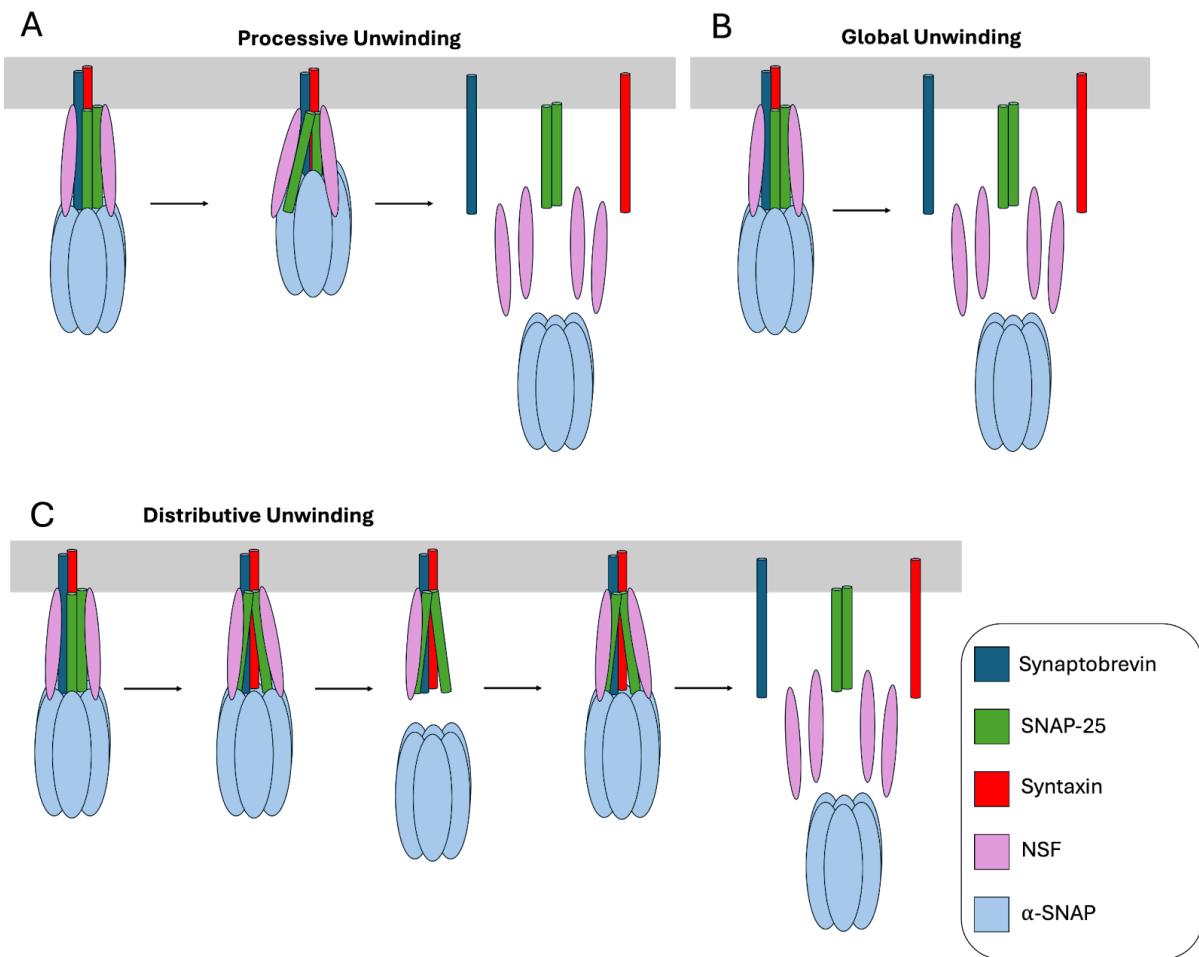


Figure 3: Overview of possible strategies of a classical neuronal SNARE complex disassembly, adapted from figure by Cipriano et al., (11). A) Processive unwinding, B) Global Unwinding, and C) Distributive Unwinding.  
 Abbreviations: SNARE, Soluble N-ethylmaleimide-sensitive factor attachment protein receptor; SNAP-25, Synaptosome Associated Protein 25; NSF, N-Ethylmaleimide-sensitive Factor;  $\alpha$ -SNAP,  $\alpha$ -soluble NSF attachment protein

## Protein Stiffness

Recent advancements in single-molecule manipulation techniques, such as atomic force microscopy (AFM) and optical tweezers, have enabled the precise measurement of proteins' mechanical resistance to external forces applied at specific positions and uncovering a residue-/position-specific response of proteins to forces (20–23).

Previously, it has been demonstrated that coarse-grained elastic network models (ENMs), based on the inter-residue contact topology, can adequately account for the varying forces applied on different residue pairs (24). Despite their simplicity, these models capture a fundamental property of proteins: resistance to deformations, also called stiffness. The stiffness of a residue in a protein at a specific moment in time is reflective of its command over the protein's stability. A higher stiffness implies that the residue in question requires more force or energy exerted on it for its interactions to be broken. Obtaining a measurement of protein stiffness, which has been difficult until now, could offer another method, in addition to measuring unfolding force and contour length increment, for understanding the varied conformational changes in proteins. These methods rely on the spatial extent of

fluctuations in each force-induced equilibrium conformation as a unique structural identifier (25).

Coarse-grained methods face a challenge as they rely on the intrinsic topology of a protein under near-equilibrium conditions and in SNARE complex assembly/disassembly additional proteins help the complex. This challenge can be addressed by integrating molecular dynamics (MD) and ENM methods to provide a comprehensive analysis of the SNARE-complex assembly/disassembly mechanism.

## Molecular Dynamics and Steered Molecular Dynamics

Insights into the molecular details of the mechanical responses have been uncovered by theoretical investigations. Specifically, MD simulations have been proven crucial in estimating the forces required for mechanical unfolding and mapping out the unfolding pathways under controlled pulling directions. Titin, in particular, has emerged as a valuable model system for elucidating the interplay between mechanical stability and biological functionality, as evidenced by both steered and quasi-equilibrium MD simulations (26–30). Similar MD studies have also been conducted on other small globular proteins, shedding light on the significant impact of force application pathways on mechanical stability and unfolding trajectories (31,32). Although they give a detailed description and the solution of a problem, MD simulations are computationally exhaustive.

MD is a computational simulation technique to study the dynamics of atoms and molecules over time in biomolecular systems such as proteins, nucleic acids, and lipids. In an MD simulation, Newton's equations of motion are solved numerically to simulate the positions and velocities of atoms and molecules as they evolve. The simulations consider various factors such as bond vibrations, atomic interactions (such as van der Waals forces and electrostatic interactions), and solvent effects. MD simulations can provide insights into molecular structures, dynamics, conformational changes, binding events, and other molecular properties under different conditions such as temperature and pressure (33).

Steered molecular dynamics (SMD) is a specialised variant of molecular dynamics focused on studying the response of a biomolecular system to an applied force. In SMD, external forces are applied to specific atoms, or groups of atoms, in a biomolecule to induce structural changes or study mechanical properties like unfolding forces, or binding affinities. They provide atomistic details into the mechanical behaviours of biomolecules, helping to understand their functional properties under mechanical stress or during biological processes involving force-induced structural changes (34).

The computational demand increases exponentially as the number of atoms in the system increases. An all-atom MD simulation with the SNARE complex along with its disassembly proteins NSF and  $\alpha$ -SNAP would consist of 1,055,740 atoms including the water and the ions. Such a simulation would not be feasible for this project due to its computational intensity, especially considering the timescales involved (potentially hundreds of nanoseconds). This would demand significant computational resources and time. Instead, we opted for an SMD simulation focused solely on the core SNARE complex, which has

approximately 98,000 atoms including the water and the ions, mimicking the forces exerted by NSF and  $\alpha$ -SNAP. By narrowing our focus in this way, we reduced the computational load and simplified the simulation process.

## Deep Learning and Autoencoders in Biological Research

Deep learning lies under the broader categories of artificial intelligence and machine learning which have gained significant importance in biological research due to their capabilities in handling complex data, extracting meaningful features, and making predictions or classifications (35). In biological research, deep learning techniques are applied to various tasks such as image analysis, sequence analysis, drug discovery, genomics, proteomics, and more. Deep learning models can automatically learn patterns and relationships from vast amounts of data, leading to insights and discoveries that may not be easily detectable through traditional analysis methods. Deep learning methods are also being integrated into MD simulation analysis to enhance our understanding of complex biomolecular systems for molecular property prediction, conformational sampling, enhanced sampling, structure analysis, feature extraction, and model interpretability.

This project aims to detect key changes upon application of a force. This falls in the category of feature extraction, and more specifically, anomaly detection. Anomalies, also known as outliers, are rare events that deviate from the typical behaviour of the system. In a previous study, autoencoders were used to detect dynamic allosteric triggered by ligand binding (36). In the reference study, autoencoders were used to compare apo and holo MD simulations and detect anomalies in terms of the allosteric signal.

There are several deep-learning approaches for anomaly detection. Autoencoders consist of an encoder and a decoder and are adept at learning representations of normal data and detecting anomalies based on higher reconstruction errors. Variational Autoencoders (VAEs) extend this by modelling data distributions and identifying anomalies as low-probability events. Generative Adversarial Networks (GANs) generate synthetic data mimicking real data distributions; anomalies are flagged as instances significantly divergent from the learned distribution. Deep Neural Networks (DNNs), including CNNs and RNNs, excel at capturing patterns in various data types, enabling anomaly detection by detecting deviations from learned behaviour.

Autoencoders play a crucial role in anomaly detection due to their capacity to learn complex patterns in data, enabling them to identify anomalies based on deviations in data reconstruction. Their unsupervised learning nature and ability to handle high-dimensional data make autoencoders a great option for uncovering anomalies across diverse datasets. Autoencoders are a type of deep learning neural network especially designed to use the input data as validation for how well it reconstructs the data after reducing it into lower dimensions (36). The encoder part of the autoencoder initially compresses the input data to form what is called the code (see *Figure 5*). The decoder then takes this compressed data and attempts to reconstruct the original data. During the training of the neural network, it adjusts its weights

and biases to minimise the reconstruction error, i.e., the measure of the information lost between the original and the reconstructed data.

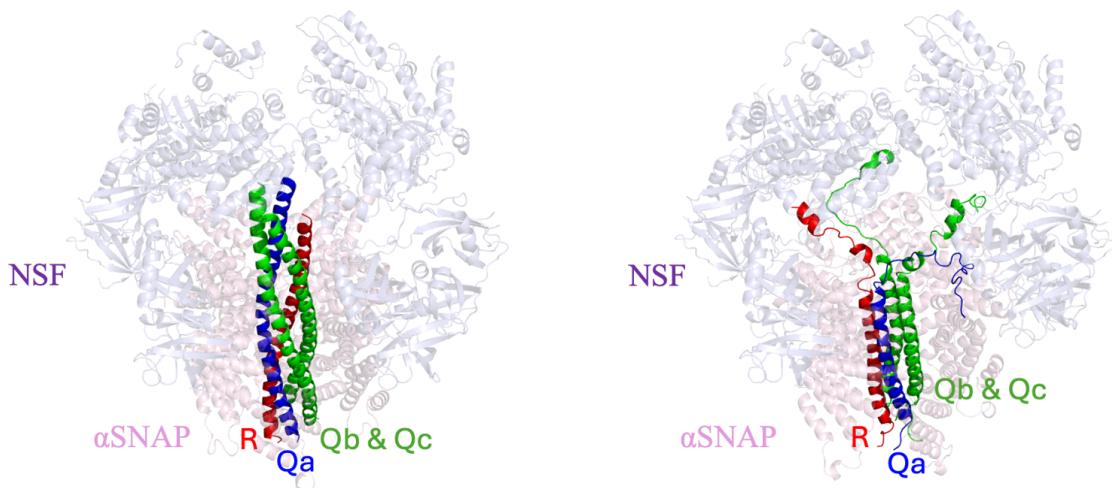
## Significance of the Project

In this project, an autoencoder-based approach will be used to determine the differences in MD and SMD simulations, indicating the residues of the protein complex seemingly most important in its disassembly process, by detecting deviations from the normal protein behaviour. This will be done by comparing the stiffness patterns from the MD simulations and SMD simulations, something that currently remains unknown.

While the mechanistic details of SNARE complex assembly are extensively studied, comparatively little is known about the mechanistic details of the disassembly. It is known that NSF and  $\alpha$ -SNAP aid the disassembly, but the atomistic details of how each SNARE protein reacts to the applied force have not yet been studied. What is known about these two proteins is that  $\alpha$ -SNAP forms salt bridges with the SNARE complex, holding it from the C-terminal, while NSF transitions from ATP-bound split-washer conformation to the ADP-bound flat washer conformation and exerts a rotational force on the SNARE complex (14).

An MD simulation study was conducted for this project prior to its start by the Fasshauer Lab on the NSF- $\alpha$ -SNAP-SNARE supercomplex, where the rotation of NSF is mimicked, comparing the NSF conformations in both the ATP-bound split washer NSF conformation (PDB id: 3j96) in the 20s supercomplex and ADP-bound flat washer NSF conformation (PDB id: 3j99) (18). This was done by employing collective variables in MD simulations. The transition of NSF from its ATP-bound conformation to the ADP-bound conformation was induced, effectively mimicking the disassembly behaviour of the SNARE complex.

*Figure 4* shows that as NSF exerts a rotational force on the SNARE complex and  $\alpha$ -SNAP clamps its C-terminal, the N-terminal portions of the complex start disassembling outside. This simulation with more than 1 million atoms took more than a week to run for only one replica and a more simplified version of the disassembly was needed. To mimic the rotational force exerted by NSF and simplify the disassembly problem we decided to take only the SNARE complex and pull from the N-terminal residues. This approach lowers the number of atoms in the system to 98,000 and the simulation time for one replica to approximately 30 minutes. With the current simplified version of the problem, we decided to inspect the disassembly of all available SNARE complexes in the PDB and identify the mechanism of disassembly and whether it is similar between different structures. Using this method, it will also allow us to confirm the pattern found to the original MD simulation with the 20s supercomplex to evaluate whether our methodology utilising SMD is an accurate mimic and hence can also be used in future studies to investigate the SNARE complex disassembly.



*Figure 4: Visualisation of the SNARE protein complex with its disassembly proteins as A) 20s supercomplex in NSF split washer conformation and B) 20s supercomplex with NSF flat washer conformation after the rotation of NSF. Coloured based on protein, the blue transparent is NSF and the pink transparent is  $\alpha$ -SNAP, while red, blue and green represent the R-SNARE,  $Q_a$ -SNARE, and  $Q_b$ - and  $Q_c$ -SNAREs respectively.*

*Abbreviations:* SNARE, Soluble N-ethylmaleimide-sensitive factor attachment protein receptor; NSF, N-Ethylmaleimide-sensitive Factor;  $\alpha$ -SNAP,  $\alpha$ -soluble NSF attachment protein

## Aim of the Project

A lot of work around a decade ago focused on trying to establish the SNARE complex disassembly, however, multiple theories were proposed, and a consensus could not be reached. With today's more advanced methods of machine learning and with the light of MD and SMD simulations, we hope to give a better understanding of the complexities of the SNARE complex disassembly. By running MD and SMD simulations of multiple SNARE complexes available in the protein data bank (PDB) we will quantify the stiffness of our complexes throughout time. An autoencoder will be built to learn hidden patterns in MD simulation stiffness matrices, which are the equilibrium stiffness properties of the proteins. Subsequently, the MD stiffness-trained autoencoder will be used to reconstruct the SMD stiffness matrices. The reconstruction error will tell us about anomalies and allow us to detect changes in the stiffness patterns during the initiation of SNARE complex disassembly.

## Materials and Methods

### PDB Structures

For this study, we considered 5 cognate SNARE complexes with the PDB IDs; 1SFC, 1GL2, 2NPS, 3B5N, and 4WY4. 1SFC is a neuronal SNARE complex (37), 1GL2 and 2NPS are endosomal SNARE complexes (38,39), 3B5N is a plasma membrane SNARE complex (40) and lastly, 4WY4 is part of the autophagosome and thus is involved in mediating autophagosome-lysosome fusion (41). Individual SNARE proteins in the structures are summarised and explained in Table 1.

*Table 1: Overview of SNARE complex structures analysed in this project. Group of the SNARE proteins are mentioned for each of the complexes with each complex' four sub-proteins and the organism the complex structure is found in.  
Abbreviations: SNARE, Soluble N-ethylmaleimide-sensitive factor attachment protein receptor*

Complex	1SFC	1GL2	2NPS	3B5N	4WY4
<b>Group</b>	IV	III	IIIb	IV	III/IV
<b>R SNARE</b>	Synaptobrevin-2	Endobrevin/ Synaptobrevin-8	Synaptobrevin-4	Snc1p	Synaptobrevin-8
<b>Q<sub>a</sub>-SNARE</b>	Syntaxin-1A	Syntaxin-7	Syntaxin-13	Sso1p	Syntaxin-17
<b>Q<sub>b</sub>-SNARE</b>	SNAP-25B	Syntaxin-8	Vti1a	Sec9b	SNAP-29
<b>Q<sub>c</sub>-SNARE</b>	SNAP-25B	Vti1b	Syntaxin-6	Sec9b	SNAP-29
<b>Organism</b>	<i>Rattus norvegicus</i>	<i>Rattus norvegicus</i> and <i>Mus musculus</i>	<i>Rattus norvegicus</i> , <i>Mus musculus</i> and <i>Homo sapiens</i>	<i>Saccharomyces cerevisiae</i>	<i>Homo sapiens</i>

## MD Simulations

The structures of SNARE complexes were downloaded from the PDB and prepared by using VMD v.1.9.4 and the plugin QwikMD (42,43). Only the core SNARE domains are considered and the chains are renumbered, the chain A (R-SNARE) residues are from 1-53, the chain B (Q<sub>a</sub>) residues are 54-106, the chain C (Q<sub>b</sub>) residues are 107-159 and the chain D (Q<sub>c</sub>) residues are 160-212. As a negative control, a four-helical bundle with the same residue length was generated by using CCBUILDER (v. 2.0.5) (44). Each system was solvated in a Transferable Intermolecular Potential with 3 Points (TIP3P) water model with 12 Å of water buffer between the protein surface and the periodic box edge (45). Sodium and chloride atoms were added to neutralise the system with a concentration of 150 mM. CHARMM36 force field was used, and the time step for integration was set to 2 fs (46,47). The simulations were performed with explicit solvent in the NpT ensemble, and a cut-off of 12.0 Å was applied to non-bonded interactions and 10.0 Å for the smothering functions. The temperature was maintained at 300 K using Langevin dynamics. The pressure was maintained at 1 atm using the Nosé-Hoover Langevin piston (48,49). All simulations were performed with NAMD3 (50). Each system was minimised for 0.1 ns, gradually heated up to 300 K for 0.25 ns, and equilibrated for 1 ns at 300 K. After the equilibration, 100 ns of simulations were run for each system. For 100 ns of simulations, 10000 stiffness matrices were produced and used to train.

## SMD Simulations

The system preparation, minimisation, annealing, and equilibration procedures were executed identically to those performed for MD simulations, and following the equilibration runs, the SMD simulations were performed (51).

The SMD was performed with an explicit solvent using the TIP3 water model in the NpT ensemble (45). The temperature was maintained at 300 K using Langevin dynamics and the pressure was kept at 1 atm using the Nosé-Hoover Langevin piston (48,49). The short- and

long-range interaction cutoffs and the methods to manage them were kept the same as the MD simulations. The time step of integration was chosen to be 2 fs for all simulations. Constant velocity pulling was applied with a pulling speed of 2.5 Å/ns and a harmonic constraint force of 7.0 kcal/mol/Å<sup>2</sup>. In this step, SMD was employed by harmonically restraining the position of the C<sub>α</sub> of residue 55 on chain B (Q<sub>a</sub>-SNARE) and moving the restraint to the C<sub>α</sub> of residue 161 on chain D (Q<sub>c</sub>-SNARE), with constant velocity. 10 replicas of the SMD simulations were performed for each system.

## Stiffness Calculations

The stiffness calculations were done by using MechStiff in the ProDy package in Python (24,52). MechStiff is used to assess the mechanical stiffness of a protein structure represented as an elastic network, typically using the Anisotropic Network Model (ANM), a type of ENM (53,54). In ENMs, the protein is represented as a network of mass and spring, where each residue is a mass and the residues within a certain cutoff (usually 7 Å) of distance are connected with a spring. This simplification captures the essential structural connectivity and dynamics of the proteins. The connectivity matrix generated by using the neighbouring information is then considered within the force constant matrix and is used to predict anisotropic motions by performing normal mode analysis. For MechStiff calculations, an external force is applied to the connectivity matrix, by simply increasing the inter-residue distance, acting as a deformation (24). Normal mode analysis was performed on the force-exerted connectivity matrix and the changes in the coordinates were used to determine the mechanically strong (stiff) or weak regions in the protein.

MechStiff calculates the stiffness between residue pairs in the protein, averaged over its interactions with other residues. The resulting stiffness map can be visualised as a 2D map or 1D profile, providing the distribution of mechanical properties across the protein. The input is a PDB file. From our MD and SMD simulations, we took screenshots of the SNARE complex and calculated the stiffness of residue pairs by using the distance cutoff value as 7 Å, and saved the NxN stiffness matrices, where N is the residue number, for further analysis.

## Autoencoder

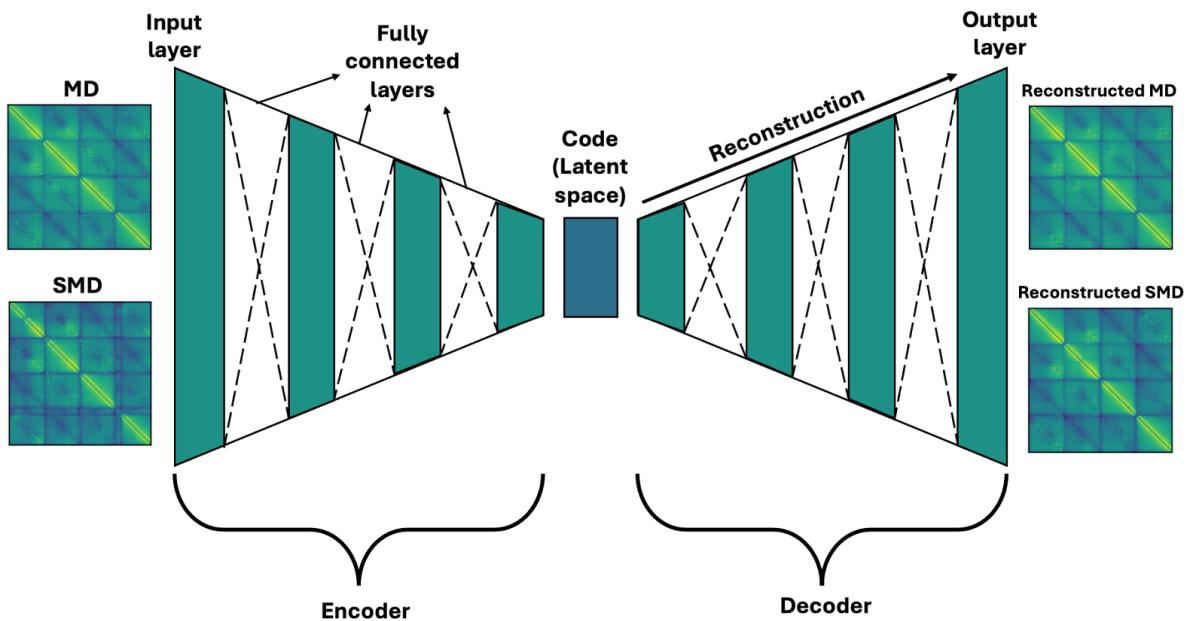
The multilayer autoencoder was implemented in this project and consisted of 2 encoder layers, of the sizes 200 and 100 neurons, and 2 decoder layers, of the sizes 100 and 200 (see *Figure 5*). The size of the code layer, i.e., the latent dimension, was 90 neurons.

The encoder first flattened the input data and passed it through fully connected layers with *sigmoid* activations, gradually reducing the dimensionality to a specified latent dimension. Batch normalisation was applied to stabilise training.

The decoder took the encoded representation and reconstructed the original input using *ReLU*-activated layers to preserve the input structure. The model was trained using mean absolute error (MAE) loss and the Adam optimizer with a learning rate of 0.01. The autoencoder was trained for 2500 epochs, adjusting weights with batch gradient descent, with a batch size of 200. 80% of the data was used to train the model and 20% was used to test, to

minimise reconstruction error, and learn meaningful representations in the latent space. The train-test split was done randomly.

To train the model, stiffness matrices generated from the MD simulations were used. Then the stiffness matrices generated from the SMD simulations were fed into the MD-trained model and the reconstructed SMD-stiffness matrices were compared with the real SMD-stiffness matrices to detect anomalies.



*Figure 5. Schematic of an autoencoder with example input used in this project. Stiffness matrices are used as input and are compressed to the code, i.e., the latent space, in the encoder through fully connected layers. The decoder reconstructs the code back to the stiffness matrices.*

*Abbreviations: MD, molecular dynamics, SMD, steered molecular dynamics*

## Model Parameter Optimisation

The autoencoder's performance was determined by its ability to reconstruct the input data. To get the best performance and reconstruct the input data with low reconstruction loss, we optimised two model parameters; latent dimension and batch size (See *Supplementary Figure 1*).

### Latent Dimension

The latent dimension is the low-dimensional space which has been generated through the encoder, i.e., the code (55). The different dimensional sizes were tested with values between 20 and 100, taking it with increments of 10. A latent dimension of 90 was decided on due to it having the lowest loss value.

### Batch Size

Model training occurs on the mini-batch of the input selected from the training data, thus it is logical that batch size is an important hyperparameter for autoencoders. Conventionally, larger batch sizes cause overgeneralisation and smaller batch sizes potentially lead to faster convergence and better generalisation. However, smaller batch sizes may also result in noisy

results and slower training. To decide on the optimum batch size for our systems we experimented on batch sizes of 200 and 900 and compared the reconstruction loss values. The batch size of 200 was chosen as it had been used in previous similar experiments in the lab and 900 was chosen based on the literature, where a larger batch size was determined as approximately 10% of the training data (56). In the end, the batch size of 200 was chosen as it yielded lower loss values. The loss values can be described as the discrepancy between what the models predict and the underlying true value (35).

### *Epochs*

Initially, the number of epochs was set to 5000 and later it was changed based on when there was a convergence of the reconstruction loss using the decided on variables. The number of epochs selected for further analyses was 2500.

### **Autoencoder Input**

The autoencoder was initially trained with the stiffness matrices generated from the MD simulations. However, for the anomaly detection, this MD-trained model was used to reconstruct SMD stiffness matrices and these SMD matrices were thus used as input.

### **Autoencoder Output**

The autoencoder tried to reconstruct the given input depending on the training data and the output was to be the *reconstructed data*. For the MD stiffness matrices the expected reconstructed data was the input stiffness matrices. However, for the SMD stiffness matrices, the model tried to reconstruct equilibrium stiffness behaviour, i.e., what it would look like if the complex had been under the same circumstances as in the MD simulations, and therefore the regions where the model failed to predict were the regions in the mechanical pathway where the force was being propagated.

### **Reconstructed SMD Stiffness Matrix Normalisation**

Since the model had a certain reconstruction loss for the cognate input, i.e., the MD stiffness matrices, while analysing the reconstructed SMD stiffness matrices, we needed to take that into account. Therefore, the reconstructed SMD-stiffness matrices were normalised with the loss value for reconstructed MD stiffness matrices. For each frame of the reconstructed MD stiffness matrices, the error of the stiffness matrix output of the decoder compared to the input matrix inserted into the encoder, and the mean error was calculated. This was named the mean MD error. The reconstructed SMD stiffness matrices were then divided by the mean MD error.

### **Packages and Software**

For a comprehensive list of all packages used in this project, please refer to the section *Code availability* under the *Materials and Methods* section. Within this section, only the major packages and software are recounted.

The code for this project was mainly written in Python, v. 3.9.18 for the autoencoder and optimisation steps, and v. 3.10.12 for investigating the stiffness propagation path. For the

machine learning used in this project, i.e., the autoencoder, the Curnagl cluster was used to host the data and run the code. Google Colaboratory was used to store the files of source code and run that which were not run on the cluster (57).

For the optimisation as well as the using thorough training and validation of the autoencoder, tensorflow (v. 2.15.0) was used (58). Additionally, prody (v. 2.4.1), mdtraj (v. 1.9.9), scipy (v. 1.12.0) and numpy (v. 1.23.5) were used (59–62). For the plotting of graphs, matplotlib (v. 3.8.3) was used (63).

## Code Availability

The source code used in this project is saved in Google Colaboratory notebooks, the links to which can be found below:

- [!\[\]\(8382ea97f7a202bfb1c791a20a742461\_img.jpg\) Optimisation.ipynb](#)
- [!\[\]\(fa2541a4a5206a1c100ea1d9cf154431\_img.jpg\) Normalisation.ipynb](#)
- [!\[\]\(d21cb8a7ab1334dbb9f024737cef676c\_img.jpg\) MD\\_Autoencoder.ipynb](#)
- [!\[\]\(c4226810c87a3277379ed1718306b4c3\_img.jpg\) SMD\\_Autoencoder.ipynb](#)

## Results

The autoencoder was trained to recognise and reconstruct stiffness matrices generated from MD simulations for 1SFC, 1GL2, 2NPS, 3B5N, 4WY4, and the negative control. After achieving a low reconstruction loss value for MD stiffness matrices, stiffness matrices from SMD were fed into the autoencoder, and anomalies for each system were detected. These anomalies in the stiffness matrices upon force application gave us the force propagation paths for the different SNARE complexes.

For each system, 100 ns MD simulations were performed. To establish a negative control and specify the four-helical bundle structural features, an additional set of simulations was conducted on a 4-helical bundle structure designed to mimic a SNARE complex, composed of a four-helical bundle with a randomised sequence, unrelated to SNARE proteins (44). By comparing the behaviour of this negative control with the SNARE complexes under identical conditions, we aimed to isolate and characterise the contributions of SNARE-specific elements to the observed dynamics and interactions. After optimising the autoencoder hyperparameters, we trained the models for all of the 6 systems.

## MD Results

### 1SFC

The autoencoder that was trained with the stiffness matrices from the MD simulations for 1SFC had a 95% accuracy. *Figure 6A* illustrates the reconstructed stiffness matrices, demonstrating the model's proficiency in capturing the equilibrium stiffness values. *Figure*

*6B* shows that the loss value converged with a low loss value of 0.05, indicating the model's good performance and ability to accurately reconstruct the data.

## 1GL2

A 95% accuracy was also seen in the reconstruction of the autoencoder which was trained with stiffness matrices from the MD simulations from 1GL2 (see *Figure 6C-D*). The loss value converged at 0.05.

## 2NPS

For the autoencoder trained on the stiffness matrices from the MD simulations for 2NPS, an accuracy of 95% and a loss value of 0.05 were once again achieved (see *Figure 6E-F*).

## 3B5N

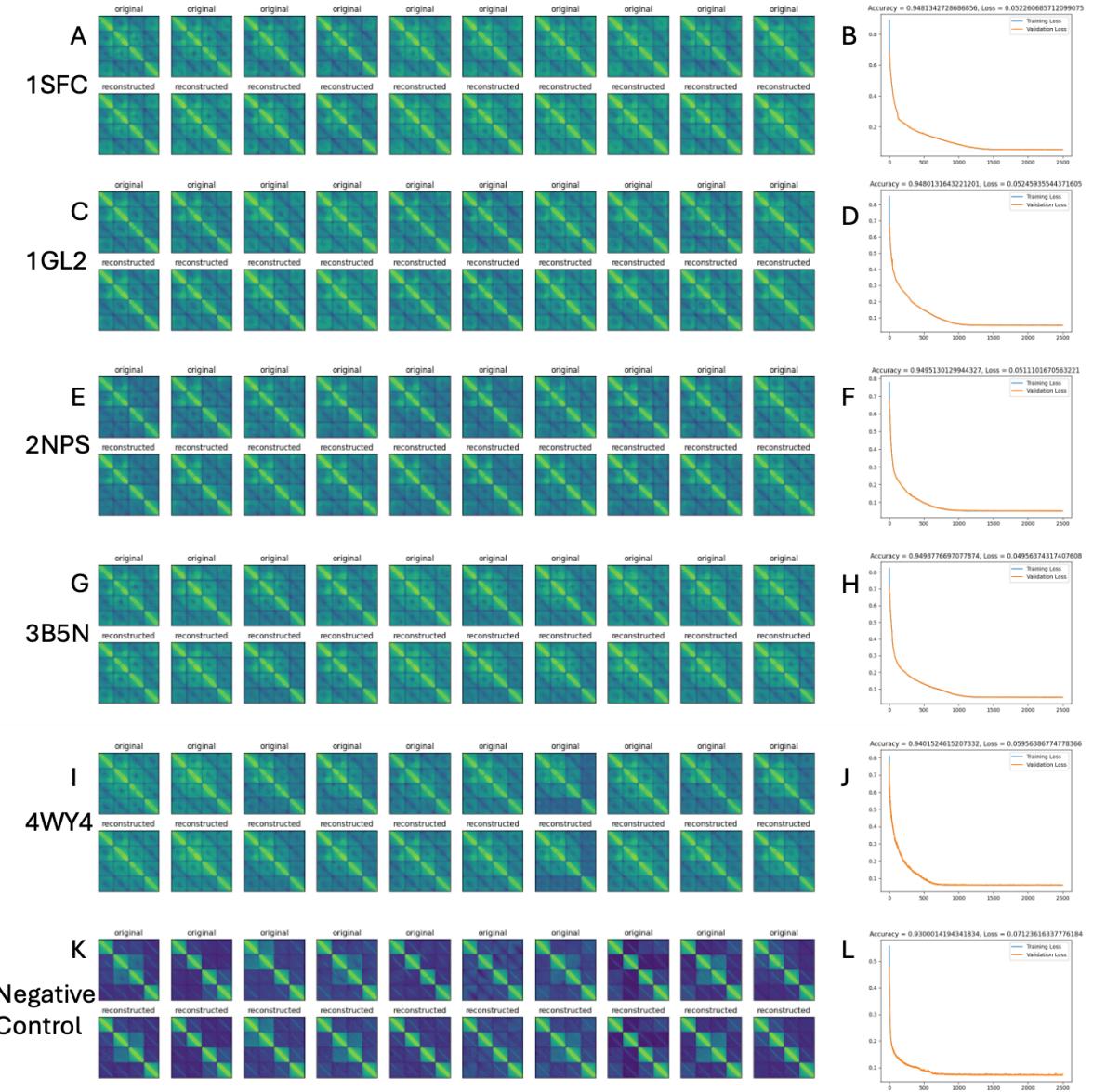
As for 1SFC, 1GL2, and 2NPS, the autoencoder which was trained with the stiffness matrices from the MD simulations for 3B5N had an accuracy of 95% and a loss value of 0.05 (see *Figure 6G-H*).

## 4WY4

The autoencoder which was trained on the stiffness matrices from the 4WY4 MD simulations reached an accuracy of 94% and its loss value converged at 0.06 (see *Figure 6I-J*).

## Negative Control

Using the mimic of the SNARE complex to train the autoencoder with its stiffness matrices of the MD simulations, an accuracy of 93% and a loss value converging at 0.07 was achieved (see *Figure 6K-L*).



*Figure 6. Original vs reconstructed stiffness matrices from the SNARE complexes' MD simulations as well as training loss and validation loss curves for the corresponding structures. Each stiffness matrix contains a 4x4 grid showing the Chain A (R-SNARE), Chain B ( $Q_a$ -SNARE), Chain C ( $Q_b$ -SNARE) and Chain D ( $Q_c$ -SNARE) from left to right and from top to bottom. Within each of the 16 squares are the 53 residues per protein. The colours are reflective of the interactions between the residues. The stiffness matrices of A) ISFC, C) IGL2, E) 2NPS, G) 3B5N, and I) 4WY4 show the stiffness matrices of the last 10 frames of the MD simulations (top), and the autoencoder reconstructed matrices (bottom). K) The negative control with 10 stochastically chosen frames from the MD simulation (top) and the corresponding autoencoder reconstructed matrices (bottom). The training loss and validation loss curves for B) ISFC, D) IGL2, F) 2NPS, H) 3B5N, J) 4WY4, and L) negative control MD autoencoder training.*

*Abbreviations: SNARE, Soluble N-ethylmaleimide-sensitive factor attachment protein receptor; MD, molecular dynamics*

## SMD Results

After training the model to reconstruct the equilibrium stiffness values we tried to reconstruct the SMD stiffness matrices by using the same model and detecting the anomalies. We performed 10 SMD simulations. *Figure 7* shows one example of reconstructions of SMD stiffness matrices per SNARE complex structure. The patterns the autoencoder could not capture, due to it being trained on MD simulations, was saved as anomalies and are thus the

residues important for the structure's disassembly process. These are further analysed in the *Anomalies Detected* section of the *Results*.

## 1SFC

The 1SFC SNARE complex structure reconstructions of SMD stiffness matrices show how the autoencoder failed to predict certain parts of the stiffness in the simulation (see *Figure 7A*).

## 1GL2

The SMD stiffness matrices of the 1GL2 SNARE complex structure can be seen in *Figure 7B* as the original as well as the decoded by the autoencoder.

## 2NPS

In *Figure 7C*, the anomalies detected by the autoencoder for the SMD simulations of the 2NPS SNARE complex structure are shown.

## 3B5N

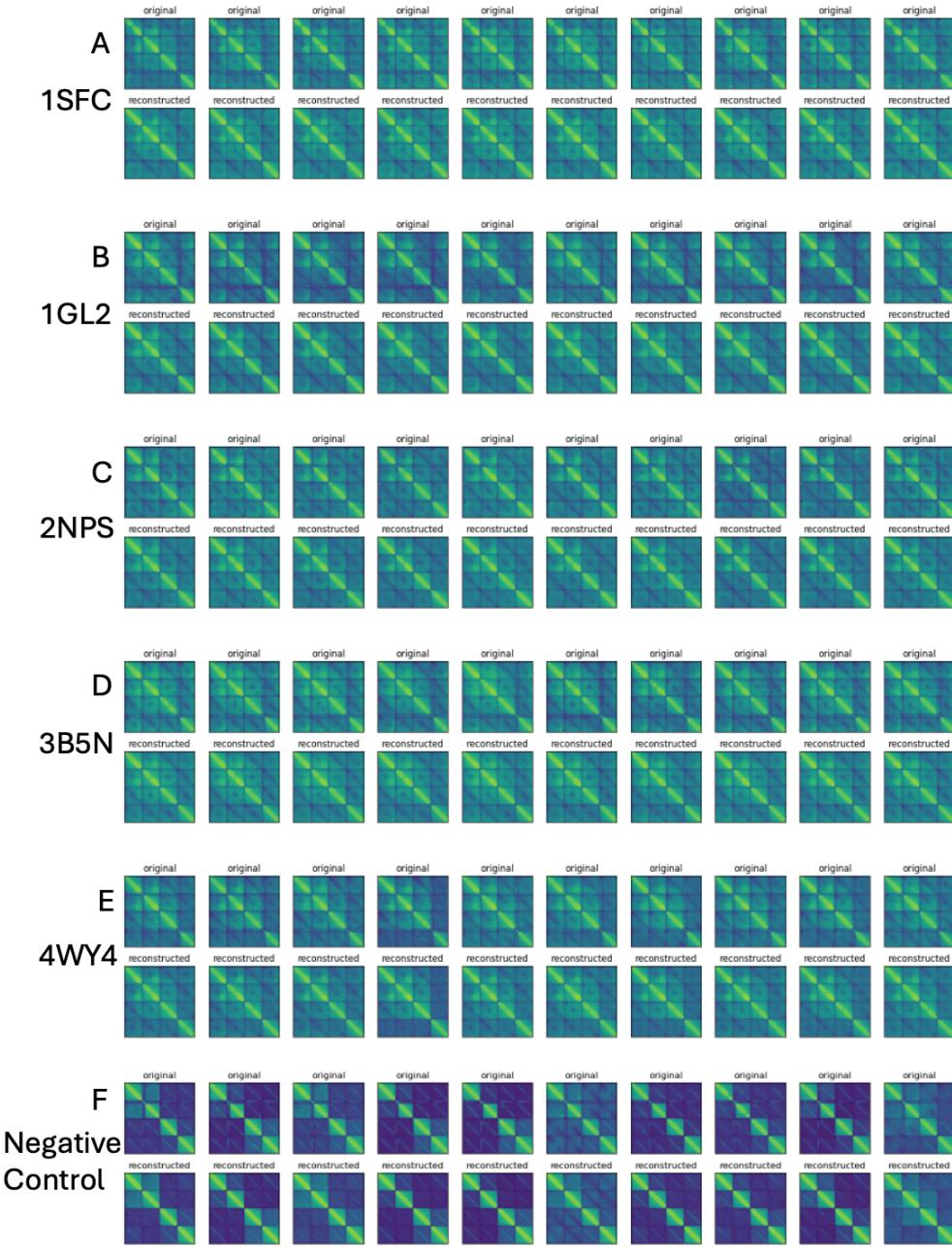
The autoencoder's reconstructed version of the SMD stiffness matrices of the 3B5N SNARE complex structure is visualised in *Figure 7D*.

## 4WY4

In *Figure 7E*, the stiffness matrices of the 4WY4 SNARE complex structure are shown as the original and decoded by the autoencoder.

## Negative Control

Like previously, the negative control was a mimic of the SNARE complex which in the SMD simulation was pulled in the same manner as the other SNARE proteins. *Figure 7F* shows the original and MD-trained autoencoder reconstructed stiffness matrices for the negative control.



*Figure 7: Original vs reconstructed stiffness matrices from the SNARE complexes' SMD simulations as well as training loss and validation loss curves for the corresponding structures. Each stiffness matrix contains a 4x4 grid showing the Chain A (R-SNARE), Chain B ( $Q_a$ -SNARE), Chain C ( $Q_b$ -SNARE) and Chain D ( $Q_c$ -SNARE) from left to right and from top to bottom. Within each of the 16 squares are the 53 residues per protein. The colours are reflective of the interactions between the residues. The stiffness matrices of A) ISFC, B) 1GL2, C) 2NPS, D) 3B5N and E) 4WY4 show the stiffness matrices of the last 10 frames of the SMD simulation (top) and the autoencoder reconstructed matrices (bottom). F) The negative control with 10 stochastically chosen frames from the SMD simulation (top) and the corresponding autoencoder reconstructed matrices (bottom).*

*Abbreviations: SNARE, Soluble N-ethylmaleimide-sensitive factor attachment protein receptor; SMD, steered molecular dynamics*

## Anomalies Detected

The reconstruction error from the SMD stiffness matrices was used to detect the anomalies. For each set of reconstructed SMD stiffness matrices, the error was measured by subtracting

the original stiffness matrix from the reconstructed stiffness matrix. The SMD reconstruction error matrices were normalised with the mean MD reconstruction error matrix. The highest reconstruction error from 10 replicate SMD stiffness values was detected and clustered to find the most common propagation path residues (see *Table 2*). The SNARE-specific residues important for the disassembly pathway could be found by concatenating the pathways from the 5 SNARE complex structures and subtracting the pathway of the negative control.

*Table 2: The common mechanical force propagation path residues for the 5 cognate SNARE complexes, 1SFC, 1GL2, 2NPS, 3B5N, and 4WY4, as well as for the one negative control. The SNARE specific path is the combination of the 5 cognate SNARE complexes, 1SFC, 1GL2, 2NPS, 3B5N, and 4WY4, minus the negative control.*

*Abbreviations: SNARE, Soluble N-ethylmaleimide-sensitive factor attachment protein receptor*

<b>1SFC</b>	160, 155, 105, 106, 163, 51, 159, 158, 104, 6, 109, 54, 7, 162, 11, 5, 27, 211, 101, 3, 150, 156, 4, 108, 166, 30, 112, 21, 170, 173, 167, 110, 53, 55, 171, 17, 9, 58, 10, 63, 62, 59, 113
<b>1GL2</b>	210, 211, 207, 59, 143, 157, 55, 113, 6, 54, 107, 116, 57, 9, 3, 163, 4, 183, 162, 160, 161, 53, 60, 7, 12, 115, 2, 112, 52, 5, 156, 61, 65, 64, 167, 68, 168, 169, 62
<b>2NPS</b>	108, 54, 53, 51, 157, 103, 194, 4, 104, 161, 3, 7, 5, 164, 162, 6, 166, 158, 57, 17, 170, 165, 201, 56, 167, 58, 59, 55, 9, 60, 63, 173, 62, 11
<b>3B5N</b>	162, 53, 55, 108, 105, 2, 160, 12, 16, 6, 43, 109, 9, 5, 19, 164, 163, 3, 57, 58, 167, 166, 150, 4, 7, 8, 186, 170, 173, 174, 56, 59, 61, 60, 13, 62, 63, 104, 171
<b>4WY4</b>	210, 155, 34, 154, 3, 6, 156, 53, 75, 5, 56, 2, 50, 163, 113, 161, 164, 54, 165, 166, 169, 167, 173, 105, 55, 60, 172, 52, 59, 170, 46, 27, 57, 58, 61, 63, 64, 62
<b>Negative Control</b>	165, 110, 11, 51, 210, 169, 55, 2, 5, 4, 47, 57, 162, 3, 53, 207, 117, 50, 167, 109, 6, 48, 7, 158, 58, 113, 62, 60, 59, 52, 61
<b>SNARE Specific</b>	8, 9, 10, 12, 13, 16, 17, 19, 21, 27, 30, 34, 43, 46, 47, 48, 54, 56, 63, 64, 65, 68, 75, 101, 103, 104, 105, 106, 107, 108, 112, 115, 116, 117, 143, 150, 154, 155, 156, 157, 159, 160, 161, 163, 164, 166, 168, 170, 171, 172, 173, 174, 183, 186, 194, 201, 211

The common mechanical force propagation paths were analysed in order to investigate which SNARE proteins were involved most heavily in each of the complexes' disassembly process (see *Table 3*). The negative control had the R-SNARE strongly involved in its disassembly propagation path. The R-SNARE was also the SNARE protein with most residues involved in the propagation path for 1SFC and 3B5N. 2NPS and 4WY4 had the Q<sub>a</sub>-SNARE which was the most involved SNARE protein. 1GL2 had a quite even distribution of residues from the SNARE proteins in its disassembly propagation path, however, the most common was the Q<sub>c</sub>-SNARE. Looking at the residues which were SNARE specific, the R-SNARE and the Q<sub>c</sub>-SNARE were majorly and equally involved, but the Q<sub>a</sub>-SNARE and Q<sub>b</sub>-SNARE were also seemingly playing a part.

*Table 3: Proportion of residues in common mechanical force propagation path belonging to specific SNARE domains for each of the SNARE complex structures, the negative control and the SNARE specific path. Values are given in percentages.*

*Abbreviations: SNARE, Soluble N-ethylmaleimide-sensitive factor attachment protein receptor*

	R-SNARE	Q <sub>a</sub> -SNARE	Q <sub>b</sub> -SNARE	Q <sub>c</sub> -SNARE
<b>1SFC</b>	0.3255814	0.2325581	0.2325581	0.2093023

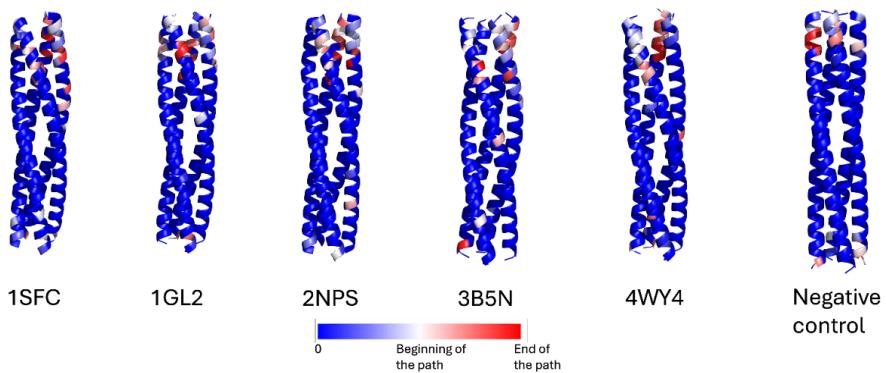
<b>1GL2</b>	0.2564103	0.2564103	0.2051282	0.2820513
<b>2NPS</b>	0.2941176	0.3235294	0.08823529	0.2941176
<b>3B5N</b>	0.3589744	0.2820513	0.07692308	0.2820513
<b>4WY4</b>	0.2631579	0.3421053	0.1052632	0.2894737
<b>Negative Control</b>	0.4193548	0.2258065	0.1612903	0.1935484
<b>SNARE Specific</b>	0.2807018	0.2105263	0.2280702	0.2807018

Using 1GL2 as a reference, the Root Mean Squared Deviation (RMSD) was calculated to investigate the similarity between the different PDB structures. 1SFC had a RMSD score of 1.004, 2NPS had a score of 0.604, 3B5N had a score of 1.093 and 4WY4 had a score of 0.586. Only the negative control was largely different from the 1GL2 structure with an RMSD score of 24.376 which was expected as it is not an actual SNARE protein. In addition to the RMSD scores, the percent identity of the SNARE complex structures was calculated with respect to the 1SFC sequences (see *Table 4*). The percent identities comparing all the SNARE protein complexes can be found in *Supplementary Table 1*.

*Table 4. Percent Identity matrix of 1SFC SNARE proteins with other SNARE proteins in different PDB structures*  
*Abbreviations: SNARE, Soluble N-ethylmaleimide-sensitive factor attachment protein receptor; PDB, protein data bank*

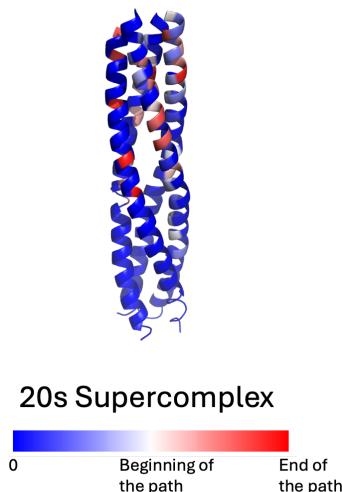
	<b>1GL2</b>	<b>2NPS</b>	<b>3B5N</b>	<b>4WY4</b>	<b>Negative Control</b>
<b>Chain A</b>	30.77	39.62	38.46	30.77	8.70
<b>Chain B</b>	37.74	41.51	47.17	35.85	2.17
<b>Chain C</b>	22.64	16.98	33.96	30.19	15.22
<b>Chain D</b>	28.30	30.19	26.42	28.30	8.70
<b>Average</b>	29.86	32.08	36.50	31.28	8.70

*Figure 8* shows the spatial arrangement of force propagation pathways coloured according to the start (light blue) to the end (red) of the path. The residues coloured in dark blue were not observed in the propagation path. For the negative control, only the N and C terminal residues were affected. For SNARE complexes there was a common pattern where most simulations are affecting the N and C terminal residues but also there's a common activation pattern in the conserved layers, especially the 0-layer.



*Figure 8: Common force propagation path residues coloured from blue to red for all systems. Force propagation starts from light blue colours and red indicates the end of the path. A dark blue colour indicates that the residue is not involved in the propagation path. All protein complexes were superposed on the same conformation.*

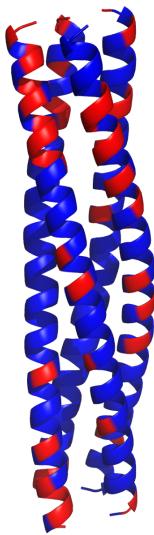
The same method of the MD simulation for the SNARE complex analysis was applied in the 20s supercomplex with the PDB ID: 3J96 (18). The mechanical force propagation pathway was analysed to verify our SMD method (see *Figure 9*). There was only 1 disassembly simulation done for the 20s supercomplex as the simulation of over 1 million atoms in the supercomplex was much more computationally heavy than the close to 100 thousand atoms in our model which did not include the NSF and  $\alpha$ -SNAP.



*Figure 9: The force propagation path residues coloured from blue to red for the SNARE complex in the 20s supercomplex, disassembled with the help of NSF and  $\alpha$ -SNAP. A dark blue colour indicates that the residue is not involved in the propagation path.*

*Abbreviations: SNARE, Soluble N-ethylmaleimide-sensitive factor attachment protein receptor; NSF, N-Ethylmaleimide-sensitive Factor;  $\alpha$ -SNAP,  $\alpha$ -soluble NSF attachment protein*

In order to compare the 20s Supercomplex with our SMD simulations, a combined path of the 5 cognate SNARE complexes minus the negative control was plotted on the 3D structure (see *Figure 10*).



*Figure 10: Combination of force propagation path of the 5 SNARE complexes, 1SFC, 1GL2, 2NPS, 3B5N, and 4WY4, minus the negative control. The SNARE specific residues, i.e., residues only observed in the force propagation paths of the SNARE actual SNARE complexes and not in the negative control, are plotted in red. A dark blue colour indicates that the residue is either not involved in any the propagation paths or that it is also present in the negative control.*

*Abbreviations: SNARE, Soluble N-ethylmaleimide-sensitive factor attachment protein receptor*

## Discussion

The vesicular transport machinery is vital for the survival of cells and organisms, and it is conserved across eukaryotes. The SNARE protein complex, a key player in this machinery, is known to instigate the required fusion between the vesicles and a lipid bilayer through its assembly process. Though the SNARE complex has been known about for several decades, the exact mechanisms of what happens following the SNARE-induced vesicular fusion remains debated. There are three main hypotheses which appear more or less likely in the discussion on the disassembly of the SNARE complex, these being if it occurs in a procesive-, global- or distributive-unwinding way (11). The differences between these three hypotheses lay in how the disassembly proteins NSF and  $\alpha$ -SNAP aid in the disassembly of the complex.

In the present study, we show support for common propagation paths of different SNARE protein complex structures and how these, though different, have similarities. In comparison to the negative control, all of the SNARE complex structures we investigated had conserved residues, through the 10 replicas made, important in the disassembly according to their stiffness. The methods applied in this project were MD and SMD simulations of 5 cognate SNARE complexes where their stiffness at each frame was saved and then analysed with an autoencoder-based approach.

### SNARE Structures vs. Negative Control

Looking at *Figure 8*, as well as in the stiffness propagation path table in *Table 2*, we could observe a similarity between the 5 cognate SNARE structures in comparison to the negative

control. Whereas the negative control only have residues of importance in its termini, the SNARE complexes showed higher stiffness in central areas of the protein complex. Some of these stiffness peaks were especially located around the 0-layer of the complex. This could suggest an added reason for the conservation seen in the 0-layer.

## SNARE-Specific vs. 20s Supercomplex

Using the patterns for the SNARE complexes in *Figure 8*, we attempted to make it comparable with the 20s Supercomplex by combining the paths of the 5 structures and subtracting the negative control to find the SNARE specific residues which were involved in the disassembly. Comparing what we see in *Figure 10*, i.e., the 3D structure with the SNARE specific residues highlighted, with the pattern shown in the 20s supercomplex in *Figure 9*, we can strongly suggest that our method utilising SMD simulations of the SNARE complex alone was successful in mimicking the effect of MD simulations with the SNARE complex together with NSF and  $\alpha$ -SNAP. In consensus with the SNARE specific pattern, the MD 20s Supercomplex simulation showed that not only are the termini important in the disassembly propagation path, but also the residues in central areas of the proteins are required for proper disassembly. However, to further validate these results, further studies would be needed, especially doing more simulations of the 20s supercomplex to definitively confirm our findings.

All in all, there are 2 main take away messages from this project. Firstly, residues of importance for the disassembly of the SNARE complex are not just located in the termini of the complex but also in central areas. Some important residues are seen in the 0-layer of the complex where the amino acid conservation is high, however, questions whether the other residues important for the disassembly also have a high amino acid conservation remain to be elucidated. Secondly, in comparing the SNARE specific residues of the combined paths of the SMD simulations of the 5 SNARE structures with the 20s Supercomplex we can see that our method of utilising the SMD of the core SNARE complex was able to accurately mimic the MD simulation of the 20s Supercomplex. This indicates that the SMD could be a good model which can be used in future studies to reduce the computational power and time needed to simulate the SNARE complex together with its disassembly proteins.

## Future Aspects

The aim of this project was to give a better picture of the complexities of the SNARE complex disassembly. To this end, analyses of stiffness in induced-disassembly structures were employed which gave an initial look into the different residues of the SNARE proteins which are involved in the disassembly of the complex. In the future, further studies are needed to investigate why certain residues in these disassembly propagation pathways are important and conserved. Investing the conservation of amino acids in the various residues, not just the 0-layer would perhaps provide an interesting insight into the role the side chains in the protein complex core play in this disassembly process.

Putting this research into a broader context, understanding the basis of the SNARE complex disassembly allows us to understand fundamental cellular processes which may be able to translate into clinical applications.

## Limitations

This project shows that there are some conserved pathways in the disassembly of SNAREs, however, there are certainly further improvements that could be made to this project and limitations which were unavoidable. The cause of these limitations was primarily time constraints and the nature of computational research.

This project concerns itself with simulating the protein complexes and as such, it remains impossible to definitively validate if the results of this study also are applicable in nature. Additionally, as the computational power required prevented us from creating simulations longer than milliseconds, the actual disassembly process with  $\alpha$ -SNAP and NSF proteins, i.e., the 20s supercomplex, could not be investigated. Additionally, it is worth noting that all simulations were done in a water model without a lipid membrane which in nature the C-terminal of the SNARE complex is attached to. As such, certain aspects of the propagation path may not align entirely with what would be observed *in vivo* or *in vitro*, however, this would be difficult to confirm as looking at this resolution in other types of experiments would be very difficult.

These limitations should not raise concerns about the impact of the results of this paper to any great extent, however. The methods used for the MD and SMD simulations are methods that have been used for decades and their results are generally trusted and comparisons with the negative control give sufficient support for our claims.

## Acknowledgements

I would first and foremost want to thank my supervisor Aysima Hacisuleyman who has guided me through this project and without whom it would have been most challenging to complete. I also want to thank Professor Dr. Dirk Fasshauer for allowing me a place in his lab group and for providing a calm and encouraging working atmosphere in the lab. Additionally, I want to extend my thanks to Clément Train for his invaluable insights and suggestions for my project. Furthermore, I want to thank all of the members of the Fasshauer lab, Carlos Pulido Quetglas, Deepak Yadav, Sévan Stroh, Michela Dall'Angelo, Iman Bentahar and Arianna Tonazzolli, as well as everyone at the Department of Computational Biology for being very kind and welcoming to me.

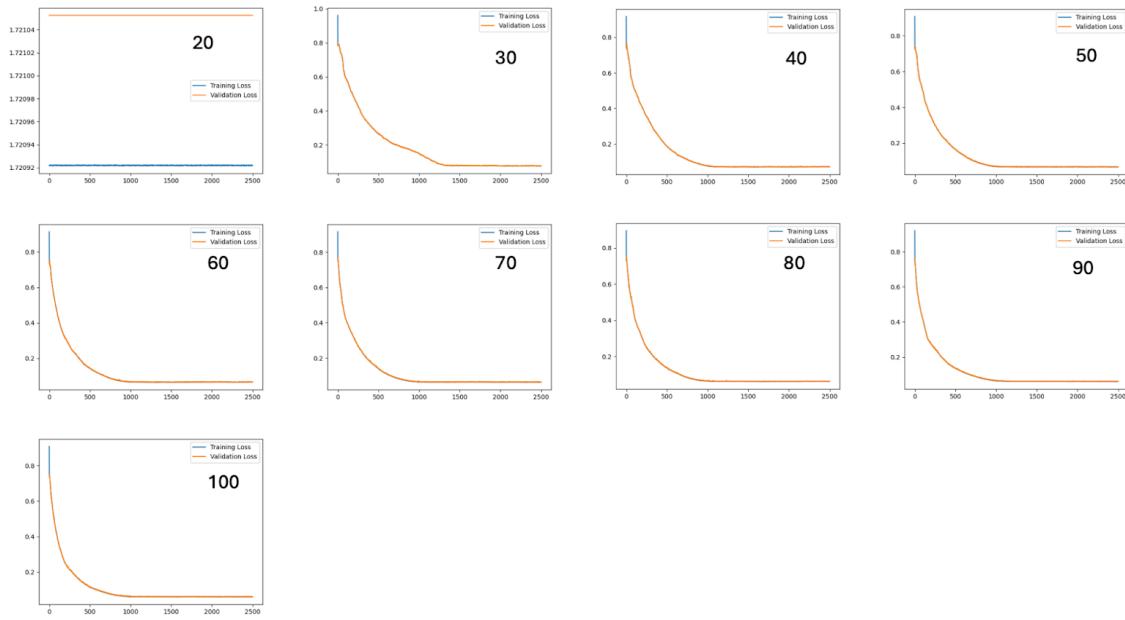
## References

1. Min D, Kim K, Hyeon C, Hoon Cho Y, Shin YK, Yoon TY. Mechanical unzipping and rezipping of a single SNARE complex reveals hysteresis as a force-generating mechanism. *Nat Commun.* 2013 Apr 16;4(1):1705.
2. Fasshauer D, Sutton RB, Brunger AT, Jahn R. Conserved structural features of the synaptic fusion complex: SNARE proteins reclassified as Q- and R-SNAREs. *Proc Natl Acad Sci U S A.* 1998 Dec 22;95(26):15781–6.
3. Ortega MA, Fraile-Martinez O, Garcia-Montero C, Alvarez-Mon MA, Gomez-Lahoz AM, Albillas A, et al. An Updated View of the Importance of Vesicular Trafficking and Transport and Their Role in Immune-Mediated Diseases: Potential Therapeutic Interventions. *Membranes.* 2022 May 25;12(6):552.
4. Malsam J, Söllner TH. Organization of SNAREs within the Golgi Stack. *Cold Spring Harb Perspect Biol.* 2011 Oct;3(10):a005249.
5. Ryu JK, Jahn R, Yoon TY. Review: Progresses in understanding N-ethylmaleimide sensitive factor (NSF) mediated disassembly of SNARE complexes. *Biopolymers.* 2016 Aug;105(8):518–31.
6. Kloepper TH, Kienle CN, Fasshauer D. An Elaborate Classification of SNARE Proteins Sheds Light on the Conservation of the Eukaryotic Endomembrane System. *Mol Biol Cell.* 2007 Sep;18(9):3463–71.
7. Yadav D, Hacisuleyman A, Dergai M, Khalifeh D, Abriata LA, Peraro MD, et al. A look beyond the QR code of SNARE proteins [Internet]. bioRxiv; 2024 [cited 2024 Apr 25]. p. 2024.04.24.590896. Available from: <https://www.biorxiv.org/content/10.1101/2024.04.24.590896v1>
8. Yoon TY, Munson M. SNARE complex assembly and disassembly. *Curr Biol.* 2018 Apr 23;28(8):R397–401.
9. Zhang Y. Energetics, kinetics, and pathway of SNARE folding and assembly revealed by optical tweezers. *Protein Sci Publ Protein Soc.* 2017 Jul;26(7):1252–65.
10. Dubuke ML, Munson M. The Secret Life of Tethers: The Role of Tethering Factors in SNARE Complex Regulation. *Front Cell Dev Biol.* 2016 May 9;4.
11. Cipriano DJ, Jung J, Vivona S, Fenn TD, Brunger AT, Bryant Z. Processive ATP-driven Substrate Disassembly by the N-Ethylmaleimide-sensitive Factor (NSF) Molecular Machine♦. *J Biol Chem.* 2013 Aug 9;288(32):23436–45.
12. Jakhanwal S, Lee C, Urlaub H, Jahn R. An activated Q-SNARE/SM protein complex as a possible intermediate in SNARE assembly. *EMBO J.* 2017 Jun 14;36(12):1788–802.
13. Baker RW, Jeffrey PD, Zick M, Phillips BP, Wickner WT, Hughson FM. A direct role for the Sec1-Munc18-family protein Vps33 as a template for SNARE assembly. *Science.* 2015 Sep 4;349(6252):1111–4.
14. Baker RW, Hughson FM. Chaperoning SNARE assembly and disassembly. *Nat Rev Mol Cell Biol.* 2016 Aug;17(8):465–79.
15. Zhang Y, Ma L, Bao H. Energetics, kinetics, and pathways of SNARE assembly in membrane fusion. *Crit Rev Biochem Mol Biol.* 2022 Aug;57(4):443–60.
16. Stein A, Weber G, Wahl MC, Jahn R. Helical extension of the neuronal SNARE complex into the membrane. *Nature.* 2009 Jul 23;460(7254):525–8.
17. Shah N, Colbert KN, Enos MD, Herschlag D, Weis WI. Three αSNAP and 10 ATP Molecules Are Used in SNARE Complex Disassembly by

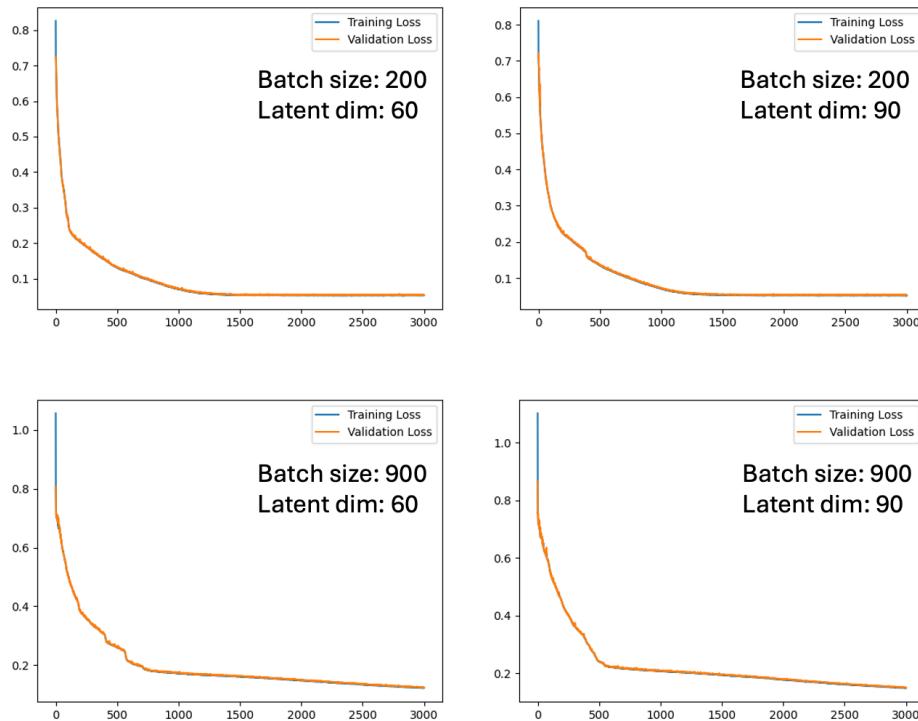
- N*-ethylmaleimide-sensitive Factor (NSF)\*. *J Biol Chem.* 2015 Jan 23;290(4):2175–88.
- 18. Zhao M, Wu S, Zhou Q, Vivona S, Cipriano DJ, Cheng Y, et al. Mechanistic insights into the recycling machine of the SNARE complex. *Nature.* 2015 Feb;518(7537):61–7.
  - 19. Naydenov NG, Feygin A, Wang L, Ivanov AI. N-ethylmaleimide-sensitive factor attachment protein  $\alpha$  ( $\alpha$ SNAP) regulates matrix adhesion and integrin processing in human epithelial cells. *J Biol Chem.* 2014 Jan 24;289(4):2424–39.
  - 20. Neuman KC, Nagy A. Single-molecule force spectroscopy: optical tweezers, magnetic tweezers and atomic force microscopy. *Nat Methods.* 2008 Jun;5(6):491–505.
  - 21. Rief M, Grubmüller H. Force spectroscopy of single biomolecules. *Chemphyschem Eur J Chem Phys Phys Chem.* 2002 Mar 12;3(3):255–61.
  - 22. Rounsevell R. Atomic force microscopy: mechanical unfolding of proteins. *Methods.* 2004 Sep;34(1):100–11.
  - 23. Sotomayor M, Schulten K. Single-Molecule Experiments in Vitro and in Silico. *Science.* 2007 May 25;316(5828):1144–8.
  - 24. Eyal E, Bahar I. Toward a Molecular Understanding of the Anisotropic Response of Proteins to External Forces: Insights from Elastic Network Models. *Biophys J.* 2008 May 1;94(9):3424–35.
  - 25. Stannard A, Mora M, Beedle AEM, Castro-López M, Board S, Garcia-Manyes S. Molecular Fluctuations as a Ruler of Force-Induced Protein Conformations. *Nano Lett.* 2021 Apr 14;21(7):2953–61.
  - 26. Lu H, Schulten K. The key event in force-induced unfolding of Titin's immunoglobulin domains. *Biophys J.* 2000 Jul;79(1):51–65.
  - 27. Lu H, Krammer A, Isralewitz B, Vogel V, Schulten K. Computer modeling of force-induced titin domain unfolding. *Adv Exp Med Biol.* 2000;481:143–60; discussion 161–162.
  - 28. Lee EH, Gao M, Pinotsis N, Wilmanns M, Schulten K. Mechanical strength of the titin Z1Z2-telethonin complex. *Struct Lond Engl.* 1993. 2006 Mar;14(3):497–509.
  - 29. Gräter F, Shen J, Jiang H, Gautel M, Grubmüller H. Mechanically induced titin kinase activation studied by force-probe molecular dynamics simulations. *Biophys J.* 2005 Feb;88(2):790–804.
  - 30. Pabón G, Amzel LM. Mechanism of Titin Unfolding by Force: Insight from Quasi-Equilibrium Molecular Dynamics Calculations. *Biophys J.* 2006 Jul;91(2):467–72.
  - 31. Carrion-Vazquez M, Li H, Lu H, Marszalek PE, Oberhauser AF, Fernandez JM. The mechanical stability of ubiquitin is linkage dependent. *Nat Struct Mol Biol.* 2003 Sep;10(9):738–43.
  - 32. Li PC, Makarov DE. Simulation of the mechanical unfolding of ubiquitin: Probing different unfolding reaction coordinates by changing the pulling geometry. *J Chem Phys.* 2004 Sep 8;121(10):4826–32.
  - 33. Collier TA, Piggot TJ, Allison JR. Molecular Dynamics Simulation of Proteins. In: Gerrard JA, Domigan LJ, editors. *Protein Nanotechnology: Protocols, Instrumentation, and Applications.* New York, NY: Springer US; 2020. p. 311–27.
  - 34. Izrailev S, Stepaniants S, Isralewitz B, Kosztin D, Lu H, Molnar F, et al. Steered Molecular Dynamics. In: Deufhard P, Hermans J, Leimkuhler B, Mark AE, Reich S, Skeel RD, editors. *Computational Molecular Dynamics: Challenges, Methods, Ideas.* Berlin, Heidelberg: Springer Berlin Heidelberg; 1999. p. 39–65. (Griebel M,

- Keyes DE, Nieminen RM, Roose D, Schlick T, editors. Lecture Notes in Computational Science and Engineering; vol. 4).
- 35. Prince SJD. Understanding Deep Learning [Internet]. The MIT Press; 2023. Available from: <http://udlbook.com>
  - 36. Tsuchiya Y, Taneishi K, Yonezawa Y. Autoencoder-Based Detection of Dynamic Allostery Triggered by Ligand Binding Based on Molecular Dynamics. *J Chem Inf Model.* 2019 Sep 23;59(9):4043–51.
  - 37. Sutton RB, Fasshauer D, Jahn R, Brunger AT. Crystal structure of a SNARE complex involved in synaptic exocytosis at 2.4 Å resolution. *Nature.* 1998 Sep;395(6700):347–53.
  - 38. Antonin W, Fasshauer D, Becker S, Jahn R, Schneider TR. Crystal structure of the endosomal SNARE complex reveals common structural principles of all SNAREs. *Nat Struct Biol.* 2002 Feb;9(2):107–11.
  - 39. Zwilling D, Cypionka A, Pohl WH, Fasshauer D, Walla PJ, Wahl MC, et al. Early endosomal SNAREs form a structurally conserved SNARE complex and fuse liposomes with multiple topologies. *EMBO J.* 2007 Jan 10;26(1):9–18.
  - 40. Strop P, Kaiser SE, Vrljic M, Brunger AT. The Structure of the Yeast Plasma Membrane SNARE Complex Reveals Destabilizing Water-filled Cavities\*. *J Biol Chem.* 2008 Jan 11;283(2):1113–9.
  - 41. Diao J, Liu R, Rong Y, Zhao M, Zhang J, Lai Y, et al. ATG14 promotes membrane tethering and fusion of autophagosomes to endolysosomes. *Nature.* 2015 Apr 23;520(7548):563–6.
  - 42. Humphrey W, Dalke A, Schulten K. VMD: Visual molecular dynamics. *J Mol Graph.* 1996 Feb;14(1):33–8.
  - 43. Ribeiro JV, Bernardi RC, Rudack T, Stone JE, Phillips JC, Freddolino PL, et al. QwikMD — Integrative Molecular Dynamics Toolkit for Novices and Experts. *Sci Rep.* 2016 May 24;6(1):26536.
  - 44. Wood CW, Bruning M, Ibarra AÁ, Bartlett GJ, Thomson AR, Sessions RB, et al. CCBUILDER: an interactive web-based tool for building, designing and assessing coiled-coil protein assemblies. *Bioinformatics.* 2014 Nov 1;30(21):3029–35.
  - 45. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. *J Chem Phys.* 1983 Jul 15;79(2):926–35.
  - 46. Best RB, Zhu X, Shim J, Lopes PEM, Mittal J, Feig M, et al. Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone  $\varphi$ ,  $\psi$  and side-chain  $\chi(1)$  and  $\chi(2)$  dihedral angles. *J Chem Theory Comput.* 2012 Sep 11;8(9):3257–73.
  - 47. MacKerell AD, Bashford D, Bellott M, Dunbrack RL, Evanseck JD, Field MJ, et al. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J Phys Chem B.* 1998 Apr 30;102(18):3586–616.
  - 48. Martyna GJ, Tobias DJ, Klein ML. Constant pressure molecular dynamics algorithms. *J Chem Phys.* 1994 Sep 1;101(5):4177–89.
  - 49. Feller SE, Zhang Y, Pastor RW, Brooks BR. Constant pressure molecular dynamics simulation: The Langevin piston method. *J Chem Phys.* 1995 Sep 15;103(11):4613–21.
  - 50. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, et al. Scalable molecular dynamics with NAMD. *J Comput Chem.* 2005 Dec;26(16):1781–802.
  - 51. Izrailev S, Stepaniants S, Balsera M, Oono Y, Schulten K. Molecular dynamics study of unbinding of the avidin-biotin complex. *Biophys J.* 1997

- Apr;72(4):1568–81.
- 52. Zhang S, Krieger JM, Zhang Y, Kaya C, Kaynak B, Mikulska-Ruminska K, et al. *ProDy* 2.0: increased scale and scope after 10 years of protein dynamics modelling with Python. Cowen L, editor. *Bioinformatics*. 2021 Oct 25;37(20):3657–9.
  - 53. Atilgan AR, Durell SR, Jernigan RL, Demirel MC, Keskin O, Bahar I. Anisotropy of fluctuation dynamics of proteins with an elastic network model. *Biophys J*. 2001 Jan;80(1):505–15.
  - 54. Eyal E, Yang LW, Bahar I. Anisotropic network model: systematic evaluation and a new web interface. *Bioinforma Oxf Engl*. 2006 Nov 1;22(21):2619–27.
  - 55. Dillon B, Plehn T, Sauer C, Sorrenson P. Better latent spaces for better autoencoders. *SciPost Phys*. 2021 Sep 17;11(3):061.
  - 56. Keskar NS, Mudigere D, Nocedal J, Smelyanskiy M, Tang PTP. On Large-Batch Training for Deep Learning: Generalization Gap and Sharp Minima [Internet]. arXiv; 2016 [cited 2024 May 22]. Available from: <https://arxiv.org/abs/1609.04836>
  - 57. Google. Google Colaboratory [Internet]. 2024. Available from: <https://colab.research.google.com/>
  - 58. TensorFlow Developers. TensorFlow [Internet]. [object Object]; 2023 [cited 2024 Apr 22]. Available from: <https://zenodo.org/doi/10.5281/zenodo.10126399>
  - 59. Bakan A, Meireles LM, Bahar I. *ProDy*: Protein Dynamics Inferred from Theory and Experiments. *Bioinformatics*. 2011 Jun 1;27(11):1575–7.
  - 60. McGibbon RT, Beauchamp KA, Harrigan MP, Klein C, Swails JM, Hernández CX, et al. MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories. *Biophys J*. 2015 Oct;109(8):1528–32.
  - 61. Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat Methods*. 2020 Mar 2;17(3):261–72.
  - 62. Harris CR, Millman KJ, Van Der Walt SJ, Gommers R, Virtanen P, Cournapeau D, et al. Array programming with NumPy. *Nature*. 2020 Sep 17;585(7825):357–62.
  - 63. Hunter JD. Matplotlib: A 2D Graphics Environment. *Comput Sci Eng*. 2007;9(3):90–5.



*Supplementary Figure 1:* Loss value curves graphs for the optimisation of latent dimensions. Values between 20 and 100 were tested with an interval of 10. The blue line represent the loss value (y-axis) at a certain epoch (x-axis) of the training data and the orange line represent the same but in the validation data.



*Supplementary Figure 2:* Loss value curves graphs for the optimisation of batch size. Two different values for the latent dimensions, 60 and 90, were used for comparing loss values with batch sizes of 200 and 900. The blue line represent the loss value (y-axis) at a certain epoch (x-axis) of the training data and the orange line represent the same but in the validation data.

*Supplementary Table 1: Percent Identity matrix for all SNARE complexes. Column names are in the same order in left to right as row names are top to bottom.*

*Abbreviations: SNARE, Soluble N-ethylmaleimide-sensitive factor attachment protein receptor*

1sfC_cut_C	100	33.96	30.19	22.64	16.98	15.38	13.46	9.62	13.21	18.87	16.98	16.98	13.21	20.75	11.32	15.22	15.22	15.22	18.87	18.87	15.09	15.09			
3b5n_C	33.96	100	20.75	15.09	15.09	7.69	7.69	11.54	9.43	5.66	5.66	15.09	13.21	11.32	7.55	10.87	10.87	10.87	16.98	15.09	9.43	9.43	13.21		
4wy4_cut_C	30.19	20.75	100	32.08	18.87	3.85	5.77	11.54	7.55	5.66	11.32	9.43	7.55	9.43	15.09	17.39	17.39	17.39	17.39	18.87	16.98	7.55	13.21	13.21	
1gl2_cut_C	22.64	15.09	32.08	100	43.4	7.69	9.62	11.54	11.32	13.21	15.09	11.32	11.32	7.55	9.43	17.39	17.39	17.39	17.39	13.21	16.98	13.21	24.53		
2nps_cut_C	16.98	15.09	18.87	43.4	100	9.62	9.62	9.62	9.43	13.21	13.21	5.66	9.43	9.43	11.32	15.22	15.22	15.22	15.22	18.87	13.21	16.98	28.3	15.09	
1gl2_cut_A	15.38	7.69	3.85	7.69	9.62	100	98.11	28.3	30.77	30.77	11.54	15.38	19.23	13.46	17.31	8.7	8.7	8.7	11.54	17.31	7.69	17.31	11.54		
4wy4_cut_A	13.46	7.69	5.77	9.62	9.62	98.11	100	28.3	30.77	28.85	9.62	13.46	17.31	11.54	15.38	8.7	8.7	8.7	11.54	15.38	5.77	15.38	13.46		
3b5n_A	9.62	11.54	11.54	9.62	28.3	28.3	100	38.46	34.62	19.23	26.92	23.08	15.38	11.54	13.04	13.04	13.04	13.04	13.46	19.23	15.38	15.38	7.69		
1sfC_cut_A	13.21	9.43	7.55	11.32	9.43	30.77	30.77	38.46	100	39.62	11.32	9.43	9.43	7.55	15.09	8.7	8.7	8.7	13.21	20.75	11.32	5.66	11.32		
2nps_cut_A	18.87	5.66	5.66	13.21	13.21	30.77	28.85	34.62	39.62	100	13.21	11.32	13.21	20.75	13.21	4.35	4.35	4.35	13.21	20.75	9.43	15.09	11.32		
4wy4_cut_B	16.98	5.66	11.32	15.09	13.21	11.54	9.62	19.23	11.32	13.21	100	32.08	33.96	35.85	39.62	8.7	8.7	8.7	15.09	15.09	18.87	15.09	13.21		
1gl2_cut_B	16.98	15.09	9.43	11.32	5.66	15.38	13.46	26.92	9.43	11.32	32.08	100	75.47	37.74	28.3	6.52	6.52	6.52	11.32	16.98	24.53	16.98	13.21		
2nps_cut_B	13.21	13.21	7.55	11.32	9.43	19.23	17.31	23.08	9.43	13.21	33.96	75.47	100	41.51	26.42	4.35	4.35	4.35	9.43	15.09	18.87	15.09	9.43		
1sfC_cut_B	20.75	11.32	9.43	7.55	9.43	13.46	11.54	15.38	7.55	20.75	35.85	37.74	41.51	100	47.17	2.17	2.17	2.17	11.32	16.98	22.64	16.98	16.98		
3b5n_B	11.32	7.55	15.09	9.43	11.32	17.31	15.38	11.54	15.09	13.21	39.62	28.3	26.42	47.17	100	6.52	6.52	6.52	11.32	15.09	15.09	16.98	15.09		
negative_ctrl_A	15.22	10.87	17.39	17.39	15.22	8.7	8.7	13.04	8.7	4.35	8.7	6.52	4.35	2.17	6.52	100	100	100	100	26.09	8.7	8.7	10.87	21.74	
negative_ctrl_B	15.22	10.87	17.39	17.39	15.22	8.7	8.7	13.04	8.7	4.35	8.7	6.52	4.35	2.17	6.52	100	100	100	100	26.09	8.7	8.7	10.87	21.74	
negative_ctrl_C	15.22	10.87	17.39	17.39	15.22	8.7	8.7	13.04	8.7	4.35	8.7	6.52	4.35	2.17	6.52	100	100	100	100	26.09	8.7	8.7	10.87	21.74	
negative_ctrl_D	15.22	10.87	17.39	17.39	15.22	8.7	8.7	13.04	8.7	4.35	8.7	6.52	4.35	2.17	6.52	100	100	100	100	26.09	8.7	8.7	10.87	21.74	
3b5n_D	18.87	16.98	18.87	13.21	18.87	11.54	11.54	13.46	13.21	13.21	15.09	11.32	9.43	11.32	11.32	26.09	26.09	26.09	26.09	100	26.42	26.42	24.53	22.64	
4wy4_cut_D	18.87	15.09	16.98	16.98	13.21	17.31	15.38	19.23	20.75	20.75	15.09	16.98	15.09	16.98	15.09	8.7	8.7	8.7	8.7	26.42	100	28.3	30.19	20.75	
1sfC_cut_D	15.09	9.43	7.55	13.21	16.98	7.69	5.77	15.38	11.32	9.43	18.87	24.53	18.87	22.64	15.09	8.7	8.7	8.7	26.42	28.3	100	28.3	30.19		
1gl2_cut_D	15.09	9.43	13.21	24.53	28.3	17.31	15.38	15.38	5.66	15.09	15.09	16.98	15.09	16.98	16.98	10.87	10.87	10.87	10.87	24.53	30.19	28.3	100	32.08	
2nps_cut_D	15.09	13.21	13.21	24.53	15.09	11.54	13.46	7.69	11.32	11.32	13.21	13.21	9.43	16.98	15.09	21.74	21.74	21.74	21.74	22.64	20.75	30.19	32.08	100	